NANO LETTERS

pubs.acs.org/NanoLett Letter

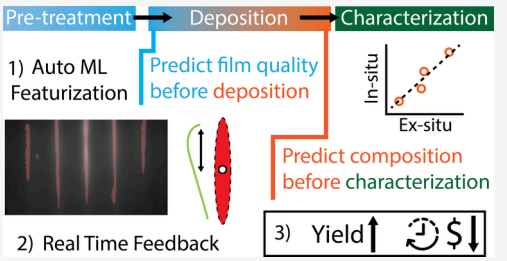
Predicting and Accelerating Nanomaterial Synthesis Using Machine Learning Featurization

Christopher C. Price,* Yansong Li, Guanyu Zhou, Rehan Younas, Spencer S. Zeng, Tim H. Scanlon, Jason M. Munro,* and Christopher L. Hinkle*

Cite This: *Nano Lett.* 2024, 24, 14862−14867 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Materials synthesis optimization is constrained by serial feedback processes that rely on manual tools and intuition across multiple siloed modes of characterization. We automate and generalize feature extraction of reflection high-energy electron diffraction (RHEED) data with machine learning to establish quantitatively predictive relationships in small sets (∼10) of expert-labeled data, saving significant time on subsequently grown samples. These predictive relationships are evaluated in a representative material system ($W_{1-x}V_xSe_2$ on c-plane sapphire (0001)) with two aims: 1) predicting grain alignment of the deposited film using pregrowth substrate data and 2) estimating vanadium dopant concentration using in situ RHEED as a proxy for ex situ methods (e.g., X-ray photoelectron spectroscopy). Both tasks are accomplished using the same materials-agnostic features, avoiding specific system retraining and leading to a potential 80% time saving over a 100-sample synthesis campaign. These predictions provide guidance to avoid doomed trials, reduce follow-on characterization, and improve control resolution for materials synthesis.

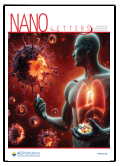


KEYWORDS: *Machine Learning, Epitaxial Growth, 2D Materials, Electron Diffraction, Synthesis Control*

Differentiated and substantial performance requirements for emerging electronics applications and the deceleration of Moore's law in silicon are driving demand for advanced materials discovery, optimization, and scale-up.[1] Engineering and development of materials platforms are difficult and time-consuming; lab-to-production timelines currently take 10 years or longer, and time to market is the primary barrier to commercialization.[2] Significant progress has been made by leveraging ab initio physical simulations (DFT)[3,4] and subsequent machine learning (interatomic potentials;[5−7] generative models[8,9]) to efficiently identify and screen stable and synthesizable material candidates in the first stage of advanced materials development. However, the theoretical assumptions of DFT, including the absence of constraints relevant to real-world synthesis, result in a large time and effort barrier to the realization of materials after promising targets are identified.[10] Although computational capabilities for design have taken off, synthesis recipe design, process optimization, and iterative improvement of material quality relies on a relatively slow, manual, and intuition-guided experimental approach. To address this bottleneck, recent efforts in both software and hardware have made advancements toward fully autonomous synthesis and optimization within the laboratory.[11−14] Advanced tools in machine learning and artificial intelligence have proven incredibly useful at targeting both the interpretation of experimental data and the subsequent decision making required as part of feedback loop-based solutions, including Bayesian optimization approaches for accelerating searches in chemical spaces.[15−19] Early versions of these autonomous systems have emphasized the importance and challenges of effective and rapid materials characterization, especially when available data sets are small and the target properties require multiple tools to assess.

Synthesis optimization is difficult because each trial is time-consuming to conduct and evaluate, especially when nanoscale properties need to be interrogated. While ultrahigh vacuum techniques such as molecular beam epitaxy (MBE) have highly controlled synthesis environments, the preparation, processing, and subsequent characterization of a single sample takes multiple days.[20] Synthesis recipes are highly sensitive, varying across equipment installations and requiring recalibration after tool maintenance, which can extend over weeks. Due to the expense in time and resources consumed per run, it is critical to maximize the information gained and chance of success for each trial in both manual and autonomously driven settings. In-situ characterization captures large volumes of abstract data with high granularity, yet these data cannot be analyzed with conventional methods in time to impact a trial in progress. An example is reflection high energy electron diffraction

Received: September 16, 2024
Revised: November 4, 2024
Accepted: November 6, 2024
Published: November 12, 2024



https://doi.org/10.1021/acs.nanolett.4c04500
Nano Lett. 2024, 24, 14862−14867

(RHEED), frequently used to qualitatively monitor MBE growth[21] by providing information on the surface structure of a sample. RHEED images contain a fingerprint of the material surface at a point in time that can take 15 min to manually extract for a single image, while processes can change in seconds and data is generated 10 to 100 times per second. Recent work has shown that machine learning can process RHEED data,[22−28] but these early demonstrations required manual tuning of hyperparameters, fitting to specific materials systems or camera settings and delivering results after the run completed. While providing significant postrun insights, these attributes hinder the general predictive capacity to modify or reduce the number of trials in synthesis optimization, since they require significant system-specific data to be acquired up front.

In this work, we develop and demonstrate fully automated and general pipelines using both supervised and unsupervised machine learning models to rapidly extract physically motivated and holistic quantitative fingerprints from RHEED data. We show that these fingerprints can speed up the synthesis feedback loop by constructing predictive models from small data sets (∼10 samples) of labeled trials to provide relevant feedback from ex situ analysis using only in situ inputs. These predictive models are demonstrated in two stages of the synthesis process for the target system, two-dimensional (2D) V-doped WSe_2 on $Al_2O_3(0001)$ (sapphire): 1) evaluating the probability of a substrate to produce grain-aligned film growth and 2) estimating the composition of a dopant in the film before ex situ X-ray photoelectron spectroscopy (XPS) is conducted. For both objectives, the success of the predictive models can save significant time and cost by avoiding doomed trials and reducing the number of steps required to assess the sample. By producing these empirically derived predictions for near-real-time feedback, we show that the synthesis optimization loop can be accelerated, leading to a higher throughput of material samples with the target characteristics.

Precise doping control in 2D materials is a difficult but necessary milestone to achieve for the next generation of power- and space-efficient semiconductors.[29] Vanadium doping in WSe_2 gives p-type doping with spin polarization, making this system a candidate as a high-mobility dilute magnetic semiconductor.[30,31] It is crucial to control the dopant concentration during codeposition of V and W to minimize domain formation and phase separation, key challenges in this material system. Illustrating the limitations of theoretical prediction, reliably synthesizing theoretically stable, uniformly distributed phases requires high-fidelity control of the kinetics connecting the input procedure to the resulting material composition and microstructure. End to end, each trial to map out these relationships can consume over 24 combined hours of tool and active operator time, even excluding the time required for sample loading and equipment standby (Figure 1a). Here, we develop a general framework that can be used to avoid doomed trials and map ex situ measurements to in situ characterization in molecular beam epitaxy to save 80% of the time over a 100 trial synthesis campaign.

The data workflow for automated generalized characterization analysis is listed in Figure 1b. The input RHEED images are passed through a featurization pipeline, which extracts, normalizes, and labels diffraction features categorized in Figure 1d. Images are first cropped to remove artifacts from the detector, and an image segmentation pipeline is composed of two models: a U-Net architecture for RHEED proposed by

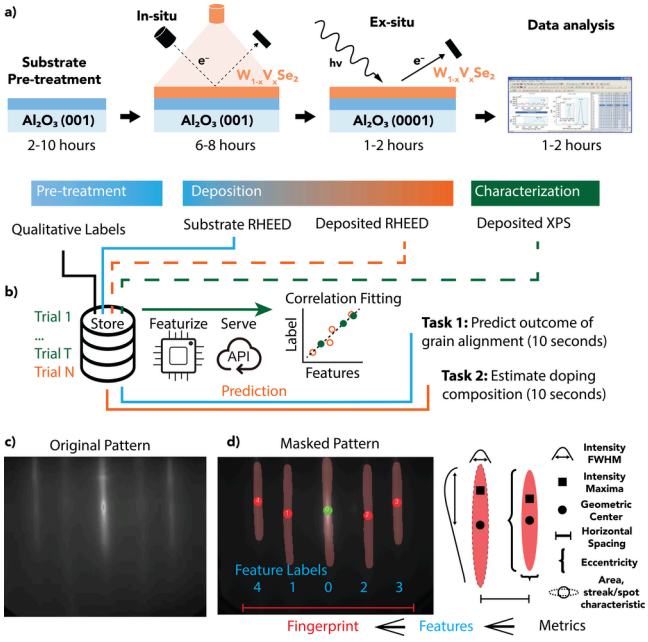


Figure 1. (a) Summary of experimental flows for sample preparation, film growth, and characterization. At the beginning and end of MBE deposition, in situ RHEED is collected and automatically fingerprinted. After synthesis, the sample is transferred for XPS characterization. (b) Summary of data analysis flows for synthesis and characterization data. Labeled trials are iteratively updated in the database, and correlation fitting is performed for the two tasks against the input labels. Next-trial predictions are generated within 10 s. (c) An image of a RHEED pattern of the as-grown film and (d) the color mask representing featurized regions. Comprehensive metrics are extracted for each diffraction feature to form a complete fingerprint unbiased by user priors. Fingerprints are input into the empirical correlation models; see Supporting Information (SI) section 1.

Liang et al.[22] followed by a transformer-based segmentation model[32] tuned for performance on low-contrast medical grayscale images.[33] Output masks from this segmentation pipeline are labeled to identify contiguous diffraction regions, and comprehensive metrics are computed for each diffraction feature (Figure 1d). A coordinate system using the specular spot, identified by position relative to the other features, as the origin is adopted to enable the comparison of diffraction features across different patterns. The original RHEED pattern in Figure 1c shows typically diffuse scattering features that need to be consistently separated from the background, highlighting the challenges of manual analysis and the need for task-specific models. In the featurization scheme, no hyperparameters are input or adjusted across different patterns or material systems to maximize the generalizability of the workflow and enable real-time result generation without operator intervention.

These automatically generated diffraction fingerprints are correlated with qualitative labels on grain orientation and the quantitative results of manual XPS analysis in an effort to 1) predict whether a growth is likely to lead to aligned or randomly oriented grains (textured growth) based on the RHEED image of the substrate wafer before deposition starts and 2) estimate the vanadium doping composition in a deposited film using a RHEED snapshot as input. These tasks were designed to solve real challenges encountered over a three year period of aiming to synthesize high-quality samples. The automated process takes 10 s to produce featurized

https://doi.org/10.1021/acs.nanolett.4c04500
Nano Lett. 2024, 24, 14862−14867

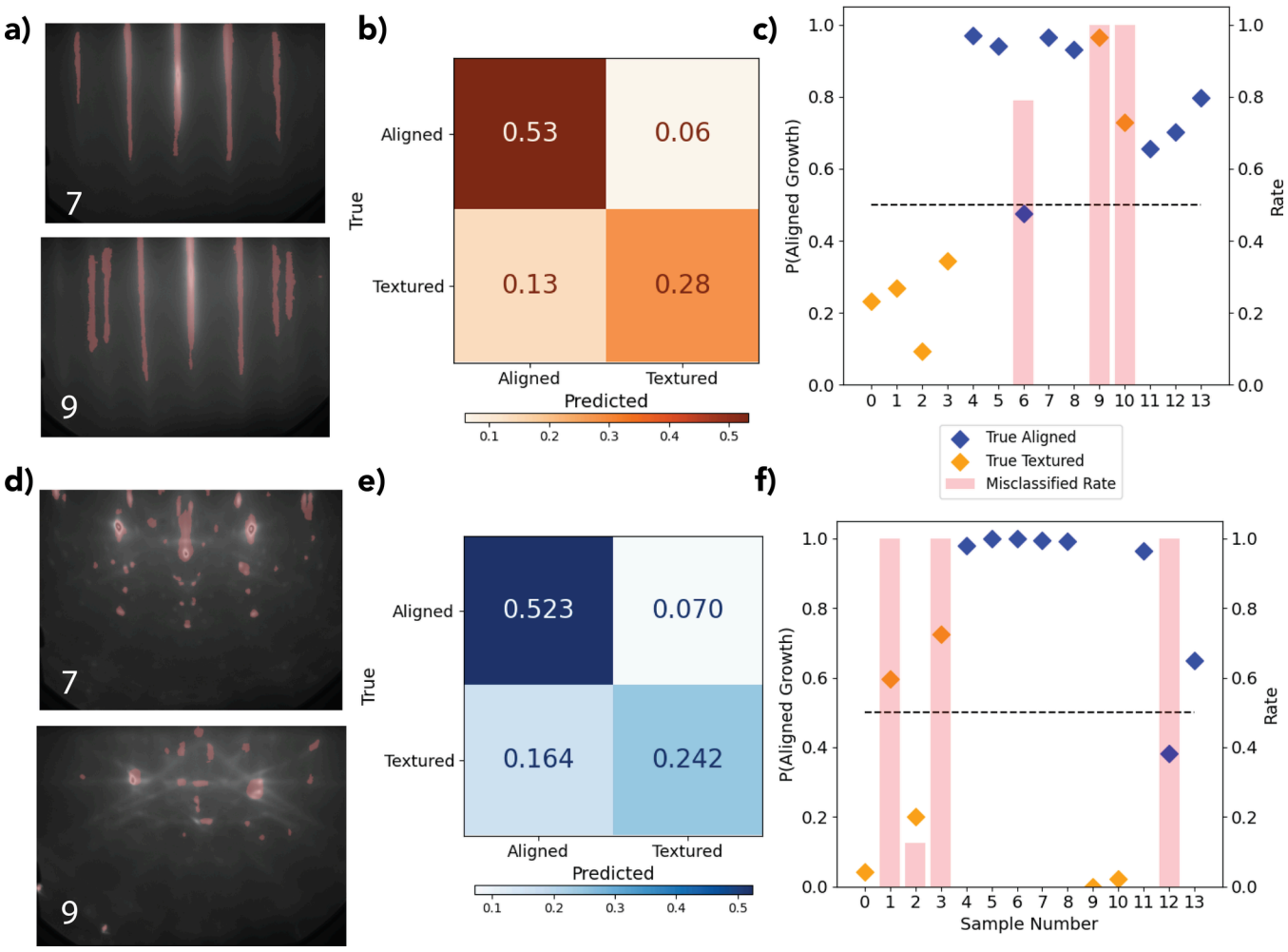


Figure 2. (a) Segmented RHEED patterns for examples of aligned (top) and textured (bottom) WSe_2 film growth. Labels in the bottom left correspond to the sample number. (b) Confusion matrix and classification accuracy for a logistic regression model fit with bootstrap aggregation to a set of 14 samples of featurized WSe_2 patterns. (c) Probability of aligned growth predictions by sample (scatters) and frequency of misclassification (bars) for the WSe_2 RHEED data. (d) Segmented RHEED patterns for examples of sapphire substrates that led to aligned (top) and textured (bottom) film growth. (e) Confusion matrix and classification accuracy for the same model structure in (b), fit to the substrate RHEED instead of the film RHEED against the film labels. (f) Same as (c) for the sapphire substrate pattern classification task.

RHEED data sets per frame and 10 s to generate predictions from the correlative models derived from the task-specific training samples, significantly shortening the feedback time relative to traditional approaches; details are given in SI section 6 (Methods).

Given the critical role of crystallinity in material synthesis and downstream device performance, identifying whether a deposited film has aligned grains can avoid doomed efforts that lead to low-quality samples and wasted time. In the case of 2D chalcogenide growth, epitaxial alignment is particularly susceptible to the surface topography of the sapphire substrate.[34] Surface reconstructions (1×1 Al-terminated or $(\sqrt{31} \times \sqrt{31})R9$) supporting aligned growth are achieved by thermal annealing, but significant variance in the results exists due to coupling of the annealing procedure with the individual wafer and furnace conditions. Figure 2 shows the classification results based on the featurized RHEED data sets and an initial label set from visual inspection which categorizes as-grown films as either textured or aligned. Figure 2a shows examples of strongly aligned (top) and strongly textured (bottom) WSe_2 films, and Figure 2b,c gives the baseline classification results for the deposited WSe_2 films; details of the classification are given in SI 6 (Methods). We restrict ourselves to small training data sets to mimic the typical data availability in the early stages of a synthesis effort and maximize the ability to provide guidance for a subsequent trial. The confusion matrix in Figure 2b gives a binary grain alignment prediction accuracy of 80%; further details of the bagging procedure are given in SI 6 (Methods). The classification probabilities for grain-aligned films are plotted in Figure 2c along with the misclassification frequency for each sample. The probability of classification serves as an uncertainty metric and a quantitative approximation of the degree of overall grain alignment. Some samples are always misclassified when held out of the training set, including canonically textured film 9; an explanation for this is given in SI section 2. Overall, the RHEED features contain enough signal to automatically match the expert-identified trends in the labels with a small set of examples. Automating this task removes operator bias from data analysis, and quantification helps to set thresholds for films that meet the quality criteria for subsequent device fabrication. However, additional operator time, tool time, and resources could be saved by avoiding low-quality film growth before it occurs.

https://doi.org/10.1021/acs.nanolett.4c04500
Nano Lett. 2024, 24, 14862−14867

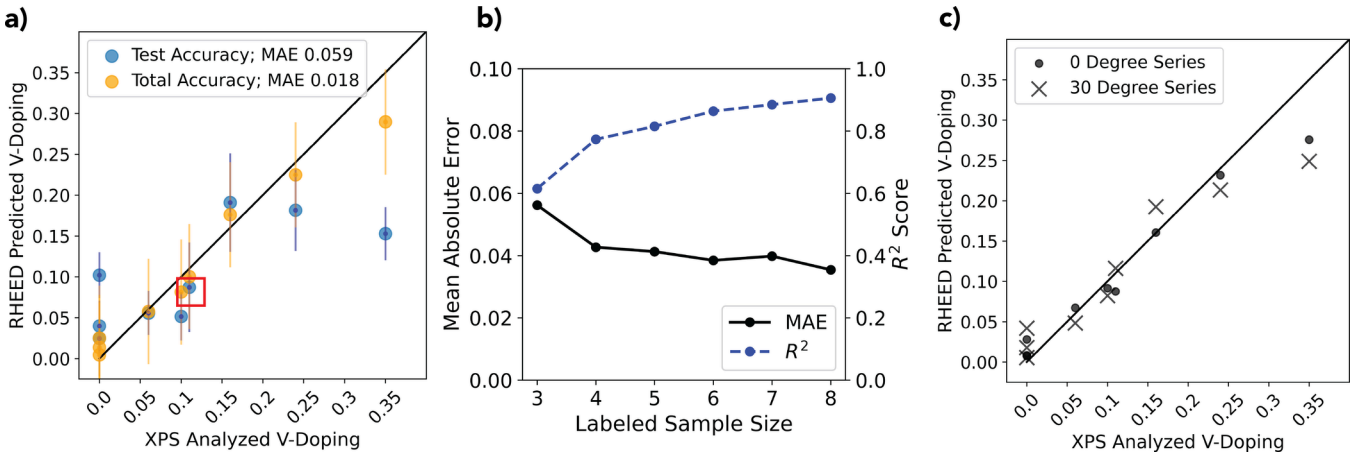


Figure 3. (a) Plot of predicted vs actual vanadium doping composition $W_{1-x}V_xSe_2$ assessed by XPS measurement (x axis) and predicted from RHEED features (y axis). Orange points show the predictions from a model fit to all 9 data points indicative of overall correlation; blue points show the composition prediction for each data point from a model generated with that point withheld. Predictions are the average of models independently fit to 0 and 30° data series. The black line is a visual guide to indicate zero absolute error between the XPS-derived composition and the RHEED-predicted composition result. Error bars give the standard deviation of predictions for the individual estimators within the bagging ensemble. MAE is the mean absolute error of x. (b) Monotonic improvement in prediction accuracy for composition with added training samples, indicating avoidance of overfitting and tunability of desired prediction precision. (c) Predictions separately generated for the two independent RHEED series collected on the same samples at two different azimuthal angles separated by 30° (dots and x's). Averaging the prediction at each labeled composition gives the orange points in (a).

Natural variance with substrates and precursors can lead to unexpected growth outcomes, even if the same recipe is programmatically followed. In this WSe_2 system, deposition alone consumed up to 8 h, and average all-in synthesis times can range from 5 to 14 h for MBE.[35] Using the same data infrastructure and feature extraction as for the WSe_2 films, we perform an identical fitting procedure using the predeposition sapphire RHEED patterns as inputs instead of the as-grown films, with the results in Figure 2d–f. Figure 2d similarly shows the different surface reconstructions of the sapphire that can lead to aligned (top) or textured (bottom) growth. The logistic regression classifier with bootstrap aggregation achieves an accuracy near 80%, similar to the results from WSe_2 RHEED, although the most-misclassified samples differ from those using the film RHEED. For the misclassified samples in Figure 2c, the quantitative probability score is close to 50%, indicating greater uncertainty of prediction. We show that the quantitative classification probability for both data sets correlates with qualitative assessment with a detailed view of sample 12 in Figure S1; this sample is labeled as grain-aligned but shows a lower classification probability for both the substrate and the film. In the sapphire RHEED data for sample 12, several features of the $(\sqrt{31} \times \sqrt{31})R9$ reconstruction are missing compared to the aligned-producing substrate 7, and the Kikuchi lines are better matched with the textured-generating substrate of sample 9. In the as-deposited film, the pattern is consistent with an aligned growth, but there are identified small features that are signatures of the textured films. This indicates that the classification probability contains information about the quantified likelihood of grain alignment in the deposited film that is conditional on the substrate. This provides a new resource for deciding whether to proceed with growth on a given substrate. If the likelihood of achieving high-quality growth is deemed low, then operators have the option to perform additional pregrowth treatment or switch to a different substrate rather than proceeding with a likely doomed trial.

After improving the yield of films with the target grain microstructure, we turn to optimization and control of film composition by estimating the dopant concentration while the sample remains in the growth chamber. Conventionally, dopant concentrations are determined ex situ by characterization techniques such as XPS and energy-dispersive X-ray spectroscopy (EDX) after the entire growth session is complete, requiring time-consuming sample relocation. The scattering factors of different elements create intensity modulations and shifts in the RHEED pattern, but this information is difficult to assess directly since it is tightly convolved with other diffraction mechanisms.[36−38] The relationships among composition, tool parameters, and growth recipe are nonlinear even for highly controlled synthesis environments, and fully mapping design space requires extensive trial and error. If compositional feedback can be quickly generated and delivered from in situ inputs, then better control can be exerted over compositional doping and process refinement.

Figure 3 gives prediction results for V-doping composition based on automatically generated RHEED features; predictions are averaged from the output of two separate models fit on two azimuthal angles collected for each model system ($W_{1-x}V_xSe_2$) sample. By labeling a small set of RHEED images with composition from XPS analysis of the V 2p peak (Figure S3), we show that substantial predictive capability can be uncovered from a small initial data set. Due to the small number of labeled samples, we apply linear regression with bagging to look for correlations between the featurized RHEED and the XPS quantification; details are given in SI 6 (Methods). Interpolative test accuracy is strongest as shown by the blue test predictions in Figure 3a, and aside from the composition end point (x = 0.35) of the target range, there are no substantially outlying predictions. This highlights the importance of combining data-driven practices with experiment design to maximize the strength of the surrogate models: given a set of labeled examples, the surrogate models provide accurate interpolative estimates for new samples. The red

https://doi.org/10.1021/acs.nanolett.4c04500
Nano Lett. 2024, 24, 14862−14867

square in Figure 3a highlights a sample synthesized and characterized by a different operator months after the initial campaign that spanned multiple tool maintenance cycles, demonstrating the persistence of the identified trend to factors that affect the consistency of growth; details are given in Figure S2. Figure 3b shows the prediction accuracy improving monotonically with each additional labeled training point in both the mean absolute error (MAE) and coefficient of determination (R^2), indicating that the correlation model is not overfit. Figure 3c shows close agreement in predictions generated independently from two high-symmetry azimuthal angles, which acts as a physical sanity check and emphasizes that the V-doping-correlated scattering changes are not an artifact of data collection. A full accounting of features input into the composition regression along with their Pearson correlation coefficients for model interpretability is given in Figure S4. The coefficients show that the first order features contain the most correlated variance in the 0° azimuthal data, while the second-order features contain more variance in the 30° series. The full width at half-maximum and feature axis lengths have the greatest correlation magnitudes, indicating that the shape of the internal diffraction intensity distribution is most correlated with the compositional change. These data properties are difficult to assess visually, even for experts without automated tools. Individual metric analysis serves as an entry point for deeper physics-based analysis, indicating atomistic mechanisms that may be correlated with the target property.

We demonstrate that machine learning models tailored for RHEED data can extract high-fidelity feature sets that reveal rich relationships across material systems, even in the limit of small sets of labeled data. Predictions based on these relationships can help avoid synthesis trials with high failure probability, reduce the amount of ex situ characterization required, and provide real-time feedback on properties traditionally measured only ex situ. With research projects needing hundreds of samples and material processing requiring thousands, delivering these predictions with a dedicated data infrastructure could save thousands of expert hours in preparation and analysis. Our approach complements intelligent experimental design algorithms for synthesis, such as Bayesian optimization, by accelerating the acquisition function and providing higher-quality inputs for adaptive searching.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.nanolett.4c04500.

Explanation of RHEED feature set; example of mixed-grain aligned substrate and film RHEED; forward predictive inference highlighting model utility; and RHEED feature label scheme and Pearson correlation coefficients between features and V-composition (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Christopher C. Price − *Atomic Data Sciences, Boston, Massachusetts 02108, United States;* orcid.org/0000-0002-4702-5817; Email: chris@atomicdatasciences.com

Jason M. Munro − *Atomic Data Sciences, Boston, Massachusetts 02108, United States*; Email: jason@atomicdatasciences.com

Christopher L. Hinkle − *Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States;* orcid.org/0000-0002-5485-6600; Email: chinkle@nd.edu

Authors

Yansong Li − *Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States*

Guanyu Zhou − *Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States;* orcid.org/0000-0002-2291-9427

Rehan Younas − *Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States*

Spencer S. Zeng − *Atomic Data Sciences, Boston, Massachusetts 02108, United States*

Tim H. Scanlon − *Atomic Data Sciences, Boston, Massachusetts 02108, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.nanolett.4c04500

Notes

The authors declare the following competing financial interest(s): Some of the authors (CCP, SSZ, THS, JMM) are affiliated with Atomic Data Sciences, a startup developing the software used in this work. We certify that the study was conducted independently and the use of the company's software was chosen based on its suitability for the specific technical requirements of this research.

■ ACKNOWLEDGMENTS

This work was supported in part by SUPREME, one of seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program sponsored by DARPA. This work was also supported in part by the DMREF program of the National Science Foundation (NSF) through the Division of Materials Research (DMR) awards nos. 2324172 and 1921818.

■ REFERENCES

(1) Kim, K. S.; Kwon, J.; Ryu, H.; Kim, C.; Kim, H.; Lee, E.-K.; Lee, D.; Seo, S.; Han, N. M.; Suh, J. M.; Kim, J.; Song, M.-K.; Lee, S.; Seol, M.; Kim, J. The future of two-dimensional semiconductors beyond Moore's law. *Nat. Nanotechnol.* **2024**, *19*, 895−906.

(2) Maine, E.; Seegopaul, P. Accelerating advanced-materials commercialization. *Nat. Mater.* **2016**, *15*, 487−491.

(3) Hegde, V. I.; Borg, C. K. H.; Del Rosario, Z.; Kim, Y.; Hutchinson, M.; Antono, E.; Ling, J.; Saxe, P.; Saal, J. E.; Meredig, B. Quantifying uncertainty in high-throughput density functional theory: A comparison of AFLOW, Materials Project, and OQMD. *Physical Review Materials* **2023**, *7*, 053805.

(4) Yang, R. X.; McCandler, C. A.; Andriuc, O.; Siron, M.; Woods-Robinson, R.; Horton, M. K.; Persson, K. A. Big Data in a Nano World: A Review on Computational, Data-Driven Design of Nanomaterials Structures, Properties, and Synthesis. *ACS Nano* **2022**, *16*, 19873−19891.

(5) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* **2023**, *5*, 1031−1041.

(6) Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science* **2022**, *2*, 718−728.

(7) Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Advances in Neural Information Processing Systems*; 2022, *35*, 11423−11436.
(8) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling deep learning for materials discovery. *Nature* **2023**, *624*, 80−85.
(9) Zeni, C. et al. MatterGen: a generative model for inorganic materials design. *arXiv* 2023, *1*, https://arxiv.org/abs/2312.03687.
(10) Lee, A.; Sarker, S.; Saal, J. E.; Ward, L.; Borg, C.; Mehta, A.; Wolverton, C. Machine learned synthesizability predictions aided by density functional theory. *Communications Materials* **2022**, *3*, 73.
(11) Choudhary, K.; DeCost, B.; Chen, C.; Jain, A.; Tavazza, F.; Cohn, R.; Park, C. W.; Choudhary, A.; Agrawal, A.; Billinge, S. J. L.; Holm, E.; Ong, S. P.; Wolverton, C. Recent advances and applications of deep learning methods in materials science. *npj Computational Materials* **2022**, *8*, 59.
(12) Delgado-Licona, F.; Abolhasani, M. Research Acceleration in Self-Driving Labs: Technological Roadmap toward Accelerated Materials and Molecular Discovery. *Advanced Intelligent Systems* **2023**, *5*, 2200331.
(13) Szymanski, N. J.; Zeng, Y.; Huo, H.; Bartel, C. J.; Kim, H.; Ceder, G. Toward autonomous design and synthesis of novel inorganic materials. *Materials Horizons* **2021**, *8*, 2169−2198.
(14) Xie, Y.; Sattari, K.; Zhang, C.; Lin, J. Toward autonomous laboratories: Convergence of artificial intelligence and experimental automation. *Prog. Mater. Sci.* **2023**, *132*, 101043.
(15) Szymanski, N. J.; et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **2023**, *624*, 86−91.
(16) Lunt, A. M.; Fakhruldeen, H.; Pizzuto, G.; Longley, L.; White, A.; Rankin, N.; Clowes, R.; Alston, B.; Gigli, L.; Day, G. M.; Cooper, A. I.; Chong, S. Y. Modular, multi-robot integration of laboratories: an autonomous workflow for solid-state chemistry. *Chemical Science* **2024**, *15*, 2456−2463.
(17) Biswas, A.; Liu, Y.; Creange, N.; Liu, Y.-C.; Jesse, S.; Yang, J.-C.; Kalinin, S. V.; Ziatdinov, M. A.; Vasudevan, R. K. A dynamic Bayesian optimized active recommender system for curiosity-driven partially Human-in-the-loop automated experiments. *npj Computational Materials* **2024**, *10*, 29.
(18) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **2021**, *590*, 89−96.
(19) Lazin, M. F.; Shelton, C. R.; Sandhofer, S. N.; Wong, B. M. High-dimensional multi-fidelity Bayesian optimization for quantum control. *Machine Learning: Science and Technology* **2023**, *4*, 045014.
(20) Ding, L.; Zhang, C.; Nærland, T. U.; Faleev, N.; Honsberg, C.; Bertoni, M. I. Silicon Minority-carrier Lifetime Degradation During Molecular Beam Heteroepitaxial III-V Material Growth. *Energy Procedia* **2016**, *92*, 617−623.
(21) Hasegawa, S. *Characterization of Materials*; John Wiley & Sons, Inc., 2012; pp 1925−1938.
(22) Liang, H.; Stanev, V.; Kusne, A. G.; Tsukahara, Y.; Ito, K.; Takahashi, R.; Lippmaa, M.; Takeuchi, I. Application of machine learning to reflection high-energy electron diffraction images for automated structural phase mapping. *Physical Review Materials* **2022**, *6*, 063805.
(23) Vasudevan, R. K.; Tselev, A.; Baddorf, A. P.; Kalinin, S. V. Big-Data Reflection High Energy Electron Diffraction Analysis for Understanding Epitaxial Film Growth Processes. *ACS Nano* **2014**, *8*, 10899−10908.
(24) Kwoen, J.; Arakawa, Y. Classification of Reflection High-Energy Electron Diffraction Pattern Using Machine Learning. *Cryst. Growth Des.* **2020**, *20*, 5289−5293.
(25) Gliebe, K.; Sehirlioglu, A. Distinct thin film growth characteristics determined through comparative dimension reduction techniques. *J. Appl. Phys.* **2021**, *130*, 125301.
(26) Yang, J.-H.; Kang, H.; Kim, H. J.; Kim, T.; Ahn, H.; Rhee, T. G.; Khim, Y. G.; Choi, B. K.; Jo, M.-H.; Chang, H.; Kim, J.; Chang, Y. J.; Lee, Y.-L. *Digital Discovery* **2024**, *3*, 573−585.
(27) Provence, S. R.; Thapa, S.; Paudel, R.; Truttmann, T. K.; Prakash, A.; Jalan, B.; Comes, R. B. Machine learning analysis of perovskite oxides grown by molecular beam epitaxy. *Physical Review Materials* **2020**, *4*, 083807.
(28) Kim, H. J.; Chong, M.; Rhee, T. G.; Khim, Y. G.; Jung, M.-H.; Kim, Y.-M.; Jeong, H. Y.; Choi, B. K.; Chang, Y. J. Machine-learning-assisted analysis of transition metal dichalcogenide thin-film growth. *Nano Convergence* **2023**, *10*, 10.
(29) Younas, R.; Zhou, G.; Hinkle, C. L. A perspective on the doping of transition metal dichalcogenides for ultra-scaled transistors: Challenges and opportunities. *Appl. Phys. Lett.* **2023**, *122*, 160504.
(30) Zhang, F.; et al. Monolayer Vanadium-Doped Tungsten Disulfide: A Room-Temperature Dilute Magnetic Semiconductor. *Advanced Science* **2020**, *7*, 2001174.
(31) Yun, S. J.; Duong, D. L.; Ha, D. M.; Singh, K.; Phan, T. L.; Choi, W.; Kim, Y.-M.; Lee, Y. H. Ferromagnetic Order at Room Temperature in Monolayer WSe2 Semiconductor via Vanadium Dopant. *Advanced Science* **2020**, *7*, 1903076.
(32) Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; Dollár, P.; Girshick, R. Segment Anything. *arXiv* 2023, *1*, https://arxiv.org/abs/2304.02643.
(33) Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; Wang, B. Segment anything in medical images. *Nat. Commun.* **2024**, *15*, 654.
(34) Mortelmans, W.; Kazzi, S. E.; Mehta, A. N.; Vanhaeren, D.; Conard, T.; Meersschaut, J.; Nuytten, T.; Gendt, S. D.; Heyns, M.; Merckling, C. Peculiar alignment and strain of 2D WSe2 grown by van der Waals epitaxy on reconstructed sapphire surfaces. *Nanotechnology* **2019**, *30*, 465601.
(35) He, L.; Wang, S. L.; Yang, J. R.; Yu, M. F.; Wu, Y.; Chen, X. Q.; Fang, W. Z.; Qiao, Y. M.; Gui, Y.; Chu, J. Molecular beam epitaxy (MBE) in situ high-temperature annealing of HgCdTe. *J. Cryst. Growth* **1999**, *201−202*, 524−529.
(36) Peng, L.-M. Electron Scattering Factors of Ions and their Parameterization. *Acta Crystallographica Section A: Foundations of Crystallography* **1998**, *54*, 481−485.
(37) Kawamura, T.; Fukaya, Y.; Fukutani, K. Finding RHEED conditions sensitive to hydrogen position on Pd(100). *Surf. Sci.* **2022**, *722*, 122098.
(38) Pawlak, J.; Przybylski, M.; Mitura, Z. An Analysis of Kikuchi Lines Observed with a RHEED Apparatus for a TiO2-Terminated SrTiO3 (001) Crystal. *Materials* **2021**, *14*, 7077.

https://doi.org/10.1021/acs.nanolett.4c04500
Nano Lett. 2024, 24, 14862−14867