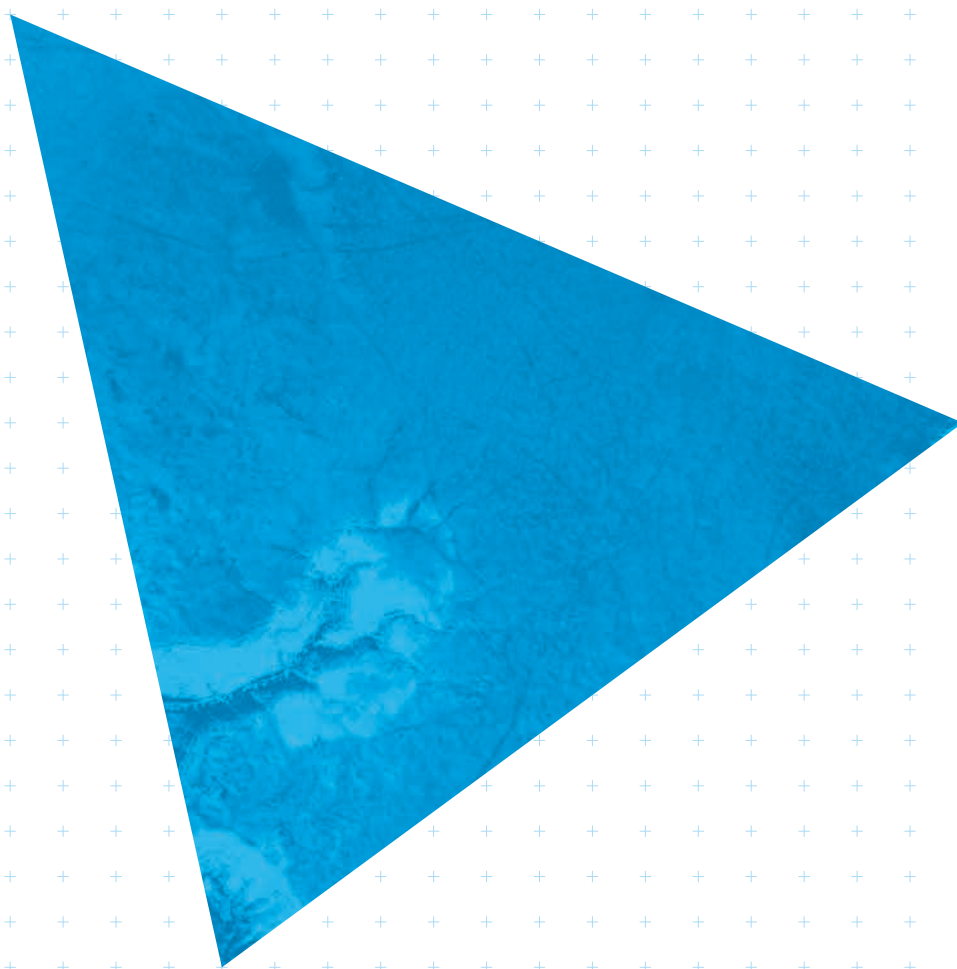


科学技術未来戦略ワークショップ報告書

材料創製技術を革新する プロセス・インフォマティクス

2021年1月24日（日）開催



エグゼクティブサマリー

本報告書は、国立研究開発法人科学技術振興機構（JST）研究開発戦略センター（CRDS）が令和3年1月24日に開催した科学技術未来戦略俯瞰ワークショップ「材料創製技術を革新するプロセス・インフォマティクスの俯瞰」に関するものである。

昨今のエネルギー・環境問題解決に向けた新たな材料開発や、COVID-19によって加速されるデジタル化の潮流を担う新たなハードウェアを実現する材料など、材料技術への要求はますます高度になっている。さらにはトレードオフの関係にあるような機能を同時に、かつ低コスト、低環境負荷で実現することが求められている。

このような材料への高度な要求に応えるために、従来の日本が得意としてきた技術蓄積を基礎にした材料開発手法に加え、高度なシミュレーション技術やデータ科学などコンピューターを用いた新たな材料設計手法であるマテリアル・インフォマティクスが活用され、すでにさまざまな成果が生み出されている。しかし、設計・予測された物質・材料を合成・作製する具体的な手法がわからず、実証できていない場合も散見されている。したがって、設計・予測された材料をどう合成・作製するのかという課題を解決する必要がある。

材料の合成プロセスは、材料ごとに手法が異なり、またプロセスを制御するパラメータも非常に多く、データ科学的にアプローチするのは一見困難に思えるが、最近進展してきているシミュレーション技術、リアルタイム・プロセス計測技術、ロボット技術なども活用することで、材料合成プロセスの世界へデータ科学的側面から足を踏み込む環境が整ってきていると言えよう。

CRDSでは、今後の物質・材料開発では、4つの科学（実験科学、理論科学、計算科学、データ科学）を融合的・統合的に活用して、目的材料の合成プロセスを効率的に探索する手法であるプロセス・インフォマティクスが必須になると考え、プロセス・インフォマティクスの共通基盤研究と個別プロセスのインフォマティクス活用研究の両方を実施することで、材料創製研究におけるプロセス・インフォマティクスの実施体制を構築することができるという仮説を提示した。

本ワークショップでは、まず、個別プロセスにおけるインフォマティクスの活用事例として、無機結晶（バルク・薄膜）、有機合成、機能性高分子、粉体プロセス、それぞれについて、現状と課題・将来展望について話題提供いただいた。高速処理可能な高精度モデリング技術により最適なプロセス条件を探索する例や、自律的最適化（Closed-Loop）により広範囲のパラメータ最適化を図る例、また、機械学習による自律的システムに人が積極的に関与する例、多数のプロセスパラメータの次元削減の際に物理的に意味のあるデータを中間データに使う例などの紹介があった。また、最適化モデルの構築に失敗したケースでは、プロセスパラメータ設定値と実測値とのずれが原因であり、データの正確さ・信頼性を担保する仕組みが重要であることが指摘された。

また、プロセス・インフォマティクスに必要な技術群として、データ科学、シミュレーション・モデリング技術、ハイスループット実験手法、先端計測技術などについて、プロセスに関連する最新の研究開発状況、課題・将来展望について話題提供をいただいた。マルチフィジックス・シミュレーションによりプロセス内で起きている現象を丸ごと把握する例、ハイスループット実験により材料特性とプロセス条件を同時に最適化する例、リアルタイム性と信頼性を両立させたモデリング自動化ワークフローの例などの紹介があった。齊一なデータを集めることやシミュレーション活用の重要性、データフォーマット統一、サンプル搬送手段共通化など共通基盤として要求される事項について把握することができた。また、プロセス・インフォマティクスの高

度活用するためには、プロセスの抽象化がカギであることが提言され、重要な示唆となった。

その上で、全体で総合討論を行い、以下のような重要な視点を得ることができた。

- ・日本の強みであるプロセス技術をさらに発展・強化するには、研究者の勘・コツを生かす仕組みを推進するべきである。その一つにはプロセスのデジタル化を推進し、共通基盤化していくことが挙げられる。
- ・産業界で量産プロセスを構築するには合成プロセスの根本的なメカニズムが明らかになっていることが必要である。
- ・アカデミアは、合成プロセスの本質的な理解を目指す研究を行うことが必要である。そのうえで、アカデミアでの基礎研究、企業での実用化研究をシームレスに実施していくことが望まれる。
- ・プロセス全体をシミュレーションしうるデジタルツイン構築が重要である。そのためには、有機合成・触媒反応では化学反応の素過程の本質的理解が必須である一方、無機結晶成長ではプロセス全体のマルチフィジックスのシミュレーションが重要課題である。
- ・プロセス・インフォマティクスの確立・普及には、研究者の意識の変革が重要である。そのためには、データ科学を用いたプロセス技術の分かりやすい成功例を、有機材料、無機材料、複合材料でそれぞれの分野で作ることが大切である。
- ・プロセスにデータ科学を活用するには、まずデータを集めることが必要である。一定の条件で実施された結果でなくても、実験条件等のデータを付加情報として収集・蓄積することが、データ活用では重要である。また、基板ホルダーの共通化、前処理やデータ登録の自動化などによりデータ収集の効率化を推進することで、機械学習モデルを構築する上で重要な、予想・仮説とは異なるネガティブデータも自然に集まってくるようになることが期待される。
- ・材料研究者が機械学習を理解して材料研究者の視点でテーマ設定をする、あるいは、データ科学の研究者に材料の基礎知識・感覚を身につけさせるなど、双方向の融合が必要であり、とくに、両方の視点を持った若手人材の育成が重要である。

CRDSでは、本ワークショップでの議論を踏まえ、今後、国として重点的に推進すべき研究領域、具体的な研究開発課題を検討し、研究開発の推進方法を含めて、戦略プロポーザルを策定し、関係府省や関連する産業界・学会等へ提案する予定である。

目次

1	趣旨説明	1
	曾根 純一、福井 弘行 (JST-CRDS)	1
2	個別プロセスにおけるインフォマティクス活用研究	4
2.1	結晶成長デジタルツインに向けた取り組み (SiC 溶液成長を中心に) 宇治原 徹 (名古屋大学)	4
2.2	ロボット - 機械学習による全自動・自律的な 無機機能性薄膜合成の現状と展望 清水 亮太 (東京工業大学)	11
2.3	精密有機合成とプロセス・インフォマティクス 甲村 長利 (産業技術総合研究所)	19
2.4	粉体プロセス開発のハイスループット化のための プロセス・インフォマティクスと材料探索ハイスループット化との関係 長藤 圭介 (東京大学)	25
2.5	データ駆動型高分子材料開発とプロセス・インフォマティクス 内藤 昌信 (物質・材料研究機構)	30
3	プロセス・インフォマティクスに必要な技術群	36
3.1	プロセス・インフォマティクスにおける技術紹介および課題と将来展望 金子 弘昌 (明治大学)	36
3.2	窒化物半導体 MOVPE プロセスの マルチフィジックス結晶成長シミュレーション 寒川 義裕 (九州大学)	42
3.3	触媒インフォマティクスの実践に見るプロセスの重要性 谷池 俊明 (北陸先端科学大学院大学)	48
3.4	計測インフォマティクスを目的とした計測スペクトルの スパースモデリングの自動化 吉川 英樹 (物質・材料研究機構)	54

4 | 総合討論（まとめ）

ファシリテータ 伊藤 聡 (JST-CRDS)..... 59

付録 76

付録 1：開催趣旨・プログラム 76

付録 2：参加者一覧 78

1 | 開催趣旨

曾根 純一、福井 弘行（JST-CRDS）

1
開催趣旨

最近の異常気象に見られるような温暖化問題、そこではエネルギー・環境問題に立ち向かう新たな材料技術が求められている。また、COVID-19の猛威によって加速されるデジタル化の流れ、ここではロボット、自動運転、遠隔操作などの新たなハードウェアの登場が必要になっている。これらを実現するため、材料への要求は高度になる一方である。

ここで材料に要求されている機能は単純な単一機能ではなく、複数機能、時としてトレードオフの関係にあるような機能の同時実現を求められる。また、その機能を低コスト、低環境負荷で実現するという非常に高度な要求が材料に対して突きつけられている。

この様な状況下、これまで日本が得意としてきた技術蓄積をベースにした経験知、それに基づく勘とコツだけで材料開発を実行することは困難になっており、高度なシミュレーションやデータ科学などコンピューターを用いた新たな材料設計が必要となっている。実際、構造や組成、それと材料の機能・特性を結び付けるマテリアルズ・インフォマティクスが登場し、今はまだ進化の過程ではあるが、既に一定の成功を収めている。

しかし、材料開発としては次のステップ、すなわち期待される材料をどう合成、作製するのか、材料合成・製造に関わる極めて難度の高い、いわば本丸の課題が待っている。材料の合成・製造、すなわちプロセスは、材料ごとに手法のバリエーション、またそれを制御するパラメータも非常に多く、一見、科学的にアプローチするのは困難に思える。しかし、マテリアルズ・インフォマティクス、さらにAIに見られるようなデータ科学の最近の進展、シミュレーションの高度化、オペランド計測のような化学反応プロセスをリアルタイム観測する技術、さらにはロボットを使ったハイスループット実験など、材料プロセスの世界へ科学的に足を踏み込む環境が整ってきている。

CRDSでは、いわゆる4つの科学、実験科学、理論科学、計算科学、データ科学をフル活用して目的材料の合成プロセスを効率的に探索する手法であるプロセス・インフォマティクスが必要になると考えている。

図 1-1-1 に示すように、現在、生まれつつある合成経路設計、プロセスのシミュレーション・モデリング技術、実験 DX によるハイスループット実験・評価やプロセス計測技術などの技術を活用して、「どう作るか」というプロセス・インフォマティクスの技術体系をつくり上げ、「何をつくるか」であるマテリアルズ・インフォマティクスと「何ができていますか」である計測インフォマティクスと強くリンクさせることでデータ駆動型物質・材料創製を次のステージに進化させる必要があると考えている。

具体的な研究の進め方としては、図 1-1-2 に示すように、プロセス・インフォマティクスの共通基盤研究①と個別プロセスのインフォマティクス活用研究②を両輪として実施すべきであると考えている。共通基盤では、プロセス・インフォマティクスを実現するために必要な仕組み・制度や研究テーマに取り組む。また、個別対象プロセスについては、まず、プロセス・インフォマティクスの必要性、実現性が高いプロセスから始めるのがよいと考える。図 1-1-2 に記載してあるプロセスはあくまでも一例ではあるが、これらはプロセス・インフォマティクスに適しており、かつプロセス・インフォマティクスを必要としているものである。

また、個別プロセスでの検討で得られた知見で共通基盤をさらに増強することで、プロセス・インフォマティクスの技術体系を強力なものにすることができ、より幅広い範囲、既存材料を含む多くのプロセスに適用できるようにすることを期待している。

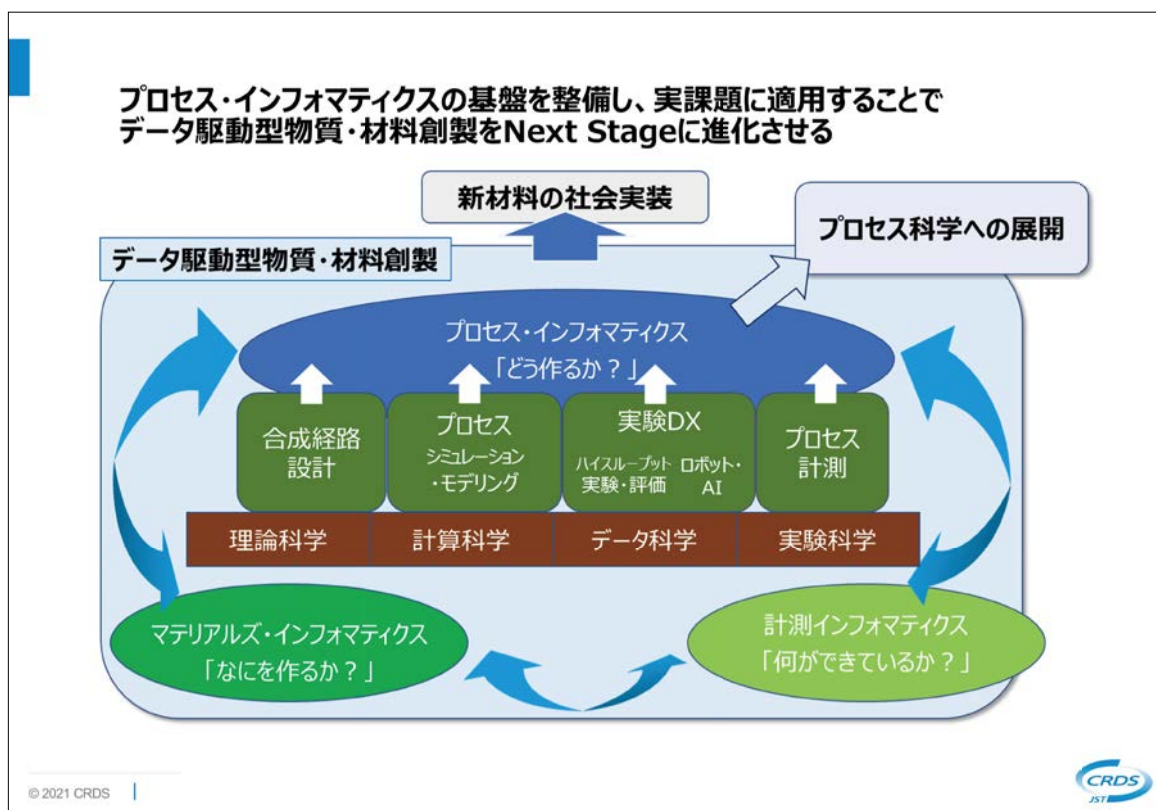


図 1-1-1 プロセス・インフォマティクスでデータ駆動型材料創製をNext Stageに

- 1) P I 共通基盤構築研究①と個別対象プロセスのインフォマティクス活用研究②を実施。
- 2) 共通基盤では、P I を実現するために必要な仕組み・制度・研究テーマに取り組む。
- 3) 個別対象プロセス②は、まずはP I の必要性・実現性が高いプロセスから開始。
- 4) 個別対象プロセス②の検討で得られた知見により、共通基盤での取り組みを拡充。
- 5) 次ステップとして、既存材料を含めた多くのプロセスにP I を広げる。

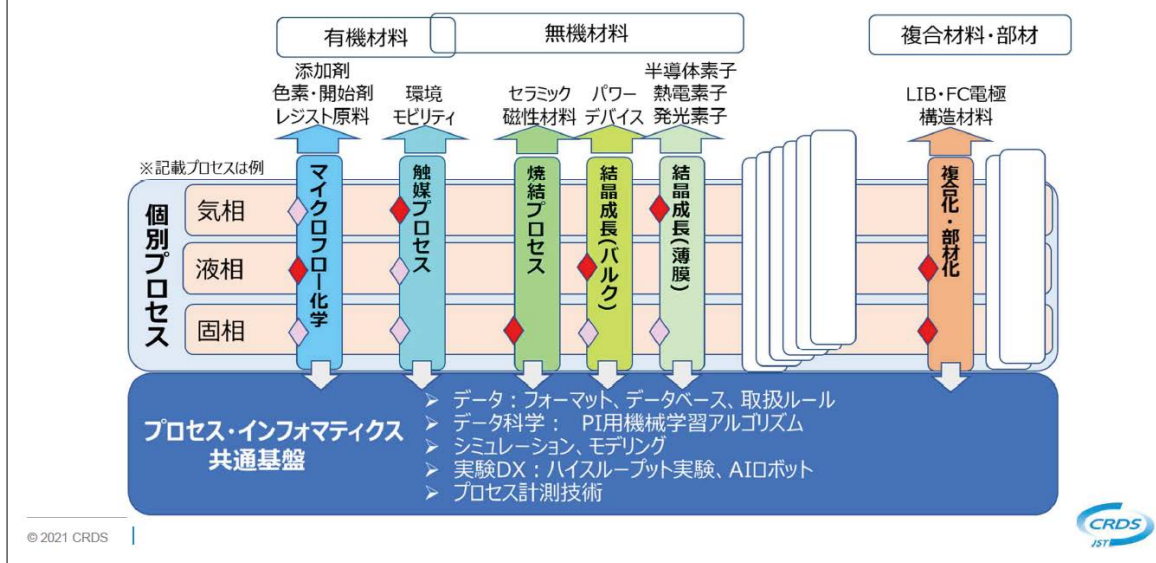


図 1-1-2 プロセス・インフォマティクス研究の推進方法案

本ワークショップは、以下のようなCRDSの仮説を様々な視点から議論し、戦略プロポーザル策定につなげることを目的として開催した。

- A) プロセス・インフォマティクス共通基盤構築研究①、個別対象プロセスのプロセス・インフォマティクス活用研究②をその両輪として実施する。また、個別対象プロセス②の検討で得られた知見をもとに、共通基盤での取り組みを拡充することにより、材料創製研究におけるプロセス・インフォマティクス実施体制の盤石な基盤を構築することができる。また、これにより、マテリアルズ・インフォマティクスで提案された物質・材料の合成プロセスを効率的に確立できるようになり、日本の材料の優位性が確保できる。
- B) プロセス・インフォマティクス共通基盤構築研究①を、拠点を中心に実施することで、プロセス・インフォマティクス立ち上げに必要な基盤を構築することができる。
- C) 個別対象プロセスのプロセス・インフォマティクス活用研究②は、対象プロセスごとにプロセスセンターを設置し、そこに実験DXの環境やデータ科学を集約して実施することによって、産官学の連携が進み、データ駆動型材料創製を加速することができる。
- D) プロセス・インフォマティクス共通基盤の強化・拡充を実現することによって、次ステップとして、新規材料における適用プロセス範囲の拡大や既存プロセスの効率化（省エネルギー、副生成物抑制）への応用にプロセス・インフォマティクスを発展させることができる。

2 | 個別プロセスにおけるインフォマティクス活用研究

2.1 結晶成長デジタルツインに向けた取り組み（SiC 溶液成長を中心に）

宇治原 徹（名古屋大学）

この講演では、SiCの溶液成長を例に、これまでどういったことを行ってきたかを説明した後に、それをどうやって横展開しているのか、さらに、それをやっていく上で次に何が課題だと考えているかを紹介する。

SiCの結晶成長に関しては、これまでに長期にわたって研究・開発が行われてきたが、品質やコストの面で未だ大きな問題がある。我々は、それらの問題を根本から解決するために、通常用いられている昇華法に変えて、溶液法を用いたSiC結晶成長を行ってきた。この方法の特徴は、昇華法のものに比べて圧倒的に高品質のものが得られるという点にある。図 2-1-1にはX線トポグラフィで撮った結晶の写真を示す。上が従来法である昇華法で成長した市販のもの、下が我々の方法で成長したSiC結晶である。従来法で作られた結晶には欠陥に起因する線や点が多く見られるが、我々のものにはほとんどそれが存在しない。数値として表してみても、我々の結晶は、市販されているものに比べて、いずれの種類の欠陥についても、桁違いに低い欠陥密度となっており、非常に高品質であることが分かる。

この高品質な結晶が本当に役に立つことを実証するためには、これを用いてデバイスを作成し、その性能評価を行う必要があった。この写真を発表した2017年当時、我々のラボレベルの実験で得られる結晶のサイズが10～15mm程度と、残念ながらデバイスメーカーと連携してデバイス試作を行うための4インチや6インチに比べるとはるかに小さかった。また、シリコンの例を見てもわかるように、一般に結晶の大口径化は非常に難しく、単に装置をスケールアップするだけでなく、成長条件等を根本的に見直していかなければならないため、一般に数十年の時間がかかるとされてきたものであった。

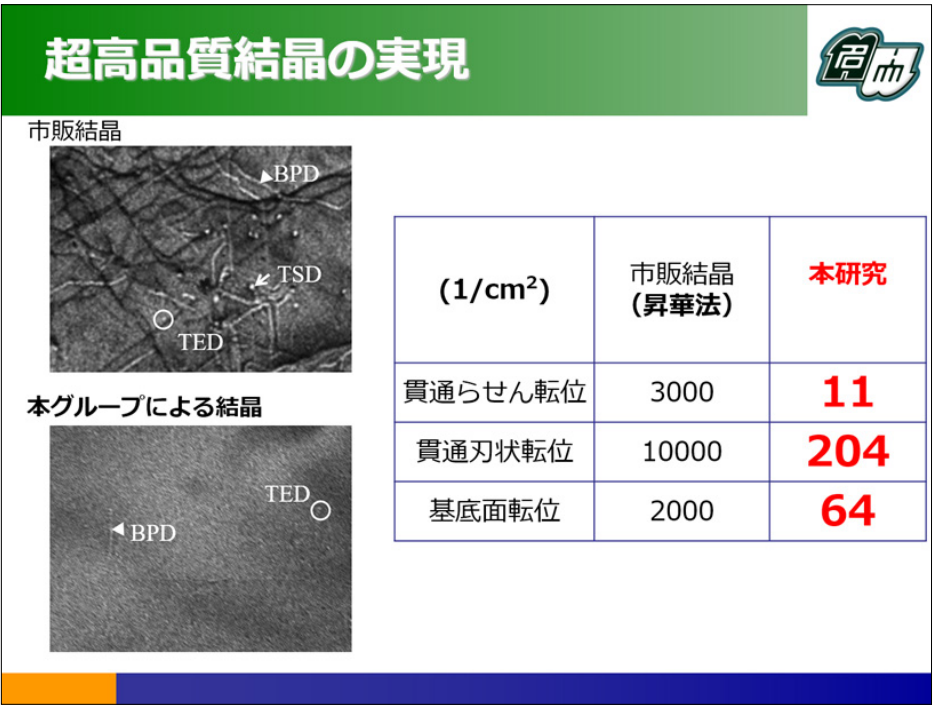


図 2-1-1 溶液法で作成した SiC 結晶

せっかく良いものをつくっても使えないというのでは全く意味がないので、そこを何とかしたい、というのが、我々自身が機械学習やインフォマティクスを使おうと思った動機である。

実際、我々が行いたいことは、以下の通りであった。まず、リアルな空間で実際に基礎研究を行い、その結果を機械学習などにより、うまくサイバー空間でモデリングする。このモデリングは、高速性と正確性の両方を兼ね備えていることが必要であり、単に正確なシミュレーションを行うだけではスピードの面で問題がある。このモデリングを使って、良い条件、良いプロセスを発見し、それをリアル空間に実装する。この一連の流れが、我々が持っていた夢である。

現在、その夢に関しては、なかなかうまく実現していると思っている。

図 2-1-2 左に示した図は、実際に我々が用いている結晶成長装置のコンピューターグラフィックスである。中央のるつぼの中にある虹色部分は、温度分布を表している。るつぼが動くにつれて、温度分布もるつぼの位置により変化するが、ここで示した温度分布以外に、組成や流れといった様々なものを計算することができる。重要な点は、従来の流体シミュレーションではある条件に関して1つ計算するだけでも、速くても1時間、場合によっては半日程度の計算時間が必要となるが、我々の方法を用いると、これを瞬時に計算することができる点である。いわば、コンピューターの中で我々は結晶成長を擬似的に再現する、先ほど述べたモデルを構築することができているということになる。

機械学習技術を使うことで、どれぐらい効果が得られたかを示したのが図 2-1-2 右である。ここでは、溶液の過飽和度を、通常の数値シミュレーションと我々の予測モデルで求め、その結果をプロットしている。厳密に見ると多少違うところがあるが、傾向の大部分は再現することができている。しかも計算スピードに関しては圧倒的な差があり、シミュレーションが3800秒かかったのに対して、我々のモデルは0.0003秒で予測を終えている。これによって、シミュレーションに基づいてAIモデルをコンピューターの中に構築できたことになる。

ひとたびこれができると、あとは目的関数をうまく設定して、それで最適値を求めることにより、「どんなプロセスであれば4インチ、6インチで結晶が作れるか」といったことをコンピューターの中で試行錯誤できることになる。

実際に用いた目的関数についてはここでは述べないが、例えば、成長界面での過飽和度分布、流れの向きのような量を目的関数として、るつぼ回転、結晶回転、るつぼや結晶の位置、温度分布、高周波パワー周波数のような制御パラメータと、るつぼや断熱材の形状やサイズ、溶液の体積といったデザインパラメータからその値を求めるようにモデルを作成する。我々が基礎研究で得た知見を使って重要な目的関数を設定し、その目的関数からプロセス最適化を行っていく。それによる結果だけを述べると、2017年に10～15mmであっ



図 2-1-2 SiC 結晶成長の高速予測モデルの構築

た結晶サイズが、2018年には3インチ結晶基板が取れる85mm、さらに、2019年末ぐらいには6インチに到達している。

このように、こういう方法を使うと良いものができるだけでなく、それを即座に使えるものにできる。大口径化する際に難しい点は、均質性をどう保つかという点であり、あらゆる場所で同じように成長させることは一般に非常に困難な課題である。我々の場合も、ある1つの条件でずっと成長し続けた場合、面内に大きな不均一性ができてしまうため、1、2分おきに5～6種類のシーケンスを繰り返し切り替え、全体が均質になるような条件を探しだし、それを行っている（図 2-1-3）。作成した結晶の表面モフォロジーを調べてみると、1つの条件で成長した場合には、場所によって平らな部分とラフな部分ができるが、シーケンスを切りかえることで全体的に平らなものができていることが確認されている。実験で1つの条件を見出すだけでも大きな困難があるため、このように幾つもの条件の繰り返しパターンを実験から見つけることは、現実的には不可能だったと思われる。高速予測が可能なモデルをつかったからこそ、こうした複雑なシーケンスも設計できたことになる。

空間的な不均一のほかに、もう一つ重要な要素として成長環境の時間的な変化がある。SiCの溶液成長では、るつぼを溶かしながら結晶成長を行っているため、結晶が成長するにつれるるつぼの形状や溶液の状態が変わっていく。結晶成長の分野では、成長の間に結晶周りの環境が時間変化することはよく知られていて、例えば結晶成長を100時間行う場合に100時間同じ条件で成長するのではなく、10時間おきなどに結晶成長パラメータを変えるといったことが行われてきている。我々の方法では、このような環境の時間変化も、モデル化できているので、より細かなパラメータ変更も可能になる。図 2-1-4 右の上下は50時間の結晶成長において、ずっと同じ条件で結晶成長した場合（上）と、途中で100回シーケンスを変えた場合（下）の比較である。こういうシーケンスの調整を行うだけで、結晶成長の平均的な速度はどんどん上がっていくことがわかる。こういう複雑なシーケンスを設計できれば、より良い材料ができていくことが示唆されている。

「我々のつくったAIモデルは正確なのか」という質問をよく受ける。これに対する答えは、「シミュレーションが正確でなければ当然正確ではない」というものである。シミュレーションの中で一番正確でないところは、いろいろな物性値である。融液の物性や、気相成長の場合には気相の物性、あるいは、表面の反応係数のよ

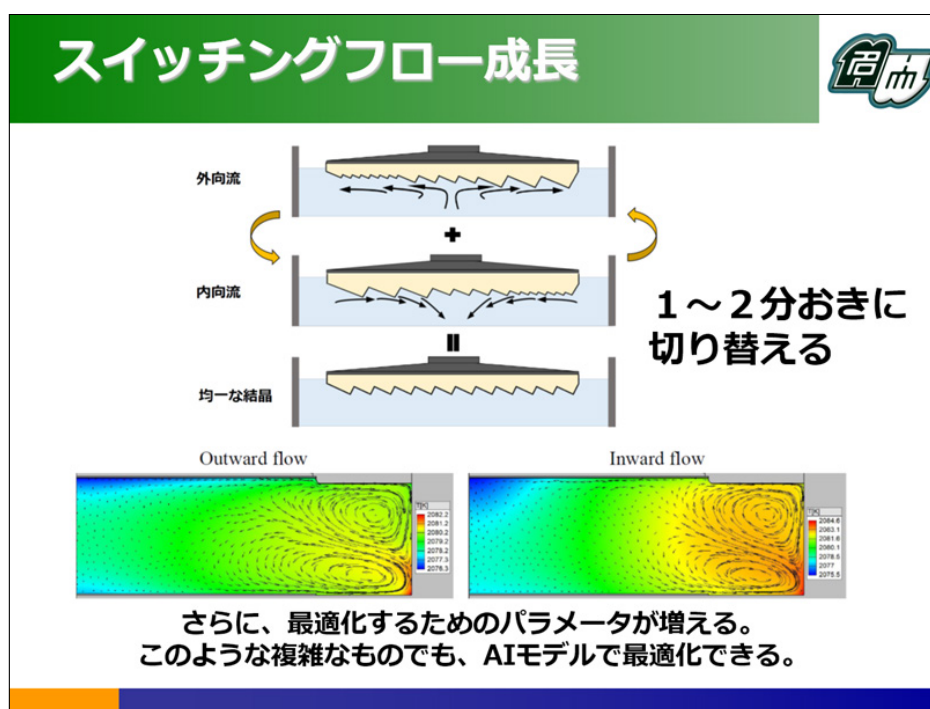


図 2-1-3 均質性確保のためのシーケンス

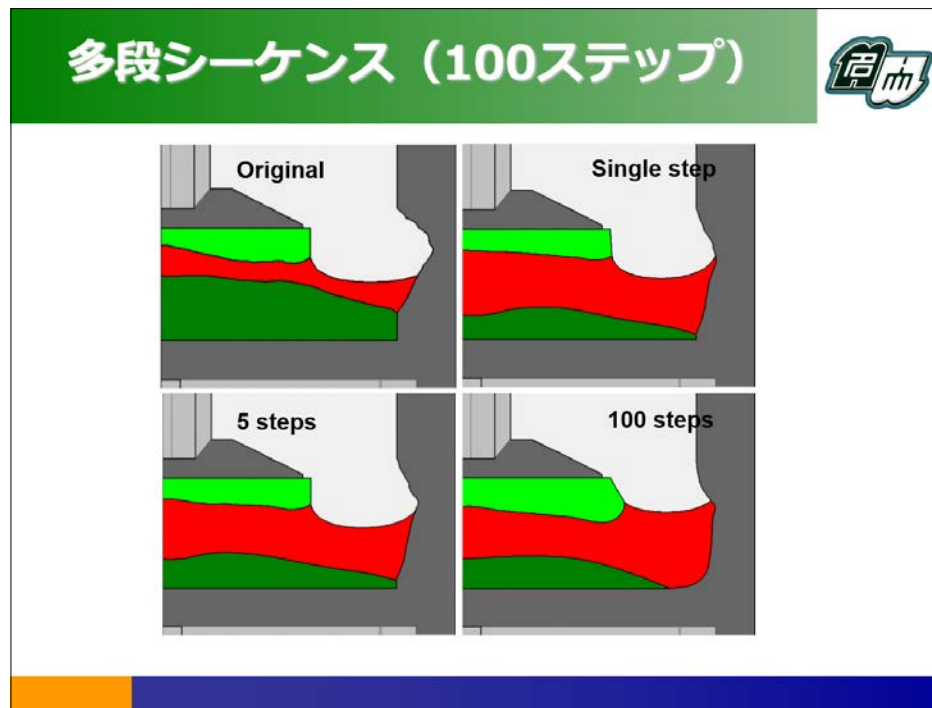


図 2-1-4 多段シーケンスの最適化

うなものを正確に決めることは非常に困難である。しかし、これまで述べたような機械学習モデルをつくってやると、そこで示された結果と実際の実験結果を比較して、どういう物性値を当てはめればうまく実験を再現できるかを調べることができる。これはデバイスシミュレーションではよく行われていて、測定したI-V特性やC-V特性が実測とうまく合うように、デバイスの物性値を人の手で合わせ込むといった操作がなされている。人手による場合には、条件を満たすパラメータセットは、たかだか数パターンぐらいしか調べられないが、AIモデルによる方法ではその候補を1,000通り～1万通りを簡単に探すことができる。うまく実験結果を再現するような融液の物性の推定を、AIモデルを用いて実際に行ってみたところ、実験的に求めることが難しいとされる「種結晶/溶液界面反応速度定数」なども、かなり確定的に推定できることも確かめられている。ほかのいろいろな物性についても同様な方法を用いて、機械学習モデルと比較することで求めることもできる。

我々は、この方法を溶液法以外のいろいろなプロセスに展開している。もともとは、我々のライバルであった、SiCの昇華法のようなガスを介した成長に関して、チャンバーの中のガスの分圧分布を予測するモデルもできている。また、窒化ガリウムのHVPEという気相成長方法において、ノズルのパターンをどうするかという、結構難しい問題に関しても最適化を行うことができおり、対称ではない複雑なノズルが設計されている。

AIモデル構築の課題を図 2-1-5に示す。我々はAIモデルをつくることで、ある意味デジタルツインを構成しようとしている。つまり、まずシミュレーションがあって、モデルをつくって、それでプロセスの構築・制御を行いたい。シミュレーションについては、正確なシミュレーションの開発が当然今後も重要となるし、シミュレーションのもととなる物性値の推定についても、未知の材料を扱おうと思うとそもそも物性値が分からないというのがあり、この推定を今後も続けていくことが重要である。

AIモデルの構築に関しても、これはすべて同じニューラルネットワークを使っているわけではなくて、プロセスごとに物理現象に基づいて機械学習手法、ニューラルネットワークの形をどんどん変えて行くことも必要である。

また、先ほども述べたように、プロセスを最適化する際には目的関数の設定が最も重要だと我々は考えていて、人の感覚や勘・コツのようなものは、ここに入れ込んでいくことになるが、経験からどういう風に目的関数を自動推定するかがチャレンジとなる。また、当然ながら、その先の最適化や逆問題の解き方、モニタリン

AIモデル構築における課題



物理シミュレーション



AIモデル



最適プロセスの構築

プロセスの制御

正確なシミュレーションの開発

物性値の推定

実験による推定（少ないデータからも推定）
マテリアルズインフォマティクスによる推定

物理現象に基づく機械学習手法の開発

特徴量の適切な設定
新たなニューラルネットワークの開発

適切な目的関数の設定

人の感覚（勘やコツ）なども、ここで考慮

最適化・逆問題

情報理論の専門家が必要
量子コンピュータの活用などもここで実施

モニタリング技術の開発

実際とモデルを一致させることが重要
制御には、リアルタイムモニタリングも必須

図 2-1-5 AIモデル構築の課題

グの技術の開発も重要となってくる。

我々が最終的に目指しているのは、まずは、結晶成長の分野で、結晶成長サイバーファクトリーのようなものを構築することである。そこでは、いろいろな材料の結晶が産学連携で検討されており、企業の人々がそこに自分たちのモデルをつくり、自分たちの作りたい結晶をサイバー空間でどんどん作成する。サイバー空間では何万台もの装置が何百万回も試行錯誤できるため、そこで多くの人々が自分たちの作りたい材料、結晶を合成していく。

さらに、図 2-1-6のように、そのようなサイバーファクトリーを他のプロセスについても構築して、それらの共通基盤となる部分を集約していく。この図は、CRDSが示しているのと同様であるが、1つ注意すべき点としては、プロセス・インフォマティクスの技術自体は、黄色の四角で書いてあるところの個々の検討の中で生まれるので、まずはここを中心にして、そこからどんどん集約し、下に下りながら横展開していくといったやり方がいいのではないかと考えている。

さらに、将来的に重要になるのは、いわゆるサプライチェーンを越えたプロセス開発体制をつくることだと思っている。結局、どんな材料が必要かは最終的にはニーズ側から決まっており、半導体があってデバイスがあってモジュールがあるとしたら、出口側からどんどん上流にニーズが上がっていく。このサプライチェーン全体を見た時には、半導体メーカー、デバイスメーカー、モジュールメーカーのそれぞれに、多くの工程があるのに加えて、そうした工程に使われる装置、原料、周辺部材も各工程のプロセスパラメータと考えなければならない。それぞれのメーカーごとに個別のAIモデルを構築するだけでなく、サプライチェーンの全工程を表現する一つのAIモデルをつくり、ニーズ側から順番に、どういう素材を作るかという観点も全部含めて、最終的には最適化されて、そのプロセスが構築されていく。そして作るものもそこで決定していく、こういうものが最終的な姿だと私は考えている。

我々は、このような複数工程が関わるような最適化、いわゆるカスケード最適化にもチャレンジしており、これを使うと、全ての工程を同時に最適化するよりも少ない試行回数によってよりよいものが見つけられるという開発もしている。

ここで、講演をまとめる。この講演ではSiC溶液成長のことを重点に紹介したが、これを他の結晶成長に

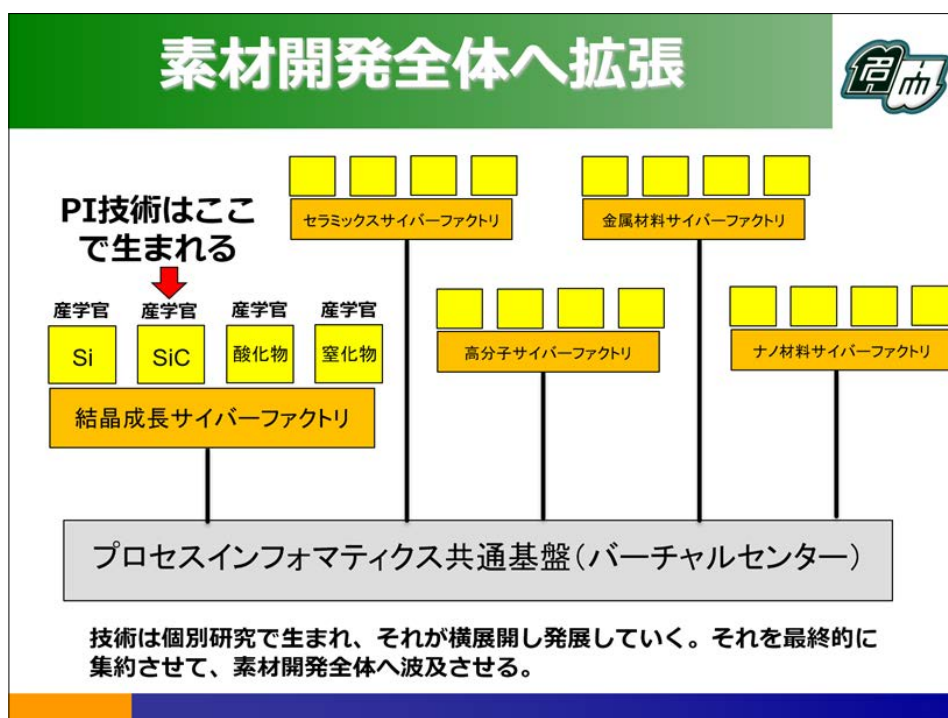


図 2-1-6 プロセス・インフォマティクス推進方策のイメージ

横展開した結果も一部紹介した。また、今のAIモデル構築においてのいくつかの課題があることを紹介した。さらに、プロセス・インフォマティクス技術は現場で生まれると実感していることを述べ、類似のプロセスであれば横展開が比較的容易であるので、そこは拠点化したほうが良いとの考えを示した。最後に、何を作るべきかをしっかり考えるため、カスケード的な最適化手法が、今後必要になると考えている。

【質疑応答】

Q：モデリングというのは、ナビエ・ストークスのような流体の方程式をベースにしてつくっているのか。そういうものは、自分の実験に合わせて手を加えられるようなオープンソース的な環境はあるのか。

A：CFD（computational fluid dynamics：数値流体力学）計算自体は、ごくごく一般的なCFDを用いている。流体に関するシミュレーションは商用でもたくさん出ているし、もちろんオープンソースでできるものもこの世の中にはたくさんある。結晶成長の分野で必要な物理モデルを入れながらシミュレーションを行うことは、今の段階でも相当できる。ただ、物性値を正しく入れないと結局正しくないの、物性値の予測を今後どうやっていくかは同時に考えないと正しいモデルがつかれないので、そこは重要である。

Q：スイッチングフローに関しては、計測が非常に重要だと思うが、温度分布や濃度等の計測はどのように行っているのか。

A：すべてを直接測るわけにはいかないの、今は複数の箇所の温度を測るとか、回転のトルクを測るとかを行っている。マルチフィジックスシミュレーション、物理シミュレーションで計算できる値をなるべく多く取ってなるべく多く測ろうという努力をしている。

Q：サイバー空間での機械学習によるパラメータ最適化をベースにしていると理解したが、フィジカル空間のデータをどのように活用しているのか。

A：先ほど結晶成長速度の話をしたが、例えばAIモデルで予測される結晶成長速度分布（場所によって

成長速度が違う)が、実際の実験結果とどうやったらうまく合うか、いろいろ物性値を変えながら一番マッチするものを探していくことをやっている。この方法の良いところは、たくさんの実験結果を必要とせず、1、2回の実験結果であったとしても物性値を一応推定することができる。プロセス・インフォマティクスでは多くのデータを取ることが難しいので、我々はなるべく少ないデータから良い推定値が得られるということで、そういう方法を取っている。

2

個別プロセスにおける インフォマティクス活用 研究

2.2 ロボット-機械学習による全自動・自律的な無機機能性薄膜合成の現状と展望

清水 亮太（東京工業大学）

本講演では、ロボットと機械学習を使った薄膜プロセスを全自動・自律化するという試みや現状を紹介し、今後の展望を述べる。特に、プロセス・インフォマティクスの技術は現場で生まれると感じており、研究の最前線で感じるポイントや課題を述べたい。

我々のグループでは酸化物を中心とした無機材料の薄膜研究を行っており、材料創製にあたってはパルスレーザー堆積法、スパッタといった成膜装置を使うことが多い。これらは非平衡プロセスであるため、どう作るか、つまりプロセスが非常に重要である。

一方で、成膜をする段階ではプロセスパラメータとして熱力学的な量を設定して議論することが多く、例えば成膜温度、酸素分圧といったマトリックスをつくり、このマトリックス上を絨毯爆撃的に材料合成し、最適なプロセス条件（成長窓）を探すという方法が従来一般的であった。図 2-2-1 はチタン水素化物薄膜合成の例であるが、頑張って30条件ほど合成評価したとしても、最適なプロセス条件は非常に粗い精度でしか探索できていない。

ベストなプロセス条件をいかに早く見つけ出すか、そのためにマトリックスをいかに早く埋めるかが課題となり、それを克服することによって、例えば卒論生が1年間かかる実験を1週間で実現し、繰り返しの単純作業を減らすことによって創造的な研究スタイルへ進もうというのが我々のモチベーションである。そこで、機械学習とロボットによる効率化を進めている。

まず我々が対象とする無機物質の物性とプロセスの関係について整理する。ここではシリコン（Si）を例にとって考える。電子伝導性は、単結晶 Si、多結晶 Si、歪み Si、アモルファス Si といった結晶構造に強く依存する。また、機械特性は欠陥の種類や量、さらに半導体特性はドーパントの種類や量に依存する。そして最終的な物性はこれらの因子がどう変化するかで決まる。そしてこの因子はプロセスに依存し、温度、圧力、原料供給スピードなど様々なパラメータに依存している。つまり、そもそも構造組成に曖昧さがあ

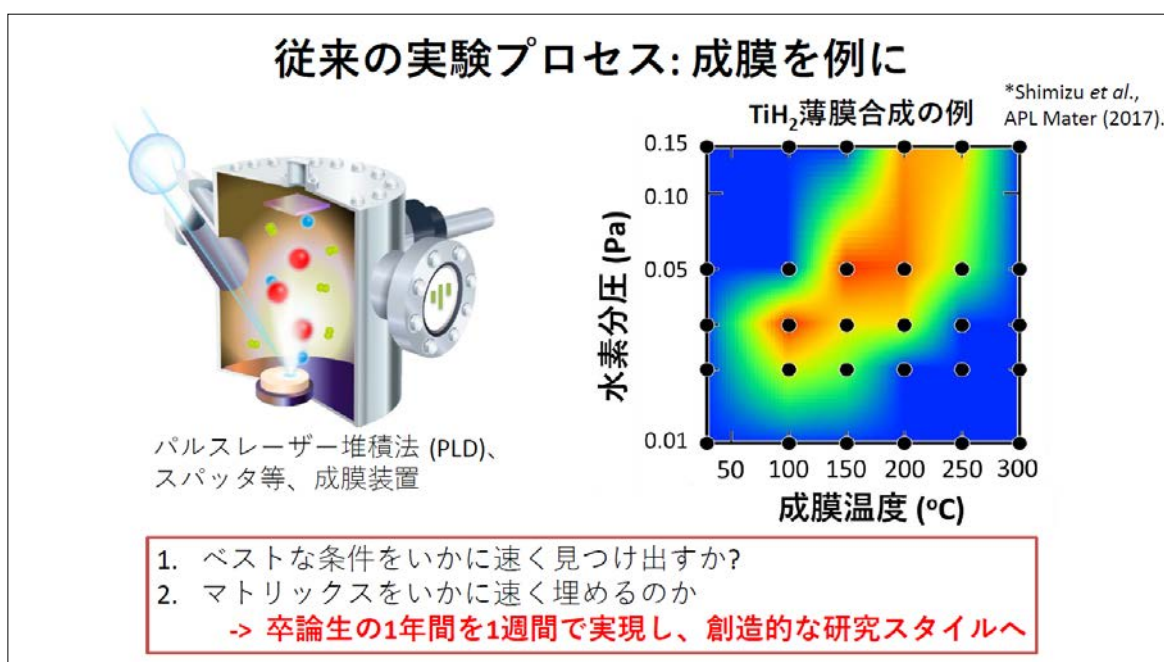


図 2-2-1 従来の薄膜合成プロセスの課題

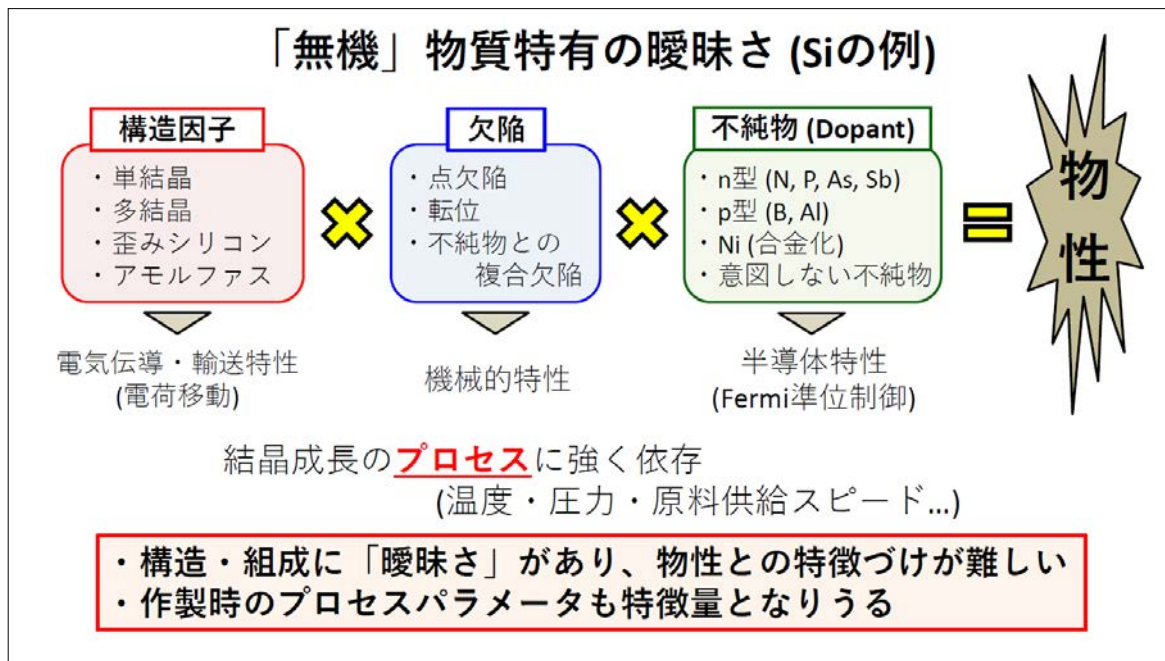


図 2-2-2 無機材料の特性とプロセスの関係

らにプロセスパラメータ自体も特徴量となるという複雑さを持つことが材料開発の難しさであり、面白いところでもあろう。

そこで我々は、このようにパラメータが多い状況で目的物性を持つ材料やそのプロセスをいかにして見つけるかということで、Closed-loop、つまり自律的な物質・材料探索に着目して、薄膜材料をターゲットとした研究を進めている。

例えば図 2-2-3 の場合は、「TiO₂の抵抗値を最小化する」という命題の下に、ベイズ最適化によって合成条件を指示し、その指示にもとづいてロボットが材料を自動で合成・評価する。その結果を返すことによって逐次的に機械学習部分が学習して、また合成条件の指示を行い、ロボットによる合成・評価が進む、という全自動サイクルを回している。(具体的な実験装置およびシステム動作のアニメーションは、「AI、ロボット、

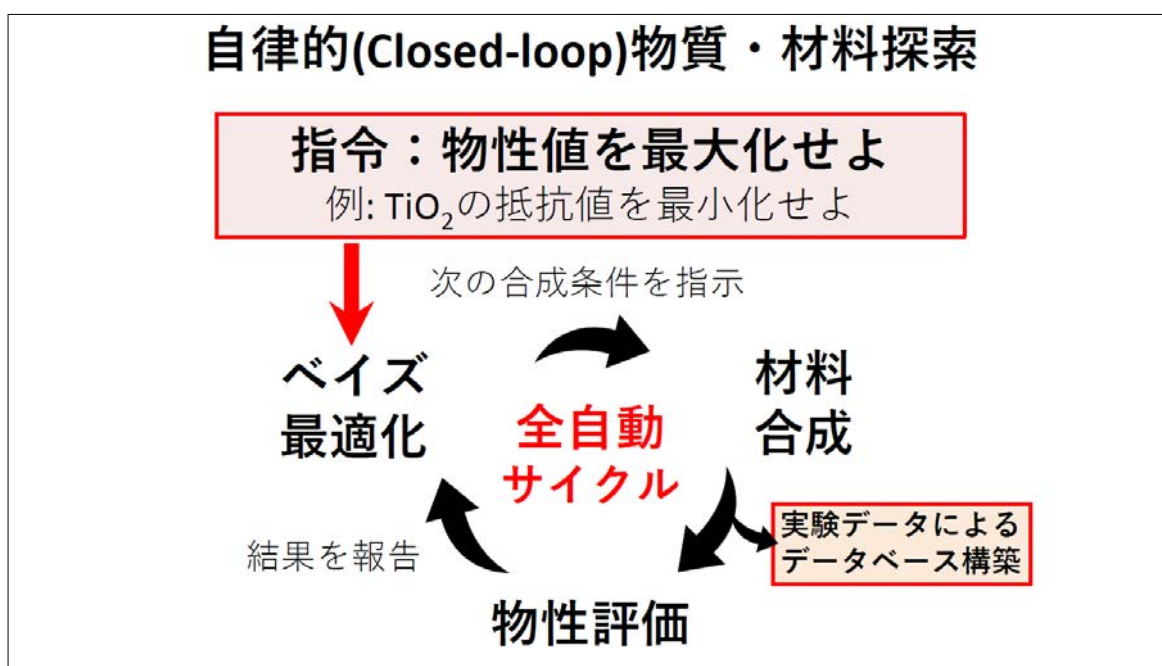


図 2-2-3 機械学習とロボットを活用した自律的 (Closed-loop) 物質・材料探索

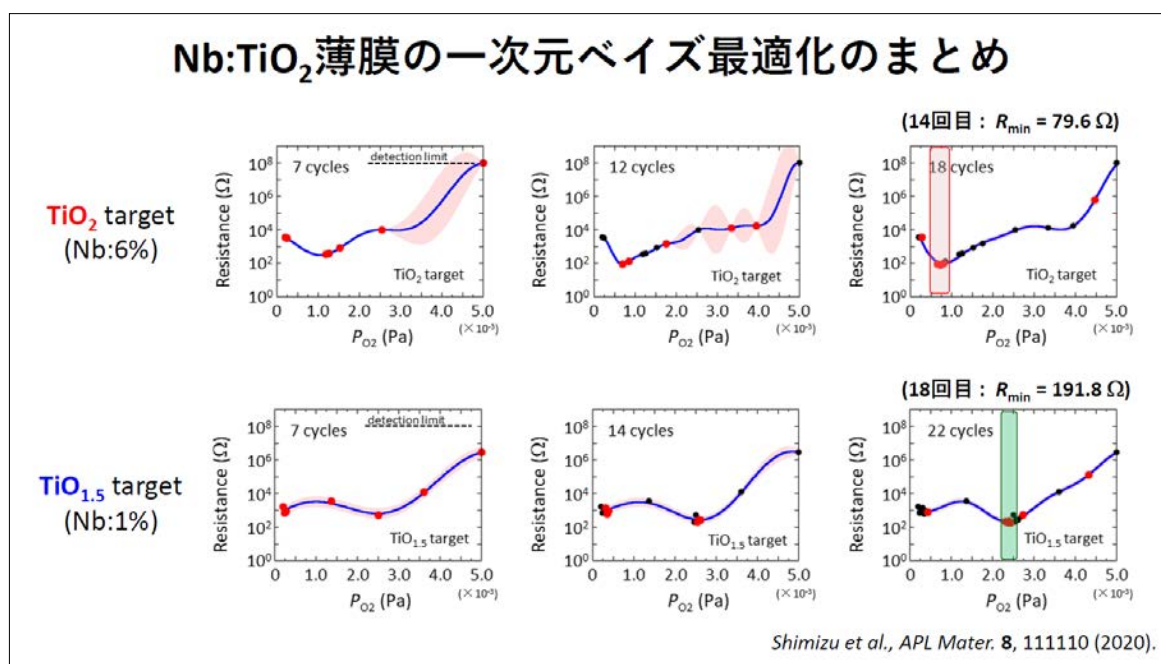


図 2-2-4 TiO₂薄膜の一次元ベイズ最適化

研究者が協働するデジタルラボトリ」<https://www.youtube.com/watch?v=VpcWW3ozarE>を参照されたい。)このような自律的材料合成の取組は、すでに創薬、有機化学、太陽電池の分野で先行しており、億レベルの選択枝のなかから合成条件を見出すことが可能になりつつある。

図 2-2-4はTiO₂薄膜に対する1次元ベイズ最適化の例である。抵抗値を最小化するような合成条件、ここでは酸素分圧の探索を行う。実際には、ロボットがある条件の薄膜を合成し、その物性値にもとづき新たな予測を行い、何度か繰り返すと予測曲線が収束し最適値が見つかるといった流れである。

図 2-2-4の上の例は18回の予測サイクルを行ったものだが、15回目ぐらいで最適な条件にたどりついている。

図 2-2-4下はTiO_{1.5}という酸素の少ない状態から始めた時の予測曲線を示している。注目すべきは、酸素リッチなTiO₂から始めた場合(図 2-2-4上)はプロセス中に供給する酸素は少なくて済み、酸素プアなTiO_{1.5}から始めた場合はプロセス中に供給すべき酸素は多くなるということで、きちんとそのずれを認識して予測できるということである。

これらをもとにClosed-loopシステムの利点をまとめる(図 2-2-5)。1点目は、局所解に落ちずに大域的に最適化ができるということである。これはベイズ最適化を活用する利点であろう。2点目は、一時的に局所解に落ちたとしても、次に述べるロボットの生産能力で探索範囲を拡大することでカバーできることである。つまり、Closed-loopシステムはグローバルミニマムを探索するのに非常に向いているシステムであると考えられる。

このロボットの生産能力という点では、～10倍のハイスループットを実現している。図 2-2-5右下のタイムスタンプによると、このシステムは大体2.5時間1個サンプルを作っている。つまり、1日に大体10個のサンプルを休みなく作れるということで、この休みなくという点がロボットの強みであろう。実際に、人が作業を行うことを想定すると、1日2個、休みを入れるとして週4日で8個のサンプルしか作製できず、全自動システムと比べると1桁ぐらい違っており、10倍のハイスループット化に成功している。

ここまでは1次元ベイズ最適化の例を紹介したが、2次元への拡張にも成功している。具体的には、酸素分圧とアニール温度をパラメータとした際にも、50点ほどのサイクルを回す、つまり5日間ほどで成長窓がどこにあるかすぐに見つけることができる。さらに、In酸化物系透明導電膜の成膜についても2次元ベイズ最適化の適用に成功している。この場合には、図 2-2-6のように成膜レート(スパッタ出力)と酸素分圧をパラ

Closed-loopシステム利用による恩恵

1. 大域的最適化の探索が可能

- ・ベイズ最適化の活用と「探索」
- ・ロボットによる生産能力(後述)



局所解に落ちることなく大域解を探索可能

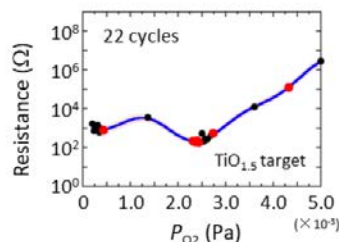
2. ~10倍のハイスループット

- ・ Closed-loop cycle:
--> ~ 10 個/日 (24 hours, 7 days)
- ・ 人力:
--> ~ 1 個/日 (休息, 休日, 会議含む)

(参考) 1週間での比較

全自動: > 70個/週

人力: 1日2個、週4日作製 -> 8 個/週



Timestamp for Nb(6%):TiO₂ target

Date	n	P _{O2} (Pa, × 10 ⁻³)	Resistance (Ω)
2020/2/27 15:45	13	4.471	6.37E+05
2020/2/27 20:05	14	0.729	7.96E+01
2020/2/27 22:36	15	0.654	9.25E+01
2020/2/28 01:00	16	0.276	3.52E+03
2020/2/28 03:24	17	0.805	9.80E+01
2020/2/28 05:47	18	0.767	9.20E+01

~2.5 h
🌀

図 2-2-5 Closed-loopシステムの利点

メータとしている。興味深いことに、成膜レートを上げた場合に成長窓が少し広がっていることが分かる。さらに少し高圧側、図で右側に寄っている。これは成膜レートの上昇に伴い酸素との衝突頻度がプロセス中で低くなった結果である。成長膜が広がる、つまりロバスト性が高くなるという意味ではもの作りの観点では重要である。

さて、ここからはプロセスデータの活用と課題について、2件の実例を通しながら紹介する。まず、先に少し触れたTiO₂の2次元ベイズ最適化の失敗例である。

図 2-2-7 左は成功例であるが、これは実はいろいろと試行錯誤を行った結果である。実際には右図のように収束しない失敗例も生じていた。図2-2-8はなぜ失敗したのかを考察した結果である。図のように薄膜合成プロセスは3段階あり、それぞれ1. クリーニング、2. 蒸着、3. 蒸着後の加熱、に分かれている。それぞれ

In酸化物系透明導電膜の成長時の原子論的考察

説明変数	変域	グリッド数*	最小刻み
酸素分圧 P _{O2} (横軸)	0.2 – 9.0 × 10 ⁻³ (Pa)	177	0.05 × 10 ⁻³ (Pa)
スパッタ出力 (縦軸)	75 – 125 (W) (堆積時間は反比例)	51	1.0 (W)

*177 × 51
= 9,027 点

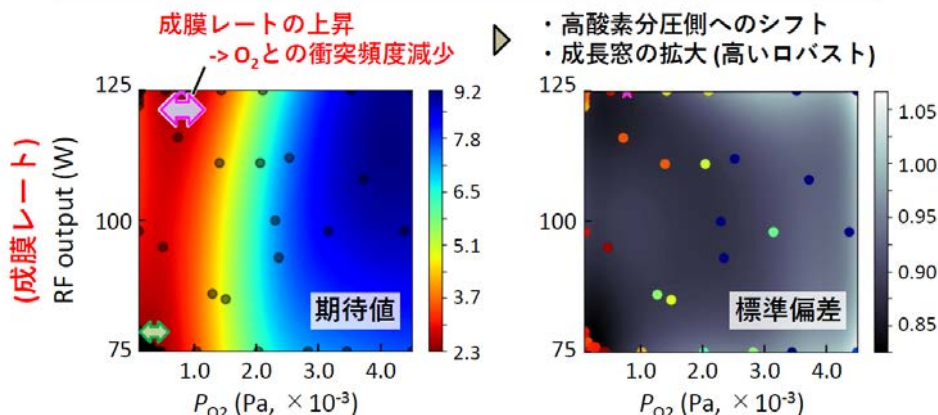


図 2-2-6 In 酸化物系透明導電膜の成長における原子論的考察

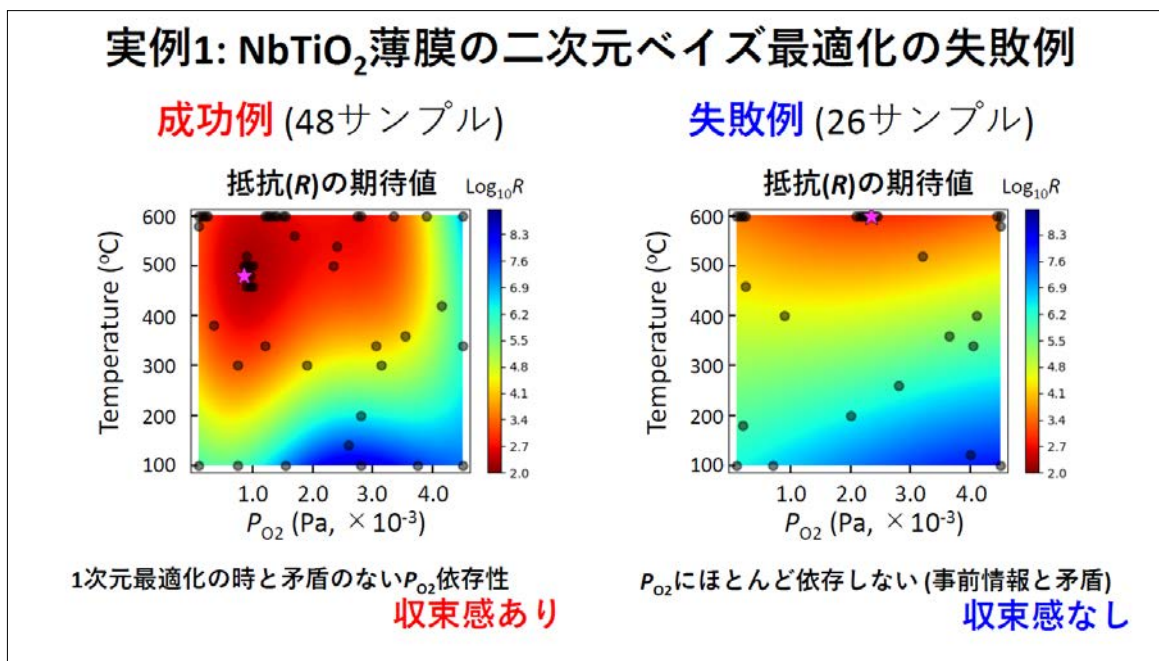


図 2-2-7 NbTiO₂ 薄膜の二次元ベイズ最適化の成功例と失敗例

の段階の系の状態をつぶさに見ていくと、蒸着の段階でガスのレートが非常に落ちていることが分かる。つまり、データベース上では「酸素が入った状態で蒸着している」と記録しているが、実際には一部のバルブが開いていなかったために、適切に酸素が導入されていないことに原因があることが分かった。

この例から分かるように、プロセス・インフォマティクスというデータ科学的手法を取るという意味では、そのデータがどれだけ正しいかをきちんと担保することは非常に重要であると考えます。特に他者が作成したデータも含めて共有、解析を進める上では、その正しさを後でトレースできることが絶対必要であろう。そのためにはプロセス中で起きているデータを逐一取得しておくべきであろう。それはトレーサビリティという観点だけでなく、モデリングやシミュレーションの精度向上にもつながるであろう。

もう1つの実例として、水素化物について紹介する。この材料はアンモニア合成・分解触媒として注目され

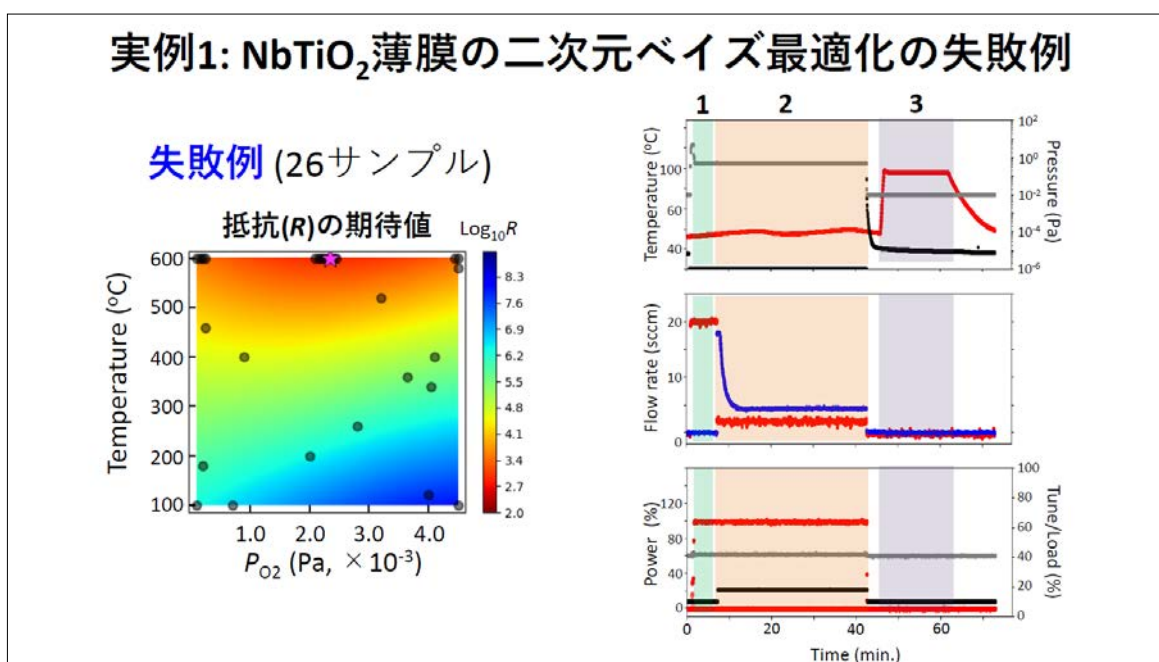
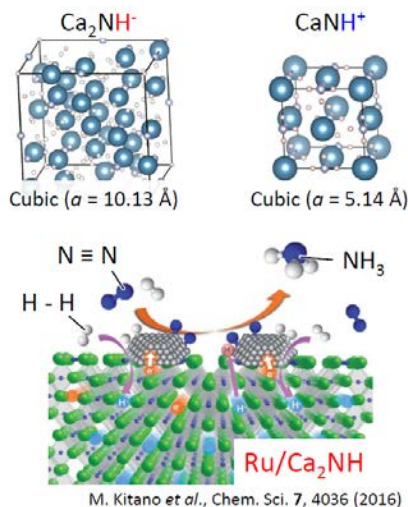
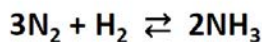


図 2-2-8 NbTiO₂ 薄膜の二次元ベイズ最適化の失敗例の分析

実例2: $\text{Ca}_2\text{NH}/\text{CaNH}$ (NH_3 合成/分解触媒)



成膜：反応性マグネトロンスパッタ

- ・ Ca金属ターゲット
- ・ $\text{Ar} + \text{H}_2 + \text{N}_2$ 混合ガス中でスパッタ成膜

Ramanスペクトルの P_{H_2} 依存性

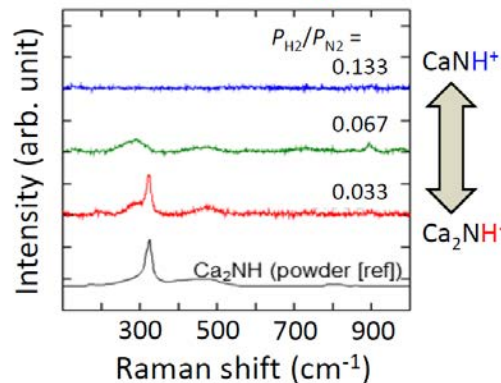
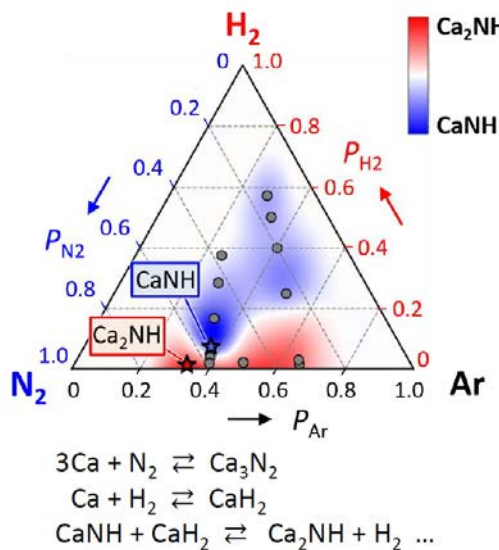


図 2-2-9 水素化物触媒の例

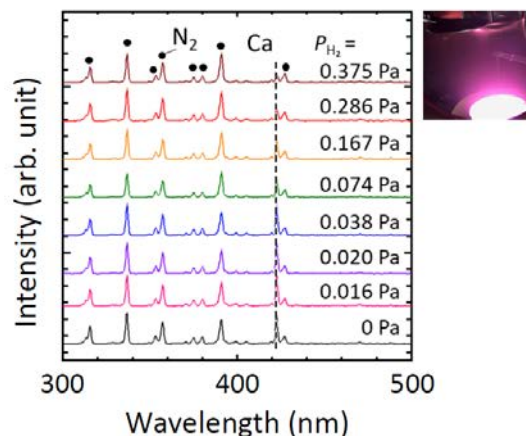
ており、水素の荷電状態が異なる構造が2つあることがポイントである（図 2-2-9）。どちらもサイズの近いCaの面心立方格子を副格子にもつ類似の構造をもつ。そのため、X線回折によるキャラクタリゼーションが面倒だが、ラマン分光を使うと比較的容易に区別できることが知られている。問題意識としては、ヒドリド状態とプロトン状態をどのように作り分けるかを調査することになった。

水素を少なめにするとヒドリド状態となり、水素を増やすとプロトン状態となるという結果となった。実際に図 2-2-10の三角相図が得られ相境界も図に示されるようになっていたが、素過程としては非常に複雑であり、現在も考察中である。その中で1つ参考になるのは、成膜中のプラズマの発光スペクトルである。水素分圧依存性を変更していった場合のプラズマの発光状態に着目した場合に、カルシウムの発光ピークが水素を増やすと小さくなり、水素を減らすと大きくなるということが分かった。これはカルシウムのフラックス量や化学

実例2: $\text{Ca}_2\text{NH}/\text{CaNH}$ 成膜中素過程の考察



スパッタ成膜中のプラズマ発光スペクトル



H_2 分圧, Ca供給量などが相決定に重要

図 2-2-10 分光スペクトルと成膜素過程の考察

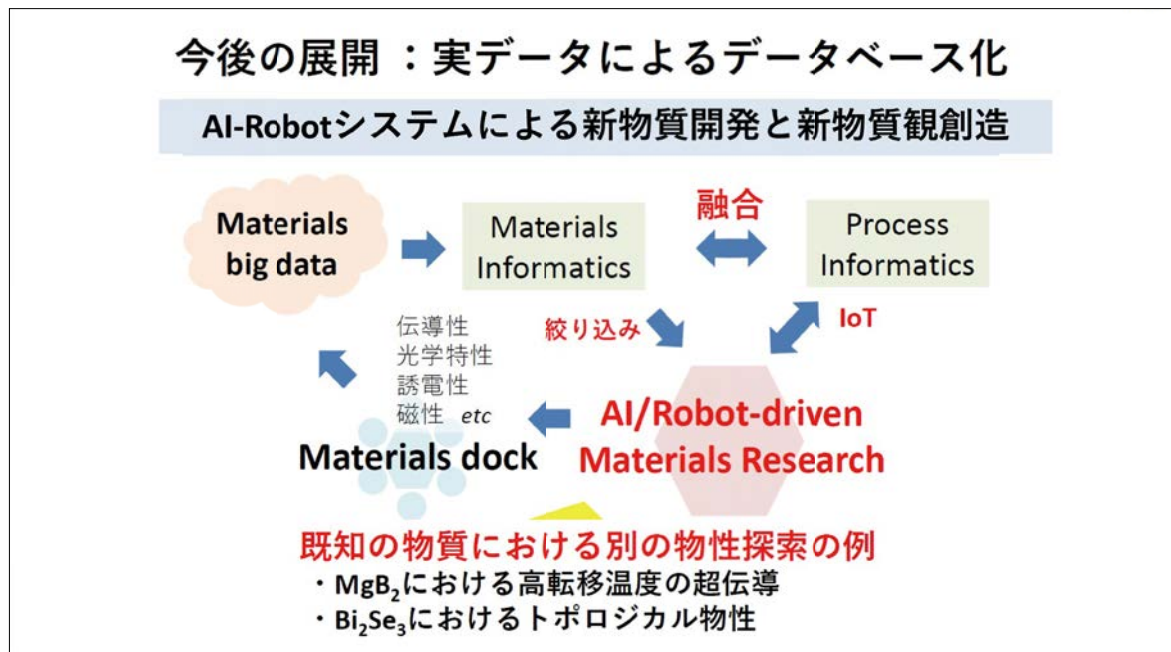


図 2-2-11 データベース拡張によるMIとPIの展望

状態が相決定に重要であるということを意味しており、やはり成膜中のプロセスを十分にモニタリングすることが重要であることを示唆している。先ほども述べたようにプロセスをリアルタイムでモニターし、取得できただけのデータを取得することが重要になると考える。

最後にマテリアルズ・インフォマティクスとプロセス・インフォマティクスの関係について展望を述べる（図 2-2-11）。マテリアルズ・インフォマティクスで予測された新物質をプロセス・インフォマティクスによって、プロセスの原理を理解したうえで効率的に合成するというのが1つのループである。さらにその際に得られる様々な物性データもビッグデータ化していくことで、既知の物質であっても、これまでに知られていなかった別の物性探索の道が開けてくる可能性があると考えられる。

最後に、データ管理の問題も重要である。「データの取得・保存形式（標準化・網羅的・信頼性・共通フォーマット）」「データの管理主体（public/private）」「データの共有/秘匿性（アクセス権、有料 無料）」といった観点が課題になるであろう。

具体的には、リアルタイムモニターといっても、どのように取るか、どれだけ取るか、それがどれだけ信頼に値するのかというところをちゃんとトレースできることが重要である。さらに、そのデータを集めるといったときに、誰が管理するのかといった問題は非常にセンシティブである。その共有をどれぐらい広げるのか、クローズにするのか、有料にするのか無料にするのかといった、拠点としての基盤をどういう仕組みでつくるのかは、やはり誰か旗振り役がいないと進まないであろう。

以上、機械学習とロボットを活用したプロセス・インフォマティクスについて現状と展望を述べたが、今後データ共有に向けた取り組みや具体的な運用に期待したい。

【質疑応答】

Q：プラズマ発光の分光分析は空間分解能はあるか。空間分解能があれば、そのデータをもとにして流体モデルが作れる。

A：今はできていないが、技術的には可能であるので検討してみたい。

Q：なぜそのプロセスが良かったのかを振り返るには、失敗例がもっとも重要だと考える。今回の発表では、実験データの質を確かめるという点で人の考察が入っていた。今後AIで自動化が進むなかで人が

どのように介入するかというヒントを得た気がした。人が介入する、つまり考察をもとに実験装置を改良するといった場合に注意点などあるか。

A：実験装置の改良自体は簡単にできるが、それによって改良前の状態を再現できるかなどを確認する必要がある。その意味では、プラズマ状態等をモニタリングし機械学習することによって、改良前状態の再現までの時間短縮もできる可能性がある。また、装置間における個体差を埋める観点でも、情報共有に有用だと思う。

Q：途中で計測のパラメータを増やす、つまり新たな計測装置を導入したような場合に、計測装置を導入する前と後のデータを併せて最適化を行うことは可能か。

A：考え方としては、できるだけ古いデータを活かす形のデータ構造やモデルを構築していきたい。それには古い虫食いのデータでも、新しい情報がリッチなデータと可能な限り等価に扱えるような仕組みが重要だと思う。それにはできるだけ取れるデータはデジタルデータとして残しておくことが重要。

Q：装置の設計、実験条件の選定、目的関数の設定はあくまで人の仕事だと認識している。これらをあえて全てAIに行わせることは可能だと考えるか。今回、例えば清水先生の装置を装置Aとして、ほかの全く別のプロセスを装置B、また全く別のプロセスを装置Cなどいろいろな装置をサンプルデータとして捉えてそれぞれの装置をうまく抽象化すれば、新たな装置を提案するAIを構築できるのではないかと、講演を聴いて少し考えた。

A：装置をどう設計するかとパラメータをどう設定するかは別の問題であると考えている。装置に関しては、現時点では人間が目的を持って組み上げるものだと思う。パラメータをどのようにして選ぶかは難しい問題で人が選ぶと熱力学的なパラメータになってしまうが、実際には非平衡なプロセスであり、AIならば独自のパラメータを選定してくれる可能性はあるかもしれない。その意味では非平衡なパラメータを解釈性のあるパラメータとして可視化することが今後重要になると考える。

2

個別プロセスにおける インフォマティクス活用 研究

2.3 精密有機合成とプロセス・インフォマティクス

甲村 長利（産業技術総合研究所）

有機合成は、身の周りの化学品を作るプロセスで使われている。石油やバイオマスなどの原料から、石油化学や化学工学などの方法を駆使して基礎化学品が作られる。その基礎化学品を、有機合成の手法を用いて様々な機能性化学品に変換していく。さらに機能性化学品がユーザー企業によって製品へと加工され私たちの手に届く。この一連の流れの中において、様々な機能性化学品を作ることが有機合成の大きな役割である。

有機物に限らず全ての材料に関して同様なことがいえるが、図 2-3-1 に示すように、機能性化学品の設計、製造では、まず、何を作るかという材料の設計があり、材料が決定すると、次にどう作るかという製造プロセスの設計になる。そのような流れで製品が作られる。

数年前まではこのような単純なスキームだったが、最近は変わってきている。実際のプロジェクトとして、マテリアルズ・インフォマティクスではNEDO「超先端材料超高速開発基盤（超超）プロジェクト」が推進され、プロセス・インフォマティクスはNEDO「機能性化学品の精密生産プロセスの開発」プロジェクトと認識しているが、現在は、この図式が変わってきている。

有機合成には、図 2-3-2 に示すように、大きく分けてバッチ法とフロー法がある。バッチ法は、一段一段、フラスコ、もしくは釜の中に原料を投入・反応させ、それを取り出して精製して作っていくステップ・バイ・ステップな方法である。それに対しフロー法は、ある筒の中に原料を投入したら、出てくるものが生成物となる方法である。現在、ほとんどの機能性化学品はバッチ法の繰り返しで製造されている。

フロー法は、石油化学など生産量の大きなものに適しているといわれ、機能性化学品製造にはなかなか適用されてこなかった。しかし、フロー法には、少量から大量まで小さいカラムの中で製造ができること、反応器だけを制御すれば良いこと、省エネルギー、安全であることなど、多くのメリットもあるため、機能性化学品製造にフロー法を適用することが検討されてきた。

そのような状況の中、東京大学的小林修教授らが、ロリプラムという医薬品の原料を、4つの触媒反応を連結させ、原料から生成物まで途中で精製せずに合成する方法を報告した（Nature 520, 329（2015））。少量スケールではあるが、非常に画期的な成果であった。

小林教授らの成果をベースにして、我々は、経産省の省エネ型化学品製造プロセス技術の開発事業の中の1

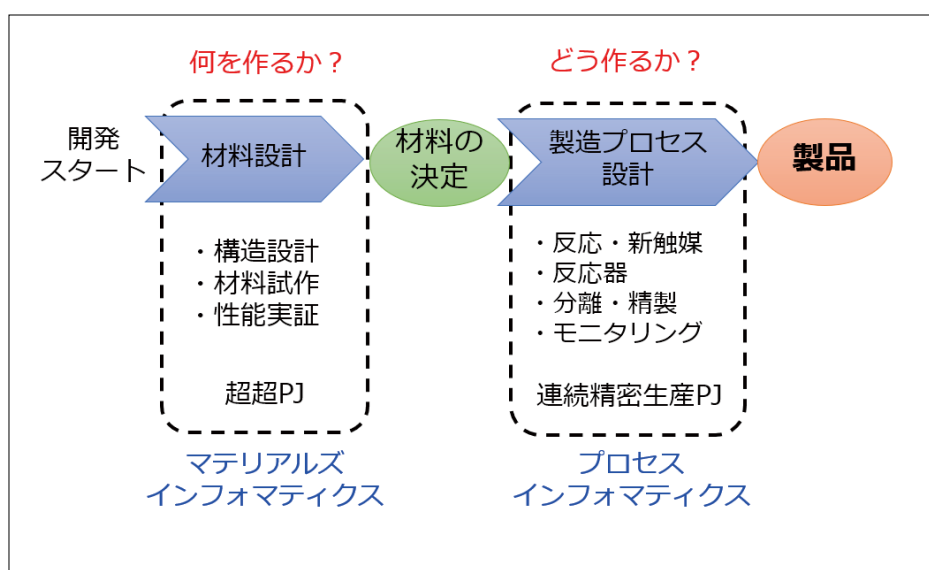


図 2-3-1 機能性化学品の設計・製造プロセス

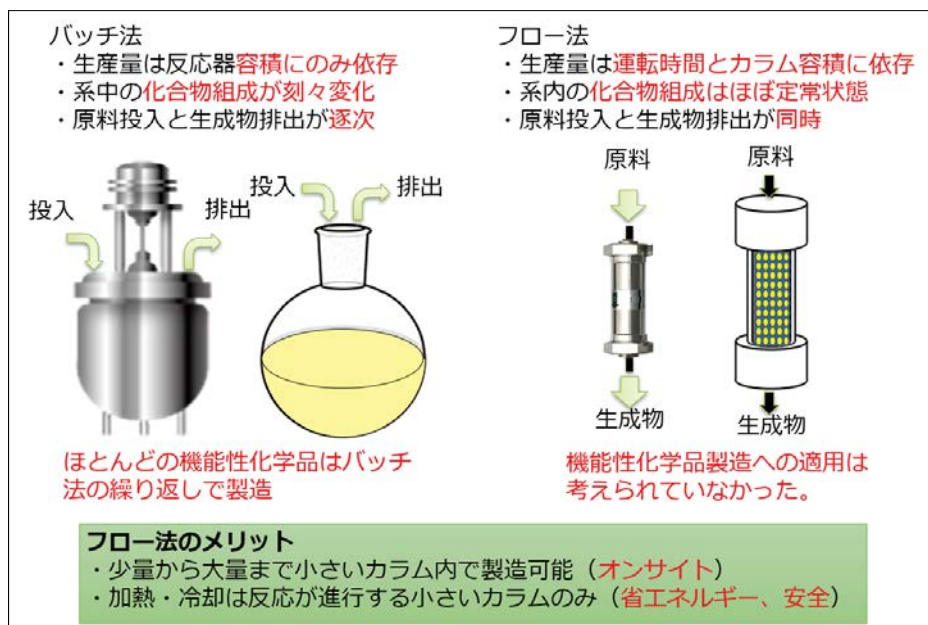


図 2-3-2 バッチ法とフロー法の違いとフロー法のメリット

つのトピックスとして、機能性化学品の連続精密生産プロセス技術の開発をプロジェクトとして取り組んでいる。

このプロジェクトの概要を図 2-3-3 に示す。これまで行われてきたバッチ法をフロー法に置き換えて、機能性化学品製造のためのプロセスの効率化、省エネ化、省コスト化を目指している。基幹にある5反応を軸に、反応・新触媒、反応器開発、連続分離・精製、システム開発を行い、オンデマンド型の連続精密生産プロセスの構築を目指す。

本プロジェクトでは、ある1つの製品を作るための装置の開発や方法の開発を行っており、どういうルートで合成するかという検討は含んでいない。機能性化学品の設計、製造として何を作るか、どう作るかという部分を対象としており、合成経路の設計は対象としていなかった。

合成経路設計は非常に大切だが、これまでは、有機合成科学者の経験と勘でカバーしてきた。つまり、合成経路設計は有機合成化学者のまさに醍醐味、腕の見せ所であり、有機合成におけるオリジナリティの源泉となるところであった。しかし、今は、ここをプロセス・インフォマティクスとして捉えて研究するべきだとい

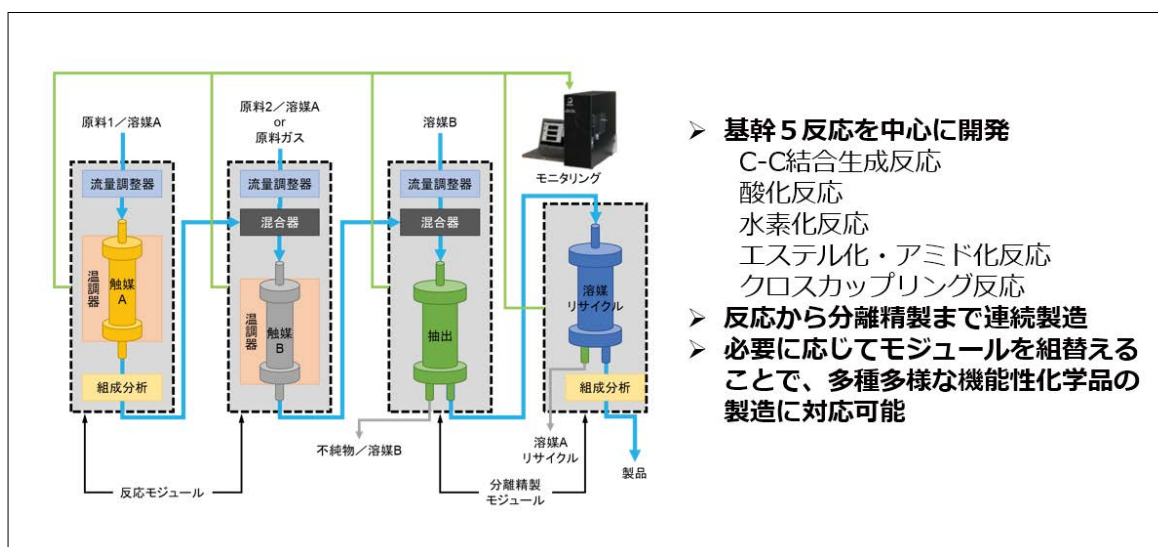


図 2-3-3 NEDO 機能性化学品の連続精密生産プロセス技術開発プロジェクトの概要

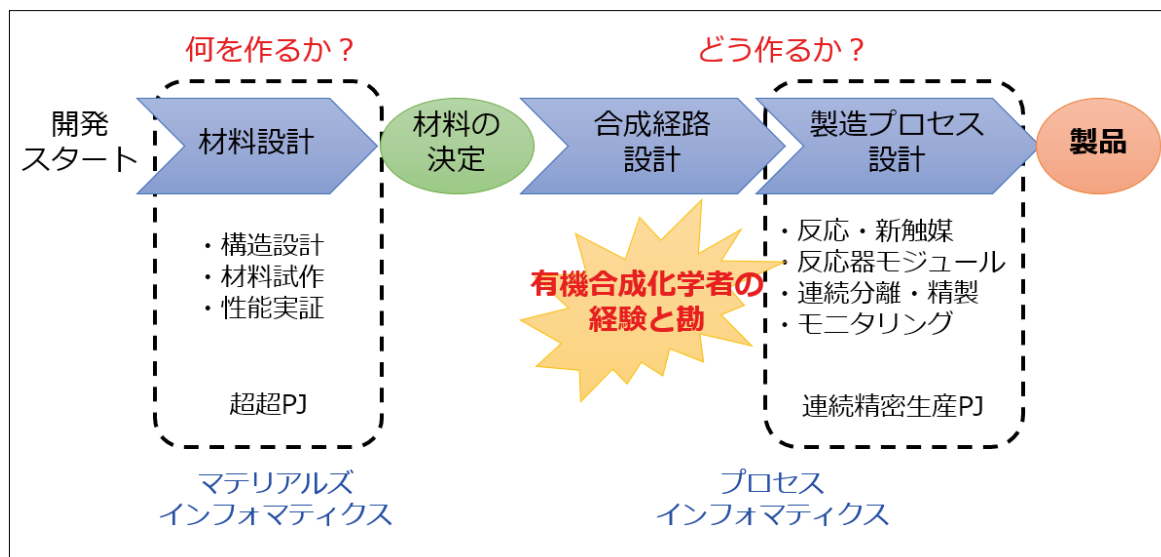


図 2-3-4 機能性化学品の設計・製造 - 合成経路設計 -

う共通の認識になってきている。合成経路設計については方法が確立しているわけではないので、つくり上げて行く必要がある。

現在ではいくつもの有機合成経路提案システムが公開されている。これらは合成反応のデータベースをもとにAIを使って合成経路、反応条件の提案を行うものであり、ツールが公開されている。合成反応のビッグデータを利用したシステムとして例えば、Synthiaが有名だが、論文などのデータの多くは海外のACSやElsevierなどが押さえており、自由に使うことができないという問題がある。今は公開されていないが、日本でも東京大学の船津先生がAIPHOSを構築している。自動的に逆合成ルート設計を行い、新規化合物の合成ルートを経験のデータから類推できるシステムである。ただし、非常に多くの合成ルートが提案され、どのルートが一番よいかを知るには、実際にプロセス実験を行なう必要があるという問題点もある。

何を作るかはマテリアルズ・インフォマティクスで設計できても、実際の合成が困難であることがあるため、合成経路設計、製造プロセスの検討を深める必要がある。昨年6月からNEDO先導研究プログラムで、AI

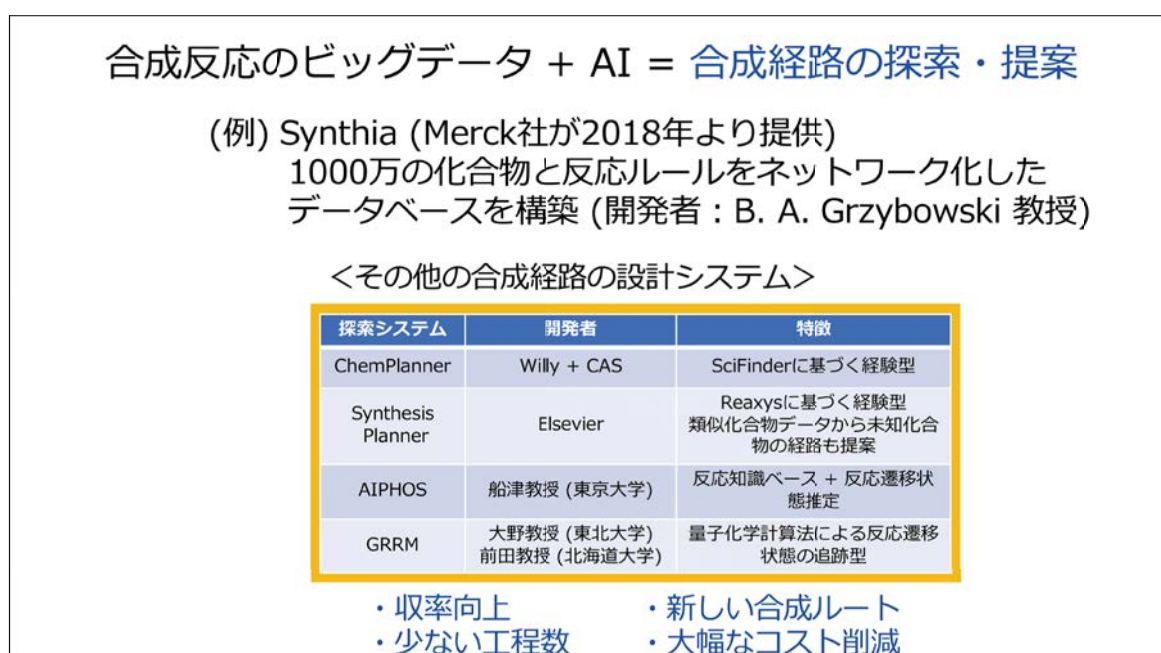


図 2-3-5 有機合成経路提案システム

による分子設計、計算化学による合成経路開発、マイクロリアクターを用いた実験化学による合成経路検証による機能性化学品創成イノベーションサイクルを構築することを目指したプロジェクトを開始した。AIにより非常に多く提案される合成経路を、計算化学によって絞り込み、絞り込んだ結果について高速で合成検証を行い最適化するというプロジェクトである。

現在進行形のプロジェクトであるため細かい内容の説明は控えるが、図 2-3-6 でコンセプトを紹介する。反応のデータベースをもとにAIから提案される合成経路は現在のシステムでは非常に多い。この絞り込みに計算化学を利用する。仮想実験として計算化学を使おうとしているが、今でも、計算にはコストや時間が非常に多くかかる。このため高速遷移状態解析を駆使して非常に速くこの経路を導き出そうというコンセプトである。この中で、自動化とUI開発が非常に重要である。計算化学によって製造プロセスの最適化を行った結果、プロセスが改善した例もあり、非常に有効な手段だと私たちは考えている。

材料を決定した後はどうするかを、NEDOの先導研究と連続精密生産プロジェクトを組み合わせで検討しており、この全体を機能性化学品製造を加速するプロセス・インフォマティクスであると捉えている。

さらに、図 2-3-8 に示すように連続生産に関する3つのプロジェクトの融合を融合させることが我々のコンセプトである。これから大きくしていきたいと思っているNEDOの先導研究、現在の連続精密生産プロセス技術に加えて、戦略的エネルギー技術革新プログラム事業を併せて有機合成におけるプロセス・インフォマティクスを充実させていきたいと考えている。

プロセス・インフォマティクスに関連した海外の事例としては、グラスゴー大学のクローニン教授がロボットと人工知能で新反応を発見したという成果を発表している。また、有機合成実験の自動化に関する研究も行っている。実験室レベルでのバッチ反応ではあるが、人の手や人の頭を借りずに機能性化学品の合成を達成できている。最近、人が寝ている間にやってくれるような、移動型ロボットのシステムも開発された。どこまで複雑な機能性化学品が作れるかは分からないが、こういったプロトタイプ的システムの開発について、日本は大分遅れているという印象を持っている。

また、令和2年2月に公開された特許庁の「特許出願技術動向調査；マテリアルズ・インフォマティクス」の中に6つの提言がある。提言1としてはMIを活用した研究開発を継続的に進めることがあげられている。また、提言2は製造プロセスとデータ科学（プロセス・インフォマティクス）に関するものであり、「材料と製

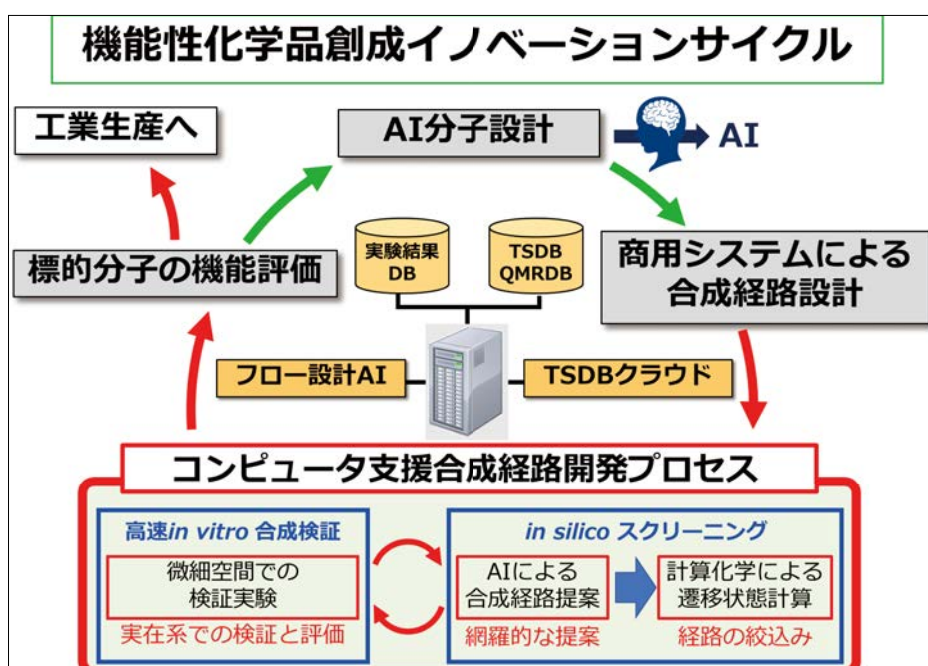


図 2-3-6 計算化学を用いた仮想実験による合成経路検索（NEDO 先導研究のコンセプト）

造プロセスとの相互関係は強く複合化しているので、両者を一体で検討することも必要」と書かれている。有機合成化学に関しても、プロセス・インフォマティクスを活用してさらなる開発の加速をしなければならないと考えている。

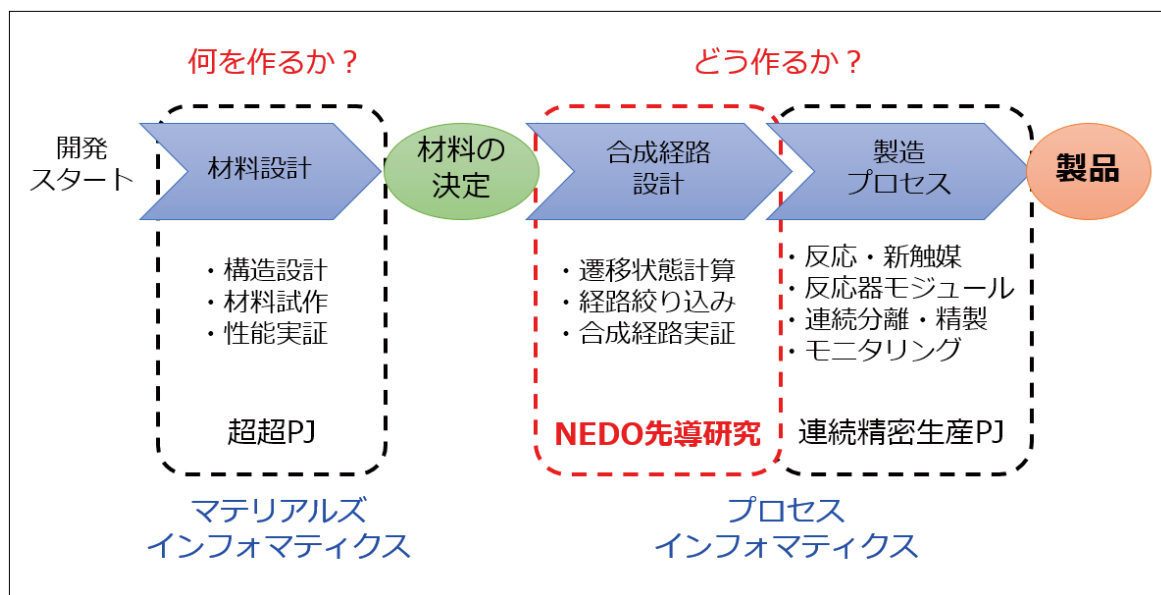
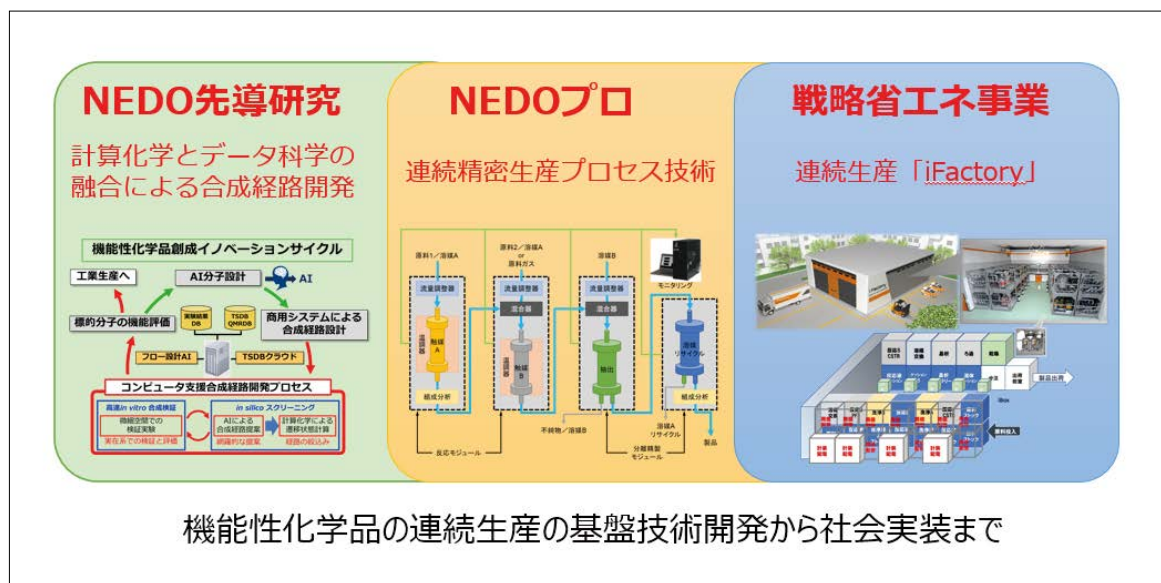


図 2-3-7 機能性化学品の設計・製造プロセスにおけるプロジェクトの位置付け



機能性化学品の連続生産の基盤技術開発から社会実装まで

図 2-3-8 機能性化学品の連続生産に関する3プロジェクトの融合

【質疑応答】

Q：計算化学で絞り込むということだったが、実際の反応では、温度や濃度などの最適化の条件により副生成物のでき方が違うなど、いろいろ変わる。そういうものはどういう仕組みで最適化するのか。

A：計算化学では、副生成物も含めて素過程を全部洗い出して、ターゲット化合物を求めていく方法を取っている。その副生成物、遷移状態から違うものに行ったり、違う遷移状態を通して変なものができたりするところも計算できる。計算化学から導き出された結果を、実際にマイクロリアクターを使って検証することも行い、収率等、計算結果との比較をする試みを行っている。

Q：マテリアルズ・インフォマティクスで何を作るかを決めて、その後マイクロフローリアクターを使って合成するという一連のパスがあり、その途中でプロセスの設計が入っていた。その場合、マイクロフローリアクターができることを制約条件としてプロセスの設計、あるいは反応経路の設計をやるのか。

A：NEDO先導研究では、マイクロフローリアクターの活用は、計算化学の検証という位置づけである。基本的には、バッチ反応でも高速遷移状態解析を使えるようにしたいと思っている。バッチからフローという流れを作ろうとしているが、実際に企業の中でそれが入れ替わっていくのには時間がかかると思うので、この考え方をバッチでも使えるようにしたい。ただし、せっかくマイクロリアクターでの検証を行っているので、フロー精密合成にも使えるようなプロセスを考えている。

Q：素反応の洗い出しと素反応にかかわる触媒の役割を解明するだけでも膨大な仕事量になるような気がする。いろいろな素反応を計算化学で洗い出す難しさを、どのように超えようとしているのか。

A：計算化学でできることは増えてきたとが、やはりまだ課題はある。例えば溶媒の効果を正確に評価することなどである。また触媒反応では、そもそも触媒表面の計算はなかなか難しいため、実はまだそこまですべてで至っておらず、今後の課題である。

Q：バッチ法からフロー法にして、それぞれをモジュール化して組み合わせるとなると、ある種のミニプラントのようなイメージを持つ。このフローの中での反応は、基本的に不均一系なのか。

A：不均一系である。この反応器、カラムの中に触媒を入れて反応剤を流すので、触媒が溶媒に溶けると出てきてしまうため、溶けない触媒を使う必要がある。このため、不均一系反応である。プロジェクトではフロー法用の触媒開発や反応開発も検討している。

Q：反応のそれぞれのモジュールでうまくいったかどうかはどのように評価しているか。最後に出てきたものはプロセス全部が終わった後の結果として分かると思うが、それぞれのモジュールでの評価もしなくてはいけないと思う。

A：現状は最終の生成物の解析だけに止まっている。反応のモニタリングはこれからの課題である。分析結果をフィードバックする方法なども今後の技術開発の1つである。

2.4 粉体プロセス開発のハイスループット化のためのプロセス・インフォマティクスと材料探索ハイスループット化との関係

長藤 圭介（東京大学）

私の専門は機械で、粉体プロセスについての研究を行っている。本日の他の講演者の先生方とはバックグラウンドがだいぶ違うので、議論をかき混ぜることになるかもしれないが、その点は御容赦いただきたい。

これまで、粉体やレーザで表面を加工したり、3Dプリンタで造形したりといったことを主にやってきたが、その中で、現場で学生や若手がやっている勘・コツ・経験の限界を感じており、それがきっかけでプロセス・インフォマティクスを去年からJSTの未来社会創造事業で取り組ませていただいている。今日の発表は、その中での取組の内容が基盤になる。

まず、日系企業の製品の世界シェアの中では、機能性材料はシェアが大きく、それを支えているがゆえにエンドユーザーの製品も強いと言える。ところが、世界的には、材料探索やプロセス探索がハイスループット化して非常に開発スピードが速くなっているの、その状況の中で、日本がどうしていくかは大きな課題になっていると考える。

私は「材料は買ってくるもの」という立ち位置だが、粉体とか成膜のプロセスは、燃料電池や二次電池をはじめとするいろいろなデバイスの量産に必須技術になっている。粉体に限らずいろいろな材料のプロセスがあるが、高機能になればなるほどこういう微細構造の作り込みとの連携の寄与度が大きくなってきている。

材料研究の立場（図 2-4-1 の上流）でも、勘・コツ・経験・すり合わせに依存しているのを脱却したいというモチベーションがあると思うが、我々にも同様のモチベーションある。下流だからこそ、パターンが多く、非常に苦労しているところである。それで、データ駆動型のプロセス・インフォマティクスということに挑戦することにした。

「マテリアルズ・インフォマティクス」という言葉の定義は、だんだん拡大解釈されつつあると個人的に感じていて、計算科学で候補物質を出していくだけではなく実際に合成できればいけないという流れに次第になっているように思う。私自身は、それよりもさらに下流の位置にいるが、それを評価し量産に持っていくこ

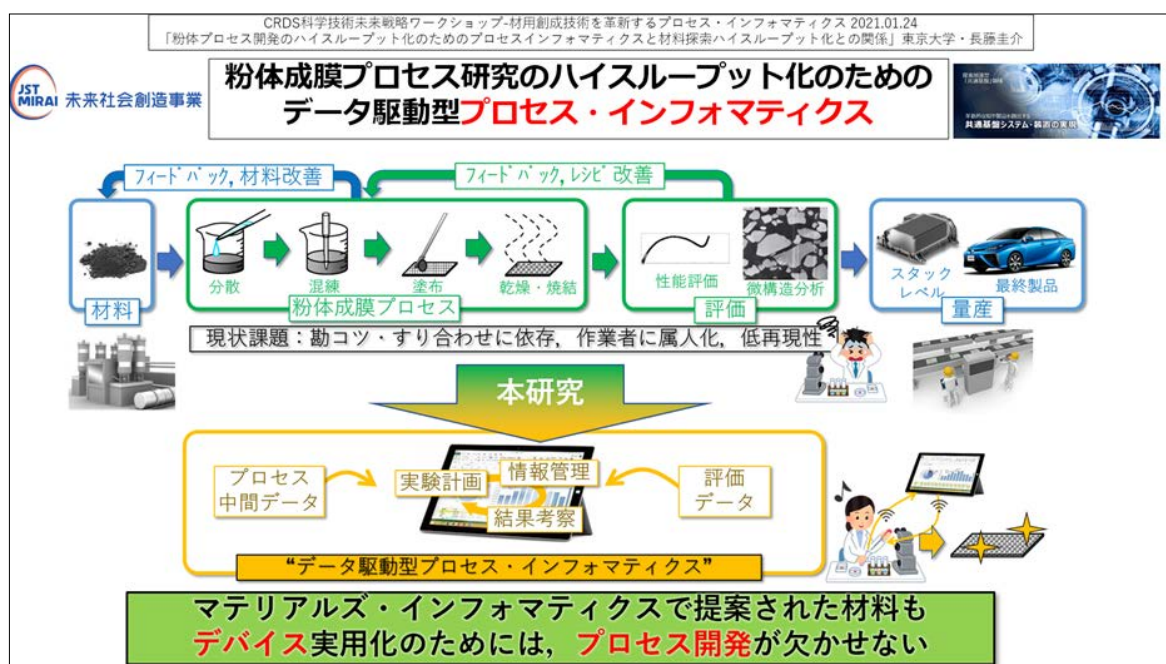


図 2-4-1 粉体プロセスにおけるプロセス・インフォマティクス

の領域は、ぽっかりと空いている。

もともと私たちはデバイス側の研究者なので、デバイス評価で勝負していた。粉体のプロセス自体に手を入れるということはアカデミアとしては論文が書けない方向に行くのでなかなか手を出しにくい領域であった。しかし、未来社会創造事業をきっかけに、そこに食い込んでいこうとしたのがこの研究のバックグラウンドである。

日本の主力製品の中で、粉体プロセスからできる製品、あるいはそれに関わる部品、そういうものは国際的に競争力が強く、電池をはじめセラミックの部品などがある。なぜ強いかというと、複雑現象であり一般化できない現象が多くあるところを勘・コツ・経験・すり合わせでなんとかモノにできているというのが背景にあると考えている。いわば「賢い絨毯爆撃」である暗黙知の中でモノを作っているわけで、それにはやはり限界があって、先ほど示したデータ駆動型のアプローチを世界がやってきたときに、ちょっと耐えられないのではないかと思っている。

粉体プロセスというのは分散、混練、塗布、乾燥がベースになっていて、これはお好み焼きを作るプロセスに似ている。多くの日本人が経験したことがあると思うが、同じ材料で同じレシピでも、人が変わるとやはりできるものが違うということがよくある。工業製品については、そういったことに対して、うまくスケールアップしてなんとかものになっているのが現状であり、今、我々が行っている固体高分子型燃料電池の作り込みに関しても、同様な経験をしている。

材料製造者が出してくるよい材料に対して、それを受けるプロセス側との組合せでよい性能を実現していくことを考えなければいけない。粉体の物理現象解明に関して、理論的に話を展開すれば、幾らでも研究テーマになり、図 2-4-2 の物理量から説明はできるが、実際の材料に対してそれをものにするとなったときに、無限ともいえるプロセスパラメータを最適化しないといけな。途中で何が起きているかをセンシングしないと、このような最適化には限界があると考え。しかも、そのセンシングは、図 2-4-2 の表の最下列にあるような物理現象を象徴するようなものでないと意味がない。そういうことで、我々は中間データを取るべきだと提唱し、それを行っている。開発のポイントは、中間データをいかに取得するか、それをいかにシステムにしてい

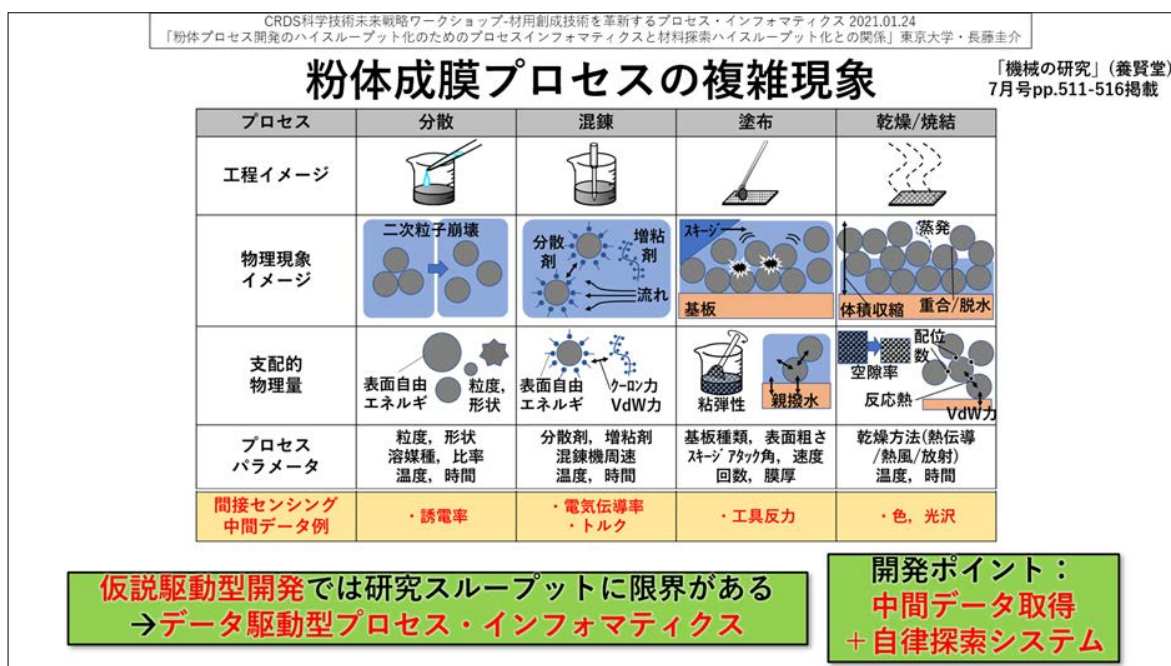


図 2-4-2 粉体成膜プロセス

仮説駆動型とデータ駆動型の開発がよく対比されるが、私は仮説駆動とデータ駆動は常にツインにあると思っている。例えば、ガリレオの斜塔の実験に関して、まず等加速度運動の仮説があり、それに対して実験を行い、結果を分析して考察するという一連の流れ（図 2-4-3 上）があるとされ、研究はこのように仮説駆動であるべきとされているが、実際には、図 2-4-3 下のように、ガリレオも最初は「よく分からないけど、やってみよう」とやってみたはずである。それをデータ解析して、あたかもそれを仮説としたかのようにストーリーにするというのが現実の研究発表の流れとなっている。

しかし、近年は、データ科学がすごく強くなってくると、「よく分からないけれどもやってみよう」というボリュームが大きくなる傾向があり、そういう意味で今は「仮説駆動型から脱却して、データ駆動へ」と言われるようになってきているが、それは結局ボリュームの違いであって、常に両方があるのではないかと考えている。

粉体の話に戻ると、イメージとしては、図 2-4-4 左側にあるような4つの粉体成膜プロセス装置を、一気にプラントにするのではなく、今のものを生かして、それをつなげるには、データ空間に持っていく中間データが必要なのではないかと考えている。その中間データには、見た目で「こうやったら良いのではないか」といった暗黙知を数値化するという意図もある。そしてできたものの評価結果も、データに上げていき、実験して、モデルを更新し、次に予測して、そして提案するサイクルをぐるぐる回すというのが、サイバー・フィジカルの機械学習システムである。

プロセスの中で中間データを取得することがなぜ必要かを説明する。我々が最終的に欲しいのは、分極抵抗や耐久性といったデバイス性能を目的に、それに対する溶媒比率や温度というような最適なプロセスパラメータを取得する方法である。だが、どんなプロセスパラメータを候補に入れるか、出来上がったものに対してどんな評価データを目的関数にするかは、いくらでもリストアップすることができるが、これが多くなり過ぎると実際に最適化が行えなくなってしまう。中間データを用いる理由は、それをモデルの中間に位置するところに入れることによって、実験の試行回数を減らせることにある。いかに優れたロボットシステムでも無限数のハイスループット実験は不可能なので、やはり次数削減のプロセスが必要になり、それが速ければ速いほど効率がよいことになる。こうした能動学習や、中間データを用いる手法は、ロボットの分野でオートエンコーダと呼ばれるが、この中間データの候補出しに人間の勘・コツを入れることが大事ではないかと考えている。現在進行中の燃料電池電極に関する開発では、中間データとして例えば「かすれている」とか、「表面粗さ」

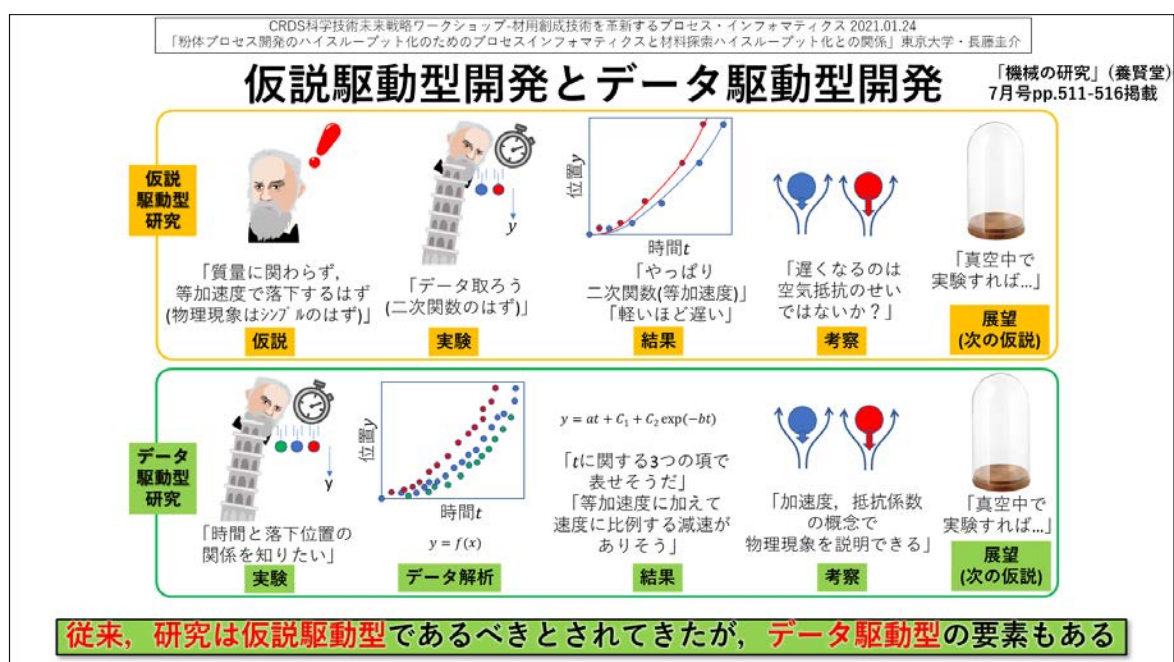


図 2-4-3 仮説駆動型とデータ駆動型

とか、「画像から得られる特徴量」といった特徴量も可能性に入れた研究を進めている。

現状のプロセス開発は、計画も作業も完全に人依存のことが多いが、プロセス・インフォマティクスでの自律システムを目指す際にも、ループの中に人を取り込むという考えが大切ではないかと思う。もちろん、そのバランスは、自動で1週間やらせて、人が1日考えるでも良いと思う。ケースにもよるが、こういうやり方が、日本ならではのプロセス・インフォマティクスではないか。これは、データ駆動-仮説駆動ハイブリッドという言い方ができるかもしれない。人の外挿力と、その中での最適解を素早く出す自律システムの内挿力、これをうまく掛け合わせたものである。Human-in-the-loopという言葉があるが、そういったパラダイムシフトが今必要なのではないか。

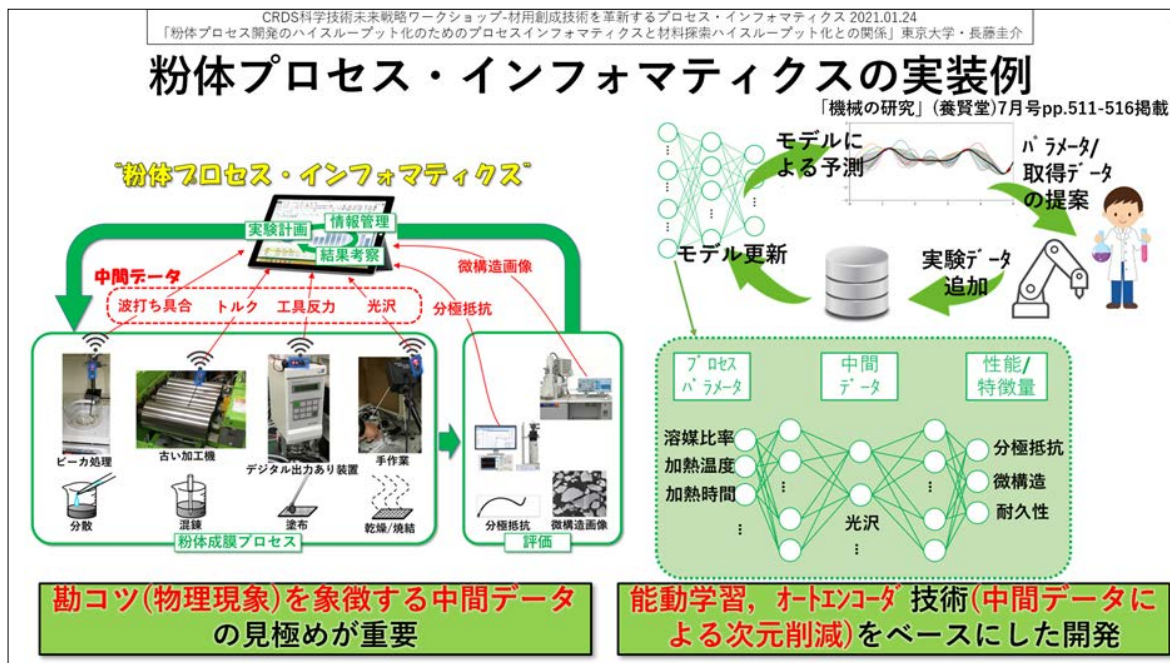


図 2-4-4 粉体プロセス用プロセス・インフォマティクスの例



図 2-4-5 研究開発現場の変遷

マテリアルズ・インフォマティクスもプロセス・インフォマティクスも、今後は、サイバーとフィジカルを一体・自律化していくものとして定義されていくべきだと考える。マテリアルズ・インフォマティクスはサイバー空間だけでというところからが始まったと思うが、プロセス・インフォマティクスはそもそもフィジカル空間を避けられないので、こうした一体化を考えていく必要がある。日本人の勘・コツ・経験をうまく取り込むような仕組みが大事である。

本日の発表内容は、未来社会創造事業で集まった機械系、化学系のメンバーなど様々な先生方との議論の中で出たものである。ここにお礼を申し上げる。

【質疑応答】

Q：物理的な意味のあるものを中間データにするということだが、ここで何を選ぶかは、やはり研究者の勘とかコツということになるのか。

A：人がやっているときに感じていることを数値にするようなものであれば何でも良いと思う。手っ取り早く画像で何とかするというのもありだろうし、工具から出てくる出力みたいなものをそのまま使うでも良い。どのデータを使ってどう組み込むかというのは人の役割だと考える

Q：最後のループの中に、例えばプロセスのシミュレーションとかそういったものを入れることは考えられないか。例えば混練とか乾燥とか塗布とかは、まだまだ不十分な部分も多いと思うが、それなりにシミュレーションは存在するのではないか。

A：自律的に実験をして、人がそれを見てというループに入れるとしたら、モデルベースのシミュレーションと解析技術がありうる。これは人がひらめくためのベースをつくるという位置づけになる。もう一つは、そもそもこのプロセスで何が起きているのかというシミュレーションもあり得る。粉の動きや乾燥の熱流体のシミュレーションなどをどう組み込んでいくかも大事だと思っている。この自律探索システム内の予測システムになるわけだが、粉体に関しては複雑過ぎて難しい面もあるが、それは大切だと思っている。

Q：今日は粉体の話をされたが、今回のプロセス・インフォマティクスはできたら共通的なもの、ある意味で普遍化したいと考える。どういうやり方が考えられるか。

A：個別のテーマから共通のテーマにしていくのは結構大変である。分野が近ければ、共用可能なプロトコルをつくってみるというトライは大切である。しかし、異分野でも使えるプロトコルにするのは難しいと考える。中間データに注目する方法自体は、共通化できる。勘・コツは人の五感からのデータを基にしているが、人の五感をDX化するというやり方の前に、物理現象を代表するようなデータ化、センシングみたいなものが有効だと思う。

C：中間の段階のパラメータを測定してスクリーニングをかけ、そこにフィジカルなところや、勘・コツなどを入れていく、そこに日本の伝統が生きていくというのは非常に心強い話である。これは、真空プロセス、プラズマプロセスなどに対しても活用できるように感じた。これをデータベースの中に取り込んで、最終的には大きなデータをつくって、その中間のパラメータをベースに記述子みたいな形でいろいろなものを見つけていく、非常に重要なコンセプトだと思う。粉体だけにとどまらず、いろいろなところ展開できる気がする。

C：欧米では、規則をつくってみんなでそれを使う、その規則は誰かがつくってくれるというようなマインドだが、日本人はどちらかというと個を大事にする。勘・コツ・経験が生きているのは諸刃の剣であり、良い面でも悪い面でもある。データ科学に勘・コツを取り込む仕組みをつくったあと、日本でどのように使うかとか、それをみんなでどう賢くしていくかのようなことが本質的な議論だと思う。

2.5 データ駆動型高分子材料開発とプロセス・インフォマティクス

内藤 昌信（物質・材料研究機構）

私は、新しい高分子機能性材料を作り、その加工プロセスや力学物性という製品に近いところまでの評価という一連の流れに取り組んでいる。

データ駆動型材料設計は、図 2-5-1 に示すように、まずはビッグデータのあり/なしでルートが異なる。ビッグデータがある場合はそれを活用して、新しい機能性高分子材料を作るところに来ている。機械が予測したものを有機合成、高分子合成の専門家が実際に合成して、それが実際に使えるのかどうか検証する流れである。

一方で、高分子材料ではビッグデータがないことも往々にしてある。その場合は、まず、高品質なデータを作り、そこから予測モデルを立てて検証するという流れになる。スモールデータセットではあるが、ビッグデータなどとうまく組み合わせて新材料を開発する。

高分子材料のビッグデータには、NIMSが所管している世界最大級の高分子データベースである PolyInfo がある。ホモポリマーの情報は約 16,000 件、物性データとしては約 33 万件が掲載されている。PolyInfo は、企業で研究していたシニアの方々が、実際に論文を読んで、その論文に書かれているデータをデータシートに写し、さらにそれを別の人がチェックしてデータベースに乗せるということを 20 年継続して集めたものである。

PolyInfo は基本的には論文ベースのデータベースであるが、往々にして論文に掲載されたデータだけでは充分でない。図 2-5-2 に PolyInfo のデータ構造を示すが、約 100 のデータタイプがあるが、多く掲載されているデータ項目もある一方で、非常に少ないデータ項目もある。熱物性や力学特性、屈折率などは比較的充実しているが、それ以外のデータはあまりない。自分が欲しいデータがない場合には、スマートラボなどでデータを集めることが必要になる。また、論文からデータだけ取り出してきたデータベースは論文をいちいち読まなくても横串を刺すような形で活用することができるというメリットがあるが、論文データには失敗データが含まれていないことが多くあるので、失敗データを含む良質なデータを自分で取って補完することが必要に

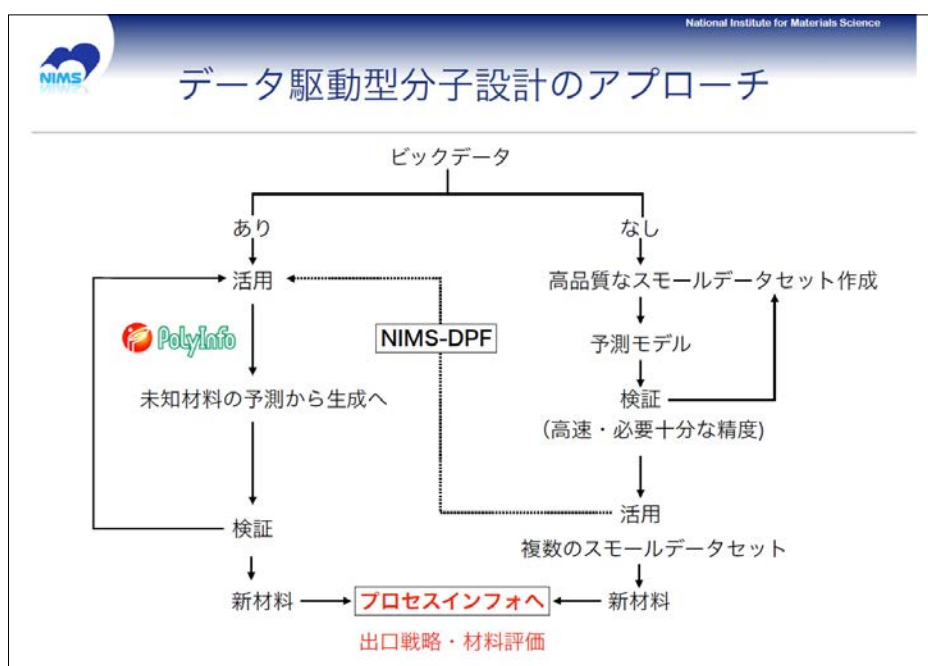


図 2-5-1 データ駆動型分子設計のアプローチ

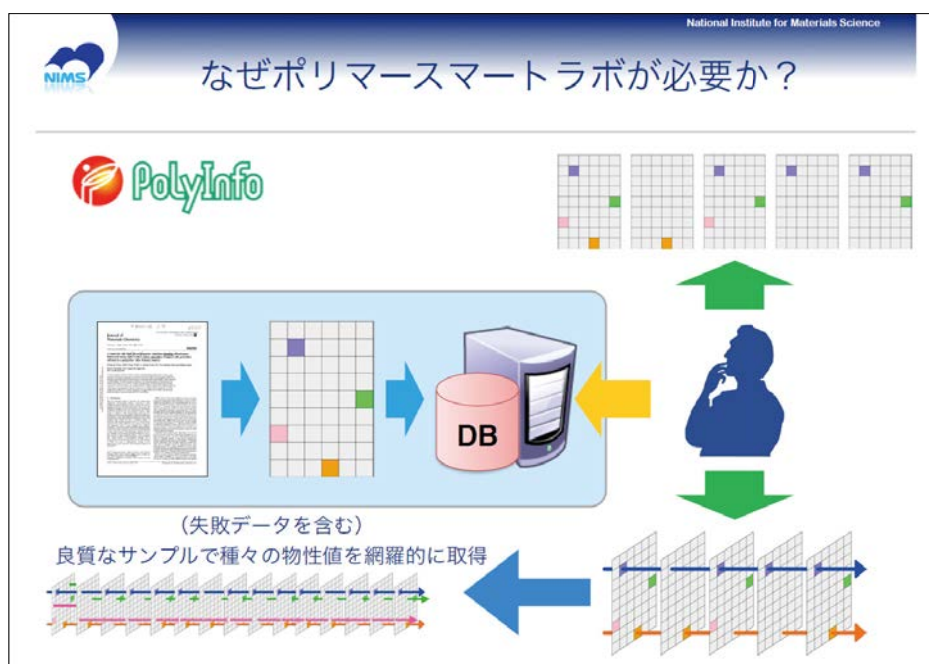


図 2-5-2 ポリマースマートラボの必要性

なることもある。

ポリマースマートラボの問題として、高分子材料では例えば力学測定やデバイスの評価をするときに、数g、数10gの規模で材料が必要になる点が挙げられる。そのため、スマートラボにも大きな装置、非常に高価な装置を使うことが必要になる。例えばヒューマノイド・ロボットやラボ用自動合成装置は市販されているが、1億円程度するような高価なものであり、なかなかみんなが使えるものではない。また、力学測定装置も数100～数1000万円と非常に高価なものであり、またサンプル作成が非常に難しい。

我々は「誰にでも手が届くスマートラボ」を目指している。ただし、必ずしも「誰にでも」にはなっていない

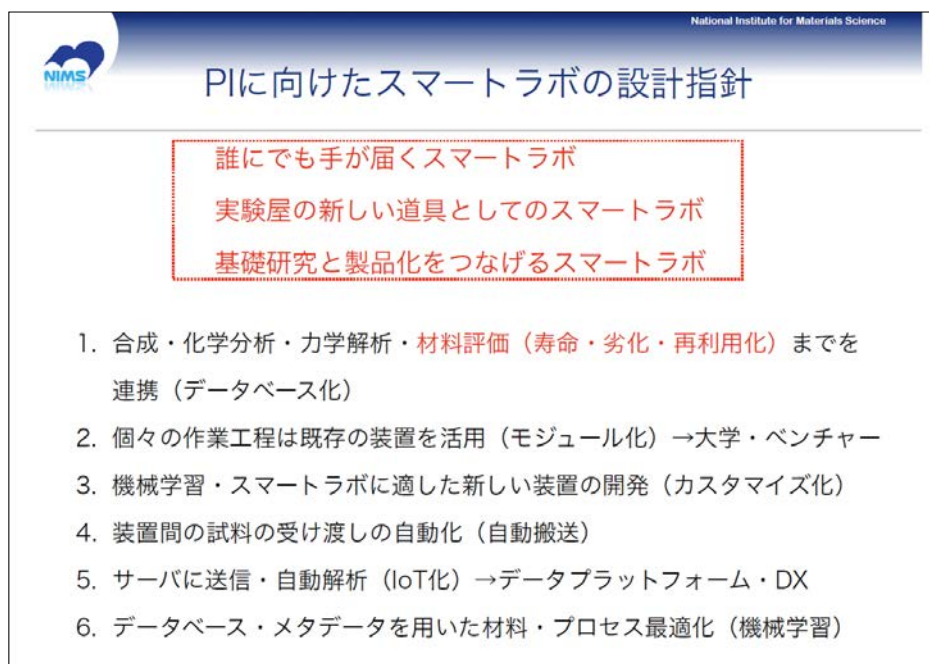


図 2-5-3 PIの向けたスマートラボの設計指針

くて、「ある程度の予算を頑張って取ってくれば」、という注釈付きではあるが。誰にでも手が届くようなスマートラボで、実験屋の新しい道具として使えるようなもの、また、基礎研究と製品化をつなぐことができるスマートラボをつくっていかうという指針でやっている（図 2-5-3）。具体的には、合成から化学分析、力学解析、材料評価としての寿命や劣化、リサイクル性までを連携させてデータベース化しようと考えている。使用する装置は、できる限り既存の装置を使い、それをモジュールと考えて連結する方法である。既存装置をスマートラボに適するようにカスタマイズしたり、機械学習に使いやすいような形にしてうまく活用することになる。装置間の受渡しを自動化する自動搬送は、実は非常に難しく、大切なノウハウである。得られたデータをサーバに自動送信して蓄積し、そのデータを基に材料とプロセスの最適化を行うという全体像である。

高分子合成は、図 2-5-4 に示すように、ステップが比較的シンプルである。原料を秤量し、加熱攪拌し、追加原料を投入する。反応終了後は、溶液のまま化学分析する場合と、溶媒除去・乾固後に力学物性や機械物性を測る場合がある。各ステップでデータを IoT 化して、データをデータプラットフォームに蓄積するような仕組みにしている。ここで得られるデータには、高分子の合成情報や、構造解析、熱物性、力学特性などの生データ、メタデータ、計算データが含まれる。

具体的には、自動分注装置を改良して合成装置として使用したり、多連で溶媒除去をできる市販装置などを活用している。各種物性評価では、オートサンプラーや自動搬送装置を活用して自動的にデータ収集ができるようにしている。これら各種装置は、USB タイプの無線 LAN 接続機器を使って、取得したデータが全て NIMS のデータサーバに蓄積し、解析できる仕組みになっている。この方法によって、年間で約 10,000 サンプル程度、精密合成され配列などが分かった高分子材料を、化学情報、熱物性、力学特性などと一緒に集めることが可能になる。

試料片作製や材料力学試験は、従来のやり方は煩雑で再現性が乏しいという問題点がある。我々は、図 2-5-5 に示すように、例えば引張試験はナノインデンテーションのように自動化可能な方法に置き換えて評価している。材料の引張試験のデータとナノインデンテーションで得られるヤング率との相関を取ることによって、異なるデータセットの拡張を行っている。このようにしてスモールデータを幾つか重ね合わせることによって、より機械学習しやすいものになっている。また、サンプル作成方法も、機械学習に適した方法を検討している。

高分子材料にプロセス・インフォマティクスを適用する際の難しさは、図 2-5-6 に示すように、高分子の構

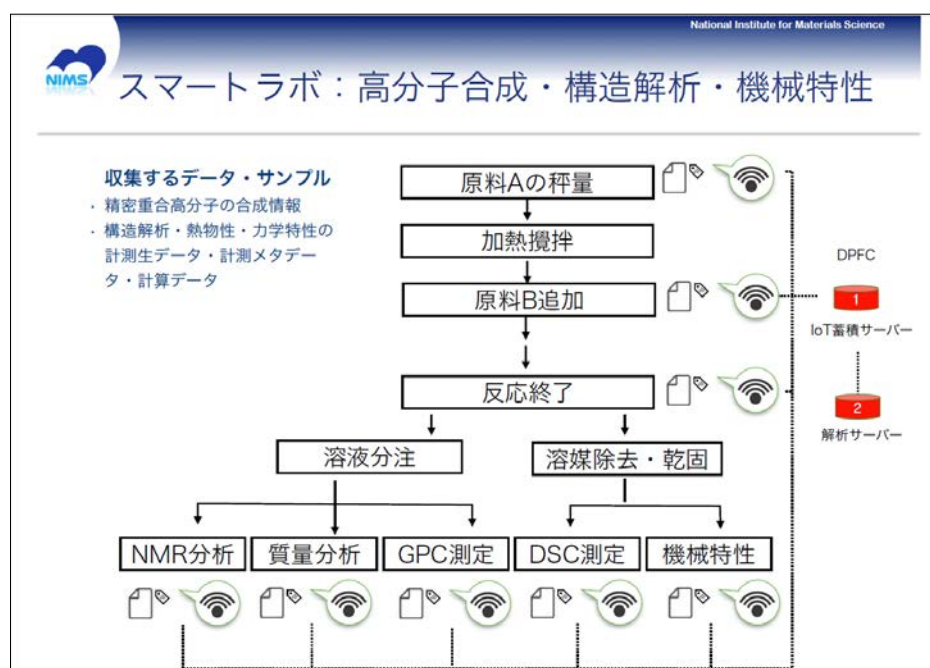


図 2-5-4 スマートラボ：高分子合成・構造解析・機械特性評価

造形成が非常に複雑であり、また、製造するときに企業のノウハウが含まれているところにある。通常、協業する場合は、秘密ではない材料を取り扱うことが多いが、秘密であっても協業できることが望ましく、次のような検討をしている。

精密保持の観点から、化学構造解析をすることは許されないの、我々は、直接イオン化質量分析（DART-MS）を使った材料評価を活用している。サンプルを前処理なしで大気圧下でプラズマジェットにかざすだけで熱分解が起き、フラグメントの分子量測定ができ、いろいろな材料の特徴を現すことができる。これをサンプルの熱分解と組み合わせると3次元にデータを取ることができ、大量のデータを一気に集めること

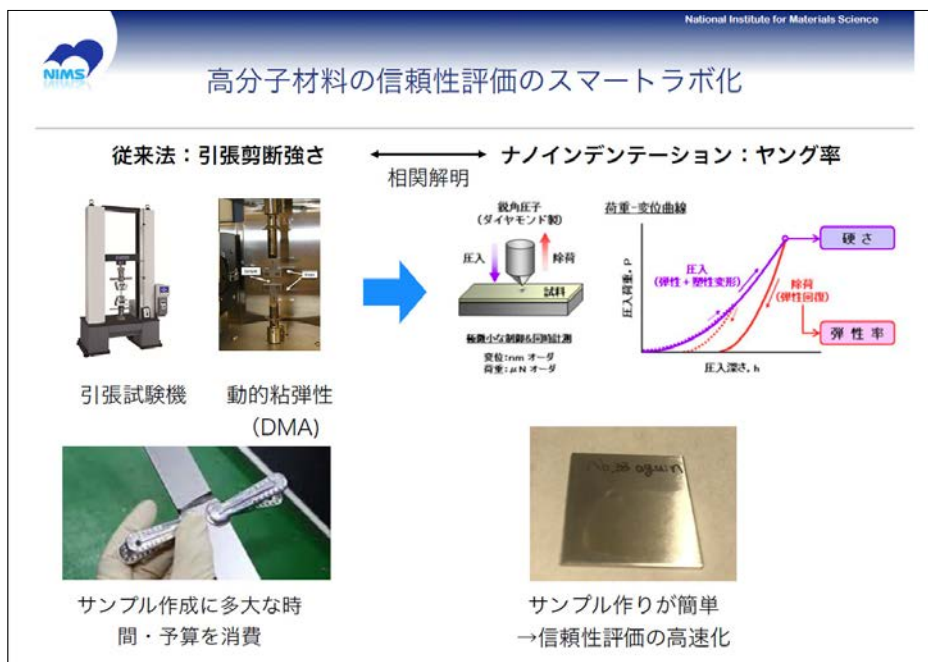


図 2-5-5 自動化可能な力学特性評価

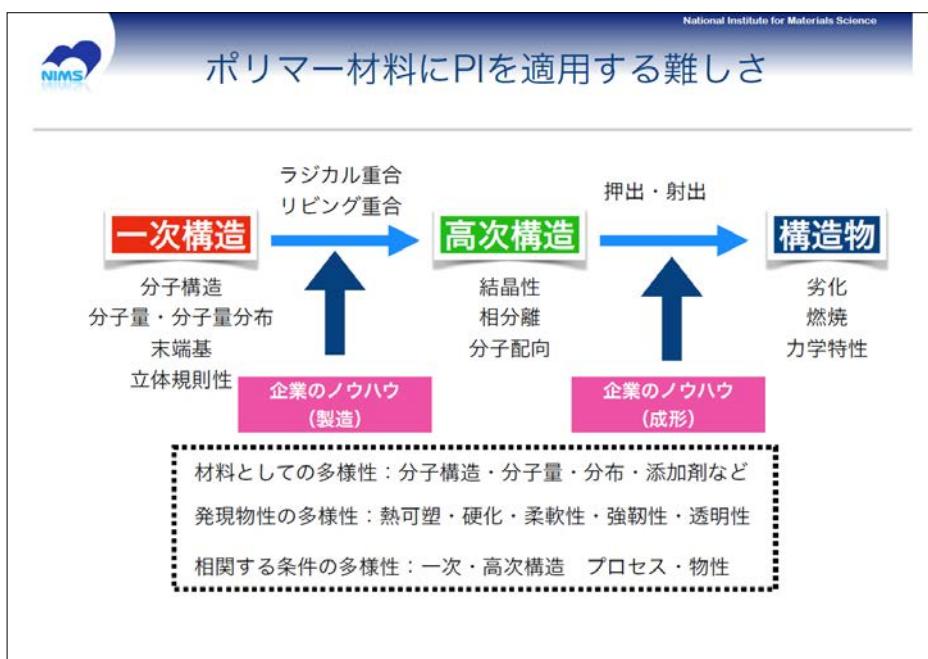


図 2-5-6 ポリマー材料にPIを適用する難しさ

ができる。

このDART-MSを接着剤の機械学習に使った例を図2-5-8に示す。MPaは接着強度の単位であり、3種類のサンプルのDART-MSの分析結果をそれぞれ示している。横軸は分子量、縦軸は温度に対応するものになっており、接着の強度に対応して何か違うパターンが見えている。機械学習を使って検討したところ、DART-MSのパターンと接着強度との相関を知ることができた。この技術は、不純物の検出など微量分析に使うこともできるので、品質管理や製品の認証等の技術に応用していくことも考えている。

データ収集のIoT化も進めている。これまでNIMSではデータをUSBで持ち運んでいたが、今では全てネットワークに組み込むことによりデータサーバに蓄積しそれを解析することができるようになっている。今まで1つのデータを手にするまでに半日以上かかっていたものが、2分程度で解析結果が手元に届くような状態になっている。また、定型的なデータ構造とすることで、集めたデータセットはすぐに機械学習に利用できるようになっている。

このように、我々が考えている高分子合成、構造解析、機械特性のスマートラボは、合成から力学特性まで一連の流れの中でデータを蓄積しながらやっていくものであるが、今後はデータを外部と連携しながら拡張していくことも考えている。

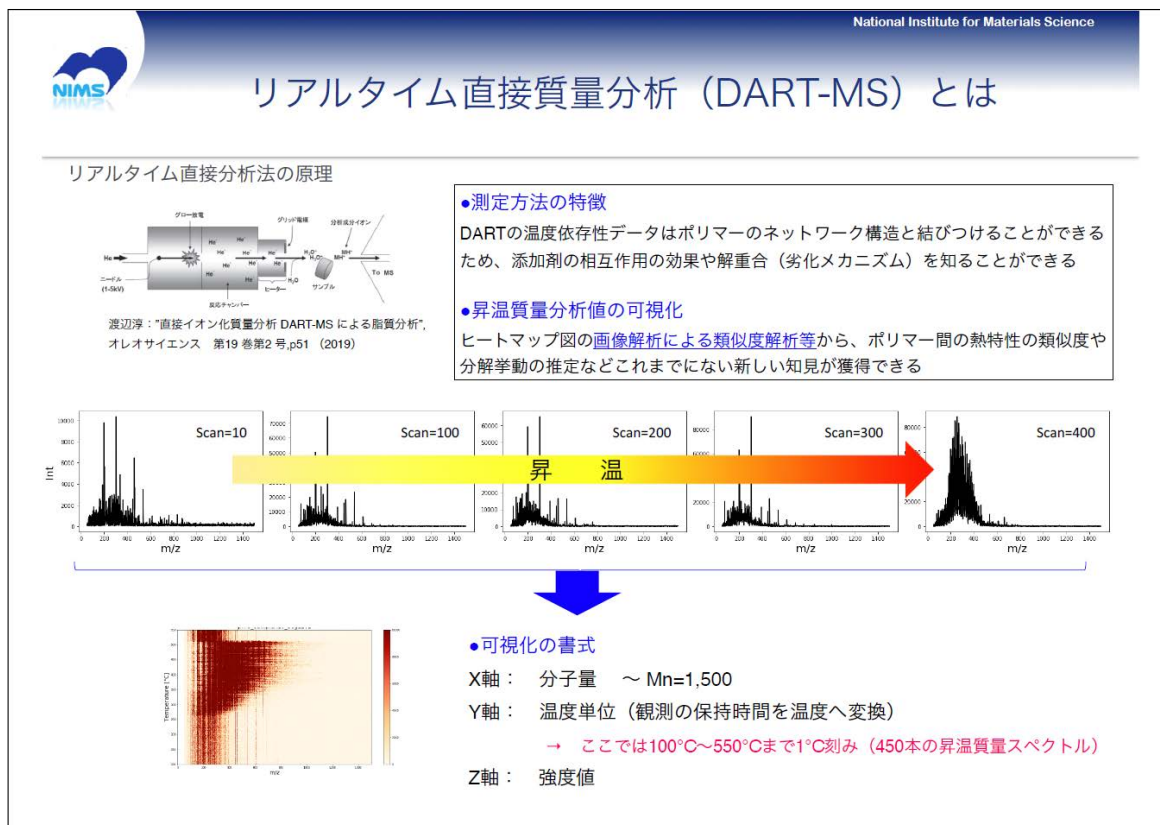


図 2-5-7 リアルタイム直接質量分析 (DART-MS)

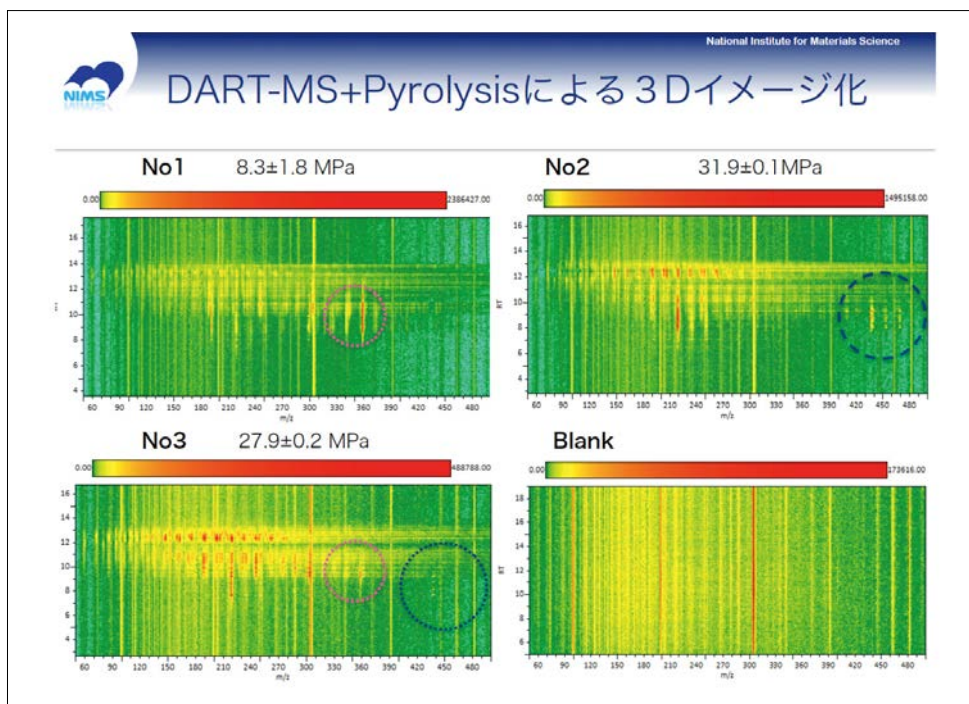


図 2-5-8 接着剤の DART-MS + Pyrolysis 分析

【質疑応答】

Q：高分子材料では高次構造が非常に重要だが、モルフォロジーと物性の相関は検討しているか。

A：我々は、微細構造に注目して特徴量を求めるというやり方はしていない。構造の詳細が分かる方法では、例えば共同研究をやるときにサンプルを出してもらいにくくなるなど支障があるので、材料自体は何が入っていてもよいというスタンスで、DART-MSを使ってデータ解析をしている。通常、粉体や熱硬化性樹脂等は構造解析が非常に難しく、固体 NMR や X 線構造解析が必要になるが、DART-MS は簡便にデータを取れるので、まずこれを使っているという意味もある。

Q：接着剤メーカーから見るとハイスピード評価は非常に魅力的だと思う。しかし、接着剤をどういうプロセスで使うかという情報も重要であり、ユーザーサイドとのコラボレーションも大切ではないか。

A：我々にはできるのは、接着の自動化装置を作ってデータを集めるところである。接着剤メーカーは、ユーザーに材料を渡して OK / NG の結果だけを聞いていることが多いのに対し、我々の方法はいわゆる中間データが得られるので有意義である。また、基礎データをたくさん持っていることによって、ユーザーサイドとの共同研究にも繋がっている。

C：様々な評価を組み合わせる際に自動搬送が意外に難しいということは同意見である。サンプルホルダーや搬送系を標準化してシームレスにつなぐ必要性を感じている。

3 | プロセス・インフォマティクスに必要な技術群

3.1 プロセス・インフォマティクスにおける技術紹介および課題と将来展望

金子 弘昌 (明治大学)

私は、インフォマティクス、つまりデータ解析や機械学習などを武器にした研究を進めている。主にこのケモインフォマティクス、マテリアルズ・インフォマティクス、プロセス・インフォマティクスの研究を推進しており、大学やいろいろな企業との共同研究も進めている。

ケモインフォマティクスで対象にするのは、主に分子、物性や薬理活性が分かっている化合物である。図3-1-1に示すように、実験しないと分からない物性・活性を y と、構造さえ分かれば得られる情報を構造記述子 X として、この X と y の間で推定モデルを構築する。 X は構造から計算できるので、コンピュータ上で構造を生成、 y を予測し、予測値が良好なものを選んで合成する。合成した結果をフィードバックして、モデルを改善することを繰り返す。いわゆる能動学習も可能である。

マテリアルズ・インフォマティクスでは、主に結晶構造等を対象とする。本ワークショップのテーマであるプロセス・インフォマティクスも一部含んだ概念で考えている。結晶構造や実験・製造条件を X として物性・活性 y との間でモデルを構築して、物性を予測する。予測結果がよければ、例えば結晶構造であれば、もう一回シミュレーションを回してみるとか、実験・製造条件であれば、実際に合成してみて狙った材料が作れるかどうか検証してみて、フィードバックしながらモデルの改善を繰り返す。

一例として、理化学研究所との共同研究している機械学習による触媒設計を紹介する。この反応は、Buchwald-Hartwigクロスカップリング反応（A）と鈴木・宮浦クロスカップリング反応（B）が協奏的に起こる系である。基質と触媒をそれぞれ変化させたときに、A反応とB反応の比率がどうなるかを予測しよう

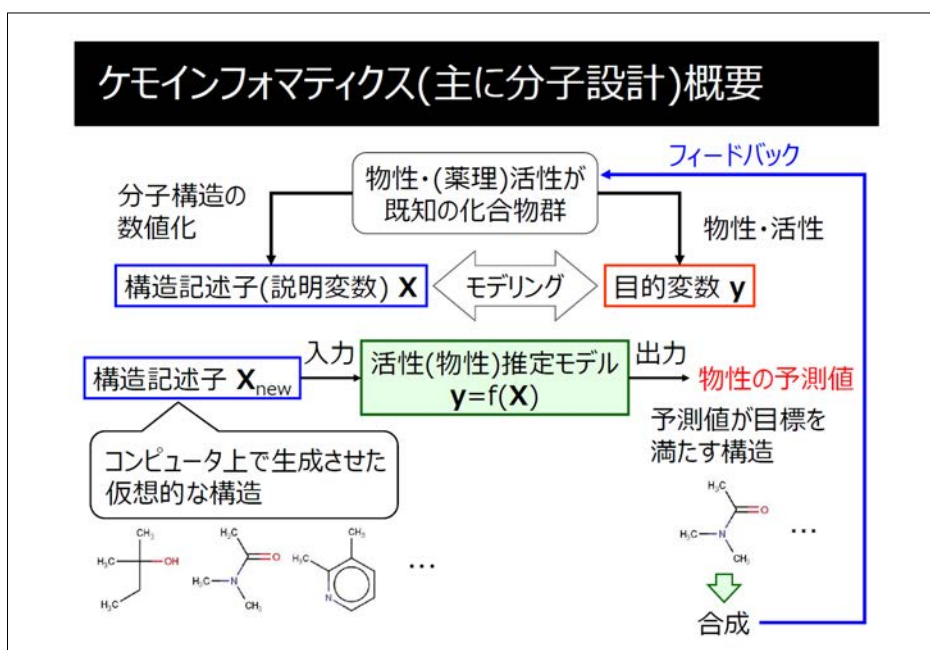


図 3-1-1 ケモインフォマティクスの概要

としている。実験データだけでは十分な数がないので、論文データも追加して転移学習をした結果、良好な推定モデルを構築することができた。このモデルを活用して、A反応の収率が高くなる触媒とB反応の収率が高くなる触媒をそれぞれ設計したところ、どちらも推定値と実験値がかなり合う結果となった。

図 3-1-3 ではプロセス・インフォマティクスの中でも主に推定制御を対象にして説明する。ここでは、いわゆる時系列データを扱う。容易に測定可能なプロセス変数 X と測定が困難なプロセス変数 y との間で推定モデルを構築する。温度や圧力などの容易に測定可能な変数、これらはリアルタイムに測定できるので、これを X として入力することで測定困難な変数 y を予測することができる。 y をあたかも実測値のように扱って運転制

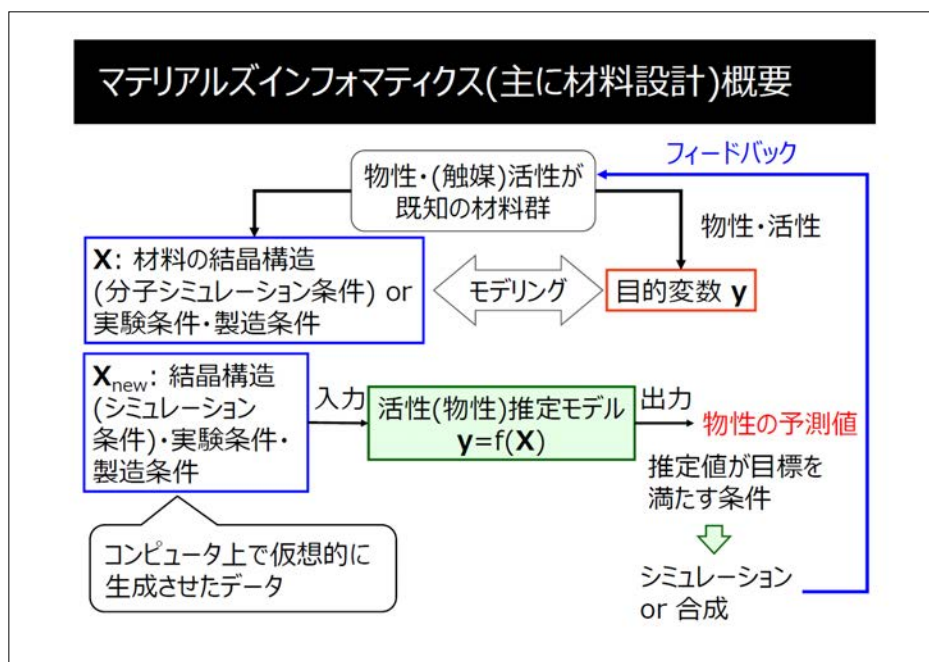


図 3-1-2 マテリアルズ・インフォマティクスの概要

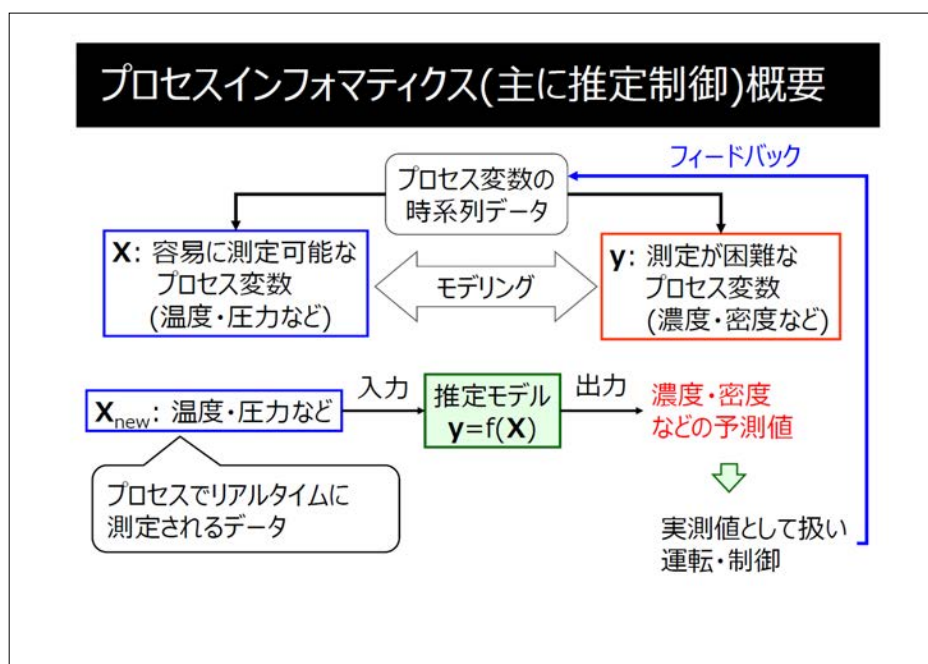


図 3-1-3 プロセス・インフォマティクス (推定制御) の概要

御することで、効率的なプロセス管理が達成できる。

プロセス・インフォマティクスの一例として、あるポリマー重合プラントの例を紹介する。ここでは、1つの重合プラントでさまざまな製品ポリマーを製造しているので、製品の切替え時に、目標物性の製品が製造できているかをなるべく早く判定する必要がある。そのために、密度やMelt Flow Rate (MFR) がある範囲に入っているかが重要になる。図 3-1-4は密度の時間変化を示しているが、サンプリングして密度計で実測する方法では密度が分かるのが数時間後である。つまり、この方法では銘柄切替え後に狙った物性になったことを確かめるのに数時間かかることになり、その間、製品が無駄になってしまう。ここに推定制御モデル（ソフトセンサー）を用いると、図 3-1-4の赤線のような推定値をリアルタイムに得ることができるので、ある時刻に目標範囲に入ったことが即座に分かる。また、運転途中で目標範囲を逸脱してしまうようなケースでも遅滞なく分かるので、迅速な制御が可能である。

材料合成におけるプロセス・インフォマティクスにおいては、目標の材料特性を持つ材料を作るために温度分布とその時間変化をどうするべきかという課題がよくある。これは、モデルの逆解析である。

温度変化のデータXと作られた材料の特性や結晶構造Yの間で推定モデル $Y=f(X)$ を構築し、このモデルを使って、狙った材料特性や結晶構造を入力することで、温度分布の時間変化をどうするべきかを解くのがモデルの逆解析である。基本的にはモデルの順解析、つまりXの値を入力してYの値を取得することが行われるが、その逆を解こうという話である。

ただし、一般的なモデルの逆解析で実際に行われているのは順解析を繰り返すことである（図 3-1-5）。つまり、Xの上限、下限などの制限を設定し、コンピュータ上で大量の仮想的なサンプルを生成する。これを全て計算することでyの値を予測して、その結果から、いいものを選ぶというアプローチである。つまり、Xを入力してYを予測するという順解析を繰り返すにすぎず、擬似的なモデル逆解析である。ベイズ最適化も同じ流れで、予測値だけでなく予測の分散を考慮して獲得関数を計算するということであり、順解析を繰り返すというのは同じことである。

最初にサンプルを生成する際に、人が予め想定したXの探索範囲でしかYを予測しないという点が問題であり、既存のYの値を超越する最適化をするには不十分である。また、やはりXの組合せが多くなればなるほど網羅的に大量のサンプルを生成しYを予測する必要があるので、短時間でXを提案はすることは難しい。

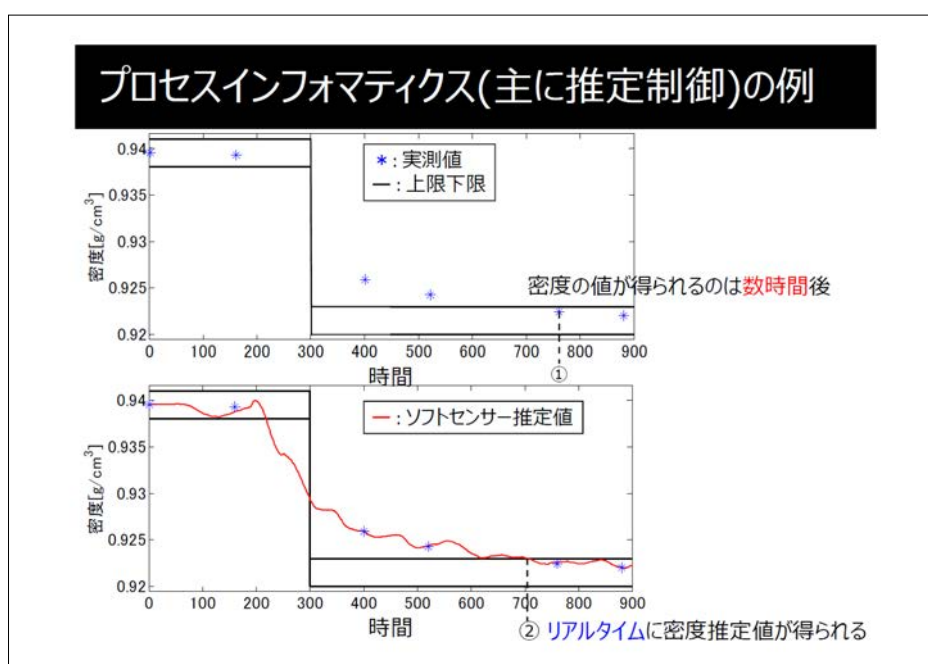


図 3-1-4 プロセス・インフォマティクス（推定制御）の一例

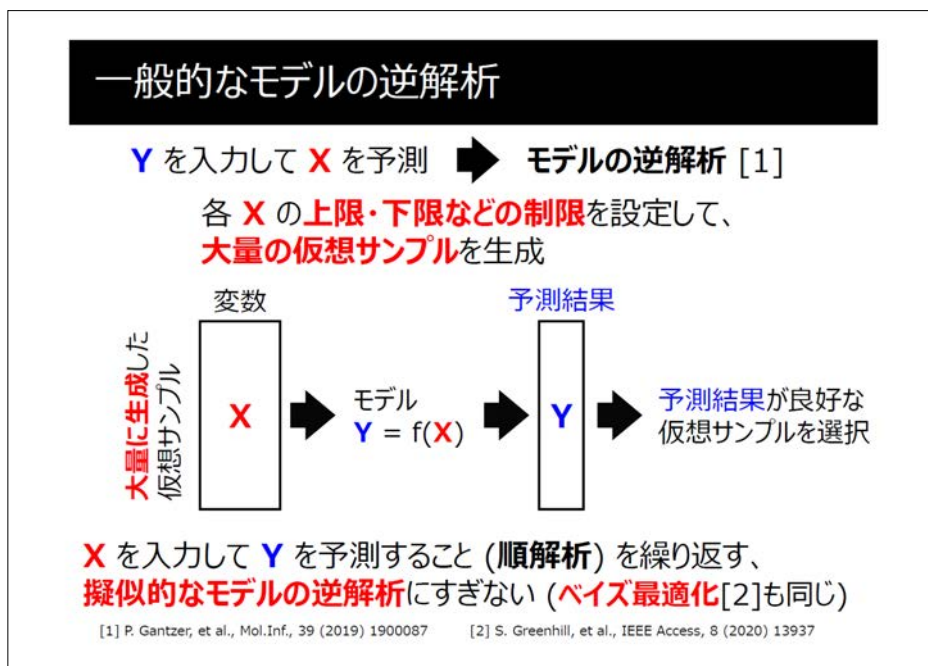


図 3-1-5 一般的なモデルの逆解析

その中で、我々は直接的モデル逆解析法（Gaussian Mixture Regression：GMR、図 3-1-6）を提案している。これはまさにYの値からXの値を予測する方法である。ケモインフォマティクスの例でいうと、物性Yと特徴量Xとの間で、確率密度で表現されるようなモデルを構築する。このモデルであれば、Yの値を入力するということは、確率密度分布をY方向に輪切りをすることであり、山の断面を見るようにXの確率密度で予測ができる。この方法は、Xの上限、下限を設定したり、大量のサンプルを生成したりする必要がなく、Yの物性を達成するXを予測するためのカギといえる。

この方法で適応的実験計画、つまり能動学習をやったところ、ベイズ最適化と比べて回数が1/3に削減さ

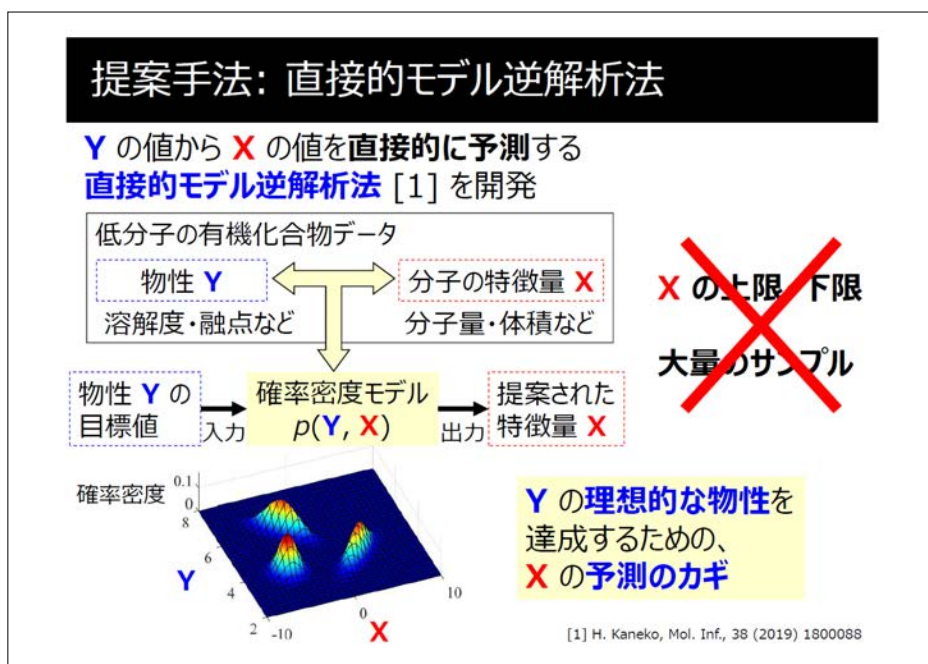


図 3-1-6 直接的モデル逆解析法の提案

れた例もある。図 3-1-7 の右下は、実際の実験回数を横軸に取り、Y の値を縦軸に取った例であるが、赤で示すベイズ最適化では Y の値があまり向上しなかったケースでも、直接的逆解析（GMR）を使うこと一気に向上するような結果も得られた。

最後に、プロセス・インフォマティクスの課題を整理する。

まず初めに、材料特性などの目的変数に関係のあるパラメータで、かつ測定可能、かつ制御可能なものを決める必要がある。結晶成長では、たとえば温度である。温度はサーモグラフィ等で測定できるが、選ぶパラメータは測定にくいものである可能性もあり、その場合には測定インフォマティクスと共同する必要がある。また、パラメータの時間変化や空間分布を測定する必要がある場合もあり、この場合も計測インフォマティクスと一緒にやる必要がある。

その上で目的変数とパラメータの関係をモデル化し、 $Y = f(X)$ の関係を得る。モデルを使って、材料特性等の目的変数が理想的な値となるためのパラメータの時間変化、分布を提案するために GMR のようなモデルの逆解析をすることになる。

提案された時間変化、分布となるようにプロセスを制御したり、必要に応じて新たなプロセス、装置を提案したりするためには、人の知見が必要になると考える。また、インフォマティクスでそれをやるためには、プロセスの抽象化が必要になる。

今後の展望としては、ケモインフォマティクス、マテリアルズ・インフォマティクスの融合が大切になると考える。プロセス・インフォマティクスでは、合成プロセスの設計を目指している。合成プロセスを個別に設計するだけでなく、図 3-1-8 に示すように材料特性を満たすような分子・材料の設計図と合成プロセスを同時に最適化するようなことも狙うのがよいのではないかと考える。

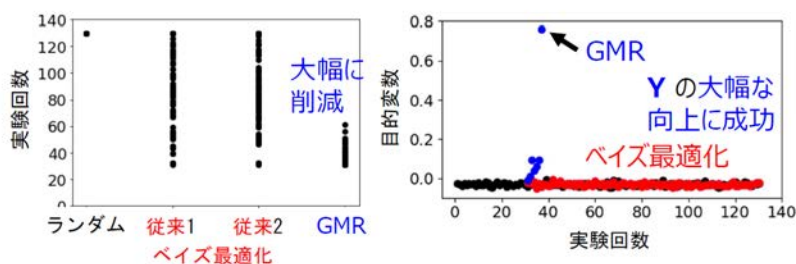
3

プロセス・インフォマティクスに必要
な技術群

直接的モデル逆解析による適応的実験計画

提案した**直接的モデル逆解析法 (GMR)** により **Y** の目標達成までにかかる実験回数を従来法より約 **1 / 3** に削減 [1]

複雑な関数で **X** (15 パラメタ) から **Y** を計算することを実験と見立て、目標達成までにかかる実験回数を**ベイズ最適化**と**GMR**で比較



[1] H. Kaneko, Chemom. Intell. Lab. Syst., 208, 104226, 2021.

図 3-1-7 直接的モデル逆解析による適応的実験計画の例

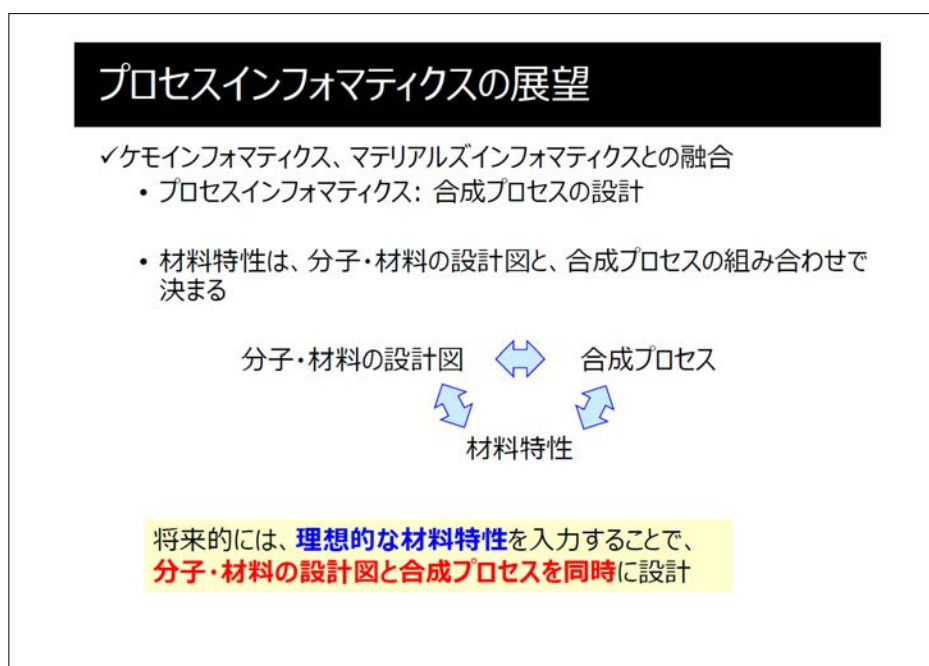


図 3-1-8 プロセス・インフォマティクスの展望

【質疑応答】

Q：プロセスの抽象化ができないと、プロセス条件の最適化はできても、インフォマティクスでプロセスそのものの最適化をすることは難しい。プロセスの抽象化として具体的にどうすればいいか。

A：たとえば温度変化は1つの抽象化になると考える。温度変化のデータセットとして、結晶成長も一つのサンプル、ポリマー重合も一つのサンプル、合金の製造も一つのサンプルというような形でサンプルとして扱えるのではないかと考える。それをうまくのデータセットとして扱って、今までに全くないような新しい装置の提案をすることができるのでないかと考える。理想的な温度プロファイルが決まったが、そのプロファイルを満たすことが今の装置ではできないときに、では、どういう装置であれば達成できるかということ、いろいろな装置のデータを活用することで新たな装置を提案できるのではないかと考えた。抽象化としては、物性や圧力、流量も可能性があると思う。

Q：抽象化には可能性を感じた。溶融プロセス一つでも、各研究機関でやり方や装置が違うので、抽象化するとよいというのはよく分かるが、全く違うプロセスに対しても、抽象化法は使えるか。

A：使えるようになるまで抽象化するというイメージである。例えば、分子は、電子密度分布で表現すれば、その原子が何かを考えることなく分子を考えることができる。装置もそこまで抽象化できれば、インフォマティクスで扱えるところに落とし込めるという展望である。

Q：材料設計と合成プロセスの設計を同時に進めなければいけないと、最後にコメントされた。例えば、どういうものを作るかというところで最適化し、その後でプロセスを最適化する場合と、2つ同時に最適化する場合で、最適の場所が変わるので、両方同時に最適化しないと本当にいいものはできないということか。

A：そのように思う。もう一つの面では、材料の設計図のデータセット、合成プロセスのデータセットが個別に解析されている状況だとすると、その個別のデータ範囲でしか最適化できないので、そこを一緒に扱うことでより広い範囲での最適化を達成できると考える。

3.2 窒化物半導体MOVPEプロセスのマルチフィジックス結晶成長シミュレーション

寒川 義裕（九州大学）

ここでは、窒化物半導体の有機金属気相成長（Metal-Organic Vapor Phase Epitaxy：MOVPE）プロセスを例に、マルチフィジックス結晶成長シミュレーションをベースとしたプロセス・インフォマティクス（PI）の試みを紹介する。

図 3-2-1 にPIの分類を示す。PIでは一般に、材料プロセス実験を行い、取得された材料物性（目的変数）を最適化するためにプロセス条件（説明変数）を探索すること〔CASE 1〕となる。例えば、ベイズ最適化や実験計画法を用いた機械学習による研究がある。詳しくは金子弘昌・准教授（明治大学）の講演（3.1 節）を参照されたい。また、材料プロセスを模擬するプロセス・シミュレーションの精度向上や計算コスト削減を目的とした機械学習〔CASE 2〕、およびその延長線上にあるデジタルツインを活用したPI〔CASE 3〕がある。ここでデジタルツインとは、現実空間（フィジカル空間）にある装置、設備などを仮想空間（サイバー空間）に再現することを表す造語である。デジタルツインの環境を活用することで、現実空間の材料プロセスを予測・探索することが可能となる。以下では、現実空間の材料プロセスを仮想空間において再現する「マルチ

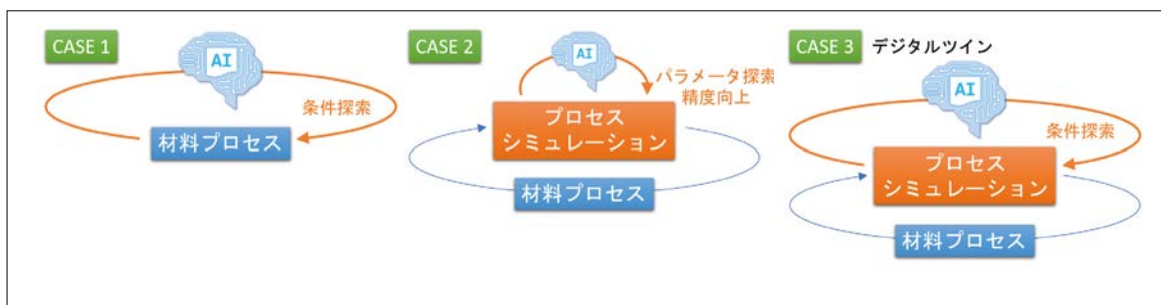


図 3-2-1 プロセス・インフォマティクスの分類

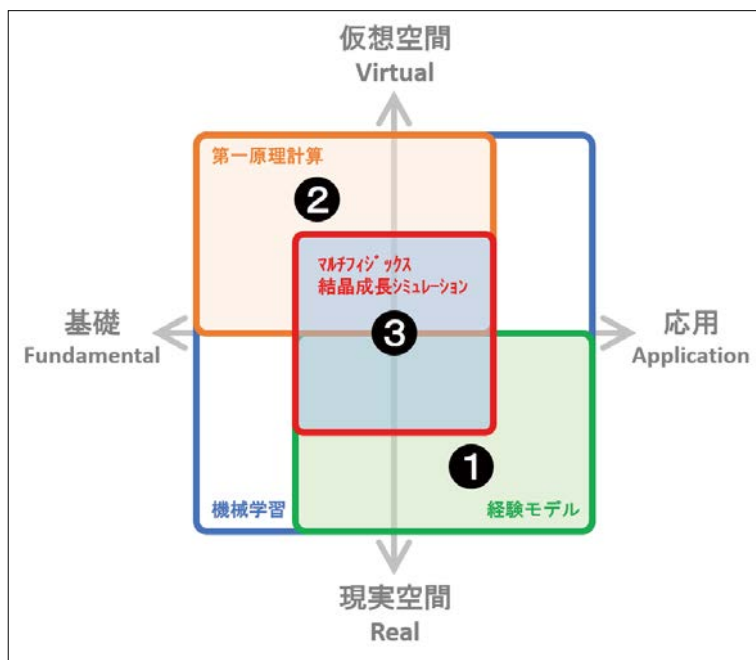


図 3-2-2 プロセス・シミュレーションの分類

フィジックス結晶成長シミュレーション」について解説する。

図 3-2-2 にプロセス・シミュレーションの分類を示す。①経験モデルは、実験により得られたフィッティング（経験）パラメータを用いた物理モデルである。実際の実験結果をベースとしているため現実空間の物理・化学現象を再現することが可能である。一方で、パラメータ構築に用いた実験条件の範囲外の現象を予測することは困難である。②第一原理計算は、経験的パラメータを用いないため未知のプロセス結果を予測することができる。しかし、絶対零度の系を模擬しているため現実空間におけるプロセス結果との乖離が生じる。第一原理分子動力学（Molecular Dynamics：MD）により有限温度を考慮できるが、計算コストの制約によりモデルの原子数または計算時間（マシンタイムではなくシミュレーションの時間発展）が限られる。最近では、第一原理計算に基づく機械学習ポテンシャルを用いたMDシミュレーションにより計算コストを削減し、より現実的な系の解析が試みられている（図 3-2-1 [CASE 2]）。例えば：W. Li et al., *J. Chem. Phys.* **153**, 164119, 2020）。③マルチフィジックス結晶成長シミュレーションでは、量子力学に基づく第一原理計算と統計熱力学との融合により、有限ガス分圧（ p ）、有限温度（ T ）の寄与を解析可能な物理モデルを構築している。現実の系を模擬できる物理モデルを構築することによりデジタルツインの環境を整備することができる（図 3-2-1 [CASE 3]）。ここでは、原子スケールの物理・化学現象を解析可能な結晶成長シミュレーションを活用して材料プロセスの予測・探索を行うことになるので、ホワイトボックス型AI（説明可能なAI（Explainable AI=XAI））を用いることで解釈可能な探索結果を得ることができる。実験室レベルのケーススタディであればブラックボックス型AIが優位であるが、将来的な産業化・大型化を目指すのであればホワイトボックス型AIによる物理的根拠に基づく探索結果の導出が有効となる。ここで得られる知見は、現在内閣府で検討が進むマテリアル革新力強化の観点からも極めて有用である。以下では、窒化ガリウム（GaN）のMOVPE成長における炭素不純物混入の解析を対象としたマルチフィジックス結晶成長シミュレーションの研究事例を紹介する。

文部科学省「省エネルギー社会の実現に資する次世代半導体研究開発」天野浩（名古屋大学）では、電気自動車用途として要求される1kV耐圧のGaNパワーデバイスの開発に取り組んでおり、仕様要件をクリアするためにはチャンネル層のキャリア密度を 10^{16}cm^{-3} 以下に抑制する必要がある。これはMOVPE成長中に意図せず混入する不純物の濃度を 10^{16}cm^{-3} 以下に抑制する必要があることを意味する。図 3-2-3 に GaN MOVPE プロセスの概略図を示す。一般に、Ga 原料としてトリメチルガリウム（TMG： $\text{Ga}(\text{CH}_3)_3$ ）が使用されており、この原料ガスがメチル基を含んでいるため、メチル基による汚染、すなわち炭素不純物の混入

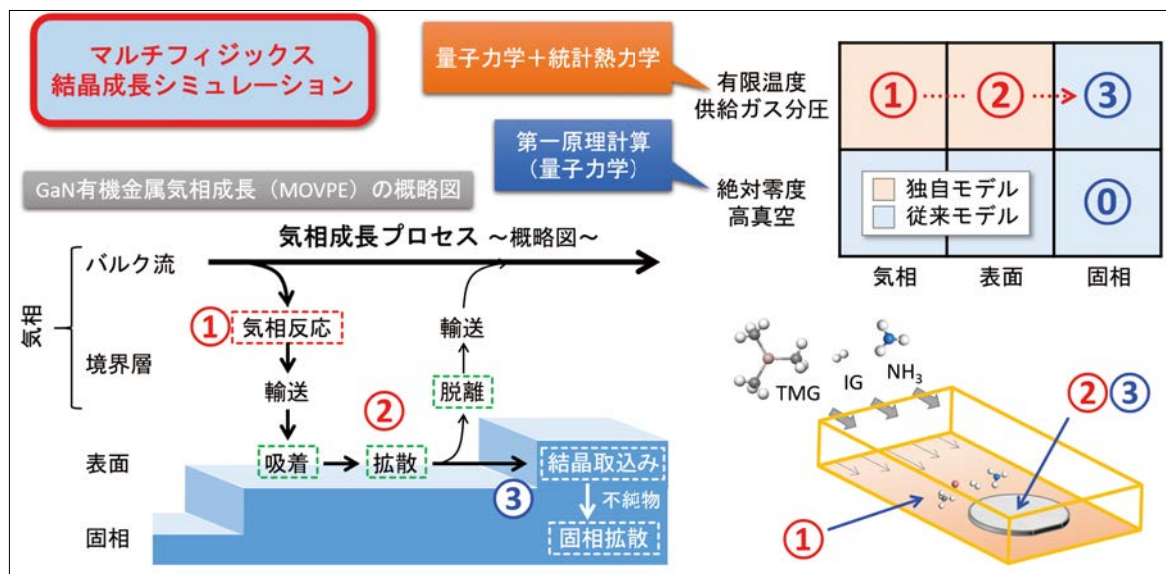


図 3-2-3 GaN MOVPE プロセスの概略図

が起こる。この研究課題における説明変数は成長条件（III族ガス分圧（ p_{III} ）、V族（N）ガス分圧（ p_V ）、成長温度（ T ）など）であり、最適化すべき目的変数は残留炭素濃度となる。すなわち、デジタルツインの構築に向けて、成長条件と残留炭素濃度の相関を解析するプロセス・シミュレーションが必要となる。ここでは、 p 、 T と残留炭素濃度の相関を解析する③マルチフィジックス結晶成長シミュレーションの開発を行った。

図 3-2-3に示すように、GaN MOVPEプロセスでは①気相反応、②表面反応、③結晶への混入および固相拡散により炭素不純物が成長薄膜中に蓄積される。すなわち、上流①から下流③までの結晶成長素過程を統合したプロセス・シミュレーションが必要となる。従来の第一原理計算では絶対零度における系を解析対象としていたが、独自モデルでは量子力学と統計熱力学に立脚した新たな理論解析手法により有限ガス分圧（ p ）、有限温度（ T ）における系の解析を可能としている。図 3-2-4に本理論解析手法のコンセプトを示す。まず、第一原理計算により原子・分子の表面吸着エネルギー（ E_{ad} ）を求める。この値と気相中の原子・分子の自由エネルギー（化学ポテンシャル（ μ ））の大小関係を求め、吸着/脱離の優位性を判断する。ここで、化学ポテンシャル（ μ ）は並進、振動、回転の分配関数を用いて統計熱力学により求めることができる。図

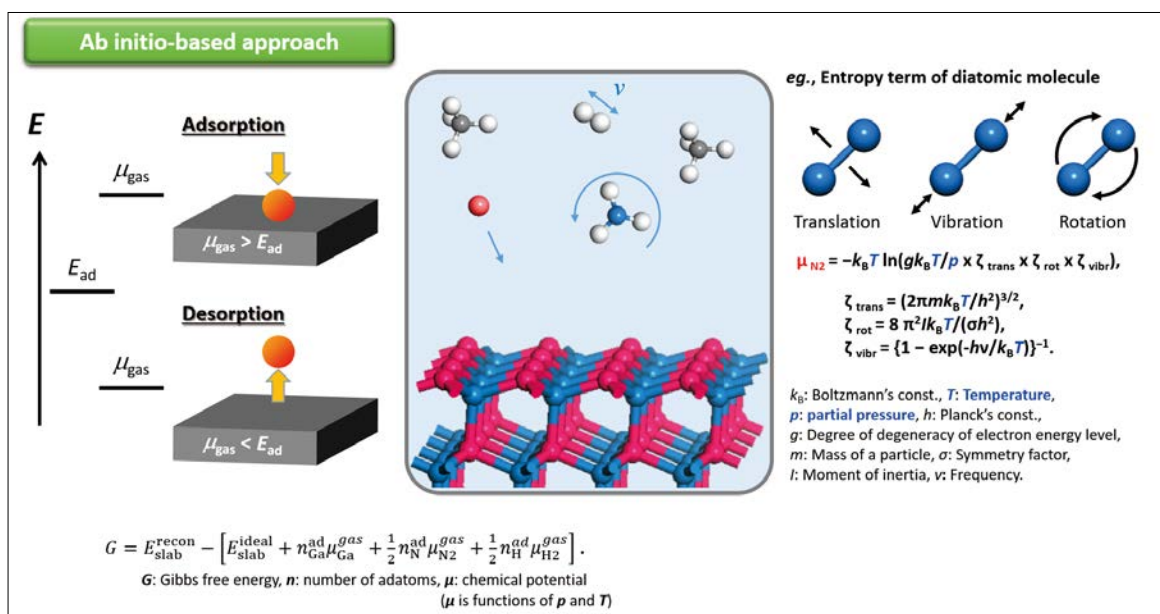


図 3-2-4 量子力学と統計熱力学に立脚した理論解析手法のコンセプト。

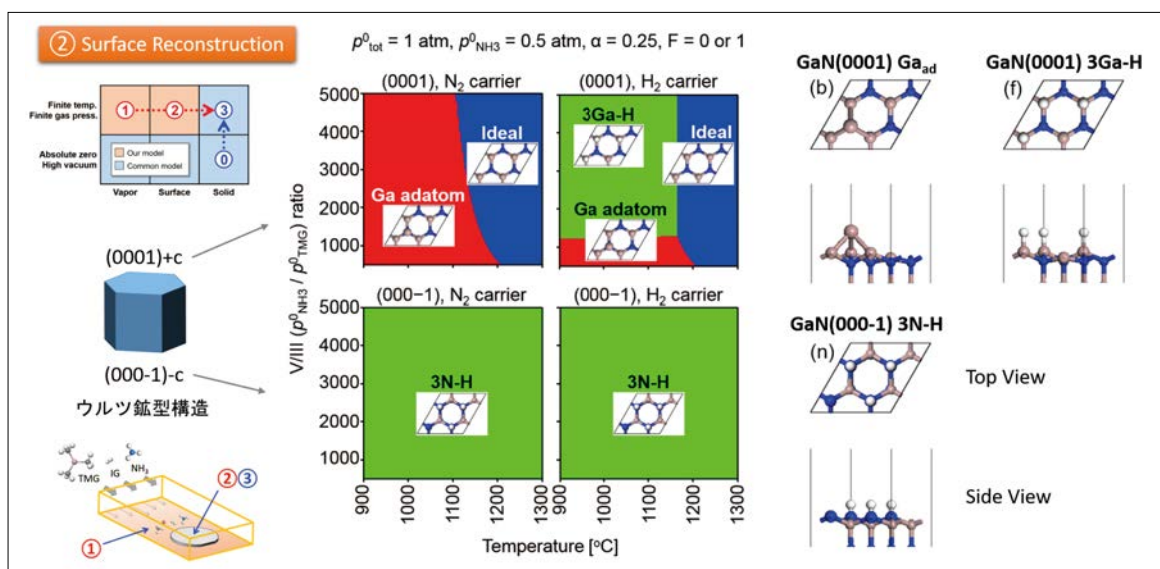


図 3-2-5 GaN MOVPE 成長表面の $p - T$ 相図

中の μ_{N_2} の表式を見ると、この表式は p 、 T のみの関数となっていることがわかる。すなわち、吸着/脱離の優位性を $p-T$ 相図によって表すことができる。

一例として、図 3-2-5 に GaN MOVPE における成長表面の $p-T$ 相図を示す。H₂ キャリアガスを用いた一般的な成長条件下では、GaN (0001) 面において H 被覆率 3/4 の 3Ga-H 表面が、GaN (000-1) 面において H 被覆率 3/4 の 3N-H 表面が出現することが予測されている。これらの再構成表面における不純物原子の吸着、固体内への拡散をシミュレートする統合モデルを構築することになる。

図 3-2-6 に統合モデルと一連の解析事例を示す。まず熱力学解析により MOVPE 炉内のガス組成を解析した①。ここでは、温度上昇に伴い TMG が炉内の水素、アンモニアと反応してメチル基が解離し、モノメチルガリウム (MMG) と GaH が成長表面に供給されることがわかった。図 3-2-5 に示す再構成表面上への MMG と GaH の吸着エネルギーを求め、ボルツマン因子により吸着確率 (C_s) を求める。 p_i 、 $C_{s,i}$ (i : MMG, GaH) を用いて MMG と GaH の被覆率比から表面炭素濃度 ($\rho_{C, surface}$) を求め、これを境界条件としてモンテカルロ (MC) シミュレーションにより固相拡散を解析する。MC シミュレーションでは、第一原理計算により求めた N 置換炭素 (C_N) 形成エネルギーの深さ方向プロファイルパラメータとして用いた。以上により、①気相反応、②表面反応、③固相拡散を統合したマルチフィジックス結晶成長シミュレーションを構築することができた。

図 3-2-7 に成長条件 (説明変数) と残留炭素濃度 (目的変数) の相関を示す。図から成長温度、V/III 比 (p_V/p_{III} 比) を上昇/増加させると残留炭素濃度が減少することがわかる。また、(0001) 面の方が (000-1) 面に比べて炭素取込量が多いことがわかる。これらの解析結果は実験結果と一致する (図 3-2-8)。また、実験結果からは、説明変数 (入力: 成長条件) と目的変数 (出力: 残留炭素濃度) の相関しかわからないが、統合モデルによるマルチフィジックス結晶成長シミュレーションにより、各結晶成長素過程 (①気相反応、②表面反応、③固相拡散) がそれぞれ出力にどのように寄与しているかを読み解くことができる。すなわち、このデジタルツイン環境を活用してホワイトボックス型 AI による条件探索を行うことにより、物理的根拠を解釈

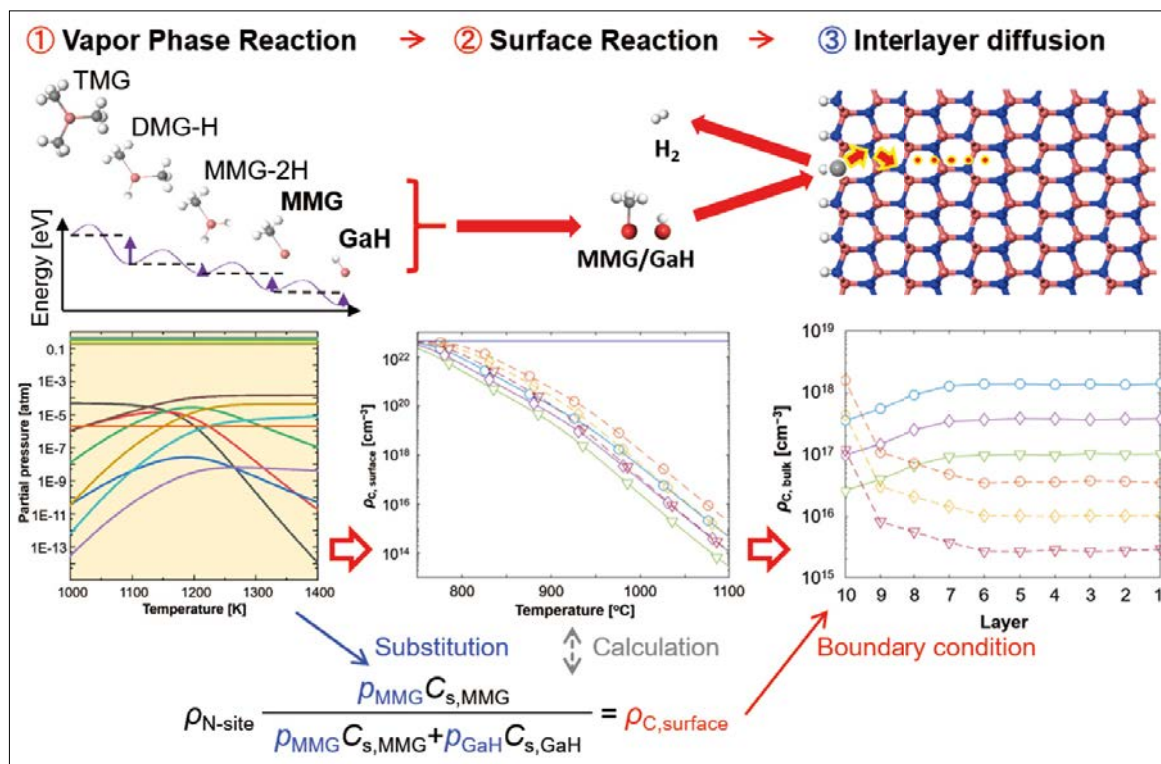


図 3-2-6 統合モデルによる解析事例

可能な探索結果を得ることができる。

現在、今回紹介したマルチフィジックス結晶成長シミュレーションを展開させ、富岳プロジェクトにおいて、ステップを含むような大規模系で表面拡散をも考慮した解析を進めている。これによりデジタルツイン環境の拡充を図っている。GaN MOVPE を例にプロセス・インフォマティクス技術確立し、その有用性を示すことで、他材料への発展・展開が期待される。

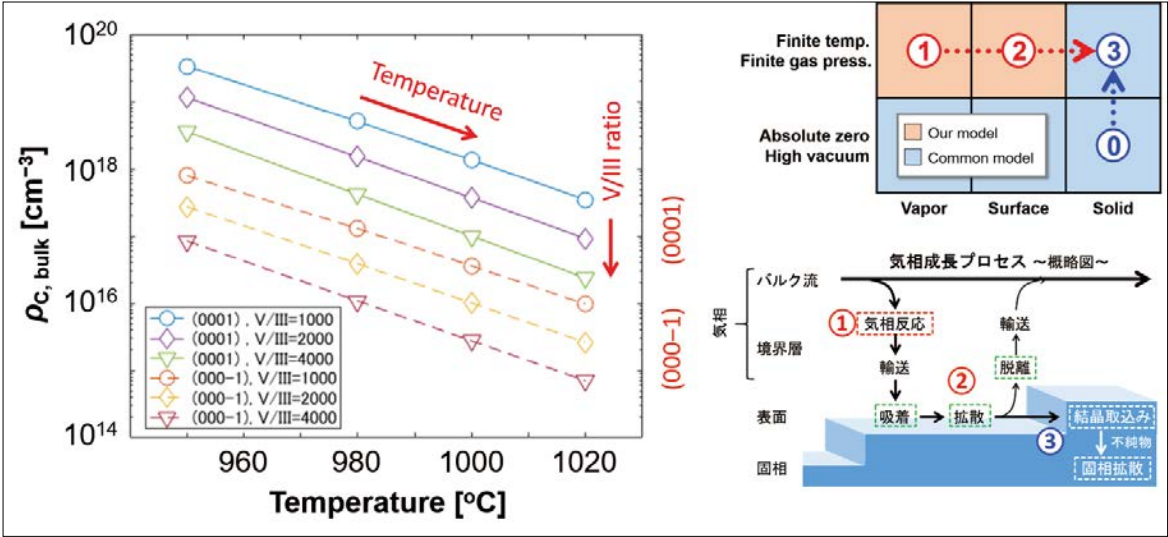


図 3-2-7 成長条件（説明変数）と残留炭素濃度（目的変数）の相関

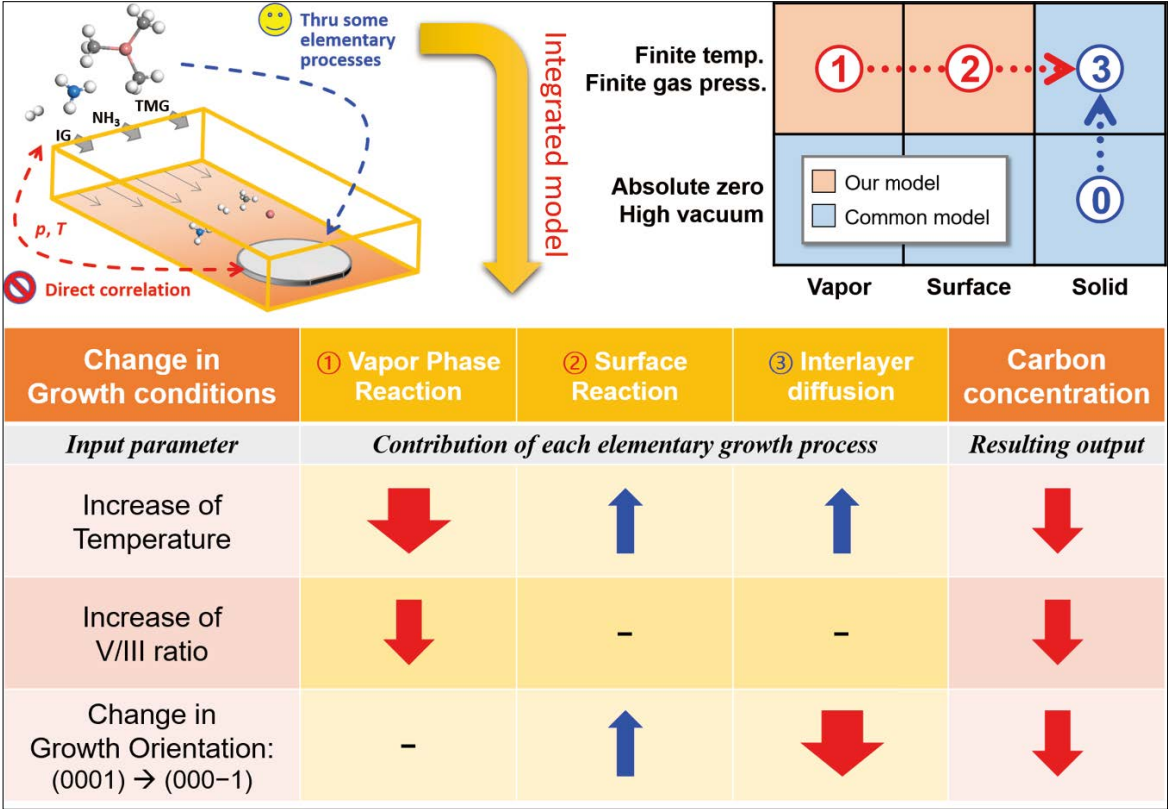


図 3-2-8 各素過程が残留炭素濃度の増減に与える影響

【質疑応答】

Q：特定の材料をうまく成膜するという目的があって、そのプロセスを探索するためのシミュレーション手法としてとらえたが、この手法で得られたメカニズム等の知見からより良い材料を提案するようなことは可能か？

A：ここでは特定の材料のシミュレーション手法とその活用例を紹介した。この取り組みを他材料に展開することにより、デジタルツイン環境を整備することができる。マテリアルズ・インフォマティクスにより予測された複数の候補材料に対して、デジタルツイン環境を活用した機械学習を行うことにより、産業化の観点から有用な材料を選択することが可能となる。

Q：重要なのは $\mu - E$ の関係から $p - T$ の関係に変換するところだと理解したが、まだ考慮できていない物理過程などはあるか。

A：今回紹介した研究事例ではガス分子のエントロピーしか考慮していなかった。最近、結晶の格子振動（フォノン）や吸着分子の配列エントロピーの効果を入れた解析を行えるように物理モデルを拡張している。今後、「富岳」を用いた大規模計算によりステップ・ダイナミクスの効果を取り入れることで、物理モデルをより現実の系に近づけることができる。

3.3 触媒インフォマティクスの実践に見るプロセスの重要性

谷池 俊明（北陸先端科学技術大学院大学）

触媒分野でのインフォマティクス研究を実践する中で感じているプロセスの重要性について、最新の研究成果をもとに話題提供をする。

ハイスループット実験とは、高度に自動化・並列化した装置と効率的な実験プロトコルのいずれか、あるいは両方を駆使し、従来の実験方法と比較してスループットを飛躍的に改善した実験を示す。目的は、科学的な仮説検証作業の飛躍的な加速やワーク・ライフ・バランスの確保、産業界においては人的資源のマネジメントなどさまざまあるが、一言でいうと研究資源の効率的な利用が根底にある。

我々は、社会課題に対するマテリアルソリューションを、人的、金銭的、時間的に最小限の投資で生み出すことを第一の目的としてハイスループット実験を行っている。また、必要に応じてハイスループット実験装置やプロトコル開発を行っているところも特徴であり、実際に装置を作って触媒やポリマーの研究を行っている。

ハイスループット実験に求める第2の目的は、MIの実践である。まず、その原資となるデータについての問題点を簡単に述べる。MIでは、取り扱うデータ量が十分に大きく、パラメータ空間にわたり均一に分散し、かつ互いに整合性の取れたデータの存在を前提としている。データ量という観点からは、文献や研究グループに散在する過去のデータを収集することも考えられるが、このようなデータはその時代の研究者の興味や実験方法、実験条件、すなわちプロセスを強く反映しているため、物質とプロセスからなる広大なパラメータ空間に対して離れ小島が点在するようなデータ構造に陥っている。このため、量的、分散性、整合性に配慮したデータを短期間で収集する技術として、ハイスループット実験が重要であると考えている。

メタンの酸化的二量化反応（OCM反応）は、豊富に存在する天然ガスの主成分であるメタンを部分的に酸化しエチレンを直接合成する反応である。高難度反応の1つであり、現在に至るまで40年にわたり研究開発が続けられているが、工業生産の目標と言われるC2化合物の収率30%に届く触媒やプロセスはほとんど

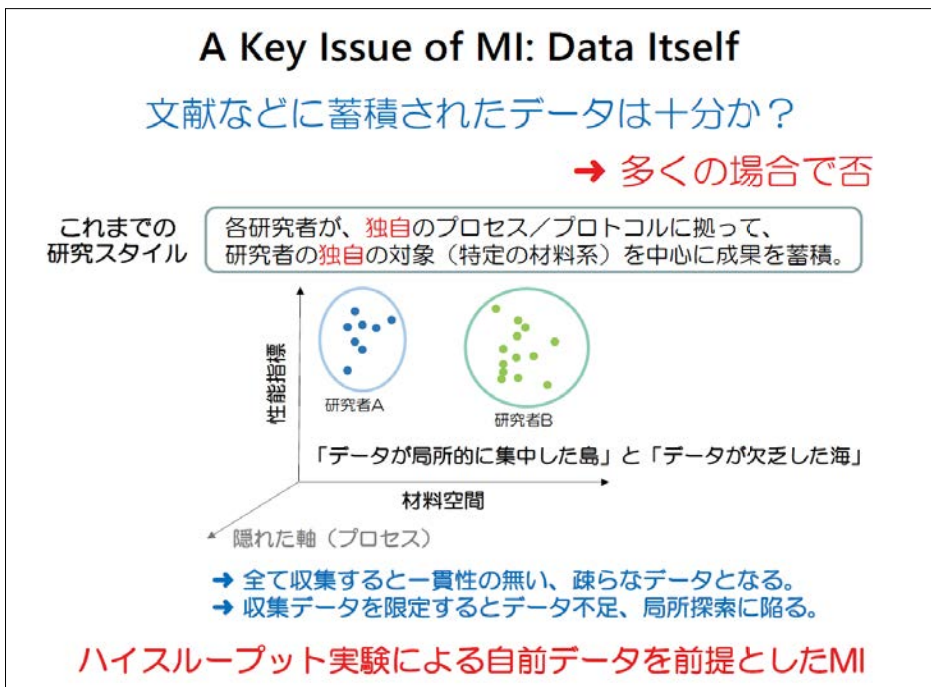


図 3-3-1 文献などに散在するデータの特徴

報告されていない。OCM反応はJST-CREST「多様な天然炭素資源の活用に資する革新的触媒と創出技術」が掲げる主要反応の1つであり、我々も当該プロジェクトの中で研究に取り組んでいる。

MIの隆盛に伴い、40年間に蓄積された文献データを収集し触媒設計に活用する研究が報告されている。触媒研究者が共通的に認識していたものを統計的に確認することはできたが、文献データが抱える重大な欠陥も明らかになった。40年間のデータとしては、量が非常に少ないこと、経験的な方法で発見された優れた触媒の周辺にデータが集中していること、また、データ分散に見られる人為的バイアス、触媒評価の方法や条件の不統一によるデータ間不整合などである。また、高性能触媒を狙っているため、低性能であった触媒の結果は文献には載せていないものが多いという問題もある。

我々の研究グループではOCM触媒開発の律速であった評価方法に対してハイスループット触媒評価システムを開発した（図3-3-2）。

①メタンや酸素などを含む混合ガスを供給し、②スプリッターで20分割後、③それぞれ触媒を含む20本の反応管に流通させる。④反応後の出口ガスを開発したオートサンプラーに順次送り、⑤質量分析計で分析する。この装置では、20触媒の性能をあらかじめ登録した反応条件で完全自動取得することができ、1日に4,000点を超える斉一的な触媒データを収集することができる。

実験の流れの例を示す。まず、現時点でベストな触媒として知られるマンガンナトリウムタングステートをシリカに担持した触媒に対し、1～2成分を置換した59触媒を図3-3-3に示す並列化した含浸法によって調製した。次に、文献のOCM反応条件範囲を参考に、温度、アルゴン（キャリアガス）濃度、メタン酸素比、ガス総量について216条件を含むプログラムシーケンスを作成した。

それぞれの触媒が性能を発揮する条件が同一でないことは明らかであり、広い条件で斉一的に触媒性能を比較することが重要である。これまでのハイスループットスクリーニングではプロセス条件スクリーニングのための自動化は行われておらず、固定した条件で触媒性能を比較するにとどまっていた。我々は、この装置を3日間運転することで、59触媒に対して12,700点のデータを取得した。これは過去40年間に取得されたデータ量と比較して1桁大きいだけでなく、触媒の善し悪しにかかわらず、同一のプロセスを用いて広範な条件で斉一的に取得された初めてのデータである。このようなデータを多面的に解析することで、この反応に対する

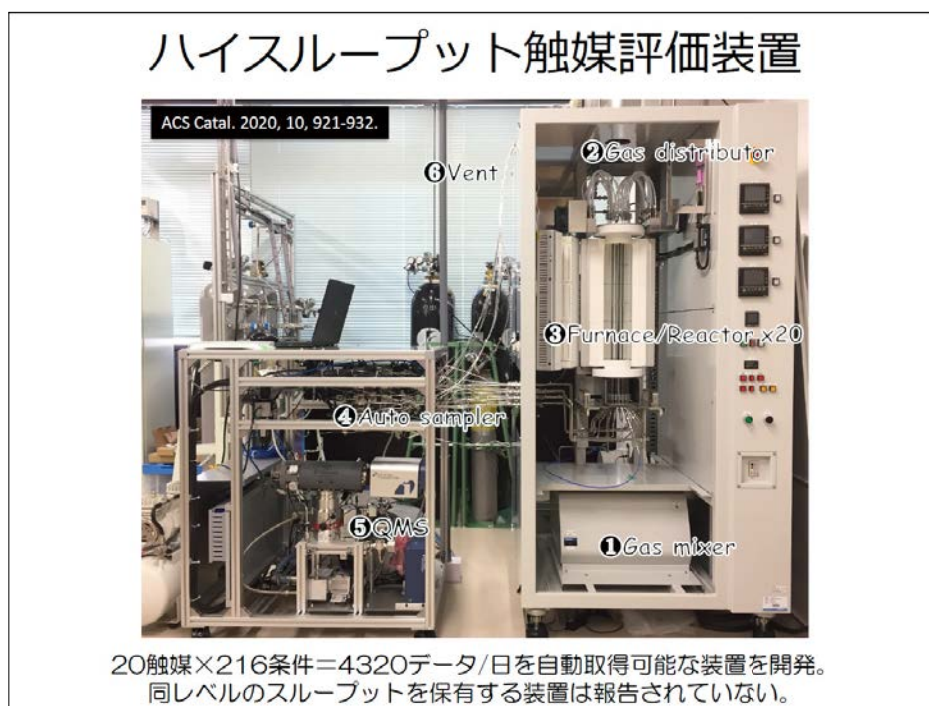


図 3-3-2 ハイスループット触媒評価装置の外観

様々な知見が得られた。

実際に得られた結果を図 3-3-4 に示す。メタン転換率を X 軸、C2 選択率を Y 軸、そして「メタン転換率」×「C2 選択率」で表される C2 収率を色彩軸で示した散布図である。

全データは C2 収率 21% 以下の領域に反比例的に閉じ込められていることが分かり、転換率と選択率が強いトレードオフの関係にあることを示している。これが、OCM 研究が 40 年間に渡って工業的な目標値に突破できなかった最大の原因である。

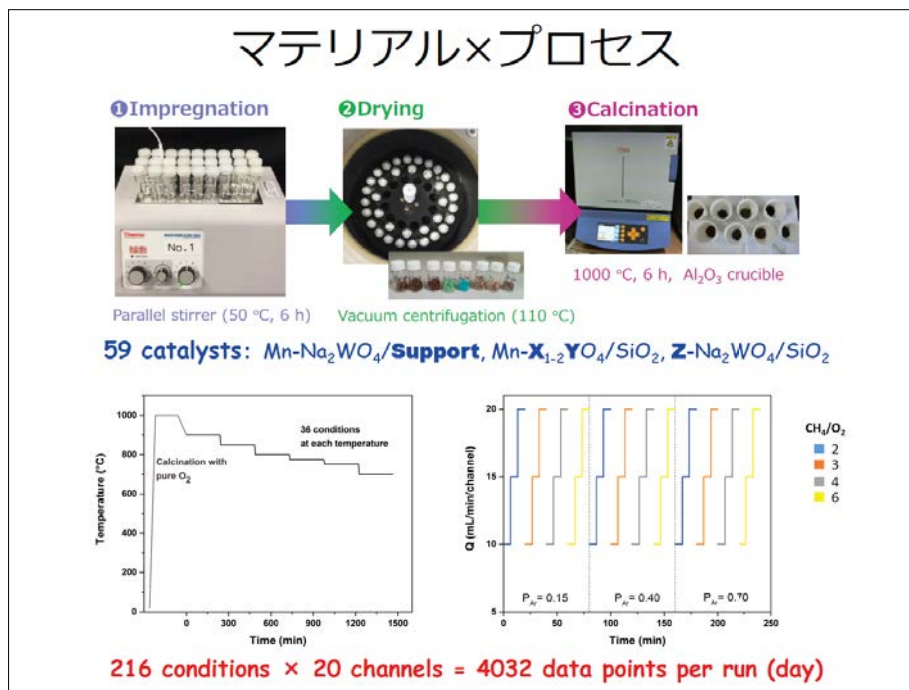


図 3-3-3 含浸法による並列触媒調製プロセス

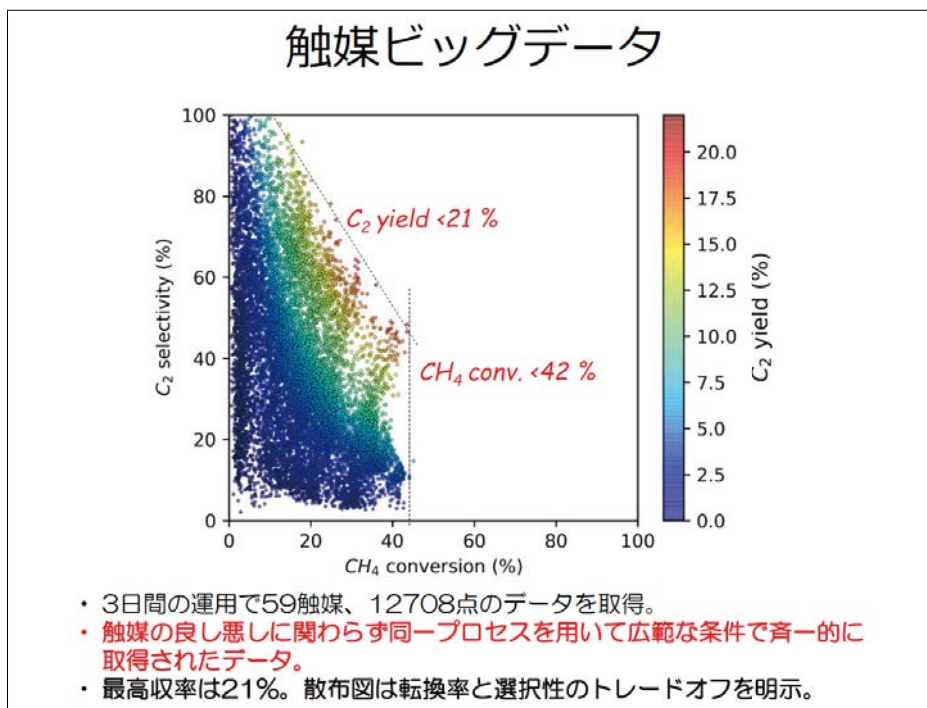


図 3-3-4 ハイスループット触媒評価装置で得られた OCM 反応結果

ハイフーループット実験により得られた大規模、かつ斉一的なデータセットは、機械学習の難度を格段に引き下げるものであり、過去40年間のデータに対するランダムフォレスト回帰のテストスコアが0.45であったのに対して、0.89という高いスコアを比較的容易に得ることができた。C₂収率が10%以下のデータを意図的に除いて同様の回帰を行った場合には、外挿に当たるC₂収率10%以下の領域は予測することができず、本来性能の低い触媒やプロセス条件での収率を高く見積もった結果となった。これらのことから、ネガティブデータの不足がインフォマティクスにとって致命的な問題になるのが明らかになった。

次に、データの分析が触媒やプロセス開発をどのように加速するかのを示す。図3-3-5は、OCM反応の副生成物である一酸化炭素の選択率をX軸に、二酸化炭素の選択率をY軸にとり、C₂収率を色彩軸で示した散布図である。

C₂収率が高い黄色や赤の点が集まった領域に着目すると、CO生成はほぼ0%に抑えることができるが、CO₂生成を10%以下に抑える触媒や反応条件は存在しない。OCM反応は、気相におけるフリーラジカル反応に触媒が介入することで進行するが、CO₂生成は気相反応が過剰であることを強く示唆しており、そのため反応プロセスの条件見直しが必要だと考えた。

具体的には、触媒評価装置は3エリア（T1、T2、T3）独立の温度制御が可能で、触媒層はT2エリアに固定されている。T1、T3エリアは均一な温度ゾーンを確保するためT2エリアと同じ温度にするのが通例であるが、T1、T3エリアの温度最適化を検討した結果、T2エリアより100℃下げた際にC₂収率が顕著に向上することが判明した。（T1 = T2 = T3）での最高収率21%から、（T1、T3 = T2-100℃）の条件では文献チャンピオンデータと同レベルの26%程度にまで向上した。

さらに、反応管の形状も極めて重要である。図3-3-6に示すように、反応管の内径やデッドボリュームを変化させたところ、同一の触媒であっても収率が10%から26%の範囲で大きく変化することが分かった。

これらの実験結果から、触媒反応の開発に当たってマテリアルとプロセスの同時最適化が非常に重要であることが示される。従来の研究では、触媒開発とプロセス開発は別々に行われてきたが、ハイスループット実験とインフォマティクスの導入によりマテリアルとプロセスからなる広大なパラメータ空間の探索が期待される。

化学反応は素反応の集合によって構成されるため、これらを制御する触媒設計は1成分では困難であり、

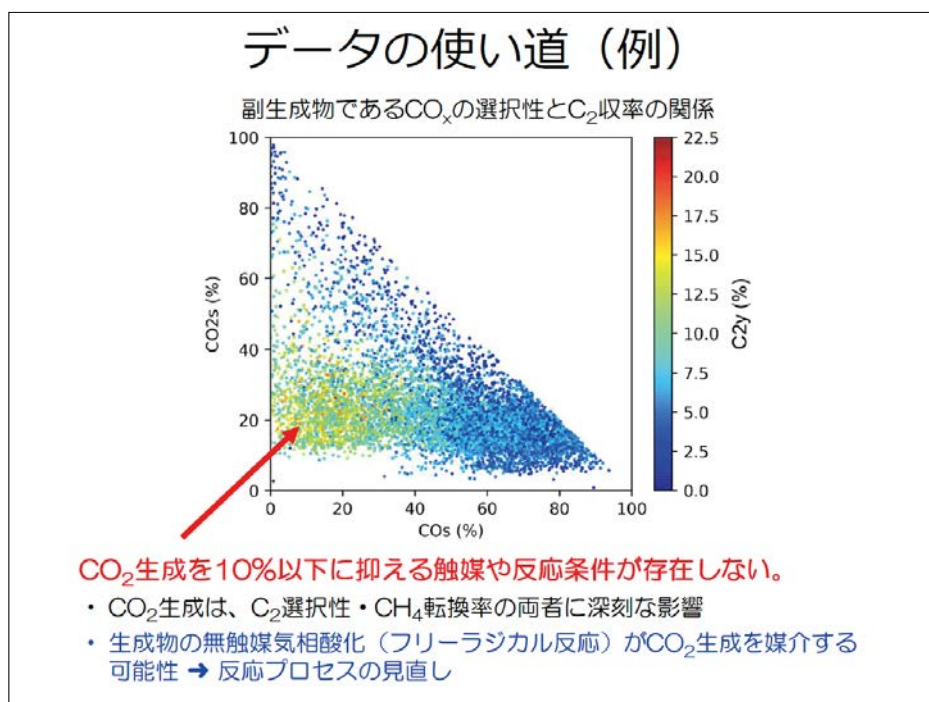


図 3-3-5 OCM 反応での副生成物（CO、CO₂）の生成状況

触媒評価におけるプロセスの重要性

ある種の材料性能は、合成だけでなくこれを成形・運用するプロセスに大きく依存（例えば高分子の成形器、触媒を用いる反応器）。

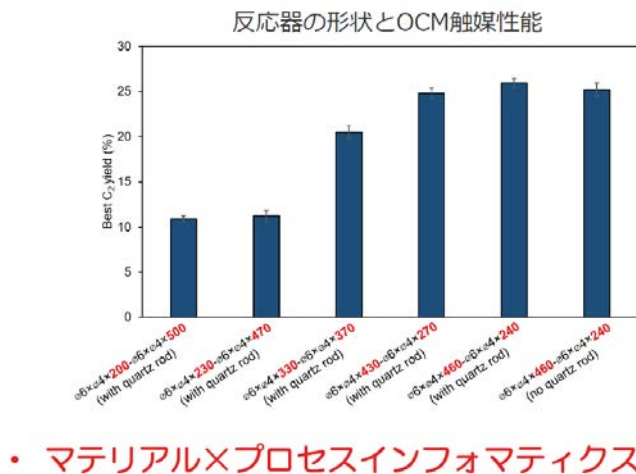


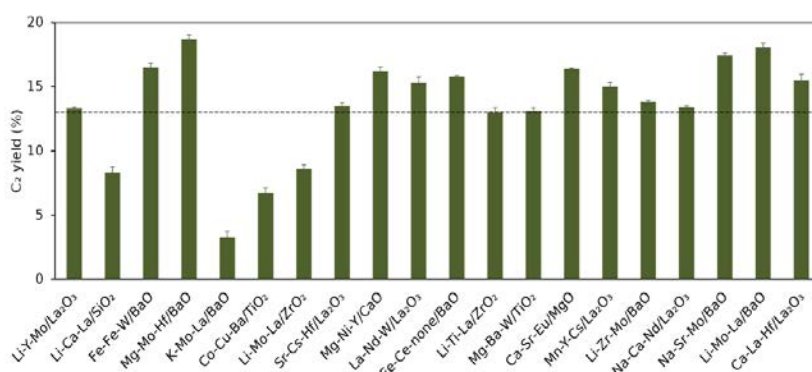
図 3-3-6 触媒評価装置の反応器の違いによるC2収率

複数の物質の組合せが基本になる。しかし、組合せの予測はMIや第一原理計算を用いても難度が高く、全ての素反応がどのようなものがあるか、また素反応全てに対して触媒がどのように介入するかという速度論計算も難しい。触媒研究は、現在も経験と直感に基づくトライ・アンド・エラーに強く依存しており、経験的な成果の積み重ねは物質空間のサンプリングに著しいバイアスを生み出している。

我々は、28元素から3元素を選び9個の単体から選出された4元系触媒 36,540種類から300種類の触媒

最新研究

バイアスフリーなデータによる非経験的触媒設計



- ・ 乱数を用いて新たに触媒を生成
- ・ 決定木が収率13%以上を予測した20個の触媒を実際に検証
 - 80%の予測精度、全ての触媒が未報告
 - 前知見を必要としない触媒設計の実現

図 3-3-7 バイアスフリーなデータによる非経験的触媒設計

を無作為に抽出、性能をハイスループット実験によって得られたバイアスフリーのデータから、決定木に基づいて触媒設計と性能の相関を学習する研究を行った。乱数を用いて新たに触媒を生成して、決定木が収率13%以上を予測した20個の触媒の性能を検証した結果を図3-3-7に示す。

20種類の触媒のうち、16種類の触媒が収率13%以上の結果を示しており、80%の予測精度が得られた。また、これらの触媒は文献では報告されていない触媒であった。今後、工業的な目標である収率30%まで高めていくことは必要であるが、前知見を必要としない触媒設計が実現できたと考えている。

インフォマティクス実践のためには、斉一的なデータが極めて重要である。触媒の例で示したように、マテリアルとプロセスは不可分であり切り離して考えるべきではない。今後、ハイスループット実験による触媒とプロセスに関するデータを使って、反応ネットワークのミクロキネティクスや関連する速度論パラメータをデータ同化することで、反応のブラックボックスの中味を明らかにすることができれば、真の意味での材料とプロセスの同時最適化が実現すると考える。

【質疑応答】

Q：過去の文献データ（回帰0.45）を一旦クラスタリングし、クラスに分けてから再度回帰し直すと、回帰がよくなるのではないか。

A：クラスタリングをすれば回帰はよくなるが、得られる結論は触媒研究者が既に知っている知見をそのまま改めて取り出しているのに過ぎない。文献データは偏在しすぎている点が問題である。

Q：データにバイアスをかけない魅力は理解できる。紹介されたもの以外に、バイアスをかけないことでうまく行った例はあるか。それによって新たな気づきにつながったような例があるか。

A：OCMとは別の反応で、それまでのチャンピオンデータで収率45%の反応を、適応学習の第2世代目で抜いた例もある。OCMでも、過去に見出された触媒よりも選択率が高い触媒が見つかった。探索空間を広げ、網を大きく張ったことによりサンプリング密度は落ちたが、大きなメリットがあった。この検討過程で得られたデータをグループクラスタリングすることで触媒設計における知見として取り出すこともできるので、それを使うことで従来型の研究も加速できると考える。

Q：ハイスループット実験による触媒とプロセス条件のデータと、ミクロキネティクス、シミュレーションをデータ同化することが、材料とプロセスの同時最適化にどうつながるのか。

A：300触媒を乱数で取りだして検討したとき、各触媒がなぜ良いのかというメカニズムが記述できないことが悩みであった。一方で、ミクロキネティクス、シミュレーション側からの悩みは、触媒上で何が起きているか分からないので反応ネットワークが描けない、速度論データが取れないことである。触媒化学研究者としてミクロキネティクスに期待することは、触媒を反応機構という点で記述することである。ハイスループット実験にシミュレーション側から期待することはデータ同化である。この2つができれば触媒の記述もプロセスの記述も同時にできる、つまり同時最適化できることになる。

Q：色々なプロセスにおけるハイスループット実験装置を設計・製作できる人材はどのように育成すべきか。

A：従来型の実験のプロセスをよく知っていること、プロセスで何が重要かをしっかり知った上で設計に落とし込めることが必要。装置設計はバリアが高いが、デザイン思考的に、まず1回造ってみようというメンタリティが重要である。

3.4 計測インフォマティクスを目的とした計測スペクトルのスパースモデリングの自動化

吉川 英樹（物質・材料研究機構）

計測インフォマティクスと材料データプラットフォームについて、現在取り組んでいることを中心に紹介する。材料データプラットフォームにおける、データをつくる、ためる、利活用するの機能のうち、データをつくることに関しては、NIMSの研究者や共同研究している各機関のデータ、あるいは産業界のデータなどを広く集め、過去の論文から抽出したデータを長い年月かけてMatNavi、PolyInfo等のデータベースとして蓄積してきた。実験装置からのデータの自動収集の検討も進めている。つまり、他機関、他者のデータをいかに効率よく収集し、利活用し、かつ、いろいろなナレッジをいかにそこから引き出すかということが我々のミッションであり、①蓄積した文献、他者・自分の過去のデータ、②自分の新データ、③AIを志向して開発した実験データ解析ツールの三位一体を実現するプラットフォーム構築を目指している。

マテリアルズ・インフォマティクスでは、理論計算データでの成功例は多い。理論計算はばらつきの少ない系統的な大量のデータを自動計算で得ることができるためである。実験データについては、自動実験は容易ではないためばらつきの少ない系統的なデータの数が充分ではなく、データ駆動型研究が難しい状況にある。文献に掲載されている実験データを集めても、装置や実験条件、実験者の違いなどによりばらつきが大きく、材料の本質が見えなくなるという問題がある。ばらつく他者のデータをどのように処理してデータ駆動型解析に利用するかという仕掛けを検討するのが大きな課題の1つである。

医学の分野でも似た課題があり、例えば、脳のMRIによりアルツハイマー病の診断をする際、頭部の骨格や測定装置の違いが問題になることがある。この場合でも画像の標準化処理により病巣に由来する画像の違いをクリアに判断する技術が使われている。材料研究においても、装置の違いに由来するデータの揺らぎを吸収して、材料由来のデータをどのように抽出するかが大きな課題になってくる。

その1例として、X線光電子分光（XPS）のスペクトル解析手法を開発した。実験装置が変わると、たとえ同じ材料を計測しても、スペクトルのピーク位置、幅、形が全く同じになることはなく、位置、幅、形が微

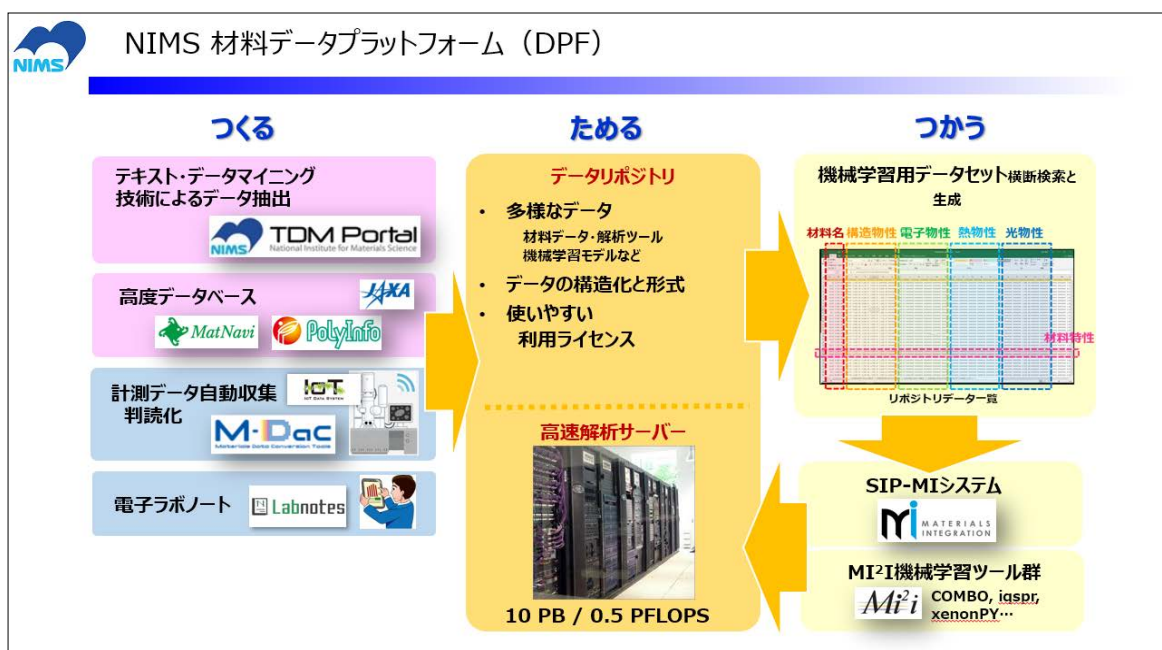


図3-4-1 NIMS 材料データプラットフォーム

妙に変化し、バックグラウンドの形も変化する。これらを統計的に処理し、装置の違いに由来するデータの揺らぎも含めてデータを解釈するモデル化（スパースモデリング）の検討をしてきた。一般にモデル化された解は多数個存在し得るため、ベイズ情報量規準を用いた解の選別を行っている。いろいろな機関の実験データをモデル化し、装置由来の揺らぎもモデル化することで、材料由来の本質的なスペクトルの違いを明らかにすることができる。

ただし、以上のデータ駆動型解析のシステムを材料データプラットフォームに構築しても、解析の前段階として、実験データを集め、形式を整えるという手作業に現場で非常に多くの手間をかけていることが分かった。この手間を回避しないと、良いデータ解析アルゴリズムを作っても使われないため、データの収集、変換、解析ツールの適用、蓄積というデータのワークフローを構築した。

そのワークフローの概念図を図3-4-2に示す。IoT技術を使って装置から生データをセキュアに自動転送、自動分類した後、生データの中からメタデータを自動抽出する。メタデータの記述には、装置ごとに異なる用語が使われているので、時には装置メーカーの協力も得ながら、メタデータを一般的な用語に変換して記録するという取組も行っている。数値データについては、グラフ化処理、数理統計処理、スパースモデリングなどが自動的に行われ、蓄積されるという流れを取っている。

ワークフローをXPSの例でもう少し詳しく示す(図3-4-3)。バイナリで得られた生データをテキストファイルに変換し、自動的にメタデータとしての実験条件を抽出し、その用語を判読化・構造化した後にXML形式で出力する。数値データは判読性の優れたフォーマットに変換し、それを解読するためのPythonパーサもセットで記録する。データだけではなく、そのデータを読み取るパーサも合わせて記録していることが特徴である。これらの変換ツールの内で計測器メーカーの理解が得られたものはM-DacとしてGitHubで公開している。

自動化されたワークフローは、NIMS内の共用サイトにて、グループ内共用やNIMS内公開をしている。一部についてはNIMSのデータリポジトリ（MDR）で外部に一般公開する予定である。

XPSを例としたスペクトルデータのスパースモデリングのワークフローの詳細について説明をする。ワークフローの実行は、基本的には1、2分で全ての処理が終わるようにし、速いものでは秒単位で済むものもあり、リアルタイム性を非常に重視している。

実行速度の早さが求められる一方で、ワークフローで得られた解の信頼性の高さを担保する必要もある。

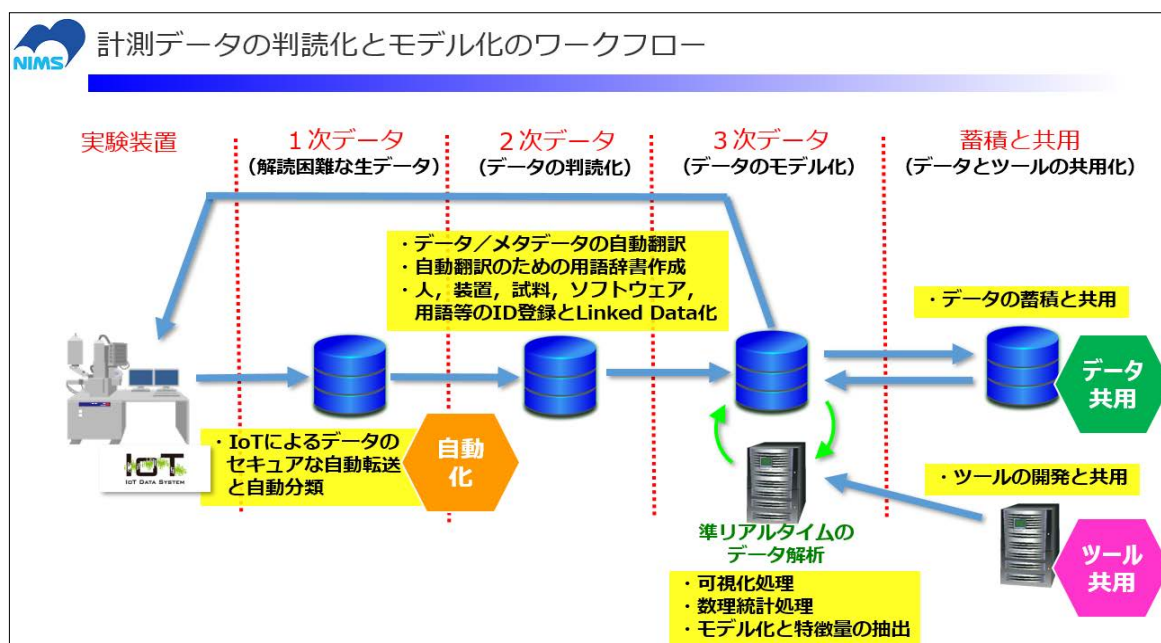


図3-4-2 計測データの判読化とモデル化のワークフロー

実はリアルタイム性と信頼性は相反する場合が多いので、それをどう両立するかが難しい。我々は図3-4-4に示すように、リアルタイムに高速計算できる手法と、計算時間はかかるが信頼性が高い手法を組み合わせている。高速計算の手法に関しては、遺伝的アルゴリズムを含むメタヒューリスティックな解の探索方法と、ベイズ情報量規準による解の選別を組み合わせた簡便法であり、秒単位から長くても分単位の計算時間で済ませる。信頼性が高い手法としては、温度交換マルコフ連鎖モンテカルロ（MCMC）法を使用している。この2つを組み合わせることによって高速性と信頼性を両立させている。簡便法で求めた解に、温度交換MCMC法で求めたエラーバーの一般式を適用する取組も行っている。

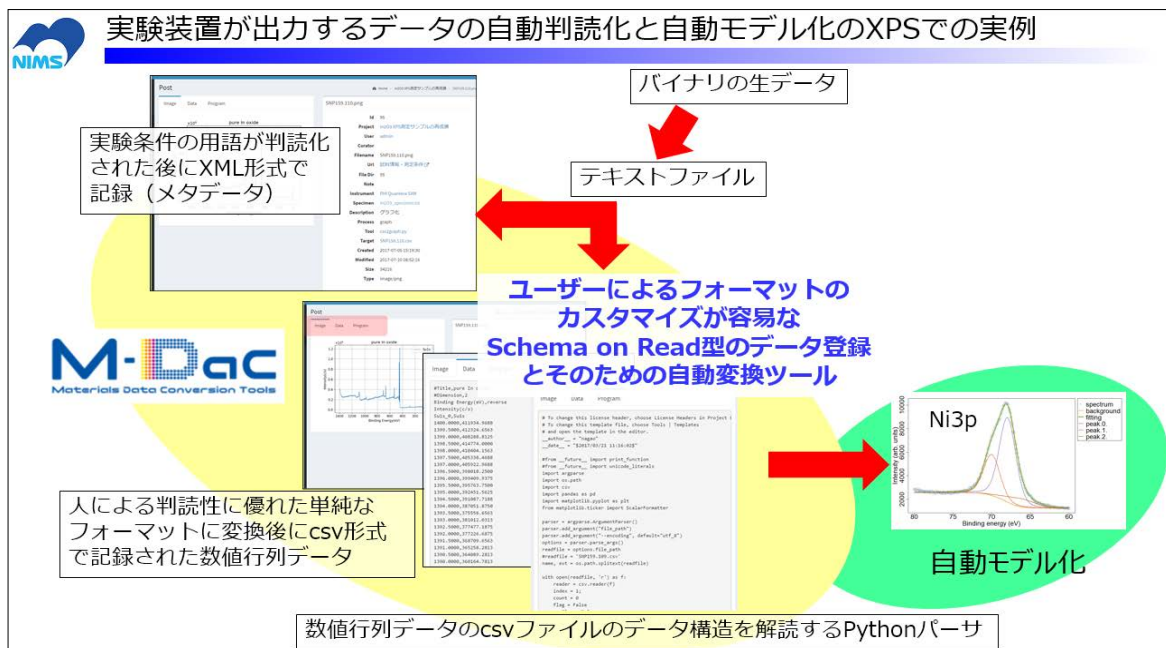


図3-4-3 XPSでのワークフローの例

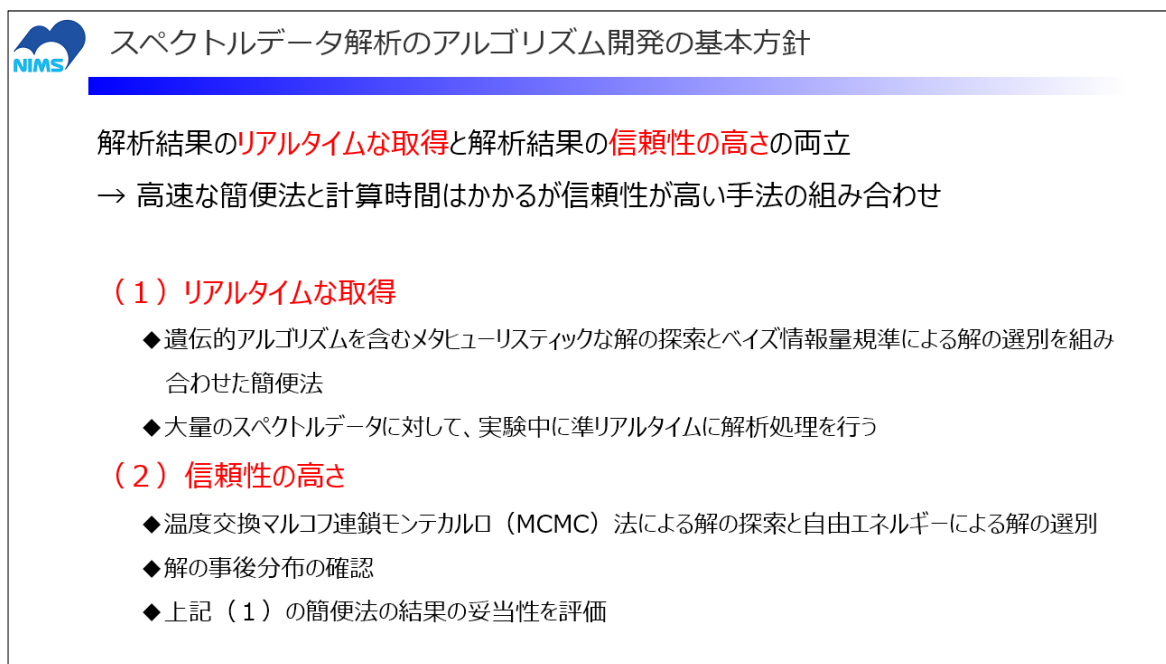
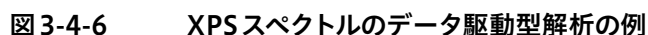


図3-4-4 リアルタイム性と信頼性の両立

3種類のストロンチウム化合物（酸化ストロンチウム、炭酸ストロンチウム、チタン酸ストロンチウム）と酸化チタンから成る混合物において、ストロンチウム3d、酸素1s、チタン2pという複数の関連スペクトル全てが成分比の整合性をもって説明できる最適な化合物種の組合せを探す問題に取り組んだ。10種類の化合物



の組合せである合計1,023通りから最適なものを探索した。メタヒューリスティックな解の探索とベイズ情報量規準（BIC）を使って求めた解の選別を組み合わせることによって、それぞれ解の候補をランキングとして出すことができた。BICが一番低いものが最も確からしいということが分かった。

この手法はXPSを例として検討したが、他の計測法にも適用し得る汎用性を持ったアルゴリズムであるので、他の計測法に随時展開していきたい。重要なポイントは、他者の参照データ群を使い、それらの実験装置や実験条件の違いを考慮したデータ駆動型のアルゴリズム開発を進めることである。このような最新のアルゴリズムを使ったツールを共用することを通してデータの共用を進める。

【質疑応答】

Q：XPSによる混合物の同定でベイズ情報量規準（BIC）を使っているが、化合物の数が少ない方が優先的に選ばれてしまうようなことはないか。

A：BICには数が少ない方がより望ましいというタームが入っているが、その影響はあまりない。複数のスペクトル（ストロチウム3d、酸素1s、チタン2p）が全て整合するようにするという拘束条件をかけるのと正しい答えになる。逆にいうと、ストロチウム3dだけで合わせようとする、少ない組み合わせを選ぶ傾向がでる。

Q：XPSスペクトルは、昔は、いわゆるピーク分割とかピークフィッティングといていた。あれと何が違うのか。

A：できるだけ広いパラメーター空間で解を探索している点が大きく異なる。例えば、半値幅、位置、ピーク形状、裾の形、バックグラウンドの形、構成化合物種など、非常に幅広いものを対象にすることができる。計算機パフォーマンス向上による部分もあるが、アルゴリズムの進歩の効果が大きい。また、参照スペクトルが数多く蓄積されることにより、装置の違いによる揺らぎの程度が見えてくるので、その情報が事前情報として入れられるようになったこともある。

Q：装置からの生データ、バイナリデータを自動的に仕分ける話があったが、装置メーカーに協力してもらうようなアプローチはしていないのか。

A：そのような動きをしている。経済産業省が進めている計測データの共通フォーマット化にも我々は関係しており、いろいろな計測機器メーカーにフォーマットの開示をお願いしようとしている。NIMS内で閉じるのではなく、オールジャパン、場合によっては海外も含めて公開して、皆さんに使っていただくようにしたい。また海外でそういう動きがあれば連携していきたい。

Q：MatNavi、PolyInfoは論文掲載の実験データが中心だが、そこに新しく取ったデータや、今回のような処理をして揺らぎを減少させ使いやすくしたデータを加えていくのか。

A：現状では、データベースを一緒にすることは難しい。しかし、スパースモデリングされたモデルパラメータは公開していきたいと考えている。スパースモデルの中にはナレッジが入っているので、そのナレッジが入ったデータを、本来の基データとどう区別して公開するかは、今後の検討課題である。非常に重要なデータベースになると思う。

4 | 総合討論

ファシリテーター 伊藤 聡（JST-CRDS、計算科学振興財団）

総合討論では、コメンテーターを、産業界から3名、アカデミア側から2名の先生にお願いしている。それぞれの方からのショートプレゼンテーションから始める。

【水野 洋（パナソニック株式会社）】

モノづくり企業という意味で、マテリアルズ・インフォマティクスだけではなくプロセス・インフォマティクスにも大きな期待を持っており、我々の考える方向性について説明する。

当初はデータ駆動科学としては、従来の計算科学に機械学習を付加することであったが、それだけでは十分ではない。現在は、図4-1-1に示すように、経験科学、理論科学、計算科学、データ駆動科学を統合したデジタルイゼーションという位置づけとして取り組んでいる。ここでマテリアルズ・インフォマティクスは広く捉えプロセス・インフォマティクスも包含すると考えている。

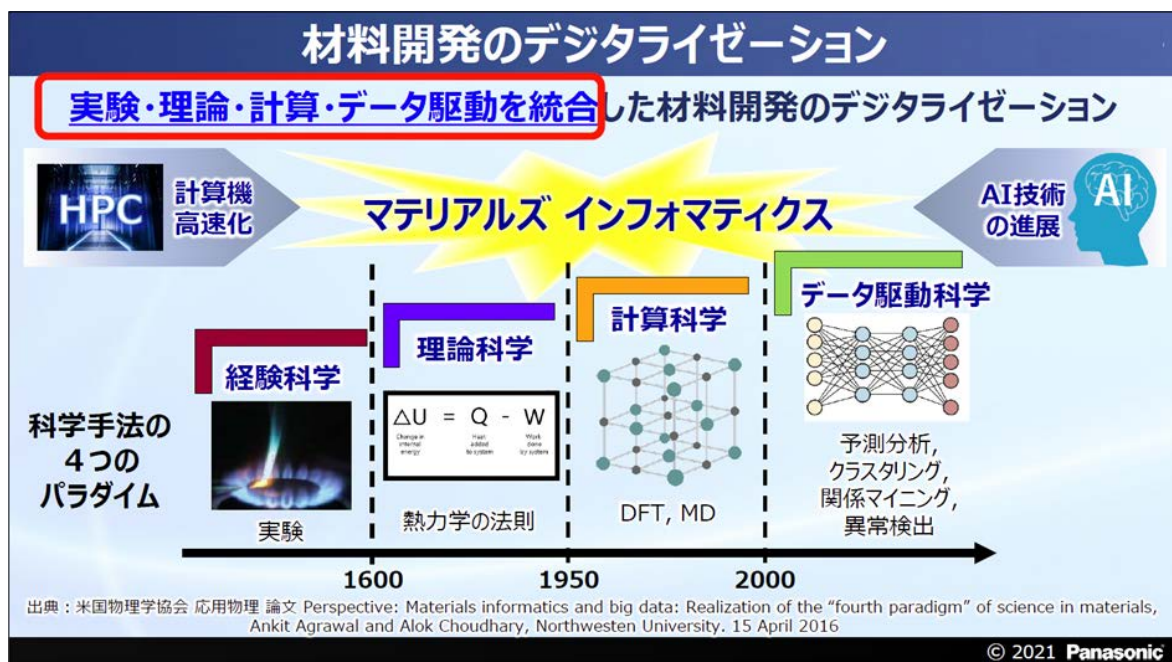


図4-1-1 材料開発のデジタルイゼーションの変遷と取り組み

材料開発デジタルイゼーションの全体像としては、図4-1-2に示すイメージである。基軸となるのはデータである。既知情報のデータは論文や特許から抽出しデータベースとして利用される。しかし、新材料探索という視点で考えると既存の論文や特許には活用できるデータは少なく、ハイスループット実験と、機械学習やベイズ推定を駆使することでデータを蓄積し、図中外側のループ（サイバー・フィジカル間ループ）をまわすことで目的に近づいていくプロセスと、また、近年はVAEやGAN、量子コンピューターを活用した生成モデルのアプローチも実用化されており、ハイスループット計算と組合せた内側のショートループ（サイバー内ループ）をまわすプロセスを並列で進めている。

プロセス・インフォマティクスに関連する我々の取り組みとして、自然言語処理を用いて論文から合成プロセスを抽出する工程を産総研、豊田工業大学と共同で進めている（図4-1-3）。自然言語処理は技術が進ん

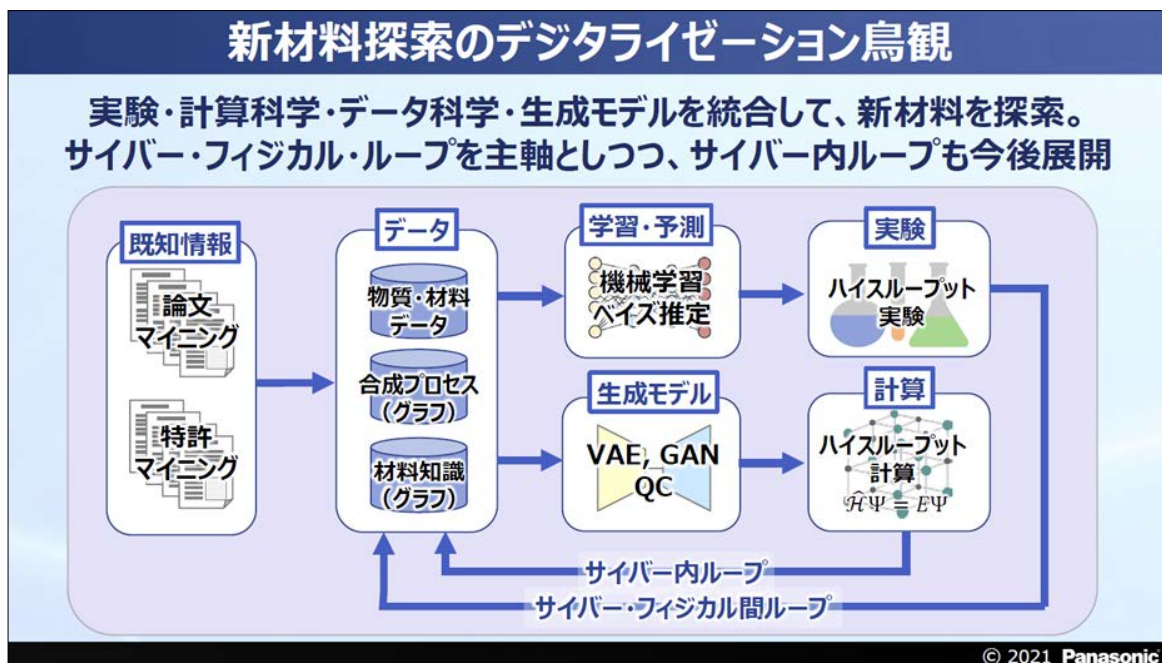


図 4-1-2 新材料探索のデジタイゼーション鳥瞰

できており、オープンソースを使うことによりさまざまなことができるようになってきている。

企業の視点から見るとマテリアルズ・インフォマティクスにおいては、マテリアルズ・最適化ゼーションとマテリアルズ・ディスカバリーの両輪が重要である。

我々はアカデミアとは主にマテリアルズ・ディスカバリーで連携しているが、材料メーカーから材料を購入し製品化していくことも重要であり、プロセス・インフォマティクスにおいても同じことが言えると思われる。すなわち、図 4-1-5 に示すように、我々は材料の開発を源流までやる場合と、材料メーカーから購入する、またはその中間のケースがあり、こういったことを視野に入れて取り組むことが企業としての取り組みとして重要である。

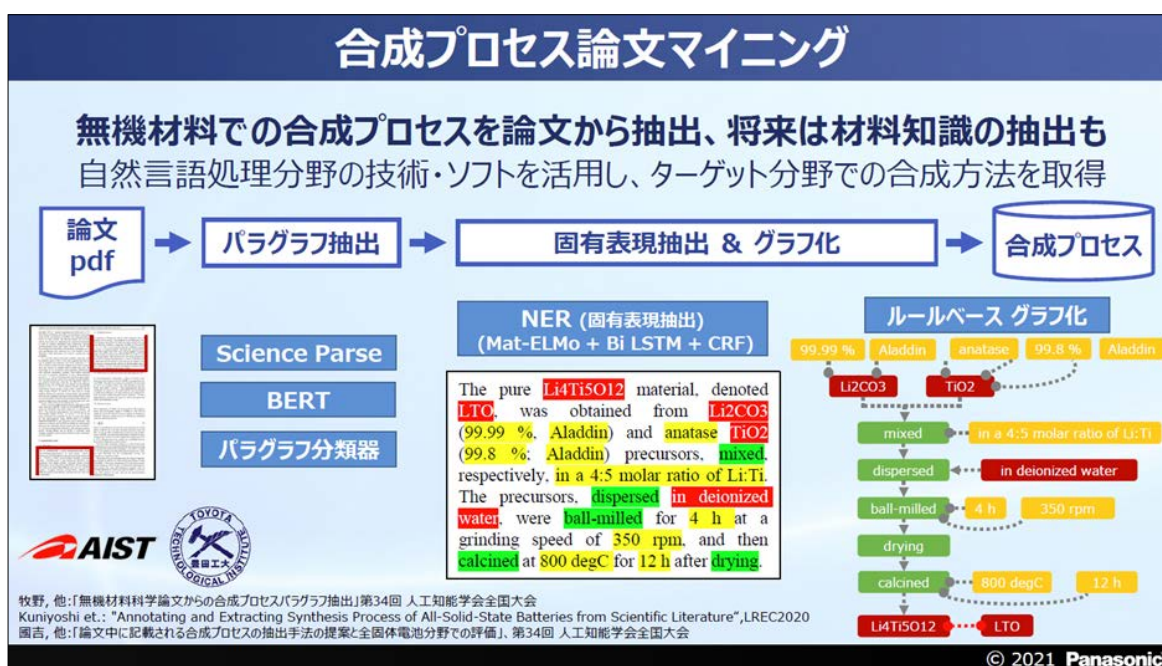
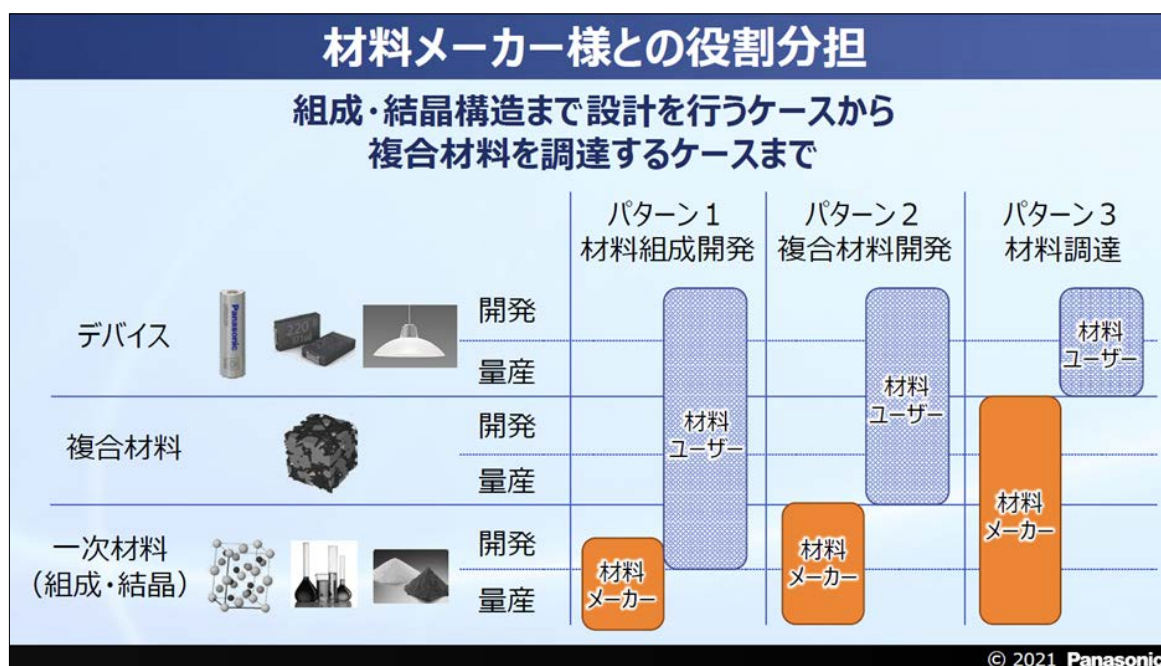


図 4-1-3 合成プロセスの論文マイニングの流れ

企業における材料開発		
“Materials Optimization” と “Materials Discovery” 材料ユーザー企業では、両方重要		
	Materials Optimization	Materials Discovery
目的	材料メーカーがラインナップした既存材料から応用に最適な複合材料を開発	従来には無い特性・性能を持つ未知の新材料を探索する
データ	・オープンデータベース ・自社の実験データ ・材料メーカーから提供されるデータ	・オープンデータベース ・自社の実験データ ・計算データ
有効なMI技術	・論文/特許マイニング（データ取得） ・統計学/機械学習 ・ベイズ最適化 ・ハイスループット実験	・論文マイニング（合成法、知識取得） ・機械学習 ・ベイズ最適化 ・生成モデル ・ハイスループット実験/計算

© 2021 Panasonic

図 4-1-4 企業における材料開発の二つの取り組み



© 2021 Panasonic

図 4-1-5 材料メーカーとの役割分担

我々はマテリアルズ・インフォマティクスという材料探索を量産まで繋げることをプロセス・インフォマティクスと呼んでいるが、本日のテーマはプロセス・インフォマティクスの初期段階に相当すると考えている。また、本日のテーマではないが、図 4-1-6 に示すように、もう一つの軸として、材料がシステムの中でどのような性能を持つのかを理解する意味でのマルチスケールシミュレーションも重視している。新材料から材料最適化といった考え方、合成プロセスから量産プロセス、企業とアカデミアの連携という視点で議論したい。

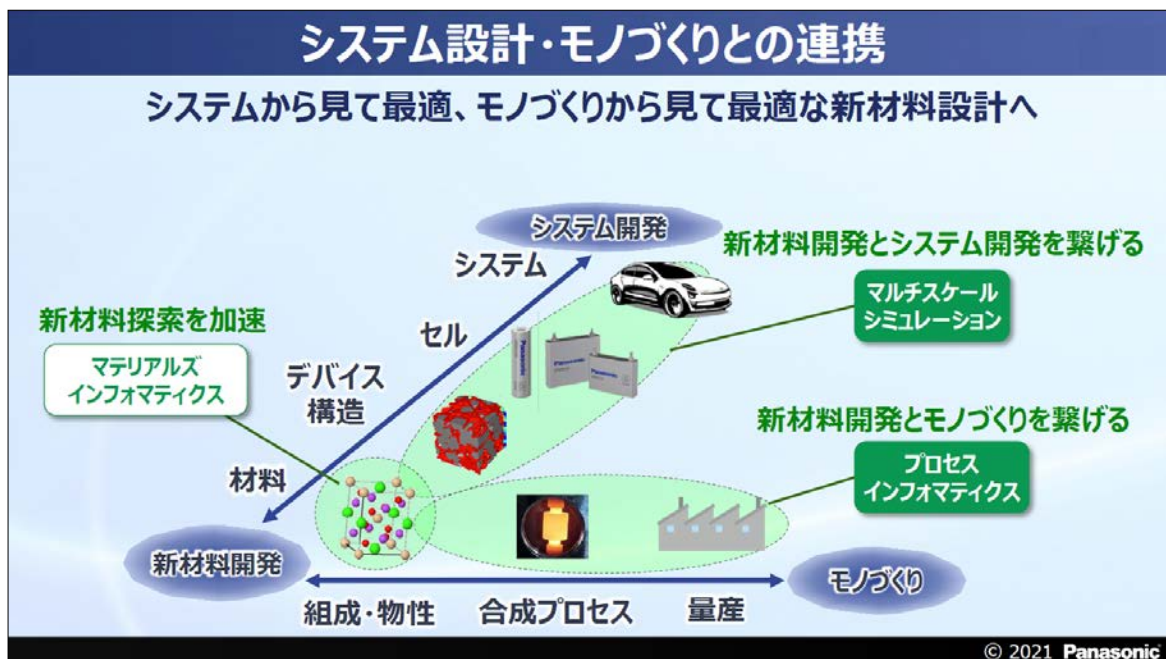


図4-1-6 システム設計・モノづくりとの連携

【富谷 茂隆（ソニー株式会社）】

材料・デバイスにおける基礎研究から製品化への道のりは、幾つもの障壁を乗り越えなければならない。まず基礎研究から技術開発の段階に魔の川がある。その次に、技術開発から事業化に向けて死の谷があり、さらに製品にするためにはスマート化された製造プロセスにする段階でダーウィンの海がある。このように3つの段階があり、それぞれに対応して、材料・デバイスにおけるX-インフォマティクスがある（図4-2-1）。材料探索では狭義のマテリアルズ・インフォマティクス、そして製品化、製造ということでは半導体業界ではバーチャルメトロジー、あるいは化学工学ではソフトセンサーと言われるマニファクチャリング・インフォマティクス、そしてそれをつなぐ非常に重要なプロセス・インフォマティクスである。また、それらを支える計測インフォマティクスが同時に重要である。

材料においては、単一の材料だけで機能を創製することはほとんどなく、材料を複合的に用いる場合がほとんどである。例えば量子ドットも、無機材料のコアに有機または無機のリガンドを複合化させる。混晶半導体の場合は、いろいろな材料の複合体で構成されている。この場合、組成の均一性が重要となる。あるいは有機太陽電池といったバルクヘテロ構造のように、有機材料が結晶構造を取るとか取らないかという点が機能発現に非常に重要となる場合もある。薬物では、いろいろな抗体薬物複合体という、かなり複雑な構造を有していることがある。

また、材料やデバイスは図4-2-2のようにそれぞれ階層構造になっている。このため、それぞれの階層構造をつなぐシミュレーションが重要である。また、階層構造をつなぐプロセス、特にアトム、クラスターからドメインをどのようなプロセスでつくり上げていくか、あるいは、ドメインからデバイスをどのようなプロセスでつくり上げていくか、が非常に重要である。

ナノテク分野での日米の特許戦略の違いを分析すると、日本は比較的材料分野が非常に強く、米国は用途、アプリケーションが強いことが分かる。米国のオバマ政権では材料部分を強くするためにマテリアルズ・ゲノム・イニシアチブができた。一方、日本は用途の部分の議論が弱いと言われている。このため、日本では、今後、用途を見据えてテーマ設定することが非常に重要になると考えている。

プロセスはどちらかというと基礎研究よりも応用寄りの研究のため、用途を意識したテーマの設定が非常に重要になってくる。材料側からの発想だけでは、用途を意識したテーマ設定は難しく、価値創造プロセスの策定が必要だと考えている。

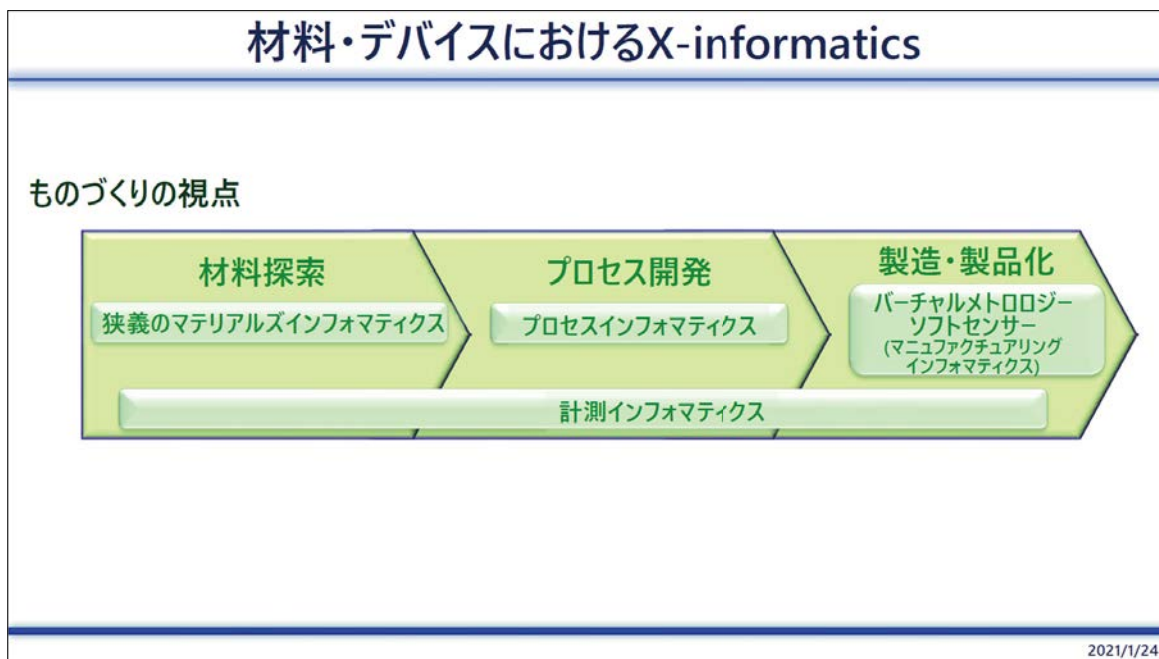


図4-2-1 材料・デバイスにおけるX-informatics

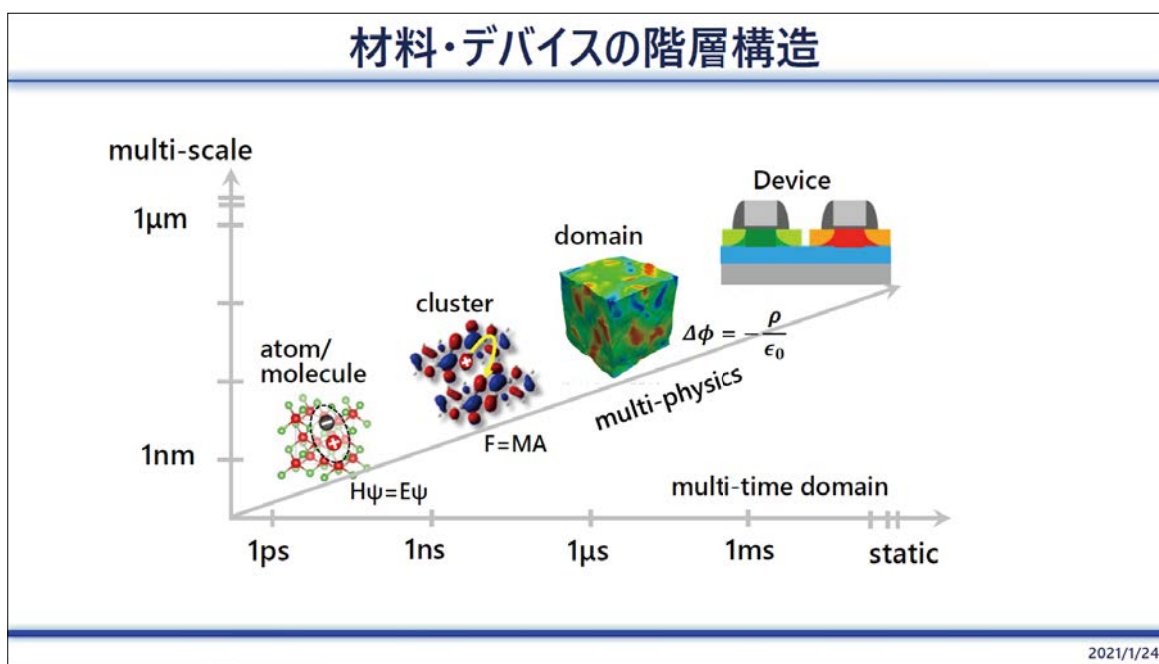


図4-2-2 材料・デバイスの階層構造

マテリアルズ・インフォマティクスについては、プラットフォームのサービスカンパニーとしていろいろなベンチャー企業ができています。日本でも幾つかあるが、米国には圧倒的な数があり、そこにベンチャー投資家からかなり資金が流れ込んでいる。新しい産業ができるのではないとも言われている。

プロセス・インフォマティクスに関するベンチャーはあまり多くはないが、比較的力を入れている例として、米国のWildcat Discovery Technologiesや、半導体系のプロセスを取り扱っているIntermolecularなどがある。

特許庁の特許出願技術動向調査では、MIに関しては、中国での特許出願も増えてきていることが分かる。こういった状況であるので、日本ではプロセス・インフォマティクスにかなり力を入れていかなければいけない

と考えている。

半導体製造工程には、成膜プロセスやエッチングプロセスなど多様なプロセスがあり、それぞれ個々のプロセスを極めることが重要であるが、さらに、これら様々なプロセスをインテグレートすることが非常に重要である。このため、半導体工場ではバーチャルメトロロジという、様々なプロセスをモニタリング、あるいは検査してデータ解析し、因果関係や加工条件の把握、装置改善とそれによる出来栄の影響の監視、装置の状態監視、あるいは加工パラメータの制御、判断の自動化する方法がかなり進んでいる。

最初に示したように、材料探索はMIでかなり進んでいる。一方、製造・製品化のほうもかなりデジタル化され、それを使ったデータ解析が相当進んできている。材料探索と製造・製品化の間のプロセス開発についてはまだ立ち遅れていると考えている。また、個別プロセス開発と同時に、プロセス間のインテグレーションを見据えてプロセス・インフォマティクスを立ち上げていけないとよい最終製品ができないと考えている。

今後はカスタマイズが重要になると考えている。今後、IoT社会が本格化し、様々なデバイスやセンサーが必須となるため、用途ごとのカスタマイズが求められる。多種多様な材料、デバイスを開発するために、プロセス・インフォマティクスは不可欠となり、プロセス開発の効率化が重要になってきていると思う。低コスト化や省エネルギー化、あるいは人的資源を含めた資源の効率化で、環境にやさしい開発につながることを期待している。

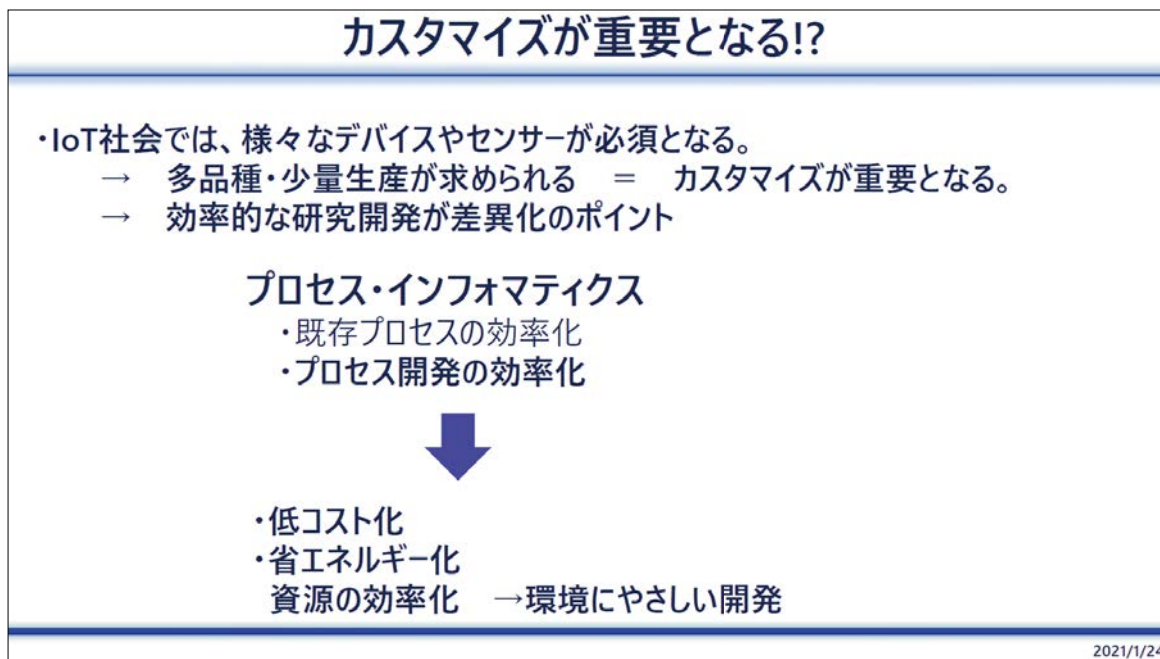


図4-2-3 カスタマイズが重要となる

【右田 啓哉（株式会社日本触媒）】

化学メーカー、素材メーカーの立場から意見を述べる。

日本触媒は、名前に「触媒」とあるが、有機、無機、複合材料といろいろなものを扱っている。さまざまな状況変化・激化に対応していかなければいけないと痛感しており、迅速に製品化して上市するためのスキームとして各種のインフォマティクスに大いに期待しており、広義には「研究開発をDXする」ことを狙っている。メーカーとしては、R&Dに対してもインフォマティクスは使うが、製造しさらに販売するところまで完遂してこそであるので、R&Dと製造はつながっているという意識でやってきている。

今回のワークショップでのプロセス・インフォマティクスにおけるターゲットは、主に合成プロセスであり、また実製造プロセスでの安定生産などは含めないということであるので、図4-3-1の左下に「ターゲット」と示



図4-3-1 プロセス・インフォマティクス (MI) とマテリアルズ・インフォマティクス (PI)

した部分だと認識している。また、MIによって標的的材料が得られる状況がすでに確立しているという前提で話を進める。

素材メーカーにおけるプロセス開発の諸問題を図4-3-2に示す。お客様もしくは最終市場からのニーズに応える材料を、マテリアルズ・インフォマティクス (MI) によって設計できたとしても、それを合成する方法はいろいろと考えられる。さらに、その合成方法に対してのプロセスの概念設計 (簡略設計) にも選択肢がある。さらには、弁やコントローラーなどの細かい装置などの詳細設計に進み、さらにその後、実機生産へと繋がる。もし、お客様の需要が僅かな変更だとしても、機能が変わると材料が変わる、そうすると製法も全く変わってしまう。ある種の相転移のようなことが起こり、これまでやってきたことが全く通用しなくなってしまうことが起きる。金子先生の講演では、プロセスの抽象化が必要だという話があったが、末端の詳細設計のレベルでの最適化を行っても、全く別の製法に対しては通用しないので、プロセスの抽象化が必要だと私も考えている。

仮に、プロセスについて前もって逆解析で検討しようとしても、その組合せは非常に大きな掛け合わせになるので、計算し尽くすことはなかなか難しい。

化学、素材においては、合成したものの物性値、状態の測定なども、一つ一つがなかなかの時間を必要とする。その時間をいかに短縮するかも課題である。さらに、必要な物性値も1個ではなく、10個程度になることはよくある。また、毒性や安全性といった項目もクリアしなければならない。もちろんコストも下げなければならない。このように多目的最適化としての性質を非常に強く持っている。

概念設計ができたとしても、スケールアップや実運転において安定的に生産できるかという面も、プロセス開発の段階から見積もっておかなければいけない。

インフォマティクスへの期待としては、分岐を手際よく絞り込む予測技術と、個別の検討用設備の高度化による高速化が挙げられる。

日本触媒社内でも取り組んできたものの中から、プロセス・インフォマティクスに関連する例をいくつか紹介する。材料としてもモノマー、ポリマー、触媒など様々なことに取り組んできた。ラボスケールから実機、バーチャル空間での最適化などに取り組んできたが、ラボスケールの条件最適化では、手法としてはベイズ最適化を使ってうまくいくことが多い。

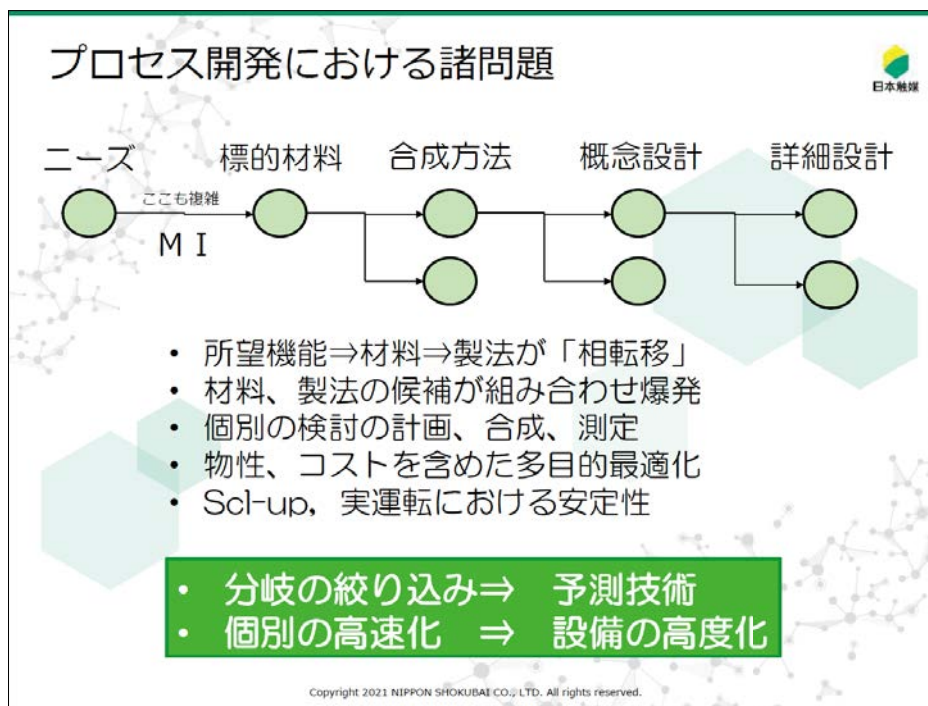


図4-3-2 プロセス開発における諸問題

例えば、アクリル系モノマーの最適化のために、データを集めて深層学習をした例がある。深層学習によって分子構造から物性値、また毒性を予測する、さらに逆解析によって、モノマー分子構造を出すということを検討し、うまくいった。合成ルートの予測も検討したところ、予備検討段階であり大規模データベースを使用しなかったのが論文にはしていないが、有機化学を学び終えた学部卒レベルの予測精度は習得できたと考えている。

実際のプロセス設計に関しても、いろいろ工夫をして取り組んでいる。よく「スモールデータ」と言われるが、そのスモールな状況でどうするかが重要である。実測値や理論計算値だけがデータではなく、実際、何でもデータだと私個人は思っている。それをつなげるのが、ベイズ推定やアノテーションというデータ取得技術である。ある多官能モノマーの脱水プロセスの例では、圧力、温度、フィード量、組成や触媒槽長などについて、それぞれがどの程度効きそうかを実験担当の研究者に聞いた。例えば圧力を1Pa上げたらどうかといったことを。そのときに「ちょっと収率が上がるんじゃないかな、でもあまり自信ないな。20%ぐらいの確信度」というような回答も一つのデータとしてモデリングに使うことができる。実測値だけでなく、直感・主観も取り入れることができるのが今の機械学習技術である。

機械学習と理論を組み合わせた検討例もある。機械学習は帰納的なモデルであり、データが必要。一方で、微分方程式、反応方程式は演繹的であり、パラメータが必要。お互いの弱点を補うため、この両者を連成させたモデルをつくり、文献データを主に使って触媒の良好な元素と組成を特定できたという事例がある。データを組み合わせることができるし、ほぼ実験をせずに予測ができる。

計測値の利用の例としては、各種ポリマー製品、モノマー製品、固体触媒など種々の材料について計測値と物性値を溜めて機械学習することによって、スペクトルデータから品質を予測することができる。ラボでモデルを作っておけば、そのモデルを実機に適用することで同じ品質のものが生産できる。測定の省力化にも利用できる。押出機的设计と運転条件の最適化の例もある。物性値とその処理量、コストを最適化するのだが、通常は実機サイズでは一発勝負となり、あまりデータが取れない。したがって、その外挿性に関して非常に慎重にならざるを得ない。このようなケースで、ラボや過去の実機のデータによる転移学習を利用して、ラボと実機で共通な装置特性と物性値などの関係性を、階層の深いところで学習して正確に予測するという技術も

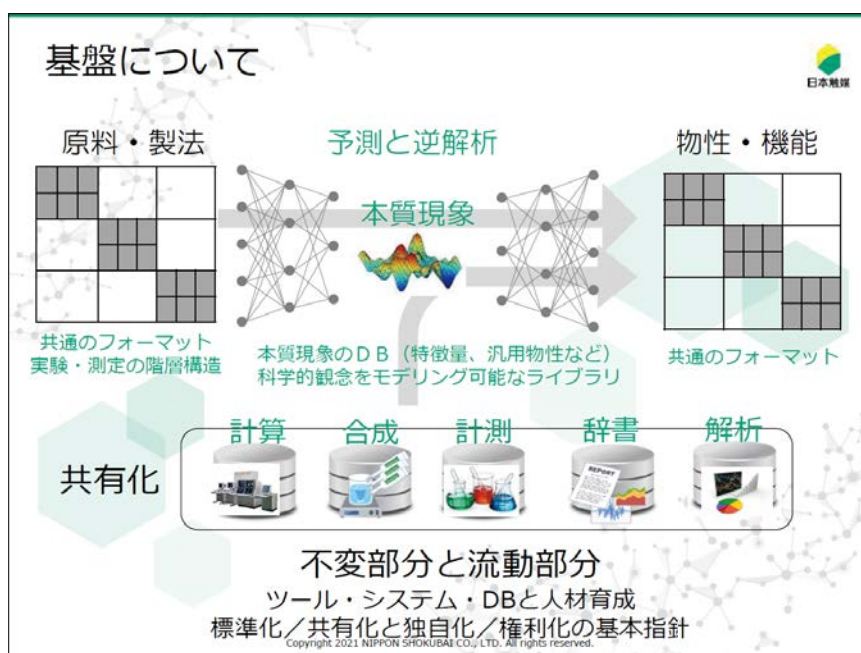


図4-3-3 プロセス・インフォマティクス基盤について

実例としてある。

プロセス開発における諸問題というのは、相転移が起こること、組合せ爆発が起こること、個別の時間がかかること、多目的最適化であること、また実生産のことも考えなければいけないことなど、諸々の難しさがある。しかし、基本的なフレームワークは、まずパラメータを定義して、実験計画法によってデータを満遍なく取り、ベイズ最適化と高速実験、計算、計測を統合したモデルをつくり、ベイズ最適化を回すということである。この際の枠組みは人間が設計するしかない。本質を捉えたモデルの設計にしないと、データがたくさんあってもなかなかうまくいかないことも確かである。

最後に、プロセス・インフォマティクス基盤についての私案を述べる。

個別企業の内部の個別のテーマでは、原料や製法などのデータは全くフォーマットが異なるので、無理やり1つのデータセットにしても穴ぼこだらけになる。図4-3-3では、グレーの固まりが1つのテーマというイメージであるが、物性や機能も対応したところだけが埋まり、それ以外はデータがないということになる。したがって、スモールデータの解析しかできないというのが現状である。しかし、本質現象に相当する中間層（ニューラルネットワークで言う中間層）を設け、そこでは関係性があるはずだという思想によりつなげることで予測も逆解析も可能になる。したがって、本質現象に関連するものは、共通のフォーマットを使用することが重要である。また、本質現象に関する特徴量、汎用物性等のデータベースは、国として基盤のデータベースをつくり、シェアすることが大事なのではないかと考える。

また、科学者の経験や習慣などの科学的観念をモデリング可能なライブラリを整備するのもいいのではないか。その際、辞書、オントロジーというか、共通言語、階層関係、包含関係などを整備して、その上でデータフォーマットを共有化することが大事になる。

【畑中 美穂（慶應義塾大）】

マテリアルズ・インフォマティクス（MI）とプロセス・インフォマティクス（PI）は切り離せないということと、アカデミアの役割はデータをつくり公開することであるということ、まず、私のメッセージとして伝えたい。

自動実験をする際、フラスコの中で反応させるバッチシステムより、カラム中を流して合成していくフローシ

システムのほうが、合成も測定も自動化しやすいというメリットがある。しかし、化合物を合成する際、反応剤だけを入れたら生成物が出てくるということはほとんどなく、いろいろな触媒や添加剤を使うことが多い。しかも、触媒や添加剤の方が価格が高い。バッチシステムでは少量添加で効果が得られることが多いが、フローシステムで一緒に流してしまうとなると、高価な触媒や添加剤を垂れ流すことになるので、壁面などに触媒・添加剤から生じる活性種を埋め込んでおくような設計が必要になる。

ところが、どういう活性種、触媒とか添加剤を埋め込んでおいたらいいかは、合成したい化合物によって違う。作りたい化合物が違えばそのたびにカラムの中に入れておくべき活性種も違う。プロセスとしてもカラム長や流量、流速などたくさんのパラメータを一度に最適化しないといけないので、組合せ爆発が起きてしまう。したがって、カラムに入れておく活性種を簡単に探せるようにすることは、PIを加速するために大切である。活性種を見つけるということは「材料を探す」ことであるので、まさにMIである。こういう意味で、PIとMIは協働しなくてはならない、切り離せないと考えられる。

実際、触媒がどういう活性種をつくって反応しているかは計算化学を使って調べることもできる。例えば、オレフィンに二酸化炭素と水素を反応させてカルボン酸を作る反応では、触媒、添加剤が計4種類必要である。触媒、添加剤それぞれがどういう役割をしているのか計算化学を使って調べようとすると、複雑な反応サイクルの計算に1、2年かかってしまうことがあるが、これでは実験屋は待ってられない。実験コストよりも計算コストが低くなるような状況にならないと、計算で新しい触媒の設計をすることにはならない。計算化学に頼るのではなく、機械学習を使って触媒性能を予測するシステムが必要である。

機械学習によって触媒性能を予測するためには、触媒を数値で表現することが必要であり、これが結構難しい。分子性触媒の配位子1つをとっても構造パターンは非常に多い。しかし、幅広い配位子に対して共通して使える数値パラメータは少なく、触媒をうまく数値表現することが難しい。

しかし、遷移金属錯体を触媒とする反応に着目すると、教科書に載っているような素反応の組合せで出ていることが分かる。この素反応は、たかだか十数種類しかない。どんなに異なる触媒反応でも、分解すれば十数種類の素反応の集まりで記述が可能である。触媒は、図4-4-1に示すように各反応のエネルギーの高いところ、遷移状態（TS）のエネルギーを下げる役割をしている。下がったときのTSのエネルギーは、どんな骨格を持つ触媒にも共通して使える数値表現になるはずである。異なる骨格を持つ触媒を素反応のエネル

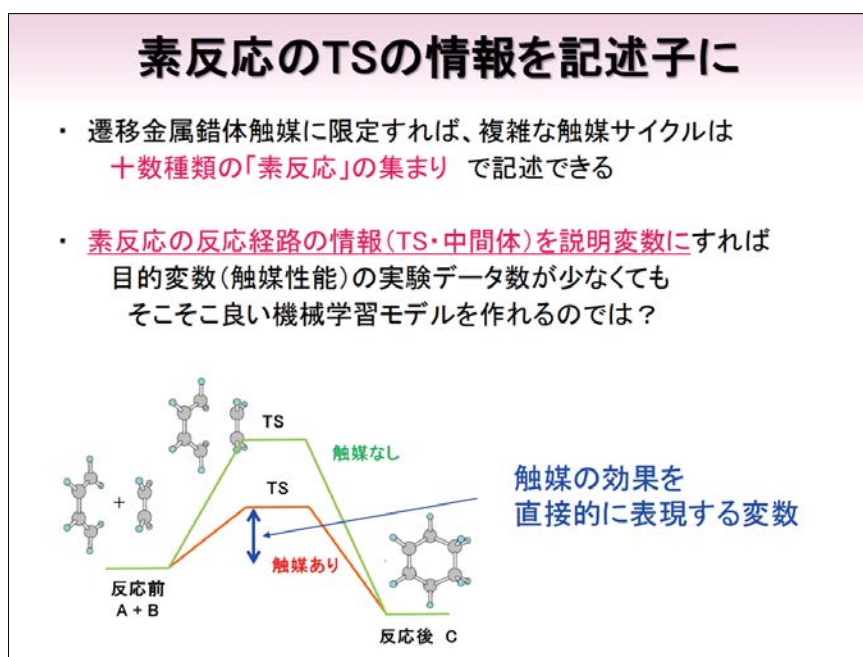


図4-4-1 素反応の遷移状態（TS）の情報を記述子に

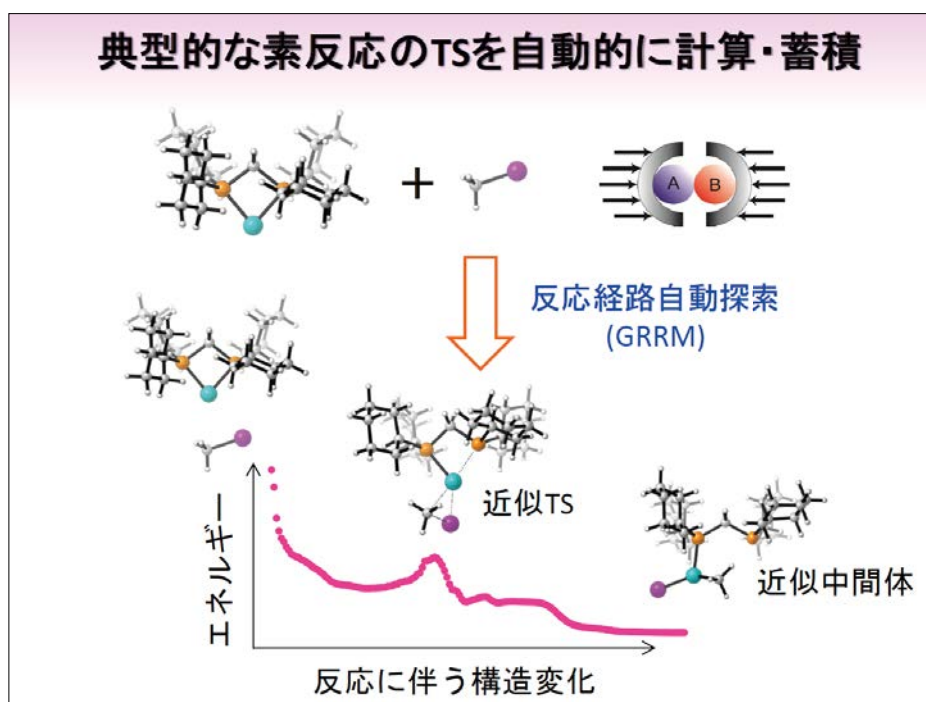


図4-4-2 反応経路自動探索 (GRRM)

ギーで数値表現し、触媒性能を予測するためのモデルづくりに使うことができる。

この素反応のエネルギーは北大WPIで検討している反応経路自動探索 (GRRM、図4-4-2) を使うと自動的に出すことができる。これによって触媒の素反応のデータベースを一度作っておく。図4-4-3に示すように、いろいろな実験グループからターゲットとしての反応の実験データを提供してもらえれば、計算データと実験データの間の機械学習モデルをつくり、我々のデータベースから適切な触媒を提案することができる。その候

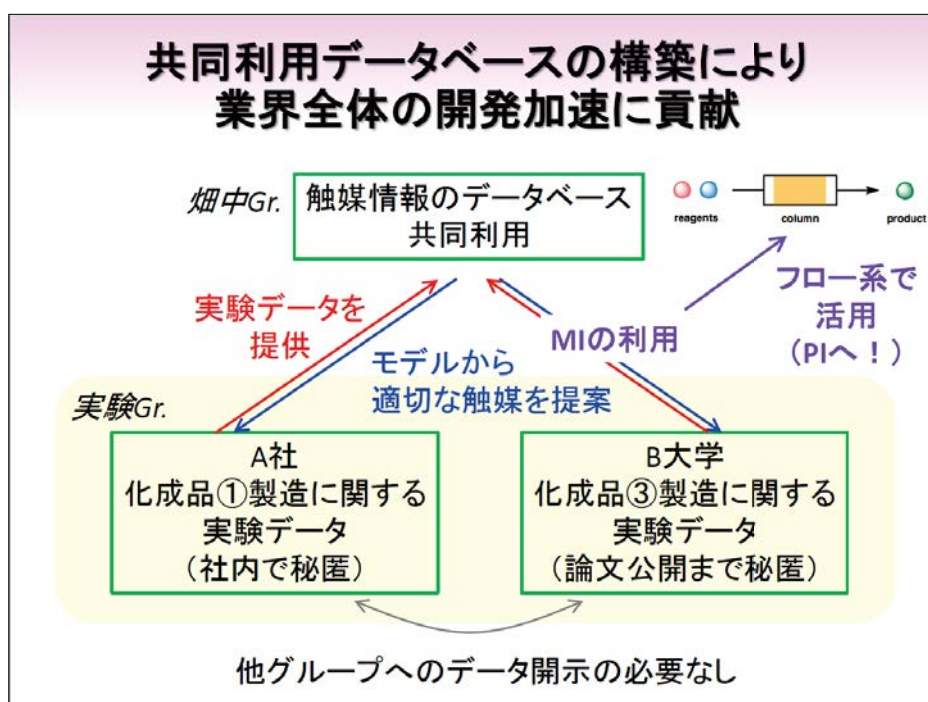


図4-4-3 触媒情報の共同利用データベース

補触媒をフローシステムの実験に使うことで、少ない実験データで触媒の選択ができるようになると考えている。

我々のデータベースには計算データしか入っていないので、競合する実験を行っているいろいろなグループが共通して使ったとしても、彼らの間で実験データをお互いに開示する必要がないというのがいい点である。競合グループが共通して使えるプラットフォームのような形であるので、業界全体の開発加速に貢献できるものと考えている。この例は、計算化学がプロセス・インフォマティクスに貢献できるよい例であり、アカデミアがデータを公開できる強みを生かしているものである。

【柳 日馨（大阪府立大）】

有機合成の立場から物づくりに携わってきた立場から、考えていることを共有したい。世界的には、反応データベースはすでにできているといえる。ただし、そのデータベースで個々の反応の評価はできていない状況である。言い換えると、地図はできたが、あくまでも平面地図で、三次元的になっていないという状況に似ている。地図によって目的地が見えるようになったが、高さの状況が分からないので、目的地へ行くまでに高い山を越えなければいけないのか、谷底に降りなければいけないのかというような情報が不足している。ここに化学情報戦略で取り組むべき重要な課題が潜んでいる。つまり、ものの作り方において、三次元地図によりどういう経路を辿って目的地に一早く至るのかという比喻と同じ事が求められている。質の高い反応情報を素早く得ることが優れた機能性化合物に先駆けて到達する肝となるに違いない。そのため、遷移状態計算が重要であると考えている。遷移状態計算によって「山の高さ」が分かり、反応への質的評価、いわば反応の格付けができるようになる。平面地図は欧米主導で完成されている状況だが、そこから立体地図へ持っていくことで日本は世界で先進的なポジションを取れるチャンスがある。

ものづくりをする際には、未知の道、見つけられていない道がまだあり、そこにも目を配る必要がある。これは優れた研究者個人のスキルや勘に関係してきた。しかし、これからは勘に頼るのではなく、遷移状態計算を基軸とした情報を蓄積し活用していく方向に移行するべきと思われる。そしてそのための準備が今われわれには求められている。

もう一つ期待したいのは、高速化を目指した有機合成の刷新である。有機合成をする際に、狙ったものができているかを確認するために、昔は最低でも数gが必要だと言われていたが、今は数mgでもよくなっている。これは有機合成が進歩したこともあろうが、むしろ測定機器の進歩によるものが大きい。合成量はあまり気にされていない感があるが、これからは機能や活性評価に必要な最小量を作るという方向に進むのではないか。廃棄コストのかかる化合物を蓄える必要もない。一方、反応で用いるバッチフラスコを小さくしていても人が扱えるには限界がある。バッチフラスコは実験しやすいから使われてきたが、いわばこれは人間の都合である。そこで以前から取り組んできたマイクロ空間での反応に改めて着目してみた。マイクロバッチすなわちドロップレット1滴で反応を実施し情報を取ることができれば非常に速く結果を得る事ができる。個々のマイクロ空間の反応は、条件を変えることで、それぞれを反応装置とした反応条件のスクリーニングができるであろうし、いろいろな誘導体や異なる試薬を入れて大規模なパラレル合成を行うことも可能であろう。マイクロ空間とはいえ、評価に必要な量ぐらいいは合成できる。マイクロ空間を反応空間として駆使する上で自動化技術との融合は欠かせないし（手動では到底不可能）、自動化した微細空間反応は高速化の大きな武器になると期待している。

私が問題提起したかったことをまとめる。有機合成化学者の優れたスキルや勘に依存する必要のない「立体的」な反応データベースを計算化学と有機合成がともに盛んな日本でつくってほしいこと。遷移状態計算の検証に微細空間を活用した高速反応技術を活用していくこと。そして微細空間反応側から計算化学のもう片方を見ればそこには連続生産が存在している。プロセス・インフォマティクスを経て機能性物質の高速開発・製造のためのインターフェースとしても超微量合成の大切な役割があると思う。

プロセス・インフォマティクスに向けて

世界で先進的なポジションを取るために

有機合成の超高速化をはかる

機能評価に必要な最小量合成を追求、候補品合成数は数多く、廃棄物を出さず、超高速で機能性物質へ到達 遷移状態計算の多用で最適系へ

『一滴の血液で病気の診断』
を可能とする時代

『一滴の化合物で評価を』



マイクロ空間反応の活用

高速化には開発研究プロセスでの「改善」や「効率化」ではなく、開発様式そのものを「刷新」することが肝要 ロボットやAI支援を利活用

計算化学の活用

ほとんどの無駄な実験を省く

機械学習とルーチン化させた遷移状態計算が関与するメソッド開発
玉石混交の化学反応データベースのハイグレード化

図 4-5-1 物質創製において世界で先進的なポジションを取るために

超微量合成からプロセス・インフォマティクスへのショートカット戦略

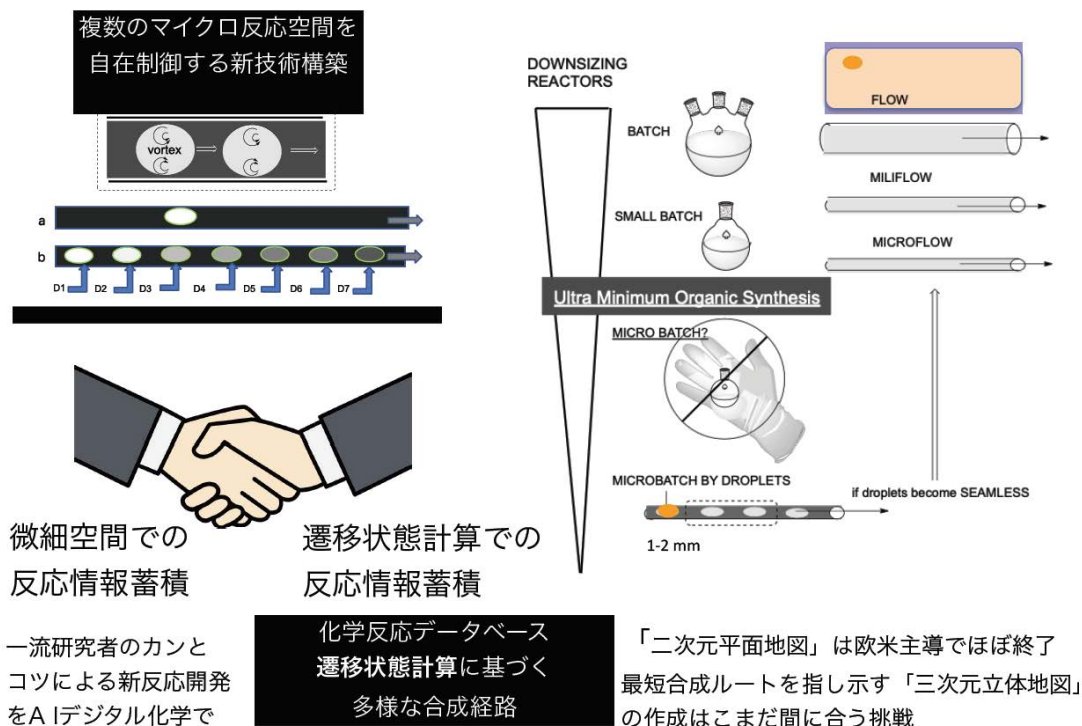


図 4-5-2 超微量合成からプロセス・インフォマティクスへのショートカット戦略

【ファシリテーターによるトーンセッティング】

本日の発表者・コメンテーターの皆さんには、事前にアンケートをお願いし、充実した回答をいただき大変感謝している。総合討論では、本日の各発表の内容、コメンテーターの方々のご意見、アンケート結果なども踏まえて議論を進めたい。

CRDSでは、現時点の仮説として、プロセス・インフォマティクス共通基盤を構築し、その上に個別プロセスを位置づける仕組みがよいのではないかと考えている。また、個別プロセスとしては、領域の柱になるようなプロセスをいくつか選ぶのがよいと考えている。この仮説に対して事前アンケートでは、まず個別プロセスに対して実績を積み上げ、その過程で得られた成果から横展開できるものを発展させて共通基盤にするのがよいという意見が多かった。

とくに化学反応データベースやシミュレーション、モデリング技術は、多くのプロセスに対して使えるので、共通基盤として整備しておく必要があり、ツールやデータの共通化・標準化も必要である。

また、本日、「プロセスの抽象化」という考え方も提案された。個々のプロセスから共通項を見出していわばプロセスを抽象化し、抽象化したものを共通基盤としてとらえて、インフォマティクスに活用するという考え方である。

ここからの総合討論では

- ①プロセス・インフォマティクス共通基盤として取り組むべき研究開発課題、研究開発体制、制度面で検討すべき事項、参画が望まれる研究者・技術者の専門分野
 - ②個別対象プロセスとして取り上げるべきプロセス。プロセスセンターにおける人材・技術・装置、制度・仕組み。参画が望まれる研究者・技術者の専門分野
 - ③アカデミアと産業界の連携の仕組み
 - ④その他、コミュニティ形成、人材育成、海外連携などについて
- など様々な視点で討論したい。

以下に総合討論で指摘された重要な事柄を項目別に整理して示す。

【人間の勘・コツ、匠の技とデータ科学】

- ・実際にプロセスをよく知っている人間の勘・コツ的なものをプロセス・インフォマティクスの共通基盤にどう落とし込むのが重要である。
- ・人間は様々なプロセスデータと特性を関係づけて自らの勘・コツとしているはずであるから、そうしたものを分析して目的関数化するようなことができるのではないか。人間が感じていることを数値化することができれば、プロセスデータと勘・コツの関係を機械学習することができるようになる。
- ・匠の技のデジタル化のようなアプローチは大変興味深い。
- ・機械学習では、Human-in-the-loop（人間参加型機械学習）と言われる考え方がある。デジタル化で自動的にできるのはあるところまでで、そこから先は人間が関与してやるべきだという議論である。
- ・日本の強みとしては、人の勘・コツをどう生かすかを重視するべきである。プロセスのデジタル化は大変なことではあるが実施し、共通基盤化していくべき。
- ・勘や経験を持つ専門家、とくに実験研究者にマテリアルズ・インフォマティクスやプロセス・インフォマティクスを理解してもらうことが必要。機械学習も説明変数と目的変数を結び付ける場面では勘とコツが必要になる。
- ・勘やコツはDXで追い抜かれることをデジタル家電で経験した。アナログ家電時代はノウハウで競争していたが、今はデジタル技術で凌駕されている。他分野でも将来そうなる可能性を考え、我々の強みをしっかり守っていくことが重要だと思う。そこを考慮した基盤をつくり、みんなの知恵を守れるようにしていくことが重要である。

- ・勘とコツもデータ科学で強化できるということを理解すること。熟練者は予備知識があるので、目的変数と因果関係がある説明変数を選ぶことができる。そうした知見をもとに機械学習の一手法である転移学習を使って別のプロセスにアプローチできる。

〔意識改革の必要性〕

- ・共通基盤の目的は、意識改革とプラットフォーム化が2つある。従来の共通基盤はプラットフォームとして箱物や仕組のような具体的なものをつくるのがメインだったと思うが、実際にはむしろ意識改革が大切である。プロトコルをつくりそれに従えばそれでうまく行くという話ではなく、研究者が自ら勘やコツのデジタル化や暗黙知の形式知化に取り組むような意識改革が大切である。
- ・共通基盤での活動としては、意識改革を目的としたワークショップ実施等も必要である。
- ・意識改革を進めるために、分かりやすい成功例を有機材料、無機材料、複合材料でそれぞれ1つずつ作ることがよい。無機材料の例を有機の人がすぐには理解できないだろうし、逆もしかり。成功事例をつかって、何度もワークショップで啓蒙していくことが一番大事な共通基盤だと考える。そういう活動によって、共通基盤として本当に必要なものの情報が集まってくるようになると思う。

〔人材育成〕

- ・プロセスの研究においても実際に研究に従事する若い人たちが頭脳循環のような形で別の研究をすることが必要である。いろいろなバックグラウンドを持った上でそれを積み上げてテーマ設定などをすることが大事。材料研究者が機械学習を理解して新たな視点でテーマ設定をする、あるいは、データ科学の研究者に材料の基礎知識・感覚を身につけさせるなど、両方を融合させることが大切。材料研究者とデータ科学研究者の単なる共同研究ではなく、両方の視点を持った若手人材の育成がこれから重要になってくる。

〔スケールアップ〕

- ・ラボレベルのプロセスはインフォマティクスを活用してうまく設計できても、スケールアップできないことがある。量産を視野に入れたプロセス・インフォマティクスを考えるべきではないか。
- ・アカデミアがスケールアップにまで取り組むかは、非常に悩ましい問題である。むしろ、アカデミアは小スケールにこだわり、できるだけ理想的な反応空間で合成し、スケールアップに必要な情報を得るようにすべきではないか。合成プロセスの科学をしっかりと作ることによって、その後の製品化の開発が速くなり、研究室で見つけたものが世に出るまでの期間が圧倒的に短縮されることを狙いたい。
- ・合成プロセスの根本的なメカニズムが明らかになれば、量産体制に持っていくことは産業界でできる。「○○℃で○○時間焼成し、次に……」というような表層的なHowだけを提示されてもスケールアップはできない。分子レベルで本当に何がおきているのか、なぜ結晶質/非晶質が合成されるのかということが明らかになれば、スケールアップ手法をつくることができる。
- ・私はアカデミアの研究者であるが、企業との共同研究との経験からは、基礎研究と企業の研究の区別はなく、ほとんど連続していると感じている。単なる表層的なHowだけではダメだというのは賛成だが、スケールアップの際に新たな科学的知見が必要になることもある。プロセス・インフォマティクスの検討を機会に、ベンチャー企業の仕組みを使い、産学連携はもっとシームレスかつ一体的になることを始めたほうがいいのではないか。
- ・産学連携や共同研究を進める際、アカデミア側が一定量のサンプルを用意できるようになっていると話が圧倒的に早く進む。
- ・有機合成プロセスでは、スケールアップを考える際、有機合成化学と化学工学を同時に検討する必要がある。反応収率が必ずしも高くなくても安定していれば、化学工学的手法で安定生産に繋がられる。

- ・ 化学反応については、素反応のネットワークが分かれば、データ同化や物理化学現象のシミュレーションなどを活用してスケールアップが可能だと思う。有機合成では化学反応を理解することが一番の本質である。
- ・ 化学反応プロセスと、結晶成長や粉体プロセスは、だいぶ異なるので、そこは分けて考えたほうがいい。
- ・ 無機材料のプロセスでは、表面反応や固体内拡散など、上流から下流までのマルチフィジックスのシミュレーションが重要である。気相、気固界面、固体中それぞれで何が起きているかを全部つなげて研究することが必要である。さらにマルチスケールにつなげられるプラットフォームをつくれば、スケールアップにもつながる。原子レベルから数cm程度の領域、ダイナミックレンジで 10^8 ぐらいのスケールをいかにシームレスにつなぐかが重要な基盤技術となる。
- ・ マイクロリアクターのように反応律速な状態でモノづくりする際は問題ないが、スケールアップによって反応律速から拡散律速に変わるようなケースもあるので、反応だけを理解すればよいということにはならない。やはりプロセス全体をシミュレーションするような目的を持った、合成プロセスのデジタルツインのようなものが必要だと思う。

[データについて]

- ・ 異なる由来のデータを混ぜて使うことも必要になる。
- ・ データの基準・ベースがずれていても、まずデータが蓄積されないとデータ科学を適用することが出来ない。そのため、ベースがずれている、装置が異なるなどの理由により単純にデータ混合してはいけないと思われる場合でも、まずはデータを収集・蓄積することが必要である。ある程度のデータ量があればクラスタリングをすることでパターンを見つけ出すことができることもある。完全にランダムではなく、例えば3パターンぐらいに分かれることがわかると、なぜ分かれているかを分析することも可能となる。さらに、最近ではデータの質を判別できるようなアルゴリズムも開発されているので、最初に思い込みをせずにどんどん集めていくのが大切である。
- ・ 高精度測定では1測定当たりの時間が長くなる傾向にあるが、新しい計測技術によって短時間化を図ることができれば、データ収集がしやすくなる。また、機械学習には、ネガティブデータも重要であるが、今はほとんど収集されていない。計測技術の進歩を組み合わせることでデータ収集が容易になれば、自然と集まってくるようになるだろう。その蓄積の結果、信頼のあるデータベースになれば、みんな使うようになり、さらにデータが集まるという好循環に入ると思う。さらに、基板ホルダーの共通化などを進めることによって、プロセスデータをシェアしやすくするような仕組みも必要である。

【まとめ】

総合討論では、CRDSの材料創製技術を革新するプロセス・インフォマティクスに関する問題意識と研究開発課題に関する仮説、および、有識者からの話題提供、事前アンケートの結果を踏まえ、様々な視点で議論し、以下のような重要な視点を得ることができた。

○人間の勘・コツ、匠の技

- ・人間は様々なプロセスデータと特性を関係づけることで自らの勘やコツを会得している。
- ・熟練者は予備知識があるので、機械学習を設計する上で、目的変数と因果関係がある説明変数を選ぶことができる。こうした結果を活用すると、自らの勘とコツもデータ科学でさらに強化できる。これは一種のHuman-in-the-loop（人間参加型機械学習）であり、勘やコツを生かすスキームとなりうる。
- ・日本の強みであるプロセス技術の発展・強化には研究者の勘・コツを生かす仕組みを推進するべきである。その一つにはプロセスのデジタル化を推進し、共通基盤化していくことが挙げられる。

○スケールアップ

- ・産業界で量産プロセスを構築するには合成プロセスの根本的なメカニズムが明らかになっていることが必要である。
- ・アカデミアは、理想的な反応空間で、スケールアップに必要な情報を得る研究に重点を置き、合成プロセスの科学を構築することが必要である。
- ・アカデミアでの基礎研究、企業での実用化研究をもっとシームレスにするべきである。スケールアップに必要な実用化研究から重要な科学的課題が顕在化し、アカデミアでの基礎研究が引き起こされることもある。
- ・プロセス全体をシミュレーションしうるデジタルツイン構築が重要である。その際、有機合成・触媒反応では化学反応の素過程の本質的理解が必須である一方、無機結晶成長ではプロセス全体のマルチフィジックスのシミュレーションが重要な課題である。

○意識改革の必要性

- ・勘やコツ、ノウハウに関する研究者の意識の変革が重要である。
- ・意識改革に有効なのは、データ科学を用いたプロセス技術の分かりやすい成功例を有機材料、無機材料、複合材料でそれぞれの分野で作ることであり、その成功事例によって啓蒙していくことが肝要である。

○データについて

- ・プロセスにデータ科学を活用するには、まずデータを集めることが必要である。一定の条件で実施された結果でなくても、実験条件等のデータを付加情報として収集・蓄積することが、データ活用のためには重要である。また、基板ホルダーの共通化、前処理やデータ登録の自動化などによりデータ収集の効率化を推進することで、機械学習モデルを構築する上で重要な、予想・仮説とは異なるネガティブデータなども自然に集まってくるようになることが期待される。

○人材育成

- ・新材料を創成するプロセスの研究に従事する若い人たちが頭脳循環のような形で複数の研究に従事する経験を積ませることが必要である。
- ・材料研究者とデータ科学者の共同研究は重要であるが、これをさらに深化させるには、両方の視点を持った若手人材の育成が重要である。

以上

付録

付録1 開催趣旨・プログラム

開催趣旨

JST 研究開発戦略センター（CRDS）は、科学技術に求められる社会的・経済的ニーズを踏まえて国として重点的に推進すべき研究領域や課題を俯瞰の視点から系統的に抽出し、その研究開発戦略を提言する公的シンクタンクとして活動しています。活動の過程において、研究開発動向の俯瞰を目的とした「俯瞰ワークショップ」や、重要テーマの研究開発戦略を議論する「科学技術未来戦略ワークショップ」などを開催してきました。本ワークショップ「材料創製技術を革新するプロセス・インフォマティクス」は、上記「科学技術未来戦略ワークショップ」の一つとして開催するものです。

我が国の貿易収支を支える製造業において、競争力の源泉は、高性能な機器や素材の製造を可能にする材料技術・ナノテクノロジーにあると言えます。しかし、近年、材料・デバイスの高性能化・高機能化に対する要求や期待がますます高まってきており、従来の手法だけでは求める機能を有する物質・材料を創出することが難しくなっています。そのため、物質・材料創製研究において新しい科学的手法の導入の必要性が高まっています。

マテリアルズ・インフォマティクスは、特定機能を有する物質・材料を効率的に「設計」する方法として様々な成果を生み出している一方で、「設計」された物質・材料を合成する方法を見つけられず、その効果を実証できていないケースも散見されます。これは、従来のマテリアルズ・インフォマティクスでは「いかに作るか」に踏む込む視点が充分ではないことに起因すると考えられます。

また、CRDS 戦略プロポーザル「未来材料開拓イニシアチブ～多様な安定相のエンジニアリング～」(2019年発行)では、多元素化による材料設計の自由度を増大や、準安定相や不安定相を含むこれまで未利用だった多様な安定相を活用することを提言していますが、このためにも材料の合成プロセスの探索空間を拡げるための新しい科学的手法を導入することが必要です。

このため、物質・材料創製研究において、実験科学、理論科学、計算科学、データ科学の4つの科学をフル活用して目的材料の合成プロセスを効率的に探索する科学的手法であるプロセス・インフォマティクスが必要になるとの仮説の下、本ワークショップでは、プロセス・インフォマティクス技術基盤を構築し、技術的優位性を確保し続けるための研究開発戦略について議論します。

本ワークショップでは、「プロセス・インフォマティクス共通基盤を支える技術群」「個別プロセスのインフォマティクス活用研究」に関連する我が国の現状・課題・将来展望についての話題提供を基に、戦略的に取り組むべき重要な研究開発課題、世界をリードするための効率的・効果的な研究実施体制および産学連携の仕組み、人材育成・社会受容などを含む社会的・経済的インパクトについて議論します。これらの議論を通して材料創製技術を革新するプロセス・インフォマティクスを実現するための研究開発の方向性を明らかにし、取り組むべき科学技術とその推進方法の共通認識を得たいと考えています。

プログラム

(敬称略)

開催日時：2021年1月24日（日）9：30～17：00

開催会場：オンライン会議（Zoom）にて開催

（JST関係者+希望者）TKP市ヶ谷カンファレンスセンター バンケットホール8A

司会 福井 弘行（JST-CRDS）

9：30～9：35 開会挨拶 曾根 純一（JST-CRDS）

9：35～9：50 ワークショップの趣旨説明 福井 弘行（JST-CRDS）

9：50～12：10 話題提供①個別プロセスにおけるインフォマティクス活用研究

9：50～10：15 結晶成長デジタルツインに向けた取り組み（SiC溶液成長を中心に）

宇治原 徹（名古屋大学）

10：15～10：40 ロボット-機械学習による全自動・自律的な無機機能性薄膜合成の現状と展望

清水 亮太（東京工業大学）

10：55～11：20 精密有機合成とプロセス・インフォマティクス

甲村 長利（産業技術総合研究所）

11：20～11：45 粉体プロセス開発のハイスループット化のためのプロセス・インフォマティクスと、
材料探索ハイスループット化との関係

長藤 圭介（東京大学）

11：45～12：10 データ駆動型高分子材料開発とプロセス・インフォマティクス

内藤 昌信（物質・材料研究機構）

13：15～14：55 話題提供②プロセス・インフォマティクスに必要な技術群

13：15～13：40 プロセス・インフォマティクスにおける技術紹介および課題と将来展望

金子 弘昌（明治大学）

13：40～14：05 窒化物半導体 MOVPE プロセスのマルチフィジックス結晶成長シミュレーション

寒川 義裕（九州大学）

14：05～14：30 触媒インフォマティクスの実践に見るプロセスの重要性

谷池 俊明（北陸先端科学大学院大学）

14：30～14：55 計測インフォマティクスを目的とした計測スペクトルのスパースモデリングの自動化

吉川 英樹（物質・材料研究機構）

15：15～16：55 総合討論 ファシリテーター

伊藤 聡（JST-CRDS）

コメンテーター

畑中 美穂（慶應義塾大学）

水野 洋（パナソニック）

富谷 茂隆（ソニー）

右田 啓哉（日本触媒）

柳 日馨（大阪府立大学）

【論点】1.「PI 共通基盤構築研究」の拠点で取り組むべき研究開発課題。研究開発体制、制度面などで検討すべき事項、参画が望まれる研究者・技術者の専門分野。

2.「個別対象プロセス」として取り上げるべきプロセス。プロセスセンターに用意すべき人材・技術・装置、制度・仕組み。参画が望まれる研究者・技術者の専門分野。

3.アカデミアと産業界の連携の仕組み

4.その他（コミュニティ形成、人材育成、海外連携）

16：55～17：00 閉会挨拶

曾根 純一（JST-CRDS）

付録2 参加者一覧

(敬称略、所属・役職はワークショップ開催日時点のもの)

招聘識者

(発表者)

- ・宇治原 徹 名古屋大学 未来材料・システム研究所 教授
- ・金子 弘昌 明治大学 理工学部 准教授
- ・寒川 義裕 九州大学 応用力学研究所 教授
- ・甲村 長利 産業技術総合研究所 触媒化学融合研究センター フロー化学チーム チーム長
- ・清水 亮太 東京工業大学 物質理工学院 准教授
- ・谷池 俊明 北陸先端科学技術大学院大学 先端科学技術研究科 教授
- ・内藤 昌信 物質・材料研究機構 データ駆動高分子設計グループ グループリーダー
- ・長藤 圭介 東京大学大学院 工学系研究科 機械工学専攻 准教授
- ・吉川 英樹 物質・材料研究機構 材料データプラットフォームセンター 副センター長

(コメンテーター)

- ・富谷 茂隆 ソニー株式会社 R&Dセンター TL27 (Materials Analysis) 主幹研究員
- ・畑中 美穂 慶應義塾大学 理工学部化学科 准教授
- ・右田 啓哉 株式会社日本触媒 データサイエンス&インフォマティクス推進室 室長
- ・水野 洋 パナソニック株式会社 マテリアル応用技術センター デジタル融合総括
- ・柳 日馨 大阪府立大学 研究推進機構 特認教授

JST-CRDS 検討チームメンバー

- ・曾根 純一 ナノテクノロジー・材料ユニット 上席フェロー
- ・福井 弘行 ナノテクノロジー・材料ユニット フェロー/チームリーダー
- ・伊藤 聡 ナノテクノロジー・材料ユニット 特任フェロー、計算科学振興財団 チーフコーディネーター
- ・酒部 健一 戦略研究推進部グリーンイノベーショングループ 主任調査員
- ・佐藤 勝昭 ナノテクノロジー・材料ユニット 特任フェロー、東京農工大学 名誉教授
- ・竹内 良昭 環境エネルギーユニット フェロー
- ・沼澤 修平 ナノテクノロジー・材料ユニット フェロー
- ・的場 正憲 システム・情報科学技術ユニット フェロー
- ・眞子 隆志 ナノテクノロジー・材料ユニット フェロー
- ・水本 邦彦 研究プロジェクト推進部 グリーンイノベーショングループ 主任調査員
- ・宮下 哲 ナノテクノロジー・材料ユニット フェロー

関係府省・機関等

- ・馬場 大輔 内閣府 政策統括官（科学技術・イノベーション担当）付 企画官
- ・増田 幸一郎 内閣府 政策統括官（科学技術・イノベーション担当）付 政策企画調査官
- ・小川 浩司 文部科学省 研究振興局参事官（ナノテクノロジー・物質・材料担当）付 課長補佐
- ・夏目 和人 文部科学省 研究振興局参事官（ナノテクノロジー・物質・材料担当）付 ナノテクノロジー・材料企画係長
- ・中尾 航 文部科学省 研究振興局参事官（ナノテクノロジー・物質・材料担当）付 学術調査官
- ・鷲見 昭英 経済産業省 産業技術環境局 研究開発課 産業技術プロジェクト推進室 室長
- ・富樫 達也 経済産業省 産業技術環境局 研究開発課 産業技術プロジェクト推進室 室長補佐
- ・関口 貴子 経済産業省 産業技術環境局 研究開発課 研究開発専門職
- ・稲垣 彰彦 経済産業省 製造産業局 素材産業課 課長補佐
- ・木村 阿紀子 物質・材料研究機構 経営企画部門 経営戦略室 専門職
- ・堀田 裕司 産業技術総合研究所 材料・化学領域 研究戦略部 研究企画室長
- ・阿多 誠介 産業技術総合研究所 材料・化学領域 研究戦略部 企画主幹
- ・井上 貴仁 産業技術総合研究所 材料・化学領域 研究戦略部 イノベーションコーディネーター
- ・矢田 陽 産業技術総合研究所 触媒化学融合研究センター 主任研究員
- ・吉岡 宏人 新エネルギー・産業技術総合開発機構 材料・ナノテクノロジー部 主査
- ・中村 大輝 新エネルギー・産業技術総合開発機構 材料・ナノテクノロジー部 職員

JST

- ・永野 智己 CRDS 総括ユニットリーダー、JST研究監
- ・赤木 浩 CRDS ナノテクノロジー・材料ユニット フェロー
- ・荒岡 礼 CRDS ナノテクノロジー・材料ユニット フェロー
- ・馬場 寿夫 CRDS ナノテクノロジー・材料ユニット フェロー
- ・渡邊 孝信 CRDS ナノテクノロジー・材料ユニット フェロー
- ・岩本 敏 CRDS ナノテクノロジー・材料ユニット 特任フェロー、東京大学生産技術研究所 教授
- ・本間 格 CRDS ナノテクノロジー・材料ユニット 特任フェロー、東北大学多元物質科学研究所 教授
- ・八巻 徹也 CRDS ナノテクノロジー・材料ユニット 特任フェロー、量子科学技術研究開発機構 高崎量子応用研究所 上席研究員
- ・永井 諭子 研究プロジェクト推進部 副調査役
- ・岡島 博司 未来創造研究開発推進部 共通基盤領域 TM
- ・小関 徳昭 未来創造研究開発推進部 主任調査員

曽根 純一	上席フェロー・総括責任者	(ナノテクノロジー・材料ユニット)
福井 弘行	フェロー・リーダー	(ナノテクノロジー・材料ユニット)
伊藤 聡	特任フェロー	(ナノテクノロジー・材料ユニット)
酒部 健一	主任調査員	(戦略研究推進部)
佐藤 勝昭	特任フェロー	(ナノテクノロジー・材料ユニット)
竹内 良昭	フェロー	(環境・エネルギーユニット)
沼澤 修平	フェロー	(ナノテクノロジー・材料ユニット)
的場 正憲	フェロー	(システム・情報科学技術ユニット)
眞子 隆志	フェロー	(ナノテクノロジー・材料ユニット)
水本 邦彦	主任調査員	(研究プロジェクト推進部)
宮下 哲	フェロー	(ナノテクノロジー・材料ユニット)

科学技術未来戦略ワークショップ報告書

CRDS-FY2020-WR-12

材料創製技術を革新するプロセス・インフォマティクス

令和 3 年 3 月 March 2021

ISBN 978-4-88890-718-7

国立研究開発法人科学技術振興機構 研究開発戦略センター

Center for Research and Development Strategy, Japan Science and Technology Agency

〒102-0076 東京都千代田区五番町7 K's 五番町

電話 03-5214-7481

E-mail crds@jst.go.jp

<https://www.jst.go.jp/crds/>

本書は著作権法等によって著作権が保護された著作物です。

著作権法で認められた場合を除き、本書の全部又は一部を許可無く複写・複製することを禁じます。

引用を行う際は、必ず出典を記述願います。

This publication is protected by copyright law and international treaties.

No part of this publication may be copied or reproduced in any form or by any means without permission of JST, except to the extent permitted by applicable law.

Any quotations must be appropriately acknowledged.

If you wish to copy, reproduce, display or otherwise use this publication, please contact crds@jst.go.jp.

FOR THE FUTURE OF
SCIENCE AND
SOCIETY



<https://www.jst.go.jp/crds/>