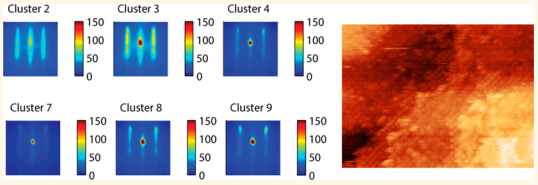
Big-Data Reflection High Energy Electron Diffraction Analysis for Understanding Epitaxial Film Growth Processes

Rama K. Vasudevan,*,†,‡ Alexander Tselev,†,‡ Arthur P. Baddorf,†,‡ and Sergei V. Kalinin*,†,‡

†Center for Nanophase Materials Sciences and ‡ORNL Institute for Functional Imaging of Materials, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States

ABSTRACT Reflection high energy electron diffraction (RHEED) has by now become a standard tool for *in situ* monitoring of film growth by pulsed laser deposition and molecular beam epitaxy. Yet despite the widespread adoption and wealth of information in RHEED images, most applications are limited to observing intensity oscillations of the specular spot, and much additional information on growth is discarded. With ease of data acquisition and increased computation speeds, statistical methods to rapidly mine the data set are now feasible. Here, we develop such an approach to the analysis of the fundamental growth processes through multivariate statistical analysis of a RHEED image sequence. This approach is illustrated for growth of $La_xCa_{1-x}MnO_3$ films grown on etched (001) $SrTiO_3$ substrates, but is universal. The multivariate methods including principal component analysis and k-means clustering provide insight into the relevant behaviors, the timing and nature of a disordered to ordered growth change, and highlight statistically significant patterns. Fourier analysis yields the harmonic components of the signal and allows separation of the relevant components and baselines, isolating the asymmetric nature of the step density function and the transmission spots from the imperfect layer-by-layer (LBL) growth. These studies show the promise of big data approaches to obtaining more insight into film properties during and after epitaxial film growth. Furthermore, these studies open the pathway to use forward prediction methods to potentially allow significantly more control over growth process and hence final film quality.



KEYWORDS: RHEED · big data · surface diffraction · multivariate statistics · oxides · epitaxial film growth

In the past few decades, epitaxial growth of materials by pulsed laser deposition (PLD) and molecular beam epitaxy (MBE) has allowed controlled deposition of films onto crystalline substrates with unit-cell precision.[1,2] This level of precision requires a robust *in situ* measurement technique for monitoring the film properties during the growth, a need which has most often been filled by reflection high energy electron diffraction (RHEED[3]) systems. Ever since oscillations in the specularly reflected RHEED intensity were observed in the early 80s for GaAs grown by MBE,[4] this has become the standard protocol for controlling film thickness during deposition. The period of the observed oscillations is related to the growth rate, while it is reasoned the maximum and minimum of each oscillation correspond (roughly) to complete and half-covered layers, respectively.[5] With the development of differentially pumped systems in more recent years, RHEED could be used in comparatively higher pressures,[6] expanding its utility, and is now commonly used for growth of oxides,[2] nitrides,[7] nanodots,[8] nanostructures,[9] *etc.*

In addition to providing information on the rate of growth, RHEED intensity oscillations can provide some information on the step density[10] or the distribution of islands on the surface.[11] Generally, kinematic theory assumes that the oscillations in intensity of the specularly reflected electrons are caused by an oscillating coverage of the growing layer. The electrons themselves interfere as they glance off of steps of different heights, leading to oscillating amplitude in intensity profile as the step coverage changes. In the alternative dynamic approach, multiscattering is considered and the reflectivity of the layer is calculated

* Address correspondence to rvv@ornl.gov, sergei2@ornl.gov.

Received for review August 22, 2014 and accepted September 30, 2014.

Published online September 30, 2014
10.1021/nn504730n

by considering refraction of the electron wave into the first layer and reflection from the bottom interface. These boundary conditions are affected by the scattering potential, which is assumed to vary linearly with the growing layer, giving rise to observables including oscillations in amplitude and phase shift by change in incident angle.[12] Yet, the exact nature of the origin of oscillations observed in RHEED is still under debate, with seemingly contradictory predictions from dynamic scattering and the kinematic or phenomenological step-density based approaches.[5,13,14] However, there is also a wealth of *static* information that is available in the obtained RHEED images pertaining to the surface crystal structure (*e.g.*, in-plane lattice parameters,[8] reconstructions[15]), degree of disorder,[3] and the type of growth mode[16] that can also be analyzed.

Given the proliferation of RHEED systems in growth chambers worldwide and the wealth of static and dynamic information that can be obtained about the surface from analysis of RHEED data, toolkits to analyze the entire RHEED image sequence should be developed. In the 1990s, systems to record, store, and analyze the data were developed,[17,18] and at present, though the common commercially available RHEED systems[19] are perfectly capable of recording the entire image sequence at appropriate capture rates and high resolutions, most routine analysis of acquired data sets is still limited to either averaging the intensity of one or several diffraction spots (thus yielding a single time-varying intensity plot) or producing a few static images of RHEED diffraction patterns at particular times, *e.g.*, at the start, middle and at the end of the deposition. Thus, the majority of information in the acquired RHEED image sequence is discarded, along with the static and dynamic information on surface properties contained therein.

Here, we demonstrate the approach for full information capture in RHEED based on multivariate statistical tools, including k-means clustering and principal component analysis, as well as Fourier methods. These techniques allow an interpretation of the entire acquired RHEED image sequence without significant loss of information and easily allow trends to be visualized by isolating RHEED patterns based on statistical significance. Fourier analysis yields the harmonic components of the signal and allows separation of the relevant components and baselines, allowing insight into growth mechanisms by filtering out Kikuchi-based interference effects. Furthermore, by providing statistically significant and periodic components, they allow a connection to parameters in a recent kinematic model of RHEED oscillations. Principal component analysis and k-means clustering allows reduction of data dimensionality and separation into clusters based on statistical variance, allowing identification of growth mode transitions. Overall, these studies show the promise of big data approaches to obtaining more insight into film properties during and after oxide growth and provide statistically significant information for theoretical analysis. If implemented *in situ* during growth, these approaches open the pathway to use forward prediction methods to potentially allow significantly more control over growth process and hence final film quality.

RESULTS

As a test case, we analyze a RHEED data set acquired for growth of a 25 unit cell-thick film of $La_{3/8}Ca_{5/8}MnO_3$ (LCMO) on an etched (001) $SrTiO_3$ substrate by pulsed laser deposition (PLD). The RHEED acquisition was carried out with commercially available RHEED software[19] at a frame rate of 30 frames/sec with PixelFLY digital hardware. More details of the growth and scanning tunneling microscopy (STM) of these films will be published elsewhere. The PLD chamber is connected to an STM chamber to allow for *in situ* transfer of samples for imaging. We chose this particular setup and sample as a test case since topographic STM images are available for films of different thicknesses, allowing us to compare the RHEED analysis with the real-space data,[10,11] allowing for unambiguous identification of growth modes.

Parts a−f of Figure 1 show RHEED images from various times of the deposition. At time t = 0 s, the specular reflection is the brightest spot, but the side spots (10) and (−10) from the zeroth Laue zone are also visible (circled in Figure 1a). As the growth of the film proceeds, streaks appear until about time t = 176s, when the pattern returns to one more reminiscent of the flat substrate. The growth lasts for ∼200s in total, with the camera recording the RHEED images shut off after 211s. The mean intensity of the specular spot during the deposition is plotted in Figure 1g, *i.e.*, what is seen during the growth. Deposition start and stop times are indicated by the dark blue arrows in Figure 1(g). Note that there are two points at which the intensity of the RHEED beam was altered; at t = ∼65 s it was increased, and at t = ∼110 s it was decreased (orange arrows in Figure 1g). There is an overall amplitude envelope to the RHEED oscillations, which is reproducible over many samples and consists of an initial decrease in the amplitude (after 7−10 UC), then a recovery in the intensity, and then a slight decrease, the latter of which is presumably due to surface roughening. After deposition, the films were transferred into an adjacent chamber for STM imaging (discussed later).

Principal Component Analysis. In order to observe the statistically relevant components in the data set, we performed principal component analysis (PCA). PCA allows examination of data structure without any underlying physical assumptions and allows identification of statistically relevant behaviors.[20] PCA is a commonly used tool in multivariate statistics[21] and is a linear

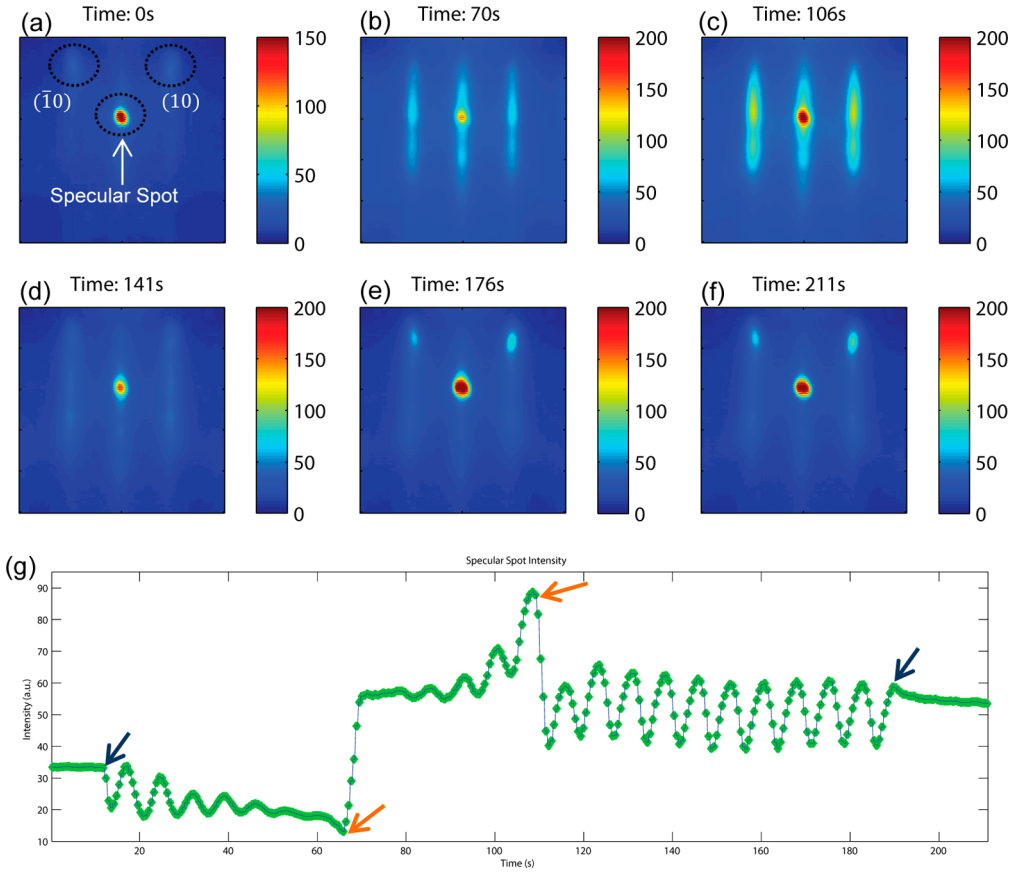


Figure 1. RHEED movie frames. (a−f) Frames from captured RHEED data of $La_{3/8}Ca_{5/8}MnO_3$ growing on (001) $SrTiO_3$ substrate at selected times. The position of the specular spot and the (−10) and (10) side spot from the zeroth Laue Zone are indicated in (a). (g) RHEED intensity from the specular spot (simple mean over the entire spot). The points at which the deposition was started and halted are indicated by blue arrows, and points where the intensity of the beam was manually changed are indicated by the orange arrows. In the first instance, the intensity was increased, and in the second instance, the intensity was decreased.

transformation where a data set is decomposed into a new set of eigenvectors based on criteria of maximum statistical variance between eigenvectors and orthogonality. The first eigenvector captures the maximum amount of variance in the data set, the second represents the second most variance with the constraint that it be orthogonal to the first, and so on for all further eigenvectors. Mathematically for the RHEED image sequence I containing n frames, $I = I(k_x,k_y,t)$, is now rewritten in the form

$$I = a_1(t)e_1(k_x, k_y) + \cdots + a_n(t)e_n(k_x, k_y) \quad (1)$$

$$= \sum_{m=1}^{n} a_m(t)e_m(k_x, k_y) \quad (2)$$

with $e_m(k_x, k_y)$ being the mth eigenvector and a_m indicates the time-dependent eigenvalue. Though the summation is carried out to n, it can be terminated after some threshold percentage of variance in the data has been accounted for as the higher order terms simply account for noise in the data set. The eigenvectors ("principal components") in this decomposition will thus be 2D images and, when multiplied by time-dependent eigenvalues ("loadings"), can reproduce the entire RHEED image sequence. In Figure 2a,b, we show the result of principal component analysis on the RHEED image sequence, with the first six eigenvectors shown in Figure 2a), with the time-dependent eigenvalues shown in Figure 2b. These six eigenvectors (out of 1054) account for 75.7% of the statistical variance in the data set, thereby representing a very large decrease in data size while still retaining the majority of the information. The eigenvalues indicate periodic oscillations, due to the growth, as well as two large changes in baselines indicated by the "x" marks on the component 2 plot in Figure 2b, where the intensity of the RHEED beam was manually adjusted. Note that there is no requirement to maintain positive PCA components during the decomposition, since eigenvectors *per se* do not have defined physical meaning. Furthermore, a negative eigenvector may simply be multiplied by a negative eigenvalue to produce a positive intensity value. The first eigenvector represents a mean intensity profile, while the second component appears to show some additional spots more suggestive of a 3D or imperfect 2D growth. The third component appears to be mostly pure streaks. No shifts could be seen in either streak spacing or width.

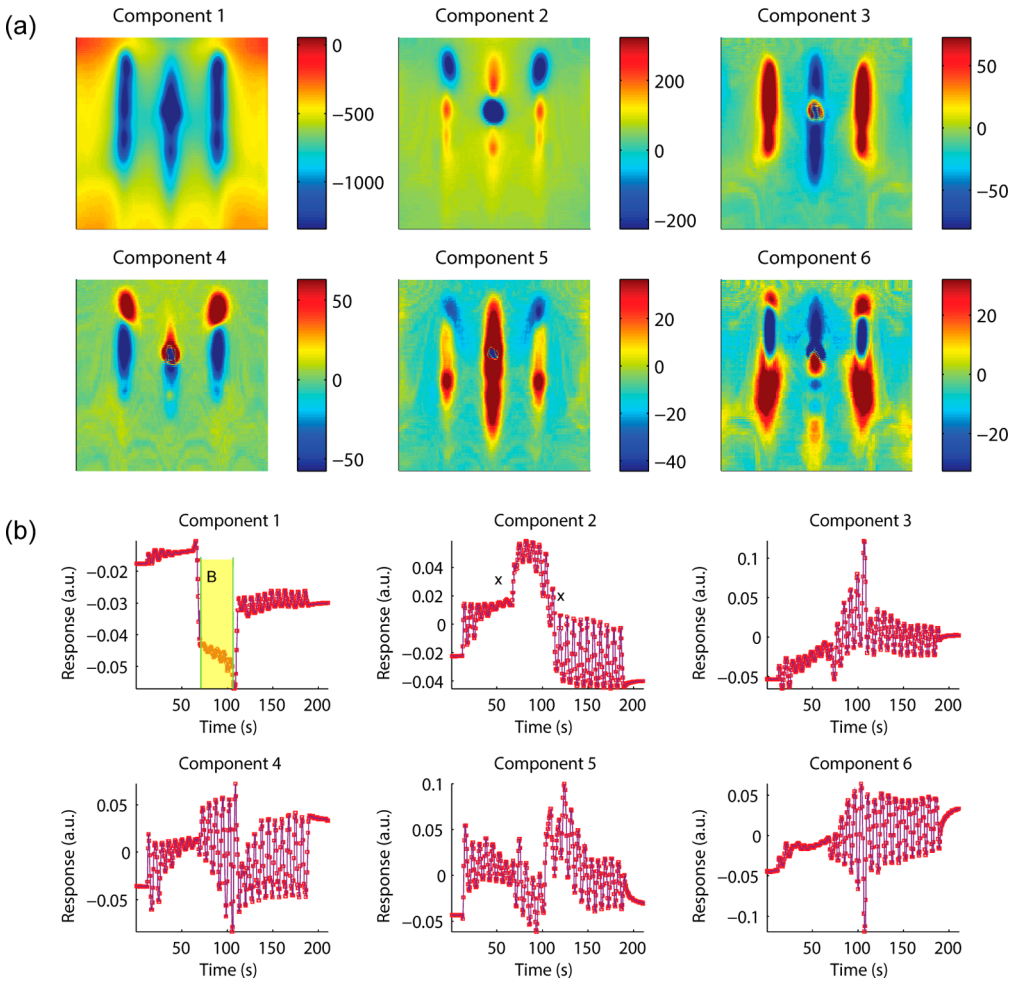


Figure 2. Principal component analysis (PCA). The PCA of the entire acquired data set is shown above with the (a) eigenvectors and (b) eigenvalues plotted. The decomposition results in a set of eigenvectors (images), which when multiplied by the respective time-dependent eigenvalues, can reproduce the entire movie. The intensity of the RHEED beam was altered twice during the acquisition; these times are marked with x and are clearly visible in the plotted eigenvalues. A small segment of the data set is highlighted in yellow in (b), and data in this time window were reanalyzed by PCA with the results shown in Figure 3.

In order to reduce the effect of the manual intensity change on the decomposition, the data were analyzed by PCA restricted to time windows $t = 70$ s and $t = 108$ s, referred to as "segment B" and outlined in the component 1 plot in Figure 2b. In this second decomposition, shown in Figure 3a,b, it is interesting to note that there is little periodicity in the sixth component, and in the eigenvector the two side spots are clearly shifted down, while the specular spot is shifted up, with respect to the other components. In our PLD setup there is a thermal expansion of the substrate suspension system over time through heating, which effectively changes the incident angle of the beam toward more gliding angles. In reciprocal space, this has the effect of tilting the Ewald sphere around a point at its edge taken as the origin of k-space,[22] thereby altering the position of the intersection of reciprocal lattice rods with the sphere, thus moving the diffraction spots vertically on the screen. Therefore, component 6 may be partly a result of an adjustment due to this thermal expansion, although there are may be other factors involved, since there is some remanent periodicity. Although this particular instrumental feature is specific to our system, we note here the ability of PCA to determine the existence of systematic variations in the data in this fashion that can, *e.g.*, be used for data correction, and the method is general. Components 4 and 5 appear to show extra transmission spots and may be related to either 3D growth modes, or imperfect layer by layer growth, as they consist of multiple spots below the (10) and ($\bar{1}$0) spots, while the pattern indicative of 2D growth mode that is more well reflected in component 2 shows a pronounced increase in $|a_2|$ in the oscillation amplitude envelope. Therefore, PCA is a useful tool to identify the relevant behaviors and modes that are present in the data set, as well as in reducing the data set to a manageable size, reducing it by ~99% while retaining ~75% of the information.

k-Means Clustering and Growth Transitions. It is less evident from the PCA analysis whether there are any significant changes in the growth modes occurring in

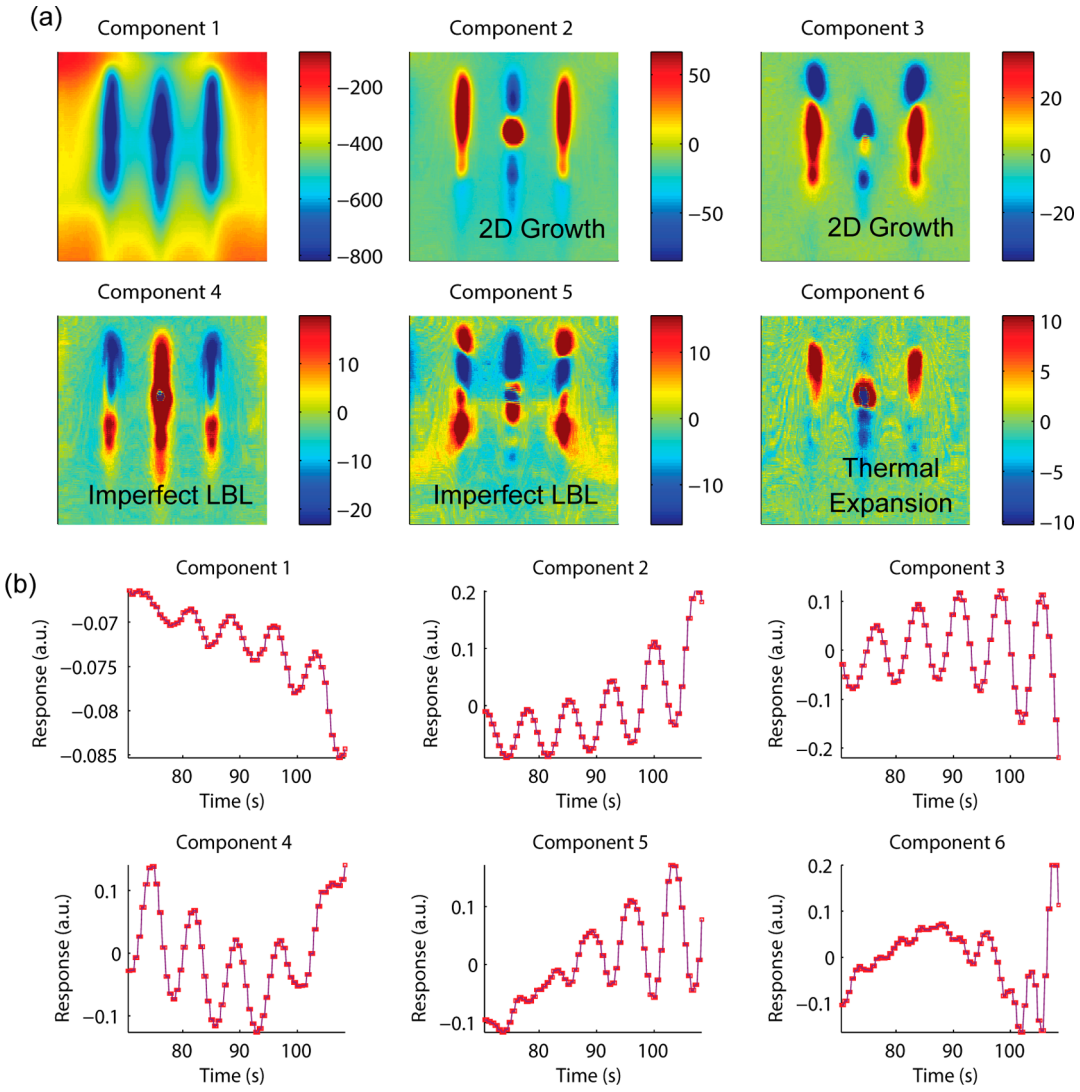


Figure 3. PCA on RHEED segment B. The smaller time-series segment marked B in Figure 2 was reanalyzed with PCA; the results are plotted above with the (a) eigenvectors and (b) eigenvalues. Components 4 and 5 show extra spots characterizing them as imperfect layer-by-layer (LBL) growth modes.

the film, even though there appear to be several components related to different types of growth. To explore this aspect we use the k-means clustering approach, which is a method to partition the RHEED image sequence into *k* clusters in which each observation (image) belongs to the cluster with the nearest mean. These clusters offer an approach to averaging based on statistical similarity of the data, and observe where these averages manifest in time. It should be noted that while it may appear at first glance that PCA and k-means are distinct approaches, the relationship between them has been considered previously,[23,24] establishing the link between the algorithm that performs unsupervised dimension reduction (PCA) and unsupervised learning (k-means). A dendrogram, shown in Figure 4a, is used to determine the optimum cluster number and indicates that this RHEED image sequence can be grouped into 10 clusters. The 10 groups are shown boxed in blue in the dendrogram. Naturally, this dendrogram could be subdivided even further, but this will result in little additional cluster separation. Note that the validity of the chosen cluster number can also be checked by observing the results of the k-means analysis, by checking whether the clusters are indeed oversampled. In case the clusters are oversampled, then the only effect will be that certain clusters will contain members that are statistically very similar; *i.e.*, the mean value of these clusters will be similar.

The k-means approach was applied to the RHEED image sequence, and the data set was divided into 10 distinct clusters. The mean of all members of the individual clusters was computed and is plotted in the upper panel in Figure 4b, while the temporal distribution of the clusters is shown in the lower panel in Figure 4b. Three important clusters between times 70 s and 110 s are bounded by blue dashed rectangles in the upper and lower panels, and they indicate a streaky pattern reminiscent of a more disordered type

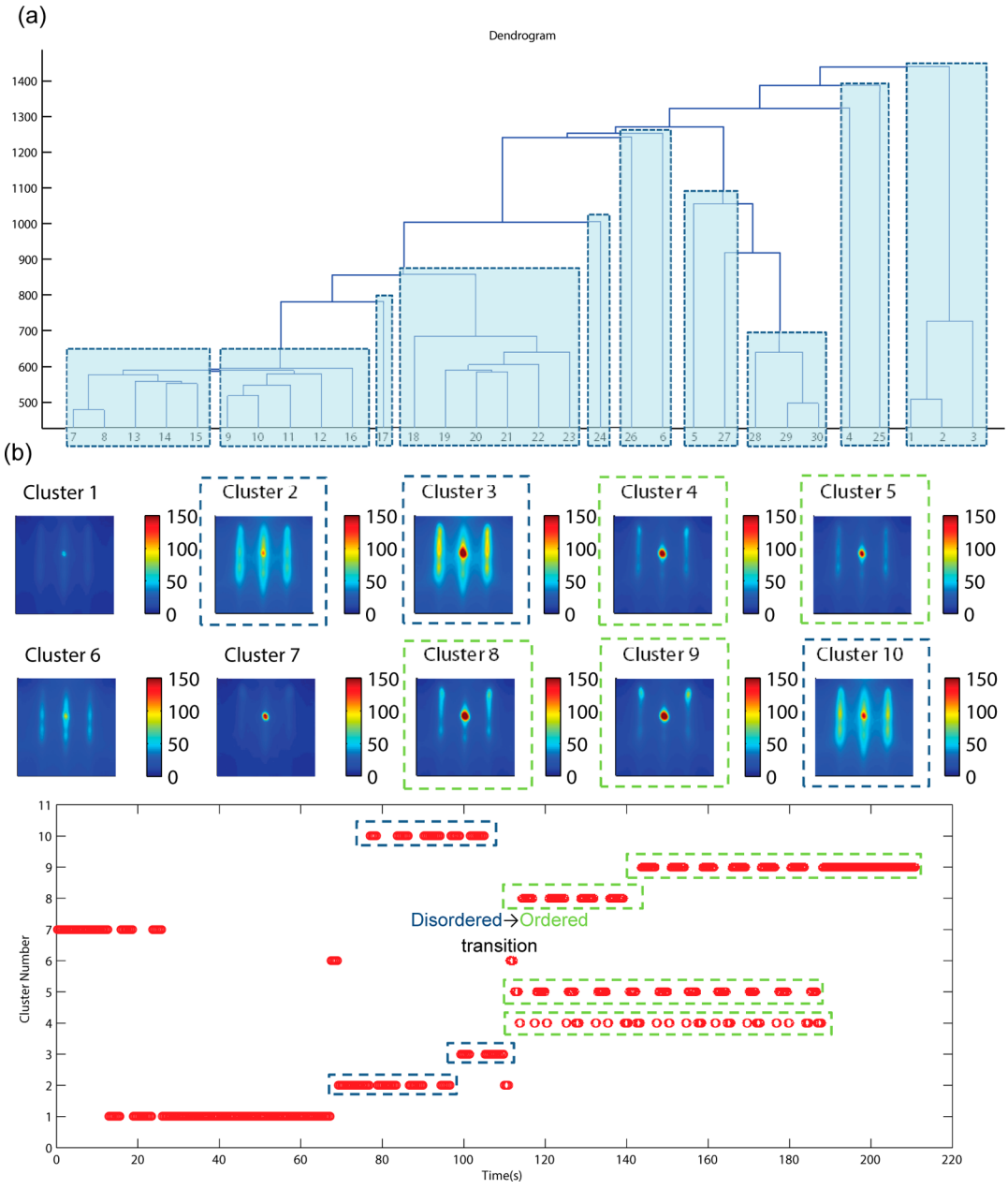


Figure 4. k-Means clustering. (a) Dendrogram of the RHEED image sequence, with leaves grouped into 10 clusters. (b) The result of k-means clustering of the entire data set is shown in the upper and lower panels in this portion of the figure. The upper panels reveal the image obtained by averaging the members of each cluster, while the timeline in the lower panel indicates the temporal dependence of the clusters. At about 115 s, the pattern changes from being dominated by streaks (indicating a more disordered three-dimensional growth) to a clear 2D growth mode dominated by specular and side spots.

of surface. The four clusters bounded by green dashed rectangles appear to show more spot-like diffraction patterns and are less streaky, indicating significantly less disorder in the film once it is growing at these thicknesses. This second set of cluster begins appearing after ~115 s based on the timeline in the lower panel in Figure 4b. This suggests that at $t \sim 115$ s there is a crossover to a more ordered growth pattern in our LCMO films, a feature that is in fact confirmed by STM topography images shown in Figure 5 for both 16 unit-cell (UC) film (corresponding to deposition stop at $t \sim 120$ s) and 25 unit cell-film (present RHEED pattern, deposition stop at $t \sim 210$s). In the thin 16UC film, as shown in the STM topography images in Figure 5a,b, there are far more layers of the surface exposed, leading to much more diffuse scattering and a streaky pattern (finite-size disorder[3]), whereas by 25UC (Figure 5c,d) the growth has transitioned to 2D mode, indicated by the fact that only three layers are exposed. This contrast is further shown through analysis of the surface roughness of the two films, shown in the inset of Figure 5a, where it is evident that the 16UC film has a much wider distribution in the heights (and a roughness RMS of 0.222 nm), as

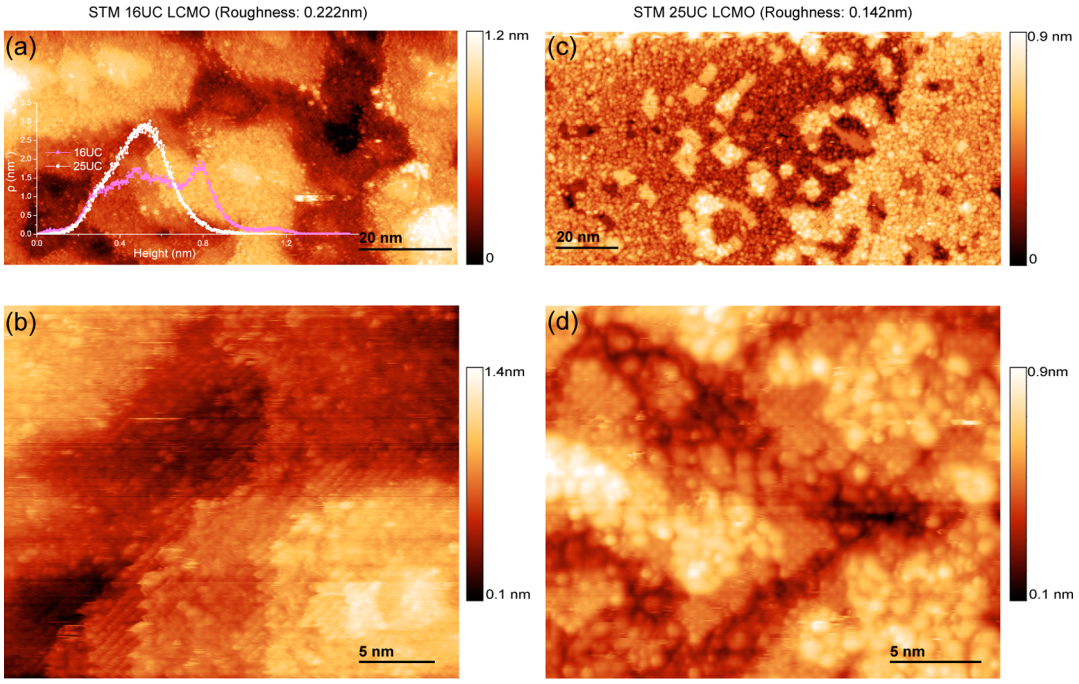


Figure 5. STM topography images of LCMO films of varying thickness. (a, b) STM topography images for a 16 unit cell (UC) LCMO sample. (c,d) STM topography images of the 25 UC LCMO sample, for which the RHEED data is shown in Figure 1. The height distributions from (a, c) are shown inset in the lower left of (a).

compared to the significantly flatter 25UC sample which shows a roughness RMS of 0.142 nm. Note that, here, the intensity of the streaks in the RHEED images is the criteria that we have used to determine the existence of the transition, however in other films it may be the, *e.g.*, appearance of clear 3D spots.[25] These STM images confirm the validity of the statistical analysis undertaken. Thus, k-means clustering can readily identify the presence of growth mode transitions and provides a statistical method to determine the transition point, and we expect this to be equally applicable to, *e.g.*, 2D to 3D growth modes,[16] strain relaxation,[26] *etc*.

FFT Analysis. The PCA revealed the presence of periodic components, and this is also expected from physics of the growth process due to oscillating coverage of the layers. Hence, we now proceed to analyze the data enforcing the periodicity clause, namely by looking for periodic components through Fourier analysis. The data set $I = (k_x, k_y, t)$ was transformed to $I' = I(k_x, k_y, \omega)$ by means of a fast Fourier transform (FFT) at each pixel location. The fast Fourier transform calculates the discrete Fourier transform F_n, which is defined as

$$F_n(k_x, k_y, \omega) = \sum_{k=0}^{N-1} I(k_x, k_y, t_n) e - \frac{2\pi i k n}{N} \qquad (3)$$

where $n = 0, 1, \ldots N - 1$ and ω is the frequency. F_n is a sequence of complex numbers, with the amplitude of each member computed by $|F_n|(Re(F_n)^2 + \mathrm{Im}(F_n)^2)^{1/2}$ and the phase computed by $\tan^{-1}((\mathrm{Im}(F_n)/(Re(F_n))$.

An example of a single-point spectrum from the data set, after application of the fast Fourier transform, is shown in Figure 6a for the pixel $x = 145$, $y = 197$; *i.e.*, we show $|F_n(k_{145}, k_{197}, \omega)|$ in this plot. Clearly, there are a number of peaks present at this particular spatial location due to the periodicity in the signal. The point of maximum intensity is $\max(|F_n|) = |F_\omega| = 987.8$, $\omega = 0.138$ Hz, and the value of the phase at this location (plot not shown) is -2.08 rad. Multiplying the frequency by n ($n = 2, 3, 4, \ldots$) and subsequent determination of the intensity (*i.e.*, amplitude), and phase can be achieved to give the nth harmonic response, *i.e.*, $|F_{n\omega}|$. Repeating this procedure for all (k_x, k_y) points yields the spatial maps of the fast Fourier transforms (separated into the amplitudes and phases for the different harmonics); the mean amplitude spectrum is plotted in Figure 6b, and the spatial FFT maps plotted up to the third harmonic are shown in Figure 6c−h.

In general, although the intensity function is known to be periodic, it is not sinusoidal, and thus, any FFT will necessarily contain higher harmonic components to account for the true shape of the curve. The frequency plot inset in Figure 6a highlights locations where the periodic signal is above the noise level and can be used for masking and reducing the number of pixels analyzed, thereby reducing computation time. The phase image in Figure 6d shows that the transmission spots associated with the imperfect LBL growth modes are significantly out of phase (but less than π rad) with the 2D growth mode diffracted spots. At the same time there is enhanced intensity at the second and third harmonic at these regions of k-space compared with the background.

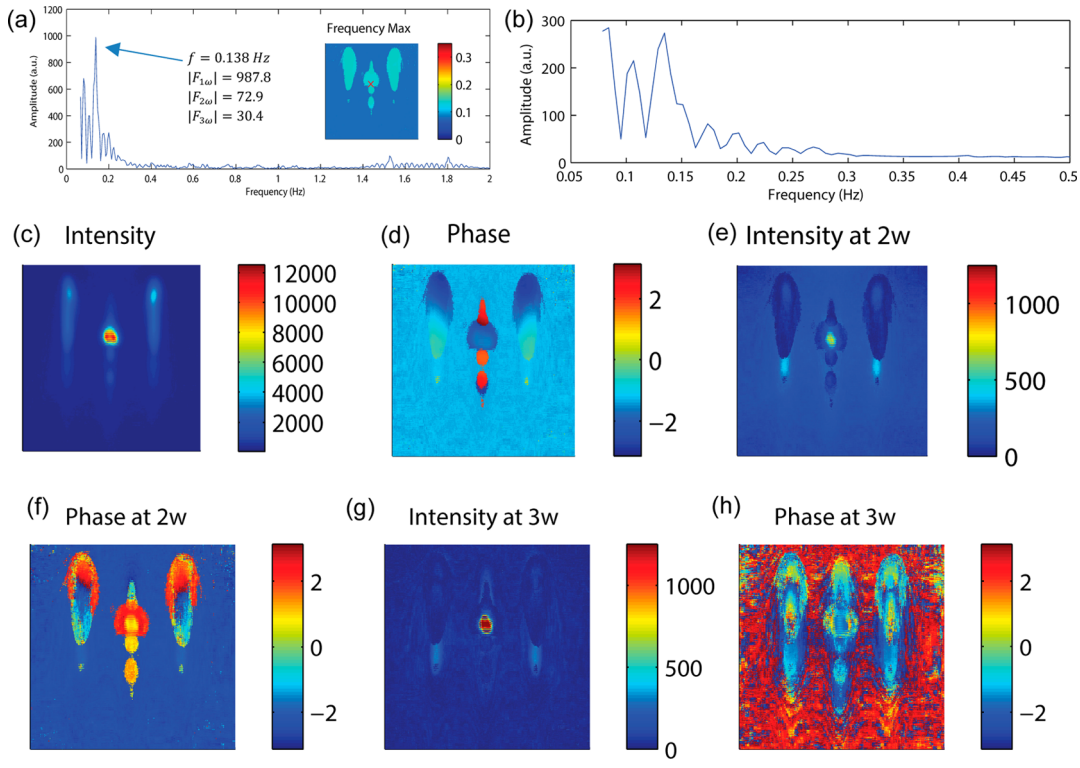


Figure 6. Fourier transformation. After FFT, the amplitude spectrum of a representative single point is plotted in (a). The average spectrogram amplitude for the entire RHEED data set is shown in (b). The frequency at which the maximum intensity occurs, calculated across all points, is plotted in the inset in (a). Correspondingly, the maximum intensities, along with phase, are shown in panels (c) and (d). Second- and third-harmonic component spatial maps are shown in (e–h).

We note that in a few cases higher harmonics have been reported previously,[27,28] although this is (to our knowledge) the first mapping of such harmonics. In those cases, the harmonics were evident in the amplitude oscillations of the specular spot and were explained by interference between diffuse scattering and elastic scattering of electrons.[5] Dobson *et al.*[28] argue that the reconstruction of the signal by filtering the out-of-phase components can allow phase-locked epitaxy conditions to be recovered, wherein intensity maxima always correspond to monolayer completion, since the phase shifts are reasoned to be due to two scattering processes that are π rad out of phase and the proportion to which they contribute to the signal vary with the incident angle. Shin *et al.* have shown experimentally that the phase shift can be caused by overlap of Kikuchi features with the specular spot.[29] Filtering in this manner can be used as a basis for more detailed modeling of surface coverage, but in this case, we have shown the technique can be applied to the entire image sequence. That is, here we can identify the enhanced nonlinearity in the regions of k-space corresponding to 3D growth modes.

Physical Descriptors. Of course, the intensity at the second and especially third harmonic at the specular spot is also of interest, and it can be considered as arising from deviations from the "sinusoidal" oscillations, due to the growth process (which is not expected to be sinusoidal, even ideally). Recently, a model of RHEED oscillations[30] was proposed that incorporates diffuse scattering from surface steps as well as interference between steps. In this case, the intensity I of the specular spot for a system with m exposed layers was approximated as

$$\frac{I}{I_0} = |\theta_m \exp(iq_z md) + \sum_{n=2}^{m} \{[1 - \beta S(\theta_n)](\theta_{n-1}\theta_n)\exp[iq_z(n-1)d]\} + [1 - \beta S(\theta_1)](1 - \theta_1)|^2 \quad (4)$$

where m is the number of incomplete layers, β is the characteristic length of electron scattering from the lower part of the steps, θ_n is the individual layer coverage, d is the interlayer spacing perpendicular to the surface, and q_z is the wavenumber corresponding to momentum transfer along the surface normal and is dependent on the incident angle, and $S(\theta_n)$ is the step density. Inspection of eq 4 reveals that harmonics are clearly to be expected from I/I_0 and will be dependent on the functional form of $S(\theta_n)$. Furthermore, more subtle features such as the period changing over growth should result in either peak splitting at $|F_\omega|$ and at higher harmonics (or perhaps the presence of a shoulder) and can be linked to physical processes. Additionally, it is evident that the spots associated with

the 3D or imperfect LBL growth mode are associated with a large second harmonic, and can be therefore filtered out spatially. Fourier decomposition can thus be used either as a filtering tool (if multiscattering is known to be a problem[28]) or to determine the degree to which the growth is nonlinear by studying the harmonics in the oscillations.

CONCLUSION AND OUTLOOK

RHEED remains an indispensable tool for *in situ* characterization of surfaces during epitaxial growth, providing invaluable insight into crystallography and growth transition modes, yet methods to analyze the entire RHEED data set have not been extensively explored in the literature. Here we have explored methods to analyze and comprehend the information contained in an entire RHEED image sequence for an LCMO film grown on etched (001) STO substrates. Multivariate methods such as PCA allow exploration of the data set, separating the data into time and k-space domains, and greatly compress the data in addition to identifying the relevant behaviors without making assumptions on the physics involved. Relatedly, k-means clustering allows a transition between disordered and more ordered growth to be pinpointed, and is confirmed with real-space STM images. The k-means clustering is useful in reducing the number of RHEED patterns that need to be inspected. Given the periodic nature of the signal, Fourier analysis was warranted and revealed the presence of higher harmonics that were localized in particular regions of k-space. The second harmonic signal was strongest in the regions associated with the nonideal growth modes, while higher harmonic components in the specular spot region arise from the nonsinusoidal nature of the growth oscillations.

In the future, the multivariate toolkit presented here presents opportunities on several fronts. The direct applications allow storing and compressing data, especially though PCA and k-means methods, which can be used to detect changes in growth modes both *via* time dynamics and changes in k-space behaviors. Furthermore, very sensitive *in operando* monitoring of the process can be directly used to adjust growth conditions and operation, *etc*. Such tools open the pathway to use forward prediction methods, *e.g.*, based on Kalman filtering to potentially give significantly more control over final film quality than is currently achieved. Also, one can use more complex statistical methods such as independent component analysis to demix the data signal, thus allowing one to separate the components while imposing physics-based constraints. For example, Bayesian linear demixing has been useful in uncovering relevant behaviors in I−V data in oxides,[31] though such computations were too expensive for this particular data set. Finally, the multivariate tools potentially provide the link to theory *via* separating time dynamics in different regions and k-spaces and finding reliable descriptors. These approaches can be readily automated with commercial software packages and show the promise of big data approaches toward obtaining more insight into film properties during and after epitaxial film growth.

METHODS

Films of LCMO were grown by pulsed laser deposition on etched (100) $SrTiO_3$ substrates. Substrates were prepared by sonication in ultrapure water for 10 min and etching for 30s in buffered oxide etch solution (pH = 4.5). They were then annealed in air for 4 h at 900 °C and allowed to cool at a rate of less than 150 °C/h. After cooling, the substrates were soaked in water and sonicated in ultrapure water at 55 °C again. This recipe yielded mostly B-terminated surfaces with occasional small patches of A-termination. Films were deposited at 750 °C in O_2 pressure of 50 mTorr, and monitoring was performed by *in situ* differentially pumped high-pressure RHEED (Staib system, Pixel Fly camera hardware, and ksa400 Software). The PLD laser was a Lambda Physik KrF laser with a 248 nm wavelength striking a target of composition $La_{0.375}Ca_{0.625}MnO_3$ at a repetition rate of 10 Hz. The cooling rate was 150 °C/min after the end of the deposition, and below 500 °C, the oxygen flow was reduced to zero with the base pressure lower than 1×10^{-6} in an attempt to prevent excess oxygen adsorption. Samples were then transferred *in situ* for STM on an Omicron VT-STM with a Nanonis controller. All analysis was carried out in Matlab v2013b.

Conflict of Interest: The authors declare no competing financial interest.

Acknowledgment. This research was sponsored by the Division of Materials Sciences and Engineering, BES, DOE (R.K.V., A. T., S.V.K.). This research was conducted and partially supported (A.P.B.) at the Center for Nanophase Materials Sciences, which is sponsored at Oak Ridge National Laboratory by the Scientific User Facilities Division, Office of Basic Energy Sciences, U.S. Department of Energy.

REFERENCES AND NOTES

1. Ingle, N.; Yuskauskas, A.; Wicks, R.; Paul, M.; Leung, S. The Structural Analysis Possibilities of Reflection High Energy Electron Diffraction. *J. Phys. D: Appl. Phys.* **2010**, *43*, 133001.
2. Eason, R. *Pulsed Laser Deposition of Thin Films: Applications-Led Growth of Functional Materials*; John Wiley & Sons: New York, 2007.
3. Ichimiya, A.; Cohen, P. I. *Reflection High-Energy Electron Diffraction*; Cambridge University Press: Cambridge, 2004.
4. Neave, J.; Joyce, B.; Dobson, P.; Norton, N. Dynamics of Film Growth of GaAs by MBE from RHEED Observations. *Appl. Phys. A: Mater. Sci. Process.* **1983**, *31*, 1–8.
5. Dobson, P.; Joyce, B.; Neave, J.; Zhang, J. Current Understanding and Applications of the RHEED Intensity Oscillation Technique. *J. Cryst. Growth* **1987**, *81*, 1–8.
6. Rijnders, G. J.; Koster, G.; Blank, D. H.; Rogalla, H. *In Situ* Monitoring During Pulsed Laser Deposition of Complex Oxides Using Reflection High Energy Electron Diffraction under High Oxygen Pressure. *Appl. Phys. Lett.* **1997**, *70*, 1888–1890.
7. Schikora, D.; Hankeln, M.; As, D.; Lischka, K.; Litz, T.; Waag, A.; Buhrow, T.; Henneberger, F. Epitaxial Growth and Optical Transitions of Cubic GaN Films. *Phys. Rev. B* **1996**, *54*, R8381.

8. Daudin, B.; Widmann, F.; Feuillet, G.; Samson, Y.; Arlery, M.; Rouviere, J.; Stranski-Krastanov Growth, Mode During the Molecular Beam Epitaxy of Highly Strained GaN. *Phys. Rev. B* **1997**, *56*, R7069.
9. Bogle, K. A.; Cheung, J.; Chen, Y. L.; Liao, S. C.; Lai, C. H.; Chu, Y. H.; Gregg, J. M.; Ogale, S. B.; Valanoor, N. Epitaxial Magnetic Oxide Nanocrystals *via* Phase Decomposition of Bismuth Perovskite Precursors. *Adv. Funct. Mater.* **2012**, *22*, 5224–5230.
10. Holmes, D.; Sudijono, J.; McConville, C.; Jones, T.; Joyce, B. Direct Evidence for the Step Density Model in the Initial Stages of the Layer-by-Layer Homoepitaxial Growth of GaAs (111) A. *Surf. Sci.* **1997**, *370*, L173–L178.
11. Stroscio, J. A.; Pierce, D.; Dragoset, R. Homoepitaxial Growth of Iron and a Real Space View of Reflection-High-Energy-Electron Diffraction. *Phys. Rev. Lett.* **1993**, *70*, 3615.
12. Braun, W.; Däweritz, L.; Ploog, K. Origin of Electron Diffraction Oscillations During Crystal Growth. *Phys. Rev. Lett.* **1998**, *80*, 4935.
13. Mitura, Z.; Dudarev, S.; Whelan, M. Interpretation of Reflection High-Energy Electron Diffraction Oscillation Phase. *J. Cryst. Growth* **1999**, *198*, 905–910.
14. Korte, U.; Maksym, P. Role of the Step Density in Reflection High-Energy Electron Diffraction: Questioning the Step Density Model. *Phys. Rev. Lett.* **1997**, *78*, 2381.
15. Braun, W.; Möller, H.; Zhang, Y.-H. Reflection High-Energy Electron Diffraction During Substrate Rotation: A New Dimension for *in Situ* Characterization. *J. Vac. Sci. Technol., B* **1998**, *16*, 1507–1510.
16. Boschker, J. E.; Folven, E.; Monsen, Å. F.; Wahlström, E.; Grepstad, J. K.; Tybell, T. Consequences of High Adatom Energy During Pulsed Laser Deposition of $La_{0.7}Sr_{0.3}MnO_3$. *Cryst. Growth Des.* **2012**, *12*, 562–566.
17. Barlett, D.; Snyder, C.; Orr, B.; Clarke, R. Ccd-Based Reflection High-Energy Electron Diffraction Detection and Analysis System. *Rev. Sci. Instrum.* **1991**, *62*, 1263–1269.
18. Gur'yanov, G.; Demidov, V.; Korneeva, N.; Petrov, V.; Samsonenko, Y. B.; Tsyrlin, G. System for Recording and Analysis of Reflection High-Energy Electron Diffraction Patterns. *Technol. Phys.* **1997**, *42*, 956–960.
19. kSA 400, K-Space Associates, Dexter, MI.
20. Jesse, S.; Kalinin, S. V. Principal Component and Spatial Correlation Analysis of Spectroscopic-Imaging Data in Scanning Probe Microscopy. *Nanotechnology* **2009**, *20*, 085714.
21. Rencher, A. C.; Christensen, W. F. *Methods of Multivariate Analysis*; Wiley: New York, 2012.
22. Mahan, J. E.; Geib, K. M.; Robinson, G.; Long, R. G. A Review of the Geometrical Fundamentals of Reflection High-Energy Electron Diffraction with Application to Silicon Surfaces. *J. Vac. Sci. Technol., A* **1990**, *8*, 3692–3700.
23. Drineas, P.; Frieze, A.; Kannan, R.; Vempala, S.; Vinay, V. Clustering Large Graphs *via* the Singular Value Decomposition. *Mach. Learn.* **2004**, *56*, 9–33.
24. Ding, C.; He, X. K-Means Clustering Via Principal Component Analysis. In *Proceedings of the 21st International Conference on Machine Learning*, St. Louis, MO; ACM, 2004; p 29.
25. Nabetani, Y.; Ishikawa, T.; Noda, S.; Sasaki, A. Initial Growth Stage and Optical Properties of a Three-Dimensional InAs Structure on GaAs. *J. Appl. Phys.* **1994**, *76*, 347–351.
26. Massies, J.; Grandjean, N. Oscillation of the Lattice Relaxation in Layer-by-Layer Epitaxial Growth of Highly Strained Materials. *Phys. Rev. Lett.* **1993**, *71*, 1411.
27. Sakamoto, T.; Kawamura, T.; Hashiguchi, G. Observation of Alternating Reconstructions of Silicon (001) 2×1 and 1×2 Using Reflection High-Energy Electron Diffraction During Molecular Beam Epitaxy. *Appl. Phys. Lett.* **1986**, *48*, 1612–1614.
28. Joyce, B.; Zhang, J.; Neave, J.; Dobson, P. The Application of RHEED Intensity Effects to Interrupted Growth and Interface Formation During MBE Growth of GaAs/(Al, Ga) as Structures. *Appl. Phys. A: Mater. Sci. Process.* **1988**, *45*, 255–260.
29. Shin, B.; Leonard, J. P.; McCamy, J. W.; Aziz, M. J. On the Phase Shift of Reflection High Energy Electron Diffraction Intensity Oscillations During Ge (001) Homoepitaxy by Molecular Beam Epitaxy. *J. Vac. Sci. Technol., A* **2007**, *25*, 221–224.
30. Shin, B.; Aziz, M. J. Modeling RHEED Intensity Oscillations in Multilayer Epitaxy: Determination of the Ehrlich-Schwoebel Barrier in Ge (001) Homoepitaxy. *Phys. Rev. B* **2007**, *76*, 165408.
31. Strelcov, E.; Belianinov, A.; Hsieh, Y.-H.; Jesse, S.; Baddorf, A. P.; Chu, Y.-H.; Kalinin, S. V. Deep Data Analysis of Conductive Phenomena on Complex Oxide Interfaces: Physics from Data Mining. *ACS Nano* **2014**, *8*, 6449–6457.