Deep Learning for Semiconductor Defect Classification

J. O'Leary (UC Berkeley), K. Sawlani (Lam Research Corp.), and A. Mesbah (UC Berkeley)

Several convolutional Neural Networks (CNN) are created to identify the composition of semiconductor defects based on a combination of SEM images of the defects and spectral data. The CNNs use image and spectral data to classify semiconductor defects with an industrially pragmatic accuracy.

INTRODUCTION

- Manual classification of semiconductor defects by process and productivity engineers can take hours or days, which leads to slow solutions and longer learning curves on product failures while still being prone to human error.
- Deep learning strategies can be used to reduce analysis time and inconsistencies due to human error, which in turn can result in systematic root cause analysis for sources of semiconductor defects.

METHODS

- An EDX spectral scan and top-view SEM image is provided with each semiconductor defect (Fig. 1):

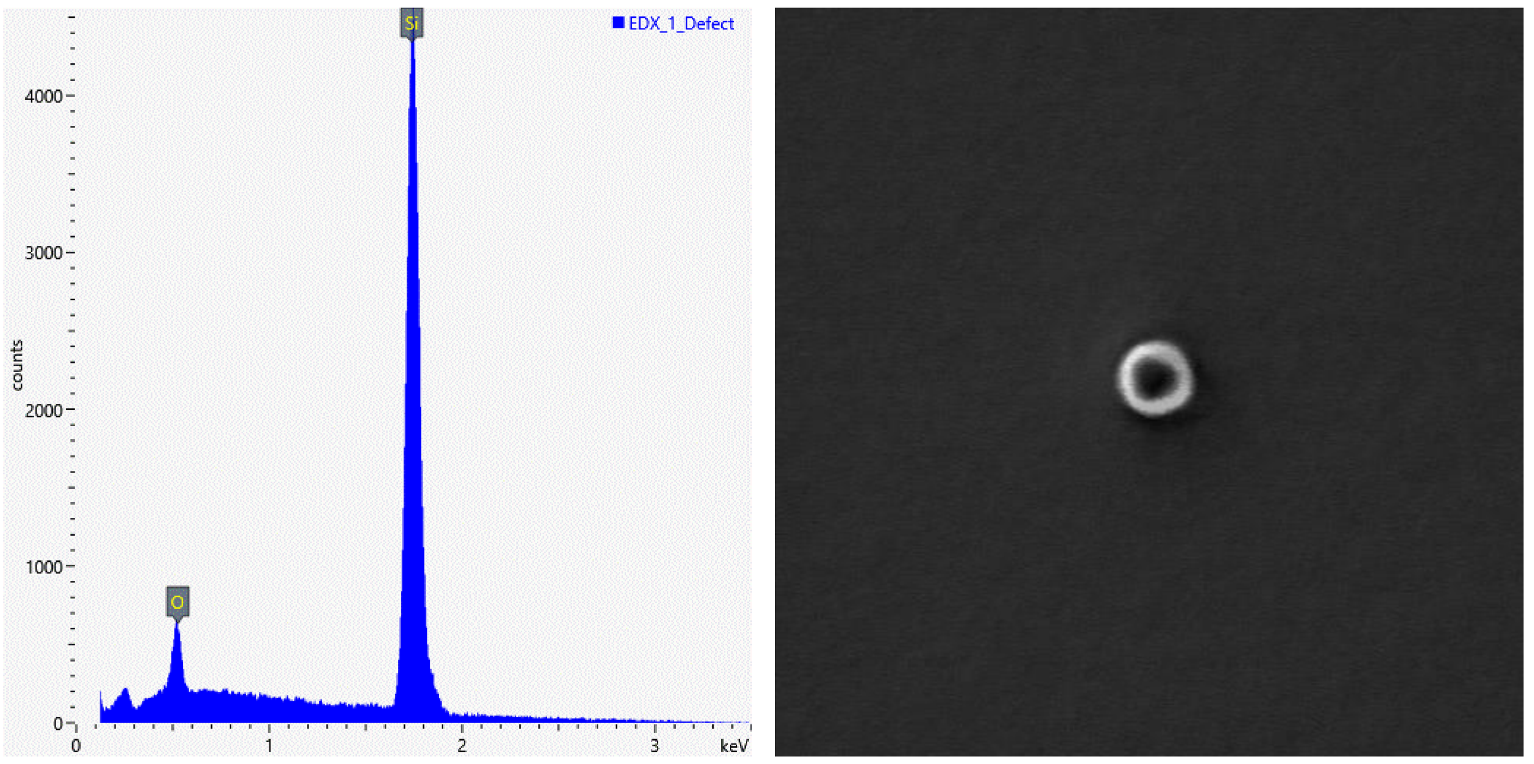

Fig. 1. **Example image and labeled spectra of SiO_2 semiconductor defect.**

- It is challenging to classify defects by *spectral data alone* because (Fig. 2):
  1. Certain defects are too small to have their peaks detected
  2. Peak overlap confounds defect classification
- It is challenging to classify defects by *image data alone* because many different defect types are similar in shape, orientation, and size

Objective: Investigate deep learning techniques that can leverage spectral and image data for defect classification

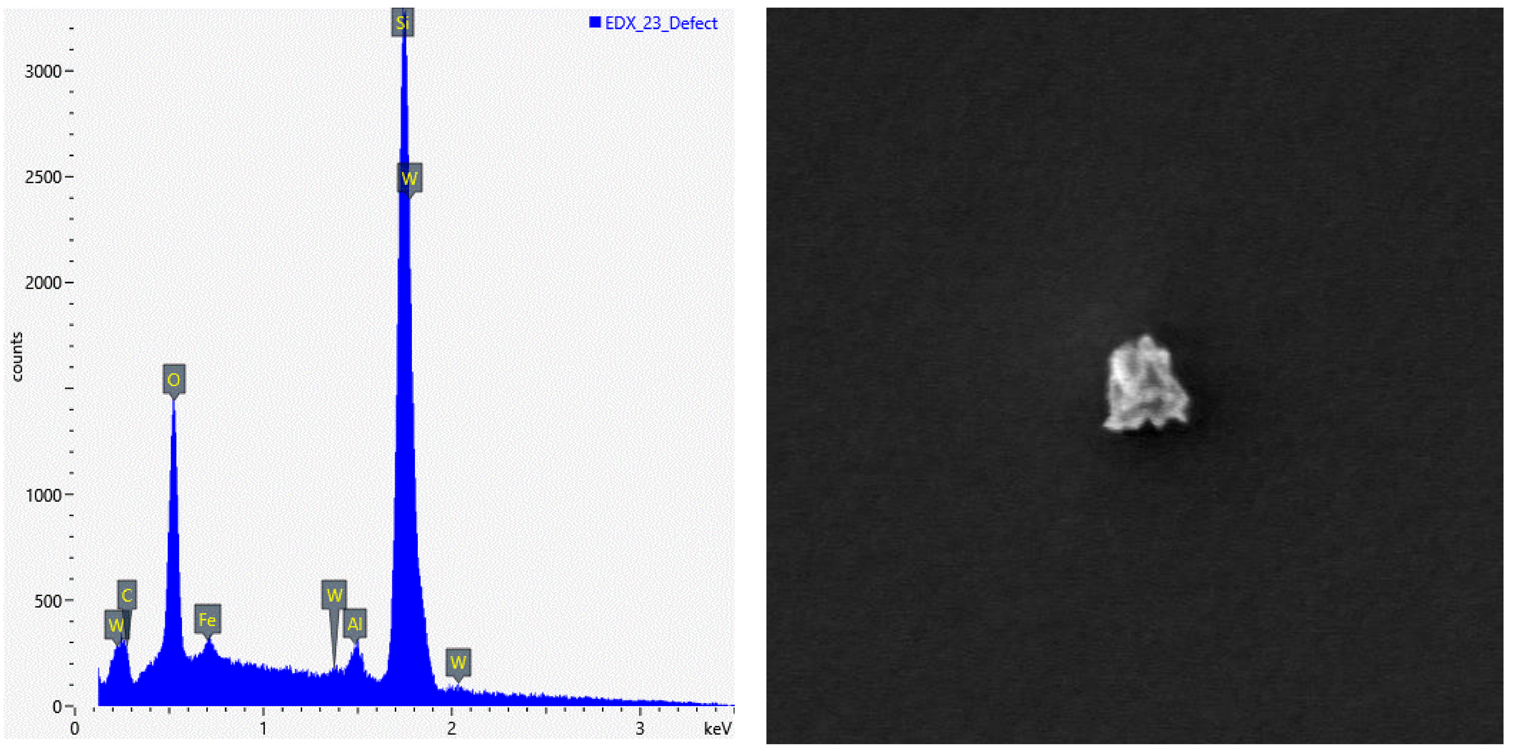

Fig. 2. **Example image and labeled spectra of $AlO_xF_yN_z$ semiconductor defect.** Note that Nitrogen and Fluorine peaks in the $AlO_xF_yN_z$ defect are either too small to detect or overlap with other peaks.

- CNNs are a class of feed-forward, artificial neural networks that are commonly used for image classification.
- Here, SEM images are fed into a CNN. Then, spectral data is merged directly with the fully connected layer. The spectral data includes two separate components:
  1. Raw intensity values for each defect (counts vs. keV)
  2. Automatically labeled peaks (some of which are incorrect due to peak overlap)

RESULTS AND DISCUSSION

- Several CNNs based on popular architectures (e.g., AlexNet, VGGNet, and ResNet) are trained using 5422 defects belonging to 13 unique classes. The same nets are also trained using a subset of 3925 defects belonging to defect types with >200 defects that comprise 7 different classes.
- Each defect is rotated 6 times to augment the size of the data set.
- The results for the best-performing CNN are reported in Table 1.

Table 1. **Summary of CNN defect classification accuracy.** Note that accuracies for CNNs without spectral data are included for reference.

Defect #	Class #	Images	Spectral Data	Top-1 Accuracy	Top-3 Accuracy
5422	13	Y	N	61%	88%
5422	13	Y	Y	67%	99.8%
3925	7	Y	N	63%	90%
3925	7	Y	Y	73%	100%

- Top one and top three defects can be classified with high accuracy
- Inconsistent labels can result in low classification accuracy for the top one defect (e.g., AlO_xF_y and $AlO_xF_yN_z$ are similar particles with different labels in Fig. 3b)

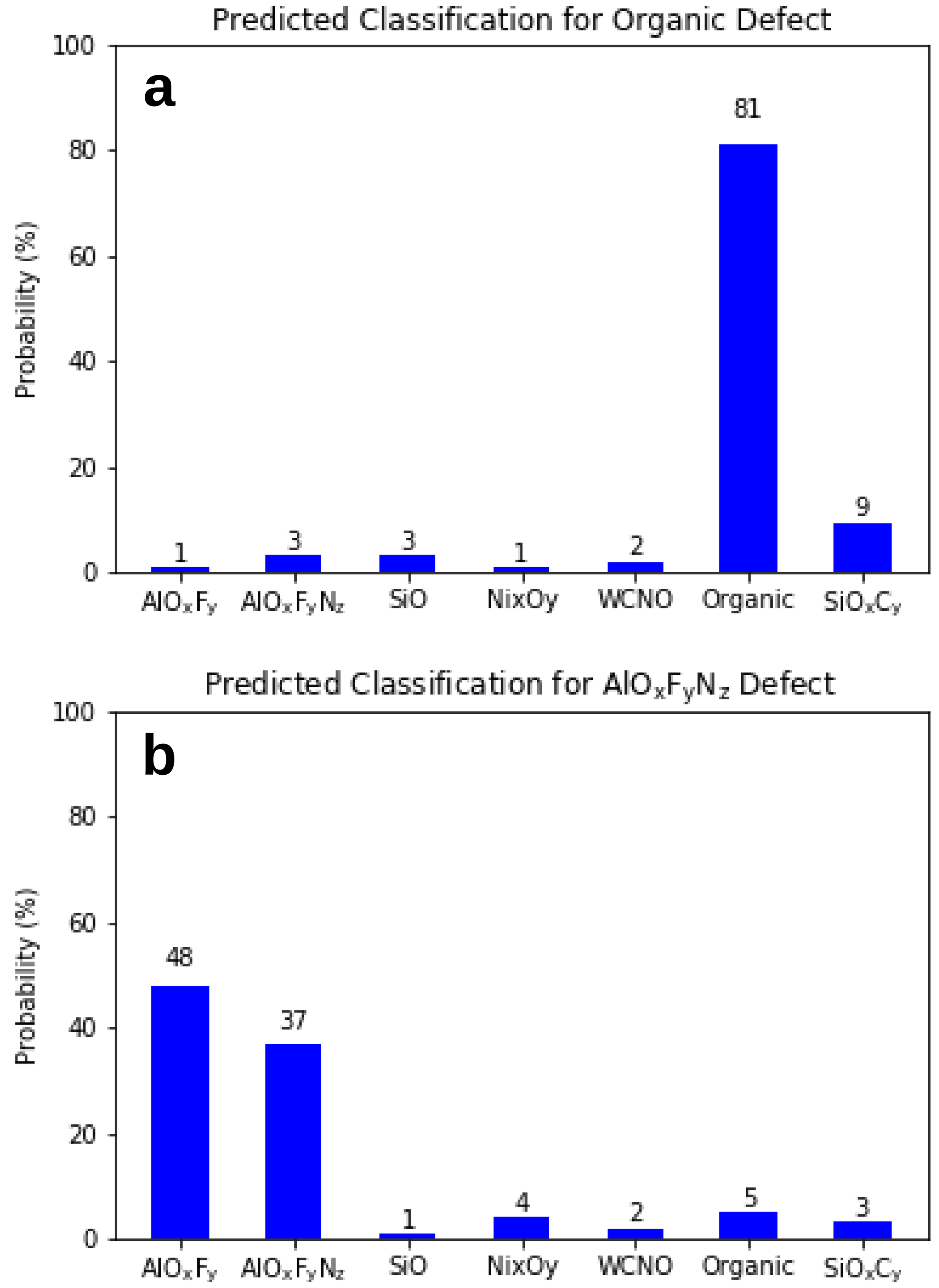

Fig. 3. **Predicted classification for a sample organic defect and a sample $AlO_xF_yN_z$ defect.** The CNN correctly identifies the organic defect, while it essentially narrows down the $AlO_xF_yN_z$ defect to one of two very similar choices. Note that these charts do not represent the defects shown in Figures 1 and 2.

FUTURE WORK

- Explore transfer learning methods for reducing computational cost, hyper-parameter optimization, and relative weight adjustment of different inputs (e.g., images, raw spectral data, labeled peaks)
- Identify minimum number of unique defects and minimum data augmentation requirement for additional defect classes to a model with minimum 95% Top-3 accuracy

Acknowledgement: This research was funded by Lam Research Corporation. **Contact:** mesbah@berkeley.edu

CNNs can use image and spectral data to classify semiconductor defects with an industrially pragmatic accuracy

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.