

ナノ知識探索プロジェクト：実験記録からの知識発見

Knowledge Exploratory Project for Nano Device Design and Manufacturing: Knowledge Discovery from Experimental Records

吉岡 真治*1
Masaharu YOSHIOKA

富岡 克広*2
Katsuhiko TOMIOKA

原 真二郎*1
Shinjiro HARA

福井 孝志*1
Takashi FUKUI

*1 北海道大学大学院 情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

*2 科学技術振興機構 さきがけ研究者

JST PRESTO

We are developing a framework for knowledge exploratory in nanodevice development based on collaboration between engineers of nanodevices and computer science researchers. Development of nanodevices requires varieties of knowledge; some of this knowledge is tacit, based on the user's experience. Therefore, it is difficult to become a good engineer in this development process. In this paper, we develop an experiment record management system for supporting a nanodevice development experiment planning process. This system applies knowledge discovery techniques to records of previous experiments to extract experienced engineers' tacit knowledge. We also propose and discuss the concept "Evidence-based experiment planning" for future research.

1. はじめに

現在の電子デバイスに対する省電力化、高効率化、小型化などの要求を受けて、新たな機能を実現する様々なナノ結晶デバイスが開発されている [Fukui 91, Noborisaka 05]。しかし、このデバイス開発に関連する知識は、十分体系化されているとは言いがたい。特に、その製造プロセスに関する情報は重要であり、同じような構造を作るためのプロセスを用いても、そのプロセスの特性の違いによって、出来上がったデバイスの性能が異なるといった状況になっている。そのため、理想的なデバイスの構造を単純に設計するだけでは不十分であり、実際に、そのデバイスを作成する製造プロセスを確立し、その性質を分析する必要がある。

新規のナノ結晶デバイスの開発のプロセスにおいて目的の結晶を生成するためには、製造プロセスに関する様々な実験条件を変更しながら実験を繰り返し行うといった試行錯誤的な作業が不可欠である。この試行錯誤の内容については、研究者の熟練度によって大きく異なり、数々の試行錯誤の経験をつんできた熟練の研究者は、これまでの経験を踏まえることによって無駄な試行錯誤を減らして、効率の良いデバイス開発の支援を行うことが可能となっている。

しかし、現状では、熟練者の知識は「匠の技」と呼ぶような状況であり、その大部分は明示化されていない暗黙知の状態である。そのため、初学者への知識の継承が困難である。よって、この分野が真の工学への転換を図るためには、この「匠の技」から明示化された知識を探索し、工学の知識として継承可能なものへと昇華する必要がある。

現在、人工知能の分野においても、このような知識継承の問題は注目を集めている [森田 07]。しかし、ドメインごとに継承すべき知識のタイプが異なっているために、ドメインごとに様々な手法が提案されると共に、実証を試みている状況である。本研究では、データマイニングやオントロジーといった新しい人工知能の技術を応用する [山口 07] という立場に立ち、

ナノ結晶デバイスの実験記録を情報源として、「匠の技」の裏側にある知識の明示化を支援する手法を提案する。

2. ナノ結晶デバイスの実験記録の活用

ナノ結晶デバイスの開発においては、目的のナノ結晶構造を得るために、数多くの実験を行いながら、結晶を作るための物質の供給方法などのパラメータを決めていく必要がある。一つのナノ結晶構造を得るために決めなければならないパラメータ数は、非常に多いが、一回の実験に、多くの時間(長いものでは、準備を入れると半日以上)とコスト(一回あたり数万円)がかかるため、そのパラメータは、主に、熟練者の経験に頼った試行錯誤によって決定されている。

このような複数のパラメータによる影響を考慮した実験計画を行う方法としては、直交表を用いて、複数のパラメータ間の影響を考慮することにより、実験を効率的に行う手法である実験計画法 [田口 76] が知られている。しかし、実験計画法は、効率のよい実験を行うことができるといっても、数多くの実験を必要とすること、作成するデバイスの目的や材料に多くのバリエーションが存在するために、一つのデバイスに対して、しっかり実験を行ったとしても、他のデバイスに使える保証がないことといった問題があることなどから、このようなアプローチはとられていない。

ただし、これまでの経験を共有するための手段として、実際にどのようなパラメータを用いてどのようなデバイスの作成実験を行ったのかという記録が作成されている。熟練した研究者は、これらの中から適切な類似事例を見つけ出し、参照することによって、パラメータの決定に役立てる。また、初学者は、熟練した研究者から、類似事例を参照するようにという指示に基づき、実際に実験を行いながら、経験的に知識を学ぶことが可能となっている。

これらの現状を踏まえ、本研究では、実験記録から得られる情報を整理し、知識発見の技術を応用することによって、「匠の技」の明示化の支援が可能になると考えた。具体的には、北海道大学量子集積エレクトロニクスセンターで行われているナノ結晶デバイスの開発に関する知識の明示化を、それに関わる

連絡先: 吉岡真治, 北海道大学大学院情報科学研究科, 札幌市北区北 14 条西 9 丁目, 011-706-7107, yoshioka@ist.hokudai.ac.jp

実験記録シートを用いて支援する手法を提案する。

2.1 実験記録シートの分析

現在、北海道大学量子集積エレクトロニクスセンターで行われているナノ結晶デバイスの開発においては、作成する全てのナノ結晶デバイスについて、その作成に用いた実験のパラメータを記入する実験記録シートが用意されている (図 1)。

| Sample structure | | Gas Source | | MO Source | | | | V/III |
|------------------|-------|------------|-------------------------|------------------------|-----------------------------|-----------------------------|------------------------------|-------|
| Growth layer | Temp. | Time | AsH ₃ (5%) | SiH ₄ 50ppm | TMGa Temp. -9°C Heater 80°C | TMAI Temp. 20°C Heater 90°C | DEZn Temp. 3.5°C Heater 50°C | |
| (1) n-GaAs | 770°C | 20' | 2.0 SCCM 2.5E-04 atm | atm | 0.7 SCCM 1.2E-04 atm | 0.7 SCCM 1.2E-04 atm | atm | atm |
| (2) n-AlGaAs | 750°C | 5' | 2.0 SCCM 1.2E-04 atm | atm | 0.7 SCCM 1.2E-04 atm | 0.69 SCCM 1.1E-04 atm | atm | atm |
| (3) p-GaAs | 680°C | 30' | 2.0 SCCM atm | atm | 0.7 SCCM atm | 1.0 SCCM 1.4E-04 atm | atm | atm |
| (4) p-AlGaAs | 700°C | 5' | 2.0 SCCM atm | atm | 0.7 SCCM atm | 1.0 SCCM atm | atm | atm |
| (5) p-GaAs | 700°C | 2' | 2.0 SCCM atm | atm | 0.7 SCCM atm | 1.0 SCCM atm | atm | atm |

図 1: ナノデバイス実験パラメータ記録シート (例)

このシートには、自由記述項目や図や表などを用いることにより、以下の情報を記載する。

1. 実験関連情報
実験者、目的、日付、実験器具内での配置情報などで、主に、自由記述項目ならびに絵として記載される。
2. 実験パラメータ
ナノ結晶デバイスの作成実験で作成した積層構造ならびに、各層を作成したガスや金属の情報とそのパラメータ (温度・圧力...) で、主に、表形式で記載される。

本分析では、表などの情報を活用することが不可欠であるため、実験記録シートは単なるテキストの情報ではなく、その構造を保持したデータベースに格納することが不可欠である。

次に、本シートに実際記入されているデータのバリエーションなどについて、熟練者にインタビューをしたところ、実際のシートには、シートの作成時点では、想定されていなかった情報 (例えば、次の積層を行う前に、温度調整のための時間をもうけることや、特定の層の繰り返しに名前をつけたりすることによって、複雑な層構造を簡略記術する方法など) が記載さ

れていることが確認された。このバリエーションのため、関係データベースに保存するためのスキーマの定義が困難である。

これらの分析結果を踏まえ、実験記録シートについては、半構造データを柔軟に記述し取り扱うことの出来る XML データベースとして作成することとした。

図 2 に図 1 に示したシートの XML 表記の一部を記す。多少冗長ではあるが、各層で用いられた条件や、層の番号などを与えることにより、間に、温度調整のための操作を含むような層の関係 (図 1 の (2) 層と (3) 層の関係) を扱えるようにしている。

```
<?xml version="1.0" encoding="UTF-8"
standalone="yes"?>
<document id="2354">
<date>2008-8-7</date>
<DewPoint>-</DewPoint>
<SampleNo>154</SampleNo>
<Program>N2354-npn AGA CMS NW on
GaAs(111)B </Program>
<Name>富岡, 小林</Name>
<Purpose><! [CDATA [p-GaAs/p-AlGaAs/
p-GaAs/n-AlGaAs/n-GaAs core-multi-shell
NW growth on GaAs(111)B]]></Purpose>
<Structure layer="8">
<Layer id="1" line="1">
<Type>n+-GaAs</Type>
<Operation temp="770" time="20m"
<GasSource type="AsH3" note="5%">
<flow>200</flow>
<pressure>2.5E-04</pressure>
</GasSource>
<GasSource type="SiH4" note="50ppm">
<flow>20</flow>
<pressure>2.5E-08</pressure>
</GasSource>
<MOSource type="TMGa" temp="-9"
<flow>0.70</flow>
<pressure>1.0E-06</pressure>
</MOSource>
<MOSource type="TMAI" temp="20"
<flow>
<pressure/>
</MOSource>
<MOSource type="DEZn" temp="3.5"
<flow>
<pressure/>
</MOSource>
<PressureRate>250</PressureRate>
</Operation>
</Layer>
....
```

図 2: XML 形式による実験記録レコードの記述 (一部)

2.2 実験記録マネジメントシステムの構築

2.1 節での分析を踏まえた実験記録マネジメントシステムの構築を行った。本システムでは、出来る限り、元の実験記録シートのイメージをそのまま残した入力インターフェースを作成することにより、従来の実験記録シートの利用者が違和感な

く使えることを目指した。

| Sample No. | Temp | Time | Class Source | | MO Source | | V/V |
|------------|------|------|------------------|------------------|-------------------|-------------------|------|
| | | | Ash ₁ | Sil ₁ | Temp ₁ | Temp ₂ | |
| 1-Data | 770 | 30m | 200 | MOCH | 1.75 | MOCH | MOCH |
| 1-Data | 750 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 730 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 710 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 690 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 670 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 650 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 630 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 610 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 590 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 570 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 550 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 530 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 510 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 490 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 470 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 450 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 430 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 410 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 390 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 370 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 350 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 330 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 310 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 290 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 270 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 250 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 230 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 210 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 190 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 170 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 150 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 130 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 110 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 90 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 70 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 50 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 30 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |
| 1-Data | 10 | 30m | 100 | MOCH | 1.50 | MOCH | MOCH |

図 3: 実験記録マネージメントシステム

また、簡単な検索機能として、自由記述項目などに対するキーワード検索、積層構造 (A on B) を含むような実験記録の検索などを実現した。

2.3 実験記録マネージメントシステムからの知識探索支援

2.2 で述べた実験記録マネージメントシステムに、知識発見の技術を適用することによって、知識探索の支援を行う方法を提案する。

まず、支援環境を構築するにあたり、探索すべき知識についての分類を行う。今回のような、実験計画を行うための知識は、大きく次の二つに分類できる。

対象知識 材料の物理学的特性などの作成するナノデバイス、実験器具、材料などのモノに関する知識

プロセス知識 実験計画の策定やパラメータの追いつみ方などのプロセスに関する知識

後者のプロセス知識については、ドメインに依存しない汎用性の高い知識が得られる期待があるものの、対象知識が十分に得られていない段階では、その獲得が難しいという問題点がある。

よって、まず、対象知識の獲得から、支援を行う。

最も簡単な知識獲得の手法としては、データマイニングで用いられる頻出パターンマイニングの手法である。

実際に、デバイスの作成時において、問題のない積層が行えるような実験パラメータについては、繰り返し利用されることが想定され、頻出パターンマイニングを行うことによって、抽出することが期待される。

また、同じような積層をする際に異なるパラメータの設定のグループが利用されている場合には、各々のグループの特徴を比較することによって、パラメータ選択の理由を説明するための要因を見つけ出すことが期待される。

これらの情報は、熟練者にとっては、既に暗黙の知識として保持されているものと考えられるが、初学者への伝承可能な知識を探索するという観点からは十分意味のあるものだと考える。

これまでの議論を踏まえて、実験記録マネージメントシステムに次のような知識探索支援機能を提案し、簡単な実装を行った。

1. 頻出アイテムの列挙

本実験記録データベースには、数多くの情報が記載されるが、定型的な要素を多く含むため、単純に頻出パターンの列挙を試みても、結果の分析が難しいと考えられる。よって、ユーザーが調べたい条件 (例えば、特定の材料で構成される層の積層条件など) を与えた上で、頻出アイテムを数える操作を実装する。

2. パラメトリック情報の分布の可視化

本実験記録には、多くの数値データが含まれる。ユーザーが分布を調べたいパラメータを指定することにより、その分布をグラフとして表示する。この結果として、異なるパラメータを用いる実験結果のグルーピングならびに、比較が可能となる。

2.4 実験記録データの遡及入力とシステムの評価

本提案システムの有効性を検証するために、2005年～2008年の間に行われた約700件の実験記録を遡及入力し、初期データベースの構築を行うと共に、データベース構築にあたる問題点の洗いだしとデータ分析の手法についての検討を行った。遡及入力の際に出てきた問題は下記の通りである。

1. 略記法の取り扱い

データシートにその値を省略する場合があることが分かった。インタビューにより、同じ条件の繰り返しなどの場合に、略記をする場合がある事が確認された。

2. 表記のバリエーションの取り扱い

積層する化合物の名前を記述するところに、化合物の名前に加え、積層する層の番号などの追記しているような事例が存在した。

1. については、入力時に省略を行わないように指示することとし、2. については、できる限り入力時の情報を保持したいという観点から、そのままの情報を入力とし、分析時などに必要に応じた表記揺れの統合処理を行うこととした。

また、遡及入力を行った結果を用いた簡単なデータ分析を行ったところ、次のようなコメントを得た。

1. 実験で使ったパラメータの傾向がグラフとして可視化されるのは望ましい。ただし、試行錯誤の課程があるので、最終のナノ結晶デバイスを作るのに使った条件と、試しただけでうまくいかなかった条件の区別をしたい。

2.5 実験記録マネージメントシステムの改良と今後の展開

前節で行ったシステムの評価コメントを受ける形で、実験記録マネージメントシステムの改良を行った。具体的には、

1. 作成したナノ結晶デバイスの評価に関する情報の入力

現在の実験シートには、実験がうまくいったのか、失敗したのかと行った情報が一切記載されていない。そのため、出来上がったデバイスが初期の目標を達成したのかが、達成していない場合は、どのような問題があるのかを記述するための項目を設定した。

2. パラメトリック情報と頻度情報の可視化機能

先に提案したパラメトリック情報と頻度情報の可視化機能を実装した (図 4)。このインターフェースを使うことにより、実験記録中の任意の 2 つのパラメータに関する分布の可視化が行われると共に、どのような条件が多く用いられているのかといった統計情報が簡単に閲覧できるようになる。

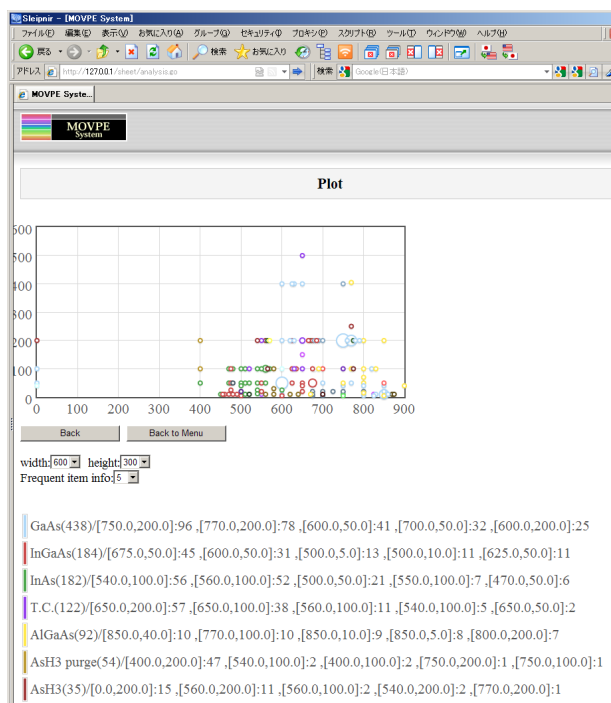


図 4: パラメトリック情報と頻度情報の可視化

このように、ナノ結晶デバイスの研究者のようなドメインの専門家と協力をして、今回のような知識探索支援のシステムを構築する際には、以下の2つのステップをサイクルとして繰り返すことが有効であることが確認された。

1. プロトタイピングを行うことにより、実際のデータに対する分析の可能性を示す。
2. 具体的なシステムの結果を見ることによって、さらなる要求を具体化することが可能となり、その要求を新しいシステムの試用として実現する。

また、現在の実験記録分析システムの有用性を知識継承という観点から検証するために、研究室に配属されて間もない初学者を対象として、利用実験を行う予定である。

3. Evidence-based Experiment Planning にむけて

前節で述べた支援システムは主に初学者を対象として、知識の伝承を支援するための知識探索であった。これに対し、熟練者を対象にした支援について検討を行う。

具体的には、個別の実験条件データのマイニングだけでなく、実験シーケンスとしてのマイニングなども視野に入れた分析をおこなう。その効果として、より複雑な検索条件に対応すると共に、うまくいった実験シーケンスや失敗したシーケンスなどをまとめて表示し、より、実験者に役立つ情報を集約して表示が可能となる。

また、実験シーケンスに関する分析を行うことにより、プロセス知識についての知見が得られることが期待される。

また、これらの知見を明示的な知識として、体系化することが望ましいが、なかなか教科書に書くような知識としてまとめることは困難であると考えている。そのため、少なくとも、実験記録分析の知見を考慮した実験条件間の類似性の定義などを

行い、適切な既存事例を提示し、実験条件の設定の Evidence として利用する Evidence-based Experiment Planning の実現を目指す。

さらに、対象物に関するモデルや、関連する知識を記述した推論システムを統合する [吉岡 98]。その結果、因果関係を考慮したデータ分析を実現し、既存の事例と第一原理を組み合わせた分析を支援する環境の構築を目指す。このような環境を構築すると、パラメータの変更が与える影響を簡単なシミュレーションで確認してから、実験を遂行できる。さらに、実際の実験結果との比較から、モデルや知識の修正へのフィードバックが可能となり、体系的な実験結果の整理、知識探索の支援につながる事が期待される。

4. おわりに

本稿では、ナノ結晶デバイスの実験記録データベースの構築とそこからの知識発見を例にとり、知識探索の支援環境を提案した。

さらにこのモデルケースは他のドメインへ展開できる。例えば作業記録の分析手法については、他のドメインでも応用可能であると考えられるので、ナノ結晶の開発といった一つのドメインでの成果を他のドメインにも転用可能となる。今後は、この成果を踏まえて、汎用性の高いシステム構築へと発展させていく予定である。

謝辞

本研究の一部は、北海道大学大学院情報科学研究科グローバル COE プログラム「知の創出を支える次世代 IT 基盤拠点 (平成 19 年度 ~ 23 年度)」によるものである。

参考文献

- [Fukui 91] Fukui, T., Ando, S., Tokura, Y., and Toriyama, T.: GaAs Tetrahedral Quantum Dot Structures Fabricated using Selective Area Metalorganic Chemical Vapor-deposition, *APPLIED PHYSICS LETTERS*, Vol. 58, pp. 2018–2020 (1991)
- [森田 07] 森田 千絵:「知識継承」特集にあたって, 人工知能学会誌, Vol. 22, No. 4, pp. 459–460 (2007)
- [Noborisaka 05] Noborisaka, J., Motohisa, J., Hara, S., and Fukui, T.: Fabrication and characterization of freestanding GaAs/AlGaAs core-shell nanowires and AlGaAs nanotubes by using selective-area metalorganic vapor phase epitaxy, *APPLIED PHYSICS LETTERS*, Vol. 87, (2005)
- [田口 76] 田口 玄一: 実験計画法 (上), 丸善 (1976), (第 3 版)
- [山口 07] 山口 高平: 知識マネジメントと AI 技術, 人工知能学会誌, Vol. 22, No. 4, pp. 461–466 (2007)
- [吉岡 98] 吉岡 真治, 富山 哲男: 設計支援のための統合モデリング環境の研究 - プラガブル・メタモデル機構の提案 -, 人工知能学会誌, Vol. 13, No. 2, pp. 312–319 (1998)