

Interactive Visualization of Massive LCLS Datasets

Russell Marasigan¹, Frédéric Poitevin²

¹University of California, Irvine, ²Linac Coherent Light Source, SLAC

Introduction

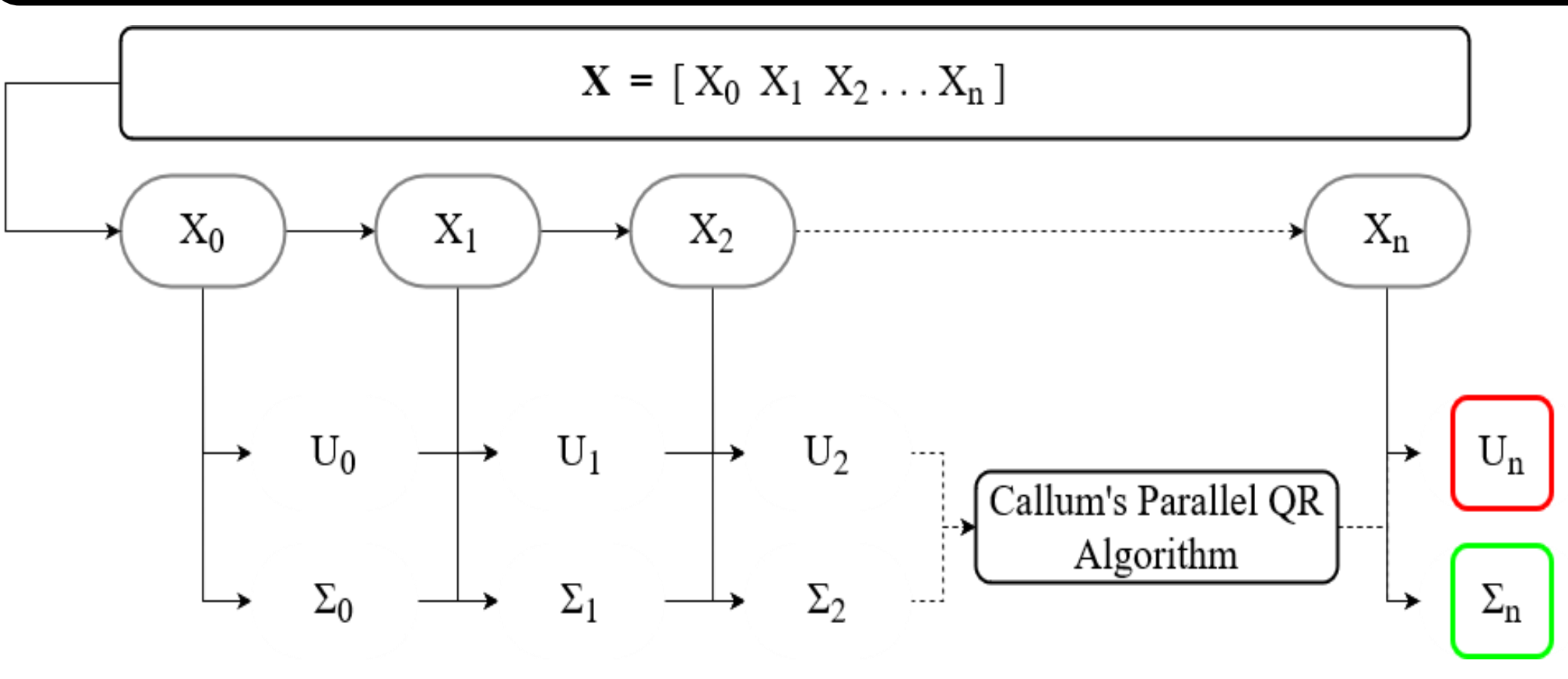
Motivation: LCLS users need interactive tools to navigate their experimental data in real time.

Principal Component Analysis (PCA): Dimensionality reduction technique that identifies the most significant modes of variation in high-dimensional datasets.

Parallelized Incremental PCA (PiPCA): An implementation of PCA initiated by Callum Hepworth (LCLS Summer Intern 2022) that allows PCA to run on massive datasets utilizing multiple CPUs.

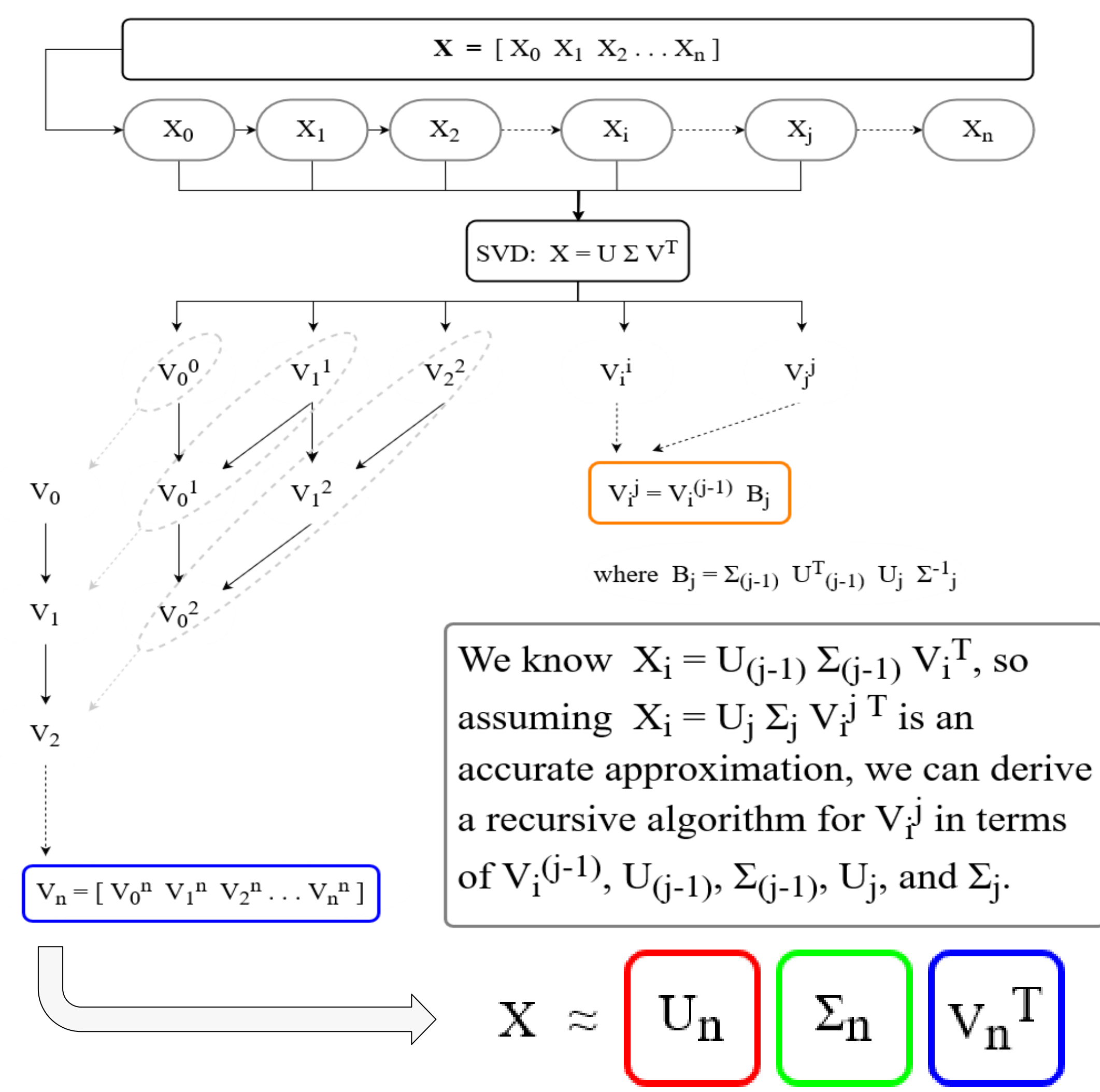
Internship Goal: Finalize the implementation of PiPCA and build an interactive dashboard to visualize its output model.

Previous PiPCA Algorithm

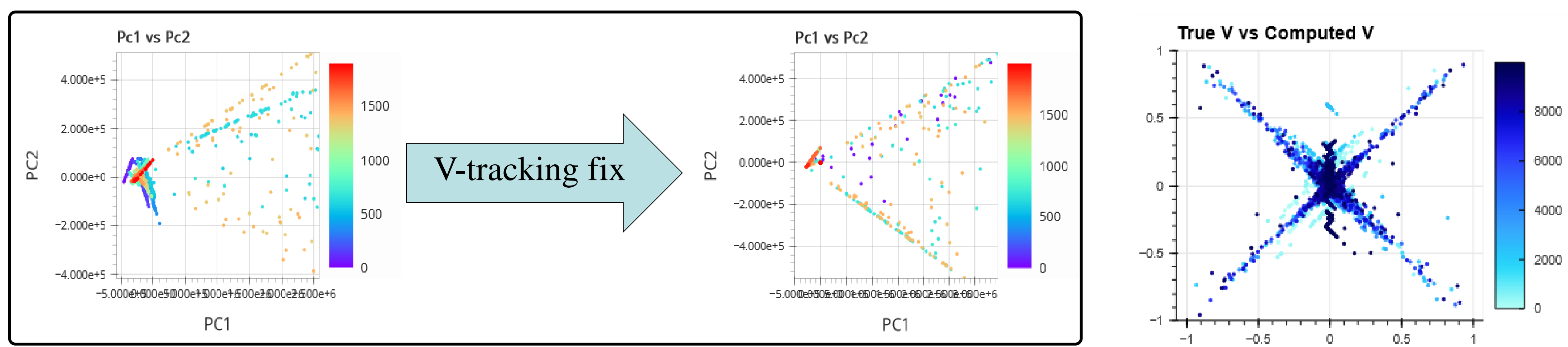
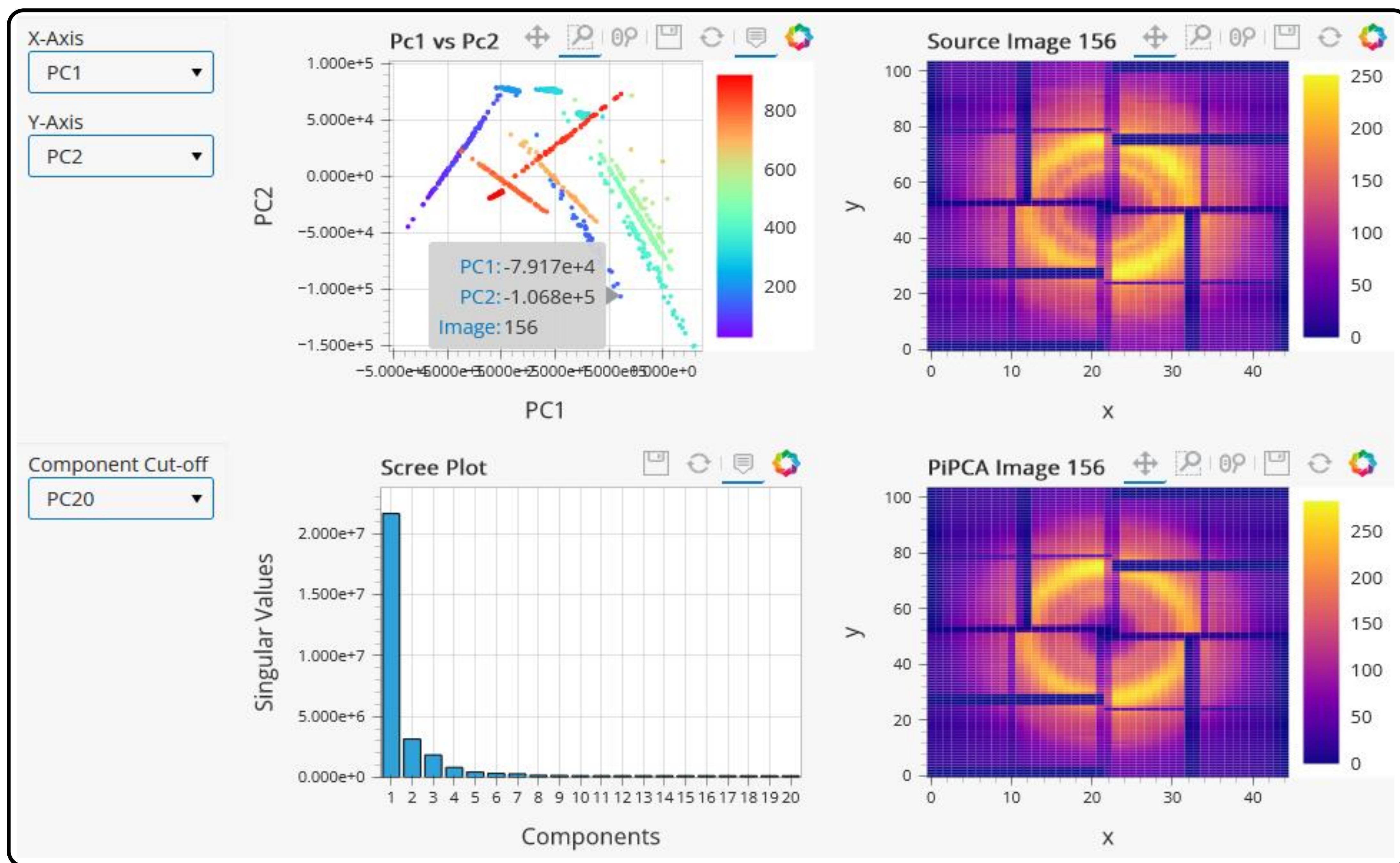


Recursively updates matrices U and Σ for each image batch

Tracking Matrix V through PiPCA



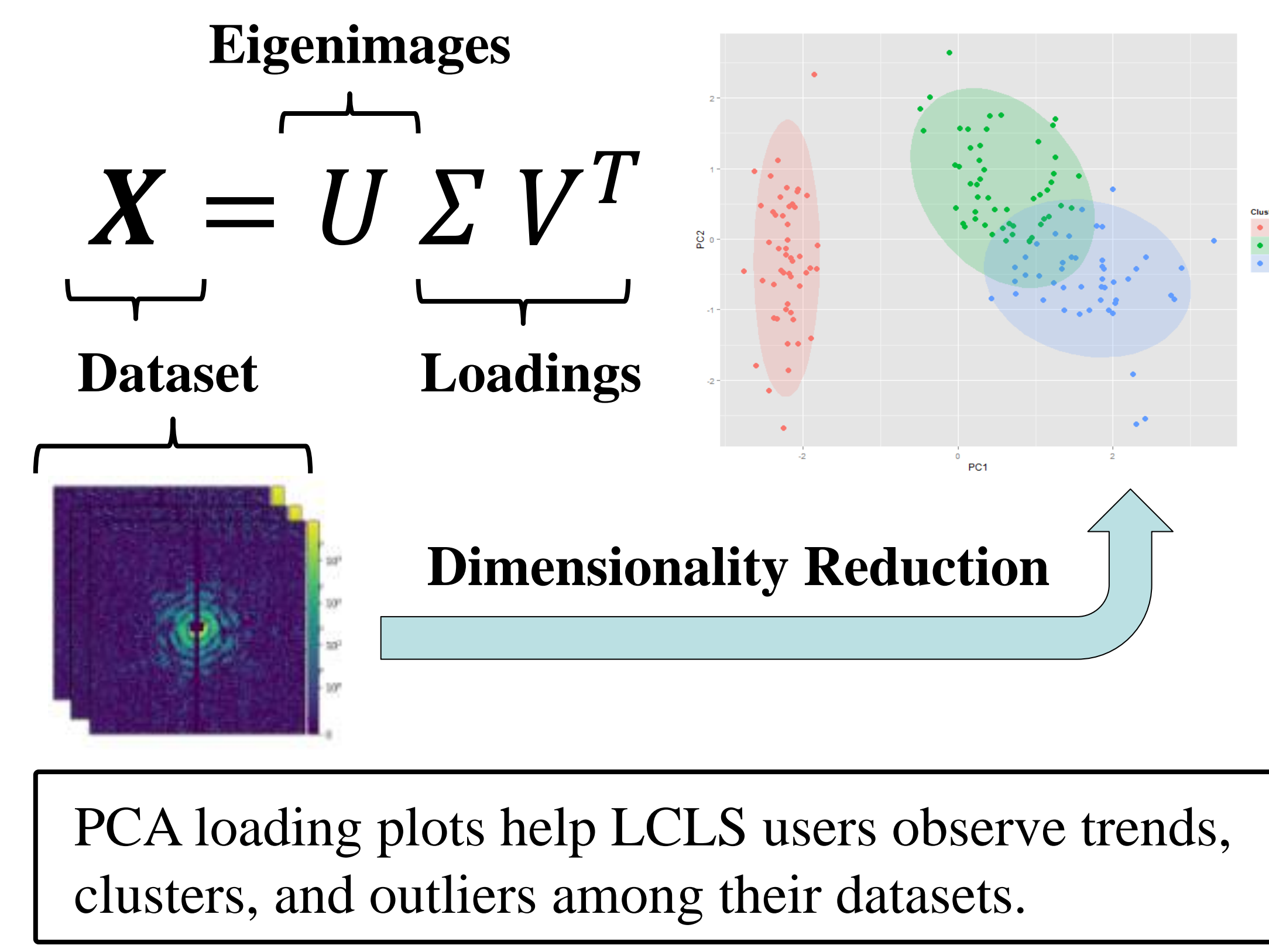
Results



- Developed an interactive dashboard enabling LCLS users to visualize and better understand the data they collect during their experiments.
- Fixed the updating of loadings (V-tracking) by deriving and implementing an incremental algorithm that updates past data with latest available model.
- Software engineering skills I have developed and applied:
 - Data management on clusters (SDF & S3DF)
 - Git, GitHub, & HDF5 file management
 - VS Code & Shell
 - Holoviews, Panel, & Bokeh
 - MPI Communicator & SLURM

Selecting a component cut-off will update the scree plot to display the singular values up to that component. It will also recalculate and update the reconstructed PiPCA image in order to see how much variation is lost in a specific truncation.

PCA Model



Future Steps

- Measure performance for real-time analysis
- Port to GPUs for LCLS-II readiness
- Add clustering capability
- Continue work on model inheritance with John Winnicki so the dashboard can be used by other dimensionality reduction techniques.
- Continue drafting a user's manual to explain the purpose, functionality, and outputs of these processing algorithms.

Accessibility

btX (BeamTime with X-rays) is a GitHub repository consisting of miscellaneous functions for aiding LCLS experiments.

You can access the pipca.py script using the QR code on the right.

Acknowledgments

Special thanks to my mentor Frédéric Poitevin for his guidance, insight, and coaching. I would also like to thank Callum Hepworth for initiating this project and John Winnicki for his collaborations. As well as Ariana Peck and Omar Quijano for their support and encouragement. Finally, thank you to the LCLS internship organizers for the opportunity to work here this summer!