Knowledge Exploratory Project for Nanodevice Design and Manufacturing

Masaharu Yoshioka
Graduate School of Information Science and Technology, Hokkaido Univeristy
N14 W9, Kita-ku, Sapporo-shi Hokkaido Japan
yoshioka@ist.hokudai.ac.jp

Katsuhiro TOMIOKA
Japan Science and Technology Agency
N14 W9, Kita-ku, Sapporo-shi Hokkaido Japan
tomioka@rciqe.hokudai.ac.jp

Shinjiroh HARA
Graduate School of IST, Hokkaido Univeristy
N14 W9, Kita-ku, Sapporo-shi Hokkaido Japan
hara@rciqe.hokudai.ac.jp

Takashi FUKUI
Graduate School of IST, Hokkaido Univeristy
N14 W9, Kita-ku, Sapporo-shi Hokkaido Japan
fukui@rciqe.hokudai.ac.jp

ABSTRACT

We are developing a framework for knowledge creation in nanodevice development, based on collaboration between nanodevice engineers and computer science researchers. Development of nanodevices requires a variety 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. We propose the concept of "Evidence-based experiment planning" and develop a process for supporting experiment planning in nanodevice development. This system applies knowledge discovery techniques to records of previous experiments to extract experienced engineers' tacit knowledge.

Categories and Subject Descriptors

H3.5 [**Information Systems**]: Online Information Services—Data Sharing

General Terms

Knowledge Management, Data mining

Keywords

Experiment planning, Nanodevice design

1. INTRODUCTION

"Nanoinformatics" is one of the emerging research fields in developing a computational framework to support nanoscale research [8]. A wide variety of research studies are conducted in nanoinformatics[6]. Especially in the field of nanomedicine development, some integrated computational frameworks have already been proposed (e.g., [5]).

However, the nanodevice development process is not well systematized, and it requires both engineering knowledge and craftsmanship skills [3, 7]. For example, nanodevice design based on knowledge of first principles, such as atomic physics, does not mean the end of the development process. Because the manufacturing process may affect the quality of the nanodevice, much trial and error is required before the final product can be realized. Skilled engineers can conduct this experiment planning more effectively than novices.

However, knowledge about this planning process is difficult to transfer from skilled engineers to novices. Novice engineers can only acquire the knowledge required for their planning through the guidance of skilled colleagues. To accelerate this nanodevice development process, it is better to make this tacit knowledge explicit.

Based on a collaboration between top-level nanodevice engineers and computer science researchers, we propose a computer support framework for knowledge transfer and transformation processes for nanocrystal device development. We introduce an experiment record management system for extracting useful information from these records. We also discuss our future plan to extract useful information from this management system.

2. EXTRACTION OF USEFUL INFORMATION FROM EXPERIMENT RECORDS

In the nanocrystal device development process, engineers conduct multiple experiments before making final products. Each experiment requires a long time (more than half a day, including preparation) and is expensive (more than a hundred dollars), making it difficult to use a comprehensive experiment planning method such as "Experimental Design" or "Taguchi methods"[9]. Therefore, in this trial and error process, the engineers modify experimental parameters step by step.

In this planning process, skilled engineers give advice to other engineers in two formats. One is direct advice that shows how to modify these parameters. The other is indirect advice that suggests that the user should read related experiment records.

For the first step of our project, we have built an experiment record management system that supports the search for useful experiment records from the available records.

2.1 Analysis of experiment records

In the Research Center for Integrated Quantum Electronics, Hokkaido University, researchers are developing various kinds of nanodevices (e.g., nanowires on Si) by using a selective-area metal-organic vapor phase epitaxy (SA-MOVPE) method [4].

Even though SA-MOVPE is a good method that can control the quality of the device, it requires many trial-and-error experiments to arrive at the final process. To keep records about these experiments, researchers use the SA-MOVPE growth parameter record sheet (Figure 1) for each experiment.

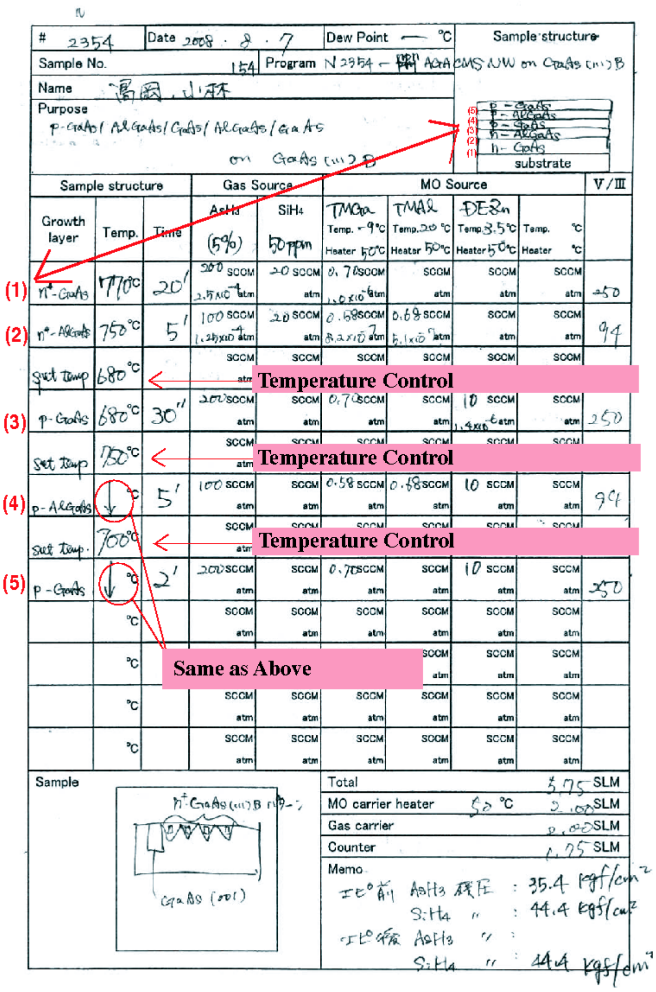

Figure 1: An example of a SA-MOVPE growth parameter record sheet

The following three types of information are recorded in these sheets.

- Background information
  - ID
  - Date
  - Name of the experimenter
  - Purpose
- Growth layer information
  - Type of growth layer
  - Gas and metal organics sources with parameters (concentration, temperature, ...)
  - Operation for growing layers (gas temperature, pressure, mixture, ...)
- Memo

To utilize the parametric information described as a table in the sheet, a database system that can store this structured information is necessary.

In addition, because there are varieties of development processes for these experiments, users have extended the original intended usage of the sheet (e.g., by introducing cyclic development processes). With these variations, it is difficult to define a relational schema to store the information. Therefore, we decided to use XML semistructured data for our management system.

2.2 Construction of an experiment record management system

Based on the analysis of Section 2.1, we have implemented a prototype system for the SA-MOVPE experiment record management system. This system has an interface that is similar to the image of the original sheet.

In this system, we can store all the information of the original sheet except diagrams. In addition, this system can use structured queries based on the XML data (e.g., find records that have the name "Tomioka" as experimenter, find records that have a layer "p-GaAs" on a layer "p-AlGaAs").

To evaluate the effectiveness of the proposed system, we constructed a database of real experiment records from 2005 to 2008 (approximately 700 records) using the system. From this construction process, we found the following problems with data input.

1. Omission (Ellipsis)
   In the data sheet, users tend to omit data elements that have the same value in the column above. It is necessary to recover such information for further analysis.
2. Variations of description
   There are variations of description for the layer information. For example, some researchers may add information about layer thickness.

Because we would like to keep almost all the information from each datasheet, the following guidelines were used for inputting the data.

- Omissions should be recovered when the user identifies them.
- Additional information about layers should be kept as it is. We plan to develop a system for handling such variation later.

```xml
<?xml version="1.0" encoding="UTF-8"
  standalone="yes"?>
<document id="12354">
  <date>2008-8-7</date>
  <DewPoint>-</DewPoint>
  <SampleNo>154</SampleNo>
  <Program>N2354-npn AGA CMS NW on
  GaAs(111)B </Program>
  <Name>Tomioka</Name>
  <Purpose><![CDATA[p-GaAs/p-AlGaAs/
  p-GaAs/n-AlGaAs/n-GaAs core-multishell
  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="TMAl" temp="20"
          <flow/>
          <pressure/>
        </MOSource>
        <MOSource type="DEZn" temp="3.5"
          <flow/>
          <pressure/>
        </MOSource>
        <MOSource type="" temp="" heater="">
          <flow/>
          <pressure/>
        </MOSource>
        <PressureRate>250</PressureRate>
      </Operation>
    </Layer>
```

Figure 2: A part of a record sheet in XML format

After constructing the database, we demonstrated the data retrieval and analysis function of the system to the nanodevice engineers. From this demonstration process, the following comments were offered by the engineers.

- Record retrieval with structural queries is helpful for finding relevant previous research experiment records.
- It is preferable to have a good visualization tool to represent parameter information, such as information about parameters that are frequently used.
- It is preferable to have an evaluation of the experiment (e.g., Succeed/Fail, analysis of the failure).

2.3 Proposal for a new experiment record management system

Based on the initial database construction experiment, we proposed a framework that uses a knowledge discovery technique to extract useful information from the records stored in the system discussed in the previous section.

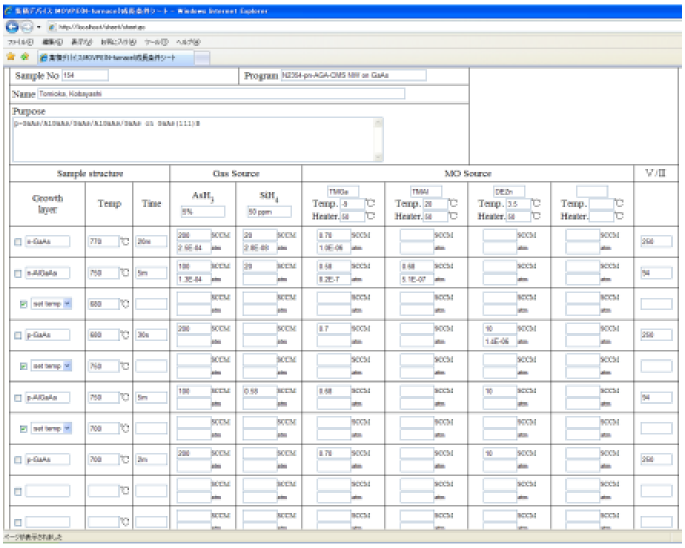

Figure 3: Screenshot of the SA-MOVPE experiment record management system

Before proceeding to the detailed explanation of creating knowledge, we first discuss the two types of knowledge being created.

Object knowledge
: Knowledge about objects, such as nanodevices, materials and experimental instruments.

Process knowledge
: Knowledge about a process, such as how to design experiment plans, and how to modify parameters to obtain optimal results.

Process knowledge may be abstract and widely applicable, but it is very difficult to extract without an understanding of object knowledge about the problem domain. We therefore started by making a computational framework for extracting object knowledge about nanodevice development. One of the best-known and useful knowledge discovery techniques is frequent-pattern mining (e.g., Apriori [2] and FREQT [1]). By using this mining technique, we can extract the most common parameter settings for a particular growth layer. In addition, when there are two or three common parameters for such a growth layer, we can expect that significantly different conditions may apply to the selection of the various parameters for these experiments. Combinations of frequent-pattern mining with such comparative analysis may extract information related to the tacit knowledge of skilled engineers.

Information extracted from such a simple mining technique may simply be common sense for a skilled engineer, but it may be useful for the novice engineer who has just started to use such experimental instruments.

Based on this discussion, we implemented a new experiment record management system with the following knowledge creation support functions.

1. Frequent-item mining
   In the XML record of each experiment, there are several routine descriptions. Therefore, simple frequent-pattern mining may find such information, even though it may be very difficult to analyze the results. We therefore implement simple frequent-item enumeration

for the records that match particular conditions (e.g., records for a particular growth layer).

2. Parametric information visualization
The system uses an interactive parametric information visualization system to form groups for comparison (Figure 4). By using this framework, the user can understand the relationship between two different parameters and the frequency associated with the label (e.g., name of the layer). Figure 4 represents the relationship between temereatrue for the process and gas flow late of AsH_3. From this plot, the user can understand it is prefereable to use high temperature (larger than 600) for layer "GaAs" and to use lower temperature (smaller than 600) for layer "InAs".

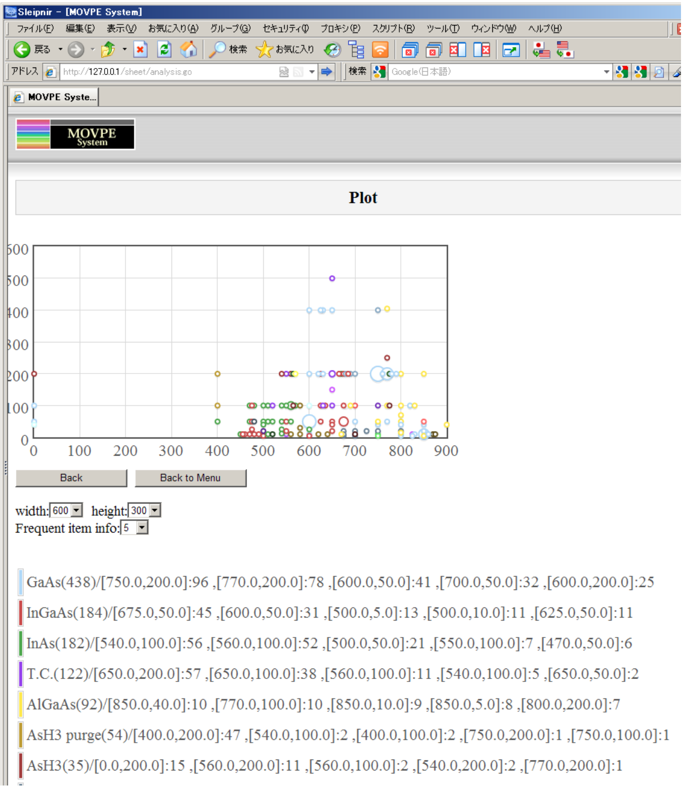

Figure 4: Parametric information visualization with frequency analysis

In addition, we updated the data sheet format to include information about the evaluation of the experiment. We are at the stage of updating the contents of database with the information of this information.

3. TOWARDS EVIDENCE-BASED EXPERIMENT PLANNING

The system discussed above aims to give novice engineers an understanding of tacit knowledge based on previous trial and error processes.

To achieve further support, analysis of data from individual experiments is not enough. For example, it is better to have ontological knowledge about the nanodevice information, and it is preferable to consider the usage of the nanoinformatics computational resources [6, 5]. In addition, it is necessary to make a framework that can represent a sequence of trial and error processes for analysis. With this sequential information, it may be possible to extract process knowledge.

However, we think that it is not easy to formalize information extracted from these raw data records as systematized knowledge. We therefore have tried to define a framework for analyzing similarity between the different experimental cases and have made a computational framework for "Evidence-based Experiment Planning" that uses previous experiment cases as evidence for new experiment planning.

4. SUMMARY

In this paper, we have proposed an experiment record management system that can extract useful information from the records based on discussions between nanocrystal device researchers and computer scientists. It is necessary to have a spiral development process (i.e., prototyping of the record management system and evaluation from the nanodevice engineers) for such system.

In the future, we plan to apply this record management system to the education of novice engineers in nanodevice development and to evaluate the effectiveness of the system.

Achknowledgement

This research was partialy supported by a grant for Hokkaido University Global COE program, "Next-Generation Information Technology Based on Knowledge Discovery and Knowledge Federation", from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

5. REFERENCES

[1] K. Abe, S. Kawasoe, T. Asai, H. Arimura, and S. Arikawa. Optimized substructure discovery for semi-structured data. In *Proc. 6th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-2002)*, pages 1–14. Springer-Verlag GmbH, 2002. LNAI 2431.

[2] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo. Fast discovery of association rules. pages 307–328, 1996.

[3] T. Fukui, S. Ando, Y. Tokura, and T. Toriyama. GaAs tetrahedral quantum dot structures fabricated using selective area metalorganic chemical vapor-deposition. *APPLIED PHYSICS LETTERS*, 58:2018–2020, 1991.

[4] K. Ikejiri, T. Sato, H. Yoshida, K. Hiruma, J. Motohisa, S. Hara, and T. Fukui. Growth characteristics of GaAs nanowires obtained by selective area metal-organic vapour-phase epitaxy. *NANOTECHNOLOGY*, 19, 2008. 265604.

[5] F. Martìn-Sanchez, V. Lòpez-Alonso, I. Hermosilla-Gimeno, and G. Lopez-Campos. *A Primer in Knowledge Management for Nanoinformatics in Medicine.* Springer-Verlag GmbH, 2008. LNCS 5178.

[6] N. N. Network. *Workshop on Nanoinformatics Strategies.* 2007. http://128.119.56.118/ nnn01/Workshop.html.

[7] J. Noborisaka, J. Motohisa, S. Hara, and T. Fukui. Fabrication and characterization of freestanding GaAs/AlGaAs core-shell nanowires and AlGaAs nanotubes by using selective-area metalorganic vapor phase epitaxy. *APPLIED PHYSICS LETTERS*, 87, 2005.

[8] K. Ruping and B. Sherman. Nanoinformatics: Emerging computational tools in nano-scale research. In *Technical Proceedings of the 2004 NSTI Nanotechnology Conference and Trade Show, Volume 3*, pages 525–528, 2004.

[9] G. Taguchi. Quality engineering (taguchi methods) for the development of electronic circuit technology. *Reliability, IEEE Transactions on*, 44(2):225–229, Jun 1995.