RHEED Universal Translator: Bidirectional Cross-Modal Learning Between RHEED Images and XPS-Derived Stoichiometry

Jawad Chowdhury, Asraful Haque, Sumner Haaris

Oak Ridge National Laboratory, Oak Ridge, TN, USA

Abstract

Reflection High-Energy Electron Diffraction (RHEED) provides real-time feedback during thin-film growth, yet its interpretation is often qualitative. In contrast, stoichiometry derived from X-ray Photoelectron Spectroscopy (XPS) provides quantitative chemical information but is typically obtained post-growth. Bridging these modalities is essential for enabling closed-loop and inverse-design workflows in thin-film synthesis. In this work, we present the RHEED Universal Translator, a deep-learning framework that learns a shared latent representation between RHEED images and XPS-derived stoichiometry for the $SrTiO_3$ (STO) material system. The model enables bidirectional inference, supporting both stoichiometry prediction from RHEED images and generation of RHEED-like patterns from stoichiometric inputs. By enforcing cross-modal consistency, this framework establishes a scalable pathway toward cross-modal learning in oxide thin films, with potential generalization to ABO_3 perovskite systems as additional data become available.

Introduction

Quantitative control of thin-film growth requires real-time diagnostics that can be directly linked to material composition. RHEED is widely used during epitaxial growth to monitor surface structure and growth dynamics, yet its connection to chemical stoichiometry is not straight forward. In contrast, stoichiometry measurements obtained from XPS are highly quantitative but are typically decoupled from in situ growth monitoring and also a time-consuming process. Establishing a principled connection between RHEED and stoichiometry would enable predictive and inverse modeling capabilities critical for autonomous synthesis and closed-loop experimentation. Learning from multiple modalities provides a powerful mechanism for capturing relationships between disparate data sources. Motivated by this idea, we develop a framework that jointly embeds RHEED images and stoichiometry into a common latent space, enabling bidirectional cross-modal translation rather than one-way prediction.

Methodology

- **Dataset**

The dataset consists of paired RHEED images and corresponding stoichiometry values derived from XPS measurements for $SrTiO_3$ thin films. RHEED images are normalized for efficient training, while stoichiometry values are scaled to a normalized range. The paired structure of the dataset enables supervised learning of cross-modal relationships.

- **Model Architecture**

The RHEED Universal Translator adopts a symmetric encoder-decoder architecture with a shared latent bottleneck (Figure 1). Separate encoders map RHEED images and stoichiometry values into a common latent dimensionality. These representations are projected into a shared latent space, enforcing alignment between the two modalities. Two decoders then reconstruct either RHEED images or stoichiometry from the latent representation, enabling both forward and inverse translation. This design supports four inference paths: image-to-image, stoichiometry-to-stoichiometry, image-to-stoichiometry, and stoichiometry-to-image.

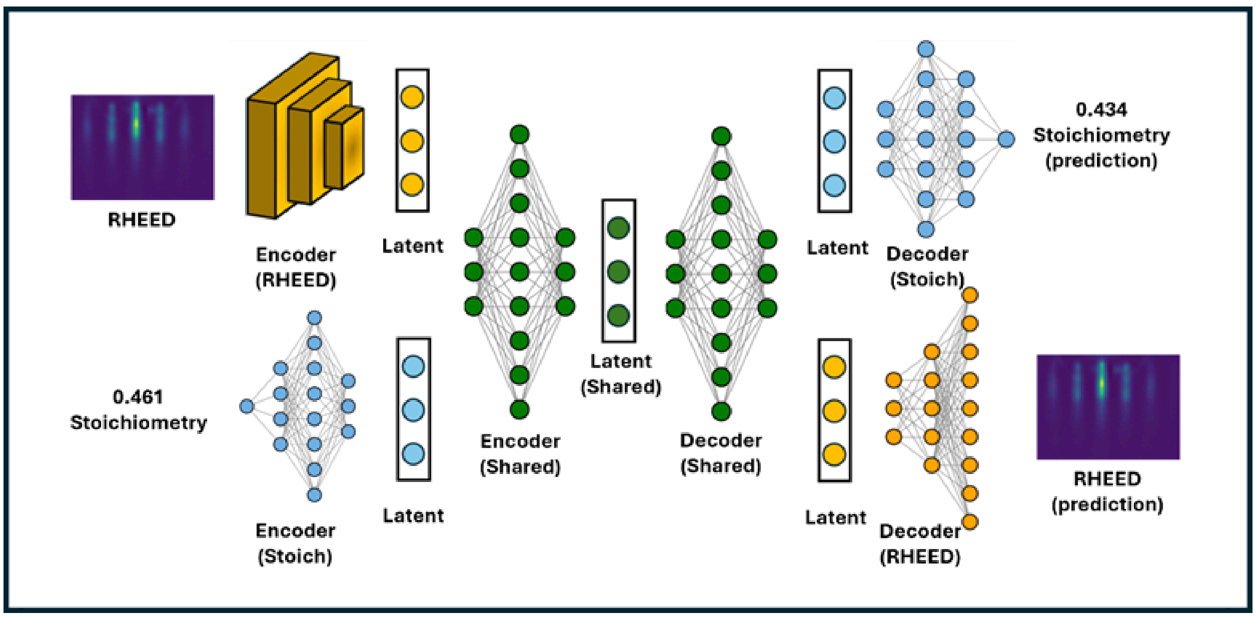

Figure 1: Architecture of the proposed RHEED Universal Translator showing bidirectional cross-modal translation through a shared latent space.

- **Training Strategy**

The model is trained using reconstruction losses for both modalities along with cross-modal consistency constraints. Mild data augmentation is applied to RHEED images during training to improve robustness, while evaluation is performed on un-augmented data to ensure unbiased assessment.

Results

The trained model demonstrates several key capabilities - including accurate prediction of film stoichiometry from RHEED images by learning systematic relationships between diffraction features and compositional variations, as well as generation of RHEED-like images from stoichiometric inputs that reproduce structured diffraction patterns consistent with experimentally observed RHEED signatures.

By enforcing a shared latent-space alignment across modalities during training, the model learns consistent representations for both images and stoichiometry, enabling physically meaningful links between structural and chemical descriptors. This enforced alignment contributes to robust cross-modal inference even when trained on a relatively small dataset, reducing reliance on superficial correlations.

Discussion and Conclusion

Our work demonstrates how cross-modal learning can bridge real-time RHEED microscopy signals and post-growth XPS-based chemical characterization through a shared learning space, enabling both forward and inverse inference pathways beyond conventional analysis. Although demonstrated on the $SrTiO_3$ material system, the framework is inherently extensible and can be generalized to ABO_3 perovskite and more complex oxide thin films as additional data become available. Overall, the RHEED Universal Translator provides a scalable foundation for bidirectional inference, supporting future inverse-design and closed-loop microscopy-guided synthesis workflows.

Resources

- **Code:** https://github.com/hasanjawad001/rheed-universal-translator
- **Hackathon:** Machine Learning for Microscopy Hackathon 2 (2025)
- **Dataset (Origin):** https://arxiv.org/abs/2501.18523, https://github.com/sumner-harris/Deep-Learning-with-RHEED

Email: chowdhurym1@ornl.gov, haquea@ornl.gov, harrissb@ornl.gov