Automated Analysis and Denoising of Neutron Scattering Data

David Bruhwiler*, Matthew Kilpatrick, Boaz Nash, Evan Carlin (RadiaSoft),

Christina Hoffmann, Zachary Morgan, Andrei T Savici, Bogdan Vacaliuc (ORNL)

A. Kuhn, J. Mensmann, P. Messmer, M. Nienhaus, S. Roemer, D. Tatulea (NVIDIA)

*bruhwiler@radiasoft.net

radiasoft

OAK RIDGE National Laboratory

NVIDIA

American Conference on Neutron Scattering (ACNS)

Knoxville, Tennessee

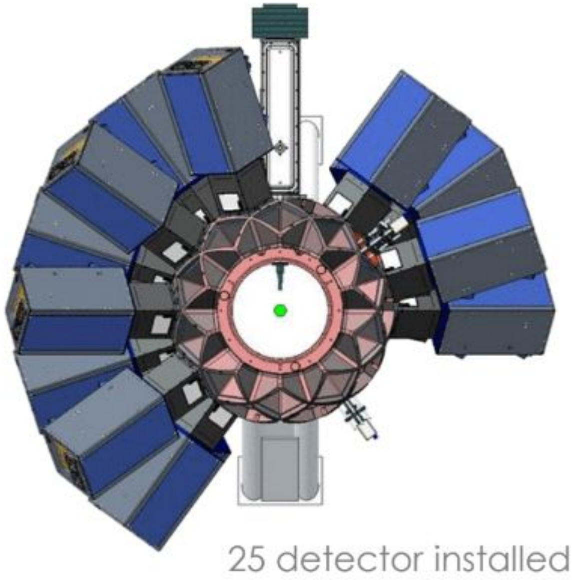
26 June 2024

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0021551.

U.S. DEPARTMENT OF ENERGY | Office of Science

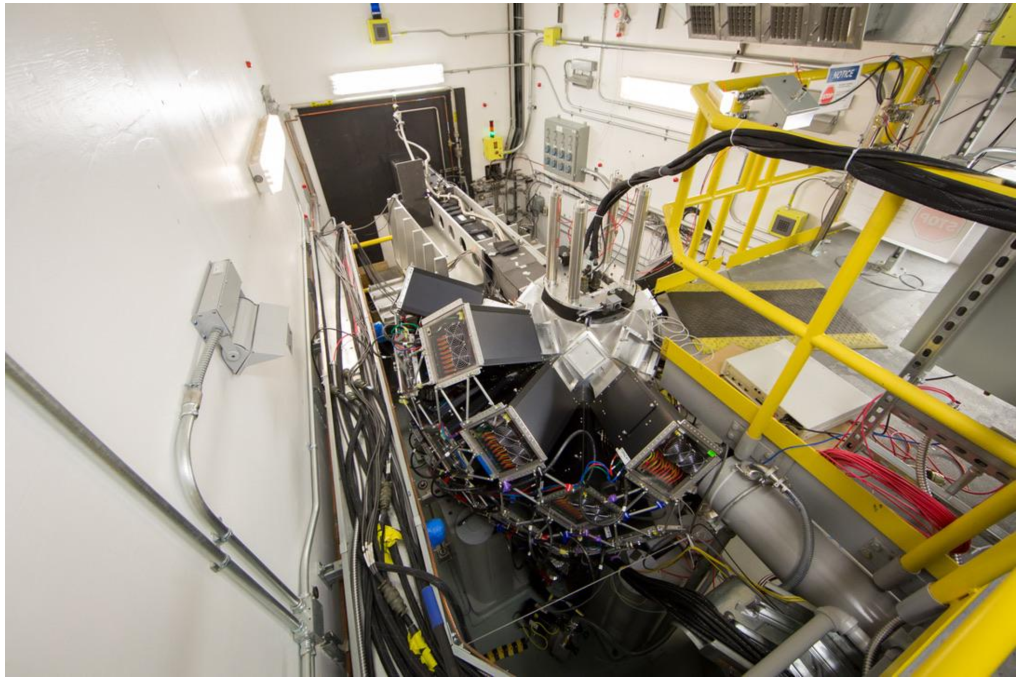
TOPAZ Single-crystal Diffractometer

- Located at ORNL and receives neutrons from the Spallation Neutron Source
- Samples can be measured with high precision for volumetric sampling in reciprocal space
  - Samples are rotated to measure all aspects of the lattice
  - Temperature control from 5 K - 450 K
  - Broad Q coverage

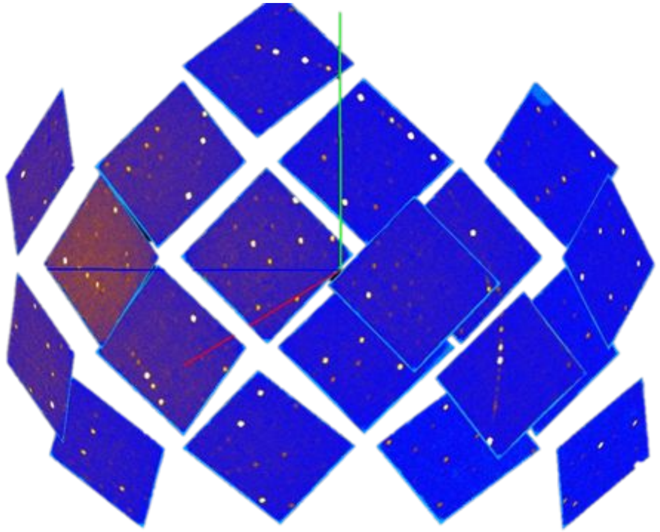


OAK RIDGE National Laboratory

https://neutrons.ornl.gov/topaz



Analysis methods for TOPAZ data

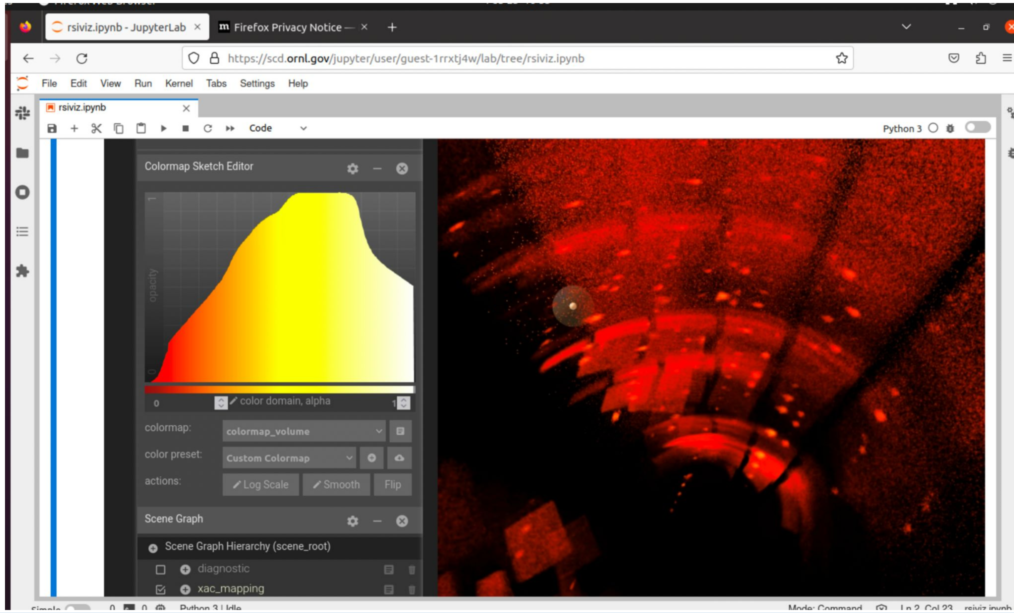


- Mantid, https://www.mantidproject.org
  - Open source community-developed application
  - Algorithms use raw data processed within 2D slices of the 3D data

- Some challenges faced by the community
  - Large datasets can be up to 100 Gb in size and current tools limit interactivity
  - Displaying data requires a high level of user interaction
  - 2D slices may miss key features of the diffraction data

- Using ML & leveraging Mantid to automate analysis & viz
  - DBSCAN used to automate & improve Bragg peak identification
  - 3D UNet used to automate detection of aluminum signature

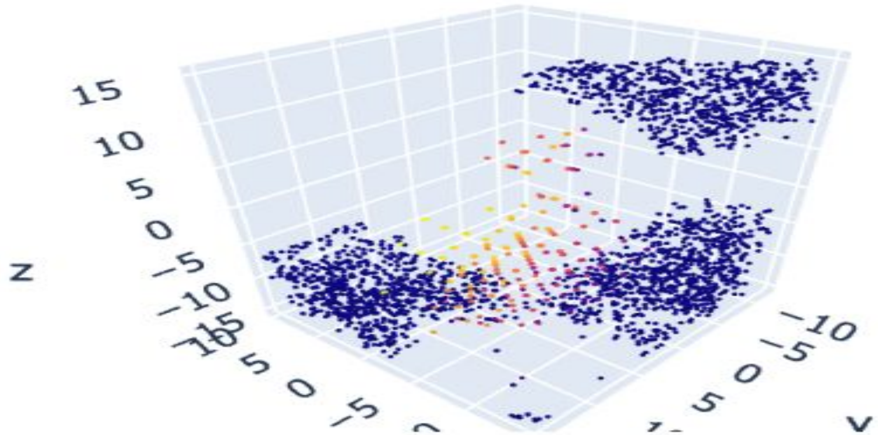
- We are improving interactivity of the data viz
  - collaboration with NVIDIA Omniverse/IndeX development team
  - working towards combination of 3D viz and automated analysis
  - IndeX 3D Scientific Data Viz, https://developer.nvidia.com/index
  - Enables 3D browser-based viz of streaming data (deployed)

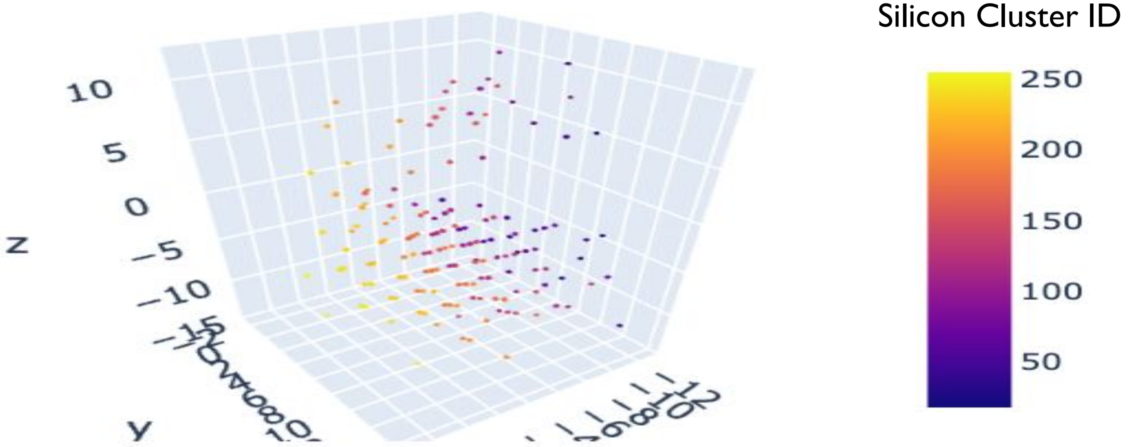


Previous work: analysis of scattering data for a Silicon crystal

- Preprocessing the data
  - Data is transformed to q-space in the usual manner
  - We consider a 3D histogram of neutron counts in q-space
  - Typically use 1,000 cubed histograms these are voxels to NVIDIA Omniverse / IndeX
  - 5,000 cubed voxels will be practical on a single, modern GPU with latest version of IndeX software

- DBSCAN (a density based clustering algorithm) is used for Bragg peak detection & analysis
  - Bragg peaks are automatically identified as oblong clusters (Mantid requires user input)
  - Mantid takes peaks and calculates UB matrix & Miller indices
  - DBSCAN / Mantid approach is 10x faster, more accurate & requires no user input
  - Timescale: 10-20 min

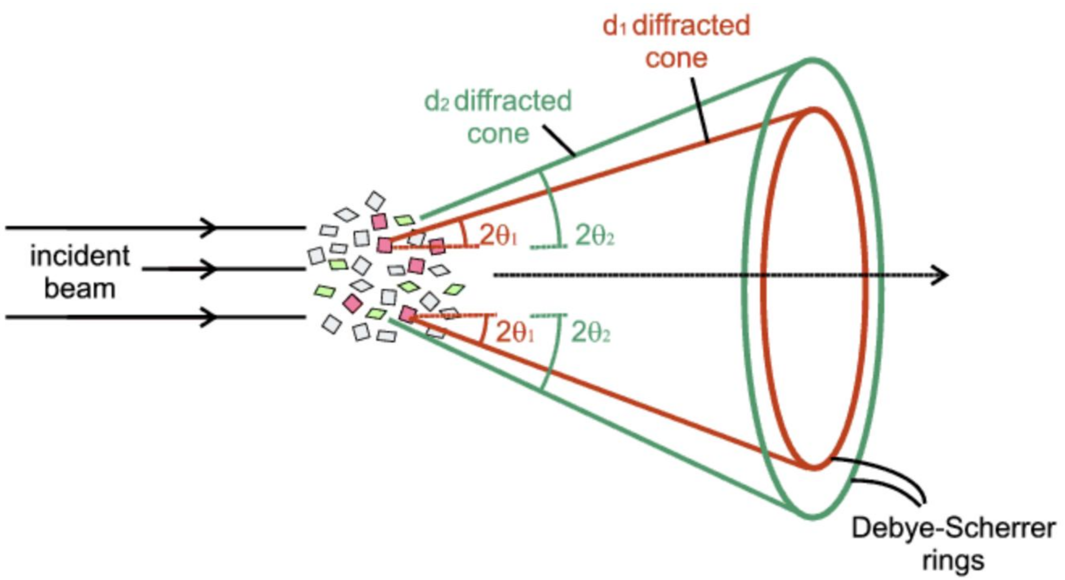


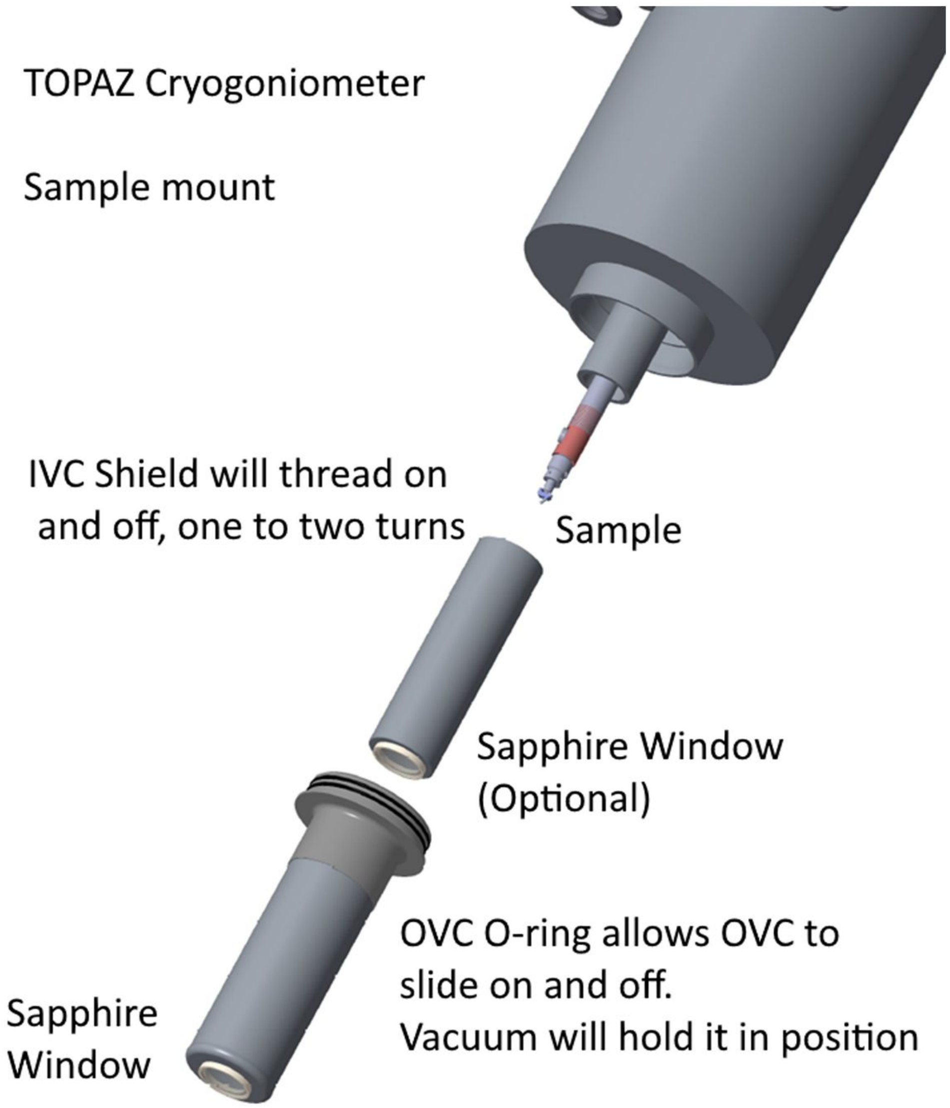


https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html#sklearn.cluster.DBSCAN

Background Removal Processes

- Sample mounting
  - Crystal sample can be tested at low temperatures
  - Aluminum cylinders have a randomly oriented crystal structure
  - These rings will scatter neutrons into a sphere or ring
- Scattered neutrons
  - Each plane orientation will create rings within reciprocal space
  - Translates into a sphere in 3D representation



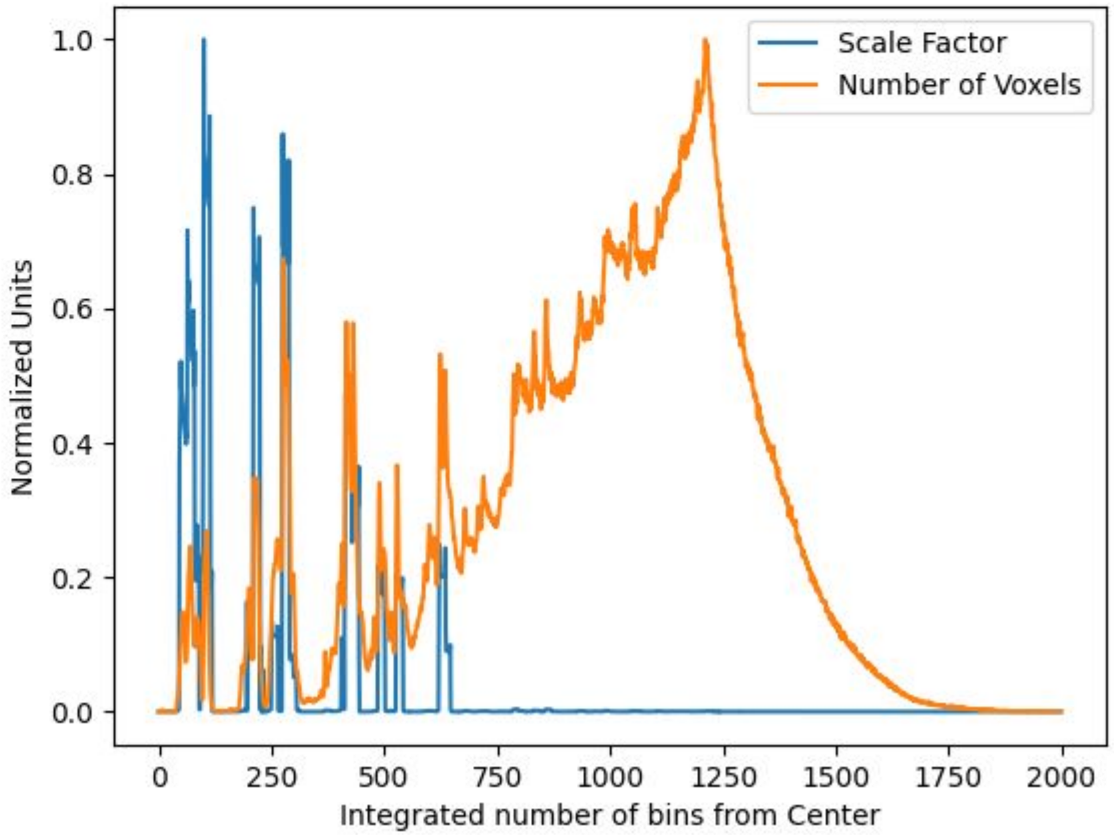


Identify Accurate Ground Truths

- Properties of Aluminum Rings
  - Monochromatic in neutron intensity
  - Concentrated at a fixed radius
  - Varies from sample to sample
- Define a scale factor
  - Simple clustering
    - Tune DBSCAN parameters to get rings
  - Gradient based definitial for clustering
    - Normalized by the mean number of voxels

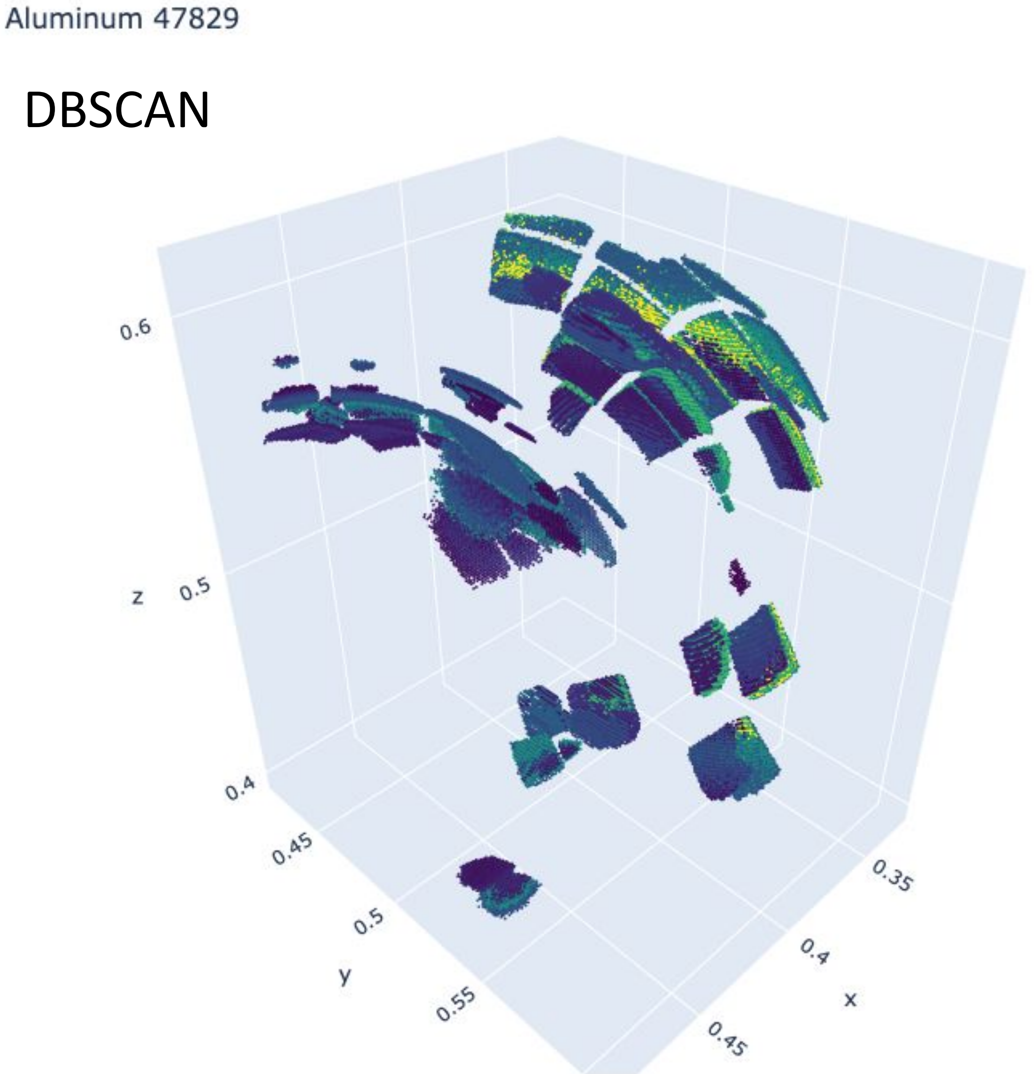
$$S_{rings}(R_w) = \frac{max(\nabla_w R) \cdot \bar{R}}{1 + r}$$

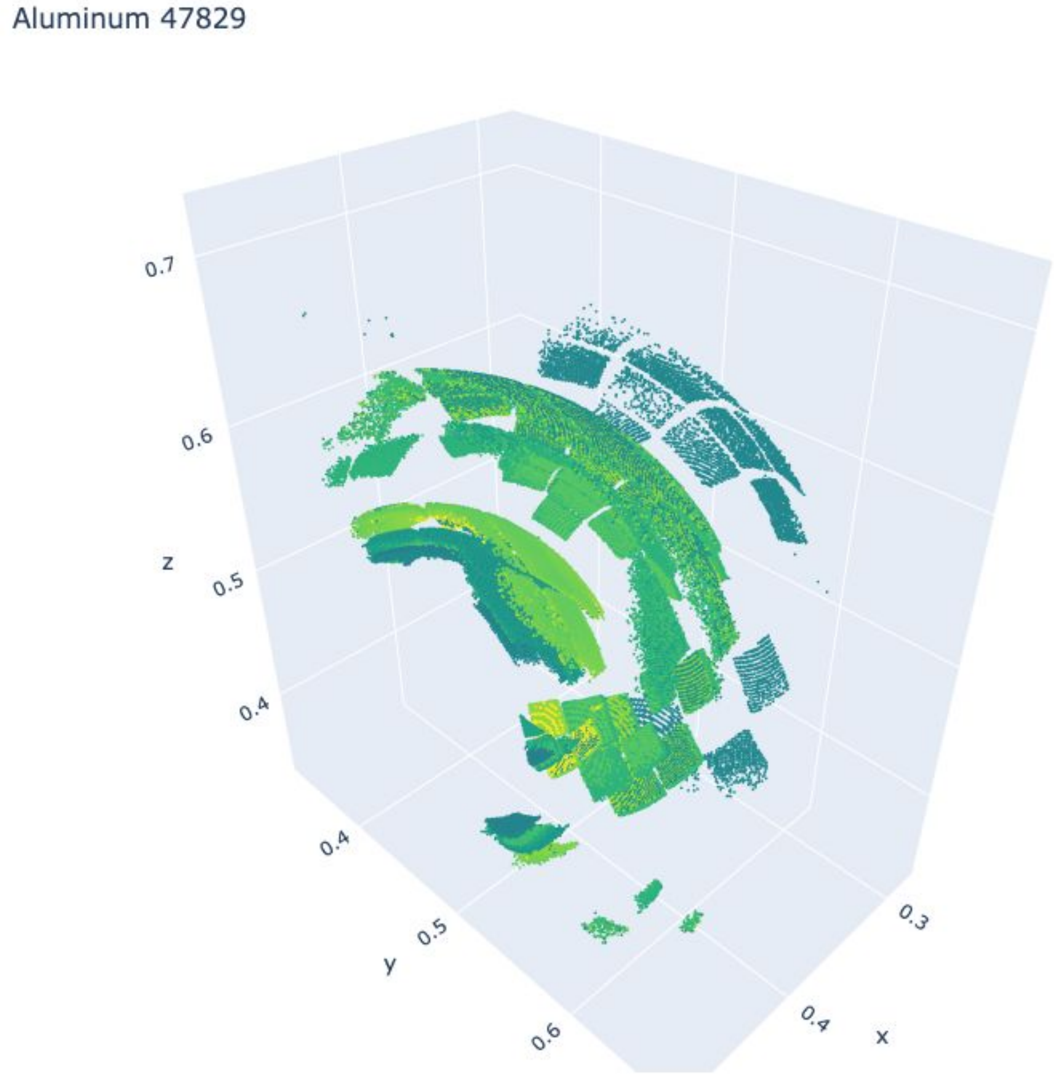
- How do these two methods compare?



Aluminum Labels

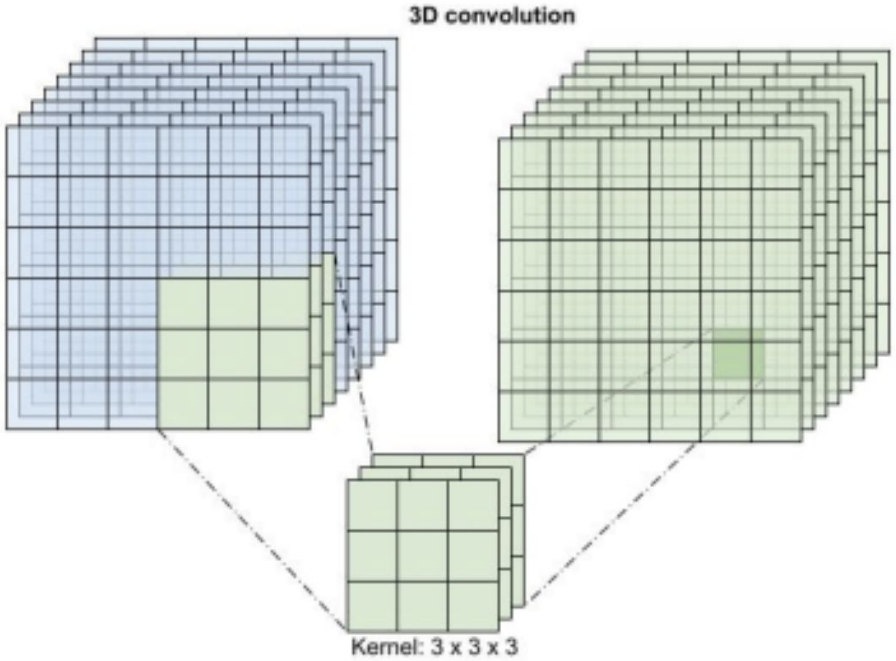
DBSCAN

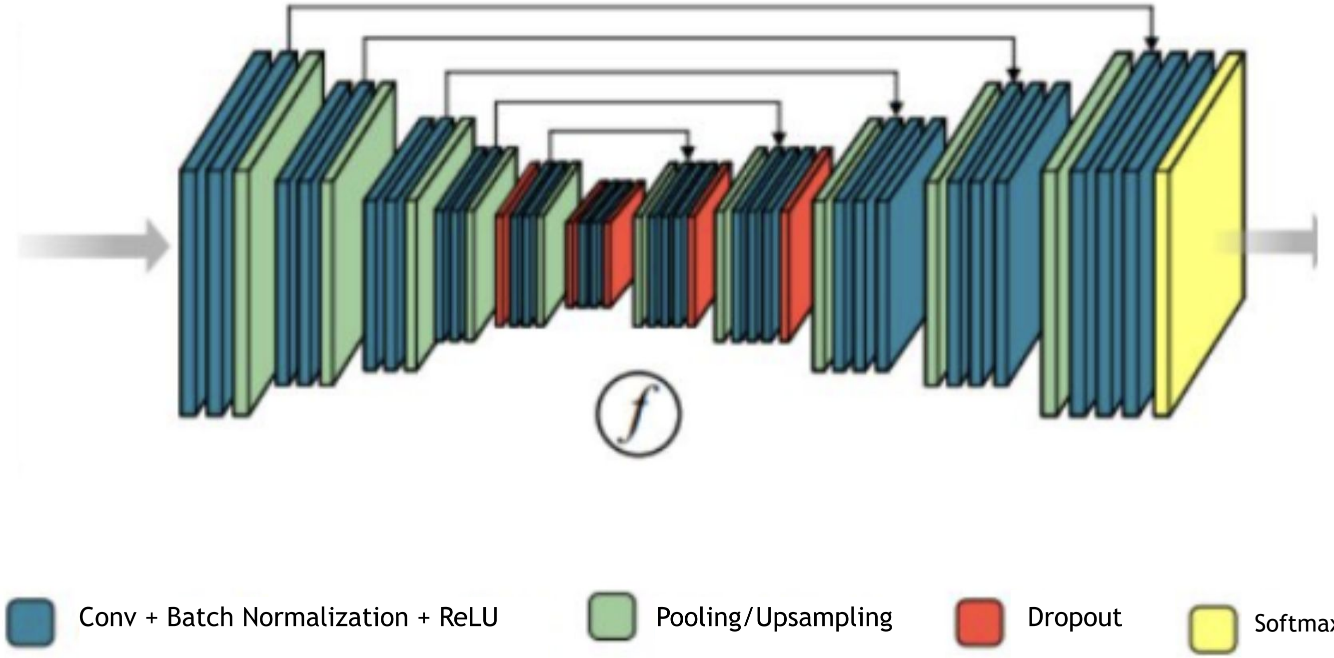




Identification of Aluminum Rings

- Building a neural network
  - Identify a 3D CNN to train and predict aluminum rings
  - General Unet mainly used for image segmentation
    - Fully connected
    - Contraction and expansion paths
    - 3D Unet used for learning dense volumetric segmentation from sparse annotations

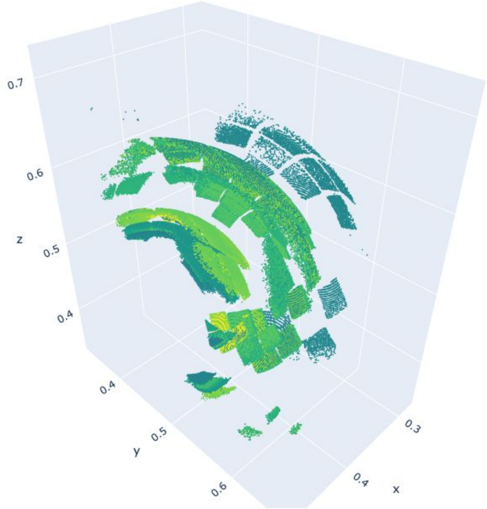




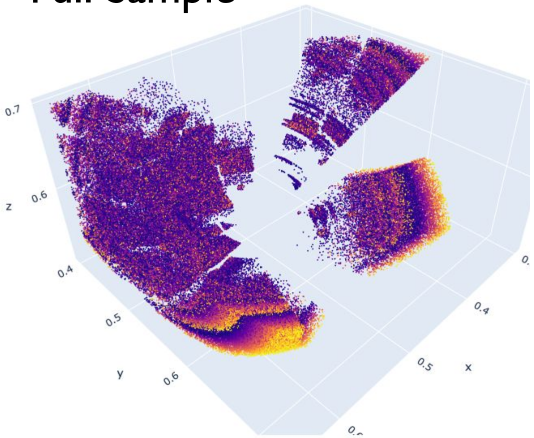
Identification of Aluminum Rings

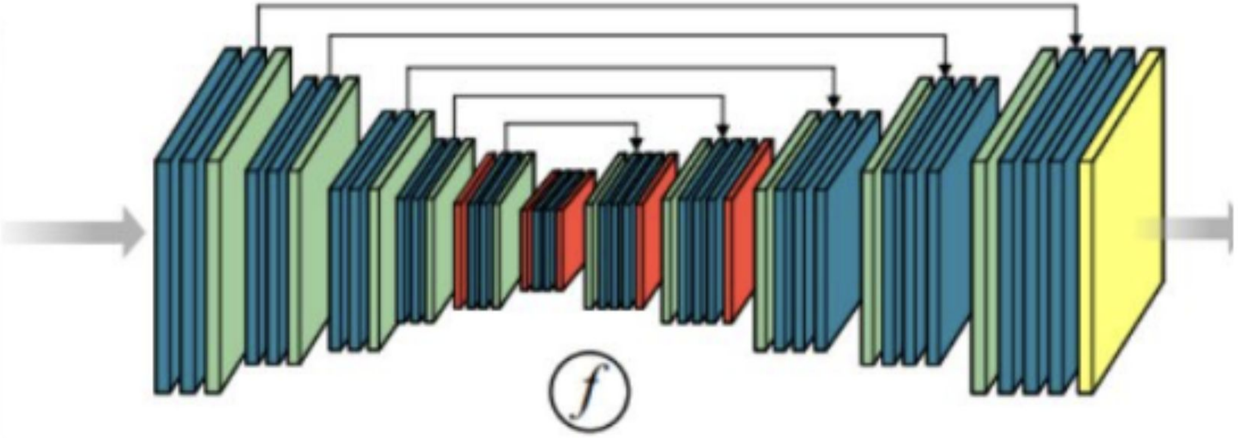
- Building a neural network
  - Identify a 3D CNN to train and predict aluminum rings
  - General Unet mainly used for image segmentation
    - Fully connected
    - Contraction and expansion paths
    - 3D Unet used for learning dense volumetric segmentation from sparse annotations

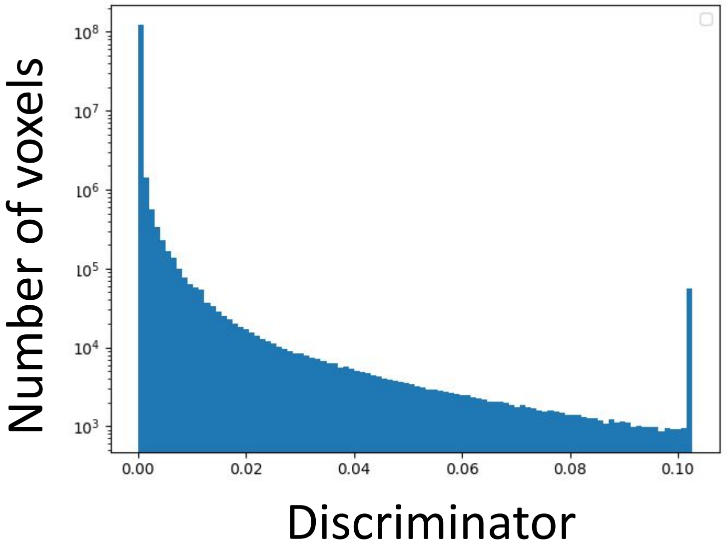
Labelled Rings



Full Sample



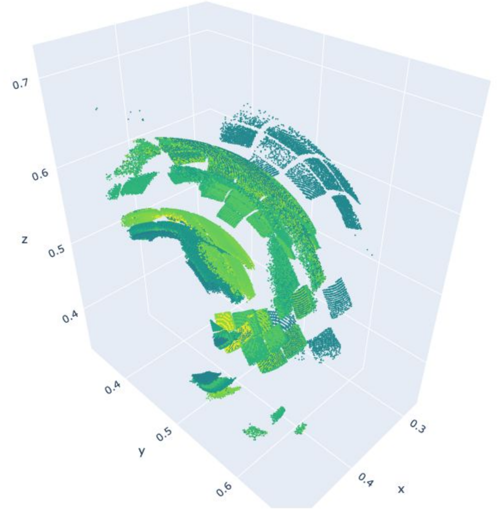




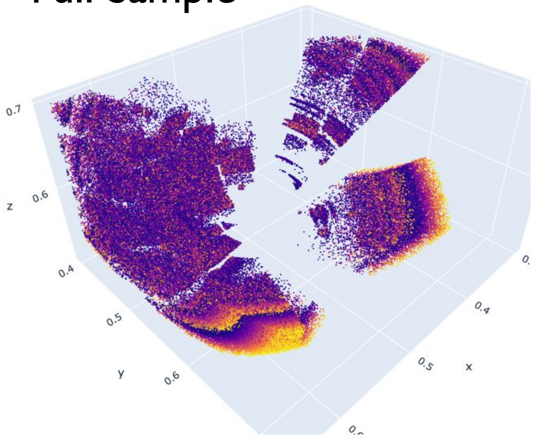
Identification of Aluminum Rings

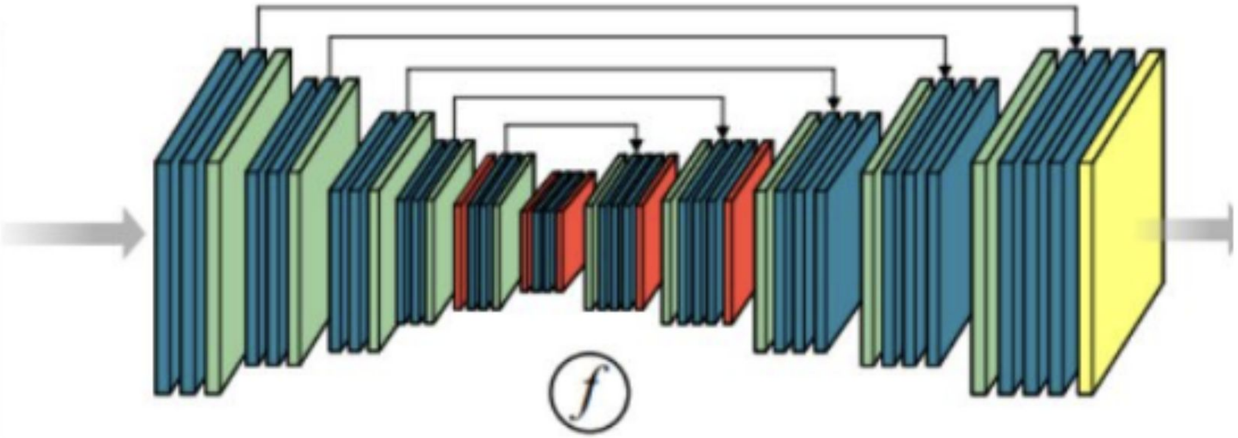
- Building a neural network
  - Identify a 3D CNN to train and predict aluminum rings
  - General Unet mainly used for image segmentation
    - Fully connected
    - Contraction and expansion paths
    - 3D Unet used for learning dense volumetric segmentation from sparse annotations

Labelled Rings

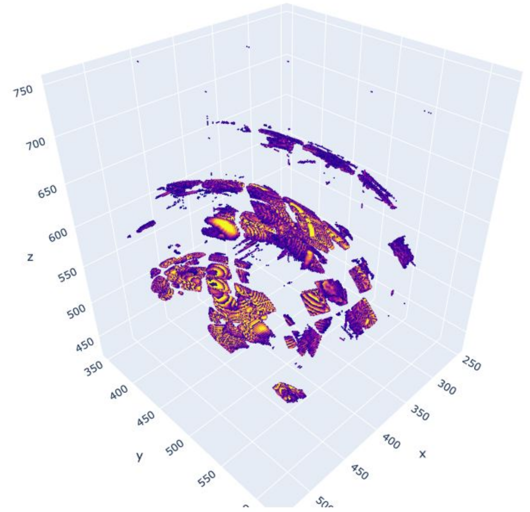


Full Sample



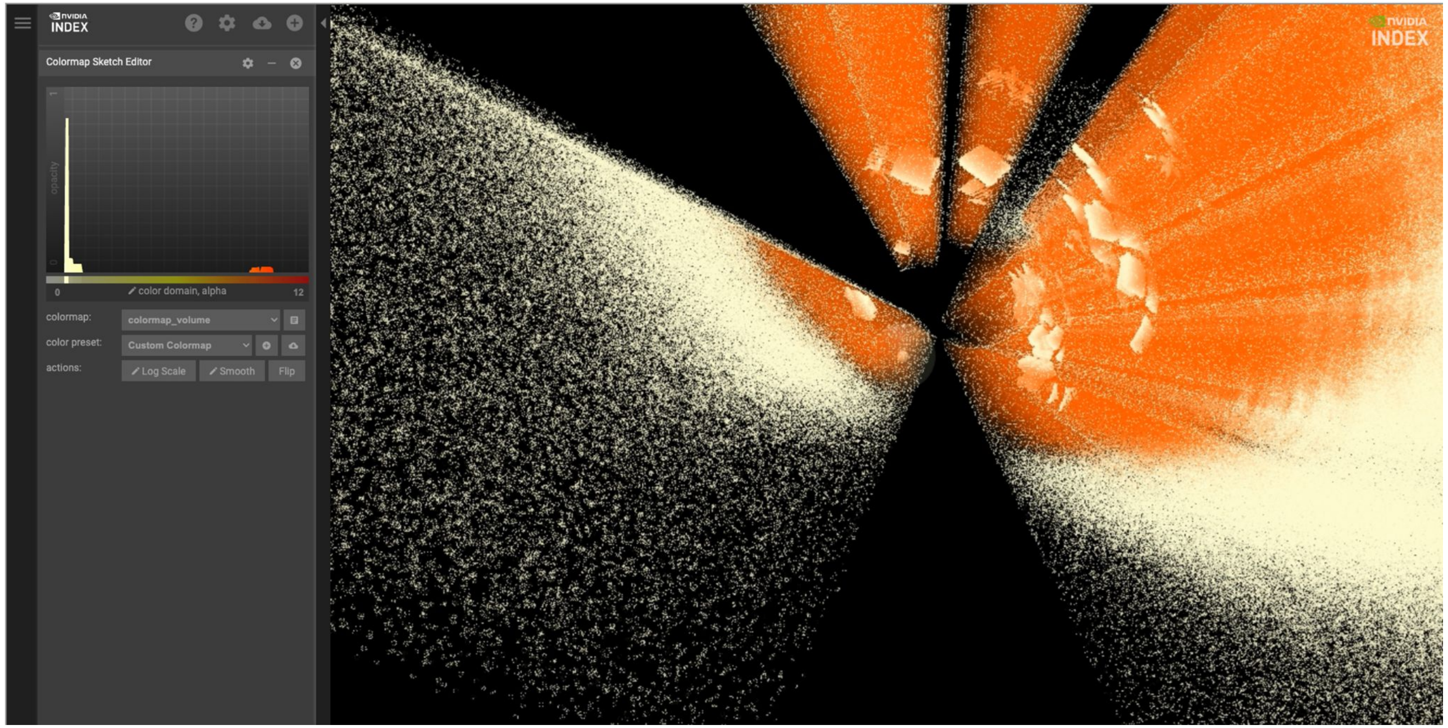


Predicted Rings



NVIDIA's Omniverse/IndeX: Browser-based 3D visualization

- Omniverse / IndeX is an interactive 3D volumetric visualization framework
  - rotate, pan, zoom into large 3D volumes
  - executes remotely on a GPU server
  - streams video to the user's browser
- TOPAZ event data is converted into a voxel database
  - voxels are a regular 3D grid
  - each is colored and/or made transparent based on the number of neutron events within it
  - colors & transparency can be dynamically edited
- The voxel database can be dense or sparse
  - large fraction of voxels are empty, so the sparse data format is very efficient
- Interactive response is robust to WiFi quality
  - even with a poor network connection, the typical delay after a mouse click is $<$ 0.25 s



Summary

- Our goal is to automate data analysis to enable scientific discovery
  - ML + leveraging Mantid
    - faster, better Bragg peak detection
    - deployed on-premises for TOPAZ
  - NVIDIA's Omniverse / IndeX technology
    - interactive 3D viz in your browser – in the control room, or at your home institution
    - an NVIDIA card is required, presently on the floor of the SNS experimental hall
    - deployed on-premises for TOPAZ
  - ML + NVIDIA/IndeX (the work presented here)
    - it is possible to automatically identify an unwanted aluminum signature
    - not yet deployed
  - ML + EPICS controls
    - has been used to automate sample alignment
    - demonstrated for TOPAZ & also at HIFR

U.S. DEPARTMENT OF ENERGY | Office of Science

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0021551.

- This lab-industry collaboration is facilitated by open-source software
  - RadiaSoft has an open source business model
  - Proprietary IndeX technology is free for scientific use, courtesy of NVIDIA
  - broad applications to other neutron beamlines