

Quantum and Classical Machine Learning Investigation of Synthesis–Structure Relationships in Epitaxially Grown Wide Band Gap Semiconductors

A. S. Messecar^{1,*}, S. M. Durbin², R. A. Makin¹

¹College of Engineering and Applied Sciences, Western Michigan University, Kalamazoo, MI 49008 USA

²College of Engineering, University of Hawai‘i at Mānoa, Honolulu, HI 96822 USA

*Corresponding author: andrew.s.messecar@wmich.edu

ABSTRACT

Several hundred plasma–assisted molecular beam epitaxy synthesis experiments of GaN and ZnO thin film crystals were organized into data sets that correlate the operating parameters selected for growth to two figures of merit: a binary determination of surface morphology, and a continuous Bragg–Williams measure of lattice ordering (S^2). Quantum as well as conventional supervised machine learning algorithms were optimized and trained on the data, enabling a comparison of their generalization performance. The models displaying the best generalization performance on each data set were subsequently used to predict each figure of merit across the ZnO and GaN processing spaces.

KEYWORDS

Machine learning, quantum information, artificial intelligence, molecular beam epitaxy (MBE), thin film

INTRODUCTION

Machine learning technologies are a top strategy for synthesis parameter identification owing to their ability to accurately recognize and determine patterns within data sets, such as those reporting thin film crystal epitaxy trials. Recent studies have demonstrated enhanced material synthesis resulting from insights gained from machine learning techniques such as tree–based models, Bayesian optimization, and unsupervised learning algorithms for a broad variety of systems, including transition metal dichalcogenides [1,2], perovskite oxides [3–7], ZnSnP₂ [8], superconducting TiN [9], and arsenide nanostructures [10].

Here, we investigate the generalization performance of both quantum and classical supervised machine learning algorithms for the prediction of synthesis–structure relationships in epitaxially grown ZnO and GaN. This study is focused on predicting two different figures of merit across a processing space of plasma–assisted molecular beam epitaxy (PAMBE) operating parameters: a binary determination of surface flatness as assessed via *in situ* reflection high–energy electron diffraction (RHEED) [11], and a continuous measure of lattice ordering using the Bragg–Williams parameter (S) [12–14]. The accuracy of the predictions made by the best performing models are evaluated against the operating parameters for PAMBE synthesis of GaN and ZnO reported in the literature.

DATA

Records describing ZnO and GaN PAMBE growth experiments have been organized into separate, material–specific data sets. There are: 430 data points correlating PAMBE operating parameters with GaN surface flatness as determined via RHEED, 294 points associating PAMBE synthesis conditions with ZnO surface morphology, 65 records pairing PAMBE growth parameters with GaN S^2 , and 38 data points correlating PAMBE synthesis parameters with ZnO S^2 . Each experiment was conducted using a Perkin–Elmer 430 molecular beam epitaxy system equipped with an Oxford Instruments model HD–25 inductively–coupled, radio frequency (RF)

plasma source in conjunction with custom E–Science elemental source effusion cells. The complete set of PAMBE operating parameters was recorded for each experiment, including substrate temperature, metal source effusion cell temperature, RF plasma forward power, initial chamber pressure and growth duration. Additionally, for ZnO experiments, the O₂ gas flow rate was recorded from a mass flow controller.

Each synthesis experiment utilized in this work was monitored via *in situ* 20 kV RHEED (see Figure 1a–1b), a technique that can provide real–time information about the deposition rate, growth mode, and crystallinity of the sample [11]. The nature of the streak–like features visible within RHEED patterns contains information regarding the present surface morphology of crystalline samples [11]. Continuous, coalesced streaks, as visible in Figure 1a, indicate an atomically–flat crystal surface. Spotty, discontinuous features, such as those in Figure 1b, are characteristic of an uneven sample surface (on an atomic scale) resulting from vertical as opposed to layer–by–layer growth. To study the relationships between PAMBE operating parameters and the resulting surface morphology of GaN and ZnO, images of RHEED patterns were manually inspected to determine the morphology during growth. RHEED patterns characterizing atomically–flat ZnO or GaN growth were numerically represented by a value of unity within the data set, while RHEED patterns indicating the presence of vertical growth were assigned a value of 0. Every data point used to study either GaN or ZnO surface morphology was then developed by pairing the complete set of PAMBE operating parameters with a corresponding binary flatness label.

Another figure of merit that is of interest for PAMBE–grown thin films is the degree of lattice ordering as measured by the Bragg–Williams order parameter S [12–14]. S has shown promise as a means of developing structure–property relationships for material properties that are dominated by single site and pairwise interactions [15–17]. This figure of merit ranges continuously between 0 and 1, with $S = 1$ corresponding to a perfectly ordered crystal lattice free of antisite defects and $S = 0$ indicating an entirely random placement of atoms upon fixed lattice sites [12–17]. The parameter can be measured as a squared value (S^2) using a variety of different characterization techniques, including scanning electron microscopy (SEM), RHEED, Raman spectroscopy, and x–ray diffraction [16–18]. In this work, S^2 has been measured from SEM images acquired from PAMBE–grown crystalline ZnO and GaN thin films. These S^2 measurements have been paired with the sample synthesis parameter values to form data sets associating PAMBE operating parameters with the resulting value of S^2 in either GaN or ZnO.

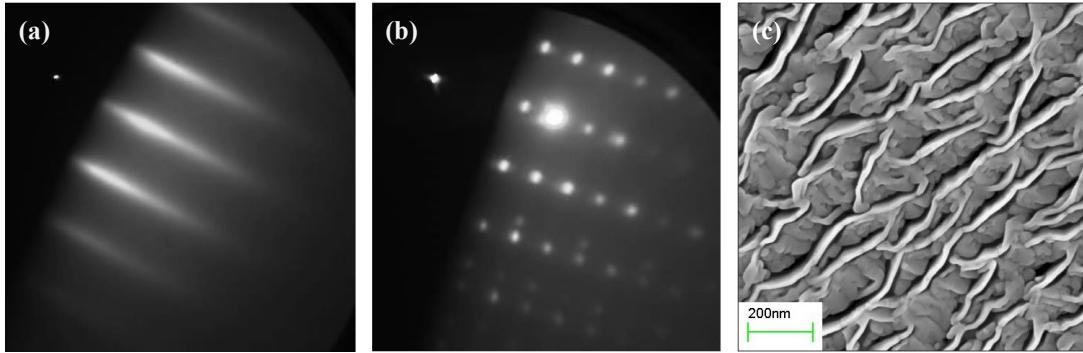


Figure 1. Images of RHEED patterns obtained during PAMBE synthesis of monocrystalline ZnO with (a) flat and (b) uneven surface morphologies. (c) Scanning electron micrograph of a PAMBE–grown ZnO thin film.

METHODOLOGY

A selection of multiple classification machine learning algorithms, consisting of six classical models and one hybrid quantum–classical algorithm, were trained on the ZnO and GaN surface morphology data sets and evaluated for generalization performance. Each data set was normalized prior to the training of the classification algorithms. All supervised learning algorithms were implemented using the Python 3 open–source programming language [19]. All of the conventional classification models investigated (k nearest neighbors, naïve Bayes classifier, random forest classifier, support vector machine, gradient boosting, and an artificial neural network classifier) were developed using the “Scikit–Learn” and “XGBoost” Python 3 libraries [20,21]. “Scikit–Optimize”

was used to implement a Bayesian optimization of hyperparameter values on the conventional algorithms using 4-fold cross validation [22].

The quantum support vector machine model was built with the combination of the “PennyLane” and “Scikit-Learn” libraries [20,23]. The quantum support vector classifier was developed by implementing a quantum variational circuit in PennyLane as a custom kernel function for a support vector classifier in Scikit-Learn. The quantum variational circuit was built using four qubits and used amplitude embedding to encode the training data into quantum mechanical states [23,24]. Unlike the conventional support vector classifier, the quantum support vector machine was not subjected to Bayesian hyperparameter optimization; instead, the L2 regularization parameter remained at the default value of 1.0 [20].

All the models investigated in this study were subjected to a bootstrapping process in order to generate statistics regarding their generalization accuracy on the surface morphology data sets. Each model was trained on 85% of the data points which were randomly selected from the total data set and then tested on the remaining 15%. The test accuracy was recorded for this random sampling of data points. This process was repeated one thousand times for each classifier algorithm, allowing for statistics to be calculated on the test performance accuracy scores resulting from each random sampling of data points. For the data sets associating PAMBE operating parameters with S^2 measured in ZnO and GaN thin film crystals, the same supervised machine learning algorithms that were optimized and trained on the surface morphology data sets, with the exception of the naïve Bayes classifier and the quantum support vector machine, were implemented as regressor models and subjected to the same Bayesian hyperparameter selection and bootstrapping processes. Using the hyperparameters selected through the Bayesian optimization processes, the classifier models as well as the regressor algorithms were evaluated for generalization performance as the percentage of data allotted to training was increased incrementally up to 90%. These analyses were conducted in order to identify any sensitivities to the quantity of training data that the algorithms might exhibit. Figures portraying the results of these analyses are available in the Supplementary Material. To aid in visualizing the predictions made by the trained models, statistical analyses were performed (calculating p-values and correlation coefficients as well as fitting decision trees) to identify the most statistically influential synthesis parameters (results in Supplemental Materials). In each scenario, the predictions (either surface morphology or S^2) were plotted against pairs of growth parameters determined to be statistically influential. Additional plots were made across pairs of variables that are of experimental interest for PAMBE experiment design. For all mappings, variables not plotted against were held constant at their median training values. Predictions made by the trained algorithms were visualized using Gnuplot [25].

RESULTS AND DISCUSSION

The mean and standard deviation of the test accuracy values resulting from the bootstrapping process are shown in Table 1 for each of the classification algorithms. The results of assessing the generalization performance of the regressor algorithms when optimized for and trained on the S^2 data sets are included in the Supplementary Material. When both optimized for and trained on the GaN surface morphology data, the naïve Bayes classifier and artificial neural network classifier tie for best generalization performance in terms of both mean test accuracy score and the standard deviation of the test accuracy scores resulting from the bootstrapping process. In the case of the ZnO surface morphology data, the naïve Bayes classifier displays a higher mean test accuracy than all the other classifiers. There are small differences between the mean test accuracy values calculated between many of the algorithms that have been selected for investigation. The magnitudes of these differences paired with the standard deviations in test accuracy calculated from the bootstrapping process indicate that there is overlap in the generalization accuracies of select classifier algorithms when training on the data sets used in this work. For both materials, the naïve Bayes classifier requires the least amount of time to train of all the classification algorithms evaluated in this study, and the quantum support vector classifier requires the greatest duration of training time. A table comparing the training durations of every algorithm implemented in this work has been made available in the Supplementary Material.

Table 1: Mean and standard deviation of the test accuracy scores calculated from a bootstrapping process performed on each of the classifier algorithms. On the GaN surface morphology data, the artificial neural network algorithm ties the naïve Bayes classifier for highest mean test accuracy. For the ZnO surface morphology data, the naïve Bayes classifier outperforms all of the other classification algorithms in terms of mean test accuracy.

Classifier Algorithm	GaN Morphology		ZnO Morphology	
	Mean Accuracy	Standard Deviation	Mean Accuracy	Standard Deviation
k Nearest Neighbors	0.776	0.025	0.834	0.029
Naïve Bayes	0.779	0.025	0.836	0.030
Random Forest	0.705	0.028	0.764	0.034
Support Vector Machine	0.777	0.024	0.832	0.029
Gradient Boosting	0.776	0.026	0.831	0.030
Artificial Neural Network	0.779	0.025	0.834	0.029
Quantum Support Vector Machine	0.773	0.022	0.834	0.028

Due to its generalization performance in the bootstrapping process, the naïve Bayes classifier was selected to generalize beyond the training data by predicting the class conditional probability of synthesizing flat (as determined by RHEED) GaN and ZnO thin film crystals for a broad range of possible PAMBE operating parameters. The naïve Bayes classifier was trained upon all the available data in either data set before making the predictions. Where applicable, operating parameters have been translated to system-agnostic experiment parameters using calibration curves developed from data obtained during system calibration runs for the growth of GaN and ZnO. The class conditional probability of obtaining a flat epitaxially-grown GaN crystal is predicted across a processing space defined by the values of gallium flux and nitrogen equivalent growth rate, as shown in Figure 2a. Gallium flux and nitrogen equivalent growth rate were converted from gallium effusion cell temperature and initial nitrogen chamber pressure using system-specific calibration curves to enable predictions in terms of system-agnostic parameters. The same class conditional probability is predicted for GaN across the axes of gallium flux and substrate temperature in Figure 2b. The class conditional probability of growing a flat ZnO thin film crystal has been predicted across a growth space consisting of the O₂ gas flow rate and impinging zinc flux values, as shown in Figure 2c. For both materials, the operating parameters not visualized in the mapping are held constant at their median training data values during the predictions made by the trained naïve Bayes classifier algorithms. A table displaying the median training data values for each PAMBE operating parameter from every data set under investigation in this work has been placed in the Supplementary Material along with predicted mappings for additional combinations of ZnO and GaN synthesis parameters.

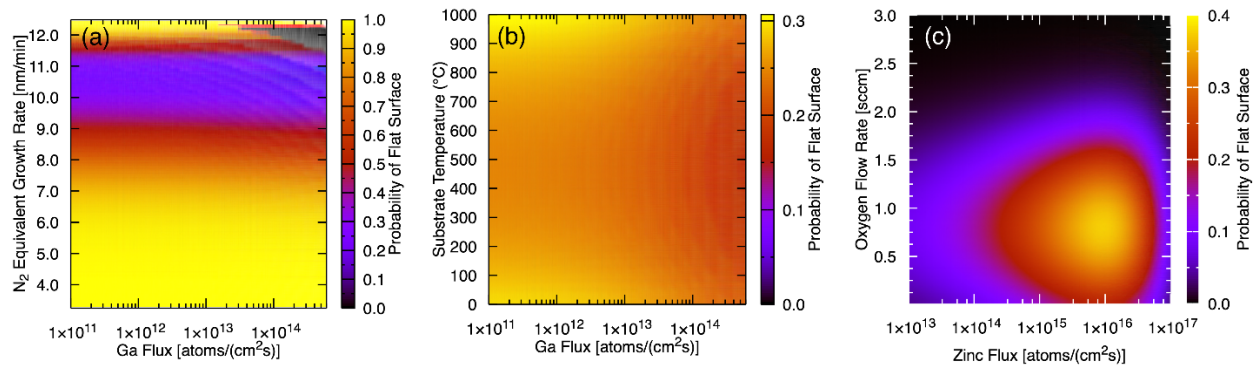


Figure 2: Class conditional probability of obtaining a flat, PAMBE-grown GaN (a–b) or ZnO (c) thin film crystal as predicted by a naïve Bayes classifier when trained on the surface morphology data sets. PAMBE operating parameters not visualized are held constant at the median training values listed in the Supplementary Material. (a) Machine learning predicted class conditional probability values of obtaining a flat GaN sample translated to a processing space defined by impinging gallium flux and nitrogen equivalent growth rate axes obtained using calibration curves. Substrate temperature and RF plasma source forward power are held constant at their median training values of 540 °C and 350 W, respectively. (b) Supervised learning forecasted class conditional probability of GaN PAMBE resulting in a flat surface predicted along axes of substrate temperature and gallium flux. Initial nitrogen chamber pressure and RF plasma source forward power are maintained at their respective median training values of 1×10^{-5} torr and 350 W. (c) Machine learning predicted class conditional probability values of obtaining a flat ZnO sample translated to a processing space defined by O₂ gas flow rate and impinging zinc flux axes as calculated using a calibration curve. Substrate temperature is held constant at 600 °C, RF plasma source forward power is maintained at 400 W, and the growth duration is also kept at the median training value of 9.5 minutes.

The naïve Bayes classifier predicts a minimum class conditional probability of a PAMBE experiment resulting in a flat GaN surface for a region of the growth space centered near 10.5 nm/min on the nitrogen equivalent growth rate axis and spanning a broad range of flux values on the gallium flux axis. The classifier predicts that the maximum probability of obtaining a flat GaN surface occurs outside of a region similarly centered near 10.5 nm/min on the nitrogen equivalent growth rate axis, spanning the gallium flux axis, and intercepting the nitrogen growth rate axis at 12 nm/min and 7.5 nm/min as shown in Figure 2a. The combination of higher impinging nitrogen growth rate and gallium flux values depicted in Figure 2a is predicted to result in a negligible probability of obtaining a flat GaN surface; this is well in agreement with published work showing vertical, columnar GaN growth to result from this region of the PAMBE processing space [26,27]. The naïve Bayes classifier forecasts that the combination of higher substrate temperatures and lower gallium flux values result in the greatest likelihood of growing an atomically-flat GaN thin film crystal, as shown in Figure 2b. Maximizing the value of impinging gallium flux is predicted to minimize the probability of obtaining an atomically-flat GaN surface for all values of substrate temperature.

When forecasting the class conditional probability of obtaining a flat, epitaxially-grown ZnO surface, the decision boundary learned by the naïve Bayes classifier is centered just below 1×10^{16} atoms/cm²s on the zinc flux axis and near 0.85 sccm on the O₂ gas flow rate axis, as shown in Figure 2c. The interior of this decision boundary corresponds to the highest predicted class conditional probability of synthesizing a flat ZnO thin film crystal via PAMBE. The lowest predicted probability of growing a flat ZnO surface occurs outside of this region of the processing space centered just below 1×10^{16} atoms/cm²s on the zinc flux axis and intercepting the O₂ gas flow rate axis near the value of 1.9 sccm. These machine learning predictions were made assuming a substrate temperature of 600 °C and an RF plasma source power of 400 W; uneven PAMBE-grown ZnO surfaces have been shown to occur when synthesized at a growth temperature of 600 °C and with an O₂ flow rate of 2.0 sccm [28], corroborating these predictions. Additionally, PAMBE of uneven ZnO has been demonstrated to result from substrate temperatures at or below 420 °C, zinc effusion cell temperatures of 310 °C, and plasma source power settings at or below 250 W [29], which agrees with the predicted trends made here by the naïve Bayes classifier.

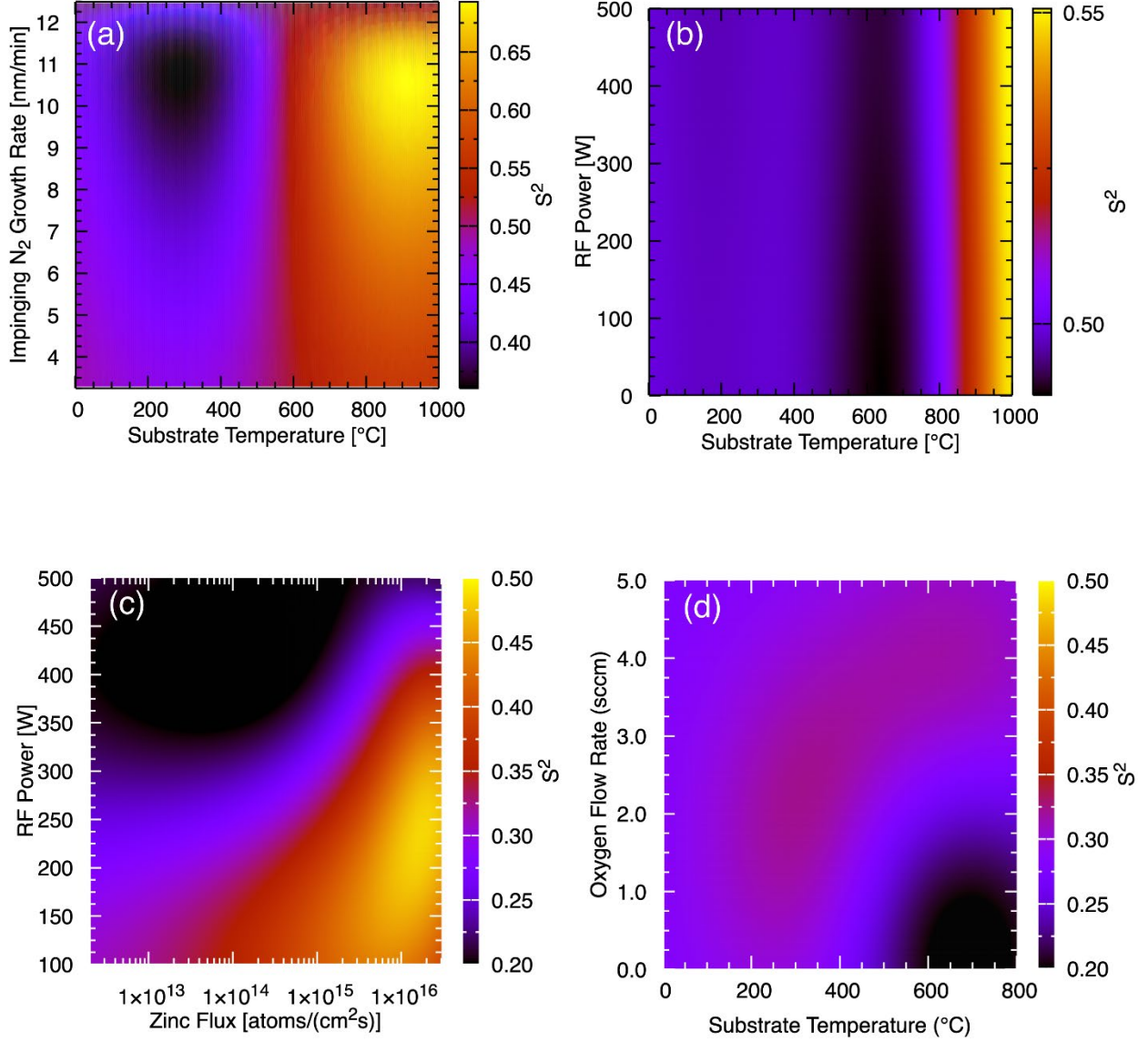


Figure 3: Forecasted value of S^2 in epitaxially-grown (a–b) GaN and (c–d) ZnO thin film crystals as predicted by classical support vector regressor algorithms optimized and trained on the S^2 data sets. (a) Predicted value of S^2 in epitaxially-grown GaN thin films across a growth space defined by impinging nitrogen growth rate and substrate temperature. Gallium effusion cell temperature is maintained at the median training value of 960 °C and the forward power on the RF plasma source is held constant at 150 W (b) Maximizing the values of both substrate temperature and RF plasma source forward power is forecasted to result in the highest predicted value of S^2 in PAMBE-grown GaN when the gallium effusion cell temperature is set at 960 °C and the initial nitrogen chamber pressure is 1×10^{-5} torr (c) Forecasted value of S^2 in epitaxially-grown ZnO across axes of zinc flux and RF plasma source forward power. These predictions are made assuming a substrate temperature of 650 °C, a growth duration of 3 hours, and an oxygen gas flow rate of 1.5 sccm. (d) S^2 in epitaxially-grown ZnO is predicted to be minimized with the combination of higher growth temperatures and lower oxygen gas flow rates when the growth duration is equal to 3 hours, the zinc effusion cell temperature is set to 345 °C, and the RF plasma source forward power is also set to the median value of 387.5 W.

Classical support vector machines were optimized and trained on the ZnO and GaN S^2 data sets; support vector machines were chosen for generalizing beyond the GaN and ZnO S^2 training data due to their superior generalization performance displayed for each material when subjected to Bayesian hyperparameter selection and

subsequent bootstrapping processes. Support vector regressors were used instead of classifiers due to S^2 being a continuously valued response rather than a discrete, binary variable as in the case of the numerical label corresponding to surface morphology. A trained support vector regressor was used to predict S^2 in GaN across a PAMBE processing space defined by the impinging nitrogen growth rate and substrate temperature growth variables as shown in Figure 3a. The highest predicted values of S^2 for GaN thin film samples are predicted to be obtained by substrate temperatures above 800 °C along with impinging nitrogen growth rates between 10 nm/min and 11 nm/min. The lowest forecasted value of S^2 in epitaxially-grown GaN is predicted to result from the combination of growth temperatures between 200 °C and 350 °C and impinging nitrogen growth rates between 9.5 nm/min and 11.5 nm/min. Additionally, as shown in Figure 3b, S^2 in GaN is forecasted to be at its highest when maximizing both RF plasma forward power and substrate temperature. As depicted in Figure 3c, the highest predicted S^2 values for ZnO are predicted to result from ZnO PAMBE synthesis involving zinc flux values on the order of 1×10^{16} atoms/cm²s as well as RF plasma forward power settings between 175 W and 300 W. The lowest values of S^2 in ZnO are predicted to occur for growth using the combination of lower impinging zinc flux values and RF plasma source power settings greater than or equal to 325 W (approximately middle of the typical operation range for the type of plasma source employed). The combination of higher growth temperatures and lower oxygen gas flow rates is forecasted to result in minimal values of S^2 in PAMBE-grown ZnO.

CONCLUSION

Quantum and classical supervised machine learning algorithms were trained on data describing past PAMBE synthesis experiments of ZnO and GaN. The generalization performances of the various algorithms were assessed and compared to one another in order to identify the supervised learning algorithms that display superior generalization performance. The naïve Bayes classifier exhibited the best generalization performance for predicting the surface morphology of both GaN and ZnO thin film crystals as determined by *in situ* RHEED patterns. The trained naïve Bayes classifiers predict decision boundaries corresponding to the maximum and minimum class conditional probabilities of growing a flat ZnO or GaN surface via PAMBE. Classical support vector regressors were optimized and trained on data describing the relationships between PAMBE operating parameters and the resulting lattice ordering of GaN and ZnO as measured by the Bragg-Williams order parameter (S^2). The optimized and trained support vector regressor predicted that the combination of higher impinging zinc flux values and RF plasma source power settings result in ZnO thin film crystals with higher values of S^2 . Consequently, we have shown how similar machine learning model development can utilize distinct figures of merit, and comparison of the resulting optimal growth spaces may provide insight into the relationship between process parameters and growth dynamics in addition to guiding the selection of parameters to achieve desired layer characteristics. In addition, such models can assist in calibrating process parameters across different growth systems, where specific values often have poorly characterized offsets, multipliers, or can vary significantly depending on where in the system they are acquired.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Phillip Anderson, Chito Kendrick, and William Lee for synthesizing the original samples used in this study. This work was supported in part by the National Science Foundation (grant number DMR-2003581) and the MacDiarmid Institute for Advanced Materials and Nanotechnology (film growth). Analysis was performed using hardware received through the NVIDIA Academic Hardware grant program.

CONFLICT OF INTEREST

The authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

Data and code can be made available upon reasonable request.

REFERENCES

- [1] 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, *Machine-Learning-Assisted Analysis of Transition Metal Dichalcogenide Thin-Film Growth*, *Nano Convergence* **10**, 10 (2023).
- [2] A. Costine, P. Delsa, T. Li, P. Reinke, and P. V. Balachandran, *Data-Driven Assessment of Chemical Vapor Deposition Grown MoS₂ Monolayer Thin Films*, *Journal of Applied Physics* **128**, (2020).
- [3] S. R. Provence, S. Thapa, R. Paudel, T. Truttmann, A. Prakash, B. Jalan, and R. B. Comes, *Machine Learning Analysis of Perovskite Oxides Grown by Molecular Beam Epitaxy*, arXiv.Org (2020).
- [4] Y. Wakabayashi, T. Otsuka, Y. Krockenberger, H. Sawada, Y. Taniyasu, and H. Yamamoto, *Machine-learning-assisted molecular beam epitaxy of SrRuO₃ thin films*, JSAP Annual Meetings Extended Abstracts 4106 (2019).
- [5] Y. K. Wakabayashi, T. Otsuka, Y. Krockenberger, H. Sawada, Y. Taniyasu, and H. Yamamoto, *Machine-Learning-Assisted Thin-Film Growth: Bayesian Optimization in Molecular Beam Epitaxy of SrRuO₃ Thin Films*, *APL Materials* **7**, 101114 (2019).
- [6] Y. Wakabayashi, T. Otsuka, Y. Krockenberger, H. Sawada, Y. Taniyasu, and H. Yamamoto, *Crystal quality improvement of SrRuO₃ thin films by machine-learning-assisted molecular beam epitaxy*, JSAP Annual Meetings Extended Abstracts 933 (2020).
- [7] Y. K. Wakabayashi, T. Otsuka, Y. Krockenberger, H. Sawada, Y. Taniyasu, and H. Yamamoto, *Bayesian Optimization with Experimental Failure for High-Throughput Materials Growth*, arXiv.Org **8**, 1 (2022).
- [8] R. Katsube, K. Terayama, R. Tamura, and Y. Nose, *Experimental Establishment of Phase Diagrams Guided by Uncertainty Sampling: An Application to the Deposition of Zn-Sn-P Films by Molecular Beam Epitaxy*, *ACS Materials Letters* **2**, 571 (2020).
- [9] I. Ohkubo, Z. Hou, J. N. Lee, T. Aizawa, M. Lippmaa, T. Chikyow, K. Tsuda, and T. Mori, *Realization of Closed-Loop Optimization of Epitaxial Titanium Nitride Thin-Film Growth via Machine Learning*, *Materials Today Physics* **16**, 100296 (2021).
- [10] C. Shen et al., *Machine-Learning-Assisted and Real-Time-Feedback-Controlled Growth of InAs/GaAs Quantum Dots*, arXiv.Org (2023).
- [11] A. Ichimiya, *Reflection High-Energy Electron Diffraction / Ayahiko Ichimiya and Philip I. Cohen*. (Cambridge University Press, 2004).
- [12] W. L. Bragg and E. J. Williams, *The Effect of Thermal Agitation on Atomic Arrangement in Alloys*, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **145**, 699 (1934).
- [13] W. L. Bragg and E. J. Williams, *The Effect of Thermal Agitation on Atomic Arrangement in Alloys—II*, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **151**, 540 (1935).
- [14] W. L. Bragg and E. J. Williams, *The Effect of Thermal Agitation on Atomic Arrangement in Alloys—III*, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **152**, 231 (1935).
- [15] D. B. Laks, S.-H. Wei, and A. Zunger, *Evolution of Alloy Properties with Long-Range Order*, *Physical Review Letters* **69**, 3766 (1992).
- [16] R. A. Makin, K. York, S. M. Durbin, and R. J. Reeves, *Revisiting Semiconductor Band Gaps through Structural Motifs: An Ising Model Perspective*, *Physical Review B* **102**, (2020).

- [17] R. A. Makin et al., *Alloy-Free Band Gap Tuning Across the Visible Spectrum*, Physical Review Letters **122**, (2019).
- [18] R. A. Makin, K. R. York, A. S. Messecar, and S. M. Durbin, *Quantitative Disorder Analysis and Particle Removal Efficiency of Polypropylene-Based Masks*, (2020).
- [19] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).
- [20] F. Pedregosa et al., *Scikit-Learn: Machine Learning in Python*, Journal of Machine Learning Research **12**, 2825 (2011).
- [21] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, NY, USA, 2016), pp. 785–794.
- [22] T. Head, M. Kumar, H. Nahrstaedt, G. Louppe, and I. Shcherbatyi, *Scikit-Optimize/Scikit-Optimize*, (2021).
- [23] V. Bergholm et al., *PennyLane: Automatic Differentiation of Hybrid Quantum-Classical Computations*, (2022).
- [24] M. Rath and H. Date, *Quantum Data Encoding: A Comparative Analysis of Classical-to-Quantum Mapping Techniques and Their Impact on Machine Learning Accuracy*, (2023).
- [25] T. Williams, C. Kelley, and many others, *Gnuplot 4.4: An Interactive Plotting Program*, (2010).
- [26] S. Fernández-Garrido, J. Grandal, E. Calleja, M. A. Sánchez-García, and D. López-Romero, *A Growth Diagram for Plasma-Assisted Molecular Beam Epitaxy of GaN Nanocolumns on Si(111)*, Journal of Applied Physics **106**, 126102 (2009).
- [27] H. Okumura, B. M. McSkimming, T. Huault, C. Chaix, and J. S. Speck, *Growth Diagram of N-Face GaN (000 1) Grown at High Rate by Plasma-Assisted Molecular Beam Epitaxy*, Applied Physics Letters **104**, 12111 (2014).
- [28] J.-H. Kim, S. Kyu Han, S.-K. Hong, J. Wook Lee, J. Yong Lee, J.-H. Song, S. Ig Hong, and T. Yao, *Growth of Epitaxial ZnO Films on Si (1 1 1) Substrates with Cr Compound Buffer Layer by Plasma-Assisted Molecular Beam Epitaxy*, Journal of Crystal Growth **312**, 2190 (2010).
- [29] Y. Li, H.-Q. Wang, H. Zhou, D. Du, W. Geng, D. Lin, X. Chen, H. Zhan, Y. Zhou, and J. Kang, *Tuning the Surface Morphologies and Properties of ZnO Films by the Design of Interfacial Layer*, Nanoscale Research Letters **12**, 551 (2017).

AUTHORS CONTRIBUTIONS

All authors contributed equally to this manuscript.

FUNDING

This work was supported in part by the National Science Foundation (grant number DMR–2003581).