Data Curation and Analysis for Semiconductor Device Research

Robert Kaufman, Patrick Su, Zhe Yang, Steve Konstanty, Todd Nicholson, John Dallesasse, and Klara Nahrstedt

University of Illinois at Urbana-Champaign
Department of Computer Science, Department of Electrical and Computer Engineering
Contact: rbkaufm2@illinois.edu, psu8@illinois.edu, klara@Illinois.edu

Abstract and Motivation

The process of fabricating and characterizing a semiconductor device can be tedious, in part because of the copious amount of variables that contribute to the final outcome. Proper documentation of process parameters, physical device progress, and preliminary data is crucial to fine-tuning the fabrication recipe and creating a quality device. A secure database system designed with this documentation in mind greatly improves the fabrication workflow. By allowing the database to connect with a networked instance of Jupyter Notebook, the processing of intermediate data is further streamlined. Future improvements planned for the system include utilizing the advanced computational methods such as image classification and object detection to identify connections between process parameters and results that are difficult for an individual to identify.

4CeeD: A Secure Research Database

4CeeD is a webservice built on top of Clowder that seeks to address common struggles with the storage and organization of research data. It is designed with 5 fundamental operations in mind:

Capture — A simple, secure file-uploader allows for quick capture of important information. Upload images of a process step, use the robust template system to note important parameters, and utilize the built in metadata extractors to capture all information available in the files.

Curate — Use the system of collections and datasets to categorize and organize research data. Collections act like a standard hierarchal folder but with additional options for descriptors. Datasets help collect data while providing quick visual representations to ease the curation process.

Coordinate — Spaces allow for collections and data to be easily shared among multiple users to help coordinate cooperative processes.

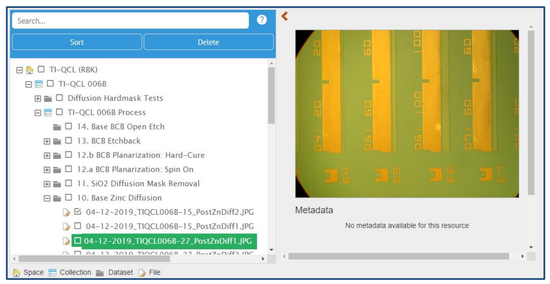
Correlate — The metadata extractor and template system makes correlating process parameters to gathered data simple. In addition, the visual preview of files can help users roughly identify the effects of process changes to result differences.

Distribute — The separation storage hierarchy of spaces, collections, and datasets makes it easy to distribute the data to only who you want while keeping the rest private.

General 4CeeD Structure

Documenting Fabrication with 4CeeD

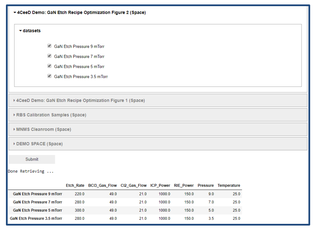
- Personally utilized 4CeeD to manage process data for fabricating a semiconductor laser over the last year
- Used a space shared with group members to aggregate data
- Separated device lots by collections and processes by datasets
- In each device collection, data was separated into process, electrical characterization, and optical characterization sub-collections
- Each process step was dedicated to a dataset in which progress images, process parameters, and general notes were stored
- When a process failed and required more development time, sub-collections were made and each new process test was stored in a dataset
- Overall, 4CeeD made managing and sharing data much easier than using a simple cloud storage option as has been used before.
- Robust note taking ability of datasets and quick process visualizations made fixing process recipes much easier

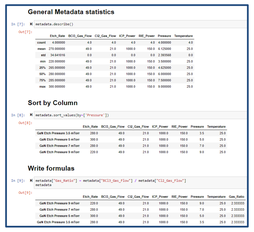


Integration with Jupyter Notebook

Interfacing with 4CeeD

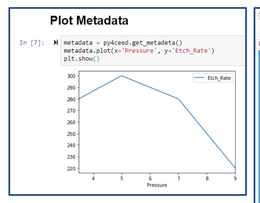
- Jupyter Notebook allows both new and more advanced users quickly create a script that fits their needs through an open source solution.
- Custom py4ceed library allows interfacing with 4CeeD data through a simple user interface to increase accessibility to the general user
- Included a quick method to retrieve and do simple processing of fabrication parameters through the metadata stored in datasets
- Integrated methods for basic operations such as retrieving standard statistics or sorting by custom formulae
- Still allows more advanced users to fully use the robustness of python for more advanced data analysis on their processes
- Step-by-step visualization of data takes advantage of Jupyter Notebook's blocked layout

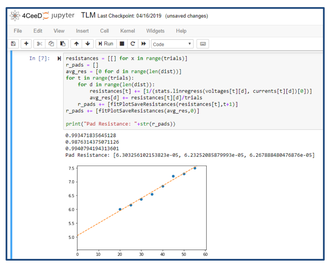




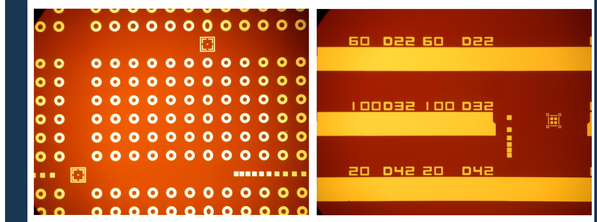
Portable Analysis and Visualization

- Hosted Jupyter Notebook servers allow for users to log in and process their data anywhere on an authorized network and on any device regardless of processing power
- Allows for immediate visualizations of either dataset metadata or data from files stored in 4CeeD
- Ability to save any scripts back to 4CeeD ensures that scripts are always accessible
- Personally, has allowed for quick analysis of etch rate changes for a process development (*left*)
- Also has saved a lot of time by having a portable script for parsing, calculating, and visualizing resistivity values for almost 50 device samples (*right*)

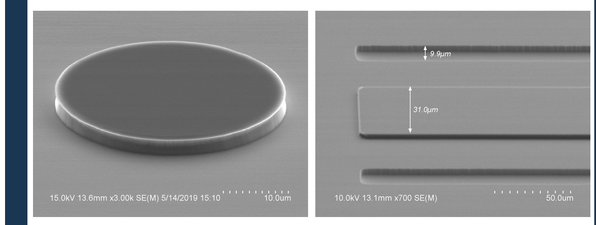




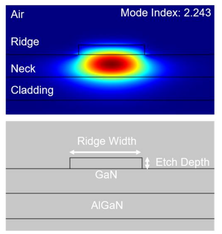
Applications to Advanced Computing

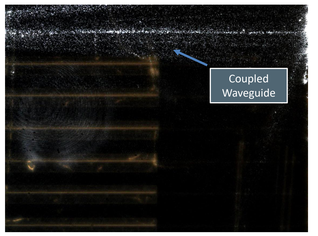


- Hundreds to thousands of semiconductor devices are made in a small area where a single device can cause the entire chip to fail.
- Large data set can be used to drive advanced computational tools such as deep learning and causal analysis for better performing devices.



- Scanning electron microscope images are used to validate and verify certain device parameters such as device size, sidewall, and smoothness that affect its device performance.
- Image verification is typically carried out by visual inspection from a researcher and therefore can only provide qualitative feedback rather than quantitative analysis.
- Advanced computational tools such as image classification and object detection can accelerate and provide a more accurate verification.





Acknowledgements

This material is based in part upon work supported by the National Science Foundation under award number 1443013 and 1659293.