4CeeD Lecture Series

Lecture #1: Introduction and Overview

March 21st, 2023

Use of 4CeeD Services for Material Scientists

Leah Espenhahn (leahe2@illinois.edu), Robert Kaufman (rbkaufm2@illinois.edu)
Beitong Tian (beitong2@illinois.edu), and Klara Nahrstedt (klara@illinois.edu)

A timely and trusted curator and coordinator of scientific data
Lecture Series Learning Objectives

• Lecture 1 (3/21): Overview of 4CeeD
  • Introduction to 4CeeD
    • Breakout Session #1: Log-In and Explore
  • Key 4CeeD Features: Templates for Fast Storage
    • Breakout Session #2: Creating and Using Templates

• Lecture 2 (3/23): Workshop (Cont.) & Advanced Features
• Lecture 3 (3/28): 4CeeD Backend Services
• Lecture 4 (3/30): SENSELET/MAINTLET
Introductory Questionnaire

• Do you describe yourself more of a computer scientist or a material scientist?

• Do you conduct your research primarily on modeling/simulation or experimental research?
What is 4CeeD and its goals?

- Address Scientific Digital Data Acquisition, Curation and Sharing prior to Scientific Publication of Results via Private Cloud Storage Facility

**Experimental Setting**
- Time: 30 min
- Temp: 425°C

**Observation Notes:**
- Oxidation depth is 12 um
- Oxidation layer is Al$_{0.98}$GaAs
- N$_2$ bubbler flowing steam
- 2” Quartz tube furnace

**Sample Experimental Dataset from SEM Imaging**

- **SEM Image**
- **Sidewall View of AlGaAs DBR**
- Hitachi S4800 SEM

**Instrument (in MRL/HMNTL/BI)**
Why is current data collection an issue?

- Consideration of National Academy Studies: **20-year gap** from discovery of new materials to implementation of next-generation devices
- Necessitates **real-time** and **trusted** processing from **materials** – to-devices digital data

Red LED (1962) Holonyak, Jr.  
Blue LED (1993) Nakamura  
Transistor (1947) Bardeen  
Graphene (2004) Geim and Novoselov
Why is University data collection unique?

- Industry concentrates on scaling and improving reliability and manufacturability to deliver an end product (smartphone, cars)

- Academia concentrates on diverse and riskier research using older equipment to prove innovative ideas and concepts
Current State of Data Capture

- Fabricate experimental sample
- Prepare analytical sample
- Bring sample to instrument for analysis
- Extract data (File conversion)
- Transport data to office computer
- Analyze data

Instrument (MRL/HMNTL)

Flash Drives

Office (MRL/MEB/MNMS)

Metadata Loss:
- Excessive file name schemes
- Manual notes
- What metadata is important?

Sneakernet:
- Security risk at both ends
- Very limited transport space
- Lost or forgotten flash drives

Current State of Data Capture:
- Only what you have
- Connect notes to data (after time delay)
- Retroactive not reactive
Scenario with 4CeeD Integration

- Fabricate experimental sample
- Prepare analytical sample
- Bring sample to instrument for analysis
- Extract data (NO FILE CONVERSION)
- Transport data to office computer (DIRECT)
- Analyze data (REAL TIME)
- Repeat per iteration

Instrument (MRL/MNTL)

Laptop

Campus PC

Collaborators

Office (ECEB)

4CeeD Data Capture

Real time - Metadata paired with data - Reactive interpretation - Easy interpretation/searching

Direct web interface

- Highly secure data storage
- Large transport volume
- Easy to use/annotate

Complete Data

- More simple file names
- Automatic note taking
- All metadata included

Collaborators

Laptop

Campus PC

Office (ECEB)
How is Metadata Currently Stored?

- Manual notetaking of complex experiments can lead to inconsistent or inadequate documentation

- Data transfer from tools is often done using flash-drive or emails that carry limitations and security risks

- Material research data is expensive and time intensive
  - Publication data is often documented well but remaining data is discarded
Metadata of Device Development Process

- Experiments can have multiple steps where each step is verified for success by various metrology methods.

### Steps

- **SiO₂ Mask Deposition** → **Lithography** → **Diffusion** → **SiNₓ Deposition** → **Plasma Etching**
- **SiNₓ Removal** → **Oxidation** → **Lithography** → **Metallization** → **Device Characterization**

### Measurement Data

- Profilometry
- Ellipsometry
- Optical microscopy
- SIMS
- SEM
- SPA
How is Metadata Currently Stored?

Experimental Setting:
Time: 11 minutes
Temperature: 425°C

Material Analysis:
Oxidation Depth: 8.93 μm
Oxidation Layer: Al_{0.98}Ga_{0.02}As
Furnace: N₂ Ambient, 2” Quartz

SEM Image Settings
Scale: 4 μm (10 divisions)
Acceleration Voltage: 10.0 kV
Working Distance: 6.8 mm

Multiple text and image files are necessary to capture all the pertinent data of a single experiment
Example of Metrology Data (SIMS, SEM, OM, SPA)

Vastly different data formats for same device
Multiple text and image files are necessary to capture all the pertinent data of a single experiment.
File explorer is limited to default “file list” where information is kept in long text names.
How this look from 4CeeD [Datasets]

- 4CeeD is designed to present only pertinent information for quick understanding of the experiment.
Data Hierarchy in 4CeeD

- "Spaces", "Collections", and "Datasets"
Data Organization Structure

1) Space

2) Collection

3) Datasets

Owner

Shared User

Shared User

Shared User

Shared User

Shared User

Shared User

2) Collection

3) Datasets

Owner

Shared User

3) Datasets

Owner

Shared User

10. Base Zn Diffusion

Process of diffusing Zn into the base region of the Ti-QCL to increase doping and improve contact resistance.

Process (See detailed process on sheet in 2111):
1. Prepare tubes
2. Flame off and dimple
3. Load diffusant (Zn3As2); load sample
4. Seal off...

Owner: Robert Kaufman
Created on Aug 21, 2019
Breakout Session #1: Log-In & Explore 4CeeD

learn.4ceed.illinois.edu

- "Sign Up" with Illinois email
- **Note:** Need to be on University connection to access (IllinoisNet or VPN)
- Try out 4CeeD yourself before next hands-on lecture
  - **Make an account**
  - Make a space, collection, dataset
  - Upload some image and add a template
Lecture Series Learning Objectives

• Lecture 1 (3/21): Overview of 4CeeD
  • Introduction to 4CeeD
    • Breakout Session #1: Log-In and Explore
  • Key 4CeeD Features: Templates for Fast Storage
    • Breakout Session #2: Upload your own data
Efficient Data Collection: Templates

• Templates and Extractors for Rapid Storage

Templates are stored as Metadata
Efficient Data Collection: Templates

- Templates provide consistency, accessibility, and enables digital processing (ex. Jupyter Notebook)

Free-text view can be unorganized, hard-to-read, and difficult to compare results
Breakout Session #2: Templates For Fast Storage

- Stored Key-Value Pairs

Incredibly useful for Jupyter Integration (Later Topic)