Cloud-connected data analysis software for silicon photovoltaic characterization

Michael Albert

Office of Science, Science Undergraduate Laboratory Internship Program

California Polytechnic University, San Luis Obispo, CA

National Renewable Energy Laboratory

Golden, Colorado

August 28, 2025

Prepared in partial fulfillment of the requirement of the Department of Energy, Office of Science's Science Undergraduate Laboratory Internship Program under the direction of David Young at the National Renewable Energy Laboratory.

Participant: Michael Albert

Research Advisor: David Young

ABSTRACT

Device characterization is a key step in Advanced TOPcon solar cell research at NREL. The data obtained using characterization tools like the Sinton Lifetime Tester and Sinton JV Flash Tester helps researchers at NREL make conclusions about the quality of solar cells, the efficacy of fabrication processes, and the design of future experiments. Currently, these characterization data are stored locally on outdated desktop devices and analyzed using external programs, which requires researchers to be in the lab to access past data, and leads to data loss and mismanagement. This paper describes the design and benefits of a cloud-connected, distributed, and no-code data analysis solution built in Python. The resulting tool allows reserachers to analyze and download past data from remotely. The program itself functions in two phases. The first is a spreadsheet menu which connects to the runsheets researchers use to organize solar cell recipes and experiments by batch and sample. Researchers can select one of these sheets within the tool or search by a sample ID across all run sheets on the cloud store. Once a sample ID is selected, the second phase of the tool begins, which allows researchers to view all data across the characterization tools for that sample, and any other samples within that batch. A key benefit of the program is automatic plotting and analysis capabilities using Python's MatPlotLib library, saving researchers time when viewing and sharing their results. These two phases help to link the fabrication and characterization steps of the cell manufacturing process, which will keep data organized for future use in machine learning applications. The tool has been deployed for use by NREL researchers, and benefits to the characterization and data analysis process are being monitored for future improvements.

I. INTRODUCTION

Photovoltaics (PV) are now the lowest cost form of electricity in many parts of the world[1], and are predicted to have the greatest global share of installed electrical capacity by 2050.[2] Rapidly advancing device innovation and manufacturing capacity spurred by global demand for renewable energy has been central to the production growth and cost descrease of solar PV.[3]

Passivated emitter and rear cell (PERC) device technologies had controlled approximately 80% of the market share since 2020 due to low production costs of US$0.22/W.[1,4] However, a newer technology, monocrystalline Si (c-Si) solar cells based on tunneling oxide SiO_x/polycrystalline-Si (SiO_x/poly-Si) passivating contacts (TOPCon), with a market share of 21% in 2023, are predicted to overcome PERC device production by 2025.[1] Accordingly, these cells are likely to become the technology of choice for new-cell manufacturing in the US.

TOPCon cells have reached energy conversion efficiencies of just over 27%[5], prompting further investment in this technology. In accordance with these developments to PV technology, the United States Department of Energy (US DOE) aims to drive innovation to make American-made solar technology affordable, and to provide relevant and objective technical information to meet the growing electricity demand.[6]

Researchers in the High Efficiency Crystalline Photovoltaics Group at the National Renewable Energy Laboratory (NREL) are developing advanced TOPCon cells. These devices feature an n-type doped c-Si wafer, double-side-textured surfaces with doped poly-Si/SiO_x passivated contacts, and i-poly-Si/SiO_x passivation.[7] A key stage in this research is experimenting with variation in the material properties of the cell and steps of the fabrication process.[8]

The success of these experiments is determined through a characterization process[9,10], which uses instruments to take cell perfomance data, namely the Sinton Wafer-Lifetime Tool (WCT-120) and Sinton Light I-V Tool (XT-80). Analysis of these data requires manual data extraction and plotting, which is a process this research seeks to improve for the group. This article will detail the development of a software tool with the goal of supporting NREL researchers by providing a scalable, robust data analysis program that reduces the uncertainty and latency of PV cell characterization and experiment results reporting. The tool provides a cloud connected user interface, built on an existing laboratory information management systems (LIMS)[11] for automatically viewing, plotting, and saving characterization data.

II. PROCESS

This program builds on existing technologies to improve the characterization process for researchers. The program itself uses a graphical user interface (GUI) on the frontend, and is split into a series of steps on the backend, allowing for a guided user experience (UX). There are two pathways a user can take within the program to access characterization data: manual dropdown search and runsheet sample selection.

A. Existing technologies

1. Sinton WCT-120 and XT-80

The WCT-120 and XT-80 are characterization tools designed by Sinton Instruments. These intruments automatically upload characterization data to a desktop computer that displays characterization results using their own internal software. Using admin settings on these devices, users can create their own file directory for saving plot data and general run results as text log files. Users are also able to input a batch and sample ID for the cell under test to organize data. Result files are dated, and contain a spreadsheet of runs done by a user, organized by the batch and sample IDs. The plot data files are named by a consistent {Batch_ID-Sample_ID}.txt structure. On each instrument's local desktop file system a file tree automatically created, seen in Figure 1.

The user, batch, and sample values from the Sinton characterization tools lend themseleves well to an automatic file searching algorithm on the backend of the cloud-connnected data analysis tool. The algorihm traverses the tree structure, looking first for the user ID. Under this user ID are all the text log files for that user. Once a text log is downloaded and a batch ID in the textlog file has been selected by the user, a batch directory can be downloaded. These directories contain all of the plot data text files for that batch. Now that the user has the corresponding textlog and plot data files locally downloaded from an experiment, they can use the tool to analyze those data.

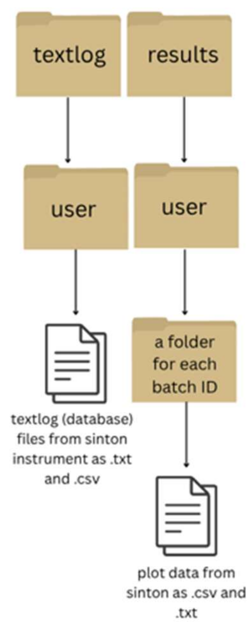


Figure 1: The file tree created by the Sinton tools for both the WCT-120 and XT-80, which is replicated for the cloud-connected data analysis tool.

2. *VAST cloud storage LIMS*

Data is harvested off of the desktop file sytem of each instrument to the LIMS at two endpoints: ./textlog/ and ./results/. The files at these endpoints contain all the necessary data for the analysis tool to function, and correspond to the ID of each experimental cell: batch ID and sample ID. Once these file names at each endpoint are compiled into Python arrays, the user may begin interacting with the tool and accessing data.

B. Graphical user interface

The GUI for the tool is built using PyQT5, a toolkit that links QT Designer, natively a C++ program, with Python code. Using PyQT5, interactive widgets can be combined to create a complete GUI, allowing the user to navigate between pages and interact with data by simply pressing buttons and selecting values from dropdown menus. These dropdowns also function as auto-completing search bars, simplifying the process of finding files. To better support researchers, the GUI was modeled after the GUI on each Sinton intrument's desktop device and adapted for laptop use. Each page of the tool's GUI is separated into frames that can resized for ease of use.

The tool is split into three independently functioning pages: a run sheets page (Figure 2), a Sinton lifetime page (Figure 3), and a Sinton JV flasher page (Figure 4). Changing user, batch, or sample values on the lifetime page, for example, is independent of the multiflash and run sheet pages. Navigation between pages is done with a navbar at the top of the GUI.

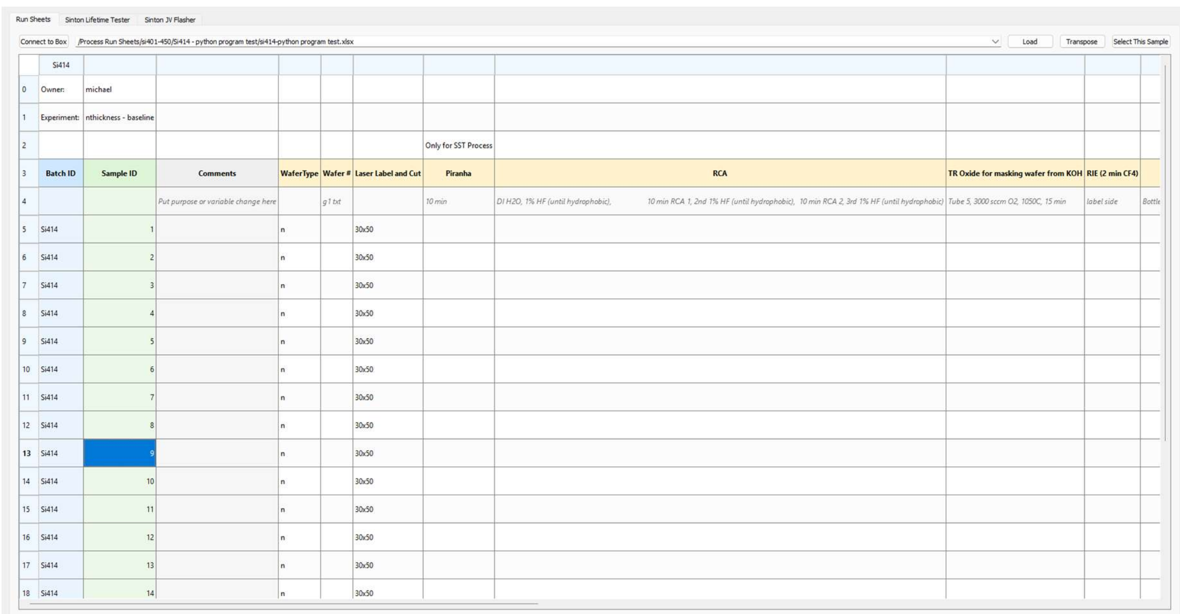


Figure 2: The Run Sheets page. A user has navigated to the run sheet file for the Batch ID Si414.

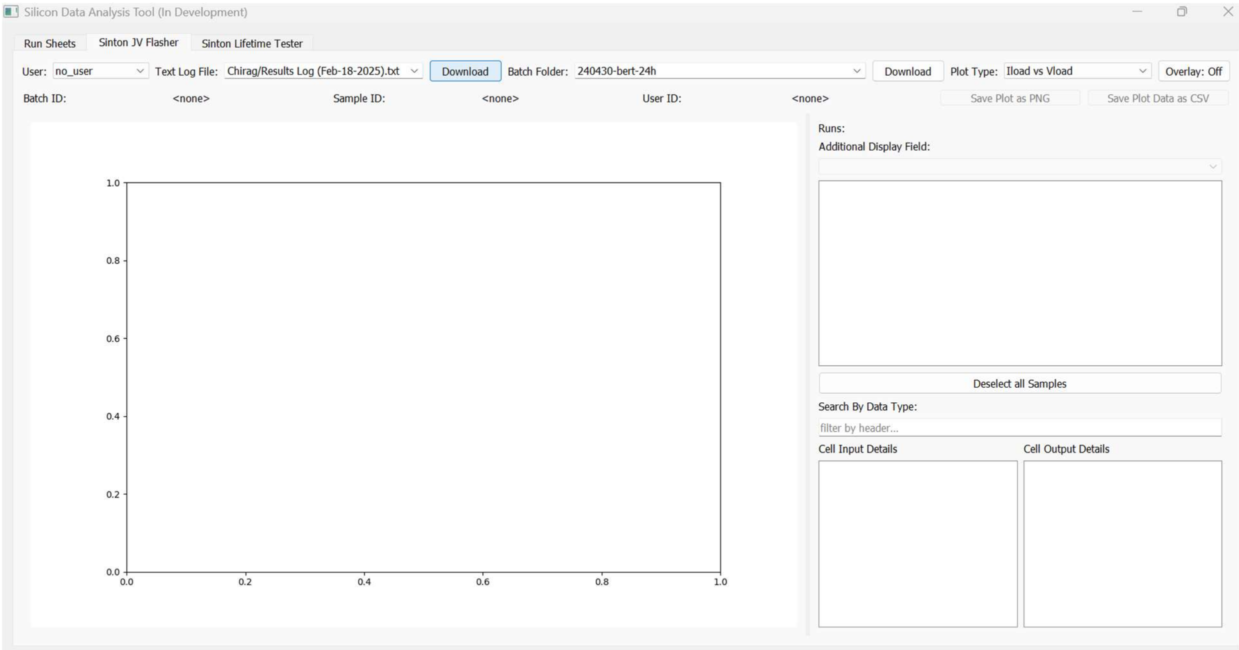


Figure 3: The default Sinton JV flasher page GUI. No files have been downloaded in this capture.

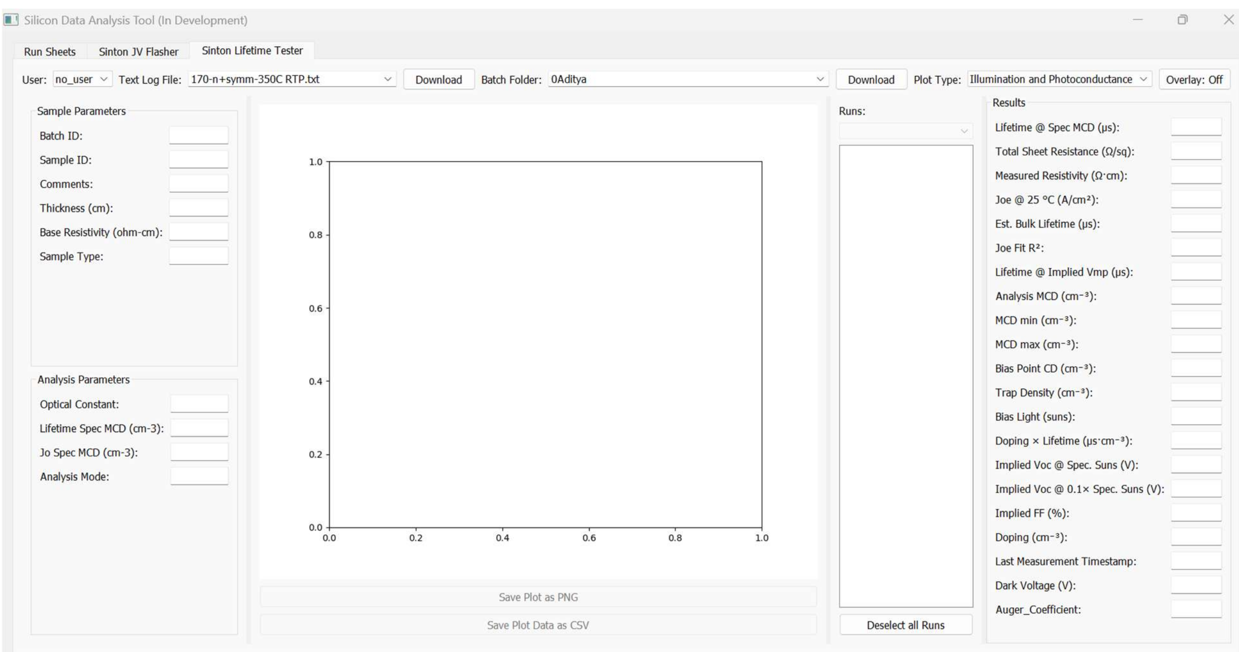


Figure 4: The The default Sinton lifetime page GUI. No files have been downloaded in this capture.

C. User experience

The design goal of tool was to make as much of the data analysis process automatic and efficient as possible. Researchers have no reason to learn how to use a tool if it won't benefit them. This step of the process brought in a major challenge: reseraachers each have their own naming habits and data organization preferences. To make automation work, a unified scheme for naming and saving data and for creating run sheets had to be established. Should this scheme be followed by users, the "run sheet sample selection" user path is almost entirely automatic and robust. The "manual dropdown search" path serves as an alternative for users who don't use run sheets or are accessing old data that does not follow the scheme.

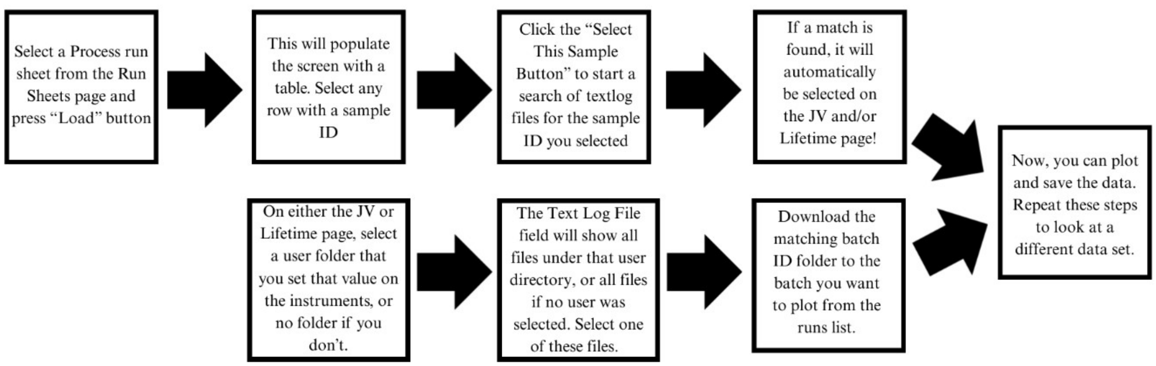


Figure 4: High level flowchart of program flow when using the tool on either path.

1. Manual dropdown search

The "manual dropdown search" user path allows users to traverse the file tree on either the multiflash or the lifetime instrument through a series of dropdowns and menus that reduce the scope of the search at each step. For example, users start by selecting the top level user directory they use on the instruments. This reduces the scope of reading files from the VAST storage to files only within that user directory.

The second step after user selection is selecting a text log file. These files are named by date, which is not necessarily helpful to users. This was the main motivation for the "run sheet" user path, but as stated previously, that is not always an option. Users select and download one of these files from the cloud server to their local device, which populates a run list. The run list shows the batch and sample id of all rows in the text log files and functions as the menu for actually plotting data.

The third step is downloading a batch folder. Similarly to the text log files, the folders are filtered to only those under the selected user directory. When the batch folder is downloaded, all the plot data .txt files under it are also downloaded. This allows the user to quickly select any sample ID from the runs list, seen in Figure 5, and instantly plot the data. So long as the naming and organization scheme is followed during characterization, there should be exactly one plot data file for each row in the text log file. On the backend, the code looks for the plot file name that matches the selected run's batch and sample ID.

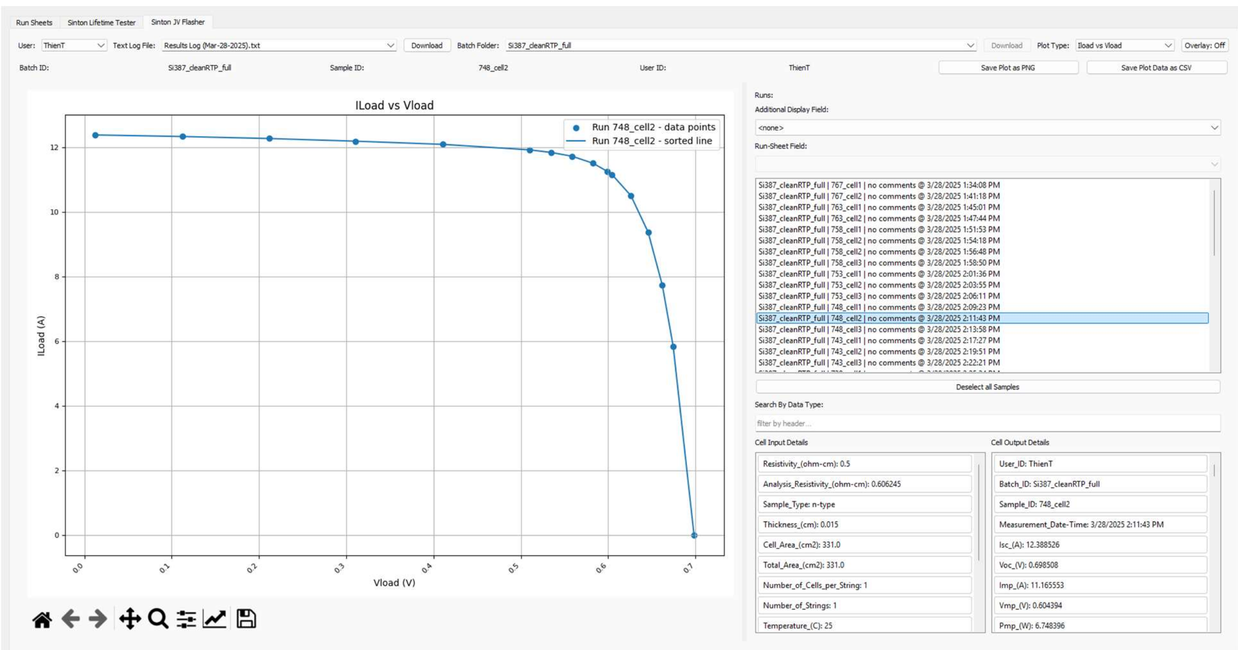


Figure 5: Sinton multiflash page with a user, textlog file, batch folder, and sample selected. The runs list (pictured right) allows users to switch between plotting samples with a click.

The tool also has features to overlay data from multiple samples in a batch at once. Users can overlay plots from the plot options menu, or select any two data values as x and y axes to create a scatter plot. Figure 6 shows an example of plot overlay, and Figure 7 shows an example of scatter plotting.

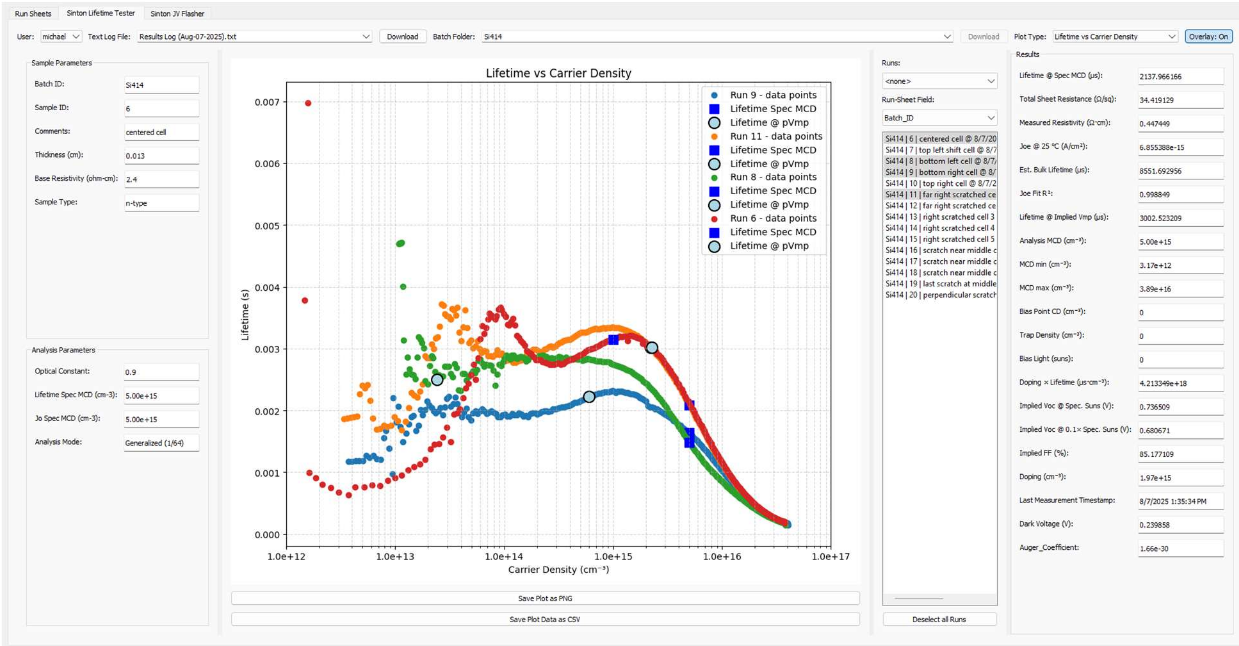


Figure 6: Overlaid sample plots of carrier lifetime vs density on the Sinton lifetime page.

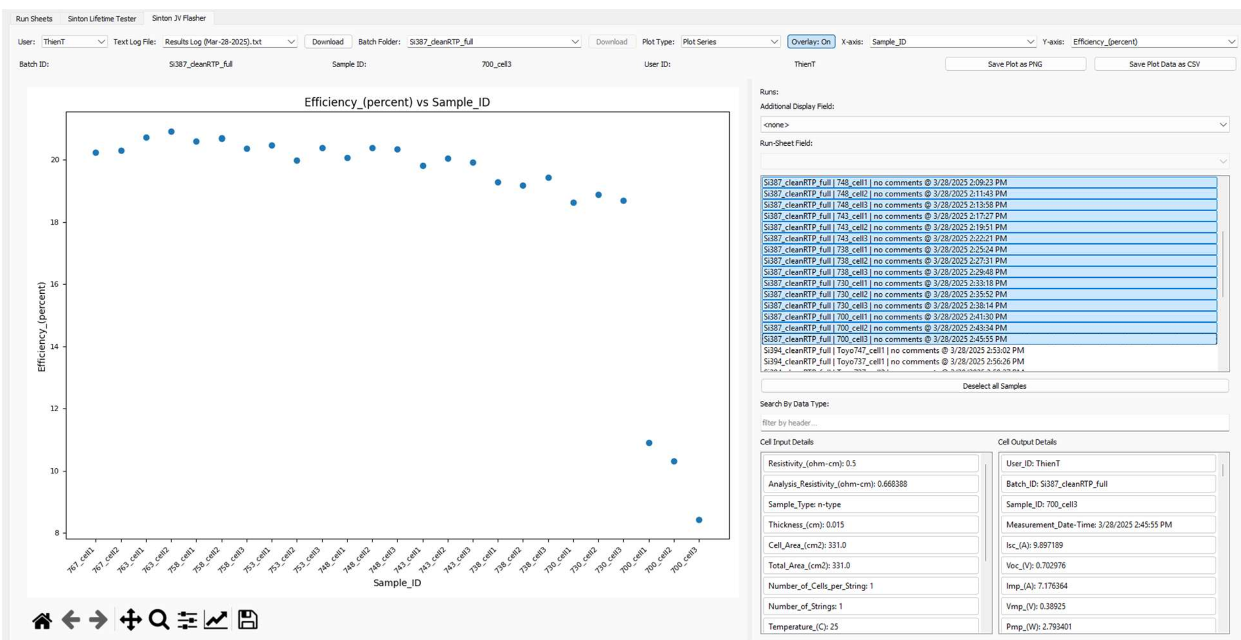


Figure 7: Automatic scatter plot of "Sample ID" (x-axis) vs "Efficiency (percent)" (y-axis) created using the series plotting tool.

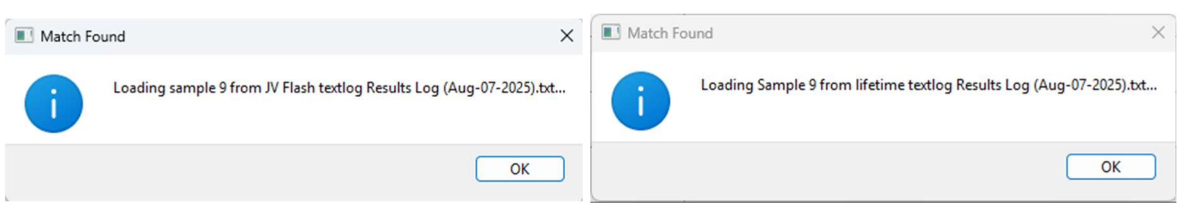
2. Run sheet sample selection

Run sheets are organizational tools used by the Silicon PV team at NREL to track the fabrication steps in experiments. These sheets are organized by user, batch ID, and sample ID just like the Sinton tools, so integration is logical. This feature also ties the whole experimental process together with characterization, keeping research organized.

The run sheets are stored on Box instead of on the same VAST cloud server as the characterization data. Accessing files in Box requires open authorization (OAuth) through the BoxSDK. When the OAuth process starts, the user is routed to a login webpage, and on successful authorization, is routed back to the app. From there, the names of all runsheet files are pulled from Box via a HTTP request and populated into a searchbar. When a user selects a file, the runsheet gets replicated inside the GUI with some additional formatting to assist users (see Figure 2).

Selecting a row on the runsheet will save the user ID, batch ID, and sample ID associated with that row as global variables. Then, the tool temporarily downloads each textlog file in the user's directory and searches for a batch and sample ID match. If a match is found, the search stops and that file is kept. If not, the search continues until all files have been checked. After a match is found, the backend will programmatically mimic the manual dropdown search steps by filling in the user, textlog and batch combos, then downloading all files. From there, the user can operate the tool the same way as they would from the manual dropdown search mode.

There are a few advantages of this feature opposed to manual dropdown search. The first is that it is typically much faster. Automatic file searching for a matching ID takes only a few seconds, and downloading the batch folder is almost instant. Additionally, manually finding the correct files can be difficult if the user does not know the textlog date. The run sheet tool will always find a match if there is one.



Figures 8,9: The messages displayed to the user when an ID match is found between the run sheet and the Sinton data files.

Another built in feature of using run sheets instead of manual search mode is that column headers from the sheet, which specify the fabrication steps that may be varied in the experiment, can be captured and used as axis labels in the series plotting tool. This allows users to easily see a correlation between experimental inputs and output with just a few clicks.

III. OUTOMES

The research result from developing this tool is a greatly improved way to access and analyze characterization data from the Silicon PV team at NREL. Previously, researchers needed to download files manually to a thumb drive and import to a tool like Excel to visualize data. Analyzing results from multiple samples at a time was even more difficult, as researchers needed to merge data sets together into excel. Now, these steps are consolidated into one tool with automation and an easy navigate GUI. As described in the process section, visualizing trends between experimental variation and characterization outputs is easily done, which is extremely valuable to developing future experiments and justifying experimental success.

A bonus of developing this tool is that it encouraged researchers to develop and follow a unified scheme for collecting and organizing data. Before this tool, there was little reason to use run sheets, and even less reason to follow one format. Now, to use the tool with its best capabilities, researchers must be consistent. This greatly increases the ease of developing tools like this one in the future, and may prove useful for future projects with artificial intelligence or retrieval augmental generation which work better with structured data.

The tool was distributed to researchers using UV as a package manager, and Github for version control.

IV. CONCLUSION & FUTURE RESEARCH

The analysis and organization of characterization data at NREL was a process that needed improvement. This tool serves that purpose, and will improve both the throughput and accuracy of results reporting from experiments by using the automation built into the tool, and by creating a structured data organization scheme for researchers to follow.

The tool is currently in use by researchers, so benefits experienced from using the tool and any bugs or limitations can be reported for future improvements. Future research will mainly include expansion of the tool to other instruments or research groups, for example in perovskite or III-V materials PV research. The tool is structured to be modular, so linking a new instrument or adding a new feature is independent of existing code and relatively simple. The only unique code would be file processing as results files from each instrument have differences.

V. ACKNOWLEDGEMENTS

I would like to thank my mentor, David Young, as well as Robert White, Thien Truong, Bennet Meyers-Im, Bill Nemeth, the rest of the Silicon PV group at NREL, Pat Kline, Rachael Schowalter, and Marcus Giron. This work was supported in part by the U.S. Department of

Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program.

VI. REFERENCES

1. P. Verlinden, D. L. Young, G. Xiong, M. O. Reese, L. M. Mansfield, M. Powalla, S. Paetel, R. M. France, P. T. Chiu, and N. M. Haegel, "Photovoltaic device innovation for a solar future," *Device* **1**, 100013 (2023).
2. A. O. M. Maka, T. Ghalut, and E. Elsaye, "The pathway towards decarbonisation and net-zero emissions by 2050: The role of solar energy technology," *Green Technologies and Sustainability* 2, 100107 (2024).
3. M. Victoria, N. Haegel, I. M. Peters, R. Sinton, A. Jäger-Waldau, C. del Cañizo, C. Breyer, M. Stocks, A. Blakers, I. Kaizuka, K. Komoto, and A. Smets, "Solar photovoltaics is ready to power a sustainable future," Joule 5, 1041–1056 (2021).
4. M. Woodhouse, D. Feldman, V. Ramasamy, B. Smith, T. Silverman, T. Barnes, J. Zuboy, and R. Margolis, *Research and Development Priorities to Advance Solar Photovoltaic Lifecycle Costs and Performance*, Tech. Rep. NREL/TP-7A40-80505 (National Renewable Energy Laboratory, Golden, CO, 2021).
5. Jinko Solar Co., Ltd., "27.02%! JinkoSolar Sets New TOPCon Conversion Efficiency Record," *JinkoSolar* (web press release), June 26, 2025, https://www.jinkosolar.com/en/site/newsdetail/2645.
6. T. J. Silverman and H. Huang, *SETO Multi-Year Program Plan* (U.S. Department of Energy, Solar Energy Technologies Office, Washington, DC, May 2021).
7. C. M. Mule, D. W. Mulder, P. C. Taylor, W. Nemeth, S. Theingi, M. Page, M. Kaupa, S. Agarwal, and P. Stradins, "Characterization of dangling bond defects at the crystalline Si/SiO_x interface in a polycrystalline Si passivating contact solar cell at room temperature with electrically detected magnetic resonance spectroscopy," *APL Mater.* **13**, 071103 (2025).
8. F. Feldmann, G. Nogay, P. Löper, D. L. Young, B. G. Lee, P. Stradins, M. Hermle, and S. W. Glunz, "Charge carrier transport mechanisms of passivating contacts studied by temperature-dependent J–V measurements," *Sol. Energy Mater. Sol. Cells* 178, 15–19 (2018).
9. C. B. Honsberg and S. G. Bowden, "Photovoltaics education website," https://www.pveducation.org (2019).
10. Y. H. Kao, L. Kazmerski, K. G. Lynn, and A. Mascarenhas, "Photovoltaics Characterization: An Overview," in *Photovoltaics for the 21st Century*, edited by V. K. Kapur, G. P. Ceasar, R. D. McConnell, A. Rohatgi, and D. Carlson, *Electrochemical Society Proceedings* Vol. 99-11 (Electrochemical Society, Pennington, NJ, 1999), pp. 289–303.
11. R. R. White and K. Munch, "Handling large and complex data in a photovoltaic research institution using a custom laboratory information management system," *MRS Online Proc. Libr.* 1654, 1104 (2014).