ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #44681

Quantum and Classical Supervised Learning Study of Epitaxially–Grown ZnO Surface Morphology

Mr. Andrew Steven Messecar, Western Michigan University

Andrew Messecar is a Ph.D. candidate at Western Michigan University's Department of Computer Science. He works with Dr. Robert Makin in the College of Engineering and Applied Sciences' Molecular Beam Epitaxy Laboratory. His research interests include materials and process informatics, the epitaxial synthesis of novel electronic materials and devices, and the simulation of physical systems using quantum and classical computation.

STEVEN DURBIN
Robert Makin, Western Michigan University

Quantum and Classical Supervised Learning Study of Epitaxially–Grown ZnO Surface Morphology

Andrew S. Messecar*
Dept. of Computer Science
Western Michigan University
Kalamazoo, MI 49008

Steven M. Durbin
College of Engineering
University of Hawai'i
Honolulu, HI 96822

Robert A. Makin
Dept. of Computer Science
Western Michigan University
Kalamazoo, MI 49008

**Email: andrew.s.messecar@wmich.edu*

Abstract

Material synthesis parameter spaces typically have a very high dimensionality and are often intractable in size. Exploring these vast, multi–dimensional processing spaces by trial–and–error experimentation – even for well–studied materials – is not feasible on reasonable time scales. Thus, considerable interest exists in the development of machine learning–based approaches for the rapid and accurate identification of optimal materials designs and synthesis conditions. In this work, data describing over 125 plasma–assisted molecular beam epitaxy (PAMBE) synthesis experiments of ZnO thin film crystals have been organized into a single data set. For each growth record, the complete set of PAMBE operating parameters for ZnO synthesis are associated with a measure of crystal surface morphology as determined by *in-situ* reflection high–energy electron diffraction (RHEED) patterns. Quantum and classical supervised learning algorithms – including logistic regression, support vector machines, and a quantum variational circuit – are trained on the data and used to study which growth parameters are most statistically important for influencing surface morphology in epitaxially–grown ZnO thin films. Comparisons are drawn between the generalization performances of the various algorithms that are trained on the data. The support vector classifier exhibited superior generalization performance among the compared algorithms and is used to predict the surface morphology of ZnO thin film crystal across processing spaces defined by the most statistically important synthesis parameters. These supervised learning–based predictions yield experiment design rules which can be used to inform future ZnO PAMBE growth trials. This analysis offers a valuable perspective on the mechanisms that are active during the PAMBE synthesis of ZnO and other related oxide compounds.

1. Introduction

The ability for machine learning techniques to accurately determine the underlying relationships between predictors and responses makes their application one of the top emerging strategies for studying a broad variety of complex systems. One field in which machine learning has begun to provide actionable insights is in the optimization of thin film material synthesis experiments. Recent studies have demonstrated enhanced material synthesis resulting from

insights gained from machine learning techniques such as tree–based algorithms, Bayesian optimization, and unsupervised learning algorithms for a broad variety of systems, including transition metal dichalcogenides[1–2], perovskite oxides[3–4], and arsenide nanostructures[5].

In this study, we apply supervised learning algorithms to data describing the synthesis of ZnO via plasma–assisted molecular beam epitaxy (PAMBE). One of the advantages to PAMBE material synthesis is the ability to perform *in–situ* reflection high–energy electron diffraction (RHEED), which offers real–time information about the growth process as well as the structure and quality of the thin film[6]. Of particular interest is an assessment of ZnO thin film crystal surface morphology via RHEED patterns, since high quality surfaces are required for integrating ZnO into the manufacturing of electronic devices[7].

Our aim in this manuscript is to examine the performance of two classical and one quantum supervised learning technique(s) when predicting the surface quality of PAMBE–grown ZnO as a function of thin film deposition parameters. This work encompasses the results of ongoing doctoral research on exploring the applicability of supervised machine learning technologies to the study of the processing spaces for epitaxially–grown thin film crystals.

2. Experimental Details

2.1 Data

Information describing ZnO PAMBE synthesis experiments was organized into a tabular data set of 294 data points. Each ZnO sample was grown inside of a single Perkin–Elmer 430 molecular beam epitaxy (MBE) chamber equipped with an oxygen plasma source. The growth experiments of interest to this project were monitored via *in–situ* RHEED diagnostics (see Figure 1). RHEED patterns are formed by electrons, incident at shallow angles, diffracting the surface of the grown sample and contain real–time information about sample crystallinity, growth mode, and deposition rate[6]. Of interest for this study is the nature of the streak–like features observed within RHEED patterns of monocrystalline material samples. These RHEED patterns contain information regarding the growth mode and corresponding nature of the sample's surface[6]. Uniform, solid streaks in the pattern, as shown in Figure 1a, are characteristic of an atomically flat sample surface. By contrast, spotty parallel lines, as can be seen in Figure 1b and Figure 1c, indicate an uneven crystal surface. By visually inspecting images of RHEED patterns acquired during ZnO deposition, a binary classification of surface quality was developed to correspond with each image. RHEED patterns representing atomically flat ZnO sample surfaces were numerically represented in the data set by a value of 1, while patterns resulting from uneven (either partially or completely) ZnO surfaces were assigned a value of 0.

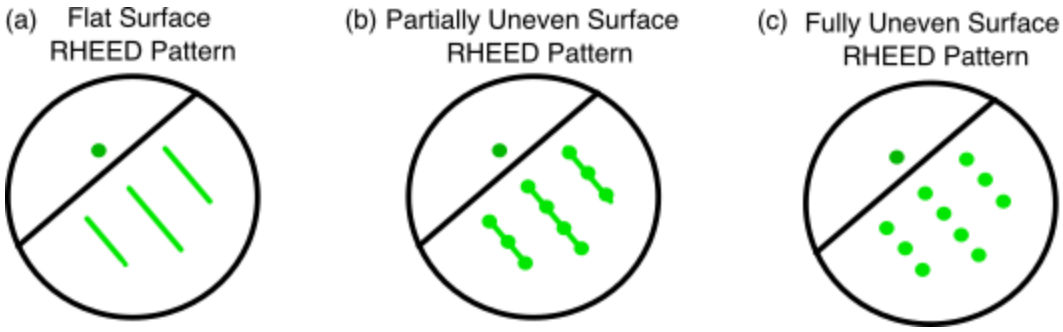

Fig 1. Diagrams of RHEED patterns corresponding to (a) flat, monocrystalline ZnO sample, (b) partially uneven, monocrystalline ZnO, and (c) fully uneven, monocrystalline ZnO sample.

For every available RHEED image, the binary classification corresponding to epitaxially–grown ZnO sample surface quality was paired with the complete set of PAMBE operating parameters that were in place when the image was acquired. This includes values for substrate temperature, zinc effusion cell temperature, oxygen gas flow rate, plasma source forward power level, and growth duration. Statistical metrics for the input variables over the complete data set are listed in Table 1. By pairing PAMBE operating parameters with the resulting surface quality in this way, the data set can be used to map the values of the synthesis parameters to the binary response variable using supervised learning.

Table 1: Characteristic values of the PAMBE operating parameter data for the 294 data points describing ZnO epitaxy that are utilized in this study. Included are the maximum, minimum, arithmetic mean, median, and mode values of each PAMBE synthesis parameter in the data set.

	Maximum	Minimum	Mean	Median	Mode
Substrate Temperature	800 °C	300 °C	559.2 °C	600 °C	650 °C
Zinc Effusion Cell Temperature	500 °C	334 °C	471.8 °C	465 °C	345 °C
Oxygen Flow Rate	5.0 sccm	0.5 sccm	1.14 sccm	1.0 sccm	0.5 sccm
RF Plasma Source Power	500 W	200 W	360.8 W	400 W	400 W
Growth Duration	240 min	15 sec	57.96 min	9.5 min	180 min

2.2 Model Development

All data analysis and development of machine learning models was done in the Python and R open–source programming languages[8–9]. Multiple supervised learning algorithms were developed in order to investigate how effectively each model can learn the underlying relationship between PAMBE operating parameters and the resulting surface morphology of epitaxially–grown ZnO as it has been numerically labeled in this project. Logistic regression was implemented in R for the purpose of assessing the statistical importance of the different input variables for influencing the binary numerical response. A series of three additional supervised learning techniques, two conventional algorithms and one quantum machine learning algorithm, were implemented in Python to map the input values of the ZnO PAMBE growth parameters to the binary response output variable.

The two conventional supervised learning models, a support vector classifier (SVC) and support vector regressor (SVR), were built using the "Scikit–Learn" Python library[10]. The hyperparameters for these three algorithms were tuned by a Bayesian optimization selection process that was completed using "Scikit–Optimize"[11]. 3–fold cross validation was implemented during the hyperparameter tuning process. The performance of each tuned algorithm was evaluated on a separate test data set consisting of 25% of the available data.

A quantum variational circuit was also developed and trained upon the data using the "Qiskit" Python library[12]. This model consisted of 5 qubits, one for each of the input values, and one classical register for measuring the output value. Each qubit in the circuit was subjected to a rotational gate at an angle corresponding to the values of the input data[13]. In a manner analogous to the adjustment of weights in artificial neural network algorithms, the difference between the output measured at the classical register and the real output value in the training data was used to iteratively adjust the rotational gate angles via stochastic gradient descent over the course of multiple training epochs.

3. Results and Discussion

P–values were first calculated between each of the operating parameters and the binary response (see Figure 2a). These values were read by fitting the data set to a logistic regression function using the generalized linear model function in R and reading the output summary of the fitted model[9]. The calculated p–values for the zinc effusion cell temperature, oxygen gas flow rate, and substrate (growth) temperature variables all fall below 0.05, with the values corresponding to zinc effusion cell temperature (0.0064) and oxygen gas flow rate (0.0358) having the smallest magnitude.

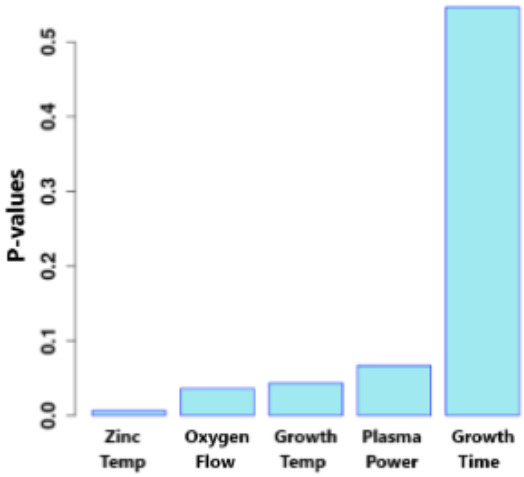

Fig 2. P–values calculated between operating parameters and the binary response. The zinc effusion cell temperature, oxygen gas flow, and substrate temperature variables all have p–values less than the 0.05 threshold.

Prior to training any of the models, the data set was normalized and then separated into a training set (75%) and test set (25%). During Bayesian hyperparameter tuning, the training data set was further partitioned into 3 subsets for cross validation. For the SVC, the values for the regularization parameter and kernel coefficient were optimized and a polynomial kernel of degree 4 was found to result in the optimal generalization performance. Similarly, for the SVR, the values for the regularization parameter and kernel coefficient were optimized and a polynomial kernel of degree 1 was found to result in the optimal generalization performance. The hyperparameters of the quantum variational circuit were not subjected to Bayesian optimization, but the learning rate of the optimizer used during training was manually optimized. Additionally, the rotation gates of the quantum variational circuit were adjusted according to a stochastic gradient descent optimization routine which involved a mean squared error (L2) loss function and a learning rate of 0.1. The circuit was evaluated 10^4 times for each data point during training in the interest of generating sufficient statistics.

For each algorithm, once the optimized parameters were selected, the performance of the algorithm was evaluated using the test data set to calculate an appropriate accuracy test score. The SVC and quantum variational circuit were both evaluated using the accuracy metric: the total number of correct classifications divided by the total number of predictions. The SVR was evaluated for generalization performance by calculating the mean squared error (MSE) of the predictions made on the test data set. To calculate the classification accuracy of the SVR, the values predicted by the trained regressor were threshold into either 0 (uneven) or 1 (flat), based upon whether the predicted value was below or above 0.5, respectively. The test accuracy scores for each trained algorithm are reported in Table 2. Between the two classical machine learning models, the SVR performed slightly worse on the test data set after training with an accuracy of 0.891 on the testing data. The SVC displayed improved generalization performance compared to

the SVR in the form of a higher test score, as shown in Table 2. The quantum variational circuit achieved a lower accuracy than all of the classical supervised learning algorithms; however, this is not unexpected since the performance of quantum machine learning techniques on classical data is particularly sensitive to the encoding strategy chosen to represent the data in a quantum mechanical state as well as the quantity and nature of the data that is available for training[13].

Table 2: Test accuracy scores of the optimized supervised learning models. Each algorithm was tested upon 25% of the available data after undergoing parameter tuning. The optimized SVC algorithm exhibits the best test performance of these algorithms while the trained quantum variational circuit implemented here displays the lowest performance on the testing data.

Algorithm	Test Score
Support Vector Classifier	0.904
Support Vector Regressor	0.891
Quantum Variational Circuit	0.581

The trained SVC algorithm displayed superior generalization performance on the data and was used to predict the morphology of an epitaxially–grown ZnO thin film crystal surface across a broad range of possible PAMBE synthesis parameters. The SVC was selected over the other algorithms to generate visualizations of predicted trends in the processing space, which are characteristic of the relationship being estimated by the machine learning algorithm, due to its superior accuracy on the testing data. The SVC was trained on the training data set and then used to predict the response variable across a broad range of possible combinations of oxygen gas flow rate and zinc effusion cell temperature, as shown in Figures 3. Other synthesis parameters are held at their median training values for the predictions which are shown in Table 1, except substrate temperature which is varied from 400 °C to 800 °C in 200 °C increments from Figures 3a to 3c.

With a substrate temperature of 400 °C, as shown in Figure 3a, a roughly parabolic area of the processing space (in blue) commencing with oxygen flow rates equal to 2.0 sccm is predicted to result in flat ZnO sample growth; this domain occurs for a plasma power setting of 400 W and a range of zinc effusion cell temperatures that broadens as the oxygen gas flow rate is increased. This region corresponding to a flat ZnO surface morphology is predicted to increase in area as the substrate temperature is increased to 600 °C and 800 °C, as seen in Figures 3b and 3c. This demonstrates the importance of growth temperature on the resulting surface morphology of epitaxially–grown ZnO.

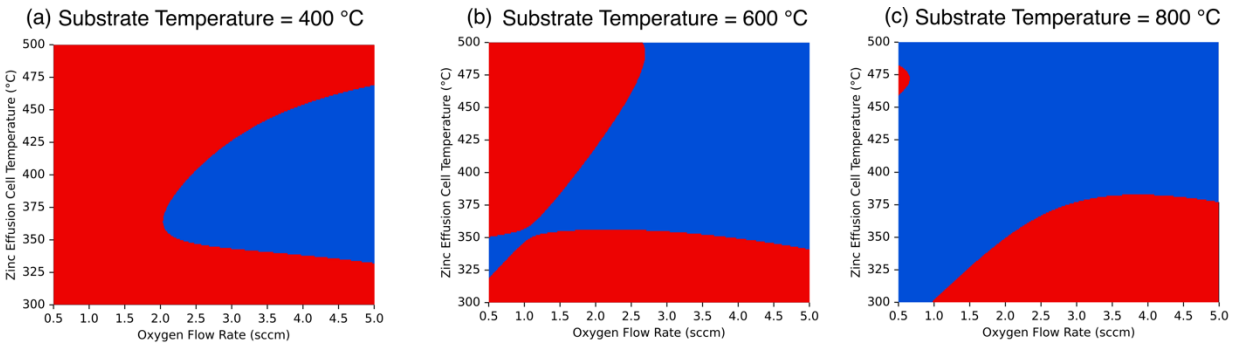

Fig 3. Predicted surface quality – flat (blue) or uneven (red) – for epitaxially–grown thin film ZnO crystals as predicted by a trained SVC algorithm. Predictions were made allowing zinc effusion cell temperature and oxygen gas flow rate to vary while holding the settings of plasma source power and growth duration constant at the median training data values listed in Table 1. (a) For a plasma source power setting of 400 W and a growth duration of 9.5 minutes, as listed in Table 1, a growth temperature of 400 °C is predicted to yield a roughly parabolic domain of the processing space corresponding to the growth of flat ZnO crystal surfaces. (b–c) As the substrate temperature is increased to 600 °C and 800 °C, the region of the processing space resulting in flat ZnO surface growth is forecasted to broaden.

4. Summary and Conclusion

294 data points describing the PAMBE operating parameters and resulting surface morphology for ZnO PAMBE synthesis experiments were organized together into a single tabular data set. P–values were calculated between operating parameter values and the binary descriptor of ZnO surface roughness as determined by RHEED patterns. The calculated p–values indicate that the values of zinc effusion cell temperature, oxygen gas flow rate, and substrate temperature are all statistically influential on the surface roughness of epitaxially–grown ZnO. Quantum and classical supervised learning models were tuned, trained, and compared for generalization performance. The combination of Bayesian hyperparameter selection and 3–fold cross validation produced a SVC algorithm displaying a test score superior to those of the other algorithms. The trained SVC was used to predict the surface morphology of PAMBE–grown thin film ZnO crystals across processing spaces defined by the most statistically significant growth parameters. Increasing the combination of substrate temperature, zinc effusion cell temperature, and oxygen gas flow rate is predicted to improve the surface morphology of PAMBE–grown crystals of ZnO. The resulting predictions offer new insight on the processing space for epitaxially–grown ZnO that can be used to inform the design of future experiments. Although the quantum variational circuit implemented in this study did not outperform the classical supervised learning models, this project granted the opportunity for doctoral research to expand into novel applications of both quantum and classical machine learning. It also broadened the range of materials under consideration in the doctoral work to include additional materials (oxide semiconductors) that are relevant for electronic devices.

Acknowledgements

The authors gratefully acknowledge 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).

Conflict of Interest

The authors have no conflicts of interest to disclose.

Data Availability Statement

Data can be made available upon reasonable request.

References

[1] 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. (2023). Machine-learning-assisted analysis of transition metal dichalcogenide thin-film growth. *Nano Convergence*, *10*(1). https://doi.org/10.1186/s40580-023-00359-5

[2] Costine, A., Delsa, P., Li, T., Reinke, P., & Balachandran, P. V. (2020). Data-driven assessment of chemical vapor deposition grown mos2 monolayer thin films. *Journal of Applied Physics*, *128*(23). https://doi.org/10.1063/5.0017507

[3] Provence, S. R., Thapa, S., Paudel, R., Truttmann, T. K., Prakash, A., Jalan, B., & Comes, R. B. (2020). Machine learning analysis of perovskite oxides grown by molecular beam epitaxy. *Physical Review Materials*, *4*(8). https://doi.org/10.1103/physrevmaterials.4.083807

[4] Wakabayashi, Y. K., Otsuka, T., Krockenberger, Y., Sawada, H., Taniyasu, Y., & Yamamoto, H. (2019). Machine-learning-assisted thin-film growth: Bayesian optimization in molecular beam epitaxy of $SrRO_3$ thin films. APL Materials, 7(10). https://doi.org/10.1063/1.5123019

[5] Shen, C., Zhan, W., Xin, K., Li, M., Sun, Z., Cong, H., . . . Wang, Z. (2023). *Machine-learning-assisted and real-time-feedback-controlled growth of InAs/GaAs quantum dots*. Ithaca: Cornell University Library, arXiv.org. http://arxiv.org/abs/2306.12898

[6] Ichimiya, A., & Cohen, P. I. (2004). *Reflection high-energy electron diffraction*. Cambridge University Press.

[7] Fortunato, E., Gonçalves, A., Pimentel, A., Barquinha, P., Gonçalves, G., Pereira, L., Ferreira, I., & Martins, R. (2009). Zinc oxide, a multifunctional material: From material to device applications. *Applied Physics A, 96*(1), 197–205. https://doi.org/10.1007/s00339-009-5086-5

[8] Van Rossum, G., & Drake Jr, F. L. (1995). *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

[9] Milborrow, S. (2022, May 21). *Package rpart.plot*. CRAN. https://cran.r-project.org/web/packages/rpart.plot/index.html

[10] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit–learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*(2825–2830). https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

[11] Head, T., Kumar, M., Nahrstaedt, H., Louppe, G., & Shcherbatyi, I. (2021). scikit-optimize/scikit-optimize (v0.9.0). Zenodo. https://doi.org/10.5281/zenodo.5565057

[12] Matthew Treinish. (2023). Qiskit/qiskit-metapackage: Qiskit 0.44.0 (0.44.0). Zenodo. https://doi.org/10.5281/zenodo.8190968

[13] Rath, M., & Date, H. (2023). *Quantum data encoding: A comparative analysis of classical-to-quantum mapping techniques and their impact on machine learning accuracy*. Ithaca: Cornell University Library, arXiv.org. https://doi.org/10.48550/arXiv.2311.10375