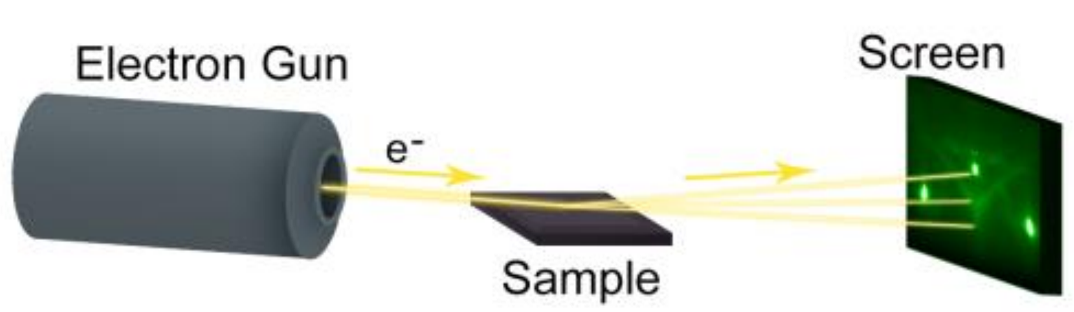
Accelerating RHEED Analysis with FPGA-Optimized Neural Networks

Matt Wilkinson, Pujan Patel, Yichen Guo, Scott Hauck, Joshua Agar, Shih-Chieh Hsu, Geoff Jones

Introduction

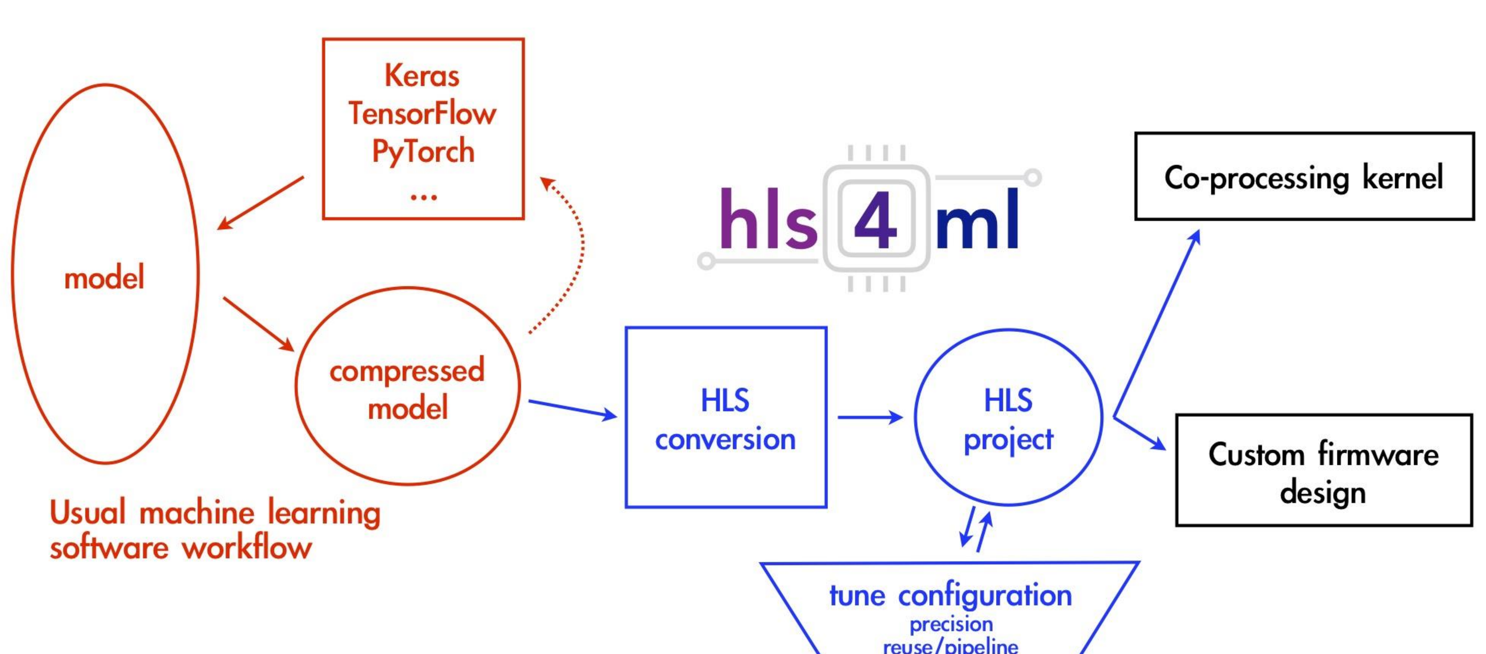
Reflection High Energy Electron Diffraction (RHEED) is a technique that uses electron diffraction to monitor thin film growth and surface structure in real-time during deposition.



An electron beam is directed at a surface, and the resulting diffraction pattern reveals information about the surface structure and crystalline quality.

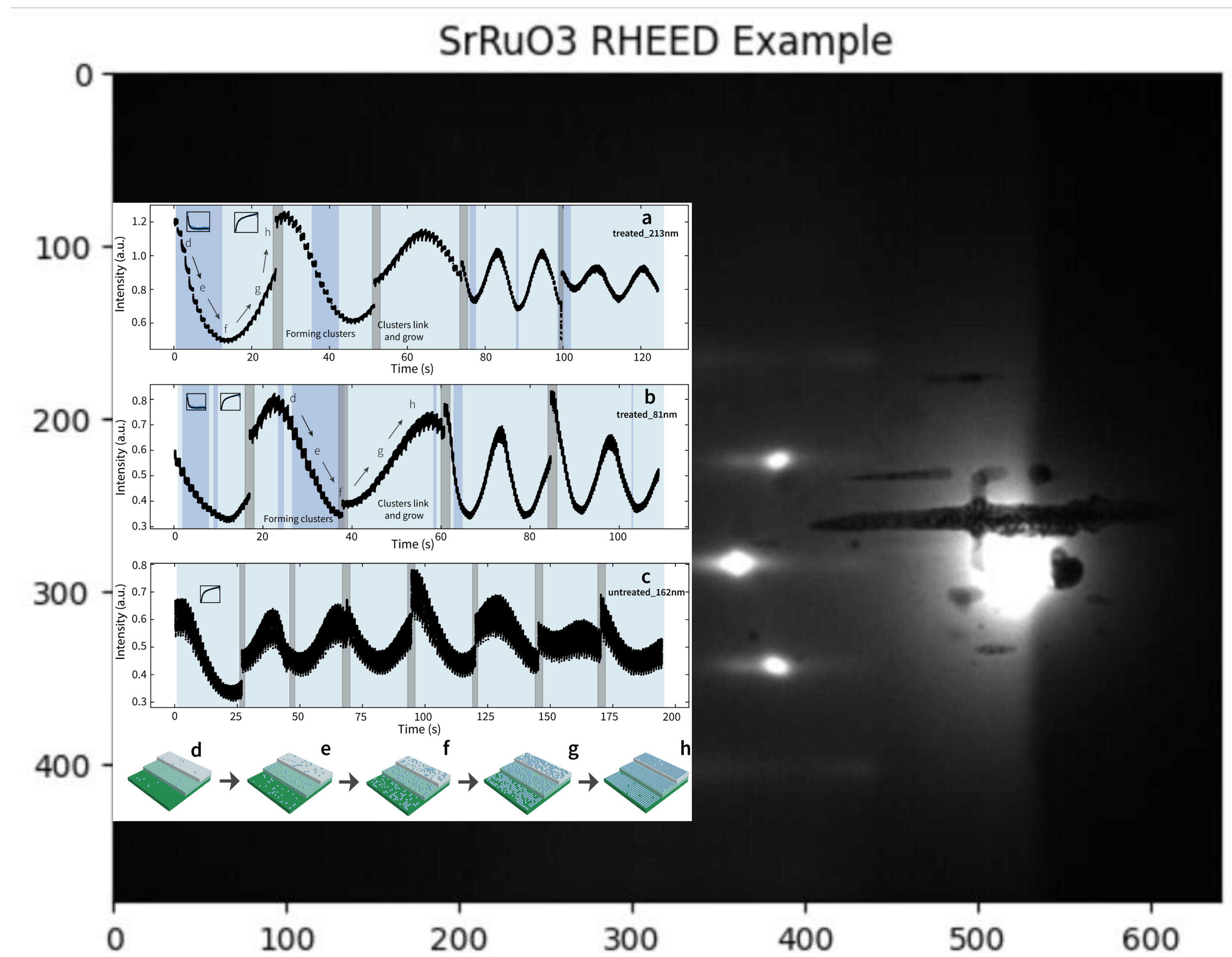
Goal

Use HLS4ML and HLS4ML optimization to deploy a machine learning system to provide real-time analysis for RHEED systems.

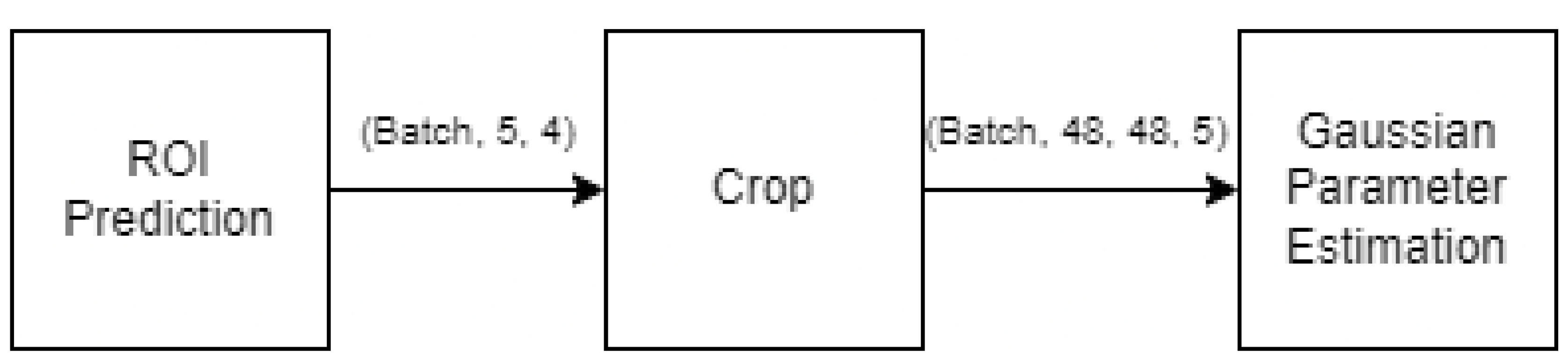


[1]

Method



The spots in RHEED images tend to be very spread out so cropping individual spots from the image before fitting parameters will save on resources.



Two cropping options:

Post-HLS4ML conversion

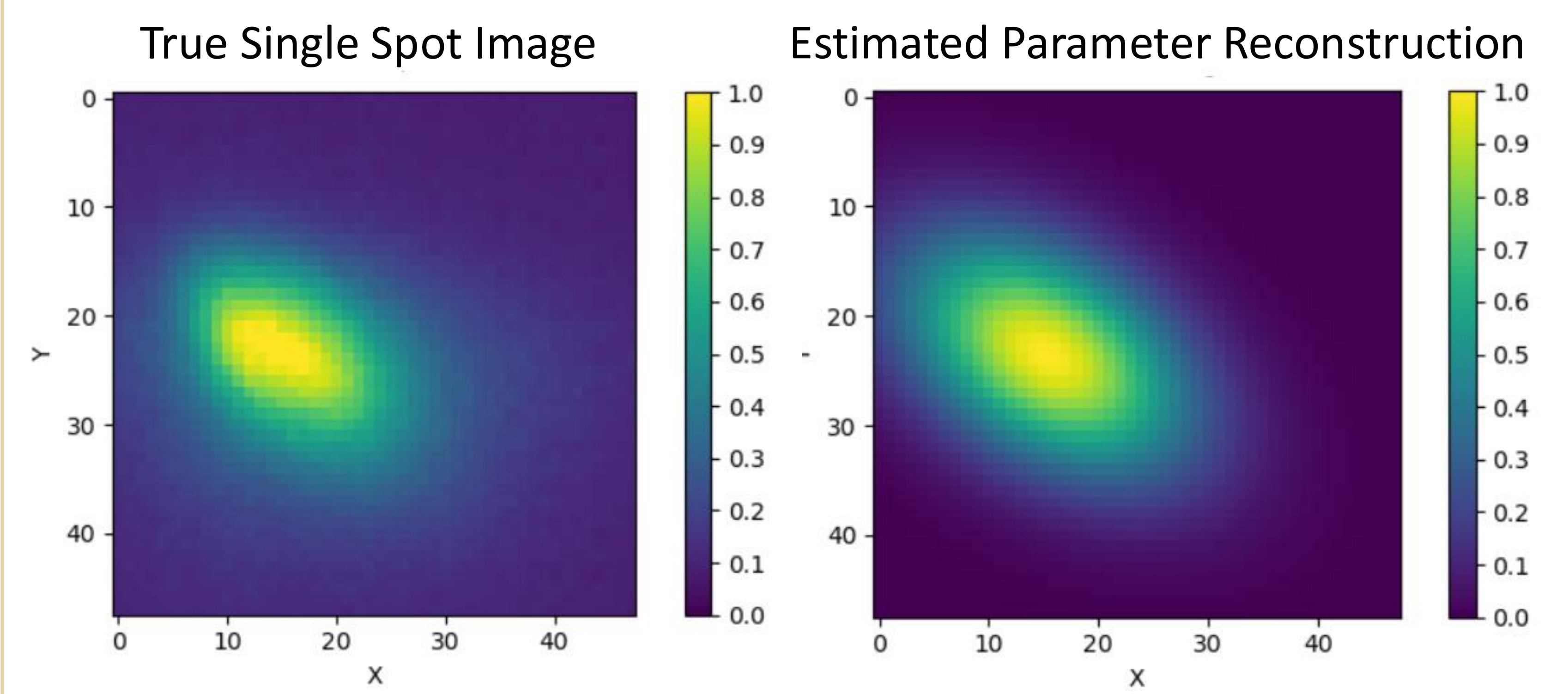
- Does not require custom HLS4ML layer
- Requires two separate neural networks
- Requires packaging in HDL to avoid overlapping naming

Pre-HLS4ML conversion

- Allows for a unified model to be trained
- Requires a custom Keras and HLS4ML layer
- Can be trained completely unsupervised on real RHEED data

Gaussian Parameter Estimation

After single spots are cropped from image, they are modeled as gaussians by predicting (mean_x, mean_y, cov_x, cov_y, rotation) with a CNN



Achieved a training loss of 0.0005566 when comparing the original image to a reconstructed image with a MSE loss

Conclusion

> FPGAs can be used to alleviate current problems with RHEED analysis like slow inference times and expensive hardware

Next Steps

> Train a unified python model and test the accuracy and resource usage against the two separate model approach

> Design a system to provide real-time control feedback to the deposition

References

[1] J. Duarte et al 2018 JINST 13 P07027

NSF: MRI: Development of Heterogeneous Edge Computing Platform for Real-Time Scientific Machine Learning (2215789)
NSF: MRI: Development of a Platform for Accessible Data-Intensive Science and Engineering (2320600)
DOE: Real-time Data Reduction Codesign at the Extreme Edge for Science
ARL: Collaborative for Hierarchical Agile and Responsive Materials (CHARM) (W911NF-19-2-0119)