Journal of Vacuum Science and Technology A

RESEARCH ARTICLE | MARCH 27 2025

Machine-learning-enabled on-the-fly analysis of RHEED patterns during thin film deposition by molecular beam epitaxy

Special Collection: Artificial Intelligence and Machine Learning for Materials Discovery, Synthesis and Characterization

Tiffany C. Kaspar ; Sarah Akers ; Henry W. Sprueill ; Arman H. Ter-Petrosyan ; Jenna A. Bilbrey ; Derek Hopkins; Ajay Harilal; Jijo Christudasjustus ; Patrick Gemperline ; Ryan B. Comes

J. Vac. Sci. Technol. A 43, 032702 (2025)
https://doi.org/10.1116/6.0004493

JVST A
Journal of Vacuum Science & Technology A

ARTICLE

pubs.aip.org/avs/jva

Machine-learning-enabled on-the-fly analysis of RHEED patterns during thin film deposition by molecular beam epitaxy

Cite as: J. Vac. Sci. Technol. A 43, 032702 (2025); doi: 10.1116/6.0004493
Submitted: 14 February 2025 · Accepted: 5 March 2025 ·
Published Online: 27 March 2025

Tiffany C. Kaspar,[1,a] Sarah Akers,[2,b] Henry W. Sprueill,[2] Arman H. Ter-Petrosyan,[2,c] Jenna A. Bilbrey,[2] Derek Hopkins,[3] Ajay Harilal,[2] Jijo Christudasjustus,[1] Patrick Gemperline,[4] and Ryan B. Comes[5]

AFFILIATIONS

[1]Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352
[2]National Security Directorate, Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352
[3]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352
[4]Department of Physics, Auburn University, 315 Roosevelt Concourse, Auburn, Alabama 36849
[5]Department of Materials Science and Engineering, University of Delaware, 201 DuPont Hall, Newark, Delaware 19716

Note: This paper is part of the Special Topic Collection on Artificial Intelligence and Machine Learning for Materials Discovery, Synthesis and Characterization.

[a]**Electronic mail:** tiffany.kaspar@pnnl.gov
[b]**Electronic mail:** sarah.akers@pnnl.gov
[c]**Current address:** Department of Materials Science and Engineering, University of California, Irvine, CA 92697

ABSTRACT

Thin film deposition is a fundamental technology for the discovery, optimization, and manufacturing of functional materials. Deposition by molecular beam epitaxy (MBE) typically employs reflection high-energy electron diffraction (RHEED) as a real-time *in situ* probe of the growing film. However, the state-of-the-art for RHEED analysis during deposition requires human observation. Here, we present an approach using machine learning (ML) methods to monitor, analyze, and interpret RHEED images on-the-fly during thin film deposition. In the analysis workflow, RHEED pattern images are collected at one frame per second and featurized using a pretrained deep convolutional neural network. The feature vectors are then statistically analyzed to identify changepoints; these changepoints can be related to changes in the deposition mode from initial film nucleation to a transition regime, smooth film deposition, and in some cases, an additional transition to a rough, islanded deposition regime. The feature vectors are additionally analyzed via graph analysis and community classification. The graph is quantified as a stabilization plot, and we show that inflection points in the stabilization plot correspond to changes in the growth regime. The full RHEED analysis workflow is termed RHAAPsody and includes data transfer and output to a visual dashboard. We demonstrate the functionality of RHAAPsody by analyzing the precaptured RHEED images from epitaxial depositions of anatase TiO_2 on $SrTiO_3$(001) and show that the analysis workflow can be executed in less than 1 s. Our approach shows promise as one component of ML-enabled real-time feedback control of the MBE deposition process.

I. INTRODUCTION

Thin film deposition is a key synthesis route to discover new materials and optimize them for use in advanced technologies. However, the discovery and optimization of new materials are hindered by the large phase space of synthesis variables available for optimization.[1] One avenue to increase the speed and efficiency of materials development is to apply artificial intelligence and machine learning (AI/ML) approaches to take full advantage of information-rich *in situ* synthesis probes[2,3] and *ex situ* characterization techniques.[4–6] Incorporating this information into the AI/ML decision-making process, both as real-time feedback control during

synthesis and as materials design guidance, would lead to a more rapid and complete optimization of material properties that would revolutionize materials discovery.

Reflection high-energy electron diffraction (RHEED) is widely utilized during physical vapor deposition synthesis processes such as molecular beam epitaxy (MBE) and pulsed laser deposition. RHEED patterns reveal real-time information on the crystalline phase, orientation(s), and morphology of the top one or few monolayers (MLs) of growing films.[7,8] During the deposition of films in the layer-by-layer growth mode, intensity oscillations of the specular RHEED spot are typically tracked to provide a real-time measure of the growth rate. Other characteristics of the RHEED pattern, such as diffraction streak spacing, width, intensity relative to background, and degree of modulation, as well as the presence or absence of diffraction intensity corresponding to surface reconstructions or secondary phases, may be quantitatively tracked or qualitatively assessed by the deposition system operator during the film growth. In limited cases,[9,10] such real-time assessment can guide manual on-the-fly control of deposition parameters to obtain a desired outcome. However, such success requires the researcher to develop significant skill and experience in the deposition of the particular material under study and this knowledge is not often directly transferable to other materials. More typically, observing and understanding these RHEED features can inform the choice of deposition parameters for the next growth run but does not lead to actionable deposition parameter changes during the current run that are able to successfully avoid a negative film quality outcome.

AI/ML approaches have been applied to the analysis of RHEED images after the deposition run to reveal more detailed information on the growth process.[11–21] Principal component analysis (PCA) and *k*-means clustering algorithms have been shown to reveal information on the film growth mode, stoichiometry, strain relaxation, and crystallinity when applied to a series of RHEED images analyzed after the deposition.[11,13,18,20] However, PCA methods are not shift-invariant and, therefore, RHEED images must be aligned before analysis. This is a particular challenge for RHEED images because the precise diffraction pattern and its location on the phosphor screen depend sensitively on the diffraction geometry, which varies between different deposition systems and, even in the same deposition system, can be impacted by subtle variations in substrate tilt from sample to sample. Kwoen and Arakawa[14,15] utilized a convolutional neural network (CNN) to avoid difficulties with image variability and was able to demonstrate the accurate classification of GaAs RHEED images into as many as three different classes based on the surface reconstruction of the GaAs film. Schaefer *et al.*[21] utilized a pretrained CNN, AlexNet, to featurize and classify the RHEED images obtained after the deposition of $(In_xGa_{1-x})_2O_3$ by MBE. A large number of depositions (109) could be executed in a high-throughput manner by utilizing the self-etching characteristics of β-Ga_2O_3 under Ga-rich conditions, and this RHEED data set was utilized to train the CNN to classify images as "streaky" or "spotty." Liang *et al.*[16] developed a toolset of several supervised and unsupervised ML approaches to classify RHEED images; a key component of their toolset is a deep learning model called U-Net that is pretrained on millions of images from ImageNet and was further fine-tuned on hand-labeled RHEED images to identify streak and spot regions within each image. This toolset was used to classify the RHEED patterns in the Fe_xO_y thin film system and demonstrated good agreement with crystalline phase analysis by x-ray diffraction (XRD). A recent study by Price *et al.*[17] utilized several supervised and unsupervised ML approaches, including U-Net, to rapidly classify RHEED features (10 s per image) relevant to the deposition by MBE of $(W_{1-x}V_x)Se_2$ on $Al_2O_3(0001)$ substrates. They demonstrated the ability to classify the orientation of WSe_2 films ("epitaxial" versus "textured") based on the final film RHEED pattern with ∼80% accuracy, and, importantly, they were able to achieve a similar level of classification success predicting the resulting film orientation solely from the initial $Al_2O_3(0001)$ substrate RHEED pattern (using fixed deposition parameters). Using a small training set of $(W_{1-x}V_x)Se_2$ film RHEED patterns labeled with the vanadium doping level obtained from *ex situ* x-ray photoelectron spectroscopy (XPS) measurements, promising results on the prediction of vanadium composition from RHEED patterns were demonstrated.

Because the diffraction data are available in real time, RHEED is a natural choice to provide on-the-fly feedback to guide the thin film deposition process. Khaireh-Walieh *et al.*[12] utilized both unsupervised and supervised ML models to classify GaAs substrate RHEED images as oxidized or deoxidized. They did not apply their model in real time but estimated that their lightweight analysis could be accomplished well within the timeframe over which the RHEED patterns typically change, which the authors estimate as on the order of tens of seconds. Shen *et al.*[19] recently demonstrated real-time feedback-controlled growth of InAs/GaAs quantum dots (QDs) based on the AI/ML analysis of RHEED videos collected while the sample was rotating. The video was preprocessed into individual images, which were then deconvoluted using a 3D ResNet-50 CNN. This CNN was trained on RHEED videos from 120 distinct deposition runs (four repeats each of 30 different deposition conditions). Once the model was trained, it was used to provide on-the-fly feedback control during QD deposition. Feedback control took the form of a continuous ramp of the substrate temperature, either increasing or decreasing, to increase the model's probability that the desired QD density (classified as "low," "middle," or "high" density) was achieved. The model autonomously determined whether a temperature ramp was necessary and whether the ramp was increasing or decreasing.

Such feedback control will be a critical component of autonomous synthesis approaches to speed material discovery and design. RHEED analysis approaches that are lightweight to facilitate on-the-fly feedback control and material-agnostic, requiring little or no training data from the material system of interest, are required to implement true autonomous synthesis. Here, we present an approach to on-the-fly RHEED pattern analysis that identifies pattern changes before they are obvious to the eye. This approach is unsupervised, material-agnostic, and can be executed within 1 s, which is of sufficiently low latency for real-time analysis during thin film deposition. Our approach combines image featurization by a pretrained CNN followed by statistical changepoint analysis and graph analysis to quantify the stability of thin film growth. We demonstrate the on-the-fly operation of this analysis routine on prerecorded data collected during the deposition of model epitaxial oxide thin films. The image analysis codes developed in this work are available online as an open-source and

extensible platform at https://github.com/pnnl/RHAAPSODY, and the prerecorded RHEED data used for demonstration in this work are available at https://data.pnnl.gov/group/nodes/dataset/34144.

II. EXPERIMENT

Model thin films of epitaxial anatase TiO_2 were deposited on $SrTiO_3(001)$ substrates by oxygen-plasma-assisted MBE.[22] The $SrTiO_3(001)$ substrates were prepared with TiO_2 termination by etching for 30 s in buffered HF, rinsing with de-ionized water, and then annealing in air at 1000 °C for 1 h. These annealed substrates were placed in a UV/O_3 cleaner for 5 min before being loaded into the vacuum system. The substrate surfaces were further cleaned *in situ* by exposure to activated oxygen from a differentially pumped electron cyclotron resonance microwave plasma source at an oxygen pressure of 3.2×10^{-6} Torr and room temperature for 30 min to ensure removal of all residual carbon from the substrate surface. The flow of activated oxygen continued as the substrate was heated to the deposition temperature of 650 °C. Ti was supplied from a high temperature effusion cell to realize an anatase TiO_2 growth rate of 0.08 Å s^{-1} (119 s/unit cell). Films were deposited to a nominal final thickness of 50 nm.

RHEED diffraction patterns were generated with a differentially pumped Staib RH 30 electron gun system operated at 15 kV and ~1.1–1.3 mA emission current. The diffraction patterns were collected from the phosphor screen using a k700-12 CCD detector [12-bit bit depth, 1/2-in. CCD, 656 × 492 pixels, 88 frames per second (fps), from k-Space Associates, Inc.] and kSA 400 software (v. 5.90, k-Space Associates, Inc.). During the deposition run, diffraction pattern images were captured at 1 fps and saved as raw image files (*.ksah) on the CPU RHEED controller computer.

Film crystallinity was assessed after the deposition using high resolution XRD in the θ–2θ geometry. XRD patterns were collected using a Rigaku SmartLab diffractometer with a rotating Cu anode operated at 45 kV and 200 mA. The incident beam (Cu K_α) was filtered using a two-bounce Ge(220) monochromator, and a matching monochromator was employed for the diffracted beam.

III. IMAGE ANALYSIS

The initial image analysis techniques were developed using RHEED videos collected from the epitaxial deposition of the complex oxide thin films La_2CuO_4,[23] $SrTiO_3$,[18] and $MnFe_2O_4$ (Ref. 24) with oxygen-plasma-assisted or ozone-assisted MBE.[25] The videos were collected as screen-capture recordings and saved in the Hierarchical Data Format HDF5. The results of these analyses are not shown, but these videos provided the initial data on which we developed the image analysis approach outlined below. This approach was fine-tuned using the RHEED images from anatase TiO_2 deposition (collected at 1 fps as *.ksah files).

The image analysis codes developed in this work are available online as an open-source and extensible platform at https://github.com/pnnl/RHAAPSODY.

A. Image preprocessing

RHEED diffraction images from TiO_2 deposition in raw format were transformed into an eight-bit grayscale Tag Image File Format (TIFF) file with a custom C# application utilizing the header information generated by the kSA400 software and the LibTiff library. Only standard image tag information was included in the TIFF header.

Each TIFF image was cropped to 280 × 138 pixels to exclude the nondiffraction regions. The cropped image was rescaled to one quarter size and then standardized using wavelet denoising followed by histogram equalization using the built-in functions and default parameters of the Python package scikit-image.[26] This image standardization equalizes the image contrast of bright and dim features in the RHEED pattern and de-emphasizes the differences in background intensity (see Fig. S1 in the supplementary material). Image features were then extracted from a VGG16 (Ref. 27) CNN pretrained on ImageNet.[28] VGG16 has demonstrated success in image analysis applications where subtle details are of key importance, such as medical imaging.[29] Because grayscale images have only one color channel and CNNs pretrained on ImageNet expect three color channels, the single channel was tiled into three color channels before ingestion into the pretrained VGG16 model. The last convolutional layer of the model is a 512-dimensional feature vector. Finally, the mean feature vector of the first 30 frames is subtracted from each subsequent frame's feature to better distinguish between vectors in the high dimensional vector space.

B. Changepoint analysis

The images are compared pair-wise using a cosine similarity kernel,

$$K(x, y) = \text{CosineSim}(x, y) = \frac{\langle x, y \rangle}{|x||y|}, \quad (1)$$

where x and y are the image features of two frames from the same video. A similarity matrix is built, with each entry equal to $K(x,y)$. As outlined in Fig. 1(a) and illustrated in Fig. 1(b), the changepoint algorithm segments the sequence of images, I, between the previous changepoint (τ_{n-1}) and the current time (t) into subintervals I_0 and I_1, where the images within each subinterval are similar and between subintervals are dissimilar. The time that separates I_0 and I_1 is identified as the changepoint time, τ_n. To support online changepoint detection and reduce the computational cost, a sliding window of width W is implemented such that I ranges from W to the current time (t) when $t - \tau_{n-1} \geq W$. In this work, a sliding window width of $W = 300$ s was found to reasonably balance the computational time with effective changepoint detection; varying the window size from 250 to 350 s did not significantly impact the computational time or the changepoints detected.

To measure the dissimilarity within intervals, we employ a segmentation cost function, adapted from the CosineCost function from the Ruptures Python library,[30–32]

$$\text{SegCost}(I_i) = \sum_{t \in I_i} K(x_t, x_t) - \frac{1}{|I_i|} \sum_{t,t' \in I_i} K(x_t, x_{t'}), \quad (2)$$

where t and t' are times in the interval I_i ($I_i = I$, I_0 or I_1), x_t and $x_{t'}$ are the associated features of the frames at those times, and $K(x_t, x_{t'})$ is the corresponding entry in the similarity matrix for those frames.

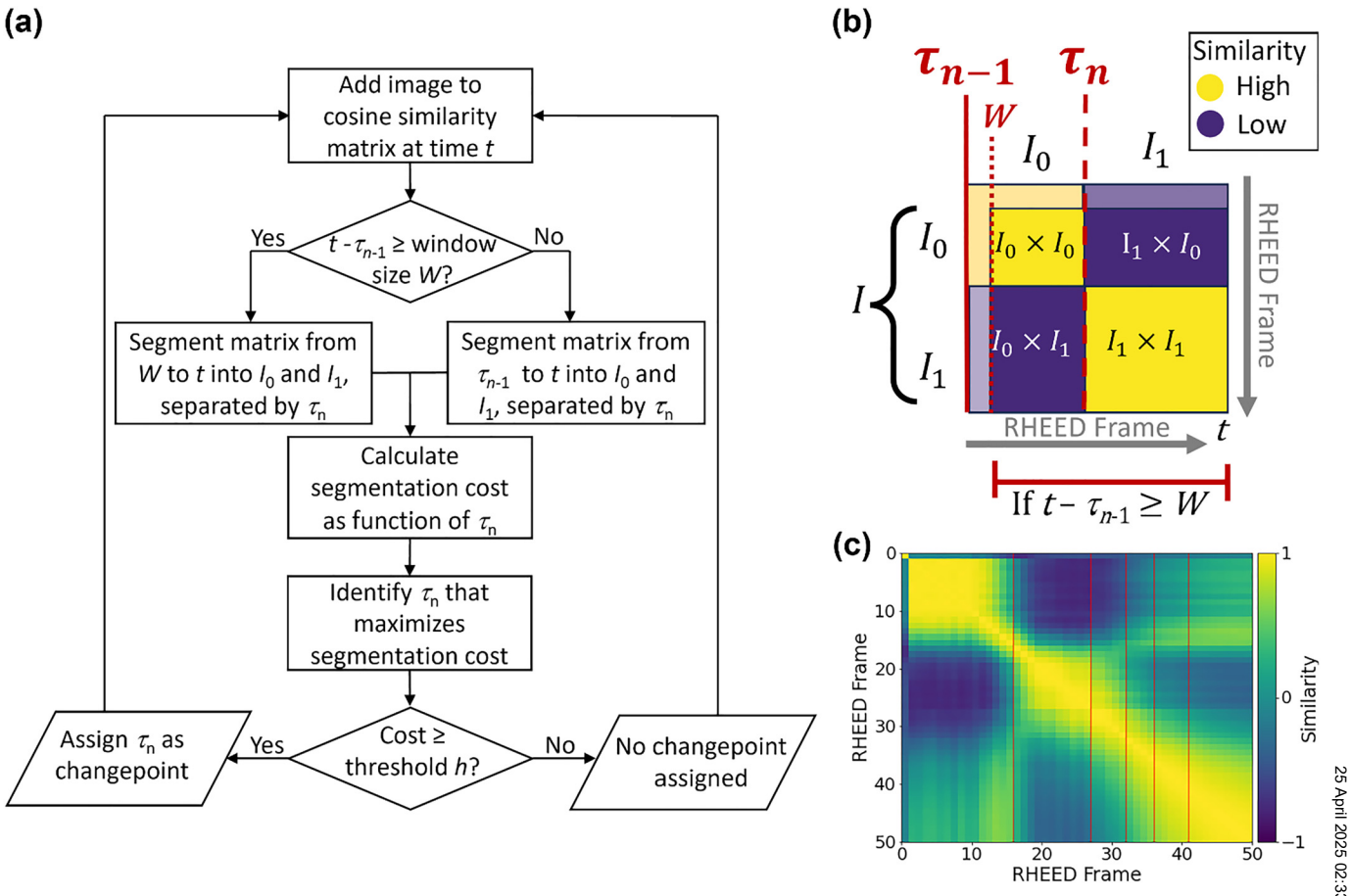

FIG. 1. (a) Schematic of the pairwise similarity matrix to illustrate the changepoint detection process. The image set I in the time interval from the previous changepoint to the current image is segmented into subintervals I_0 and I_1. The segmentation is optimized by maximizing the similarity of the diagonal blocks of the matrix (internal similarity scores $I_0 \times I_0$ and $I_1 \times I_1$), and the optimal value is declared the changepoint τ_n. Off-diagonal blocks of the matrix (cross-similarity scores $I_0 \times I_1$ and $I_1 \times I_0$) have low similarity values. (b) The similarity matrix of deposition B through frame 50 and changepoint analysis with h = 0.05. (c) The assigned changepoints are indicated by vertical red lines (the equivalent horizontal lines are omitted for visual clarity).

$|I_i|$ denotes the length of I_i. Changepoints are detected in the RHEED video sequence by optimizing the choice of τ to maximize the segmentation cost of the full time interval, I, compared against the segmentation cost of the subintervals I_0 and I_1 separated by τ_n,

$$\tau_n = \underset{\tau_n \in I}{\operatorname{argmax}} \left\{ \frac{\text{SegCost}(I) - \text{SegCost}(I_0) - \text{SegCost}(I_1)}{|I|} \right\} \quad (3)$$

if

$$\frac{\text{SegCost}(I) - \text{SegCost}(I_0) - \text{SegCost}(I_1)}{|I|} \geq h, \quad (4)$$

where h is a threshold value chosen to ensure that only significant changepoints are detected. In this work, if Eq. (4) exceeds a threshold of $h = 0.025$, τ_n is declared a changepoint.

C. Graph analysis

The similarity matrix used during statistical changepoint analysis can be represented via a graph data structure.[33] In this case, the similarity matrix is treated as an adjacency matrix for graph $\mathbf{G}_t$ at time t. Each node in $\mathbf{G}_t$ represents an image at time t. Each edge connects two nodes and is weighted by the corresponding value in the similarity matrix. After the initial warm-up phase, edges with similarities below the mean similarity are filtered out to reduce the connectivity of $\mathbf{G}_t$. The Louvain method for community detection[34] is then applied. To cluster similar nodes into groups within $\mathbf{G}_t$; this is equivalent to clustering groups of similar images during the current thin film growth.

Singular value decomposition (SVD) is then performed on $\mathbf{G}_t$ to quantify the graph structure and connectivity.[35] Singular values measure the importance of orthogonal bases in capturing information or approximating data. The first two singular values, SV0 and

SV1, correspond to the orthogonal basis vectors that capture the greatest information (variance) in the data. The first singular value (SV0) indicates how strongly the network is dominated by a single cluster of nodes. The second singular value (SV1) reflects the strength of the second-most dominant cluster. Their ratio, SV0/SV1, quantifies this structural diversity such that high values indicate a single dominant cluster, while low values indicate two comparatively sized clusters. Measured over time, the nodes that fall into the dominant cluster as $\mathbf{G}_t$ grows (i.e., images that are similar to those that came before) will cause SV0/SV1 to increase, whereas the nodes that do not (i.e., images are sufficiently different from those that came before) will cause SV0/SV1 to decrease. In this context, we call SV0/SV1 the stabilization ratio, as it effectively describes the stability of the RHEED pattern during thin film growth. If the stabilization ratio produces an inflection point when measured against time, then a change in the film will have occurred.

SVD analysis can also be used to generate embedding representations of each node in the graph by truncating the decomposition matrices to incorporate only the basis vectors that capture the most variance in the data. This method can condense high dimensional information about the structure and connectivity of a given node into a low-dimensional vector space. By generating these node embeddings using the top two components for the image-based graph $\mathbf{G}_t$, we can visualize the latent relationships between images over time as the deposition proceeds.

D. Hardware components

The RHEED images were collected on a CPU tower computer supplied by the vendor (k-Space Associates, Inc.) with an Intel Core i5-10400 processor and 16 GB RAM. The control program, Pilot, was run on a CPU laptop with an Intel Core i7-8550H processor and 16 GB RAM. Image featurization and analysis were executed on a Lambda Vector GPU workstation (processor: AMD Ryzen Threadripper PRO 5955WX: 16 cores, 4.0–4.5 GHz, 64 MB cache, PCIe 4.0; GPU: 2x NVIDIA RTX A6000: 48 GB memory, 10752 CUDA cores, 336 Tensor cores, NVLink).

IV. RESULTS

A. Deposition of model anatase TiO_2 films

For the development and testing of the RHEED analysis workflow, epitaxial anatase TiO_2 deposited on $SrTiO_3$(001) was chosen as the model thin film system. Anatase TiO_2 has a simple stoichiometry, thereby reducing the number of deposition parameters to consider, but diverse outcomes can be expected under different deposition conditions: smooth versus rough anatase[36] and phase-pure epitaxial anatase versus inclusions of epitaxial rutile,[37] polycrystalline rutile, or amorphous TiO_2 are also possible. In the current study, we focus on the transition from an epitaxial film with a smooth surface to an epitaxial film with a rough, islanded surface. This transition is typically observed in RHEED diffraction patterns as a change from diffraction streaks to diffraction spots, with the spots lying along the streak positions.[8]

Physical vapor deposition processes, as exemplified by MBE, operate on multiple time scales. Adatoms on the growing film surface interact with the surface via adsorption and positional hopping on a timescale on the order of their atomic vibrations (ν, typically[38] 10^{13} s^{-1}). It is these rapid interactions that lead to the formation of the thin film. Increasing the substrate temperature allows the adatoms to overcome larger energy barriers more often, and, therefore, affects film crystallinity and morphology. On the other hand, the rate of impinging adatoms on the surface is many orders of magnitude slower. For the TiO_2 films studied here, the rate of Ti adatom deposition is only 0.033 ML s^{-1}, which equates to the arrival of one Ti adatom on a given square nanometer of the surface approximately every 4.3 s. Due to this low adatom flux, the film develops slowly over time, typically on the order of seconds or minutes. Given this, it was deemed unnecessary to collect and analyze RHEED images at the frame rate of the camera (88 fps). A rate of 1 fps was chosen and found to capture the changes in surface morphology with sufficient temporal resolution for on-the-fly analysis.

RHEED diffraction image snapshots from two anatase TiO_2 deposition runs are presented in Figs. 2 and 3, respectively. The full RHEED image series used for the analysis of deposition A and deposition B are available as uncropped TIFF files at https://data.pnnl.gov/group/nodes/dataset/34144. Both runs were executed nominally under the same deposition conditions, as specified in Sec. II. For deposition A shown in Fig. 2, in the initial stage of film growth, the RHEED pattern transitioned from that of the $SrTiO_3$(001) substrate [Fig. 2(a)] to a streaky pattern corresponding to smooth anatase TiO_2 deposition [Fig. 2(b)]. As the film deposition progressed, some slight modulation of intensity appeared within the diffraction streaks [Fig. 2(c)]. These are diffraction spots arising from slight surface roughening. Over time, these spots diminished in intensity [Fig. 2(d)], and they were barely detectable by frame 6000 of the deposition [Fig. 2(e)]. After the deposition finished (frame 6327) and the film had cooled to room temperature, the final RHEED pattern in this azimuth indicated that the film was smooth and well-crystallized, as shown in Fig. 2(f), but some modulation indicative of surface roughening was observed in the RHEED diffraction pattern collected along the TiO_2[110] direction (Fig. S2 in the supplementary material).

The intensity of the specular RHEED beam was extracted from the series of RHEED images after deposition and is plotted in Fig. 2(g). Surprisingly, no intensity oscillations are observed during this deposition, despite the persistence of the streaky RHEED pattern in the TiO_2[110] azimuth throughout the growth. The primary feature of the RHEED intensity plot is a significant decrease in intensity beginning at frame 12, when the Ti shutter was opened and deposition commenced, and continuing until approximately frame 150. This decrease in RHEED intensity has been well documented for epitaxial anatase TiO_2 deposition,[36,39] where it was observed to occur for the first ~50–150 s after the start of deposition. At the deposition rate employed by Chambers *et al.*,[39] this timeframe was equivalent to the deposition of approximately 1–1.5 unit cells of anatase. Similarly, for deposition A, frame 150 corresponds to the deposition of 1.2 unit cells of anatase.

Deposition B (Fig. 3) started off similarly to deposition A, with a transition from the $SrTiO_3$(001) substrate [Fig. 3(a)] to a streaky pattern from a smooth film of anatase [Fig. 3(b)]. As the deposition progressed, however, the RHEED pattern transitioned

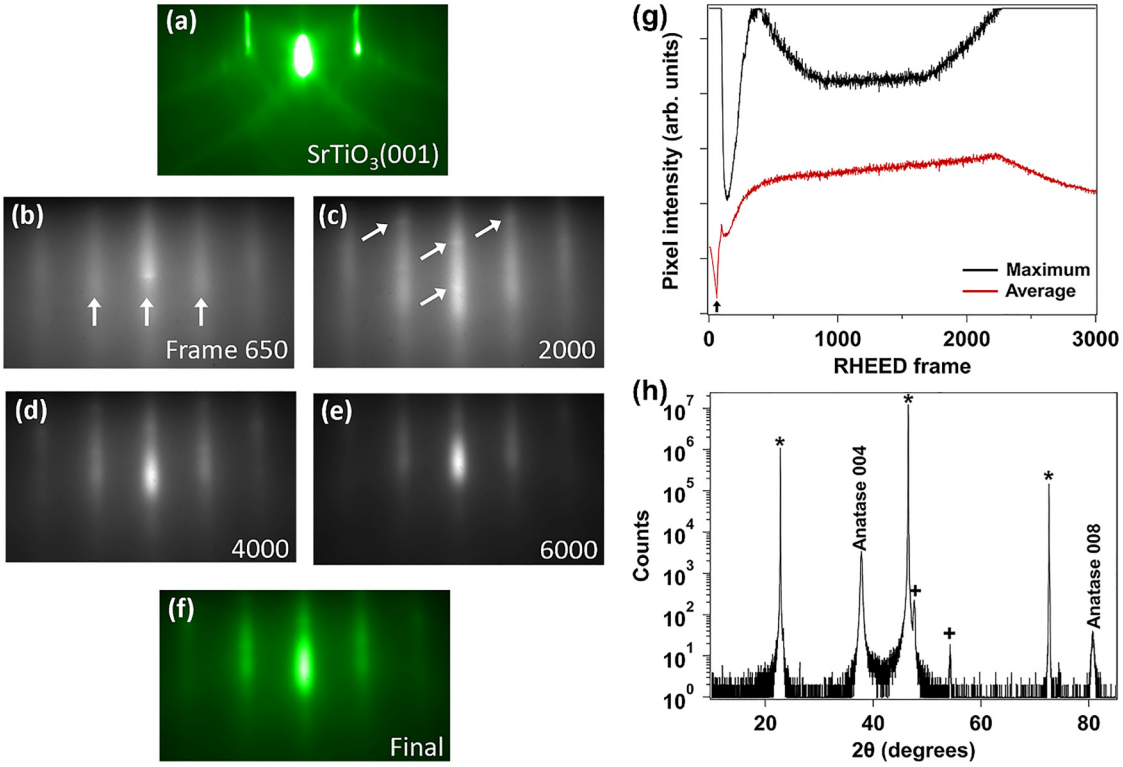

FIG. 2. Deposition A of an epitaxial anatase TiO_2 film on $SrTiO_3$(001). (a)–(f) RHEED diffraction pattern images collected along $SrTiO_3$/TiO_2[100]. (a) and (f) were collected before and after deposition, respectively, at room temperature. Green color scale added for clarity (online article only). (b)–(e) were collected during deposition with the substrate at 650 °C; the image frame is indicated on each image. White arrows in (b) indicate diffraction streaks that arise from the smooth anatase film; white arrows in (c) indicate the spots appearing along the streaks due to film roughness. (g) Intensity oscillations of the specular RHEED beam. (h) θ–2θ XRD pattern of the final film. Reflections from $SrTiO_3$⟨001⟩ planes indicated by (*). (+) denotes reflections of unknown origin that appear in all $SrTiO_3$ patterns.

from a streaky pattern to one exhibiting well-defined spots along the streaks [Figs. 3(c)–3(e)] over the course of <150 s. Retrospective analysis of the image series places the point at which the spots become visually evident at ~frame 718 [i.e., between Figs. 3(c) and 3(d)]. After ending the deposition (frame 6280) and cooling the film to room temperature, the spots are strongly pronounced and the underlying streaks are faint, indicating that the anatase film is significantly islanded [Figs. 3(f) and S3 in the supplementary material].

The intensity of the specular RHEED beam exhibited oscillations during the initial stages of deposition, as shown in Fig. 3(g). Interestingly, these oscillations begin from the moment that growth commences (frame 10), although the overall intensity of the specular RHEED beam decreases over the first ~125–150 frames in a similar manner to the transition regime observed for deposition A [Fig. 2(g)] and during previous anatase depositions.[36,39] In addition, the oscillation frequency is found to be 32 s, which corresponds to 0.25 unit cells (1 ML) of anatase. Previous studies of anatase deposition have found that the RHEED intensity oscillation frequency corresponds to 0.5 unit cells (2 ML) of anatase.[36,39] This discrepancy may be due to a difference in the growth mode for the anatase film in deposition B compared to the previous depositions on $LaAlO_3$(001) substrates or may be attributable to a difference in the incident angle of the RHEED electron beam on the film surface in the current study relative to previous reports; such variations in the incident angle are known to affect RHEED diffraction intensities in nonintuitive ways.[10,40]

Postdeposition analysis of both films by high resolution XRD [deposition A in Fig. 2(g) and deposition B in Fig. 3(g)] confirms that both have crystallized in the anatase TiO_2 polymorph, with no evidence of rutile formation. The out-of-plane crystalline orientation is (001), as expected.

B. On-the-fly RHEED analysis

Our approach for real-time RHEED pattern analysis consists of three components: preprocessing, changepoint detection, and graph analysis. The initial assessment of image preprocessing approaches was performed using pre-existing RHEED video data sets from the epitaxial deposition of complex oxides by oxide MBE.[18,23,24] From these initial studies, it was determined that the features produced with a CNN and adaptive histogram equalization

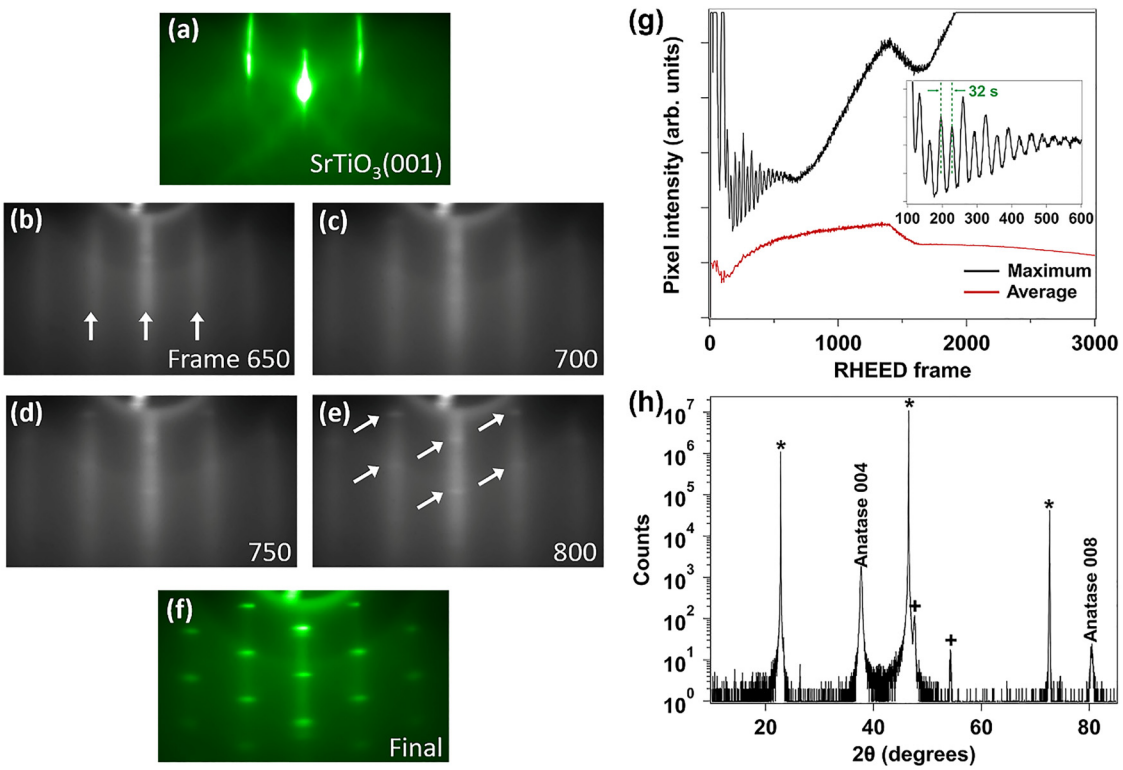

FIG. 3. Deposition B of the epitaxial anatase TiO_2 film on $SrTiO_3$(001). (a)–(f) RHEED diffraction pattern images collected along $SrTiO_3$/TiO_2[100]. (a) and (f) were collected before and after deposition, respectively, at room temperature. Green color scale added for clarity (online article only). (b)–(e) were collected during deposition with the substrate at 650 °C; the image frame is indicated on each image. White arrows in (b) indicate diffraction streaks that arise from the smooth anatase film; white arrows in (c) indicate the spots appearing along the streaks due to film roughness. (g) Intensity oscillations of the specular RHEED beam. The inset shows intensity oscillations with 32 s spacing. (h) θ–2θ XRD pattern of the final film. Reflections from $SrTiO_3$ ⟨001⟩ planes indicated by (*). (+) denotes reflections of unknown origin that appear in all $SrTiO_3$ patterns.

performed better for changepoint detection than the features constructed with PCA. We also tested several changepoint detection methods from the Ruptures[32] Python package and chose the CosineCost function, which we modified to allow for online changepoint detection, for subsequent changepoint analyses.

After preprocessing and unsupervised featurization into a multidimensional feature vector using the deep CNN model VGG16, each image is compared to all previous images using Eq. (1) and added to a pairwise similarity matrix for changepoint analysis. A similarity matrix based on the first 50 RHEED images from deposition B is presented in Fig. 1(c). As detailed in Sec. III A, the changepoint time τ_n of interval I is determined by maximizing the difference in segmentation cost functions between the full interval I and the subintervals I_0 and I_1 [Eq. (3)]. When this difference is maximized, the diagonal blocks of the similarity matrix (internal similarity scores) have high overall values, while the off-diagonal blocks (cross-similarity scores) have low overall values. Therefore, the images in subinterval I_0 are more similar to each other and less similar to those in subinterval I_1. In Fig. 1(c), the similarity between images [Eq. (1)] is represented by a color scale ranging from yellow to violet, where yellow denotes fully similar images ($K = 1$) and violet denotes fully dissimilar images ($K = -1$).

In any given interval I, there will always be a value of τ that maximizes Eq. (3). However, the difference in similarity between the images in subintervals I_0 and I_1 may be very small or even within the noise of the images, and, therefore, the changepoint at τ would not represent a significant change in the RHEED pattern itself. To avoid the detection of nonsignificant changepoints, a threshold value, h, is utilized to select only segmentation cost function differences in Eq. (3) that are deemed "significant." The optimum value of h will result in a set of changepoints that match a "ground truth" set of changepoints within a RHEED image. However, such a set of ground truth changepoints is not available because there is no independent, accurate measure of changepoints; human observation of the series of RHEED images can identify some changepoints but may miss changes that are not obvious to the eye but are indicative of actual surface morphology changes. Therefore, in this work, the value of h was chosen in an *ad hoc* manner to produce a set of changepoints that reasonably matched those expected based on expert knowledge of the deposition

process. Figure S4 in the supplementary material presents the changepoint analysis of deposition B with different values of h. As expected, the number of changepoints decreases, and the time to detect the changepoint increases, as the threshold value increases. A value of $h = 0.025$ was chosen as a balance between changepoint sensitivity and the identification of spurious changepoints; in particular, the chosen threshold value of 0.025 avoids the spurious changepoint at frame ~900, while detecting the changepoint at ~718 as early as possible.

The similarity matrix after on-the-fly analysis of deposition B is shown through frame 1400 in Fig. 4(a). Here, 17 changepoints are identified. Most of the changepoints occurred in the initial stages of film deposition (≤frame 394), but two are identified later in the deposition at frames 739 and 825. In contrast, the similarity matrix of deposition A, shown through frame 2100 in Fig. 4(b), exhibits less variation in the degree of similarity of the images. During the deposition, ten changepoints are identified. All of these changepoints occurred in the initial stages of film deposition; the last changepoint was identified at frame 211.

As shown schematically in Fig. 1(a), the changepoints are assigned retrospectively. For deposition B, Fig. 4(c) plots each identified changepoint versus the time (which is equivalent to the number of frames past the changepoint) required to make the assignment, and the equivalent plot for deposition A is presented

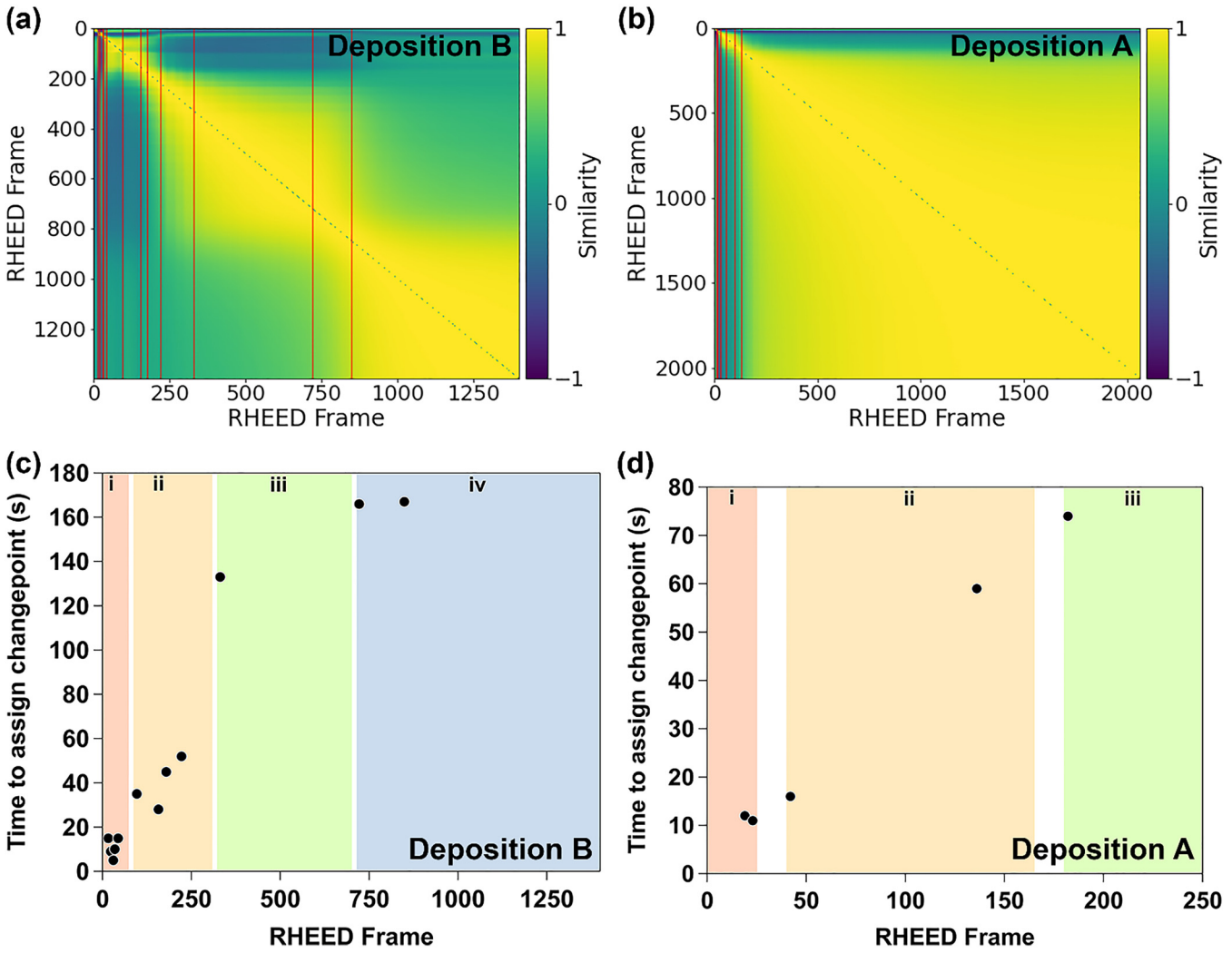

FIG. 4. Similarity matrices for (a) deposition B through frame 1400 and (b) deposition A through frame 2100. The assigned changepoints are indicated by vertical lines. The time required to detect and assign each changepoint during on-the-fly analysis of (c) deposition B and (d) deposition A. Four regimes of the thin film deposition process are indicated by shading: initial film nucleation (i), transition regime (ii), deposition of a smooth film (iii), and deposition of a rough film (iv).

in Fig. 4(d). In Fig. 4(c), the changepoints for deposition B can be grouped into four regimes based on both their frequency of occurrence and the time required to make the assignment. Comparison of these regimes with the corresponding RHEED images facilitates the assignment of these regimes to different stages of the thin film deposition process: (i) film nucleation, (ii) transition regime, (iii) growth of a smooth film, and (iv) growth of a rough film. At the beginning of the deposition, the RHEED pattern changes dramatically as the anatase TiO_2 film nucleates on the $SrTiO_3$ surface. The changes occurring during the initial film nucleation phase (denoted as region i) are detected frequently (spaced approximately every 5–10 s) and assigned within 5–15 s of their occurrence. The transition between this nucleation regime and the next stage occurs between the changepoints at frames 94 and 138, which suggests that the transition from nucleation occurs after approximately half a unit cell of TiO_2 has been deposited (which requires 60 s of deposition time). Starting with the changepoint identified at frame 138, the delay before changepoint assignment increases to 15–25 s as the growth progresses into the transition regime (ii). In Fig. 4(d), the transition regime is assigned as falling between the changepoints at frames 138 and 293, which approximately corresponds to the decrease and recovery of RHEED intensity in Fig. 3(g). After the changepoint at frame 293, changes to the RHEED pattern occur less frequently and more slowly with time, and this regime correlates well with the time over which the RHEED pattern exhibited streaks [Fig. 3(b)]. Therefore, this stage is associated with the deposition of a smooth film (iii). As illustrated in Figs. 3(b)–3(e), the smooth film transitions slowly to a rough surface morphology. The intensity oscillations of the RHEED specular beam cease around frames 650–680, and spots become apparent to the eye around frame 718. This change in the deposition mode to the growth of a rough film (iv) is identified as a changepoint at frame 739, with a second changepoint at frame 825 as the rough surface morphology is established. The changepoints in this regime require ~80–100 s for assignment; the changepoint at frame 739 that represents the initial appearance of spots within the diffraction streaks was assigned 98 s later.

For the changepoints assigned during deposition A in Fig. 4(c), the regimes corresponding to initial nucleation (i), transition (ii), and growth of a smooth film (iii) can be identified; there is no transition to the growth of a rough film (iv) in this case because the RHEED pattern remained streaky. The delineation between growth regimes is somewhat harder to determine for this deposition because, overall, fewer changepoints were identified, but reasonable delineations can be made by comparing the changepoints in Fig. 4(c) with the stabilization plot shown below in Fig. 5(d).

The graph structure generated from the similarity matrix was also investigated to determine if accurate changepoints can be extracted on-the-fly. The singular value embeddings for nodes in the graph captured at frame 550 of deposition B, $\mathbf{G}_{t=550}$, are shown in Fig. 5(a). Note that the values of the SVD node embeddings are relative to all nodes in the graph, i.e., all frames taken up to time $t = 550$. This causes RHEED frames that are more similar to each other to be plotted close together, and frames that are less similar to be plotted further apart. The most recently added frame is indicated by the red symbol on the plot. The symbol colors correspond to the community classification of each image from Louvain clustering. In Fig. 5(a), the film growth has progressed until a smooth, streaky RHEED pattern has been established. Three regimes can be identified in the plot: a clustered set of frames representing the initial nucleation of the film on the substrate, a more dispersed cluster that represents the transition regime, and a tight cluster of frames that represent the growth of a smooth film with a streaky RHEED pattern. As the film growth progresses, the transition from a smooth film with a streaky RHEED pattern to a rough film with spots in the RHEED pattern [Fig. 2(b)] alters the graph structure. The node embeddings after the spots have been established ($\mathbf{G}_{t=850}$, through frame 850) are plotted in Fig. 5(b), which reveals that the incoming images are not similar to the cluster of streaky images collected through frame 550. Note that the values of the individual node embeddings change as t increases from 550 to 850, as the embeddings are derived from the similarity relations between all frames in $\mathbf{G}_t$.

To extract quantifiable, actionable information from the graph data, we employ stabilization plots, as shown in Figs. 5(c) and 5(d). As described in Sec. III C, stabilization corresponds to the ratio of the two largest singular values from $\mathbf{G}_t$, SV0/SV1. An increase in the stabilization ratio occurs when the new frames added to the graph are similar to the previous frames, whereas the ratio decreases when the new frames differ from the previous frames. Therefore, a change in slope of the stabilization plot indicates a slight change in the similarity of the newly arriving images. An inflection point reflects a significant change in the similarity of newly arriving images, which indicates that the film has undergone a transition.

The behavior of the stabilization plot through frame 1400 of deposition B is shown in Fig. 5(c). Slope changes and inflection points (in the form of peaks and valleys) are clearly distinguishable. By comparing the shape of the stabilization plot with the growth regime assignments determined from the changepoint plot in Fig. 4(d), a strong correlation between inflection points and growth regime changes can be observed. Transitions between different film growth regimes are identified by valleys. Leading to these valleys, the incoming images exhibit differences from the previous images, resulting in decreased stabilization. The valley occurs when the transition is complete and the incoming images begin to exhibit similarity to the previous images. In the stabilization plot in Fig. 5(c), the valley between the transition regime and the growth of a smooth film occurs at frame 309, which matches well with the changepoint identified at frame 293 in Fig. 4(d). In contrast, neither the valley in the stabilization plot at frame 52 nor the peak at frame 152 corresponds to an obvious change in the behavior of the changepoints, although several changepoints occur in that region. Instead, from changepoint analysis, the transition between the nucleation and transition regimes is placed between frames 94 and 138. Lacking an independent measure of the growth mode, we cannot determine whether the changepoint analysis or the stabilization plot more accurately reflects the change in the growth mode.

The graph structure and the stabilization plot begin to change as the film begins to roughen and the RHEED pattern develops spots. Analysis of the RHEED pattern images retrospectively places the appearance of spots in the RHEED pattern at approximately frame 718. However, the graph structure begins to change before that frame (see the video in the supplementary material). This change is captured in the stabilization plot as a change in slope at

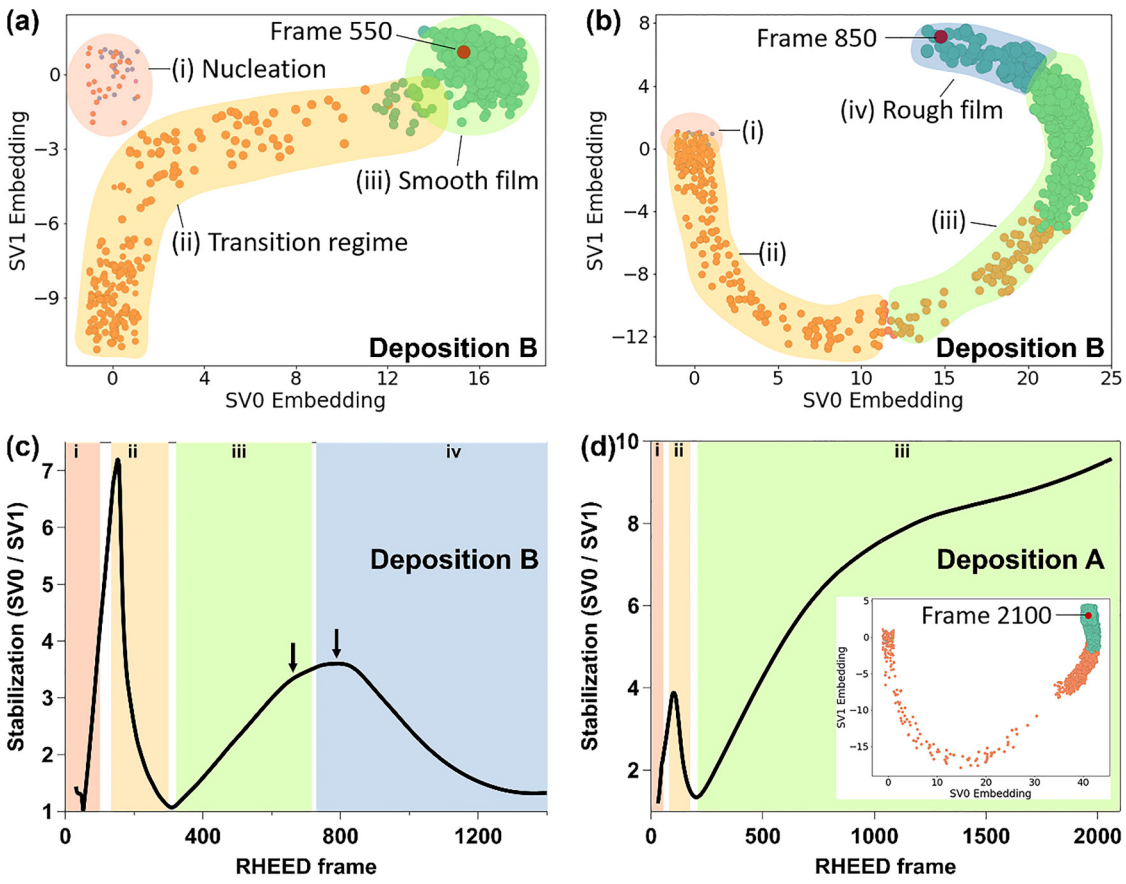

FIG. 5. SVD node embeddings of the graph of deposition B at (a) frame 550 ($\mathbf{G}_{t=550}$) and (b) frame 850 ($\mathbf{G}_{t=850}$). The red (dark gray) symbol indicates the most recent frame; other colors (shades of gray) indicate the community classification of each image. Shading has been added to approximately delineate the growth regimes based on statistical changepoint analysis. (c) Stabilization plot of deposition B at frame 1400. Shading has been added to delineate the growth regimes determined by changepoint statistics analysis [Fig. 4(d)]. Arrows indicate changes in the slope associated with film roughening. (d) Stabilization plot of deposition A at frame 2100. Shading has been added to delineate the growth regimes determined by changepoint statistics analysis [Fig. 4(c)]. Inset: graph data structure of deposition A at frame 2100. Red (dark gray) symbol indicates the most recent frame; other symbol colors (shades of gray) indicate the community classification of each image.

frame 662. It is notable that this slope change, indicative of a change in the RHEED pattern, occurs 56 frames (i.e., nearly 1 min) before the RHEED spots become apparent to the eye. The peak of the stabilization plot occurs at frame 789. These two slope changes in the stabilization plot bracket the appearance of spots in the RHEED pattern.

The stabilization plot for deposition A is shown in Fig. 5(d). The plot has a similar shape to that for deposition B [Fig. 5(c)], although the initial nucleation regime is not captured. An initial peak is observed at frame 101 as the film growth mode is within the transition regime, and the growth of a smooth film with a streaky RHEED pattern is established as the stabilization begins to increase at approximately frame 200. The location of the inflection points in the stabilization plot correlate well with the growth regimes identified by the changepoint analysis of deposition A [Fig. 4(c)]. In contrast to deposition B, deposition A does not transition to a rough film with well-defined spots in the RHEED pattern. Therefore, the stabilization plot does not exhibit strong inflection points past frame 200, although subtle slope changes are observed that may correlate with the weak spots that appear in the RHEED pattern, then disappear [see Figs. 2(c)–2(e)]].

C. Demonstration of analysis workflow: RHAAPsody

A prototype of the analysis workflow described above was implemented as an automated data transfer, analysis, and output dashboard program named the RHEED Heuristic Adaptive Automation Platform (RHAAPsody). A schematic of the RHAAPsody components and the data transfer between them is provided in Fig. 6. The RHEED images are collected in raw format from the instrument at 1 fps and saved to a shared folder on the CPU-based computer supplied by the RHEED system vendor. The images are then accessed by a custom, centralized control application named Pilot, which is housed on a separate CPU computer.

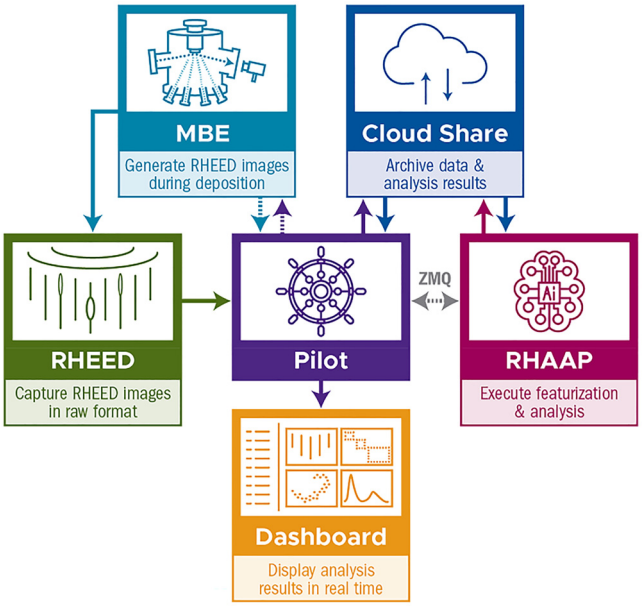

FIG. 6. Schematic representation of the RHAAPsody on-the-fly RHEED analysis data flow. Solid arrows indicate data transfers. Dashed gray arrows indicate ZMQ message communication. Dotted arrows between Pilot and MBE indicate connectivity that exists but was not used in the current analysis workflow.

Pilot converts the raw RHEED images to TIFF format and distributes them to a cloud-based file share location. A ZeroMQ (ZMQ) publish-subscribe model[41] is utilized by Pilot to trigger the featurization and analysis program, named RHAAP, which is housed on a GPU computer (RHAAP is available online at https://github.com/pnnl/RHAAPSODY), to download the TIFF image for analysis. Each image is preprocessed, then fed into the pretrained CNN (VGG16), and converted into a 512-dimensional feature vector. The feature vector is compared against previous images with statistical analysis of the changepoints. Subsequently, graph analysis determines any change in clustering of the new image. Analytics from the changepoint and graph clustering methods are uploaded as images to the cloud-based file share, and a ZMQ message is sent from RHAAP to Pilot that triggers Pilot to download them. Pilot renders the analytics images and the current RHEED image in a dashboard-style display interface, as shown in Fig. 7. On this dashboard, the RHEED image filename, index (frame number), rate of image capture (in fps), and timestamp from the beginning of data acquisition are displayed in real time. In addition, the most recent changepoint assignment and the time at which it was assigned are displayed. The underlying data management structure utilizes "loops" for organization, in which each loop (as a separate folder) contains the analysis output images and changepoint assignments for a period during deposition in which the deposition parameters are not changed by feedback control. Because feedback control is not yet implemented, all data are stored in Loop 0. ZMQ communication between Pilot and RHAAP is confirmed by displaying a count of "Received Messages" and "Sent Messages."

A video of the RHAAPsody dashboard during the analysis of deposition B is available in the supplementary material. In this video, the RHEED images had been captured previously and saved to a shared folder on the RHEED computer in the raw format. For the purposes of demonstrating on-the-fly capability, the raw .ksah images were pulled by Pilot at 1 fps. For each image, all steps of the RHAAPsody analysis workflow were accomplished in <1 s [Fig. 7(b)].

V. DISCUSSION

RHEED is a key *in situ* diagnostic tool to assess and guide thin film deposition. It provides rich information on film crystallinity, orientation, morphology, and phase purity, and the current standard practice during epitaxial thin film deposition by MBE is to manually monitor the RHEED pattern during deposition. This manual approach has not changed significantly over the decades that MBE and RHEED have been in use.[9] Here, we have shown that automated analysis of RHEED patterns on the fly can perform similarly to or better than human observation in identifying features that signal a change in the film.

A comparison of change detection among the different RHEED analysis approaches in this work can be instructive. Such a comparison in the early stages of deposition is hindered by the numerous changes occurring in RHEED pattern shape and intensity over a short time period. Therefore, we assess the accuracy of change detection in the later stages of deposition for deposition B, focusing on the transition from a smooth anatase film with a streaky RHEED pattern to a rough, islanded film with a spotty RHEED pattern. In Fig. 8, a comparison of three on-the-fly analysis results is plotted: the intensity oscillations of the specular RHEED beam, the changepoints assigned by statistical analysis, and the stabilization plot calculated from the graph structure. All three analyses exhibit a demonstrable change as the film growth mode transitions from smooth to rough, with the visual appearance of spots in the RHEED pattern occurring around frame 718. The RHEED intensity oscillations have ceased by frame 680, and the stabilization plot exhibits a change in slope at frame 682. A statistical changepoint is identified at frame 739, which matches reasonably well with the visual assessment of the appearance of spots at frame 718.

The results presented in Fig. 8 show the promise of on-the-fly analysis of RHEED image information during deposition, but the approaches also have drawbacks. RHEED intensity oscillations have been utilized for several decades as a real-time monitor of film quality and as the basis for manual feedback control to guide the deposition to a desired film outcome.[10] Indeed, for deposition B, diminution and termination of the RHEED specular beam intensity oscillations occurred as the film transitioned from a smooth surface morphology to a rough morphology. These oscillations could have been used independently, without employing ML image analysis approaches, to identify the change in the growth regime. However, this approach relies on the presence of RHEED intensity oscillations for analysis, and oscillations do not occur for every material or under every deposition condition. This limitation is illustrated by the RHEED intensity data for deposition A, which does not

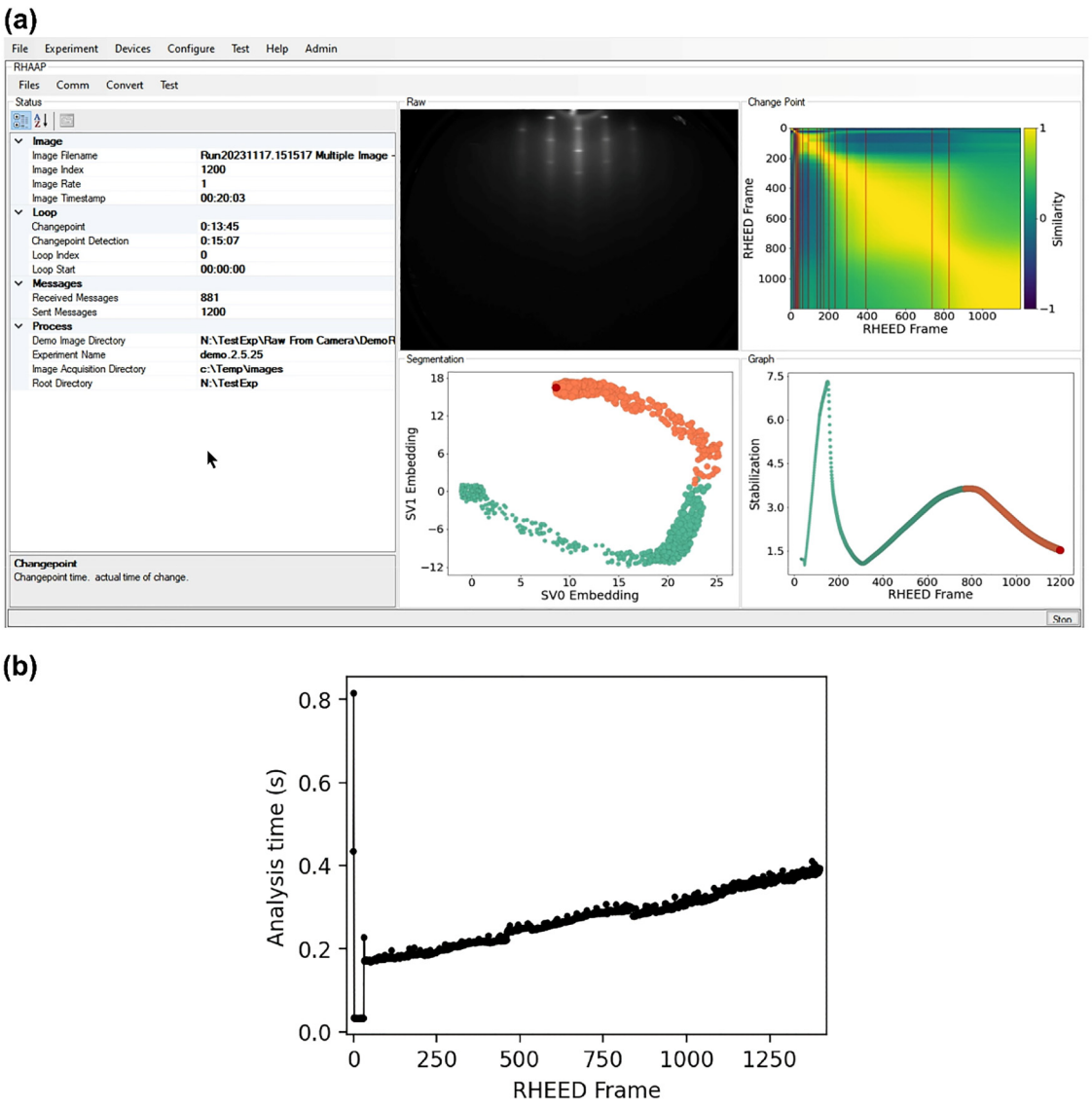

FIG. 7. (a) RHAAPsody output dashboard during on-the-fly analysis of deposition B. The RHEED image (raw), similarity matrix with the changepoints marked (change point), graph structure (segmentation), and stabilization plot (graph) are displayed as images and updated once a second. The left-hand panel displays real-time data pertaining to the analysis, including information on the current RHEED image (image), the assigned time of the most recent changepoint and the time at which it was assigned (loop), ZMQ communication confirmation (messages), and file information (process). (b) Analysis time for the RHAAPsody workflow (excluding RHEED image capture by the RHEED computer) after the addition of each RHEED frame for deposition B.

exhibit oscillations despite the persistence of a streaky RHEED pattern throughout the deposition.

The statistical changepoint analysis based on the similarity matrix of the incoming RHEED images did an excellent job of identifying the point at which the RHEED images began to exhibit spots. The drawbacks to this method, however, are that it is sensitive to the threshold parameter, h, set by the user, and the time (or, equivalently, number of frames) past the changepoint that are required before the changepoint can be detected and assigned. The changepoint at frame 739 was not assigned until 98 s later (frame 837). This delay may render the changepoint assessment unactionable if the film has established the negative outcome (in this case, surface roughness) before the changepoint is detected. Both drawbacks might be partially alleviated by further optimization of the threshold parameter, the frame collection rate, and the sliding window width. A third drawback, which

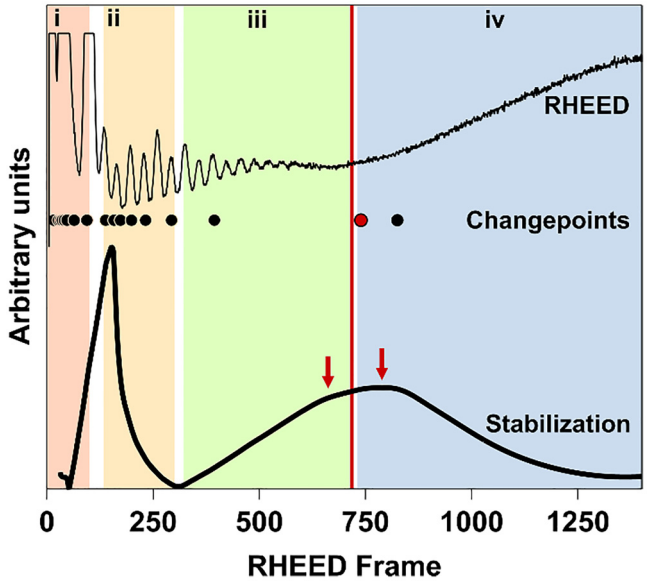

FIG. 8. Comparison of on-the-fly RHEED image analysis approaches for deposition B: RHEED specular beam maximum intensity (top), statistical changepoint analysis (middle), and stabilization plot from graph analysis (bottom). Vertical line indicates the frame at which spots become apparent in the RHEED pattern, as determined retrospectively by the eye. The changepoint associated with the transition to spots in the RHEED pattern is highlighted in red (gray), and the changes in the slope of the stabilization plot associated with the transition are indicated by arrows.

is inherent to this analysis, is that changepoints cannot be categorized in terms of their physical meaning. Therefore, additional information is required before the analysis can be acted on during the deposition.

The graph structure and the associated stabilization plot show great promise for on-the-fly analysis of RHEED images. The stabilization plot provides a real-time, interpretable analysis of the graph structure and, therefore, the evolution of the RHEED images. In Fig. 8, we show that the first change in slope of the stabilization plot, which can be interpreted as being indicative of the early degradation of the RHEED images, occurs at frame 662, 56 s before visible spots appear at frame 718. Such early detection is critical for actionable decision-making during the deposition process. However, we note that the appearance of the slope change was found to be highly dependent on the warm-up window used to generate the graph structure and, therefore, may not give an advance indication of change during every deposition run. Further optimization of the parameters of the stabilization calculation (such as warm-up time) may be beneficial but are beyond the scope of the current work.

On-the-fly analysis must be executed within the time between arriving data. In the case of the RHEED analysis presented here, the RHEED frames arrive once per second and, therefore, the analysis must occur within 1 s. Figure 7(b) confirms that the RHAAPsody workflow, including data transfer, analysis, and output to the dashboard, occurs well under one second per frame. However, the time to execute the workflow increases with increasing deposition time, and for sufficiently long depositions, the workflow may require more than 1 s to execute. Such an increase could be alleviated by implementing a rolling analysis window in place of the current approach of analyzing all frames of the deposition. Careful optimization of such a window is expected to provide accurate analysis results while maintaining a rapid execution time.

The automated, on-the-fly RHEED analysis presented here is one promising step toward realizing real-time autonomous deposition control during MBE. Such control would dramatically reduce the number of growth interations required to optimize and reproducibly deposit high quality epitaxial films, accelerating materials discovery for advanced technologies. However, the necessary control would be complex, involving multiple deposition parameters and accounting for multiple deposition outcomes detectable by RHEED (such as surface roughness, secondary phase formation, and polycrystallinity). Successful implementation will require both advanced on-the-fly RHEED analysis and a sophisticated control framework. Advances in scientific machine learning, such as differentiable predictive control,[42] could enable these tasks by leveraging experimental data to learn both the dynamics of the deposition process and optimization of process parameter controls.

VI. SUMMARY AND CONCLUSIONS

We have demonstrated an automated, ML-enabled RHEED analysis workflow that can be executed on-the-fly during thin film deposition by MBE. The deposition of epitaxial anatase TiO_2 on $SrTiO_3$(001) by plasma-assisted MBE was utilized as a use case to develop and test the approach. In the workflow, RHEED pattern images are collected at 1 fps, which is sufficiently fast to capture most observable changes. After preprocessing, image featurization is accomplished using a VGG16 CNN. The feature vectors are compiled into a similarity matrix for changepoint identification. Statistical analysis of the similarity matrix retrospectively identifies changepoints that can be correlated with the evolution of the anatase TiO_2 growth mode from initial nucleation to a transition regime to the growth of a smooth film characterized by a streaky RHEED pattern; during some depositions, this pattern deteriorates into a spotty pattern as the film evolves into an islanded growth mode. Graph analysis was utilized to cluster images based on similarity. We showed that the ratio of the first two singular values, SV0/SV1, resulting from SVD of the graph provides a quantification of the stability of the graph. Inflections in the ratio indicate changes in the RHEED images, and the inflection points correlated well with those identified by statistical analysis and postgrowth visual observation of the RHEED images to identify the point at which the smooth anatase deposition became islanded. Such an automated analysis approach is a key component necessary to realize autonomous MBE deposition control.

SUPPLEMENTARY MATERIAL

See the supplementary material for RHEED image preprocessing examples, initial substrate and final anatase film RHEED images, changepoint statistics threshold value sensitivity analysis,

and a video of the RHAAPsody dashboard showing the on-the-fly analysis of deposition B (using prerecorded RHEED image data).

ACKNOWLEDGMENTS

The authors acknowledge the support and assistance of k-Space Associates, Inc. in the development of RHEED image capture and conversion approaches. This research was supported by the AT SCALE Initiative under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). The research used resources available through Research Computing at PNNL. PNNL is a multiprogram national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830. P.G. and R.C. gratefully acknowledge support for synthesis and characterization from the National Science Foundation (NSF) under Award No. DMR-2045993.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Tiffany C. Kaspar: Conceptualization (lead); Funding acquisition (lead); Project administration (lead); Supervision (equal); Writing – original draft (lead); Writing – review & editing (lead). **Sarah Akers:** Conceptualization (equal); Data curation (lead); Formal analysis (lead); Methodology (lead); Software (lead); Writing – original draft (supporting); Writing – review & editing (supporting). **Henry W. Sprueill:** Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Writing – original draft (equal); Writing – review & editing (equal). **Arman H. Ter-Petrosyan:** Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Writing – original draft (supporting). **Jenna A. Bilbrey:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Writing – original draft (supporting). **Derek Hopkins:** Data curation (equal); Methodology (supporting); Software (equal); Writing – original draft (supporting). **Ajay Harilal:** Formal analysis (supporting); Methodology (supporting). **Jijo Christudasjustus:** Investigation (supporting); Writing – original draft (supporting). **Patrick Gemperline:** Formal analysis (supporting); Methodology (supporting). **Ryan B. Comes:** Formal analysis (supporting); Methodology (supporting); Writing – original draft (supporting).

DATA AVAILABILITY

The image analysis codes developed in this work are available online at https://github.com/pnnl/RHAAPSODY. The RHEED image series for deposition A and deposition B are available as TIFF files at https://data.pnnl.gov/group/nodes/dataset/34144. Additional data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

[1]Y. K. Wakabayashi, T. Otsuka, Y. Krockenberger, H. Sawada, Y. Taniyasu, and H. Yamamoto, APL Mater. **7**, 101114 (2019).
[2]S. B. Harris, A. Biswas, S. J. Yun, K. M. Roccapriore, C. M. Rouleau, A. A. Puretzky, R. K. Vasudevan, D. B. Geohegan, and K. Xiao, Small Methods **8**, 2301763 (2024).
[3]S. B. Harris, C. M. Rouleau, K. Xiao, and R. K. Vasudevan, npj Comput. Mater. **10**, 105 (2024).
[4]S. V. Kalinin *et al.*, npj Comput. Mater. **9**, 227 (2023).
[5]A. S. Messecar, S. M. Durbin, and R. A. Makin, MRS Commun. **14**, 660 (2024).
[6]K. L. Bassett *et al.*, Integr. Mater. Manuf. Innov. **12**, 430 (2023).
[7]J. E. Mahan, K. M. Geib, G. Y. Robinson, and R. G. Long, J. Vac. Sci. Technol. A **8**, 3692 (1990).
[8]S. Hasegawa, in *Characterization of Materials*, edited by E. N. Kaufmann (Wiley, Hoboken, NJ, 2012), pp. 1925–1938.
[9]I. Bozovic and J. N. Eckstein, MRS Bull. **20**, 32 (1995).
[10]J. H. Haeni, C. D. Theis, and D. G. Schlom, J. Electroceram. **4**, 385 (2000).
[11]K. Gliebe and A. Sehirlioglu, J. Appl. Phys. **130**, 125301 (2021).
[12]A. Khaireh-Walieh, A. Arnoult, S. Plissard, and P. R. Wiecha, Cryst. Growth Des. **23**, 892 (2023).
[13]H. J. Kim, M. Chong, T. G. Rhee, Y. G. Khim, M. H. Jung, Y. M. Kim, H. Y. Jeong, B. K. Choi, and Y. J. Chang, Nano Converg. **10**, 10 (2023).
[14]J. Kwoen and Y. Arakawa, Cryst. Growth Des. **20**, 5289 (2020).
[15]J. Kwoen and Y. Arakawa, J. Cryst. Growth **593**, 126780 (2022).
[16]H. T. Liang, V. Stanev, A. G. Kusne, Y. Tsukahara, K. Ito, R. Takahashi, M. Lippmaa, and I. Takeuchi, Phys. Rev. Mater. **6**, 063805 (2022).
[17]C. C. Price, Y. Li, G. Zhou, R. Younas, S. S. Zeng, T. H. Scanlon, J. M. Munro, and C. L. Hinkle, Nano Lett. **24**, 14862 (2024).
[18]S. R. Provence, S. Thapa, R. Paudel, T. K. Truttmann, A. Prakash, B. Jalan, and R. B. Comes, Phys. Rev. Mater. **4**, 083807 (2020).
[19]C. Shen *et al.*, Nat. Commun. **15**, 2724 (2024).
[20]R. K. Vasudevan, A. Tselev, A. P. Baddorf, and S. V. Kalinin, ACS Nano **8**, 10899 (2014).
[21]S. Schaefer, D. Febba, K. Egbo, G. Teeter, A. Zakutayev, and B. Tellekamp, J. Mater. Chem. A **12**, 5508 (2024).
[22]T. C. Kaspar and Y. Du, Acc. Mater. Res. **5**, 1013 (2024).
[23]Y. E. Suyolcu, G. Christiani, P. T. Gemperline, S. R. Provence, A. Bussmann-Holder, R. B. Comes, P. A. van Aken, and G. Logvenov, J. Vac. Sci. Technol. A **40**, 10.1116/6.0001473 (2022).
[24]A. R. C. Bredar, M. D. Blanchet, A. R. Burton, B. E. Matthews, S. R. Spurgeon, R. B. Comes, and B. H. Farnum, ACS Catal. **12**, 3577 (2022).
[25]G. Rimal and R. B. Comes, J. Phys. D: Appl. Phys. **57**, 193001 (2024).
[26]S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and C. Scikit Image, PeerJ **2**, e453 (2014).
[27]S. Y. Liu and W. H. Deng, *Proceedings 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Kuala Lumpur, Malaysia, 3–6 November 2015 (IEEE, New York, 2015).
[28]J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, June 20–25, 2009 (IEEE, New York, 2009).
[29]X. X. Chen *et al.*, Med. Image Anal. **79**, 102444 (2022).
[30]D. Garreau and S. Arlot, Electron. J. Stat. **12**, 4440 (2018).
[31]Z. Harchaoui and O. Cappe, *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, Madison, WI, August 26–29, 2007 (IEEE, New York, 2007).
[32]C. Truong, L. Oudre, and N. Vayatis, Signal Process. **167**, 107299 (2020).
[33]T. Bonald, N. de Lara, Q. Lutz, and B. Charpentier, J. Mach. Learn. Res. **21**, 185 (2020).
[34]V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, J. Stat. Mech. Theory Exp. **2008**, P10008 (2008).
[35]E. Andreotti, D. Edelmann, N. Guglielmi, and C. Lubich, Linear Algebra Appl. **610**, 673 (2021).
[36]T. C. Kaspar *et al.*, Phys. Rev. B **73**, 155327 (2006).

[37]R. Shao, C. M. Wang, D. E. McCready, T. C. Droubay, and S. A. Chambers, Surf. Sci. **601**, 1582 (2007).
[38]M. Ohring, *Materials Science of Thin Films: Deposition and Structure* (Academic, San Diego, CA, 2002).
[39]S. A. Chambers, C. M. Wang, S. Thevuthasan, T. Droubay, D. E. McCready, A. S. Lea, V. Shutthanandan, and C. F. Windish, Jr., Thin Solid Films **418**, 197 (2002).
[40]T. W. Zhang, Z. W. Mao, Z. B. Gu, Y. F. Nie, and X. Q. Pan, Appl. Phys. Lett. **111**, 011601 (2017).
[41]P. Hintjens, *ZeroMQ: Messaging for Many Applications* (O'Reilly Media, Sebastopol, CA, 2013).
[42]J. Drgoňa, K. Kiš, A. Tuor, D. Vrabie, and M. Klauco, J. Process Control **116**, 80 (2022).