

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Accelerating Material Discovery and Analysis Using Machine Learning

### Permalink

<https://escholarship.org/uc/item/6jc281x0>

### Author

Kaufmann, Kevin Richard

### Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Accelerating Material Discovery and Analysis Using Machine Learning

A dissertation submitted in partial satisfaction of the requirements for the Doctor of Philosophy

in

NanoEngineering

by

Kevin Richard Kaufmann

Committee in charge:

Professor Kenneth Vecchio, Chair  
Professor Virginia de Sa  
Professor William Gerwick  
Professor Vlado Lubarda  
Professor Jian Luo

2021

©

Kevin Richard Kaufmann, 2021

All rights reserved

The dissertation of Kevin Richard Kaufmann is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

University of California San Diego

2021

iii

## **DEDICATION**

This dissertation is dedicated to my wife, Kelsie, who has stood by me through this journey. I know the path was arduous to me, and I cannot imagine it was any easier for you. Thank you for your unwavering love and support. May the next adventure involve a lot less time apart.

## EPIGRAPH

An incomplete collection of wise and motivating words

It does not matter where you start the human race.

*Wayne Neilson*

When wireless is perfectly applied the whole earth will be converted into a huge brain, which in fact it is, all things being particles of a real and rhythmic whole.

*Nikola Tesla*

Every morning in Africa, a gazelle wakes up. It knows it must run faster than the fastest lion or it will be killed. Every morning a lion wakes up. It knows it must outrun the slowest gazelle or it will starve to death. It does not matter whether you are a lion or a gazelle: when the sun comes up, you had better be running.

*Christopher McDougall*

You're braver than you believe, stronger than you seem, and smarter than you think.

*Christopher Robin*

A short pencil is better than a long memory.

*Unknown*

## TABLE OF CONTENTS

DISSERTATION APPROVAL PAGE .....	iii
DEDICATION .....	iv
EPIGRAPH.....	v
TABLE OF CONTENTS.....	vi
LIST OF FIGURES .....	x
LIST OF TABLES.....	xvi
ACKNOWLEDGEMENTS .....	xviii
VITA.....	xxi
ABSTRACT OF THE DISSERTATION .....	xxv
Chapter 1 Introduction .....	1
1.1 Motivation.....	1
1.2 Background and Objective.....	2
1.3 References.....	7
Chapter 2 Discovery of high-entropy ceramics via machine learning.....	12
Abstract.....	12
2.1 Introduction.....	13
2.2 Results.....	17
2.2.1 Model Performance.....	17
2.2.2 Feature importance.....	19
2.2.3 Experimental and computational validation .....	23
2.3 Discussion .....	30
2.4 Methods.....	31
2.4.1 Machine learning architecture.....	31
2.4.2 From chemistry to features .....	32
2.4.3 Interpreting the random forest algorithm.....	33
2.4.4 Sample preparation .....	33
2.4.5 Sample analysis.....	34
2.4.6 Calculation of the entropy-forming-ability .....	35
2.5 Acknowledgements.....	35
2.6 References.....	37
Chapter 3 Searching for high entropy alloys: A machine learning approach .....	43

Abstract .....	43
3.1 Introduction .....	44
3.2 Methods .....	47
3.2.1 Material selection .....	47
3.2.2 Model architecture .....	47
3.2.3 From chemistry to features .....	48
3.2.4 Evaluating the random forest algorithm .....	49
3.3 Results .....	49
3.3.1 Random forest model .....	49
3.3.2 Model performance .....	52
3.3.3 Gauging model certainty .....	53
3.4 Conclusions .....	56
3.5 Acknowledgements .....	58
3.6 References .....	59
Chapter 4 Development of ultrahigh-entropy ceramics with tailored oxidation behavior .....	64
Abstract .....	64
4.1 Introduction .....	65
4.2 Methods .....	68
4.2.1 Entropy forming ability .....	68
4.2.2 Predicting EFA .....	69
4.2.3 Sample preparation .....	69
4.2.4 X-ray diffraction and electron microscopy .....	70
4.2.4 Property testing .....	71
4.2.5 Ellingham diagram calculations .....	72
4.2.6 Oxidation Test .....	73
4.3 Results .....	73
4.3.1 Entropic contributions .....	73
4.3.2 EFA predictions .....	75
4.3.3 Structure determination and chemistry analysis .....	77
4.3.4 Material Properties .....	79
4.3.5 Ellingham diagram .....	80
4.3.6 Oxidation analysis .....	81
4.4 Discussion .....	89
4.5 Conclusion .....	91
4.6 Acknowledgements .....	91
4.7 References .....	93
Chapter 5 Crystal symmetry determination in electron diffraction using machine learning .....	99
Abstract .....	99
5.1 Main Text .....	100
5.2 Materials and Methods .....	110
5.2.1 Electron backscatter diffraction pattern collection .....	111
5.2.2 Neural network architecture and training procedure .....	111
5.2.3 Validation studies .....	113

5.2.4 The classification model .....	113
5.3 Acknowledgements .....	114
5.4 References .....	115
Chapter 6 Deep neural network enabled space group identification in EBSD .....	121
Abstract .....	121
6.1 Introduction .....	122
6.2 Materials and methods .....	126
6.2.1 Materials .....	126
6.2.2 Electron backscatter diffraction pattern collection .....	126
6.2.3 Neural network architecture .....	127
6.2.4 Neural network training .....	128
6.2.5 Diffraction pattern classification .....	128
6.2.6 Neural network interpretability .....	129
6.3 Results and discussion .....	129
6.4 Conclusion .....	143
6.5 Acknowledgements .....	144
6.6 References .....	146
Chapter 7 Phase Mapping in EBSD using Convolutional Neural Networks .....	151
Abstract .....	151
7.1 Introduction .....	152
7.2 Materials and methods .....	155
7.2.1 Electron backscatter diffraction pattern collection .....	155
7.2.2 Neural network architecture and phase mapping procedure .....	156
7.3 Results and discussion .....	159
7.4 Conclusion .....	170
7.5 Acknowledgements .....	171
7.6 References .....	173
Chapter 8 Efficient <i>few-shot</i> machine learning for classification of EBSD patterns .....	178
Abstract .....	178
8.1 Introduction .....	179
8.2 Results and Discussion .....	183
8.2.1 Training metrics .....	183
8.2.2 Performance on holdout data .....	185
8.2.3 Visualization .....	186
8.2.4 Feature importance .....	189
8.2.5 Performance in practice .....	191
8.3 Materials and methods .....	196
8.3.1 Materials .....	196
8.3.2 Electron backscatter diffraction pattern collection .....	197
8.3.3 Neural network architecture .....	198
8.3.4 Neural network training .....	198

8.3.5 Diffraction pattern classification.....	199
8.3.6 Neural network insight.....	200
8.4 Acknowledgements.....	201
8.5 References.....	202
Chapter 9 An Acquisition Parameter Study for Machine-Learning-Enabled Electron Backscatter Diffraction.....	210
Abstract.....	210
9.1 Introduction.....	211
9.2 Materials and methods .....	214
9.2.1 Materials .....	214
9.2.2 Electron backscatter diffraction pattern collection .....	215
9.2.3 Convolutional neural network.....	216
9.2.4 Diffraction pattern classification.....	217
9.3 Results.....	217
9.3.1 Equipment setup and Hough-based results .....	217
9.3.2 Effects on the EBSD patterns .....	220
9.3.3 Frame averaging.....	222
9.3.4 Detector tilt .....	225
9.3.5 Detector Distance.....	226
9.3.6 Accelerating voltage .....	227
9.3.7 Pattern resolution .....	228
9.4 Discussion.....	229
9.5 Conclusions.....	231
9.6 Acknowledgements.....	232
9.7 References.....	233
Appendix A Supplementary Information for Chapter 2 .....	239
Appendix B Supplementary Information for Chapter 3 .....	246
Appendix C Supplementary Information for Chapter 4 .....	328
Appendix D Supplementary Information for Chapter 5 .....	346
Appendix E Supplementary Information for Chapter 6.....	357
Appendix F Supplementary Information for Chapter 7 .....	372
Appendix G Supplementary Information for Chapter 8 .....	375
Appendix H Supplementary Information for Chapter 9 .....	382

## LIST OF FIGURES

Figure 2.1 Evaluation of the ML models fit to available data. ....	19
Figure 2.2 Correlation between EFA and top two features .....	23
Figure 2.3 Machine learning model compared to <i>ab-initio</i> results for high-entropy carbides. ....	26
Figure 2.4 The X-ray diffraction patterns for the same 7 five-metal carbides. ....	28
Figure 2.5 Microstructural analysis of the synthesized materials. ....	30
Figure 3.1 Example decision tree from the fitted model. ....	51
Figure 4.1 Modeling ideal configurational entropy. ....	75
Figure 4.2 X-ray diffraction patterns for the 6 to 9 cation metal carbides.....	78
Figure 4.3 Chemistry analysis of the 6-cation UHECs.....	79
Figure 4.4 Moduli comparison for carbides.....	80
Figure 4.5 Ellingham diagrams for each composition. ....	81
Figure 4.6 Overview of oxide layer formation. ....	83
Figure 4.7 Analysis of the oxidation interfaces. ....	85
Figure 4.8 X-ray diffraction data post-oxidation. ....	88
Figure 5.1 Illustration of the inner workings of a convolutional neural network. ....	104
Figure 5.2 Confusion matrix displaying the ResNet50 algorithm’s classification results.....	106
Figure 5.3 Visualizing the features utilized for classification of a diffraction pattern to face-centered cubic (FCC). ....	109
Figure 5.4 Comparison of phase mapping techniques. ....	110
Figure 6.1 Effect of atomic scattering factors on observable Kikuchi bands. ....	130
Figure 6.2 Plots of normalized classification accuracy after fitting the model with the high atomic number materials.....	132
Figure 6.3 Comparison of diffraction patterns from B2 FeAl (space group 221) and BCC Fe (space group 229). ....	133

Figure 6.4 Plots of normalized classification accuracy after fitting the model with a single low atomic number material. ....	134
Figure 6.5 Plots of normalized classification accuracy after fitting the model with a single low atomic number material. ....	135
Figure 6.6 Plots of normalized classification accuracy after fitting the model with a single low atomic number material. ....	136
Figure 6.7 Plots of normalized classification accuracy after fitting the model. ....	138
Figure 6.8 Comparison of feature detection with Hough-based EBSD and the trained CNN....	140
Figure 6.9 Dynamically simulated EBSPs.....	141
Figure 6.10 Plots of normalized classification accuracy after fitting the model with data from each material. ....	143
Figure 7.1 Phase mapping rutilated quartz based on Bravais lattices. ....	160
Figure 7.2 Phase mapping a meteorite. ....	162
Figure 7.3 Phase mapping a metal matrix composite. ....	163
Figure 7.4 Phase mapping an Fe-Al MIL composite.....	165
Figure 7.5 Phase mapping a thermal barrier coating. ....	168
Figure 7.6 Phase mapping 430 stainless steel. ....	170
Figure 8.1 Training and validation history statistics for the two models.....	185
Figure 8.2 Kernel pairs with lowest Euclidean distance.....	187
Figure 8.3 Selected feature maps from the first convolution layer.....	189
Figure 8.4 Visual explanation of feature contributions. ....	191
Figure 8.5 Phase mapping a dual-phase sample. ....	193
Figure 9.1 EBSD setup and variable parameters. ....	218
Figure 9.2 Hough-based EBSD results with different phase lists.....	219
Figure 9.3 Impact of operating conditions on the EBSD patterns. ....	221
Figure 9.4 Visual overview of frame averaging on CNN classification for duplex steel. ....	223
Figure 9.5 Effect of frame averaging on classification accuracy.....	224

Figure 9.6 Effect of detector tilt on classification accuracy. ....	225
Figure 9.7 Effect of sample-to-detector distance on classification accuracy.....	226
Figure 9.8 Effect of accelerating voltage on CNN classification. ....	228
Figure 9.9 Effect of pattern resolution on CNN classification. ....	229
Supplementary Figure 1 Cr-C binary phase diagram from ThermoCalc Software SSOL6 database version 6.1.....	240
Supplementary Figure 2 Mo-C binary phase diagram from ThermoCalc Software SSOL6 database version 6.1.....	241
Supplementary Figure 3 W-C binary phase diagram from ThermoCalc Software SSOL6 database version 6.1.....	242
Supplementary Figure 4 Example of a decision tree from the fitted model. ....	243
Supplementary Figure 5 Phase evolution diagrams for single and multi-phase compositions...	244
Supplementary Figure 6 Model performance as features are eliminated. ....	247
Supplementary Figure 7 Statistical analysis of RFECV over 100 runs.....	248
Supplementary Figure 8 Statistical analysis of elemental impact on predicted EFA. ....	329
Supplementary Figure 9 EDS maps for each experimental composition. ....	330
Supplementary Figure 10 EDS maps for each experimental composition. ....	331
Supplementary Figure 11 EDS maps for experimental compositions.....	332
Supplementary Figure 12 EDS maps for each experimental composition. ....	333
Supplementary Figure 13 Post-oxidation images of the samples. ....	334
Supplementary Figure 14 Analysis of the oxide edge and interface for HfMoNbTaZrC <sub>5</sub> . ....	335
Supplementary Figure 15 Analysis of the oxide edge and interface for HfMoNbTaZrC <sub>5</sub> . ....	336
Supplementary Figure 16 Analysis of the oxide edge and interface for HfMoNbTaZrC <sub>5</sub> . ....	337
Supplementary Figure 17 Analysis of the transition layers within oxidized HfMoNbTaVWZrC <sub>7</sub> . ....	338

Supplementary Figure 18 Analysis of the transition layers within oxidized HfMoNbTaVWZrC <sub>7</sub> . .....	339
Supplementary Figure 19 Analysis of the transition layers within oxidized HfMoNbTaVWZrC <sub>7</sub> . .....	340
Supplementary Figure 20 Confusion matrices displaying the classification results for the 14 Bravais lattices. ....	347
Supplementary Figure 21 Performance of the Xception machine learning model on nine different materials. ....	348
Supplementary Figure 22 Demonstration of machine learning aided EBSD's capability to autonomously identify space groups in new materials. ....	349
Supplementary Figure 23 Heatmaps elucidating the normalized importance of features (dark blue to dark red) in each diffraction pattern for classification of the correct Bravais lattice. ....	350
Supplementary Figure 24 Heatmaps demonstrating the classification of diopside. ....	351
Supplementary Figure 25 Diagram of the materials and their Bravais lattice. ....	351
Supplementary Figure 26 Annotated image of the experimental setup within the SEM chamber. .....	352
Supplementary Figure 27 An example diffraction pattern for each of the 28 materials used for training the machine learning model. ....	353
Supplementary Figure 28 Schematic describing the process of training a machine learning algorithm to recognize the symmetry features present in diffraction patterns. ....	354
Supplementary Figure 29 Crystal structure relationships from the seven crystal systems to the 230 space groups. ....	355
Supplementary Figure 30 Representative diffraction patterns from each material studied. ....	358
Supplementary Figure 31 Inverse pole figures of the entire dataset. ....	359
Supplementary Figure 32 Probability plots of mean angular deviation and band contrast for the entire dataset. ....	360
Supplementary Figure 33 Schematic of the neural network. ....	361
Supplementary Figure 34 Histograms of mean angular deviation and band contrast in the training set for the associated model. ....	362
Supplementary Figure 35 Inverse pole figures for each space group training set in the associated model. ....	363

Supplementary Figure 36 Inverse pole figures for patterns based on correct or incorrect classification. ....	363
Supplementary Figure 37 Histograms of mean angular deviation and band contrast separated by correct or incorrect classification. ....	364
Supplementary Figure 38 Feature comparison for correct and incorrect classifications. ....	364
Supplementary Figure 39 Representative diffraction patterns from each phase in the six materials studied. ....	373
Supplementary Figure 40 Schematic of the neural network. ....	374
Supplementary Figure 41 Histogram of band contrast values for the 430 stainless steel map. ....	374
Supplementary Figure 42 The categorical cross-entropy loss function. ....	376
Supplementary Figure 43 Shapley value analysis for handwritten digits. ....	376
Supplementary Figure 44 Visual explanation of feature contributions. ....	377
Supplementary Figure 45 Example EBSPs from the Ni <sub>90</sub> Al <sub>10</sub> sample. ....	378
Supplementary Figure 46 XRD Pattern for the Ni <sub>90</sub> Al <sub>10</sub> sample. ....	378
Supplementary Figure 47 Schematic of Neural Network Operating on an EBSP. ....	379
Supplementary Figure 48 Diffraction pattern with increasing frame averaging. ....	383
Supplementary Figure 49 Diffraction pattern with different detector tilt. ....	384
Supplementary Figure 50 Diffraction pattern with decreasing sample-to-detector distance. ....	385
Supplementary Figure 51 Diffraction pattern with variable accelerating voltage. ....	386
Supplementary Figure 52 A visual explanation of the observed changes based on the resolution of the EBSD detector. ....	386
Supplementary Figure 53 Visual overview of frame averaging on CNN classification for duplex steel. ....	387
Supplementary Figure 54 Visual overview of detector tilt on CNN classification for duplex steel. ....	388
Supplementary Figure 55 Visual overview of detector distance on CNN classification for duplex steel. ....	389
Supplementary Figure 56 Visual overview of accelerating voltage on CNN classification for duplex steel. ....	390

Supplementary Figure 57 Visual overview of pattern resolution on CNN classification for duplex steel. .... 391

## LIST OF TABLES

Table 2.1 Identification of the important features for predicting EFA values.....	21
Table 2.2 Results for the ML predicted EFA for seventy new compositions.....	25
Table 3.1 Ranking feature importance.....	52
Table 3.2 ML-HEA algorithm agreement with known data and other models. ....	55
Table 3.3 Quantifying uncertainty in the ML-HEA model.....	56
Table 4.1 Compositions selected for fabrication. ....	76
Table 4.2 Compositions for each phase in HfMoNbTaZrC <sub>5</sub> post-oxidation. ....	86
Table 4.3 Compositions for each phase in HfMoNbTaVWZrC <sub>7</sub> post-oxidation.....	87
Table 8.1 Classification metrics.....	186
Table 8.2 Tabulated predictions for the dual-phase sample. ....	195
Supplementary Table 1 Model performance on existing data. Units: EFA in (eV/atom) <sup>-1</sup> . ....	245
Supplementary Table 2 ML-HEA results for binary alloys.....	249
Supplementary Table 3 ML-HEA results for ternary alloys.....	251
Supplementary Table 4 ML-HEA results for quaternary alloys.....	257
Supplementary Table 5 ML-HEA results for quinary alloys.....	279
Supplementary Table 6 Percentage votes for each class by alloy. ....	281
Supplementary Table 7 Machine learning predicted EFA values for 6-9 cation compositions. ....	341
Supplementary Table 8 ICSD structure information for the matched phases. ....	345
Supplementary Table 9 The point groups and their respective lattices are listed with the symmetry elements that are present for the crystal structure.....	356
Supplementary Table 10 Material acquisition and processing. ....	365
Supplementary Table 11 Number of diffraction patterns classified to each space group. ....	366
Supplementary Table 12 Number of diffraction patterns classified to each space group. ....	367

Supplementary Table 13 Number of diffraction patterns classified to each space group. ....	368
Supplementary Table 14 Number of diffraction patterns classified to each space group. ....	369
Supplementary Table 15 Number of diffraction patterns classified to each space group. ....	370
Supplementary Table 16 Number of diffraction patterns classified to each space group. ....	371
Supplementary Table 17 Classification metrics by space group for the transfer learning model. .....	379
Supplementary Table 18 Classification metrics by space group for the model trained from scratch. .....	380
Supplementary Table 19 Comparison of the symmetry elements that describe each space group. .....	381
Supplementary Table 20 Pattern acquisition rates.....	392

## ACKNOWLEDGEMENTS

I would, first and foremost, like to acknowledge the support and mentorship of my advisor and committee chair, Dr. Kenneth Vecchio. It has been an honor to work with you and I appreciate your support and guidance of my research. I am also grateful for the advice and assistance from my other committee members, Dr. Jian Luo, Dr. Vlado Lubarda, Dr. Virginia de Sa, and Dr. William Gerwick for their part in advancing this research.

I would like to thank all my collaborators at UCSD, Dr. Chaoyi Zhu, Alexander S. Rosengarten, William Mellor, Dr. Tyler Harrington, Daniel Maryanovsky, Eduardo Marin, Haoren Wang, Hobson Lane, Dr. Mojtaba Samiee, Xiao Lui, Dr. Joshua Gild, Olivia Dippo, Matthew Quinn, and Lucas Borowski. Additionally, I would like to thank the rest of the lab members of the Vecchio group for their thoughtful discussion and assistance throughout the years.

I would like to acknowledge the financial support from the Department of Defense (DoD) for their support through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program. I would also like to thank the ARCS Foundation, San Diego Chapter for their support and commitment to supporting graduate students in pursuit of advancing science.

I would also like to acknowledge the staff of the NanoEngineering Materials Research Center: Wayne Neilson, Sabine Faulhaber, and Dr. Stephen Horvath. Each of you were a valuable resource, whether we were discussing science or life. And to Dana Jiminez, I do not know how the NanoEngineering graduate program will manage without you.

Last, but not least, I would like to thank my wife, parents, siblings, loved ones, and friends for their endless support through my twenty-five years of studies. To my parents, Bertram and

Diane, thank you for encouraging me to go on every adventure, especially this one. Oh, and for always picking me up on time.

Chapter 2, in full, is a reprint of the material “Discovery of high-entropy ceramics via machine learning” as it appears in *npj Computational Materials*. K. Kaufmann, D. Maryanovsky, W. Mellor, C. Zhu, A. S. Rosengarten, T. J. Harrington, C. Oses, C. Toher, S. Curtarolo, and K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

Chapter 3, in full, is a reprint of the material “Searching for High Entropy Alloys: A Machine Learning Approach” as it appears in *Acta Materialia*. K. Kaufmann & K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

Chapter 4, in full, is currently in submission as “Development of ultrahigh-entropy ceramics with tailored oxidation behavior”, W. M. Mellor<sup>†</sup>, K. Kaufmann<sup>†</sup>, O. F. Dippo, S. D. Figueroa, G. D. Schrader, and K. S. Vecchio. The dissertation author was the primary investigator and co-author of this material.

Chapter 5, in full, is a reprint of the material “Crystal symmetry determination in electron diffraction using machine learning”, as it appears in *Science*. K. Kaufmann, C. Zhu, A. S. Rosengarten, D. Maryanovsky, T. J. Harrington, E. Marin, and K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

Chapter 6, in full, is a reprint of the material “Deep Neural Network Enabled Space Group Identification in EBSD”, as it appears in *Microscopy and Microanalysis*. K. Kaufmann, C. Zhu, A. S. Rosengarten, A. S. Rosengarten, and K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

Chapter 7, in full, is a reprint of the material “Phase Mapping in EBSD Using Convolutional Neural Networks”, as it appears in *Microscopy and Microanalysis*. K. Kaufmann, C. Zhu, A. S. Rosengarten, D. Maryanovsky, H. Wang, and K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

Chapter 8, in full, is a reprint of the material “Efficient few-shot machine learning for classification of EBSD patterns”, as it appears in *Scientific Reports*. K. Kaufmann, H. Lane, X. Liu, & K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

Chapter 9, in full, is currently in submission as “An Acquisition Parameter Study for Machine-Learning-Enabled Electron Backscatter Diffraction”, K. Kaufmann & K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

## VITA

- 2015 Bachelor of Science, NanoEngineering, University of California San Diego
- 2016 Master of Science, NanoEngineering, University of California San Diego
- 2021 Doctor of Philosophy, NanoEngineering, University of California San Diego

## PUBLICATIONS

**Kaufmann, K.**, and Vecchio, K. S. “An Acquisition Parameter Study for Machine-Learning-Enabled Electron Backscatter Diffraction”. *In Review*.

Mellor, W. M. †, **Kaufmann, K.** †, Dippo, O. F., Figueroa, S. D., Schrader, G. D., & Vecchio, K. S. “Development of ultrahigh-entropy ceramics with tailored oxidation behavior”. *In Review*.

**Kaufmann, K.**, Lane, H., Liu, X., & Vecchio, K. S. (2021). Efficient few-shot machine learning for classification of EBSD patterns. *Scientific Reports*, 11 (1), 1-12.  
doi: [10.1038/s41598-021-87557-5](https://doi.org/10.1038/s41598-021-87557-5)

Dippo, O. F., **Kaufmann, K. R.**, & Vecchio, K. S. (2021). High-Throughput Rapid Experimental Alloy Development (HT-READ). *arXiv preprint arXiv:2102.06180*.

Sangiovanni, D. G., Mellor, W., Harrington, T., **Kaufmann, K.**, & Vecchio, K. S. (2021). Optimizing hardness and toughness in high-entropy refractory ceramics via tuning the valence electron concentration. *arXiv preprint arXiv:2102.02455*.

**Kaufmann, K.**, Zhu, C., Rosengarten, A. S., Maryanovsky, D., Harrington, T. J., Marin, E., & Vecchio, K. S. (2020). Crystal symmetry determination in electron diffraction using machine learning. *Science*, 367 (6477), 564-568.  
doi: [10.1126/science.aay3062](https://doi.org/10.1126/science.aay3062).

**Kaufmann, K.**, Maryanovsky, D., Mellor, W. M., Zhu, C., Rosengarten, A. S., Harrington, T. J., Oses, C., Toher, C., Curtarolo, S., & Vecchio, K. S. (2020). Discovery of high-entropy ceramics via machine learning. *Npj Computational Materials*, 6(1), 1-9.  
doi: [10.1038/s41524-020-0317-6](https://doi.org/10.1038/s41524-020-0317-6)

**Kaufmann, K.**, & Vecchio, K. S. (2020). Searching for high entropy alloys: A machine learning approach. *Acta Materialia*, 198, 178-222.  
doi: [10.1016/j.actamat.2020.07.065](https://doi.org/10.1016/j.actamat.2020.07.065)

**Kaufmann, K.**, Zhu, C., Rosengarten, A. S., & Vecchio, K. S. (2020). Deep neural network enabled space group identification in EBSD. *Microscopy and Microanalysis*, 26(3), 447-457.  
doi: [10.1017/S1431927620001506](https://doi.org/10.1017/S1431927620001506)

**Kaufmann, K.**, Zhu, C., Rosengarten, A. S., Maryanovsky, D., Wang, H., & Vecchio, K. S. (2020). Phase Mapping in EBSD Using Convolutional Neural Networks. *Microscopy and Microanalysis*, 26(3), 458-468.  
doi: [10.1017/S1431927620001488](https://doi.org/10.1017/S1431927620001488)

Zhu, C., **Kaufmann, K.**, & Vecchio, K. S. (2020). Novel remapping approach for HR-EBSD based on demons registration. *Ultramicroscopy*, 208, 112851.  
doi: [10.1016/j.ultramic.2019.112851](https://doi.org/10.1016/j.ultramic.2019.112851).

Zhang, C., Zhu, C., Cao, P., Wang, X., Ye, F., **Kaufmann, K.**, Casalena, L., MacDonald, B.E., Pan, X., Vecchio, K. & Lavernia, E. J. (2020). Aged metastable high-entropy alloys with heterogeneous lamella structure for superior strength-ductility synergy. *Acta Materialia*, 199, 602-612.  
doi: [10.1016/j.actamat.2020.08.043](https://doi.org/10.1016/j.actamat.2020.08.043)

Zhu, C., Wang, H., **Kaufmann, K.**, & Vecchio, K. S. (2020). A computer vision approach to study surface deformation of materials. *Measurement Science and Technology*, 31(5), 055602.  
doi: [10.1088/1361-6501/ab65d9](https://doi.org/10.1088/1361-6501/ab65d9).

Gild, J., **Kaufmann, K.**, Vecchio, K., & Luo, J. (2019). Reactive flash spark plasma sintering of high-entropy ultrahigh temperature ceramics. *Scripta Materialia*, 170, 106-110.  
doi: [10.1016/j.scriptamat.2019.05.039](https://doi.org/10.1016/j.scriptamat.2019.05.039).

Gild, J., Braun, J., **Kaufmann, K.**, Marin, E., Harrington, T., Hopkins, P., Vecchio, K., & Luo, J. (2019). A high-entropy silicide:(Mo<sub>0.2</sub>Nb<sub>0.2</sub>Ta<sub>0.2</sub>Ti<sub>0.2</sub>W<sub>0.2</sub>)Si<sub>2</sub>. *Journal of Materiomics*, 5(3), 337-343.  
doi: [10.1016/j.jmat.2019.03.002](https://doi.org/10.1016/j.jmat.2019.03.002).

**Kaufmann, K.**, Zhu, C., Rosengarten, A. S., Maryanovsky, D., Harrington, T., Marin, E., & Vecchio, K. (2019). High-Throughput Identification of Crystal Structures Via Machine Learning. *Microscopy and Microanalysis*, 25(S2), 2258-2259.  
doi: [10.1017/S1431927619012029](https://doi.org/10.1017/S1431927619012029).

Zhu, C., **Kaufmann, K.**, & Vecchio, K. (2019). Automated Reconstruction of Spherical Kikuchi Maps. *Microscopy and Microanalysis*, 25(4), 912-923.  
doi: [10.1017/S1431927619000710t](https://doi.org/10.1017/S1431927619000710t).

Harrington, T. J., Gild, J., Sarker, P., Toher, C., Rost, C. M., Dippo, O. F., McElfresh, C., **Kaufmann, K.**, Marin, E., Borowski, L., Hopkins, P. E., Luo, J., Curtarolo, S., Brenner, D. W., & Vecchio, K. S. (2019). Phase stability and mechanical properties of novel high entropy transition metal carbides. *Acta Materialia*, 166, 271-280.  
doi: [10.1016/j.actamat.2018.12.054](https://doi.org/10.1016/j.actamat.2018.12.054).

Singh, V. V., **Kaufmann, K.**, de Ávila, B. E. F., Karshalev, E., & Wang, J. (2016). Molybdenum Disulfide-Based Tubular Microengines: Toward Biomedical Applications. *Advanced Functional Materials*, 26(34), 6270-6278.  
doi: [10.1002/adfm.201602005](https://doi.org/10.1002/adfm.201602005).

Esteban-Fernández de Ávila, B., Lopez-Ramirez, M. A., Báez, D. F., Jodra, A., Singh, V. V., **Kaufmann, K.**, & Wang, J. (2016). Aptamer-modified graphene-based catalytic micromotors: Off-on fluorescent detection of ricin. *ACS Sensors*, 1(3), 217-221.  
doi: [10.1021/acssensors.5b00300](https://doi.org/10.1021/acssensors.5b00300).

Singh, V. V., **Kaufmann, K.**, de Ávila, B. E. F., Uygun, M., & Wang, J. (2016). Nanomotors responsive to nerve-agent vapor plumes. *Chemical communications*, 52(16), 3360-3363.  
doi: [10.1039/C5CC10670B](https://doi.org/10.1039/C5CC10670B).

Singh, V. V., Martin, A., **Kaufmann, K.**, DS de Oliveira, S., & Wang, J. (2015). Zirconia/graphene oxide hybrid micromotors for selective capture of nerve agents. *Chemistry of Materials*, 27(23), 8162-8169.  
doi: [10.1021/acs.chemmater.5b03960](https://doi.org/10.1021/acs.chemmater.5b03960).

Uygun, M., Singh, V. V., **Kaufmann, K.**, Uygun, D. A., de Oliveira, S. D., & Wang, J. (2015). Micromotor-Based Biomimetic Carbon Dioxide Sequestration: Towards Mobile Microscrubbers. *Angewandte Chemie International Edition*, 54(44), 12900-12904.  
doi: [10.1002/anie.201505155](https://doi.org/10.1002/anie.201505155).

Singh, V. V., Soto, F., **Kaufmann, K.**, & Wang, J. (2015). Micromotor-Based Energy Generation. *Angewandte Chemie International Edition*, 54(23), 6896-6899.  
doi: [10.1002/anie.201501971](https://doi.org/10.1002/anie.201501971).

Singh, V. V., **Kaufmann, K.**, Orozco, J., Li, J., Galarnyk, M., Arya, G., & Wang, J. (2015). Micromotor-based on-off fluorescence detection of sarin and soman simulants. *Chemical Communications*, 51(56), 11190-11193.  
doi: [10.1039/C5CC04120A](https://doi.org/10.1039/C5CC04120A).

Li, J., Singh, V. V., Sattayasamitsathit, S., Orozco, J., **Kaufmann, K.**, Dong, R., ... & Wang, J. (2014). Water-driven micromotors for rapid photocatalytic degradation of biological and chemical warfare agents. *ACS nano*, 8(11), 11118-11125.  
doi: [10.1021/nn505029k](https://doi.org/10.1021/nn505029k).

Sattayasamitsathit, S., Kou, H., Gao, W., Thavarajah, W., **Kaufmann, K.**, Zhang, L., & Wang, J. (2014). Fully loaded micromotors for combinatorial delivery and autonomous release of cargoes. *Small*, 10(14), 2830-2833.  
doi: [10.1002/smll.201303646](https://doi.org/10.1002/smll.201303646).

Sattayasamitsathit, S., **Kaufmann, K.**, Galarnyk, M., Vazquez-Duhalt, R., & Wang, J. (2014). Dual-enzyme natural motors incorporating decontamination and propulsion capabilities. *RSC Advances*, 4(52), 27565-27570.  
doi: [10.1039/C4RA04341C](https://doi.org/10.1039/C4RA04341C).

Gu, Y., Sattayasamitsathit, S., Jia, W., **Kaufmann, K.**, Wang, C., & Wang, J. (2013). High-Power Low-Cost Tissue-Based Biofuel Cell. *Electroanalysis*, 25(4), 838-844.  
doi: [10.1002/elan.201200568](https://doi.org/10.1002/elan.201200568).

Forbes, L. M., Sattayasamitsathit, S., Xu, P. F., O'Mahony, A., Samek, I. A., **Kaufmann, K.**, Wang, J., & Cha, J. N. (2013). Improved oxygen reduction reaction activities with amino acid R group functionalized PEG at platinum surfaces. *Journal of Materials Chemistry A*, 1(35), 10267-10273.  
doi: [10.1039/C3TA12133J](https://doi.org/10.1039/C3TA12133J).

Sattayasamitsathit, S., Gu, Y., **Kaufmann, K.**, Minter, S., Polsky, R., & Wang, J. (2013). Tunable hierarchical macro/mesoporous gold microwires fabricated by dual-templating and dealloying processes. *Nanoscale*, 5(17), 7849-7854.  
doi: [10.1039/C3NR02940A](https://doi.org/10.1039/C3NR02940A).

Gu, Y., Sattayasamitsathit, S., **Kaufmann, K.**, Vazquez-Duhalt, R., Gao, W., Wang, C., & Wang, J. (2013). Self-propelled chemically-powered plant-tissue biomotors. *Chemical Communications*, 49(66), 7307-7309.  
doi: [10.1039/C3CC42782J](https://doi.org/10.1039/C3CC42782J).

Sattayasamitsathit, S., Gu, Y., **Kaufmann, K.**, Jia, W., Xiao, X., Rodriguez, M., & Polsky, R. (2013). Highly ordered multilayered 3D graphene decorated with metal nanoparticles. *Journal of Materials Chemistry A*, 1(5), 1639-1645.  
doi: [10.1039/C2TA00954D](https://doi.org/10.1039/C2TA00954D).

## **ABSTRACT OF THE DISSERTATION**

Accelerating Material Discovery and Analysis Using Machine Learning

by

Kevin Richard Kaufmann

Doctor of Philosophy in NanoEngineering

University of California San Diego, 2021

Professor Kenneth Vecchio, Chair

The big data revolution is only just beginning in the materials science and engineering field, offering the promise to enable high-throughput workflows and accelerate material development. For this to be realized, a new set of tools capable of using this data for identifying better material candidates and assisting in the analysis of samples must be developed. Currently, most material development projects require manually searching vast composition space while relying primarily

on expert domain knowledge. While this strategy has been reasonably effective throughout history, emerging technologies are placing rigorous demands on material performance. Furthermore, many of the simplest combinations of elements (i.e. typically one primary element and one to three minor alloying elements) have been thoroughly evaluated. Within the last two decades, the materials science community has become interested in high entropy materials, typically containing five or more cations in near-equimolar amounts. This compositional space is largely unexplored and challenging to model with traditional tools, making it an excellent use case for machine learning. In the first part of the dissertation, machine learning tools are developed and demonstrated for predicting the relative synthesizability of high entropy and ultrahigh ceramics as well as the prediction of crystal structure for alloys. These machine learning approaches were validated by experiments and comparison with computational predictions where possible. The second part of the dissertation details progress in accelerating post-fabrication aspects of material design frameworks. For material analysis routines, deep neural networks were constructed to analyze diffraction patterns and determine symmetry and/or structure of the phases present. The electron backscatter diffraction (EBSD) platform was used to demonstrate the ability for such algorithms to not only identify symmetry, but that neural networks are also a potential solution to some of the grand challenges state of the art EBSD software cannot solve. This includes space group identification and the phase mapping of materials with similar crystal structures. Further validation is performed to demonstrate the model's reliability with changing acquisition parameters. Combined, these machine learning-based tools represent an opportunity to reduce the time and expertise required for material development and are likely to become valuable components in high-throughput workflow.

## Chapter 1 Introduction

### 1.1 Motivation

Material discovery plays a vital role in the technological progress of human civilization. Therefore, independent of the industry or potential applications (e.g. medicine, aerospace, or energy), the demand for materials with improved performance and better tradeoffs will likely never cease. In modern material design and analysis frameworks, there are several interrelated challenges when it comes to designing, fabricating, and analyzing a new material with targeted properties. The first is the design and analysis of materials relies primarily on the domain knowledge of experts in their field, sometimes coupled with theoretical models. In the same vein, the amount of data available is often limited to tens or hundreds of points, particularly for inorganic materials. The second challenge is the amount of data collected for each trial that requires analysis by a researcher with expertise in the material field and analytical tool. If the experiment is unsuccessful, it must be determined what caused the material not to meet the target and what changes could yield improvements if the next iteration is to move closer to the goal. This means that restrictive limitations on the search space are typically necessary to keep the dimensionality of the problem tractable and that the amount of information that is truly gleaned from each successive experiment can be limited by the available methods for finding patterns in the data. Lastly, there is the challenge of the overwhelming search space where the best answers may be found. Despite continuous effort since the bronze age, only a small fraction of the possible candidates have been systematically evaluated. However, many of the simplest combinations of inorganic materials (i.e., minor additions of alloying elements to a base element) have been studied and the field has turned to more complex combinations of the elements where there is little existing

data and fewer models. Clearly, a new set of tools are needed that will accelerate the discovery and characterization of the novel materials needed to solve current and future challenges.

## **1.2 Background and Objective**

The current material design cycle is highly inefficient with respect to the number of human-driven steps that are critical for success. The process typically begins when an experienced metallurgist considers the opportunity to improve an existing alloy or develop a new one. Data about similar materials is gathered and computational tools, such as thermodynamic software or first principles calculations, are employed. Potential candidates are typically fabricated and evaluated in a singular fashion with several different characterization techniques often required to develop the most complete understanding of the processing-microstructure-properties relationship. At this stage, the metallurgist uses the collected data to evaluate why the candidate material exhibits this combination of properties and how they can be further tuned in successive trials. This cycle can be repeated hundreds or thousands of times as the complex, multi-element (typically only 3-4 elements) space is experimentally evaluated in search of a material that meets the design goals.

Research into the development of methods and tools to improve upon the slow and tedious nature of this process have been of considerable interest since the announcement of the Materials Genome Initiative in 2011. The federal initiative was launched with the goal of discovering, manufacturing, and deploying advanced materials faster and with significantly less capital expenditure [1]. Early research stemming from this initiative was focused on the development of highly-accessible database infrastructure such as Automatic FLOW (AFLOW) [2], the Open Quantum Materials Database (OQMD) [3], and the Materials Project [4]. Each of these platforms focus primarily on the reliable automation of first principles calculations for inorganic materials

on a large-scale and making the results available to the community. As the databases grew, the potential for leveraging data and algorithms (i.e. materials informatics) became apparent with successful undertakings including the design of lithium-ion battery cathodes [5], predicting new auxetic materials [6], or searching for topological insulators [7]. While these platforms and the materials screening methods they offered significantly reduced the reliance on trial-and-error strategies, first principles calculations remain costly and time-consuming owing to the need for significant access to a supercomputer. Furthermore, these computational materials discovery tools require information about the materials of interest, such as atomistic structural information, which is commonly one of the unknowns in materials discovery. Additionally, first principles calculation of material properties can only be applied to single-phase materials while multi-phase materials can be of interest for the combination of properties they offer [8]. As a result, the compositional and structural search space that can be reasonably handled by these techniques remains limited.

With data generation rates accelerating and increasing access to data, interest in applying techniques from the field of data science began to emerge. The application of these tools to material design and discovery will be discussed first, followed by their utilization in material characterization. In 2014, Meredig *et al.* were one of the first to suggest data-driven design of materials using computer models trained to find patterns in materials datasets and apply the information to identify promising candidates in unconstrained composition space [9]. The key components employed in such works are a clean, well-labeled dataset, a feature engineering strategy for numerically representing the material composition to a computer, an educated choice of learner, and some strategy for model optimization [10,11]. Of these, creating meaningful numerical descriptors has proven to be the most challenging owing to the fact that traditional representations of crystal structures, i.e. translational vectors and fractional coordinates, are not

invariant [11]. Thus far, feature engineering strategies have primarily relied on statistical and rule-of-mixtures style weighting of a handful of general attributes for the atoms that make up the material [10,12,13]. In some cases, more precise features based on domain knowledge are applied; the use of valence electron concentration, Pauling electronegativity, or mixing entropy for the prediction of single-phase high entropy materials is one such example [14]. Using these strategies, data-driven design of multiple classes of materials have been demonstrated [11,15,16], including bulk metallic glasses [12], magnetic Heusler compounds [17], shape memory alloys [18], and high entropy materials [14,19,20]. In these works, the proposed materials presented to the trained machine learning algorithm were ranked on their likelihood of meeting the target goals. The decision of which proposed materials to test experimentally is frequently based on a combination of the algorithm's rankings, researchers' domain knowledge, insight into how the model made its predictions, and other experimental and calculated data. While the initial predictions from the model may not yield novel materials that meet the design goal [21] or every prediction may not be correct [18], machine learning models have the advantage of being readily updated to incorporate the new results and can be applied in a rapid, iterative fashion that adapts at the pace of the research. The rate at which machine learning models can be updated and re-assess materials compared to more traditional modeling efforts has become of significant interest. Recently, there have been several proposals and a handful of research efforts to develop automated experimental workflows that leverage machine learning models to quickly determine which experiments to run and apply the results to make better decisions in successive iterations [22–25].

An important component of these proposed frameworks will be the ability to rapidly characterize the resulting materials. In fact, the correlation between structure and properties is one of the primary relationships of interest in the development or application of any material: proteins

[26,27], molecules [28–30], geological materials [31,32], and inorganic materials [33–38]. Herein, the focus will be on the development of tools applied primarily to inorganic materials. In 2015, DeCost *et al.* hypothesized that digital image techniques would eventually be able to segment, characterize, and compare microstructures [39]. Applying techniques from the field of computer vision, a subset of artificial intelligence concerned with teaching computers to understand and potentially respond to information from the visual world, image-based tasks in materials science including classification [40–45], segmentation [46,47], and other analyses [48–50] have since been accomplished. Of recent interest has been the application of convolutional neural networks (CNNs), a class of deep learning models designed for processing data that comes in multiple arrays (e.g. color images composed of three 2D arrays) [51]. CNNs use a training dataset to learn to assign importance to features that maximize classification accuracy. This importance is learned autonomously by tuning feature detectors (i.e. filters or weights), which is accomplished by computing the gradient of the loss function with respect to all weights and a given input-output example (i.e. backpropagation), an efficient approach to minimizing the difference between the neural network’s predictions and the ground truth [52]. These filters operate in groups called filter banks to form feature maps containing distinctive local motifs found somewhere within the input image [51]. Efforts into visualizing the knowledge learned by these deep neural networks demonstrates that they are indeed capable of learning similar motifs to what would be described by a human for a given output category [53–55]. In materials science, these models can accelerate characterization efforts for multiple techniques such as optical and electron microscopy [39–43,45,47,48,56]. Further development of these tools can assist with and eventually even automate steps within high-throughput frameworks.

The objective of this work is to expand the research and development and application of data-driven strategies toward materials design and analysis. Throughout the development of these algorithms, we seek to understand how and why the models make their respective predictions. Within material discovery, this means decoding both the algorithm's decisions and the fundamental physics that are dictating material synthesizability and properties. This can reveal previously unknown relationships and spark new insight for solving future material challenges. The materials characterization work done herein is applied to improve the analysis of electron diffraction data. This relies on deep neural networks that are commonly referred to as 'black box' models, which by definition humans cannot fully assess; however, techniques from the burgeoning field of explainable AI are applied to validate the classification performance of these models. Additionally, for both materials design and analysis, comparisons of the developed approaches compared to state-of-the-art models and software will be discussed. By applying these tools in experimental studies, we aim to demonstrate their potential as components of high-throughput frameworks.

### 1.3 References

- [1] National Science and Technology Council (US), Materials genome initiative for global competitiveness, Exec. Off. Pres. Natl. Sci. Technol. Counc. (2011).
- [2] S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R. V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, D. Morgan, AFLOW: An automatic framework for high-throughput materials discovery, *Comput. Mater. Sci.* 58 (2012) 218–226. doi:10.1016/J.COMMATSCI.2012.02.005.
- [3] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD), *JOM.* 65 (2013) 1501–1509. doi:10.1007/s11837-013-0755-4.
- [4] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.* 1 (2013) 011002. doi:10.1063/1.4812323.
- [5] G. Hautier, A. Jain, S.P. Ong, B. Kang, C. Moore, R. Doe, G. Ceder, Phosphates as lithium-ion battery cathodes: An evaluation based on high-throughput ab initio calculations, *Chem. Mater.* 23 (2011) 3495–3508. doi:10.1021/cm200949v.
- [6] J. Dagdelen, J. Montoya, M. De Jong, K. Persson, Computational prediction of new auxetic materials, *Nat. Commun.* 8 (2017) 1–8. doi:10.1038/s41467-017-00399-6.
- [7] K. Yang, W. Setyawan, S. Wang, M. Buongiorno Nardelli, S. Curtarolo, A search model for topological insulators with high-throughput robustness descriptors, *Nat. Mater.* 11 (2012) 614–619. doi:10.1038/nmat3332.
- [8] D.C. Hofmann, J.-Y. Suh, A. Wiest, G. Duan, M.-L. Lind, M.D. Demetriou, W.L. Johnson, Designing metallic glass matrix composites with high toughness and tensile ductility, *Nature.* 451 (2008) 1085–1089. doi:10.1038/nature06598.
- [9] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, *Phys. Rev. B - Condens. Matter Mater. Phys.* 89 (2014) 1–7. doi:10.1103/PhysRevB.89.094104.
- [10] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: Recent applications and prospects, *Npj Comput. Mater.* 3 (2017). doi:10.1038/s41524-017-0056-5.
- [11] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, *Nature.* 559 (2018) 547–555. doi:10.1038/s41586-018-0337-2.
- [12] F. Ren, L. Ward, T. Williams, K.J. Laws, C. Wolverton, J. Hattrick-Simpers, A. Mehta, Accelerated discovery of metallic glasses through iteration of machine learning and high-

- throughput experiments, *Sci. Adv.* 4 (2018) eaaq1566. doi:10.1126/sciadv.aaq1566.
- [13] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, *Npj Comput. Mater.* 2 (2016) 16028. doi:10.1038/npjcompumats.2016.28.
- [14] W. Huang, P. Martin, H.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, *Acta Mater.* 169 (2019) 225–236. doi:10.1016/j.actamat.2019.03.012.
- [15] A. Agrawal, A. Choudhary, Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science, *APL Mater.* 4 (2016). doi:10.1063/1.4946894.
- [16] J. Schmidt, M.R.G. Marques, S. Botti, M.A.L. Marques, Recent advances and applications of machine learning in solid-state materials science, *Npj Comput. Mater.* 5 (2019) 1–36. doi:10.1038/s41524-019-0221-0.
- [17] S. Sanvito, C. Oses, J. Xue, A. Tiwari, M. Zic, T. Archer, P. Tozman, M. Venkatesan, M. Coey, S. Curtarolo, Accelerated discovery of new magnets in the Heusler alloy family, *Sci. Adv.* 3 (2017). doi:10.1126/sciadv.1602241.
- [18] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, Accelerated search for materials with targeted properties by adaptive design, *Nat. Commun.* 7 (2016) 11241. doi:10.1038/ncomms11241.
- [19] K. Kaufmann, D. Maryanovsky, W.M. Mellor, C. Zhu, A.S. Rosengarten, T.J. Harrington, C. Oses, C. Toher, S. Curtarolo, K.S. Vecchio, Discovery of high-entropy ceramics via machine learning, *Npj Comput. Mater.* 6 (2020) 42. doi:10.1038/s41524-020-0317-6.
- [20] K. Kaufmann, K.S. Vecchio, Searching for high entropy alloys: A machine learning approach, *Acta Mater.* 198 (2020) 178–222. doi:10.1016/j.actamat.2020.07.065.
- [21] E. Antono, N.N. Matsuzawa, J. Ling, J.E. Saal, H. Arai, M. Sasago, E. Fujii, Machine-Learning Guided Quantum Chemical and Molecular Dynamics Calculations to Design Novel Hole-Conducting Organic Materials, *J. Phys. Chem. A.* 124 (2020) 8330–8340. doi:10.1021/acs.jpca.0c05769.
- [22] O.F. Dippo, K.R. Kaufmann, K.S. Vecchio, High-Throughput Rapid Experimental Alloy Development (HT-READ), (2021). <http://arxiv.org/abs/2102.06180> (accessed April 22, 2021).
- [23] R.C. Kurchin, E. Muckley, L. Kavalsky, V. Hegde, D. Gandhi, X. Sun, M. Johnson, A. Edelman, J. Saal, C.V. Rackauckas, B. Meredig, V. Shah, V. Viswanathan, ACED: Accelerated Computational Electrochemical systems Discovery, *ArXiv.* (2020). <http://arxiv.org/abs/2011.04426> (accessed April 22, 2021).
- [24] E.M. Chan, C. Xu, A.W. Mao, G. Han, J.S. Owen, B.E. Cohen, D.J. Milliron, Reproducible, high-throughput synthesis of colloidal nanocrystals for optimization in multidimensional parameter space, *Nano Lett.* 10 (2010) 1874–1885. doi:10.1021/nl100669s.

- [25] D.P. Tabor, L.M. Roch, S.K. Saikin, C. Kreisbeck, D. Sheberla, J.H. Montoya, S. Dwaraknath, M. Aykol, C. Ortiz, H. Tribukait, C. Amador-Bedolla, C.J. Brabec, B. Maruyama, K.A. Persson, A. Aspuru-Guzik, Accelerating the discovery of materials for clean energy in the era of smart automation, *Nat. Rev. Mater.* 3 (2018) 5–20. doi:10.1038/s41578-018-0005-z.
- [26] T. Walz, T. Hirai, K. Murata, J.B. Heymann, K. Mitsuoka, Y. Fujiyoshi, B.L. Smith, P. Agre, A. Engel, The three-dimensional structure of aquaporin-1, *Nature*. 387 (1997) 624–627. doi:10.1038/42512.
- [27] J.A. Rodriguez, M.I. Ivanova, M.R. Sawaya, D. Cascio, F.E. Reyes, D. Shi, S. Sangwan, E.L. Guenther, L.M. Johnson, M. Zhang, L. Jiang, M.A. Arbing, B.L. Nannenga, J. Hattne, J. Whitelegge, A.S. Brewster, M. Messerschmidt, S. Boutet, N.K. Sauter, T. Gonen, D.S. Eisenberg, Structure of the toxic core of  $\alpha$ -synuclein from invisible crystals, *Nature*. 525 (2015) 486–490. doi:10.1038/nature15368.
- [28] K. Hedberg, L. Hedberg, D.S. Bethune, C.A. Brown, H.C. Dorn, R.D. Johnson, M. DE Vries, Bond lengths in free molecules of buckminsterfullerene, c60, from gas-phase electron diffraction., *Science*. 254 (1991) 410–2. doi:10.1126/science.254.5030.410.
- [29] S.W. Hui, D.F. Parsons, Electron diffraction of wet biological membranes., *Science*. 184 (1974) 77–8. doi:10.1126/SCIENCE.184.4132.77.
- [30] L. Palatinus, P. Brázda, P. Boullay, O. Perez, M. Klementová, S. Petit, V. Eigner, M. Zaarour, S. Mintova, Hydrogen positions in single nanocrystals revealed by electron diffraction., *Science*. 355 (2017) 166–169. doi:10.1126/science.aak9652.
- [31] R. Bernard, J. Day, E. Chin, Strong Olivine Lattice Preferred Orientation in Brachinite-Like Achondrites, in: *Lunar Planet. Sci. Conf.*, 2019.
- [32] L.J. Hufford, E.J. Chin, S.E. Perry, E.A. Miranda, J.M. Hanchar, Xenolith Constraints on Rheology of Heterogenous Deep Crust Beneath the Eastern Mojave Desert, California, in: *GSA Annu. Meet. Phoenix, Arizona, USA-2019*, 2019.
- [33] S.R. Spurgeon, C. Ophus, L. Jones, A. Petford-Long, S. V. Kalinin, M.J. Olszta, R.E. Dunin-Borkowski, N. Salmon, K. Hattar, W.C.D. Yang, R. Sharma, Y. Du, A. Chiamonti, H. Zheng, E.C. Buck, L. Kovarik, R.L. Penn, D. Li, X. Zhang, M. Murayama, M.L. Taheri, Towards data-driven next-generation transmission electron microscopy, *Nat. Mater.* (2020). doi:10.1038/s41563-020-00833-z.
- [34] C. Zhang, C. Zhu, P. Cao, X. Wang, F. Ye, K. Kaufmann, L. Casalena, B.E. Macdonald, X. Pan, K. Vecchio, E.J. Lavernia, Aged metastable high-entropy alloys with heterogeneous lamella structure for superior strength-ductility synergy, *Acta Mater.* 199 (2020) 602–612. doi:10.1016/j.actamat.2020.08.043.
- [35] H. Wang, T. Harrington, C. Zhu, K.S. Vecchio, Design, fabrication and characterization of FeAl-based metallic-intermetallic laminate (MIL) composites, *Acta Mater.* 175 (2019) 445–456. doi:10.1016/j.actamat.2019.06.039.

- [36] K.S. Vecchio, D.B. Williams, Convergent beam electron diffraction study of Al<sub>3</sub>Zr in Al-Zr AND Al-Li-Zr alloys, *Acta Metall.* 35 (1987) 2959–2970. doi:10.1016/0001-6160(87)90295-1.
- [37] K.S. Vecchio, D.B. Williams, Convergent beam electron diffraction analysis of the T<sub>1</sub> (Al<sub>2</sub>CuLi) phase in Al-Li-Cu alloys, *Metall. Trans. A.* 19 (1988) 2885–2891. doi:10.1007/BF02647714.
- [38] A.J. Schwartz, M. Kumar, B.L. Adams, D.P. Field, *Electron backscatter diffraction in materials science*, Springer Science+Business Media, LLC, New York, 2009. doi:10.1007/978-0-387-88136-2.
- [39] B.L. DeCost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, *Comput. Mater. Sci.* 110 (2015) 126–133. doi:10.1016/J.COMMATSCI.2015.08.011.
- [40] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, T.J. Harrington, E. Marin, K.S. Vecchio, Crystal symmetry determination in electron diffraction using machine learning, *Science (80-. )*. 367 (2020) 564–568. doi:10.1126/science.aay3062.
- [41] K. Kaufmann, C. Zhu, A.S. Rosengarten, K.S. Vecchio, Deep Neural Network Enabled Space Group Identification in EBSD, *Microsc. Microanal.* 26 (2020) 447–457. doi:10.1017/S1431927620001506.
- [42] K. Kaufmann, H. Lane, X. Liu, K.S. Vecchio, Efficient few-shot machine learning for classification of EBSD patterns, *Sci. Rep.* 11 (2021) 8172. doi:10.1038/s41598-021-87557-5.
- [43] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, H. Wang, K.S. Vecchio, Phase Mapping in EBSD Using Convolutional Neural Networks, *Microsc. Microanal.* 26 (2020) 458–468. doi:10.1017/S1431927620001488.
- [44] A. Ziletti, D. Kumar, M. Scheffler, L.M. Ghiringhelli, Insightful classification of crystal structures using deep learning, *Nat. Commun.* 9 (2018) 2775. doi:10.1038/s41467-018-05169-6.
- [45] M.H. Modarres, R. Aversa, S. Cozzini, R. Ciancio, A. Leto, G.P. Brandino, Neural Network for Nanoscience Scanning Electron Microscope Image Recognition, *Sci. Rep.* 7 (2017) 1–12. doi:10.1038/s41598-017-13565-z.
- [46] T. Stan, Z.T. Thompson, P.W. Voorhees, Optimizing convolutional neural networks to perform semantic segmentation on large materials imaging datasets: X-ray tomography and serial sectioning, *Mater. Charact.* 160 (2020) 110119. doi:10.1016/j.matchar.2020.110119.
- [47] B.L. DeCost, B. Lei, T. Francis, E.A. Holm, High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel, *Microsc. Microanal.* 25 (2019) 21–29. doi:10.1017/S1431927618015635.
- [48] Z. Ding, E. Pascal, M. De Graef, Indexing of electron back-scatter diffraction patterns using a convolutional neural network, *Acta Mater.* 199 (2020) 370–382.

doi:10.1016/j.actamat.2020.08.046.

- [49] K. de Haan, Z.S. Ballard, Y. Rivenson, Y. Wu, A. Ozcan, Resolution enhancement in scanning electron microscopy using deep learning, *Sci. Rep.* 9 (2019) 1–7. doi:10.1038/s41598-019-48444-2.
- [50] T. Xie, J.C. Grossman, Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties, *Phys. Rev. Lett.* 120 (2018) 145301. doi:10.1103/PhysRevLett.120.145301.
- [51] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature.* 521 (2015) 436–444. doi:10.1038/nature14539.
- [52] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, *Nature.* 323 (1986) 533–536. doi:10.1038/323533a0.
- [53] A. Mahendran, A. Vedaldi, Visualizing Deep Convolutional Neural Networks Using Natural Pre-images, *Int. J. Comput. Vis.* 120 (2016) 233–255. doi:10.1007/s11263-016-0911-8.
- [54] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, in: *Proc. IEEE Int. Conf. Comput. Vis.*, 2017: pp. 618–626. doi:10.1109/ICCV.2017.74.
- [55] R. Kotikalapudi, keras-vis, (2017). <https://github.com/raghakot/keras-vis>.
- [56] M. Ge, F. Su, Z. Zhao, D. Su, Deep learning analysis on microscopic imaging in materials science, *Mater. Today Nano.* 11 (2020) 100087. doi:10.1016/j.mtnano.2020.100087.

## Chapter 2 Discovery of high-entropy ceramics via machine learning

Kevin Kaufmann<sup>1</sup>, Daniel Maryanovsky<sup>1,2,6</sup>, William M. Mellor<sup>3,6</sup>, Chaoyi Zhu<sup>3</sup>, Alexander S. Rosengarten<sup>1</sup>, Tyler J. Harrington<sup>3</sup>, Corey Oses<sup>4</sup>, Cormac Toher<sup>4</sup>, Stefano Curtarolo<sup>4,5</sup> and Kenneth S. Vecchio<sup>1\*</sup>

<sup>1</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA.

<sup>2</sup>Department of Cognitive Science, UC San Diego, La Jolla, CA 92093, USA.

<sup>3</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA.

<sup>4</sup>Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708, USA.

<sup>5</sup>Materials Science, Electrical Engineering, Physics and Chemistry, Duke University, Durham, NC 27708, USA.

<sup>6</sup>These authors contributed equally: Daniel Maryanovsky, William M. Mellor.

\* Corresponding Author

### Abstract

Although high-entropy materials are attracting considerable interest due to a combination of useful properties and promising applications, predicting their formation remains a hindrance for rational discovery of new systems. Experimental approaches are based on physical intuition and/or expensive trial and error strategies. Most computational methods rely on the availability of sufficient experimental data and computational power. Machine learning (ML) applied to materials science can accelerate development and reduce costs. In this study, we propose an ML method, leveraging thermodynamic and compositional attributes of a given material for predicting the synthesizability (i.e., entropy-forming ability) of disordered metal carbides. The relative

importance of the thermodynamic and compositional features for the predictions are then explored. The approach's suitability is demonstrated by comparing values calculated with density functional theory to ML predictions. Finally, the model is employed to predict the entropy-forming ability of 70 new compositions; several predictions are validated by additional density functional theory calculations and experimental synthesis, corroborating the effectiveness in exploring vast compositional spaces in a high-throughput manner. Importantly, seven compositions are selected specifically, because they contain all three of the Group VI elements (Cr, Mo, and W), which do not form room temperature-stable rock-salt monocarbides. Incorporating the Group VI elements into the rock-salt structure provides further opportunity for tuning the electronic structure and potentially material performance.

## **2.1 Introduction**

Traditional alloys have been developed utilizing one principal element with minor additions of other alloying elements as a means of achieving a desired combination of properties and/or microstructures. Recently, research efforts have been directed toward the study of materials with significant atomic fractions of multiple elements, thus opening a richer composition space [1–3]. This class of materials typically contains four or more elements that do not necessarily result in a single-phase (multi-principle element alloys), and often greater than five elements to maximize the configurational entropy and improve the stability of the single-phase solid solution (high-entropy alloys) [4]. High-entropy offers increased solubility of components, drawing new attention to unexplored center regions of phase diagrams. Novel high-entropy materials that exist as a single, highly disordered, crystalline phase have been of particular research interest [5–10]. As this field has continued to evolve, a number of fascinating combinations of material properties have begun to emerge [11–16].

Finding these materials is often challenging though, owing to the sheer size of these unexplored regions away from the corners of phase diagrams. The search for effective scientific strategies and models has thus far required time- and cost-intensive experimental evaluations of many candidate single-phase high-entropy materials. The disordered configuration presents a challenge for most computational approaches [17], and there is not always sufficient experimental data for validation of positive and negative calculated results. Phase diagram calculations, often combined with other rules and models, have been applied successfully [6,18–20] but the underlying databases lack significant experimental underpinnings. High-throughput computational materials design combines thermodynamic and electronic-structure methods with data mining capabilities to more quickly evaluate material compositions for novel properties [21–23]. These *ab-initio* computing efforts have recently yielded a descriptor known as entropy-forming-ability (EFA), which has shown considerable promise for predicting the ease of synthesizability and homogeneity of such materials [5,12]. A high EFA value for a specific composition signifies a small energy penalty to incorporate disorder, *i.e.* this descriptor can be a sorting parameter for likely single-phase, disordered, high-entropy materials. This descriptor was previously calculated for 56 high-entropy carbide compositions and the single-phase cutoff was experimentally validated to exist between an EFA value of 45 and 50 [5,12]. The highest EFA materials have been demonstrated, via EXAFS, to exhibit minimal short-range chemical order [12], a concern in the high entropy materials community [24,25]. While this method is high throughput in comparison to other *ab-initio* efforts, calculation of EFA values remains a computationally intensive, time-consuming task. Thoroughly searching this new composition space, conservatively estimated to comprise hundreds of billions of new alloys [1], is simply not feasible with this approach alone. Herein, we propose applying data science tools, specifically

machine learning (ML), in order to guide more expensive computational and experimental search strategies toward promising candidate materials and therefore accelerate materials discovery.

Recently, the materials science field has embraced the big data revolution as large databases become cost effective and data generation rates continue to accelerate [26–30]. This has resulted in the development of a number of powerful data science tools to assist material scientists [31–36]. In the realm of materials discovery, data science tools have aided in the accelerated discovery or identification of new compositions for bulk metallic glasses [37], shape memory alloys [38], Heusler compounds [39,40], and photocatalysts for CO<sub>2</sub> reduction [41]. Other work has focused on the development of ML methods to establish structure-property linkages [35,42], or predict the crystal stability of new materials [43,44]. In 2016, Ward *et al.* [45] proposed a chemically diverse list of attributes, primarily data-mined from the periodic table, as a general set of features for broad material property prediction. These data-driven models can be fit to existing experimental data and continuously refined as new data is collected [46]. This inherent flexibility of ML-based decision-making frameworks provides an advantage given the dynamic nature of phase formation and stability. Moreover, when compared to density functional theory (DFT), the state-of-the-art toolbox for quantum mechanical modeling, ML models can perform well with reduced computational cost and without the need for atomic structure information [44,47]. This provides an opportunity to search materials space in an unconstrained manner without concern for the combinatorial explosion of higher-order compositions (ternaries, quaternaries, quinary, etc.) [48].

Through this work, we aim to accelerate materials innovation by developing a rapid predictor of the stability of high-entropy materials and demonstrating the model's capability to predict single- or multi-phase results. With regard to speed, our ML model can evaluate the EFA

of a single composition in under a millisecond, compared to hundreds of hours per composition with DFT, even using efficient automatic frameworks such as AFLOW [49]. The robustness of the model is investigated by focusing on locating successful five component compositions containing all three of the Group VI metals (Cr, Mo, and W) as 60% of the cation sublattice. The interest in using the ML model to locate single-phase compositions containing all three Group VI metals stems from the relationship between the electronic structure and mechanical/physical properties of transition metal carbides [50,51]. Prior studies have revealed that the transition metal carbides can be more effectively tuned by the enhanced metallic bonding owing to valence filling instead of conventional microstructural engineering principles [50,52]. For example, the Group IV and V monocarbides readily form the rock-salt structure and demonstrate improved mechanical properties, such as fracture toughness, with changing directionality of the bonding as more valence electrons become available in Group V [50,53]. Computationally, the trend in increasing toughness is expected to continue to the right on the periodic table; however, the Group VI metals do not form a room temperature stable rock-salt phase [54–56]. By employing high-entropy effects (*i.e.* increased solubility), we proposed that the three Group VI metals can be incorporated into a room temperature stable rock-salt structure, resulting in an increased number of available electrons, and a novel group of materials with the potential to overturn previous material engineering limitations.

In this work, several single-phase, rock-salt crystal structure, five-metal cation carbides – for which three of the precursors have different structures and stoichiometric ratios of anions to cations from the resultant face centered cubic high-entropy material – are evaluated. The available precursors for the Group VI metals are hexagonal Mo<sub>2</sub>C, hexagonal WC, orthorhombic W<sub>2</sub>C, and orthorhombic Cr<sub>3</sub>C<sub>2</sub>). Rock-salt MoC and WC are only stable at temperatures above 1940°C and

2500°C, respectively. The only FCC system in the Cr-C phase diagram is Cr<sub>23</sub>C<sub>6</sub>. See Supplementary Figures 1-3 for the binary phase diagrams. To date, the authors are unaware of any previously explored high-entropy carbides containing Cr or the prior calculation of the EFA value by DFT for any Cr-containing compounds. The formation of a rock-salt structured monocarbide, wherein 60% of the cation species (Cr, Mo, and W) do not form this structure as their stable room temperature phase, is neither obvious nor readily predictable based on current theories.

These design goals are accomplished by supplementing the set of chemical descriptors of each composition with information from the calculated phase diagrams and utilizing a ML framework to rapidly predict the entropy-forming-ability of seventy previously unstudied high-entropy carbides containing Cr, an element not considered in the original composition space[5]. Complete information on the construction, training, and implementation of the ML model is included in the Methods section. Based on the validation against previously reported high-entropy metal carbides, comparison with DFT calculations for several new compositions, and the ability to locate and synthesize several otherwise unintuitive materials, we find that this screening strategy is aptly designed to identify promising high-entropy systems. The successful outcome demonstrates the synergy between thermodynamics, chemical descriptors, and ML methods for rapidly evaluating new materials based on prior experiments and computation.

## **2.2 Results**

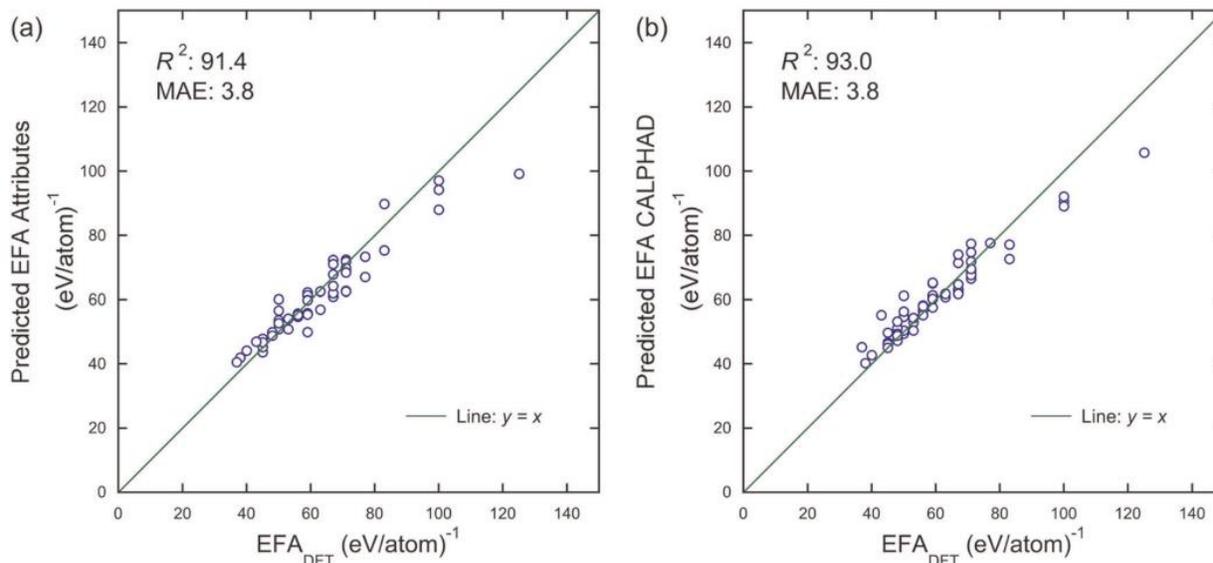
### **2.2.1 Model Performance**

The search for new high-entropy ceramics begins with fitting a random forest [57], a type of ML model, on 56 previously reported EFA values [5]. This data set includes nine synthesized compositions, six single-phase and three multi-phase. The previous study only utilizes eight

carbide forming metal elements (Hf, Nb, Ta, Ti, Mo, V, W, and Zr). As will be demonstrated, even this sparse data set with relatively few compositions with high entropic contributions is very useful in guiding subsequent experiments toward the best candidates and away from the multi-phase materials.

Because our goal is to select the best model hyperparameters for predicting new compositions outside our training set, we evaluate the ML model's performance using 5-fold cross validation and a grid search across selected hyperparameters (see the Methods section for further details). The final model hyperparameters selected for both models, with and without CALPHAD data, are 10 predictor trees and mean absolute error (MAE) for scoring. The best hyperparameters were used to fit models to the labeled data. Supplementary Figure 4 shows an example predictor tree from the model with CALPHAD data and demonstrates the complex relationships between the predictor variables. Figure 1 compares the performance of the ML model fit with only chemical attributes (Figure 1a) and the model fit with chemical attributes and information from CALPHAD (Fig. 1b). The DFT calculated and ML predicted values for each model are listed in Supplementary Table 1. While the mean absolute errors for all models are equivalent ( $3.8 \text{ (eV/atom)}^{-1}$ ), the coefficient of determination ( $R^2$ ) suggests the observed outcomes are better replicated by the ML model with access to the CALPHAD data. However, both models have a systematic error in which the compositions with known EFA less than 50 are overestimated and, more noticeably, compositions with an EFA above 80 are underestimated. The small number of samples above 80 (6 total) coupled with bootstrapping (~66% of the data is used per tree) results in a low probability for them to be included in the construction of each decision tree. Further, the average EFA of the materials in each tree (approx. 58 depending on the tree) is in line with the average for the dataset. With only 1 sample above 100 in the dataset, and these samples having a

low probability of being used in tree construction, the averaging process in random forest pulls down the predicted values for the highest EFA materials. It will be demonstrated that the improved  $R^2$  performance of the model toward fitting the starting dataset will provide improved extrapolation on the Cr-containing systems in the search for high-entropy ceramics containing all three Group VI precursors.



**Figure 2.1 Evaluation of the ML models fit to available data.** a) The ML-predicted EFA using a random forest fit with 108 chemical attributes evaluated against the labels of the data set from DFT. b) The ML predicted EFA values for a random forest fit with 108 chemical attributes plus 8 features from CALPHAD evaluated against the known EFA from DFT. The line  $y = x$  is plotted to show the deviation from perfect predictions.

### 2.2.2 Feature importance

The permutation importance of the chemical attribute and CALPHAD features is studied to provide interpretability to the ML model. Details for each chemical attribute can be found in the Supplementary Information. The rationale for selecting permutation importance is the following: randomly permuting the value of predictor variable  $X_i$  and computing the EFA together with the unpermuted predictor variables, will result in significantly reduced prediction accuracy if the original variable  $X_i$  was significantly associated with the output value. Permutation importance

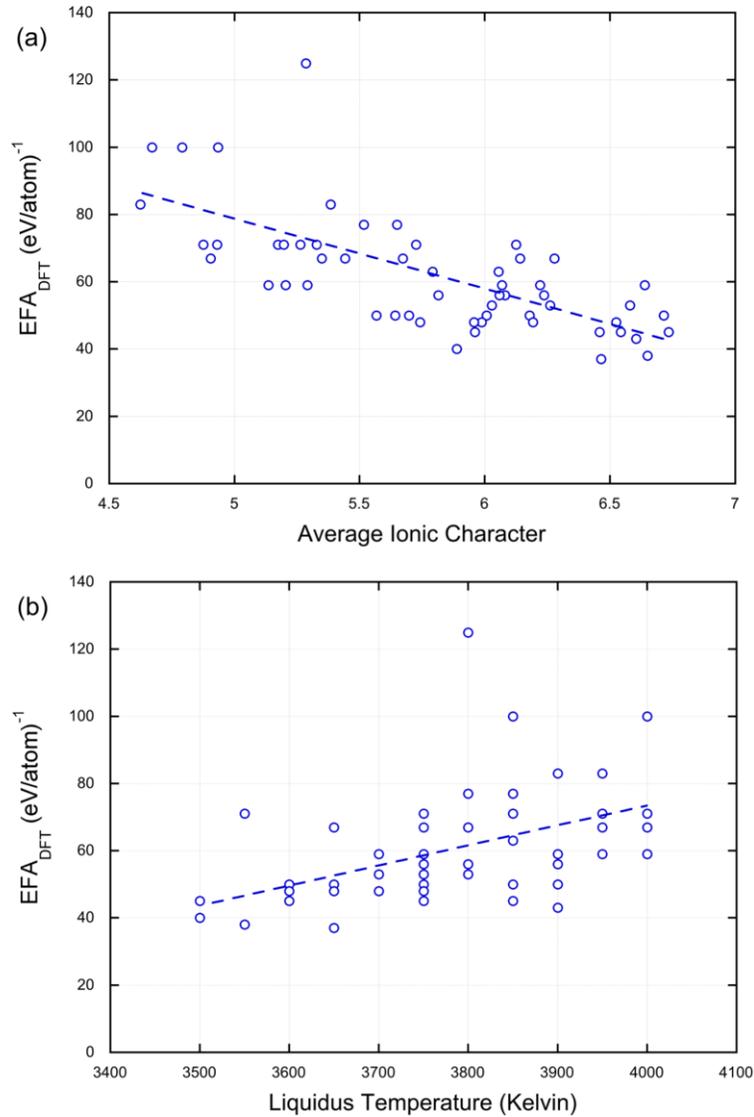
also has the advantage, compared to univariate screening methods, in that it assesses the impact of each predictor variable individually as well as with the other unpermuted predictor variables[58]. Table 1 shows the top ten features and their importance rank for fitting each random forest model to the EFA values calculated from DFT. As evidenced in the model performance and feature importance, the eight additional CALPHAD features provide valuable information about the EFA of a given composition, particularly the liquidus temperature (ranked second). However, CALPHAD diagrams alone would also be insufficient for determining the ability to fabricate a single-phase material. Supplementary Figure 5 demonstrates this by comparing ThermoCalc SSOL6 database computed diagrams of compositions known to form single or multi-phase carbides. For each of these compositions, CALPHAD alone would predict rock-salt to be the primary structure to evolve from the liquid, which would be stable down to nearly 1500K before forming a secondary metal carbide. In reality, only MoNbTaVWC<sub>5</sub> (Supplementary Figure 5a) readily forms a single phase experimentally, while the other three compositions (Supplementary Figure 5b-d) have been demonstrated previously to be multi-phase materials[5]. However, including some CALPHAD data as features improves the ML model via this thermodynamic-based preview of what is likely to occur and improves its extrapolation capabilities beyond that of the chemical attributes alone.

**Table 2.1 Identification of the important features for predicting EFA values.** The top ten features for the ML model with only the chemical attributes are on the left. The top ten features for the ML model including CALPHAD features are on the right. Both models rely on similar features regarding electronegativity, ionic character, and electron orbitals for making the best predictions. The avg(x) and avg. dev(x) denote the composition-weighted average and average deviation, respectively, calculated over the vector of elemental values for each compound. The min(x), max(x), fwm(x) and range(x) correspond to the minimum, maximum, fraction-weighted mean, and range of an attribute for each compound. \* denotes a predictor variable from CALPHAD.

Predictor rank	Model	
	Stoichiometric Attributes	CALPHAD
1	avg(ionic character)	avg(ionic character)
2	min(electrons)	Liquidus temperature*
3	avg. dev(s-valence electrons)	range(electronegativity)
4	max(atomic weight)	avg. dev(d-valence electrons)
5	max(covalent radius)	max(atomic weight)
6	fwm(covalent radius)	fwm(f-valence electrons)
7	range(Mendeleev number)	max(covalent radius)
8	avg. dev(melting temp)	max(unfilled valence electrons)
9	fwm(unfilled s-valence)	fwm(covalent radius)
10	fwm(f-electrons)	range(unfilled valence electrons)

Important features from the chemical attributes are the average ionic character between each of the atomic species, the maximum and fraction-weighted covalent radius, and a few features representing the valence electrons or unfilled orbitals. These chemical attributes quantify the expected bonding nature and local environment each atom will experience if single phase (*i.e.* homogeneously disordered). Along the same lines, these metrics also assist the ML model to determine what atomic environments are unfavorable, resulting in multi-phase materials. Further

analysis of the relationship between the entropy-forming-ability of a composition and the top ranked predictors reveals there is noticeable correlation (Figure 2). A plot of average ionic character versus EFA reveals that increasing the average ionic character between the pairs of atoms is more likely to result in a multi-phase material (Figure 2a). This property has been previously suggested to play a role in determining single or multi-phase outcomes, but has not yet been extensively studied and its contribution not well understood[10,59]. A parameter not previously studied in the high-entropy literature, the liquidus temperature derived from CALPHAD also provides insight into the magnitude of the expected EFA for a given composition (Figure 2b). Intuitively, the compositions with the highest EFA values lie furthest away from the trendlines, highlighting the need for multi-variable approaches, like those offered in ML, to locate the best compositions.



**Figure 2.2 Correlation between EFA and top two features .** a) Increasing the average ionic character of each atom pair in the composition is correlated with a decrease in the entropy-forming ability. b) A positive correlation exists between the increasing liquidus temperature and higher EFA values. Trendlines shown in blue dashes. Note that ten compositions completely overlap when comparing EFA with liquidus temperature.

### 2.2.3 Experimental and computational validation

New experimental compositions were chosen utilizing all nine of the Group IV, V, and VI refractory metals (Cr, Hf, Nb, Mo, Ta, Ti, V, W, and Zr) in equiatomic amounts plus carbon occupying the anion lattice. The ML model fitted to the 56 compositions with DFT derived EFA

values is used to rapidly screen the Cr-containing compositions for high and low expected EFA values. The ML calculated EFA values for the full set of 70 new five-metal compositions are provided in Table 2. Seven candidates are selected from this list for analysis by DFT and experimental synthesis: (i) three candidates with predicted EFA values over  $100 \text{ (eV/atom)}^{-1}$  and each containing all of the Group VI metals ( $\text{CrMoNbVWC}_5$ ,  $\text{CrMoTaVWC}_5$ ,  $\text{CrMoNbTaWC}_5$ ), (ii) three candidates with a predicted EFA less than or equal to  $50 \text{ (eV/atom)}^{-1}$  ( $\text{CrHfMoTiWC}_5$ ,  $\text{CrMoTiWZrC}_5$ , and  $\text{CrHfTaWZrC}_5$ ), and (iii) one composition with an intermediate entropy-forming-ability ( $\text{CrMoTiVWC}_5$ ) also containing the 3 Group VI metals. Computing the EFA from DFT and fabricating the selected compositions with a low predicted EFA serves two purposes: (i) to demonstrate the model performs well at finding both the best and worst candidates, and (ii) to establish that not every system containing all three Group VI metals will form a single phase.

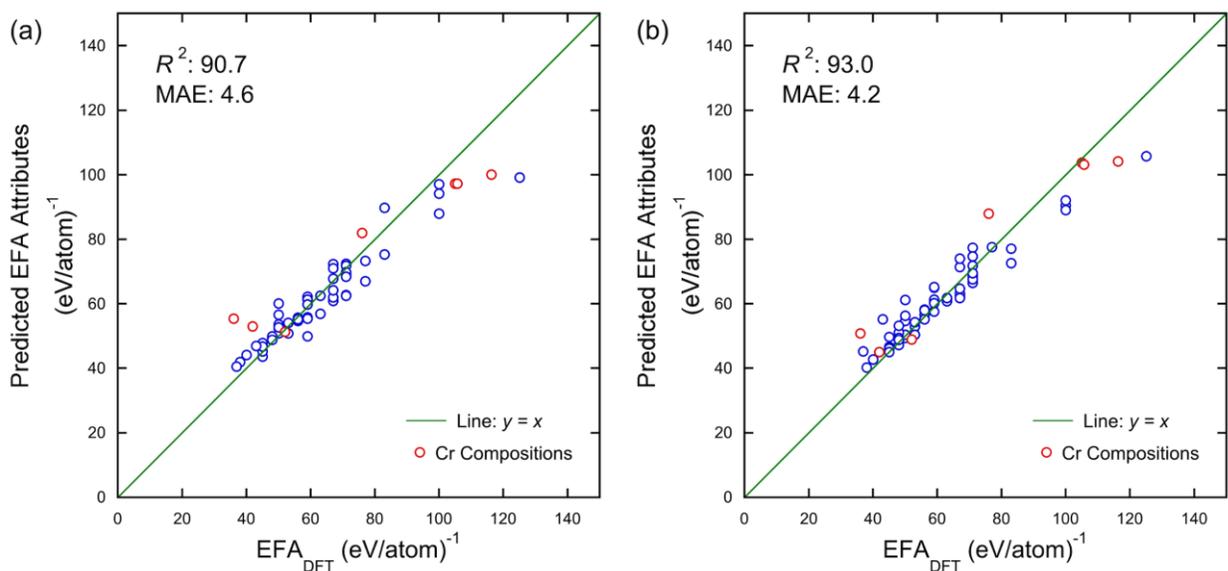
Several of the compositions in Table 2 have ML predicted EFA values that suggest they will readily form a single-phase high-entropy carbide, despite containing the three Group VI refractory metal elements. If successfully synthesized into a single phase, these novel materials would contain three carbides that do not exist as room temperature stable rock-salt monocarbides (refer to the binary phase diagrams in Supplementary Figures 1-3). Several fundamentally interesting compositions are those where one of the rock-salt stable precursors (*i.e.* NbC, TaC, or VC) is substituted from  $\text{MoNbTaVWC}_5$  ( $\text{EFA}_{\text{DFT}}$  of  $125 \text{ (eV/atom)}^{-1}$ )[5] with an orthorhombic ( $\text{Cr}_3\text{C}_2$  or  $\text{W}_2\text{C}$ ) or hexagonal ( $\text{Mo}_2\text{C}$  or WC) precursor that do not form stable rock-salt structures.

**Table 2.2 Results for the ML predicted EFA for seventy new compositions.** Results for both ML models are provided for each composition. For the selected compositions, a DFT computed EFA value is listed in the next column. In the experimental result, “S” and “M” stand for single- and multi-phase, respectively. Units: EFA in (eV/atom)<sup>-1</sup>.

Composition	EFA <sub>Attributes</sub>	EFA <sub>CALPHAD</sub>	EFA <sub>DFT</sub>	Exp.	Composition	EFA <sub>Attributes</sub>	EFA <sub>CALPHAD</sub>	EFA <sub>DFT</sub>	Exp.
CrNbTaVWC <sub>5</sub>	94	105			CrHfTiVZrC <sub>5</sub>	76	70		
CrMoNbTaVC <sub>5</sub>	107	105			CrMoTaVZrC <sub>5</sub>	70	70		
CrMoNbVWC <sub>5</sub>	100	104	116	S	CrMoTiVZrC <sub>5</sub>	65	67		
CrMoNbTaWC <sub>5</sub>	97	104	105	S	CrHfMoTiVC <sub>5</sub>	72	66		
CrMoTaVWC <sub>5</sub>	97	103	106	S	CrMoNbTiZrC <sub>5</sub>	68	66		
CrMoNbTaTiC <sub>5</sub>	98	93			CrTaTiVWC <sub>5</sub>	69	66		
CrMoTaTiVC <sub>5</sub>	98	93			CrHfMoNbZrC <sub>5</sub>	63	62		
CrMoNbTiVC <sub>5</sub>	95	92			CrHfNbTaWC <sub>5</sub>	63	62		
CrHfNbTaVC <sub>5</sub>	97	91			CrHfMoVZrC <sub>5</sub>	69	62		
CrMoTiVWC <sub>5</sub>	82	88	76	S	CrHfMoTaZrC <sub>5</sub>	61	60		
CrMoTaTiWC <sub>5</sub>	82	88			CrMoVWZrC <sub>5</sub>	65	59		
CrHfNbTaTiC <sub>5</sub>	93	88			CrMoNbWZrC <sub>5</sub>	64	59		
CrNbTaTiVC <sub>5</sub>	95	88			CrNbTaWZrC <sub>5</sub>	59	58		
CrHfTaTiVC <sub>5</sub>	91	87			CrHfMoVWC <sub>5</sub>	65	57		
CrHfMoNbVC <sub>5</sub>	75	84			CrHfMoTaWC <sub>5</sub>	61	56		
CrNbTaVZrC <sub>5</sub>	93	83			CrHfTaVWC <sub>5</sub>	54	56		
CrTaTiVZrC <sub>5</sub>	91	83			CrHfTaTiWC <sub>5</sub>	56	55		
CrMoNbTiWC <sub>5</sub>	74	81			CrHfMoNbWC <sub>5</sub>	52	54		
CrMoNbVZrC <sub>5</sub>	80	81			CrMoTaWZrC <sub>5</sub>	54	54		
CrNbTaTiZrC <sub>5</sub>	91	79			CrHfNbTiWC <sub>5</sub>	58	52		
CrHfMoNbTaC <sub>5</sub>	76	79			CrHfMoTiZrC <sub>5</sub>	65	51		
CrHfNbTiVC <sub>5</sub>	91	78			CrHfTaWZrC <sub>5</sub>	55	51	36	M
CrMoNbTaZrC <sub>5</sub>	70	78			CrHfNbVWC <sub>5</sub>	58	51		
CrHfTaVZrC <sub>5</sub>	77	78			CrTaTiWZrC <sub>5</sub>	54	50		
CrHfNbTaZrC <sub>5</sub>	70	77			CrNbVWZrC <sub>5</sub>	51	50		
CrNbTiVZrC <sub>5</sub>	94	76			CrMoTiWZrC <sub>5</sub>	51	49	52	M
CrNbTiVWC <sub>5</sub>	72	74			CrTaVWZrC <sub>5</sub>	52	48		
CrHfTaTiZrC <sub>5</sub>	82	74			CrTiVWZrC <sub>5</sub>	50	48		
CrHfMoTaVC <sub>5</sub>	73	73			CrNbTiWZrC <sub>5</sub>	51	48		
CrHfNbTiZrC <sub>5</sub>	76	72			CrHfVWZrC <sub>5</sub>	52	47		
CrHfMoNbTiC <sub>5</sub>	69	72			CrHfNbWZrC <sub>5</sub>	52	46		
CrHfNbVZrC <sub>5</sub>	88	71			CrHfTiWZrC <sub>5</sub>	51	46		
CrHfMoTaTiC <sub>5</sub>	73	71			CrHfTiVWC <sub>5</sub>	46	46		
CrMoTaTiZrC <sub>5</sub>	72	70			CrHfMoWZrC <sub>5</sub>	53	45		
CrNbTaTiWC <sub>5</sub>	67	70			CrHfMoTiWC <sub>5</sub>	53	45	42	M

As the first step in validating the ML model’s extrapolation into the Cr-containing chemical space, the entropy-forming-ability of the seven selected compositions were subsequently

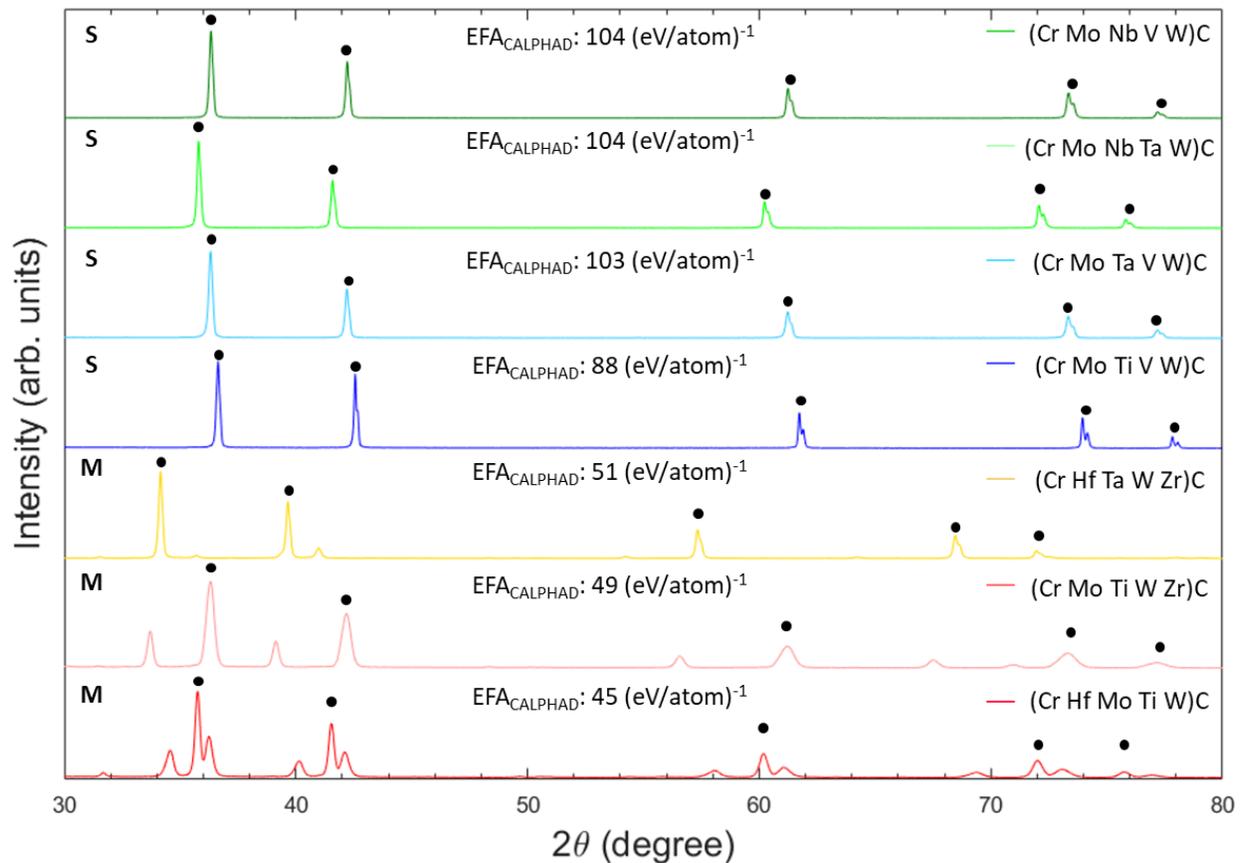
computed by DFT. The *ab-initio* EFA values are located in Table 2 and plots comparing the ML model with chemical attributes (Figure 3a) and the ML model including CALPHAD data (Figure 3b) illustrate the improved regression performance of the model after inclusion of the CALPHAD features. The red circles in Figure 3 are the predicted EFA values for the seven Cr-containing compositions compared to their DFT calculated value. While the ML models were not refit with the new DFT-computed EFA values, the  $R^2$  and mean absolute error of each ML model can be re-evaluated after including the extrapolated data. In comparison to the chemical attributes alone, the  $R^2$  value remains the same, and the MAE increases only slightly.



**Figure 2.3 Machine learning model compared to *ab-initio* results for high-entropy carbides.** a) The ML predicted EFA using a random forest fit with 108 chemical attributes evaluated against the labels of the data set from DFT. b) The ML predicted EFA values for a random forest fit with 108 chemical attributes plus 8 features from CALPHAD evaluated against the known EFA from DFT. The line  $y = x$  is plotted to show the deviation from perfect predictions. Red circles are used to mark the newly calculated Cr-containing compositions.

As a secondary method of validating the ML model, the seven selected materials were fabricated following conventional fabrication processes described in detail in the Methods section. Successful fabrication of the rock-salt structure after full densification was verified via X-ray diffraction (XRD) (Figure 4). Results of XRD analysis for each sample following spark

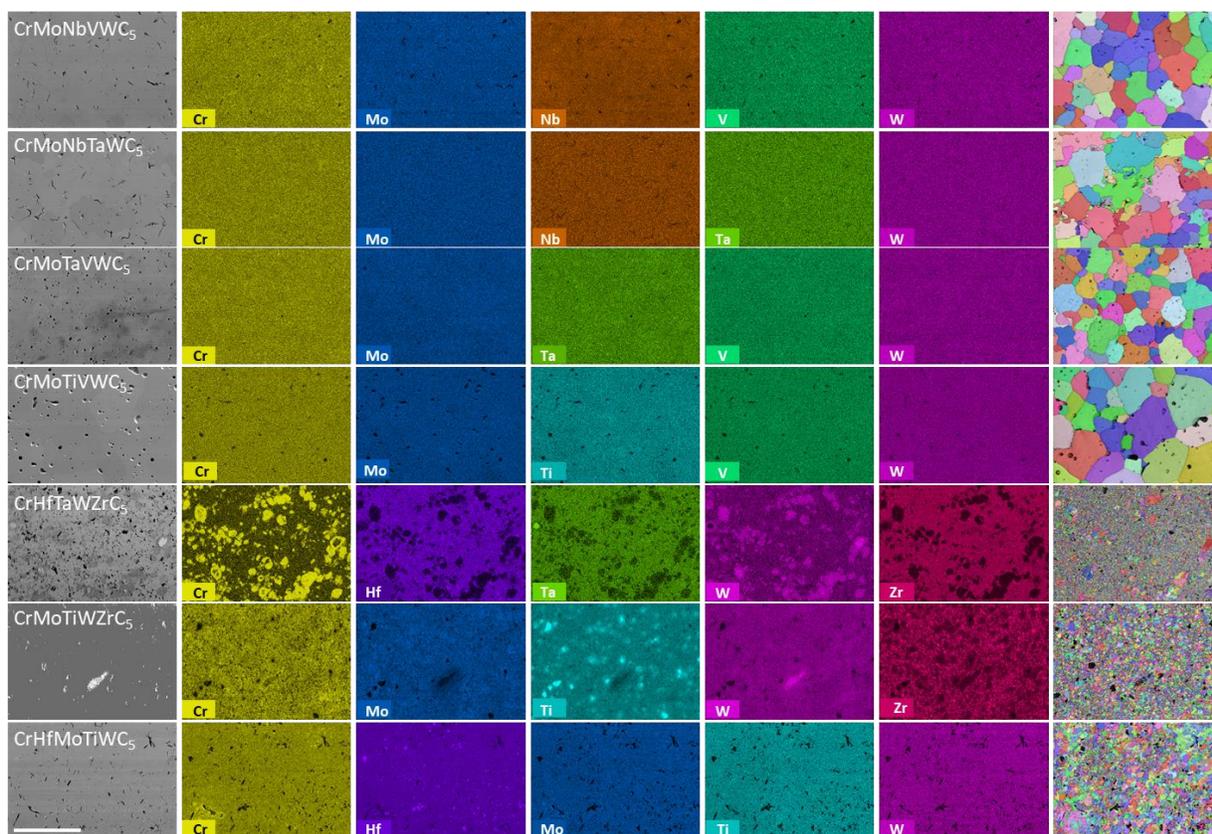
plasma sintering (SPS) demonstrate that compositions  $\text{CrMoNbVWC}_5$ ,  $\text{CrMoNbTaWC}_5$ ,  $\text{CrMoTaVWC}_5$ , and  $\text{CrMoTiVWC}_5$  (the top 4) only exhibit a single set of FCC peaks of the desired rock-salt high-entropy phase. Conversely, XRD of  $\text{CrHfTaWZrC}_5$ ,  $\text{CrMoTiWZrC}_5$ , and  $\text{CrHfMoTiWC}_5$ , (bottom 3) reveal the presence of multiple structures. In the event there are multiple FCC structures present, the majority FCC phase is indexed. In  $\text{CrMoTiWZrC}_5$  and  $\text{CrHfMoTiWC}_5$  the secondary phase is also FCC. The  $\text{CrHfTaWZrC}_5$  system contains a secondary hexagonal phase. The XRD pattern for  $\text{CrHfMoTiWC}_5$  and  $\text{CrHfTaWZrC}_5$  also contain a small amount (<5%) of  $\text{HfO}_2$  that remains due to processing. This is determined not to significantly alter the composition of the carbide phase.



**Figure 2.4 The X-ray diffraction patterns for the same 7 five-metal carbides.** The first four compositions (from the top) exhibit only the desired FCC structure peaks, whereas the remaining compositions have additional peaks indicating the presence of extra phases. The primary FCC phase is indexed with black circles. Compositions are listed from largest to smallest ML predicted EFA. S: single-phase formed; M: multi-phase formed.

Microstructure analysis and energy dispersive X-ray spectroscopy (EDX) were then utilized to determine the homogeneity of the sintered pellets as shown in Figure 5. Coupling the results of both techniques verified that the as-processed samples were either single-phase and chemically homogenous or underwent chemical segregation. For example, in the  $\text{CrMoNbVWC}_5$  microstructure, only grain contrast is present, and no notable indication of clustering or segregation is visible in the elemental maps. On the contrary, the  $\text{CrHfTaWZrC}_5$  sample has observable chemical contrast in the microstructure, and the chemical maps demonstrate that the secondary

phase present in XRD is rich in Cr and W. The  $\text{CrMoTiWZrC}_5$  and  $\text{CrHfTaWZrC}_5$  samples displayed were sintered at  $1600^\circ\text{C}$  to prevent the loss of Cr. When sintered at  $1800^\circ\text{C}$ , EDS revealed the Cr content in these samples was as low as 2 at%, and the chrome carbide was found to have reacted with the graphite tooling. The medium entropy composition,  $\text{CrMoTiVWC}_5$ , resulted in a single FCC rock-salt structure after initial sintering, but required annealing as described in the Methods section to reach chemical homogeneity. Subsequently, electron backscatter diffraction (EBSD) was utilized to study the resulting microstructure of the samples. The single-phase, homogenous samples are observed to contain large, nearly equiaxed grains with some deviation owing to the remaining pores. This furthers the assertion these compositions are single-phase, since they allow for the kinetics of grain growth. In stark contrast, the multi-phase materials have a significantly reduced grain size, owing to the competing phases preventing further grain growth during sintering.



**Figure 2.5** Microstructural analysis of the synthesized materials. The first column is an electron micrograph for each of the synthesized compositions. Columns 2-6 are selected EDS chemistry maps are present for each of the five metal cations present in each system. Column 7 is an EBSD map of the grain structure, revealing the effect on grain size in multi-phase compared to single-phase compositions. Compositions are listed from largest to smallest ML predicted EFA. Scale bar 100  $\mu\text{m}$ .

## 2.3 Discussion

A powerful data-driven approach to estimating the synthesizability of high-entropy materials, based on data from previous DFT calculations and experimental results, is detailed and demonstrated on 70 new chromium containing compositions. The ML framework is found to be improved by the inclusion of data from CALPHAD and robust toward extrapolating outside the starting chemical space. The ML model enhancement achieved by combining general features and thermodynamic data from CALPHAD is explored via assessing the impact of each predictor variable individually as well as with the other predictor variables (permutation importance) and

evaluating compositions outside the original chemical space. The predictive capability of this method is validated by *ab-initio* calculations and experimental fabrication of several previously unreported compositions, including four single-phase rock-salt materials that would not be obvious candidates given the stable precursors and binary phase diagrams of the Group VI transition metals. These novel materials, of which 60% of the cation lattice contains Group VI metals, represent a step forward in electronic structure engineering of transition metal carbides: prior modeling of the bonding nature with increased valence electrons[54–56] suggests that future material property studies are likely to yield useful combinations for practical engineering applications. Furthermore, the experimentally studied compositions result in single or multi-phase materials in agreement with their predicted EFA values. The remaining predicted materials include diverse chemistries and present ample opportunity for materials discovery. Moreover, the methodology designed opens the door to locating other high-entropy materials, not just ceramics, in a similar manner.

## **2.4 Methods**

### **2.4.1 Machine learning architecture**

Random forests are a combination of decision trees that individually make predictions on each input and the overall prediction determined by a majority voting process[57,60]. Random forest was selected for its utility and performance on diverse problems when compared to other supervised learning models [60]. The random forest regressor is implemented with Scikit-learn [61]. Model hyperparameters are selected via an exhaustive 5-fold cross-validated grid search using the following parameters: number of tree predictors in range 10 to 110 in steps of 10, mean squared error and mean absolute error as criterion, and the number of features to consider when looking for the best split from one to the total number of features available. Each fold is scored

using the mean absolute error between the labels from DFT and the predicted values. To obtain a deterministic behavior during model fitting, the random state is seeded. The best performing hyperparameters are selected to fit a model using the entire training set, with bootstrapping, to maximize the amount of information available for making future predictions.

#### **2.4.2 From chemistry to features**

Each composition is converted to a set of features with the goal of creating a quantitative representation that relates to the essential chemistry, physics, and thermodynamics of each material in a data set. The attributes utilized in this work should not be considered an exhaustive list, but instead a step toward creating a synergistic set of attributes that capture the knowledge of chemistry and experimentally robust thermodynamics. The 108 compositional attributes, defined in the Supplementary Information, are a subset of the general ML framework demonstrated previously to perform well on diverse material problems [45]. The elemental data used to compute the compositional features is sourced from Magpie [45,62]. These chemical attributes are augmented with select data about the number of phases and phase fractions calculated in 100 Kelvin steps as well as the liquidus and solidus temperature from ThermoCalc Software SSOL6 database version 6.1 [63]. The ~800 CALPHAD features are reduced to 8 predictor variables (1% of those available) using the Select From Model method in Scikit-learn [61] to avoid the “curse-of-dimensionality” and find the most relevant subset [64–66]. Select From Model was chosen in this study for its rapid reduction of features in one step in comparison to other multi-step methods such as recursive feature elimination. The max number of features was set to 8 for this study to target 1% of the available data. We do not intend for this feature list to be exhaustive or concrete. The selected features are defined in the Supplementary Information. The data for the predictor variables for the training data and new compositions is contained in the GitHub repository.

### **2.4.3 Interpreting the random forest algorithm**

The random forest model is analyzed to provide clarity to how the ML model evaluated these materials. The variable importance is extracted for the fit model using the “rfpimp” package in Python (available at <https://github.com/parrt/random-forest-importances>, last access: 15 August 2019). The predictor variable importance is ranked on the permutation importance, which directly measures importance by observing the effect on model accuracy by randomly permuting the values of each predictor variable [67]. That is to say, the permutation importance is measuring the impact on output EFA of swapping the value of a selected feature from one composition with the value from a different composition. This method has recently been introduced as an improvement to the mean decrease in impurity metric [58].

### **2.4.4 Sample preparation**

All samples were prepared using the same methods and tools utilized in the previous EFA and HEC studies [5,12]. Initial powders of each of the five binary precursor carbides (NbC, HfC, TiC, ZrC, VC, TaC, Mo<sub>2</sub>C, W<sub>2</sub>C, WC, and Cr<sub>3</sub>C<sub>2</sub>) are obtained in >99% purity and –325 mesh (<44 μm) particle size (Alfa Aesar). The sample is weighed out in 12 g batches and mixed to achieve the desired five-metal carbide compositions. To ensure adequate mixing, each sample is high energy ball milled under argon in a shaker pot mill for a total of 2 h in individual 30-min intervals intersected by 15-min rest times to avoid heating and consequent oxide formation. All milling is done in tungsten carbide-lined stainless-steel milling jars with tungsten carbide grinding media. Bulk sample pellets are synthesized via solid-state processing routes. The field-assisted sintering technique (FAST), also called spark plasma sintering (SPS), is employed to simultaneously densify and react the compositions into single-phase materials. Sintering of each composition is performed at 1800°C with a heating rate of 100°C/min, 60 MPa uniaxial pressure

applied at temperature, with a 10-min dwell at temperature. Subsequent samples of CrMoTiWZrC<sub>5</sub> and CrHfTaWZrC<sub>5</sub> were necessarily sintered at 1600°C instead to prevent the loss of Cr. The composition with a medium EFA value, CrMoTiVWC<sub>5</sub>, is annealed at 1800°C for 3 hours followed by 2000°C for 3 hours to attain chemical homogeneity. All samples are heated in a vacuum environment of less than 20 mtorr with additional holds throughout for adequate off-gassing of the powder materials. Sintering is done in 20 mm graphite die and plunger sets surrounded by carbon-based heat shielding. Additionally, graphite foil surrounds the samples on all sides to prevent reaction with the die. The compositions listed are nominal since actual synthesized compositions can vary due to carbon vacancies in the anion sublattice.

#### **2.4.5 Sample analysis**

Microstructural and elemental analysis is performed using a Thermo Fischer (formerly FEI) Apreo field emission scanning electron microscope (SEM) equipped with an Oxford X-Max<sup>N</sup> EDS detector and an Oxford Symmetry electron backscatter diffraction (EBSD) detector. A combination of secondary and back-scattered electron detectors are utilized for imaging. EDS scans are conducted at length scales of 500x and 1000x to verify multi-length scale homogeneity in the resulting microstructure. EDS quantification confirmed the resulting ratio of metal ions are nearly equiatomic. Crystal structure analysis is implemented using a Rigaku Miniflex X-ray Diffractometer with a 1D detector using a step size of 0.02° and 5 degrees per minute scan rate, using Cu K $\alpha$  radiation (wavelength  $\lambda = 1.54059 \text{ \AA}$ ) for all measurements. The lattice parameter is calculated utilizing a combinatorial method from both MDI Jade and Match! Phase identification software. This is subsequently utilized to model and create a theoretical diffraction profile to be utilized in EBSD.

#### **2.4.6 Calculation of the entropy-forming-ability**

The EFA is calculated using the AFLOW-POCC module [17] implemented in the Automatic-Flow (AFLOW) Framework for Materials Discovery[68]. For each disordered composition, a set of representative ordered supercells is resolved. First, AFLOW-POCC determines the smallest supercell size accommodating the stoichiometry exactly (for the five-metal rock-salt carbides, the value is 5). The unique superlattices of this size are then constructed based on the Hermite Normal Form matrices. The lattices are decorated to generate all viable configurations. To identify unique configurations and their degeneracies rapidly, the Universal Force Field method is employed. The energies of the unique configurations are then calculated using density functional theory having input parameters/settings in accordance with the AFLOW Standard [29]. K-point meshes are generated using the Monkhorst-Pack scheme (Gamma-centered for all materials belonging to the hP and hR Bravais lattice) having at least 6,000 k-points per reciprocal atom. Project-Augmented Wavefunction (PAW) potentials are constructed according to the Perdew-Berke-Ernzerhof (PBE) exchange-correlation functional as implemented in VASP. The plane-wave basis has a kinetic energy cut-off 1.4 times larger than that recommended for each species. Spin polarization is considered. The electronic and ionic convergence criteria are  $10^{-3}$  and  $10^{-2}$  eV, respectively. The EFA is defined as the inverse of the spread of these energies [5].

#### **2.5 Acknowledgements**

The authors acknowledge support through the Office of Naval Research ONR-MURI (grant No. N00014-15-1-2863). K. Kaufmann would like to acknowledge support by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program. K. Kaufmann would also like to acknowledge the financial support of the ARCS Foundation, San Diego

Chapter. KSV would like to acknowledge the financial generosity of the Oerlikon Group in support of his research group.

Chapter 2, in full, is a reprint of the material “Discovery of high-entropy ceramics via machine learning” as it appears in *npj Computational Materials*. K. Kaufmann, D. Maryanovsky, W. Mellor, C. Zhu, A. S. Rosengarten, T. J. Harrington, C. Oses, C. Toher, S. Curtarolo, and K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

## 2.6 References

- [1] D.B. Miracle, High entropy alloys as a bold step forward in alloy development, *Nat. Commun.* 10 (2019) 1805. doi:10.1038/s41467-019-09700-1.
- [2] C. Oses, C. Toher, S. Curtarolo, High-entropy ceramics, *Nat. Rev. Mater.* (2020) 1–15. doi:10.1038/s41578-019-0170-8.
- [3] E.P. George, D. Raabe, R.O. Ritchie, High-entropy alloys, *Nat. Rev. Mater.* 4 (2019) 515–534. doi:10.1038/s41578-019-0121-4.
- [4] C. Toher, C. Oses, D. Hicks, S. Curtarolo, Unavoidable disorder and entropy in multi-component systems, *Npj Comput. Mater.* 5 (2019). doi:10.1038/s41524-019-0206-z.
- [5] P. Sarker, T. Harrington, C. Toher, C. Oses, M. Samiee, J.P. Maria, D.W. Brenner, K.S. Vecchio, S. Curtarolo, High-entropy high-hardness metal carbides discovered by entropy descriptors, *Nat. Commun.* 9 (2018) 4980. doi:10.1038/s41467-018-07160-7.
- [6] O.N. Senkov, J.D. Miller, D.B. Miracle, C. Woodward, Accelerated exploration of multi-principal element alloys with solid solution phases, *Nat. Commun.* 6 (2015) 6529. doi:10.1038/ncomms7529.
- [7] B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Microstructural development in equiatomic multicomponent alloys, *Mater. Sci. Eng. A.* 375–377 (2004) 213–218. doi:10.1016/J.MSEA.2003.10.257.
- [8] J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, S.-Y. Chang, Nanostructured High-Entropy Alloys with Multiple Principal Elements: Novel Alloy Design Concepts and Outcomes, *Adv. Eng. Mater.* 6 (2004) 299–303. doi:10.1002/adem.200300567.
- [9] J. Gild, J. Braun, K. Kaufmann, E. Marin, T. Harrington, P. Hopkins, K. Vecchio, J. Luo, A high-entropy silicide: (Mo 0.2 Nb 0.2 Ta 0.2 Ti 0.2 W 0.2 )Si<sub>2</sub>, *J. Mater.* (2019). doi:10.1016/j.jmat.2019.03.002.
- [10] C.M. Rost, E. Sachet, T. Borman, A. Moballegh, E.C. Dickey, D. Hou, J.L. Jones, S. Curtarolo, J.P. Maria, Entropy-stabilized oxides, *Nat. Commun.* 6 (2015) 8485. doi:10.1038/ncomms9485.
- [11] B. Gludovatz, A. Hohenwarter, D. Catoor, E.H. Chang, E.P. George, R.O. Ritchie, A fracture-resistant high-entropy alloy for cryogenic applications, *Science* (80-. ). 345 (2014) 1153–1158. doi:10.1126/science.1254581.
- [12] T.J. Harrington, J. Gild, P. Sarker, C. Toher, C.M. Rost, O.F. Dippo, C. McElfresh, K. Kaufmann, E. Marin, L. Borowski, P.E. Hopkins, J. Luo, S. Curtarolo, D.W. Brenner, K.S. Vecchio, Phase stability and mechanical properties of novel high entropy transition metal carbides, *Acta Mater.* 166 (2019). doi:10.1016/j.actamat.2018.12.054.
- [13] X. Lim, Mixed-up metals make for stronger, tougher, stretchier alloys, *Nature.* 533 (2016) 306–307. doi:10.1038/533306a.

- [14] Z. Li, C.C. Tasan, H. Springer, B. Gault, D. Raabe, Interstitial atoms enable joint twinning and transformation induced plasticity in strong and ductile high-entropy alloys, *Sci. Rep.* 7 (2017) 40704. doi:10.1038/srep40704.
- [15] T.-K. Tsao, A.-C. Yeh, C.-M. Kuo, K. Kakehi, H. Murakami, J.-W. Yeh, S.-R. Jian, The High Temperature Tensile and Creep Behaviors of High Entropy Superalloy, *Sci. Rep.* 7 (2017) 12658. doi:10.1038/s41598-017-13026-7.
- [16] O.N. Senkov, G.B. Wilks, J.M. Scott, D.B. Miracle, Mechanical properties of Nb<sub>25</sub>Mo<sub>25</sub>Ta<sub>25</sub>W<sub>25</sub> and V<sub>20</sub>Nb<sub>20</sub>Mo<sub>20</sub>Ta<sub>20</sub>W<sub>20</sub> refractory high entropy alloys, *Intermetallics*. 19 (2011) 698–706. doi:10.1016/J.INTERMET.2011.01.004.
- [17] K. Yang, C. Oses, S. Curtarolo, Modeling Off-Stoichiometry Materials with a High-Throughput Ab-Initio Approach, *Chem. Mater.* 28 (2016) 6484–6492. doi:10.1021/acs.chemmater.6b01449.
- [18] M. Gao, D. Alman, Searching for Next Single-Phase High-Entropy Alloy Compositions, *Entropy*. 15 (2013) 4504–4519. doi:10.3390/e15104504.
- [19] F. Zhang, C. Zhang, S.L. Chen, J. Zhu, W.S. Cao, U.R. Kattner, An understanding of high entropy alloys from phase diagram calculations, *Calphad*. 45 (2014) 1–10. doi:10.1016/J.CALPHAD.2013.10.006.
- [20] Y. Lederer, C. Toher, K.S. Vecchio, S. Curtarolo, The search for high entropy alloys: A high-throughput ab-initio approach, *Acta Mater.* 159 (2018) 364–383. doi:10.1016/j.actamat.2018.07.042.
- [21] S. Curtarolo, G.L.W. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, *Nat. Mater.* 12 (2013) 191–201. doi:10.1038/nmat3568.
- [22] G.S. Rohrer, M. Affatigato, M. Backhaus, R.K. Bordia, H.M. Chan, S. Curtarolo, A. Demkov, J.N. Eckstein, K.T. Faber, J.E. Garay, Y. Gogotsi, L. Huang, L.E. Jones, S. V. Kalinin, R.J. Lad, C.G. Levi, J. Levy, J.-P. Maria, L. Mattos, A. Navrotsky, N. Orlovskaya, C. Pantano, J.F. Stebbins, T.S. Sudarshan, T. Tani, K. Scott Weil, Challenges in Ceramic Science: A Report from the Workshop on Emerging Research Areas in Ceramic Science, *J. Am. Ceram. Soc.* 95 (2012) 3699–3712. doi:10.1111/jace.12033.
- [23] O. Levy, G.L.W. Hart, S. Curtarolo, Structure maps for hcp metals from first-principles calculations, *Phys. Rev. B - Condens. Matter Mater. Phys.* 81 (2010). doi:10.1103/PhysRevB.81.174106.
- [24] P. Singh, A. V. Smirnov, D.D. Johnson, Atomic short-range order and incipient long-range order in high-entropy alloys, *Phys. Rev. B - Condens. Matter Mater. Phys.* 91 (2015) 224204. doi:10.1103/PhysRevB.91.224204.
- [25] S. Yin, J. Ding, M. Asta, R.O. Ritchie, Ab initio modeling of the role of local chemical short-range order on the Peierls potential of screw dislocations in body-centered cubic high-entropy alloys, (2019).

- [26] A.A. White, Big data are shaping the future of materials science, *MRS Bull.* 38 (2013) 594–595. doi:10.1557/mrs.2013.187.
- [27] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD), *JOM.* 65 (2013) 1501–1509. doi:10.1007/s11837-013-0755-4.
- [28] S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R. V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, D. Morgan, AFLOW: An automatic framework for high-throughput materials discovery, *Comput. Mater. Sci.* 58 (2012) 218–226. doi:10.1016/J.COMMATSCI.2012.02.005.
- [29] C.E. Calderon, J.J. Plata, C. Toher, C. Oses, O. Levy, M. Fornari, A. Natan, M.J. Mehl, G. Hart, M. Buongiorno Nardelli, S. Curtarolo, The AFLOW standard for high-throughput materials science calculations, *Comput. Mater. Sci.* 108 (2015) 233–238. doi:10.1016/J.COMMATSCI.2015.07.019.
- [30] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.* 1 (2013) 011002. doi:10.1063/1.4812323.
- [31] B.L. DeCost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, *Comput. Mater. Sci.* 110 (2015) 126–133. doi:10.1016/J.COMMATSCI.2015.08.011.
- [32] B.L. DeCost, B. Lei, T. Francis, E.A. Holm, High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel, *Microsc. Microanal.* 25 (2019) 21–29. doi:10.1017/S1431927618015635.
- [33] C. Zhu, H. Wang, K. Kaufmann, K.S. Vecchio, A computer vision approach to study surface deformation of materials, *Meas. Sci. Technol.* 31 (2020) 055602. doi:10.1088/1361-6501/ab65d9.
- [34] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, T.J. Harrington, E. Marin, K.S. Vecchio, Crystal symmetry determination in electron diffraction using machine learning, *Science (80-. )*. 367 (2020) 564–568. doi:10.1126/science.aay3062.
- [35] W.B. Park, J. Chung, J. Jung, K. Sohn, S.P. Singh, M. Pyo, N. Shin, K.S. Sohn, Classification of crystal structure using a convolutional neural network, *IUCrJ.* 4 (2017) 486–494. doi:10.1107/S205225251700714X.
- [36] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K.A. Persson, G. Ceder, A. Jain, Unsupervised word embeddings capture latent knowledge from materials science literature, *Nature.* 571 (2019) 95–98. doi:10.1038/s41586-019-1335-8.
- [37] F. Ren, L. Ward, T. Williams, K.J. Laws, C. Wolverton, J. Hattrick-Simpers, A. Mehta, Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments, *Sci. Adv.* 4 (2018) eaaq1566. doi:10.1126/sciadv.aaq1566.

- [38] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, Accelerated search for materials with targeted properties by adaptive design, *Nat. Commun.* 7 (2016) 11241. doi:10.1038/ncomms11241.
- [39] S. Sanvito, C. Oses, J. Xue, A. Tiwari, M. Zic, T. Archer, P. Tozman, M. Venkatesan, M. Coey, S. Curtarolo, Accelerated discovery of new magnets in the Heusler alloy family, *Sci. Adv.* 3 (2017). doi:10.1126/sciadv.1602241.
- [40] J. Carrete, N. Mingo, S. Wang, S. Curtarolo, Nanograined Half-Heusler Semiconductors as Advanced Thermoelectrics: An Ab Initio High-Throughput Statistical Study, *Adv. Funct. Mater.* 24 (2014) 7427–7432. doi:10.1002/adfm.201401201.
- [41] A.K. Singh, J.H. Montoya, J.M. Gregoire, K.A. Persson, Robust and synthesizable photocatalysts for CO<sub>2</sub> reduction: a data-driven materials discovery, *Nat. Commun.* 10 (2019) 443. doi:10.1038/s41467-019-08356-1.
- [42] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, Universal fragment descriptors for predicting properties of inorganic crystals, *Nat. Commun.* 8 (2017). doi:10.1038/ncomms15679.
- [43] W. Ye, C. Chen, Z. Wang, I.-H. Chu, S.P. Ong, Deep neural networks for accurate predictions of crystal stability, *Nat. Commun.* 9 (2018) 3800. doi:10.1038/s41467-018-06322-x.
- [44] F. Legrain, J. Carrete, A. van Roekeghem, S. Curtarolo, N. Mingo, How Chemical Composition Alone Can Predict Vibrational Free Energies and Entropies of Solids, *Chem. Mater.* 29 (2017) 6220–6227. doi:10.1021/acs.chemmater.7b00789.
- [45] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, *Npj Comput. Mater.* 2 (2016) 16028. doi:10.1038/npjcompumats.2016.28.
- [46] J. Ling, M. Hutchinson, E. Antono, S. Paradiso, B. Meredig, High-Dimensional Materials and Process Optimization Using Data-Driven Experimental Design with Well-Calibrated Uncertainty Estimates, *Integr. Mater. Manuf. Innov.* 6 (2017) 207–217. doi:10.1007/s40192-017-0098-z.
- [47] V. Stanev, C. Oses, A.G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, I. Takeuchi, Machine learning modeling of superconducting critical temperature, *Npj Comput. Mater.* 4 (2018). doi:10.1038/s41524-018-0085-8.
- [48] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, *Phys. Rev. B - Condens. Matter Mater. Phys.* 89 (2014) 1–7. doi:10.1103/PhysRevB.89.094104.
- [49] C. Oses, C. Toher, S. Curtarolo, Data-driven design of inorganic materials with the Automatic Flow Framework for Materials Discovery, *MRS Bull.* 43 (2018) 670–675. doi:10.1557/mrs.2018.207.

- [50] N. De Leon, X.X. Yu, H. Yu, C.R. Weinberger, G.B. Thompson, Bonding effects on the slip differences in the B1 monocarbides, *Phys. Rev. Lett.* 114 (2015). doi:10.1103/PhysRevLett.114.165502.
- [51] E. Wuchina, E. Opila, M. Opeka, W. Fahrenholtz, I. Talmy, UHTCs: ultra-high temperature ceramics for extreme environment applications, *Electrochem. Soc. Interface.* 16 (2007) 30–36.
- [52] S.H. Jhi, J. Ihm, S.G. Loule, M.L. Cohen, Electronic mechanism of hardness enhancement in transition-metal carbonitrides, *Nature.* 399 (1999) 132–134. doi:10.1038/20148.
- [53] L.E. Toth, *Transition Metal Carbides and Nitrides*, Academic Press, New York, 1971.
- [54] M. Kavitha, G. Sudha Priyanga, R. Rajeswarapalanichamy, K. Iyakutti, Structural stability, electronic, mechanical and superconducting properties of CrC and MoC, *Mater. Chem. Phys.* 169 (2016) 71–81. doi:10.1016/j.matchemphys.2015.11.031.
- [55] K. Balasubramanian, S. V. Khare, D. Gall, Valence electron concentration as an indicator for mechanical properties in rocksalt structure nitrides, carbides and carbonitrides, *Acta Mater.* 152 (2018) 175–185. doi:10.1016/j.actamat.2018.04.033.
- [56] T. Das, S. Deb, A. Mookerjee, Study of electronic structure and elastic properties of transition metal and actinide carbides, *Phys. B Condens. Matter.* 367 (2005) 6–18. doi:10.1016/j.physb.2005.05.041.
- [57] L. Breiman, Random Forests, *Mach. Learn.* 45 (2001) 5–32. doi:10.1023/A:1010933404324.
- [58] C. Strobl, A.L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: Illustrations, sources and a solution, *BMC Bioinformatics.* 8 (2007) 25. doi:10.1186/1471-2105-8-25.
- [59] E.W. Huang, D. Yu, J.W. Yeh, C. Lee, K. An, S.Y. Tu, A study of lattice elasticity from low entropy metals to medium and high entropy alloys, *Scr. Mater.* 101 (2015) 32–35. doi:10.1016/j.scriptamat.2015.01.011.
- [60] R. Caruana, A. Niculescu-Mizil, An empirical comparison of supervised learning algorithms, *ACM Int. Conf. Proceeding Ser.* 148 (2006) 161–168. doi:10.1145/1143844.1143865.
- [61] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830. <http://www.jmlr.org/papers/v12/pedregosa11a> (accessed August 15, 2019).
- [62] wolverton / magpie — Bitbucket, (n.d.). <https://bitbucket.org/wolverton/magpie/src/master/> (accessed November 21, 2019).
- [63] J.O. Andersson, T. Helander, L. Höglund, P. Shi, B. Sundman, Thermo-Calc & DICTRA,

- computational tools for materials science, *Calphad Comput. Coupling Phase Diagrams Thermochem.* 26 (2002) 273–312. doi:10.1016/S0364-5916(02)00037-8.
- [64] J. Ramírez, J.M. Górriz, F. Segovia, R. Chaves, D. Salas-Gonzalez, M. López, I. Álvarez, P. Padilla, Computer aided diagnosis system for the Alzheimer’s disease based on partial least squares and random forest SPECT image classification, *Neurosci. Lett.* 472 (2010) 99–103. doi:10.1016/j.neulet.2010.01.056.
- [65] Y. Qi, Random Forest for Bioinformatics BT - Ensemble Machine Learning: Methods and Applications, in: C. Zhang, Y. Ma (Eds.), Springer US, 2012: pp. 307–323. doi:10.1007/978-1-4419-9326-7\_11.
- [66] R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L.M. Ghiringhelli, SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates, *Phys. Rev. Mater.* 2 (2018) 083802. doi:10.1103/PhysRevMaterials.2.083802.
- [67] A. Altmann, L. Toloşi, O. Sander, T. Lengauer, Permutation importance: a corrected feature importance measure, *Bioinformatics.* 26 (2010) 1340–1347. doi:10.1093/bioinformatics/btq134.
- [68] C. Toher, C. Oses, D. Hicks, E. Gossett, F. Rose, P. Nath, D. Usanmaz, D.C. Ford, E. Perim, C.E. Calderon, J.J. Plata, Y. Lederer, M. Jahnátek, W. Setyawan, S. Wang, J. Xue, K. Rasch, R. V. Chepulskaa, R.H. Taylor, G. Gomez, H. Shi, A.R. Supka, R.A.R. Al Orabi, P. Gopal, F.T. Cerasoli, L. Liyanage, H. Wang, I. Siloi, L.A. Agapito, C. Nyshadham, G.L.W. Hart, J. Carrete, F. Legrain, N. Mingo, E. Zurek, O. Isayev, A. Tropsha, S. Sanvito, R.M. Hanson, I. Takeuchi, M.J. Mehl, A.N. Kolmogorov, K. Yang, P. D’Amico, A. Calzolari, M. Costa, R. De Gennaro, M.B. Nardelli, M. Fornari, O. Levy, S. Curtarolo, The AFLOW Fleet for Materials Discovery, in: *Handb. Mater. Model.*, Springer International Publishing, Cham, 2018: pp. 1–28. doi:10.1007/978-3-319-42913-7\_63-1.

## Chapter 3 Searching for high entropy alloys: A machine learning approach

Kevin Kaufmann and Kenneth S. Vecchio\*

Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA

\*Corresponding author

### Abstract

For the past decade, considerable research effort has been devoted toward computationally identifying and experimentally verifying single phase, high-entropy systems. However, predicting the resultant crystal structure(s) “in silico” remains a major challenge. Previous studies have primarily used density functional theory to obtain correlated parameters and fit them to existing data, but this is impractical given the extensive regions of unexplored composition space and considerable computational cost. A rapidly developing area of materials science is the application of machine learning to accelerate materials discovery and reduce computational and experimental costs. Machine learning has inherent advantages over traditional modeling owing to its flexibility as new data becomes available and its rapid ability to construct relationships between input data and target outputs. In this article, we propose a novel high-throughput approach, called “ML-HEA”, for coupling thermodynamic and chemical features with a random forest machine learning model for predicting the solid solution forming ability. The model can be a primary tool or integrated into existing alloy discovery workflows. The ML-HEA method is validated by comparing the results with reliable experimental data for binary, ternary, quaternary, and quinary

systems. Comparison to other modeling approaches, including CALPHAD and the LTVC model, are also made to assess the performance of the machine learning model on labeled and unlabeled data. The uncertainty of the model in predicting the resultant phase of each composition is explored via the output of individual predictor trees. Importantly, the developed model can be immediately applied to explore material space in an unconstrained manner and is readily updated to reflect the results of new experiments.

### **3.1 Introduction**

Alloying has traditionally been performed using one principal element with relatively small amounts of other elements to confer desirable material properties. Recently, research efforts have been directed toward a new alloying strategy combining significant atomic fractions of multiple elements, thus opening a largely unexplored composition space [1]. Researchers commonly refer to this class of materials as complex concentrated alloys, multi-principal element alloys, or high-entropy alloys (HEAs). High entropy alloys typically contain five or more elements to maximize the configurational entropy and improve the stability of the highly-disordered, fully-homogeneous, single-phase solid solution. The intrinsic properties of an HEA are highly dependent on the resultant phase(s) [2,3]. The most common microstructures for high entropy alloys to form are either multi-phase, a single face-centered cubic (fcc) phase, or a single body-centered cubic (bcc) phase, while other single-phase crystal structures occur very rarely [2]. When an HEA does form a single-phase solid solution, the bcc phase is favored owing to its ability to accommodate larger ranges of atomic size in the same lattice [4]. HEAs that exist as a single, highly-disordered, crystalline phase have been of particular research interest [5–11], since they often exhibit the most desirable properties [3,12–18].

Determining, *a priori*, which compositions are most likely to form a single-phase solid solution has remained a major challenge in the community. The original hypothesis that many multi-component systems would be entropically stabilized into a single-phase solid solution has proven untrue [19]. This has launched exhaustive searches for models and screening tools to better understand HEA phase formation [20]. Several features, including enthalpy of mixing, atomic size ratios, intrinsic strain, and valence electron configuration, have been demonstrated to provide insight into HEA phase formation when fit to the limited experimental data [21–23]. However, many of these parameters are derived from the Hume-Rothery rules, which are expectedly limited by their origination from observing patterns in a small set of binary alloys [24]. Phase diagram calculations, often combined with other rules and models, have also been previously applied [6,25,26], but the underlying databases lack significant experimental underpinnings in this vastly unexplored space. Recent high-throughput computational materials design strategies combine thermodynamic and electronic-structure methods in density functional theory (DFT) with data mining capabilities to more quickly evaluate material compositions for novel properties [5,27,28]. While these recent DFT-based methods are high throughput in comparison to other ab-initio efforts, this method can require 100s of hours of computation per composition. Furthermore, DFT calculations are impractical to deal with the large simulation cells required and the increased uncertainty surrounding the d-orbitals of transition metal atoms commonly found in HEAs [29]. The last challenge is adapting these models as new compositions are found that disagree with the predicted results. Thoroughly searching this new composition space, conservatively estimated to comprise hundreds of billions of new alloys [1], is simply not feasible with these approaches, especially given the time required to update these models to reflect new experimental results. Herein, we propose applying machine learning (ML), combined with thermodynamic data

from ThermoCalc [30] and composition-based features [31], as a powerful tool in the rapid search for single-phase solid solution HEAs.

Many aspects of materials science that have been challenging to model or difficult to translate into code have benefitted from the recent adoption of machine learning [32–41]. These tools have provided useful models in the search for new bulk metallic glasses [42], shape memory alloys [43], photocatalysts for CO<sub>2</sub> reduction [44], and high entropy ceramics [45]. The successful application of ML in complex material space has also motivated recent work utilizing ML for the phase prediction of HEAs [46,47]. In comparison to the previous studies [46,47], this work employs the more interpretable random forest model and provides significantly more initial information about each composition for the machine learning model to determine the most important parameters for mapping each composition to its most-likely phase formation. This work also provides an uncertainty metric by tallying and reporting the percentage of predictor trees that “vote” for each of the output phases (multi-phase, face-centered cubic, or body-centered cubic) for each input composition. This novel, high-throughput method — called “ML-HEA”—is validated by comparing the results with the most reliable experimental data for binary, ternary, quaternary, and quinary systems as well as its performance compared to thermodynamic (CALPHAD) and state-of-the-art DFT (LTVc [28]) based approaches for determining the expected phase. The ML-HEA method provides an opportunity to search materials space in a rapid and unconstrained manner without concern for the combinatorial explosion of higher-order compositions (ternaries, quaternaries, quinary, etc.) [48]. In fact, modern hardware can evaluate each composition using our machine learning framework in under a millisecond. Thoroughly calculating all alloys in this space in one pass using an “off the shelf” computer may still be intractable due to time constraints, but this prediction rate offers a clear advantage compared to DFT performed on modern

supercomputers [42,45,48]. Furthermore, the model can readily be extended to the search for non-equiatomic HEAs. The ML-HEA model can be used as a supplementary model in existing workflows, as a preliminary screening tool to constrain material space before performing more computationally expensive analyses, or as a standalone module for material discovery.

## **3.2 Methods**

### **3.2.1 Material selection**

To provide a fair comparison with the DFT-based results, the material dataset obtained from recent DFT work developing the LTVC model [28] was utilized from time of publication. While there may be more up-to-date results for some of the compositions listed with unknown solid solutions, fitting our machine learning model to these updated results would provide an unfair advantage by means of having more information with which to construct the model. The dataset contains a total of 1,798 unique equiatomic compositions studied by Lederer *et al.* for the construction of a DFT-based model [28]. The 1,798 compositions are comprised of 117 binaries, 441 ternaries, 1,110 quaternaries, and 130 quinary. There is a total of 134 labeled compositions: 117 binaries, 8 quaternaries, and 9 quinary.

### **3.2.2 Model architecture**

Random forests are a combination of decision trees that individually make predictions on each input and the overall prediction determined by a majority voting process [49,50]. This model was selected for its well-balanced approach to bias and variance as well as its performance in other materials science studies where data is often limited [45,48,51]. The random forest classifier was implemented with Scikit-learn [52]. Model hyper-parameters are selected via an exhaustive 5-fold cross-validated grid search using the following parameters: number of tree predictors in range 11 to 111 in steps of 10, the number of features considered at each split, and scoring based on gini (a

measure of impurity) or entropy (a measure of information gain) as model performance criterion. Other than the number of decision trees, which can be infinite, this searches over all possible hyperparameter combinations. To avoid the effects of collinearity (highly correlated features) and dimensionality, 5-fold cross validated recursive feature elimination (RFECV) is performed before the grid search through the hyperparameters [53]. Each fold in the 5-fold cross validation was scored using prediction accuracy on the test fold. Eliminating the low-performance features with RFECV can improve the model and its interpretability (e.g. feature ranking). The best performing hyperparameters were selected to fit a model using the entire training set, with bootstrap aggregation applying the 632+ rule [54] (each decision tree is built using a random subset containing approximately 2/3 of the data) [49], to maximize the amount of information available for making future predictions; similar to what has been done in other materials science works using random forest [42,55].

### **3.2.3 From chemistry to features**

Each composition was converted to numerical representation that relates to the essential chemistry, physics, and thermodynamics of each material in the data set. The features utilized in this work should not be considered an exhaustive list, but instead a step toward creating a synergistic set of features that capture a significant amount of information from chemistry and experimentally robust thermodynamics. The 108 compositional features, detailed in the Supplementary Information, are a subset of the general machine learning framework demonstrated previously to perform well on diverse material problems [31]. These chemical features are augmented with 244 thermodynamic features regarding the number of phases and phase fractions in 50 Kelvin steps as well as the liquidus temperature, solidus temperature, and the single phase temperature range for bcc, fcc, and hexagonal from the ThermoCalc Software SSOL6 database

version 6.1 [30]. The single phase temperature range was defined as the difference between the formation temperature of the first solid phase present at the solidus and the formation temperature of the second solid phase. The features for each chemistry studied are provided in the GitHub repository containing the python code, training data, and new compositions.

### **3.2.4 Evaluating the random forest algorithm**

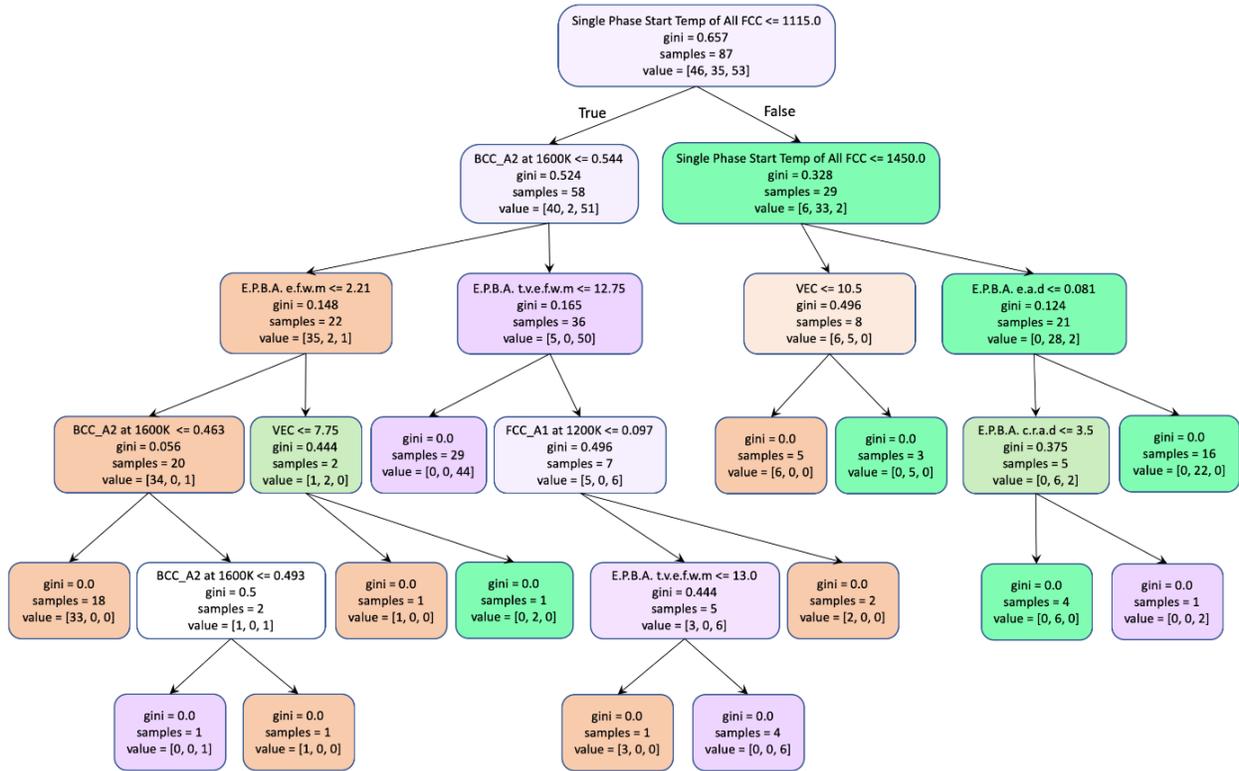
The random forest model was first analyzed to improve scientific understanding of how the machine learning model translates the compositional data to predicted phases. The feature importance was extracted for the model using the “rfpimp” package in Python (available at <https://github.com/parrt/random-forest-importances>, last access: 15 August 2019). The feature importance was ranked on the permutation importance, which directly measures importance by observing the effect on model accuracy by randomly permuting the values of each feature [56]. That is to say, the permutation importance is measuring the impact of swapping the value of a selected feature from one composition with the value from a different composition. This method has recently been introduced as an improvement to the mean decrease in impurity metric [57]. The uncertainty of the model is extracted by tallying and reporting the percentage of predictor trees that “vote” for each of the output phases (none/neither, face-centered cubic, or body-centered cubic) for each input composition.

## **3.3 Results**

### **3.3.1 Random forest model**

The ML-HEA machine learning model is fit with the most current and reliable data [28] for 117 binaries, 8 quaternaries, and 9 quinary. This dataset was selected in order to allow for a fair comparison with a DFT-based model constructed and assessed using the same materials (LTVC [28]). The resultant random forest model is first analyzed to determine the

hyperparameters and chemical features that resulted in the most effective model in 5-fold cross-validation. The best performing set of hyperparameters was 11 decision trees with 3 features made available at each split from the available 13 features remaining after recursive feature elimination (Supplementary Figure 6). A statistical analysis of the RFECV results was performed by executing the feature selection process one hundred times and recording each score. The average accuracy in combination with the first and second standard deviation are plotted in Supplementary Figure 7. The observed standard deviation is observed to always be less than 3.5%. The use of the 13 identified features from Supplementary Figure 6 appears reasonable given these results. With increased data, it is possible that the number of optimal features identified may change. Fundamentally, this approach combines many individual decision trees (weak learners), where each tree is a non-parametric supervised learning method, into a large number of “voting” trees that individually predict the value or class of a target descriptor and democratically decide the most likely result (strong learners). Each tree learns simple decision rules inferred from the available features and with the goal of maximizing classification accuracy. An example decision tree from the fitted model is shown in Fig. 1.



**Figure 3.1 Example decision tree from the fitted model.** Each new alloy composition is assessed by each fitted tree in the random forest model. The decision tree makes a structure prediction by starting at the top of the tree and answering the true or false question at each node until a leaf is reached. A random forest consists of at least two such decision trees. Color scale is orange for neither fcc or bcc single-phase, green for fcc, and purple for bcc. Value order is [Neither single-phase fcc nor bcc, fcc, bcc].

Table 1 shows the top ten features and their importance rank after fitting the random forest model to the 134 compositions with known structure. The list of features contains a mix of information from chemical features and CALPHAD that are both logical and in good agreement with previous studies into predicting HEA's resultant structures [5,6,21–23,25–28]. In contrast to these works, the ML-HEA method can incorporate any number of initial input features from multiple sources, autonomously determine the most relevant features for solving the problem, and subsequently construct non-linear relationships between these features for maximizing accuracy

at structure prediction. Furthermore, the model is readily able to incorporate new features or to include new experimental results, each time updating the decision trees and feature importance.

**Table 3.1 Ranking feature importance.** The top ten features ranked by their permutation importance. The fwm(x) and range(x) correspond to the fraction-weighted mean and range of the feature for each compound. An \* denotes a feature from CALPHAD.

Feature rank	Feature	Feature rank	Feature
1	fwm (unfilled total valence electrons)	6	*fcc fraction at 1200K
2	*bcc fraction at 1600K	7	fwm (covalent radii)
3	*Single phase start temperature of fcc	8	range (electronegativity)
4	range (covalent radii)	9	Valence Electron Concentration (VEC)
5	fwm (electronegativity)	10	*Single phase start temperature of bcc

### 3.3.2 Model performance

First, the model's performance on the binaries, quaternaries, and quaternaries that make up the labeled training set is assessed (Table 2; known solid solution column). There are no ternaries with known solid solution in the dataset and therefore the training set. Since the model was fit with bootstrap aggregation, each decision tree was constructed using  $\sim 2/3$  of the training data (refer to section 2.1 for further information). Thus, it is not a trivial result that ML-HEA correctly predicted the solid solutions for all 134 alloys in the training set. Moreover, since all 117 of the binaries in the dataset have known phase formation, it is possible to make a direct accuracy comparison with CALPHAD and the LTVC model. While ML-HEA correctly identifies the solid solution for each of the binaries, CALPHAD and LTVC only score 94% and 87.2% respectively. Supplementary Table 2 in the Appendix lists the ML-HEA, CALPHAD, LTVC, and experimental results for each of the binary compositions.

Second, the predictions from ML-HEA method can be compared to the predictions made using CALPHAD or the LTVC model on all 1,798 alloys (Table 2; CALPHAD and LTVC columns). The ML-HEA model's predictions for the ternary compositions agree with CALPHAD and LTVC, 82.1% and 63.3% of the time, respectively. Supplementary Table 3 in the Appendix lists the ML-HEA, CALPHAD, LTVC, and experimental results for each of the ternary compositions. Within the quaternaries (Supplementary Table 4 in the Appendix) and quinary compositions (Supplementary Table 5 in the Appendix), both ML-HEA and LTVC achieve 100% accuracy on the known data (17 samples total); however, the two models only agree with each other 62.2% of the time for quaternaries and 72.3% of the time for quinary compositions. It is reasonable to speculate that the small amount of experimental data available to fit these two models is the primary reason for disagreement. Until much of this phase space is solved, it is impossible to ascertain which model, if either since there are three available options, is correct in their differing predictions of these alloys. As this phase space is explored further, it will be interesting to study how and why a model was (in)correct and will likely yield new design rules and insights, as it has done in other works [58]. However, the ML-HEA model has the advantage that each new experimental result can be added directly to the training set and the model refit with the goal of improving its prediction capabilities. The ease with which this is performed allows for new results to be added individually or in batches, and the model refitting to occur before new compositions are selected. This ability for the ML-HEA model to be rapidly updated and return new predictions is likely to make it a valuable tool in the search for new HEAs.

### **3.3.3 Gauging model certainty**

The ML-HEA approach has an advantage over CALPHAD and DFT-based methods in that we can estimate and quantify (un)certainty in the model using the output from individual decision

trees. Each decision tree that composes the ML-HEA model makes a prediction separately from each of the other decision trees. The percentage of trees voting for each solid solution class is recorded and used as a metric of how certain the model is of its answer. Refer to Supplementary Table 6 in the Appendix for the decision tree prediction percentages by composition. Table 3 (part a) details the average certainty of the model in predicting each of the classes of solid solutions for the binaries, quaternaries, and quaternaries for which the answer is known. There were no ternaries with known solid solutions in the dataset. It has already been established that the model is able to predict the correct solid solution with 100% accuracy for the 134 alloys in the training set. In addition, ML-HEA displays a high degree of average certainty, typically greater than 90%, in its predictions. It is not surprising that the fcc class contains the most uncertainty provided the small amount of labeled data and the unlikelihood of multi-component materials adopting an fcc structure. Table 3 (part b) details the average certainty of the model in its predictions for compositions outside the training set. In addition to the good agreement with CALPHAD and LTVC for these compositions, the model also maintains a reasonably high degree of average certainty (~75%). These two factors combined suggest the model will perform reasonably well as a standalone tool or in conjunction with other candidate alloy screening methods.

**Table 3.2 ML-HEA algorithm agreement with known data and other models.** The predictions from the ML-HEA model are first compared with the 134 alloys that have known solid solution. The ML-HEA predictions for all 1,798 alloys are then compared with the predictions from CALPHAD and the LVTC model [28]. The percentage of ML-HEA predictions that are in agreement is reported in parentheses for each comparison.

Material Systems	Known solid solution (Neither, FCC, or BCC)	CALPHAD	LVTC
Binaries	117 of 117 (100%)	110 of 117 (94%)	102 of 117 (87.2%)
Ternaries	N/A	362 of 441 (82.1%)	279 of 441 (63.3%)
Quaternaries	8 of 8 (100%)	N/A	690 of 1110 (62.2%)
Quinaries	9 of 9 (100%)	N/A	94 of 130 (72.3%)

**Table 3.3 Quantifying uncertainty in the ML-HEA model.** (a) The average percentage of decision trees that voted for the correct class (certainty) for materials with known phase formation. (b) The average percentage of decision trees that voted for the correct class (certainty) for materials with unknown phase formation. “N/A” means no materials satisfied the known (a) or unknown (b) criteria and were subsequently predicted to belong to the class. See Supplementary Table 6 in the Appendix for composition specific vote tallies.

**a**

	<b>Neither FCC or BCC Single Phase</b>	<b>FCC</b>	<b>BCC</b>
<b>Binaries</b>	48 out of 117 with 90.5% certainty	36 out of 117 with 89.6% certainty	33 out of 117 with 91.7% certainty
<b>Ternaries</b>	N/A	N/A	N/A
<b>Quaternaries</b>	N/A	3 out of 8 with 87.9% certainty	5 out of 8 with 98.2% certainty
<b>Quinaries</b>	N/A	1 out of 9 with 100% certainty	8 out of 9 with 96.6% certainty

**b**

	<b>Neither FCC or BCC Single Phase</b>	<b>FCC</b>	<b>BCC</b>
<b>Binaries</b>	N/A	N/A	N/A
<b>Ternaries</b>	206 out of 441 with 75.0% certainty	123 out of 441 with 82.9% certainty	112 out of 441 with 80.7% certainty
<b>Quaternaries</b>	674 out of 1102 with 73.8% certainty	232 out of 1102 with 76.0% certainty	out of 1102 with 76.4% certainty
<b>Quinaries</b>	39 out of 121 with 67.4% certainty	N/A	82 out of 121 with 74.4% certainty

### 3.4 Conclusions

Herein, a data-driven workflow for predicting whether a given composition will form a single- or multi-phase solid solution is presented. The result is an adaptive learning model that can progressively improve through systematic infusion of new data as the high entropy alloy space continues to be explored. Notably, the uncertainty of the model can be quantified by recording

each decision tree's vote and reporting the percentage of votes for each class. The resultant ML-HEA model can be used as a standalone module or as a preliminary tool for reducing the list to the most likely compositions for further, more computationally expensive analysis.

An in-depth analysis of the model, particularly the decision trees, the remaining features after recursive feature elimination, and the importance rank of those features, reveals the model is using similar parameters to those found useful in prior modeling works. It also reveals the opportunity for synergy between chemical features and information from CALPHAD.

The ML-HEA model identifies the correct lattice for the 134 systems (binary, ternary, quaternary, and quinary) with known phase formation. When compared to the predictions from CALPHAD, the ML-HEA method is in good agreement with the predictions for binary and ternary systems (94% and 82.1%, respectively). The machine learning model's predictions for all 1,798 of the binary, ternary, quaternary, and quinary systems are further corroborated using results from the DFT-based LTVC method (87.2%, 63.3%, 62.2%, and 72.3% agreement, respectively). Cases found to disagree with CALPHAD or LTVC, and later found to be incorrect predictions, could be based in the lack of training data for the ternary compositions as well as the absence of compositions that are not bcc or fcc in the quaternaries and quinaries. This is not to state that CALPHAD or LTVC will always be correct, but is an explanation of the likely root causes of our model's disagreement with other approaches.

Like most other models including CALPHAD and LTVC, our approach primarily returns a discrete answer that may suggest to the user that the model is 100% confident in the output, while in reality the decision tree voting may have been decided by a small fraction. For this reason, the user should also review the decision tree voting records provided. When searching through a

continuous composition space, it is also possible that asymptotic changes in the predictions could be observed.

As further data is collected, the ML-HEA model can continuously be updated with the intent of improving accuracy. Of the 1,110 quaternary and 130 quinary systems, the ML-HEA model predicts 436 quaternaries and 91 quinary systems will form an fcc or bcc solid solution. This suggests there is ample opportunity for material discovery in this composition space. Furthermore, this same methodology can be applied to explore non-equiatomic composition ranges without concern for the resultant combinatorial explosion.

### **3.5 Acknowledgements**

This work has no direct funding support. The authors would like to thank Dr. Chaoyi Zhu and Daniel Maryanovsky for their assistance with the code development. K. Kaufmann would like to acknowledge support by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program. K. Kaufmann would also like to acknowledge the financial support of the ARCS Foundation, San Diego Chapter. K. Vecchio would like to acknowledge the financial generosity of the Oerlikon Group in support of his research group.

Chapter 3, in full, is a reprint of the material “Searching for High Entropy Alloys: A Machine Learning Approach” as it appears in *Acta Materialia*. K. Kaufmann & K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

### 3.6 References

- [1] D.B. Miracle, High entropy alloys as a bold step forward in alloy development, *Nat. Commun.* 10 (2019) 1805. doi:10.1038/s41467-019-09700-1.
- [2] D.B. Miracle, O.N. Senkov, A critical review of high entropy alloys and related concepts, *Acta Mater.* 122 (2017) 448–511. doi:10.1016/j.actamat.2016.08.081.
- [3] Z. Li, K.G. Pradeep, Y. Deng, D. Raabe, C.C. Tasan, Metastable high-entropy dual-phase alloys overcome the strength–ductility trade-off, *Nature.* 534 (2016) 227–230. doi:10.1038/nature17981.
- [4] B.S. Murty, J.-W. Yeh, S. Ranganathan, P. Bhattacharjee, *High-entropy alloys*, 2nd ed., Elsevier, 2019.
- [5] P. Sarker, T. Harrington, C. Toher, C. Oses, M. Samiee, J.P. Maria, D.W. Brenner, K.S. Vecchio, S. Curtarolo, High-entropy high-hardness metal carbides discovered by entropy descriptors, *Nat. Commun.* 9 (2018) 4980. doi:10.1038/s41467-018-07160-7.
- [6] O.N. Senkov, J.D. Miller, D.B. Miracle, C. Woodward, Accelerated exploration of multi-principal element alloys with solid solution phases, *Nat. Commun.* 6 (2015) 6529. doi:10.1038/ncomms7529.
- [7] B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Microstructural development in equiatomic multicomponent alloys, *Mater. Sci. Eng. A.* 375–377 (2004) 213–218. doi:10.1016/J.MSEA.2003.10.257.
- [8] J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, S.-Y. Chang, Nanostructured High-Entropy Alloys with Multiple Principal Elements: Novel Alloy Design Concepts and Outcomes, *Adv. Eng. Mater.* 6 (2004) 299–303. doi:10.1002/adem.200300567.
- [9] J. Gild, J. Braun, K. Kaufmann, E. Marin, T. Harrington, P. Hopkins, K. Vecchio, J. Luo, A high-entropy silicide: (Mo 0.2 Nb 0.2 Ta 0.2 Ti 0.2 W 0.2 )Si 2, *J. Mater.* (2019). doi:10.1016/j.jmat.2019.03.002.
- [10] C.M. Rost, E. Sachet, T. Borman, A. Moballegh, E.C. Dickey, D. Hou, J.L. Jones, S. Curtarolo, J.-P. Maria, Entropy-stabilized oxides, *Nat. Commun.* 6 (2015) 8485. doi:10.1038/ncomms9485.
- [11] J. Gild, K. Kaufmann, K. Vecchio, J. Luo, Reactive flash spark plasma sintering of high-entropy ultrahigh temperature ceramics, *Scr. Mater.* 170 (2019). doi:10.1016/j.scriptamat.2019.05.039.
- [12] B. Gludovatz, A. Hohenwarter, D. Catoor, E.H. Chang, E.P. George, R.O. Ritchie, A fracture-resistant high-entropy alloy for cryogenic applications, *Science* (80-. ). 345 (2014) 1153–1158. doi:10.1126/science.1254581.
- [13] T.J. Harrington, J. Gild, P. Sarker, C. Toher, C.M. Rost, O.F. Dippo, C. McElfresh, K. Kaufmann, E. Marin, L. Borowski, P.E. Hopkins, J. Luo, S. Curtarolo, D.W. Brenner, K.S.

- Vecchio, Phase stability and mechanical properties of novel high entropy transition metal carbides, *Acta Mater.* 166 (2019). doi:10.1016/j.actamat.2018.12.054.
- [14] X. Lim, Mixed-up metals make for stronger, tougher, stretchier alloys, *Nature.* 533 (2016) 306–307. doi:10.1038/533306a.
- [15] Z. Li, C.C. Tasan, H. Springer, B. Gault, D. Raabe, Interstitial atoms enable joint twinning and transformation induced plasticity in strong and ductile high-entropy alloys, *Sci. Rep.* 7 (2017) 40704. doi:10.1038/srep40704.
- [16] T.-K. Tsao, A.-C. Yeh, C.-M. Kuo, K. Kakehi, H. Murakami, J.-W. Yeh, S.-R. Jian, The High Temperature Tensile and Creep Behaviors of High Entropy Superalloy, *Sci. Rep.* 7 (2017) 12658. doi:10.1038/s41598-017-13026-7.
- [17] O.N. Senkov, G.B. Wilks, J.M. Scott, D.B. Miracle, Mechanical properties of Nb<sub>25</sub>Mo<sub>25</sub>Ta<sub>25</sub>W<sub>25</sub> and V<sub>20</sub>Nb<sub>20</sub>Mo<sub>20</sub>Ta<sub>20</sub>W<sub>20</sub> refractory high entropy alloys, *Intermetallics.* 19 (2011) 698–706. doi:10.1016/J.INTERMET.2011.01.004.
- [18] F. von Rohr, M.J. Winiarski, J. Tao, T. Klimczuk, R.J. Cava, Effect of electron count and chemical complexity in the Ta-Nb-Hf-Zr-Ti high-entropy alloy superconductor., *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) E7144–E7150. doi:10.1073/pnas.1615926113.
- [19] S. GUO, C.T. LIU, Phase stability in high entropy alloys: Formation of solid-solution phase or amorphous phase, *Prog. Nat. Sci. Mater. Int.* 21 (2011) 433–446. doi:10.1016/S1002-0071(12)60080-X.
- [20] M.C. Gao, Design of high-entropy alloys, in: *High-Entropy Alloy. Fundam. Appl.*, Springer International Publishing, 2016: pp. 369–398. doi:10.1007/978-3-319-27013-5\_11.
- [21] Y. Zhang, Y.J. Zhou, J.P. Lin, G.L. Chen, P.K. Liaw, Solid-Solution Phase Formation Rules for Multi-component Alloys, *Adv. Eng. Mater.* 10 (2008) 534–538. doi:10.1002/adem.200700240.
- [22] M.G. Poletti, L. Battezzati, Electronic and thermodynamic criteria for the occurrence of high entropy alloys in metallic systems, *Acta Mater.* 75 (2014) 297–306. doi:10.1016/J.ACTAMAT.2014.04.033.
- [23] Y.F. Ye, C.T. Liu, Y. Yang, A geometric model for intrinsic residual strain and phase stability in high entropy alloys, *Acta Mater.* 94 (2015) 152–161. doi:10.1016/J.ACTAMAT.2015.04.051.
- [24] W. Hume-Rothery, *Atomic theory for students of metallurgy*, Institute of Metals, London, 1952.
- [25] M. Gao, D. Alman, Searching for Next Single-Phase High-Entropy Alloy Compositions, *Entropy.* 15 (2013) 4504–4519. doi:10.3390/e15104504.
- [26] F. Zhang, C. Zhang, S.L. Chen, J. Zhu, W.S. Cao, U.R. Kattner, An understanding of high entropy alloys from phase diagram calculations, *Calphad.* 45 (2014) 1–10. doi:10.1016/J.CALPHAD.2013.10.006.

- [27] S. Curtarolo, G.L.W. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, *Nat. Mater.* 12 (2013) 191–201. doi:10.1038/nmat3568.
- [28] Y. Lederer, C. Toher, K.S. Vecchio, S. Curtarolo, The search for high entropy alloys: A high-throughput ab-initio approach, *Acta Mater.* 159 (2018) 364–383. doi:10.1016/j.actamat.2018.07.042.
- [29] R. Feng, P.K. Liaw, M.C. Gao, M. Widom, First-principles prediction of high-entropy-alloy stability, *Npj Comput. Mater.* 3 (2017). doi:10.1038/s41524-017-0049-4.
- [30] J.O. Andersson, T. Helander, L. Höglund, P. Shi, B. Sundman, Thermo-Calc & DICTRA, computational tools for materials science, *Calphad Comput. Coupling Phase Diagrams Thermochem.* 26 (2002) 273–312. doi:10.1016/S0364-5916(02)00037-8.
- [31] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, *Npj Comput. Mater.* 2 (2016) 16028. doi:10.1038/npjcompumats.2016.28.
- [32] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: Recent applications and prospects, *Npj Comput. Mater.* 3 (2017). doi:10.1038/s41524-017-0056-5.
- [33] B.L. DeCost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, *Comput. Mater. Sci.* 110 (2015) 126–133. doi:10.1016/J.COMMATSCI.2015.08.011.
- [34] B.L. DeCost, T. Francis, E.A. Holm, High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel, (2018). <http://arxiv.org/abs/1805.08693> (accessed August 29, 2018).
- [35] C. Zhu, H. Wang, K. Kaufmann, K.S. Vecchio, A computer vision approach to study surface deformation of materials, *Meas. Sci. Technol.* 31 (2020) 055602. doi:10.1088/1361-6501/ab65d9.
- [36] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, T.J. Harrington, E. Marin, K.S. Vecchio, Crystal symmetry determination in electron diffraction using machine learning, *Science* (80-. ). 367 (2020) 564–568. doi:10.1126/science.aay3062.
- [37] K. Kaufmann, C. Zhu, A.S. Rosengarten, K.S. Vecchio, Deep Neural Network Enabled Space Group Identification in EBSD, *Microsc. Microanal.* (2020) 1–11. doi:10.1017/s1431927620001506.
- [38] W.B. Park, J. Chung, J. Jung, K. Sohn, S.P. Singh, M. Pyo, N. Shin, K.S. Sohn, Classification of crystal structure using a convolutional neural network, *IUCrJ.* 4 (2017) 486–494. doi:10.1107/S205225251700714X.
- [39] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K.A. Persson, G. Ceder, A. Jain, Unsupervised word embeddings capture latent knowledge from materials science literature, *Nature.* 571 (2019) 95–98. doi:10.1038/s41586-019-1335-8.

- [40] J. Schmidt, M.R.G. Marques, S. Botti, M.A.L. Marques, Recent advances and applications of machine learning in solid-state materials science, *Npj Comput. Mater.* 5 (2019) 1–36. doi:10.1038/s41524-019-0221-0.
- [41] F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N.T.P. Hartono, S. Ramasamy, B.L. DeCost, S.I.P. Tian, G. Romano, A. Gilad Kusne, T. Buonassisi, Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks, *Npj Comput. Mater.* 5 (2019). doi:10.1038/s41524-019-0196-x.
- [42] F. Ren, L. Ward, T. Williams, K.J. Laws, C. Wolverton, J. Hattrick-Simpers, A. Mehta, Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments, *Sci. Adv.* 4 (2018) eaaq1566. doi:10.1126/sciadv.aaq1566.
- [43] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, Accelerated search for materials with targeted properties by adaptive design, *Nat. Commun.* 7 (2016) 11241. doi:10.1038/ncomms11241.
- [44] A.K. Singh, J.H. Montoya, J.M. Gregoire, K.A. Persson, Robust and synthesizable photocatalysts for CO<sub>2</sub> reduction: a data-driven materials discovery, *Nat. Commun.* 10 (2019) 443. doi:10.1038/s41467-019-08356-1.
- [45] K. Kaufmann, D. Maryanovsky, W.M. Mellor, C. Zhu, A.S. Rosengarten, T.J. Harrington, C. Oses, C. Toher, S. Curtarolo, K.S. Vecchio, Discovery of high-entropy ceramics via machine learning, *Npj Comput. Mater.* 6 (2020) 42. doi:10.1038/s41524-020-0317-6.
- [46] W. Huang, P. Martin, H.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, *Acta Mater.* 169 (2019) 225–236. doi:10.1016/j.actamat.2019.03.012.
- [47] A. Agarwal, A.K.P. Rao, Artificial Intelligence Predicts Body-Centered-Cubic and Face-Centered-Cubic Phases in High-Entropy Alloys, *JOM.* 71 (2019) 3424–3432. doi:10.1007/s11837-019-03712-4.
- [48] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, *Phys. Rev. B - Condens. Matter Mater. Phys.* 89 (2014) 1–7. doi:10.1103/PhysRevB.89.094104.
- [49] L. Breiman, Random Forests, *Mach. Learn.* 45 (2001) 5–32. doi:10.1023/A:1010933404324.
- [50] R. Caruana, A. Niculescu-Mizil, An empirical comparison of supervised learning algorithms, *ACM Int. Conf. Proceeding Ser.* 148 (2006) 161–168. doi:10.1145/1143844.1143865.
- [51] A.O. Oliynyk, E. Antono, T.D. Sparks, L. Ghadbeigi, M.W. Gaultois, B. Meredig, A. Mar, High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds, *Chem. Mater.* 28 (2016) 7324–7331. doi:10.1021/acs.chemmater.6b02724.
- [52] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher,

- M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830. <http://www.jmlr.org/papers/v12/pedregosa11a> (accessed August 15, 2019).
- [53] P.M. Granitto, C. Furlanello, F. Biasioli, F. Gasperi, Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products, *Chemom. Intell. Lab. Syst.* 83 (2006) 83–90. doi:10.1016/j.chemolab.2006.01.007.
- [54] B. Efron, R. Tibshirani, Improvements on cross-validation: The .632+ bootstrap method, *J. Am. Stat. Assoc.* 92 (1997) 548–560. doi:10.1080/01621459.1997.10474007.
- [55] A.O. Oliynyk, E. Antono, T.D. Sparks, L. Ghadbeigi, M.W. Gaultois, B. Meredig, A. Mar, High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds, *Chem. Mater.* 28 (2016) 7324–7331. doi:10.1021/acs.chemmater.6b02724.
- [56] A. Altmann, L. Toloşi, O. Sander, T. Lengauer, Permutation importance: a corrected feature importance measure, *Bioinformatics.* 26 (2010) 1340–1347. doi:10.1093/bioinformatics/btq134.
- [57] C. Strobl, A.L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: Illustrations, sources and a solution, *BMC Bioinformatics.* 8 (2007) 25. doi:10.1186/1471-2105-8-25.
- [58] E.A. Holm, In defense of the black box, *Science* (80-. ). 364 (2019) 26–27. doi:10.1126/science.aax0162.

## Chapter 4 Development of ultrahigh-entropy ceramics with tailored oxidation behavior

William M. Mellor<sup>a†</sup>, Kevin Kaufmann<sup>b†</sup>, Olivia F. Dippo<sup>a</sup>, Samuel D. Figueroa<sup>b</sup>, Grant D. Schrader<sup>b</sup>, Kenneth S. Vecchio<sup>a,b‡</sup>

<sup>a</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA.

<sup>b</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA.

<sup>†</sup>These authors contributed equally. <sup>‡</sup>Corresponding author.

### Abstract

In the last decade, single-phase high-entropy materials have attracted considerable research interest owing to their unexpected existence and unique combinations of properties. Recent development of 5-cation high-entropy carbides (HECs) has demonstrated alluring properties compared to the rule of mixtures and binary carbides. Proposed here is the development of ultrahigh-entropy carbides (UHECs) containing 6+ principal elements with greater combinatorial possibilities. The use of 6+ multi-cation compositions allows for the design of ceramics with further tunable properties, while likely possessing higher orders of entropic stabilization. There are 133 possible carbide compositions containing 6, 7, 8, or 9 refractory metal cations in equiatomic ratios. Candidate selection for fabrication and material testing was accelerated by using a machine learning model that was originally trained to predict the synthesizability of five cation disordered metal carbides. Two compositions from each category of six through eight cations, one containing Cr and one without, plus the one possible nine cation carbide were

fabricated and characterized. The potential for these 6+ cation UHECs as improved materials for oxidative environments is demonstrated by comparing the oxidation performance of a 5- and 7-cation system after 10 minutes at 1973K in air. The oxidation behavior is correlated with Ellingham diagrams, and it is demonstrated that the 7-cation carbide has the ability to form a transitional stable 5+ cation HEC layer as elements preferentially form oxides, which results in significantly improved oxidation resistance.

#### **4.1 Introduction**

In recent years, high-entropy alloys (HEAs) have attracted much attention, with the first publication appearing in 2004 [1]. The novelty of these material systems lies primarily in the potential to drive complex compositions into stable single-phase solid-solution materials through configurational entropy. The high configurational entropy is thought to help combat the formation of thermodynamically competing intermetallic phases [2,3]. Maximum configurational entropy is attained when constituent elements are held in equiatomic amounts in a single phase. The observed entropic effect is thought to be significant, particularly concerning phase formation and stability, when the composition consists of at least five principal alloying elements held in roughly equiatomic amounts [2,3]. Since the inception of this material class, the exact role that entropy plays in the stabilization of the material has remained uncertain [4,5]. The high entropy ceramics community has recognized the calculation of configurational entropy per mol of atoms by considering the metal cations and the light element anions in their distinct atom positions, or ‘sublattices.’ [5–8] However, this method is inconsistent with the historical definition of the high-entropy moniker (i.e.  $S > 1.5R$ ) [9]. Gild *et al.* [10] and Wright *et al.* [7,8] have asserted that the configurational entropy requirements for high-entropy must be satisfied on at least one of the sublattices in ceramics, and Toher *et al.* [4] furthered this argument referring to the anion lattice

as primarily a “spectator species” with regard to high-entropy materials. Given the ongoing discussions in the community, both calculations are presented here: i) entropy calculated per mole of formula units, as well as ii) entropy calculated per mole of atoms. Regardless of the calculation method, the configurational entropy of these novel 6+ cation ceramics is greater than that of any previous 5-cation ceramics and are referred to hereafter as ultrahigh-entropy carbides (UHEC). For example, an equiatomic 9-cation carbide will have a configurational entropy that is more than 30% higher than an equiatomic 5-cation carbide.

Recent interest in high entropy ceramics has yielded a significant number of single-phase materials of varying types (e.g. carbides, nitrides, oxides, and borides) [5–7,11–21]. The carbides of group IVB, VBs, and VIB transition metals are promising candidates for extreme environment applications [22,23] in large part due to their high hardness, excellent corrosion resistance, and high melting temperatures [22]. These materials are subjected to extreme environments to meet the demands of shielding for nuclear reactors, rocket nozzles, and hypersonic vehicle leading edges, where these properties are critical design constraints. As examples, recent work in high-entropy ceramics has demonstrated interesting properties including hardness, [11,16,18,20,21,24] moduli, [15,16,24–26] and thermal conductivity [18,20,27] that exceed rule of mixtures predictions. The properties observed in high-entropy materials have been partially explained through phenomena such as atomic size differences and stabilization of higher symmetry Bravais lattices [16,28,29]. Some ceramic properties are also known to be dependent on valence electron concentration (VEC) [26,30–32]. The ability to extend the number of cations above five metal atoms presents an opportunity for further increasing the configurational entropy, as well as fine-tuning of the VEC and therefore the material properties of this novel class of ceramics. [24,25,30–32] The oxidation behavior of these novel materials is also of considerable interest for enabling

high temperature, oxidative atmosphere applications, such as hypersonic vehicle leading edges. In this work, the oxidation behavior of bulk 5- and 7-cation compositions are compared. These results are linked to, and corroborated by, Ellingham diagrams [33,34]. Additional intriguing combinations of material properties are also likely to be observed as the constituent elements become more diverse.

Despite the unique and desirable combinations of thermo-mechanical properties that high-entropy alloys and ceramics have demonstrated, [16,35–39] there is a significant lack of logical, rapid and effective methodologies for determining suitable candidates for synthesis [2]. Approaches relying upon phase diagram calculations, as well as other rules and models, have demonstrated some success; [40–44] however, the underlying databases lack substantial experimental results for optimal fitting and validation. A descriptor known as the entropy-forming-ability (EFA) has demonstrated significant promise in terms of predicting the synthesizability of high-entropy materials [15]. A high EFA value translates to a relatively small energy penalty for a given composition exhibit disorder, and a low EFA value suggests challenges to achieving single phase random structures (i.e. may favor lattice ordering) [15]. As successful as this approach has proven to be in probing likely candidates within the HEC space, there are significant drawbacks in terms of relying solely upon this approach. *Ab-initio* calculations required for EFA determination are computationally intensive, time-consuming, and subsequently expensive. Machine learning (ML) is gaining traction in multiple fields within materials science, owing to the ability to learn from available data sets, even relatively small ones, to assist in prediction, design, and analysis of materials [45–57]. A machine learning based approach to predicting the EFA descriptor for high-entropy carbides has recently been demonstrated on 5-component carbides and validated by experimental synthesis and follow-up comparison with DFT [12]. The previously

demonstrated machine learning model [12] is used here to rapidly probe the vast compositional space of 6-9 component UHECs in an unconstrained manner.

Herein, a data-driven approach for the discovery of 6-9 component ultrahigh-entropy carbides (UHEC) is presented and utilized. Nine bulk samples that have not previously been reported are systematically studied for phase formation and mechanical properties. The specific mechanical properties studied are the elastic, bulk, and shear moduli. The ability to combine 6+ metal cations into the carbide lattice is a valuable tool for further enhancing chemical tuning for desired properties among diverse applications. As proof of concept, the capacity to tailor the progression of the oxidation process through compositional design is presented. To demonstrate this capability, the oxidation behavior of a 5-component HEC and 7-component UHEC are compared, and the underlying thermodynamic mechanisms elucidated using relevant Ellingham diagrams.

## **4.2 Methods**

### **4.2.1 Entropy forming ability**

The entropy forming ability parameter was introduced by Sarker and Harrington *et al.* [15] to quantify the energy distribution of all possible metastable configurations for a given system. The EFA descriptor for the N-species system is the inverse of the standard deviation of the energy distribution at a finite temperature above the ground state. Thus, the wider the energy distribution of these metastable states, the less likely the system will result in a disordered single-phase material. Refer to Sarker and Harrington *et al.* [15] for further details regarding the EFA formalism.

### 4.2.2 Predicting EFA

Although many machine learning algorithms are available and potentially suitable for this task, random forest was selected for its solution to the bias-variance tradeoff and generally good performance on materials science datasets [47,48,57–59]. Random forest is an ensemble method consisting of decision tree predictors that are individually fit to a subset of the data, called bootstrap aggregating. While each individual decision tree is potentially overfit to its subset of the training data, bootstrap aggregating solves this using the insight that a suitably large number of uncorrelated errors average out to zero [60]. Since each tree learns from different subsets of the data, they are thus fairly uncorrelated. The correlation between decision trees can be further reduced by tuning hyperparameters such as the number of descriptors considered at each split in the decision tree; however, this can come at the cost of an increase in bias since less data is made available to the model, and the fit becomes more specific to the data subset. By averaging the results of each decision tree, the random forest model attempts to achieve the same low-variance as an individual decision tree with the added benefit of low-bias. The model used in this work was trained (fit) using all available DFT calculated EFA data for high-entropy carbides [12,15]. Five-fold cross-validated recursive feature elimination and five-fold cross-validated grid search were utilized to select the optimal feature set and hyperparameters when fit to the 5-metal cation dataset. Refer to the previous work by Kaufmann *et al.* for in-depth details about the model hyperparameter tuning and insight gained from its application to 5-cation HECs [12]. The random forest regressor, feature selection, and hyperparameter tuning were implemented with Scikit-learn [61].

### 4.2.3 Sample preparation

Starting powders of each individual carbide compound precursor (NbC, HfC, TiC, ZrC, VC, TaC, Mo<sub>2</sub>C, WC, and Cr<sub>3</sub>C<sub>2</sub>) (Alfa Aesar, >99.5% purity, -325 mesh) were weighed out in

equiatomic amounts, and hand mixed. The powders were then inserted into a tungsten carbide lined stainless steel jar containing 10mm tungsten carbide milling media to be high energy ball milled (HEBM) in a SPEX 8000D shaker pot high energy ball mill (SPEX CertiPrep, NJ, USA) for 5 hours in 12 g batches. The milling occurred in 10 cycles of 30 minutes with 15 minutes of rest after each individual cycle to allow for thermal cooling and to prevent significant oxide formation. The entire HEBM process is carried out in an argon atmosphere to further minimize possible oxidation of the powders. Following the HEBM, the sample powders were encapsulated within 20mm graphite dies lined with graphite foil on all sides and readied for spark plasma sintering (SPS). The samples were then spark plasma sintered using a Thermal Technologies 3000 series SPS (Thermal Technologies, CA, USA). Samples void of chromium carbide were sintered at 2473K, while samples containing chromium carbide were sintered at a temperature of 2073K; a heating rate of 100 K/min was used for both processes. This difference is due to the low melting temperature of the  $\text{Cr}_3\text{C}_2$  precursor powder. All samples were processed with an applied uniaxial load of 20MPa. Samples were brought to desired applied pressure from a starting value of 5 MPa up to 20 MPa at 15 MPa/min. The result of this process is a 20mm diameter pellet approximately 4 mm in thickness. The sample is left to slow cool in a vacuum environment until room temperature is reached.

#### **4.2.4 X-ray diffraction and electron microscopy**

Crystal structure analysis is implemented using a Rigaku Miniflex X-ray Diffractometer with a 1D detector using a step size of  $0.02^\circ$  and 5 degrees per minute scan rate. Copper  $\text{K}_\alpha$  radiation (wavelength  $\lambda = 1.54059 \text{ \AA}$ ) is used for all measurements. The lattice parameter is calculated utilizing both MDI Jade and Match! phase identification software. This is subsequently

utilized to model and create the theoretical diffraction profile utilized in EBSD to identify the high-entropy phases.

Microstructural and elemental analysis is performed using a Thermo Scientific Apreo field emission scanning electron microscope (SEM) equipped with an Oxford X-Max<sup>N</sup> EDS detector and an Oxford Symmetry electron backscatter diffraction (EBSD) detector. A combination of secondary and back-scattered electron detectors are utilized for imaging. EDS scans are conducted to verify multi-length scale homogeneity in the resulting microstructure.

#### **4.2.4 Property testing**

Samples underwent a series of mechanical testing methods to determine modulus properties (elastic, shear, and bulk). Prior to any property measurement, sample density was measured via Archimedes principle. The ratio to theoretical density was computed wherein the lattice parameter was determined from respective x-ray diffraction patterns. Only samples that could be fabricated with >95% density were subjected to property measurements to ensure the observed properties are not influenced by porosity effects. While all seven chosen compositions fabricated could be made single phase, only three were suitable for mechanical testing as they exhibited density >95%.

The acoustic wave speed method was utilized to determine the elastic, bulk, and shear moduli. Using a transducer coupled to a signal processor and oscilloscope, acoustic wave profiles of each sample were collected, and the wave velocity determined. This wave speed was then used to calculate respective modulus values of a sample. This method was repeated across all samples in accordance with ASTM standard E494-15.

#### 4.2.5 Ellingham diagram calculations

An Ellingham diagram is a useful tool to study the temperature dependence of various compounds, usually oxides, via reactions with various metals. In an Ellingham diagram, the Gibbs free energy is plotted as a function of temperature for the reaction between a gaseous species (O, N, S, or F as common examples), a series of metals, and the compounds formed between the gaseous species and the metals. Ellingham diagrams are a thermodynamic calculation that ignore any potential kinetic effects. The free energy curves of a given reaction species typically begin at large negative values at low temperature and typically possess a positive slope with increasing temperature. For the vast majority of Ellingham diagrams in the literature, the diagrams involve reactions between pure metals and the gaseous species, making the chemical potential simply the molar Gibbs free energy of the given metal. In this study, we are concerned with the oxidation reaction of HECs, some containing many more than 5 metal species, each of which is associated with a carbon atom (M-C pairs). Therefore, to obtain a representative Ellingham diagram for the oxidation of the HECs, the chemical potential of the metals in combination with their bonded carbon atoms within a composition is considered. The free energy of a compound is thus given by Equation 1, where  $x$  and  $y$  are the number of cation and anion atoms in a compound molecule, respectively.  $\mu_m$  and  $\mu_o$  are the chemical potential of cation and anion (in their compound form).

$$\Delta G = [\Delta G_{formation} \times (x + y) - (x\mu_M + y\mu_o)]/y \quad \text{Equation 1}$$

The difference between the chemical potential of the pure elements and the elements in the compound form is given by Equation 2.

$$\mu_{M_{compound}} = \mu_{M_{pure}} + \text{mixing entropy} + \text{mixing enthalpy} \quad \text{Equation 2}$$

This Ellingham diagram then can be used to determine the likely sequence of oxidation of the various metallic elements present in a given HEC composition. The chemical potential of each composition was performed using Thermocalc™ software and the SSOL7 thermodynamic database. Once the chemical potential for the monocarbide pairings in a given carbide compound is determined, then the Gibbs Free Energy of complex carbide is determined for the reaction with oxygen. Since many of the metal species involved form multiple possible oxide stoichiometries, the most common, and often lowest free energy curve for each metal species oxidation is shown.

#### **4.2.6 Oxidation Test**

The 5-cation (HfMoNbTaZrC<sub>5</sub>) HEC and 7-cation (HfMoNbTaVWZrC<sub>7</sub>) UHEC were each placed within an ZrO<sub>2</sub> crucible and inserted into a box furnace at a temperature of 1973K and a pressure of 1 atm, where the samples remained for a duration of 10 minutes in air. The samples were removed, mounted, and cross sectioned. The through thickness cross-section of the sample was ground to 4000 grit and polished to 0.05μm colloidal silica and characterized.

### **4.3 Results**

#### **4.3.1 Entropic contributions**

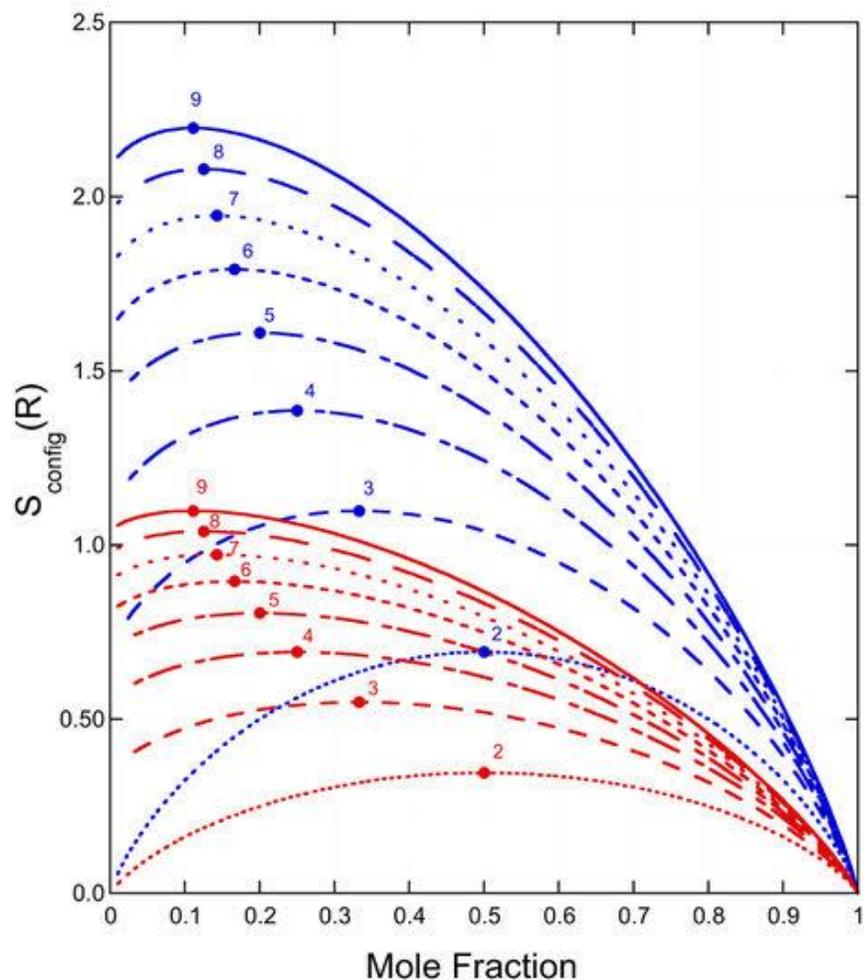
Ideal configurational entropy calculations in the field of high-entropy ceramics have been published using two different sets of units: entropy per mole of formula units and entropy per mole of atoms [5,7,10,15,24]. The ideal configurational entropy for two to nine cation ceramics is plotted per mole of formula units (blue curves) and per mole of atoms (red curves), considering the cation and anion lattices in the rock-salt FCC crystal structure as two distinct interpenetrating sublattices (Figure 1). Treating the structure as being comprised of sublattices allows us to fully capture the contribution of all distinct atom positions to configurational entropy. In Equation 3,  $a^s$  is the number of sites on the  $s$  sublattice, and  $X_i^s$  is the site fraction of element species  $i$  randomly

distributed on the  $s$  sublattice, and in each case, the maximum configurational entropy is achieved when the number of cations is equimolar. For rock-salt structured ceramics, the calculated contribution from the cation lattice is half of the total potential configurational entropy of the system, as this sublattice only occupies one half of all available lattice sites in the system. Moreover, for rock-salt structured ceramics with one anion (i.e. carbon in a carbide), the entropic contribution for the anion sublattice is zero, as it displays long range ordering. The resulting calculation, as shown in Equation 3, will be on a per mole of atoms basis. To instead solve for configurational entropy on a per formula unit basis, Equation 3 can be multiplied by  $\sum_S a^S$  (total number of atoms per formula unit), resulting in Equation 4. Both units of entropy, per mole of atoms and per mole formula units, are in line with the sublattice model of ideal configurational entropy [2,62,63].

$$\Delta S_m^{ideal} = \frac{-R \sum_S \sum_i a^S X_i^S \ln(X_i^S)}{\sum_S a^S} \quad \text{Equation 3}$$

$$\Delta S_m^{ideal} = -R \sum_S \sum_i a^S X_i^S \ln(X_i^S) \quad \text{Equation 4}$$

It should be noted that there is an apparent contradiction in the literature related to the definition of ‘high-entropy’ as being greater than  $1.5R$  [9,64] when ideal configurational entropy is calculated per mole of atoms in high-entropy ceramic materials. As shown in Figure 1, none of the red curves (entropy per mole of atoms) reach the value of  $1.5R$ , and therefore perhaps the metric for ‘high-entropy’ should be redefined for ceramic materials. Regardless of the method of calculation or ‘high-entropy’ definition used, the configurational entropy of the proposed UHECs can be significantly greater than that of any 5-component HECs and highlights the interest in increasing the number of metal cations in the rock-salt carbides beyond five.



**Figure 4.1 Modeling ideal configurational entropy.** Configurational entropy per mole of formula units (Equation 3) is plotted in blue. Configurational entropy per mole of atoms (Equation 4) is plotted in red.

### 4.3.2 EFA predictions

There are 133 possible compositions using equiatomic combinations of group IVB, VB, and VIB transition metal carbides; this is a prohibitively large search space for DFT or trial and error experimental strategies. The choice of sample compositions was accelerated using the entropy forming ability (EFA) determined via a machine learning (ML) model trained on the known DFT-based EFA values for five component HECs [12]. Supplementary Table 7 details all possible combinations of 6, 7, 8, and 9 transition metal cations, their ML predicted EFA, as well

as the sample compositions selected for experiments. Feature importance for the machine learning model is explored in depth in Kaufmann *et al.* [12]. Supplementary Figure 8 shows statistics for each element’s impact on predicted EFA for a composition. Interestingly, individual elements do not appear to play a very significant role in the final EFA; however, the elements Ti, Zr, and W are somewhat less likely to be in compositions with high predicted EFA values. It has been observed that high-entropy carbides and nitrides with lower EFA values can exhibit chromium loss [12,24]. Therefore, where applicable, two sample compositions were chosen, wherein one composition contains chromium, and the other composition is without chromium. There appears to be ample opportunity for material discovery in this space, as many of the 6-9 cation carbides listed have an ML-predicted EFA above the experimentally observed cutoff shown to exist between an EFA value of 45 and 50 in previous works [15,16]. Table 1 lists the compositions selected for fabrication, the number of metal cations in each composition, and their ML-predicted EFA values. The individual compositions were selected starting from the high EFA 5-cation compositions MoNbTaVWC<sub>5</sub> [15] and CrMoNbVWC<sub>5</sub> [12] and successively adding one cation to each composition. The decision of which metal cation to add was guided by the ML-predicted EFA values until converging at all 9 refractory metal cations.

**Table 4.1 Compositions selected for fabrication.** The ML model’s predictions are provided for each composition selected for fabrication. The number of metal cations is listed for each composition. Units: EFA in (eV/atom)<sup>-1</sup>.

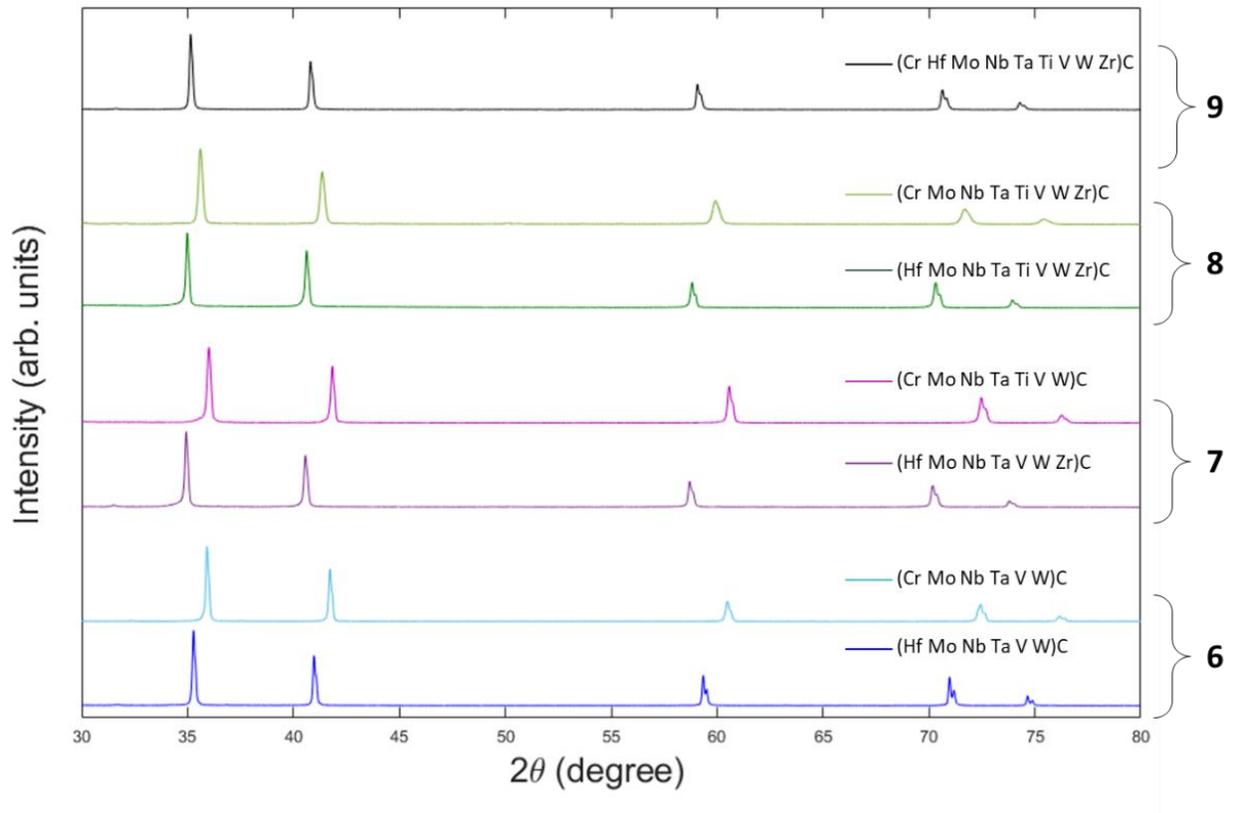
Material	# of Cations	EFA	Material	# of Cations	EFA
CrMoNbTaVWC <sub>6</sub>	6	105	HfMoNbTaTiVWZrC <sub>8</sub>	8	55
CrMoNbTaTiVWC <sub>7</sub>	7	85	CrHfMoNbTaTiVWZrC <sub>9</sub>	9	51
HfMoNbTaVWZrC <sub>7</sub>	7	57	CrMoNbTaTiVWZrC <sub>8</sub>	8	47
HfMoNbTaVWC <sub>6</sub>	6	55			

### 4.3.3 Structure determination and chemistry analysis

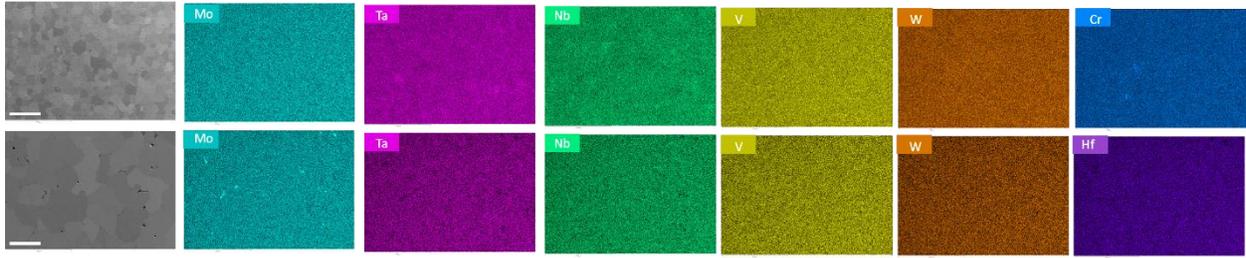
For the first time, bulk ultrahigh-entropy carbide samples that contain six, seven, and nine metal carbides were successfully synthesized. Wang *et al.* previously synthesized three eight-metal carbides, including HfMoNbTaTiVWZrC<sub>8</sub>, which is also synthesized in the present work [65]. CrMoNbTaTiVWZrC<sub>8</sub> is synthesized in this work for the first time, an eight-metal carbide containing all three carbide-forming group 6 elements (Cr, Mo, W). Due to the low melting temperature of Cr<sub>3</sub>C<sub>2</sub>, one of two composition-dependent processing pathways was prescribed (Methods section). Successful fabrication of the rock-salt structure after SPS densification and homogenization was verified via X-ray diffraction (Figure 2). Each of the seven samples exhibit only a single set of FCC peaks of the desired ultrahigh-entropy phase. Even CrMoNbTaTiVWZrC<sub>8</sub> resulted in a single FCC phase in XRD despite the composition having the lowest ML predicted EFA studied experimentally. The ML<sub>EFA</sub> value of 47 does lie near the previously shown experimental cutoff between 45 and 50; however, the exact cutoff remains uncertain, and it is possible the actual EFA is higher than predicted by the ML model for this composition, or that the increased configurational entropy plays a role not captured by the ML model. Additional DFT-based studies on these systems are outside the scope of this work, due to the DFT computation time for 6+ cation structures.

Figure 3 depicts the EDS chemistry maps from two fabricated ultrahigh-entropy carbides each containing six cations. The compositions are CrMoNbTaVWC<sub>6</sub> and HfMoNbTaVWC<sub>6</sub>, and their ML-predicted EFA values are 105 and 55, respectively. The EDS maps for all seven compositions are shown in Supplementary Figures 9-12 owing to the large number of elemental maps. For each sample, the XRD analysis indicates the successful capture of a single-phase rock-

salt structure and EDS analysis indicates that each is chemically homogeneous as well. The SEM micrographs for the fabricated samples show large, nearly equiaxed grains, with some small deviation in grain size, likely owing to the remaining pores. If the samples contained multiple phases, it would be expected that the grains would be less equiaxed and the microstructure much finer as observed in prior works [12,15].



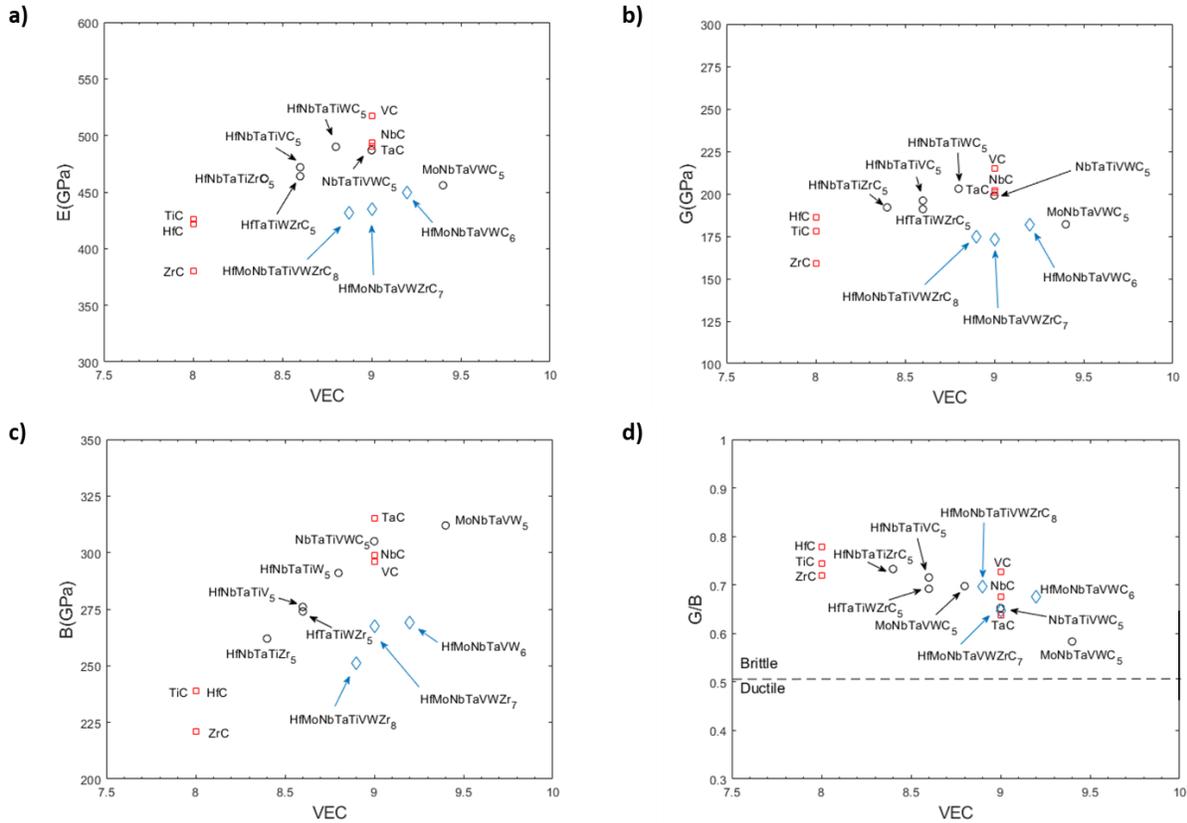
**Figure 4.2 X-ray diffraction patterns for the 6 to 9 cation metal carbides.** The seven compositions studied exhibit only the desired FCC structure peaks. The compositions are grouped by number of cations, indicated to the right of the plot.



**Figure 4.3 Chemistry analysis of the 6-cation UHECs.** The first column is an electron micrograph for each of the synthesized compositions. Columns 2-7 are EDS chemistry maps for each of the six metal cations present in each system. Scale bars 50  $\mu\text{m}$ .

#### 4.3.4 Material Properties

Figure 4 depicts the experimental findings of the corresponding moduli of the experimentally fabricated materials as a function of the valence electron concentration (VEC). The data for binary and 5-cation systems is also reported to serve as a juxtaposition of the effects of additional cations in the system. VEC calculations for each composition were computed from the nominal compositions. The observed trends from the elastic, bulk, and shear moduli of these 6+ cation UHECs are in good agreement with those previously measured for 5-cation HECs as a function of VEC [16]. It is thus reasonable to speculate that other mechanical properties correlated with VEC can also be tuned via the cation ratios [31,32,66].

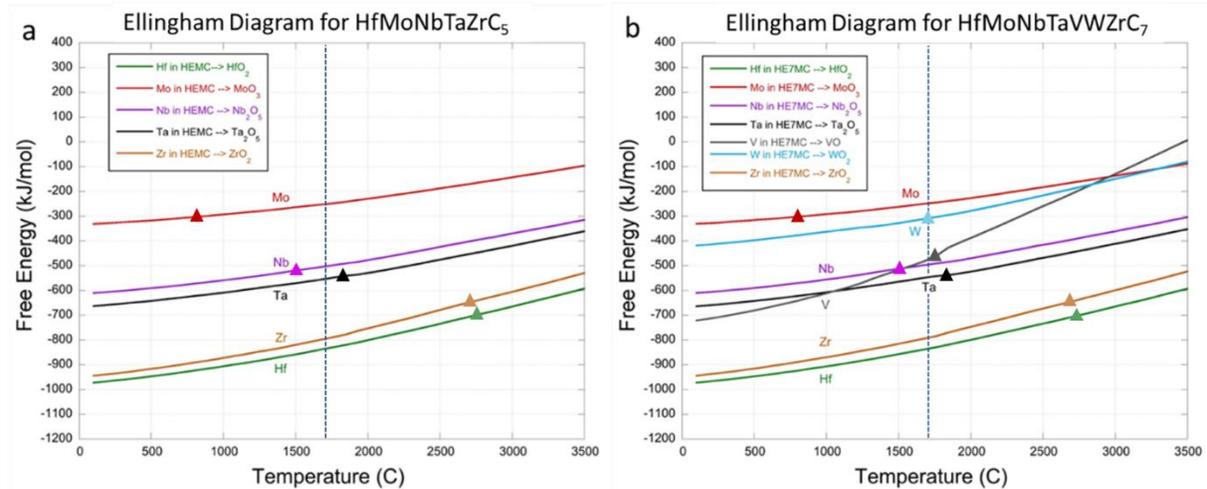


**Figure 4.4 Moduli comparison for carbides.** (a) Elastic, (b) shear, (c) bulk, and (d) Pugh's ratio, each as a function of valence electron concentration (VEC) for 6+ cation UHEC compositions (blue diamonds) 5-cation HEC compositions (black circles) and refractory binary materials (red squares). Properties for the binary materials are from Dipppo et al.[6] and are representative of their respective FCC phase. Properties for the 5-cation HECs are from Harrington et al.[16]. All properties were determined at room temperature.

### 4.3.5 Ellingham diagram

To obtain a representative Ellingham diagram for the oxidation of the HECs and UHECs, the complete composition of the HEC and UHEC materials were used to determine the proper chemical potentials of the M-C compounds prior to calculating the oxidation free energy curves (Methods section). In prior work in high-entropy ceramics, the Ellingham diagrams only considered the chemical potential of metals reacting with a gaseous species [33,34]. Figure 5 details the relative propensity for oxidation of each metal (M) plus carbon pair in each complex carbide composition up to 3500°C. Since many of the metal species involved form multiple

possible oxide stoichiometries, the most common, and often most negative free energy curve for each metal species oxidation is shown in Figure 5. These diagrams can be used to relate the oxidation species sequence expected in the two corresponding sintered samples discussed in the following section. In fact, the Ellingham diagrams can be used to design the UHEC compositions that enable the early oxidation of certain elemental species (e.g. Hf and Zr), while retaining a highly stable HEC composition, for which the remaining entropy is greater than would be present in any 5-cation HEC following oxidation.



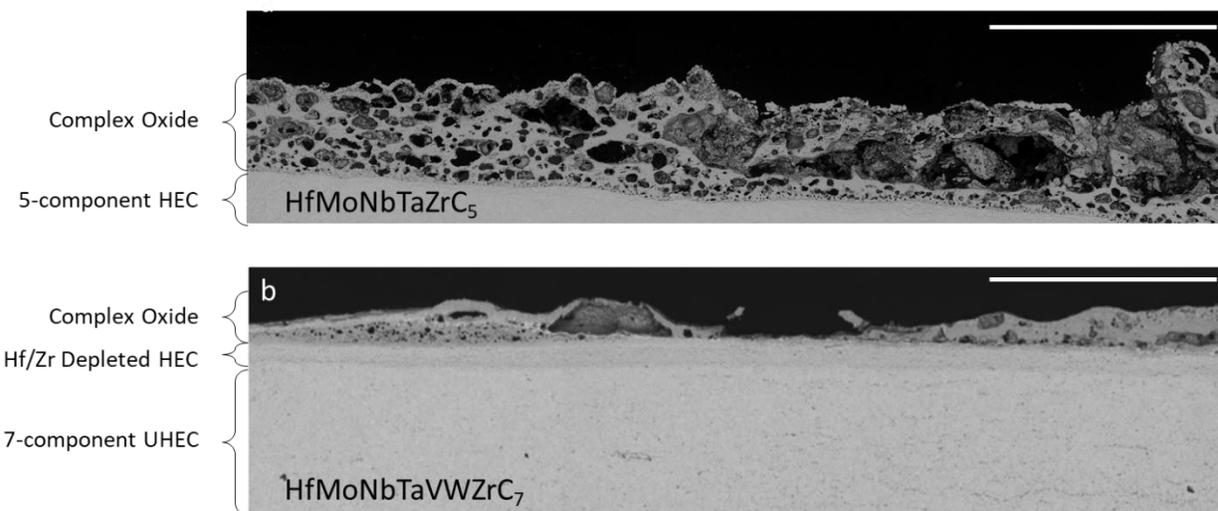
**Figure 4.5 Ellingham diagrams for each composition.** The Ellingham diagrams for (a) HfMoNbTaZrC<sub>5</sub> and (b) HfMoNbTaVWZrC<sub>7</sub>. The Gibbs Free Energy was calculated after determining the chemical potential of each species in the respective carbide compound. A larger (i.e. more negative) free energy indicates a preferred reaction HEMC compared to a M-C pairing with a smaller free energy at the same temperature.

#### 4.3.6 Oxidation analysis

Two sample compositions, 5-cation HfMoNbTaZrC<sub>5</sub> and 7-cation HfMoNbTaVWZrC<sub>7</sub>, were chosen using the analysis of the Ellingham diagrams to demonstrate the effects of adding cations with low free energies of oxide formation at high temperature (i.e. VC and WC) to high-

entropy carbides. The group IV refractory oxides exhibit the most negative free energy of formation [33,34] (Fig. 5). Of these, hafnium and zirconium were selected with the goal of serving as primary oxide formers and are thus present in both compositions. Group VI refractory oxides exhibit the highest free energy of formation and may remain as carbides in the system as other elements preferentially oxidize. At the oxidation temperature in this study (1700 °C), the free energies of the Group V oxides (Nb, Ta, and V) are nearly equal and lie in between the Group IV and VI oxides (Fig. 5). These elements are thus expected to oxidize collectively after the Group IV elements. The additional two elements in the 7-cation system (vanadium and tungsten) are ideal additions in that vanadium has a similar free energy of oxidation to niobium and tantalum, while tungsten is only slightly more reactive than molybdenum (Fig. 5). Assuming that Hf and Zr are both the primary initial oxidizing species, the initial 5-cation composition would become a 3-cation MoNbTa-carbide, no longer high-entropy, and likely inherently less stable at elevated temperature. In contrast, the 7-cation UHEC would reduce to a 5-cation MoNbTaVW-carbide, retaining the high-entropy form, and likely stable to higher temperatures. It was with these principles in mind that the two similar compositions were selected to study the experimental order and rate of oxidation. Macroscopic images of the two samples post-oxidation are provided to display the differences in porosity and oxide growth (Supplementary Figure S13). Fig. 6 compares the oxide evolution between a  $\text{HfMoNbTaZrC}_5$  and  $\text{HfMoNbTaVWZrC}_7$  across a 4mm wide area. It can be readily observed that the oxide layer in the 5-cation system is approximately 3 times thicker than its 7-cation counterpart. This alone is evidence that the oxidation rate is significantly reduced through the chemical tuning of sacrificial oxide formers and subsequent phase formation (e.g., the depleted UHEC transition layer). Additionally, the oxide layer that forms in the 5-cation system is extremely porous, likely reducing the overall structural stability of this layer.

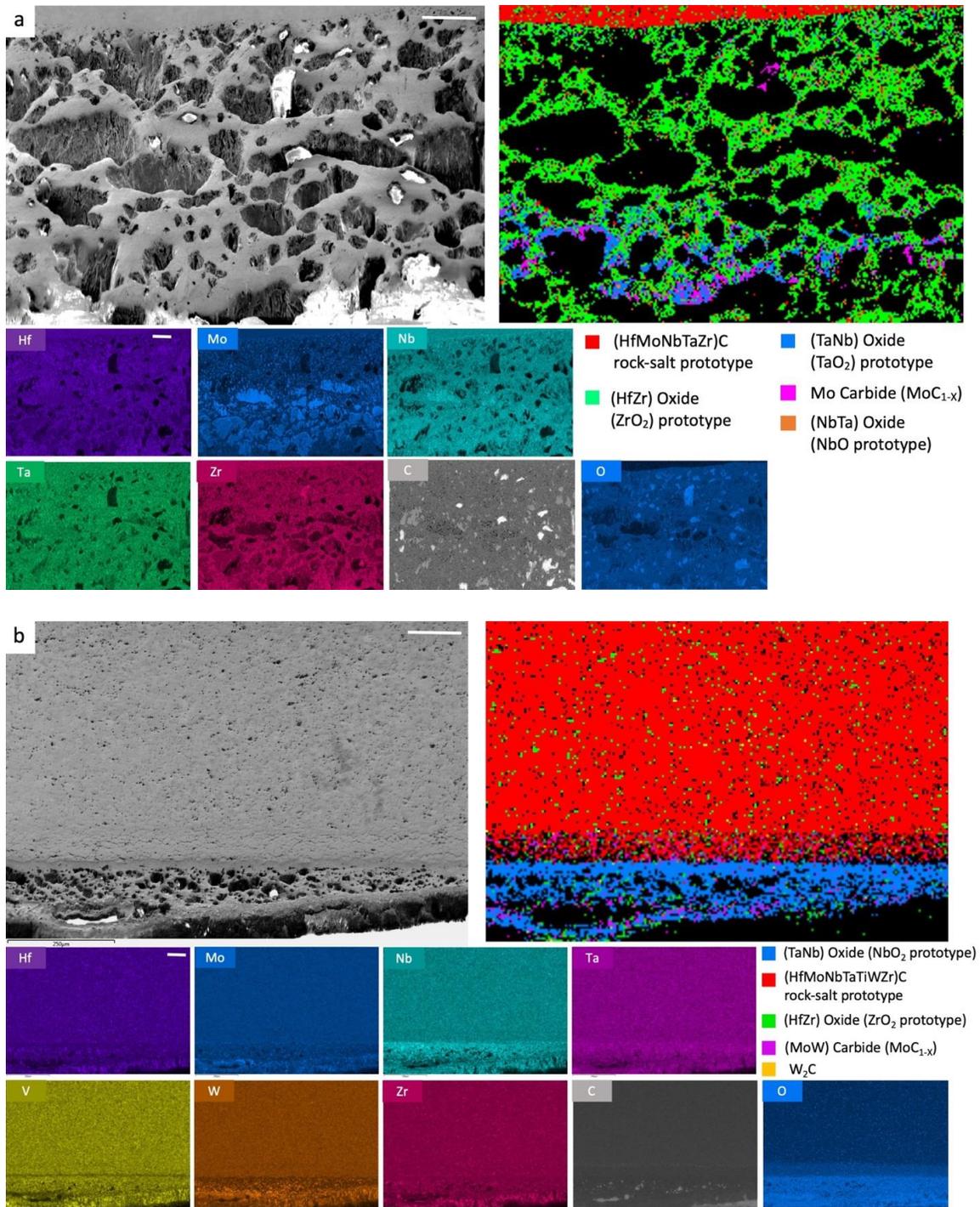
Conversely, the 7-cation system is relatively dense suggesting more structural stability. It is hypothesized that the thick transition layer present in the 7-cation system facilitates the structural stability between the oxide and substrate as well as slows the oxide growth by maintaining a high configurational entropy phase as an oxygen diffusion barrier.



**Figure 4.6 Overview of oxide layer formation.** Large area electron images for the oxidation interfaces of (a)  $\text{HfMoNbTaZrC}_5$  and (b)  $\text{HfMoNbTaVWZrC}_7$ . While the 5-cation HEC is observed to rapidly grow a highly porous oxide layer, the 7-cation UHEC oxidizes much more slowly when exposed to the same oxidative environment (1973 K, 1 atm for 10 minutes). Scale bars 1 mm.

Fig. 7 depicts the results of the oxidation test of  $\text{HfMoNbTaZrC}_5$  and  $\text{HfMoNbTaVWZrC}_7$  at the interface of oxide formation and the surviving high-entropy substrate. Electron backscatter diffraction was used in tandem with EDS (Table 2 and Table 3) and XRD (Fig. 8) to accurately determine phases present as well as their compositions. The legend for each EBSD phase map is given in Fig. 7 denoting the primary elements in each phase, not their structure. Structure information for each phase can be found in the XRD patterns (Fig. 8), the prototype structures in Table 2 and Table 3, and is further detailed in Supplementary Table 8. Supplementary Figures

S14-S16 and Supplementary Figures S17-S19 further detail the microstructure of the interfacial and oxide layers.



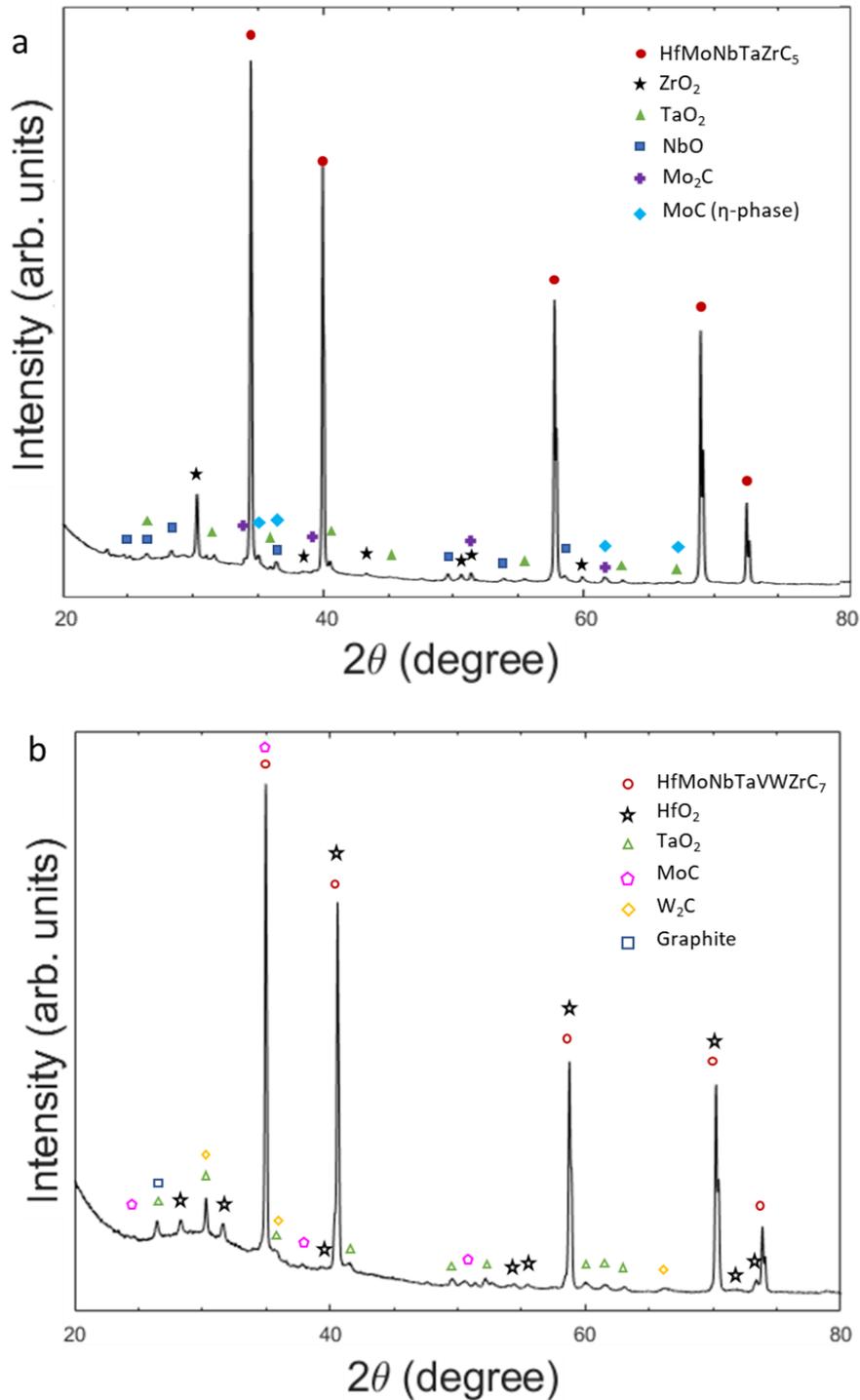
**Figure 4.7 Analysis of the oxidation interfaces.** EDS and EBSD maps of the oxidation interface for (a)  $\text{HfMoNbTaZrC}_5$  and (b)  $\text{HfMoNbTaVWZrC}_7$ . The legends for each EBSD map denote the primary chemistry for each phase, while the structures are correlated in Supplementary Table 8. Scale bars 125  $\mu\text{m}$ . Samples tested at 1973K and 1 atm pressure for 10 minutes.

**Table 4.2 Compositions for each phase in HfMoNbTaZrC<sub>5</sub> post-oxidation.** The average and standard deviation are reported for each phase not including the bulk HEC. The phase labels denote the major elements and correlate with the EBSD maps. Each average is computed using ten EDS measurements.

<b>Phase Label</b>		<b>Zr</b>	<b>Nb</b>	<b>Mo</b>	<b>Hf</b>	<b>Ta</b>	<b>Prototype</b>
<b>(HfZr) Oxide</b>	Average	29.70	12.24	10.11	31.54	16.41	ZrO <sub>2</sub> (Baddeleyite)
	Std. Dev	5.63	2.04	15.63	5.34	3.11	
<b>(TaNb) Oxide</b>	Average	6.98	38.04	2.83	7.48	44.67	TaO <sub>2</sub>
	Std. Dev	1.57	4.11	0.89	1.08	3.66	
<b>(NbTa) Oxide</b>	Average	6.73	60.04	10.31	7.32	15.60	NbO
	Std. Dev	1.09	6.91	9.56	0.99	2.36	
<b>Mo Carbide</b>	Average	1.67	2.00	92.50	1.73	2.10	MoC <sub>1-x</sub>
	Std. Dev	0.22	0.24	0.39	0.06	0.10	

**Table 4.3 Compositions for each phase in HfMoNbTaVWZrC<sub>7</sub> post-oxidation.** The average and standard deviation are reported for each phase not including the bulk UHEC. The phase labels denote the major elements and correlate with the EBSD maps. Each average is computed using ten EDS measurements.

<b>Phase Label</b>		<b>V</b>	<b>Zr</b>	<b>Nb</b>	<b>Mo</b>	<b>Hf</b>	<b>Ta</b>	<b>W</b>	<b>Prototype</b>
<b>(TaNb) Oxide</b>	Average	16.94	15.06	22.43	4.07	12.53	22.83	6.15	NbO <sub>2</sub>
	Std. Dev	0.28	1.14	4.10	1.51	0.57	2.20	1.31	
<b>(HfZr) Oxide</b>	Average	5.48	39.64	4.59	3.42	34.27	6.42	6.18	ZrO <sub>2</sub> (Baddeleyite)
	Std. Dev	3.46	8.29	5.71	4.74	11.12	4.65	3.38	
<b>(MoW) Carbide</b>	Average	0.54	1.31	1.00	55.53	0.58	0.23	40.82	MoC <sub>1-x</sub>
	Std. Dev	0.36	1.24	1.72	1.62	0.53	0.47	2.75	
<b>Depleted UHEC</b>	Average	14.40	10.88	14.51	16.43	10.41	15.55	17.82	NaCl
	Std. Dev	1.04	2.46	2.92	2.54	2.84	1.72	1.21	



**Figure 4.8 X-ray diffraction data post-oxidation.** Labeled XRD data for post-oxidation samples of (a) HfMoNbTaZrC<sub>5</sub> and (b) HfMoNbTaVWZrC<sub>7</sub>. The phase labels denote the crystallographic structures that match each phase, not the chemistry. Refer to Supplementary Table 8 for complete structure information.

In each of these materials, it is observed that hafnium and zirconium are indeed the initial elements to oxidize, indicated by their presence on the outer edges of the oxide layers in Figure 7. Along with the oxidation of these two elements, the 5-cation HEC forms a complex amalgamation of oxides with retained islands of  $\text{MoC}_{1-x}$  dispersed throughout the oxide layer. There is no transitional barrier formed that separates the 5-cation HEC substrate from the oxide, allowing unimpeded diffusion, and ultimately resulting in a thicker oxide layer (Fig. 7a). The 7-cation UHEC (Fig. 7b) primarily forms an oxide layer rich in Nb and Ta interspersed with oxides rich in Hf and Zr. This region also contains (Mo,W) carbides, predominantly of the MoC structure accompanied by a few fine scale  $\text{W}_2\text{C}$  particles that are too small for reliable EDS measurements in the SEM. Between the oxide and 7-cation substrate exists an approximately  $75\mu\text{m}$  thick HEC transition layer deficient in Hf and Zr. Refer to the oxygen EDS map in Fig. 7b for correlation of the layers to the EBSD phase map. The Hf and Zr form fine-scale (Hf,Zr) oxides within this depleted UHEC layer (Supplementary Fig. S12), consistent with oxidation predictions.

#### **4.4 Discussion**

Herein, the synthesis of 6-9 metal cation ultrahigh-entropy carbides is demonstrated with compositional guidance from machine learning predicted entropy forming ability, and their tailored oxidation performance is evaluated. The ML model suggests a significant fraction of these previously unexplored compositions will form single-phase ceramics, which is validated on seven selected chemistries using XRD and EDS chemistry analysis. While the ML-predicted EFA is reported herein for the possible equiatomic combinations of the studied carbide-forming refractory metal cations, there is likely ample opportunity for material discovery in the non-equiatomic combinations as well. The introduction of additional metal cations, above that of previously identified 5-cation high-entropy carbide structures, presents a new realm of opportunity in terms

of chemical tunability in ceramics. Regarding mechanical properties, the bulk and shear moduli are observed to follow previously reported trends corresponding to their valence electron concentration (VEC), wherein higher VEC corresponds to a lower G/B ratio, approaching the brittle to ductile transition for rock-salt-structured ceramics [6,16,32,66,67]. Compositionally, the inclusion of greater than 5 refractory elements in carbide materials marks a significant step forward in the field of entropic stabilization. Importantly for many next generation applications, this is expected to yield a net increase in the overall melting temperature of the system owing to the increasingly negative slope (*i.e.* entropy) of the Gibbs free energy equation.

While much of the previous work involving high-entropy ceramics has been focused on the discovery of viable single-phase systems, the present work is also interested in tailoring the performance of such materials for extreme conditions. In this work, we calculate Ellingham diagrams that accurately describe the complexity and ceramic nature of these materials, demonstrate that the oxidation process correlates well with these Ellingham diagrams. Furthermore, we illustrate that a greater number of cations can be employed to create a transitional, and possibly protective, HEC layer deficient in the elements that oxidize first. The ability to slow oxidation kinetics using high-entropy alloy nanoparticles compared to monometallic and bimetallic nanoparticles has recently been reported, [68] supporting the presented findings. The presence of the Hf- and Zr-rich oxides forming first within the HEC and UHEC corroborates the thermodynamically controlled oxidation order. The absence of other oxides within each carbide substrate further corroborates this. Continued research is required to firmly establish the degree to which the oxidation rate is controlled by the same thermodynamic principles and the precise role of the depleted UHEC transition layer. Moreover, there may exist an opportunity to tailor the

properties of the resultant oxide(s) and carbide(s); this strategy could pave the way for functional ceramics with designed properties, thus extending technologies for extreme environments.

#### **4.5 Conclusion**

Seven single phase ultrahigh-entropy rock-salt carbides were fabricated via high-energy ball milling and spark plasma sintering. The observed mechanical properties agree with trends previously observed for 5-cation HECs as a function of VEC. Ellingham diagrams are employed to understand the potential for tunability in an oxidative environment. These findings are demonstrated through comparison of a 5-cation and 7-cation system in a box furnace at 1973K in atmospheric conditions. It is observed that the 7-cation system creates a more uniform oxide layer in comparison to the 5-cation system, as well as an oxide transition layer that lead to an overall decrease in the oxidation rate.

#### **4.6 Acknowledgements**

The authors acknowledge support through the Office of Naval Research ONR-MURI (grant No. N00014-15-1-2863). K.K. was supported by the Department of Defense through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. K.K. would also like to acknowledge the generous support of the ARCS Foundation, San Diego Chapter. K.S.V. would like to acknowledge the financial generosity of the Oerlikon Group in support of his research group. This work was partially supported through access and utilization of the UC San Diego, Dept. of NanoEngineering's Materials Research Center (NEMRC). K.S.V. would like to acknowledge the contributions of Naixie Zhou to the Ellingham diagram calculations.

Chapter 4, in full, is currently in submission as “Development of ultrahigh-entropy ceramics with tailored oxidation behavior”, W. M. Mellor<sup>†</sup>, K. Kaufmann<sup>†</sup>, O. F. Dippo, S. D.

Figuroa, G. D. Schrader, and K. S. Vecchio. The dissertation author was the primary investigator and co-author of this material.

## 4.7 References

- [1] B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Microstructural development in equiatomic multicomponent alloys, *Mater. Sci. Eng. A.* 375–377 (2004) 213–218. <https://doi.org/10.1016/J.MSEA.2003.10.257>.
- [2] D.B. Miracle, O.N. Senkov, A critical review of high entropy alloys and related concepts, *Acta Mater.* 122 (2017) 448–511. <https://doi.org/10.1016/j.actamat.2016.08.081>.
- [3] M.H. Tsai, J.W. Yeh, High-entropy alloys: A critical review, *Mater. Res. Lett.* 2 (2014) 107–123. <https://doi.org/10.1080/21663831.2014.912690>.
- [4] C. Toher, C. Oses, D. Hicks, S. Curtarolo, Unavoidable disorder and entropy in multi-component systems, *Npj Comput. Mater.* 5 (2019). <https://doi.org/10.1038/s41524-019-0206-z>.
- [5] S.J. McCormack, A. Navrotsky, Thermodynamics of high entropy oxides, *Acta Mater.* 202 (2021) 1–21. <https://doi.org/10.1016/j.actamat.2020.10.043>.
- [6] O.F. Dippo, N. Mesgarzadeh, T.J. Harrington, G.D. Schrader, K.S. Vecchio, Bulk high - entropy nitrides and carbonitrides, *Sci. Rep.* 10 (2020) 1–11. <https://doi.org/10.1038/s41598-020-78175-8>.
- [7] A.J. Wright, J. Luo, A step forward from high-entropy ceramics to compositionally complex ceramics: a new perspective, *J. Mater. Sci.* 55 (2020) 9812–9827. <https://doi.org/10.1007/s10853-020-04583-w>.
- [8] A.J. Wright, Q. Wang, C. Huang, A. Nieto, R. Chen, J. Luo, From high-entropy ceramics to compositionally-complex ceramics: A case study of fluorite oxides, *J. Eur. Ceram. Soc.* 40 (2020) 2120–2129. <https://doi.org/10.1016/j.jeurceramsoc.2020.01.015>.
- [9] J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, S.-Y. Chang, Nanostructured High-Entropy Alloys with Multiple Principal Elements: Novel Alloy Design Concepts and Outcomes, *Adv. Eng. Mater.* 6 (2004) 299–303. <https://doi.org/10.1002/adem.200300567>.
- [10] J. Gild, M. Samiee, J.L. Braun, T. Harrington, H. Vega, P.E. Hopkins, K. Vecchio, J. Luo, High-entropy fluorite oxides, *J. Eur. Ceram. Soc.* 38 (2018) 3578–3584. <https://doi.org/10.1016/J.JEURCERAMSOC.2018.04.010>.
- [11] M. Qin, Q. Yan, H. Wang, C. Hu, K.S. Vecchio, J. Luo, High-entropy monoborides: Towards superhard materials, *Scr. Mater.* 189 (2020) 101–105. <https://doi.org/10.1016/j.scriptamat.2020.08.018>.
- [12] K. Kaufmann, D. Maryanovsky, W.M. Mellor, C. Zhu, A.S. Rosengarten, T.J. Harrington, C. Oses, C. Toher, S. Curtarolo, K.S. Vecchio, Discovery of high-entropy ceramics via machine learning, *Npj Comput. Mater.* 6 (2020) 1–9. <https://doi.org/10.1038/s41524-020-0317-6>.

- [13] M. Qin, Q. Yan, Y. Liu, J. Luo, A new class of high-entropy M<sub>3</sub>B<sub>4</sub> borides, *J. Adv. Ceram.* 2021 (2020) 0–0. <https://doi.org/10.1007/s40145-020-0438-x>.
- [14] M. Qin, Q. Yan, H. Wang, K.S. Vecchio, J. Luo, High-entropy rare earth tetraborides, *J. Eur. Ceram. Soc.* (2020). <https://doi.org/10.1016/j.jeurceramsoc.2020.12.019>.
- [15] P. Sarker, T. Harrington, C. Toher, C. Oses, M. Samiee, J.P. Maria, D.W. Brenner, K.S. Vecchio, S. Curtarolo, High-entropy high-hardness metal carbides discovered by entropy descriptors, *Nat. Commun.* 9 (2018) 4980. <https://doi.org/10.1038/s41467-018-07160-7>.
- [16] T.J. Harrington, J. Gild, P. Sarker, C. Toher, C.M. Rost, O.F. Dippo, C. McElfresh, K. Kaufmann, E. Marin, L. Borowski, P.E. Hopkins, J. Luo, S. Curtarolo, D.W. Brenner, K.S. Vecchio, Phase stability and mechanical properties of novel high entropy transition metal carbides, *Acta Mater.* 166 (2019). <https://doi.org/10.1016/j.actamat.2018.12.054>.
- [17] J. Gild, K. Kaufmann, K. Vecchio, J. Luo, Reactive flash spark plasma sintering of high-entropy ultrahigh temperature ceramics, *Scr. Mater.* 170 (2019). <https://doi.org/10.1016/j.scriptamat.2019.05.039>.
- [18] J. Gild, J. Braun, K. Kaufmann, E. Marin, T. Harrington, P. Hopkins, K. Vecchio, J. Luo, A high-entropy silicide: (Mo<sub>0.2</sub> Nb<sub>0.2</sub> Ta<sub>0.2</sub> Ti<sub>0.2</sub> W<sub>0.2</sub>)Si<sub>2</sub>, *J. Mater.* (2019). <https://doi.org/10.1016/j.jmat.2019.03.002>.
- [19] C.M. Rost, E. Sachet, T. Borman, A. Moballegh, E.C. Dickey, D. Hou, J.L. Jones, S. Curtarolo, J.P. Maria, Entropy-stabilized oxides, *Nat. Commun.* 6 (2015) 1–8. <https://doi.org/10.1038/ncomms9485>.
- [20] J. Gild, A. Wright, K. Quiambao-Tomko, M. Qin, J.A. Tomko, M. Shafkat bin Hoque, J.L. Braun, B. Bloomfield, D. Martinez, T. Harrington, K. Vecchio, P.E. Hopkins, J. Luo, Thermal conductivity and hardness of three single-phase high-entropy metal diborides fabricated by borocarbothermal reduction and spark plasma sintering, *Ceram. Int.* 46 (2020) 6906–6913. <https://doi.org/10.1016/j.ceramint.2019.11.186>.
- [21] M. Qin, J. Gild, H. Wang, T. Harrington, K.S. Vecchio, J. Luo, Dissolving and stabilizing soft WB<sub>2</sub> and MoB<sub>2</sub> phases into high-entropy borides via boron-metals reactive sintering to attain higher hardness, *J. Eur. Ceram. Soc.* 40 (2020) 4348–4353. <https://doi.org/10.1016/j.jeurceramsoc.2020.03.063>.
- [22] L.E. Toth, *Transition Metal Carbides and Nitrides*, Academic Press, New York, 1971.
- [23] M.A. Abdelkareem, T. Wilberforce, K. Elsaid, E.T. Sayed, E.A.M. Abdelghani, A.G. Olabi, Transition metal carbides and nitrides as oxygen reduction reaction catalyst or catalyst support in proton exchange membrane fuel cells (PEMFCs), *Int. J. Hydrogen Energy.* (2020). <https://doi.org/10.1016/j.ijhydene.2020.08.250>.
- [24] O.F. Dippo, N. Mesgarzadeh, T.J. Harrington, G.D. Schrader, K.S. Vecchio, Bulk high-entropy nitrides and carbonitrides, *Sci. Rep.* 10 (2020) 21288. <https://doi.org/10.1038/s41598-020-78175-8>.

- [25] X. Yan, L. Constantin, Y. Lu, J.F. Silvain, M. Nastasi, B. Cui, (Hf<sub>0.2</sub>Zr<sub>0.2</sub>Ta<sub>0.2</sub>Nb<sub>0.2</sub>Ti<sub>0.2</sub>)C high-entropy ceramics with low thermal conductivity, *J. Am. Ceram. Soc.* 101 (2018) 4486–4491. <https://doi.org/10.1111/jace.15779>.
- [26] D.G. Sangiovanni, F. Tasnadi, T. Harrington, K.S. Vecchio, I.A. Abrikosov, Temperature-dependent elastic properties of binary and multicomponent high-entropy refractory carbides, (2020).
- [27] A.J. Wright, Q. Wang, S.T. Ko, K.M. Chung, R. Chen, J. Luo, Size disorder as a descriptor for predicting reduced thermal conductivity in medium- and high-entropy pyrochlore oxides, *Scr. Mater.* 181 (2020) 76–81. <https://doi.org/10.1016/j.scriptamat.2020.02.011>.
- [28] D.B. Miracle, High entropy alloys as a bold step forward in alloy development, *Nat. Commun.* 10 (2019) 1805. <https://doi.org/10.1038/s41467-019-09700-1>.
- [29] D. Miracle, J. Miller, O. Senkov, C. Woodward, M. Uchic, J. Tiley, Exploration and Development of High Entropy Alloys for Structural Applications, *Entropy.* 16 (2014) 494–525. <https://doi.org/10.3390/e16010494>.
- [30] H. Kindlund, D.G. Sangiovanni, L. Martínez-De-Olcoz, J. Lu, J. Jensen, J. Birch, I. Petrov, J.E. Greene, V. Chirita, L. Hultman, Toughness enhancement in hard ceramic thin films by alloy design, *APL Mater.* 1 (2013) 042104. <https://doi.org/10.1063/1.4822440>.
- [31] D.G. Sangiovanni, Inherent toughness and fracture mechanisms of refractory transition-metal nitrides via density-functional molecular dynamics, *Acta Mater.* 151 (2018) 11–20. <https://doi.org/10.1016/j.actamat.2018.03.038>.
- [32] K. Balasubramanian, S. V. Khare, D. Gall, Valence electron concentration as an indicator for mechanical properties in rocksalt structure nitrides, carbides and carbonitrides, *Acta Mater.* 152 (2018) 175–185. <https://doi.org/10.1016/j.actamat.2018.04.033>.
- [33] L. Backman, J. Gild, J. Luo, E.J. Opila, Selective Oxidation in Refractory High Entropy Materials, n.d.
- [34] L. Backman, E.J. Opila, Thermodynamic assessment of the group IV, V and VI oxides for the design of oxidation resistant multi-principal component materials, *J. Eur. Ceram. Soc.* 39 (2019) 1796–1802. <https://doi.org/10.1016/j.jeurceramsoc.2018.11.004>.
- [35] Y. Zhang, T.T. Zuo, Z. Tang, M.C. Gao, K.A. Dahmen, P.K. Liaw, Z.P. Lu, Microstructures and properties of high-entropy alloys, *Prog. Mater. Sci.* 61 (2014) 1–93. <https://doi.org/10.1016/j.pmatsci.2013.10.001>.
- [36] A. Poulia, E. Georgatis, A. Lekatou, A.E. Karantzalis, Microstructure and wear behavior of a refractory high entropy alloy, *Int. J. Refract. Met. Hard Mater.* 57 (2016) 50–63. <https://doi.org/10.1016/j.ijrmhm.2016.02.006>.
- [37] E. Fazakas, V. Zadorozhnyy, L.K. Varga, A. Inoue, D. V. Louzguine-Luzgin, F. Tian, L. Vitos, Experimental and theoretical study of Ti<sub>20</sub>Zr<sub>20</sub>Hf<sub>20</sub>Nb<sub>20</sub>X<sub>20</sub> (X = V or Cr)

- refractory high-entropy alloys, *Int. J. Refract. Met. Hard Mater.* 47 (2014) 131–138. <https://doi.org/10.1016/j.ijrmhm.2014.07.009>.
- [38] Z.K. Liu, B. Li, H. Lin, Multiscale Entropy and Its Implications to Critical Phenomena, Emergent Behaviors, and Information, *J. Phase Equilibria Diffus.* (2019). <https://doi.org/10.1007/s11669-019-00736-w>.
- [39] Y. Wu, J. Si, D. Lin, T. Wang, W.Y. Wang, Y. Wang, Z.K. Liu, X. Hui, Phase stability and mechanical properties of AlHfNbTiZr high-entropy alloys, *Mater. Sci. Eng. A.* 724 (2018) 249–259. <https://doi.org/10.1016/j.msea.2018.03.071>.
- [40] O.N. Senkov, J.D. Miller, D.B. Miracle, C. Woodward, Accelerated exploration of multi-principal element alloys with solid solution phases, *Nat. Commun.* 6 (2015) 6529. <https://doi.org/10.1038/ncomms7529>.
- [41] M. Gao, D. Alman, Searching for Next Single-Phase High-Entropy Alloy Compositions, *Entropy.* 15 (2013) 4504–4519. <https://doi.org/10.3390/e15104504>.
- [42] F. Zhang, C. Zhang, S.L. Chen, J. Zhu, W.S. Cao, U.R. Kattner, An understanding of high entropy alloys from phase diagram calculations, *Calphad.* 45 (2014) 1–10. <https://doi.org/10.1016/J.CALPHAD.2013.10.006>.
- [43] Z.K. Liu, Ocean of Data: Integrating First-Principles Calculations and CALPHAD Modeling with Machine Learning, *J. Phase Equilibria Diffus.* 39 (2018) 635–649. <https://doi.org/10.1007/s11669-018-0654-z>.
- [44] J.M. Rickman, H.M. Chan, M.P. Harmer, J.A. Smeltzer, C.J. Marvel, A. Roy, G. Balasubramanian, Materials informatics for the screening of multi-principal elements and high-entropy alloys, *Nat. Commun.* 10 (2019) 1–10. <https://doi.org/10.1038/s41467-019-10533-1>.
- [45] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T.D. Hirzel, D. Duvenaud, D. Maclaurin, M.A. Blood-Forsythe, H.S. Chae, M. Einzinger, D.G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S.I. Hong, M. Baldo, R.P. Adams, A. Aspuru-Guzik, Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach, *Nat. Mater.* 15 (2016) 1120–1127. <https://doi.org/10.1038/nmat4717>.
- [46] Y. Liu, T. Zhao, W. Ju, S. Shi, S. Shi, S. Shi, Materials discovery and design using machine learning, *J. Mater.* 3 (2017) 159–177. <https://doi.org/10.1016/j.jmat.2017.08.002>.
- [47] K. Kaufmann, K.S. Vecchio, Searching for high entropy alloys: A machine learning approach, *Acta Mater.* 198 (2020) 178–222. <https://doi.org/10.1016/j.actamat.2020.07.065>.
- [48] A. Dunn, Q. Wang, A. Ganose, D. Dopp, A. Jain, Benchmarking Materials Property Prediction Methods: The Matbench Test Set and Automatminer Reference Algorithm, (2020).

- [49] S.R. Spurgeon, C. Ophus, L. Jones, A. Petford-Long, S. V. Kalinin, M.J. Olszta, R.E. Dunin-Borkowski, N. Salmon, K. Hattar, W.C.D. Yang, R. Sharma, Y. Du, A. Chiramonti, H. Zheng, E.C. Buck, L. Kovarik, R.L. Penn, D. Li, X. Zhang, M. Murayama, M.L. Taheri, Towards data-driven next-generation transmission electron microscopy, *Nat. Mater.* (2020). <https://doi.org/10.1038/s41563-020-00833-z>.
- [50] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, *Phys. Rev. B - Condens. Matter Mater. Phys.* 89 (2014) 094104. <https://doi.org/10.1103/PhysRevB.89.094104>.
- [51] A. Ziletti, D. Kumar, M. Scheffler, L.M. Ghiringhelli, Insightful classification of crystal structures using deep learning, *Nat. Commun.* 9 (2018) 2775. <https://doi.org/10.1038/s41467-018-05169-6>.
- [52] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, T.J. Harrington, E. Marin, K.S. Vecchio, Crystal symmetry determination in electron diffraction using machine learning, *Science* (80-. ). 367 (2020) 564–568. <https://doi.org/10.1126/science.aay3062>.
- [53] F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N.T.P. Hartono, S. Ramasamy, B.L. DeCost, S.I.P. Tian, G. Romano, A. Gilad Kusne, T. Buonassisi, Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks, *Npj Comput. Mater.* 5 (2019). <https://doi.org/10.1038/s41524-019-0196-x>.
- [54] W.B. Park, J. Chung, J. Jung, K. Sohn, S.P. Singh, M. Pyo, N. Shin, K.S. Sohn, Classification of crystal structure using a convolutional neural network, *IUCrJ.* 4 (2017) 486–494. <https://doi.org/10.1107/S205225251700714X>.
- [55] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, Universal fragment descriptors for predicting properties of inorganic crystals, *Nat. Commun.* 8 (2017). <https://doi.org/10.1038/ncomms15679>.
- [56] W. Ye, C. Chen, Z. Wang, I.-H. Chu, S.P. Ong, Deep neural networks for accurate predictions of crystal stability, *Nat. Commun.* 9 (2018) 3800. <https://doi.org/10.1038/s41467-018-06322-x>.
- [57] F. Ren, L. Ward, T. Williams, K.J. Laws, C. Wolverton, J. Hattrick-Simpers, A. Mehta, Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments, *Sci. Adv.* 4 (2018) eaaq1566. <https://doi.org/10.1126/sciadv.aaq1566>.
- [58] L. Breiman, Random Forests, *Mach. Learn.* 45 (2001) 5–32. <https://doi.org/10.1023/A:1010933404324>.
- [59] R. Caruana, A. Niculescu-Mizil, An empirical comparison of supervised learning algorithms, *ACM Int. Conf. Proceeding Ser.* 148 (2006) 161–168. <https://doi.org/10.1145/1143844.1143865>.

- [60] B. Efron, R. Tibshirani, Improvements on cross-validation: The .632+ bootstrap method, *J. Am. Stat. Assoc.* 92 (1997) 548–560. <https://doi.org/10.1080/01621459.1997.10474007>.
- [61] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [62] M. Temkin, Mixtures of Fused Salts as Ionic Solutions, *Acta Physicochim. USSR.* 20 (1945) 411–420.
- [63] M. Hillert, Modelling of disorder, in: *Phase Equilibria, Phase Diagrams Phase Transform. Their Thermodyn. Basis*, 2nd ed., Cambridge University Press, 2008: pp. 420–437.
- [64] J.W. Yeh, Y.L. Chen, S.J. Lin, S.K. Chen, High-Entropy Alloys – A New Era of Exploitation, *Mater. Sci. Forum.* 560 (2007) 1–9. <https://doi.org/10.4028/www.scientific.net/msf.560.1>.
- [65] Y. Wang, T. Csanádi, H. Zhang, J. Dusza, M.J. Reece, R. Zhang, Enhanced Hardness in High-Entropy Carbides through Atomic Randomness, *Adv. Theory Simulations.* 3 (2020) 2000111. <https://doi.org/10.1002/adts.202000111>.
- [66] A.B. Mei, H. Kindlund, E. Broitman, L. Hultman, I. Petrov, J.E. Greene, D.G. Sangiovanni, Adaptive hard and tough mechanical response in single-crystal B1 VN<sub>x</sub> ceramics via control of anion vacancies, *Acta Mater.* 192 (2020) 78–88. <https://doi.org/10.1016/j.actamat.2020.03.037>.
- [67] D.G. Sangiovanni, F. Tasnádi, L.J.S. Johnson, M. Odén, I.A. Abrikosov, Strength, transformation toughening, and fracture dynamics of rocksalt-structure T<sub>i</sub>1-x A<sub>1x</sub>N (0≤x≤0.75) alloys STRENGTH, TRANSFORMATION TOUGHENING, AND.. D. G. SANGIOVANNI et al., *Phys. Rev. Mater.* 4 (2020) 033605. <https://doi.org/10.1103/PhysRevMaterials.4.033605>.
- [68] B. Song, Y. Yang, M. Rabbani, T.T. Yang, K. He, X. Hu, Y. Yuan, P. Ghildiyal, V.P. Dravid, M.R. Zachariah, W.A. Saidi, Y. Liu, R. Shahbazian-Yassar, In Situ Oxidation Studies of High-Entropy Alloy Nanoparticles, *ACS Nano.* 16 (2020) 45. <https://doi.org/10.1021/acsnano.0c05250>.

## Chapter 5 Crystal symmetry determination in electron diffraction using machine learning

Kevin Kaufmann<sup>1</sup>, Chaoyi Zhu<sup>2†</sup>, Alexander S. Rosengarten<sup>1†</sup>, Daniel Maryanovsky<sup>3</sup>, Tyler J. Harrington<sup>2</sup>, Eduardo Marin<sup>1</sup>, and Kenneth S. Vecchio<sup>1,2\*</sup>

<sup>1</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA.

<sup>2</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA.

<sup>3</sup>Department of Cognitive Science, UC San Diego, La Jolla, CA 92093, USA.

\*Correspondence to: kvecchio@eng.ucsd.edu.

† These authors contributed equally to this work.

### Abstract

Electron backscatter diffraction (EBSD) is one of the primary tools for crystal structure determination. However, this method requires human input to select potential phases for Hough-based or dictionary pattern matching and is not well-suited for phase identification. Automated phase identification is the first step in making EBSD into a high-throughput technique. We utilized a machine learning-based approach and developed a general methodology for rapid and autonomous identification of the crystal symmetry from EBSD patterns. We evaluated our algorithm with diffraction patterns from materials outside the training set. The neural network assigned importance to the same symmetry features a crystallographer uses for structure identification.

## 5.1 Main Text

Identifying structure is a crucial step for the analysis of proteins [1–3], micro- [4,5] and macro-molecules [6], pharmaceuticals [7], geological specimens [8], synthetic materials [9–11], and for many other types of materials. Crystal structure plays an important role in the material properties exhibited [12,13]. Determining the crystal symmetry, lattice parameters, and atom positions of the crystal phases is a challenging task, especially for low symmetry phases and multi-phase materials. The most common techniques involve either X-ray diffraction (XRD) or transmission electron microscopy (TEM)-based convergent beam electron diffraction (CBED) [14–16]. X-ray diffraction only requires a sample powder or a polished bulk sample and in most cases only a few hours to collect diffraction intensities over a range of angles. Researchers must refine diffraction patterns to match the experimentally collected pattern to one in a database or from a theoretical model. This process has drawbacks as structural misclassification occurs due to lattice parameter shifts, overlapping XRD peaks in multi-phase samples, texture effects, and the matching thresholds set by researchers. TEM studies employing CBED are more precise than XRD in their ability to pinpoint the location of individual crystals, produce singular diffraction patterns for a given phase, and capture subtle symmetry information. However, sample preparation, data collection rates, data analysis rates, and the requirement of substantial operator experience limits the throughput for CBED based studies [17–19].

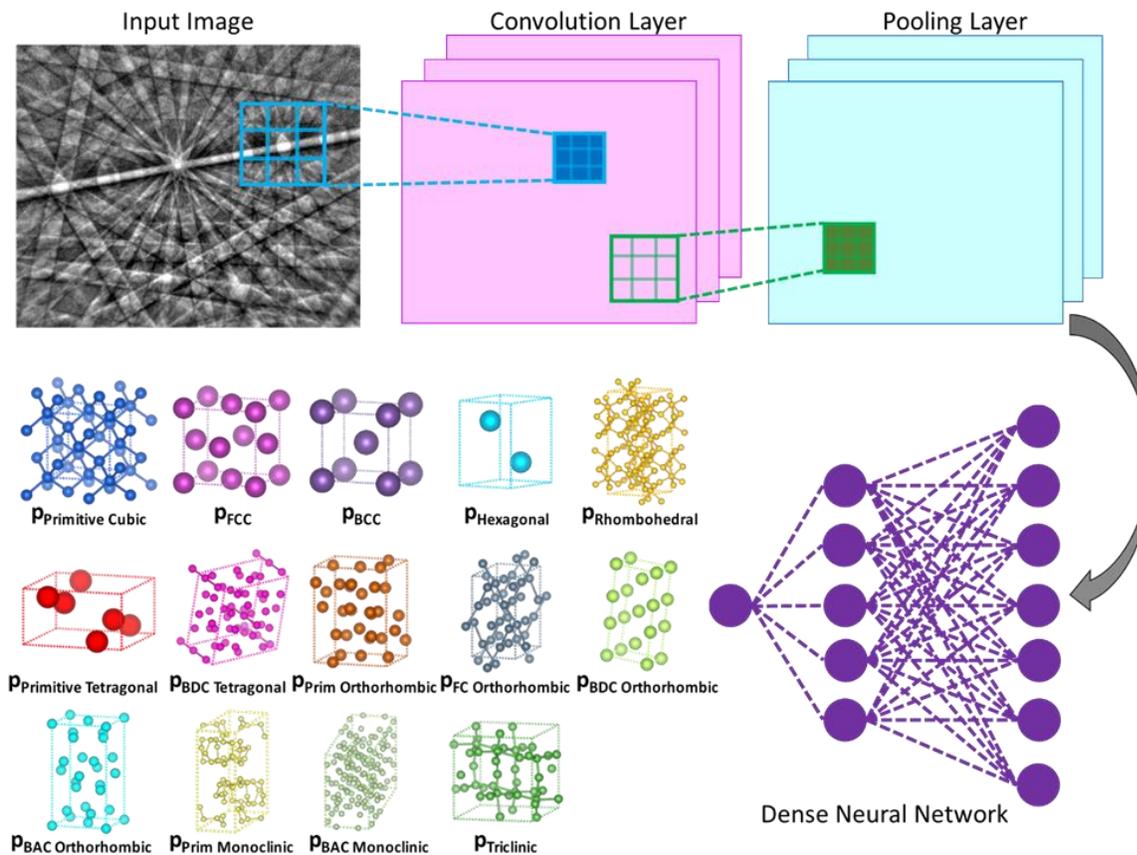
A scanning electron microscope (SEM), equipped with an electron backscatter diffraction (EBSD) system, has become important for characterization for crystalline materials and geological samples [8]. Nishikawa and Kikuchi discovered EBSD patterns in 1928 [20]. Early research by Alam *et al.* [21], Venables *et al.* [22], and Dingley [23], lead to the emergence of commercial EBSD systems. The development of fully-automated, image analysis methods occurred in the

early 1990's [24,25]. Because of the introduction of automated EBSD, commercial software and hardware have evolved to capture more than 3000 patterns per second, which expands the applicability of the technique to assist researchers with more complex problems [26]. For example, the high-throughput capability of a modern EBSD system enables determination of fine-scale grain structures, sample texture, point-to-point crystal orientation, residual stress/strain, geometrically necessary dislocation densities, and other information. [27–31]. The relative ease of sample preparation compared with TEM samples and the larger sample area analysis in less time makes the SEM-EBSD an attractive technique for studies of location specific orientation with high precision ( $\sim 2^\circ$ ), misorientation resolution ( $0.2^\circ$ ) and spatial resolution ( $\sim 40$  nm) [32]. One of the most common applications of EBSD in multi-phase samples is phase differentiation along with orientation determination. A user selects the phases presumed to be in the sample and a program finds the best fit phase and orientation to the diffraction pattern [33]. Selected libraries of simulated diffraction patterns of phases can be utilized in a dictionary indexing approach to assist with phase differentiation, including when working with deformed or fine grain materials [34]. Phase identification is possible when combined with other analytical techniques such as energy dispersive X-ray spectroscopy (EDS) or wavelength dispersive X-ray spectroscopy (WDS) [33,35,36]. This requires that the chemical and structural information of the phase exists in a theoretical model or crystal database, such as the Inorganic Crystal Structure Database (ICSD). A method was developed to determine the crystal structure using EBSD without EDS data but requires hand-drawn lines to be overlaid with a high degree of accuracy on individual Kikuchi bands [35,37]. This method is slow and tedious as it requires manual annotation of each individual pattern. In general, EBSD has been limited to elucidating orientation of user-defined crystal structures.

Recently, the materials science field has begun to embrace the big data revolution [38]. Researchers have shown the ability to predict new compositions for bulk metallic glasses [39], shape memory alloys [40], Heusler compounds [41,42], and ultra-incompressible superhard materials [43]. Other groups are developing machine learning methods to establish structure-property linkages [44–46], or predict the crystal stability of new materials [47]. Holm *et al.* [48,49], have demonstrated the classification of optical microscopy images into one of seven groups with greater than 80% accuracy [48], as well as microconstituent segmentation using the PixelNet convolutional neural network (CNN) architecture trained on manually annotated micrographs of ultrahigh carbon steel [49]. These machine learning driven analysis techniques represent important developments in the materials science toolbox. Previous studies have attempted crystal symmetry identification using deep neural networks and TEM diffraction, however, the developed model’s practical use is hindered by the choice to use images simulated in RGB color, while real TEM diffraction patterns are captured in greyscale [50]. Another study utilized full XRD pattern images for single-phase materials [51]. These techniques only provide point (TEM) or global (XRD) information about the sample compared to EBSD’s mapping capabilities, which can provide spatially-relevant crystallographic information across many length scales. Herein, we demonstrate a hybrid methodology, EBSD coupled with a machine learning algorithm, to identify the Bravais lattice or space group of a bulk sample from diffraction patterns. The trained machine learning model is subsequently applied to a distinct set of materials it was not trained on, but which contain the same crystal symmetry, and identifies the correct Bravais lattice or space group with a high degree of accuracy.

We used two convolutional neural networks (CNN) in this work. The two image classification model architectures are ResNet50 [52] and Xception [53] (Fig. 1). We started by

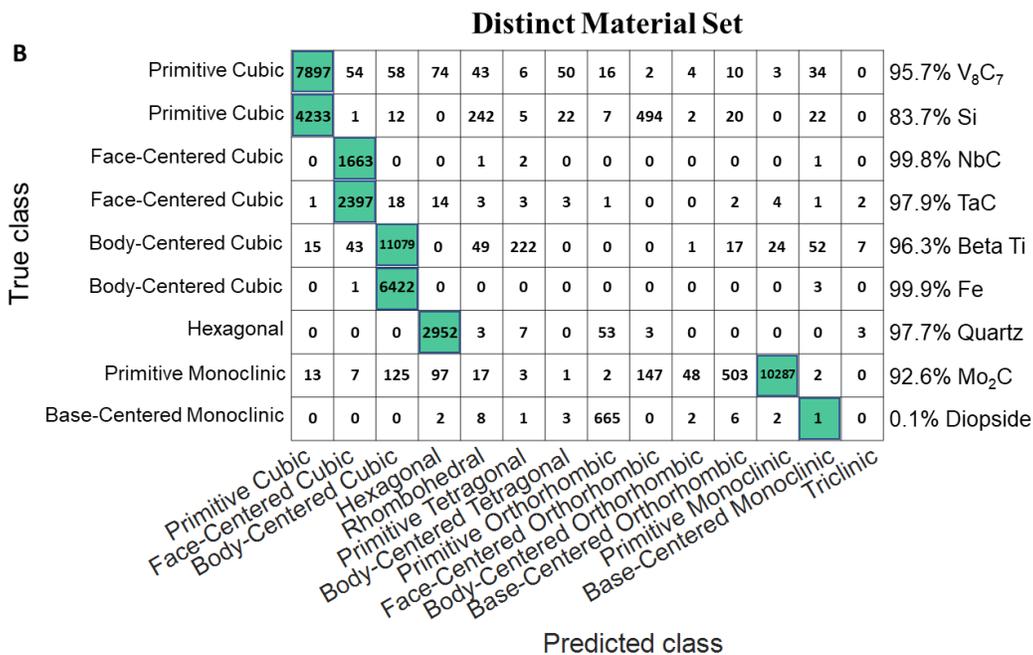
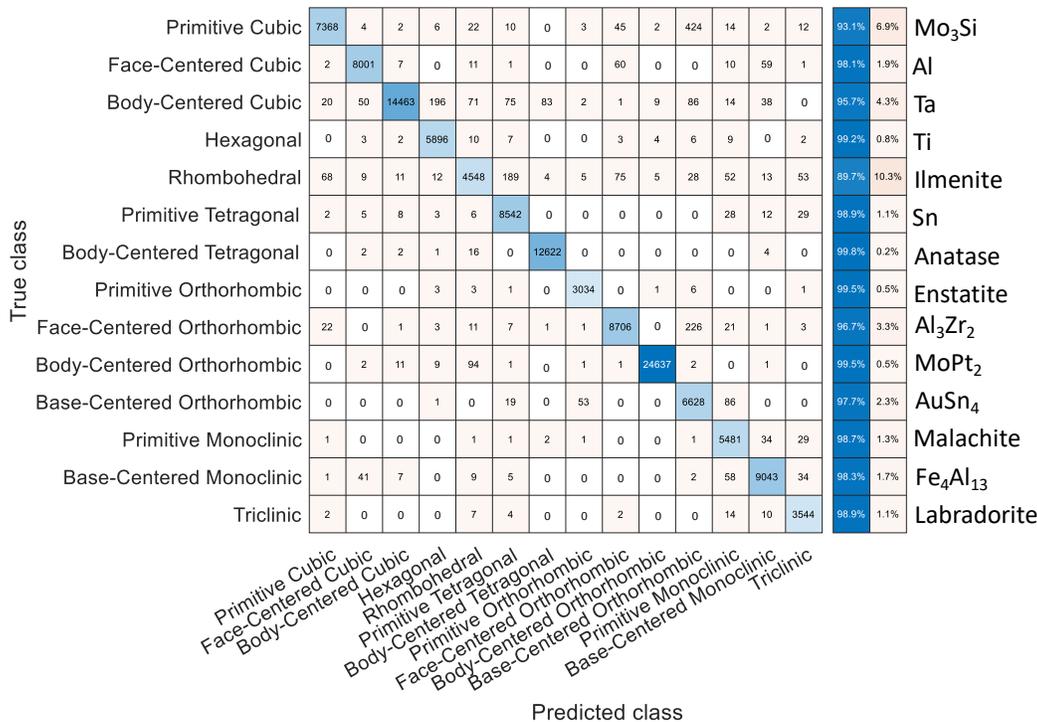
constructing convolution layer, where a learnable filter is convolved across the image. We computed the scalar product between the filter and the input at every position, or ‘patch’, to form a feature map. Next, we stacked sequentially a series of alternating convolutional and pooling layers. We organized the feature maps with the units in a convolutional layer and connected each feature map to local patches in the previous layer through a set of weights called a filter bank. All units in a feature map shared the same filter banks (also called kernels), while different feature maps in a convolutional layer used different filter banks. We placed pooling layers after convolutional layers to down sample the feature maps. This produced coarse grain representations and spatial information about the features in the data. The trained layers of feature detection nodes are ‘learned’ from the data as the algorithm finds motifs encoding the underlying crystallographic symmetry present in the diffraction patterns.



**Figure 5.1 Illustration of the inner workings of a convolutional neural network.** Convolutional neural networks are composed of a series of alternating convolutional and pooling layers. Each convolutional layer extracts features from its preceding layer, using filters (or kernels) learned from training the model, to form feature maps. These feature maps are then down-sampled by a pooling layer to exploit data locality. A traditional dense neural network, a simple type of classification network, is placed as the last layer of the CNN, where the probability that the input diffraction pattern belongs to a given class (e.g. Bravais lattice or space group) is computed.

We found that both ResNet50 and Xception [53] CNNs performed similarly well at classifying EBSD patterns. We applied the trained model to diffraction patterns that are ‘new’ to the algorithm. This means the patterns were not part of the training set but a random mix of orientations that may or may not be similar to the training patterns. Both the ResNet50 (Fig. 2) and Xception (Supplementary Figure 20) architecture correctly classified nearly 300,000 diffraction patterns with >90% overall accuracy for each architecture. Specifically, this means no user input was required for the algorithm to identify to which of the fourteen Bravais lattices each

individual EBSD belonged. The main crystal structure misclassification was jadeite, a monoclinic mineral often assigned to structures containing the same symmetry elements (Supplementary Figure 20A, C). This specific misclassification resulted in an overall decrease of the algorithm's performance. An in-depth analysis of this misclassification type is performed to understand the cause. The model displayed much higher accuracy on all other materials, typically greater than 95% for individual materials.



**Figure 5.2 Confusion matrix displaying the ResNet50 algorithm’s classification results.** (A) A second set of diffraction patterns from fourteen of the materials were classified by the trained algorithm. The diagonal (blue shaded boxes) represents the successful matching of the CNN predictions to the true Bravais lattices of the sample. (B) The algorithm classifies electron backscatter diffraction patterns collected from materials not used during training of the model. Correct classification is identified by the green squares instead of along the diagonal.

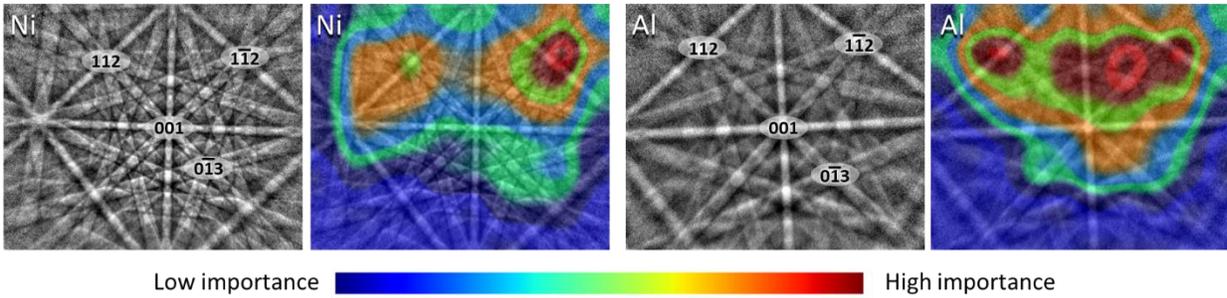
We collected 50,000 EBSD patterns from nine completely different materials for blind testing of our algorithms' crystal symmetry identification. Each architecture correctly classifies the Bravais lattice of the unknown material with 93.5% (Fig. 2B) and 91.2% (Supplementary Figure 21) overall accuracy for ResNet50 and Xception, respectively. The base-centered monoclinic crystal structure has a propensity to be incorrectly classified as primitive orthorhombic or rhombohedral. The base-centered monoclinic, primitive orthorhombic, and rhombohedral Bravais lattices utilized in training belong to the  $2/m$ ,  $mmm$ , and  $\bar{3}m$  point group, respectively. The  $2/m$  and  $mmm$  point groups each only have 2-fold axis symmetry, mirror plane symmetry, and inversion center symmetry in different multiplicity (SupplementaryTable 9). The rhombohedral  $\bar{3}m$  point group shares these same symmetry elements, with the addition of one 3-fold axis symmetry. As expected for a well fit model, of the misclassification to these two point groups, the model displays a lesser degree of misclassification to the  $\bar{3}m$  point group containing the extra symmetry element. This sort of misclassification event presents itself as a potential source of error, especially in a low symmetry phase, where one pattern from the Kikuchi sphere may not contain enough symmetry elements for the best possible classification.

We used our method to classify the small changes in atomic arrangement that distinguish space groups within the  $4/m \bar{3} 2/m$  (cubic) point group (Supplementary Figure 22) at a rate of 1 pattern per second. We found the ResNet50 algorithm only misclassified a small portion of our as-collected dataset between the selected cubic space groups.

Walking through a specific example of feature identification by the algorithm helps us understand how it arrives at a correct classification. We start with diffraction patterns of nickel

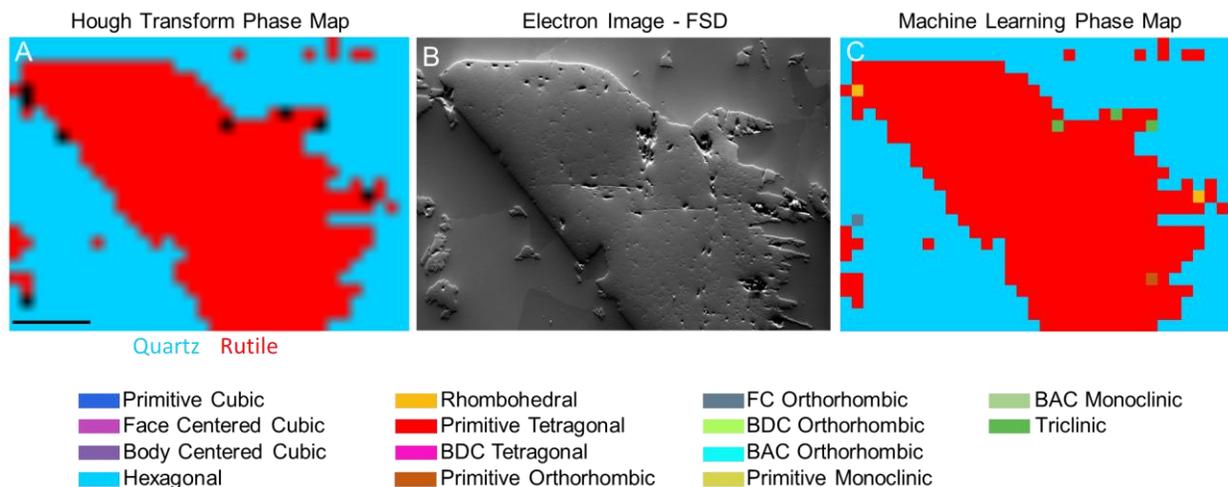
and aluminum with similar crystallographic orientation (Fig. 3). The importance of features in each image are determined by the learned filter banks in the algorithm. The trained neural network architecture and a set of tools called Grad-CAM [54] that provides weighting to local regions in the images. After the algorithm computes the ‘importance’ of these local regions, Grad-CAM maps the normalized weights from 0 (dark blue) to 1 (dark red). These heatmaps are similar for nickel and aluminum and show an intense interest of the network in symmetry located at the zone axes. the regions of greatest interest are the  $[1\bar{1}2]$  and  $[112]$  zone axes (2-fold symmetry). The machine learning algorithm couples this information with the presence of the  $[001]$  (4-fold symmetry) and  $[0\bar{1}3]$  (2-fold symmetry) and their spatial relationship, owing to pooling layers, to correctly identify the Bravais lattice as face-centered cubic. We observed a similar interest in information nearest the zone axes for the other materials.

We determined heatmaps for the 28 materials we used in the training set (Supplementary Figure 23). This allowed us to investigate where the algorithm has difficulty with identifications. We used Grad-CAM to investigate the misidentification of diopside (Supplementary Figure 24). The base centered monoclinic and primitive orthorhombic class both result in similar activations with interest centered around the only ‘x-fold’ symmetry present in diopside, two-fold symmetry  $[11\bar{2}]$  zone axis. Because the base-centered monoclinic and primitive orthorhombic structures only differ on the number of 2-fold axes and not having higher symmetry elements, the algorithm has difficulty 15 distinguishing between the structures. It is observed that the area of greatest interest is not always centered around the bright spot of a zone axis, like for  $\text{Cr}_3\text{Si}$  or  $\text{Sn}$ , and instead favors the side with other zone axes nearby in the diffraction pattern.



**Figure 5.3 Visualizing the features utilized for classification of a diffraction pattern to face-centered cubic (FCC).** An electron backscatter diffraction pattern from nickel (Ni) and aluminum (Al) were selected from nearly identical orientations. In the diffraction patterns, four of the zone axes present in each are labeled. The corresponding heatmap displays the importance of information in the image for correctly classifying it as FCC. It is observed that for each of these two images, the symmetry information near the  $[\bar{1}\bar{1}2]$  zone axis produces the highest activation, followed by the  $[112]$  zone axis, and symmetry shared by the  $[001]$  and  $[0\bar{1}3]$  zone axes.

Our algorithm reduced the amount of prior sample knowledge required for crystal structure identification. A common approach for crystal identification is to run 20 the diffraction images through a Hough transform, which help extract diffraction maxima at Kikuchi band intersections. This method can lead to misclassification of similar crystal structures that have similar diffraction maxima [55–57]. In contrast, our algorithm autonomously utilizes all the information in each diffraction pattern. We demonstrate how this helps with a multi-phase sample using rutiled quartz (Fig. 4), containing a phase that was not in our training set. Our machine learning generated phase map is nearly identical to the one generated by the Hough transform method. Of the seven errors, five are located where the traditional method could not index the structure.



**Figure 5.4 Comparison of phase mapping techniques.** (A) The phase map on the left was generated by traditional Hough-transform EBSD where the user had to select quartz and rutile as the two specific phases present in the sample. Black pixels could not be identified. (B) The center electron image is the region of the sample from which the diffraction patterns were collected. The quartz appears recessed and rutile appearing raised above the surface. (C) The image on right is a phase map generated via machine learning determination of the Bravais lattice for each diffraction pattern. Black scale bar in (A) is 100 $\mu$ m.

Our methodology enables high-throughput and autonomous determination of crystal symmetry in electron backscatter diffraction. We found that the CNN identifies specific features resulting from unique crystal symmetry operations within the diffraction pattern images. We believe the method can be expanded to encompass a multi-tiered model to determine the complete crystal structure. Improvements building off of our methodology include neural network architectures specifically designed for specific multi-phase samples or through incorporating other data like phase chemistry into the algorithm. We believe a wide range of research areas including pharmacology, structural biology, and geology would benefit by using automated algorithms that reduce the amount of time required for structural identification.

## 5.2 Materials and Methods

### **5.2.1 Electron backscatter diffraction pattern collection**

Elements or alloys of known crystal structure were collected and polished for EBSD. High purity samples of each metal, alloy, or ceramic material ( $\text{Cr}_3\text{Si}$ ,  $\text{Mo}_3\text{Si}$ ,  $\text{V}_8\text{C}_7$ , Si, Ni, Al, NbC, TaC, W, Ta,  $\beta$ -Ti, Fe, WC, Ti,  $\text{Al}_2\text{O}_3$ , Sn,  $\text{Mo}_5\text{Si}_3$ ,  $\text{Mo}_2\text{C}$ ,  $\text{TiSi}_2$ ,  $\text{Al}_3\text{Zr}_2$ ,  $\text{Ni}_2\text{V}$ ,  $\text{MoPt}_2$ ,  $\text{AuSn}_4$ ,  $\text{ZrSi}_2$ ,  $\text{Nb}_2\text{O}_5$ ,  $\text{Fe}_4\text{Al}_{13}$ ,  $\text{FeNi}_3$ , TiC, NiAl, and  $\text{Ni}_3\text{Al}$ ), and carefully selected geological specimens (quartz, ilmenite, rutile, anatase, forsterite, enstatite, malachite, jadeite, diopside, and anorthite) were selected. Supplementary Figure 25 displays the materials and their Bravais lattice. In multi-phase geological specimens, only patterns identifiable to one of the phases were analyzed. EBSPs were collected in a Thermo-Fisher (formerly FEI) APREO scanning electron microscope (SEM) equipped with an Oxford Symmetry EBSD detector. Supplementary Figure 26 details the experimental setup within the SEM chamber. The Oxford Symmetry EBSD detector was utilized in high resolution ( $1244 \times 1024$ ) mode with frame averaging. EBSPs were collected from multiple large areas with large step sizes chosen to maximize the number of differently oriented patterns taken from unique grains. After collecting high resolution EBSPs from each material, all patterns collected were exported as tiff images. Supplementary Figure 27 displays example images of the high-resolution diffraction patterns collected. None of the collected patterns were excluded from training, testing, or validation studies unless their origin was uncertain (i.e. non-indexable patterns from multi-phase geological specimens). That is to say, the collected data was not filtered for pattern quality via any means, and the library of images for each phase may contain partial or low-quality diffraction patterns, which will decrease the accuracy of their identification.

### **5.2.2 Neural network architecture and training procedure**

Two well-studied CNN architectures designed for broad image classification tasks, ResNet50 [52] and Xception [53], were utilized in this work. These two architectures were

selected for their ease of replication by the scientific community and as baselines for future model exploration and development. They also serve as an excellent starting point for efficient pruning to a more efficient model via the Lottery Ticket Hypothesis [58]. The Lottery Ticket Hypothesis states that there exists a smaller subnetwork within an architecture capable of achieving a similar accuracy on the specific task and that the optimal subnetwork can only be located by pruning after training the model on that task. Training was performed using Adam optimization [59], with batches of 32 images, and a minimum delta of 0.001 in the validation loss for early stopping criteria. The CNNs were implemented with TensorFlow [60] and Keras [61]. The minimum pixel resolution required for the trained models has been set to  $246 \times 299$  (height  $\times$  width). Refer to Supplementary Figure 28 for a detailed schematic of the test-train workflow and refer to Supplementary Figure 29 for the connection between the 7 crystal structures, 14 Bravais lattices, 32 crystal point groups, and 230 crystal space groups.

For the 14 Bravais lattices, two materials from each class were selected as training materials (e.g. nickel and aluminum for face centered cubic). Two materials were selected for training to increase the network's focus on common symmetry elements in the diffraction patterns. For training the model, only 640 diffraction patterns from each material were randomly selected, and the remaining diffraction patterns from those same materials were employed as the test set to gauge the model's accuracy after training. An additional distinct set of materials, not used in training, were also utilized to demonstrate the utility of the model.

In the space group classification example, the ResNet50 convolutional neural network was trained as a binary classifier between two space groups within the same point group. The two models were trained using diffraction patterns from FeNi<sub>3</sub> (L1<sub>2</sub>, space group 221) and Cr<sub>3</sub>Si (primitive cubic, space group 223); and NiAl (B2, space group 221) and Ta (BCC, space group

229). The trained model was then blind tested on two materials that were new to the machine learning model but within the same space groups: Ni<sub>3</sub>Al (L1<sub>2</sub>, space group 221) and Mo<sub>3</sub>Si (primitive cubic, space group 223); and FeAl (B2, space group 221) and W (BCC, space group 229).

### **5.2.3 Validation studies**

The accuracy of the trained models was evaluated using new diffraction patterns collected from each of the materials utilized to train the deep neural network as well as diffraction patterns from materials the model had not previously encountered. Furthermore, heatmap overlays giving class-specific gradient information at the classification layer of the Xception architecture were produced using Grad-CAM [54]. This technique reveals the importance of regions and features present in the diffraction pattern to assist in the determination of why the algorithm classified the image to a particular structure.

### **5.2.4 The classification model**

Having collected approximately 400,000 EBSD patterns from 40 different materials, it is necessary to classify each of these images based solely on the information contained in the image. The classical computer vision approach is to manually engineer features and use a discriminatory model to make the ultimate decision about the Bravais lattice to which the source image belongs. Such an approach would require a multitude of heuristics—such as detecting Kikuchi bands, accounting for their slight orientation changes, looking for symmetry in the image, etc., and carrying the burden of developing the logic that defines these abstract qualities. The process of generating the corresponding computer logic is even more challenging when one considers the smallest changes in diffraction patterns resulting from orientation changes or minute atomic position differences and defects in materials of the same crystal structure. Thus, the traditional

computer vision approach becomes less feasible as the number of categories grows, cannot be generalized to other crystal classes (including expanding these capabilities to point and space groups), and lacks a procedure to systematically improve prediction capabilities. On the other hand, one could use a technique that determines its own internal representation of the data so long as it performs well at the discrimination task. This is the underlying principle behind deep representation learning (i.e. deep neural networks) [62]. Such methods allow the model to find patterns that may be unintuitive or too nuanced for humans to discern. Other discovered features might be obvious to experts, but difficult to translate into specific logic.

These deep learning systems take in raw data and automatically discover, through filters learned via backpropagation [63], the abstract representations that maximize classification performance. In convolutional neural networks, such as the ones used in this work, the early layers typically learn to look for the presence or absence of edges or curves, while the later layers assemble these motifs into representative combinations and eventually familiar objects. In this case, these more familiar objects are the symmetry elements present in the diffraction pattern encoding a Bravais lattice. A schematic representation of the CNN utilized in this work is addressed in Figure 5.1 of the main text.

### **5.3 Acknowledgements**

The authors would like to thank Dr. Emily Chin of Scripps Institute of Oceanography and her students for their assistance collecting and identifying mineral and geological samples.

Chapter 5, in full, is a reprint of the material “Crystal symmetry determination in electron diffraction using machine learning”, as it appears in *Science*. K. Kaufmann, C. Zhu, A. S. Rosengarten, D. Maryanovsky, T. J. Harrington, E. Marin, and K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

## 5.4 References

- [1] I.M. Abrams, J.W. McBain, D.F. Parsons, A closed cell for electron microscopy., *Science*. 100 (1944) 273–4. doi:10.1126/science.100.2595.273.
- [2] T. Walz, T. Hirai, K. Murata, J.B. Heymann, K. Mitsuoka, Y. Fujiyoshi, B.L. Smith, P. Agre, A. Engel, The three-dimensional structure of aquaporin-1, *Nature*. 387 (1997) 624–627. doi:10.1038/42512.
- [3] J.A. Rodriguez, M.I. Ivanova, M.R. Sawaya, D. Cascio, F.E. Reyes, D. Shi, S. Sangwan, E.L. Guenther, L.M. Johnson, M. Zhang, L. Jiang, M.A. Arbing, B.L. Nannenga, J. Hattne, J. Whitelegge, A.S. Brewster, M. Messerschmidt, S. Boutet, N.K. Sauter, T. Gonen, D.S. Eisenberg, Structure of the toxic core of  $\alpha$ -synuclein from invisible crystals, *Nature*. 525 (2015) 486–490. doi:10.1038/nature15368.
- [4] K. Hedberg, L. Hedberg, D.S. Bethune, C.A. Brown, H.C. Dorn, R.D. Johnson, M. DE Vries, Bond lengths in free molecules of buckminsterfullerene, c60, from gas-phase electron diffraction., *Science*. 254 (1991) 410–2. doi:10.1126/science.254.5030.410.
- [5] C.I. Blaga, J. Xu, A.D. DiChiara, E. Sistrunk, K. Zhang, P. Agostini, T.A. Miller, L.F. DiMauro, C.D. Lin, Imaging ultrafast molecular dynamics with laser-induced electron diffraction, *Nature*. 483 (2012) 194–197. doi:10.1038/nature10820.
- [6] S.W. Hui, D.F. Parsons, Electron diffraction of wet biological membranes., *Science*. 184 (1974) 77–8. doi:10.1126/SCIENCE.184.4132.77.
- [7] L. Palatinus, P. Brázda, P. Boullay, O. Perez, M. Klementová, S. Petit, V. Eigner, M. Zaarour, S. Mintova, Hydrogen positions in single nanocrystals revealed by electron diffraction., *Science*. 355 (2017) 166–169. doi:10.1126/science.aak9652.
- [8] M. Bestmann, S. Piazzolo, C.J. Spiers, D.J. Prior, Microstructural evolution during initial stages of static recovery and recrystallization: new insights from in-situ heating experiments combined with electron backscatter diffraction analysis, *J. Struct. Geol.* 27 (2005) 447–457. doi:10.1016/J.JSG.2004.10.006.
- [9] W.J. Huang, R. Sun, J. Tao, L.D. Menard, R.G. Nuzzo, J.M. Zuo, Coordination-dependent surface atomic contraction in nanocrystals revealed by coherent diffraction, *Nat. Mater.* 7 (2008) 308–313. doi:10.1038/nmat2132.
- [10] P.A. Midgley, R.E. Dunin-Borkowski, Electron tomography and holography in materials science, *Nat. Mater.* 8 (2009) 271–280. doi:10.1038/nmat2406.
- [11] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, G. Ceder, Predicting Crystal Structures with Data Mining of Quantum Calculations, *Phys. Rev. Lett.* 91 (2003) 135503. doi:10.1103/PhysRevLett.91.135503.
- [12] J.F. Nye, *Physical properties of crystals: their representation by tensors and matrices*, Clarendon Press, Oxford, Great Britain, Oxford, 1985. <https://global.oup.com/academic/product/physical-properties-of-crystals->

9780198511656?cc=us&lang=en& (accessed October 7, 2018).

- [13] J.C. Tan, A.K. Cheetham, Mechanical properties of hybrid inorganic–organic framework materials: establishing fundamental structure–property relationships, *Chem. Soc. Rev.* 40 (2011) 1059. doi:10.1039/c0cs00163e.
- [14] B. Ollivier, R. Retoux, P. Lacorre, D. Massiot, G. Férey, Crystal structure of  $\kappa$ -alumina: an X-ray powder diffraction, TEM and NMR study, *J. Mater. Chem.* 7 (1997) 1049–1056. doi:10.1039/a700054e.
- [15] J.E. Post, D.R. Veblen, Crystal structure determinations of synthetic sodium, magnesium, and potassium birnessite using TEM and the Rietveld method, 1990. [http://rruff.info/uploads/AM75\\_477.pdf](http://rruff.info/uploads/AM75_477.pdf) (accessed October 7, 2018).
- [16] S.J. Andersen, H.W. Zandbergen, J. Jansen, C. Træholt, U. Tundal, O. Reiso, The crystal structure of the  $\beta''$  phase in Al–Mg–Si alloys, *Acta Mater.* 46 (1998) 3283–3298. doi:10.1016/S1359-6454(97)00493-X.
- [17] K.S. Vecchio, D.B. Williams, Convergent beam electron diffraction analysis of the T 1 (Al<sub>2</sub>CuLi) phase in Al–Li–Cu alloys, *Metall. Trans. A.* 19 (1988) 2885–2891. doi:10.1007/BF02647714.
- [18] K.S. Vecchio, D.B. Williams, Convergent beam electron diffraction study of Al<sub>3</sub>Zr in Al–Zr AND Al–Li–Zr alloys, *Acta Metall.* 35 (1987) 2959–2970. doi:10.1016/0001-6160(87)90295-1.
- [19] D.B. Williams, A.R. Pelton, R. Gronsky, eds., *Images of Materials*, Oxford University Press, 1991.
- [20] S. Nishikawa, S. Kikuchi, Diffraction of cathode rays by calcite, *Nature.* 122 (1928) 726. doi:10.1038/122726a0.
- [21] M.N. Alam, M. Blackman, D.W. Pashley, High-Angle Kikuchi Patterns, *Proc. R. Soc. A Math. Phys. Eng. Sci.* 221 (1954) 224–242. doi:10.1098/rspa.1954.0017.
- [22] J.A. Venables, C.J. Harland, Electron back-scattering patterns—a new technique for obtaining crystallographic information in the scanning electron microscope, *Philos. Mag.* 27 (1973) 1193–1200. doi:10.1080/14786437308225827.
- [23] D.J. Dingley, A comparison of diffraction techniques for the SEM, *Scan. Electron Microsc.* 4 (1981) 273–286.
- [24] B.L. Adams, S.I. Wright, K. Kunze, Orientation imaging: The emergence of a new microscopy, *Metall. Trans. A.* 24 (1993) 819–831. doi:10.1007/BF02656503.
- [25] N.C. Krieger Lassen, D. Juul Jensen, K. Conradsen, Image processing procedures for analysis of electron back scattering patterns, *Scan. Microsc.* 6 (1992) 115–121. <http://ci.nii.ac.jp/naid/10030703203/en/> (accessed October 20, 2018).
- [26] J. Goulden, P. Trimby, A. Bewick, The Benefits and Applications of a CMOS-based EBSD

- Detector, *Microsc. Microanal.* 24 (2018) 1128–1129. doi:10.1017/S1431927618006128.
- [27] A.J. Schwartz, M. Kumar, B.L. Adams, D.P. Field, Electron backscatter diffraction in materials science, *Electron Backscatter Diffr. Mater. Sci.* (2009) 1–403. doi:10.1007/978-0-387-88136-2.
- [28] V. Randle, O. Engler, O. Engler, *Introduction to Texture Analysis*, CRC Press, 2014. doi:10.1201/9781482287479.
- [29] C. Zhu, V. Livescu, T. Harrington, O. Dippo, G.T. Gray, K.S. Vecchio, Investigation of the shear response and geometrically necessary dislocation densities in shear localization in high-purity titanium, *Int. J. Plast.* 92 (2017) 148–163. doi:10.1016/J.IJPLAS.2017.03.009.
- [30] C. Zhu, T. Harrington, G.T. Gray, K.S. Vecchio, Dislocation-type evolution in quasi-statically compressed polycrystalline nickel, *Acta Mater.* 155 (2018) 104–116. doi:10.1016/J.ACTAMAT.2018.05.022.
- [31] C. Zhu, T. Harrington, V. Livescu, G.T. Gray Iii, K.S. Vecchio, Determination of geometrically necessary dislocations in large shear strain localization in aluminum, *Acta Mater.* 118 (2016) 383–394. doi:10.1016/j.actamat.2016.07.051.
- [32] D. Chen, J.C. Kuo, W.T. Wu, Effect of microscopic parameters on EBSD spatial resolution, *Ultramicroscopy.* 111 (2011) 1488–1494. doi:10.1016/j.ultramic.2011.06.007.
- [33] M.M. Nowell, S.I. Wright, Phase differentiation via combined EBSD and XEDS, *J. Microsc.* 213 (2004) 296–305. doi:10.1111/j.0022-2720.2004.01299.x.
- [34] F. Ram, M. De Graef, Phase differentiation by electron backscatter diffraction using the dictionary indexing approach, *Acta Mater.* 144 (2018) 352–364. doi:10.1016/J.ACTAMAT.2017.10.069.
- [35] D.J. Dingley, S.I. Wright, Phase Identification Through Symmetry Determination in EBSD Patterns, in: A. Schwartz, M. Kumar, B. Adams, D. Field (Eds.), *Electron Backscatter Diffr. Mater. Sci.*, Springer, Boston, MA, 2009: pp. 97–107. doi:10.1007/978-0-387-88136-2.
- [36] R.P. Goehner, J.R. Michael, Phase Identification in a Scanning Electron Microscope Using Backscattered Electron Kikuchi Patterns., *J. Res. Natl. Inst. Stand. Technol.* 101 (1996) 301–308. doi:10.6028/jres.101.031.
- [37] L. Li, M. Han, Determining the Bravais lattice using a single electron backscatter diffraction pattern, *J. Appl. Crystallogr.* 48 (2015) 107–115. doi:10.1107/S1600576714025989.
- [38] A.A. White, Big data are shaping the future of materials science, *MRS Bull.* 38 (2013) 594–595. doi:10.1557/mrs.2013.187.
- [39] F. Ren, L. Ward, T. Williams, K.J. Laws, C. Wolverton, J. Hattrick-Simpers, A. Mehta, Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments, *Sci. Adv.* 4 (2018) eaaq1566. doi:10.1126/sciadv.aaq1566.
- [40] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, Accelerated search

- for materials with targeted properties by adaptive design, *Nat. Commun.* 7 (2016) 11241. doi:10.1038/ncomms11241.
- [41] A.O. Oliynyk, E. Antono, T.D. Sparks, L. Ghadbeigi, M.W. Gaultois, B. Meredig, A. Mar, High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds, *Chem. Mater.* 28 (2016) 7324–7331. doi:10.1021/acs.chemmater.6b02724.
- [42] S. Sanvito, C. Oses, J. Xue, A. Tiwari, M. Zic, T. Archer, P. Tozman, M. Venkatesan, M. Coey, S. Curtarolo, Accelerated discovery of new magnets in the Heusler alloy family, *Sci. Adv.* 3 (2017). doi:10.1126/sciadv.1602241.
- [43] A. Mansouri Tehrani, A.O. Oliynyk, M. Parry, Z. Rizvi, S. Couper, F. Lin, L. Miyagi, T.D. Sparks, J. Brgoch, Machine Learning Directed Search for Ultraincompressible, Superhard Materials, *J. Am. Chem. Soc.* 140 (2018) 9844–9853. doi:10.1021/jacs.8b02717.
- [44] T. Xie, J.C. Grossman, Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties, *Phys. Rev. Lett.* 120 (2018) 145301. doi:10.1103/PhysRevLett.120.145301.
- [45] J. Jung, J.I. Yoon, H.K. Park, J.Y. Kim, H.S. Kim, An efficient machine learning approach to establish structure-property linkages, *Comput. Mater. Sci.* 156 (2019) 17–25. doi:10.1016/J.COMMATSCI.2018.09.034.
- [46] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, Universal fragment descriptors for predicting properties of inorganic crystals, *Nat. Commun.* 8 (2017). doi:10.1038/ncomms15679.
- [47] W. Ye, C. Chen, Z. Wang, I.-H. Chu, S.P. Ong, Deep neural networks for accurate predictions of crystal stability, *Nat. Commun.* 9 (2018) 3800. doi:10.1038/s41467-018-06322-x.
- [48] B.L. DeCost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, *Comput. Mater. Sci.* 110 (2015) 126–133. doi:10.1016/J.COMMATSCI.2015.08.011.
- [49] B.L. DeCost, T. Francis, E.A. Holm, High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel, (2018). doi:10.2118/162526-PA.
- [50] A. Ziletti, D. Kumar, M. Scheffler, L.M. Ghiringhelli, Insightful classification of crystal structures using deep learning, *Nat. Commun.* 9 (2018) 2775. doi:10.1038/s41467-018-05169-6.
- [51] F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N.T.P. Hartono, S. Ramasamy, B.L. DeCost, S.I.P. Tian, G. Romano, A. Gilad Kusne, T. Buonassisi, Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks, *Npj Comput. Mater.* 5 (2019). doi:10.1038/s41524-019-0196-x.
- [52] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: *Proc. IEEE Int. Conf. Comput. Vis.*, 2016: pp. 770–778. [118](http://image-</a></p></div><div data-bbox=)

- net.org/challenges/LSVRC/2015/ (accessed February 1, 2019).
- [53] F. Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, in: Proc. IEEE Int. Conf. Comput. Vis., 2017: pp. 1251–1258. [http://openaccess.thecvf.com/content\\_cvpr\\_2017/papers/Chollet\\_Xception\\_Deep\\_Learning\\_CVPR\\_2017\\_paper.pdf](http://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf) (accessed February 1, 2019).
- [54] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, in: Proc. IEEE Int. Conf. Comput. Vis., 2017: pp. 618–626. doi:10.1109/ICCV.2017.74.
- [55] T. Karthikeyan, M.K. Dash, S. Saroja, M. Vijayalakshmi, Evaluation of misindexing of EBSD patterns in a ferritic steel, J. Microsc. 249 (2013) 26–35. doi:10.1111/j.1365-2818.2012.03676.x.
- [56] C.L. Chen, R.C. Thomson, The combined use of EBSD and EDX analyses for the identification of complex intermetallic phases in multicomponent Al-Si piston alloys, J. Alloys Compd. 490 (2010) 293–300. doi:10.1016/j.jallcom.2009.09.181.
- [57] S. McLaren, S.M. Reddy, Automated mapping of K-feldspar by electron backscatter diffraction and application to <sup>40</sup>Ar/<sup>39</sup>Ar dating, J. Struct. Geol. 30 (2008) 1229–1241. doi:10.1016/J.JSG.2008.05.008.
- [58] K. He, X. Zhang, S. Ren, J. Sun, in *Proceedings of the IEEE International Conference on Computer Vision* (2016; <http://image-net.org/challenges/LSVRC/2015/>), pp. 770–778.
- [59] F. Chollet, in *Proceedings of the IEEE International Conference on Computer Vision* (2017; [http://openaccess.thecvf.com/content\\_cvpr\\_2017/papers/Chollet\\_Xception\\_Deep\\_Learning\\_CVPR\\_2017\\_paper.pdf](http://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf)), pp. 1251–1258.
- [60] J. Frankle, G. K. Dziugaite, D. M. Roy, M. Carbin, The Lottery Ticket Hypothesis at Scale (2019) (available at <http://arxiv.org/abs/1903.01611>).
- [61] D. P. Kingma, J. Ba, in *International Conference on Learning Representations* (2014; <http://arxiv.org/abs/1412.6980>), pp. 1–15.
- [62] M. Abadi *et al.*, in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation* (2016; <https://tensorflow.org>), pp. 265–283.
- [63] F. Chollet, Keras (2015), (available at <https://github.com/keras-team/keras>).
- [64] Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature*. **521**, 436–444 (2015).
- [65] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors. *Nature*. **323**, 533–536 (1986).
- [66] F. Hoffmann, The Space Group List Project (2014), (available at <https://crystalsymmetry.wordpress.com/230-2/>).
- [67] K. Momma, F. Izumi, IUCr, VESTA 3 for three-dimensional visualization of crystal,

volumetric and morphology data. *J. Appl. Crystallogr.* **44**, 1272–1276 (2011).

## Chapter 6 Deep neural network enabled space group identification in EBSD

Kevin Kaufmann<sup>1</sup>, Chaoyi Zhu<sup>2</sup>, Alexander S. Rosengarten<sup>1</sup>, and Kenneth S. Vecchio<sup>1,2</sup>

<sup>1</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA

<sup>2</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA

### Abstract

Electron backscatter diffraction (EBSD) is one of the primary tools in materials development and analysis. The technique can perform simultaneous analyses at multiple length scales, providing local sub-micron information mapped globally to centimeter scale. Recently, a series of technological revolutions simultaneously increased diffraction pattern quality and collection rate. After collection, current EBSD pattern indexing techniques (whether Hough-based or dictionary pattern matching based) are capable of reliably differentiating between a ‘user selected’ set of phases, if those phases contain sufficiently different crystal structures. EBSD is currently less well-suited for the problem of phase identification where the phases in the sample are unknown. A pattern analysis technique capable of phase identification, utilizing the information rich diffraction patterns potentially coupled with other data, would enable EBSD to become a high-throughput technique replacing many slower (x-ray diffraction) or more expensive (neutron diffraction) methods. We utilize a machine learning technique to develop a general methodology for the space group classification of diffraction patterns. This is demonstrated within the  $(4/m, \bar{3}, 2/m)$  point group. We evaluate the machine learning algorithm’s performance in

real-world situations using materials outside the training set, simultaneously elucidating the role of atomic scattering factors, orientation, and pattern quality on classification accuracy.

## **6.1 Introduction**

Conventional electron backscatter diffraction (EBSD) is a standard scanning electron microscope (SEM)-based technique used to determine the three-dimensional orientation of individual grains in crystalline materials. Phase differentiation is a necessary component of this technique for analysis of multi-phase samples and has been of particular interest in the community [1–4]. However, determining the underlying structure of unknown materials (phase identification) has remained a challenge in EBSD. Currently, Hough- or dictionary-based pattern matching approaches require a ‘user-defined’ set of phases at the onset of analysis [5–9]. The Hough transform-based method is the most common approach to pattern matching used in commercial systems. Hough-based indexing looks for the diffraction maxima and creates a sparse representation of the diffraction pattern [10]. The sparse representation is used with a look-up table of interplanar angles constructed from the set of selected reflectors for phases specified by the user. Beyond requiring the user to have sufficient knowledge of the sample before beginning analysis, the process remains susceptible to structural misclassification [11–13]. Potential phase identification solutions leveraging energy dispersive X-ray spectroscopy (EDS) or wavelength dispersive X-ray spectroscopy (WDS) have been previously demonstrated and adopted commercially [14–16]. These strategies are effective for single-point identification of crystal structure subject to an expert user’s ability to select the correct phase from the potential matches. Methods utilizing hand-drawn lines overlaid on individual Kikuchi diffraction patterns have been developed for determining the Bravais lattice or point group [15,17–19]. These represent important milestones for phase identification from EBSD patterns; however, they remain limited

by at least one of the following: analysis time per pattern, the need for an expert crystallographer, or necessitating multiples of the same diffraction pattern with different SEM settings [18].

Going beyond Bravais lattice and point group identification to determine the space group of the crystal phases is in general a challenging task. X-ray diffraction (XRD) and transmission electron microscopy (TEM)-based convergent beam electron diffraction (CBED) are the most common solutions [20,21]. X-ray diffraction becomes challenging in multi-phase samples owing to overlapping peaks, texture effects, large numbers of peaks for low-symmetry phases, and pattern refinement. Moreover, the information in XRD patterns does not provide detailed microstructure information, such as morphology, location, or grain statistics that can be observed using EBSD's better spatial resolution. However, with careful analysis it is possible to extract crystal symmetry, phase fractions, average grain size, and dimensionality from XRD [22]. On the other hand, CBED is limited by intense sample preparation, small areas of analysis, and substantial operator experience [23–25]. In comparison, EBSD can be performed on large samples [26–29], including three-dimensional EBSD [30], with high precision ( $\sim 2^\circ$ ), high misorientation resolution ( $0.2^\circ$ ) and high spatial resolution ( $\sim 40$  nm) [31]. Furthermore, the diffraction patterns collected in EBSD contain many of the same features observed in CBED, including excess and deficiency lines and higher-order Laue zone (HOLZ) rings [17,32]. Identification of a space group dependent property (chirality) was recently demonstrated in quartz using single experimental EBSD patterns [33]. However, state of the art EBSD software cannot currently classify the collected diffraction patterns to their space group and can misidentify the Bravais lattice. Common examples encountered in EBSD include the difficulty distinguishing  $L1_2$  (space group 221) from FCC (space group 225) or B2 (space group 221) from BCC (space group 229) [34–36]. Inspired by the similarities between CBED and EBSD patterns [17,24,37], we propose applying an image recognition technique from

the machine learning field to provide an opportunity for real-time space group recognition in EBSD. Given that CBED patterns contain sufficient 3-D structural diffraction detail to allow structure symmetry determination to the space group level (Vecchio and Williams, 1987), and given the considerable similarity between CBED and EBSD patterns, along with the much larger angular view captured in EBSD patterns, and the demonstration of chirality determination in experimental patterns [33], it is reasonable to consider space group differentiation in EBSD patterns.

The recent advent of deep neural networks, such as the convolutional neural network (CNN) designed for image data, offer an opportunity to address many of the challenges to autonomously extracting information from diffraction data [38,39]. CNNs are of particular interest owing to multiple advantages over classical computer vision techniques, which require a multitude of heuristics [40–44] — such as detecting Kikuchi bands, accounting for orientation changes, determining band width, etc., — and carrying the burden of developing the logic that defines these abstract qualities. Instead, this deep learning technique (e.g. CNNs) determines its own internal representation of the data, via backpropagation [45], such that it maximizes performance at the discrimination task. This is the underlying principle behind deep representation learning (i.e. deep neural networks) [46]. CNNs operate by convolving learnable filters across the image, and the scalar product between the filter and the input at every position, or ‘patch’, is computed to form a feature map. The units in a convolutional layer are organized as feature maps, and each feature map is connected to local patches in the previous layer through a set of weights called a filter bank. All units in a feature map share the same filter banks, while different feature maps in a convolutional layer use different filter banks. Pooling layers are placed after convolutional layers to down sample the feature maps and produce coarse grain representations and spatial information

about the features in the data. The key aspect of deep learning is that these layers of feature detection nodes are not programmed into lengthy scripts or hand-designed feature extractors, but instead are ‘learned’ from the training data. CNNs are further advantageous over other machine learning models since they can operate on the unprocessed image data and the same architectures are applicable to diverse problems. For example, a similar methodology was recently demonstrated to identify the Bravais lattice of an experimental EBSD pattern [47]. Another recent example has applied a CNN to simulated EBSD patterns from eight materials that are typically confused in conventional analyses [4] with exceptional success.

In the present work, it is demonstrated that convolutional neural networks can be constructed to rapidly classify the space group of singular EBSD patterns. This process is capable of being utilized in a real-time analysis and high-throughput manner in line with recent advancements in EBSD technology [48]. The method is described in detail and demonstrated on a dataset of samples within the  $(4/m, \bar{3}, 2/m)$  point group. The dataset utilized further allows for studying the impact that scattering intensity factors [49,50] have on classification accuracy. Heavier materials tend to have higher atomic scattering factors in electron diffraction, resulting in more visible reflectors for a diffraction pattern from the same space group and three-dimensional orientation. Training on only low or high atomic number materials is found to reduce future classification accuracy. Increasing the range of atomic scattering factors utilized in the training set, even if the minimum and maximum  $\bar{Z}$  materials within the space group are not both included, alleviates the effects of this physical phenomenon on the neural network’s classification abilities. Furthermore, this work provides a brief analysis on the effect of two common pattern quality metrics and train/test orientation differences on classification accuracy. The inner workings of this deep neural network-based method are studied using visual feature importance analysis. By

allowing a machine learning algorithm to perform EBSD pattern classification to the space group level, a significant advancement in the utilization and accuracy of phase identification by EBSD can be achieved. When combined with chemical information on phases, for example from energy dispersive x-ray spectroscopy, this approach can lead to automated phase identification.

## **6.2 Materials and methods**

### **6.2.1 Materials**

Eighteen different single-phase materials, comprising 6 of the 10 space groups within the  $(4/m, \bar{3}, 2/m)$  point group were selected for demonstrating the proposed space group classification methodology. Suitable samples for the remaining 4 space groups could not be obtained. This point group was chosen as it contains space groups that are very similar structurally and represent a significant classification challenge for conventional EBSD. The method of fabrication and homogenization (if applicable) for each sample are listed in Supplementary Table 10. The homogenization heat treatments were performed for three weeks in an inert atmosphere at temperatures guided by each phase diagram.

### **6.2.2 Electron backscatter diffraction pattern collection**

EBSD patterns (EBSPs) were collected in a Thermo Fisher (formerly FEI) Apreo scanning electron microscope (SEM) equipped with an Oxford Symmetry EBSD detector. The Oxford Symmetry EBSD detector was utilized in high resolution (1244x1024) mode. The geometry of the setup was held constant for each experiment. The working distance was  $18.1 \text{ mm} \pm 0.1 \text{ mm}$ . Aztec was used to set the detector insertion distance to 160.2 and the detector tilt to -3.1. The imaging parameters were 20kV accelerating voltage, 51nA beam current,  $0.8 \text{ ms} \pm 0.1 \text{ ms}$  dwell time, and 30 pattern averaging.

After collecting high resolution EBSPs from each material, all patterns collected were exported as tiff images. Supplementary Figure 30 in the Appendix shows example images of the high-resolution diffraction patterns collected from the materials utilized in this study, organized by space group. All collected data for each material was individually assessed by the neural network, and the collection of images for each sample may contain partial or low-quality diffraction patterns, which will decrease the accuracy of their identification. See Supplementary Figure 31 for the inverse pole figures (IPFs) for each material. The IPFs were constructed using the MTEX software package [51]. The data in Supplementary Figure 31 has been plotted first using the scale bars set by MTEX to show the fine distribution of the data, and then with the scale bar fixed from 0 to 5 times random for the purpose of demonstrating the data does not approach medium texture levels. Analysis shows the experimental datasets have very low texture, typically in the range of 2-3 multiples of uniform distribution (M.U.D.) also referred to as times random. Typically, 5-10 is considered medium texture and greater than 10 is considered strong texture. See Supplementary Figure 32 for histograms of mean angular deviation (MAD) and band contrast (BC) to compare pattern quality for each material. Each plot is also annotated with the mean ( $\mu$ ) and standard deviation ( $\sigma$ ). The purpose for not filtering the test data was to assess the model as it would be applied in practice. Only the training sets were visually inspected to confirm high-quality diffraction patterns (no partial patterns) were utilized in fitting the neural network.

### **6.2.3 Neural network architecture.**

The well-studied convolutional neural network architecture Xception [52] was selected as the basis architecture for fitting a model that determines which space group a diffraction pattern originated from. Supplementary Figure 33 in the appendix details a schematic of the convolutional

neural network operating on an EBSP. For a complete description of the Xception architecture, please refer to Figure 5 in [52].

#### **6.2.4 Neural network training**

Training was performed using 400 diffraction patterns per space group, evenly divided between the number of materials the model had access to during training. For example, if the model was given two materials during training, 200 diffraction patterns per material were made available. The validation set contained 100 diffraction patterns per space group, equivalent to the standard 80:20 train/validation split. The test set contains the rest of the patterns that were not used for training or validation. Model hyperparameters were selected or tuned as follows. Adam optimization with a learning rate of 0.001 [53], and a minimum delta of 0.001 as the validation loss were employed for stopping criteria. The weight decay was set to  $1e^{-5}$  following previous optimization work [52]. The CNNs were implemented with TensorFlow [54] and Keras [55]. The code for implementing these models can be found at <https://github.com/krkaufma/Electron-Diffraction-CNN> or Zenodo (DOI: 10.5281/zenodo.3564937).

#### **6.2.5 Diffraction pattern classification**

Each diffraction pattern collected, but not used in training (>140,000 images), was evaluated in a random order by the corresponding trained CNN model without further information. The output classification of each diffraction pattern was recorded and saved in a (.csv) file and are tabulated in the Appendix. All corresponding bar plots of these data were generated with MATLAB. Precision and recall were calculated for each material and each space group using Scikit-learn [56]. Precision (equation 5) for each class (e.g. 225) is defined as the number of correctly predicted images out of all photos predicted to belong to that class (e.g. 225). Recall is

the number of correctly predicted images for each class divided by the actual number of photos for the class (equation 6).

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \quad \text{Equation 5}$$

$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \quad \text{Equation 6}$$

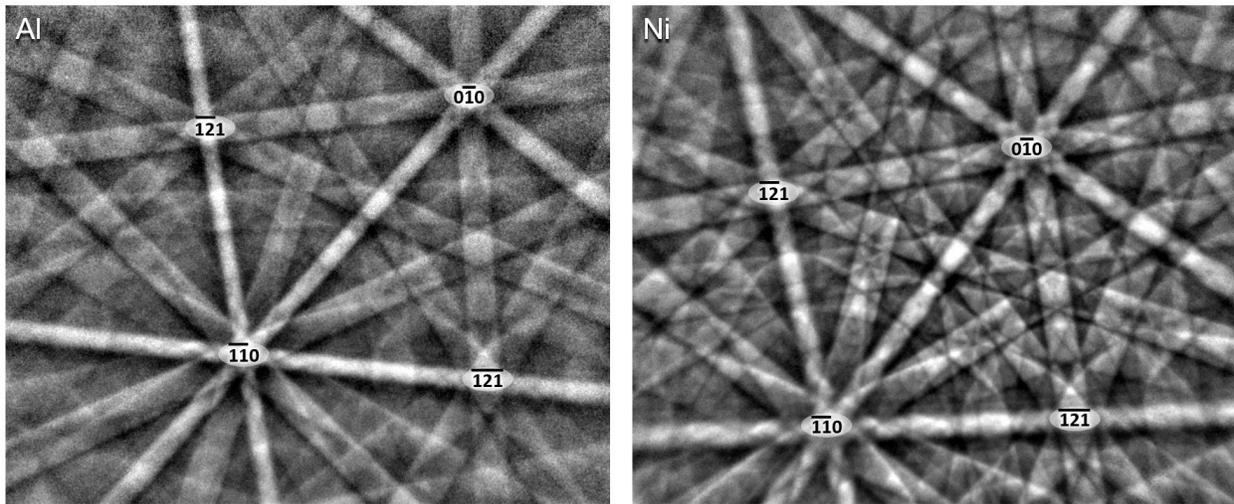
### 6.2.6 Neural network interpretability.

Gradient-weighted class activation mapping (Grad-CAM) was employed to provide insight into the deep neural network [57]. This method computes the importance of local regions in the diffraction image, normalizes them from 0 to 1, and is overlaid as a localization heatmap highlighting the important regions in the image. The ‘guided’ backpropagation modifier was used to achieve pixel-space gradient visualizations and filter out information that suppresses the neurons. This information can be safely filtered out since we are only interested in what image features the neuron detects with respect to the target space group. The gradients flowing into the final convolution layer were targeted for two reasons: (i) convolutional layers naturally retain spatial information unlike the fully-connected layers and (ii) previous works have asserted that with increasing depth of a CNN, higher-level visual constructs are captured [58,59].

## 6.3 Results and discussion

The model is first trained on one material from each of six space groups in the  $(4/m, \bar{3}, 2/m)$  point group; there are 10 space groups within the  $(4/m, \bar{3}, 2/m)$  point group, but suitable samples for 4 of these space groups could not be obtained. The first iteration uses the material with the largest formula weighted atomic number ( $\bar{Z}$ ) (i.e. atomic scattering factor) in order to establish a baseline performance. Ta was used as the training material for space group

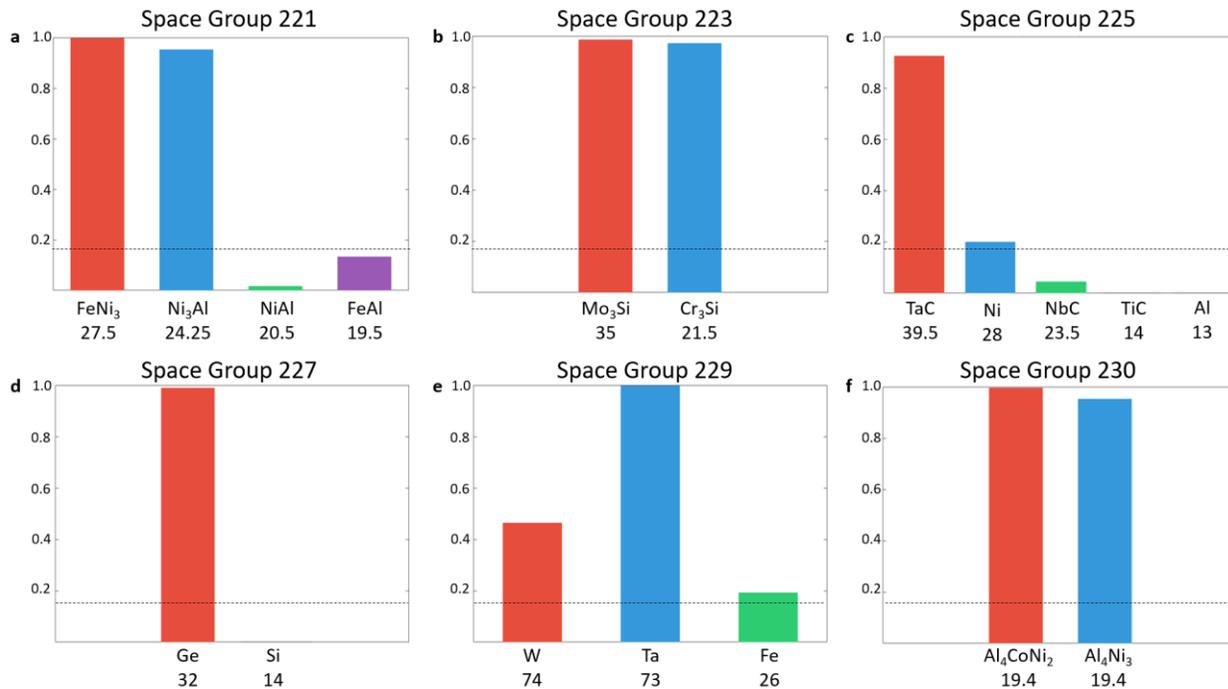
229 instead of W since training would require nearly 50% of the available W diffraction patterns. Since W and Ta only differ by one atomic number, Ta will serve as a good baseline measurement and provide information about small steps in  $\bar{Z}$ . This baseline will be compared to models trained using materials with a reduced number of visible reflectors resulting from lower atomic scattering factors. Fig. 1 demonstrates this effect using a similarly oriented experimental diffraction pattern from Al and Ni with the same four zone axes labeled. The diffraction pattern from Ni has significantly more observable information (Kikuchi bands) since the atomic scattering factor is modulating their intensity. Using a limited number of materials (e.g. one) to represent the entire population of the space group when the neural network is learning the filters that maximize classification accuracy is likely to yield representations that are not as effective when Kikuchi bands are more or less visible based on diffraction intensity.



**Figure 6.1** Effect of atomic scattering factors on observable Kikuchi bands. Al ( $Z = 13$ ) and Ni ( $Z = 28$ ). Despite belonging to the same space group, the symmetry information visible for the lower atomic number Al is noticeably reduced compared with Ni.

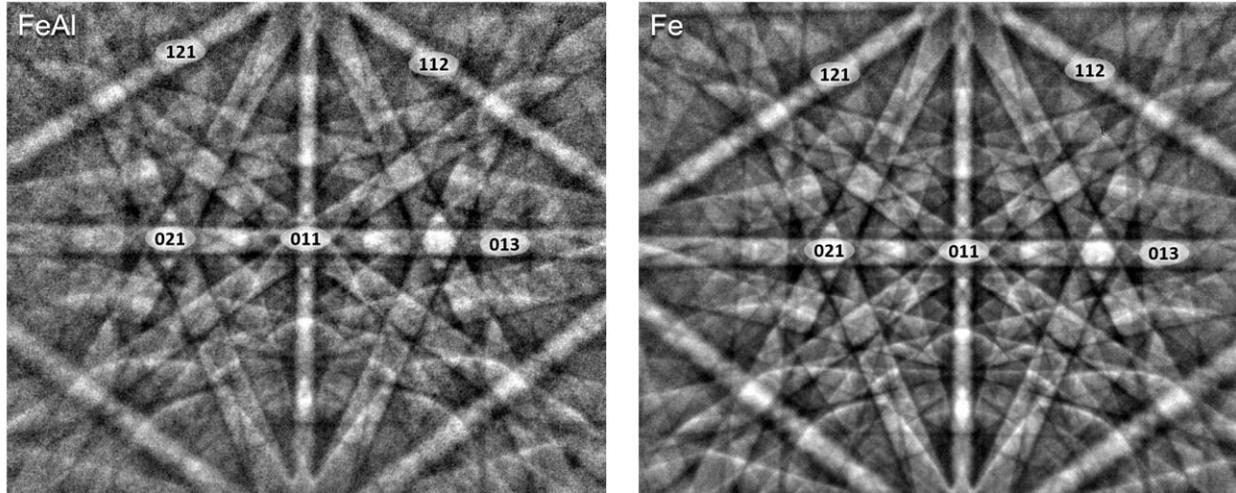
Fig. 2 shows the normalized accuracy of the first iteration of the model at classifying each of the 18 materials to the correct space group after learning from one material per class. Using

only one material during training creates a significant opportunity for the neural network to "invent" representations that are not based on space group symmetry. The number of patterns from a given material that were classified to each space group is given in Supplementary Table 11. As shown in Fig. 2, the model performs significantly better than random guessing (the dashed line at 16.7%) on multiple materials, including 6 of the 12 materials that were not in the training set. Of those 6, it achieves better than 45% accuracy on 4 of them: (Ni<sub>3</sub>Al, Cr<sub>3</sub>Si, W, and Al<sub>4</sub>Ni<sub>3</sub>). This is a good indicator that the model is learning useful features for differentiating the space groups, instead of trivial ones that only work during the training process. Furthermore, analysis of the IPFs in Supplementary Figure 31 shows that these materials can have distinctly different orientation distributions. As an example, the Al<sub>4</sub>Ni<sub>3</sub> sample has a greater distribution of data with orientations near [001] in X, Y, and Z than the material used in training (Al<sub>4</sub>CoNi<sub>2</sub>). Yet, the model achieves 95% accuracy on Al<sub>4</sub>Ni<sub>3</sub>. From Supplementary Table 10, the overall accuracy of the model is 51.2%, well above the 16.7% chance of guessing correctly. It is observed that new materials with similar atomic scattering factors to the training material tend to have higher classification accuracy. Mo<sub>3</sub>Si and Cr<sub>3</sub>Si demonstrate that large differences in atomic scattering factor is not the penultimate factor and does not prevent accurate classification when evaluating new materials. There are also several materials with similar average  $Z$  that are rarely misclassified as one another. These include NiAl ( $\bar{Z} = 20.5$ ), FeAl ( $\bar{Z} = 19.5$ ), and Cr<sub>3</sub>Si ( $\bar{Z} = 21.5$ ) as well as Ni<sub>3</sub>Al ( $\bar{Z} = 24.25$ ) and NbC ( $\bar{Z} = 23.5$ ).



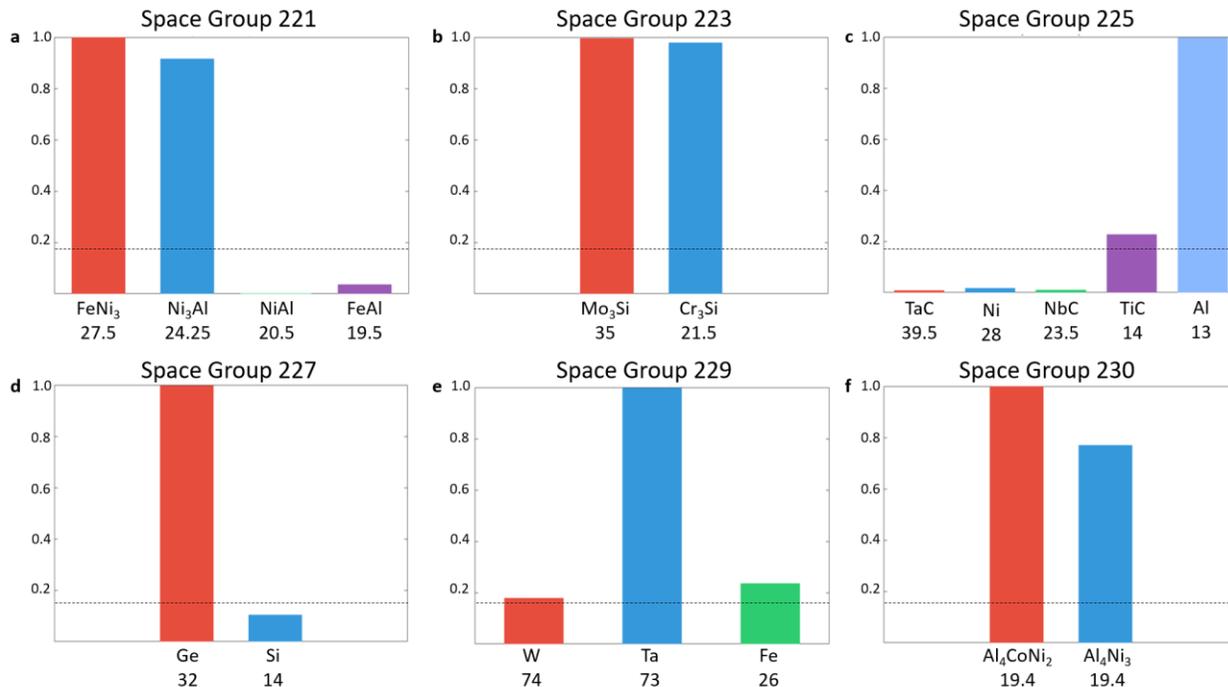
**Figure 6.2** Plots of normalized classification accuracy after fitting the model with the high atomic number materials. (a) Space group 221; trained on FeNi<sub>3</sub>. (b) Space group 223; trained on Mo<sub>3</sub>Si. (c) Space group 225; trained on TaC. (d) Space group 227; trained on Ge. (e) Space group 229; trained on Ta. (f) Space group 230; trained on Al<sub>4</sub>CoNi<sub>2</sub>. The dashed line represents the chance of randomly guessing the correct space group. The formula weighted atomic number is located below each material.

Studying the misclassification events in Supplementary Table 10 provides further valuable insight. For example, the FeAl and NiAl samples are both B2, an ordered derivative of the BCC lattice, and the highest number of misclassifications for these two materials belong to space group 229 and 230. Fig. 3 shows the difficulty of distinguishing these two space groups, a common problem in the literature [35,36].



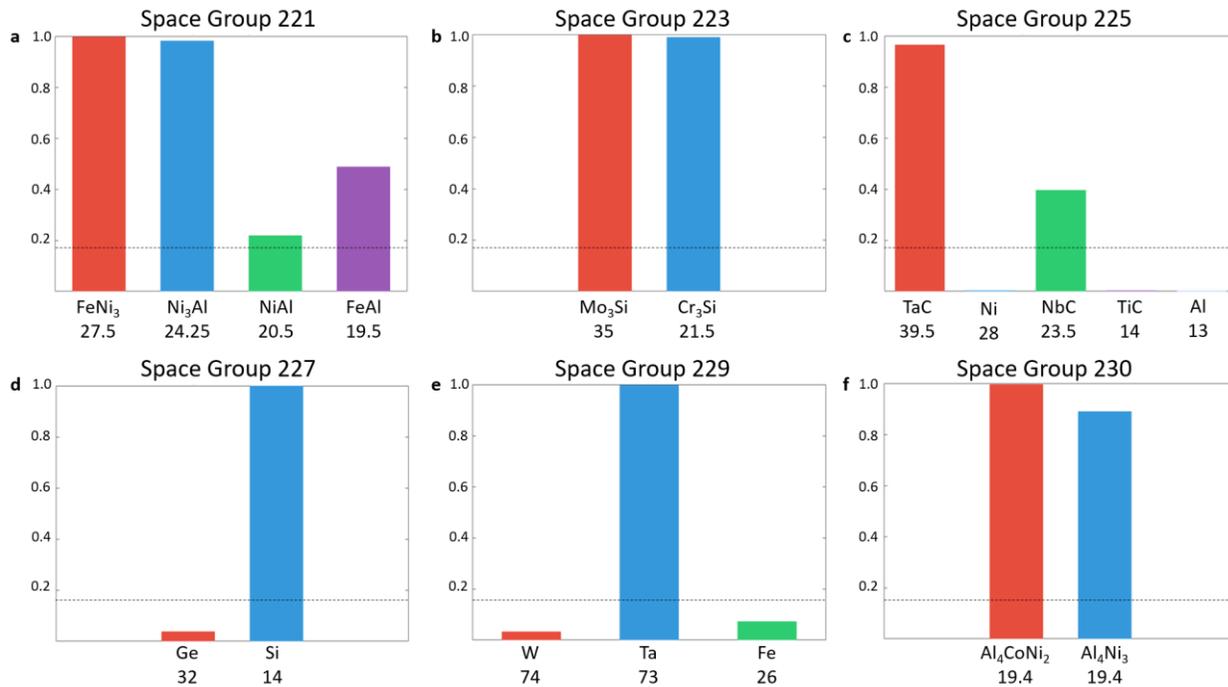
**Figure 6.3 Comparison of diffraction patterns from B2 FeAl (space group 221) and BCC Fe (space group 229).** B2 FeAl and BCC Fe can produce nearly identical diffraction patterns despite belonging to two different space groups.

Using the previous model as a baseline, the next three iterations study the effect of swapping the material used to represent a space group with the lowest average atomic number material in the dataset. First, TaC was replaced with Al in space group 225. The results of the exchange is shown in Fig. 4 and Supplementary Table 12 in the Appendix. The plots in Fig. 4 look largely unchanged except for space group 225 (Fig. 4c), for which materials with the lowest atomic scattering factors are now the most accurately classified. Furthermore, Supplementary Table 12 shows that the incorrect classifications for all other space groups are now primarily space group 225.



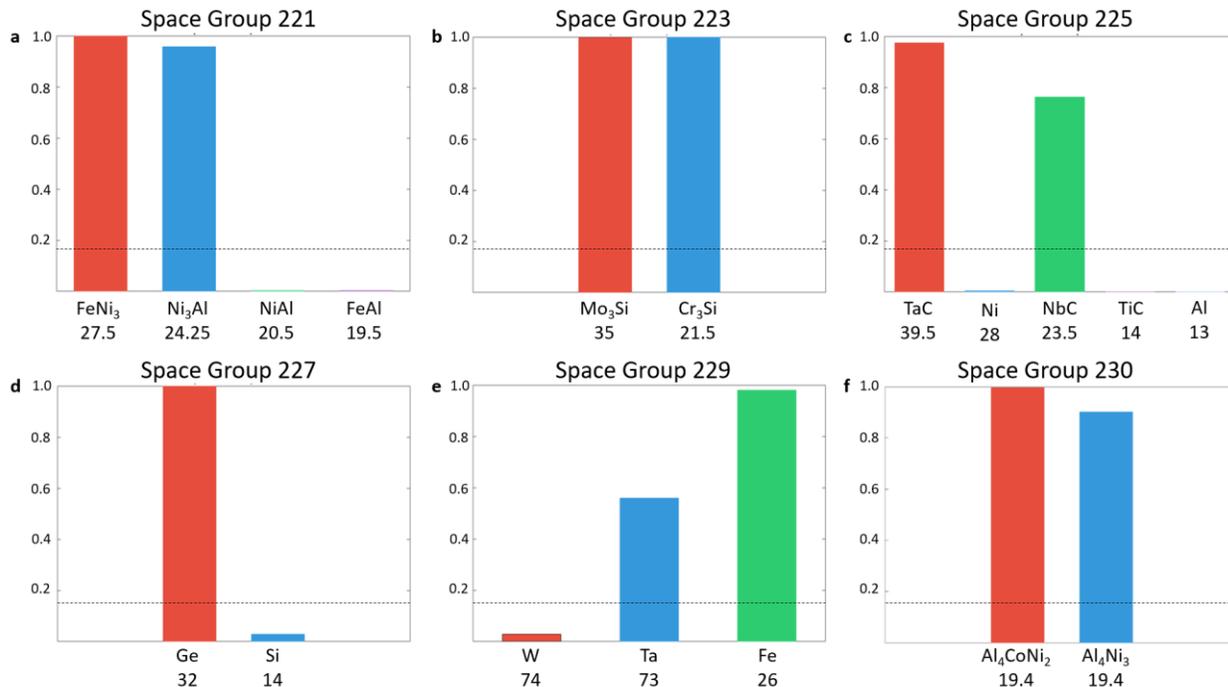
**Figure 6.4** Plots of normalized classification accuracy after fitting the model with a single low atomic number material. (a) Space group 221; trained on FeNi<sub>3</sub>. (b) Space group 223; trained on Mo<sub>3</sub>Si. (c) Space group 225; trained on Al. (d) Space group 227; trained on Ge. (e) Space group 229; trained on Ta. (f) Space group 230; trained on Al<sub>4</sub>CoNi<sub>2</sub>. The dashed line represents the chance of randomly guessing the correct space group. The formula weighted atomic number is located below each material.

In order to confirm the effects observed by exchanging Al for TaC in space group 225 previously, Si is exchanged with Ge in space group 227 (Fig. 5 and Supplementary Table 13). The training set for space group 225 is returned to TaC. The two space groups most affected by this change are 225 and 227, which both have FCC symmetry elements. For both space groups, as well as Fe, the use of Si as the training material causes diffraction patterns to be primarily misclassified as space group 221. The substitution of Si does result in the beneficial effect of increasing the classification accuracy of materials belonging to space group 221.



**Figure 6.5** Plots of normalized classification accuracy after fitting the model with a single low atomic number material. (a) Space group 221; trained on FeNi<sub>3</sub>. (b) Space group 223; trained on Mo<sub>3</sub>Si. (c) Space group 225; trained on TaC. (d) Space group 227; trained on Si. (e) Space group 229; trained on Ta. (f) Space group 230; trained on Al<sub>4</sub>CoNi<sub>2</sub>. The dashed line represents the chance of randomly guessing the correct space group. The formula weighted atomic number is located below each material.

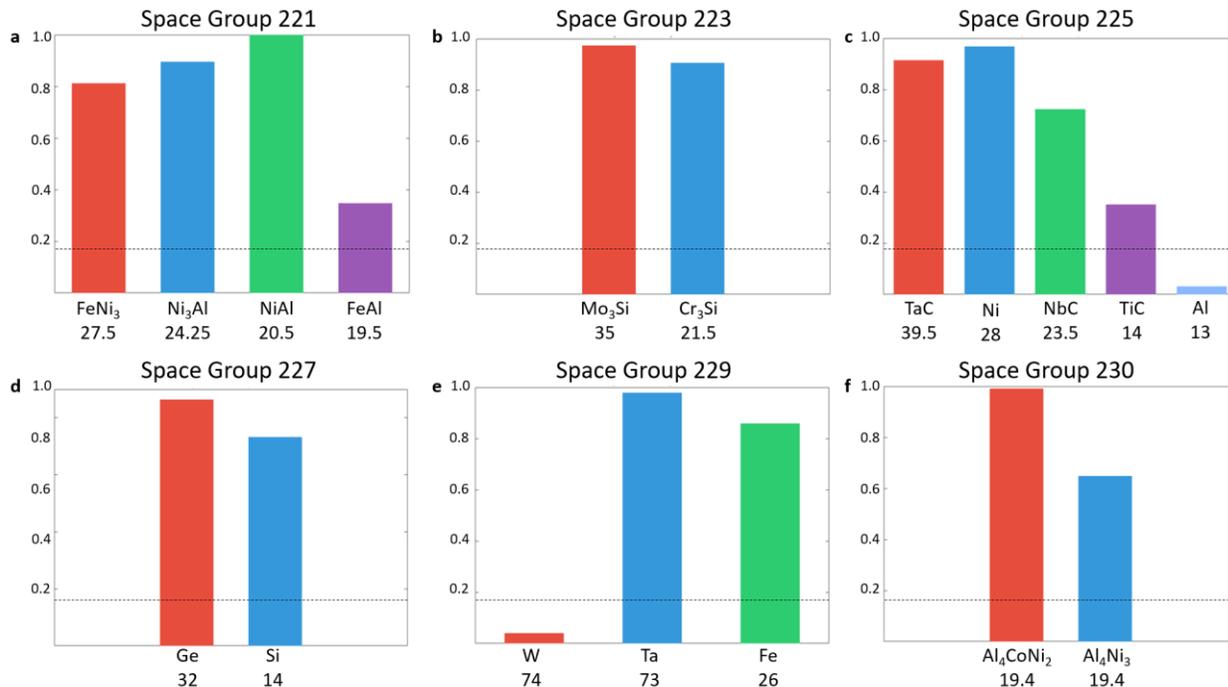
The last swap studied was the lower atomic number Fe in place of Ta (Fig. 6). Similar to what was observed when Al was used in space group 225, the increased correct classifications have shifted toward the low atomic number materials within space group 229. Furthermore, the B2 materials are now primarily misclassified as belonging to the BCC symmetry space group 229 (Supplementary Table 14). In comparison to Supplementary Table 10, almost half of the NiAl diffraction patterns were classified as 229 instead of 230. These three examples confirm that providing the neural network with only the lightest or heaviest materials will result in avoidable misclassification events.



**Figure 6.6** Plots of normalized classification accuracy after fitting the model with a single low atomic number material. (a) Space group 221; trained on FeNi<sub>3</sub>. (b) Space group 223; trained on Mo<sub>3</sub>Si. (c) Space group 225; trained on TaC. (d) Space group 227; trained on Ge. (e) Space group 229; trained on Fe. (f) Space group 230; trained on Al<sub>4</sub>CoNi<sub>2</sub>. The dashed line represents the chance of randomly guessing the correct space group. The formula weighted atomic number is located below each material.

Previously, the model has only been supplied with information from one material to learn from. It should not be surprising that the limited data representing the population of all materials in each class results in misclassification events when extrapolating further away from the training data. To demonstrate the increased accuracy of the technique when the data begins to better represent the population, we add a second material to the training set for space groups 221, 225, 227, and 229. Materials were not added to space group 223 and 230 since performance on those two groups was already well above the probability of the neural network guessing correctly by chance (16.7%) for the second material. The total number of diffraction patterns available to the neural network during training remained fixed as described in the Methods section. Supplementary Figure 34 and Supplementary Figure 35 display the IPFs and histograms, respectively, for the

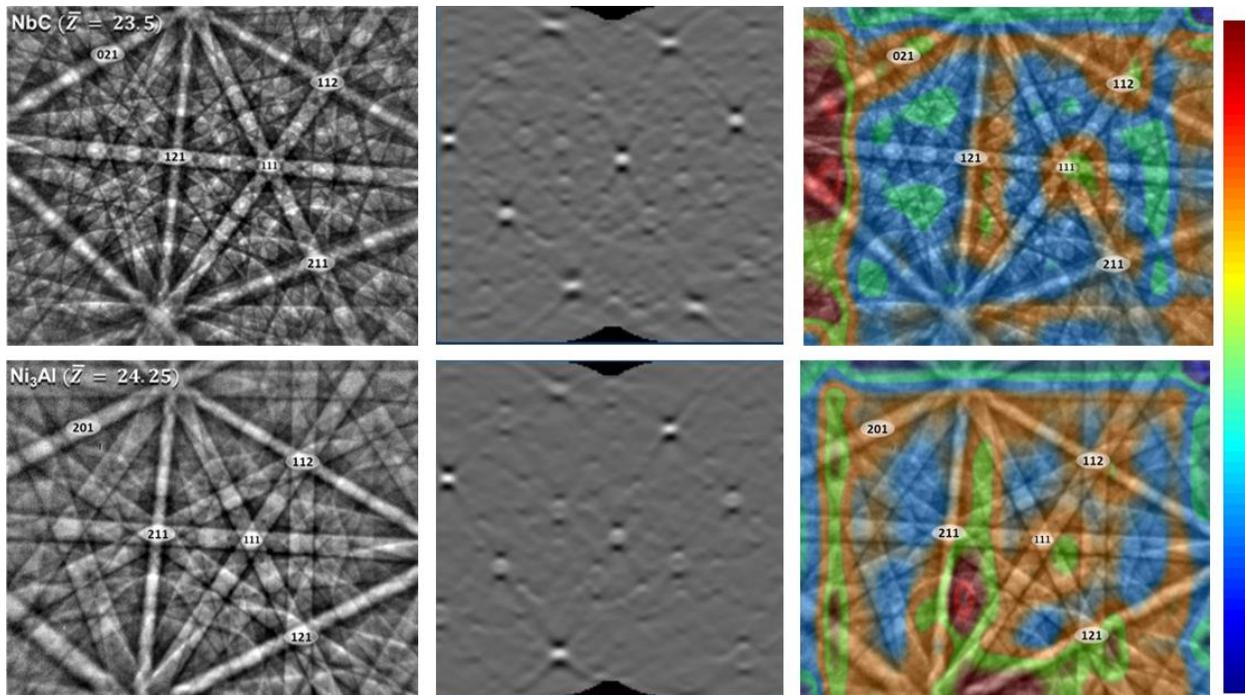
training data used in this model. As evidenced by the low M.U.D. and wide range of pattern quality, the data provided to the model during training is quite diverse. The classification results with new diffraction patterns are shown in Fig. 7 and Supplementary Table 15 in the Appendix. Compared to the neural networks trained with only one material, this model has 29% higher accuracy (now 80% correct) and demonstrates significantly improved accuracy on materials not utilized in the training set. For example, diffraction patterns from  $\text{Ni}_3\text{Al}$ ,  $\text{FeAl}$ ,  $\text{Cr}_3\text{Si}$ ,  $\text{NbC}$ ,  $\text{TiC}$ , and  $\text{Al}_4\text{CoNi}_2$  are correctly classified significantly more than random guessing even though the neural network was not provided diffraction patterns from these materials to learn from. Moreover, significantly increasing the range of atomic scattering factors, such as in Fig 7a and Fig. 7c, improves the neural network's performance on materials with scattering factors further outside the range. Aluminum and tungsten are the only two materials where the neural network's classification accuracy is below the probability of randomly guessing the correct answer. An analysis of the orientations and pattern quality for patterns that were correctly identified and misclassified was performed on several materials outside the training set.  $\text{NbC}$ ,  $\text{Al}$ , and  $\text{W}$  were selected in order to make a number of comparisons including materials with varying classification accuracy or within the same space group. Supplementary Figure 36 shows that for each material, the IPFs for the correctly classified patterns resembles the IPFs for patterns that were misclassified. In other words, specific orientations do not seem to be more likely to be classified correctly. Supplementary Figure 37 shows the corresponding MAD and BC histograms. While the plots and associated statistics for  $\text{NbC}$  show patterns with lower band contrast or higher MAD have a slight tendency to be misclassified, there does not appear to be a strong relationship between these pattern quality metrics and classification accuracy.



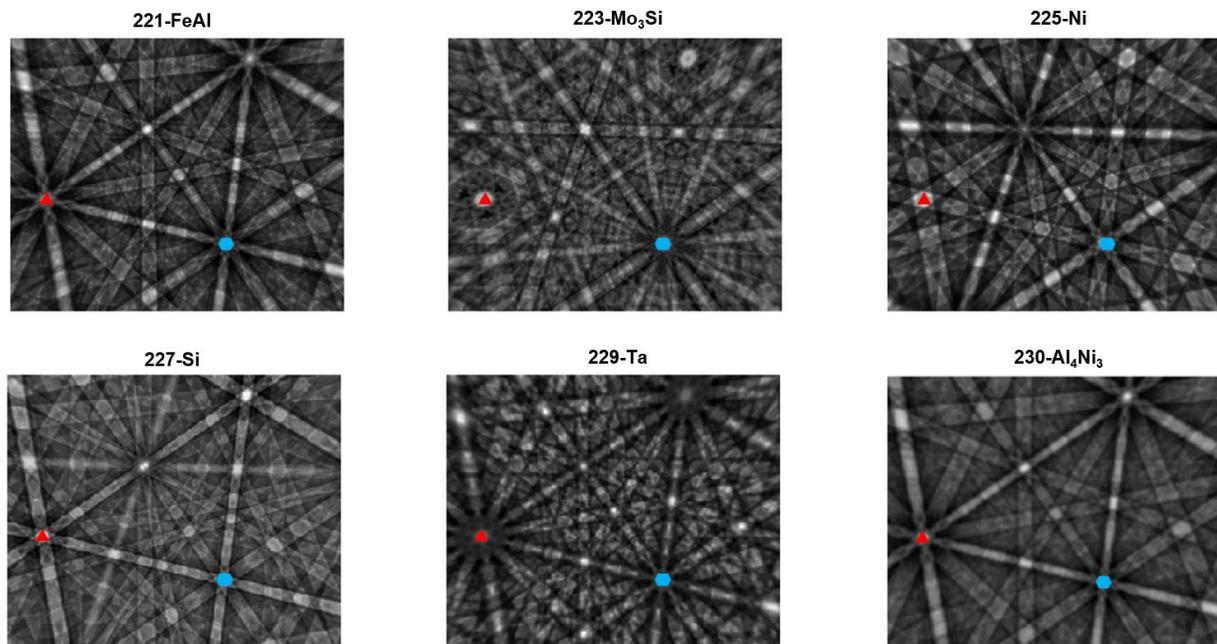
**Figure 6.7 Plots of normalized classification accuracy after fitting the model.** (a) Space group 221; trained on FeNi<sub>3</sub> and NiAl. (b) Space group 223; trained on Mo<sub>3</sub>Si. (c) Space group 225; trained on TaC and Ni. (d) Space group 227; trained on Ge and Si. (e) Space group 229; trained on Ta and Fe. (f) Space group 230; trained on Al<sub>4</sub>CoNi<sub>2</sub>. The dashed line represents the chance of randomly guessing the correct space group. The formula weighted atomic number is located below each material.

The excellent overall performance of the neural network in correctly identifying the space group necessitates an understandable interpretation of what information is important to the model, since orientation and pattern quality are not seemingly biasing the results of this study. Fig. 8 is a set of “visual explanations” for the decisions made by the Hough-based method and the trained CNN from Fig. 7. The selected diffraction patterns (Fig. 8 left-side) from NbC and Ni<sub>3</sub>Al are of nearly the same orientation and average atomic number but are from space group 225 and 221; respectively. Importantly, the model had not trained on any EBSPs from these two materials at this point. The Hough-based method produces the “butterfly peak” [10] representation (Fig. 8 middle) for matching the interplanar angles to user selected libraries. The resulting Hough-transforms are nearly identical for both NbC and Ni<sub>3</sub>Al, save for minor differences owing to the

small orientation shift. Furthermore, the Hough-based method creates a sparse representation that does not capture the finer detail found within the EBSPs. For example, the [111] and [112] zone axes are each encircled by a discernable HOLZ ring [17] as well as excess and deficiency features [32]. In comparison to the Hough method, the convolutional neural network can learn these features, or lack of them, and associate them with the correct space group. Using gradient-weighted class activation mapping (Grad-CAM) [57], we produce a localization heatmap highlighting the important regions for predicting the target class (Fig. 8 right-side). When visually inspecting the importance of local regions to the neural network, it is immediately observed that the neural network finds each of the labeled zone axes to be of high importance (orange in color) in both EBSPs. In fact, the heatmaps look remarkably similar and are concentrated about the same features a crystallographer would use. These features clearly include the zone axes and HOLZ rings. To investigate this further, a diffraction pattern is simulated for each material using EMSOFT [60]. In Fig. 9, it is first observed that many of the features of the patterns are similar, such as Kikuchi bands and diffraction maxima. These similarities explain why the Hough transform-based method cannot distinguish the two. However, it is also immediately noticeable that equivalent zone axes and surrounding regions have very different appearances. As examples, two sets of equivalent zone axes have been marked with either a red triangle or blue hexagon. The red triangle is the equivalent to the [111] zone axis studied in Fig. 8. While the fidelity of the simulated patterns with experimental patterns is not perfect, the existence of the discussed features can be confirmed by comparing them with the experimental patterns previously discussed. For example, compare the red and blue labeled zone axes in Fig. 9 for Ni and Mo<sub>3</sub>Si with the experimental NbC (Fig. 8) and Cr<sub>3</sub>Si (Supplementary Figure 30) diffraction patterns. The structure and features visible are clearly well correlated.



**Figure 6.8 Comparison of feature detection with Hough-based EBSD and the trained CNN.** (Top row from left to right) Experimental EBSP from NbC (space group 225; FCC structure), Hough-based feature detection, and gradient-weighted class activated map. (Bottom row from left to right) Experimental EBSP from Ni<sub>3</sub>Al (space group 221; L12 structure), Hough-based feature detection, and gradient-weighted class activated map. The importance scale for the heatmaps goes from dark blue (low) to dark red (high).

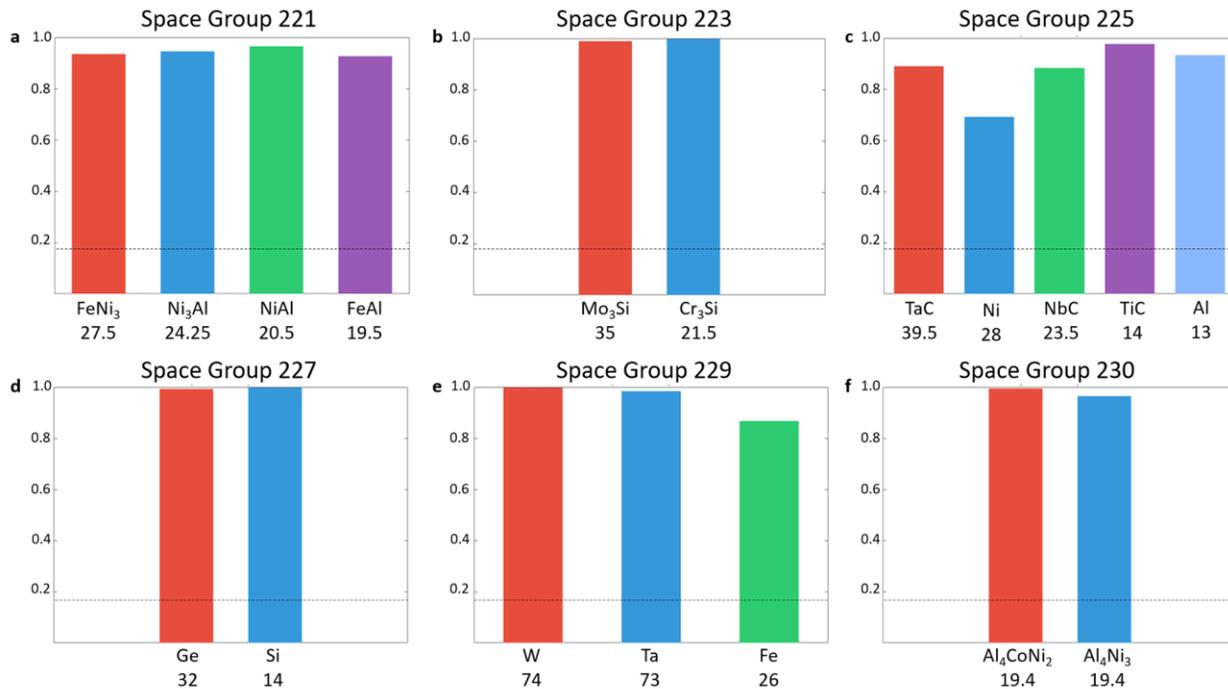


**Figure 6.9 Dynamically simulated EBSPs.** One diffraction pattern per space group was simulated to study the expected differences and assess the feature importance observed in experimental patterns. The observed reflectors are similar for each space group; however, attributes nearby the zone axes can vary significantly. Two sets of equivalent zone axes have been indicated with either a red triangle or blue hexagon.

Supplementary Figure 38 helps to elucidate the failure mechanism of this model by studying the activations of the current model, which incorrectly identified the pattern, compared to the activations in the first model, which had correctly identified the same EBSP. In each case, the heatmaps for space group 229 are overlaid onto the diffraction pattern. When correctly identified as belonging to space group 229 (Supplementary Figure 38, center), significantly more zone axes are given higher importance scores than when misclassified to space group 225 (Supplementary Figure 38, right). The same is observed with features within the HOLZ rings, such as near the [001] and [101] axis. It is important to note, the similarity between the activations at zone axes, such as [001], [012], and [102], should not be surprising since the activations for class 229 are being studied. Failures such as this can potentially be alleviated as the number of samples, experimental or simulated, used in fitting these models continues to grow. Previous

works using experimental [47] and simulated [4] patterns have further demonstrated the attentiveness of the last layers of the model on the zone axes and surrounding features. The study using simulated patterns also elucidated each convolutional layer's attention to specific aspects of EBSPs including edges and major Kikuchi bands. These studies into the “visual” perception of the model suggest that the network has learned relevant and intuitive features for identifying space groups.

By providing the neural network with diffraction patterns from many materials, the neural network can improve its resiliency to small changes within a space group, simultaneously developing a better understanding of what elements in the image are most useful and learning filters that better capture the information. Figure 10 demonstrates this by supplying the same number of diffraction patterns to learn from as used previously, but evenly divided between all available materials in each space group. Supplementary Table 16 shows the number of images classified to each space group from individual materials. The classification accuracy has increased to 93%, compared to 40-65% for the models that were only provided with one training material per space group (Figs. 2, 4, 5, 6, 7) and 80% when using two materials for some of the classes (Fig. 8).



**Figure 6.10** Plots of normalized classification accuracy after fitting the model with data from each material. (a) Space group 221. (b) Space group 223. (c) Space group 225. (d) Space group 227. (e) Space group 229. (f) Space group 230. The dashed line represents the chance of randomly guessing the correct space group. The formula weighted atomic number is located below each material.

## 6.4 Conclusion

In this paper, a high-throughput CNN-based approach to classifying electron backscatter diffraction patterns at the space group level is developed and demonstrated. In each study, the CNN is shown to be able to classify at least several materials outside the training set with much better probability than random guessing. Several investigations are conducted to explore the potential for biases owing to crystallographic orientation, pattern quality, or physical phenomena. The number of visible reflectors, directly correlated with atomic scattering factors, is found to have an impact classification accuracy when only low or high atomic number materials are used to fit the model. Increasing the range of atomic scattering factors used in training each class is found to reduce the number of misclassification events caused by large differences in the number of visible

reflectors. The dataset for each material and space group were of very low texture and good distributions of pattern quality metrics. Continued inclusion of data, particularly from more materials, will likely increase the robustness of the model when presented with new data. Investigation of the convolutional neural network's inner workings, through visualization of feature importance, strongly indicates the network is using the same features a crystallographer would use to manually identify the structure, particularly the information within the HOLZ rings. This analysis, combined with IPFs, suggests there is minimal orientation bias present. We believe this method can be expanded to the remaining space groups and implemented as part of a multi-tiered model for determining the complete crystal structure. There are no algorithmic challenges to extending this framework to all 230 space groups, it is only currently limited by the lack of data, which simulated EBSD patterns may help to resolve. This technique should benefit from continued advancements in detectors, such as direct electron detectors, and the framework is expected to be immediately applicable to similar techniques such as electron channeling patterns and CBED. A wide range of other research areas including pharmacology, structural biology, and geology are expected to benefit by using similar automated algorithms to reduce the amount of time required for structural identification.

## **6.5 Acknowledgements**

The authors would like to thank Wenyong Jiang, William Mellor, and Xiao Liu for their assistance. K. Kaufmann was supported by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program. K. Kaufmann would also like to acknowledge the support of the ARCS Foundation, San Diego Chapter. KV would like to acknowledge the financial generosity of the Oerlikon Group in support of his research group.

Chapter 6, in full, is a reprint of the material “Deep Neural Network Enabled Space Group Identification in EBSD”, as it appears in *Microscopy and Microanalysis*. K. Kaufmann, C. Zhu, A. S. Rosengarten, A. S. Rosengarten, and K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

## 6.6 References

- [1] T.B. Britton, C. Maurice, R. Fortunier, J.H. Driver, A.P. Day, G. Meaden, D.J. Dingley, K. Mingard, A.J. Wilkinson, Factors affecting the accuracy of high resolution electron backscatter diffraction when using simulated patterns, *Ultramicroscopy*. 110 (2010) 1443–1453. doi:10.1016/j.ultramic.2010.08.001.
- [2] R. Hielscher, F. Bartel, T.B. Britton, Gazing at crystal balls: Electron backscatter diffraction pattern analysis and cross correlation on the sphere, *Ultramicroscopy*. 207 (2019). doi:10.1016/j.ultramic.2019.112836.
- [3] A. Foden, D.M. Collins, A.J. Wilkinson, T.B. Britton, Indexing electron backscatter diffraction patterns with a refined template matching approach, *Ultramicroscopy*. 207 (2019) 112845. doi:10.1016/j.ultramic.2019.112845.
- [4] A. Foden, A. Previero, T.B. Britton, Advances in electron backscatter diffraction, (2019). <http://arxiv.org/abs/1908.04860> (accessed December 9, 2019).
- [5] Y.H. Chen, S.U. Park, D. Wei, G. Newstadt, M.A. Jackson, J.P. Simmons, M. De Graef, A.O. Hero, A Dictionary Approach to Electron Backscatter Diffraction Indexing, *Microsc. Microanal.* (2015). doi:10.1017/S1431927615000756.
- [6] S. Singh, Y. Guo, B. Winiarski, T.L. Burnett, P.J. Withers, M. De Graef, High resolution low kV EBSD of heavily deformed and nanocrystalline Aluminium by dictionary-based indexing, *Sci. Rep.* (2018). doi:10.1038/s41598-018-29315-8.
- [7] S. Singh, M. De Graef, Dictionary Indexing of Electron Channeling Patterns, *Microsc. Microanal.* 23 (2017) 1–10. doi:10.1017/S1431927616012769.
- [8] G. Nolze, R. Hielscher, A. Winkelmann, Electron backscatter diffraction beyond the mainstream, *Cryst. Res. Technol.* 52 (2017) 1600252. doi:10.1002/crat.201600252.
- [9] V.S. Tong, A.J. Knowles, D. Dye, T. Ben Britton, Rapid electron backscatter diffraction mapping: Painting by numbers, *Mater. Charact.* 147 (2019) 271–279. doi:10.1016/J.MATCHAR.2018.11.014.
- [10] N.C.K. Lassen, Automated Determination of Crystal Orientations from Electron Backscattering Patterns, The Technical University of Denmark, 1994.
- [11] T. Karthikeyan, M.K. Dash, S. Saroja, M. Vijayalakshmi, Evaluation of misindexing of EBSD patterns in a ferritic steel, *J. Microsc.* 249 (2013) 26–35. doi:10.1111/j.1365-2818.2012.03676.x.
- [12] C.L. Chen, R.C. Thomson, The combined use of EBSD and EDX analyses for the identification of complex intermetallic phases in multicomponent Al-Si piston alloys, *J. Alloys Compd.* 490 (2010) 293–300. doi:10.1016/j.jallcom.2009.09.181.
- [13] S. McLaren, S.M. Reddy, Automated mapping of K-feldspar by electron backscatter diffraction and application to <sup>40</sup>Ar/<sup>39</sup>Ar dating, *J. Struct. Geol.* 30 (2008) 1229–1241. doi:10.1016/J.JSG.2008.05.008.

- [14] M.M. Nowell, S.I. Wright, Phase differentiation via combined EBSD and XEDS, in: *J. Microsc.*, 2004: pp. 296–305. doi:10.1111/j.0022-2720.2004.01299.x.
- [15] R.P. Goehner, J.R. Michael, Phase Identification in a Scanning Electron Microscope Using Backscattered Electron Kikuchi Patterns., *J. Res. Natl. Inst. Stand. Technol.* 101 (1996) 301–308. doi:10.6028/jres.101.031.
- [16] D.J. Dingley, S.I. Wright, Phase Identification Through Symmetry Determination in EBSD Patterns, in: A. Schwartz, M. Kumar, B. Adams, D. Field (Eds.), *Electron Backscatter Diffraction Mater. Sci.*, Springer, Boston, MA, 2009: pp. 97–107. doi:10.1007/978-0-387-88136-2.
- [17] J.R. Michael, J.A. Eades, Use of reciprocal lattice layer spacing in electron backscatter diffraction pattern analysis, *Ultramicroscopy*. 81 (2000) 67–81.
- [18] L. Li, M. Han, Determining the Bravais lattice using a single electron backscatter diffraction pattern, *J. Appl. Crystallogr.* (2015). doi:10.1107/S1600576714025989.
- [19] K.Z. Baba-Kishi, D.J. Dingley, Backscatter Kikuchi Diffraction in the SEM for Identification of Crystallographic Point Groups, *Scanning*. 11 (1989) 305–312. doi:10.1002/sca.4950110605.
- [20] B. Ollivier, R. Retoux, P. Lacorre, D. Massiot, G. Férey, Crystal structure of  $\kappa$ -alumina: an X-ray powder diffraction, TEM and NMR study, *J. Mater. Chem.* 7 (1997) 1049–1056. doi:10.1039/a700054e.
- [21] J.E. Post, D.R. Veblen, Crystal structure determinations of synthetic sodium, magnesium, and potassium birnessite using TEM and the Rietveld method, *Am. Mineral.* 75 (1990) 477–489. [http://rruff.info/uploads/AM75\\_477.pdf](http://rruff.info/uploads/AM75_477.pdf) (accessed October 7, 2018).
- [22] E. Garnier, *Powder Diffraction. Theory and Practice*. Edited by R. E. Dinnebier and S. J. L. Billinge. Cambridge: RSC Publishing, 2008. Pp. xxi + 582. Price (hardcover): GBP 59.00. ISBN (online): 978-1-84755-823-7; ISBN (print): 978-0-85404-231-9. , *Acta Crystallogr. Sect. A Found. Crystallogr.* 65 (2009) 51–51. doi:10.1107/s010876730802850x.
- [23] K.S. Vecchio, D.B. Williams, Convergent beam electron diffraction analysis of the T 1 (Al<sub>2</sub>CuLi) phase in Al-Li-Cu alloys, *Metall. Trans. A.* 19 (1988) 2885–2891. doi:10.1007/BF02647714.
- [24] K.S. Vecchio, D.B. Williams, Convergent beam electron diffraction study of Al<sub>3</sub>Zr in Al-Zr AND Al-Li-Zr alloys, *Acta Metall.* 35 (1987) 2959–2970. doi:10.1016/0001-6160(87)90295-1.
- [25] D.B. Williams, A.R. Pelton, R. Gronsky, eds., *Images of Materials*, Oxford University Press, New York, 1991.
- [26] H. Wang, T. Harrington, C. Zhu, K.S. Vecchio, Design, fabrication and characterization of FeAl-based metallic-intermetallic laminate (MIL) composites, *Acta Mater.* 175 (2019) 445–456. doi:10.1016/j.actamat.2019.06.039.
- [27] C. Zhu, K. Kaufmann, K.S. Vecchio, Novel remapping approach for HR-EBSD based on

- demons registration, *Ultramicroscopy*. 208 (2020) 112851. doi:10.1016/j.ultramic.2019.112851.
- [28] L.J. Hufford, E.J. Chin, S.E. Perry, E.A. Miranda, J.M. Hanchar, Xenolith Constraints on Rheology of Heterogenous Deep Crust Beneathe the Eastern Mojave Desert, California, in: *GSA Annu. Meet. Phoenix, Arizona, USA-2019*, 2019.
- [29] R. Bernard, J. Day, E. Chin, Strong Olivine Lattice Preferred Orientation in Brachinite-Like Achondrites, in: *Lunar Planet. Sci. Conf.*, 2019.
- [30] M. Calcagnotto, D. Ponge, E. Demir, D. Raabe, Orientation gradients and geometrically necessary dislocations in ultrafine grained dual-phase steels studied by 2D and 3D EBSD, *Mater. Sci. Eng. A*. 527 (2010) 2738–2746. doi:10.1016/j.msea.2010.01.004.
- [31] D. Chen, J.C. Kuo, W.T. Wu, Effect of microscopic parameters on EBSD spatial resolution, *Ultramicroscopy*. (2011). doi:10.1016/j.ultramic.2011.06.007.
- [32] A. Winkelmann, Dynamical effects of anisotropic inelastic scattering in electron backscatter diffraction, *Ultramicroscopy*. 108 (2008) 1546–1550. doi:10.1016/J.ULTRAMIC.2008.05.002.
- [33] A. Winkelmann, G. Nolze, Chirality determination of quartz crystals using Electron Backscatter Diffraction, *Ultramicroscopy*. 149 (2015) 58–63. doi:10.1016/J.ULTRAMIC.2014.11.013.
- [34] T. Wang, M. Komarasamy, S. Shukla, R.S. Mishra, Simultaneous enhancement of strength and ductility in an AlCoCrFeNi<sub>2.1</sub> eutectic high-entropy alloy via friction stir processing, *J. Alloys Compd.* 766 (2018) 312–317. doi:10.1016/j.jallcom.2018.06.337.
- [35] M. Li, J. Gazquez, A. Borisevich, R. Mishra, K.M. Flores, Evaluation of microstructure and mechanical property variations in Al<sub>x</sub>CoCrFeNi high entropy alloys produced by a high-throughput laser deposition method, *Intermetallics*. 95 (2018) 110–118. doi:10.1016/j.intermet.2018.01.021.
- [36] M.C. Gao, J.-W. Yeh, P.K. Liaw, Y. Zhang, eds., *High-Entropy Alloys*, Springer International Publishing, Cham, 2016. doi:10.1007/978-3-319-27013-5.
- [37] J.M. Cowley, *Diffraction Physics*, 2nd ed., North-Holland, Amsterdam, 1990.
- [38] A. Ziletti, D. Kumar, M. Scheffler, L.M. Ghiringhelli, Insightful classification of crystal structures using deep learning, *Nat. Commun.* 9 (2018) 2775. doi:10.1038/s41467-018-05169-6.
- [39] F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N.T.P. Hartono, S. Ramasamy, B.L. DeCost, S.I.P. Tian, G. Romano, A. Gilad Kusne, T. Buonassisi, Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks, *Npj Comput. Mater.* 5 (2019). doi:10.1038/s41524-019-0196-x.
- [40] H. Lombaert, L. Grady, X. Pennec, N. Ayache, F. Cheriet, Spectral log-demons: Diffeomorphic image registration with very large deformations, *Int. J. Comput. Vis.* 107

- (2014) 254–271. doi:10.1007/s11263-013-0681-5.
- [41] C. Zhu, H. Wang, K. Kaufmann, K.S. Vecchio, A computer vision approach to study surface deformation of materials, *Meas. Sci. Technol.* 31 (2020) 055602. doi:10.1088/1361-6501/ab65d9.
- [42] H. Wang, L. Dong, J. O’Daniel, R. Mohan, A.S. Garden, K. Kian Ang, D.A. Kuban, M. Bonnen, J.Y. Chang, R. Cheung, Validation of an accelerated “demons” algorithm for deformable image registration in radiation therapy, *Phys. Med. Biol.* 50 (2005) 2887–2905. doi:10.1088/0031-9155/50/12/011.
- [43] B.L. DeCost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, *Comput. Mater. Sci.* 110 (2015) 126–133. doi:10.1016/J.COMMATSCI.2015.08.011.
- [44] E. Alegre, J. Barreiro, H. Cáceres, L.K. Hernández, R.A. Fernández, M. Castejón, Design of a Computer Vision System to Estimate Tool Wearing, *Mater. Sci. Forum.* 526 (2006) 61–66. doi:10.4028/www.scientific.net/MSF.526.61.
- [45] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, *Nature.* 323 (1986) 533–536. doi:10.1038/323533a0.
- [46] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature.* 521 (2015) 436–444. doi:10.1038/nature14539.
- [47] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, T.J. Harrington, E. Marin, K.S. Vecchio, Crystal symmetry determination in electron diffraction using machine learning, *Science (80-. ).* 367 (2020) 564–568. doi:10.1126/science.aay3062.
- [48] J. Goulden, P. Trimby, A. Bewick, The Benefits and Applications of a CMOS-based EBSD Detector, *Microsc. Microanal.* 24 (2018) 1128–1129.
- [49] H.P. Hanson, F. Herman, J.E. Lea, S. Skillman, HFS atomic scattering factors, *Acta Crystallogr.* 17 (1964) 1040–1044. doi:10.1107/s0365110x64002638.
- [50] S.I. Wright, M.M. Nowell, EBSD image quality mapping, *Microsc. Microanal.* 12 (2006) 72–84. doi:10.1017/S1431927606060090.
- [51] F. Bachmann, R. Hielscher, H. Schaeben, Texture analysis with MTEX- Free and open source software toolbox, in: *Solid State Phenom.*, 2010: pp. 63–68. doi:10.4028/www.scientific.net/SSP.160.63.
- [52] F. Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, in: *Proc. IEEE Int. Conf. Comput. Vis.*, The Computer Vision Foundation, Seattle, 2017: pp. 1251–1258. [http://openaccess.thecvf.com/content\\_cvpr\\_2017/papers/Chollet\\_Xception\\_Deep\\_Learning\\_CVPR\\_2017\\_paper.pdf](http://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf) (accessed February 1, 2019).
- [53] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, in: *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 2015: pp. 1–15. <http://arxiv.org/abs/1412.6980>

(accessed October 8, 2018).

- [54] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, G. Brain, TensorFlow: A System for Large-Scale Machine Learning, in: Proc. 12th USENIX Symp. Oper. Syst. Des. Implement., USENIX, Savannah, Georgia, 2016: pp. 265–283. <https://tensorflow.org>. (accessed October 8, 2018).
- [55] F. Chollet, Keras, (2015). <https://github.com/keras-team/keras>.
- [56] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830. <http://www.jmlr.org/papers/v12/pedregosa11a> (accessed August 15, 2019).
- [57] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, in: Proc. IEEE Int. Conf. Comput. Vis., 2017: pp. 618–626. doi:10.1109/ICCV.2017.74.
- [58] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (2013) 1798–1828. doi:10.1109/TPAMI.2013.50.
- [59] A. Mahendran, A. Vedaldi, Visualizing Deep Convolutional Neural Networks Using Natural Pre-images, *Int. J. Comput. Vis.* 120 (2016) 233–255. doi:10.1007/s11263-016-0911-8.
- [60] P.G. Callahan, M. De Graef, Dynamical electron backscatter diffraction patterns. Part I: Pattern simulations, *Microsc. Microanal.* 19 (2013) 1255–1265. doi:10.1017/S1431927613001840.

## Chapter 7 Phase Mapping in EBSD using Convolutional Neural Networks

Kevin Kaufmann<sup>1</sup>, Chaoyi Zhu<sup>2</sup>, Alexander S. Rosengarten<sup>1</sup>, Daniel Maryanovsky<sup>3</sup>, Haoren Wang<sup>1</sup>, and Kenneth S. Vecchio<sup>1,2</sup>

<sup>1</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA

<sup>2</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA

<sup>3</sup>Department of Cognitive Science, UC San Diego, La Jolla, CA 92093, USA

### Abstract

The emergence of commercial electron backscatter diffraction (EBSD) equipment ushered in an era of information rich maps produced by determining the orientation of user selected crystal structures. Since then, a technological revolution has occurred in the quality of and the rate these diffractions patterns can be collected and analyzed for orientation using the Hough transform. The next revolution in EBSD is the ability to directly utilize the information rich diffraction patterns in a high-throughput manner. Aided by machine learning techniques, this new methodology is, as demonstrated herein, capable of accurately separating phases in a material by crystal symmetry, chemistry, and even lattice parameters with fewer human decisions. This work is the first demonstration of such capabilities and addresses many of the major challenges faced in modern EBSD. Diffraction patterns are collected from a variety of samples and a convolutional neural network, a type of machine learning algorithm, is trained to autonomously recognize the subtle differences in the diffraction patterns and output phase maps of the material. This study offers a

path to machine learning coupled phase mapping as databases of EBSD patterns encompass an increasing amount of the space groups, chemistry changes, and lattice parameter variations.

## **7.1 Introduction**

Conventional electron backscatter diffraction (EBSD) is a scanning electron microscope (SEM)-based technique used to determine the three-dimensional orientation of individual grains in crystalline materials. At present, the collected electron backscatter diffraction patterns (EBSPs) are primarily utilized to construct representative maps of the microstructure, given user defined phases. The resulting maps and EBSPs can be further utilized to study dislocation evolution [1], geometrically necessary dislocations [2], among others [3–6]. These analyses are currently achieved in commercial EBSD utilizing the Hough transform on each diffraction pattern combined with a look-up table of interplanar angles constructed from a set of selected reflectors as specified by the user (typically up to 5 phases). In the process, the Hough transform strips most of the information about crystal structure and chemistry from the image [7–10]. The small subset of information being utilized for differentiating the phases often results in mis-classification of similar crystal structures [11] and orientations that produce similar diffraction maxima in different crystal systems [12,13]. Despite its shortcomings, the Hough transform has continued to be the standard in pattern indexing even though computing power has increased exponentially, largely because computer algorithms have previously been incapable of Kikuchi band detection, or the more challenging problem of autonomously parsing relevant information for determining chemistry and crystal structure.

The inability for state-of-the-art Hough-based EBSD systems to make determinations about individual components of or the entire crystal structure of a phase has not gone unnoticed in the scientific community. One proposed solution is dictionary-based approaches [14–17].

Libraries of selected materials can be selected and simulated with software such as EMSoft [18] for the purpose of comparing the experimentally collected diffraction patterns with those simulated. Through this process, the most similar patterns in the dictionary determines the phase and orientation of each EBSD pattern, even for deformed or fine-grained materials [14]. These approaches have helped to allay major challenges in phase differentiation, such as separating martensite from cementite [14]. Other types of phase identification methodologies have been demonstrated wherein EBSD is combined with other analytical techniques, such as energy dispersive X-ray spectroscopy (EDS) or wavelength dispersive X-ray spectroscopy (WDS), given that the chemical and structural information of the phase exists in a theoretical model or crystal database [10,19,20]. The problem remains that the collected electron backscatter diffraction images are traditionally processed via the Hough transform when attempting to perform phase identification. As mentioned previously, the down sampling of this data allows for patterns from multiple crystal structures to be mistaken for others since the computer algorithms are capable of finding orientations, which may appear to be correct [12]. In commercial phase identification systems, the user must also select a limited number of elements, typically up to 3, from the EDS results before the computer narrows down the list of possible candidates. The combination of these two reductions in information results in a multitude of phases from a variety of crystal structures being returned for the user to decide is the best fit for each phase present in the microstructure.

The recent advent of machine learning techniques, such as the convolutional neural network (CNN), offer an opportunity to address many of the challenges to autonomously extracting information from diffraction data [21,22]. The reason CNNs are of particular interest is due to multiple advantages over classical computer vision techniques, which require a multitude

of heuristics [23–27] — such as detecting Kikuchi bands, accounting for orientation changes, determining band width, searching for Higher Order Laue Zones (HOLZ) rings, etc., and carrying the burden of developing the logic that defines these abstract qualities. Instead, this deep learning technique determines its own internal representation of the data, via backpropagation [28] potentially coupled with one of many possible hyperparameter tuning strategies [29], such that it maximizes performance at the discrimination task. This is the underlying principle behind deep representation learning (i.e. deep neural networks) [30]. Such methods allow the model to find patterns that may be unintuitive or too nuanced for humans to discern. Other discovered features might be obvious to experts, but difficult to translate into specific logic (e.g. linking the background signal to the chemical composition). CNNs are further advantageous over other machine learning models since they operate on the image data directly. As an example of other machine learning models applied to EBSD, a nearest neighbor machine learning model has been previously explored by Goulden *et al.* [31] to address the challenge of separating ferrite and martensite by machine learning aided pattern quality analysis, but pattern quality is too rigid a metric for general use, and the method was reliant on human analysis and confirmation over multiple rounds of indexing a single map. Another advantage is the flexibility of these CNNs, allowing for the transfer of knowledge learned from discriminating images in other contexts [32,33], the development of models suitable for application in a highly specific materials space, or deployment in an application where the phases present are completely unknown. For example, during the initial analysis of a material, a pre-trained CNN could be utilized for the determination of which Bravais lattices or space groups are present [34]. Other studies utilizing CNNs in EBSD have demonstrated orientation determination [35] and phase classification of several materials that may be easily confused in traditional EBSD [36]. The work by Foden *et al.* demonstrates the

application of simulated patterns for achieving this goal. Simulated EBSPs may alleviate the challenges associated with finding suitable materials for data collection, similar to what is being done in other applications [21,22]. In well-known classes of materials, application specific models could be developed to distinguish phases traditional Hough-based EBSD finds nearly indistinguishable, such as martensite and ferrite. Moreover, modern computing hardware allows for real time classification in line with recent advancements in EBSD technology [37].

Herein, it is demonstrated that convolutional neural networks can be constructed to recognize phases based on their structure/symmetry, chemistry, and even lattice parameter variations; some of these well beyond the scope of what the Hough transform is capable of, even with user supplied information about the sample. Several of these CNN-based demonstrations are meant to serve as proof of concept models for application specific use cases, while others such as the Bravais lattice identification model can readily be applied to materials outside the training set. Further automation and improved accuracy of the proposed process could be achieved by the development of diffraction pattern databases, inclusion of simulated diffraction patterns into the training process, and the development of standard models for various use cases.

## **7.2 Materials and methods**

### **7.2.1 Electron backscatter diffraction pattern collection**

Six different multi-phase materials were selected for demonstrating the proposed phase-mapping methodology: (1) a rutiled quartz sample, (2) a sample of the Campo del Cielo meteorite, (3) an arc-melted ingot of  $\text{Ni}_{80.8}\text{B}_{13.6}\text{Si}_{5.4}\text{Fe}_{0.2}$  (at%) blended with 40 wt% eutectic tungsten carbide (a metal matrix composite) [38], (4) an Fe-Al metallic-intermetallic laminate (MIL) composite [39], (5) a thermally-cycled MCrAlY-based thermal barrier coating [40–42], and (6) a sample of 430 stainless steel. Diffraction patterns from twenty-eight other materials, detailed

in [34], were utilized for training the CNN-based model for Bravais lattices demonstrated herein. EBSD patterns (EBSPs) were collected in a Thermo-Fisher (formerly FEI) Apreo scanning electron microscope (SEM) equipped with an Oxford Symmetry EBSD detector and Oxford X-Max<sup>N</sup> EDS detector. The Oxford Symmetry EBSD detector was utilized in high resolution (1244x1024) mode with frame averaging and EDS maps were collected simultaneously. The Hough indexing parameters were 12 Kikuchi bands, a Hough resolution of 250, and band center indexing. After collecting high resolution EBSPs from each material, all patterns collected were exported as tiff images. Supplementary Figure 39 in the Appendix shows example images of the high-resolution diffraction patterns collected from the six materials utilized in this study. The collected data was not filtered for pattern quality via any means, each pattern was individually assessed by the neural network, and the collection of images for each sample may contain partial or low-quality diffraction patterns, which will decrease the accuracy of their identification.

### **7.2.2 Neural network architecture and phase mapping procedure**

The well-studied convolutional neural network architecture Xception [43] was selected for fitting a model that determines which phase or crystal structure a diffraction pattern originated from. Supplementary Figure 40 in the appendix details a schematic of the convolutional neural network operating on an EBSP. Training was performed using Adam optimization [44], with batches of 32 images, and a minimum delta of 0.001 as the validation loss employed for early stopping criteria. Categorical cross-entropy was used as the loss function. All training data labels were created using validated phases in combination with combined EBSD/EDS. In the cases where training data was collected from multi-phase materials, such as the MIL composite, it is possible that a small number of diffraction patterns could have been attributed to the incorrect class. If a substantial number of patterns were mislabeled, we expect to see poor performance

from the model. The CNNs were implemented using Python 3.5 with TensorFlow [45] and Keras [46]. The conversion from neural network predictions to phase maps is accomplished using MATLAB R2018B. The histogram plot and pattern quality phase map were also made using MATLAB R2018B.

The rutilated quartz specimen was mapped using a convolutional neural network previously trained to differentiate the crystal symmetry of the 14 Bravais lattices [34]. This model was trained using the same hyperparameters used for all other models in this work. The model was not trained using quartz diffraction patterns.

For the Campo del Cielo meteorite sample, EBSD patterns were collected from taenite, schreibersite, iron, and regions without any evidence of a diffraction pattern (i.e. containing only background noise). The model was trained on five hundred randomly selected patterns from each of these phases.

Diffraction patterns for three of the five phases present in the  $\text{Ni}_{80.8}\text{B}_{13.6}\text{Si}_{5.4}\text{Fe}_{0.2}$  plus 40 wt% eutectic tungsten carbide sample were collected from separate samples of the pure phases (Ni,  $\text{W}_2\text{C}$ , and WC), similar to a ‘standard’ library. Diffraction patterns from the other two phases,  $\text{Ni}_3\text{B}$  and  $\text{W}_2\text{NiB}_2$ , were collected directly from the sample owing to the challenge of making pure specimens of these two phases. The model was trained to differentiate five hundred randomly selected patterns from each of these phases.

The Fe-Al MIL composite is presented herein as four distinct phases. This material was fabricated via diffusion controlled growth from a “multiple-thin foil” configuration and a two-stage reaction process described elsewhere [39]. This fabrication strategy yields layers of pure iron, an Al-enriched  $\alpha$ -Fe layer, an Fe-enriched FeAl B2 layer, and a near equiatomic (~48 at.%

Fe) FeAl B2 layer. Multiple layers of the sample were mapped using Hough-based EBSD alone, with reference chemistries from EDS for each phase, and separately with the convolutional neural network (CNN-based) EBSD approach. Diffraction patterns for training the CNN were collected from each of these regions and separated into four groups based on the chemistry at that location. The model was trained to recognize the subtle differences caused by chemistry and structure changes using five hundred randomly selected patterns from each of these phases.

The thermal barrier coating sample contains five face-centered-cubic (FCC) phases, all space group 225, and one rhombohedral structure. The FCC phases are Ni, yttria-stabilized zirconia (YS-ZrO<sub>2</sub>), TaC, Cr<sub>23</sub>C<sub>6</sub>, and Hastelloy X. The rhombohedral phase is the thermally grown oxide (TGO) Al<sub>2</sub>O<sub>3</sub>. The sample was phase mapped using standalone Hough-based EBSD, EBSD combined with chemistry information from EDS assigned to each phase, and separately with the convolutional neural network (CNN-based) EBSD approach. Five hundred diffraction patterns were randomly selected from each phase to train the model.

A section of 430 stainless steel (max 0.12 wt% carbon) was selected to study the methodology's ability to separate martensite from ferrite. The lattice parameters for phase mapping via traditional EBSD were determined from XRD and the well-defined dependence of the lattice parameter  $c$  on carbon content [47]. Diffraction patterns used to train the model were collected from out-of-sample 'standards' of pure Fe and rapidly quenched MMFX steel [48]. The training patterns from the MMFX steel (max 0.15 wt% carbon) were filtered such that they were of the same high quality as the pure Fe patterns. Five hundred high quality diffraction patterns from each phase were then selected to train the model.

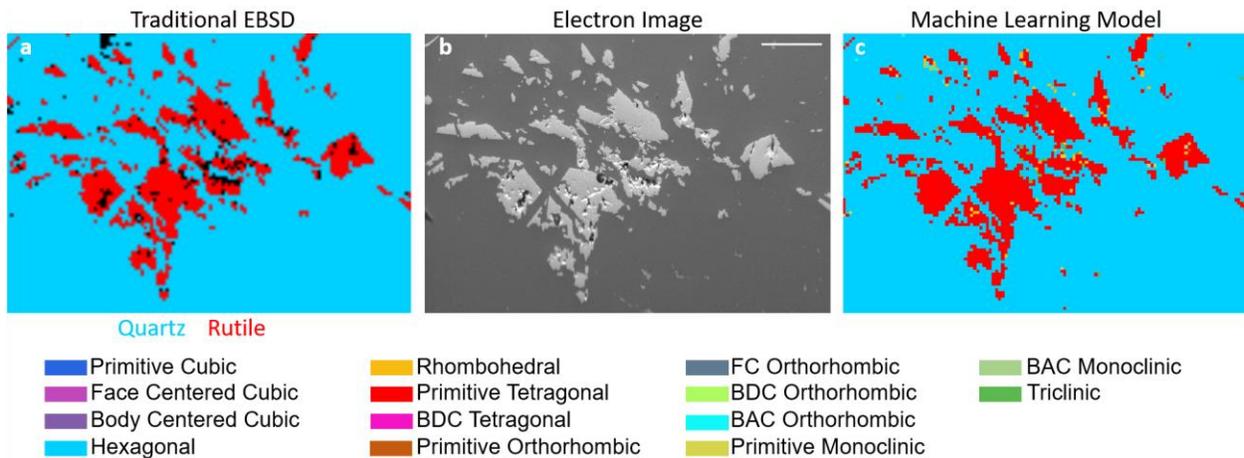
For each of the six materials, each diffraction pattern collected in an EBSD map was evaluated in a random order by the corresponding trained CNN model without further information.

The output classification of each diffraction pattern was recorded and saved in a (.csv) file. A custom Matlab script was used to assign an RGB color value to the predicted class for each diffraction pattern and assign the color value to the corresponding pixel in the EBSD map. Unlike the results obtained from the commercial Oxford Aztec software, no smoothing algorithm was employed for interface boundaries between phases, making the CNN-based maps look more pixelated than their commercial counterpart.

### **7.3 Results and discussion**

Fig. 1 demonstrates the CNN-based EBSD method's ability to overcome the first major challenge traditional Hough-based EBSD is presented with: "What crystal symmetries are present in this material?" The electron image in Fig. 1b clearly displays the two-phase nature of a rutiled quartz sample. When using a commercially available EBSD software package, such as Oxford's Aztec application, the user must first select a list of phases, which serve as the Hough transform libraries, for determining the phase and crystallographic orientation of each diffraction pattern. Assuming the phases are known, the user can select the quartz and rutile libraries to produce the phase map in Fig. 1a. However, there is often uncertainty as to which crystal structures are present in a sample. To alleviate this concern, a CNN-based model that analyzes each diffraction pattern and returns the most likely Bravais lattice, out of the fourteen possible choices, is developed. Fig. 1c demonstrates the model applied to the same high-resolution diffraction patterns collected from the rutiled quartz specimen. The trained CNN can determine, autonomously, the correct Bravais lattice with a high degree of accuracy, reproducing the original user-selected phase map with a high degree of fidelity. This achievement is made more significant by the fact that the CNN was not trained on any diffraction patterns from quartz. When comparing the pixels that differ between the phase maps generated by traditional Hough-based EBSD (Fig. 1a) and the CNN-based EBSD

(Fig. 1c), it is immediately evident that the diffraction patterns classified as neither hexagonal nor primitive tetragonal tend to be located on the boundary between the two phases or in pores. This also tends to be the case for the non-indexed pixels (black) in the traditional Hough-based phase map. The CNN-based approach is also able to identify a number of these non-indexed pixels as belonging to primitive tetragonal (rutile) or hexagonal (quartz). Unlike the Hough-based method, this implementation is required to choose from one of the 14 Bravais lattices; it does not yet have the option to leave a pixel non-indexed based on pattern quality or other metrics. This is a feature that will be explored later in this work. It is important to note that this CNN-based EBSD approach is autonomous, wherein the user is not involved in any phase selection or crystal structure decision making process.



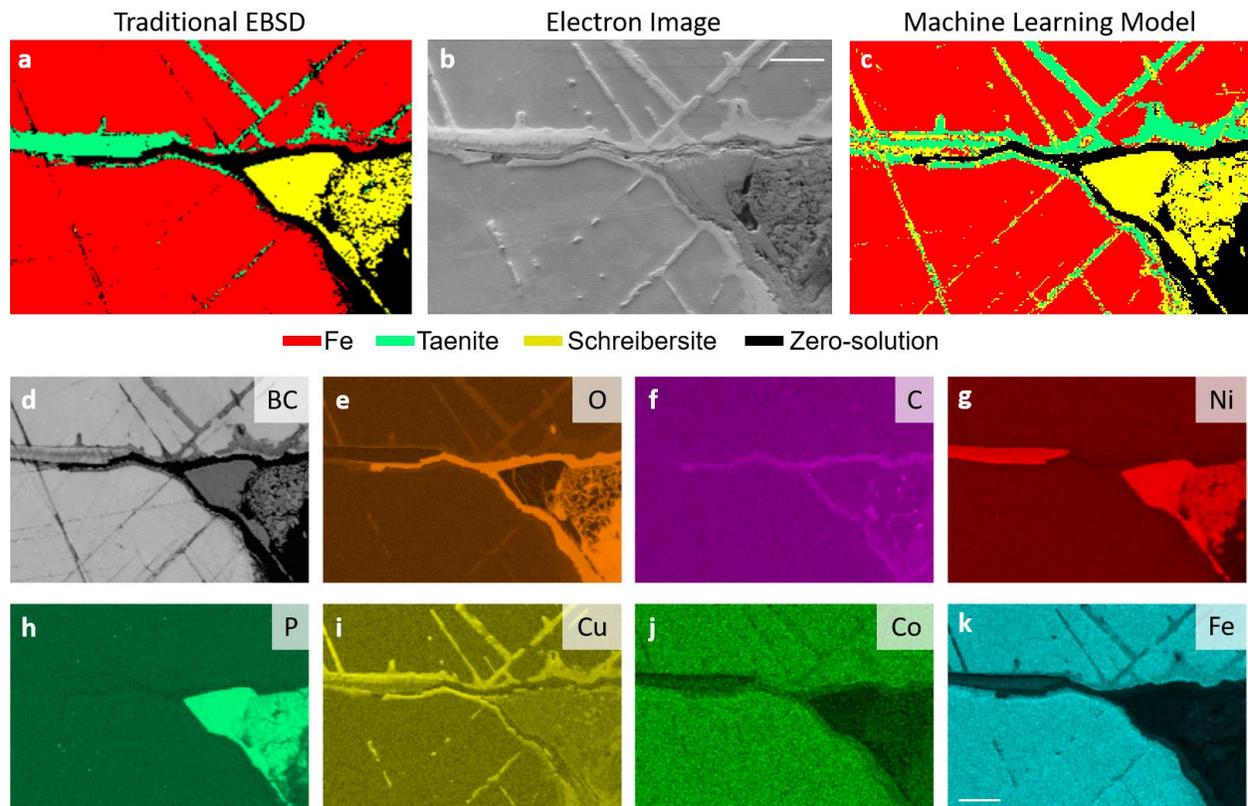
**Figure 7.1 Phase mapping rutiled quartz based on Bravais lattices.** (a) Hough transform EBSD map with rutile and quartz as user-selected phases. (b) Electron image showing the two phase microstructure. (c) Phase map generated from the CNN-based model's analysis of each diffraction pattern individually. Scale bar 250  $\mu\text{m}$ . There are 11,700 total EBSPs (pixels).

In contrast to the previous example, which is designed to be the most broadly applicable, the following examples demonstrate the versatility of convolutional neural networks through a

series of application specific demonstrations, and their ability to address even more challenging crystal structure classification problems encountered in EBSD.

The first example explores a geological specimen with large variation in pattern quality. Fig. 2 compares the performance of the traditional Hough-based approach versus our CNN-based EBSD approach on a sample of the Campo del Cielo meteorite. This specimen contains an Fe rich matrix with isolated regions of taenite and schreibersite throughout. The first challenge is that the sample becomes deeply recessed in a section of the schreibersite phase as seen in the lower right corner of the electron image (Fig. 2b). This significantly impacts the pattern quality in this region and eventually the collected patterns contain only background noise. Therefore, it is important that the CNN-based model has a method for deciding when there is insufficient information in a captured diffraction pattern to make an accurate classification. This is accomplished by adding a zero-solution class as one of the available outputs, as detailed in the methods section. Comparing the output phase map from each method, it is immediately evident that the overall features are very similar. Upon further inspection, it is observed that the CNN-based method is able to fill in a greater percentage of the data from the recessed schreibersite phase, up to the point when the CNN method determines the diffraction patterns do not contain sufficient information to make a classification (Fig. 2c). Moreover, the sections of taenite are more complete along the zero-solution boundaries, including the section below the schreibersite phase, which traditional Hough-based EBSD was unable to map. In the middle left section of the electron image, it is evident that there exists a two-phase structure that traditional Hough-based EBSD identifies only as taenite. The EDS maps (Fig. 2d-k), particularly Cu, Co, Fe, and Ni confirm that the chemistry in this region is indicative of two phases. The CNN identifies the second phase regions as schreibersite, which is a good match chemically and upon inspection of the diffraction patterns captured from that

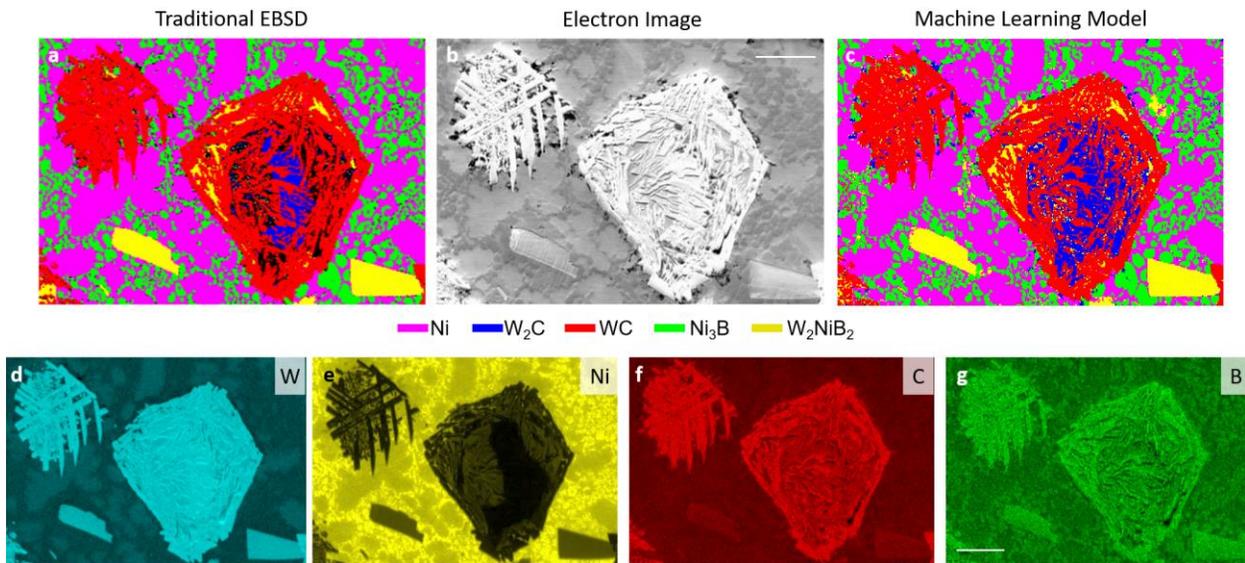
region. In the Hough-based phase map, there are four thin black lines that are not identified due to low quality patterns and diffraction overlap (Fig. 2a). The CNN-based EBSD phase map identifies these regions as a mixture of taenite and schreibersite, and the EDS maps confirm these thin sections to be primarily taenite. In summary, the machine learning-based method is found to be robust to large variations in pattern quality and can be architected with a no-solution class.



**Figure 7.2 Phase mapping a meteorite.** (a) Hough transform EBSD map with user-selected phases. (b) Electron image showing the two phase microstructure. (c) Phase map generated from the CNN-based model's analysis of each diffraction pattern individually. (d) Pattern quality map. (e-k) EDS maps of the analyzed region. Scale bars 50 $\mu$ m. There are 40,784 total EBSPs (pixels).

Fig. 3 compares the performance of traditional Hough-based approach versus our CNN-based EBSD approach on an arc-melted metal matrix composite. This example demonstrates the ability to train the CNN-based model on diffraction patterns collected from 'standards' of each phase (Ni, W<sub>2</sub>C, and WC) and apply the model to identify these phases in a complex multi-phase

sample (described in the Methods section). The 89,280 diffraction patterns collected for Fig. 3 (a,c) are independently analyzed by the CNN, which results in a high fidelity map of the material. Between Fig. 3a and 3c, only ~5,000 pixels (or 5.6%) have a different label and most of these differences are located where the traditional Hough-based method produced no-solution. The CNN-based model excels at applying the information it learned from other systems, or ‘standards’ onto a very different system with a high degree of accuracy.



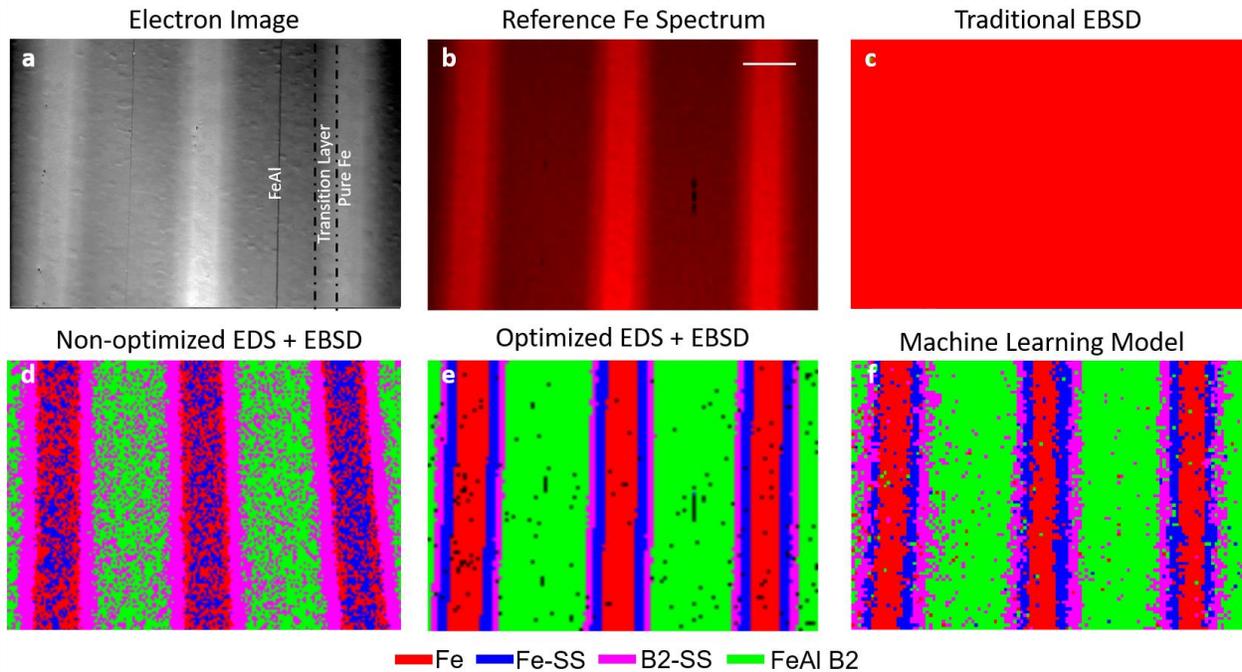
**Figure 7.3 Phase mapping a metal matrix composite.** (a) Hough transform EBSD map with user-selected libraries. (b) Electron image showing the multi-phase microstructure. (c) Phase map generated from the CNN-based model’s analysis of each diffraction pattern individually. (d-g) EDS maps of the analyzed region. Scale bars 25µm. There are 89,280 total EBSPs (pixels).

Furthermore, due to the significant difference in hardness [38], the phases in the nickel-based matrix and tungsten carbide particles are susceptible to vastly different polishing rates. This typically reduces pattern quality in protected areas, such as within the WC particles seen in the electron image. The phase map generated via the CNN-based approach shows that much of this region produces partial EBSD patterns that the Hough-based approach cannot match to a look-up table of interplanar angles. However, the CNN-based model can utilize the information that is

present to identify these EBSPs and primarily classifies them as either  $W_2C$  or  $W_2NiB_2$ . The EDS maps in Fig. 3 (d-g), as well as the measured chemistry (not shown), support these classifications. The map of W and C show that neither of the regions identified as  $W_2C$  or  $W_2NiB_2$  contain as much carbon as the regions mapped to WC in either phase map. Furthermore, the map of Ni shows that it does indeed diffuse into the large WC particles, where it causes the  $W_2NiB_2$  phase to form near the perimeter. The combined achievements on this material show the CNN is robust to partial patterns and can be trained using other samples as ‘standards’.

The next demonstration of the CNN-based EBSD approach is perhaps one of the most challenging types of problems encountered in EBSD. The four phases present in the Fe-Al sample are formed as the result of a diffusion couple style experiment between Fe and Al foils (described in the Methods section). This results in two body centered cubic phases (space group 229), pure Fe and  $\alpha$ -Fe from the Al substitution in the lattice, and two primitive cubic phases (space group 221), nearly equiatomic, FeAl, B2 and the non-equiatomic, B2 solid solution. The backscattered electron image in Fig. 4a shows the alternating layers of the material and gives an early indication that multiple phases are present in the sample. Fig. 4b provides a reference Fe map and the bright columns indicate where pure Fe layers remain after the reaction. If the FeAl B2 and pure Fe diffraction libraries are selected by the user, the Hough-based EBSD is unable to differentiate between the two phases, and a phase map as shown in Fig. 4c is produced. Because the diffraction maxima are so similar between these two phases, every diffraction pattern is identified as both phases with equal confidence, and the software selects, by default, the second phase on the user-selected list. Fig. 4d demonstrates a representative result from a user assigning a reference chemistry from within each phase, using point EDS, to each of the four phases present. If the user selects the reference chemistry for each phase, with the best of intentions and accuracy, a resultant

map as shown in Fig. 4d can be achieved. While the four phases are now each present, the phase map does not accurately describe what is observed in the chemistry analysis. Instead, the pure Fe and  $\alpha$ -Fe (Fe-SS) layers are interspersed randomly (rather than layers), and the B2-SS is scattered throughout the region known to be near-equiatomic FeAl B2. After considerable “optimization” (repeatedly changing the reference chemistry selected in point EDS for each phase), the user can eventually reach a phase map (as shown in Fig. 4e) that looks similar to the known microstructure; this of course can only be achieved because the microstructure in this case is known in advance.



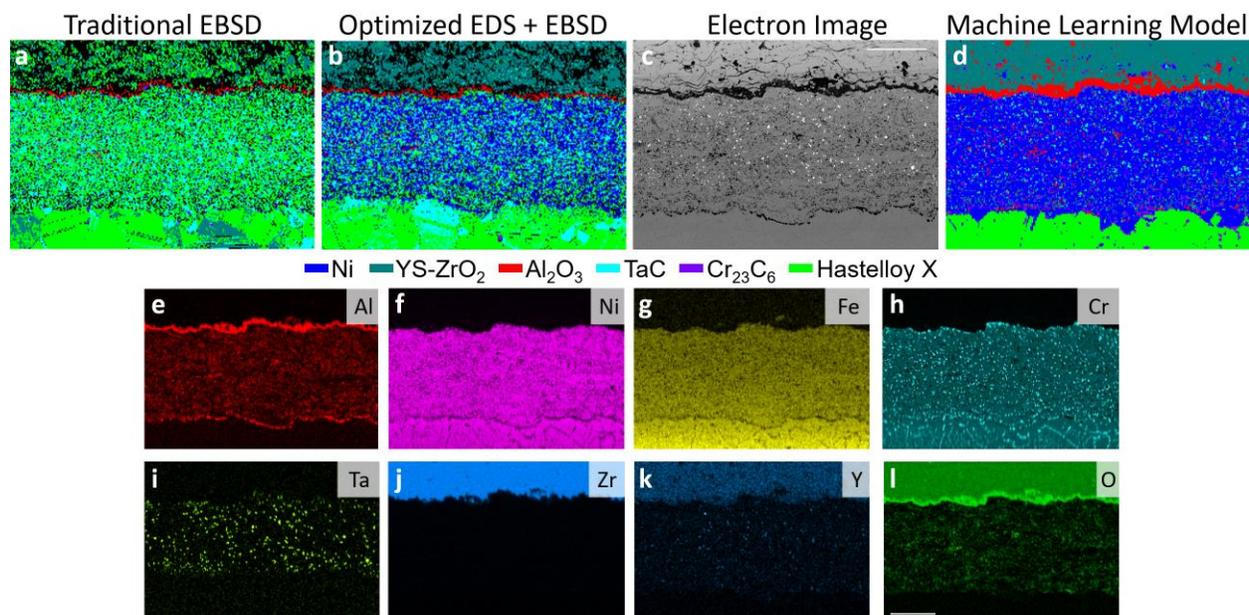
**Figure 7.4 Phase mapping an Fe-Al MIL composite.** (a) Electron image showing the multi-phase microstructure. (b) EDS map of Fe (c) Hough transform EBSD map with user-selected libraries. (d) Phase map generated by Hough transform combined with non-optimized EDS measurements. (e) Phase map generated by Hough transform combined with optimized EDS measurements as reference chemistry. (f) Phase map generated from the CNN-based model’s analysis of each diffraction pattern individually. Scale bar 500 $\mu$ m. There are 10,164 total EBSPs (pixels).

By comparison, a convolutional neural network trained on a small subset of patterns belonging to each phase produces the phase map seen in Fig. 4f. While the phases are not as

perfectly linear as in Fig. 4e, it is likely that the merging of phases at their borders is real to some degree according to the analysis performed by Wang *et al.* [39] and their observation that crystal orientation influences the diffusion rates. Otherwise, the phase map is in good agreement with the expected results, and the CNN-based EBSD method is adept at separating diffraction patterns based on space groups and small changes in chemistry within the same space group.

Convolutional neural networks applied to EBSD can resolve another challenging problem that phase identification including EDS reference chemistry for each phase does not completely resolve. The electron image for a cycled thermal barrier coating containing five FCC phases and one rhombohedral phase is shown in Fig. 5c. The top layer is yttria-stabilized zirconia (YS-ZrO<sub>2</sub>), followed by the thermally grown oxide layer (Al<sub>2</sub>O<sub>3</sub>). In the middle section is a complex bond coat containing Ni, a chrome-carbide phase (Cr<sub>23</sub>C<sub>6</sub>), TaC, and a small amount of Al<sub>2</sub>O<sub>3</sub>. The bottom layer is the nickel-based superalloy Hastelloy X and Cr<sub>23</sub>C<sub>6</sub> at the grain boundaries from thermal cycling. Except for the Al<sub>2</sub>O<sub>3</sub> phase, each of these phases belongs to space group 225, meaning their crystal symmetries are the same, but the lattice parameters can be different. As seen in Fig. 5a, the Hough transform method alone is unable to reliably differentiate the FCC phases and indexes them primarily as Hastelloy X with some interspersed YS-ZrO<sub>2</sub>. The phase fraction of TaC (the brightest spots in the electron image) is much too large (based on the known chemistry) and is incorrectly included in the substrate. The Hough transform performs reasonably well at identifying the band of Al<sub>2</sub>O<sub>3</sub>, which is not surprising given its rhombohedral structure. Including reference chemistry for each of the phases improves the overall quality of the Hough-based EBSD phase map shown in Fig. 5b. The band of TGO is further defined and the YS-ZrO<sub>2</sub> layer contains much fewer erroneous pixels. However, the Hastelloy X substrate has large grains and twins that are being identified as TaC and the Hastelloy X is indexed as approximately 40% of the bond coat

layer (center section). Furthermore, the phase fraction of TaC in the bond coat is still much higher than what is observed to be present in the electron image. Fig. 5d is the result of training a CNN-based model to differentiate these phases. While not free of errors, the resulting phase map is observed to be a higher fidelity mapping of the area shown in the electron image than traditional Hough-based EBSD offers. Comparison with the EDS maps in Fig. 5 (e-l) further confirms the increased plausibility of the phase fractions and the credible identification of each diffraction pattern. The YS-ZrO<sub>2</sub> layer is nearly completely indexed as such with most of the errors present residing at pores and cracks in the oxide. The Al<sub>2</sub>O<sub>3</sub> is now indexed as a continuous band of thermally grown oxide in good agreement with the morphology of the Al EDS map. The central region is almost entirely indexed as Ni with interspersed TaC, Cr<sub>23</sub>C<sub>6</sub>, and Al<sub>2</sub>O<sub>3</sub>. Careful comparison with the EDS maps for Ta, Cr, Al, and O reveal the chemistry to overlap well with the presence, or absence, of the respective phase in the CNN-based phase map. The Hastelloy X substrate is well confined and interlaced with Cr<sub>23</sub>C<sub>6</sub> particles that were not identified in Figs 5a and 5b, but are clearly present in the Cr EDS map.

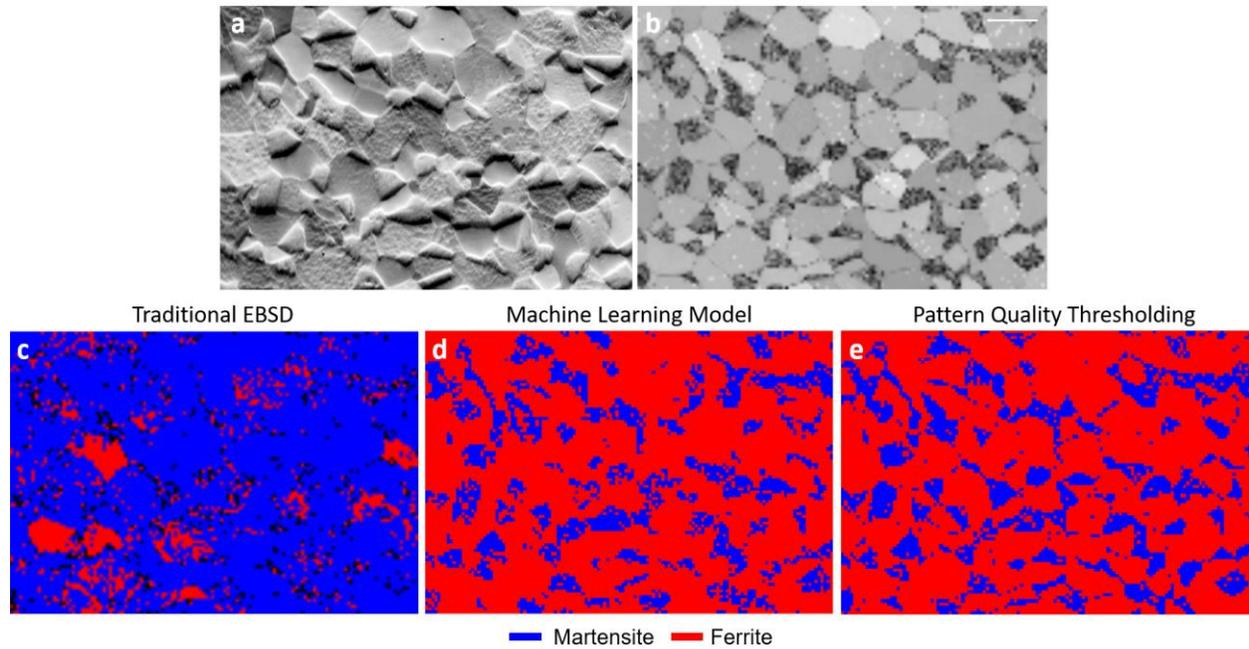


**Figure 7.5 Phase mapping a thermal barrier coating.** (a) Hough transform EBSD map with user-selected libraries. (b) Phase map generated by Hough transform combined with optimized EDS measurements as reference chemistry. (c) Electron image showing the multi-phase microstructure. The Hastelloy X substrate is at the bottom of the image. (d) Phase map generated from the CNN-based model's analysis of each diffraction pattern individually. (e-l) EDS maps of the analyzed region. Scale bars 100 $\mu$ m. There are 45,052 total EBSPs (pixels).

The final demonstration of the potential for CNNs to revolutionize EBSD technology is perhaps the most universally known limitation: the challenge of separating martensite from ferrite. This problem has traditionally been resolved by setting band contrast (or pattern quality) thresholds and accepting some degree of mis-indexing. The determination of the threshold value is biased by human criterion and can result in drastically different maps and phase fractions depending on the user selected threshold. Another applicable method for mapping tetragonality in martensitic steels, called High (Angular) Resolution EBSD (HR-EBSD), uses a reference pattern to measure the relative strain rate of all other patterns [49]. While this technique achieves high sensitivity ( $\sim 10^{-4}$ ) and can be used to back calculate relative  $c/a$  ratios, the absolute strain and  $c/a$  ratio of the reference pattern is typically unknown. This method is not compared herein owing to the need for a dictionary of simulated patterns for the initial calibration. The microstructure of commercial 430

stainless steel is displayed in Fig. 6a. While not yet obvious from the electron image alone, the martensite regions appear raised compared to the ferrite. The band contrast map in Fig. 6b confirms both the location of the martensite as well as the overlap in pattern quality between the martensite and ferrite. This is the same problem encountered in previous work [31] and explains why their pattern quality-based model was dependent on iterative phase discrimination and ‘locking’ pixels the user deemed correctly identified before re-analysis. The region depicted in Fig. 6a is mapped via traditional EBSD (Fig. 6c) using ferrite and martensite libraries generated based on the lattice parameter from XRD and well-established equations from literature [47] (refer to the Methods section for more information). The Hough-based method results in nearly all the 12,236 collected diffraction patterns being identified as martensite. In contrast, the machine learning model trained on ‘standards’ for martensite and ferrite produces a phase map (Fig. 6d) with strong correlation to what is observed in the band contrast map. As seen in Supplementary Figure 39, the martensite and iron diffraction patterns used to train the model as out-of-sample ‘standards’ are of considerable quality. Therefore, unlike the previous study [31], our CNN-based model is not reliant on pattern quality metrics as a singular variable and each diffraction pattern is indexed in one pass through the data. A histogram of the band contrast values is shown in Supplementary Figure 41. A comparison to the common thresholding method is made in Fig. 6e. A threshold of band contrast  $< 45$  is applied to map the lowest quality patterns to martensite. This threshold was selected using the histogram as a guide. The machine learning model compares favorably to this method and many of the most obvious regions overlap. However, the patterns the two models disagree on are not strictly band contrast dependent. In fact, several medium band contrast regions are identified as martensite by the machine learning model (Fig. 6d), while other medium band contrast regions are mapped to martensite when classified by pattern quality (Fig.

6e). The distinction between ferrite and martensite in EBSD has routinely been considered one of the grand challenges in EBSD. With support from well-designed machine learning models, the technique is demonstrated to be capable of highly accurate, autonomous classification of these two phases.



**Figure 7.6 Phase mapping 430 stainless steel.** (a) Electron image showing the microstructure. (b) Band contrast map for the region depicted in (a). (c) Hough transform EBSD map with user-selected libraries for martensite and ferrite. (d) Phase map generated from the machine learning model’s analysis of each diffraction pattern individually, using “out-of-sample” standards. (e) Phase map constructed using a threshold value ( $< 45$ ) for the band contrast in each EBSP. Scale bar  $25\mu\text{m}$ . There are 12,236 total EBSPs (pixels).

## 7.4 Conclusion

In this paper, a CNN-based approach to accessing latent signals in electron backscatter diffraction patterns is developed and demonstrated. The flexibility and utility of this methodology is established by designing training sets for a variety of cases that are challenging problems in traditional Hough-based EBSD, but for which the information is present in the diffraction pattern. The output of the trained convolutional neural networks is compared to phase maps generated

using state-of-the-art Hough-based EBSD systems and practices. In each case, the model is demonstrated to perform well at the task of properly identifying phases, while validating the flexibility of the new methodology.

Major improvements offered by the convolutional neural network-based phase mapping approach are the identification of crystal symmetry (e.g. Bravais lattices or space groups), the correct identification of phases from incomplete or otherwise low-quality diffraction patterns, and the ability to distinguish crystallographically similar phases with subtle differences in chemical composition or lattice parameter. Even martensite and ferrite, which have nearly equivalent crystal structure, can be distinguished using this methodology. Moreover, these are accomplished without the need for additional inputs to the model. Further development of this technology is expected to yield a significant number of advancements to the EBSD platform, specifically this approach can move EBSD from a user-dependent methodology, to an autonomous phase identifying microstructure characterization platform. This work demonstrates these potential capabilities using in-house development of models and available diffraction patterns. The methodology can readily be applied by other researchers for the development of new models. Inclusion of simulated diffraction patterns may help accelerate the process. In the future, we foresee commercial EBSD systems being equipped with standardized models, and potentially even on-demand modelling capabilities, for working with specific use cases or elucidating information about unknown materials.

## **7.5 Acknowledgements**

K. Kaufmann was supported by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program. K. Kaufmann would also like to acknowledge the support of the ARCS Foundation, San Diego Chapter. KV would

like to acknowledge the financial generosity of the Oerlikon Group in support of his research group.

Chapter 7, in full, is a reprint of the material “Phase Mapping in EBSD Using Convolutional Neural Networks”, as it appears in *Microscopy and Microanalysis*. K. Kaufmann, C. Zhu, A. S. Rosengarten, D. Maryanovsky, H. Wang, and K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

## 7.6 References

- [1] C. Zhu, T. Harrington, G.T. Gray, K.S. Vecchio, Dislocation-type evolution in quasi-statically compressed polycrystalline nickel, *Acta Mater.* 155 (2018) 104–116. doi:10.1016/j.actamat.2018.05.022.
- [2] C. Zhu, T. Harrington, V. Livescu, G.T. Gray, K.S. Vecchio, Determination of geometrically necessary dislocations in large shear strain localization in aluminum, *Acta Mater.* 118 (2016) 383–394. doi:10.1016/J.ACTAMAT.2016.07.051.
- [3] A.J. Schwartz, M. Kumar, B.L. Adams, D.P. Field, *Electron backscatter diffraction in materials science*, Springer Science+Business Media, LLC, New York, 2009. doi:10.1007/978-0-387-88136-2.
- [4] O. Engler, V. Randle, *Introduction to Texture Analysis*, Imperial College Press, London, 2010.
- [5] A.J. Wilkinson, G. Meaden, D.J. Dingley, High-resolution elastic strain measurement from electron backscatter diffraction patterns: New levels of sensitivity, *Ultramicroscopy*. (2006). doi:10.1016/j.ultramic.2005.10.001.
- [6] A.J. Wilkinson, T.B. Britton, Strains, planes, and EBSD in materials science, *Mater. Today*. 15 (2012) 366–376. doi:10.1016/S1369-7021(12)70163-3.
- [7] G. Nolze, R. Hielscher, A. Winkelmann, Electron backscatter diffraction beyond the mainstream, *Cryst. Res. Technol.* 52 (2017) 1600252. doi:10.1002/crat.201600252.
- [8] L. Li, M. Han, Determining the Bravais lattice using a single electron backscatter diffraction pattern, *J. Appl. Crystallogr.* (2015). doi:10.1107/S1600576714025989.
- [9] J.R. Michael, R.P. Goehner, Ab-initio primitive cell calculations from EBSD patterns, in: W. D. S. R (Eds.), *Microbeam Anal. 2000, Proc.*, Institute of Physics, Philadelphia, 2000: pp. 203–204. [https://books.google.com/books?hl=en&lr=&id=Ay6KGRmX-3gC&oi=fnd&pg=PA203&dq=Michael,+J.+R.,+%26+Goehner,+R.+P.+\(2000,+January\).+Ab-initio+primitive+cell+calculations+from+EBSD+patterns.+In+CONFERENCE+SERIES-INSTITUTE+OF+PHYSICS+\(Vol.+165,+pp.+203-204\).+Ph](https://books.google.com/books?hl=en&lr=&id=Ay6KGRmX-3gC&oi=fnd&pg=PA203&dq=Michael,+J.+R.,+%26+Goehner,+R.+P.+(2000,+January).+Ab-initio+primitive+cell+calculations+from+EBSD+patterns.+In+CONFERENCE+SERIES-INSTITUTE+OF+PHYSICS+(Vol.+165,+pp.+203-204).+Ph) (accessed January 3, 2019).
- [10] D.J. Dingley, S.I. Wright, Phase Identification Through Symmetry Determination in EBSD Patterns, in: A. Schwartz, M. Kumar, B. Adams, D. Field (Eds.), *Electron Backscatter Diffraction. Mater. Sci.*, Springer, Boston, MA, 2009: pp. 97–107. doi:10.1007/978-0-387-88136-2.
- [11] C.L. Chen, R.C. Thomson, The combined use of EBSD and EDX analyses for the identification of complex intermetallic phases in multicomponent Al-Si piston alloys, *J. Alloys Compd.* 490 (2010) 293–300. doi:10.1016/j.jallcom.2009.09.181.
- [12] T. Karthikeyan, M.K. Dash, S. Saroja, M. Vijayalakshmi, Evaluation of misindexing of EBSD patterns in a ferritic steel, *J. Microsc.* 249 (2013) 26–35. doi:10.1111/j.1365-2818.2012.03676.x.

- [13] S. McLaren, S.M. Reddy, Automated mapping of K-feldspar by electron backscatter diffraction and application to  $^{40}\text{Ar}/^{39}\text{Ar}$  dating, *J. Struct. Geol.* 30 (2008) 1229–1241. doi:10.1016/J.JSG.2008.05.008.
- [14] F. Ram, M. De Graef, Phase differentiation by electron backscatter diffraction using the dictionary indexing approach, *Acta Mater.* 144 (2018) 352–364. doi:10.1016/j.actamat.2017.10.069.
- [15] Y.H. Chen, S.U. Park, D. Wei, G. Newstadt, M.A. Jackson, J.P. Simmons, M. De Graef, A.O. Hero, A Dictionary Approach to Electron Backscatter Diffraction Indexing, *Microsc. Microanal.* (2015). doi:10.1017/S1431927615000756.
- [16] S.U. Park, D. Wei, M. De Graef, M. Shah, J. Simmons, A.O. Hero, EBSD image segmentation using a physics-based forward model, in: 2013 IEEE Int. Conf. Image Process. ICIP 2013 - Proc., 2013: pp. 3780–3784. doi:10.1109/ICIP.2013.6738779.
- [17] S. Singh, M. De Graef, Dictionary Indexing of Electron Channeling Patterns, *Microsc. Microanal.* 23 (2017) 1–10. doi:10.1017/S1431927616012769.
- [18] P.G. Callahan, M. De Graef, Dynamical electron backscatter diffraction patterns. Part I: Pattern simulations, *Microsc. Microanal.* 19 (2013) 1255–1265. doi:10.1017/S1431927613001840.
- [19] M.M. Nowell, S.I. Wright, Phase differentiation via combined EBSD and XEDS, in: *J. Microsc.*, 2004: pp. 296–305. doi:10.1111/j.0022-2720.2004.01299.x.
- [20] R.P. Goehner, J.R. Michael, Phase Identification in a Scanning Electron Microscope Using Backscattered Electron Kikuchi Patterns., *J. Res. Natl. Inst. Stand. Technol.* 101 (1996) 301–308. doi:10.6028/jres.101.031.
- [21] A. Ziletti, D. Kumar, M. Scheffler, L.M. Ghiringhelli, Insightful classification of crystal structures using deep learning, *Nat. Commun.* 9 (2018) 2775. doi:10.1038/s41467-018-05169-6.
- [22] F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N.T.P. Hartono, S. Ramasamy, B.L. DeCost, S.I.P. Tian, G. Romano, A. Gilad Kusne, T. Buonassisi, Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks, *Npj Comput. Mater.* 5 (2019). doi:10.1038/s41524-019-0196-x.
- [23] H. Lombaert, L. Grady, X. Pennec, N. Ayache, F. Cheriet, Spectral log-demons: Diffeomorphic image registration with very large deformations, *Int. J. Comput. Vis.* 107 (2014) 254–271. doi:10.1007/s11263-013-0681-5.
- [24] C. Zhu, H. Wang, K. Kaufmann, K.S. Vecchio, A computer vision approach to study surface deformation of materials, *Meas. Sci. Technol.* 31 (2020) 055602. doi:10.1088/1361-6501/ab65d9.
- [25] H. Wang, L. Dong, J. O’Daniel, R. Mohan, A.S. Garden, K. Kian Ang, D.A. Kuban, M. Bonnen, J.Y. Chang, R. Cheung, Validation of an accelerated “demons” algorithm for deformable image registration in radiation therapy, *Phys. Med. Biol.* 50 (2005) 2887–2905.

doi:10.1088/0031-9155/50/12/011.

- [26] B.L. DeCost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, *Comput. Mater. Sci.* 110 (2015) 126–133. doi:10.1016/J.COMMATSCI.2015.08.011.
- [27] E. Alegre, J. Barreiro, H. Cáceres, L.K. Hernández, R.A. Fernández, M. Castejón, Design of a Computer Vision System to Estimate Tool Wearing, *Mater. Sci. Forum.* 526 (2006) 61–66. doi:10.4028/www.scientific.net/MSF.526.61.
- [28] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, *Nature.* 323 (1986) 533–536. doi:10.1038/323533a0.
- [29] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, *J. Mach. Learn. Res.* 13 (2012) 281–305. <http://scikit-learn.sourceforge.net>. (accessed December 12, 2019).
- [30] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature.* 521 (2015) 436–444. doi:10.1038/nature14539.
- [31] J. Goulden, K. Mehnert, K. Thomsen, H. Jiang, A Method for Separating Crystallographically Similar Phases in Steels using EBSD and Machine Learning, *Microsc. Microanal.* 23 (2017) 110–111. doi:10.1017/S1431927617001234.
- [32] J. Gonzalez, D. Bhowmick, C. Beltran, K. Sankaran, Y. Bengio, Applying Knowledge Transfer for Water Body Segmentation in Peru, (2019). <http://arxiv.org/abs/1912.00957> (accessed March 24, 2020).
- [33] J.A.F. Thompson, M. Schonwiesner, Y. Bengio, D. Willett, How Transferable Are Features in Convolutional Neural Network Acoustic Models across Languages?, in: *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, Institute of Electrical and Electronics Engineers Inc., 2019: pp. 2827–2831. doi:10.1109/ICASSP.2019.8683043.
- [34] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, T.J. Harrington, E. Marin, K.S. Vecchio, Crystal symmetry determination in electron diffraction using machine learning, *Science (80-. ).* 367 (2020) 564–568. doi:10.1126/science.aay3062.
- [35] D. Jha, S. Singh, R. Al-Bahrani, W.K. Liao, A. Choudhary, M. De Graef, A. Agrawal, Extracting grain orientations from EBSD patterns of polycrystalline materials using convolutional neural networks, *Microsc. Microanal.* 24 (2018) 497–502. doi:10.1017/S1431927618015131.
- [36] A. Foden, A. Previero, T.B. Britton, Advances in electron backscatter diffraction, (2019). <http://arxiv.org/abs/1908.04860> (accessed December 9, 2019).
- [37] J. Goulden, P. Trimby, A. Bewick, The Benefits and Applications of a CMOS-based EBSD Detector, *Microsc. Microanal.* 24 (2018) 1128–1129.
- [38] B. Maroli, C. Liu, Overlay welding of NiSiB mixes with tungsten carbides, 2017. [https://www.hoganas.com/globalassets/download-media/technical-papers/pm/pm17\\_08\\_apma\\_overlay-welding-of-nisib-mixes-with-tungsten-](https://www.hoganas.com/globalassets/download-media/technical-papers/pm/pm17_08_apma_overlay-welding-of-nisib-mixes-with-tungsten-)

- carbides\_maroli.pdf (accessed August 19, 2019).
- [39] H. Wang, T. Harrington, C. Zhu, K.S. Vecchio, Design, fabrication and characterization of FeAl-based metallic-intermetallic laminate (MIL) composites, *Acta Mater.* 175 (2019) 445–456. doi:10.1016/j.actamat.2019.06.039.
- [40] C. Mercer, S. Faulhaber, N. Yao, K. McIlwrath, O. Fabrichnaya, Investigation of the chemical composition of the thermally grown oxide layer in thermal barrier systems with NiCoCrAlY bond coats, *Surf. Coatings Technol.* 201 (2006) 1495–1502. doi:10.1016/J.SURFCOAT.2006.02.024.
- [41] T.M. Pollock, D.M. Lipkin, K.J. Hemker, Multifunctional coating interlayers for thermal-barrier systems, *MRS Bull.* 37 (2012) 923–931. doi:10.1557/mrs.2012.238.
- [42] A.G. Evans, D.R. Mumm, J.W. Hutchinson, G.H. Meier, F.S. Pettit, Mechanisms controlling the durability of thermal barrier coatings, *Prog. Mater. Sci.* 46 (2001) 505–553. doi:10.1016/S0079-6425(00)00020-7.
- [43] F. Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, in: *Proc. IEEE Int. Conf. Comput. Vis., The Computer Vision Foundation, Seattle, 2017*: pp. 1251–1258. [http://openaccess.thecvf.com/content\\_cvpr\\_2017/papers/Chollet\\_Xception\\_Deep\\_Learning\\_CVPR\\_2017\\_paper.pdf](http://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf) (accessed February 1, 2019).
- [44] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, in: *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., 2015*: pp. 1–15. <http://arxiv.org/abs/1412.6980> (accessed October 8, 2018).
- [45] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, G. Brain, TensorFlow: A System for Large-Scale Machine Learning, in: *Proc. 12th USENIX Symp. Oper. Syst. Des. Implement., USENIX, Savannah, Georgia, 2016*: pp. 265–283. <https://tensorflow.org> (accessed October 8, 2018).
- [46] F. Chollet, Keras, (2015). <https://github.com/keras-team/keras>.
- [47] L. Xiao, Z. Fan, Z. Jinxiu, Z. Mingxing, K. Mokuang, G. Zhenqi, Lattice-parameter variation with carbon content of martensite. I. X-ray-diffraction experimental study, *Phys. Rev. B.* 52 (1995) 9970–9978. doi:10.1103/PhysRevB.52.9970.
- [48] L.M. Dougherty, E.K. Cerreta, G.T. Gray, C.P. Trujillo, M.F. Lopez, K.S. Vecchio, G.J. Kusinski, Mechanical behavior and microstructural development of low-carbon steel and microcomposite steel reinforcement bars deformed under quasi-static and dynamic shear loading, *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* 40 (2009) 1835–1850. doi:10.1007/s11661-009-9869-2.
- [49] T. Tanaka, A.J. Wilkinson, High Angular Resolution Electron Backscatter Diffraction Studies of Tetragonality in Fe-C Martensitic Steels, *Microsc. Microanal.* 24 (2018) 962–

963. doi:10.1017/s1431927618005305.

## Chapter 8 Efficient *few-shot* machine learning for classification of EBSD patterns

Kevin Kaufmann<sup>1</sup>, Hobson Lane<sup>2,3</sup>, Xiao Liu<sup>4</sup>, and Kenneth S. Vecchio<sup>1,4\*</sup>

<sup>1</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA

<sup>2</sup>Tangible AI LLC, San Diego, CA 92037, USA

<sup>3</sup>Department of Healthcare Research and Policy, UC San Diego – Extension, San Diego, CA 92037, USA

<sup>4</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA

\*Corresponding author.

### Abstract

Deep learning is quickly becoming a standard approach to solving a range of materials science objectives, particularly in the field of computer vision. However, labeled datasets large enough to train neural networks from scratch can be challenging to collect. One approach to accelerating the training of deep learning models such as convolutional neural networks is the transfer of weights from models trained on unrelated image classification problems, commonly referred to as transfer learning. The powerful feature extractors learned previously can potentially be fine-tuned for a new classification problem without hindering performance. Transfer learning can also improve the results of training a model using a small amount of data, known as *few-shot* learning. Herein, we test the effectiveness of a *few-shot* transfer learning approach for the classification of electron backscatter diffraction (EBSD) pattern images to six space groups within

the  $(4/m\bar{3}2/m)$  point group. Training history and performance metrics are compared with a model of the same architecture trained from scratch. In an effort to make this approach more explainable, visualization of filters, activation maps, and Shapley values are utilized to provide insight into the model's operations. The applicability to real-world phase identification and differentiation is demonstrated using dual phase materials that are challenging to analyze with traditional methods.

## 8.1 Introduction

Data science-based methods to materials development and analysis have gained great popularity in recent years [1–13]. Deep learning algorithms are of significant interest owing to their excellent performance without significant feature engineering, and the ubiquity of these methods will likely continue owing to the outperformance of systems directly designed by humans. While often difficult to assess how and why these ‘black box algorithms’ are capable of performing these tasks, these methods can provide significant value or spark new insights [14,15]. Application of these tools to image-based tasks in materials science has proved to be useful for classification [16–19], segmentation [20–22], and other objectives [23–25]. While deep learning provides significant opportunities for the advancement of materials science, robust application of these tools often requires much larger datasets than are typically available within the materials science community. Utilizing the knowledge deep neural networks have learned from other domains offers an opportunity to develop models in these domains, where data is sparse and further collection and labeling could be slow or tedious [26–28].

Convolutional neural networks (CNNs) are a class of deep learning models that have proven effective for analyzing image data [29]. Before a CNN can be applied to a given task, it must learn to assign importance (learnable weights and biases) to various aspects of the image that

maximize the network's differentiation capabilities. Two general strategies exist for training convolutional neural networks: 1) the weights can be randomly initialized, or 2) the weights can be transferred from a model pre-trained on a separate but related task, often in a nearby domain with significantly more data, and then refined for the current objective. The first approach, commonly referred to as "training from scratch", requires a large dataset to avoid overfitting and perform robustly on new, real-world examples. The second approach, referred to as "transfer learning", can significantly reduce the number of training examples required, accelerate the training process, and retain or exceed the performance garnered by training from scratch [27,30–32]. The transfer learning method is motivated by the human ability to intelligently apply previously learned knowledge to solve new problems faster or with better solutions [33]. Despite the potential, knowledge transfer from a given source domain is not guaranteed to improve performance in the target domain and can in fact hinder performance [26,32]. Furthermore, one of the requirements to use this approach is that the images in the new domain must conform to the processed shape and structure determined at the outset of the previous training. For models pre-trained on ImageNet [34], a library of over one million images labeled with one thousand classes, the expected input is usually  $299 \times 299$  pixels with 3-channels (one each for RGB). When dealing with one-channel grayscale images, such as those typically collected from electron diffraction studies, the decision to use transfer learning necessitates the stacking of a single image into a pseudo-color image [31].

The number of labeled images that can reasonably be collected must also be considered for the appropriate training and application of a CNN. Computer vision research has recently been motivated by children's ability to learn novel visual concepts almost effortlessly after accumulating sufficient past knowledge [35]. In deep learning and computer vision, learning

visual models of object categories has notoriously required tens of thousands of training examples [36]; however, recent research has demonstrated that it is possible to classify images accurately using relatively few labeled examples with the appropriate combination of pretraining of the CNN layers on unrelated image classification training sets [37,38], adversarial or unsupervised learning [39,40], network pruning [41], and micro architecture tuning [42]. Once several categories have been learned the hard way, learning new categories should become more efficient. The increased efficiency allows for a lesser number of images to be used in training, referred to as a “*few shots*”.

Electron backscatter diffraction (EBSD) patterns (EBSPs) are an excellent case study for the use of *few-shot* transfer learning toward accelerating analysis of electron diffraction data. The scanning electron microscope (SEM)-based method involves the capture of 2D diffraction patterns produced from an incident electron beam scattering, diffracting, and escaping from a well-polished ‘bulk’ sample [43]. The collected diffraction patterns contain significant structural and chemical information and are similar to those collected in other techniques such as convergent beam electron diffraction (CBED) [44,45]. Despite the vast amount of information in the patterns, conventional EBSD has primarily focused on determining the three-dimensional orientation of individual grains in crystalline materials [43,46–48]. Furthermore, the commercial technique typically relies on Hough-based indexing with a look-up table of interplanar angles constructed from the set of selected reflectors for phases specified by the user [49]. This generally allows for phase differentiation of sufficiently distinct crystal structures [50–52], but the process remains susceptible to structural misclassification [53–55]. Improvements to phase differentiation have been proposed and developed including dictionary indexing [56–59], spherical indexing [60–62], and more recently machine learning [63]. While each offers significant advantages over the Hough-based method, these tools continue to require assumptions about the number of phases

and/or their structure. For example, the dictionary-based approach requires simulation of a “master” pattern for each potential phase and every experimental pattern is matched against a dictionary of simulated patterns for all potential phases [56]. The highest similarity match is selected as the most likely phase and orientation. Another available solution to phase differentiation is combining an electron backscattered pattern (EBSP) with information from energy dispersive X-ray spectroscopy (EDS) or wavelength dispersive X-ray spectroscopy (WDS) [64–66]. While these have been adopted commercially, the singular EBSP is still analyzed with the Hough-based method and an expert user is required to evaluate the plausibility of returned matches. Applications using hand-drawn lines overlaid on individual Kikuchi diffraction patterns have been developed for determining the Bravais lattice or point group; however, they remain limited by at least one of the following: analysis time per pattern, the need for an expert crystallographer, or necessitating multiples of the same diffraction pattern with different SEM settings [67]. Clearly, there remains a need for a rapid EBSP classification tool capable of functioning with one EBSD pattern while remaining suitable for even the most novice user.

Recently, the EBSD community has begun to explore the use of convolutional neural networks as a foundation for addressing modern EBSD challenges [16,68–70]. Despite the marked advances these works have made, the requirements for simulating [71] or collecting experimental diffraction patterns from a significant number of materials and crystallographic orientations remains a limiting factor. In this work, we test the validity of *few-shot* transfer learning, starting from ImageNet weights, applied to classify EBSPs to one of six space groups. Herein, space group refers to the symmetry group of a configuration in three-dimensional space. We compare the time to converge, the individual kernels (weights of the neurons), activation maps, and the performance of models trained from scratch or by transfer learning. Though there has been considerable

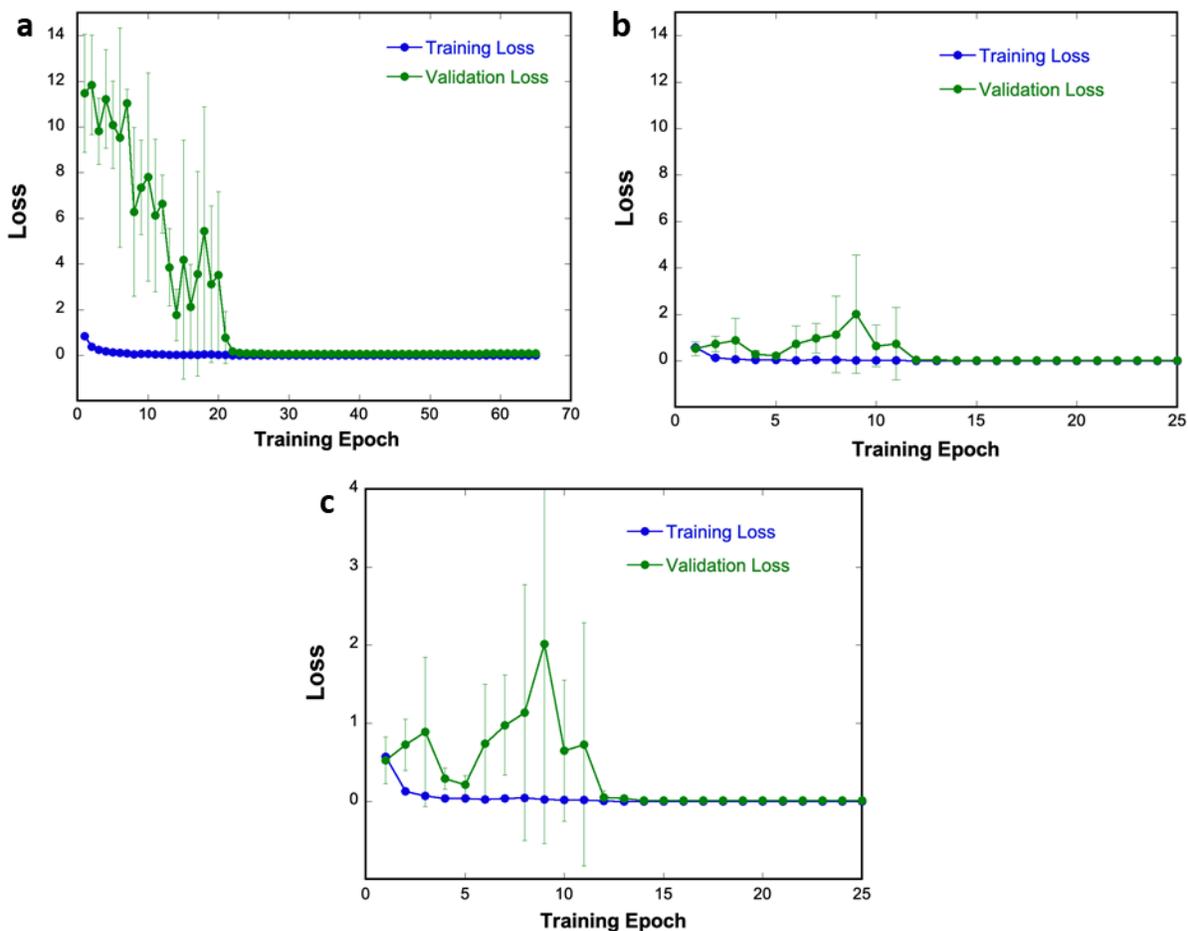
progress on interpretability of machine learning systems, mostly in the field of eXplainable AI (XAI) [72], fully understanding the internal mechanisms of deep neural networks is still an area of active research [15,73]. Visualization of the most similar weights each model independently learned, their respective activation maps, and Shapley values can increase understanding of how the artificial intelligence models accomplishes its task. Building these XAI foundations can increase trust in the model's later predictions and help identify when the prediction is incorrect. In addition to evaluating each model's performance on holdout data, each model is also tested with 6,900 EBSPs collected from a Ni<sub>90</sub>Al<sub>10</sub> sample outside the training, testing, and validation data, and a space group map is generated from the individual classifications.

## **8.2 Results and Discussion**

### **8.2.1 Training metrics.**

Each model is first trained until the validation loss converged using a small number of the available diffraction patterns from each material in the six available space groups. The training and validation loss were recorded at the end of each pass through the training set (i.e. an epoch) to monitor the model's performance as the weights are updated. The validation loss is also monitored to determine when training should cease to prevent model underfitting or overfitting owing to a fixed number of training epochs. The loss function (categorical cross-entropy) is plotted in Supplementary Figure 42 for reference. The goal of the neural network is to minimize the loss function by maximizing its prediction of the correct class. Fig. 1a shows statistics for the training and validation loss for the model trained from scratch in grayscale. While the average training loss is low ( $0.84 \pm .04$ ) by the end of the first epoch, the average validation loss is observed to be  $11.5 \pm 2.5$ . In comparison, by initializing the model with weights learned from ImageNet, the average training and validation loss are both observed to be small  $0.56 \pm .02$  and  $0.52 \pm .3$ ;

respectively) from the start (Fig. 1b). Fig. 1c shows a magnified view of Fig. 1b to show the detailed training history. The improved starting loss, presumably owing to the strong filters learned in pre-training, also aids in rapid model convergence. Including the fifteen epochs used to establish convergence, the model trained from scratch required  $50 \pm 12$  epochs, while the transfer learning model required only  $26 \pm 3$  epochs; this represents a 2-fold reduction in the average number of passes through the training set. Since each epoch with 2,400 diffraction patterns requires two minutes on this hardware, a time savings of nearly an hour per training event is gained. With the amount of training data available to the CNN expected to grow, and therefore the time per epoch expected to increase, the time savings will become more pronounced.



**Figure 8.1 Training and validation history statistics for the two models.** (a) The model trained from scratch using grayscale images. (b) The model trained starting from ImageNet weights using 3-channel stacked images. Data is plotted on the same y-axis as in (b). (c) The data in (b) for the transfer learning model plotted on a different scale. Error bars are one standard deviation from five trials per approach.

### 8.2.2 Performance on holdout data

The time saving advantages garnered by fine-tuning the weights of an existing model are only valuable if the model performs similarly well or exceeds the model trained from scratch. Table 1 shows the class-weighted *Precision* and *Recall* for the best performing model from each approach, for which further analysis into the internal operations will be studied. See Supplementary Table 17 and Supplementary Table 18 for a breakdown of *Precision* and *Recall* by space group for the transfer learned and trained from scratch models, respectively. Both metrics

are improved for the model using transfer learning even though the training, validation, and test sets were held constant. This implies that the feature extractors learned from ImageNet are not only relevant to this new domain, but also at least as valuable as those learned from scratch. This likely results from the more general nature of the feature extractors necessary for optimal performance on ImageNet [74]. It is possible that increasing the size of the dataset used to train the model from scratch could increase its performance to near that of the *few-shot* transfer learning model; however, the cost to training time would be notable. For this study, it was also important to keep the dataset and its partitions fixed for more deterministic comparisons.

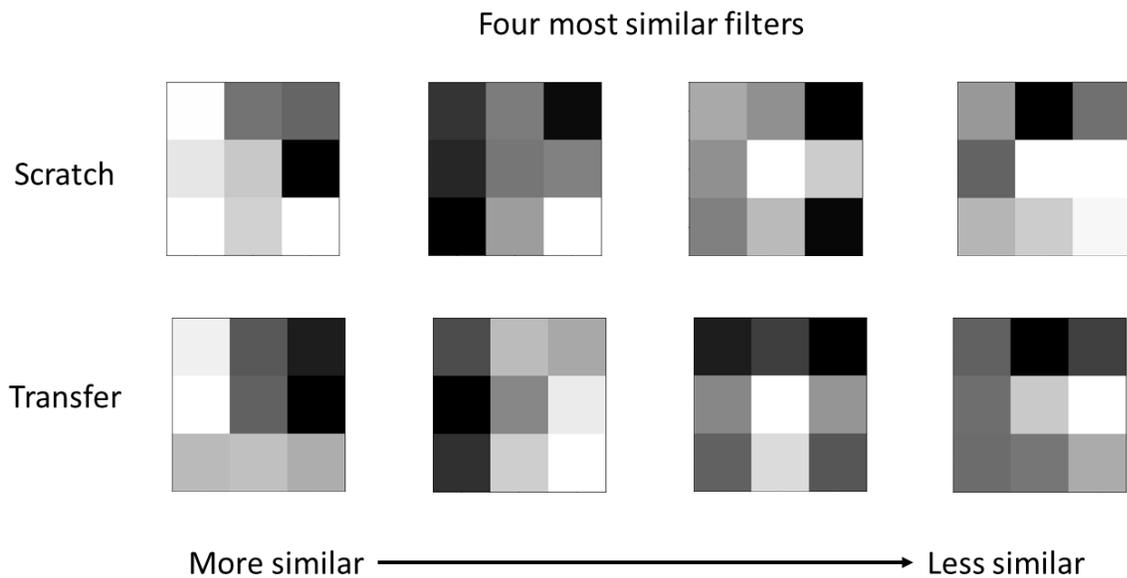
**Table 8.1 Classification metrics.** The class-weighted average Precision and Recall on the test data for each model. Both metrics are improved using the transfer learning approach to training the model. The same test data was used to benchmark each model.

	<i>Precision</i>	<i>Recall</i>
<b>Trained from scratch</b>	0.93	0.92
<b>Transfer learning</b>	0.97	0.96

### 8.2.3 Visualization

Since deep neural networks are used in this study, meaningful information about the internal mathematical operations performed necessitates studying several aspects of the model's inner workings. The first study involves visualization of the filters (or kernels) and corresponding feature maps. Filters and feature maps from the earliest layers in the model are most useful since deeper layers become more abstract. To identify corresponding filters in the two models, the Euclidean distance between all possible pairs of filters in a selected layer from the grayscale and pseudo RGB models were calculated and the similarity was ranked. The four most similar kernels between the two models in the first convolution layer are shown in Fig. 2. The pairs are grouped

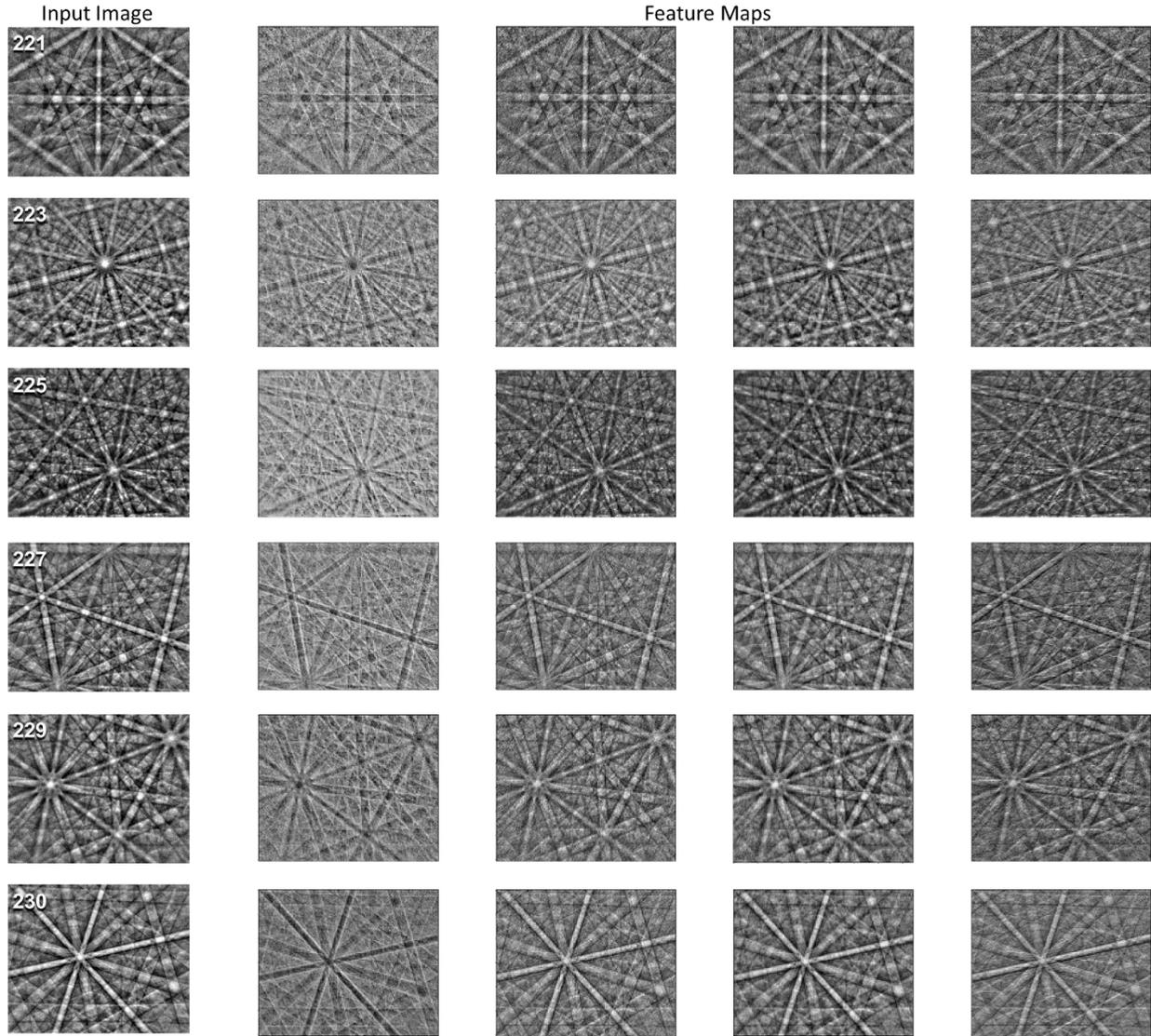
in columns between the model trained from scratch and using transfer learning. The independent co-learning of these filters suggest they are very valuable for feature tuning and selection on EBSPs. At this low-level in the model, the filters will predominantly be designed to identify edges at various orientations. The next layer will likely combine those edges into corners and small points, and the subsequent layers will figure out larger and larger shapes/features, such as the number of and relative angles between Kikuchi lines. Two of these four filters are recognizable as classical edge detection operators. The second filter has converged close to the x-direction Sobel edge detection operator [75] and the third filter resembles the Gabor filters with theta  $\theta = 135^\circ$ . Visualization of the feature maps resulting from these four filters can provide insight into their function.



**Figure 8.2 Kernel pairs with lowest Euclidean distance.** Kernel pairs are grouped by column between the two models. The leftmost pair has the lowest Euclidean distance, with the kernels becoming less similar moving to the right.

Fig. 3 shows the result of each filter from Fig. 2 being individually convolved over an input image from the six space groups studied. All feature maps shown are from the grayscale model.

The first filter, and therefore the most similar filter between the two models, is primarily observed to perform an inversion on the input image. As a result, the band edges in each image are activated (white regions) and become more distinct. It is reasonable to speculate that deeper parts of the network extract the angle and relative locations of these intersections, a function that was also suggested in Ding *et al.*'s recent work[70]. The fourth filter is observed to have normalized the contrast of the input image, perhaps to reduce the effects of atomic scattering factors (Z-contrast) observed in prior work [68]. Further analysis of the class-specific feature importance can further improve understanding of the neural network's methodology.



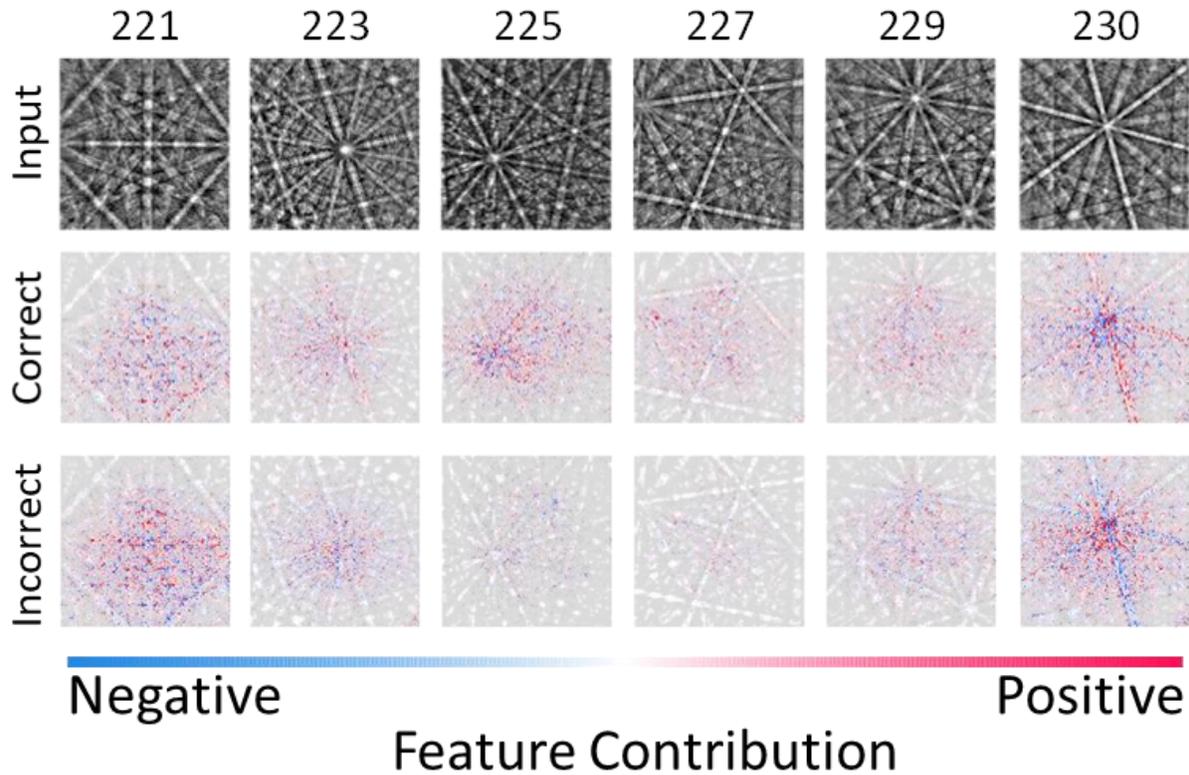
**Figure 8.3 Selected feature maps from the first convolution layer.** Feature maps extracted from the grayscale model corresponding to the filters shown previously in Fig. 2. One input image per space group is shown. From top to bottom, the six space groups are  $Pm\bar{3}m$ ,  $Pm\bar{3}n$ ,  $Fm\bar{3}m$ ,  $Fd\bar{3}m$ ,  $Im\bar{3}m$ , and  $Ia\bar{3}d$ .

### 8.2.4 Feature importance

Measurement and visualization of feature importance is another method to increase understanding of the deep neural network. This type of analysis is performed to help determine whether one should trust a prediction and why. There are a number of techniques available for CNNs including Gradient-weighted Class Activation Mapping (Grad-CAM) [76], activation

maximization [77], LIME [78], and Shapley values [79]. Several of those listed have been effectively demonstrated in other works involving EBSD patterns [16,17,68,70]; however, this is the first to use Shapley values. In game theory, Shapley values are a solution to fairly distributing the gains and costs of several actors working in coalition. By definition, each Shapley value is the average expected marginal contribution of one actor after all possible combinations have been considered. In this case, the actors are the features in the images. Essentially, the Shapley value is the average expected marginal contribution of one actor after all possible combinations have been considered. While not perfect, this has proven a fair approach to allocating value in a variety of fields [80–82]. The results of this analysis, further described in the Methods section, are shown in Fig. 4. Refer to Supplementary Figure 43 for a demonstration of SHAP analysis on handwritten numbers.

The previously unseen input images are shown in the top row, and as semi-transparent grayscale backings behind each of the explanations. After random selection of the input image, it was verified that the model correctly identified the space group (thus pseudo-random selection). The middle row in Fig. 4 corresponds to the explanations for the correct prediction, while the bottom row displays the explanations for the next most likely (i.e. incorrect) space group. All 6 Shapley explanations ordered by most to least likely space group for each input are displayed in Supplementary Figure 44. The positive (red) and negative (blue) contributions to each prediction are primarily at diffraction maxima (e.g. zone axes), band intersections, and outlining band edges. This lends credence that the model is indeed utilizing information grounded in the physics of EBSD. The clustering of Shapley values near zone axes is further reaffirming given the abundance of information and their role in classical diffraction pattern indexing [83].



**Figure 8.4 Visual explanation of feature contributions.** Shapley values are computed for each input image to gauge the importance of features in the EBSPs. The first row is the raw input image. The second row corresponds to the Shapley values for the correct prediction. Row three corresponds to the first incorrect classification as ranked by softmax probability. From left to right, the six space groups are  $Pm\bar{3}m$ ,  $Pm\bar{3}n$ ,  $Fm\bar{3}m$ ,  $Fd\bar{3}m$ ,  $Im\bar{3}m$ , and  $Ia\bar{3}d$ .

### 8.2.5 Performance in practice

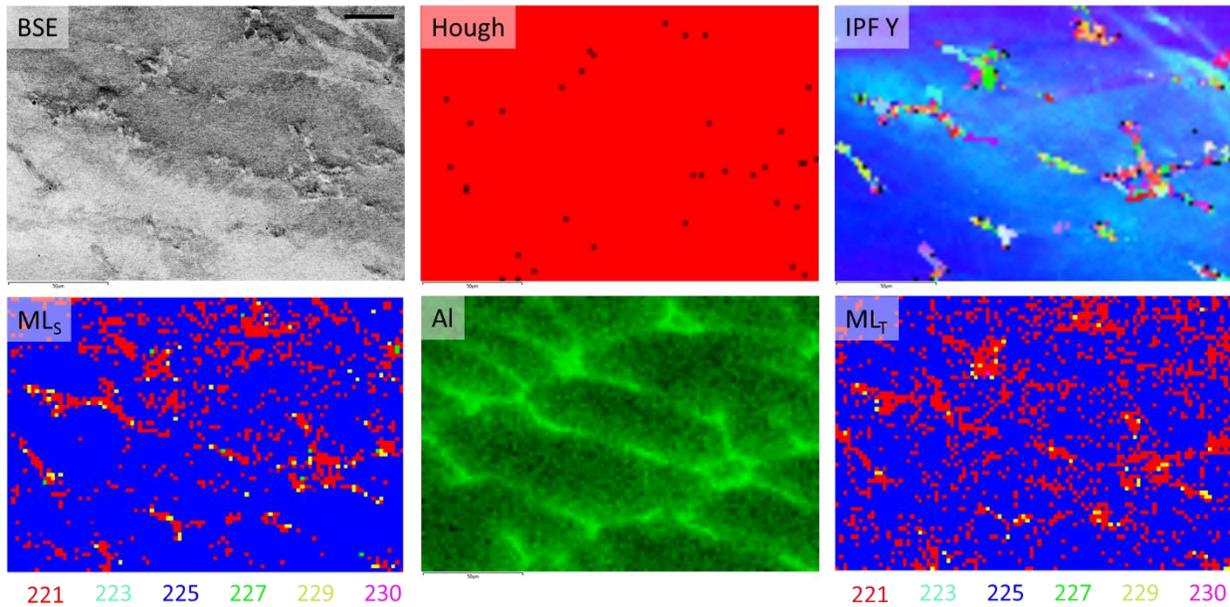
It is also of importance to demonstrate and compare the efficacy of both models in a real-world context, not only the patterns set aside for testing. In this case, the EBSD mapping of a dual phase sample serves to demonstrate that both model training strategies are capable in situations where the diffraction patterns are not collected from an ideal, single-phase material. Fig. 5 top-left shows a backscattered electron image of a  $Ni_{90}Al_{10}$  (wt%) sample containing a Ni-rich matrix (space group 225) along with  $Ni_3Al$  precipitates (space group 221) appearing raised from the surface. It is of importance to note that the model has not yet encountered a solid-solution phase

such as the Ni with Al matrix present in this sample. The Al content in the Ni matrix was determined to be 19.7% +/- 0.53% (at%). The Al chemistry of the Ni<sub>3</sub>Al precipitates was determined to be 25.7% +/- 0.74% (at%). While partially selected for this reason, the material was primarily chosen since space groups 221 and 225 have two of the lowest F1-scores in each model and can readily be produced in this singular sample. Furthermore, it was important that the phases could be differentiated visually and easily to the reader, such as with EDS maps. While this means the phases could potentially be differentiated if the EBSD operator assigned reference chemistries to the phases in advance of collecting the EBSPs, the purpose of this demonstration is really to compare the transfer learning and trained from scratch models' abilities to identify the space group without further information. For the most complete phase identification in EBSD (i.e. lattice parameters), multiple analysis methods (e.g. XRD and EDS) may need to be employed.

The Hough-based phase map in Fig. 5 is a representative image of the expected results from commercial systems. The phase map consists of entirely one phase (shown in red) and Oxford Aztec software is observed to predict Ni and Ni<sub>3</sub>Al with an equal number of bands and mean angular deviation (MAD), effectively a combined measure of certainty, for each diffraction pattern. It is the order of phase selection by the user in Oxford Aztec software that ultimately determines whether Ni or Ni<sub>3</sub>Al is selected in this case. Example diffraction patterns from each phase are shown in Supplementary Figure 45. Black pixels remain unindexed by the Hough method, typically a result of poor diffraction pattern quality. The inverse pole figure (IPF) in the Y-direction is provided along with the EDS map for aluminum to elucidate where the Ni<sub>3</sub>Al is expected to be found within the Ni matrix.

The last two images in Fig. 5 show the phase maps produced by the scratch (ML<sub>S</sub>) and *few-shot* transfer learned (ML<sub>T</sub>) models. Over a statistical number of diffraction patterns, the two

models are expected to produce similar answers with some variance, although the test set would suggest that the transfer learning model will generally outperform. Comparing the results with the Al EDS map and IPF Y, both models perform well at identifying the space group of each EBSP with a low false positive rate for the four space groups known not to be present. There are almost certainly some errors with regard to the EBSPs classified as space group 221 ( $Pm\bar{3}m$ ) or 255 ( $Fm\bar{3}m$ ), the *Precision* and *Recall* of each model alert the user of this in advance; however, the results over these 6,900 EBSPs demonstrate the improvement over the Hough-based approach and, more importantly for this study, the robustness of a *few-shot* transfer learning approach.



**Figure 8.5 Phase mapping a dual-phase sample.** Top left shows a backscattered electron image of the area to be mapped. The  $Ni_3Al$  precipitates appear raised in the Ni-rich matrix. Top middle shows the Hough transform-based phase map. Red pixels are identified as Ni or  $Ni_3Al$  with equal certainty, while black pixels were not solved. Top right is the inverse pole figure map in the Y-direction. Bottom left is the phase map produced using the predictions from the grayscale CNN. Bottom middle shows the aluminum EDS map. Bottom right is the phase map produced using the predictions from the *few-shot* transfer learning approach. There are 6,900 total EBSPs (pixels). Scale bar =  $25\mu m$ .

The number of EBSPs each model classifies to the available space groups is tallied in Table 2. A total of 6,900 diffraction patterns were individually identified by each neural network without any other information provided. The transfer learning model is observed to have a reduced misclassification rate to the space groups 227 ( $Fd\bar{3}m$ ), 229 ( $Im\bar{3}m$ ), and 230 ( $Ia\bar{3}d$ ) in this phase map. The largest difference between the two models is for the diffraction patterns classified to space group 221; the class with the lowest *Precision* for each of the two models. The total difference of 517 diffraction patterns only equates to 7% of the total diffraction patterns; well within reason and the expected margin of error between the two models, particularly between these two space groups for the current models. Of those 517, only 480 of these predictions differ between space groups 221 and 225, the other 37 differences are due to false positives (i.e. 227, 229, and 230) in the model trained from scratch. The phase fraction of  $\text{Ni}_3\text{Al}$  likely lies somewhere between what is predicted by these two models. The results of this comparison also suggests that there is future opportunity to construct an approach leveraging Bayesian deep learning or an ensemble of (i.e. at least two) individually trained models making individual classifications combined with model averaging (e.g. voting) to reduce variance, provide insight into overall uncertainty, and identify when “no solution” is an appropriate answer [84–86].

**Table 8.2 Tabulated predictions for the dual-phase sample.** The number of patterns classified to each space group by the respective model. The difference is calculated by subtracting the number predicted by the *few-shot* transfer learning model from that predicted by the model trained from scratch. Parentheses denotes the transfer learning model predicted fewer EBSPs belonging to the respective class. From left to right, the six space groups are  $Pm\bar{3}m$ ,  $Fm\bar{3}m$ ,  $Fd\bar{3}m$ ,  $Im\bar{3}m$ , and  $Ia\bar{3}d$ .

	<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>
<b>Scratch</b>	984	0	5,822	10	72	12
<b>Transfer</b>	1,501	0	5,342	0	56	1
<b>Difference</b>	517	0	480	(10)	(16)	(11)

Thus, a *few-shot* transfer learning approach to classifying electron backscatter diffraction patterns is an attractive method for leveraging the knowledge a deep neural network has attained in a previous context. The convolutional neural network-based approach to diffraction pattern classification is advantageous in that it requires little or no *a-priori* knowledge of the phases in a new sample and can readily be improved or expanded to new classes with the inclusion of new data. The similarity of EBSD patterns to those from techniques such as CBED suggests the *few-shot* transfer learning approach could also apply and potentially be more beneficial given the slower rate of data collection with other electron diffraction methods. Limitations of the current models exist in the number of space groups currently differentiable and the “black box” nature of neural networks. The number of space groups the model can learn to differentiate can be continuously expanded as more data becomes available for training. One of the goals of this work is to discern whether the *few-shot* transfer learning approach can be used to reduce the amount of data necessary for robust expansion to all 230 space groups or other diffraction pattern classification tasks. Indeed, we find that this approach does not hinder the model’s performance

on holdout or entirely new data and offers accelerated training time. While it can be difficult to precisely determine how the CNN performs this task, recent advances in eXplainable AI (e.g. SHAP) provides tools for developing insight and trust in the model's predictions. The combination of ease of scaling, flexibility of the framework, and ability to assess aspects of a model's decision process support the utilization of CNNs and *few-shot* transfer learning as another tool for phase differentiation and symmetry identification in electron diffraction.

## 8.3 Materials and methods

### 8.3.1 Materials

Eighteen different single-phase materials, comprising 6 of the 10 space groups within the  $(4/m\bar{3}2/m)$  point group, were selected for training the space group classification CNN. Suitable samples for the remaining 4 space groups could not be obtained. The six space groups are  $Pm\bar{3}m$ ,  $Pm\bar{3}n$ ,  $Fm\bar{3}m$ ,  $Fd\bar{3}m$ ,  $Im\bar{3}m$ , and  $Ia\bar{3}d$ . Numerically, these are space groups 221, 223, 225, 227, 229, and 230. Space groups 221 and 223 are primitive cubic, 225 and 227 are face centered cubic, and 229 and 230 are body centered cubic. Each of the six space groups share the 3-fold rotary inversion necessary for inclusion in the  $(4/m\bar{3}2/m)$  point group. Supplementary Table 19 details the similarities and differences between the symmetry operations of the six space groups. The materials were FeAl, NiAl, Ni<sub>3</sub>Al, Fe<sub>3</sub>Ni, Cr<sub>3</sub>Si, Mo<sub>3</sub>Si, Ni, Al, NbC, TaC, TiC, Si, Ge, W, Ta, Fe, Al<sub>4</sub>CoNi<sub>2</sub>, and Al<sub>4</sub>Ni<sub>3</sub>. These materials were of low texture, typically less than 2 times random in any direction. Refer to Kaufmann *et al.* for the distributions of orientation, band contrast, and mean angular deviation for these samples [69].

A dual-phase material known to challenge Hough-based EBSD was fabricated to demonstrate and compare the capabilities of each CNN training approach. An additional constraint for the material selected was that the two space groups be identifiable within an EDS map, even

though this meant an operator could force the Hough-based method to differentiate the phases by chemistry if they knew the phases in advance. An ingot of Ni<sub>90</sub>Al<sub>10</sub> (wt%) was arc melted and processed via hot rolling at 600°C to 45% reduction in thickness followed by aging at 600°C for 4 hours and air cooled. X-ray diffraction (XRD) using a Rigaku Miniflex X-ray Diffractometer with a 1D detector, a step size of 0.02°, 5° per minute scan rate, and Cu K $\alpha$  radiation (wavelength  $\lambda = 1.54059 \text{ \AA}$ ) was performed to confirm the existence of phases belonging to space groups 221 ( $Pm\bar{3}m$ ) and 225 ( $Fm\bar{3}m$ ) (Supplementary Figure 46).

### 8.3.2 Electron backscatter diffraction pattern collection

EBSD patterns (EBSPs) were collected as previously described in Kaufmann *et al.* [68]. Diffraction patterns were collected using a Thermo Scientific (formerly FEI) Apreo scanning electron microscope (SEM) equipped with an Oxford Symmetry EBSD detector utilized in high resolution (1244x1024) mode. The geometry of the setup was held constant as follows. The working distance was 18.1 mm  $\pm$  0.1 mm. Oxford Aztec software was used to set the detector insertion distance to 160.2 and the detector tilt to -3.1. The imaging parameters were 20kV accelerating voltage, 51nA beam current, 0.8ms  $\pm$  0.1ms dwell time, and 30-pattern averaging. The Hough indexing parameters were 12 Kikuchi bands, a Hough resolution of 250, and band center indexing.

After collecting high resolution EBSPs from each material, all patterns collected were exported as tiff images. The images were resized for the CNN using the `resize` function in `scikit-image`. All collected data for each material was individually assessed by the neural network, and the collection of images for each sample may contain partial or low-quality diffraction patterns, which could decrease the accuracy of their identification. The test data was not filtered to better assess the model as it would be applied in practice.

### **8.3.3 Neural network architecture**

The well-studied convolutional neural network architecture Xception [87] was selected as the basis architecture for fitting a model that determines which space group a diffraction pattern originated from. The Xception architecture was used without modifications for training the model from scratch and the transfer learning process to facilitate comparison of the training metrics, performance, and internal workings. Selection of this network was partially based on Xception or derivatives of Xception being used previously in the EBSD community [16,68–70]. Xception is also a standard model with ImageNet weights readily available in deep learning APIs such as Keras [88]. A schematic of the convolutional neural network operating on an EBSP is provided in Supplementary Figure 47. Due to space constraints, only the resultant feature maps from selected convolutional layers are shown after image input and before the 2048-dimensional vector. For a complete description of the Xception architecture, please refer to Fig. 5 in Xception: Deep Learning with Depthwise Separable Convolutions [87].

### **8.3.4 Neural network training**

For both the transfer learning and from scratch approaches, training was performed using 400 diffraction patterns per space group. The diffraction patterns supplied at training were evenly divided between the number of materials per space group that the model had access to during training. For example, if the model was given two materials of the same space group during training, 200 diffraction patterns per material were made available. The validation set contained 100 diffraction patterns per space group, equivalent to the standard 80:20 train/validation split. The validation set was only used to monitor the training progress and model convergence. The test set contains the rest of the patterns (a total of 145,453 images; refer to Supplementary Table 17 for class distribution) that were not used for training or validation. The images selected for

training, testing, and validation were the exact same for transfer learning and for from scratch learning.

Model hyperparameters were selected or tuned as follows. Adam (adaptive moment estimation) optimization with a learning rate of 0.001 [89], and a minimum delta of 0.001 as the validation loss were employed for convergence criteria. Adam is chosen for its ability to work well with little hyperparameter tuning, relatively low memory requirements, and its ability to smooth the steps of gradient descent using momentum. Monitoring of validation loss, i.e. early stopping criteria, was employed instead of a fixed number of epochs to allow both models the necessary epochs to converge while keeping the risk of overfitting to the training data low. The patience criteria for validation loss convergence was set to 15 epochs to allow for sufficient certainty that the model had converged and was unlikely to meaningfully improve. The weight decay was set to  $1e^{-5}$  following previous optimization work[87]. The CNNs were implemented with TensorFlow [90] and the Keras API [88] and model training was performed using an NVIDIA Titan V.

### **8.3.5 Diffraction pattern classification**

Each diffraction pattern collected, but not used in training (>140,000 images), was evaluated in a random order by the corresponding trained CNN model without further information. The output classification of each diffraction pattern was recorded, saved in a (.csv) file, and tabulated. *Precision* and *Recall* were calculated for each material and each space group using Scikit-learn [91]. *Precision* (equation 7) for each class (e.g. 230) is defined as the number of correctly predicted images out of all photos predicted to belong to that class (e.g. 230). *Recall* (equation 8) is the number of correctly predicted images for each class divided by the actual number of images for the class. F1-score is the weighted harmonic of the *Precision* and *Recall*

and is particularly valuable in situations where the number of test images per class is variable. A high F1-score means the model has low false positives and low false negatives.

$$Precision = \frac{true\ positives}{true\ positives + false\ positives} \quad \text{Equation 7}$$

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives} \quad \text{Equation 8}$$

### 8.3.6 Neural network insight.

Comparisons between the resultant models from pseudo RGB transfer learning and training the model from scratch offer an opportunity to understand how the CNNs go about their given task. Visualization techniques are implemented using the keras-vis package v0.4.1 [77]. Filters from the first layer of each model are extracted and plotted as 3×3 matrices with matplotlib [92]. The first layer was targeted since the earliest filters represent lower level features such as colors and edges. The Euclidean distance between the individual filter arrays in each model was computed using the NumPy linear algebra toolbox to compute L2 norms [93] and the four most similar learned filters between the two models were identified. The outputs (a.k.a. feature maps) of the first layer corresponding to these four filters are also extracted to examine the activations of the two approaches. The feature maps from earlier convolution layers are more useful since deeper layers operate in feature space and are therefore more difficult to understand [29,70]. Lastly, Shapley values, deeply rooted in game theory [94,95], are estimated using the DeepSHAP tools in SHAP [96]. SHAP uses a distribution of background samples, approximates the model with a linear function between each background data sample and the current input to be explained, and assumes the input features are independent to compute approximate SHAP values. The sum of the SHAP values equals the difference between the expected model output (averaged over the background dataset) and the current model output. One hundred images per space group were

used as background samples in conjunction with one new input image per space group. The total number of diffraction patterns used as a background follows the SHAP software protocols. When plotted as an overlay, red pixels represent positive SHAP values that increase the probability of the class, while blue pixels represent negative SHAP values that reduce the probability of the class. An example of SHAP analysis on handwritten digits from the MNIST database [97] is shown in Supplementary Figure 43. For a given image, the presence and absence of features that positively correlate with a class are shown in red, while negative correlations are shown in blue. As an example, in the image of a ‘four’ the lack of a connection on top makes it a four instead of a nine. Combined, these insights into the operations of the neural network can further substantiate the validity of the transfer learning approach, increase trust by better understanding the model’s methods, and provide indications in cases where the model is incorrect about future predictions.

#### **8.4 Acknowledgements**

The authors would like to thank William M. Mellor for composing Supplementary Table 19.

Chapter 8, in full, is a reprint of the material “Efficient few-shot machine learning for classification of EBSD patterns”, as it appears in *Scientific Reports*. K. Kaufmann, H. Lane, X. Liu, & K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

## 8.5 References

- [1] J. O'Mara, B. Meredig, K. Michel, Materials Data Infrastructure: A Case Study of the Citrination Platform to Examine Data Import, Storage, and Access, *JOM*. 68 (2016) 2031–2034. doi:10.1007/s11837-016-1984-0.
- [2] L. Ward, A. Dunn, A. Faghaninia, N.E.R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K.A. Persson, G.J. Snyder, I. Foster, A. Jain, Matminer: An open source toolkit for materials data mining, *Comput. Mater. Sci.* 152 (2018) 60–69. doi:10.1016/j.commatsci.2018.05.018.
- [3] T.P. McAuliffe, A. Foden, C. Bilsland, D. Daskalaki Mountanou, D. Dye, T.B. Britton, Advancing characterisation with statistics from correlative electron diffraction and X-ray spectroscopy, in the scanning electron microscope, *Ultramicroscopy*. 211 (2020) 112944. doi:10.1016/j.ultramic.2020.112944.
- [4] K. Kaufmann, K.S. Vecchio, Searching for high entropy alloys: A machine learning approach, *Acta Mater.* 198 (2020) 178–222. doi:10.1016/j.actamat.2020.07.065.
- [5] Z. Qiao, M. Welborn, A. Anandkumar, F.R. Manby, T.F. Miller, OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features, *J. Chem. Phys.* 153 (2020) 124111. doi:10.1063/5.0021955.
- [6] D. Mrdjenovich, M.K. Horton, J.H. Montoya, C.M. Legaspi, S. Dwaraknath, V. Tshitoyan, A. Jain, K.A. Persson, propnet: A Knowledge Graph for Materials Science, *Matter*. 2 (2020) 464–480. doi:10.1016/j.matt.2019.11.013.
- [7] D.P. Tabor, L.M. Roch, S.K. Saikin, C. Kreisbeck, D. Sheberla, J.H. Montoya, S. Dwaraknath, M. Aykol, C. Ortiz, H. Tribukait, C. Amador-Bedolla, C.J. Brabec, B. Maruyama, K.A. Persson, A. Aspuru-Guzik, Accelerating the discovery of materials for clean energy in the era of smart automation, *Nat. Rev. Mater.* 3 (2018) 5–20. doi:10.1038/s41578-018-0005-z.
- [8] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, *Comput. Mater. Sci.* 68 (2013) 314–319. doi:10.1016/j.commatsci.2012.10.028.
- [9] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD), *JOM*. 65 (2013) 1501–1509. doi:10.1007/s11837-013-0755-4.
- [10] D. Jha, L. Ward, A. Paul, W. Liao, A. Choudhary, C. Wolverton, A. Agrawal, ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition, *Sci. Rep.* 8 (2018) 17593. doi:10.1038/s41598-018-35934-y.
- [11] F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N.T.P. Hartono, S. Ramasamy, B.L. DeCost, S.I.P. Tian, G. Romano, A. Gilad Kusne, T. Buonassisi, Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks, *Npj*

- Comput. Mater. 5 (2019). doi:10.1038/s41524-019-0196-x.
- [12] K. Kaufmann, D. Maryanovsky, W.M. Mellor, C. Zhu, A.S. Rosengarten, T.J. Harrington, C. Oses, C. Toher, S. Curtarolo, K.S. Vecchio, Discovery of high-entropy ceramics via machine learning, *Npj Comput. Mater.* 6 (2020) 42. doi:10.1038/s41524-020-0317-6.
- [13] B.L. DeCost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, *Comput. Mater. Sci.* 110 (2015) 126–133. doi:10.1016/J.COMMATSCI.2015.08.011.
- [14] E.A. Holm, In defense of the black box, *Science* (80-. ). 364 (2019) 26–27. doi:10.1126/science.aax0162.
- [15] A. Adadi, M. Berrada, Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), *IEEE Access.* 6 (2018) 52138–52160. doi:10.1109/ACCESS.2018.2870052.
- [16] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, T.J. Harrington, E. Marin, K.S. Vecchio, Crystal symmetry determination in electron diffraction using machine learning, *Science* (80-. ). 367 (2020) 564–568. doi:10.1126/science.aay3062.
- [17] A. Foden, A. Previero, T.B. Britton, Advances in electron backscatter diffraction, (2019). <http://arxiv.org/abs/1908.04860> (accessed December 9, 2019).
- [18] A. Ziletti, D. Kumar, M. Scheffler, L.M. Ghiringhelli, Insightful classification of crystal structures using deep learning, *Nat. Commun.* 9 (2018) 2775. doi:10.1038/s41467-018-05169-6.
- [19] M.H. Modarres, R. Aversa, S. Cozzini, R. Ciancio, A. Leto, G.P. Brandino, Neural Network for Nanoscience Scanning Electron Microscope Image Recognition, *Sci. Rep.* 7 (2017) 1–12. doi:10.1038/s41598-017-13565-z.
- [20] T. Stan, Z.T. Thompson, P.W. Voorhees, Optimizing convolutional neural networks to perform semantic segmentation on large materials imaging datasets: X-ray tomography and serial sectioning, *Mater. Charact.* 160 (2020) 110119. doi:10.1016/j.matchar.2020.110119.
- [21] B.L. DeCost, B. Lei, T. Francis, E.A. Holm, High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel, *Microsc. Microanal.* 25 (2019) 21–29. doi:10.1017/S1431927618015635.
- [22] G. Roberts, S.Y. Haile, R. Sainju, D.J. Edwards, B. Hutchinson, Y. Zhu, Deep Learning for Semantic Segmentation of Defects in Advanced STEM Images of Steels, *Sci. Rep.* 9 (2019) 1–12. doi:10.1038/s41598-019-49105-0.
- [23] K. de Haan, Z.S. Ballard, Y. Rivenson, Y. Wu, A. Ozcan, Resolution enhancement in scanning electron microscopy using deep learning, *Sci. Rep.* 9 (2019) 1–7. doi:10.1038/s41598-019-48444-2.
- [24] T. Xie, J.C. Grossman, Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties, *Phys. Rev. Lett.* 120 (2018) 145301.

doi:10.1103/PhysRevLett.120.145301.

- [25] S.R. Spurgeon, C. Ophus, L. Jones, A. Petford-Long, S. V. Kalinin, M.J. Olszta, R.E. Dunin-Borkowski, N. Salmon, K. Hattar, W.C.D. Yang, R. Sharma, Y. Du, A. Chiaramonti, H. Zheng, E.C. Buck, L. Kovarik, R.L. Penn, D. Li, X. Zhang, M. Murayama, M.L. Taheri, Towards data-driven next-generation transmission electron microscopy, *Nat. Mater.* (2020). doi:10.1038/s41563-020-00833-z.
- [26] M.T. Rosenstein, Z. Marx, L.P. Kaelbling, T.G. Dietterich, To Transfer or Not To Transfer, in: *Neural Inf. Process. Syst. (NIPS '05) Work. Inductive Transf. 10 Years Later, 2005*. <http://web.engr.oregonstate.edu/~tgd/publications/rosenstein-marx-kaelbling-dietterich-hnb-nips2005-transfer-workshop.pdf> (accessed October 12, 2020).
- [27] D. Jha, K. Choudhary, F. Tavazza, W. keng Liao, A. Choudhary, C. Campbell, A. Agrawal, Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning, *Nat. Commun.* 10 (2019) 1–12. doi:10.1038/s41467-019-13297-w.
- [28] J.A.F. Thompson, M. Schonwiesner, Y. Bengio, D. Willett, How Transferable Are Features in Convolutional Neural Network Acoustic Models across Languages?, in: *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, Institute of Electrical and Electronics Engineers Inc., 2019: pp. 2827–2831. doi:10.1109/ICASSP.2019.8683043.
- [29] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature*. 521 (2015) 436–444. doi:10.1038/nature14539.
- [30] X. Pan, L. Li, H. Yang, Z. Liu, Y. He, Z. Li, Y. Fan, Z. Cao, L. Zhang, Multi-task Deep Learning for Fine-Grained Classification/Grading in Breast Cancer Histopathological Images, in: *Stud. Comput. Intell.*, Springer Verlag, 2020: pp. 85–95. doi:10.1007/978-3-030-04946-1\_10.
- [31] Y. Xie, D. Richmond, Pre-training on Grayscale ImageNet Improves Medical Image Classification, in: *Eur. Conf. Comput. Vis. Work.*, Springer, Cham, 2018: pp. 0–0.
- [32] J. Gonzalez, D. Bhowmick, C. Beltran, K. Sankaran, Y. Bengio, Applying Knowledge Transfer for Water Body Segmentation in Peru, (2019). <http://arxiv.org/abs/1912.00957> (accessed March 24, 2020).
- [33] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (2010) 1345–1359. doi:10.1109/TKDE.2009.191.
- [34] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, Li Fei-Fei, ImageNet: A large-scale hierarchical image database, in: *IEEE Conf. Comput. Vis. Pattern Recognit.*, Institute of Electrical and Electronics Engineers (IEEE), Miami, FL, 2010: pp. 248–255. doi:10.1109/cvpr.2009.5206848.
- [35] P. Bloom, *How children learn the meanings of words*, MIT Press, Cambridge, MA, 2000.
- [36] P.F. Felzenszwalb, D.P. Huttenlocher, Pictorial structures for object recognition, *Int. J. Comput. Vis.* 61 (2005) 55–79. doi:10.1023/B:VISI.0000042934.15159.49.

- [37] L. Fei-Fei, R. Fergus, P. Perona, One-shot learning of object categories, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (2006) 594–611. doi:10.1109/TPAMI.2006.79.
- [38] B.M. Lake, R. Salakhutdinov, J.B. Tenenbaum, Human-level concept learning through probabilistic program induction, *Science* (80-. ). 350 (2015) 1332–1338. doi:10.1126/science.aab3050.
- [39] R. Zhang, T. Che, Z. Ghahramani, Y. Bengio, Y. Song, MetaGAN: An Adversarial Approach to Few-Shot Learning, in: *NeurIPS 2018*, 2018: pp. 2365–2374.
- [40] M.-Y. Liu, X. Huang, A. Mallya, T. Karras, T. Aila, J. Lehtinen, J. Kautz, Few-Shot Unsupervised Image-to-Image Translation, 2019. <https://github.com/NVLabs/FUNIT>. (accessed July 20, 2020).
- [41] H. Li, A. Kadav, I. Durdanovic, H. Samet, H.P. Graf, Pruning Filters for Efficient ConvNets, 5th Int. Conf. Learn. Represent. (2016) 1–13. <http://arxiv.org/abs/1608.08710> (accessed July 20, 2020).
- [42] Z. Guo, X. Zhang, H. Mu, W. Heng, Z. Liu, Y. Wei, J. Sun, Single Path One-Shot Neural Architecture Search with Uniform Sampling, (2019). <http://arxiv.org/abs/1904.00420> (accessed July 20, 2020).
- [43] A.J. Schwartz, M. Kumar, B.L. Adams, D.P. Field, *Electron backscatter diffraction in materials science*, Springer Science+Business Media, LLC, New York, 2009. doi:10.1007/978-0-387-88136-2.
- [44] K.S. Vecchio, D.B. Williams, Convergent beam electron diffraction study of Al<sub>3</sub>Zr in Al-Zr AND Al-Li-Zr alloys, *Acta Metall.* 35 (1987) 2959–2970. doi:10.1016/0001-6160(87)90295-1.
- [45] K.S. Vecchio, D.B. Williams, Convergent beam electron diffraction analysis of the T<sub>1</sub> (Al<sub>2</sub>CuLi) phase in Al-Li-Cu alloys, *Metall. Trans. A.* 19 (1988) 2885–2891. doi:10.1007/BF02647714.
- [46] V.S. Tong, A.J. Knowles, D. Dye, T. Ben Britton, Rapid electron backscatter diffraction mapping: Painting by numbers, *Mater. Charact.* 147 (2019) 271–279. doi:10.1016/J.MATCHAR.2018.11.014.
- [47] K. Thomsen, N.H. Schmidt, A. Bewick, K. Larsen, J. Goulden, Improving the accuracy of orientation measurements using EBSD, *Microsc. Microanal.* 19 (2013) 724–725.
- [48] C. Zhu, K. Kaufmann, K.S. Vecchio, Novel remapping approach for HR-EBSD based on demons registration, *Ultramicroscopy.* 208 (2020) 112851. doi:10.1016/j.ultramic.2019.112851.
- [49] N.C.K. Lassen, *Automated Determination of Crystal Orientations from Electron Backscattering Patterns*, The Technical University of Denmark, 1994.
- [50] T.B. Britton, C. Maurice, R. Fortunier, J.H. Driver, A.P. Day, G. Meaden, D.J. Dingley, K. Mingard, A.J. Wilkinson, Factors affecting the accuracy of high resolution electron

- backscatter diffraction when using simulated patterns, *Ultramicroscopy*. 110 (2010) 1443–1453. doi:10.1016/j.ultramic.2010.08.001.
- [51] R. Hielscher, F. Bartel, T.B. Britton, Gazing at crystal balls: Electron backscatter diffraction pattern analysis and cross correlation on the sphere, *Ultramicroscopy*. 207 (2019). doi:10.1016/j.ultramic.2019.112836.
- [52] A. Foden, D.M. Collins, A.J. Wilkinson, T.B. Britton, Indexing electron backscatter diffraction patterns with a refined template matching approach, *Ultramicroscopy*. 207 (2019) 112845. doi:10.1016/j.ultramic.2019.112845.
- [53] T. Karthikeyan, M.K. Dash, S. Saroja, M. Vijayalakshmi, Evaluation of misindexing of EBSD patterns in a ferritic steel, *J. Microsc.* 249 (2013) 26–35. doi:10.1111/j.1365-2818.2012.03676.x.
- [54] C.L. Chen, R.C. Thomson, The combined use of EBSD and EDX analyses for the identification of complex intermetallic phases in multicomponent Al-Si piston alloys, *J. Alloys Compd.* 490 (2010) 293–300. doi:10.1016/j.jallcom.2009.09.181.
- [55] S. McLaren, S.M. Reddy, Automated mapping of K-feldspar by electron backscatter diffraction and application to <sup>40</sup>Ar/<sup>39</sup>Ar dating, *J. Struct. Geol.* 30 (2008) 1229–1241. doi:10.1016/J.JSG.2008.05.008.
- [56] F. Ram, M. De Graef, Phase differentiation by electron backscatter diffraction using the dictionary indexing approach, *Acta Mater.* 144 (2018) 352–364. doi:10.1016/j.actamat.2017.10.069.
- [57] Y.H. Chen, S.U. Park, D. Wei, G. Newstadt, M.A. Jackson, J.P. Simmons, M. De Graef, A.O. Hero, A Dictionary Approach to Electron Backscatter Diffraction Indexing, *Microsc. Microanal.* (2015). doi:10.1017/S1431927615000756.
- [58] S. Singh, Y. Guo, B. Winiarski, T.L. Burnett, P.J. Withers, M. De Graef, High resolution low kV EBSD of heavily deformed and nanocrystalline Aluminium by dictionary-based indexing, *Sci. Rep.* (2018). doi:10.1038/s41598-018-29315-8.
- [59] F. Ram, S. Wright, S. Singh, M. De Graef, Error analysis of the crystal orientations obtained by the dictionary approach to EBSD indexing, *Ultramicroscopy*. 181 (2017) 17–26. doi:10.1016/J.ULTRAMIC.2017.04.016.
- [60] A.P. Day, Spherical EBSD, in: *J. Microsc.*, 2008. doi:10.1111/j.1365-2818.2008.02011.x.
- [61] W.C. Lenthe, S. Singh, M. De Graef, A spherical harmonic transform approach to the indexing of electron back-scattered diffraction patterns, *Ultramicroscopy*. 207 (2019) 112841. doi:10.1016/j.ultramic.2019.112841.
- [62] C. Zhu, K. Kaufmann, K. Vecchio, Automated Reconstruction of Spherical Kikuchi Maps, *Microsc. Microanal.* (2019). doi:10.1017/S1431927619000710.
- [63] T.P. McAuliffe, D. Dye, T. Ben Britton, Spherical-angular dark field imaging and sensitive microstructural phase clustering with unsupervised machine learning, (2020).

<http://arxiv.org/abs/2005.10581> (accessed June 1, 2020).

- [64] M.M. Nowell, S.I. Wright, Phase differentiation via combined EBSD and XEDS, in: *J. Microsc.*, 2004: pp. 296–305. doi:10.1111/j.0022-2720.2004.01299.x.
- [65] R.P. Goehner, J.R. Michael, Phase Identification in a Scanning Electron Microscope Using Backscattered Electron Kikuchi Patterns., *J. Res. Natl. Inst. Stand. Technol.* 101 (1996) 301–308. doi:10.6028/jres.101.031.
- [66] D.J. Dingley, S.I. Wright, Phase Identification Through Symmetry Determination in EBSD Patterns, in: A. Schwartz, M. Kumar, B. Adams, D. Field (Eds.), *Electron Backscatter Diffr. Mater. Sci.*, Springer, Boston, MA, 2009: pp. 97–107. doi:10.1007/978-0-387-88136-2.
- [67] L. Li, M. Han, Determining the Bravais lattice using a single electron backscatter diffraction pattern, *J. Appl. Crystallogr.* (2015). doi:10.1107/S1600576714025989.
- [68] K. Kaufmann, C. Zhu, A.S. Rosengarten, K.S. Vecchio, Deep Neural Network Enabled Space Group Identification in EBSD, *Microsc. Microanal.* 26 (2020) 447–457. doi:10.1017/S1431927620001506.
- [69] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, H. Wang, K.S. Vecchio, Phase Mapping in EBSD Using Convolutional Neural Networks, *Microsc. Microanal.* 26 (2020) 458–468. doi:10.1017/S1431927620001488.
- [70] Z. Ding, E. Pascal, M. De Graef, Indexing of electron back-scatter diffraction patterns using a convolutional neural network, *Acta Mater.* 199 (2020) 370–382. doi:10.1016/j.actamat.2020.08.046.
- [71] P.G. Callahan, M. De Graef, Dynamical electron backscatter diffraction patterns. Part I: Pattern simulations, *Microsc. Microanal.* 19 (2013) 1255–1265. doi:10.1017/S1431927613001840.
- [72] W. Samek, G. Montavon, A. Vedaldi, L.K. Hansen, K.-R. Müller, eds., *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, Springer International Publishing, Cham, 2019. doi:10.1007/978-3-030-28954-6.
- [73] S. Carter, Z. Armstrong, L. Schubert, I. Johnson, C. Olah, Activation Atlas, *Distill.* 4 (2019) e15. doi:10.23915/distill.00015.
- [74] C.A. Graff, J. Ellen, Correlating Filter Diversity with Convolutional Neural Network Accuracy, in: *Institute of Electrical and Electronics Engineers (IEEE)*, 2017: pp. 75–80. doi:10.1109/icmla.2016.0021.
- [75] N. Kanopoulos, N. Vasanthavada, R.L. Baker, Design of an Image Edge Detection Filter Using the Sobel Operator, *IEEE J. Solid-State Circuits.* 23 (1988) 358–367. doi:10.1109/4.996.
- [76] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, in: *Proc. IEEE Int. Conf. Comput. Vis.*, 2017: pp. 618–626. doi:10.1109/ICCV.2017.74.

- [77] R. Kotikalapudi, keras-vis, (2017). <https://github.com/raghakot/keras-vis>.
- [78] M.T. Ribeiro, S. Singh, C. Guestrin, “Why Should I Trust You?” Explaining the Predictions of Any Classifier, (2016). doi:10.1145/2939672.2939778.
- [79] S.M. Lundberg, P.G. Allen, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Adv. Neural Inf. Process. Syst. 30, Curran Associates, Inc., 2017: pp. 4765–4774. <https://github.com/slundberg/shap> (accessed July 1, 2020).
- [80] H. Chen, J.D. Janizek, S. Lundberg, S.-I. Lee, True to the Model or True to the Data?, (2020). <http://arxiv.org/abs/2006.16234> (accessed July 6, 2020).
- [81] A. Ghorbani, J. Zou, Data shapley: Equitable valuation of data for machine learning, in: 36th Int. Conf. Mach. Learn. ICML 2019, International Machine Learning Society (IMLS), 2019: pp. 4053–4065. <https://arxiv.org/abs/1904.02868v2> (accessed July 6, 2020).
- [82] M. Wu, M. Wicker, W. Ruan, X. Huang, M. Kwiatkowska, A game-based approximate verification of deep neural networks with provable guarantees, Theor. Comput. Sci. 807 (2020) 298–329. doi:10.1016/j.tcs.2019.05.046.
- [83] A. Winkelmann, T. Ben Britton, G. Nolze, Constraints on the effective electron energy spectrum in backscatter Kikuchi diffraction, Phys. Rev. B. 99 (2019) 064115. doi:10.1103/PhysRevB.99.064115.
- [84] C. Yuanyuan, W. Zhibin, Quantitative analysis modeling of infrared spectroscopy based on ensemble convolutional neural networks, Chemom. Intell. Lab. Syst. 181 (2018) 1–10. doi:10.1016/j.chemolab.2018.08.001.
- [85] C. Blundell, J. Cornebise, K. Kavukcuoglu, D. Wierstra, Weight Uncertainty in Neural Networks, 32nd Int. Conf. Mach. Learn. ICML 2015. 2 (2015) 1613–1622. <http://arxiv.org/abs/1505.05424> (accessed July 23, 2020).
- [86] G. Hinton, O. Vinyals, J. Dean, Distilling the Knowledge in a Neural Network, (2015). <http://arxiv.org/abs/1503.02531> (accessed October 14, 2020).
- [87] F. Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, in: Proc. IEEE Int. Conf. Comput. Vis., The Computer Vision Foundation, Seattle, 2017: pp. 1251–1258. [http://openaccess.thecvf.com/content\\_cvpr\\_2017/papers/Chollet\\_Xception\\_Deep\\_Learning\\_CVPR\\_2017\\_paper.pdf](http://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf) (accessed February 1, 2019).
- [88] F. Chollet, Keras, (2015). <https://github.com/keras-team/keras>.
- [89] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, in: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., 2015: pp. 1–15. <http://arxiv.org/abs/1412.6980> (accessed October 8, 2018).
- [90] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner,

- P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, G. Brain, TensorFlow: A System for Large-Scale Machine Learning, in: Proc. 12th USENIX Symp. Oper. Syst. Des. Implement., USENIX, Savannah, Georgia, 2016: pp. 265–283. <https://tensorflow.org>. (accessed October 8, 2018).
- [91] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830. <http://www.jmlr.org/papers/v12/pedregosa11a> (accessed August 15, 2019).
- [92] J.D. Hunter, Matplotlib: A 2D graphics environment, *Comput. Sci. Eng.* 9 (2007) 99–104. doi:10.1109/MCSE.2007.55.
- [93] S. Van Der Walt, S.C. Colbert, G. Varoquaux, The NumPy array: A structure for efficient numerical computation, *Comput. Sci. Eng.* 13 (2011) 22–30. doi:10.1109/MCSE.2011.37.
- [94] L.S. Shapley, Stochastic Games, *Proc. Natl. Acad. Sci. U. S. A.* 39 (1953) 1095–1100.
- [95] L.S. Shapley, A value for n-person games, in: H.W. Kuhn, A.W. Tucker (Eds.), *Contrib. to Theory Games (AM-28)*, Vol. II, Princeton University Press, New Jersey, 1953: pp. 307–317. [https://books.google.com/books?hl=en&lr=&id=Pd3TCwAAQBAJ&oi=fnd&pg=PA307&dq=Lloyd+S+Shapley.+“A+value+for+n-person+games”.+In:+Contributions+to+the+Theory+of+Games+2.28+\(1953\),+pp.+307–317.&ots=gswVJ9afv\\_&sig=JNbFjQIT72s2sdfisW\\_5FpqmPGc#v=onepage&q&f=false](https://books.google.com/books?hl=en&lr=&id=Pd3TCwAAQBAJ&oi=fnd&pg=PA307&dq=Lloyd+S+Shapley.+“A+value+for+n-person+games”.+In:+Contributions+to+the+Theory+of+Games+2.28+(1953),+pp.+307–317.&ots=gswVJ9afv_&sig=JNbFjQIT72s2sdfisW_5FpqmPGc#v=onepage&q&f=false) (accessed July 1, 2020).
- [96] S.M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, (2017) 4765–4774. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions> (accessed September 15, 2018).
- [97] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE.* 86 (1998) 2278–2323. doi:10.1109/5.726791.
- [98] krkaufma, krkaufma/Electron-Diffraction-CNN v1.0.1, (2019). doi:10.5281/ZENODO.3564937.

## **Chapter 9 An Acquisition Parameter Study for Machine-Learning-Enabled Electron**

### **Backscatter Diffraction**

Kevin Kaufmann and Kenneth S. Vecchio\*

Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA

\* Corresponding Author

#### **Abstract**

Methods within the domain of artificial intelligence are gaining traction for solving a range of materials science objectives, notably in the field of computer vision. A rapidly growing application of deep neural networks for computer vision is the analysis of electron diffraction patterns. An important component of deploying these models is an understanding of the performance as experimental diffraction conditions are varied. This knowledge can inspire confidence in the classifications over a range of operating conditions and identify where performance is degraded. Elucidating the relative impact of each parameter will suggest the most important parameters to vary during the collection of future training data. Knowing which data collection efforts to prioritize is of concern given the time required to collect or simulate vast libraries of diffraction patterns for a wide variety of materials without considering varying any parameters. In this work, five parameters essential to electron diffraction are individually varied during collection of electron backscatter diffraction patterns to explore the effect on the classifications produced by a deep neural network trained from diffraction patterns captured using a fixed set of parameters. The five parameters studied are frame averaging, detector tilt, sample-

to-detector distance, accelerating voltage, and pattern resolution. Ultimately, the model is found to be resilient to nearly all the individual changes studied in this work.

## 9.1 Introduction

The fields of materials development and analysis have recently begun to explore the possibility of applying data-driven strategies and artificial intelligence (AI) for accelerating or automating a variety of tasks [1,2,11–15,3–10]. Computer vision is a subset of AI with the goal of training computers to understand the visual world and potentially act on that information [16,17]. Deep learning algorithms enable many computer vision applications and are of particular interest owing to their excellent performance without significant feature engineering [18]. While it can be difficult to precisely determine how and why these ‘black-box algorithms’ are capable of performing these tasks, these methods can automate routine tasks, improve upon existing solutions, or provide understanding [19,20]. While deep learning provides significant opportunities for the advancement of materials science, robust application of these tools requires an understanding of the conditions under which optimal performance is achieved. Unlike most cases involving natural images [21], scientific images are less likely to be collected utilizing a wide range of parameters. For instance, micrographs may be captured with only 1 or 2 magnifications, while photographs of animals at different distances and from different angles are typically abundant [21]. Moreover, knowledge of the conditions for which a model fails can guide the future collection of labeled training data to improve subsequent versions. However, in the context of electron diffraction, there likely exists some cases (e.g. no frame averaging) where reduced quality or lack of sufficient information in the images may limit performance regardless of training data available for a given technique.

Convolutional neural networks (CNNs) are the current standard deep learning models for processing image data [18]. Before a CNN can be applied to a given task, it must learn to assign importance to various aspects of the image that maximize the network's differentiation capabilities. In doing so, deep learning methods glean intricate functions of the inputs that are sensitive to minute details, yet ignore irrelevant information such as backgrounds [18]. However, it should be noted that it is only with careful design of a training set and rigorous validation that practitioners can be confident that the model has truly learned relevant information, is robust to new conditions, and has not found an unscientific approach to solving the problem (such as learning the presence of a ruler means a lesion is more likely cancerous [22]) [23,24]. Application of these tools to image-based tasks in materials science has proved to be useful for classification [4,11,25,26], segmentation [3,14], and other objectives [27,28]. Examples of techniques where interest in developing artificial intelligence agents for image-based tasks include optical microscopy [3,29], STEM [30,31], TEM [32], and EBSD [4,33–36]. These efforts are motivated by accelerating data generation rates and the traditional need for tedious or arduous analysis of the data by well-trained individuals with sufficient knowledge of the material domain. Thus, it is important that these tools can be applied robustly as imaging parameters are varied to prevent increasing, instead of alleviating, researcher workloads. While different diffraction techniques often use distinct terminology, parameters such as accelerating voltage or detector distance are common among them and have similar effects on the collected diffraction patterns. Several of these parameters play a role in the amount of time required to collect each diffraction pattern and therefore complete the analysis of a sample. In this work, the electron backscatter diffraction (EBSD) technique is used owing to the relatively high rate of data collection and ease of changing each parameter.

Electron backscatter diffraction is a scanning electron microscope (SEM)-based method that involves the capture of 2D diffraction patterns produced from an incident electron beam scattering, diffracting, and escaping from a well-polished ‘bulk’ sample [37]. Despite the vast amount of information in the patterns [37–39], conventional EBSD has primarily focused on determining three-dimensional orientation [17,37,40,41]. Furthermore, the technique typically relies on a user-defined phase list and Hough-based indexing [42]. Hough-based indexing generally allows for phase differentiation of sufficiently distinct crystal structures [43–45], but the process remains susceptible to structural misclassification [46–48]. Improvements to phase differentiation have been proposed and developed including dictionary indexing [49–52] and spherical indexing [53–55], although each still requires a user to pre-select phases and further requires simulating the Kikuchi sphere for each selected phase. Recently, the EBSD community has begun to investigate the use of convolutional neural networks for indexing, phase differentiation, and determining components of crystal structure [4,25,33–35]. It is a goal of several of these efforts that the onus of phase selection and/or structure determination can be at least partially alleviated from the user [4,33]. However, to date there has not been a systematic study of CNN performance when the EBSD patterns (EBSPs) are collected using different experimental geometry than was used during collection or simulation of the training data. Knowledge of how these changes to the diffraction patterns influence proper pattern identification is paramount to widespread adoption of these machine learning-based techniques.

This work seeks to develop an understanding of model performance as several of the most common EBSD operating conditions are varied. The specific parameters are frame averaging, detector tilt, sample-to-detector distance, accelerating voltage, and pattern resolution. With regard to parameters that directly affect the time to collect each pattern and therefore complete a map,

such as frame averaging and pattern resolution, the ability to collect the data more rapidly without a significant reduction in performance is of interest. With respect to parameters such as detector tilt, it is important to determine if the model is susceptible to minor or major changes from the training conditions. Parameters such as detector distance and accelerating voltage can cause much more dramatic changes to the EBSPs, and it is therefore necessary to assess their influence. The CNN model tested in this work was trained to classify EBSPs to one of six space groups using patterns collected from a fixed EBSD setup [33]. The effect of changing these parameters is tested using new EBSD patterns collected from one material from each space group and a dual-phase 2205 duplex steel for visual demonstration. Each time one parameter is varied, the EBSPs are re-collected, and the CNN used to reassess the proper space group identification. Ultimately, the model is found to retain a high classification accuracy even with significant changes to the diffraction conditions and therefore the EBSPs.

## 9.2 Materials and methods

### 9.2.1 Materials

Eighteen different single-phase materials, comprising 6 of the 10 space groups within the  $(4/m\bar{3}2/m)$  point group were selected for training the space group classification CNN. The space groups are  $Pm\bar{3}m$  (221),  $Pm\bar{3}n$  (223),  $Fm\bar{3}m$  (225),  $Fd\bar{3}m$  (227),  $Im\bar{3}m$  (229), and  $Ia\bar{3}d$  (230). Suitable samples for the remaining 4 space groups could not be obtained. The materials were [221: FeAl, NiAl, Ni<sub>3</sub>Al, Fe<sub>3</sub>Ni], [223: Cr<sub>3</sub>Si, Mo<sub>3</sub>Si], [225: Ni, Al, NbC, TaC, TiC], [227: Si, Ge], [229: W, Ta, Fe], and [230: Al<sub>4</sub>CoNi<sub>2</sub>, and Al<sub>4</sub>Ni<sub>3</sub>]. The collected diffraction patterns from these materials were of low texture, typically less than 2 times random in any direction. Refer to Kaufmann *et al.* for the distributions of orientation, band contrast, and mean angular deviation for these samples [33].

A dual-phase material with phases that are easily differentiable by Hough-based EBSD was used to visually demonstrate and compare the classification accuracy of the model as parameters vary. A longitudinal cross-section of cold worked 2205 duplex steel was mounted, ground, and polished to  $0.05\mu\text{m}$  using colloidal silica. Looking at the longitudinal cross-section, cold worked 2205 duplex stainless steel has a ferrite matrix (space group 229) and elongated austenite islands (space group 225) [56].

### **9.2.2 Electron backscatter diffraction pattern collection**

EBSD patterns were collected in a Thermo Scientific (formerly FEI) Apreo scanning electron microscope (SEM) equipped with an Oxford Symmetry EBSD detector.

All EBSPs utilized in training were collected with the following fixed geometry. The EBSD detector was utilized in high resolution ( $1244 \times 1024$ ) mode, the working distance was  $18.1 \text{ mm} \pm 0.1 \text{ mm}$ , the sample-to-detector distance was  $19.1 \text{ mm}$ , and the detector tilt to  $13.7$  degrees above horizontal. The imaging parameters for the training set EBSPs were  $20\text{kV}$  accelerating voltage,  $51\text{nA}$  beam current, and 30-pattern averaging. This fixed geometry will be referred to as the “default” operating conditions and diffraction geometry. Refer to Figure 1 for an annotated image of the diffraction setup in the SEM.

Diffraction patterns from six single-phase materials ( $\text{Ni}_3\text{Al}$ ,  $\text{Cr}_3\text{Si}$ ,  $\text{TiC}$ ,  $\text{Si}$ ,  $\text{Fe}$ , and  $\text{Al}_4\text{CoNi}_2$ ) and the dual-phase 2205 duplex steel were collected separately from the training data to determine a baseline accuracy for the model when using the default diffraction geometry. Approximately 3,000 individual patterns were collected from a large area of each sample (i.e. low magnification) to capture as many unique orientations as possible over the fixed region. After collecting data using the default diffraction geometry, each one of the parameters was

systematically varied one at a time and the same ~3,000 patterns re-collected from the same location.

The parameters were varied as follows. Frame averaging was set to 1, 5, 10, 20, or 30. The software detector tilt below zero ranged from 1 to 5 in steps of 1. These values correspond to the detector being 14.2, 14.0, 13.7, 13.5, and 13.3 degrees above the horizontal plane. The software detector insertion distance ranged from 156 mm to 164 mm in steps of 2mm. These values correspond to sample-to-detector distances of 24.3, 21.8, 19.1, 16.8, and 14.3 mm. The sample to detector distances were calculated following the methods outlined in [57]. Accelerating voltage options were 10, 20, or 30 kV. The pattern resolution options were 156×128 (low), 622×512 (medium), or 1244×1024 (high). The working distance was held constant, since moving the sample up or down while at 70 degrees sample tilt would change the location on the sample. The beam current also remained fixed and the exposure time adjusted accordingly by the Oxford Aztec software to offset the increase or decrease in signal resulting from a varied parameter (e.g. detector distance). Supplementary Table 20 summarizes the EBSD pattern acquisition rate for each of the varied parameters compared to the “default” conditions.

### **9.2.3 Convolutional neural network**

The Xception convolutional neural network architecture [58] was selected for fitting the model. Selection of this network was based on Xception or derivatives of Xception being used previously in the EBSD community [4,33–35]. Refer to Figure 5 in Xception: Deep Learning with Depthwise Separable Convolutions [58] for a complete description of the Xception architecture. Further details about the training process can be found in Kaufmann *et al.* [33]. The CNNs were implemented with TensorFlow[59] and Keras [60].

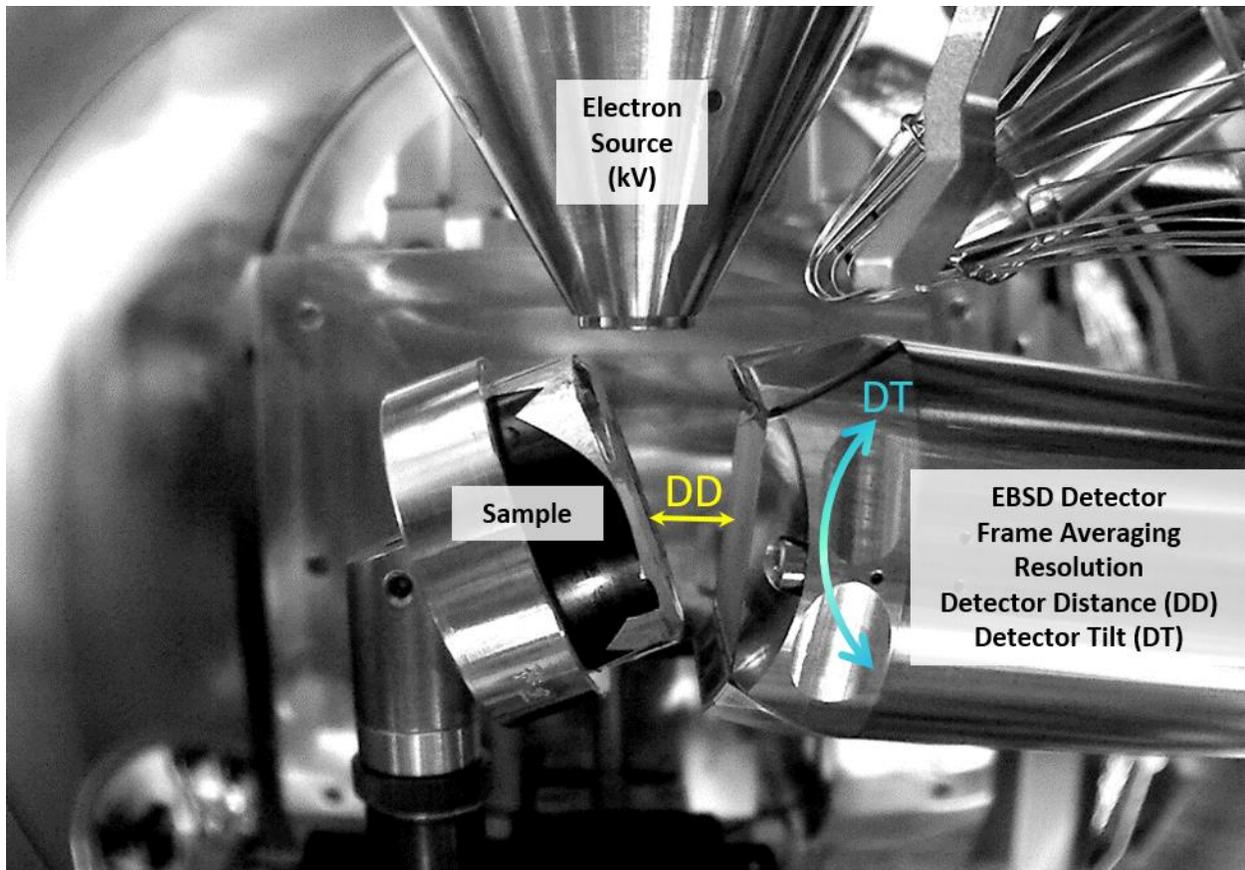
#### **9.2.4 Diffraction pattern classification**

During pattern collection, Hough-based indexing was performed with only three options: ferrite, austenite, or non-indexed. After collecting high resolution EBSPs from each material, all patterns collected were exported as tiff images. Diffraction patterns were evaluated in a random order by the trained CNN model without further information to assess the model as it would be applied in practice. The output classification of each diffraction pattern was recorded and saved in a (.csv) file. These csv files were utilized to calculate the normalized accuracy of the model for each trial. In the case of 2205 duplex steel, the predictions were converted to a space group map using the plotting tools in MATLAB R2018B following the methods established in Kaufmann *et al.* [34].

### **9.3 Results**

#### **9.3.1 Equipment setup and Hough-based results**

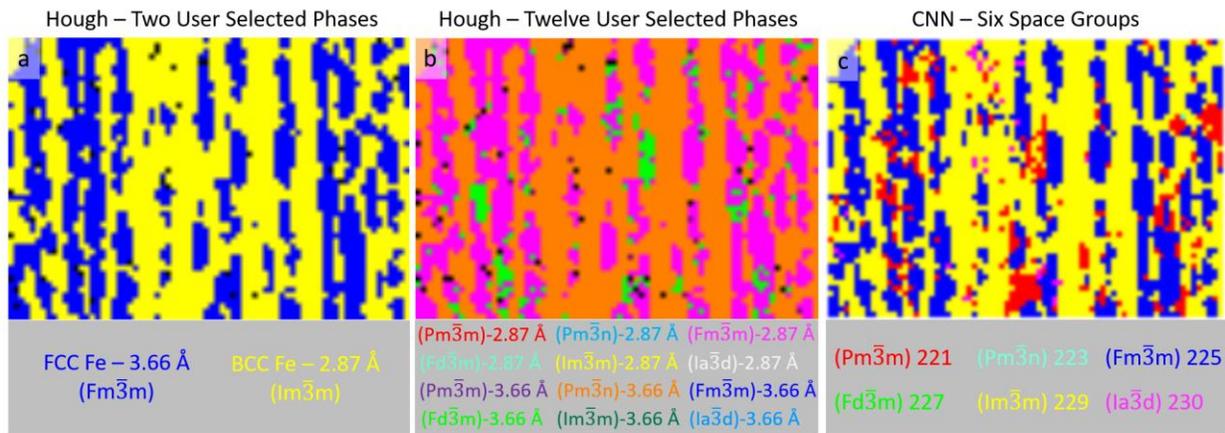
The choice of operating parameters for EBSD are not fixed but are instead valid over a range of values depending on the manufacturer's calibration. The specific parameters defined as "default" parameters in this work are those from which all EBSPs in the training set were collected (refer to section 2.2 or prior work by Kaufmann *et al.* [33]). The equipment setup for EBSD is shown and labeled in Figure 1. The detector distance from the sample and detector tilt are further detailed using arrows that describe their geometric role. EBSPs are collected from a fixed region of the specimen each time one parameter is varied.



**Figure 9.1 EBSD setup and variable parameters.** An annotated view of the EBSD setup. Parameters that vary in this study are listed below the equipment label. The sample-to-detector distance (DD) and detector tilt (DT) are further detailed by arrows describing their function.

Figure 2 highlights the need for a reliable tool to assist an operator with symmetry determination using a section of 2205 duplex steel. In the case where the phases are known in advance (Fig. 2a), Hough-based EBSD is shown to produce a high-fidelity phase map of the austenite islands (blue) in the ferrite matrix (yellow). However, in cases where the phases are unknown, there exists ample opportunity for misclassification in the Hough-based method even when using what could be considered ideal acquisition parameters. If the six space groups used in the CNN study were selected in duplicate, one copy with lattice parameters matching the austenite's lattice parameters and symmetry and the other copy matching the ferrite's, the results can be strikingly different (Fig. 2b). In fact, not a single pattern is indexed to either of the two

correct phases. In fact, two of the three space groups selected,  $Pm\bar{3}n$  (orange) and  $Fd\bar{3}m$  (lime green), are not present in the sample. Thus, a robust classifier of EBSD patterns would be a useful tool for assisting with selecting the proper phases or alerting users to potential errors in selected phase lists. Figure 2c demonstrates how a CNN-based classifier could be used for such a purpose. The CNN has predicted the space group of each high-quality EBSD pattern that Hough-based EBSD used for generating Figures 2a,b, yet the CNN-derived phase map is a much higher fidelity mapping of Figure 2a than is Figure 2b, albeit without the lattice parameter information at this stage of the CNN model. These results are significant since Figure 2b demonstrates a marked failure wherein Hough-based EBSD cannot distinguish between the two correct answers plus ten phases with the same lattice parameters but different space group symmetry. Given the potential the CNN-based method has for improving the EBSD process, this work sets out to study how reliable the CNN's classifications are as new data captured under different diffraction conditions is presented to it.

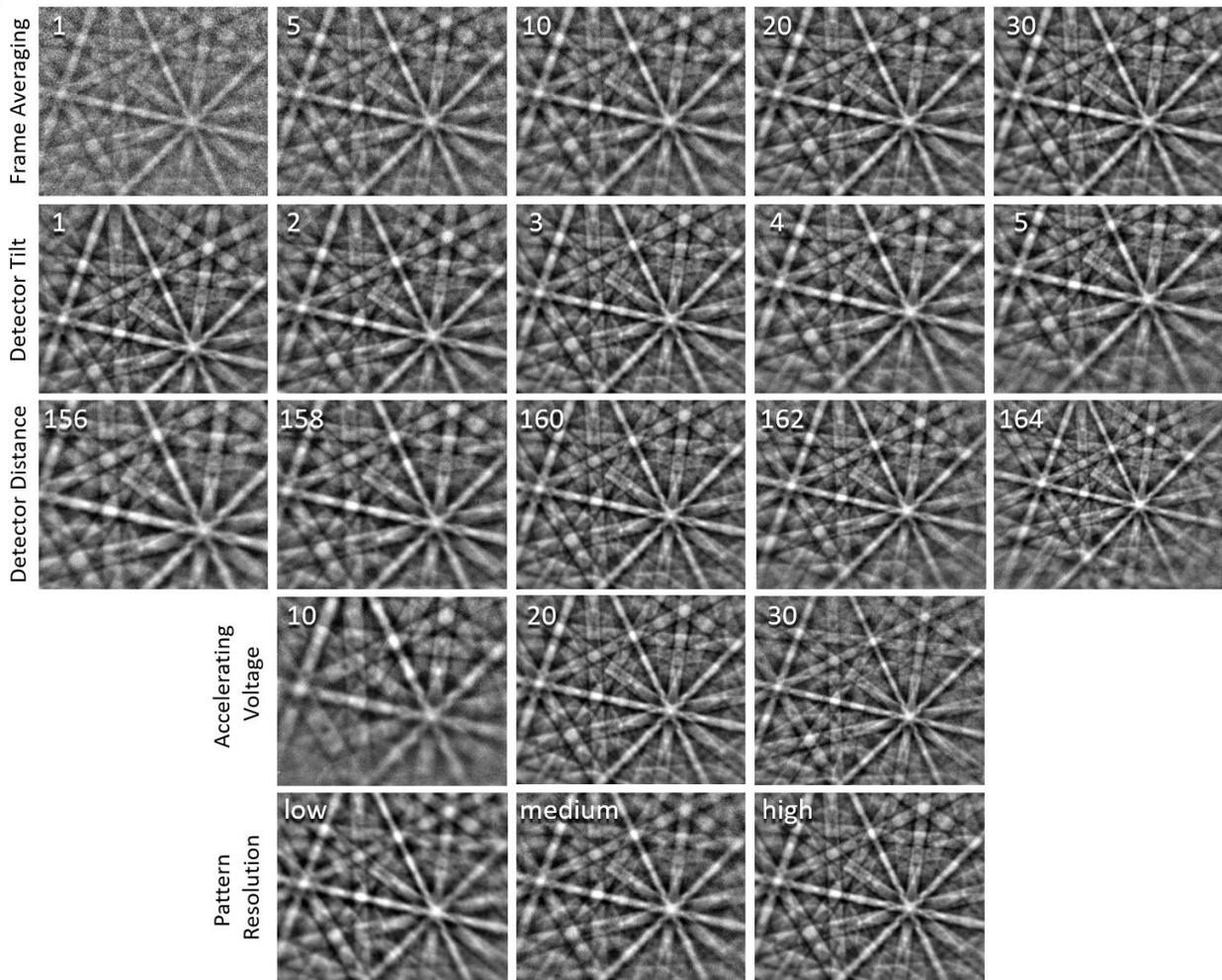


**Figure 9.2 Hough-based EBSD results with different phase lists.** A comparison of the Hough-based EBSD method is performed where (a) only the two correct phases are provided as options and (b) twelve options are made available based on the six space groups used in this work and the two lattice parameters found in the duplex steel sample. The CNN-derived space group map (c) details the classifier's predictions for the same EBSD patterns used in parts (a) and (b). The data for each map is collected using the ideal parameters defined as “default” in this work.

### 9.3.2 Effects on the EBSD patterns

It is important to understand the effect of varying the operating conditions of the SEM and EBSD detector on the electron diffraction patterns. Figure 3 visually details the effects by displaying the same diffraction pattern observed under different operating conditions. Larger versions of these images are contained in Supplementary Figures 48-52 in the Supplementary Materials. Increasing the number of frames averaged for each diffraction pattern increases the signal to noise ratio and results in better resolution of finer details in the diffraction patterns. Changing the detector tilt changes the relative position of the image with respect to the interaction volume of the sample, referred to as the pattern center in EBSD [37,55,61]. For tilt angles that are far from the ideal conditions, the top or bottom edge of the EBSP may display blurring. The most significant blurring is observed when using a tilt of 13.3 degrees. Otherwise, the changes are observed to be limited to small differences in the region of the Kikuchi sphere captured. Decreasing the sample-to-detector distance, and thus moving closer to the sample, results in the capture of significantly more solid angle. Since the EBSD detector is capturing a gnomonic projection, the increase in the solid angle means a greater area of the Kikuchi sphere is observed. This likely must be balanced with the ability to resolve the finer details and eventual blurring of the pattern edges. On the other hand, moving the detector further away captures less of the Kikuchi sphere but “magnifies” the finer details. The diffraction elements also appear slightly blurred at the furthest distance away from the sample. Changes to the SEM accelerating voltage affect the EBSPs by altering the wavelength of the incoming electrons. A decrease in the accelerating voltage increases the electron wavelength and therefore causes the Kikuchi bands to appear wider and vice versa. This results in more diffraction information from the same region of the Kikuchi sphere condensed within the viewing window for higher accelerating voltages. Note that it does

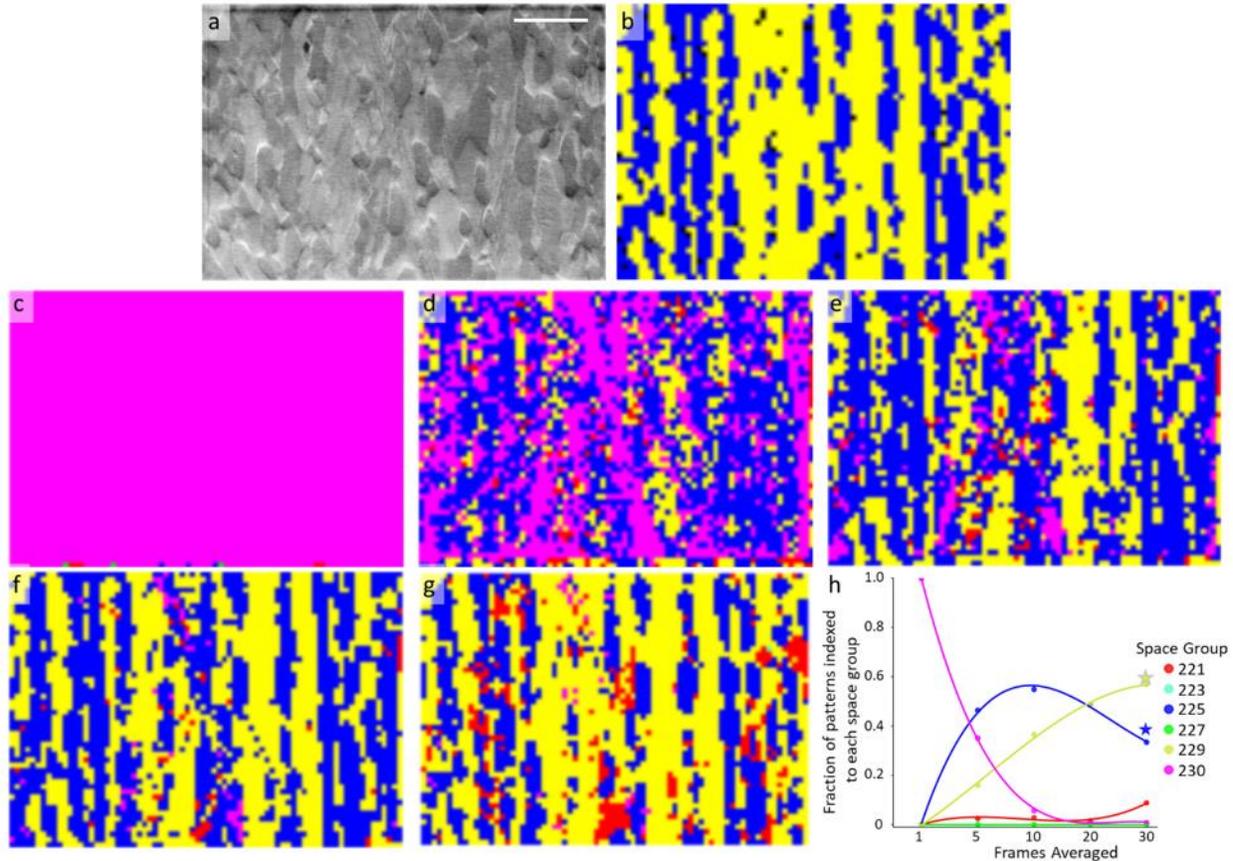
not have the same effect as changing the detector distance. Lastly, the Oxford Symmetry detector can capture patterns at one of three resolutions:  $156 \times 128$ ,  $622 \times 512$ , or  $1244 \times 1024$ . These are described as low, medium, and high resolution, respectively, in this work. For each of the different imaging conditions shown, approximately 3,000 diffraction patterns are collected from the same region of each sample and the CNNs performance analyzed.



**Figure 9.3 Impact of operating conditions on the EBSD patterns.** The diffraction pattern for a point on the sample is displayed for different imaging conditions. The effect of number of frames averaged, the tilt of the EBSD detector, the sample-to-detector distance for the EBSD detector, the SEM accelerating voltage, and the resolution at which patterns are captured are each visually described.

### 9.3.3 Frame averaging

Figure 4 presents a visual overview of the pattern classification results for each frame averaging condition studied using a sample of 2205 duplex stainless steel, a material with a ferrite matrix (space group 229; yellow) and austenite islands (space group 225; blue) [56]. Similar space group maps for each of the parameters studied are shown in the Supplementary Materials (Supplementary Figures 53-57). The electron image (Fig. 4a) and Hough-based phase map (Fig. 4b) are provided as the ground truth, since these phases can be differentiated by Hough with relative ease, assuming the operator selects the correct phases at the start (i.e., user knows the phases). Without any frame averaging, the EBSPs are lacking much of the available details and essentially all the EBSPs from this sample are misidentified (Fig. 4c). Increasing the frame averaging to 5 improves the result (Fig. 4d), but it is not until 10 frames are averaged that a reasonable number of classifications appear to be correct (Fig. 4e) when compared to the Hough-based map (Fig. 4b). Figures 4f,g each show sequential improvements over Figure 4e as nearly all of the patterns become correctly classified to the correct space group (i.e. matching the Hough-based phase map). As seen in the plot (Fig. 4h), the number of patterns classified to each space group begins to improve with 10 frames averaged. While the plot shows the relative number of patterns classified to each space group compared to Hough-based EBSD, the space group maps made from the CNN's predictions establish whether the individual EBSPs are correctly classified based on their correlation with the Hough-based phase map. Since each of the other materials studied in this work is known to be single phase, normalized classification accuracy precisely describes the model performance.

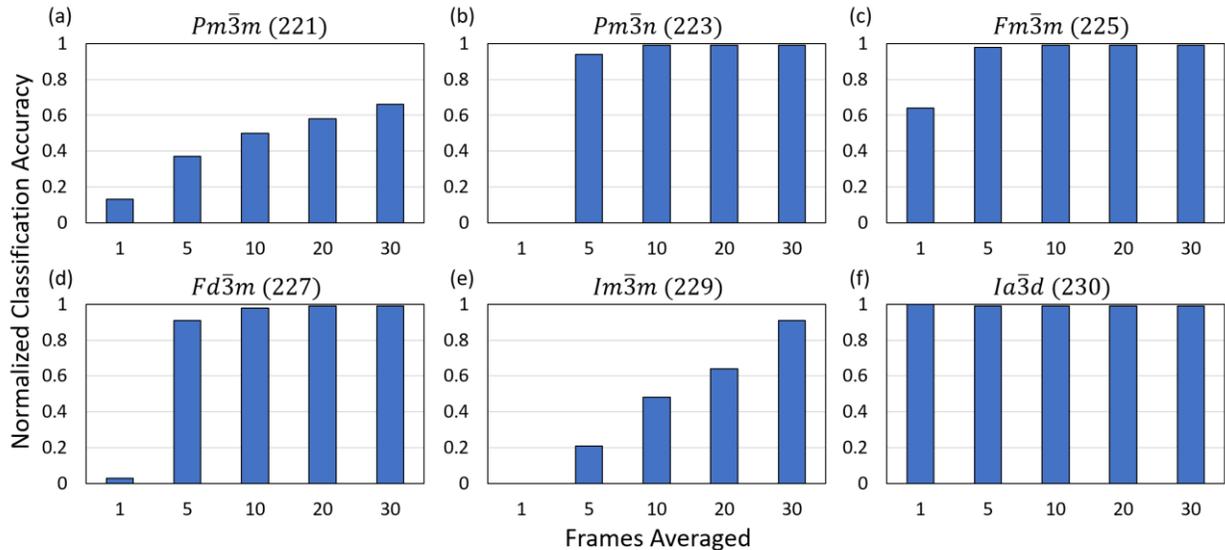


**Figure 9.4 Visual overview of frame averaging on CNN classification for duplex steel.** (a) electron image of the region of dual-phase 2205 duplex steel. (b) Hough-based EBSD phase map of the fcc (225) austenite (blue) and bcc (229) ferrite (yellow). (c) phase map generated from EBSD patterns collected with no frame averaging applied (i.e. one frame). (d) phase map generated from EBSD patterns collected with five frame averaging applied. (e) phase map generated from EBSD patterns collected with ten frame averaging applied. (f) phase map generated from EBSD patterns collected with twenty frame averaging applied. (g) phase map generated from EBSD patterns collected with thirty frame averaging applied. (h) Plot showing the fraction of patterns indexed to each space group as a function of frame averaging. Thirty frame averaging is the default parameter and is designated as such by the blue star for space group 225 and a yellow star for space group 229. Trend lines are fit with a 3<sup>rd</sup> order polynomial. Scale bar 25 $\mu$ m. There are 3,848 diffraction patterns (pixels) in each phase map.

It is worth pointing out that the Hough-based result is considered the ground truth (or correct answer) because the phases in the sample are known, and the known answer is provided as input to the Hough-based solution. So in this context, the EBSD solution for phase ID must already be known for the Hough-based approach. The power of the machine learning approach being

implemented here, and previously presented in Kaufmann *et al.* [32], is to perform symmetry identification, to the space group level, for samples for which the phases present are not known a priori (i.e. enable true phase ID by EBSD).

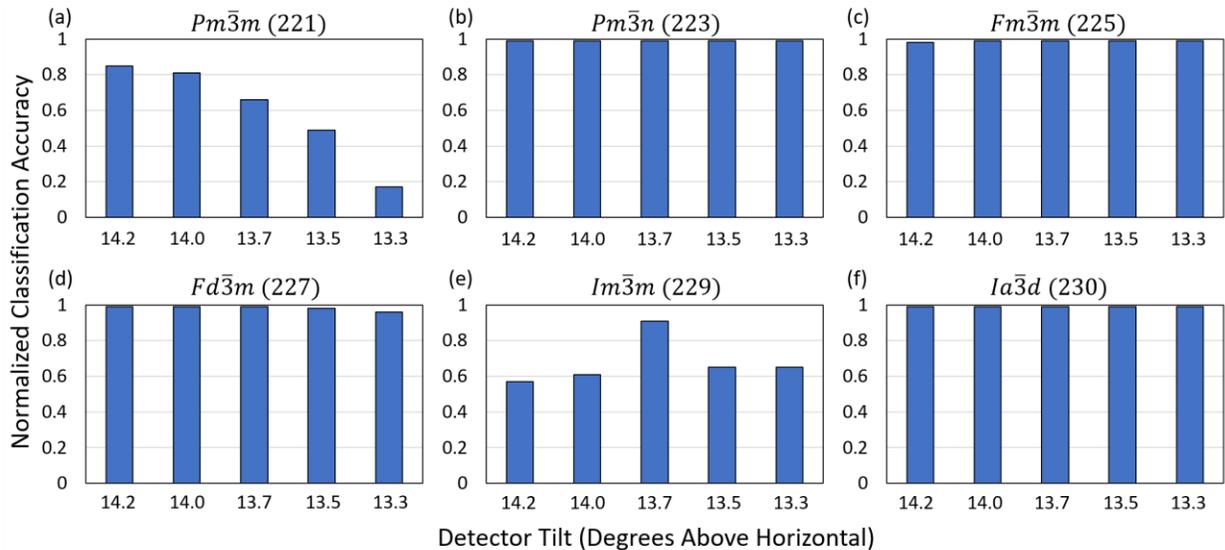
Figure 5 details the normalized accuracy of the CNN for each space group as the number of frames averaged is varied in subsequent collections of the same EBSD patterns. The default frame averaging is 30 patterns. It is observed that a high overall classification accuracy, compared to the accuracy obtained for the default parameter, is generally retained as low as 5-10 frames averaged. An exception to this is observed for space groups 221 and 229 which, along with space group 225, are the most difficult for the CNN to differentiate owing to the strong similarities between the fcc and L1<sub>2</sub> structures and bcc and B2 structures used in training the model [33].



**Figure 9.5 Effect of frame averaging on classification accuracy.** The normalized classification accuracy of the trained CNN for each space group based on the number of patterns averaged during data collection. The space groups are (a)  $Pm\bar{3}m$  (b)  $Pm\bar{3}n$  (c)  $Fm\bar{3}m$  (d)  $Fd\bar{3}m$  (e)  $Im\bar{3}m$  and (f)  $Ia\bar{3}d$ . The default number of frames averaged is 30.

### 9.3.4 Detector tilt

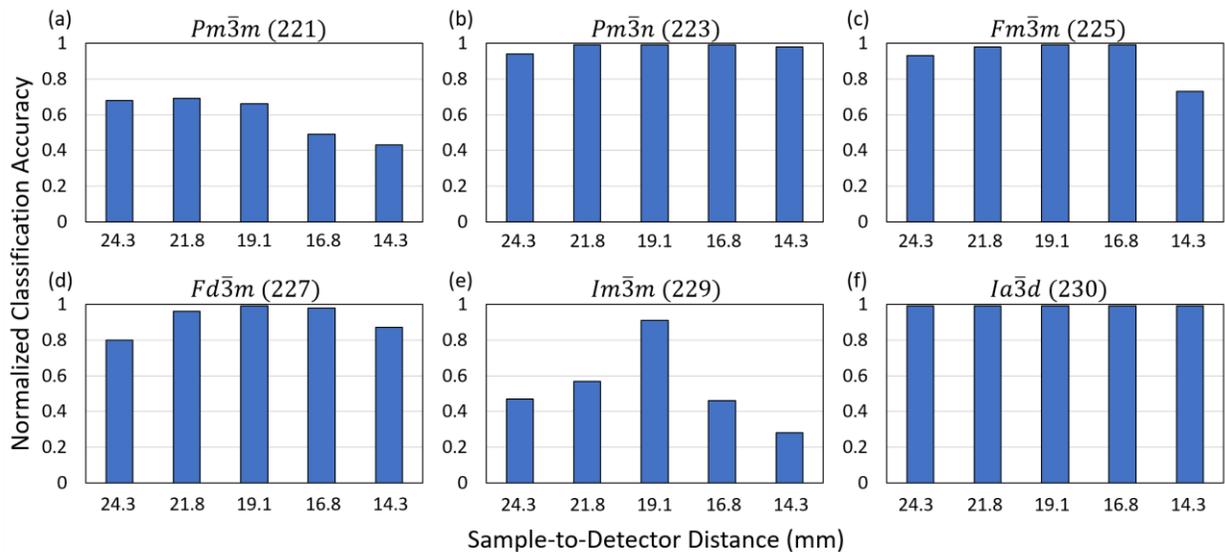
Figure 6 details the effect of the detector tilt on the CNN's classification of the collected EBSPs. Note that the detector tilt is reported as the angle of the detector above horizontal. Referring to the EBSPs in Fig. 3 and their enlarged counterparts shown in Supplementary Figure 49 in the Supplementary Materials, detector tilts of 14.2 to 13.7 degrees show ideal patterns with little or no blurring at the edges. The blurring becomes more apparent at the bottom of the EBSPs when a tilt of 13.5 degrees is applied and increases in severity at a tilt of 13.3 degrees. For most space groups in Figure 6, the model is found to be highly resilient, and it is only at or below a tilt of 13.5 degrees that any notable changes in accuracy occur. Again, space groups 221 and 229 are the exception to this observation. Referring to the symmetry maps of the duplex steel, the fractions do not vary significantly across the individual maps (Supplementary Figure 54) and the space group classification of each EBSP aligns well with the Hough-based phase map (ground truth).



**Figure 9.6 Effect of detector tilt on classification accuracy.** The normalized classification accuracy of the trained CNN for each space group based on the tilt of the detector during data collection. The space groups are (a)  $Pm\bar{3}m$  (b)  $Pm\bar{3}n$  (c)  $Fm\bar{3}m$  (d)  $Pd\bar{3}m$  (e)  $Im\bar{3}m$  and (f)  $Ia\bar{3}d$ . The detector tilts are reported as the detector's angle above horizontal. The default value for detector tilt is 13.7 degrees.

### 9.3.5 Detector Distance

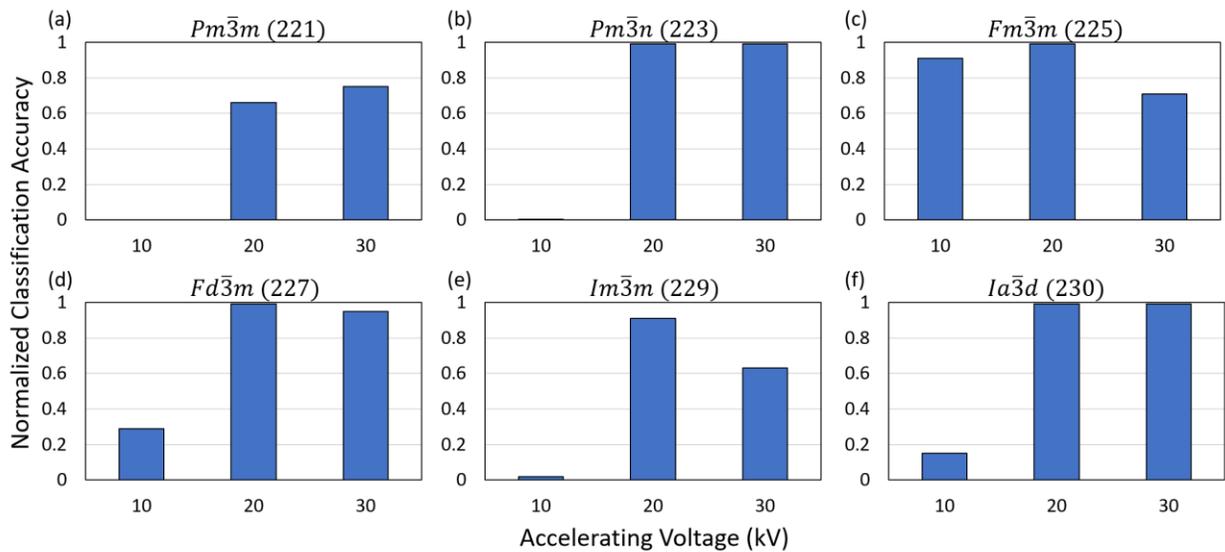
Figure 7 presents the changes in the CNN's performance as the EBSD detector collects the diffraction patterns at different sample-to-detector distances. The number of mis-classified pixels, however minute the difference is, is observed to scale with increased absolute distance from the default condition of 19.1 mm. When the detector is much further away from the sample, less solid angle is captured, and the diffraction patterns become distorted (Fig. 3 and Supplementary Figure 50). Moving closer to the sample increases the solid angle, but at the expense of the finer details. This likely contributes to the noticeably reduced performance for the closest possible setting (a sample-to-detector distance of 14.3 mm). Referring to the phase maps of the duplex steel (Supplementary Figure 55), similar effects are observed. As the sample-to-detector distance gets further from the default condition, the CNN-derived maps increasingly differ from the Hough-based phase map.



**Figure 9.7 Effect of sample-to-detector distance on classification accuracy.** The normalized classification accuracy of the trained CNN for each space group based on the sample-to-detector distance during data collection. The space groups are (a)  $Pm\bar{3}m$  (b)  $Pm\bar{3}n$  (c)  $Fm\bar{3}m$  (d)  $Pd\bar{3}m$  (e)  $Im\bar{3}m$  and (f)  $Ia\bar{3}d$ . The default value for sample-to-detector distance is 19.1 mm.

### 9.3.6 Accelerating voltage

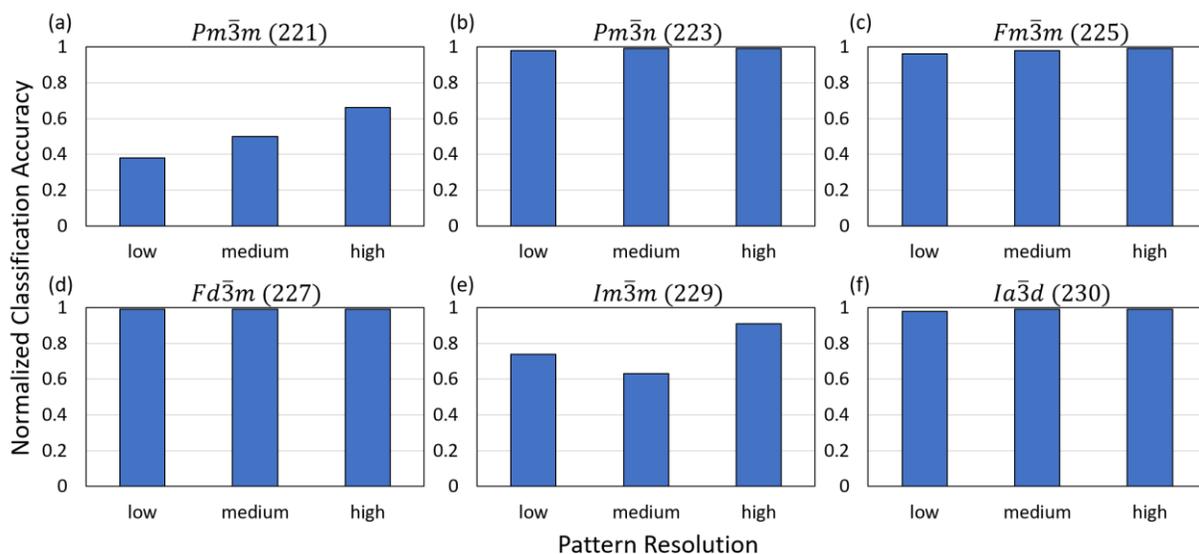
The effect of changing the wavelength of the incoming electrons by modifying the accelerating voltage is explored in Figure 8 and Figure A9. The primary effect of changing the wavelength of the incoming electrons is a change in the width of the observed Kikuchi bands (Fig. 3 and Supplementary Figure 51). The resulting effect on the collected diffraction patterns is similar to changing the detector distance, but notably does not alter the amount of solid angle captured on the phosphor screen. Instead, changing the accelerating voltage effectively changes the magnification of the diffraction data within the same screen area. For example, decreasing to 10kV causes the details of the diffraction pattern to increase in size, effectively making the data appear expanded. Note how the same zone axes are present for each accelerating voltage, but the distance between the zone axes and the width of the diffraction lines appears to change (Supplementary Figure 51). These changes are observed to have appreciable effects on the classification performance of the CNN. At 10kV, many of the patterns from each space group are misclassified (Fig. 8). On the other hand, increasing to 30kV accelerating voltage yields reasonably good classification performance, perhaps because the Kikuchi bands are narrower and more information appears to be visible. The same effects on performance are visually evident in the CNN-derived space group maps of the duplex steel (Supplementary Figure 56).



**Figure 9.8 Effect of accelerating voltage on CNN classification.** The normalized classification accuracy of the trained CNN for each space group based on the SEM accelerating voltage. The space groups are (a)  $Pm\bar{3}m$  (b)  $Pm\bar{3}n$  (c)  $Fm\bar{3}m$  (d)  $Fd\bar{3}m$  (e)  $Im\bar{3}m$  and (f)  $Ia\bar{3}d$ . An accelerating voltage of 20kV is the default.

### 9.3.7 Pattern resolution

Reducing the initial resolution of the collected diffraction patterns bins the information from neighboring pixels to accelerate the rate of collection. Figure 9 details the CNN's performance after collecting the diffraction patterns at each of the three available resolutions for the Oxford Symmetry EBSD detector. The CNN is capable of achieving a high degree of accuracy even when patterns are collected at the lowest resolution setting. Only space groups 221 and 229 are appreciably impacted owing to the strong similarities between the fcc and  $L1_2$  structures and bcc and B2 structures used in training the CNN. Supplementary Figure 57 visually demonstrates the reliable performance of the CNN at each pattern resolution by creating structure maps for the 2205 duplex steel.



**Figure 9.9 Effect of pattern resolution on CNN classification.** The normalized classification accuracy of the trained CNN for each space group based on the EBSD detector resolution. Available resolutions are  $156\times 128$  (low),  $622\times 512$  (medium), and  $1244\times 1024$  (high). The space groups are (a)  $Pm\bar{3}m$  (b)  $Pm\bar{3}n$  (c)  $Fm\bar{3}m$  (d)  $Fd\bar{3}m$  (e)  $Im\bar{3}m$  and (f)  $Ia\bar{3}d$ . High resolution ( $1244\times 1024$ ) is the default setting.

## 9.4 Discussion

The electron diffraction community has recently begun to consider artificial intelligence a necessary component of next-generation microscopy [32]. Much of this is driven by the rate at which modern microscopes can generate high quality data [62]; as a result, human knowledge and experience will no longer be an efficient means of analysis. If AI-based tools are to be implemented effectively, it is necessary for the community to identify potential areas of fragility, e.g. changing diffraction conditions, and assess the impact in order to further development and increase trust in these ‘black-box’ models.

Electron backscatter diffraction was selected as an ideal technique owing to the ease of varying each parameter and the rates of data collection achievable [37,62]. Furthermore, several recent studies have explored the use of CNNs applied to components of EBSD analysis [33–36,63];

however, these efforts have thus far only been tested with new data collected or simulated using identical geometry and diffraction conditions. Herein, a parametric analysis of five parameters commonly found among electron diffraction techniques was performed to examine the performance of a convolutional neural network trained to classify diffraction patterns. One material was selected per space group and new diffraction patterns were collected starting with diffraction conditions matching the training set and then the same patterns were recollected sixteen more times after setting just one parameter to a value different from the default conditions. The same analysis was performed for a sample of 2205 duplex steel to demonstrate CNN-based symmetry mapping [34] a sample with each varied parameter. If the trained convolutional neural network is highly sensitive to the diffraction conditions, we would have expected to see large decreases in performance with the smallest changes. For example, by changing the sample-to-detector distance, the Kikuchi bands from the same material can appear wider or narrower and the distance between diffraction information (e.g. zone axes) appears to change. However, the CNN's classification accuracy is observed to be quite stable (i.e. small reductions in classification accuracy) in comparison to the results achieved with the default conditions, suggesting the features detectors learned by the model are not biased to these characteristics. This was one of the intended goals of using multiple materials with different z-contrast and lattice parameters for the same space group in the training set [33] and this study indicates its effectiveness. Moreover, the CNN is also observed to be highly dependable after decreasing the signal to noise ratio of the captured diffraction pattern by reducing the frame averaging. Decreasing the number of frames averaged will allow for faster data collection, as much as 6 times faster if averaging five frames compared to thirty, while maintaining a high degree of classification accuracy for most materials. In each of the studies within this work, the space groups most likely to be misclassified by the model were

221 ( $Pm\bar{3}m$ ) and 229 ( $Im\bar{3}m$ ). As previously mentioned, the misclassification of these patterns can be at least partially attributed to the strong similarities between diffraction patterns from the fcc ( $Fm\bar{3}m$ ) and  $L1_2$  ( $Pm\bar{3}m$ ) structures and bcc ( $Im\bar{3}m$ ) and B2 ( $Pm\bar{3}m$ ) structures used in training the CNN [33]. Inclusion of more diverse data for these space groups may help alleviate this concern in addition to being a practical advancement toward commercial adoption. While the model's dependability and trustworthiness with respect to equipment parameters has now been evaluated for phase differentiation and identification problems, it is still important to test this hypothesis for EBSD orientation indexing CNNs [35,36,63]. A much larger study should also be performed to investigate the effects of co-varying operating parameters; however, we expect that similar tolerances will be observed based on the reliability observed under diffraction conditions far from the training data. This expectation stems from the observation that changing individual parameters can cause drastic changes to the patterns and the CNN generally maintains exceptional performance.

## 9.5 Conclusions

In this work, a systematic study of the EBSD operating parameters and their individual effects on the classification performance of a convolutional neural network is performed. Despite the CNN being trained from diffraction patterns captured with a fixed geometry and SEM settings, it is found to be resilient over a wide range of conditions. Markedly decreased performance is generally only observed for the most challenging materials to differentiate (e.g. B2 and bcc or  $L1_2$  and fcc). Furthermore, it is encouraging to verify that parameters that effect the time to map an area (e.g. frame averaging or pattern resolution) can be modified to accelerate the process without substantially degrading model performance. For parameters such as tilt, it is reassuring to validate the model performs well over a variety of reasonable parameters. Although the CNN may not

achieve high accuracy under all conditions, such as a low number of frames averaged, when used appropriately it remains a highly capable method for assisting the user with phase identification and provides a level of phase differentiation markedly above what state of the art commercial methods are currently capable of. In the current version of the CNN, the parameter settings that cause noteworthy reductions in performance across a majority of space groups are a frame averaging of 1 or utilizing 10kV accelerating voltage. In the future, training models with patterns collected using a wider variety of operating conditions, particularly those with the largest effect on performance, could result in a model that is even more resilient to some of these types of perturbations to the diffraction patterns; although it may still not be possible to overcome all limitations, such as no frame averaging, owing to significant reductions in the signal to noise ratio. A future study should also investigate the co-variation of parameters from the default conditions. Ultimately, we expect the results of this research to encourage the continued development of these tools given the reliability observed and their potential to assist with or automate the analysis of electron diffraction patterns.

## **9.6 Acknowledgements**

K. Kaufmann was supported by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program. K. Kaufmann would also like to acknowledge the support of the ARCS Foundation, San Diego Chapter. K. Vecchio would like to acknowledge the financial generosity of the Oerlikon Group in support of his research group.

Chapter 9, in full, is currently in submission as “An Acquisition Parameter Study for Machine-Learning-Enabled Electron Backscatter Diffraction”, K. Kaufmann & K. S. Vecchio. The dissertation author was the primary investigator and author of this material.

## 9.7 References

- [1] A. Dunn, Q. Wang, A. Ganose, D. Dopp, A. Jain, Benchmarking Materials Property Prediction Methods: The Matbench Test Set and Automatminer Reference Algorithm, (2020). <http://arxiv.org/abs/2005.00707> (accessed August 4, 2020).
- [2] L. Ward, A. Dunn, A. Faghaninia, N.E.R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K.A. Persson, G.J. Snyder, I. Foster, A. Jain, Matminer: An open source toolkit for materials data mining, *Comput. Mater. Sci.* 152 (2018) 60–69. doi:10.1016/j.commatsci.2018.05.018.
- [3] B.L. DeCost, B. Lei, T. Francis, E.A. Holm, High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel, *Microsc. Microanal.* 25 (2019) 21–29. doi:10.1017/S1431927618015635.
- [4] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, T.J. Harrington, E. Marin, K.S. Vecchio, Crystal symmetry determination in electron diffraction using machine learning, *Science* (80-. ). 367 (2020) 564–568. doi:10.1126/science.aay3062.
- [5] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, *Comput. Mater. Sci.* 68 (2013) 314–319. doi:10.1016/j.commatsci.2012.10.028.
- [6] P. Avery, X. Wang, C. Oses, E. Gossett, D.M. Proserpio, C. Toher, S. Curtarolo, E. Zurek, Predicting superhard materials via a machine learning informed evolutionary structure search, *Npj Comput. Mater.* 5 (2019) 1–11. doi:10.1038/s41524-019-0226-8.
- [7] F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N.T.P. Hartono, S. Ramasamy, B.L. DeCost, S.I.P. Tian, G. Romano, A. Gilad Kusne, T. Buonassisi, Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks, *Npj Comput. Mater.* 5 (2019). doi:10.1038/s41524-019-0196-x.
- [8] K. Kaufmann, D. Maryanovsky, W.M. Mellor, C. Zhu, A.S. Rosengarten, T.J. Harrington, C. Oses, C. Toher, S. Curtarolo, K.S. Vecchio, Discovery of high-entropy ceramics via machine learning, *Npj Comput. Mater.* 6 (2020) 42. doi:10.1038/s41524-020-0317-6.
- [9] K. Kaufmann, K.S. Vecchio, Searching for high entropy alloys: A machine learning approach, *Acta Mater.* 198 (2020) 178–222. doi:10.1016/j.actamat.2020.07.065.
- [10] J. O’Mara, B. Meredig, K. Michel, Materials Data Infrastructure: A Case Study of the Citrination Platform to Examine Data Import, Storage, and Access, *JOM.* 68 (2016) 2031–2034. doi:10.1007/s11837-016-1984-0.
- [11] A. Ziletti, D. Kumar, M. Scheffler, L.M. Ghiringhelli, Insightful classification of crystal structures using deep learning, *Nat. Commun.* 9 (2018) 2775. doi:10.1038/s41467-018-05169-6.
- [12] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K.A. Persson, G.

- Ceder, A. Jain, Unsupervised word embeddings capture latent knowledge from materials science literature, *Nature*. 571 (2019) 95–98. doi:10.1038/s41586-019-1335-8.
- [13] J. Dagdelen, J. Montoya, M. De Jong, K. Persson, Computational prediction of new auxetic materials, *Nat. Commun.* 8 (2017) 1–8. doi:10.1038/s41467-017-00399-6.
- [14] T. Stan, Z.T. Thompson, P.W. Voorhees, Optimizing convolutional neural networks to perform semantic segmentation on large materials imaging datasets: X-ray tomography and serial sectioning, *Mater. Charact.* 160 (2020) 110119. doi:10.1016/j.matchar.2020.110119.
- [15] C. Chen, Y. Zuo, W. Ye, X. Li, S.P. Ong, Learning properties of ordered and disordered materials from multi-fidelity data, *Nat. Comput. Sci.* 1 (2021) 46–53. doi:10.1038/s43588-020-00002-x.
- [16] C. Zhu, H. Wang, K. Kaufmann, K.S. Vecchio, A computer vision approach to study surface deformation of materials, *Meas. Sci. Technol.* 31 (2020) 055602. doi:10.1088/1361-6501/ab65d9.
- [17] C. Zhu, K. Kaufmann, K.S. Vecchio, Novel remapping approach for HR-EBSD based on demons registration, *Ultramicroscopy*. 208 (2020) 112851. doi:10.1016/j.ultramic.2019.112851.
- [18] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature*. 521 (2015) 436–444. doi:10.1038/nature14539.
- [19] E.A. Holm, In defense of the black box, *Science* (80-. ). 364 (2019) 26–27. doi:10.1126/science.aax0162.
- [20] A. Adadi, M. Berrada, Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), *IEEE Access*. 6 (2018) 52138–52160. doi:10.1109/ACCESS.2018.2870052.
- [21] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, Li Fei-Fei, ImageNet: A large-scale hierarchical image database, in: *IEEE Conf. Comput. Vis. Pattern Recognit.*, Institute of Electrical and Electronics Engineers (IEEE), Miami, FL, 2010: pp. 248–255. doi:10.1109/cvpr.2009.5206848.
- [22] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature*. 542 (2017) 115–118. doi:10.1038/nature21056.
- [23] J.R. Zech, M.A. Badgeley, M. Liu, A.B. Costa, J.J. Titano, E.K. Oermann, Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study, *PLOS Med.* 15 (2018) e1002683. doi:10.1371/journal.pmed.1002683.
- [24] P. Riley, Three pitfalls to avoid in machine learning, *Nature*. 572 (2019) 27–29. doi:10.1038/d41586-019-02307-y.
- [25] A. Foden, A. Previero, T.B. Britton, *Advances in electron backscatter diffraction*, (2019).

- <http://arxiv.org/abs/1908.04860> (accessed December 9, 2019).
- [26] M.H. Modarres, R. Aversa, S. Cozzini, R. Ciancio, A. Leto, G.P. Brandino, Neural Network for Nanoscience Scanning Electron Microscope Image Recognition, *Sci. Rep.* 7 (2017) 1–12. doi:10.1038/s41598-017-13565-z.
- [27] K. de Haan, Z.S. Ballard, Y. Rivenson, Y. Wu, A. Ozcan, Resolution enhancement in scanning electron microscopy using deep learning, *Sci. Rep.* 9 (2019) 1–7. doi:10.1038/s41598-019-48444-2.
- [28] T. Xie, J.C. Grossman, Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties, *Phys. Rev. Lett.* 120 (2018) 145301. doi:10.1103/PhysRevLett.120.145301.
- [29] B.L. DeCost, E.A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, *Comput. Mater. Sci.* 110 (2015) 126–133. doi:10.1016/J.COMMATSCI.2015.08.011.
- [30] G. Roberts, S.Y. Haile, R. Sainju, D.J. Edwards, B. Hutchinson, Y. Zhu, Deep Learning for Semantic Segmentation of Defects in Advanced STEM Images of Steels, *Sci. Rep.* 9 (2019) 1–12. doi:10.1038/s41598-019-49105-0.
- [31] N. Laanait, Q. He, A.Y. Borisevich, Reconstruction of 3-D Atomic Distortions from Electron Microscopy with Deep Learning, (2019). <http://arxiv.org/abs/1902.06876> (accessed September 29, 2020).
- [32] S.R. Spurgeon, C. Ophus, L. Jones, A. Petford-Long, S. V. Kalinin, M.J. Olszta, R.E. Dunin-Borkowski, N. Salmon, K. Hattar, W.C.D. Yang, R. Sharma, Y. Du, A. Chiaramonti, H. Zheng, E.C. Buck, L. Kovarik, R.L. Penn, D. Li, X. Zhang, M. Murayama, M.L. Taheri, Towards data-driven next-generation transmission electron microscopy, *Nat. Mater.* (2020). doi:10.1038/s41563-020-00833-z.
- [33] K. Kaufmann, C. Zhu, A.S. Rosengarten, K.S. Vecchio, Deep Neural Network Enabled Space Group Identification in EBSD, *Microsc. Microanal.* 26 (2020) 447–457. doi:10.1017/S1431927620001506.
- [34] K. Kaufmann, C. Zhu, A.S. Rosengarten, D. Maryanovsky, H. Wang, K.S. Vecchio, Phase Mapping in EBSD Using Convolutional Neural Networks, *Microsc. Microanal.* 26 (2020) 458–468. doi:10.1017/S1431927620001488.
- [35] Z. Ding, E. Pascal, M. De Graef, Indexing of electron back-scatter diffraction patterns using a convolutional neural network, *Acta Mater.* 199 (2020) 370–382. doi:10.1016/j.actamat.2020.08.046.
- [36] Y.F. Shen, R. Pokharel, T.J. Nizolek, A. Kumar, T. Lookman, Convolutional neural network-based method for real-time orientation indexing of measured electron backscatter diffraction patterns, *Acta Mater.* 170 (2019) 118–131. doi:10.1016/j.actamat.2019.03.026.
- [37] A.J. Schwartz, M. Kumar, B.L. Adams, D.P. Field, *Electron backscatter diffraction in materials science*, Springer Science+Business Media, LLC, New York, 2009.

doi:10.1007/978-0-387-88136-2.

- [38] K.S. Vecchio, D.B. Williams, Convergent beam electron diffraction study of Al<sub>3</sub>Zr in Al-Zr AND Al-Li-Zr alloys, *Acta Metall.* 35 (1987) 2959–2970. doi:10.1016/0001-6160(87)90295-1.
- [39] K.S. Vecchio, D.B. Williams, Convergent beam electron diffraction analysis of the T<sub>1</sub> (Al<sub>2</sub>CuLi) phase in Al-Li-Cu alloys, *Metall. Trans. A.* 19 (1988) 2885–2891. doi:10.1007/BF02647714.
- [40] V.S. Tong, A.J. Knowles, D. Dye, T. Ben Britton, Rapid electron backscatter diffraction mapping: Painting by numbers, *Mater. Charact.* 147 (2019) 271–279. doi:10.1016/J.MATCHAR.2018.11.014.
- [41] K. Thomsen, N.H. Schmidt, A. Bewick, K. Larsen, J. Goulden, Improving the accuracy of orientation measurements using EBSD, *Microsc. Microanal.* 19 (2013) 724–725.
- [42] N.C.K. Lassen, *Automated Determination of Crystal Orientations from Electron Backscattering Patterns*, The Technical University of Denmark, 1994.
- [43] T.B. Britton, C. Maurice, R. Fortunier, J.H. Driver, A.P. Day, G. Meaden, D.J. Dingley, K. Mingard, A.J. Wilkinson, Factors affecting the accuracy of high resolution electron backscatter diffraction when using simulated patterns, *Ultramicroscopy.* 110 (2010) 1443–1453. doi:10.1016/j.ultramic.2010.08.001.
- [44] R. Hielscher, F. Bartel, T.B. Britton, Gazing at crystal balls: Electron backscatter diffraction pattern analysis and cross correlation on the sphere, *Ultramicroscopy.* 207 (2019). doi:10.1016/j.ultramic.2019.112836.
- [45] A. Foden, D.M. Collins, A.J. Wilkinson, T.B. Britton, Indexing electron backscatter diffraction patterns with a refined template matching approach, *Ultramicroscopy.* 207 (2019) 112845. doi:10.1016/j.ultramic.2019.112845.
- [46] T. Karthikeyan, M.K. Dash, S. Saroja, M. Vijayalakshmi, Evaluation of misindexing of EBSD patterns in a ferritic steel, *J. Microsc.* 249 (2013) 26–35. doi:10.1111/j.1365-2818.2012.03676.x.
- [47] C.L. Chen, R.C. Thomson, The combined use of EBSD and EDX analyses for the identification of complex intermetallic phases in multicomponent Al-Si piston alloys, *J. Alloys Compd.* 490 (2010) 293–300. doi:10.1016/j.jallcom.2009.09.181.
- [48] S. McLaren, S.M. Reddy, Automated mapping of K-feldspar by electron backscatter diffraction and application to <sup>40</sup>Ar/<sup>39</sup>Ar dating, *J. Struct. Geol.* 30 (2008) 1229–1241. doi:10.1016/J.JSG.2008.05.008.
- [49] F. Ram, M. De Graef, Phase differentiation by electron backscatter diffraction using the dictionary indexing approach, *Acta Mater.* 144 (2018) 352–364. doi:10.1016/j.actamat.2017.10.069.
- [50] Y.H. Chen, S.U. Park, D. Wei, G. Newstadt, M.A. Jackson, J.P. Simmons, M. De Graef,

- A.O. Hero, A Dictionary Approach to Electron Backscatter Diffraction Indexing, *Microsc. Microanal.* (2015). doi:10.1017/S1431927615000756.
- [51] S. Singh, Y. Guo, B. Winiarski, T.L. Burnett, P.J. Withers, M. De Graef, High resolution low kV EBSD of heavily deformed and nanocrystalline Aluminium by dictionary-based indexing, *Sci. Rep.* (2018). doi:10.1038/s41598-018-29315-8.
- [52] F. Ram, S. Wright, S. Singh, M. De Graef, Error analysis of the crystal orientations obtained by the dictionary approach to EBSD indexing, *Ultramicroscopy*. 181 (2017) 17–26. doi:10.1016/J.ULTRAMIC.2017.04.016.
- [53] A.P. Day, Spherical EBSD, in: *J. Microsc.*, 2008. doi:10.1111/j.1365-2818.2008.02011.x.
- [54] W.C. Lenthe, S. Singh, M. De Graef, A spherical harmonic transform approach to the indexing of electron back-scattered diffraction patterns, *Ultramicroscopy*. 207 (2019) 112841. doi:10.1016/j.ultramic.2019.112841.
- [55] C. Zhu, K. Kaufmann, K. Vecchio, Automated Reconstruction of Spherical Kikuchi Maps, *Microsc. Microanal.* (2019). doi:10.1017/S1431927619000710.
- [56] A. Momeni, K. Dehghani, X.X. Zhang, Mechanical and microstructural analysis of 2205 duplex stainless steel under hot working condition, *J. Mater. Sci.* 47 (2012) 2966–2974. doi:10.1007/s10853-011-6130-3.
- [57] M.A. Jackson, E. Pascal, M. De Graef, Dictionary Indexing of Electron Back-Scatter Diffraction Patterns: a Hands-On Tutorial, *Integr. Mater. Manuf. Innov.* 8 (2019) 226–246. doi:10.1007/s40192-019-00137-4.
- [58] F. Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, in: *Proc. IEEE Int. Conf. Comput. Vis.*, The Computer Vision Foundation, Seattle, 2017: pp. 1251–1258.  
[http://openaccess.thecvf.com/content\\_cvpr\\_2017/papers/Chollet\\_Xception\\_Deep\\_Learning\\_CVPR\\_2017\\_paper.pdf](http://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf) (accessed February 1, 2019).
- [59] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, G. Brain, TensorFlow: A System for Large-Scale Machine Learning, in: *Proc. 12th USENIX Symp. Oper. Syst. Des. Implement.*, USENIX, Savannah, Georgia, 2016: pp. 265–283. <https://tensorflow.org> (accessed October 8, 2018).
- [60] F. Chollet, Keras, (2015). <https://github.com/keras-team/keras>.
- [61] J. Basinger, D. Fullwood, J. Kacher, B. Adams, Pattern Center Determination in Electron Backscatter Diffraction Microscopy, *Microsc. Microanal.* 17 (2011) 330–340. doi:10.1017/S1431927611000389.
- [62] J. Goulden, P. Trimby, A. Bewick, The Benefits and Applications of a CMOS-based EBSD Detector, *Microsc. Microanal.* 24 (2018) 1128–1129.

- [63] D. Jha, S. Singh, R. Al-Bahrani, W.K. Liao, A. Choudhary, M. De Graef, A. Agrawal, Extracting grain orientations from EBSD patterns of polycrystalline materials using convolutional neural networks, *Microsc. Microanal.* 24 (2018) 497–502. doi:10.1017/S1431927618015131.
- [64] krkaufma, krkaufma/Electron-Diffraction-CNN v1.0.1, (2019). doi:10.5281/ZENODO.3564937.

## Appendix A Supplementary Information for Chapter 2

### Discovery of high-entropy ceramics via machine learning

Kevin Kaufmann<sup>1</sup>, Daniel Maryanovsky<sup>1,2,6</sup>, William M. Mellor<sup>3,6</sup>, Chaoyi Zhu<sup>3</sup>, Alexander S. Rosengarten<sup>1</sup>, Tyler J. Harrington<sup>3</sup>, Corey Oses<sup>4</sup>, Cormac Toher<sup>4</sup>, Stefano Curtarolo<sup>4,5</sup> and Kenneth S. Vecchio<sup>1\*</sup>

<sup>1</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA.

<sup>2</sup>Department of Cognitive Science, UC San Diego, La Jolla, CA 92093, USA.

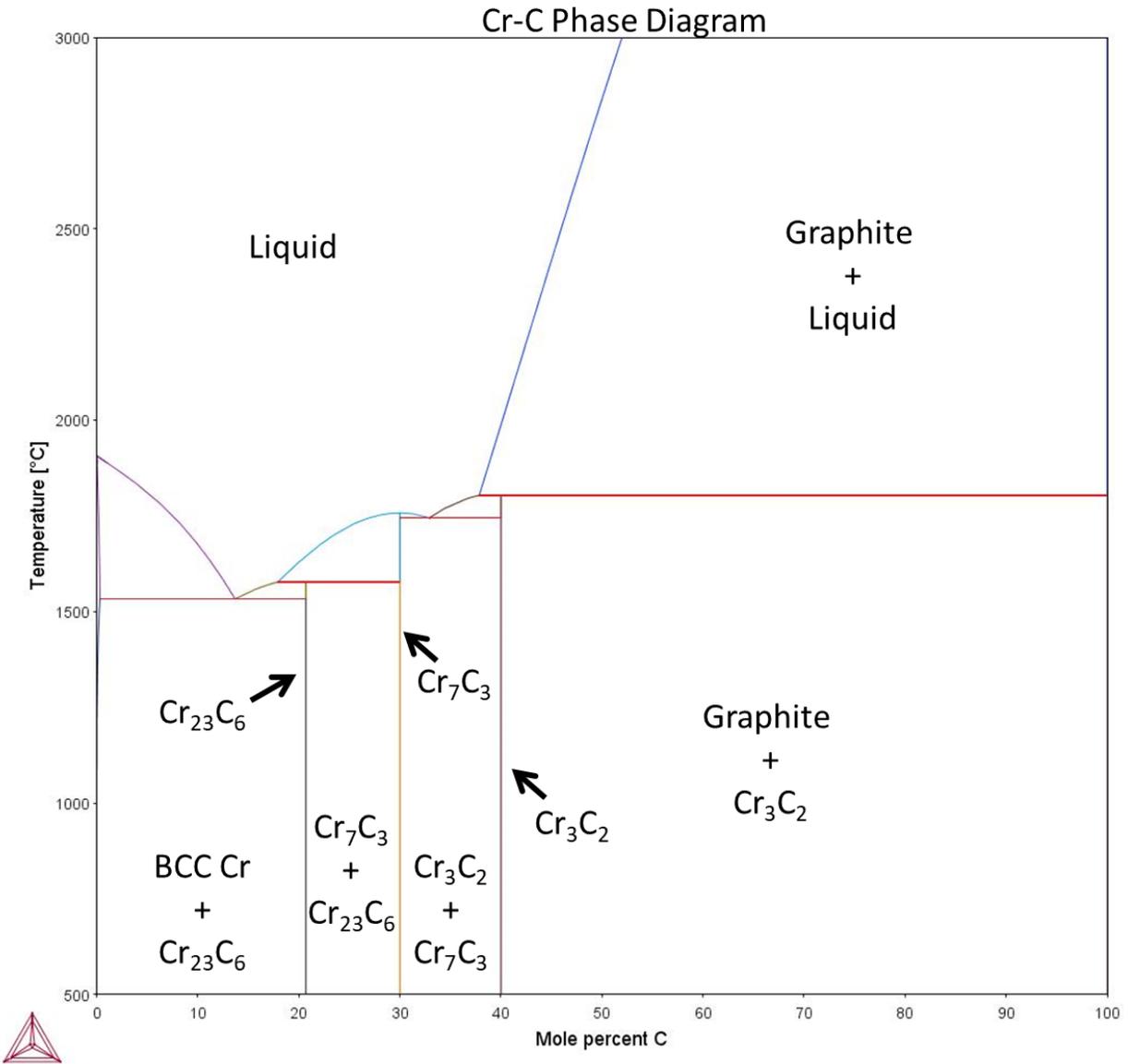
<sup>3</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA.

<sup>4</sup>Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708, USA.

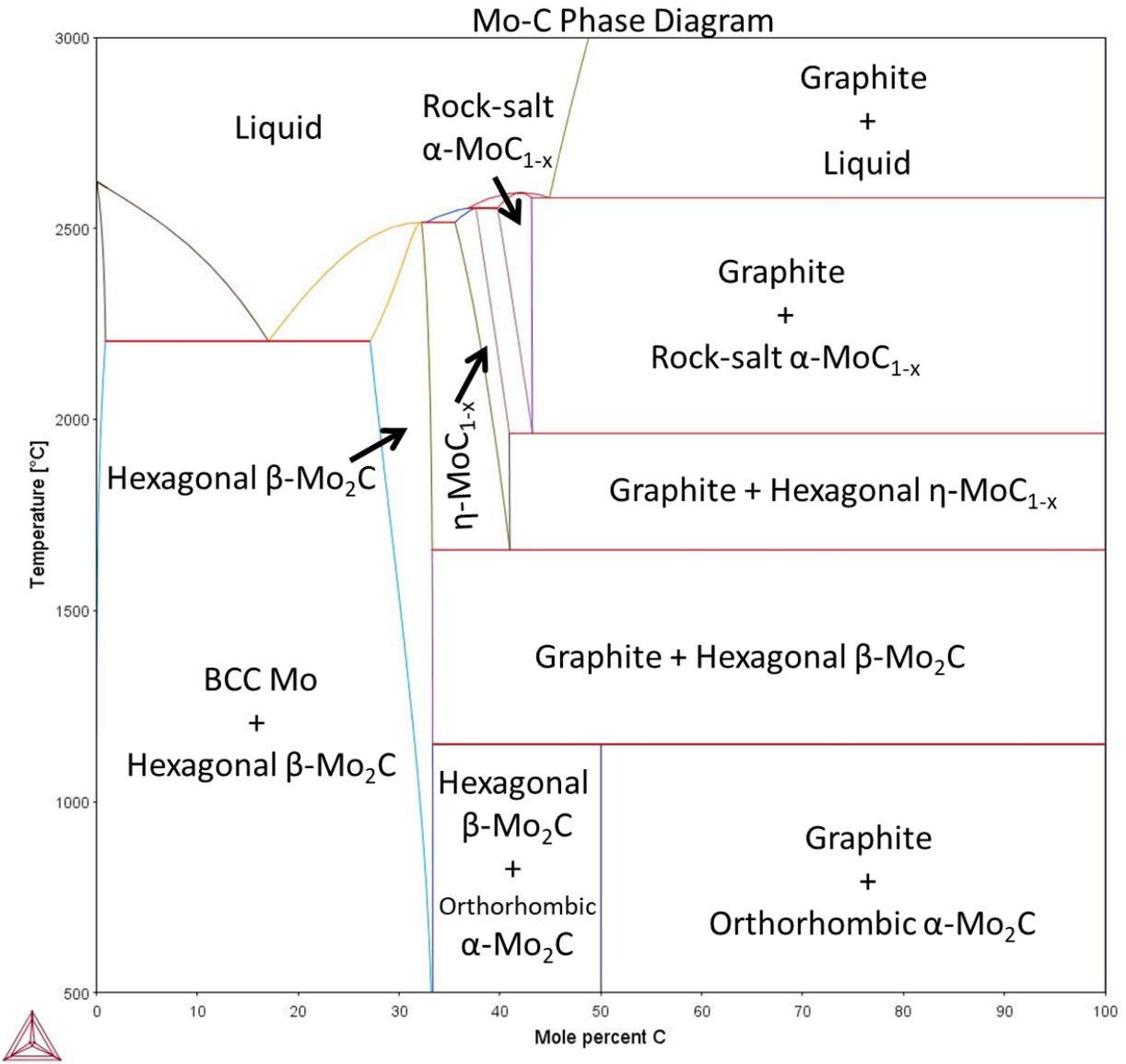
<sup>5</sup>Materials Science, Electrical Engineering, Physics and Chemistry, Duke University, Durham, NC 27708, USA.

<sup>6</sup>These authors contributed equally: Daniel Maryanovsky, William M. Mellor.

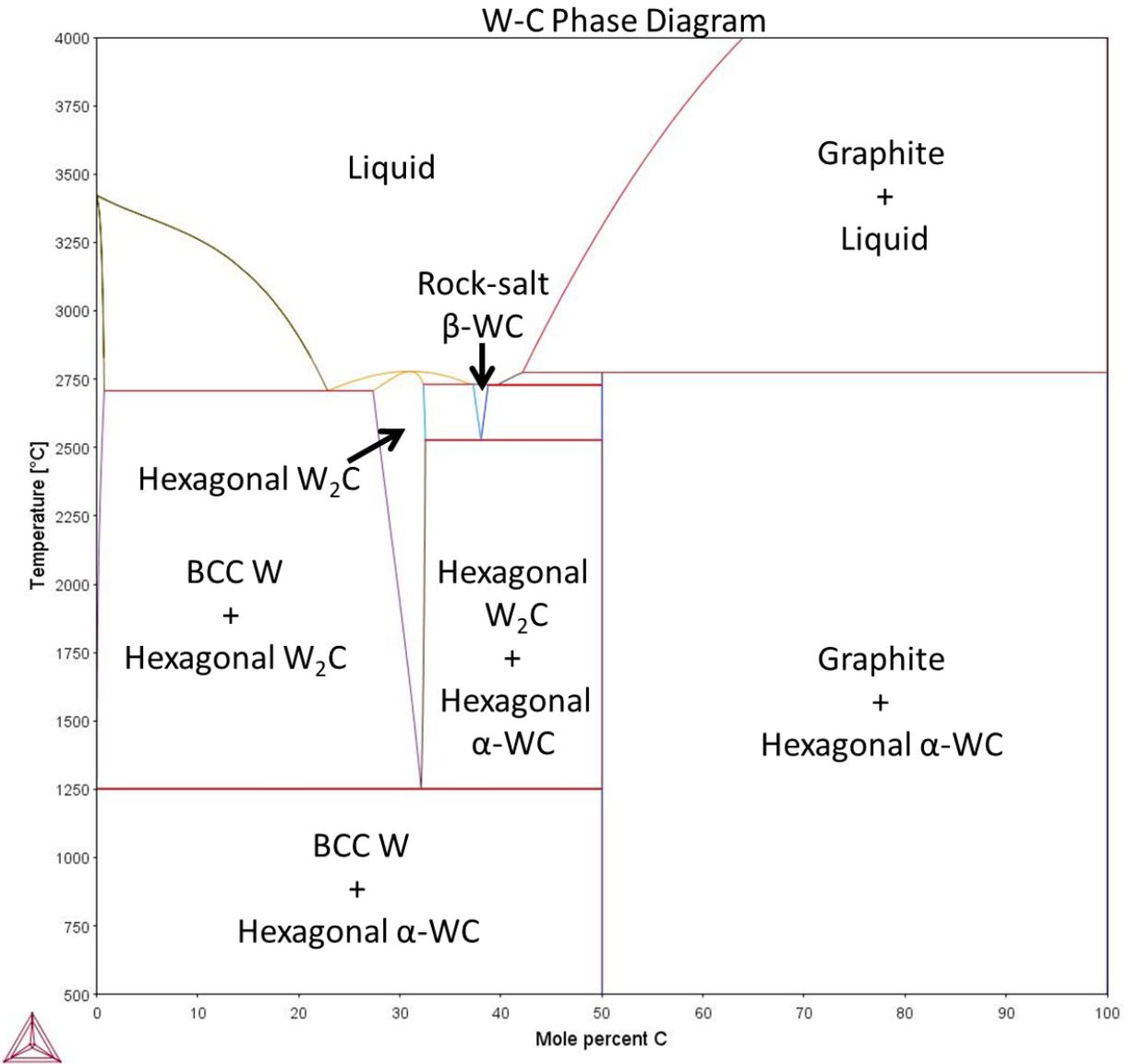
\* Corresponding Author



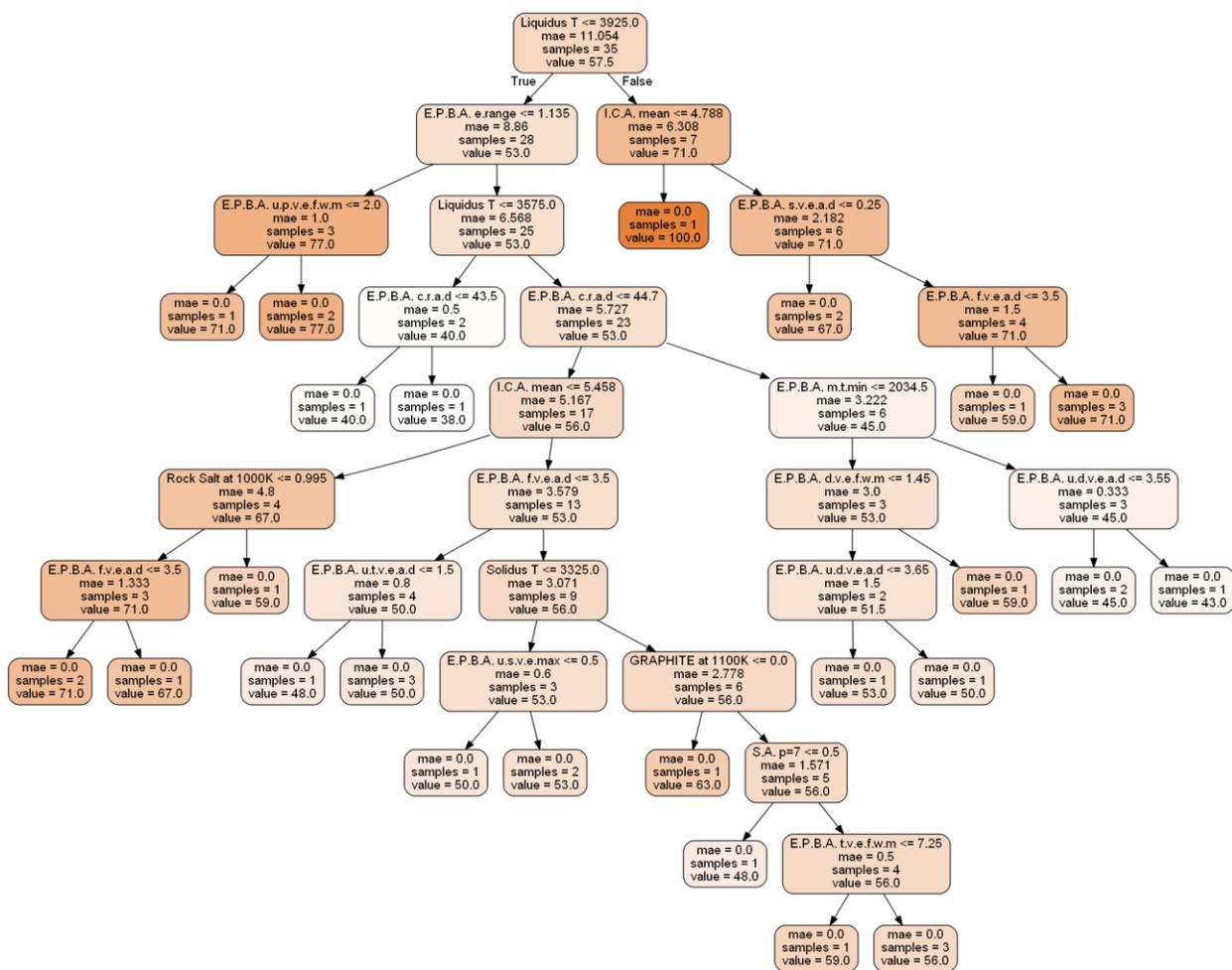
Supplementary Figure 1 Cr-C binary phase diagram from ThermoCalc Software SSOL6 database version 6.1.



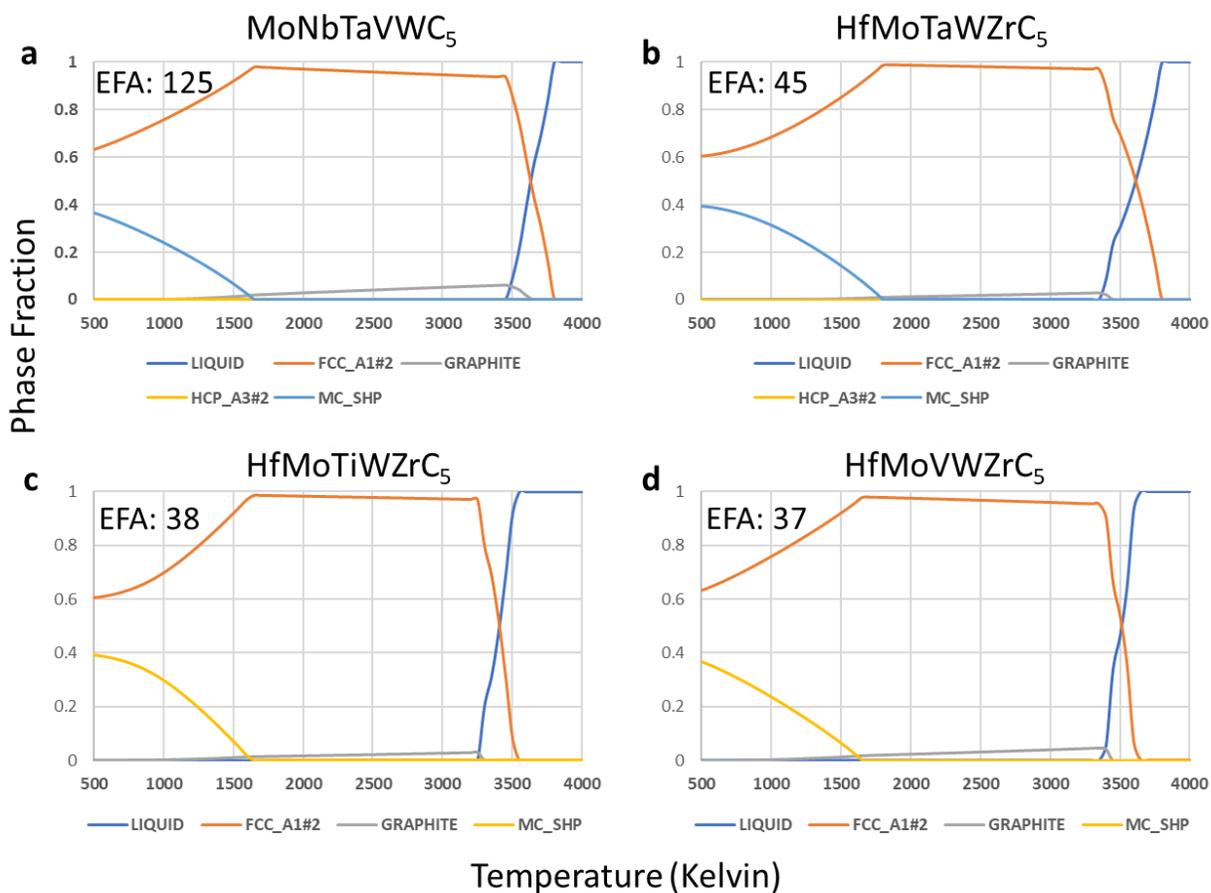
Supplementary Figure 2 Mo-C binary phase diagram from ThermoCalc Software SSOL6 database version 6.1.



Supplementary Figure 3 W-C binary phase diagram from ThermoCalc Software SSOL6 database version 6.1.



**Supplementary Figure 4 Example of a decision tree from the fitted model.** Each decision tree utilizes a subset of the labeled data and features in an order that maximizes the “unmixing” of the data. The rules leading to each subset are written inside individual rectangles. The subset population percentage is given by “samples”, the mean absolute error (mae) within each node is recorded, and the node shade represents the average EFA of that subset, i.e., dark orange illustrates a high proportion of materials with a large EFA.



**Supplementary Figure 5 Phase evolution diagrams for single and multi-phase compositions.** a) The only one of the four compositions found to form a single phase. b-d) These three compositions would appear to form a single phase; however, DFT and experimental results confirm their multi-phase nature.

**Supplementary Table 1 Model performance on existing data. Units: EFA in (eV/atom)<sup>-1</sup>.**

Composition	EFA <sub>Attrib</sub>	EFA <sub>CALP</sub>	EFA <sub>D</sub>	Ex	Composition	EFA <sub>Attribu</sub>	EFA <sub>CALP</sub>	EFA <sub>D</sub>	Exp
MoNbTaVWC	99	106	125	S	HfNbTiVWC <sub>5</sub>	50	58	59	
HfNbTaTiZrC	88	91	100	S	HfNbTaWZr	55	60	59	
HfNbTaTiVC <sub>5</sub>	97	92	100	S	NbTaVWZrC <sub>5</sub>	55	57	56	
MoNbTaTiVC	94	89	100		HfTaTiVWC <sub>5</sub>	55	55	56	
NbTaTiVZrC <sub>5</sub>	90	77	83		HfMoTaVWC	56	58	56	
HfMoNbTaTi	75	73	83		HfMoNbVWC	55	58	56	
NbTaTiVWC <sub>5</sub>	67	78	77	S	HfNbTiWZrC	51	53	53	
MoNbTaTiW	73	78	77		HfMoTaTiWC	54	54	53	
MoNbTiVWC <sub>5</sub>	72	77	71		HfMoNbTiW	54	50	53	
MoNbTaTiZr	69	69	71		HfTaTiWZrC <sub>5</sub>	54	50	50	S
HfTaTiVZrC <sub>5</sub>	72	75	71		TaTiVWZrC <sub>5</sub>	52	50	50	
HfNbTiVZrC <sub>5</sub>	69	72	71		NbTiVWZrC <sub>5</sub>	51	50	50	
HfMoNbTiVC <sub>5</sub>	70	69	71		HfMoTiVZrC <sub>5</sub>	57	55	50	
HfMoNbTaZr	63	67	71		HfMoTaVZrC	53	61	50	
HfMoNbTaW	63	68	71		HfMoNbVZrC	60	56	50	
HfMoNbTaVC	68	70	71		MoTaVWZrC <sub>5</sub>	49	51	48	
HfNbTaTiWC <sub>5</sub>	61	63	67	S	MoTaTiWZrC	49	50	48	
MoTaTiVWC <sub>5</sub>	72	71	67		MoNbVWZrC	49	53	48	
HfNbTaVZrC <sub>5</sub>	68	74	67		MoNbTiWZr	50	47	48	
HfNbTaVWC <sub>5</sub>	62	63	67		HfMoNbWZr	49	49	48	
HfMoTaTiVC <sub>5</sub>	71	65	67		HfTiVWZrC <sub>5</sub>	48	47	45	
HfMoNbTiZr	64	62	67		HfNbVWZrC <sub>5</sub>	44	46	45	
MoNbTaWZr	57	61	63		HfMoTiVWC <sub>5</sub>	45	45	45	
HfMoTaTiZrC	63	62	59		HfMoTaWZr	47	50	45	M
NbTaTiWZrC <sub>5</sub>	56	61	59		HfTaVWZrC <sub>5</sub>	47	55	43	
MoTaTiVZrC <sub>5</sub>	62	61	59		MoTiVWZrC <sub>5</sub>	44	43	40	
MoNbTiVZrC <sub>5</sub>	61	65	59		HfMoTiWZrC	42	40	38	M
MoNbTaVZrC	60	65	59		HfMoVWZrC <sub>5</sub>	41	45	37	M

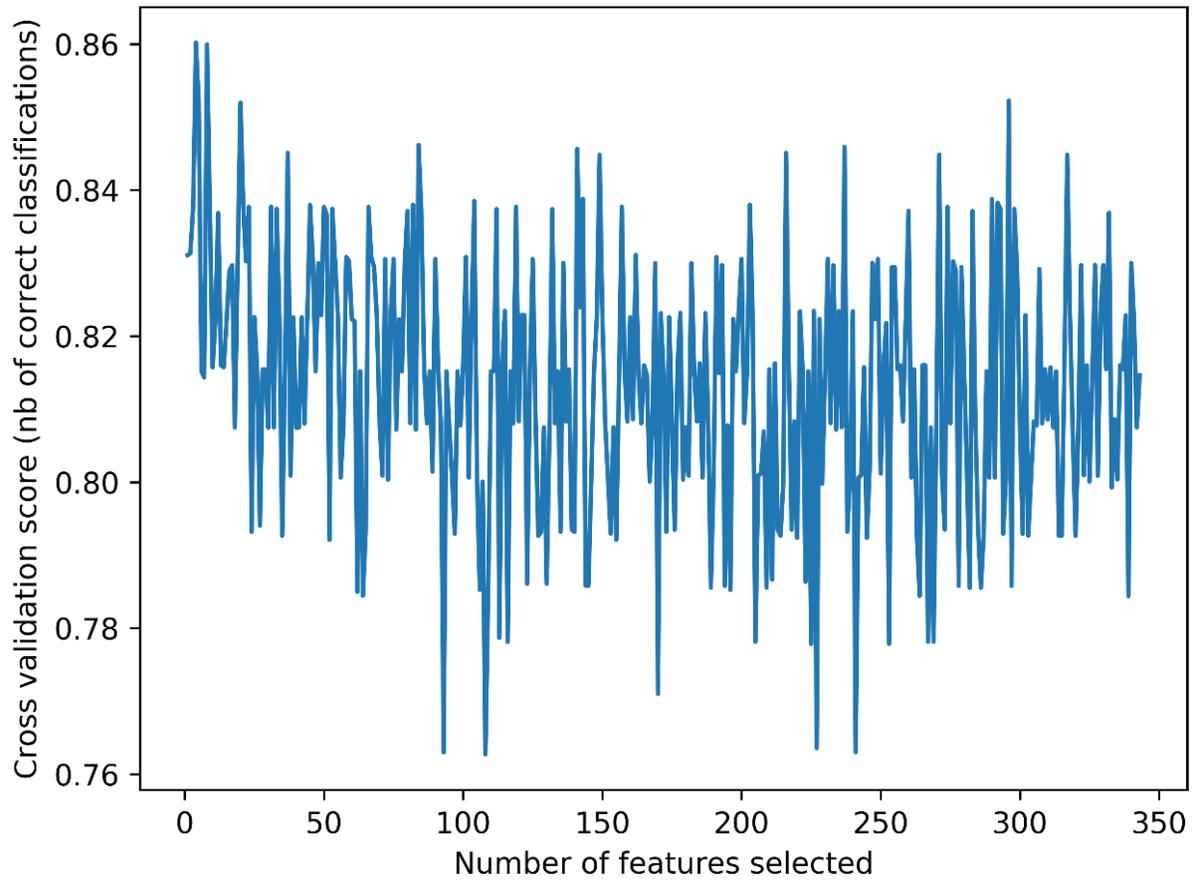
## **Appendix B Supplementary Information for Chapter 3**

Searching for high entropy alloys: A machine learning approach

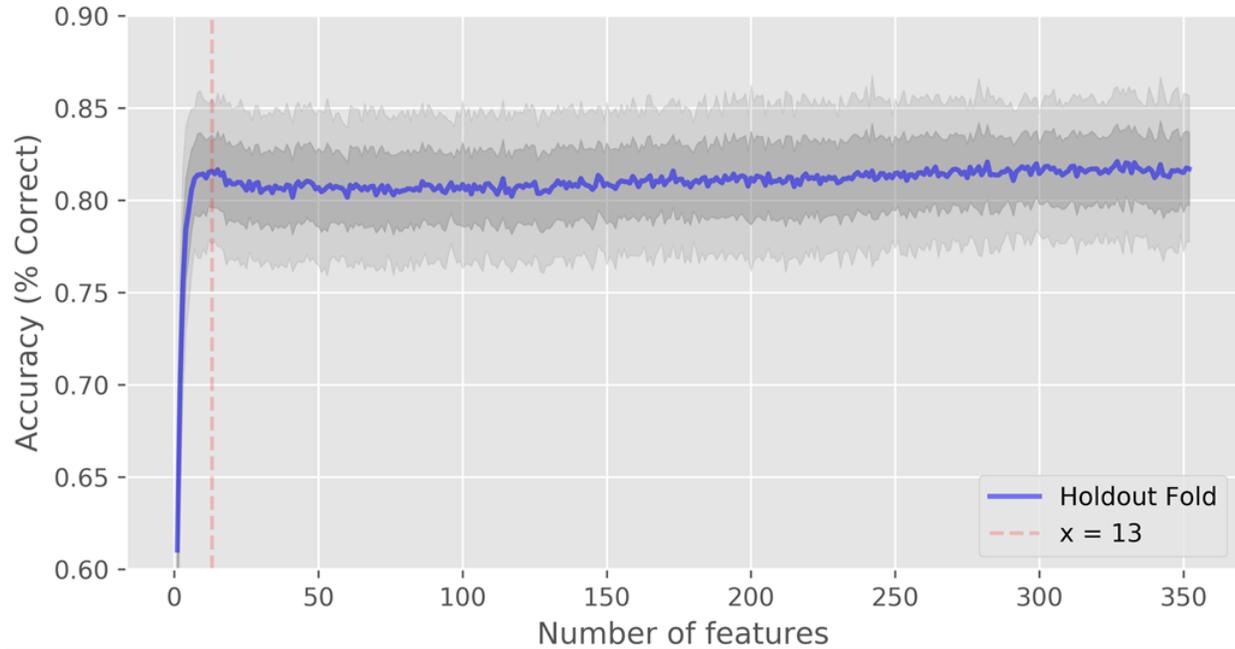
Kevin Kaufmann and Kenneth S. Vecchio\*

Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA

\*Corresponding author



**Supplementary Figure 6 Model performance as features are eliminated.** The 5-fold cross validation scores computed during the cross validated recursive feature elimination (RFECV) are plotted. After the current number of features is used to construct a model, the average score is recorded and the least important feature is removed. This process is repeated until all each number of features is scored. The highest score is achieved when thirteen features remain.



**Supplementary Figure 7 Statistical analysis of RFECV over 100 runs.** The average 5-fold cross validation scores computed over 100 individually seeded RFECV runs are plotted in blue. The dark and light gray shaded regions identify one and two standard deviations away from the average, respectively. The horizontal red line marks thirteen features, the number of features ultimately used in this study.

**Supplementary Table 2 ML-HEA results for binary alloys.** S.S.<sup>(ML)</sup> class of single solid solution phase predicted by the ML-HEA model (“none” = other or not found); S.S. class of single solid solution phase found using LTVC model from DFT [28] (“none” = other or not found); S.S.\* class of single solid solution phase found using CALPHAD (“none” = other or not found); S.S. Exp class of single solid solution phase found experimentally (“none” = other or not found).

	S.S. <sup>(ML)</sup>	S.S.	S.S. <sup>(*)</sup>	S.S. Exp.		S.S. <sup>(ML)</sup>	S.S.	S.S. <sup>(*)</sup>	S.S. Exp.
AgAu	fcc	fcc	fcc	fcc	HfNb	bcc	bcc	bcc	bcc
AgCu	none	none	none	none	HfTa	bcc	bcc	bcc	bcc
AgFe	none	none	none	none	HfTi	bcc	none	bcc	bcc
AgNi	none	none	none	none	HfV	none	none	none	none
AgPd	fcc	fcc	fcc	fcc	HfZr	bcc	none	bcc	bcc
AgPt	fcc	fcc	fcc	fcc	IrMo	none	none	none	none
AgSi	none	none	none	none	IrNi	fcc	fcc	fcc	fcc
AgTi	none	none	none	none	IrPd	fcc	none	fcc	fcc
AlMn	none	none	none	none	IrPt	fcc	fcc	fcc	fcc
AlCu	none	none	none	none	IrRh	fcc	fcc	fcc	fcc
AlHf	none	none	none	none	IrRu	fcc	fcc	none	fcc
AlSi	none	none	none	none	MnNb	none	none	bcc	none
AlTa	none	none	none	none	MnNi	fcc	none	fcc	fcc
AlTi	none	none	none	none	MnTa	none	none	bcc	none
AlW	none	none	none	none	MnV	bcc	none	bcc	bcc
AlZr	none	none	none	none	MoNb	bcc	bcc	bcc	bcc
AuCo	none	none	none	none	MoNi	none	none	none	none
AuCu	fcc	fcc	fcc	fcc	MoPd	fcc	fcc	none	fcc
AuNi	fcc	none	fcc	fcc	MoRh	none	none	none	none
AuPd	fcc	fcc	fcc	fcc	MoRu	none	none	none	none
AuPt	fcc	fcc	fcc	fcc	MoTa	bcc	bcc	bcc	bcc
AuTi	none	none	none	none	MoTi	bcc	bcc	bcc	bcc
AuV	fcc	fcc	bcc	fcc	MoV	bcc	bcc	bcc	bcc
CoCr	bcc	none	bcc	bcc	MoW	bcc	bcc	bcc	bcc
CoCu	none	none	none	none	MoZr	none	none	none	none
CoIr	fcc	fcc	fcc	fcc	NbTa	bcc	bcc	bcc	bcc
CoMn	fcc	none	fcc	fcc	NbTi	bcc	bcc	bcc	bcc
CoMo	none	none	none	none	NbV	bcc	bcc	bcc	bcc
CoNi	fcc	fcc	fcc	fcc	NbW	bcc	bcc	bcc	bcc
CoPd	fcc	none	fcc	fcc	NbZr	bcc	bcc	bcc	bcc
CoPt	fcc	fcc	fcc	fcc	NiPd	bcc	bcc	fcc	bcc
CoRh	fcc	fcc	fcc	fcc	NiPt	fcc	fcc	fcc	fcc
CoW	none	none	none	none	NiRh	fcc	fcc	fcc	fcc
CrCu	none	none	none	none	NiRu	none	fcc	none	none
CrFe	bcc	none	bcc	bcc	NiV	none	none	none	none

Supplementary Table 2, Continued.

	<b>S.S.<sup>(ML)</sup></b>	<b>S.S.</b>	<b>S.S.<sup>(*)</sup></b>	<b>S.S. Exp.</b>		<b>S.S.<sup>(ML)</sup></b>	<b>S.S.</b>	<b>S.S.<sup>(*)</sup></b>	<b>S.S. Exp.</b>
CrIr	none	none	none	none	NiW	none	none	none	none
CrMn	bcc	bcc	bcc	bcc	PdPt	fcc	fcc	fcc	fcc
CrMo	bcc	bcc	bcc	bcc	PdRh	fcc	fcc	fcc	fcc
CrNb	none	none	none	none	PdRu	none	none	none	none
CrNi	fcc	fcc	fcc	fcc	PdW	none	none	none	none
CrPt	fcc	fcc	fcc	fcc	PtRh	fcc	fcc	fcc	fcc
CrRh	none	none	none	none	PtRu	fcc	fcc	fcc	fcc
CrTa	bcc	none	bcc	bcc	PtTi	none	none	none	none
CrTi	bcc	none	bcc	bcc	PtW	none	none	none	none
CrV	bcc	bcc	bcc	bcc	RhRu	fcc	fcc	none	fcc
CrW	bcc	bcc	bcc	bcc	RhV	none	fcc	none	none
CuFe	none	none	none	none	RhW	none	none	none	none
CuIr	none	none	none	none	RuW	none	none	none	none
CuNi	fcc	fcc	fcc	fcc	TaTi	bcc	bcc	bcc	bcc
CuPd	fcc	fcc	fcc	fcc	TaV	bcc	bcc	bcc	bcc
CuPt	fcc	fcc	fcc	fcc	TaW	bcc	bcc	bcc	bcc
CuRh	fcc	fcc	fcc	fcc	TaZr	bcc	bcc	bcc	bcc
CuSi	none	none	none	none	TiV	bcc	bcc	bcc	bcc
CuV	none	none	none	none	TiW	bcc	bcc	bcc	bcc
FeMn	fcc	fcc	fcc	fcc	TiZr	bcc	bcc	bcc	bcc
FeMo	none	none	none	none	VW	bcc	bcc	bcc	bcc
FeNb	none	none	none	none	VZr	none	none	none	none
FeNi	bcc	none	bcc	bcc	WZr	none	none	none	none
FePd	fcc	fcc	fcc	fcc					

**Supplementary Table 3 ML-HEA results for ternary alloys.** S.S.<sup>(ML)</sup> class of single solid solution phase predicted by the ML-HEA model (“none” = other or not found); S.S. class of single solid solution phase found using LTVC model from DFT [28] (“none” = other or not found); S.S.\* class of single solid solution phase found using CALPHAD (“none” = other or not found).

	S.S. <sup>(ML)</sup>	S.S.	S.S.*		S.S. <sup>(ML)</sup>	S.S.	S.S.*
AgAlNi	none	none	none	AlCuSi	none	none	none
AgAlAu	none	none	fcc	AlFeMn	none	none	none
AgAlFe	none	none	none	AlHfSi	none	none	none
AgAlV	none	none	none	AlHfTa	none	none	none
AgAsAu	none	none	none	AlHfTi	none	fcc	bcc
AgAsCu	none	none	none	AlMnNi	none	none	none
AgAsPd	none	none	none	AlMnSi	none	none	none
AgAuCo	none	none	none	AlTaZr	none	none	none
AgAuCu	none	none	fcc	AlTiV	bcc	none	bcc
AgAuFe	none	none	none	AlTiZr	bcc	fcc	bcc
AgAuMo	none	none	none	AlVW	bcc	none	bcc
AgAuNi	none	none	none	AsAuCu	none	none	none
AgAuPd	fcc	fcc	fcc	AsAuPd	none	none	none
AgAuPt	fcc	fcc	fcc	AsCuPd	none	none	none
AgAuRh	fcc	none	none	AuCoCu	none	none	none
AgAuSi	none	none	none	AuCoIr	fcc	none	fcc
AgAuTi	none	none	none	AuCoMo	none	none	none
AgAuV	none	none	none	AuCoNi	none	none	none
AgCoCu	none	none	none	AuCoPd	fcc	none	fcc
AgCoPd	fcc	none	fcc	AuCoPt	fcc	none	none
AgCoPt	none	none	none	AuCoRh	fcc	none	none
AgCoRh	none	none	none	AuCoV	none	none	none
AgCuFe	none	none	none	AuCuFe	none	none	none
AgCuIr	none	none	none	AuCuIr	fcc	none	none
AgCuMn	none	none	none	AuCuNi	fcc	none	fcc
AgCuNi	none	none	none	AuCuPd	fcc	bcc	fcc
AgCuPd	none	none	none	AuCuPt	fcc	fcc	fcc
AgCuPt	none	none	none	AuCuRh	fcc	none	none
AgCuRh	none	none	none	AuCuSi	none	none	none
AgCuSi	none	none	none	AuCuTi	none	none	none
AgCuTi	none	none	none	AuCuV	none	fcc	none
AgCuV	none	none	none	AuCuW	none	none	none
AgCuZr	none	none	none	AuFeNi	fcc	none	fcc
AgFeNi	none	none	none	AuFePd	fcc	fcc	fcc
AgFePd	none	none	none	AuHfTi	none	fcc	none
AgHfZr	none	fcc	bcc	AuIrNi	fcc	none	fcc

Supplementary Table 3, Continued.

	<b>S.S<sup>(ML)</sup></b>	<b>S.S.</b>	<b>S.S.<sup>(*)</sup></b>		<b>S.S<sup>(ML)</sup></b>	<b>S.S.</b>	<b>S.S.<sup>(*)</sup></b>
AgIrNi	none	none	none	AuIrPd	fcc	none	fcc
AgIrPd	fcc	none	none	AuIrPt	fcc	none	fcc
AgIrPt	fcc	none	none	AuIrRh	fcc	none	none
AgIrRh	fcc	none	none	AuMoNi	none	none	none
AgNiPd	none	none	none	AuMoPd	fcc	none	none
AgNiPt	none	none	none	AuMoRh	none	none	none
AgNiRh	none	none	none	AuNiPd	fcc	none	fcc
AgPdPt	fcc	fcc	fcc	AuNiPt	fcc	none	fcc
AgPdRh	fcc	none	none	AuNiRh	fcc	none	none
AgPdV	none	none	none	AuNiV	none	none	fcc
AgPtRh	fcc	none	none	AuPdPt	fcc	fcc	fcc
AgPtRu	fcc	fcc	none	AuPdRh	fcc	none	none
AgPtTi	none	none	none	AuPdTi	none	none	fcc
AgPtV	none	none	none	AuPdV	none	fcc	none
AgTiV	none	none	none	AuPdW	none	none	none
AgTiZr	none	none	bcc	AuPtRh	fcc	none	none
AlAuCu	none	none	fcc	AuPtRu	fcc	none	none
AlAuNi	none	none	none	AuPtTi	none	none	none
AlAuSi	none	none	none	AuRhV	none	none	none
AlCoCr	none	none	none	AuRhW	none	none	none
AlCoCu	none	none	none	AuTaTi	none	fcc	none
AlCoNi	none	none	none	AuTiV	none	fcc	none
AlCrCu	none	none	none	AuTiZr	none	fcc	fcc
AlCrMn	bcc	none	bcc	CoCrCu	none	none	none
AlCrV	bcc	none	bcc	CoCrIr	fcc	fcc	fcc
AlCrW	none	none	none	CoCrMn	none	none	bcc
AlCuFe	none	none	none	CoCrNb	none	none	none
AlCuMn	none	none	bcc	CoCrNi	fcc	fcc	fcc
AlCuNi	none	none	none	CoCrPd	fcc	none	fcc
CoCrPt	fcc	none	fcc	CrMoTi	bcc	bcc	bcc
CoCrRh	fcc	fcc	fcc	CrMoV	bcc	bcc	bcc
CoCrW	none	none	none	CrMoW	bcc	bcc	bcc
CoCuIr	none	none	none	CrMoZr	none	none	none
CoCuMn	none	none	none	CrNbTa	bcc	none	none
CoCuNi	fcc	bcc	none	CrNbTi	bcc	none	none
CoCuPd	fcc	none	fcc	CrNbV	bcc	none	bcc
CoCuPt	fcc	none	none	CrNbW	bcc	none	bcc
CoCuRh	fcc	none	none	CrNbZr	none	none	none
CoCuV	none	none	none	CrNiPd	fcc	none	fcc

Supplementary Table 3, Continued.

	<b>S.S<sup>(ML)</sup></b>	<b>S.S.</b>	<b>S.S.<sup>(*)</sup></b>		<b>S.S<sup>(ML)</sup></b>	<b>S.S.</b>	<b>S.S.<sup>(*)</sup></b>
CoFeMn	<b>fcc</b>	<b>none</b>	<b>fcc</b>	CrNiPt	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>
CoIrMn	<b>fcc</b>	<b>none</b>	<b>fcc</b>	CrNiRh	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>
CoIrMo	<b>fcc</b>	<b>none</b>	<b>fcc</b>	CrPdPt	<b>fcc</b>	<b>none</b>	<b>none</b>
CoIrNi	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	CrPdRh	<b>fcc</b>	<b>none</b>	<b>fcc</b>
CoIrPd	<b>fcc</b>	<b>none</b>	<b>fcc</b>	CrPtRh	<b>fcc</b>	<b>none</b>	<b>fcc</b>
CoIrPt	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	CrReTa	<b>none</b>	<b>none</b>	<b>bcc</b>
CoIrRh	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	CrReW	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>
CoIrV	<b>none</b>	<b>none</b>	<b>fcc</b>	CrTaTi	<b>bcc</b>	<b>bcc</b>	<b>none</b>
CoIrW	<b>none</b>	<b>none</b>	<b>none</b>	CrTaV	<b>none</b>	<b>bcc</b>	<b>bcc</b>
CoMnNi	<b>fcc</b>	<b>none</b>	<b>fcc</b>	CrTaW	<b>none</b>	<b>bcc</b>	<b>bcc</b>
CoMnRh	<b>fcc</b>	<b>none</b>	<b>fcc</b>	CrTaZr	<b>none</b>	<b>none</b>	<b>none</b>
CoMoNi	<b>none</b>	<b>fcc</b>	<b>none</b>	CrTcW	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>
CoMoPt	<b>none</b>	<b>fcc</b>	<b>none</b>	CrTiV	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>
CoMoRh	<b>fcc</b>	<b>none</b>	<b>fcc</b>	CrTiW	<b>none</b>	<b>bcc</b>	<b>none</b>
CoNiPd	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	CrTiZr	<b>none</b>	<b>none</b>	<b>none</b>
CoNiPt	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	CrVW	<b>none</b>	<b>bcc</b>	<b>bcc</b>
CoNiRh	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	CrWZr	<b>none</b>	<b>none</b>	<b>none</b>
CoNiV	<b>fcc</b>	<b>none</b>	<b>fcc</b>	CuFeMn	<b>none</b>	<b>none</b>	<b>none</b>
CoPdPt	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	CuFeNi	<b>fcc</b>	<b>none</b>	<b>fcc</b>
CoPdRh	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	CuFePd	<b>none</b>	<b>fcc</b>	<b>none</b>
CoPdV	<b>none</b>	<b>none</b>	<b>none</b>	CuFeRh	<b>none</b>	<b>none</b>	<b>none</b>
CoPdW	<b>none</b>	<b>none</b>	<b>none</b>	CuFeV	<b>none</b>	<b>none</b>	<b>none</b>
CoPtRh	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	CuHfTi	<b>none</b>	<b>fcc</b>	<b>none</b>
CoPtV	<b>none</b>	<b>none</b>	<b>fcc</b>	CuHfZr	<b>none</b>	<b>fcc</b>	<b>none</b>
CoRhV	<b>none</b>	<b>none</b>	<b>fcc</b>	CuIrMn	<b>none</b>	<b>none</b>	<b>none</b>
CoRhW	<b>none</b>	<b>none</b>	<b>none</b>	CuIrNi	<b>fcc</b>	<b>fcc</b>	<b>none</b>
CrCuFe	<b>none</b>	<b>none</b>	<b>none</b>	CuIrPd	<b>fcc</b>	<b>none</b>	<b>none</b>
CrCuMo	<b>none</b>	<b>none</b>	<b>none</b>	CuIrPt	<b>fcc</b>	<b>none</b>	<b>none</b>
CrCuNi	<b>none</b>	<b>none</b>	<b>none</b>	CuIrRh	<b>fcc</b>	<b>none</b>	<b>none</b>
CrCuPd	<b>none</b>	<b>none</b>	<b>none</b>	CuMnNi	<b>fcc</b>	<b>none</b>	<b>fcc</b>
CrCuPt	<b>none</b>	<b>none</b>	<b>none</b>	CuNiPd	<b>fcc</b>	<b>none</b>	<b>fcc</b>
CrCuRh	<b>none</b>	<b>none</b>	<b>none</b>	CuNiPt	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>
CrCuV	<b>none</b>	<b>none</b>	<b>none</b>	CuNiRh	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>
CrCuW	<b>none</b>	<b>none</b>	<b>none</b>	CuNiSi	<b>none</b>	<b>none</b>	<b>none</b>
CrFeIr	<b>none</b>	<b>fcc</b>	<b>fcc</b>	CuNiV	<b>none</b>	<b>none</b>	<b>none</b>
CrFeMn	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>	CuPdPt	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>
CrFeMo	<b>none</b>	<b>none</b>	<b>none</b>	CuPdRh	<b>fcc</b>	<b>none</b>	<b>fcc</b>
CrFeNb	<b>none</b>	<b>none</b>	<b>none</b>	CuPdTi	<b>none</b>	<b>none</b>	<b>fcc</b>
CrFeNi	<b>fcc</b>	<b>none</b>	<b>fcc</b>	CuPdV	<b>none</b>	<b>fcc</b>	<b>none</b>
CrFeRh	<b>none</b>	<b>none</b>	<b>fcc</b>	CuPtRh	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>

Supplementary Table 3, Continued.

	S.S. <sup>(ML)</sup>	S.S.	S.S. <sup>(*)</sup>		S.S. <sup>(ML)</sup>	S.S.	S.S. <sup>(*)</sup>
CrFeTa	none	none	none	CuPtRu	fcc	none	fcc
CrFeTi	none	none	none	CuPtTi	none	none	none
CrFeV	bcc	bcc	bcc	CuPtV	none	none	none
CrFeW	none	none	none	CuRhV	none	none	none
CrHfNb	bcc	none	bcc	CuTaTi	none	fcc	bcc
CrHfTa	none	bcc	bcc	CuTaV	none	none	none
CrHfTi	bcc	none	bcc	CuTiV	none	none	none
CrHfW	none	none	none	CuTiZr	none	none	none
CrIrMn	none	none	fcc	CuVW	none	none	bcc
CrIrNi	fcc	fcc	fcc	FelrMn	fcc	none	fcc
CrIrPd	fcc	none	fcc	FelrMo	none	none	none
CrIrPt	fcc	none	none	FelrNi	fcc	fcc	fcc
CrIrRh	fcc	fcc	fcc	FelrRh	fcc	fcc	fcc
CrMnMo	none	none	bcc	FelrRu	fcc	none	fcc
CrMnNb	bcc	none	bcc	FeMnMo	none	none	none
CrMnNi	none	none	none	FeMnNb	none	none	none
CrMnRh	none	none	none	FeMnNi	fcc	none	fcc
CrMnTa	bcc	none	bcc	FeMnRh	fcc	none	fcc
CrMnTi	none	none	none	FeMnTa	none	none	none
CrMnV	bcc	none	bcc	FeMnTi	none	none	none
CrMnW	none	none	none	FeMnV	bcc	none	bcc
CrMoNb	bcc	none	bcc	FeMoNb	none	none	none
CrMoRe	none	bcc	bcc	FeMoNi	none	fcc	none
CrMoTa	none	none	bcc	FeMoPd	fcc	none	none
CrMoTc	none	bcc	bcc	FeMoRh	none	none	none
FeMoTa	none	none	none	MoNbTa	bcc	bcc	bcc
FeMoTi	none	none	none	MoNbTc	bcc	none	bcc
FeMoV	bcc	none	bcc	MoNbTi	bcc	bcc	bcc
FeMoW	none	none	none	MoNbV	bcc	bcc	bcc
FeNbTa	none	none	none	MoNbW	bcc	bcc	bcc
FeNbTi	none	none	none	MoNbZr	bcc	none	bcc
FeNbV	none	none	none	MoNiPd	none	none	none
FeNiRh	fcc	none	fcc	MoNiPt	none	fcc	none
FeTaTi	none	none	none	MoNiRh	none	fcc	none
FeTaV	none	none	none	MoPdPt	fcc	fcc	fcc
FeTaW	none	none	none	MoPdRh	fcc	fcc	fcc
FeTiV	none	none	bcc	MoReTa	bcc	none	bcc
FeVW	none	none	none	MoReTi	bcc	bcc	bcc

Supplementary Table 3, Continued.

	S.S. <sup>(ML)</sup>	S.S.	S.S. <sup>(*)</sup>		S.S. <sup>(ML)</sup>	S.S.	S.S. <sup>(*)</sup>
HfMoNb	bcc	bcc	bcc	MoReV	bcc	bcc	bcc
HfMoRe	none	none	bcc	MoReW	bcc	bcc	bcc
HfMoTa	none	bcc	bcc	MoTaTc	bcc	none	bcc
HfMoTi	bcc	bcc	bcc	MoTaTi	bcc	bcc	bcc
HfMoV	bcc	bcc	bcc	MoTaV	bcc	bcc	bcc
HfMoW	none	none	none	MoTaW	bcc	bcc	bcc
HfMoZr	bcc	none	bcc	MoTaZr	bcc	none	none
HfNbTa	bcc	bcc	bcc	MoTcTi	bcc	none	bcc
HfNbTi	bcc	bcc	bcc	MoTcV	bcc	none	bcc
HfNbV	bcc	bcc	bcc	MoTcW	bcc	bcc	bcc
HfNbW	bcc	bcc	bcc	MoTiV	bcc	bcc	bcc
HfNbZr	bcc	bcc	bcc	MoTiW	bcc	bcc	bcc
HfReTa	bcc	none	bcc	MoTiZr	bcc	bcc	bcc
HfTaTc	none	bcc	bcc	MoVW	bcc	bcc	bcc
HfTaTi	bcc	bcc	bcc	MoVZr	bcc	none	bcc
HfTaV	bcc	bcc	bcc	MoWZr	none	none	none
HfTaW	bcc	bcc	bcc	NbPdRh	none	fcc	none
HfTaZr	bcc	bcc	bcc	NbReTa	bcc	none	bcc
HfTiV	bcc	bcc	bcc	NbReTi	bcc	bcc	bcc
HfTiW	none	bcc	none	NbReV	bcc	bcc	bcc
HfTiZr	bcc	bcc	bcc	NbReW	bcc	bcc	bcc
HfVW	bcc	bcc	bcc	NbReZr	bcc	none	bcc
HfVZr	bcc	none	bcc	NbTaTc	bcc	none	bcc
HfWZr	none	none	none	NbTaTi	bcc	bcc	bcc
IrMnNi	fcc	none	fcc	NbTaV	bcc	bcc	bcc
IrMnRh	fcc	none	fcc	NbTaW	bcc	bcc	bcc
IrMoNi	none	none	none	NbTaZr	bcc	bcc	bcc
IrMoPd	fcc	none	fcc	NbTcV	bcc	none	bcc
IrMoPt	fcc	fcc	none	NbTcW	bcc	none	bcc
IrNbPd	none	none	none	NbTiV	bcc	bcc	bcc
IrNiPd	fcc	none	fcc	NbTiW	bcc	bcc	bcc
IrNiPt	fcc	fcc	fcc	NbTiZr	bcc	bcc	bcc
IrNiRh	fcc	fcc	fcc	NbVW	bcc	bcc	bcc
IrNiRu	fcc	fcc	fcc	NbVZr	bcc	none	bcc
IrNiW	none	none	none	NbWZr	bcc	none	none
IrOsPt	fcc	fcc	fcc	NiPdPt	fcc	fcc	fcc
IrPdPt	fcc	fcc	fcc	NiPdRh	fcc	fcc	fcc
IrPdRh	fcc	fcc	fcc	NiPdV	none	none	fcc
IrPdRu	fcc	none	fcc	NiPtRh	fcc	fcc	fcc
IrPdV	fcc	none	fcc	NiPtRu	fcc	fcc	fcc

Supplementary Table 3, Continued.

	<b>S.S<sup>(ML)</sup></b>	<b>S.S.</b>	<b>S.S.<sup>(*)</sup></b>		<b>S.S<sup>(ML)</sup></b>	<b>S.S.</b>	<b>S.S.<sup>(*)</sup></b>
IrPdW	<b>none</b>	<b>none</b>	<b>none</b>	NiPtV	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>
IrPtRe	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	NiRhV	<b>fcc</b>	<b>none</b>	<b>fcc</b>
IrPtRh	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	NiRhW	<b>none</b>	<b>fcc</b>	<b>none</b>
IrPtRu	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	PdPtRh	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>
IrPtTc	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>	PdPtRu	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>
IrRhV	<b>fcc</b>	<b>none</b>	<b>fcc</b>	PdPtTi	<b>none</b>	<b>fcc</b>	<b>none</b>
MnMoNb	<b>bcc</b>	<b>none</b>	<b>bcc</b>	PdRhV	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>
MnMoTa	<b>none</b>	<b>none</b>	<b>bcc</b>	PdRhW	<b>none</b>	<b>fcc</b>	<b>none</b>
MnMoTi	<b>none</b>	<b>none</b>	<b>none</b>	PdRuW	<b>none</b>	<b>fcc</b>	<b>bcc</b>
MnMoV	<b>bcc</b>	<b>none</b>	<b>bcc</b>	PdTiV	<b>none</b>	<b>none</b>	<b>bcc</b>
MnMoW	<b>none</b>	<b>bcc</b>	<b>none</b>	PtRhRu	<b>fcc</b>	<b>fcc</b>	<b>fcc</b>
MnNbTa	<b>none</b>	<b>none</b>	<b>bcc</b>	PtRhV	<b>fcc</b>	<b>none</b>	<b>none</b>
MnNbTi	<b>bcc</b>	<b>none</b>	<b>bcc</b>	PtRhW	<b>none</b>	<b>fcc</b>	<b>none</b>
MnNbV	<b>bcc</b>	<b>none</b>	<b>bcc</b>	ReTaTi	<b>bcc</b>	<b>none</b>	<b>bcc</b>
MnNbW	<b>none</b>	<b>none</b>	<b>none</b>	ReTaV	<b>bcc</b>	<b>none</b>	<b>bcc</b>
MnTaTi	<b>bcc</b>	<b>none</b>	<b>bcc</b>	ReTaW	<b>bcc</b>	<b>none</b>	<b>bcc</b>
MnTaV	<b>bcc</b>	<b>none</b>	<b>bcc</b>	ReTaZr	<b>none</b>	<b>none</b>	<b>bcc</b>
MnTaW	<b>none</b>	<b>none</b>	<b>none</b>	ReTiW	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>
MnTiV	<b>bcc</b>	<b>none</b>	<b>bcc</b>	ReVW	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>
MnTiW	<b>none</b>	<b>none</b>	<b>none</b>	TaTcTi	<b>bcc</b>	<b>none</b>	<b>bcc</b>
MnVW	<b>bcc</b>	<b>none</b>	<b>none</b>	TaTcV	<b>bcc</b>	<b>none</b>	<b>bcc</b>
MoNbRe	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>	TaTcW	<b>bcc</b>	<b>none</b>	<b>bcc</b>
TaTiV	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>	TcTiW	<b>bcc</b>	<b>none</b>	<b>bcc</b>
TaTiW	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>	TcVW	<b>bcc</b>	<b>none</b>	<b>bcc</b>
TaTiZr	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>	TiVZr	<b>none</b>	<b>none</b>	<b>bcc</b>
TaVW	<b>bcc</b>	<b>bcc</b>	<b>bcc</b>	TiWZr	<b>bcc</b>	<b>none</b>	<b>none</b>
TaVZr	<b>bcc</b>	<b>none</b>	<b>none</b>	VWZr	<b>bcc</b>	<b>none</b>	<b>none</b>
TaWZr	<b>bcc</b>	<b>none</b>	<b>none</b>				

**Supplementary Table 4 ML-HEA results for quaternary alloys.** S.S.<sup>(ML)</sup> class of single solid solution phase predicted by the ML-HEA model (“none” = other or not found); S.S. class of single solid solution phase found using LTVC model from DFT [28] (“none” = other or not found); S.S.\* class of single solid solution phase found using CALPHAD (“none” = other or not found).

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
AgAlAuCo	none	none	AgAuNiRu	none	none	AgCrNiRh	none	none
AgAlAuCu	none	fcc	AgAuNiSi	none	none	AgCrNiRu	none	none
AgAlAuFe	none	none	AgAuPdPt	fcc	fcc	AgCrOsPt	none	none
AgAlAuNi	none	none	AgAuPdRe	fcc	fcc	AgCrPdPt	fcc	none
AgAlAuSi	none	none	AgAuPdRh	fcc	none	AgCrPdRh	fcc	none
AgAlAuZr	none	none	AgAuPdRu	fcc	fcc	AgCrPdRu	none	none
AgAlCoCu	none	none	AgAuPtRh	fcc	fcc	AgCrPtRh	none	none
AgAlCoNi	none	none	AgAuPtRu	fcc	fcc	AgCrPtRu	none	fcc
AgAlCuFe	none	none	AgAuRhRu	fcc	fcc	AgCrRhRu	none	fcc
AgAlCuMn	none	none	AgCoCrCu	none	none	AgCuFeIr	none	none
AgAlCuNi	none	none	AgCoCrNi	none	none	AgCuFeNi	none	none
AgAlCuSi	none	none	AgCoCrOs	none	none	AgCuFeOs	none	fcc
AgAlCuZr	none	none	AgCoCrPd	none	none	AgCuFePd	none	fcc
AgAlFeNi	none	none	AgCoCrPt	none	none	AgCuFeRe	none	none
AgAlNiSi	none	none	AgCoCrRh	none	none	AgCuFeRh	none	none
AgAuCoCr	none	none	AgCoCrRu	none	none	AgCuFeRu	none	none
AgAuCoCu	none	none	AgCoCuFe	none	none	AgCuIrPd	none	none
AgAuCoIr	none	fcc	AgCoCuPd	none	none	AgCuIrPt	none	none
AgAuCoMo	none	none	AgCoCuPt	none	none	AgCuNbPd	none	none
AgAuCoNi	none	fcc	AgCoIrMo	none	none	AgCuNiPd	none	none
AgAuCoPd	fcc	none	AgCoIrPd	fcc	fcc	AgCuNiPt	none	none
AgAuCoPt	none	none	AgCoIrPt	none	fcc	AgCuNiSi	none	none
AgAuCoRe	none	none	AgCoMoNi	none	none	AgCuPdPt	fcc	fcc
AgAuCoRh	none	fcc	AgCoMoOs	none	none	AgCuPdRe	none	none
AgAuCoRu	none	none	AgCoMoPd	none	fcc	AgCuPdRh	none	none
AgAuCrCu	none	none	AgCoMoPt	none	none	AgCuPdRu	none	none
AgAuCrNi	none	fcc	AgCoMoRe	none	none	AgCuPtRe	none	none
AgAuCuFe	none	fcc	AgCoMoRh	none	none	AgCuPtRh	fcc	none
AgAuCuIr	fcc	none	AgCoMoRu	none	none	AgCuPtRu	none	none
AgAuCuNi	none	fcc	AgCoNiPd	none	fcc	AgFeIrNi	none	none

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
AgAuCuPd	fcc	fcc	AgCoNiPt	none	none	AgFeIrRu	none	none
AgAuCuPt	fcc	none	AgCoOsPd	none	none	AgFeMoNi	none	none
AgAuCuRh	none	none	AgCoOsPt	none	none	AgFeMoOs	none	fcc
AgAuCuRu	fcc	none	AgCoPdPt	fcc	fcc	AgFeMoPd	none	fcc
AgAuCuSi	none	none	AgCoPdRe	none	fcc	AgFeMoRe	none	fcc
AgAuCuZr	none	none	AgCoPdRh	fcc	fcc	AgFeMoRh	none	none
AgAuFeMo	none	none	AgCoPdRu	none	none	AgFeMoRu	none	none
AgAuFeNi	none	none	AgCoPtRe	none	fcc	AgFeNiOs	none	none
AgAuFePd	none	fcc	AgCoPtRh	none	fcc	AgFeNiPd	none	fcc
AgAuFeRh	none	fcc	AgCoPtRu	none	none	AgFeNiRe	none	none
AgAuIrMo	none	none	AgCrCuFe	none	none	AgFeNiRh	none	none
AgAuIrNi	none	none	AgCrCuNi	none	none	AgFeNiRu	none	none
AgAuIrPd	fcc	fcc	AgCrCuPd	none	none	AgFeOsPd	none	none
AgAuIrPt	fcc	fcc	AgCrCuPt	none	none	AgFeOsRe	none	fcc
AgAuIrRh	fcc	none	AgCrCuRh	none	none	AgFeOsRh	none	none
AgAuIrRu	fcc	none	AgCrCuRu	none	none	AgFeOsRu	none	fcc
AgAuMoNi	none	none	AgCrFeNi	none	none	AgFePdRe	none	none
AgAuMoPd	fcc	fcc	AgCrIrNi	none	none	AgFePdRh	none	fcc
AgAuMoRh	none	none	AgCrIrPd	fcc	none	AgFePdRu	none	none
AgAuNbRh	none	none	AgCrIrRh	none	none	AgFeReRu	none	fcc
AgAuNiPd	none	none	AgCrIrRu	none	fcc	AgFeRhRu	none	none
AgAuNiPt	none	none	AgCrNiOs	none	none	AgIrMoNi	none	none
AgAuNiRe	none	fcc	AgCrNiPd	none	none	AgIrMoPd	none	none
AgAuNiRh	none	fcc	AgCrNiPt	none	none	AgIrMoRh	fcc	none
AgIrNbPt	none	none	AlCrMnV	bcc	none	AuCuPtRu	fcc	fcc
AgIrNiPd	fcc	fcc	AlCrMnW	none	none	AuFeMoPd	none	none
AgIrNiPt	none	fcc	AlCrMoRe	bcc	bcc	AuFeNiPd	fcc	fcc
AgIrOsPd	fcc	fcc	AlCrMoTi	bcc	bcc	AuFeOsPd	fcc	none
AgIrOsPt	fcc	none	AlCrMoV	bcc	bcc	AuFePdRe	none	fcc
AgIrPdPt	fcc	none	AlCrMoW	bcc	bcc	AuFePdRh	fcc	none
AgIrPdRe	none	none	AlCrNbV	bcc	none	AuFePdRu	fcc	none
AgIrPdRh	fcc	none	AlCrReTi	none	none	AuIrMoPd	fcc	none

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
AgIrPdRu	none	fcc	AlCrReV	none	bcc	AuIrNiPd	fcc	fcc
AgIrPtRe	none	fcc	AlCrReW	bcc	bcc	AuIrNiPt	fcc	none
AgIrPtRh	fcc	fcc	AlCrTiV	bcc	none	AuIrOsPd	fcc	fcc
AgIrPtRu	fcc	fcc	AlCrTiW	bcc	bcc	AuIrOsPt	fcc	fcc
AgMoNiOs	none	none	AlCrVW	bcc	bcc	AuIrPdPt	fcc	fcc
AgMoNiPd	none	none	AlCuFeMn	none	none	AuIrPdRe	fcc	none
AgMoNiPt	none	none	AlCuFeNi	none	bcc	AuIrPdRh	fcc	fcc
AgMoNiRe	none	none	AlCuFeV	none	none	AuIrPdRu	fcc	fcc
AgMoNiRh	none	none	AlFeMnRe	none	none	AuIrPtRe	fcc	none
AgMoNiRu	none	fcc	AlFeMnTi	none	none	AuIrPtRh	fcc	fcc
AgMoOsPd	none	none	AlFeMnV	bcc	none	AuIrPtRu	fcc	fcc
AgMoOsRe	none	fcc	AlFeTiV	bcc	none	AuMoNiPd	none	fcc
AgMoOsRh	fcc	none	AlHfTiV	bcc	bcc	AuMoNiPt	fcc	none
AgMoPdPt	none	none	AlHfTiW	none	bcc	AuMoOsPd	fcc	none
AgMoPdRe	none	fcc	AlHfVW	none	none	AuMoPdPt	fcc	none
AgMoPdRh	none	none	AlMnMoRe	none	none	AuMoPdRe	none	fcc
AgMoPdRu	none	none	AlMnReW	none	none	AuMoPdRh	fcc	none
AgMoPtRe	none	none	AlMnTiV	bcc	bcc	AuMoPdRu	fcc	fcc
AgMoPtRh	none	none	AlMnVW	none	none	AuMoPtRe	none	none
AgMoPtRu	none	none	AlMoTiV	bcc	bcc	AuMoPtRh	fcc	fcc
AgMoReRh	none	none	AlMoVW	bcc	bcc	AuMoPtRu	none	none
AgMoRhRu	fcc	none	AlReVW	bcc	bcc	AuNbPdRh	none	none
AgNbPdPt	none	none	AlTiVW	bcc	bcc	AuNbPtRh	none	none
AgNbPdRh	none	none	AuCoCrPd	fcc	fcc	AuNiOsPd	none	fcc
AgNbPtRh	none	none	AuCoCrPt	none	none	AuNiOsPt	none	none
AgNiOsPd	none	fcc	AuCoCuPd	none	none	AuNiPdPt	fcc	none
AgNiOsPt	none	none	AuCoCuPt	none	none	AuNiPdRe	none	none
AgNiPdPt	fcc	fcc	AuCoIrPd	fcc	fcc	AuNiPdRh	fcc	none
AgNiPdRe	none	fcc	AuCoIrPt	fcc	none	AuNiPdRu	none	fcc
AgNiPdRh	fcc	fcc	AuCoMoPd	fcc	fcc	AuNiPtRe	none	fcc
AgNiPdRu	none	none	AuCoMoPt	none	none	AuNiPtRh	fcc	none
AgNiPtRe	none	fcc	AuCoNiPd	fcc	fcc	AuNiPtRu	none	none

Supplementary Table 4, Continued.

	S.S.(ML)	S.S.		S.S.(ML)	S.S.		S.S.(ML)	S.S.
AgNiPtRh	none	fcc	AuCoNiPt	fcc	none	AuOsPdPt	fcc	fcc
AgNiPtRu	none	fcc	AuCoOsPd	none	none	AuOsPdRe	none	fcc
AgOsPdPt	fcc	none	AuCoOsPt	none	none	AuOsPdRh	fcc	fcc
AgOsPdRe	none	fcc	AuCoPdPt	fcc	none	AuOsPtRh	fcc	fcc
AgOsPdRh	fcc	fcc	AuCoPdRe	none	none	AuOsPtRu	fcc	none
AgOsPdRu	none	fcc	AuCoPdRh	fcc	none	AuPdPtRe	fcc	fcc
AgOsPtRe	none	fcc	AuCoPdRu	none	none	AuPdPtRh	fcc	none
AgOsPtRh	fcc	fcc	AuCoPtRe	none	none	AuPdPtRu	fcc	fcc
AgOsPtRu	fcc	fcc	AuCoPtRh	fcc	none	AuPdReRh	fcc	fcc
AgPdPtRe	none	fcc	AuCoPtRu	none	none	AuPdReRu	none	fcc
AgPdPtRh	fcc	none	AuCrCuPd	fcc	none	AuPdRhRu	fcc	fcc
AgPdPtRu	fcc	fcc	AuCrCuPt	none	none	AuPtReRh	fcc	none
AgPdReRh	none	fcc	AuCrIrPd	fcc	none	AuPtReRu	fcc	fcc
AgPdReRu	none	none	AuCrIrPt	fcc	none	AuPtRhRu	fcc	fcc
AgPdRhRu	none	fcc	AuCrNiPd	fcc	none	CoCrCuNi	none	none
AgPtReRh	none	fcc	AuCrNiPt	none	none	CoCrCuPd	none	none
AgPtReRu	none	fcc	AuCrOsPt	none	none	CoCrCuPt	none	none
AgPtRhRu	none	fcc	AuCrPdPt	fcc	fcc	CoCrFeNi	fcc	fcc
AlNbTiV	bcc	bcc	AuCrPdRh	fcc	none	CoCrIrPd	fcc	none
AlNbVW	bcc	bcc	AuCrPdRu	none	none	CoCrIrPt	fcc	none
AlCoCrCu	none	none	AuCrPtRh	none	none	CoCrMnMo	none	none
AlCrCuFe	none	none	AuCrPtRu	none	none	CoCrMnNb	none	none
AlCrFeMn	bcc	none	AuCuFePd	none	fcc	CoCrMnNi	fcc	fcc
AlCrFeMo	bcc	none	AuCuIrPd	fcc	none	CoCrMoNb	none	bcc
AlCrFeNi	none	none	AuCuIrPt	fcc	none	CoCrMoW	none	bcc
AlCrFeRe	none	none	AuCuNiPd	fcc	none	CoCrNbW	none	bcc
AlCrFeTi	bcc	none	AuCuNiPt	fcc	none	CoCrNiPd	fcc	none
AlCrFeV	bcc	none	AuCuOsPt	fcc	none	CoCrNiPt	fcc	none
AlCrFeW	none	none	AuCuPdPt	fcc	none	CoCrOsPd	none	none
AlCrHfTi	bcc	bcc	AuCuPdRe	none	fcc	CoCrOsPt	none	none
AlCrHfW	none	none	AuCuPdRh	fcc	none	CoCrPdPt	fcc	none
AlCrMnMo	bcc	none	AuCuPdRu	fcc	fcc	CoCrPdRh	fcc	none

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
AlCrMnRe	none	none	AuCuPtRe	none	fcc	CoCrPdRu	none	none
AlCrMnTi	bcc	none	AuCuPtRh	fcc	none	CoCrPtRh	fcc	none
CoCrPtRu	none	fcc	CrCuMnV	none	none	CrHfReTi	none	none
CoCuIrPd	fcc	none	CrCuMnW	none	none	CrHfReV	none	none
CoCuIrPt	fcc	none	CrCuMoTa	none	none	CrHfReW	none	none
CoCuNiPd	fcc	none	CrCuMoTi	none	none	CrHfReZr	none	none
CoCuNiPt	fcc	fcc	CrCuMoV	none	none	CrHfTaTi	bcc	bcc
CoCuOsPd	none	none	CrCuMoW	none	none	CrHfTaV	none	bcc
CoCuOsPt	none	none	CrCuNbTa	none	none	CrHfTaW	none	bcc
CoCuPdPt	fcc	fcc	CrCuNbTi	none	none	CrHfTaZr	none	bcc
CoCuPdRe	none	fcc	CrCuNbV	none	none	CrHfTiV	bcc	bcc
CoCuPdRh	fcc	none	CrCuNbW	none	none	CrHfTiW	none	bcc
CoCuPdRu	none	none	CrCuNiPd	none	none	CrHfTiZr	bcc	bcc
CoCuPtRe	none	fcc	CrCuNiPt	fcc	none	CrHfVW	none	none
CoCuPtRh	fcc	fcc	CrCuOsPd	none	fcc	CrHfVZr	bcc	none
CoCuPtRu	fcc	none	CrCuOsPt	none	fcc	CrHfWZr	none	bcc
CoFeMnNi	fcc	fcc	CrCuPdPt	fcc	none	CrIrNiPd	fcc	none
CoIrMoPd	fcc	none	CrCuPdRh	fcc	none	CrIrNiPt	fcc	none
CoIrMoPt	fcc	none	CrCuPdRu	none	none	CrIrOsPd	none	none
CoIrNiPd	fcc	none	CrCuPtRh	fcc	none	CrIrOsPt	none	fcc
CoIrNiPt	fcc	none	CrCuPtRu	none	none	CrIrPdPt	fcc	none
CoIrOsPd	none	fcc	CrCuTaTi	none	none	CrIrPdRh	fcc	fcc
CoIrOsPt	fcc	fcc	CrCuTaV	none	none	CrIrPdRu	none	none
CoIrPdPt	fcc	none	CrCuTaW	none	none	CrIrPtRh	fcc	fcc
CoIrPdRe	none	fcc	CrCuTiV	none	none	CrIrPtRu	none	fcc
CoIrPdRh	fcc	none	CrCuTiW	none	none	CrMnMoNb	bcc	none
CoIrPdRu	none	none	CrCuVW	none	none	CrMnMoNi	none	none
CoIrPtRe	fcc	fcc	CrFeMnMo	bcc	none	CrMnMoRe	none	bcc
CoIrPtRh	fcc	fcc	CrFeMnNb	none	none	CrMnMoTa	bcc	none
CoIrPtRu	fcc	none	CrFeMnNi	fcc	none	CrMnMoTi	none	none
CoMoNbW	none	bcc	CrFeMnRe	none	bcc	CrMnMoV	bcc	none
CoMoNiPd	fcc	none	CrFeMnTa	none	none	CrMnMoW	none	bcc

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
CoMoNiPt	<b>fcc</b>	<b>none</b>	CrFeMnTi	<b>none</b>	<b>none</b>	CrMnMoZr	<b>none</b>	<b>none</b>
CoMoOsPd	<b>none</b>	<b>none</b>	CrFeMnV	<b>bcc</b>	<b>bcc</b>	CrMnNbNi	<b>none</b>	<b>none</b>
CoMoOsPt	<b>none</b>	<b>fcc</b>	CrFeMnW	<b>none</b>	<b>bcc</b>	CrMnNbRe	<b>none</b>	<b>none</b>
CoMoPdPt	<b>fcc</b>	<b>none</b>	CrFeMoNb	<b>none</b>	<b>none</b>	CrMnNbTa	<b>bcc</b>	<b>none</b>
CoMoPdRe	<b>none</b>	<b>fcc</b>	CrFeMoNi	<b>none</b>	<b>bcc</b>	CrMnNbTi	<b>bcc</b>	<b>none</b>
CoMoPdRh	<b>fcc</b>	<b>none</b>	CrFeMoRe	<b>none</b>	<b>bcc</b>	CrMnNbV	<b>bcc</b>	<b>none</b>
CoMoPdRu	<b>none</b>	<b>none</b>	CrFeMoTa	<b>none</b>	<b>none</b>	CrMnNbW	<b>none</b>	<b>none</b>
CoMoPtRe	<b>none</b>	<b>fcc</b>	CrFeMoTi	<b>bcc</b>	<b>none</b>	CrMnNbZr	<b>none</b>	<b>none</b>
CoMoPtRh	<b>fcc</b>	<b>fcc</b>	CrFeMoV	<b>bcc</b>	<b>bcc</b>	CrMnNiRe	<b>none</b>	<b>bcc</b>
CoMoPtRu	<b>none</b>	<b>fcc</b>	CrFeMoW	<b>none</b>	<b>bcc</b>	CrMnNiV	<b>none</b>	<b>none</b>
CoNiOsPd	<b>none</b>	<b>none</b>	CrFeNbRe	<b>none</b>	<b>none</b>	CrMnReTa	<b>none</b>	<b>none</b>
CoNiOsPt	<b>none</b>	<b>none</b>	CrFeNbTa	<b>none</b>	<b>none</b>	CrMnReTi	<b>none</b>	<b>none</b>
CoNiPdPt	<b>fcc</b>	<b>none</b>	CrFeNbTi	<b>none</b>	<b>none</b>	CrMnReV	<b>none</b>	<b>bcc</b>
CoNiPdRe	<b>none</b>	<b>fcc</b>	CrFeNbV	<b>none</b>	<b>none</b>	CrMnReW	<b>none</b>	<b>bcc</b>
CoNiPdRh	<b>fcc</b>	<b>none</b>	CrFeNbW	<b>none</b>	<b>bcc</b>	CrMnReZr	<b>none</b>	<b>none</b>
CoNiPdRu	<b>none</b>	<b>none</b>	CrFeNiPd	<b>fcc</b>	<b>none</b>	CrMnTaTi	<b>bcc</b>	<b>none</b>
CoNiPtRe	<b>none</b>	<b>fcc</b>	CrFeNiRe	<b>none</b>	<b>bcc</b>	CrMnTaV	<b>bcc</b>	<b>none</b>
CoNiPtRh	<b>fcc</b>	<b>fcc</b>	CrFeNiV	<b>none</b>	<b>bcc</b>	CrMnTaW	<b>none</b>	<b>none</b>
CoNiPtRu	<b>fcc</b>	<b>none</b>	CrFeOsPd	<b>none</b>	<b>fcc</b>	CrMnTaZr	<b>none</b>	<b>none</b>
CoOsPdPt	<b>none</b>	<b>none</b>	CrFePdRu	<b>none</b>	<b>fcc</b>	CrMnTiV	<b>none</b>	<b>none</b>
CoOsPdRe	<b>none</b>	<b>fcc</b>	CrFeReTa	<b>none</b>	<b>none</b>	CrMnTiW	<b>none</b>	<b>none</b>
CoOsPdRh	<b>none</b>	<b>fcc</b>	CrFeReTi	<b>none</b>	<b>none</b>	CrMnTiZr	<b>none</b>	<b>none</b>
CoOsPdRu	<b>none</b>	<b>fcc</b>	CrFeReV	<b>none</b>	<b>bcc</b>	CrMnVW	<b>none</b>	<b>bcc</b>
CoOsPtRe	<b>none</b>	<b>fcc</b>	CrFeReW	<b>none</b>	<b>bcc</b>	CrMnWZr	<b>none</b>	<b>none</b>
CoOsPtRh	<b>none</b>	<b>fcc</b>	CrFeTaTi	<b>none</b>	<b>none</b>	CrMoNbNi	<b>none</b>	<b>none</b>
CoOsPtRu	<b>none</b>	<b>fcc</b>	CrFeTaV	<b>none</b>	<b>none</b>	CrMoNbRe	<b>bcc</b>	<b>bcc</b>
CoPdPtRe	<b>none</b>	<b>fcc</b>	CrFeTaW	<b>none</b>	<b>none</b>	CrMoNbTa	<b>none</b>	<b>bcc</b>
CoPdPtRh	<b>fcc</b>	<b>fcc</b>	CrFeTiV	<b>none</b>	<b>none</b>	CrMoNbTi	<b>bcc</b>	<b>bcc</b>
CoPdPtRu	<b>fcc</b>	<b>fcc</b>	CrFeTiW	<b>none</b>	<b>none</b>	CrMoNbV	<b>none</b>	<b>bcc</b>
CoPdReRh	<b>none</b>	<b>none</b>	CrFeVW	<b>none</b>	<b>bcc</b>	CrMoNbW	<b>none</b>	<b>none</b>
CoPdReRu	<b>none</b>	<b>fcc</b>	CrHfMoNb	<b>bcc</b>	<b>bcc</b>	CrMoNbZr	<b>none</b>	<b>none</b>
CoPdRhRu	<b>none</b>	<b>none</b>	CrHfMoRe	<b>none</b>	<b>none</b>	CrMoNiRe	<b>none</b>	<b>bcc</b>

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
CoPtReRh	none	fcc	CrHfMoTa	none	bcc	CrMoReTa	none	none
CoPtReRu	none	fcc	CrHfMoTi	bcc	bcc	CrMoReTi	bcc	bcc
CoPtRhRu	fcc	fcc	CrHfMoV	bcc	none	CrMoReV	bcc	bcc
CrCuFeMn	none	none	CrHfMoW	none	bcc	CrMoReW	bcc	bcc
CrCuFeMo	none	none	CrHfMoZr	none	bcc	CrMoReZr	none	none
CrCuFeNb	none	none	CrHfNbRe	none	none	CrMoTaTi	bcc	bcc
CrCuFeNi	none	none	CrHfNbTa	bcc	bcc	CrMoTaV	bcc	bcc
CrCuFePd	none	none	CrHfNbTi	bcc	bcc	CrMoTaW	none	bcc
CrCuFeV	none	none	CrHfNbV	bcc	none	CrMoTaZr	none	none
CrCuFeW	none	none	CrHfNbW	none	bcc	CrMoTiV	bcc	bcc
CrCuIrPd	fcc	none	CrHfNbZr	none	bcc	CrMoTiW	bcc	bcc
CrCuIrPt	none	none	CrHfReTa	none	none	CrMoTiZr	bcc	none
CrMoVW	bcc	bcc	CuFeVW	none	none	FeMnMoV	bcc	none
CrMoVZr	none	none	CuIrNiPd	fcc	none	FeMnMoW	none	none
CrMoWZr	none	none	CuIrNiPt	fcc	none	FeMnNbRe	none	none
CrNbNiV	none	none	CuIrOsPd	none	fcc	FeMnNbTa	none	none
CrNbReTa	bcc	none	CuIrOsPt	fcc	none	FeMnNbTi	none	none
CrNbReTi	bcc	bcc	CuIrPdPt	fcc	none	FeMnNbV	none	none
CrNbReV	bcc	none	CuIrPdRe	none	none	FeMnNbW	none	none
CrNbReW	bcc	bcc	CuIrPdRh	fcc	none	FeMnReTa	none	none
CrNbReZr	none	none	CuIrPdRu	none	none	FeMnReTi	none	bcc
CrNbTaTi	bcc	bcc	CuIrPtRe	fcc	none	FeMnReV	none	bcc
CrNbTaV	bcc	bcc	CuIrPtRh	fcc	none	FeMnReW	none	bcc
CrNbTaW	none	bcc	CuIrPtRu	fcc	none	FeMnTaTi	none	none
CrNbTaZr	none	none	CuMnMoTa	none	none	FeMnTaV	none	none
CrNbTiV	bcc	bcc	CuMnMoV	none	none	FeMnTaW	none	none
CrNbTiW	bcc	bcc	CuMnMoW	none	none	FeMnTiV	none	none
CrNbTiZr	none	none	CuMnNbTa	none	none	FeMnTiW	none	none
CrNbVW	bcc	bcc	CuMnNbV	none	none	FeMnVW	none	bcc
CrNbVZr	none	none	CuMnTaV	none	none	FeMoNbRe	bcc	none
CrNbWZr	none	none	CuMnTaW	none	none	FeMoNbTa	none	none
CrNiOsPd	none	fcc	CuMnTiV	none	none	FeMoNbTi	none	none

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
CrNiOsPt	none	none	CuMnVW	none	none	FeMoNbV	bcc	bcc
CrNiPdPt	fcc	none	CuMoNbTa	none	none	FeMoNbW	none	none
CrNiPdRh	fcc	none	CuMoNbV	none	none	FeMoNiPd	fcc	fcc
CrNiPdRu	none	none	CuMoNbW	none	none	FeMoNiV	none	none
CrNiPtRh	fcc	none	CuMoTaTi	none	none	FeMoOsPd	none	none
CrNiPtRu	none	none	CuMoTaW	none	none	FeMoPdRe	none	fcc
CrNiReV	none	bcc	CuMoTiV	none	none	FeMoPdRh	fcc	none
CrOsPdPt	none	none	CuMoVW	none	none	FeMoPdRu	none	fcc
CrOsPdRh	none	none	CuNbPdPt	none	none	FeMoReTa	none	none
CrOsPdRu	none	none	CuNbPdRh	none	fcc	FeMoReTi	none	bcc
CrOsPtRh	none	fcc	CuNbPtRh	none	none	FeMoReV	none	bcc
CrOsPtRu	fcc	none	CuNbTaTi	none	none	FeMoReW	none	bcc
CrPdPtRh	fcc	none	CuNbTaV	none	none	FeMoTaTi	none	none
CrPdPtRu	none	none	CuNbTaW	none	none	FeMoTaV	none	none
CrPdRhRu	none	none	CuNbTiV	none	bcc	FeMoTaW	none	none
CrPtRhRu	none	fcc	CuNbVW	none	none	FeMoTiV	bcc	none
CrReTaTi	bcc	none	CuNiOsPd	none	fcc	FeMoTiW	none	none
CrReTaV	bcc	none	CuNiOsPt	none	fcc	FeMoVW	none	none
CrReTaW	bcc	none	CuNiPdPt	fcc	fcc	FeNbReTa	none	none
CrReTaZr	none	none	CuNiPdRe	none	fcc	FeNbReTi	none	none
CrReTiV	none	bcc	CuNiPdRh	fcc	none	FeNbReV	bcc	none
CrReTiW	bcc	none	CuNiPdRu	none	none	FeNbReW	none	none
CrReTiZr	none	none	CuNiPtRe	none	fcc	FeNbTaTi	none	none
CrReVW	bcc	bcc	CuNiPtRh	fcc	fcc	FeNbTaV	none	bcc
CrReVZr	none	none	CuNiPtRu	fcc	fcc	FeNbTaW	none	none
CrReWZr	none	none	CuOsPdPt	fcc	fcc	FeNbTiV	none	none
CrTaTiV	bcc	bcc	CuOsPdRe	none	none	FeNbTiW	none	none
CrTaTiW	bcc	bcc	CuOsPdRh	fcc	none	FeNbVW	none	bcc
CrTaTiZr	none	none	CuOsPdRu	none	none	FeNiOsPd	none	none
CrTaVW	none	bcc	CuOsPtRe	none	fcc	FeNiPdRe	none	fcc
CrTaVZr	none	none	CuOsPtRh	fcc	fcc	FeNiPdRh	fcc	none
CrTaWZr	none	none	CuOsPtRu	fcc	fcc	FeNiPdRu	none	none

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
CrTiVW	bcc	bcc	CuPdPtRe	none	fcc	FeOsPdRe	none	fcc
CrTiVZr	none	none	CuPdPtRh	fcc	none	FeOsPdRh	none	fcc
CrTiWZr	none	none	CuPdPtRu	fcc	fcc	FeOsPdRu	none	fcc
CrVWZr	none	none	CuPdReRh	none	fcc	FePdReRh	none	none
CuFeIrPd	fcc	none	CuPdReRu	none	none	FePdReRu	none	fcc
CuFeMnNb	none	none	CuPdRhRu	none	fcc	FePdRhRu	none	none
CuFeMnTa	none	none	CuPtReRh	none	fcc	FeReTaTi	none	none
CuFeMnTi	none	none	CuPtReRu	none	fcc	FeReTaV	none	none
CuFeMnV	none	none	CuPtRhRu	fcc	fcc	FeReTiV	none	none
CuFeMoTi	none	none	CuTaTiV	bcc	none	FeReVW	none	bcc
CuFeMoV	none	none	CuTaTiW	none	none	FeTaTiV	none	none
CuFeNbTa	none	none	CuTaVW	none	none	FeTaTiW	none	none
CuFeNbTi	none	none	CuTiVW	none	none	FeTaVW	none	none
CuFeNbV	none	none	FeIrMoPd	fcc	none	FeTiVW	none	none
CuFeNiPd	fcc	none	FeIrNiPd	fcc	fcc	HfMoNbRe	bcc	none
CuFeNiV	none	none	FeIrOsPd	none	fcc	HfMoNbTa	bcc	bcc
CuFeOsPd	none	fcc	FeIrPdRe	none	none	HfMoNbTi	bcc	bcc
CuFePdRe	none	fcc	FeIrPdRu	none	fcc	HfMoNbV	bcc	bcc
CuFePdRh	fcc	none	FeMnMoNb	none	none	HfMoNbW	bcc	bcc
CuFePdRu	none	none	FeMnMoRe	none	bcc	HfMoNbZr	bcc	bcc
CuFeTaV	none	none	FeMnMoTa	none	none	HfMoReTa	none	none
CuFeTiV	none	none	FeMnMoTi	none	none	HfMoReTi	bcc	bcc
HfMoReV	bcc	none	IrOsPdRe	none	none	MnTiVW	bcc	none
HfMoReW	none	none	IrOsPdRh	fcc	none	MnTiVZr	none	none
HfMoReZr	none	none	IrOsPdRu	none	none	MnTiWZr	none	none
HfMoTaTi	bcc	bcc	IrOsPtRe	none	fcc	MnVWZr	none	none
HfMoTaV	none	bcc	IrOsPtRh	fcc	fcc	MoNbNiV	none	bcc
HfMoTaW	bcc	bcc	IrOsPtRu	fcc	fcc	MoNbReTa	bcc	none
HfMoTaZr	bcc	bcc	IrPdPtRe	fcc	fcc	MoNbReTi	bcc	bcc
HfMoTiV	bcc	bcc	IrPdPtRh	fcc	fcc	MoNbReV	bcc	bcc
HfMoTiW	bcc	bcc	IrPdPtRu	fcc	none	MoNbReW	bcc	bcc
HfMoTiZr	bcc	bcc	IrPdReRh	fcc	none	MoNbReZr	bcc	none

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
HfMoVW	none	bcc	IrPdReRu	none	none	MoNbTaTi	bcc	bcc
HfMoVZr	none	bcc	IrPdRhRu	fcc	none	MoNbTaV	bcc	bcc
HfMoWZr	none	bcc	IrPtReRh	fcc	fcc	MoNbTaW	bcc	bcc
HfNbReTa	bcc	bcc	IrPtReRu	fcc	fcc	MoNbTaZr	bcc	none
HfNbReTi	bcc	bcc	IrPtRhRu	fcc	fcc	MoNbTiV	bcc	bcc
HfNbReV	bcc	bcc	MnMoNbRe	none	bcc	MoNbTiW	bcc	bcc
HfNbReW	bcc	bcc	MnMoNbTa	bcc	none	MoNbTiZr	bcc	bcc
HfNbReZr	none	bcc	MnMoNbTi	bcc	none	MoNbVW	bcc	bcc
HfNbTaTi	bcc	bcc	MnMoNbV	bcc	none	MoNbVZr	bcc	bcc
HfNbTaV	bcc	bcc	MnMoNbW	none	bcc	MoNbWZr	none	bcc
HfNbTaW	bcc	bcc	MnMoNbZr	none	none	MoNiOsPd	none	fcc
HfNbTaZr	bcc	bcc	MnMoNiV	none	none	MoNiOsPt	none	fcc
HfNbTiV	bcc	bcc	MnMoReTa	none	none	MoNiPdPt	fcc	fcc
HfNbTiW	bcc	bcc	MnMoReTi	none	none	MoNiPdRe	none	fcc
HfNbTiZr	bcc	bcc	MnMoReV	bcc	bcc	MoNiPdRh	fcc	none
HfNbVW	none	bcc	MnMoReW	none	bcc	MoNiPdRu	none	none
HfNbVZr	bcc	bcc	MnMoReZr	none	none	MoNiPtRe	none	fcc
HfNbWZr	bcc	bcc	MnMoTaTi	bcc	none	MoNiPtRh	fcc	fcc
HfReTaTi	bcc	bcc	MnMoTaV	bcc	none	MoNiPtRu	none	fcc
HfReTaV	bcc	bcc	MnMoTaW	none	none	MoNiReV	none	bcc
HfReTaW	none	none	MnMoTaZr	none	none	MoOsPdPt	none	fcc
HfReTaZr	none	none	MnMoTiV	bcc	none	MoOsPdRe	none	fcc
HfReTiV	none	bcc	MnMoTiW	none	none	MoOsPdRh	none	fcc
HfReTiW	none	bcc	MnMoTiZr	none	none	MoOsPdRu	none	fcc
HfReTiZr	none	bcc	MnMoVW	bcc	bcc	MoOsPtRe	none	fcc
HfReVW	bcc	bcc	MnMoVZr	none	none	MoOsPtRh	fcc	fcc
HfReVZr	none	none	MnMoWZr	none	none	MoOsPtRu	none	fcc
HfReWZr	none	bcc	MnNbNiV	none	none	MoPdPtRe	none	fcc
HfTaTiV	bcc	bcc	MnNbReTa	none	none	MoPdPtRh	fcc	fcc
HfTaTiW	bcc	bcc	MnNbReTi	bcc	none	MoPdPtRu	none	fcc
HfTaTiZr	bcc	bcc	MnNbReV	bcc	none	MoPdReRh	none	fcc
HfTaVW	bcc	bcc	MnNbReW	bcc	bcc	MoPdReRu	none	fcc

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
HfTaVZr	bcc	bcc	MnNbReZr	bcc	none	MoPdRhRu	none	fcc
HfTaWZr	bcc	bcc	MnNbTaTi	bcc	none	MoPtReRh	none	fcc
HfTiVW	bcc	bcc	MnNbTaV	bcc	none	MoPtReRu	none	fcc
HfTiVZr	bcc	none	MnNbTaW	none	none	MoPtRhRu	fcc	fcc
HfTiWZr	none	bcc	MnNbTaZr	none	none	MoReTaTi	bcc	none
HfVWZr	none	none	MnNbTiV	bcc	none	MoReTaV	bcc	none
IrMoNiPd	fcc	none	MnNbTiW	none	none	MoReTaW	bcc	bcc
IrMoNiPt	fcc	fcc	MnNbTiZr	none	none	MoReTaZr	none	none
IrMoOsPd	none	fcc	MnNbVW	bcc	none	MoReTiV	bcc	bcc
IrMoOsPt	none	fcc	MnNbVZr	none	none	MoReTiW	bcc	bcc
IrMoPdPt	fcc	fcc	MnNbWZr	none	none	MoReTiZr	bcc	bcc
IrMoPdRe	none	none	MnNiOsPd	none	none	MoReVW	bcc	bcc
IrMoPdRh	fcc	none	MnNiPdRe	none	none	MoReVZr	none	none
IrMoPdRu	none	none	MnNiPdRu	none	fcc	MoReWZr	bcc	bcc
IrMoPtRe	none	fcc	MnOsPdRe	none	none	MoTaTiV	bcc	bcc
IrMoPtRh	fcc	fcc	MnOsPdRu	none	fcc	MoTaTiW	bcc	bcc
IrMoPtRu	fcc	fcc	MnPdReRu	none	none	MoTaTiZr	bcc	bcc
IrNbPdPt	none	none	MnReTaTi	bcc	none	MoTaVW	bcc	bcc
IrNbPdRh	fcc	fcc	MnReTaV	bcc	none	MoTaVZr	none	none
IrNbPdRu	none	none	MnReTaW	none	none	MoTaWZr	none	bcc
IrNbPtRh	none	none	MnReTaZr	none	none	MoTiVW	bcc	bcc
IrNbPtRu	fcc	none	MnReTiV	none	none	MoTiVZr	bcc	bcc
IrNiOsPd	none	fcc	MnReTiW	none	none	MoTiWZr	bcc	bcc
IrNiOsPt	fcc	fcc	MnReTiZr	none	none	MoVWZr	bcc	none
IrNiPdPt	fcc	none	MnReVW	none	bcc	NbPdPtRh	none	fcc
IrNiPdRe	none	fcc	MnReVZr	none	none	NbPdRhRu	none	none
IrNiPdRh	fcc	fcc	MnReWZr	none	none	NbPtRhRu	none	none
IrNiPdRu	fcc	fcc	MnTaTiV	bcc	none	NbReTaTi	bcc	bcc
IrNiPtRe	fcc	fcc	MnTaTiW	none	none	NbReTaV	bcc	bcc
IrNiPtRh	fcc	fcc	MnTaVW	none	none	NbReTaW	bcc	bcc
IrNiPtRu	fcc	fcc	MnTaVZr	none	none	NbReTaZr	bcc	bcc
IrOsPdPt	fcc	none	MnTaWZr	none	none	NbReTiV	bcc	bcc

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
NbReTiW	bcc	bcc	NiOsPtRh	none	fcc	PdPtReRh	fcc	fcc
NbReTiZr	bcc	bcc	NiOsPtRu	none	fcc	PdPtReRu	fcc	fcc
NbReVW	bcc	bcc	NiPdPtRe	none	fcc	PdPtRhRu	fcc	fcc
NbReVZr	none	bcc	NiPdPtRh	fcc	fcc	PdReRhRu	none	none
NbReWZr	bcc	bcc	NiPdPtRu	fcc	fcc	PtReRhRu	fcc	fcc
NbTaTiV	bcc	bcc	NiPdReRh	none	fcc	ReTaTiV	bcc	bcc
NbTaTiW	bcc	bcc	NiPdReRu	none	fcc	ReTaTiW	bcc	bcc
NbTaTiZr	bcc	bcc	NiPdRhRu	none	fcc	ReTaTiZr	none	bcc
NbTaVW	bcc	bcc	NiPtReRh	none	fcc	ReTaVW	bcc	bcc
NbTaVZr	bcc	bcc	NiPtReRu	none	fcc	ReTaVZr	none	bcc
NbTaWZr	bcc	bcc	NiPtRhRu	fcc	fcc	ReTaWZr	bcc	none
NbTiVW	bcc	bcc	OsPdPtRe	fcc	fcc	ReTiVW	bcc	bcc
NbTiVZr	bcc	bcc	OsPdPtRh	fcc	fcc	ReTiVZr	none	none
NbTiWZr	bcc	bcc	OsPdPtRu	none	fcc	ReTiWZr	bcc	bcc
NbVWZr	bcc	none	OsPdReRh	fcc	fcc	ReVWZr	none	bcc
NiOsPdPt	none	none	OsPdReRu	none	fcc	TaTiVW	bcc	bcc
NiOsPdRe	none	fcc	OsPdRhRu	none	none	TaTiVZr	bcc	bcc
NiOsPdRh	none	fcc	OsPtReRh	fcc	fcc	TaTiWZr	bcc	bcc
NiOsPdRu	none	fcc	OsPtReRu	fcc	fcc	TaVWZr	bcc	none
NiOsPtRe	none	fcc	OsPtRhRu	fcc	fcc	TiVWZr	bcc	none
AgAlAuCo	none	none	AgAuNiRu	none	none	AgCrNiRh	none	none
AgAlAuCu	none	fcc	AgAuNiSi	none	none	AgCrNiRu	none	none
AgAlAuFe	none	none	AgAuPdPt	fcc	fcc	AgCrOsPt	none	none
AgAlAuNi	none	none	AgAuPdRe	fcc	fcc	AgCrPdPt	fcc	none
AgAlAuSi	none	none	AgAuPdRh	fcc	none	AgCrPdRh	fcc	none
AgAlAuZr	none	none	AgAuPdRu	fcc	fcc	AgCrPdRu	none	none
AgAlCoCu	none	none	AgAuPtRh	fcc	fcc	AgCrPtRh	none	none
AgAlCoNi	none	none	AgAuPtRu	fcc	fcc	AgCrPtRu	none	fcc
AgAlCuFe	none	none	AgAuRhRu	fcc	fcc	AgCrRhRu	none	fcc
AgAlCuMn	none	none	AgCoCrCu	none	none	AgCuFeIr	none	none
AgAlCuNi	none	none	AgCoCrNi	none	none	AgCuFeNi	none	none
AgAlCuSi	none	none	AgCoCrOs	none	none	AgCuFeOs	none	fcc
AgAlCuZr	none	none	AgCoCrPd	none	none	AgCuFePd	none	fcc
AgAlFeNi	none	none	AgCoCrPt	none	none	AgCuFeRe	none	none
AgAlNiSi	none	none	AgCoCrRh	none	none	AgCuFeRh	none	none

Supplementary Table 4, Continued

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
AgAuCoCr	none	none	AgCoCrRu	none	none	AgCuFeRu	none	none
AgAuCoCu	none	none	AgCoCuFe	none	none	AgCuIrPd	none	none
AgAuCoIr	none	fcc	AgCoCuPd	none	none	AgCuIrPt	none	none
AgAuCoMo	none	none	AgCoCuPt	none	none	AgCuNbPd	none	none
AgAuCoNi	none	fcc	AgCoIrMo	none	none	AgCuNiPd	none	none
AgAuCoPd	fcc	none	AgCoIrPd	fcc	fcc	AgCuNiPt	none	none
AgAuCoPt	none	none	AgCoIrPt	none	fcc	AgCuNiSi	none	none
AgAuCoRe	none	none	AgCoMoNi	none	none	AgCuPdPt	fcc	fcc
AgAuCoRh	none	fcc	AgCoMoOs	none	none	AgCuPdRe	none	none
AgAuCoRu	none	none	AgCoMoPd	none	fcc	AgCuPdRh	none	none
AgAuCrCu	none	none	AgCoMoPt	none	none	AgCuPdRu	none	none
AgAuCrNi	none	fcc	AgCoMoRe	none	none	AgCuPtRe	none	none
AgAuCuFe	none	fcc	AgCoMoRh	none	none	AgCuPtRh	fcc	none
AgAuCuIr	fcc	none	AgCoMoRu	none	none	AgCuPtRu	none	none
AgAuCuNi	none	fcc	AgCoNiPd	none	fcc	AgFeIrNi	none	none
AgAuCuPd	fcc	fcc	AgCoNiPt	none	none	AgFeIrRu	none	none
AgAuCuPt	fcc	none	AgCoOsPd	none	none	AgFeMoNi	none	none
AgAuCuRh	none	none	AgCoOsPt	none	none	AgFeMoOs	none	fcc
AgAuCuRu	fcc	none	AgCoPdPt	fcc	fcc	AgFeMoPd	none	fcc
AgAuCuSi	none	none	AgCoPdRe	none	fcc	AgFeMoRe	none	fcc
AgAuCuZr	none	none	AgCoPdRh	fcc	fcc	AgFeMoRh	none	none
AgAuFeMo	none	none	AgCoPdRu	none	none	AgFeMoRu	none	none
AgAuFeNi	none	none	AgCoPtRe	none	fcc	AgFeNiOs	none	none
AgAuFePd	none	fcc	AgCoPtRh	none	fcc	AgFeNiPd	none	fcc
AgAuFeRh	none	fcc	AgCoPtRu	none	none	AgFeNiRe	none	none
AgAuIrMo	none	none	AgCrCuFe	none	none	AgFeNiRh	none	none
AgAuIrNi	none	none	AgCrCuNi	none	none	AgFeNiRu	none	none
AgAuIrPd	fcc	fcc	AgCrCuPd	none	none	AgFeOsPd	none	none
AgAuIrPt	fcc	fcc	AgCrCuPt	none	none	AgFeOsRe	none	fcc
AgAuIrRh	fcc	none	AgCrCuRh	none	none	AgFeOsRh	none	none
AgAuIrRu	fcc	none	AgCrCuRu	none	none	AgFeOsRu	none	fcc
AgAuMoNi	none	none	AgCrFeNi	none	none	AgFePdRe	none	none
AgAuMoPd	fcc	fcc	AgCrIrNi	none	none	AgFePdRh	none	fcc
AgAuMoRh	none	none	AgCrIrPd	fcc	none	AgFePdRu	none	none
AgAuNbRh	none	none	AgCrIrRh	none	none	AgFeReRu	none	fcc
AgAuNiPd	none	none	AgCrIrRu	none	fcc	AgFeRhRu	none	none
AgAuNiPt	none	none	AgCrNiOs	none	none	AgIrMoNi	none	none
AgAuNiRe	none	fcc	AgCrNiPd	none	none	AgIrMoPd	none	none
AgAuNiRh	none	fcc	AgCrNiPt	none	none	AgIrMoRh	fcc	none

Supplementary Table 4, Continued.

	S.S(ML)	S.S.		S.S(ML)	S.S.		S.S(ML)	S.S.
AgIrNbPt	none	none	AlCrMnV	bcc	none	AuCuPtRu	fcc	fcc
AgIrNiPd	fcc	fcc	AlCrMnW	none	none	AuFeMoPd	none	none
AgIrNiPt	none	fcc	AlCrMoRe	bcc	bcc	AuFeNiPd	fcc	fcc
AgIrOsPd	fcc	fcc	AlCrMoTi	bcc	bcc	AuFeOsPd	fcc	none
AgIrOsPt	fcc	none	AlCrMoV	bcc	bcc	AuFePdRe	none	fcc
AgIrPdPt	fcc	none	AlCrMoW	bcc	bcc	AuFePdRh	fcc	none
AgIrPdRe	none	none	AlCrNbV	bcc	none	AuFePdRu	fcc	none
AgIrPdRh	fcc	none	AlCrReTi	none	none	AuIrMoPd	fcc	none
AgIrPdRu	none	fcc	AlCrReV	none	bcc	AuIrNiPd	fcc	fcc
AgIrPtRe	none	fcc	AlCrReW	bcc	bcc	AuIrNiPt	fcc	none
AgIrPtRh	fcc	fcc	AlCrTiV	bcc	none	AuIrOsPd	fcc	fcc
AgIrPtRu	fcc	fcc	AlCrTiW	bcc	bcc	AuIrOsPt	fcc	fcc
AgMoNiOs	none	none	AlCrVW	bcc	bcc	AuIrPdPt	fcc	fcc
AgMoNiPd	none	none	AlCuFeMn	none	none	AuIrPdRe	fcc	none
AgMoNiPt	none	none	AlCuFeNi	none	bcc	AuIrPdRh	fcc	fcc
AgMoNiRe	none	none	AlCuFeV	none	none	AuIrPdRu	fcc	fcc
AgMoNiRh	none	none	AlFeMnRe	none	none	AuIrPtRe	fcc	none
AgMoNiRu	none	fcc	AlFeMnTi	none	none	AuIrPtRh	fcc	fcc
AgMoOsPd	none	none	AlFeMnV	bcc	none	AuIrPtRu	fcc	fcc
AgMoOsRe	none	fcc	AlFeTiV	bcc	none	AuMoNiPd	none	fcc
AgMoOsRh	fcc	none	AlHfTiV	bcc	bcc	AuMoNiPt	fcc	none
AgMoPdPt	none	none	AlHfTiW	none	bcc	AuMoOsPd	fcc	none
AgMoPdRe	none	fcc	AlHfVW	none	none	AuMoPdPt	fcc	none
AgMoPdRh	none	none	AlMnMoRe	none	none	AuMoPdRe	none	fcc
AgMoPdRu	none	none	AlMnReW	none	none	AuMoPdRh	fcc	none
AgMoPtRe	none	none	AlMnTiV	bcc	bcc	AuMoPdRu	fcc	fcc
AgMoPtRh	none	none	AlMnVW	none	none	AuMoPtRe	none	none
AgMoPtRu	none	none	AlMoTiV	bcc	bcc	AuMoPtRh	fcc	fcc
AgMoReRh	none	none	AlMoVW	bcc	bcc	AuMoPtRu	none	none
AgMoRhRu	fcc	none	AlReVW	bcc	bcc	AuNbPdRh	none	none
AgNbPdPt	none	none	AlTiVW	bcc	bcc	AuNbPtRh	none	none
AgNbPdRh	none	none	AuCoCrPd	fcc	fcc	AuNiOsPd	none	fcc
AgNbPtRh	none	none	AuCoCrPt	none	none	AuNiOsPt	none	none
AgNiOsPd	none	fcc	AuCoCuPd	none	none	AuNiPdPt	fcc	none
AgNiOsPt	none	none	AuCoCuPt	none	none	AuNiPdRe	none	none
AgNiPdPt	fcc	fcc	AuCoIrPd	fcc	fcc	AuNiPdRh	fcc	none
AgNiPdRe	none	fcc	AuCoIrPt	fcc	none	AuNiPdRu	none	fcc
AgNiPdRh	fcc	fcc	AuCoMoPd	fcc	fcc	AuNiPtRe	none	fcc

Supplementary Table 4, Continued.

	<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>
AgNiPdRu	<b>none</b>	<b>none</b>	AuCoMoPt	none	none	AuNiPtRh	fcc	none
AgNiPtRe	<b>none</b>	<b>fcc</b>	AuCoNiPd	fcc	fcc	AuNiPtRu	none	none
AgNiPtRh	<b>none</b>	<b>fcc</b>	AuCoNiPt	fcc	none	AuOsPdPt	fcc	fcc
AgNiPtRu	<b>none</b>	<b>fcc</b>	AuCoOsPd	none	none	AuOsPdRe	none	fcc
AgOsPdPt	<b>fcc</b>	<b>none</b>	AuCoOsPt	none	none	AuOsPdRh	fcc	fcc
AgOsPdRe	<b>none</b>	<b>fcc</b>	AuCoPdPt	fcc	none	AuOsPtRh	fcc	fcc
AgOsPdRh	<b>fcc</b>	<b>fcc</b>	AuCoPdRe	none	none	AuOsPtRu	fcc	none
AgOsPdRu	<b>none</b>	<b>fcc</b>	AuCoPdRh	fcc	none	AuPdPtRe	fcc	fcc
AgOsPtRe	<b>none</b>	<b>fcc</b>	AuCoPdRu	none	none	AuPdPtRh	fcc	none
AgOsPtRh	<b>fcc</b>	<b>fcc</b>	AuCoPtRe	none	none	AuPdPtRu	fcc	fcc
AgOsPtRu	<b>fcc</b>	<b>fcc</b>	AuCoPtRh	fcc	none	AuPdReRh	fcc	fcc
AgPdPtRe	<b>none</b>	<b>fcc</b>	AuCoPtRu	none	none	AuPdReRu	none	fcc
AgPdPtRh	<b>fcc</b>	<b>none</b>	AuCrCuPd	fcc	none	AuPdRhRu	fcc	fcc
AgPdPtRu	<b>fcc</b>	<b>fcc</b>	AuCrCuPt	none	none	AuPtReRh	fcc	none
AgPdReRh	<b>none</b>	<b>fcc</b>	AuCrIrPd	fcc	none	AuPtReRu	fcc	fcc
AgPdReRu	<b>none</b>	<b>none</b>	AuCrIrPt	fcc	none	AuPtRhRu	fcc	fcc
AgPdRhRu	<b>none</b>	<b>fcc</b>	AuCrNiPd	fcc	none	CoCrCuNi	none	none
AgPtReRh	<b>none</b>	<b>fcc</b>	AuCrNiPt	none	none	CoCrCuPd	none	none
AgPtReRu	<b>none</b>	<b>fcc</b>	AuCrOsPt	none	none	CoCrCuPt	none	none
AgPtRhRu	<b>none</b>	<b>fcc</b>	AuCrPdPt	fcc	fcc	CoCrFeNi	fcc	fcc
AlNbTiV	<b>bcc</b>	<b>bcc</b>	AuCrPdRh	fcc	none	CoCrIrPd	fcc	none
AlNbVW	<b>bcc</b>	<b>bcc</b>	AuCrPdRu	none	none	CoCrIrPt	fcc	none
AlCoCrCu	<b>none</b>	<b>none</b>	AuCrPtRh	none	none	CoCrMnMo	none	none
AlCrCuFe	<b>none</b>	<b>none</b>	AuCrPtRu	none	none	CoCrMnNb	none	none
AlCrFeMn	<b>bcc</b>	<b>none</b>	AuCuFePd	none	fcc	CoCrMnNi	fcc	fcc
AlCrFeMo	<b>bcc</b>	<b>none</b>	AuCuIrPd	fcc	none	CoCrMoNb	none	bcc
AlCrFeNi	<b>none</b>	<b>none</b>	AuCuIrPt	fcc	none	CoCrMoW	none	bcc
AlCrFeRe	<b>none</b>	<b>none</b>	AuCuNiPd	fcc	none	CoCrNbW	none	bcc
AlCrFeTi	<b>bcc</b>	<b>none</b>	AuCuNiPt	fcc	none	CoCrNiPd	fcc	none
AlCrFeV	<b>bcc</b>	<b>none</b>	AuCuOsPt	fcc	none	CoCrNiPt	fcc	none
AlCrFeW	<b>none</b>	<b>none</b>	AuCuPdPt	fcc	none	CoCrOsPd	none	none
AlCrHfTi	<b>bcc</b>	<b>bcc</b>	AuCuPdRe	none	fcc	CoCrOsPt	none	none
AlCrHfW	<b>none</b>	<b>none</b>	AuCuPdRh	fcc	none	CoCrPdPt	fcc	none
AlCrMnMo	<b>bcc</b>	<b>none</b>	AuCuPdRu	fcc	fcc	CoCrPdRh	fcc	none
AlCrMnRe	<b>none</b>	<b>none</b>	AuCuPtRe	none	fcc	CoCrPdRu	none	none
AlCrMnTi	<b>bcc</b>	<b>none</b>	AuCuPtRh	fcc	none	CoCrPtRh	fcc	none
CoCrPtRu	<b>none</b>	<b>fcc</b>	CrCuMnV	none	none	CrHfReTi	none	none
CoCuIrPd	<b>fcc</b>	<b>none</b>	CrCuMnW	none	none	CrHfReV	none	none
CoCuIrPt	<b>fcc</b>	<b>none</b>	CrCuMoTa	none	none	CrHfReW	none	none

Supplementary Table 4, Continued.

	<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>
CoCuNiPd	<b>fcc</b>	<b>none</b>	CrCuMoTi	none	none	CrHfReZr	none	none
CoCuNiPt	<b>fcc</b>	<b>fcc</b>	CrCuMoV	none	none	CrHfTaTi	bcc	bcc
CoCuOsPd	<b>none</b>	<b>none</b>	CrCuMoW	none	none	CrHfTaV	none	bcc
CoCuOsPt	<b>none</b>	<b>none</b>	CrCuNbTa	none	none	CrHfTaW	none	bcc
CoCuPdPt	<b>fcc</b>	<b>fcc</b>	CrCuNbTi	none	none	CrHfTaZr	none	bcc
CoCuPdRe	<b>none</b>	<b>fcc</b>	CrCuNbV	none	none	CrHfTiV	bcc	bcc
CoCuPdRh	<b>fcc</b>	<b>none</b>	CrCuNbW	none	none	CrHfTiW	none	bcc
CoCuPdRu	<b>none</b>	<b>none</b>	CrCuNiPd	none	none	CrHfTiZr	bcc	bcc
CoCuPtRe	<b>none</b>	<b>fcc</b>	CrCuNiPt	<b>fcc</b>	none	CrHfVW	none	none
CoCuPtRh	<b>fcc</b>	<b>fcc</b>	CrCuOsPd	none	<b>fcc</b>	CrHfVZr	bcc	none
CoCuPtRu	<b>fcc</b>	<b>none</b>	CrCuOsPt	none	<b>fcc</b>	CrHfWZr	none	bcc
CoFeMnNi	<b>fcc</b>	<b>fcc</b>	CrCuPdPt	<b>fcc</b>	none	CrIrNiPd	<b>fcc</b>	none
CoIrMoPd	<b>fcc</b>	<b>none</b>	CrCuPdRh	<b>fcc</b>	none	CrIrNiPt	<b>fcc</b>	none
CoIrMoPt	<b>fcc</b>	<b>none</b>	CrCuPdRu	none	none	CrIrOsPd	none	none
CoIrNiPd	<b>fcc</b>	<b>none</b>	CrCuPtRh	<b>fcc</b>	none	CrIrOsPt	none	<b>fcc</b>
CoIrNiPt	<b>fcc</b>	<b>none</b>	CrCuPtRu	none	none	CrIrPdPt	<b>fcc</b>	none
CoIrOsPd	<b>none</b>	<b>fcc</b>	CrCuTaTi	none	none	CrIrPdRh	<b>fcc</b>	<b>fcc</b>
CoIrOsPt	<b>fcc</b>	<b>fcc</b>	CrCuTaV	none	none	CrIrPdRu	none	none
CoIrPdPt	<b>fcc</b>	<b>none</b>	CrCuTaW	none	none	CrIrPtRh	<b>fcc</b>	<b>fcc</b>
CoIrPdRe	<b>none</b>	<b>fcc</b>	CrCuTiV	none	none	CrIrPtRu	none	<b>fcc</b>
CoIrPdRh	<b>fcc</b>	<b>none</b>	CrCuTiW	none	none	CrMnMoNb	bcc	none
CoIrPdRu	<b>none</b>	<b>none</b>	CrCuVW	none	none	CrMnMoNi	none	none
CoIrPtRe	<b>fcc</b>	<b>fcc</b>	CrFeMnMo	bcc	none	CrMnMoRe	none	bcc
CoIrPtRh	<b>fcc</b>	<b>fcc</b>	CrFeMnNb	none	none	CrMnMoTa	bcc	none
CoIrPtRu	<b>fcc</b>	<b>none</b>	CrFeMnNi	<b>fcc</b>	none	CrMnMoTi	none	none
CoMoNbW	<b>none</b>	<b>bcc</b>	CrFeMnRe	none	bcc	CrMnMoV	bcc	none
CoMoNiPd	<b>fcc</b>	<b>none</b>	CrFeMnTa	none	none	CrMnMoW	none	bcc
CoMoNiPt	<b>fcc</b>	<b>none</b>	CrFeMnTi	none	none	CrMnMoZr	none	none
CoMoOsPd	<b>none</b>	<b>none</b>	CrFeMnV	bcc	bcc	CrMnNbNi	none	none
CoMoOsPt	<b>none</b>	<b>fcc</b>	CrFeMnW	none	bcc	CrMnNbRe	none	none
CoMoPdPt	<b>fcc</b>	<b>none</b>	CrFeMoNb	none	none	CrMnNbTa	bcc	none
CoMoPdRe	<b>none</b>	<b>fcc</b>	CrFeMoNi	none	bcc	CrMnNbTi	bcc	none
CoMoPdRh	<b>fcc</b>	<b>none</b>	CrFeMoRe	none	bcc	CrMnNbV	bcc	none
CoMoPdRu	<b>none</b>	<b>none</b>	CrFeMoTa	none	none	CrMnNbW	none	none
CoMoPtRe	<b>none</b>	<b>fcc</b>	CrFeMoTi	bcc	none	CrMnNbZr	none	none
CoMoPtRh	<b>fcc</b>	<b>fcc</b>	CrFeMoV	bcc	bcc	CrMnNiRe	none	bcc
CoMoPtRu	<b>none</b>	<b>fcc</b>	CrFeMoW	none	bcc	CrMnNiV	none	none
CoNiOsPd	<b>none</b>	<b>none</b>	CrFeNbRe	none	none	CrMnReTa	none	none
CoNiOsPt	<b>none</b>	<b>none</b>	CrFeNbTa	none	none	CrMnReTi	none	none

Supplementary Table 4, Continued.

	<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>
CoNiPdPt	<b>fcc</b>	<b>none</b>	CrFeNbTi	none	none	CrMnReV	none	bcc
CoNiPdRe	<b>none</b>	<b>fcc</b>	CrFeNbV	none	none	CrMnReW	none	bcc
CoNiPdRh	<b>fcc</b>	<b>none</b>	CrFeNbW	none	bcc	CrMnReZr	none	none
CoNiPdRu	<b>none</b>	<b>none</b>	CrFeNiPd	fcc	none	CrMnTaTi	bcc	none
CoNiPtRe	<b>none</b>	<b>fcc</b>	CrFeNiRe	none	bcc	CrMnTaV	bcc	none
CoNiPtRh	<b>fcc</b>	<b>fcc</b>	CrFeNiV	none	bcc	CrMnTaW	none	none
CoNiPtRu	<b>fcc</b>	<b>none</b>	CrFeOsPd	none	fcc	CrMnTaZr	none	none
CoOsPdPt	<b>none</b>	<b>none</b>	CrFePdRu	none	fcc	CrMnTiV	none	none
CoOsPdRe	<b>none</b>	<b>fcc</b>	CrFeReTa	none	none	CrMnTiW	none	none
CoOsPdRh	<b>none</b>	<b>fcc</b>	CrFeReTi	none	none	CrMnTiZr	none	none
CoOsPdRu	<b>none</b>	<b>fcc</b>	CrFeReV	none	bcc	CrMnVW	none	bcc
CoOsPtRe	<b>none</b>	<b>fcc</b>	CrFeReW	none	bcc	CrMnWZr	none	none
CoOsPtRh	<b>none</b>	<b>fcc</b>	CrFeTaTi	none	none	CrMoNbNi	none	none
CoOsPtRu	<b>none</b>	<b>fcc</b>	CrFeTaV	none	none	CrMoNbRe	bcc	bcc
CoPdPtRe	<b>none</b>	<b>fcc</b>	CrFeTaW	none	none	CrMoNbTa	none	bcc
CoPdPtRh	<b>fcc</b>	<b>fcc</b>	CrFeTiV	none	none	CrMoNbTi	bcc	bcc
CoPdPtRu	<b>fcc</b>	<b>fcc</b>	CrFeTiW	none	none	CrMoNbV	none	bcc
CoPdReRh	<b>none</b>	<b>none</b>	CrFeVW	none	bcc	CrMoNbW	none	none
CoPdReRu	<b>none</b>	<b>fcc</b>	CrHfMoNb	bcc	bcc	CrMoNbZr	none	none
CoPdRhRu	<b>none</b>	<b>none</b>	CrHfMoRe	none	none	CrMoNiRe	none	bcc
CoPtReRh	<b>none</b>	<b>fcc</b>	CrHfMoTa	none	bcc	CrMoReTa	none	none
CoPtReRu	<b>none</b>	<b>fcc</b>	CrHfMoTi	bcc	bcc	CrMoReTi	bcc	bcc
CoPtRhRu	<b>fcc</b>	<b>fcc</b>	CrHfMoV	bcc	none	CrMoReV	bcc	bcc
CrCuFeMn	<b>none</b>	<b>none</b>	CrHfMoW	none	bcc	CrMoReW	bcc	bcc
CrCuFeMo	<b>none</b>	<b>none</b>	CrHfMoZr	none	bcc	CrMoReZr	none	none
CrCuFeNb	<b>none</b>	<b>none</b>	CrHfNbRe	none	none	CrMoTaTi	bcc	bcc
CrCuFeNi	<b>none</b>	<b>none</b>	CrHfNbTa	bcc	bcc	CrMoTaV	bcc	bcc
CrCuFePd	<b>none</b>	<b>none</b>	CrHfNbTi	bcc	bcc	CrMoTaW	none	bcc
CrCuFeV	<b>none</b>	<b>none</b>	CrHfNbV	bcc	none	CrMoTaZr	none	none
CrCuFeW	<b>none</b>	<b>none</b>	CrHfNbW	none	bcc	CrMoTiV	bcc	bcc
CrCulrPd	<b>fcc</b>	<b>none</b>	CrHfNbZr	none	bcc	CrMoTiW	bcc	bcc
CrCulrPt	<b>none</b>	<b>none</b>	CrHfReTa	none	none	CrMoTiZr	bcc	none
CrMoVW	<b>bcc</b>	<b>bcc</b>	CuFeVW	none	none	FeMnMoV	bcc	none
CrMoVZr	<b>none</b>	<b>none</b>	CulrNiPd	fcc	none	FeMnMoW	none	none
CrMoWZr	<b>none</b>	<b>none</b>	CulrNiPt	fcc	none	FeMnNbRe	none	none
CrNbNiV	<b>none</b>	<b>none</b>	CulrOsPd	none	fcc	FeMnNbTa	none	none
CrNbReTa	<b>bcc</b>	<b>none</b>	CulrOsPt	fcc	none	FeMnNbTi	none	none
CrNbReTi	<b>bcc</b>	<b>bcc</b>	CulrPdPt	fcc	none	FeMnNbV	none	none

Supplementary Table 4, Continued.

	<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>
CrNbReV	<b>bcc</b>	<b>none</b>	CuIrPdRe	none	none	FeMnNbW	none	none
CrNbReW	<b>bcc</b>	<b>bcc</b>	CuIrPdRh	fcc	none	FeMnReTa	none	none
CrNbReZr	<b>none</b>	<b>none</b>	CuIrPdRu	none	none	FeMnReTi	none	bcc
CrNbTaTi	<b>bcc</b>	<b>bcc</b>	CuIrPtRe	fcc	none	FeMnReV	none	bcc
CrNbTaV	<b>bcc</b>	<b>bcc</b>	CuIrPtRh	fcc	none	FeMnReW	none	bcc
CrNbTaW	<b>none</b>	<b>bcc</b>	CuIrPtRu	fcc	none	FeMnTaTi	none	none
CrNbTaZr	<b>none</b>	<b>none</b>	CuMnMoTa	none	none	FeMnTaV	none	none
CrNbTiV	<b>bcc</b>	<b>bcc</b>	CuMnMoV	none	none	FeMnTaW	none	none
CrNbTiW	<b>bcc</b>	<b>bcc</b>	CuMnMoW	none	none	FeMnTiV	none	none
CrNbTiZr	<b>none</b>	<b>none</b>	CuMnNbTa	none	none	FeMnTiW	none	none
CrNbVW	<b>bcc</b>	<b>bcc</b>	CuMnNbV	none	none	FeMnVW	none	bcc
CrNbVZr	<b>none</b>	<b>none</b>	CuMnTaV	none	none	FeMoNbRe	bcc	none
CrNbWZr	<b>none</b>	<b>none</b>	CuMnTaW	none	none	FeMoNbTa	none	none
CrNiOsPd	<b>none</b>	<b>fcc</b>	CuMnTiV	none	none	FeMoNbTi	none	none
CrNiOsPt	<b>none</b>	<b>none</b>	CuMnVW	none	none	FeMoNbV	bcc	bcc
CrNiPdPt	<b>fcc</b>	<b>none</b>	CuMoNbTa	none	none	FeMoNbW	none	none
CrNiPdRh	<b>fcc</b>	<b>none</b>	CuMoNbV	none	none	FeMoNiPd	fcc	fcc
CrNiPdRu	<b>none</b>	<b>none</b>	CuMoNbW	none	none	FeMoNiV	none	none
CrNiPtRh	<b>fcc</b>	<b>none</b>	CuMoTaTi	none	none	FeMoOsPd	none	none
CrNiPtRu	<b>none</b>	<b>none</b>	CuMoTaW	none	none	FeMoPdRe	none	fcc
CrNiReV	<b>none</b>	<b>bcc</b>	CuMoTiV	none	none	FeMoPdRh	fcc	none
CrOsPdPt	<b>none</b>	<b>none</b>	CuMoVW	none	none	FeMoPdRu	none	fcc
CrOsPdRh	<b>none</b>	<b>none</b>	CuNbPdPt	none	none	FeMoReTa	none	none
CrOsPdRu	<b>none</b>	<b>none</b>	CuNbPdRh	none	fcc	FeMoReTi	none	bcc
CrOsPtRh	<b>none</b>	<b>fcc</b>	CuNbPtRh	none	none	FeMoReV	none	bcc
CrOsPtRu	<b>fcc</b>	<b>none</b>	CuNbTaTi	none	none	FeMoReW	none	bcc
CrPdPtRh	<b>fcc</b>	<b>none</b>	CuNbTaV	none	none	FeMoTaTi	none	none
CrPdPtRu	<b>none</b>	<b>none</b>	CuNbTaW	none	none	FeMoTaV	none	none
CrPdRhRu	<b>none</b>	<b>none</b>	CuNbTiV	none	bcc	FeMoTaW	none	none
CrPtRhRu	<b>none</b>	<b>fcc</b>	CuNbVW	none	none	FeMoTiV	bcc	none
CrReTaTi	<b>bcc</b>	<b>none</b>	CuNiOsPd	none	fcc	FeMoTiW	none	none
CrReTaV	<b>bcc</b>	<b>none</b>	CuNiOsPt	none	fcc	FeMoVW	none	none
CrReTaW	<b>bcc</b>	<b>none</b>	CuNiPdPt	fcc	fcc	FeNbReTa	none	none
CrReTaZr	<b>none</b>	<b>none</b>	CuNiPdRe	none	fcc	FeNbReTi	none	none
CrReTiV	<b>none</b>	<b>bcc</b>	CuNiPdRh	fcc	none	FeNbReV	bcc	none
CrReTiW	<b>bcc</b>	<b>none</b>	CuNiPdRu	none	none	FeNbReW	none	none
CrReTiZr	<b>none</b>	<b>none</b>	CuNiPtRe	none	fcc	FeNbTaTi	none	none
CrReVW	<b>bcc</b>	<b>bcc</b>	CuNiPtRh	fcc	fcc	FeNbTaV	none	bcc
CrReVZr	<b>none</b>	<b>none</b>	CuNiPtRu	fcc	fcc	FeNbTaW	none	none

Supplementary Table 4, Continued.

	<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>
CrReWZr	<b>none</b>	<b>none</b>	CuOsPdPt	fcc	fcc	FeNbTiV	none	none
CrTaTiV	<b>bcc</b>	<b>bcc</b>	CuOsPdRe	none	none	FeNbTiW	none	none
CrTaTiW	<b>bcc</b>	<b>bcc</b>	CuOsPdRh	fcc	none	FeNbVW	none	bcc
CrTaTiZr	<b>none</b>	<b>none</b>	CuOsPdRu	none	none	FeNiOsPd	none	none
CrTaVW	<b>none</b>	<b>bcc</b>	CuOsPtRe	none	fcc	FeNiPdRe	none	fcc
CrTaVZr	<b>none</b>	<b>none</b>	CuOsPtRh	fcc	fcc	FeNiPdRh	fcc	none
CrTaWZr	<b>none</b>	<b>none</b>	CuOsPtRu	fcc	fcc	FeNiPdRu	none	none
CrTiVW	<b>bcc</b>	<b>bcc</b>	CuPdPtRe	none	fcc	FeOsPdRe	none	fcc
CrTiVZr	<b>none</b>	<b>none</b>	CuPdPtRh	fcc	none	FeOsPdRh	none	fcc
CrTiWZr	<b>none</b>	<b>none</b>	CuPdPtRu	fcc	fcc	FeOsPdRu	none	fcc
CrVWZr	<b>none</b>	<b>none</b>	CuPdReRh	none	fcc	FePdReRh	none	none
CuFeIrPd	<b>fcc</b>	<b>none</b>	CuPdReRu	none	none	FePdReRu	none	fcc
CuFeMnNb	<b>none</b>	<b>none</b>	CuPdRhRu	none	fcc	FePdRhRu	none	none
CuFeMnTa	<b>none</b>	<b>none</b>	CuPtReRh	none	fcc	FeReTaTi	none	none
CuFeMnTi	<b>none</b>	<b>none</b>	CuPtReRu	none	fcc	FeReTaV	none	none
CuFeMnV	<b>none</b>	<b>none</b>	CuPtRhRu	fcc	fcc	FeReTiV	none	none
CuFeMoTi	<b>none</b>	<b>none</b>	CuTaTiV	bcc	none	FeReVW	none	bcc
CuFeMoV	<b>none</b>	<b>none</b>	CuTaTiW	none	none	FeTaTiV	none	none
CuFeNbTa	<b>none</b>	<b>none</b>	CuTaVW	none	none	FeTaTiW	none	none
CuFeNbTi	<b>none</b>	<b>none</b>	CuTiVW	none	none	FeTaVW	none	none
CuFeNbV	<b>none</b>	<b>none</b>	FelrMoPd	fcc	none	FeTiVW	none	none
CuFeNiPd	<b>fcc</b>	<b>none</b>	FelrNiPd	fcc	fcc	HfMoNbRe	bcc	none
CuFeNiV	<b>none</b>	<b>none</b>	FelrOsPd	none	fcc	HfMoNbTa	bcc	bcc
CuFeOsPd	<b>none</b>	<b>fcc</b>	FelrPdRe	none	none	HfMoNbTi	bcc	bcc
CuFePdRe	<b>none</b>	<b>fcc</b>	FelrPdRu	none	fcc	HfMoNbV	bcc	bcc
CuFePdRh	<b>fcc</b>	<b>none</b>	FeMnMoNb	none	none	HfMoNbW	bcc	bcc
CuFePdRu	<b>none</b>	<b>none</b>	FeMnMoRe	none	bcc	HfMoNbZr	bcc	bcc
CuFeTaV	<b>none</b>	<b>none</b>	FeMnMoTa	none	none	HfMoReTa	none	none
CuFeTiV	<b>none</b>	<b>none</b>	FeMnMoTi	none	none	HfMoReTi	bcc	bcc
HfMoReV	<b>bcc</b>	<b>none</b>	IrOsPdRe	none	none	MnTiVW	bcc	none
HfMoReW	<b>none</b>	<b>none</b>	IrOsPdRh	fcc	none	MnTiVZr	none	none
HfMoReZr	<b>none</b>	<b>none</b>	IrOsPdRu	none	none	MnTiWZr	none	none
HfMoTaTi	<b>bcc</b>	<b>bcc</b>	IrOsPtRe	none	fcc	MnVWZr	none	none
HfMoTaV	<b>none</b>	<b>bcc</b>	IrOsPtRh	fcc	fcc	MoNbNiV	none	bcc
HfMoTaW	<b>bcc</b>	<b>bcc</b>	IrOsPtRu	fcc	fcc	MoNbReTa	bcc	none
HfMoTaZr	<b>bcc</b>	<b>bcc</b>	IrPdPtRe	fcc	fcc	MoNbReTi	bcc	bcc
HfMoTiV	<b>bcc</b>	<b>bcc</b>	IrPdPtRh	fcc	fcc	MoNbReV	bcc	bcc
HfMoTiW	<b>bcc</b>	<b>bcc</b>	IrPdPtRu	fcc	none	MoNbReW	bcc	bcc

Supplementary Table 4, Continued.

	<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>
HfMoTiZr	<b>bcc</b>	<b>bcc</b>	IrPdReRh	<b>fcc</b>	<b>none</b>	MoNbReZr	<b>bcc</b>	<b>none</b>
HfMoVW	<b>none</b>	<b>bcc</b>	IrPdReRu	<b>none</b>	<b>none</b>	MoNbTaTi	<b>bcc</b>	<b>bcc</b>
HfMoVZr	<b>none</b>	<b>bcc</b>	IrPdRhRu	<b>fcc</b>	<b>none</b>	MoNbTaV	<b>bcc</b>	<b>bcc</b>
HfMoWZr	<b>none</b>	<b>bcc</b>	IrPtReRh	<b>fcc</b>	<b>fcc</b>	MoNbTaW	<b>bcc</b>	<b>bcc</b>
HfNbReTa	<b>bcc</b>	<b>bcc</b>	IrPtReRu	<b>fcc</b>	<b>fcc</b>	MoNbTaZr	<b>bcc</b>	<b>none</b>
HfNbReTi	<b>bcc</b>	<b>bcc</b>	IrPtRhRu	<b>fcc</b>	<b>fcc</b>	MoNbTiV	<b>bcc</b>	<b>bcc</b>
HfNbReV	<b>bcc</b>	<b>bcc</b>	MnMoNbRe	<b>none</b>	<b>bcc</b>	MoNbTiW	<b>bcc</b>	<b>bcc</b>
HfNbReW	<b>bcc</b>	<b>bcc</b>	MnMoNbTa	<b>bcc</b>	<b>none</b>	MoNbTiZr	<b>bcc</b>	<b>bcc</b>
HfNbReZr	<b>none</b>	<b>bcc</b>	MnMoNbTi	<b>bcc</b>	<b>none</b>	MoNbVW	<b>bcc</b>	<b>bcc</b>
HfNbTaTi	<b>bcc</b>	<b>bcc</b>	MnMoNbV	<b>bcc</b>	<b>none</b>	MoNbVZr	<b>bcc</b>	<b>bcc</b>
HfNbTaV	<b>bcc</b>	<b>bcc</b>	MnMoNbW	<b>none</b>	<b>bcc</b>	MoNbWZr	<b>none</b>	<b>bcc</b>
HfNbTaW	<b>bcc</b>	<b>bcc</b>	MnMoNbZr	<b>none</b>	<b>none</b>	MoNiOsPd	<b>none</b>	<b>fcc</b>
HfNbTaZr	<b>bcc</b>	<b>bcc</b>	MnMoNiV	<b>none</b>	<b>none</b>	MoNiOsPt	<b>none</b>	<b>fcc</b>
HfNbTiV	<b>bcc</b>	<b>bcc</b>	MnMoReTa	<b>none</b>	<b>none</b>	MoNiPdPt	<b>fcc</b>	<b>fcc</b>
HfNbTiW	<b>bcc</b>	<b>bcc</b>	MnMoReTi	<b>none</b>	<b>none</b>	MoNiPdRe	<b>none</b>	<b>fcc</b>
HfNbTiZr	<b>bcc</b>	<b>bcc</b>	MnMoReV	<b>bcc</b>	<b>bcc</b>	MoNiPdRh	<b>fcc</b>	<b>none</b>
HfNbVW	<b>none</b>	<b>bcc</b>	MnMoReW	<b>none</b>	<b>bcc</b>	MoNiPdRu	<b>none</b>	<b>none</b>
HfNbVZr	<b>bcc</b>	<b>bcc</b>	MnMoReZr	<b>none</b>	<b>none</b>	MoNiPtRe	<b>none</b>	<b>fcc</b>
HfNbWZr	<b>bcc</b>	<b>bcc</b>	MnMoTaTi	<b>bcc</b>	<b>none</b>	MoNiPtRh	<b>fcc</b>	<b>fcc</b>
HfReTaTi	<b>bcc</b>	<b>bcc</b>	MnMoTaV	<b>bcc</b>	<b>none</b>	MoNiPtRu	<b>none</b>	<b>fcc</b>
HfReTaV	<b>bcc</b>	<b>bcc</b>	MnMoTaW	<b>none</b>	<b>none</b>	MoNiReV	<b>none</b>	<b>bcc</b>
HfReTaW	<b>none</b>	<b>none</b>	MnMoTaZr	<b>none</b>	<b>none</b>	MoOsPdPt	<b>none</b>	<b>fcc</b>
HfReTaZr	<b>none</b>	<b>none</b>	MnMoTiV	<b>bcc</b>	<b>none</b>	MoOsPdRe	<b>none</b>	<b>fcc</b>
HfReTiV	<b>none</b>	<b>bcc</b>	MnMoTiW	<b>none</b>	<b>none</b>	MoOsPdRh	<b>none</b>	<b>fcc</b>
HfReTiW	<b>none</b>	<b>bcc</b>	MnMoTiZr	<b>none</b>	<b>none</b>	MoOsPdRu	<b>none</b>	<b>fcc</b>
HfReTiZr	<b>none</b>	<b>bcc</b>	MnMoVW	<b>bcc</b>	<b>bcc</b>	MoOsPtRe	<b>none</b>	<b>fcc</b>
HfReVW	<b>bcc</b>	<b>bcc</b>	MnMoVZr	<b>none</b>	<b>none</b>	MoOsPtRh	<b>fcc</b>	<b>fcc</b>
HfReVZr	<b>none</b>	<b>none</b>	MnMoWZr	<b>none</b>	<b>none</b>	MoOsPtRu	<b>none</b>	<b>fcc</b>
HfReWZr	<b>none</b>	<b>bcc</b>	MnNbNiV	<b>none</b>	<b>none</b>	MoPdPtRe	<b>none</b>	<b>fcc</b>
HfTaTiV	<b>bcc</b>	<b>bcc</b>	MnNbReTa	<b>none</b>	<b>none</b>	MoPdPtRh	<b>fcc</b>	<b>fcc</b>
HfTaTiW	<b>bcc</b>	<b>bcc</b>	MnNbReTi	<b>bcc</b>	<b>none</b>	MoPdPtRu	<b>none</b>	<b>fcc</b>
HfTaTiZr	<b>bcc</b>	<b>bcc</b>	MnNbReV	<b>bcc</b>	<b>none</b>	MoPdReRh	<b>none</b>	<b>fcc</b>
HfTaVW	<b>bcc</b>	<b>bcc</b>	MnNbReW	<b>bcc</b>	<b>bcc</b>	MoPdReRu	<b>none</b>	<b>fcc</b>
HfTaVZr	<b>bcc</b>	<b>bcc</b>	MnNbReZr	<b>bcc</b>	<b>none</b>	MoPdRhRu	<b>none</b>	<b>fcc</b>
HfTaWZr	<b>bcc</b>	<b>bcc</b>	MnNbTaTi	<b>bcc</b>	<b>none</b>	MoPtReRh	<b>none</b>	<b>fcc</b>
HfTiVW	<b>bcc</b>	<b>bcc</b>	MnNbTaV	<b>bcc</b>	<b>none</b>	MoPtReRu	<b>none</b>	<b>fcc</b>
HfTiVZr	<b>bcc</b>	<b>none</b>	MnNbTaW	<b>none</b>	<b>none</b>	MoPtRhRu	<b>fcc</b>	<b>fcc</b>
HfTiWZr	<b>none</b>	<b>bcc</b>	MnNbTaZr	<b>none</b>	<b>none</b>	MoReTaTi	<b>bcc</b>	<b>none</b>
HfVWZr	<b>none</b>	<b>none</b>	MnNbTiV	<b>bcc</b>	<b>none</b>	MoReTaV	<b>bcc</b>	<b>none</b>

Supplementary Table 4, Continued.

	<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>
IrMoNiPd	<b>fcc</b>	<b>none</b>	MnNbTiW	<b>none</b>	<b>none</b>	MoReTaW	<b>bcc</b>	<b>bcc</b>
IrMoNiPt	<b>fcc</b>	<b>fcc</b>	MnNbTiZr	<b>none</b>	<b>none</b>	MoReTaZr	<b>none</b>	<b>none</b>
IrMoOsPd	<b>none</b>	<b>fcc</b>	MnNbVW	<b>bcc</b>	<b>none</b>	MoReTiV	<b>bcc</b>	<b>bcc</b>
IrMoOsPt	<b>none</b>	<b>fcc</b>	MnNbVZr	<b>none</b>	<b>none</b>	MoReTiW	<b>bcc</b>	<b>bcc</b>
IrMoPdPt	<b>fcc</b>	<b>fcc</b>	MnNbWZr	<b>none</b>	<b>none</b>	MoReTiZr	<b>bcc</b>	<b>bcc</b>
IrMoPdRe	<b>none</b>	<b>none</b>	MnNiOsPd	<b>none</b>	<b>none</b>	MoReVW	<b>bcc</b>	<b>bcc</b>
IrMoPdRh	<b>fcc</b>	<b>none</b>	MnNiPdRe	<b>none</b>	<b>none</b>	MoReVZr	<b>none</b>	<b>none</b>
IrMoPdRu	<b>none</b>	<b>none</b>	MnNiPdRu	<b>none</b>	<b>fcc</b>	MoReWZr	<b>bcc</b>	<b>bcc</b>
IrMoPtRe	<b>none</b>	<b>fcc</b>	MnOsPdRe	<b>none</b>	<b>none</b>	MoTaTiV	<b>bcc</b>	<b>bcc</b>
IrMoPtRh	<b>fcc</b>	<b>fcc</b>	MnOsPdRu	<b>none</b>	<b>fcc</b>	MoTaTiW	<b>bcc</b>	<b>bcc</b>
IrMoPtRu	<b>fcc</b>	<b>fcc</b>	MnPdReRu	<b>none</b>	<b>none</b>	MoTaTiZr	<b>bcc</b>	<b>bcc</b>
IrNbPdPt	<b>none</b>	<b>none</b>	MnReTaTi	<b>bcc</b>	<b>none</b>	MoTaVW	<b>bcc</b>	<b>bcc</b>
IrNbPdRh	<b>fcc</b>	<b>fcc</b>	MnReTaV	<b>bcc</b>	<b>none</b>	MoTaVZr	<b>none</b>	<b>none</b>
IrNbPdRu	<b>none</b>	<b>none</b>	MnReTaW	<b>none</b>	<b>none</b>	MoTaWZr	<b>none</b>	<b>bcc</b>
IrNbPtRh	<b>none</b>	<b>none</b>	MnReTaZr	<b>none</b>	<b>none</b>	MoTiVW	<b>bcc</b>	<b>bcc</b>
IrNbPtRu	<b>fcc</b>	<b>none</b>	MnReTiV	<b>none</b>	<b>none</b>	MoTiVZr	<b>bcc</b>	<b>bcc</b>
IrNiOsPd	<b>none</b>	<b>fcc</b>	MnReTiW	<b>none</b>	<b>none</b>	MoTiWZr	<b>bcc</b>	<b>bcc</b>
IrNiOsPt	<b>fcc</b>	<b>fcc</b>	MnReTiZr	<b>none</b>	<b>none</b>	MoVWZr	<b>bcc</b>	<b>none</b>
IrNiPdPt	<b>fcc</b>	<b>none</b>	MnReVW	<b>none</b>	<b>bcc</b>	NbPdPtRh	<b>none</b>	<b>fcc</b>
IrNiPdRe	<b>none</b>	<b>fcc</b>	MnReVZr	<b>none</b>	<b>none</b>	NbPdRhRu	<b>none</b>	<b>none</b>
IrNiPdRh	<b>fcc</b>	<b>fcc</b>	MnReWZr	<b>none</b>	<b>none</b>	NbPtRhRu	<b>none</b>	<b>none</b>
IrNiPdRu	<b>fcc</b>	<b>fcc</b>	MnTaTiV	<b>bcc</b>	<b>none</b>	NbReTaTi	<b>bcc</b>	<b>bcc</b>
IrNiPtRe	<b>fcc</b>	<b>fcc</b>	MnTaTiW	<b>none</b>	<b>none</b>	NbReTaV	<b>bcc</b>	<b>bcc</b>
IrNiPtRh	<b>fcc</b>	<b>fcc</b>	MnTaVW	<b>none</b>	<b>none</b>	NbReTaW	<b>bcc</b>	<b>bcc</b>
IrNiPtRu	<b>fcc</b>	<b>fcc</b>	MnTaVZr	<b>none</b>	<b>none</b>	NbReTaZr	<b>bcc</b>	<b>bcc</b>
IrOsPdPt	<b>fcc</b>	<b>none</b>	MnTaWZr	<b>none</b>	<b>none</b>	NbReTiV	<b>bcc</b>	<b>bcc</b>
NbReTiW	<b>bcc</b>	<b>bcc</b>	NiOsPtRh	<b>none</b>	<b>fcc</b>	PdPtReRh	<b>fcc</b>	<b>fcc</b>
NbReTiZr	<b>bcc</b>	<b>bcc</b>	NiOsPtRu	<b>none</b>	<b>fcc</b>	PdPtReRu	<b>fcc</b>	<b>fcc</b>
NbReVW	<b>bcc</b>	<b>bcc</b>	NiPdPtRe	<b>none</b>	<b>fcc</b>	PdPtRhRu	<b>fcc</b>	<b>fcc</b>
NbReVZr	<b>none</b>	<b>bcc</b>	NiPdPtRh	<b>fcc</b>	<b>fcc</b>	PdReRhRu	<b>none</b>	<b>none</b>
NbReWZr	<b>bcc</b>	<b>bcc</b>	NiPdPtRu	<b>fcc</b>	<b>fcc</b>	PtReRhRu	<b>fcc</b>	<b>fcc</b>
NbTaTiV	<b>bcc</b>	<b>bcc</b>	NiPdReRh	<b>none</b>	<b>fcc</b>	ReTaTiV	<b>bcc</b>	<b>bcc</b>
NbTaTiW	<b>bcc</b>	<b>bcc</b>	NiPdReRu	<b>none</b>	<b>fcc</b>	ReTaTiW	<b>bcc</b>	<b>bcc</b>
NbTaTiZr	<b>bcc</b>	<b>bcc</b>	NiPdRhRu	<b>none</b>	<b>fcc</b>	ReTaTiZr	<b>none</b>	<b>bcc</b>
NbTaVW	<b>bcc</b>	<b>bcc</b>	NiPtReRh	<b>none</b>	<b>fcc</b>	ReTaVW	<b>bcc</b>	<b>bcc</b>
NbTaVZr	<b>bcc</b>	<b>bcc</b>	NiPtReRu	<b>none</b>	<b>fcc</b>	ReTaVZr	<b>none</b>	<b>bcc</b>
NbTaWZr	<b>bcc</b>	<b>bcc</b>	NiPtRhRu	<b>fcc</b>	<b>fcc</b>	ReTaWZr	<b>bcc</b>	<b>none</b>
NbTiVW	<b>bcc</b>	<b>bcc</b>	OsPdPtRe	<b>fcc</b>	<b>fcc</b>	ReTiVW	<b>bcc</b>	<b>bcc</b>

Supplementary Table 4, Continued.

	<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>		<b>S.S(ML)</b>	<b>S.S.</b>
NbTiVZr	<b>bcc</b>	<b>bcc</b>	OsPdPtRh	<b>fcc</b>	<b>fcc</b>	ReTiVZr	<b>none</b>	<b>none</b>
NbTiWZr	<b>bcc</b>	<b>bcc</b>	OsPdPtRu	<b>none</b>	<b>fcc</b>	ReTiWZr	<b>bcc</b>	<b>bcc</b>
NbVWZr	<b>bcc</b>	<b>none</b>	OsPdReRh	<b>fcc</b>	<b>fcc</b>	ReVWZr	<b>none</b>	<b>bcc</b>
NiOsPdPt	<b>none</b>	<b>none</b>	OsPdReRu	<b>none</b>	<b>fcc</b>	TaTiVW	<b>bcc</b>	<b>bcc</b>
NiOsPdRe	<b>none</b>	<b>fcc</b>	OsPdRhRu	<b>none</b>	<b>none</b>	TaTiVZr	<b>bcc</b>	<b>bcc</b>
NiOsPdRh	<b>none</b>	<b>fcc</b>	OsPtReRh	<b>fcc</b>	<b>fcc</b>	TaTiWZr	<b>bcc</b>	<b>bcc</b>
NiOsPdRu	<b>none</b>	<b>fcc</b>	OsPtReRu	<b>fcc</b>	<b>fcc</b>	TaVWZr	<b>bcc</b>	<b>none</b>
NiOsPtRe	<b>none</b>	<b>fcc</b>	OsPtRhRu	<b>fcc</b>	<b>fcc</b>	TiVWZr	<b>bcc</b>	<b>none</b>

**Supplementary Table 5 ML-HEA results for quinary alloys.** S.S.<sup>(ML)</sup> class of single solid solution phase predicted by the ML-HEA model (“none” = other or not found); S.S. class of single solid solution phase found using LTVC model from DFT [28] (“none” = other or not found); S.S.\* class of single solid solution phase found using CALPHAD (“none” = other or not found).

	S.S. <sup>(ML)</sup>	S.S.		S.S. <sup>(ML)</sup>	S.S.
AlCrMoTiW	bcc	bcc	CrMoNbTaW	bcc	none
CoCrFeMnNi	fcc	fcc	CrMoNbTaZr	none	none
CrNbTaVW	bcc	bcc	CrMoNbTiV	bcc	bcc
CrNbTaVZr	none	none	CrMoNbTiW	bcc	bcc
CrNbTaWZr	none	none	CrMoNbTiZr	bcc	bcc
CrHfMoNbTa	bcc	bcc	CrMoNbVW	bcc	bcc
CrHfMoNbTi	bcc	bcc	CrMoNbVZr	bcc	none
CrHfMoNbV	bcc	none	CrMoNbWZr	none	none
CrHfMoNbW	none	bcc	CrMoTaTiV	bcc	bcc
CrHfMoNbZr	bcc	bcc	CrMoTaTiW	bcc	bcc
CrHfMoTaTi	bcc	bcc	CrMoTaTiZr	none	none
CrHfMoTaV	none	none	CrMoTaVW	bcc	bcc
CrHfMoTaW	none	bcc	CrMoTaVZr	none	none
CrHfMoTaZr	none	bcc	CrMoTaWZr	none	none
CrHfMoTiV	bcc	none	CrMoTiVW	bcc	bcc
CrHfMoTiW	bcc	bcc	CrMoTiVZr	bcc	none
CrHfMoTiZr	none	bcc	CrMoTiWZr	none	none
CrHfMoVW	bcc	bcc	CrMoVWZr	none	none
CrHfMoVZr	bcc	none	CrNbTaTiV	bcc	bcc
CrHfMoWZr	none	bcc	CrNbTaTiW	bcc	bcc
CrHfNbTaTi	bcc	bcc	CrNbTaTiZr	none	bcc
CrHfNbTaV	none	none	CrNbTiVW	bcc	bcc
CrHfNbTaW	none	bcc	CrNbTiVZr	none	none
CrHfNbTaZr	bcc	none	CrNbTiWZr	none	none
CrHfNbTiV	bcc	none	CrNbVWZr	none	none
CrHfNbTiW	bcc	bcc	CrTaTiVW	bcc	bcc
CrHfNbTiZr	bcc	bcc	CrTaTiVZr	none	none
CrHfNbVW	none	bcc	CrTaTiWZr	none	none
CrHfNbVZr	bcc	none	CrTaVWZr	none	none
CrHfNbWZr	none	bcc	CrTiVWZr	none	none
CrHfTaTiV	bcc	bcc	HfMoNbTaTi	bcc	bcc
CrHfTaTiW	bcc	bcc	HfMoNbTaV	bcc	bcc
CrHfTaTiZr	none	bcc	HfMoNbTaW	bcc	bcc
CrHfTaVW	none	bcc	HfMoNbTaZr	bcc	bcc
CrHfTaVZr	bcc	none	HfMoNbTiV	bcc	bcc
CrHfTaWZr	none	bcc	HfMoNbTiW	bcc	bcc

Supplementary Table 5, Continued.

	<b>S.S<sup>(ML)</sup></b>	<b>S.S.</b>		<b>S.S<sup>(ML)</sup></b>	<b>S.S.</b>
CrHfTiVW	<b>bcc</b>	<b>bcc</b>	HfMoNbTiZr	<b>bcc</b>	<b>bcc</b>
CrHfTiVZr	<b>bcc</b>	<b>none</b>	HfMoNbVW	<b>bcc</b>	<b>bcc</b>
CrHfTiWZr	<b>none</b>	<b>bcc</b>	HfMoNbVZr	<b>bcc</b>	<b>bcc</b>
CrHfVWZr	<b>none</b>	<b>none</b>	HfMoNbWZr	<b>none</b>	<b>bcc</b>
CrMoNbTaTi	<b>bcc</b>	<b>bcc</b>	HfMoTaTiV	<b>bcc</b>	<b>bcc</b>
CrMoNbTaV	<b>bcc</b>	<b>bcc</b>	HfMoTaTiW	<b>bcc</b>	<b>bcc</b>
HfMoTaTiZr	<b>bcc</b>	<b>bcc</b>	MoNbReTaW	<b>bcc</b>	<b>bcc</b>
HfMoTaVW	<b>bcc</b>	<b>bcc</b>	MoNbTaTiV	<b>bcc</b>	<b>bcc</b>
HfMoTaVZr	<b>bcc</b>	<b>bcc</b>	MoNbTaTiW	<b>bcc</b>	<b>bcc</b>
HfMoTaWZr	<b>none</b>	<b>bcc</b>	MoNbTaTiZr	<b>bcc</b>	<b>bcc</b>
HfMoTiVW	<b>bcc</b>	<b>bcc</b>	MoNbTaVW	<b>bcc</b>	<b>bcc</b>
HfMoTiVZr	<b>bcc</b>	<b>bcc</b>	MoNbTaVZr	<b>bcc</b>	<b>none</b>
HfMoTiWZr	<b>bcc</b>	<b>bcc</b>	MoNbTaWZr	<b>none</b>	<b>bcc</b>
HfMoVWZr	<b>none</b>	<b>bcc</b>	MoNbTiVW	<b>bcc</b>	<b>bcc</b>
HfNbTaTiV	<b>bcc</b>	<b>bcc</b>	MoNbTiVZr	<b>bcc</b>	<b>bcc</b>
HfNbTaTiW	<b>bcc</b>	<b>bcc</b>	MoNbTiWZr	<b>bcc</b>	<b>bcc</b>
HfNbTaTiZr	<b>bcc</b>	<b>bcc</b>	MoNbVWZr	<b>bcc</b>	<b>none</b>
HfNbTaVW	<b>bcc</b>	<b>bcc</b>	MoTaTiVW	<b>bcc</b>	<b>bcc</b>
HfNbTaVZr	<b>bcc</b>	<b>bcc</b>	MoTaTiVZr	<b>bcc</b>	<b>bcc</b>
HfNbTaWZr	<b>bcc</b>	<b>bcc</b>	MoTaTiWZr	<b>none</b>	<b>bcc</b>
HfNbTiVW	<b>bcc</b>	<b>bcc</b>	MoTaVWZr	<b>bcc</b>	<b>none</b>
HfNbTiVZr	<b>bcc</b>	<b>bcc</b>	MoTiVWZr	<b>none</b>	<b>none</b>
HfNbTiWZr	<b>bcc</b>	<b>bcc</b>	NbReTaTiV	<b>bcc</b>	<b>bcc</b>
HfNbVWZr	<b>bcc</b>	<b>bcc</b>	NbTaTiVW	<b>bcc</b>	<b>bcc</b>
HfTaTiVW	<b>bcc</b>	<b>bcc</b>	NbTaTiVZr	<b>bcc</b>	<b>none</b>
HfTaTiVZr	<b>bcc</b>	<b>none</b>	NbTaTiWZr	<b>bcc</b>	<b>bcc</b>
HfTaTiWZr	<b>bcc</b>	<b>bcc</b>	NbTaVWZr	<b>none</b>	<b>none</b>
HfTaVWZr	<b>bcc</b>	<b>bcc</b>	NbTiVWZr	<b>bcc</b>	<b>bcc</b>
HfTiVWZr	<b>bcc</b>	<b>none</b>	TaTiVWZr	<b>bcc</b>	<b>none</b>

**Supplementary Table 6 Percentage votes for each class by alloy.** The certainty of the model is quantified by reporting the percent of trees that vote for each option (none, fcc, or bcc). The closer to 1, the more certain the ML-HEA model is of the resultant solid solution. The S.S.<sup>(ML)</sup> predictions are provided for quick reference. Experimental results (S.S. Exp.) are provided when available. The voting percentages are displayed to three decimal points and therefore may not add up to unity.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S.<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AgAu	0	1	0	fcc	fcc
AgCu	0.909	0.090	0	none	none
AgFe	0.909	0.090	0	none	none
AgNi	0.909	0	0.090	none	none
AgPd	0	1	0	fcc	fcc
AgPt	0	1	0	fcc	fcc
AgSi	1	0	0	none	none
AgTi	1	0	0	none	none
AlMn	1	0	0	none	none
AlCu	1	0	0	none	none
AlHf	0.909	0	0.090	none	none
AlSi	1	0	0	none	none
AlTa	1	0	0	none	none
AlTi	1	0	0	none	none
AlW	0.909	0	0.090	none	none
AlZr	0.909	0	0.090	none	none
AuCo	0.727	0.272	0	none	none
AuCu	0.272	0.727	0	fcc	fcc
AuNi	0.181	0.818	0	fcc	fcc
AuPd	0	1	0	fcc	fcc
AuPt	0	1	0	fcc	fcc
AuTi	1	0	0	none	none
AuV	0.272	0.545	0.181	fcc	fcc
CoCr	0.272	0	0.727	bcc	bcc
CoCu	0.818	0.090	0.090	none	none
CoIr	0	1	0	fcc	fcc
CoMn	0.090	0.909	0	fcc	fcc
CoMo	1	0	0	none	none
CoNi	0	0.909	0.090	fcc	fcc
CoPd	0	1	0	fcc	fcc
CoPt	0.090	0.909	0	fcc	fcc
CoRh	0.090	0.818	0.090	fcc	fcc
CoW	1	0	0	none	none
CrCu	0.818	0.090	0.090	none	none
CrFe	0.181	0	0.818	bcc	bcc
CrIr	1	0	0	none	none

Supplementary Table 6, Continued

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrMn	0.090	0.090	0.818	bcc	bcc
CrMo	0	0	1	bcc	bcc
CrNb	0.818	0.090	0.090	none	none
CrNi	0	1	0	fcc	fcc
CrPt	0.090	0.909	0	fcc	fcc
CrRh	1	0	0	none	none
CrTa	0.454	0	0.545	bcc	bcc
CrTi	0	0	1	bcc	bcc
CrV	0.090	0	0.909	bcc	bcc
CrW	0.363	0	0.636	bcc	bcc
CuFe	0.636	0.272	0.090	none	none
CuIr	0.727	0.272	0	none	none
CuNi	0	0.727	0.272	fcc	fcc
CuPd	0	1	0	fcc	fcc
CuPt	0	1	0	fcc	fcc
CuRh	0.090	0.909	0	fcc	fcc
CuSi	1	0	0	none	none
CuV	1	0	0	none	none
FeMn	0.181	0.818	0	fcc	fcc
FeMo	1	0	0	none	none
FeNb	1	0	0	none	none
FeNi	0	0.090	0.909	bcc	bcc
FePd	0	1	0	fcc	fcc
HfNb	0	0	1	bcc	bcc
HfTa	0	0	1	bcc	bcc
HfTi	0	0	1	bcc	bcc
HfV	0.727	0	0.272	none	none
HfZr	0.181	0.090	0.727	bcc	bcc
IrMo	1	0	0	none	none
IrNi	0	0.909	0.090	fcc	fcc
IrPd	0.090	0.909	0	fcc	fcc
IrPt	0.090	0.909	0	fcc	fcc
IrRh	0.090	0.909	0	fcc	fcc
IrRu	0.454	0.545	0	fcc	fcc
MnNb	0.818	0	0.181	none	none
MnNi	0	1	0	fcc	fcc
MnTa	0.909	0	0.090	none	none
MnV	0	0	1	bcc	bcc
MoNb	0	0	1	bcc	bcc

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
MoNi	0.909	0.090	0	none	none
MoPd	0.272	0.727	0	fcc	fcc
MoRh	0.818	0.090	0.090	none	none
MoRu	1	0	0	none	none
MoTa	0	0	1	bcc	bcc
MoTi	0	0	1	bcc	bcc
MoV	0	0	1	bcc	bcc
MoW	0	0	1	bcc	bcc
MoZr	1	0	0	none	none
NbTa	0	0	1	bcc	bcc
NbTi	0.181	0	0.818	bcc	bcc
NbV	0	0	1	bcc	bcc
NbW	0	0	1	bcc	bcc
NbZr	0	0	1	bcc	bcc
NiPd	0	0.454	0.545	bcc	bcc
NiPt	0.090	0.909	0	fcc	fcc
NiRh	0.090	0.909	0	fcc	fcc
NiRu	0.818	0.181	0	none	none
NiV	1	0	0	none	none
NiW	0.818	0.181	0	none	none
PdPt	0	1	0	fcc	fcc
PdRh	0	0.909	0.090	fcc	fcc
PdRu	0.727	0.272	0	none	none
PdW	0.636	0.181	0.181	none	none
PtRh	0	1	0	fcc	fcc
PtRu	0	0.818	0.181	fcc	fcc
PtTi	1	0	0	none	none
PtW	0.818	0.181	0	none	none
RhRu	0.181	0.818	0	fcc	fcc
RhV	1	0	0	none	none
RhW	0.909	0.090	0	none	none
RuW	0.727	0.090	0.181	none	none
TaTi	0	0	1	bcc	bcc
TaV	0	0	1	bcc	bcc
TaW	0.090	0	0.909	bcc	bcc
TaZr	0	0.090	0.909	bcc	bcc
TiV	0	0	1	bcc	bcc
TiW	0	0	1	bcc	bcc
TiZr	0	0	1	bcc	bcc

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
VW	0	0	1	bcc	bcc
VZr	0.909	0	0.090	none	none
WZr	0.909	0	0.090	none	none
AgAlNi	0.909	0.090	0	none	
AgAlAu	0.727	0.272	0	none	
AgAlFe	0.909	0.090	0	none	
AgAlV	0.909	0.090	0	none	
AgAsAu	0.909	0.090	0	none	
AgAsCu	0.818	0.181	0	none	
AgAsPd	0.818	0.181	0	none	
AgAuCo	0.727	0.272	0	none	
AgAuCu	0.636	0.363	0	none	
AgAuFe	0.636	0.363	0	none	
AgAuMo	0.727	0.272	0	none	
AgAuNi	0.636	0.363	0	none	
AgAuPd	0	1	0	fcc	
AgAuPt	0	0.909	0.090	fcc	
AgAuRh	0.363	0.636	0	fcc	
AgAuSi	0.818	0.181	0	none	
AgAuTi	0.727	0.272	0	none	
AgAuV	0.636	0.363	0	none	
AgCoCu	0.909	0	0.090	none	
AgCoPd	0.181	0.727	0.090	fcc	
AgCoPt	0.818	0.181	0	none	
AgCoRh	0.818	0.181	0	none	
AgCuFe	0.909	0.090	0	none	
AgCuIr	0.636	0.363	0	none	
AgCuMn	0.909	0.090	0	none	
AgCuNi	0.909	0	0.090	none	
AgCuPd	0.636	0.363	0	none	
AgCuPt	0.636	0.363	0	none	
AgCuRh	0.636	0.363	0	none	
AgCuSi	0.909	0.090	0	none	
AgCuTi	1	0	0	none	
AgCuV	0.909	0.090	0	none	
AgCuZr	1	0	0	none	
AgFeNi	0.909	0.090	0	none	
AgFePd	0.818	0.181	0	none	
AgHfZr	0.909	0	0.090	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AgIrNi	0.818	0.181	0	none	
AgIrPd	0	1	0	fcc	
AgIrPt	0.363	0.636	0	fcc	
AgIrRh	0.272	0.727	0	fcc	
AgNiPd	0.727	0.272	0	none	
AgNiPt	0.727	0.272	0	none	
AgNiRh	0.818	0.181	0	none	
AgPdPt	0	1	0	fcc	
AgPdRh	0	1	0	fcc	
AgPdV	0.909	0.090	0	none	
AgPtRh	0.454	0.545	0	fcc	
AgPtRu	0.454	0.545	0	fcc	
AgPtTi	0.909	0.090	0	none	
AgPtV	0.727	0.272	0	none	
AgTiV	0.818	0	0.181	none	
AgTiZr	0.727	0.090	0.181	none	
AlAuCu	0.727	0.272	0	none	
AlAuNi	0.727	0.272	0	none	
AlAuSi	0.727	0.181	0.090	none	
AlCoCr	0.818	0.090	0.090	none	
AlCoCu	0.909	0.090	0	none	
AlCoNi	0.727	0.272	0	none	
AlCrCu	0.909	0.090	0	none	
AlCrMn	0.272	0	0.727	bcc	
AlCrV	0.181	0	0.818	bcc	
AlCrW	0.636	0	0.363	none	
AlCuFe	0.727	0.181	0.090	none	
AlCuMn	0.727	0.181	0.090	none	
AlCuNi	0.818	0.181	0	none	
AlCuSi	0.909	0.090	0	none	
AlFeMn	0.818	0.090	0.090	none	
AlHfSi	1	0	0	none	
AlHfTa	0.818	0	0.181	none	
AlHfTi	0.636	0	0.363	none	
AlMnNi	0.818	0.181	0	none	
AlMnSi	0.909	0	0.090	none	
AlTaZr	0.818	0	0.181	none	
AlTiV	0.181	0	0.818	bcc	
AlTiZr	0.363	0	0.636	bcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AlVW	0.454	0	0.545	bcc	
AsAuCu	0.636	0.363	0	none	
AsAuPd	0.636	0.363	0	none	
AsCuPd	0.818	0.181	0	none	
AuCoCu	0.636	0.363	0	none	
AuCoIr	0.090	0.909	0	fcc	
AuCoMo	0.727	0.181	0.090	none	
AuCoNi	0.636	0.363	0	none	
AuCoPd	0.181	0.818	0	fcc	
AuCoPt	0.272	0.727	0	fcc	
AuCoRh	0.454	0.545	0	fcc	
AuCoV	0.818	0.181	0	none	
AuCuFe	0.636	0.363	0	none	
AuCuIr	0.272	0.727	0	fcc	
AuCuNi	0.181	0.818	0	fcc	
AuCuPd	0.181	0.818	0	fcc	
AuCuPt	0	1	0	fcc	
AuCuRh	0.363	0.636	0	fcc	
AuCuSi	0.818	0.181	0	none	
AuCuTi	0.909	0.090	0	none	
AuCuV	0.727	0.272	0	none	
AuCuW	0.636	0.363	0	none	
AuFeNi	0.272	0.636	0.090	fcc	
AuFePd	0.090	0.909	0	fcc	
AuHfTi	0.727	0.181	0.090	none	
AuIrNi	0.090	0.909	0	fcc	
AuIrPd	0.090	0.909	0	fcc	
AuIrPt	0	1	0	fcc	
AuIrRh	0.090	0.909	0	fcc	
AuMoNi	0.727	0.181	0.090	none	
AuMoPd	0.181	0.818	0	fcc	
AuMoRh	0.545	0.454	0	none	
AuNiPd	0.181	0.727	0.090	fcc	
AuNiPt	0.090	0.909	0	fcc	
AuNiRh	0.454	0.545	0	fcc	
AuNiV	0.636	0.363	0	none	
AuPdPt	0	1	0	fcc	
AuPdRh	0.090	0.909	0	fcc	
AuPdTi	0.545	0.454	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AuPdV	0.636	0.363	0	none	
AuPdW	0.727	0.272	0	none	
AuPtRh	0.090	0.909	0	fcc	
AuPtRu	0.090	0.909	0	fcc	
AuPtTi	0.818	0.181	0	none	
AuRhV	0.545	0.454	0	none	
AuRhW	0.727	0.272	0	none	
AuTaTi	0.909	0	0.090	none	
AuTiV	1	0	0	none	
AuTiZr	0.727	0.181	0.090	none	
CoCrCu	0.727	0.272	0	none	
CoCrIr	0.090	0.909	0	fcc	
CoCrMn	0.818	0.090	0.090	none	
CoCrNb	0.909	0	0.090	none	
CoCrNi	0	0.818	0.181	fcc	
CoCrPd	0.090	0.909	0	fcc	
CoCrPt	0.090	0.909	0	fcc	
CoCrRh	0.090	0.909	0	fcc	
CoCrW	0.909	0	0.090	none	
CoCuIr	0.727	0.181	0.090	none	
CoCuMn	0.727	0.181	0.090	none	
CoCuNi	0.272	0.636	0.090	fcc	
CoCuPd	0.181	0.727	0.090	fcc	
CoCuPt	0.272	0.636	0.090	fcc	
CoCuRh	0.272	0.727	0	fcc	
CoCuV	1	0	0	none	
CoFeMn	0.090	0.909	0	fcc	
CoIrMn	0.090	0.727	0.181	fcc	
CoIrMo	0.363	0.636	0	fcc	
CoIrNi	0	0.909	0.090	fcc	
CoIrPd	0.090	0.909	0	fcc	
CoIrPt	0.090	0.909	0	fcc	
CoIrRh	0.090	0.909	0	fcc	
CoIrV	0.636	0.363	0	none	
CoIrW	0.454	0.454	0.090	none	
CoMnNi	0	0.909	0.090	fcc	
CoMnRh	0.090	0.909	0	fcc	
CoMoNi	0.909	0.090	0	none	
CoMoPt	0.545	0.363	0.090	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CoMoRh	0.454	0.545	0	fcc	
CoNiPd	0	0.909	0.090	fcc	
CoNiPt	0.090	0.909	0	fcc	
CoNiRh	0.181	0.818	0	fcc	
CoNiV	0.363	0.636	0	fcc	
CoPdPt	0.090	0.909	0	fcc	
CoPdRh	0.090	0.909	0	fcc	
CoPdV	0.545	0.454	0	none	
CoPdW	0.727	0.272	0	none	
CoPtRh	0.090	0.909	0	fcc	
CoPtV	0.545	0.454	0	none	
CoRhV	0.636	0.363	0	none	
CoRhW	0.545	0.363	0.090	none	
CrCuFe	0.818	0.181	0	none	
CrCuMo	0.909	0	0.090	none	
CrCuNi	0.818	0.090	0.090	none	
CrCuPd	1	0	0	none	
CrCuPt	0.818	0.181	0	none	
CrCuRh	0.727	0.272	0	none	
CrCuV	0.909	0	0.090	none	
CrCuW	0.727	0	0.272	none	
CrFeIr	0.454	0.454	0.090	none	
CrFeMn	0.181	0	0.818	bcc	
CrFeMo	0.909	0	0.090	none	
CrFeNb	0.909	0	0.090	none	
CrFeNi	0	0.909	0.090	fcc	
CrFeRh	0.454	0.454	0.090	none	
CrFeTa	0.818	0	0.181	none	
CrFeTi	0.818	0	0.181	none	
CrFeV	0.454	0	0.545	bcc	
CrFeW	0.909	0	0.090	none	
CrHfNb	0.454	0	0.545	bcc	
CrHfTa	0.545	0	0.454	none	
CrHfTi	0.272	0	0.727	bcc	
CrHfW	0.727	0	0.272	none	
CrIrMn	0.454	0.363	0.181	none	
CrIrNi	0.090	0.909	0	fcc	
CrIrPd	0.181	0.818	0	fcc	
CrIrPt	0.181	0.818	0	fcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrIrRh	0.090	0.909	0	fcc	
CrMnMo	0.636	0	0.363	none	
CrMnNb	0.272	0	0.727	bcc	
CrMnNi	0.727	0.272	0	none	
CrMnRh	0.818	0.090	0.090	none	
CrMnTa	0.454	0	0.545	bcc	
CrMnTi	0.818	0	0.181	none	
CrMnV	0	0	1	bcc	
CrMnW	0.818	0.090	0.090	none	
CrMoNb	0.181	0	0.818	bcc	
CrMoRe	0.545	0	0.454	none	
CrMoTa	0.545	0	0.454	none	
CrMoTc	0.545	0	0.454	none	
CrMoTi	0.090	0	0.909	bcc	
CrMoV	0	0	1	bcc	
CrMoW	0.272	0	0.727	bcc	
CrMoZr	0.727	0	0.272	none	
CrNbTa	0.454	0	0.545	bcc	
CrNbTi	0.363	0	0.636	bcc	
CrNbV	0.181	0	0.818	bcc	
CrNbW	0.454	0	0.545	bcc	
CrNbZr	0.636	0	0.363	none	
CrNiPd	0.090	0.818	0.090	fcc	
CrNiPt	0.181	0.818	0	fcc	
CrNiRh	0.181	0.818	0	fcc	
CrPdPt	0.181	0.818	0	fcc	
CrPdRh	0.090	0.909	0	fcc	
CrPtRh	0.181	0.818	0	fcc	
CrReTa	0.636	0	0.363	none	
CrReW	0.454	0	0.545	bcc	
CrTaTi	0.454	0	0.545	bcc	
CrTaV	0.636	0	0.363	none	
CrTaW	0.636	0	0.363	none	
CrTaZr	0.727	0	0.272	none	
CrTcW	0.454	0	0.545	bcc	
CrTiV	0	0	1	bcc	
CrTiW	0.727	0	0.272	none	
CrTiZr	0.636	0	0.363	none	
CrVW	0.545	0	0.454	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrWZr	0.818	0	0.181	none	
CuFeMn	0.818	0.181	0	none	
CuFeNi	0.272	0.454	0.272	fcc	
CuFePd	0.727	0.272	0	none	
CuFeRh	0.727	0.272	0	none	
CuFeV	0.818	0.090	0.090	none	
CuHfTi	0.909	0	0.090	none	
CuHfZr	1	0	0	none	
CuIrMn	0.818	0.181	0	none	
CuIrNi	0.090	0.818	0.090	fcc	
CuIrPd	0.090	0.909	0	fcc	
CuIrPt	0.090	0.909	0	fcc	
CuIrRh	0.090	0.909	0	fcc	
CuMnNi	0.363	0.636	0	fcc	
CuNiPd	0	0.636	0.363	fcc	
CuNiPt	0	1	0	fcc	
CuNiRh	0.090	0.909	0	fcc	
CuNiSi	0.909	0.090	0	none	
CuNiV	1	0	0	none	
CuPdPt	0	1	0	fcc	
CuPdRh	0.090	0.909	0	fcc	
CuPdTi	0.727	0.272	0	none	
CuPdV	1	0	0	none	
CuPtRh	0	1	0	fcc	
CuPtRu	0	1	0	fcc	
CuPtTi	1	0	0	none	
CuPtV	0.818	0.181	0	none	
CuRhV	0.727	0.181	0.090	none	
CuTaTi	0.636	0.090	0.272	none	
CuTaV	0.545	0	0.454	none	
CuTiV	0.818	0	0.181	none	
CuTiZr	0.909	0	0.090	none	
CuVW	0.636	0	0.363	none	
FeIrMn	0	0.818	0.181	fcc	
FeIrMo	0.727	0.272	0	none	
FeIrNi	0	0.909	0.090	fcc	
FeIrRh	0.090	0.818	0.090	fcc	
FeIrRu	0	1	0	fcc	
FeMnMo	0.909	0	0.090	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
FeMnNb	0.909	0	0.090	none	
FeMnNi	0	1	0	fcc	
FeMnRh	0.090	0.727	0.181	fcc	
FeMnTa	0.909	0	0.090	none	
FeMnTi	0.909	0	0.090	none	
FeMnV	0.454	0	0.545	bcc	
FeMoNb	0.818	0	0.181	none	
FeMoNi	0.909	0.090	0	none	
FeMoPd	0.454	0.545	0	fcc	
FeMoRh	0.818	0.090	0.090	none	
FeMoTa	0.818	0	0.181	none	
FeMoTi	0.727	0	0.272	none	
FeMoV	0.454	0	0.545	bcc	
FeMoW	1	0	0	none	
FeNbTa	0.636	0	0.363	none	
FeNbTi	0.818	0	0.181	none	
FeNbV	0.818	0	0.181	none	
FeNiRh	0.181	0.727	0.090	fcc	
FeTaTi	0.818	0	0.181	none	
FeTaV	0.636	0	0.363	none	
FeTaW	0.818	0	0.181	none	
FeTiV	0.727	0.090	0.181	none	
FeVW	0.545	0	0.454	none	
HfMoNb	0.272	0	0.727	bcc	
HfMoRe	0.727	0	0.272	none	
HfMoTa	0.545	0	0.454	none	
HfMoTi	0.090	0	0.909	bcc	
HfMoV	0.272	0	0.727	bcc	
HfMoW	0.909	0	0.090	none	
HfMoZr	0.454	0	0.545	bcc	
HfNbTa	0	0	1	bcc	
HfNbTi	0	0	1	bcc	
HfNbV	0.181	0	0.818	bcc	
HfNbW	0.363	0	0.636	bcc	
HfNbZr	0	0	1	bcc	
HfReTa	0.454	0	0.545	bcc	
HfTaTc	0.818	0	0.181	none	
HfTaTi	0	0	1	bcc	
HfTaV	0.181	0	0.818	bcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
HfTaW	0.272	0	0.727	bcc	
HfTaZr	0.181	0.090	0.727	bcc	
HfTiV	0	0	1	bcc	
HfTiW	0.727	0	0.272	none	
HfTiZr	0	0	1	bcc	
HfVW	0.454	0	0.545	bcc	
HfVZr	0.272	0	0.727	bcc	
HfWZr	0.818	0	0.181	none	
IrMnNi	0.090	0.818	0.090	fcc	
IrMnRh	0.090	0.909	0	fcc	
IrMoNi	0.545	0.454	0	none	
IrMoPd	0.181	0.727	0.090	fcc	
IrMoPt	0.181	0.727	0.090	fcc	
IrNbPd	0.818	0.090	0.090	none	
IrNiPd	0.181	0.818	0	fcc	
IrNiPt	0	0.909	0.090	fcc	
IrNiRh	0.090	0.909	0	fcc	
IrNiRu	0.272	0.727	0	fcc	
IrNiW	0.636	0.272	0.090	none	
IrOsPt	0.090	0.909	0	fcc	
IrPdPt	0.090	0.909	0	fcc	
IrPdRh	0	1	0	fcc	
IrPdRu	0.272	0.727	0	fcc	
IrPdV	0.090	0.909	0	fcc	
IrPdW	0.636	0.363	0	none	
IrPtRe	0	1	0	fcc	
IrPtRh	0.090	0.909	0	fcc	
IrPtRu	0	0.909	0.090	fcc	
IrPtTc	0	0.909	0.090	fcc	
IrRhV	0.454	0.545	0	fcc	
MnMoNb	0.272	0	0.727	bcc	
MnMoTa	0.636	0	0.363	none	
MnMoTi	0.636	0	0.363	none	
MnMoV	0	0	1	bcc	
MnMoW	0.727	0.090	0.181	none	
MnNbTa	0.636	0	0.363	none	
MnNbTi	0.363	0	0.636	bcc	
MnNbV	0.181	0	0.818	bcc	
MnNbW	0.727	0	0.272	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
MnTaTi	0.363	0	0.636	bcc	
MnTaV	0.181	0	0.818	bcc	
MnTaW	0.636	0	0.363	none	
MnTiV	0.181	0	0.818	bcc	
MnTiW	0.818	0	0.181	none	
MnVW	0.363	0	0.636	bcc	
MoNbRe	0.181	0	0.818	bcc	
MoNbTa	0	0	1	bcc	
MoNbTc	0.181	0	0.818	bcc	
MoNbTi	0	0	1	bcc	
MoNbV	0	0	1	bcc	
MoNbW	0	0	1	bcc	
MoNbZr	0.363	0	0.636	bcc	
MoNiPd	0.545	0.454	0	none	
MoNiPt	0.545	0.454	0	none	
MoNiRh	0.545	0.454	0	none	
MoPdPt	0.090	0.818	0.090	fcc	
MoPdRh	0.181	0.727	0.090	fcc	
MoReTa	0.181	0	0.818	bcc	
MoReTi	0.181	0	0.818	bcc	
MoReV	0.181	0.090	0.727	bcc	
MoReW	0.363	0	0.636	bcc	
MoTaTc	0.454	0	0.545	bcc	
MoTaTi	0	0	1	bcc	
MoTaV	0.454	0	0.545	bcc	
MoTaW	0	0	1	bcc	
MoTaZr	0.454	0	0.545	bcc	
MoTcTi	0.272	0	0.727	bcc	
MoTcV	0.181	0.090	0.727	bcc	
MoTcW	0.363	0	0.636	bcc	
MoTiV	0	0	1	bcc	
MoTiW	0	0	1	bcc	
MoTiZr	0.090	0	0.909	bcc	
MoVW	0	0	1	bcc	
MoVZr	0.090	0	0.909	bcc	
MoWZr	0.545	0	0.454	none	
NbPdRh	0.727	0.272	0	none	
NbReTa	0	0	1	bcc	
NbReTi	0.090	0	0.909	bcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
NbReV	0	0	1	bcc	
NbReW	0.181	0	0.818	bcc	
NbReZr	0.363	0	0.636	bcc	
NbTaTc	0.181	0	0.818	bcc	
NbTaTi	0	0	1	bcc	
NbTaV	0	0	1	bcc	
NbTaW	0.090	0	0.909	bcc	
NbTaZr	0	0	1	bcc	
NbTcV	0	0	1	bcc	
NbTcW	0	0	1	bcc	
NbTiV	0	0	1	bcc	
NbTiW	0	0	1	bcc	
NbTiZr	0	0	1	bcc	
NbVW	0	0	1	bcc	
NbVZr	0.090	0	0.909	bcc	
NbWZr	0.363	0	0.636	bcc	
NiPdPt	0	0.909	0.090	fcc	
NiPdRh	0.090	0.909	0	fcc	
NiPdV	0.545	0.454	0	none	
NiPtRh	0.090	0.909	0	fcc	
NiPtRu	0.272	0.727	0	fcc	
NiPtV	0.454	0.545	0	fcc	
NiRhV	0.454	0.545	0	fcc	
NiRhW	0.727	0.272	0	none	
PdPtRh	0	1	0	fcc	
PdPtRu	0.090	0.818	0.090	fcc	
PdPtTi	0.727	0.272	0	none	
PdRhV	0.090	0.909	0	fcc	
PdRhW	0.636	0.363	0	none	
PdRuW	0.363	0.272	0.363	none	
PdTiV	0.727	0.090	0.181	none	
PtRhRu	0	0.818	0.181	fcc	
PtRhV	0.181	0.818	0	fcc	
PtRhW	0.636	0.363	0	none	
ReTaTi	0.090	0	0.909	bcc	
ReTaV	0	0	1	bcc	
ReTaW	0.090	0	0.909	bcc	
ReTaZr	0.545	0	0.454	none	
ReTiW	0.090	0	0.909	bcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
ReVW	0	0	1	bcc	
TaTcTi	0.272	0	0.727	bcc	
TaTcV	0.181	0	0.818	bcc	
TaTcW	0.363	0	0.636	bcc	
TaTiV	0	0	1	bcc	
TaTiW	0.090	0	0.909	bcc	
TaTiZr	0	0	1	bcc	
TaVW	0	0	1	bcc	
TaVZr	0.363	0	0.636	bcc	
TaWZr	0.272	0	0.727	bcc	
TcTiW	0.181	0	0.818	bcc	
TcVW	0	0	1	bcc	
TiVZr	0.545	0.090	0.363	none	
TiWZr	0.363	0	0.636	bcc	
VWZr	0.454	0	0.545	bcc	
AgAlAuCo	0.909	0.090	0	none	
AgAlAuCu	0.909	0.090	0	none	
AgAlAuFe	0.909	0.090	0	none	
AgAlAuNi	0.909	0.090	0	none	
AgAlAuSi	0.909	0	0.090	none	
AgAlAuZr	1	0	0	none	
AgAlCoCu	0.909	0.090	0	none	
AgAlCoNi	0.909	0.090	0	none	
AgAlCuFe	0.909	0.090	0	none	
AgAlCuMn	0.909	0.090	0	none	
AgAlCuNi	0.909	0.090	0	none	
AgAlCuSi	1	0	0	none	
AgAlCuZr	1	0	0	none	
AgAlFeNi	0.909	0.090	0	none	
AgAlNiSi	1	0	0	none	
AgAuCoCr	0.818	0.181	0	none	
AgAuCoCu	0.636	0.363	0	none	
AgAuCoIr	0.636	0.363	0	none	
AgAuCoMo	0.818	0.181	0	none	
AgAuCoNi	0.727	0.272	0	none	
AgAuCoPd	0.272	0.727	0	fcc	
AgAuCoPt	0.636	0.363	0	none	
AgAuCoRe	0.727	0.272	0	none	
AgAuCoRh	0.636	0.363	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AgAuCoRu	0.727	0.272	0	none	
AgAuCrCu	0.636	0.363	0	none	
AgAuCrNi	0.727	0.272	0	none	
AgAuCuFe	0.636	0.363	0	none	
AgAuCuIr	0.454	0.545	0	fcc	
AgAuCuNi	0.636	0.363	0	none	
AgAuCuPd	0.181	0.636	0.181	fcc	
AgAuCuPt	0.181	0.727	0.090	fcc	
AgAuCuRh	0.636	0.363	0	none	
AgAuCuRu	0.454	0.545	0	fcc	
AgAuCuSi	0.818	0.181	0	none	
AgAuCuZr	1	0	0	none	
AgAuFeMo	0.727	0.181	0.090	none	
AgAuFeNi	0.727	0.272	0	none	
AgAuFePd	0.636	0.363	0	none	
AgAuFeRh	0.545	0.454	0	none	
AgAuIrMo	0.727	0.272	0	none	
AgAuIrNi	0.636	0.363	0	none	
AgAuIrPd	0.181	0.818	0	fcc	
AgAuIrPt	0.454	0.545	0	fcc	
AgAuIrRh	0.272	0.727	0	fcc	
AgAuIrRu	0.454	0.545	0	fcc	
AgAuMoNi	0.818	0.181	0	none	
AgAuMoPd	0.363	0.636	0	fcc	
AgAuMoRh	0.636	0.363	0	none	
AgAuNbRh	0.818	0.181	0	none	
AgAuNiPd	0.636	0.363	0	none	
AgAuNiPt	0.636	0.363	0	none	
AgAuNiRe	0.727	0.272	0	none	
AgAuNiRh	0.727	0.272	0	none	
AgAuNiRu	0.727	0.272	0	none	
AgAuNiSi	0.818	0.181	0	none	
AgAuPdPt	0	1	0	fcc	
AgAuPdRe	0.454	0.545	0	fcc	
AgAuPdRh	0.090	0.909	0	fcc	
AgAuPdRu	0.272	0.636	0.090	fcc	
AgAuPtRh	0.363	0.636	0	fcc	
AgAuPtRu	0.272	0.727	0	fcc	
AgAuRhRu	0.272	0.727	0	fcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AgCoCrCu	1	0	0	none	
AgCoCrNi	0.909	0	0.090	none	
AgCoCrOs	0.909	0.090	0	none	
AgCoCrPd	0.909	0.090	0	none	
AgCoCrPt	0.909	0.090	0	none	
AgCoCrRh	0.818	0.181	0	none	
AgCoCrRu	0.909	0.090	0	none	
AgCoCuFe	0.909	0	0.090	none	
AgCoCuPd	0.818	0.181	0	none	
AgCoCuPt	0.909	0.090	0	none	
AgCoIrMo	0.727	0.272	0	none	
AgCoIrPd	0.363	0.636	0	fcc	
AgCoIrPt	0.727	0.272	0	none	
AgCoMoNi	0.909	0.090	0	none	
AgCoMoOs	0.727	0.272	0	none	
AgCoMoPd	0.818	0.181	0	none	
AgCoMoPt	0.818	0.181	0	none	
AgCoMoRe	0.818	0.181	0	none	
AgCoMoRh	0.727	0.272	0	none	
AgCoMoRu	0.727	0.272	0	none	
AgCoNiPd	0.454	0.454	0.090	none	
AgCoNiPt	0.909	0.090	0	none	
AgCoOsPd	0.818	0.181	0	none	
AgCoOsPt	0.727	0.272	0	none	
AgCoPdPt	0.272	0.727	0	fcc	
AgCoPdRe	0.909	0.090	0	none	
AgCoPdRh	0.272	0.727	0	fcc	
AgCoPdRu	0.727	0.272	0	none	
AgCoPtRe	0.909	0.090	0	none	
AgCoPtRh	0.727	0.272	0	none	
AgCoPtRu	0.818	0.181	0	none	
AgCrCuFe	0.818	0.090	0.090	none	
AgCrCuNi	1	0	0	none	
AgCrCuPd	0.909	0.090	0	none	
AgCrCuPt	0.727	0.272	0	none	
AgCrCuRh	0.818	0.181	0	none	
AgCrCuRu	0.909	0.090	0	none	
AgCrFeNi	0.909	0.090	0	none	
AgCrIrNi	1	0	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AgCrIrPd	0.181	0.818	0	fcc	
AgCrIrRh	0.636	0.363	0	none	
AgCrIrRu	0.818	0.181	0	none	
AgCrNiOs	0.909	0.090	0	none	
AgCrNiPd	1	0	0	none	
AgCrNiPt	0.909	0.090	0	none	
AgCrNiRh	0.818	0.181	0	none	
AgCrNiRu	0.909	0.090	0	none	
AgCrOsPt	0.636	0.363	0	none	
AgCrPdPt	0.181	0.818	0	fcc	
AgCrPdRh	0.272	0.636	0.090	fcc	
AgCrPdRu	0.909	0.090	0	none	
AgCrPtRh	0.727	0.272	0	none	
AgCrPtRu	0.636	0.363	0	none	
AgCrRhRu	0.727	0.272	0	none	
AgCuFeIr	0.818	0.181	0	none	
AgCuFeNi	0.909	0	0.090	none	
AgCuFeOs	0.818	0.181	0	none	
AgCuFePd	0.818	0.181	0	none	
AgCuFeRe	1	0	0	none	
AgCuFeRh	0.818	0.181	0	none	
AgCuFeRu	0.818	0.181	0	none	
AgCuIrPd	0.727	0.272	0	none	
AgCuIrPt	0.545	0.454	0	none	
AgCuNbPd	0.909	0	0.090	none	
AgCuNiPd	0.727	0.272	0	none	
AgCuNiPt	0.545	0.363	0.090	none	
AgCuNiSi	0.909	0.090	0	none	
AgCuPdPt	0.090	0.909	0	fcc	
AgCuPdRe	0.909	0.090	0	none	
AgCuPdRh	0.545	0.454	0	none	
AgCuPdRu	0.727	0.272	0	none	
AgCuPtRe	0.818	0.181	0	none	
AgCuPtRh	0.272	0.727	0	fcc	
AgCuPtRu	0.636	0.363	0	none	
AgFeIrNi	0.727	0.272	0	none	
AgFeIrRu	0.727	0.272	0	none	
AgFeMoNi	0.909	0.090	0	none	
AgFeMoOs	0.727	0.272	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AgFeMoPd	0.818	0.181	0	none	
AgFeMoRe	0.818	0.181	0	none	
AgFeMoRh	0.727	0.272	0	none	
AgFeMoRu	0.727	0.272	0	none	
AgFeNiOs	0.909	0.090	0	none	
AgFeNiPd	0.909	0.090	0	none	
AgFeNiRe	1	0	0	none	
AgFeNiRh	0.727	0.272	0	none	
AgFeNiRu	0.909	0.090	0	none	
AgFeOsPd	0.727	0.272	0	none	
AgFeOsRe	0.909	0.090	0	none	
AgFeOsRh	0.727	0.272	0	none	
AgFeOsRu	0.636	0.363	0	none	
AgFePdRe	0.909	0.090	0	none	
AgFePdRh	0.727	0.272	0	none	
AgFePdRu	0.727	0.272	0	none	
AgFeReRu	0.818	0.181	0	none	
AgFeRhRu	0.727	0.272	0	none	
AgIrMoNi	0.818	0.181	0	none	
AgIrMoPd	0.636	0.363	0	none	
AgIrMoRh	0.454	0.545	0	fcc	
AgIrNbPt	0.818	0.181	0	none	
AgIrNiPd	0.454	0.545	0	fcc	
AgIrNiPt	0.727	0.272	0	none	
AgIrOsPd	0.363	0.636	0	fcc	
AgIrOsPt	0.454	0.545	0	fcc	
AgIrPdPt	0.090	0.909	0	fcc	
AgIrPdRe	0.727	0.272	0	none	
AgIrPdRh	0	1	0	fcc	
AgIrPdRu	0.636	0.363	0	none	
AgIrPtRe	0.636	0.363	0	none	
AgIrPtRh	0.272	0.727	0	fcc	
AgIrPtRu	0.363	0.636	0	fcc	
AgMoNiOs	0.727	0.272	0	none	
AgMoNiPd	0.818	0.181	0	none	
AgMoNiPt	0.818	0.181	0	none	
AgMoNiRe	0.909	0.090	0	none	
AgMoNiRh	0.818	0.181	0	none	
AgMoNiRu	0.727	0.272	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AgMoOsPd	0.636	0.363	0	none	
AgMoOsRe	0.727	0.272	0	none	
AgMoOsRh	0.454	0.545	0	fcc	
AgMoPdPt	0.636	0.363	0	none	
AgMoPdRe	0.727	0.272	0	none	
AgMoPdRh	0.545	0.454	0	none	
AgMoPdRu	0.636	0.363	0	none	
AgMoPtRe	0.818	0.181	0	none	
AgMoPtRh	0.636	0.363	0	none	
AgMoPtRu	0.545	0.454	0	none	
AgMoReRh	0.636	0.363	0	none	
AgMoRhRu	0.454	0.545	0	fcc	
AgNbPdPt	0.909	0.090	0	none	
AgNbPdRh	0.727	0.181	0.090	none	
AgNbPtRh	0.636	0.272	0.090	none	
AgNiOsPd	0.818	0.181	0	none	
AgNiOsPt	0.818	0.181	0	none	
AgNiPdPt	0.272	0.727	0	fcc	
AgNiPdRe	0.909	0.090	0	none	
AgNiPdRh	0.454	0.545	0	fcc	
AgNiPdRu	0.818	0.181	0	none	
AgNiPtRe	0.818	0.181	0	none	
AgNiPtRh	0.727	0.272	0	none	
AgNiPtRu	0.818	0.181	0	none	
AgOsPdPt	0.454	0.545	0	fcc	
AgOsPdRe	0.727	0.272	0	none	
AgOsPdRh	0.272	0.727	0	fcc	
AgOsPdRu	0.636	0.363	0	none	
AgOsPtRe	0.636	0.363	0	none	
AgOsPtRh	0.454	0.545	0	fcc	
AgOsPtRu	0.454	0.545	0	fcc	
AgPdPtRe	0.636	0.363	0	none	
AgPdPtRh	0.090	0.909	0	fcc	
AgPdPtRu	0.090	0.909	0	fcc	
AgPdReRh	0.636	0.363	0	none	
AgPdReRu	0.727	0.272	0	none	
AgPdRhRu	0.545	0.454	0	none	
AgPtReRh	0.636	0.363	0	none	
AgPtReRu	0.636	0.363	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AgPtRhRu	0.545	0.454	0	none	
AlNbTiV	0.090	0	0.909	bcc	bcc
AlNbVW	0.181	0	0.818	bcc	
AlCoCrCu	0.909	0.090	0	none	
AlCrCuFe	0.727	0.181	0.090	none	
AlCrFeMn	0.272	0	0.727	bcc	
AlCrFeMo	0.454	0	0.545	bcc	
AlCrFeNi	0.727	0.181	0.090	none	
AlCrFeRe	0.818	0	0.181	none	
AlCrFeTi	0.363	0	0.636	bcc	
AlCrFeV	0.363	0	0.636	bcc	
AlCrFeW	0.545	0	0.454	none	
AlCrHfTi	0.272	0	0.727	bcc	
AlCrHfW	0.636	0	0.363	none	
AlCrMnMo	0.363	0	0.636	bcc	
AlCrMnRe	0.818	0	0.181	none	
AlCrMnTi	0.454	0	0.545	bcc	
AlCrMnV	0.363	0	0.636	bcc	
AlCrMnW	0.545	0	0.454	none	
AlCrMoRe	0.363	0	0.636	bcc	
AlCrMoTi	0.090	0	0.909	bcc	
AlCrMoV	0.181	0	0.818	bcc	
AlCrMoW	0.272	0	0.727	bcc	
AlCrNbV	0.181	0	0.818	bcc	
AlCrReTi	0.727	0	0.272	none	
AlCrReV	0.636	0	0.363	none	
AlCrReW	0.454	0	0.545	bcc	
AlCrTiV	0	0	1	bcc	
AlCrTiW	0.454	0	0.545	bcc	
AlCrVW	0.454	0	0.545	bcc	
AlCuFeMn	0.727	0.181	0.090	none	
AlCuFeNi	0.818	0.181	0	none	
AlCuFeV	0.727	0.181	0.090	none	
AlFeMnRe	1	0	0	none	
AlFeMnTi	0.818	0.090	0.090	none	
AlFeMnV	0.454	0	0.545	bcc	
AlFeTiV	0	0	1	bcc	
AlHfTiV	0.363	0	0.636	bcc	
AlHfTiW	0.727	0	0.272	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AlHfVW	0.727	0	0.272	none	
AlMnMoRe	0.909	0	0.090	none	
AlMnReW	0.818	0	0.181	none	
AlMnTiV	0.181	0	0.818	bcc	
AlMnVW	0.545	0	0.454	none	
AlMoTiV	0.090	0	0.909	bcc	
AlMoVW	0.181	0	0.818	bcc	
AlReVW	0.363	0	0.636	bcc	
AlTiVW	0.454	0	0.545	bcc	
AuCoCrPd	0.090	0.909	0	fcc	
AuCoCrPt	0.636	0.363	0	none	
AuCoCuPd	0.636	0.363	0	none	
AuCoCuPt	0.636	0.363	0	none	
AuCoIrPd	0.090	0.909	0	fcc	
AuCoIrPt	0.090	0.909	0	fcc	
AuCoMoPd	0.363	0.636	0	fcc	
AuCoMoPt	0.636	0.272	0.090	none	
AuCoNiPd	0.272	0.636	0.090	fcc	
AuCoNiPt	0.272	0.636	0.090	fcc	
AuCoOsPd	0.545	0.454	0	none	
AuCoOsPt	0.545	0.454	0	none	
AuCoPdPt	0.090	0.909	0	fcc	
AuCoPdRe	0.727	0.272	0	none	
AuCoPdRh	0.090	0.909	0	fcc	
AuCoPdRu	0.727	0.272	0	none	
AuCoPtRe	0.727	0.272	0	none	
AuCoPtRh	0.181	0.818	0	fcc	
AuCoPtRu	0.545	0.454	0	none	
AuCrCuPd	0.272	0.727	0	fcc	
AuCrCuPt	0.636	0.363	0	none	
AuCrIrPd	0.090	0.909	0	fcc	
AuCrIrPt	0.181	0.818	0	fcc	
AuCrNiPd	0.090	0.909	0	fcc	
AuCrNiPt	0.636	0.363	0	none	
AuCrOsPt	0.545	0.454	0	none	
AuCrPdPt	0.181	0.818	0	fcc	
AuCrPdRh	0.181	0.818	0	fcc	
AuCrPdRu	0.636	0.363	0	none	
AuCrPtRh	0.545	0.454	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AuCrPtRu	0.545	0.454	0	none	
AuCuFePd	0.727	0.272	0	none	
AuCuIrPd	0.181	0.818	0	fcc	
AuCuIrPt	0.090	0.909	0	fcc	
AuCuNiPd	0.181	0.818	0	fcc	
AuCuNiPt	0.181	0.727	0.090	fcc	
AuCuOsPt	0.454	0.545	0	fcc	
AuCuPdPt	0	1	0	fcc	
AuCuPdRe	0.727	0.272	0	none	
AuCuPdRh	0.181	0.818	0	fcc	
AuCuPdRu	0.363	0.636	0	fcc	
AuCuPtRe	0.727	0.272	0	none	
AuCuPtRh	0.363	0.636	0	fcc	
AuCuPtRu	0.363	0.636	0	fcc	
AuFeMoPd	0.636	0.363	0	none	
AuFeNiPd	0.090	0.909	0	fcc	
AuFeOsPd	0.272	0.727	0	fcc	
AuFePdRe	0.545	0.454	0	none	
AuFePdRh	0.090	0.909	0	fcc	
AuFePdRu	0.090	0.818	0.090	fcc	
AuIrMoPd	0.090	0.909	0	fcc	
AuIrNiPd	0.090	0.909	0	fcc	
AuIrNiPt	0	1	0	fcc	
AuIrOsPd	0.363	0.636	0	fcc	
AuIrOsPt	0	1	0	fcc	
AuIrPdPt	0	1	0	fcc	
AuIrPdRe	0.454	0.545	0	fcc	
AuIrPdRh	0.090	0.909	0	fcc	
AuIrPdRu	0.272	0.727	0	fcc	
AuIrPtRe	0	1	0	fcc	
AuIrPtRh	0.090	0.909	0	fcc	
AuIrPtRu	0.090	0.909	0	fcc	
AuMoNiPd	0.636	0.363	0	none	
AuMoNiPt	0.363	0.636	0	fcc	
AuMoOsPd	0.363	0.636	0	fcc	
AuMoPdPt	0.090	0.909	0	fcc	
AuMoPdRe	0.636	0.363	0	none	
AuMoPdRh	0.181	0.818	0	fcc	
AuMoPdRu	0.454	0.545	0	fcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
AuMoPtRe	0.636	0.363	0	none	
AuMoPtRh	0.454	0.545	0	fcc	
AuMoPtRu	0.545	0.454	0	none	
AuNbPdRh	0.818	0.181	0	none	
AuNbPtRh	0.818	0.181	0	none	
AuNiOsPd	0.636	0.363	0	none	
AuNiOsPt	0.545	0.454	0	none	
AuNiPdPt	0	1	0	fcc	
AuNiPdRe	0.727	0.272	0	none	
AuNiPdRh	0.181	0.818	0	fcc	
AuNiPdRu	0.545	0.454	0	none	
AuNiPtRe	0.818	0.181	0	none	
AuNiPtRh	0.181	0.818	0	fcc	
AuNiPtRu	0.545	0.454	0	none	
AuOsPdPt	0.363	0.636	0	fcc	
AuOsPdRe	0.545	0.454	0	none	
AuOsPdRh	0.363	0.636	0	fcc	
AuOsPtRh	0.272	0.727	0	fcc	
AuOsPtRu	0.272	0.727	0	fcc	
AuPdPtRe	0.363	0.636	0	fcc	
AuPdPtRh	0.090	0.909	0	fcc	
AuPdPtRu	0.090	0.909	0	fcc	
AuPdReRh	0.363	0.636	0	fcc	
AuPdReRu	0.545	0.454	0	none	
AuPdRhRu	0.272	0.727	0	fcc	
AuPtReRh	0.272	0.727	0	fcc	
AuPtReRu	0.363	0.636	0	fcc	
AuPtRhRu	0.181	0.818	0	fcc	
CoCrCuNi	0.727	0.181	0.090	none	
CoCrCuPd	0.909	0.090	0	none	
CoCrCuPt	0.909	0.090	0	none	
CoCrFeNi	0	0.727	0.272	fcc	fcc
CoCrIrPd	0.181	0.818	0	fcc	
CoCrIrPt	0.181	0.818	0	fcc	
CoCrMnMo	0.818	0.090	0.090	none	
CoCrMnNb	0.909	0	0.090	none	
CoCrMnNi	0	1	0	fcc	fcc
CoCrMoNb	0.909	0	0.090	none	
CoCrMoW	0.727	0.090	0.181	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CoCrNbW	0.818	0	0.181	none	
CoCrNiPd	0.090	0.818	0.090	fcc	
CoCrNiPt	0.181	0.818	0	fcc	
CoCrOsPd	0.818	0.181	0	none	
CoCrOsPt	0.727	0.272	0	none	
CoCrPdPt	0.181	0.818	0	fcc	
CoCrPdRh	0.090	0.818	0.090	fcc	
CoCrPdRu	0.818	0.181	0	none	
CoCrPtRh	0.181	0.818	0	fcc	
CoCrPtRu	0.727	0.272	0	none	
CoCuIrPd	0.181	0.818	0	fcc	
CoCuIrPt	0.181	0.818	0	fcc	
CoCuNiPd	0	0.818	0.181	fcc	
CoCuNiPt	0.181	0.818	0	fcc	
CoCuOsPd	0.818	0.181	0	none	
CoCuOsPt	0.818	0.181	0	none	
CoCuPdPt	0.181	0.818	0	fcc	
CoCuPdRe	0.909	0.090	0	none	
CoCuPdRh	0.181	0.818	0	fcc	
CoCuPdRu	0.818	0.181	0	none	
CoCuPtRe	0.909	0.090	0	none	
CoCuPtRh	0.090	0.909	0	fcc	
CoCuPtRu	0.454	0.545	0	fcc	
CoFeMnNi	0	0.909	0.090	fcc	fcc
CoIrMoPd	0.454	0.545	0	fcc	
CoIrMoPt	0.363	0.636	0	fcc	
CoIrNiPd	0	0.909	0.090	fcc	
CoIrNiPt	0.090	0.909	0	fcc	
CoIrOsPd	0.727	0.272	0	none	
CoIrOsPt	0.181	0.818	0	fcc	
CoIrPdPt	0.090	0.909	0	fcc	
CoIrPdRe	0.818	0.181	0	none	
CoIrPdRh	0.090	0.909	0	fcc	
CoIrPdRu	0.727	0.272	0	none	
CoIrPtRe	0.454	0.545	0	fcc	
CoIrPtRh	0.090	0.909	0	fcc	
CoIrPtRu	0.363	0.636	0	fcc	
CoMoNbW	0.727	0	0.272	none	
CoMoNiPd	0.363	0.636	0	fcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CoMoNiPt	0.454	0.545	0	fcc	
CoMoOsPd	0.727	0.272	0	none	
CoMoOsPt	0.727	0.272	0	none	
CoMoPdPt	0.363	0.636	0	fcc	
CoMoPdRe	0.818	0.181	0	none	
CoMoPdRh	0.363	0.636	0	fcc	
CoMoPdRu	0.727	0.272	0	none	
CoMoPtRe	0.727	0.272	0	none	
CoMoPtRh	0.363	0.636	0	fcc	
CoMoPtRu	0.727	0.272	0	none	
CoNiOsPd	0.727	0.181	0.090	none	
CoNiOsPt	0.818	0.181	0	none	
CoNiPdPt	0.090	0.909	0	fcc	
CoNiPdRe	0.909	0.090	0	none	
CoNiPdRh	0.181	0.818	0	fcc	
CoNiPdRu	0.727	0.272	0	none	
CoNiPtRe	0.909	0.090	0	none	
CoNiPtRh	0.181	0.818	0	fcc	
CoNiPtRu	0.363	0.636	0	fcc	
CoOsPdPt	0.545	0.454	0	none	
CoOsPdRe	0.818	0.181	0	none	
CoOsPdRh	0.636	0.363	0	none	
CoOsPdRu	0.727	0.272	0	none	
CoOsPtRe	0.818	0.181	0	none	
CoOsPtRh	0.545	0.454	0	none	
CoOsPtRu	0.727	0.272	0	none	
CoPdPtRe	0.818	0.181	0	none	
CoPdPtRh	0.090	0.909	0	fcc	
CoPdPtRu	0.363	0.636	0	fcc	
CoPdReRh	0.818	0.181	0	none	
CoPdReRu	0.818	0.181	0	none	
CoPdRhRu	0.727	0.272	0	none	
CoPtReRh	0.818	0.181	0	none	
CoPtReRu	0.818	0.181	0	none	
CoPtRhRu	0.363	0.636	0	fcc	
CrCuFeMn	0.909	0.090	0	none	
CrCuFeMo	1	0	0	none	
CrCuFeNb	1	0	0	none	
CrCuFeNi	0.636	0.363	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrCuFePd	1	0	0	none	
CrCuFeV	0.909	0	0.090	none	
CrCuFeW	0.727	0	0.272	none	
CrCuIrPd	0.272	0.727	0	fcc	
CrCuIrPt	0.636	0.363	0	none	
CrCuMnV	0.909	0	0.090	none	
CrCuMnW	0.636	0.090	0.272	none	
CrCuMoTa	0.818	0	0.181	none	
CrCuMoTi	0.818	0	0.181	none	
CrCuMoV	0.818	0	0.181	none	
CrCuMoW	0.727	0	0.272	none	
CrCuNbTa	0.909	0	0.090	none	
CrCuNbTi	0.818	0	0.181	none	
CrCuNbV	0.818	0	0.181	none	
CrCuNbW	0.818	0	0.181	none	
CrCuNiPd	0.909	0.090	0	none	
CrCuNiPt	0.181	0.818	0	fcc	
CrCuOsPd	0.818	0.181	0	none	
CrCuOsPt	0.636	0.363	0	none	
CrCuPdPt	0.272	0.727	0	fcc	
CrCuPdRh	0.181	0.818	0	fcc	
CrCuPdRu	0.818	0.181	0	none	
CrCuPtRh	0.181	0.818	0	fcc	
CrCuPtRu	0.636	0.363	0	none	
CrCuTaTi	0.909	0	0.090	none	
CrCuTaV	0.818	0	0.181	none	
CrCuTaW	0.727	0	0.272	none	
CrCuTiV	0.818	0	0.181	none	
CrCuTiW	0.818	0	0.181	none	
CrCuVW	0.636	0	0.363	none	
CrFeMnMo	0.454	0	0.545	bcc	
CrFeMnNb	0.909	0	0.090	none	
CrFeMnNi	0.363	0.636	0	fcc	
CrFeMnRe	0.909	0	0.090	none	
CrFeMnTa	0.818	0	0.181	none	
CrFeMnTi	0.909	0	0.090	none	
CrFeMnV	0.454	0	0.545	bcc	
CrFeMnW	0.636	0	0.363	none	
CrFeMoNb	0.727	0	0.272	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrFeMoNi	1	0	0	none	
CrFeMoRe	0.636	0	0.363	none	
CrFeMoTa	0.909	0	0.090	none	
CrFeMoTi	0.363	0.090	0.545	bcc	
CrFeMoV	0.454	0	0.545	bcc	
CrFeMoW	0.818	0	0.181	none	
CrFeNbRe	0.909	0	0.090	none	
CrFeNbTa	0.818	0	0.181	none	
CrFeNbTi	0.818	0	0.181	none	
CrFeNbV	0.727	0	0.272	none	
CrFeNbW	0.818	0	0.181	none	
CrFeNiPd	0.090	0.909	0	fcc	
CrFeNiRe	0.909	0.090	0	none	
CrFeNiV	1	0	0	none	
CrFeOsPd	0.636	0.363	0	none	
CrFePdRu	0.636	0.363	0	none	
CrFeReTa	0.909	0	0.090	none	
CrFeReTi	0.909	0	0.090	none	
CrFeReV	0.636	0	0.363	none	
CrFeReW	0.727	0	0.272	none	
CrFeTaTi	0.818	0	0.181	none	
CrFeTaV	0.636	0	0.363	none	
CrFeTaW	0.818	0	0.181	none	
CrFeTiV	0.636	0.090	0.272	none	
CrFeTiW	0.818	0	0.181	none	
CrFeVW	0.727	0	0.272	none	
CrHfMoNb	0.272	0	0.727	bcc	
CrHfMoRe	0.545	0	0.454	none	
CrHfMoTa	0.727	0	0.272	none	
CrHfMoTi	0.272	0	0.727	bcc	
CrHfMoV	0.181	0	0.818	bcc	
CrHfMoW	0.818	0	0.181	none	
CrHfMoZr	0.727	0	0.272	none	
CrHfNbRe	0.636	0	0.363	none	
CrHfNbTa	0.454	0	0.545	bcc	
CrHfNbTi	0.272	0	0.727	bcc	
CrHfNbV	0.363	0	0.636	bcc	
CrHfNbW	0.636	0	0.363	none	
CrHfNbZr	0.727	0	0.272	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrHfReTa	0.636	0	0.363	none	
CrHfReTi	0.636	0	0.363	none	
CrHfReV	0.545	0	0.454	none	
CrHfReW	0.636	0	0.363	none	
CrHfReZr	0.818	0	0.181	none	
CrHfTaTi	0.454	0	0.545	bcc	
CrHfTaV	0.545	0	0.454	none	
CrHfTaW	0.727	0	0.272	none	
CrHfTaZr	0.727	0	0.272	none	
CrHfTiV	0.272	0	0.727	bcc	
CrHfTiW	0.545	0	0.454	none	
CrHfTiZr	0.454	0	0.545	bcc	
CrHfVW	0.636	0	0.363	none	
CrHfVZr	0.363	0	0.636	bcc	
CrHfWZr	0.818	0	0.181	none	
CrIrNiPd	0.181	0.818	0	fcc	
CrIrNiPt	0.181	0.818	0	fcc	
CrIrOsPd	0.636	0.363	0	none	
CrIrOsPt	0.545	0.454	0	none	
CrIrPdPt	0.090	0.909	0	fcc	
CrIrPdRh	0.090	0.909	0	fcc	
CrIrPdRu	0.727	0.272	0	none	
CrIrPtRh	0.090	0.909	0	fcc	
CrIrPtRu	0.545	0.454	0	none	
CrMnMoNb	0.181	0	0.818	bcc	
CrMnMoNi	1	0	0	none	
CrMnMoRe	0.545	0	0.454	none	
CrMnMoTa	0.363	0	0.636	bcc	
CrMnMoTi	0.727	0	0.272	none	
CrMnMoV	0.181	0	0.818	bcc	
CrMnMoW	0.818	0	0.181	none	
CrMnMoZr	0.727	0	0.272	none	
CrMnNbNi	0.909	0	0.090	none	
CrMnNbRe	0.636	0	0.363	none	
CrMnNbTa	0.272	0	0.727	bcc	
CrMnNbTi	0.090	0	0.909	bcc	
CrMnNbV	0.272	0	0.727	bcc	
CrMnNbW	0.545	0	0.454	none	
CrMnNbZr	0.727	0	0.272	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrMnNiRe	0.909	0	0.090	none	
CrMnNiV	0.545	0.090	0.363	none	
CrMnReTa	0.727	0	0.272	none	
CrMnReTi	0.909	0	0.090	none	
CrMnReV	0.818	0	0.181	none	
CrMnReW	0.818	0	0.181	none	
CrMnReZr	0.909	0	0.090	none	
CrMnTaTi	0.181	0	0.818	bcc	
CrMnTaV	0.181	0	0.818	bcc	
CrMnTaW	0.545	0	0.454	none	
CrMnTaZr	0.727	0	0.272	none	
CrMnTiV	0.636	0.090	0.272	none	
CrMnTiW	0.727	0	0.272	none	
CrMnTiZr	0.818	0	0.181	none	
CrMnVW	0.545	0	0.454	none	
CrMnWZr	0.636	0	0.363	none	
CrMoNbNi	0.909	0	0.090	none	
CrMoNbRe	0.363	0	0.636	bcc	
CrMoNbTa	0.545	0	0.454	none	
CrMoNbTi	0.090	0	0.909	bcc	
CrMoNbV	0.545	0	0.454	none	
CrMoNbW	0.545	0	0.454	none	
CrMoNbZr	0.545	0	0.454	none	
CrMoNiRe	1	0	0	none	
CrMoReTa	0.545	0	0.454	none	
CrMoReTi	0.181	0	0.818	bcc	
CrMoReV	0	0	1	bcc	
CrMoReW	0.454	0	0.545	bcc	
CrMoReZr	0.727	0	0.272	none	
CrMoTaTi	0.272	0	0.727	bcc	
CrMoTaV	0.454	0	0.545	bcc	
CrMoTaW	0.545	0	0.454	none	
CrMoTaZr	0.727	0	0.272	none	
CrMoTiV	0.090	0	0.909	bcc	
CrMoTiW	0.090	0	0.909	bcc	
CrMoTiZr	0.363	0	0.636	bcc	
CrMoVW	0.181	0	0.818	bcc	
CrMoVZr	0.545	0	0.454	none	
CrMoWZr	0.818	0	0.181	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrNbNiV	0.818	0	0.181	none	
CrNbReTa	0.454	0	0.545	bcc	
CrNbReTi	0.090	0	0.909	bcc	
CrNbReV	0.090	0	0.909	bcc	
CrNbReW	0.363	0	0.636	bcc	
CrNbReZr	0.636	0	0.363	none	
CrNbTaTi	0.454	0	0.545	bcc	
CrNbTaV	0.090	0	0.909	bcc	
CrNbTaW	0.636	0	0.363	none	
CrNbTaZr	0.727	0	0.272	none	
CrNbTiV	0	0	1	bcc	
CrNbTiW	0.090	0	0.909	bcc	
CrNbTiZr	0.636	0	0.363	none	
CrNbVW	0.363	0	0.636	bcc	
CrNbVZr	0.545	0	0.454	none	
CrNbWZr	0.818	0	0.181	none	
CrNiOsPd	0.818	0.181	0	none	
CrNiOsPt	0.727	0.272	0	none	
CrNiPdPt	0.181	0.818	0	fcc	
CrNiPdRh	0.181	0.727	0.090	fcc	
CrNiPdRu	0.909	0.090	0	none	
CrNiPtRh	0.272	0.727	0	fcc	
CrNiPtRu	0.727	0.272	0	none	
CrNiReV	0.636	0	0.363	none	
CrOsPdPt	0.545	0.454	0	none	
CrOsPdRh	0.545	0.454	0	none	
CrOsPdRu	0.727	0.272	0	none	
CrOsPtRh	0.636	0.363	0	none	
CrOsPtRu	0.454	0.545	0	fcc	
CrPdPtRh	0.090	0.909	0	fcc	
CrPdPtRu	0.545	0.454	0	none	
CrPdRhRu	0.636	0.363	0	none	
CrPtRhRu	0.545	0.454	0	none	
CrReTaTi	0.272	0	0.727	bcc	
CrReTaV	0.181	0	0.818	bcc	
CrReTaW	0.454	0	0.545	bcc	
CrReTaZr	0.636	0	0.363	none	
CrReTiV	0.636	0	0.363	none	
CrReTiW	0.090	0	0.909	bcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrReTiZr	0.727	0	0.272	none	
CrReVW	0.363	0	0.636	bcc	
CrReVZr	0.636	0	0.363	none	
CrReWZr	0.727	0	0.272	none	
CrTaTiV	0.181	0	0.818	bcc	
CrTaTiW	0.181	0	0.818	bcc	
CrTaTiZr	0.727	0	0.272	none	
CrTaVW	0.636	0	0.363	none	
CrTaVZr	0.727	0	0.272	none	
CrTaWZr	0.818	0	0.181	none	
CrTiVW	0.090	0	0.909	bcc	
CrTiVZr	0.636	0	0.363	none	
CrTiWZr	0.727	0	0.272	none	
CrVWZr	0.727	0	0.272	none	
CuFeIrPd	0.181	0.818	0	fcc	
CuFeMnNb	1	0	0	none	
CuFeMnTa	0.909	0.090	0	none	
CuFeMnTi	1	0	0	none	
CuFeMnV	0.909	0	0.090	none	
CuFeMoTi	0.909	0	0.090	none	
CuFeMoV	0.909	0	0.090	none	
CuFeNbTa	1	0	0	none	
CuFeNbTi	0.909	0	0.090	none	
CuFeNbV	0.909	0	0.090	none	
CuFeNiPd	0	0.818	0.181	fcc	
CuFeNiV	0.909	0.090	0	none	
CuFeOsPd	0.727	0.272	0	none	
CuFePdRe	0.909	0.090	0	none	
CuFePdRh	0.181	0.818	0	fcc	
CuFePdRu	0.727	0.272	0	none	
CuFeTaV	0.909	0	0.090	none	
CuFeTiV	0.818	0	0.181	none	
CuFeVW	0.636	0	0.363	none	
CuIrNiPd	0.090	0.818	0.090	fcc	
CuIrNiPt	0.090	0.909	0	fcc	
CuIrOsPd	0.545	0.454	0	none	
CuIrOsPt	0.090	0.909	0	fcc	
CuIrPdPt	0.090	0.909	0	fcc	
CuIrPdRe	0.818	0.181	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CuIrPdRh	0.090	0.909	0	fcc	
CuIrPdRu	0.636	0.363	0	none	
CuIrPtRe	0.363	0.636	0	fcc	
CuIrPtRh	0.090	0.909	0	fcc	
CuIrPtRu	0.090	0.909	0	fcc	
CuMnMoTa	0.909	0.090	0	none	
CuMnMoV	0.909	0	0.090	none	
CuMnMoW	0.909	0	0.090	none	
CuMnNbTa	0.727	0.090	0.181	none	
CuMnNbV	0.818	0	0.181	none	
CuMnTaV	0.545	0	0.454	none	
CuMnTaW	0.727	0.090	0.181	none	
CuMnTiV	0.818	0	0.181	none	
CuMnVW	0.636	0.090	0.272	none	
CuMoNbTa	0.727	0	0.272	none	
CuMoNbV	0.818	0	0.181	none	
CuMoNbW	0.727	0.090	0.181	none	
CuMoTaTi	0.818	0	0.181	none	
CuMoTaW	0.727	0	0.272	none	
CuMoTiV	0.818	0	0.181	none	
CuMoVW	0.727	0	0.272	none	
CuNbPdPt	0.727	0.181	0.090	none	
CuNbPdRh	0.909	0.090	0	none	
CuNbPtRh	0.636	0.272	0.090	none	
CuNbTaTi	0.727	0.090	0.181	none	
CuNbTaV	0.727	0.090	0.181	none	
CuNbTaW	0.636	0	0.363	none	
CuNbTiV	0.727	0	0.272	none	
CuNbVW	0.636	0.090	0.272	none	
CuNiOsPd	0.818	0.181	0	none	
CuNiOsPt	0.636	0.363	0	none	
CuNiPdPt	0	1	0	fcc	
CuNiPdRe	0.909	0.090	0	none	
CuNiPdRh	0.090	0.909	0	fcc	
CuNiPdRu	0.636	0.363	0	none	
CuNiPtRe	0.909	0.090	0	none	
CuNiPtRh	0.090	0.909	0	fcc	
CuNiPtRu	0.272	0.727	0	fcc	
CuOsPdPt	0.454	0.545	0	fcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CuOsPdRe	0.818	0.181	0	none	
CuOsPdRh	0.090	0.909	0	fcc	
CuOsPdRu	0.636	0.363	0	none	
CuOsPtRe	0.818	0.181	0	none	
CuOsPtRh	0.454	0.545	0	fcc	
CuOsPtRu	0.454	0.545	0	fcc	
CuPdPtRe	0.818	0.181	0	none	
CuPdPtRh	0.090	0.909	0	fcc	
CuPdPtRu	0	1	0	fcc	
CuPdReRh	0.818	0.181	0	none	
CuPdReRu	0.818	0.181	0	none	
CuPdRhRu	0.545	0.454	0	none	
CuPtReRh	0.727	0.272	0	none	
CuPtReRu	0.818	0.181	0	none	
CuPtRhRu	0	1	0	fcc	
CuTaTiV	0.454	0	0.545	bcc	
CuTaTiW	0.545	0	0.454	none	
CuTaVW	0.545	0	0.454	none	
CuTiVW	0.545	0	0.454	none	
FeIrMoPd	0.363	0.636	0	fcc	
FeIrNiPd	0.090	0.909	0	fcc	
FeIrOsPd	0.636	0.363	0	none	
FeIrPdRe	0.818	0.181	0	none	
FeIrPdRu	0.636	0.363	0	none	
FeMnMoNb	0.727	0.090	0.181	none	
FeMnMoRe	0.909	0	0.090	none	
FeMnMoTa	0.909	0	0.090	none	
FeMnMoTi	0.909	0	0.090	none	
FeMnMoV	0.454	0	0.545	bcc	
FeMnMoW	0.818	0.090	0.090	none	
FeMnNbRe	0.727	0	0.272	none	
FeMnNbTa	0.909	0	0.090	none	
FeMnNbTi	0.909	0	0.090	none	
FeMnNbV	0.454	0.090	0.454	none	
FeMnNbW	0.818	0	0.181	none	
FeMnReTa	0.818	0	0.181	none	
FeMnReTi	0.909	0	0.090	none	
FeMnReV	0.818	0	0.181	none	
FeMnReW	0.818	0	0.181	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
FeMnTaTi	0.818	0	0.181	none	
FeMnTaV	0.727	0	0.272	none	
FeMnTaW	0.818	0	0.181	none	
FeMnTiV	0.818	0	0.181	none	
FeMnTiW	0.818	0	0.181	none	
FeMnVW	0.636	0	0.363	none	
FeMoNbRe	0.454	0	0.545	bcc	
FeMoNbTa	0.545	0	0.454	none	
FeMoNbTi	0.727	0	0.272	none	
FeMoNbV	0.363	0	0.636	bcc	
FeMoNbW	0.727	0	0.272	none	
FeMoNiPd	0.272	0.727	0	fcc	
FeMoNiV	1	0	0	none	
FeMoOsPd	0.636	0.363	0	none	
FeMoPdRe	0.818	0.181	0	none	
FeMoPdRh	0.363	0.636	0	fcc	
FeMoPdRu	0.727	0.272	0	none	
FeMoReTa	0.909	0	0.090	none	
FeMoReTi	0.636	0	0.363	none	
FeMoReV	0.636	0	0.363	none	
FeMoReW	0.636	0	0.363	none	
FeMoTaTi	0.636	0	0.363	none	
FeMoTaV	0.727	0	0.272	none	
FeMoTaW	0.727	0	0.272	none	
FeMoTiV	0.090	0	0.909	bcc	
FeMoTiW	0.727	0	0.272	none	
FeMoVW	0.636	0	0.363	none	
FeNbReTa	0.636	0	0.363	none	
FeNbReTi	0.818	0	0.181	none	
FeNbReV	0.363	0	0.636	bcc	
FeNbReW	0.545	0	0.454	none	
FeNbTaTi	0.636	0	0.363	none	
FeNbTaV	0.545	0	0.454	none	
FeNbTaW	0.818	0	0.181	none	
FeNbTiV	0.636	0.090	0.272	none	
FeNbTiW	0.636	0	0.363	none	
FeNbVW	0.636	0	0.363	none	
FeNiOsPd	0.727	0.272	0	none	
FeNiPdRe	0.727	0.272	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
FeNiPdRh	0.181	0.818	0	fcc	
FeNiPdRu	0.636	0.363	0	none	
FeOsPdRe	0.818	0.181	0	none	
FeOsPdRh	0.636	0.363	0	none	
FeOsPdRu	0.636	0.363	0	none	
FePdReRh	0.818	0.181	0	none	
FePdReRu	0.727	0.272	0	none	
FePdRhRu	0.636	0.363	0	none	
FeReTaTi	0.727	0	0.272	none	
FeReTaV	0.727	0	0.272	none	
FeReTiV	0.636	0	0.363	none	
FeReVW	0.636	0	0.363	none	
FeTaTiV	0.727	0	0.272	none	
FeTaTiW	0.727	0	0.272	none	
FeTaVW	0.818	0	0.181	none	
FeTiVW	0.545	0.090	0.363	none	
HfMoNbRe	0.363	0	0.636	bcc	
HfMoNbTa	0.272	0	0.727	bcc	
HfMoNbTi	0.090	0	0.909	bcc	
HfMoNbV	0.181	0	0.818	bcc	
HfMoNbW	0.454	0	0.545	bcc	
HfMoNbZr	0.272	0	0.727	bcc	
HfMoReTa	0.636	0	0.363	none	
HfMoReTi	0.363	0	0.636	bcc	
HfMoReV	0.363	0	0.636	bcc	
HfMoReW	0.545	0	0.454	none	
HfMoReZr	0.636	0	0.363	none	
HfMoTaTi	0.363	0	0.636	bcc	
HfMoTaV	0.545	0	0.454	none	
HfMoTaW	0.454	0	0.545	bcc	
HfMoTaZr	0.454	0	0.545	bcc	
HfMoTiV	0.090	0	0.909	bcc	
HfMoTiW	0.363	0	0.636	bcc	
HfMoTiZr	0.272	0	0.727	bcc	
HfMoVW	0.545	0	0.454	none	
HfMoVZr	0.545	0	0.454	none	
HfMoWZr	0.909	0	0.090	none	
HfNbReTa	0.454	0	0.545	bcc	
HfNbReTi	0.272	0	0.727	bcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
HfNbReV	0.272	0	0.727	bcc	
HfNbReW	0.363	0	0.636	bcc	
HfNbReZr	0.545	0	0.454	none	
HfNbTaTi	0	0	1	bcc	
HfNbTaV	0.272	0	0.727	bcc	
HfNbTaW	0.272	0	0.727	bcc	
HfNbTaZr	0	0	1	bcc	
HfNbTiV	0	0	1	bcc	
HfNbTiW	0.181	0	0.818	bcc	
HfNbTiZr	0	0	1	bcc	bcc
HfNbVW	0.545	0	0.454	none	
HfNbVZr	0.181	0	0.818	bcc	
HfNbWZr	0.272	0	0.727	bcc	
HfReTaTi	0.454	0	0.545	bcc	
HfReTaV	0.454	0	0.545	bcc	
HfReTaW	0.545	0	0.454	none	
HfReTaZr	0.545	0	0.454	none	
HfReTiV	0.636	0	0.363	none	
HfReTiW	0.545	0	0.454	none	
HfReTiZr	0.909	0	0.090	none	
HfReVW	0.454	0	0.545	bcc	
HfReVZr	0.909	0	0.090	none	
HfReWZr	0.727	0	0.272	none	
HfTaTiV	0.181	0	0.818	bcc	
HfTaTiW	0.272	0	0.727	bcc	
HfTaTiZr	0	0	1	bcc	
HfTaVW	0.363	0	0.636	bcc	
HfTaVZr	0.090	0	0.909	bcc	
HfTaWZr	0.272	0	0.727	bcc	
HfTiVW	0.181	0	0.818	bcc	
HfTiVZr	0.181	0	0.818	bcc	
HfTiWZr	0.727	0	0.272	none	
HfVWZr	0.818	0	0.181	none	
IrMoNiPd	0.454	0.545	0	fcc	
IrMoNiPt	0.272	0.727	0	fcc	
IrMoOsPd	0.545	0.454	0	none	
IrMoOsPt	0.636	0.363	0	none	
IrMoPdPt	0.272	0.636	0.090	fcc	
IrMoPdRe	0.636	0.363	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
IrMoPdRh	0	0.727	0.272	fcc	
IrMoPdRu	0.545	0.454	0	none	
IrMoPtRe	0.636	0.363	0	none	
IrMoPtRh	0.090	0.727	0.181	fcc	
IrMoPtRu	0.272	0.636	0.090	fcc	
IrNbPdPt	0.727	0.272	0	none	
IrNbPdRh	0.363	0.636	0	fcc	
IrNbPdRu	0.636	0.363	0	none	
IrNbPtRh	0.545	0.454	0	none	
IrNbPtRu	0.454	0.545	0	fcc	
IrNiOsPd	0.636	0.363	0	none	
IrNiOsPt	0.272	0.727	0	fcc	
IrNiPdPt	0	1	0	fcc	
IrNiPdRe	0.818	0.181	0	none	
IrNiPdRh	0.090	0.909	0	fcc	
IrNiPdRu	0.363	0.636	0	fcc	
IrNiPtRe	0.272	0.727	0	fcc	
IrNiPtRh	0.090	0.909	0	fcc	
IrNiPtRu	0.272	0.727	0	fcc	
IrOsPdPt	0.181	0.818	0	fcc	
IrOsPdRe	0.636	0.363	0	none	
IrOsPdRh	0.181	0.818	0	fcc	
IrOsPdRu	0.545	0.454	0	none	
IrOsPtRe	0.545	0.454	0	none	
IrOsPtRh	0	0.909	0.090	fcc	
IrOsPtRu	0.272	0.727	0	fcc	
IrPdPtRe	0.090	0.909	0	fcc	
IrPdPtRh	0.090	0.909	0	fcc	
IrPdPtRu	0.272	0.727	0	fcc	
IrPdReRh	0.181	0.727	0.090	fcc	
IrPdReRu	0.636	0.363	0	none	
IrPdRhRu	0.272	0.727	0	fcc	
IrPtReRh	0	1	0	fcc	
IrPtReRu	0.090	0.909	0	fcc	
IrPtRhRu	0	0.909	0.090	fcc	
MnMoNbRe	0.545	0	0.454	none	
MnMoNbTa	0.272	0	0.727	bcc	
MnMoNbTi	0.090	0	0.909	bcc	
MnMoNbV	0.090	0	0.909	bcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
MnMoNbW	0.545	0	0.454	none	
MnMoNbZr	0.727	0	0.272	none	
MnMoNiV	0.727	0.181	0.090	none	
MnMoReTa	0.636	0	0.363	none	
MnMoReTi	0.727	0	0.272	none	
MnMoReV	0.454	0	0.545	bcc	
MnMoReW	0.818	0	0.181	none	
MnMoReZr	0.636	0	0.363	none	
MnMoTaTi	0.090	0	0.909	bcc	
MnMoTaV	0.454	0	0.545	bcc	
MnMoTaW	0.545	0	0.454	none	
MnMoTaZr	0.818	0	0.181	none	
MnMoTiV	0.090	0	0.909	bcc	
MnMoTiW	0.636	0	0.363	none	
MnMoTiZr	0.636	0	0.363	none	
MnMoVW	0.454	0	0.545	bcc	
MnMoVZr	0.545	0	0.454	none	
MnMoWZr	0.818	0	0.181	none	
MnNbNiV	0.909	0	0.090	none	
MnNbReTa	0.545	0	0.454	none	
MnNbReTi	0.272	0	0.727	bcc	
MnNbReV	0.181	0	0.818	bcc	
MnNbReW	0.454	0	0.545	bcc	
MnNbReZr	0.363	0	0.636	bcc	
MnNbTaTi	0.272	0	0.727	bcc	
MnNbTaV	0.181	0	0.818	bcc	
MnNbTaW	0.727	0	0.272	none	
MnNbTaZr	0.636	0	0.363	none	
MnNbTiV	0.181	0	0.818	bcc	
MnNbTiW	0.636	0	0.363	none	
MnNbTiZr	0.545	0.090	0.363	none	
MnNbVW	0.363	0	0.636	bcc	
MnNbVZr	0.636	0	0.363	none	
MnNbWZr	0.818	0	0.181	none	
MnNiOsPd	0.909	0.090	0	none	
MnNiPdRe	0.909	0.090	0	none	
MnNiPdRu	0.909	0.090	0	none	
MnOsPdRe	0.636	0.363	0	none	
MnOsPdRu	0.727	0.272	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
MnPdReRu	0.636	0.363	0	none	
MnReTaTi	0.272	0	0.727	bcc	
MnReTaV	0.272	0	0.727	bcc	
MnReTaW	0.636	0	0.363	none	
MnReTaZr	0.545	0	0.454	none	
MnReTiV	0.636	0	0.363	none	
MnReTiW	0.636	0	0.363	none	
MnReTiZr	0.727	0	0.272	none	
MnReVW	0.545	0	0.454	none	
MnReVZr	0.727	0	0.272	none	
MnReWZr	0.818	0	0.181	none	
MnTaTiV	0.090	0	0.909	bcc	
MnTaTiW	0.636	0	0.363	none	
MnTaVW	0.545	0	0.454	none	
MnTaVZr	0.727	0	0.272	none	
MnTaWZr	0.818	0	0.181	none	
MnTiVW	0.363	0	0.636	bcc	
MnTiVZr	0.636	0	0.363	none	
MnTiWZr	0.636	0	0.363	none	
MnVWZr	0.727	0	0.272	none	
MoNbNiV	0.909	0	0.090	none	
MoNbReTa	0	0	1	bcc	
MoNbReTi	0	0	1	bcc	
MoNbReV	0	0	1	bcc	
MoNbReW	0.090	0	0.909	bcc	
MoNbReZr	0.363	0	0.636	bcc	
MoNbTaTi	0	0	1	bcc	
MoNbTaV	0	0	1	bcc	
MoNbTaW	0	0	1	bcc	bcc
MoNbTaZr	0.454	0	0.545	bcc	
MoNbTiV	0	0	1	bcc	
MoNbTiW	0	0	1	bcc	
MoNbTiZr	0	0	1	bcc	
MoNbVW	0	0	1	bcc	
MoNbVZr	0	0	1	bcc	
MoNbWZr	0.545	0	0.454	none	
MoNiOsPd	0.818	0.181	0	none	
MoNiOsPt	0.727	0.272	0	none	
MoNiPdPt	0.272	0.727	0	fcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
MoNiPdRe	0.818	0.181	0	none	
MoNiPdRh	0.272	0.636	0.090	fcc	
MoNiPdRu	0.818	0.181	0	none	
MoNiPtRe	0.818	0.181	0	none	
MoNiPtRh	0.272	0.727	0	fcc	
MoNiPtRu	0.818	0.181	0	none	
MoNiReV	1	0	0	none	
MoOsPdPt	0.636	0.363	0	none	
MoOsPdRe	0.727	0.272	0	none	
MoOsPdRh	0.545	0.454	0	none	
MoOsPdRu	0.545	0.454	0	none	
MoOsPtRe	0.727	0.272	0	none	
MoOsPtRh	0.454	0.545	0	fcc	
MoOsPtRu	0.636	0.363	0	none	
MoPdPtRe	0.727	0.272	0	none	
MoPdPtRh	0.090	0.727	0.181	fcc	
MoPdPtRu	0.636	0.363	0	none	
MoPdReRh	0.545	0.454	0	none	
MoPdReRu	0.636	0.363	0	none	
MoPdRhRu	0.545	0.454	0	none	
MoPtReRh	0.545	0.454	0	none	
MoPtReRu	0.636	0.363	0	none	
MoPtRhRu	0.454	0.545	0	fcc	
MoReTaTi	0.272	0	0.727	bcc	
MoReTaV	0.272	0	0.727	bcc	
MoReTaW	0.272	0	0.727	bcc	
MoReTaZr	0.818	0	0.181	none	
MoReTiV	0	0	1	bcc	
MoReTiW	0.090	0	0.909	bcc	
MoReTiZr	0.272	0	0.727	bcc	
MoReVW	0.181	0	0.818	bcc	
MoReVZr	0.545	0	0.454	none	
MoReWZr	0.363	0	0.636	bcc	
MoTaTiV	0	0	1	bcc	
MoTaTiW	0	0	1	bcc	
MoTaTiZr	0.363	0	0.636	bcc	
MoTaVW	0	0	1	Bcc	
MoTaVZr	0.636	0.090	0.272	none	
MoTaWZr	0.545	0	0.454	none	

Supplementary Table 6, Continued.

	none	fcc	bcc	S.S <sup>(ML)</sup>	S.S. Exp.
MoTiVW	0	0	1	bcc	
MoTiVZr	0	0	1	bcc	
MoTiWZr	0.454	0	0.545	bcc	
MoVWZr	0.363	0	0.636	bcc	
NbPdPtRh	0.636	0.363	0	none	
NbPdRhRu	0.636	0.272	0.090	none	
NbPtRhRu	0.636	0.363	0	none	
NbReTaTi	0	0	1	bcc	
NbReTaV	0	0	1	bcc	
NbReTaW	0	0	1	bcc	
NbReTaZr	0.363	0	0.636	bcc	
NbReTiV	0	0	1	bcc	
NbReTiW	0	0	1	bcc	
NbReTiZr	0.272	0	0.727	bcc	
NbReVW	0.090	0	0.909	bcc	
NbReVZr	0.636	0	0.363	none	
NbReWZr	0.272	0	0.727	bcc	
NbTaTiV	0	0	1	bcc	bcc
NbTaTiW	0.090	0	0.909	bcc	
NbTaTiZr	0	0	1	bcc	
NbTaVW	0	0	1	bcc	
NbTaVZr	0.181	0	0.818	bcc	
NbTaWZr	0.272	0	0.727	bcc	
NbTiVW	0	0	1	bcc	
NbTiVZr	0	0	1	bcc	bcc
NbTiWZr	0.181	0	0.818	bcc	
NbVWZr	0.454	0	0.545	bcc	
NiOsPdPt	0.636	0.363	0	none	
NiOsPdRe	0.818	0.181	0	none	
NiOsPdRh	0.727	0.272	0	none	
NiOsPdRu	0.818	0.181	0	none	
NiOsPtRe	0.818	0.181	0	none	
NiOsPtRh	0.636	0.363	0	none	
NiOsPtRu	0.727	0.272	0	none	
NiPdPtRe	0.818	0.181	0	none	
NiPdPtRh	0.090	0.909	0	fcc	
NiPdPtRu	0.363	0.636	0	fcc	
NiPdReRh	0.818	0.181	0	none	
NiPdReRu	0.818	0.181	0	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
NiPdRhRu	0.727	0.272	0	none	
NiPtReRh	0.818	0.181	0	none	
NiPtReRu	0.818	0.181	0	none	
NiPtRhRu	0.272	0.727	0	fcc	
OsPdPtRe	0.454	0.545	0	fcc	
OsPdPtRh	0.363	0.636	0	fcc	
OsPdPtRu	0.545	0.454	0	none	
OsPdReRh	0.454	0.545	0	fcc	
OsPdReRu	0.636	0.363	0	none	
OsPdRhRu	0.636	0.363	0	none	
OsPtReRh	0.454	0.545	0	fcc	
OsPtReRu	0.454	0.545	0	fcc	
OsPtRhRu	0.181	0.818	0	fcc	
PdPtReRh	0.272	0.727	0	fcc	
PdPtReRu	0.454	0.545	0	fcc	
PdPtRhRu	0	1	0	fcc	
PdReRhRu	0.545	0.454	0	none	
PtReRhRu	0.454	0.545	0	fcc	
ReTaTiV	0	0	1	bcc	
ReTaTiW	0	0	1	bcc	
ReTaTiZr	0.727	0	0.272	none	
ReTaVW	0.090	0	0.909	bcc	
ReTaVZr	0.636	0	0.363	none	
ReTaWZr	0.454	0	0.545	bcc	
ReTiVW	0	0	1	bcc	
ReTiVZr	0.727	0	0.272	none	
ReTiWZr	0.363	0	0.636	bcc	
ReVWZr	0.636	0	0.363	none	
TaTiVW	0	0	1	bcc	
TaTiVZr	0.090	0	0.909	bcc	
TaTiWZr	0.272	0	0.727	bcc	
TaVWZr	0.454	0	0.545	bcc	
TiVWZr	0.272	0	0.727	bcc	
AlCrMoTiW	0.090	0	0.909	bcc	bcc
CoCrFeMnNi	0	1	0	fcc	fcc
CrNbTaVW	0.363	0	0.636	bcc	
CrNbTaVZr	0.727	0	0.272	none	
CrNbTaWZr	0.818	0	0.181	none	
CrHfMoNbTa	0.454	0	0.545	bcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S.<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrHfMoNbTi	0.090	0	0.909	bcc	
CrHfMoNbV	0.272	0	0.727	bcc	
CrHfMoNbW	0.636	0	0.363	none	
CrHfMoNbZr	0.181	0	0.818	bcc	
CrHfMoTaTi	0.363	0	0.636	bcc	
CrHfMoTaV	0.545	0	0.454	none	
CrHfMoTaW	0.727	0	0.272	none	
CrHfMoTaZr	0.818	0	0.181	none	
CrHfMoTiV	0.090	0	0.909	bcc	
CrHfMoTiW	0.272	0	0.727	bcc	
CrHfMoTiZr	0.545	0	0.454	none	
CrHfMoVW	0.454	0	0.545	bcc	
CrHfMoVZr	0.181	0	0.818	bcc	
CrHfMoWZr	0.909	0	0.090	none	
CrHfNbTaTi	0.181	0	0.818	bcc	
CrHfNbTaV	0.545	0	0.454	none	
CrHfNbTaW	0.545	0	0.454	none	
CrHfNbTaZr	0.454	0	0.545	bcc	
CrHfNbTiV	0.272	0	0.727	bcc	
CrHfNbTiW	0.363	0	0.636	bcc	
CrHfNbTiZr	0.181	0	0.818	bcc	
CrHfNbVW	0.727	0	0.272	none	
CrHfNbVZr	0.363	0	0.636	bcc	
CrHfNbWZr	0.818	0	0.181	none	
CrHfTaTiV	0.363	0	0.636	bcc	
CrHfTaTiW	0.454	0	0.545	bcc	
CrHfTaTiZr	0.545	0	0.454	none	
CrHfTaVW	0.636	0	0.363	none	
CrHfTaVZr	0.454	0	0.545	bcc	
CrHfTaWZr	0.909	0	0.090	none	
CrHfTiVW	0.181	0	0.818	bcc	
CrHfTiVZr	0.272	0	0.727	bcc	
CrHfTiWZr	0.727	0	0.272	none	
CrHfVWZr	0.727	0	0.272	none	
CrMoNbTaTi	0.181	0	0.818	bcc	
CrMoNbTaV	0.363	0	0.636	bcc	
CrMoNbTaW	0.181	0	0.818	bcc	
CrMoNbTaZr	0.545	0	0.454	none	
CrMoNbTiV	0.090	0	0.909	bcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S.<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
CrMoNbTiW	0.090	0	0.909	bcc	
CrMoNbTiZr	0.363	0	0.636	bcc	
CrMoNbVW	0.454	0	0.545	bcc	
CrMoNbVZr	0.454	0	0.545	bcc	
CrMoNbWZr	0.727	0	0.272	none	
CrMoTaTiV	0.272	0	0.727	bcc	
CrMoTaTiW	0.090	0	0.909	bcc	
CrMoTaTiZr	0.636	0	0.363	none	
CrMoTaVW	0.363	0	0.636	bcc	
CrMoTaVZr	0.636	0	0.363	none	
CrMoTaWZr	0.909	0	0.090	none	
CrMoTiVW	0.090	0	0.909	bcc	
CrMoTiVZr	0.363	0	0.636	bcc	
CrMoTiWZr	0.636	0	0.363	none	
CrMoVWZr	0.636	0	0.363	none	
CrNbTaTiV	0.090	0	0.909	bcc	
CrNbTaTiW	0.090	0	0.909	bcc	
CrNbTaTiZr	0.636	0	0.363	none	
CrNbTiVW	0.090	0	0.909	bcc	
CrNbTiVZr	0.636	0	0.363	none	
CrNbTiWZr	0.636	0	0.363	none	
CrNbVWZr	0.727	0	0.272	none	
CrTaTiVW	0.090	0	0.909	bcc	
CrTaTiVZr	0.818	0	0.181	none	
CrTaTiWZr	0.727	0	0.272	none	
CrTaVWZr	0.727	0	0.272	none	
CrTiVWZr	0.727	0	0.272	none	
HfMoNbTaTi	0.272	0	0.727	bcc	
HfMoNbTaV	0.181	0	0.818	bcc	
HfMoNbTaW	0.272	0	0.727	bcc	
HfMoNbTaZr	0.363	0	0.636	bcc	
HfMoNbTiV	0.181	0	0.818	bcc	
HfMoNbTiW	0.090	0	0.909	bcc	
HfMoNbTiZr	0.272	0	0.727	bcc	
HfMoNbVW	0.181	0	0.818	bcc	
HfMoNbVZr	0.272	0	0.727	bcc	
HfMoNbWZr	0.545	0	0.454	none	
HfMoTaTiV	0.181	0	0.818	bcc	
HfMoTaTiW	0.272	0	0.727	bcc	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
HfMoTaTiZr	0.181	0	0.818	bcc	
HfMoTaVW	0.181	0	0.818	bcc	
HfMoTaVZr	0.454	0	0.545	bcc	
HfMoTaWZr	0.545	0	0.454	none	
HfMoTiVW	0.272	0	0.727	bcc	
HfMoTiVZr	0.090	0	0.909	bcc	
HfMoTiWZr	0.454	0	0.545	bcc	
HfMoVWZr	0.636	0	0.363	none	
HfNbTaTiV	0.181	0	0.818	bcc	
HfNbTaTiW	0.181	0	0.818	bcc	
HfNbTaTiZr	0	0	1	bcc	bcc
HfNbTaVW	0.363	0	0.636	bcc	
HfNbTaVZr	0.090	0	0.909	bcc	
HfNbTaWZr	0.181	0	0.818	bcc	
HfNbTiVW	0.454	0	0.545	bcc	
HfNbTiVZr	0.090	0	0.909	bcc	bcc
HfNbTiWZr	0.272	0	0.727	bcc	
HfNbVWZr	0.454	0	0.545	bcc	
HfTaTiVW	0.363	0	0.636	bcc	
HfTaTiVZr	0.090	0	0.909	bcc	
HfTaTiWZr	0.272	0	0.727	bcc	
HfTaVWZr	0.454	0	0.545	bcc	
HfTiVWZr	0.272	0	0.727	bcc	
MoNbReTaW	0.090	0	0.909	bcc	bcc
MoNbTaTiV	0	0	1	bcc	bcc
MoNbTaTiW	0	0	1	bcc	
MoNbTaTiZr	0.272	0	0.727	bcc	
MoNbTaVW	0	0	1	bcc	bcc
MoNbTaVZr	0.363	0	0.636	bcc	
MoNbTaWZr	0.636	0	0.363	none	
MoNbTiVW	0	0	1	bcc	
MoNbTiVZr	0	0	1	bcc	bcc
MoNbTiWZr	0.363	0	0.636	bcc	
MoNbVWZr	0.454	0	0.545	bcc	
MoTaTiVW	0	0	1	bcc	
MoTaTiVZr	0.181	0	0.818	bcc	
MoTaTiWZr	0.545	0	0.454	none	
MoTaVWZr	0.454	0	0.545	bcc	
MoTiVWZr	0.545	0	0.454	none	

Supplementary Table 6, Continued.

	<b>none</b>	<b>fcc</b>	<b>bcc</b>	<b>S.S<sup>(ML)</sup></b>	<b>S.S. Exp.</b>
NbReTaTiV	0	0	1	bcc	bcc
NbTaTiVW	0	0	1	bcc	
NbTaTiVZr	0.090	0	0.909	bcc	
NbTaTiWZr	0.181	0	0.818	bcc	
NbTaVWZr	0.545	0	0.454	none	
NbTiVWZr	0.363	0	0.636	bcc	
TaTiVWZr	0.454	0	0.545	bcc	

## **Appendix C Supplementary Information for Chapter 4**

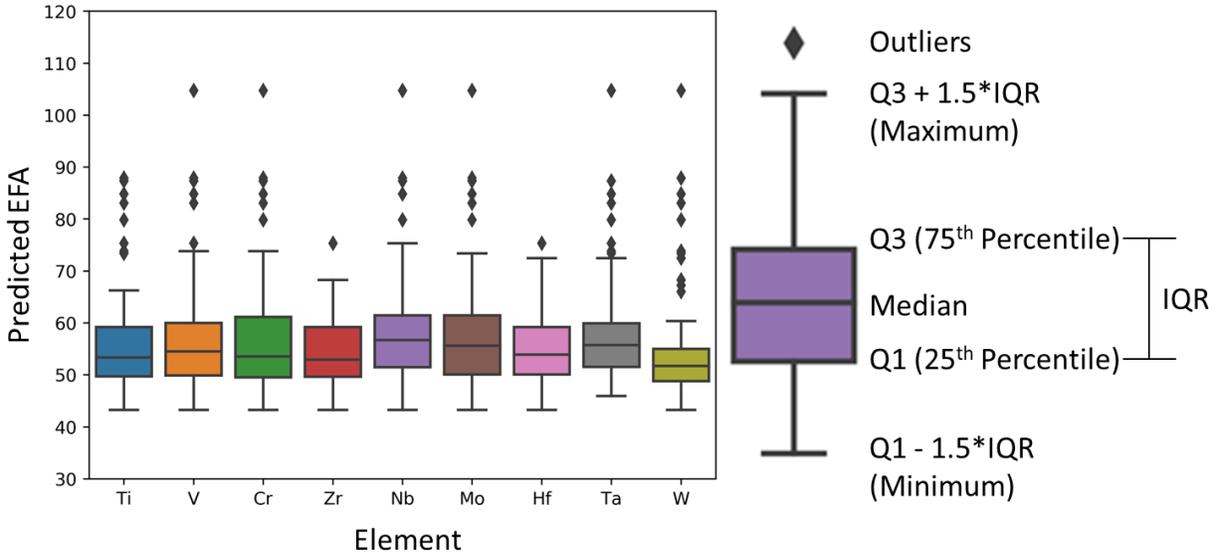
Development of ultrahigh-entropy ceramics with tailored oxidation behavior

William M. Mellor<sup>a†</sup>, Kevin Kaufmann<sup>b†</sup>, Olivia F. Dippo<sup>a</sup>, Samuel D. Figueroa<sup>b</sup>, Grant D. Schrader<sup>b</sup>, Kenneth S. Vecchio<sup>a,b‡</sup>

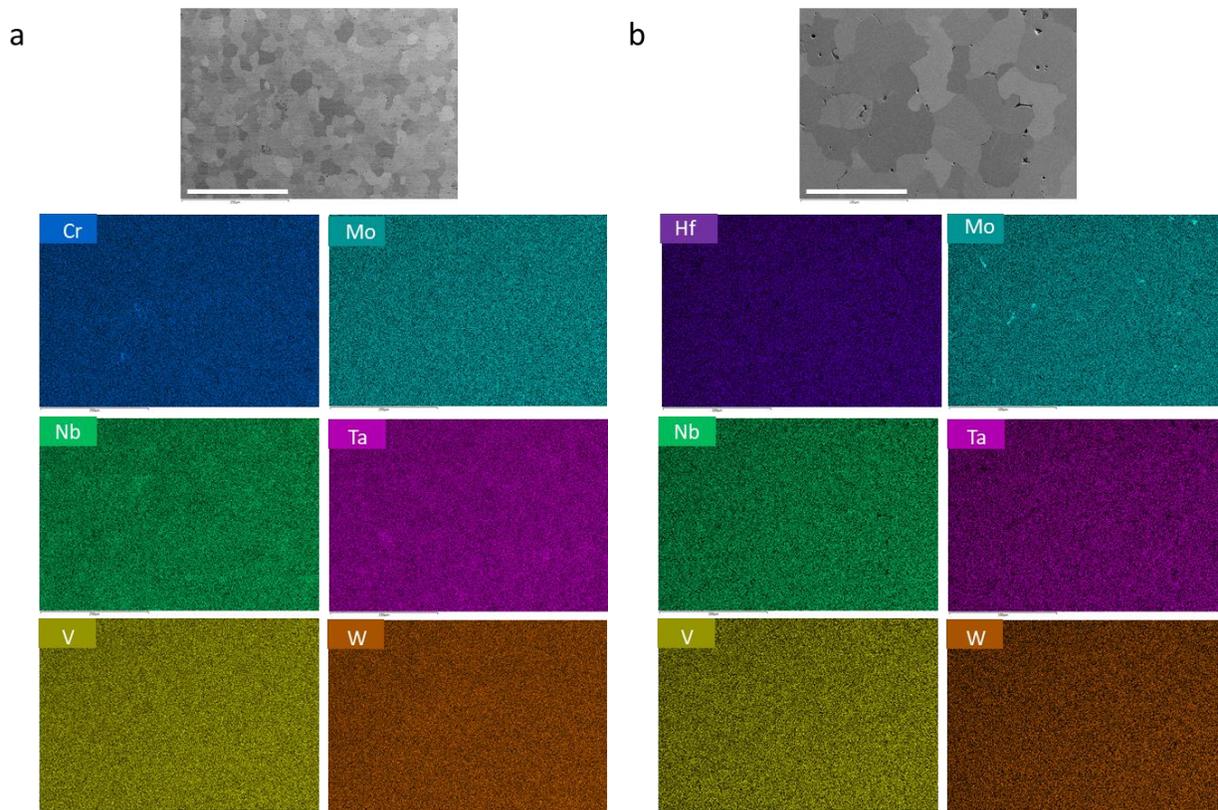
<sup>a</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA.

<sup>b</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA.

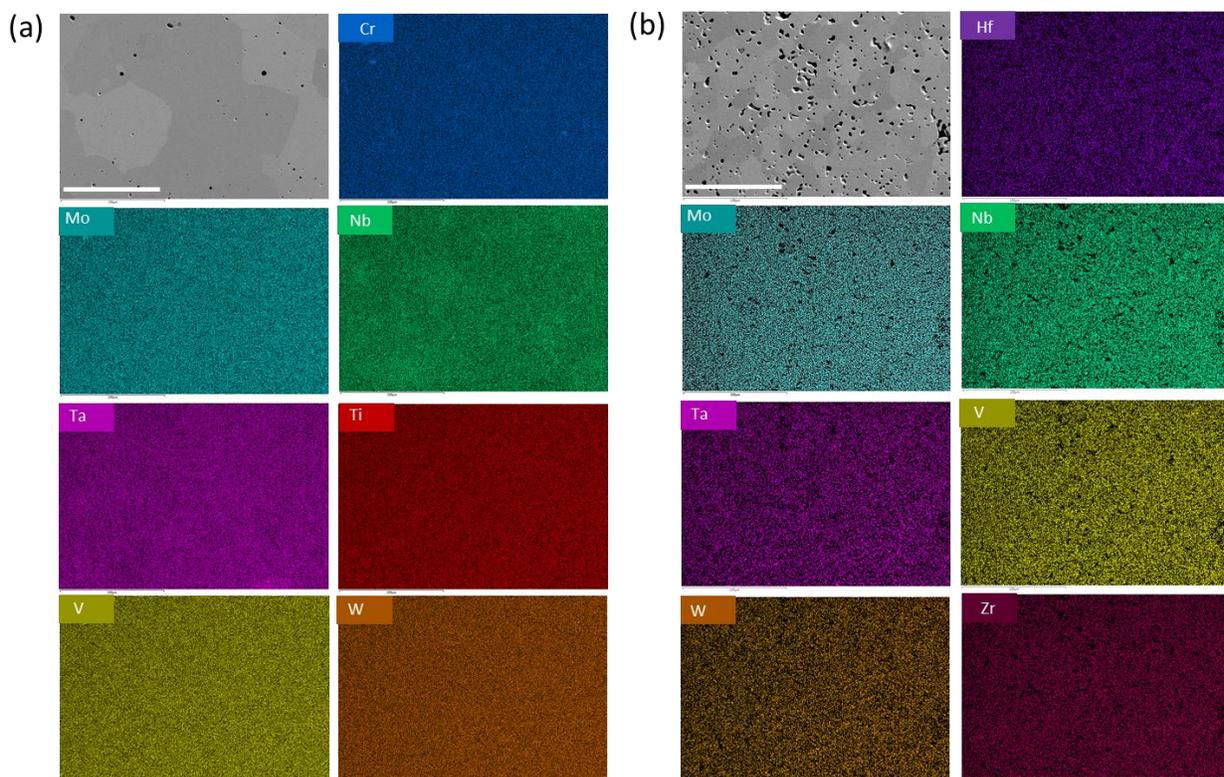
<sup>†</sup>These authors contributed equally. <sup>‡</sup>Corresponding author.



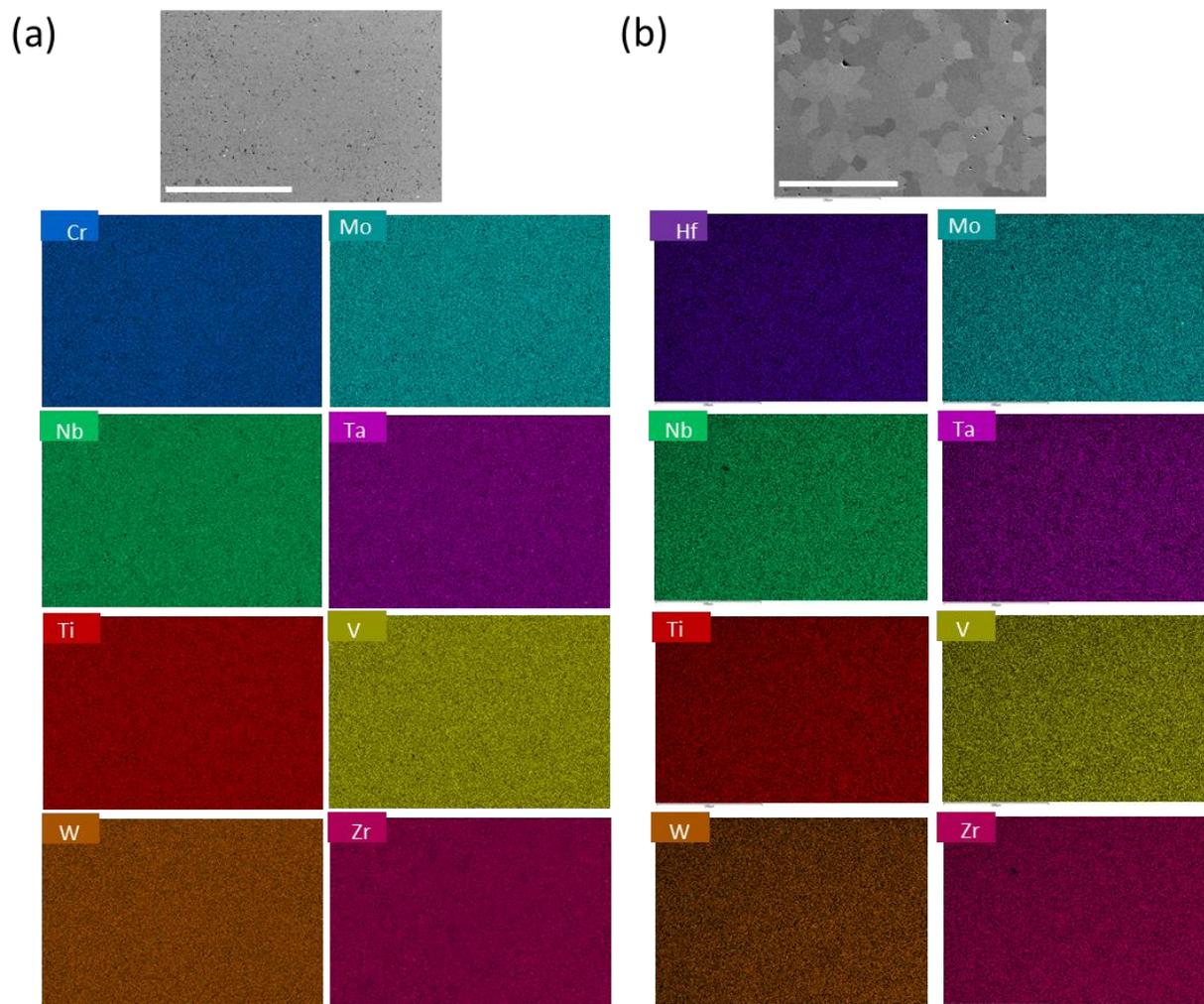
**Supplementary Figure 8 Statistical analysis of elemental impact on predicted EFA.** On the left, boxplots show the distribution of predicted EFA values for compositions containing the individual element. Notably, the ML model predicts Ti, Zr, and W to be slightly less likely to be present in compositions with a higher EFA. The box plot for each element is an analysis of 93 data points. On the right, the meaning of each component of a boxplot is denoted. Diamonds denote outliers in the dataset. The maximum is defined as the largest value in the dataset that is less than or equal to 1.5 times the interquartile range (IQR). Next is Q3, which marks the middle number between the median and the largest value in the dataset. The middle value of the data is marked by a bar within Q1 and Q3. The middle number between median and the smallest value is Q1. The range between Q1 and Q3 is called the interquartile range and describes the range of the middle 50% of the data. Lastly, the minimum marks the smallest value in the dataset that is greater than or equal to Q1 minus 1.5 times the IQR.



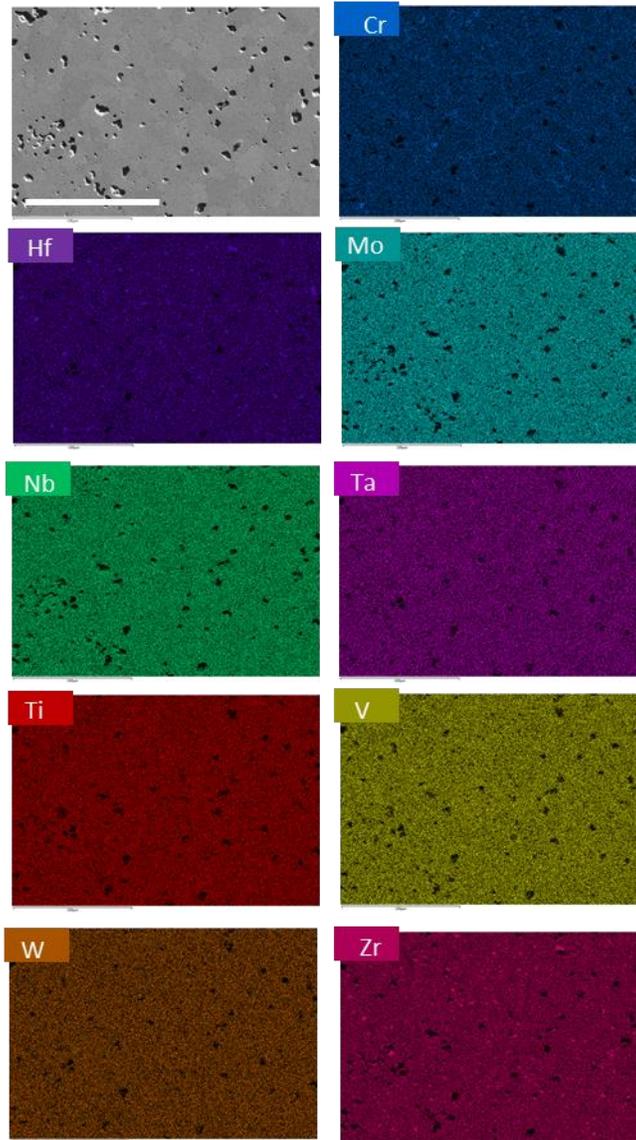
**Supplementary Figure 9 EDS maps for each experimental composition.** An electron image and chemical maps for the corresponding 6 cations are shown for UHEC compositions: (a) CrMoNbTaVWC<sub>6</sub>, (b) HfMoNbTaVWC<sub>6</sub>.



**Supplementary Figure 10 EDS maps for each experimental composition.** An electron image and chemical maps for the corresponding 7 cations are shown for UHEC compositions: (a) CrMoNbTaTiVWC<sub>7</sub>, (b) HfMoNbTaVWZrC<sub>7</sub>.



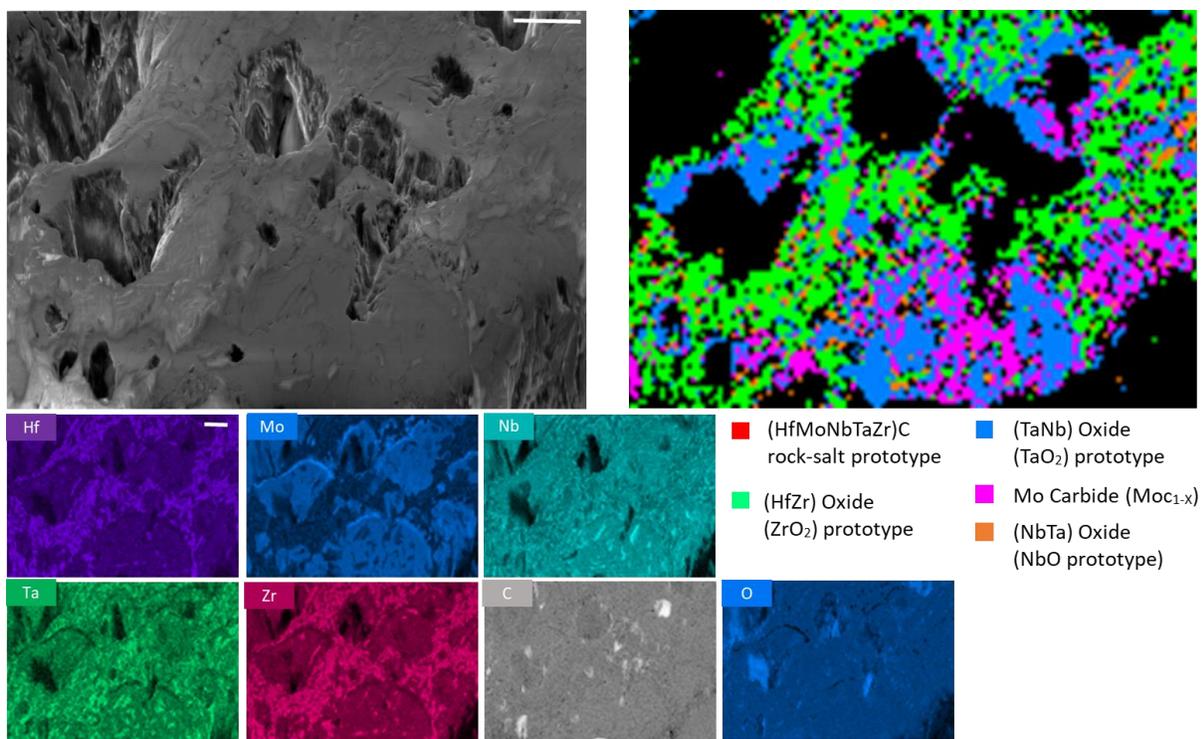
**Supplementary Figure 11 EDS maps for experimental compositions.** An electron image and chemical maps for the corresponding 8 cations are shown for UHEC compositions: (a) CrMoNbTaTiVWZrC<sub>8</sub>, (b) HfMoNbTaTiVWZrC<sub>8</sub>.



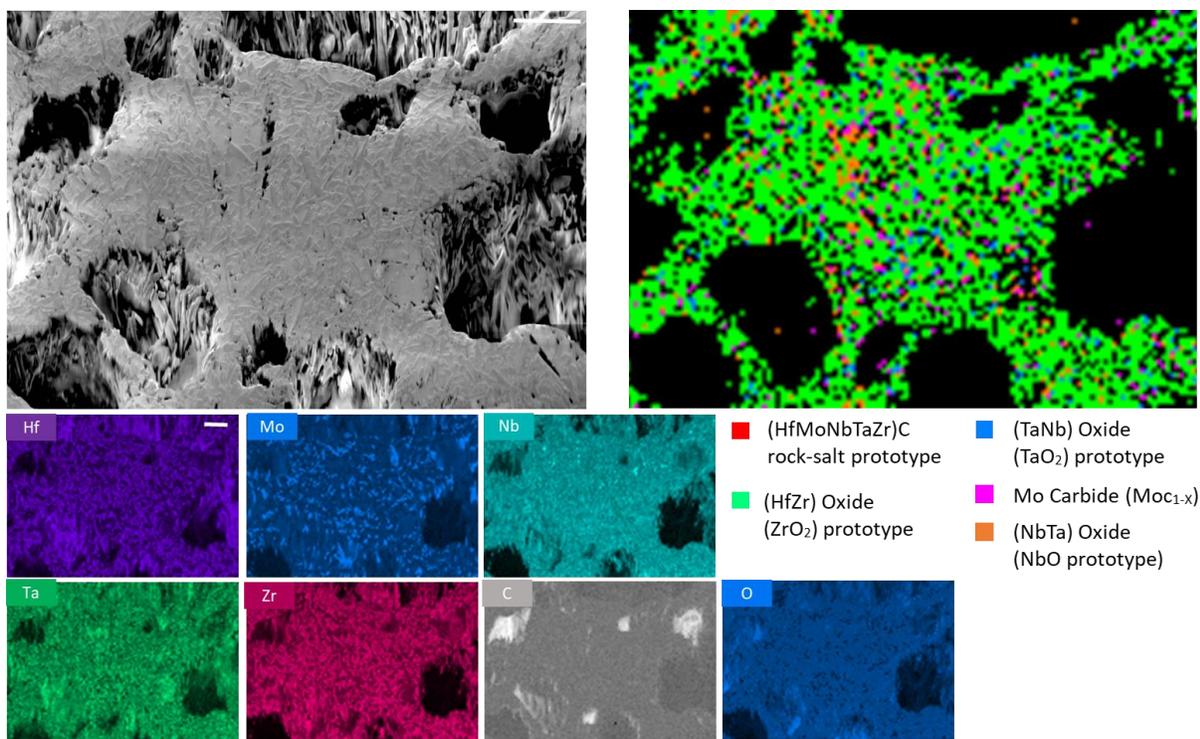
**Supplementary Figure 12 EDS maps for each experimental composition.** An electron image and chemical maps for the corresponding 9 cations are shown for each composition  $\text{CrHfMoNbTaTiVWZrC}_9$ . The small regions enriched in Hf, Zr, and Cr evident in the maps are due to small amounts of oxides present of these three elements.



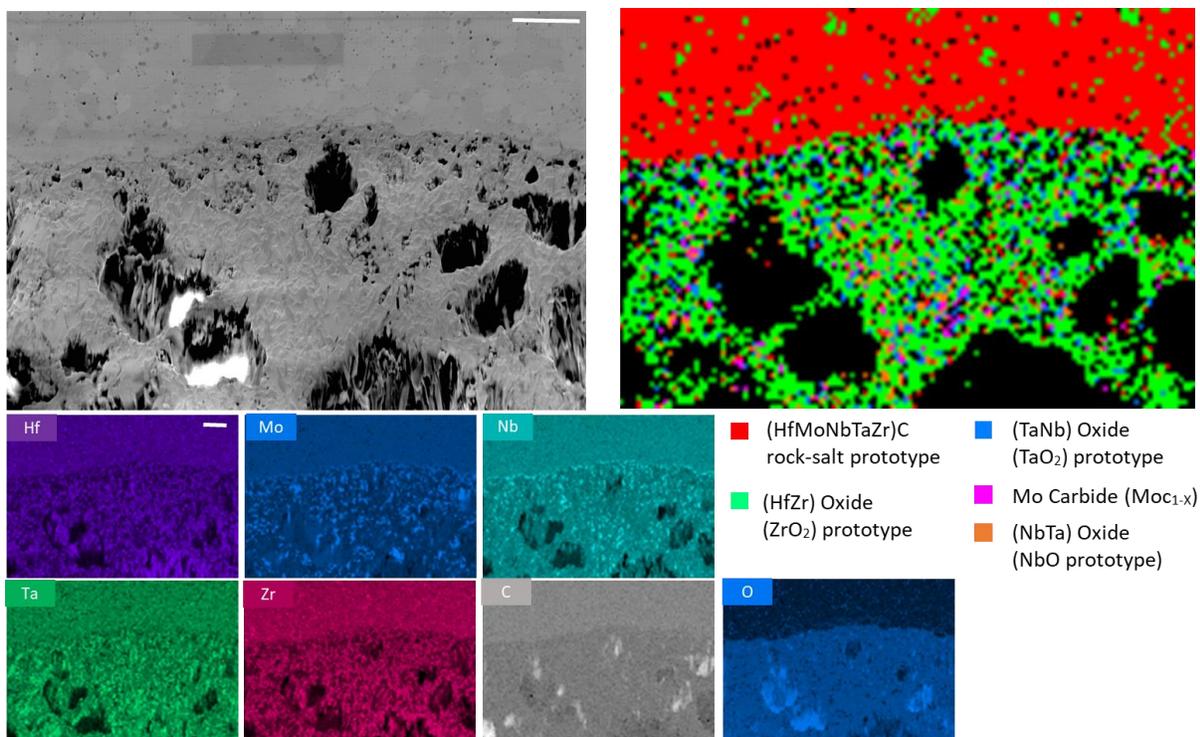
**Supplementary Figure 13 Post-oxidation images of the samples.** Macroscopic images of the two samples show differences in the oxidation behavior. The  $\text{HfMoNbTaZrC}_5$  (left) is compared with  $\text{HfMoNbTaVWZrC}_7$  (right). Scale bars 5mm.



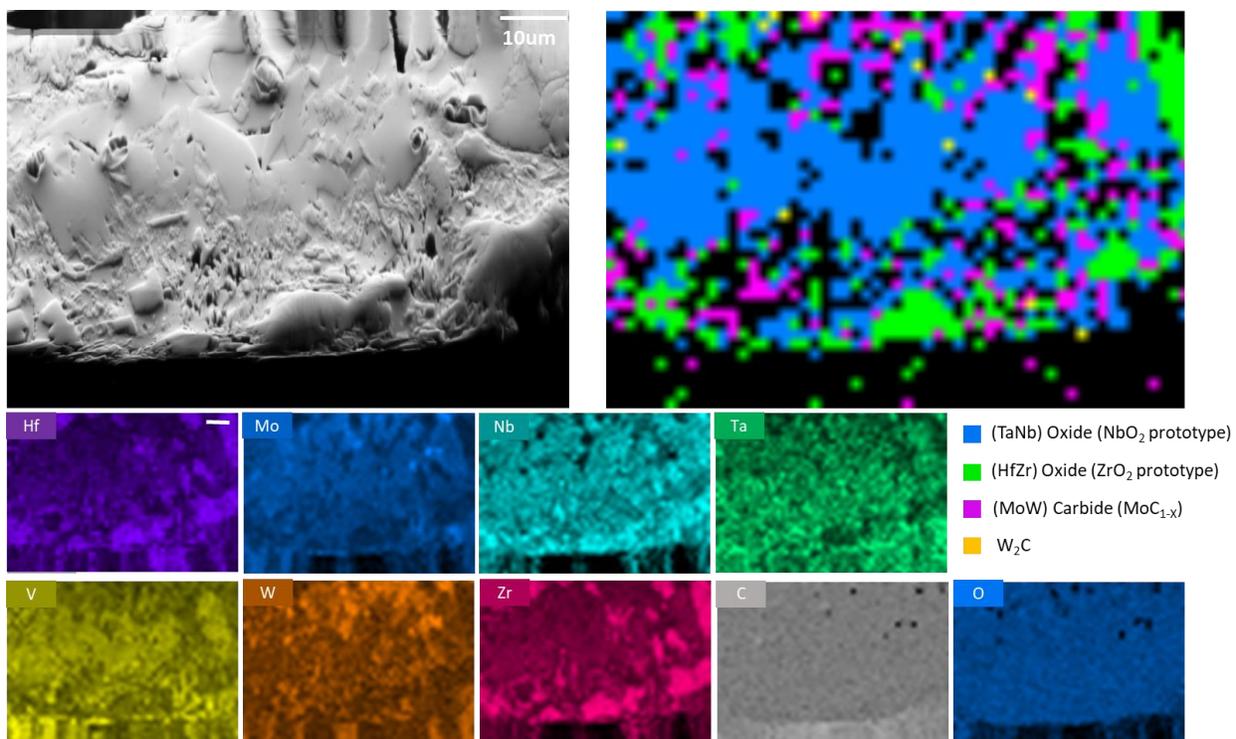
**Supplementary Figure 14 Analysis of the oxide edge and interface for HfMoNbTaZrC<sub>5</sub>.** EBSD and EDS maps for the edge of the oxide layer facing out to the environment Scale bars 25μm. Note that phases are described by their primary cations instead of crystal structure.



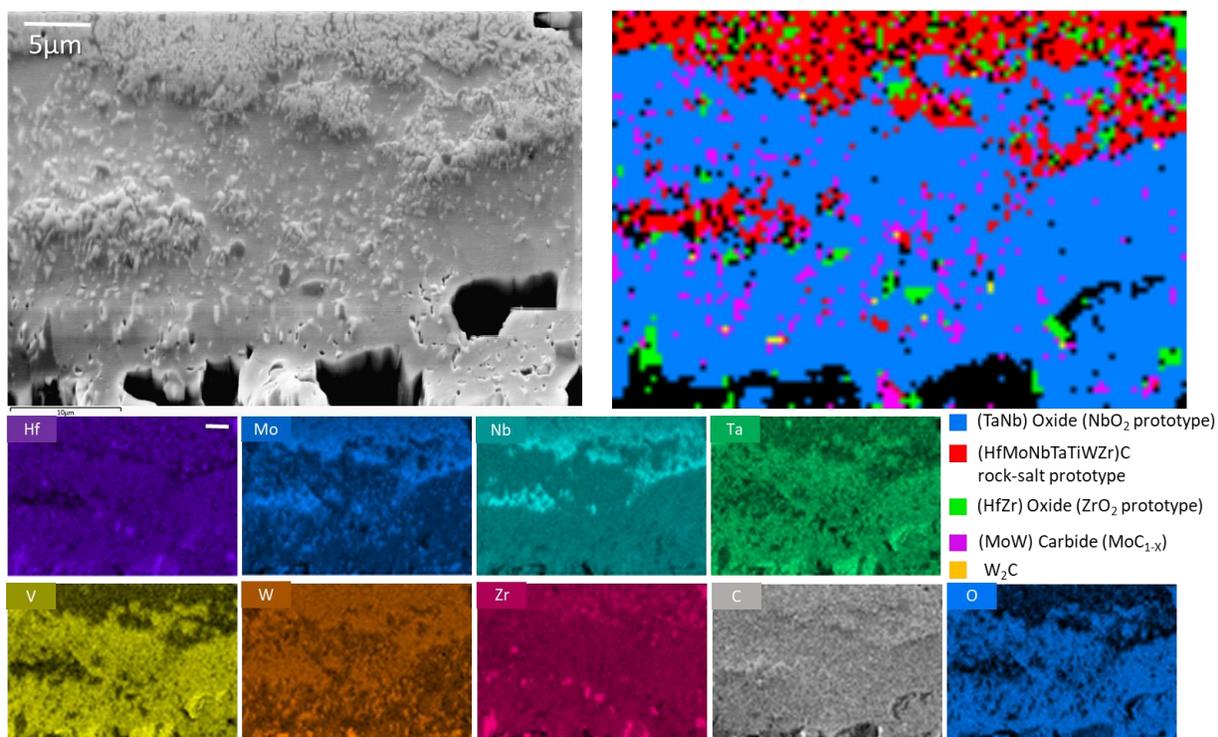
**Supplementary Figure 15** Analysis of the oxide edge and interface for  $\text{HfMoNbTaZrC}_5$ . EBSD and EDS maps for the middle of the oxide layer and (9) the interface between the oxide layer and the HEC substrate. Scale bars  $25\mu\text{m}$ . Note that phases are described by their primary cations instead of crystal structure.



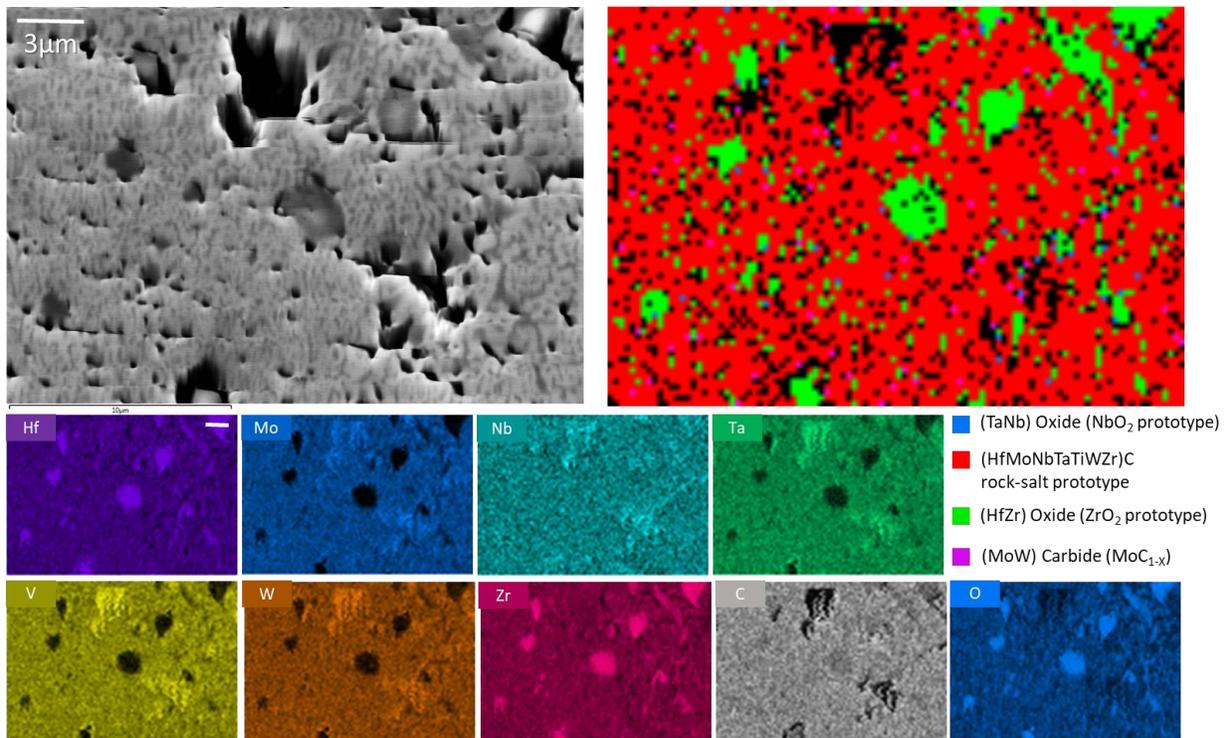
**Supplementary Figure 16 Analysis of the oxide edge and interface for HfMoNbTaZrC<sub>5</sub>.** EBSD and EDS maps for the interface between the oxide layer and the HEC substrate. Scale bars 25 $\mu$ m. Note that phases are described by their primary cations instead of crystal structure.



**Supplementary Figure 17** Analysis of the transition layers within oxidized HfMoNbTaVWZrC<sub>7</sub>. EBSD and EDS maps for the edge of the oxide layer facing out to the environment. Note the scale bar and that the phases are described by their primary cations instead of crystal structure.



**Supplementary Figure 18** Analysis of the transition layers within oxidized HfMoNbTaVWZrC<sub>7</sub>. EBSD and EDS maps for the interface between the oxide layer and the UHEC transition layer. Note scale change and that the phases are described by their primary cations instead of crystal structure.



**Supplementary Figure 19** Analysis of the transition layers within oxidized  $\text{HfMoNbTaVWZrC}_7$ . EBSD and EDS maps within the (Hf,Zr)-deficient HEC transition layer. Note scale bars vary between S10, S11, and S12 based on microstructure feature sizes. Also note that phases are described by their primary cations instead of crystal structure.

**Supplementary Table 7 Machine learning predicted EFA values for 6-9 cation compositions.** Results of the ML model's predictions are provided for each composition. Compositions in bold were selected for experimental fabrication and characterization. Units: EFA in (eV/atom)<sup>-1</sup>.

Material	# of Cations	EFA	Material	# of Cations	EFA
<b>CrMoNbTaVWC<sub>6</sub></b>	6	105	MoNbTaVWZrC <sub>6</sub>	6	54
CrMoNbTiVWC <sub>6</sub>	6	88	HfTaTiVWZrC <sub>6</sub>	6	54
CrMoNbTaTiVC <sub>6</sub>	6	87	CrHfNbTaTiVWZrC <sub>8</sub>	8	54
<b>CrMoNbTaTiVWC<sub>7</sub></b>	7	85	CrHfMoNbWZrC <sub>6</sub>	6	54
CrMoTaTiVWC <sub>6</sub>	6	83	HfNbTiVWZrC <sub>6</sub>	6	53
CrMoNbTaTiWC <sub>6</sub>	6	80	NbTaTiVWZrC <sub>6</sub>	6	53
HfNbTaTiVZrC <sub>6</sub>	6	75	CrHfMoTiVZrC <sub>6</sub>	6	53
CrNbTaTiVWC <sub>6</sub>	6	74	HfMoNbVWZrC <sub>6</sub>	6	53
MoNbTaTiVWC <sub>6</sub>	6	73	CrHfTaTiVWZrC <sub>7</sub>	7	53
CrHfMoNbTaWC <sub>6</sub>	6	73	CrMoTiVWZrC <sub>6</sub>	6	53
CrMoNbVWZrC <sub>6</sub>	6	68	CrHfNbTaTiWZrC <sub>6</sub>	7	53
CrHfMoNbVWC <sub>6</sub>	6	67	CrNbTaTiWZrC <sub>6</sub>	6	53
CrHfMoNbTaTiC <sub>6</sub>	6	66	CrNbTaTiVWZrC <sub>7</sub>	7	52
CrMoNbTaTiZrC <sub>6</sub>	6	66	HfNbTaTiVWC <sub>6</sub>	6	52
CrMoNbTaWZrC <sub>6</sub>	6	66	MoNbTaTiWZrC <sub>6</sub>	6	52
CrHfMoNbTaTiVC <sub>7</sub>	7	65	CrHfMoNbTaTiWZrC <sub>8</sub>	8	52
CrMoNbTaTiVZrC <sub>7</sub>	7	64	HfMoTaVWZrC <sub>6</sub>	6	52
CrHfMoNbTiVC <sub>6</sub>	6	64	CrHfMoTaVWZrC <sub>7</sub>	7	52
CrMoNbTiVZrC <sub>6</sub>	6	64	CrHfMoNbTaTiVWC <sub>8</sub>	8	52
CrHfMoNbTaVC <sub>6</sub>	6	63	CrHfMoNbTaWZrC <sub>7</sub>	7	52

Supplementary Table 7, Continued.

Material	# of Cations	EFA	Material	# of Cations	EFA
CrHfMoNbTaVZrC <sub>7</sub>	7	63	CrHfNbTaVWZrC <sub>7</sub>	7	52
CrMoNbTaVZrC <sub>6</sub>	6	63	CrHfMoNbTaVWZrC <sub>8</sub>	8	51
CrHfMoTaVZrC <sub>6</sub>	6	63	CrHfMoNbTiWC <sub>6</sub>	6	51
CrHfMoNbTaZrC <sub>6</sub>	6	62	CrHfMoTiVWC <sub>6</sub>	6	51
HfMoNbTiVZrC <sub>6</sub>	6	62	HfMoTaTiVWZrC <sub>7</sub>	7	51
HfMoNbTaVZrC <sub>6</sub>	6	62	CrHfTaTiVWC <sub>6</sub>	6	51
CrHfMoNbTaTiVZrC <sub>8</sub>	8	61	CrHfMoTaTiWZrC <sub>7</sub>	7	51
CrHfMoNbVZrC <sub>6</sub>	6	61	<b>CrHfMoNbTaTiVWZrC<sub>9</sub></b>	9	51
CrHfMoNbTiZrC <sub>6</sub>	6	61	CrMoNbTiWZrC <sub>6</sub>	6	50
CrHfMoNbTaTiZrC <sub>7</sub>	7	61	HfMoTaTiWZrC <sub>6</sub>	6	50
HfNbTaVWZrC <sub>6</sub>	6	60	HfMoNbTiWZrC <sub>6</sub>	6	50
CrHfNbTiVZrC <sub>6</sub>	6	60	CrHfNbTiVWZrC <sub>7</sub>	7	50
CrHfMoNbTaVWC <sub>7</sub>	7	60	CrMoNbTaTiWZrC <sub>7</sub>	7	50
CrHfNbTaTiVZrC <sub>7</sub>	7	60	CrHfNbTaVWC <sub>6</sub>	6	50
CrMoTaTiVZrC <sub>6</sub>	6	60	HfMoNbTiVWC <sub>6</sub>	6	50
HfMoNbTaTiVZrC <sub>7</sub>	7	60	CrHfNbTaWZrC <sub>6</sub>	6	50
CrHfNbTaVZrC <sub>6</sub>	6	60	MoNbTiVWZrC <sub>6</sub>	6	50
HfMoNbTaTiZrC <sub>6</sub>	6	59	HfMoNbTiVWZrC <sub>7</sub>	7	50
HfMoTaTiVZrC <sub>6</sub>	6	59	CrNbTaVWZrC <sub>6</sub>	6	50
CrHfNbTaTiVC <sub>6</sub>	6	59	CrHfMoTaTiVWC <sub>7</sub>	7	49
CrHfTaTiVZrC <sub>6</sub>	6	59	CrHfNbVWZrC <sub>6</sub>	6	49

Supplementary Table 7, Continued.

Material	# of Cations	EFA	Material	# of Cations	EFA
CrMoNbTaVWZrC <sub>7</sub>	7	59	CrMoNbTiVWZrC <sub>7</sub>	7	49
HfMoNbTaTiVWC <sub>7</sub>	7	58	CrHfMoNbTiVWC <sub>7</sub>	7	49
CrHfMoTaTiVC <sub>6</sub>	6	58	CrTaTiVWZrC <sub>6</sub>	6	49
HfMoNbTaWZrC <sub>6</sub>	6	58	CrHfTiVWZrC <sub>6</sub>	6	49
HfMoNbTaTiWZrC <sub>7</sub>	7	58	CrHfMoNbVWZrC <sub>7</sub>	7	49
HfMoNbTaTiVC <sub>6</sub>	6	58	HfMoTaTiVWC <sub>6</sub>	6	48
CrHfMoNbTiVZrC <sub>7</sub>	7	58	CrHfMoNbTiWZrC <sub>7</sub>	7	48
CrHfNbTaTiZrC <sub>6</sub>	6	58	CrHfTaVWZrC <sub>6</sub>	6	48
HfMoNbTaTiWC <sub>6</sub>	6	57	CrHfMoTaTiWC <sub>6</sub>	6	48
CrHfMoNbTaTiWC <sub>7</sub>	7	57	CrHfMoTiVWZrC <sub>7</sub>	7	48
CrNbTaTiVZrC <sub>6</sub>	6	57	CrNbTiVWZrC <sub>6</sub>	6	48
<b>HfMoNbTaVWZrC<sub>7</sub></b>	7	57	MoTaTiVWZrC <sub>6</sub>	6	48
CrHfMoTaVWC <sub>6</sub>	6	56	CrHfMoTaWZrC <sub>6</sub>	6	47
CrMoTaVWZrC <sub>6</sub>	6	56	CrHfMoVWZrC <sub>6</sub>	6	47
CrHfNbTaTiVWC <sub>7</sub>	7	56	CrHfNbTiVWC <sub>6</sub>	6	47
CrHfMoTaTiVZrC <sub>7</sub>	7	56	CrMoTaTiVWZrC <sub>7</sub>	7	47
MoNbTaTiVZrC <sub>6</sub>	6	56	<b>CrMoNbTaTiVWZrC<sub>8</sub></b>	8	47
CrHfNbTaTiWC <sub>6</sub>	6	55	CrHfMoTiWZrC <sub>6</sub>	6	47
<b>HfMoNbTaVWC<sub>6</sub></b>	6	55	CrHfNbTiWZrC <sub>6</sub>	6	47
<b>HfMoNbTaTiVWZrC<sub>8</sub></b>	8	55	HfMoTiVWZrC <sub>6</sub>	6	46
MoNbTaTiVWZrC <sub>7</sub>	7	55	CrHfMoTaTiVWZrC <sub>8</sub>	8	46

Supplementary Table 7, Continued.

<b>Material</b>	<b># of Cations</b>	<b>EFA</b>	<b>Material</b>	<b># of Cations</b>	<b>EFA</b>
HfNbTaTiWZrC <sub>6</sub>	6	55	CrMoTaTiWZrC <sub>6</sub>	6	46
HfNbTaTiVWZrC <sub>7</sub>	7	54	CrHfTaTiWZrC <sub>6</sub>	6	46
CrHfMoTaTiZrC <sub>6</sub>	6	54	CrHfMoNbTiVWZrC <sub>8</sub>	8	43

**Supplementary Table 8 ICSD structure information for the matched phases.** The phase labels from the EBSD maps are linked with the structure prototype, space group, symmetry, and the lattice parameters in angstroms. The last column denotes which oxidized sample the phase was found in the sample.

Phase Label	ICSD COD#	Space Group	Symmetry	a (Å)	b (Å)	c (Å)	Present in Sample
(NbTa) Oxide	237585	<i>Pmmm</i> (47)	Orthorhombic	3.94	6.15	3.66	HfMoNbTaZrC <sub>5</sub>
(TaNb) Oxide	60627	<i>I4<sub>1</sub>/aZ</i> (88)	Tetragonal	13.32	13.32	6.12	HfMoNbTaZrC <sub>5</sub>
(HfZr) Oxide	67004	<i>Pbcm</i> (57)	Orthorhombic	5.04	5.25	5.09	HfMoNbTaZrC <sub>5</sub>
MoC <sub>1-x</sub>	1326	<i>Pbcn</i> (60)	Orthorhombic	4.72	6.00	5.20	HfMoNbTaZrC <sub>5</sub>
HfMoNbTaZrC <sub>5</sub>	N/A	<i>Fm<math>\bar{3}m</math></i> (225)	FCC	4.52	4.52	4.52	HfMoNbTaZrC <sub>5</sub>
(NbTa) Oxide	96	<i>I4<sub>1</sub>/aZ</i> (88)	Tetragonal	13.32	13.32	6.12	HfMoNbTaVWZrC <sub>7</sub>
(HfZr) Oxide	41010	<i>P12<sub>1</sub>/c1</i> (14)	Monoclinic	5.12	5.19	5.28	HfMoNbTaVWZrC <sub>7</sub>
(MoW) Carbide	156478	<i>P63/mmc</i> (194)	Hexagonal	3.01	3.01	14.61	HfMoNbTaVWZrC <sub>7</sub>
W <sub>2</sub> C	77567	<i>P63/mmc</i> (194)	Hexagonal	2.99	2.99	4.72	HfMoNbTaVWZrC <sub>7</sub>
Depleted UHEC	N/A	<i>Fm<math>\bar{3}m</math></i> (225)	FCC	4.44	4.44	4.44	HfMoNbTaVWZrC <sub>7</sub>
HfMoNbTaVWZrC <sub>7</sub>	N/A	<i>Fm<math>\bar{3}m</math></i> (225)	FCC	4.44	4.44	4.44	HfMoNbTaVWZrC <sub>7</sub>

## Appendix D Supplementary Information for Chapter 5

Crystal symmetry determination in electron diffraction using machine learning

Kevin Kaufmann<sup>1</sup>, Chaoyi Zhu<sup>2†</sup>, Alexander S. Rosengarten<sup>1†</sup>, Daniel Maryanovsky<sup>3</sup>, Tyler J. Harrington<sup>2</sup>, Eduardo Marin<sup>1</sup>, and Kenneth S. Vecchio<sup>1,2\*</sup>

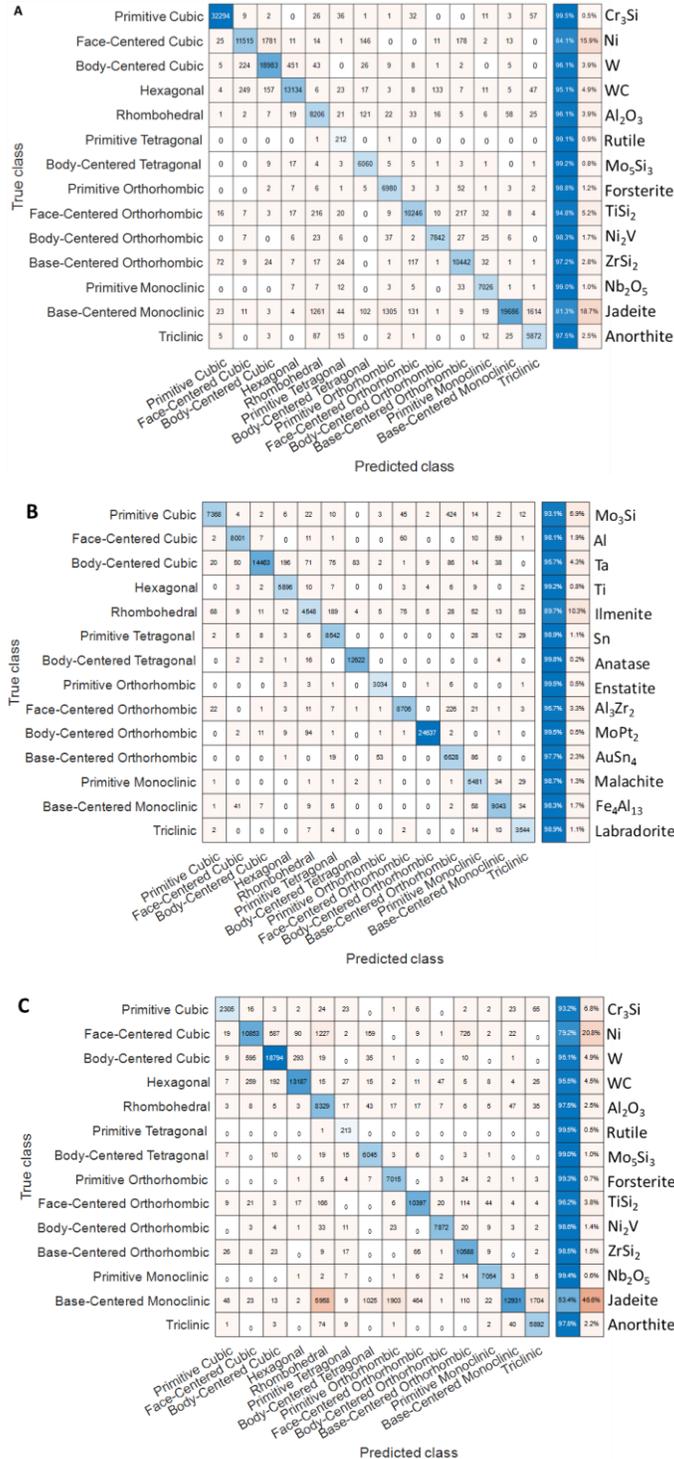
<sup>1</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA.

<sup>2</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA.

<sup>3</sup>Department of Cognitive Science, UC San Diego, La Jolla, CA 92093, USA.

\*Correspondence to: kvecchio@eng.ucsd.edu.

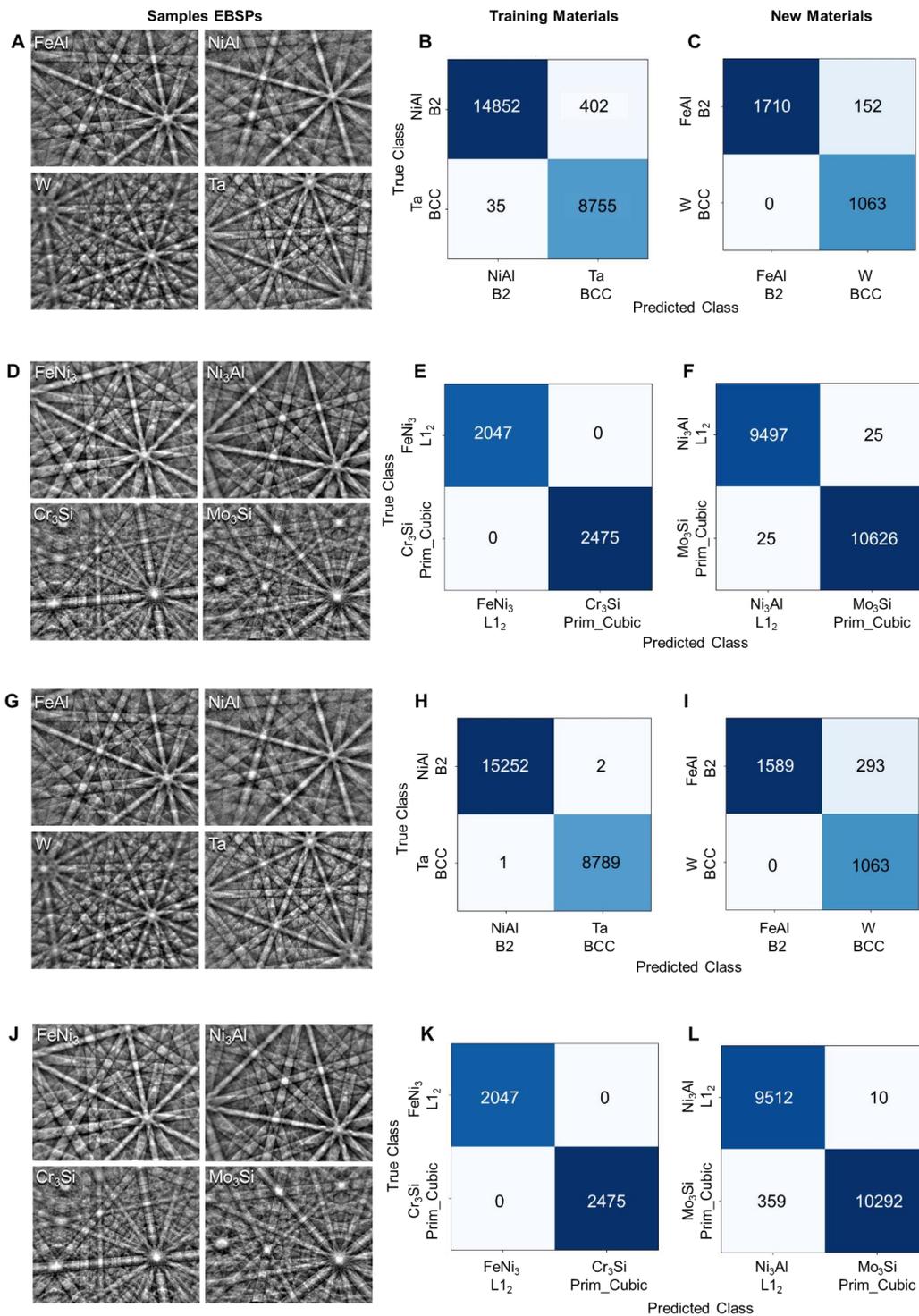
† These authors contributed equally to this work.



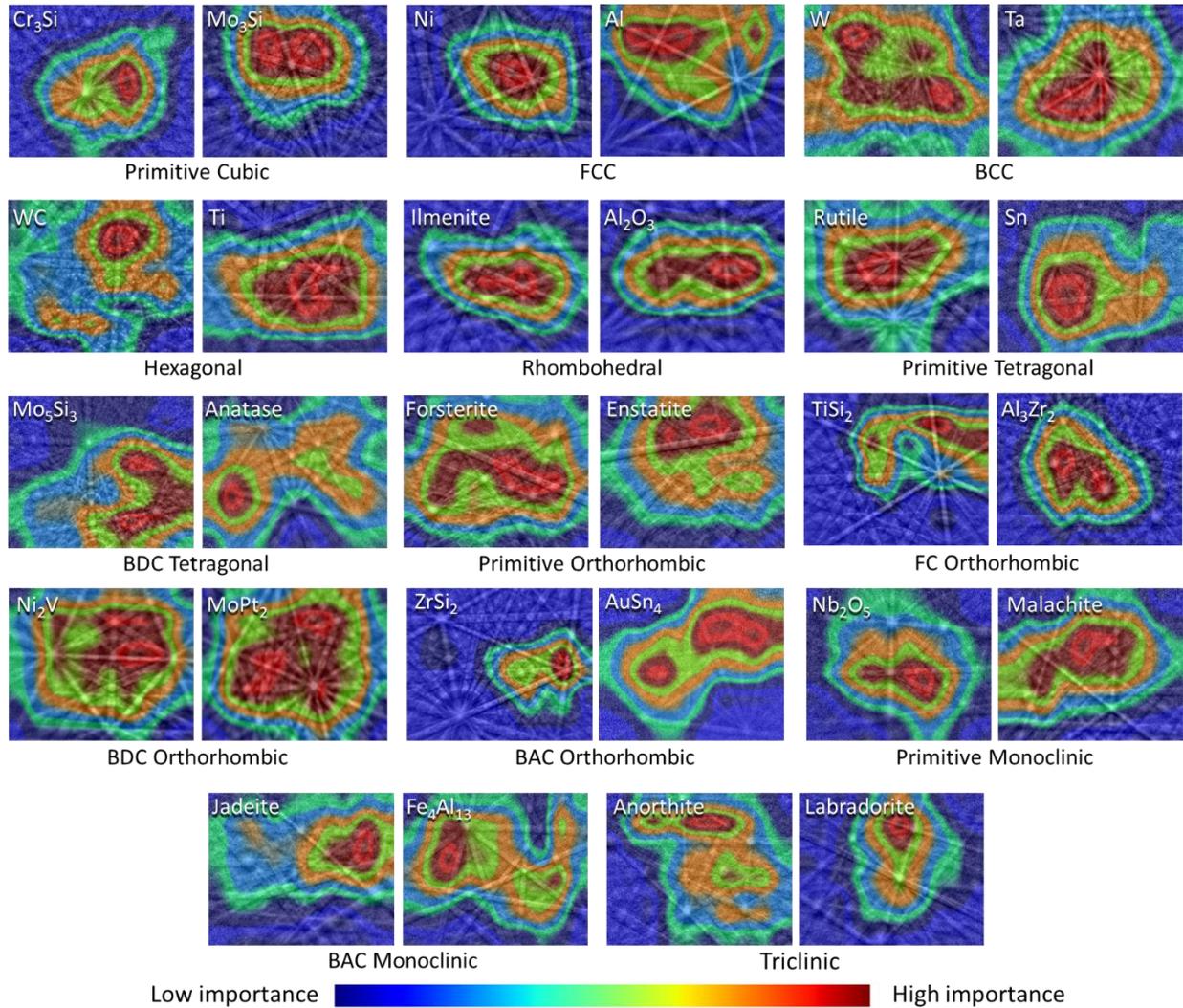
**Supplementary Figure 20 Confusion matrices displaying the classification results for the 14 Bravais lattices.** A new set of diffraction patterns were classified by the ResNet50 (A) and Xception (B, C) architecture. The two trained models were tested using newly collected diffraction patterns from the twenty-eight materials used to train the machine learning models. The diagonal (blue shaded boxes) in these tables represent the successful matching of the CNN predictions to the true Bravais lattices of the sample.

True class	Primitive Cubic	7881	106	49	5	27	34	43	18	5	6	33	7	33	4	95.5% V <sub>8</sub> C <sub>7</sub>
	Primitive Cubic	4906	1	4	0	2	0	0	2	133	7	1	0	4	0	97.0% Si
	Face-Centered Cubic	0	1663	4	0	0	0	0	0	0	0	0	0	0	0	99.8% NbC
	Face-Centered Cubic	1	2414	20	0	4	1	2	0	0	0	1	4	0	2	98.6% TaC
	Body-Centered Cubic	1	240	10956	0	27	230	0	0	0	0	5	2	46	2	95.2% Beta Ti
	Body-Centered Cubic	1	30	6158	0	0	0	0	0	5	26	206	0	0	0	95.8% Fe
	Hexagonal	0	0	0	2988	2	20	0	8	0	0	0	0	0	3	98.9% Quartz
	Primitive Monoclinic	1306	2	167	35	26	2	8	0	580	50	284	8783	9	0	78.1% Mo <sub>2</sub> C
	Base-Centered Monoclinic	18	0	0	0	6	2	2	613	12	14	22	1	0	0	0% Diopside
		Primitive Cubic	Face-Centered Cubic	Body-Centered Cubic	Hexagonal	Rhombohedral	Primitive Tetragonal	Body-Centered Tetragonal	Primitive Orthorhombic	Face-Centered Orthorhombic	Body-Centered Orthorhombic	Base-Centered Orthorhombic	Primitive Monoclinic	Base-Centered Monoclinic	Triclinic	
		Predicted class														

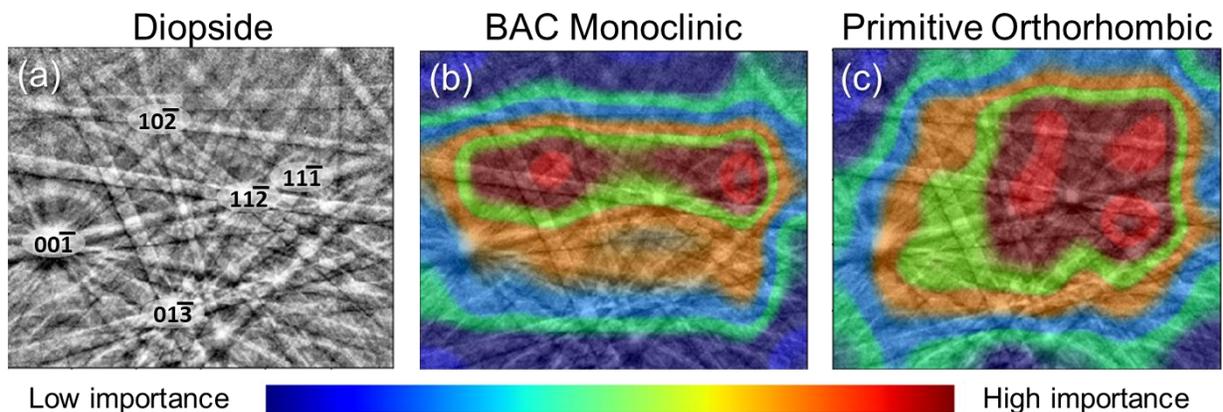
**Supplementary Figure 21 Performance of the Xception machine learning model on nine different materials.** The convolutional neural network architecture classifies electron backscatter diffraction patterns collected from materials not used to train the model. Correct classification is identified by the green squares instead of along the diagonal.



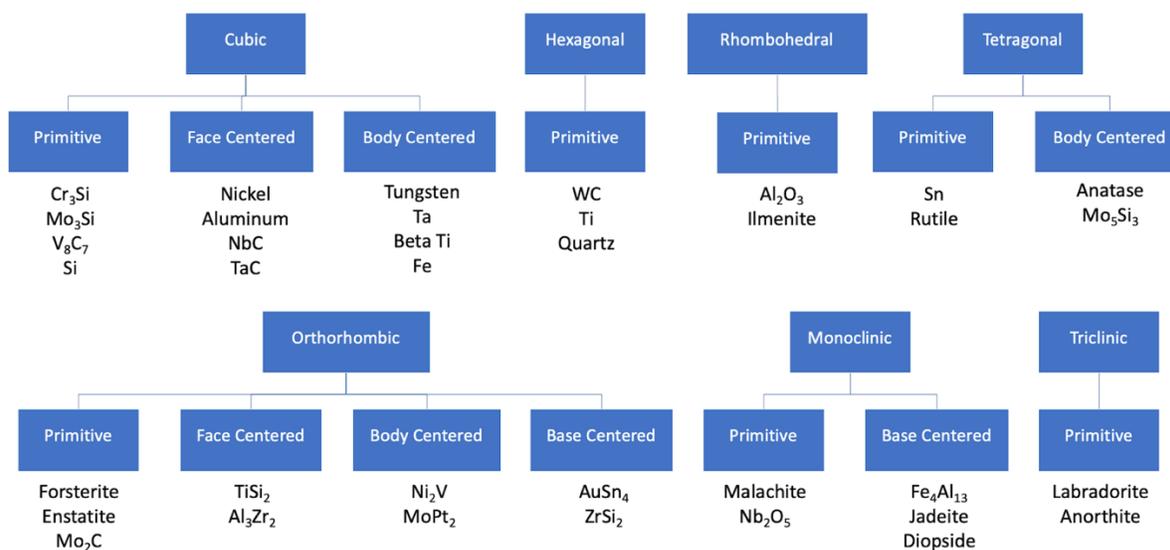
**Supplementary Figure 22 Demonstration of machine learning aided EBSD's capability to autonomously identify space groups in new materials.** (A, D, G, J) An example electron backscatter diffraction pattern for each of the materials. Zone axes with the same symmetry can be seen in each diffraction pattern. (B, E, H, K) Results of training the algorithm to discriminate between space group 221, space group 223, and space group 229 crystal structures. (C, F, I, L) Resultant confusion matrix after testing the algorithm blindly on two different materials from the same space groups.



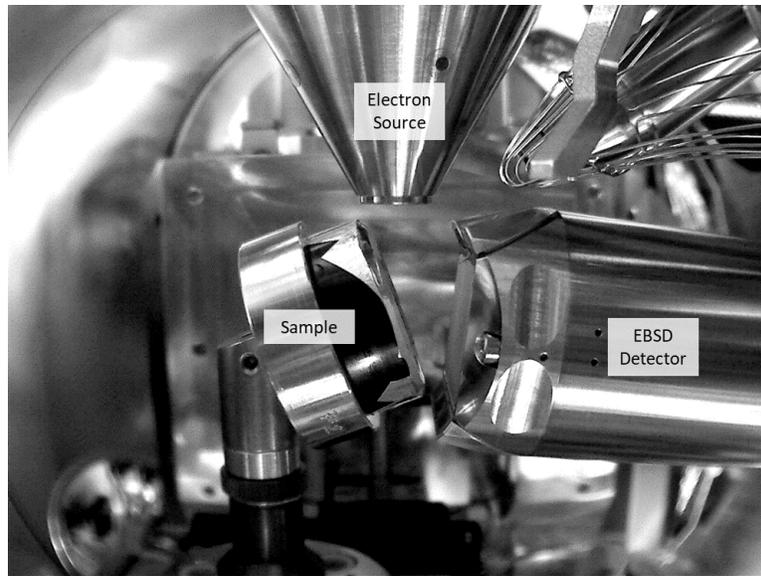
**Supplementary Figure 23 Heatmaps elucidating the normalized importance of features (dark blue to dark red) in each diffraction pattern for classification of the correct Bravais lattice.** One example is presented for each material utilized in the training set. It is observed that the symmetry features, areas nearest the zone axes, present in the image produce the highest activations for determining their origin Bravais lattice.



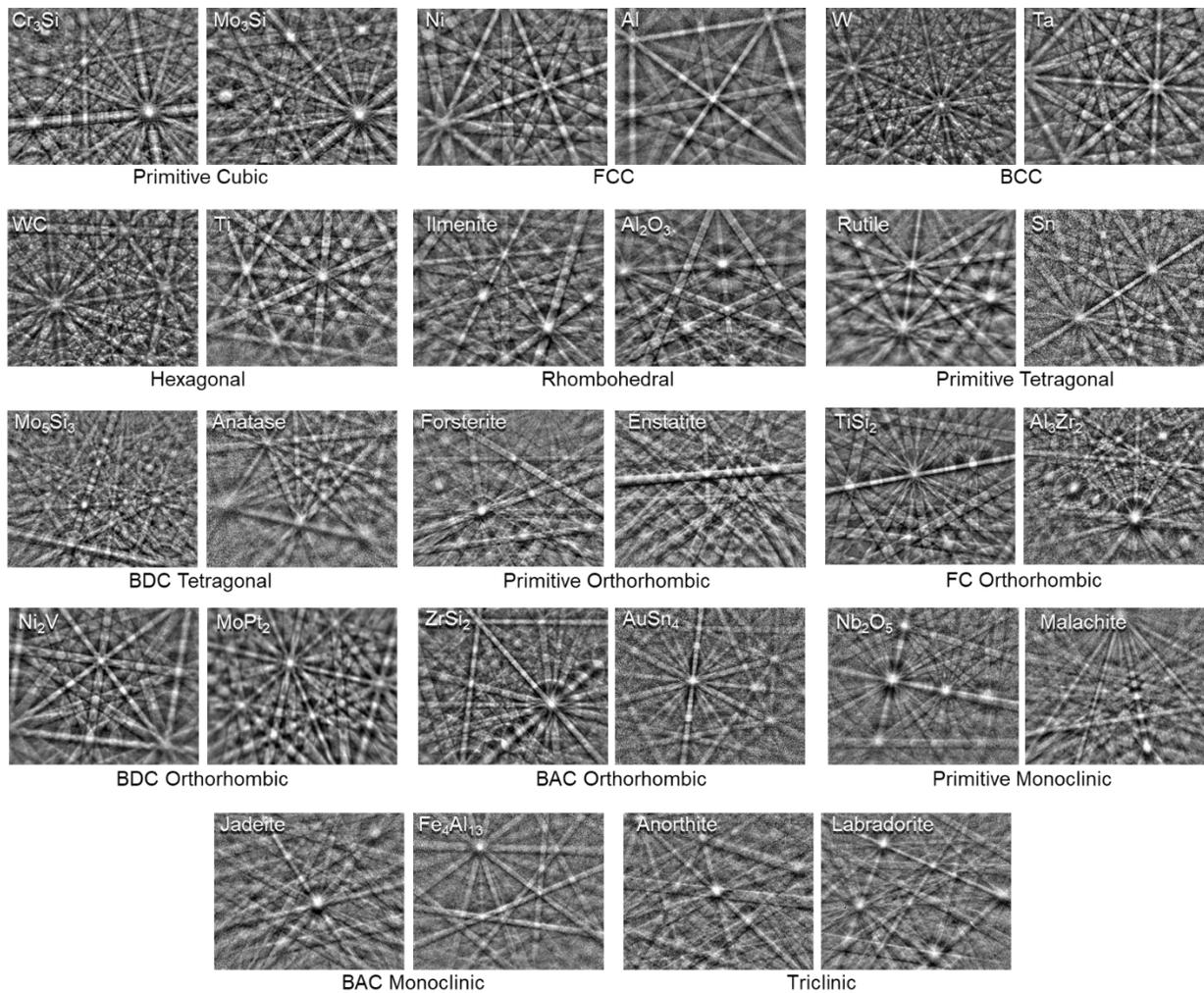
**Supplementary Figure 24 Heatmaps demonstrating the classification of diopside.** (a) An electron backscatter diffraction pattern from the base centered monoclinic phase with 2-fold axes marked. (b) The heatmap for the activations of the diffraction pattern for the base centered (BAC) monoclinic class. (c) The heatmap for the same diffraction pattern with respect to the primitive orthorhombic class. The symmetry features near the 2-fold axis produce the largest activations for each of these two classes.



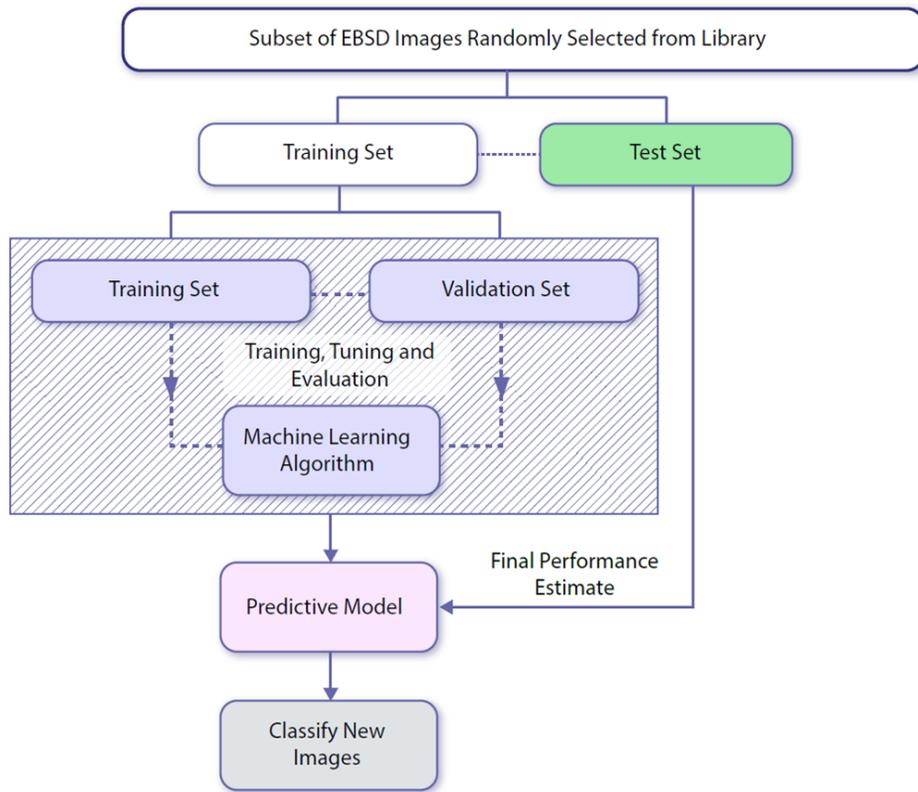
**Supplementary Figure 25 Diagram of the materials and their Bravais lattice.** For each of the fourteen Bravais lattices, diffraction patterns from at least two materials were collected. Diffraction patterns from supplementary materials in each Bravais lattice were utilized to test the performance of the model without it having any prior knowledge of the material.



**Supplementary Figure 26 Annotated image of the experimental setup within the SEM chamber.** The sample is at 70 degrees from horizontal and is facing the EBSD detector. Some of the electrons which enter the sample may backscatter and exit the sample at the Bragg condition of the periodic atomic lattice planes. These electrons diffract to form Kikuchi bands corresponding to each of the lattice diffracting crystal planes. Most commercial systems then use look-up tables of interplanar angles for user selected phases for orientation indexing.



**Supplementary Figure 27** An example diffraction pattern for each of the 28 materials used for training the machine learning model. This represents just one of the nearly infinite number of orientations, and therefore diffraction patterns, the crystal can occupy in three-dimensional space.



**Supplementary Figure 28 Schematic describing the process of training a machine learning algorithm to recognize the symmetry features present in diffraction patterns.** In the presented approach, a small subset of the EBSD patterns were randomly selected from two materials per class and then randomly subdivided into a training and testing set. The training set is then randomly divided again into a training set from which the machine learning model extracts features, and a validation set is used to test the usefulness of the features in correctly classifying the diffraction patterns. The feature bank is constantly tuned and evaluated such that the learned features produce the highest accuracy in classifying the validation set. When the machine learning algorithm determines it has found the best features for accurate classification, the learned filters are saved, and the predictive model can be applied to the classification of new diffraction patterns.

## 7 Crystal Systems

## 14 Bravais Lattices

(a)

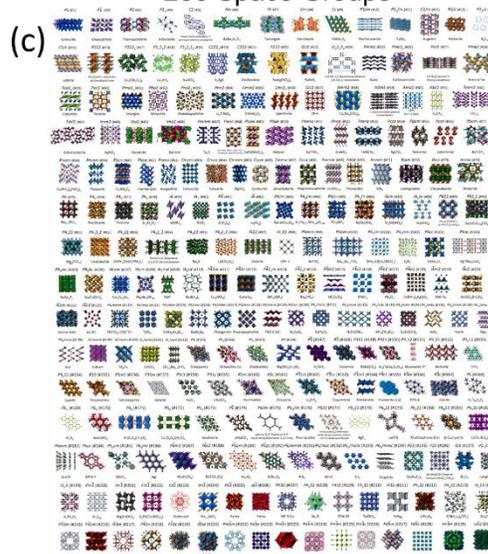
Lattice System	Lengths	Angles	Primitive (P)	Base-centered (C)	Body-centered (I)	Face-centered (F)
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$				
Hexagonal	$a = b \neq c$	$\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$				
Rhombohedral	$a = b = c$	$\alpha = \beta = \gamma \neq 90^\circ$				
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$				
Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$				
Monoclinic	$a \neq b \neq c$	$\alpha = \beta = 90^\circ \neq \gamma$				
Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$				

## 32 Point Groups

(b)

Class	Group names					
Cubic	23	$m\bar{3}$		432	$\bar{4}3m$	$m\bar{3}m$
Hexagonal	6	$\bar{6}$	$6/m$	622	$6mm$	$\bar{6}m2$ $6/mmm$
Trigonal	3	$\bar{3}$		32	$3m$	$\bar{3}m$
Tetragonal	4	$\bar{4}$	$4/m$	422	$4mm$	$\bar{4}2m$ $4/mmm$
Orthorhombic				222	$mm2$	$mmm$
Monoclinic	2		$2/m$		$m$	
Triclinic	1	$\bar{1}$				

## 230 Space Groups



**Supplementary Figure 29 Crystal structure relationships from the seven crystal systems to the 230 space groups.** (a) Illustration of the fourteen Bravais lattices and their associated symmetries. (b) A table of the 32 point groups and the crystal family they belong to. (c) An example of each of the 230 space groups. Space groups represent the minute details of atomic arrangement within a point group. (c) has been reproduced with permission from creator Frank Hoffmann under a creative commons (CC BY-NC-SA) license (64) and the crystal structures drawn with VESTA.

**Supplementary Table 9** The point groups and their respective lattices are listed with the symmetry elements that are present for the crystal structure.

Point Group	Lattice	2-fold	3-fold	4-fold	6-fold	Planes	Center
1	Triclinic						
$\bar{1}$	Triclinic						yes
2	Monoclinic	1					
<i>m</i>	Monoclinic					1	
2/ <i>m</i>	Monoclinic	1				1	yes
222	Orthorhombic	3					
<i>mm</i> 2	Orthorhombic	1				2	
<i>mmm</i>	Orthorhombic	3				3	yes
4	Tertragonal			1			
$\bar{4}$	Tertragonal	1					
4/ <i>m</i>	Tertragonal			1		1	yes
422	Tertragonal	4		1			
4 <i>mm</i>	Tertragonal					4	
$\bar{4}2m$	Tertragonal	3				2	
4/ <i>mmm</i>	Tertragonal	4		1		5	yes
23	Cubic	3	4				
<i>m</i> 3	Cubic	3	4			3	yes
432	Cubic	6	4	3			
$\bar{4}3m$	Cubic	3	4			6	
<i>m</i> 3 <i>m</i>	Cubic	6	4	3		9	yes
3	Trigonal		1				
$\bar{3}$	Trigonal		1				yes
32	Trigonal	3	1				
3 <i>m</i>	Trigonal		1			3	
$\bar{3}m$	Trigonal	3	1			3	yes
6	Hexagonal				1		
$\bar{6}$	Hexagonal		1			1	
6/ <i>m</i>	Hexagonal				1	1	yes
622	Hexagonal	6			1		
6 <i>mm</i>	Hexagonal				1	6	
$\bar{6}m2$	Hexagonal	3	1			4	
6/ <i>mmm</i>	Hexagonal	6			1	7	yes

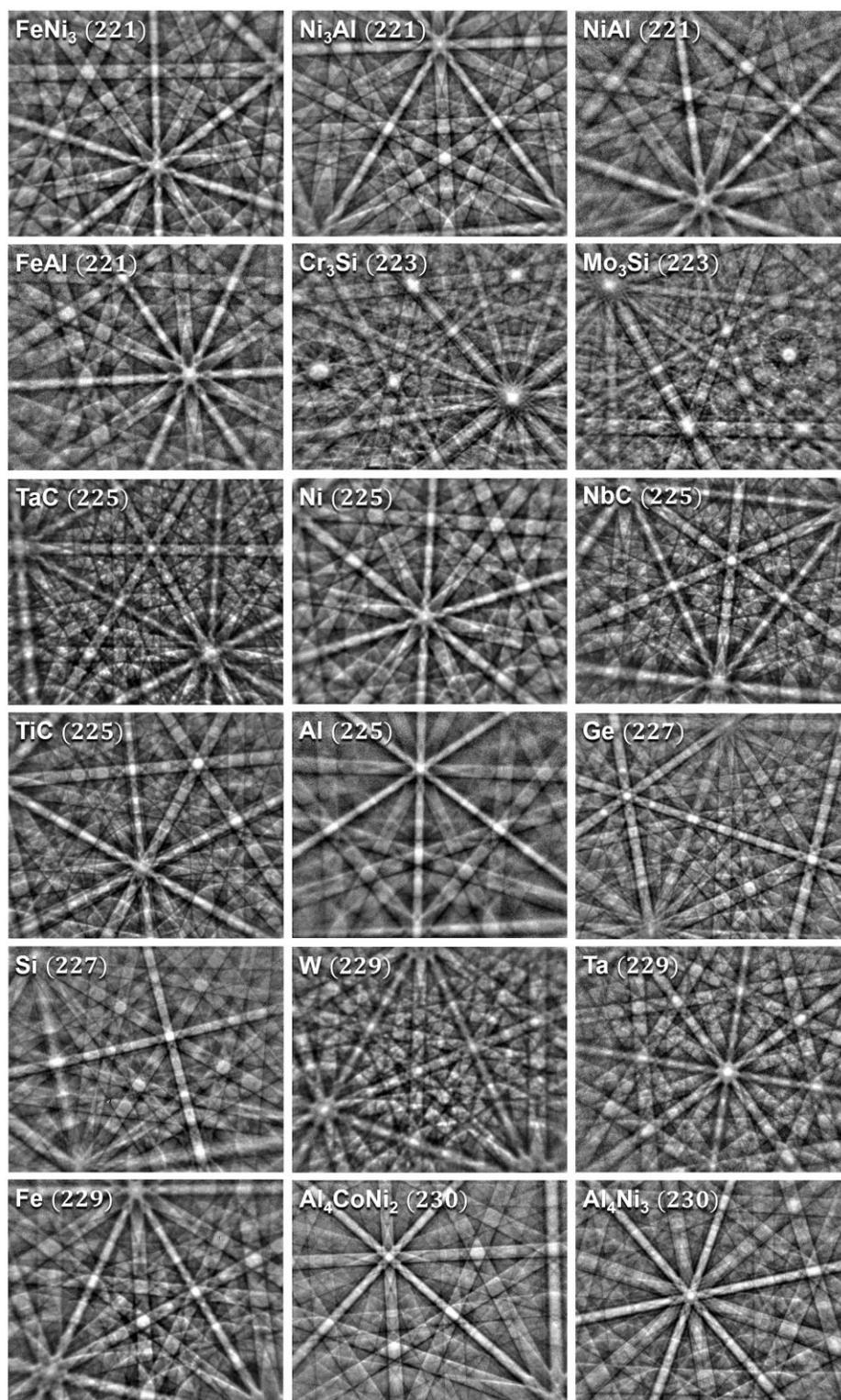
## **Appendix E Supplementary Information for Chapter 6**

Deep neural network enabled space group identification in EBSD

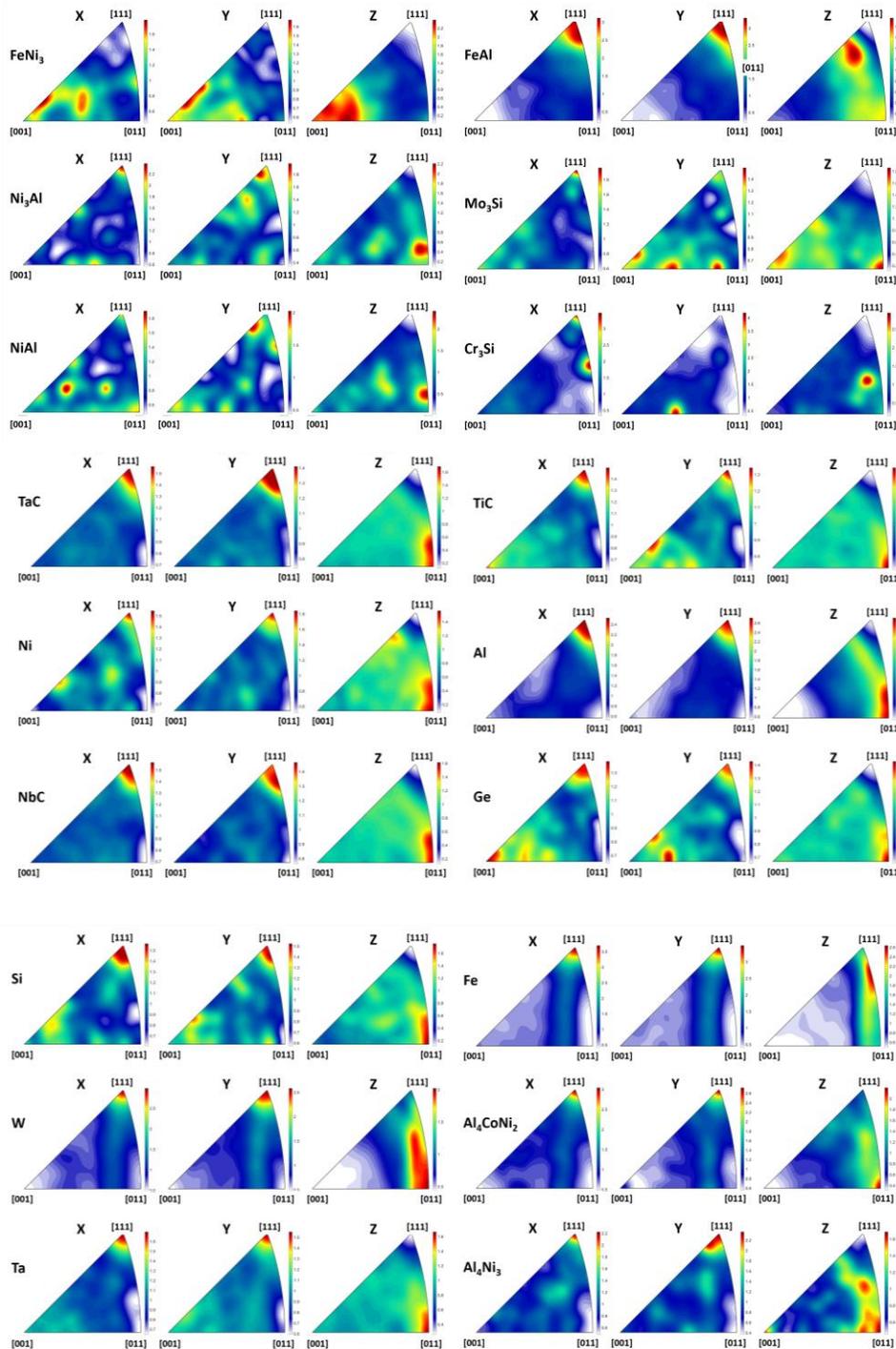
Kevin Kaufmann<sup>1</sup>, Chaoyi Zhu<sup>2</sup>, Alexander S. Rosengarten<sup>1</sup>, and Kenneth S. Vecchio<sup>1,2</sup>

<sup>1</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA

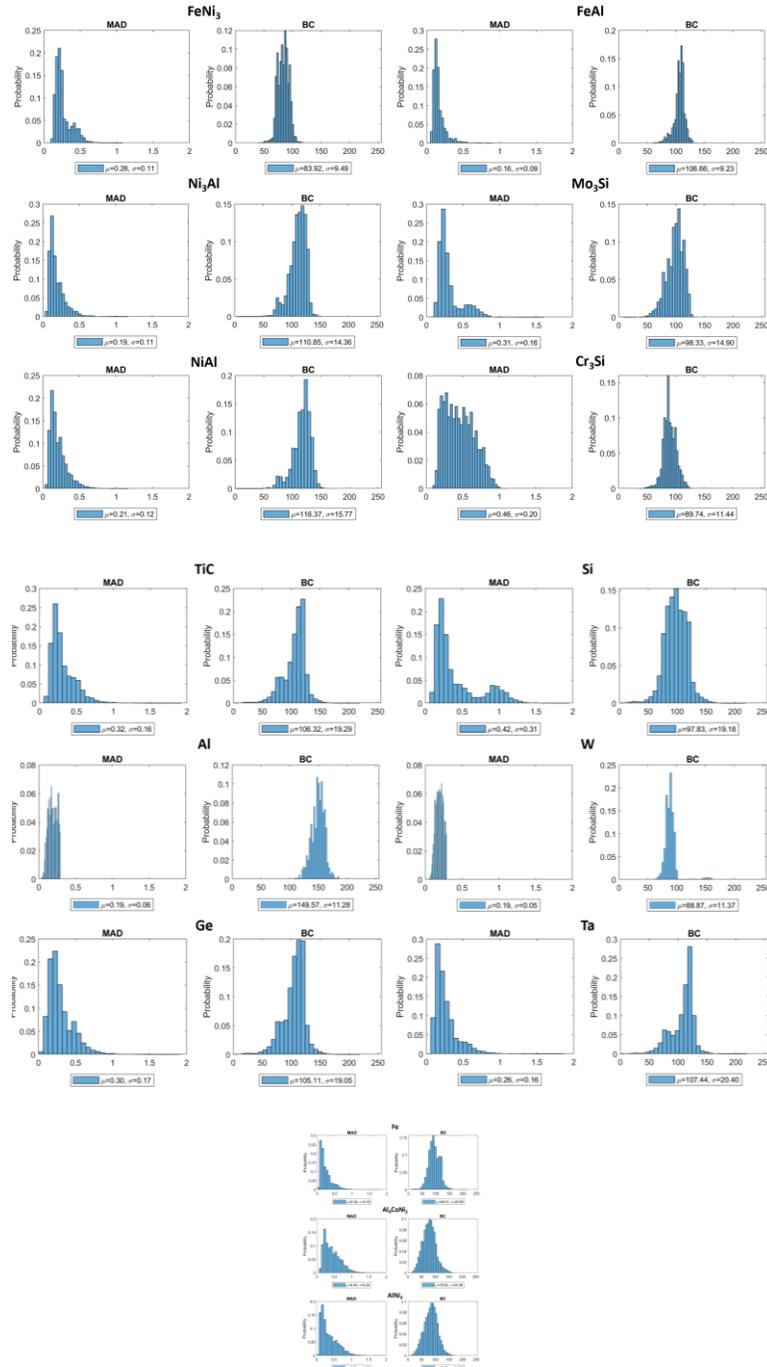
<sup>2</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA



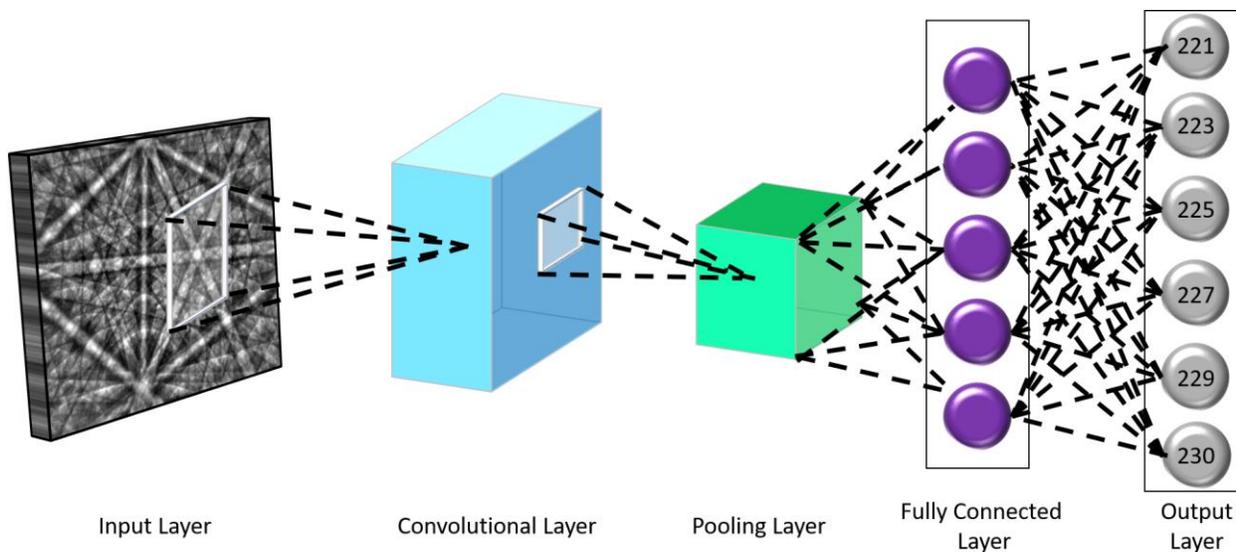
**Supplementary Figure 30 Representative diffraction patterns from each material studied. The space group and material are denoted in the upper left corner of each diffraction pattern.**



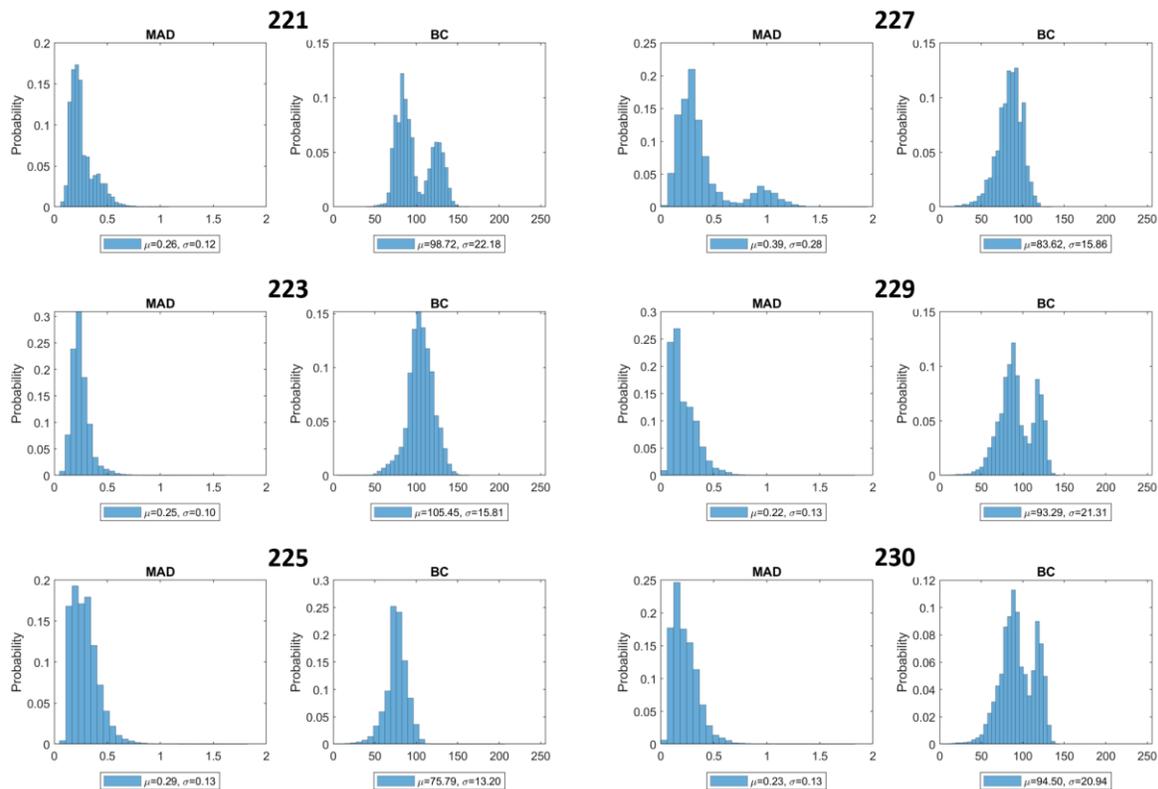
**Supplementary Figure 31 Inverse pole figures of the entire dataset.** Orientation analysis shows the materials are of very low texture, typically in the range of 2-3 multiples of uniform density (M.U.D.). Note that the scale bars are all below 5 M.U.D. The data is first plotted with the scale bars automatically determined by MTEX to show the data distribution. The second set of plots uses the fixed scale of 0 to 5 M.U.D. to demonstrate that the data does not approach medium texture levels.



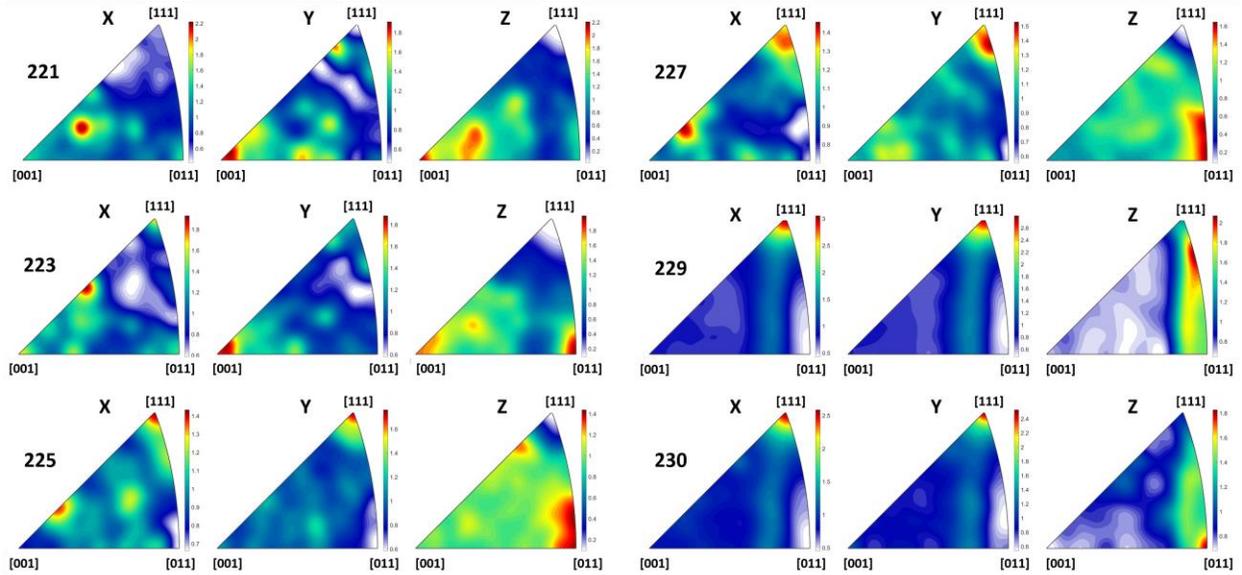
**Supplementary Figure 32 Probability plots of mean angular deviation and band contrast for the entire dataset.** The pattern quality distribution of each material is assessed using mean angular deviation (MAD) and band contrast (BC) as descriptors. Each plot is also annotated with the mean ( $\mu$ ) and standard deviation ( $\sigma$ ).



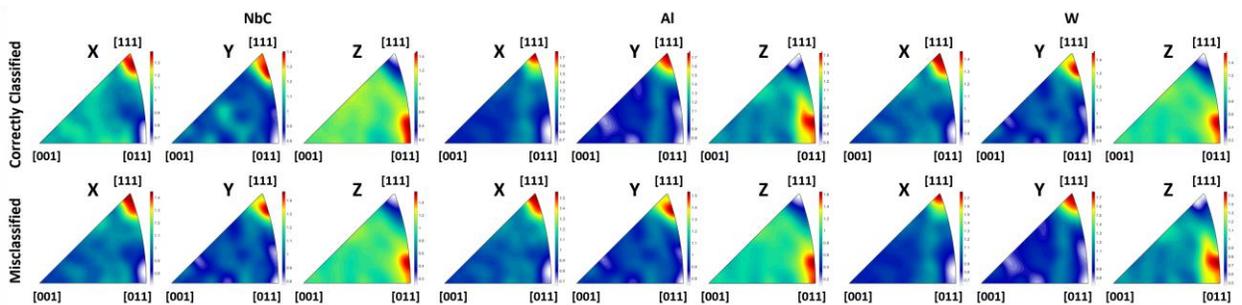
**Supplementary Figure 33 Schematic of the neural network.** In convolutional layers, a learnable filter is convolved across the image and the scalar product between the filter and the input at every position is computed to form a feature map. Pooling layers are placed after convolutional layers to down sample the feature maps and produce coarse grain representations and spatial information about the features in the data. A traditional dense neural network is placed as the last layer, where the probability that the input diffraction pattern belongs to each space group is computed.



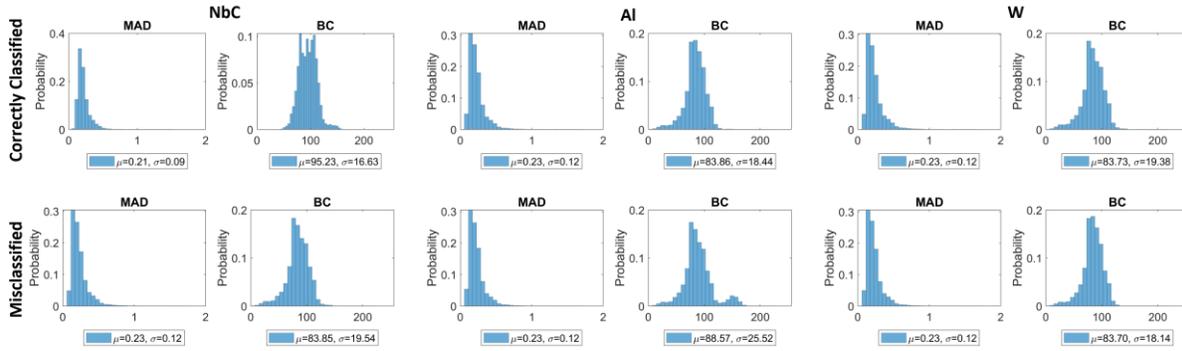
**Supplementary Figure 34 Histograms of mean angular deviation and band contrast in the training set for the associated model.** The pattern quality distribution of each material is assessed using mean angular deviation (MAD) and band contrast (BC) as descriptors. Each plot is also annotated with the mean ( $\mu$ ) and standard deviation ( $\sigma$ ).



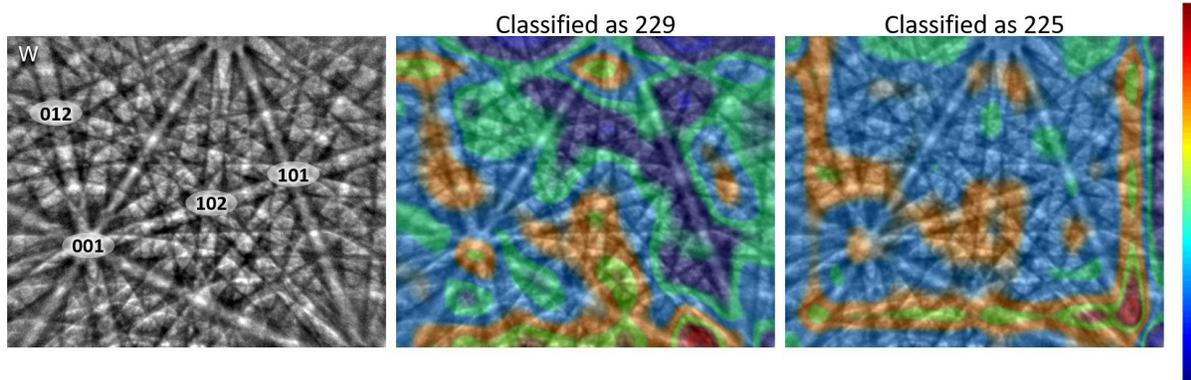
**Supplementary Figure 35 Inverse pole figures for each space group training set in the associated model.** The range of possible orientations are well represented for each class. Note that the scale bars are all below 5 M.U.D.



**Supplementary Figure 36 Inverse pole figures for patterns based on correct or incorrect classification.** The distribution of orientations that were correctly classified and misclassified are very similar, suggesting texture is not having a profound effect. Note that the scale bars are all below 5 M.U.D.



**Supplementary Figure 37 Histograms of mean angular deviation and band contrast separated by correct or incorrect classification.** The pattern quality distribution of each material is assessed using mean angular deviation (MAD) and band contrast (BC) as descriptors. Each plot is also annotated with the mean ( $\mu$ ) and standard deviation ( $\sigma$ ).



**Supplementary Figure 38 Feature comparison for correct and incorrect classifications.** The activations for the 229 class are studied when the pattern is correctly identified (middle) and misclassified to 225 (right). Similar information is identified; however, the zone axis activations are weaker for the misclassified pattern.

**Supplementary Table 10 Material acquisition and processing.** The method of fabrication is listed for each material studied. SPS denotes spark plasma sintering from a commercial powder. The homogenization heat treatments were performed for one week in an inert atmosphere.

Space Group	Material	Formula Weighted	Method of Fabrication			
		Atomic Number				
<b>221</b>			<b>Wrought</b>	<b>Arc Melt</b>	<b>SPS</b>	<b>Heat Treatment</b>
	FeNi3	27.5		X		X
	Ni3Al	24.25		X		X
	NiAl	20.5		X		X
	FeAl	19.5		X		X
<b>223</b>						
	Mo3Si	35		X		X
	Cr3Si	21.5		X		X
<b>225</b>						
	TaC	39.5			X	
	Ni	28	X			
	NbC	23.5			X	
	TiC	14			X	
	Al	13	X			
<b>227</b>						
	Ge	32		X		
	Si	14		X		
<b>229</b>						
	W	74			X	
	Ta	73			X	
	Fe	26		X		
<b>230</b>						
	Al4CoNi2	19.4		X		X
	Al4Ni3	19.4		X		X

**Supplementary Table 11 Number of diffraction patterns classified to each space group.** Space group 221; trained on FeNi<sub>3</sub>. Space group 223; trained on Mo<sub>3</sub>Si. Space group 225; trained on TaC. Space group 227; trained on Ge. Space group 229; trained on Ta. Space group 230; trained on Al<sub>4</sub>CoNi<sub>2</sub>.

Space Group	Material	Formula Weighted Atomic Number	Number of Patterns Classified to Each Space Group						Accuracy	Precision	Recall
			221	223	225	227	229	230			
<b>221</b>			<b>221</b>	223	225	227	229	230			
	FeNi <sub>3</sub>	27.5	<b>9,987</b>	0	0	0	31	1	99.7%	0.24	1.00
	Ni <sub>3</sub> Al	24.25	<b>9,085</b>	0	6	0	42	400	95.3%	0.22	0.95
	NiAl	20.5	<b>79</b>	1	71	1	52	4,336	1.7%	0	0.02
	FeAl	19.5	<b>248</b>	0	0	1	1,538	76	13.3%	0.01	0.13
<b>223</b>			221	<b>223</b>	225	227	229	230			
	Cr <sub>3</sub> Si	21.5	11	<b>12,219</b>	0	0	357	1	97.1%	0.61	0.97
	Mo <sub>3</sub> Si	35	27	<b>10,498</b>	11	0	118	6	98.5%	0.57	0.99
<b>225</b>			221	223	<b>225</b>	227	229	230			
	TaC	39.5	527	97	<b>14,143</b>	10	462	44	92.5%	0.97	0.93
	Ni	28	11,170	8	<b>29</b>	1	424	82	0.2%	0.06	0.02
	NbC	23.5	776	698	<b>832</b>	15	16,222	57	4.5%	0.66	0.05
	TiC	14	12,381	398	<b>0</b>	0	2,778	35	0.0%	0	0
	Al	13	0	0	<b>0</b>	0	3	1,085	0.0%	0	0
<b>227</b>			221	223	225	<b>227</b>	229	230			
	Ge	32	1	14	0	<b>2,212</b>	12	1	98.8%	0.98	0.99
	Si	14	1,330	6,455	0	<b>1</b>	315	637	0.0%	0.03	0
<b>229</b>			221	223	225	227	<b>229</b>	230			
	W	74	150	140	281	0	<b>492</b>	0	46.3%	0.02	0.46
	Ta	73	1	1	17	1	<b>8,771</b>	4	99.7%	0.28	1
	Fe	26	5,899	64	25	3	<b>1,835</b>	1,690	19.3%	0.08	0.19
<b>230</b>			221	223	225	227	229	<b>230</b>			
	Al <sub>4</sub> CoNi <sub>2</sub>	19.4	1	1	2	1	2	<b>1,822</b>	99.6%	0.18	1
	Al <sub>4</sub> Ni <sub>3</sub>	19.4	30	4	16	5	31	<b>1,706</b>	95.2%	0.17	0.95
Total Patterns	144,465		221	223	225	227	229	230	Average		
# Correct	73,959	Precision	0.38	0.74	0.97	0.98	0.33	0.29	0.73		
Accuracy (%)	51.20	Recall	0.75	0.98	0.24	0.2	0.57	0.97	0.51		

**Supplementary Table 12 Number of diffraction patterns classified to each space group.** Space group 221; trained on FeNi<sub>3</sub>. Space group 223; trained on Mo<sub>3</sub>Si. Space group 225; trained on Al. Space group 227; trained on Ge. Space group 229; trained on Ta. Space group 230; trained on Al<sub>4</sub>CoNi<sub>2</sub>.

Space Group	Material	Formula Weighted Atomic Number	Number of Patterns Classified to Each Space Group						Accuracy	Precision	Recall
			221	223	225	227	229	230			
<b>221</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	FeNi <sub>3</sub>	27.5	<b>10,008</b>	3	2	0	6	0	99.9%	0.33	1.00
	Ni <sub>3</sub> Al	24.25	<b>8,733</b>	9	613	45	34	99	91.6%	0.3	0.92
	NiAl	20.5	<b>2</b>	0	4,345	2	30	161	0.0%	0	0
	FeAl	19.5	<b>67</b>	34	334	4	1,389	35	3.6%	0	0.04
<b>223</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Mo <sub>3</sub> Si	35	0	<b>10,614</b>	15	0	22	9	99.6%	0.28	1
	Cr <sub>3</sub> Si	21.5	0	<b>12,328</b>	1	0	259	0	97.9%	0.32	0.98
<b>225</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	TaC	39.5	593	7,430	<b>115</b>	31	6,802	312	0.8%	0.01	0.01
	Ni	28	11,377	5	<b>196</b>	0	93	43	1.7%	0.02	0.02
	NbC	23.5	497	4,293	<b>179</b>	76	12,877	678	1.0%	0.02	0.01
	TiC	14	3,356	8,532	<b>3,537</b>	10	157	0	22.7%	0.25	0.23
	Al	13	0	0	<b>1,086</b>	1	0	1	99.8%	0.09	1
<b>227</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Ge	32	0	0	3	<b>2,236</b>	0	1	99.8%	0.86	1
	Si	14	402	4,829	2,566	<b>896</b>	15	30	10.3%	0.71	0.1
<b>229</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	W	74	505	369	0	0	<b>189</b>	0	17.8%	0.01	0.18
	Ta	73	1	6	19	1	<b>8,766</b>	2	99.7%	0.29	1
	Fe	26	3,591	1,130	2,308	38	<b>2,236</b>	213	23.5%	0.09	0.23
<b>230</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Al <sub>4</sub> CoNi <sub>2</sub>	19.4	0	0	0	1	1	<b>1,827</b>	99.9%	0.54	1
	Al <sub>4</sub> Ni <sub>3</sub>	19.4	6	0	229	164	14	<b>1,379</b>	77.0%	0.47	0.77
Total Patterns	144,465		<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>	Average		
# Correct	64,394	Precision	0.48	0.46	0.33	0.89	0.34	0.67	0.43		
Accuracy (%)	44.57	Recall	0.72	0.99	0.08	0.29	0.58	0.89	0.44		

**Supplementary Table 13 Number of diffraction patterns classified to each space group.** Space group 221; trained on FeNi<sub>3</sub>. Space group 223; trained on Mo<sub>3</sub>Si. Space group 225; trained on TaC. Space group 227; trained on Si. Space group 229; trained on Ta. Space group 230; trained on Al<sub>4</sub>CoNi<sub>2</sub>.

Space Group	Material	Formula Weighted Atomic Number	Number of Patterns Classified to Each Space Group						Accuracy	Precision	Recall
			221	223	225	227	229	230			
<b>221</b>			<b>221</b>	223	225	227	229	230			
	FeNi <sub>3</sub>	27.5	<b>10,000</b>	4	1	10	4	0	99.8%	0.22	1.00
	Ni <sub>3</sub> Al	24.25	<b>9,357</b>	25	4	41	30	76	98.2%	0.21	0.98
	NiAl	20.5	<b>990</b>	0	66	65	174	3,245	21.8%	0.03	0.22
	FeAl	19.5	<b>909</b>	55	0	15	776	108	48.8%	0.02	0.49
<b>223</b>			221	<b>223</b>	225	227	229	230			
	Mo <sub>3</sub> Si	35	0	<b>10,629</b>	7	13	9	2	99.7%	0.75	1
	Cr <sub>3</sub> Si	21.5	0	<b>12,454</b>	7	121	6	0	98.9%	0.78	0.99
<b>225</b>			221	223	<b>225</b>	227	229	230			
	TaC	39.5	192	26	<b>14,752</b>	115	176	22	96.5%	0.93	0.97
	Ni	28	11,477	2	<b>28</b>	103	84	20	0.2%	0.03	0
	NbC	23.5	3,014	470	<b>7,353</b>	255	7,478	30	39.5%	0.88	0.4
	TiC	14	12,916	53	<b>53</b>	2,554	16	0	0.3%	0.05	0
	Al	13	479	0	<b>0</b>	17	9	583	0.0%	0	0
<b>227</b>			221	223	225	<b>227</b>	229	230			
	Ge	32	1	2,076	0	<b>83</b>	72	8	3.7%	0.02	0.04
	Si	14	2	9	0	<b>8,726</b>	0	1	99.9%	0.69	1
<b>229</b>			221	223	225	227	<b>229</b>	230			
	W	74	119	2	901	7	<b>34</b>	0	3.2%	0	0.03
	Ta	73	2	3	10	3	<b>8,775</b>	2	99.8%	0.5	1
	Fe	26	7,259	688	20	655	<b>689</b>	205	7.2%	0.07	0.07
<b>230</b>			221	223	225	227	229	<b>230</b>			
	Al <sub>4</sub> CoNi <sub>2</sub>	19.4	0	0	2	0	5	<b>1,822</b>	99.6%	0.3	1
	Al <sub>4</sub> Ni <sub>3</sub>	19.4	65	79	16	13	24	<b>1,595</b>	89.0%	0.27	0.89
Total Patterns	144,465		221	223	225	227	229	230	Average		
# Correct	88,249	Precision	0.37	0.87	0.96	0.69	0.52	0.44	0.75		
Accuracy (%)	61.09	Recall	0.82	0.99	0.36	0.8	0.49	0.94	0.61		

**Supplementary Table 14 Number of diffraction patterns classified to each space group.** Space group 221; trained on FeNi<sub>3</sub>. Space group 223; trained on Mo<sub>3</sub>Si. Space group 225; trained on TaC. Space group 227; trained on Ge. Space group 229; trained on Fe. Space group 230; trained on Al<sub>4</sub>CoNi<sub>2</sub>.

Space Group	Material	Formula Weighted Atomic Number	Number of Patterns Classified to Each Space Group						Accuracy	Precision	Recall
			221	223	225	227	229	230			
<b>221</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	FeNi <sub>3</sub>	27.5	<b>10,004</b>	3	0	0	12	0	99.9%	0.38	1.00
	Ni <sub>3</sub> Al	24.25	<b>9,129</b>	11	1	3	902	87	95.8%	0.35	0.96
	NiAl	20.5	<b>13</b>	9	68	67	1,901	2,482	0.3%	0	0
	FeAl	19.5	<b>3</b>	19	0	26	1,811	4	0.2%	0	0
<b>223</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Mo <sub>3</sub> Si	35	2	<b>10,630</b>	12	4	10	2	99.7%	0.46	1
	Cr <sub>3</sub> Si	21.5	10	<b>12,575</b>	0	3	0	0	99.9%	0.5	1
<b>225</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	TaC	39.5	112	116	<b>14,889</b>	34	118	14	97.4%	0.89	0.97
	Ni	28	11,016	11	<b>46</b>	10	609	22	0.4%	0.02	0
	NbC	23.5	126	1,123	<b>14,173</b>	167	2,965	46	76.2%	0.88	0.76
	TiC	14	4,634	1,922	<b>11</b>	68	8,957	0	0.1%	0.01	0
	Al	13	9	1	<b>0</b>	14	67	997	0.0%	0	0
<b>227</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Ge	32	0	0	0	<b>2,237</b>	1	2	99.9%	0.8	1
	Si	14	504	6,358	1	<b>242</b>	1,282	351	2.8%	0.3	0.03
<b>229</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	W	74	26	44	963	0	<b>30</b>	0	2.8%	0	0.03
	Ta	73	52	2,934	788	106	<b>4,913</b>	2	55.9%	0.21	0.56
	Fe	26	113	31	32	2	<b>9,329</b>	9	98.0%	0.34	0.98
<b>230</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Al <sub>4</sub> CoNi <sub>2</sub>	19.4	0	0	5	2	0	<b>1,822</b>	99.6%	0.31	1
	Al <sub>4</sub> Ni <sub>3</sub>	19.4	9	10	14	52	90	<b>1,617</b>	90.2%	0.29	0.9
Total Patterns	144,465		221	223	225	227	229	230	Average		
# Correct	91,663	Precision	0.54	0.65	0.94	0.82	0.44	0.46	0.73		
Accuracy (%)	63.45	Recall	0.74	1	0.47	0.23	0.74	0.95	0.63		

**Supplementary Table 15 Number of diffraction patterns classified to each space group.** Space group 221; trained on FeNi<sub>3</sub> and NiAl. Space group 223; trained on Mo<sub>3</sub>Si. Space group 225; trained on TaC and Ni. Space group 227; trained on Ge and Si. Space group 229; trained on Ta and Fe. Space group 230; trained on Al<sub>4</sub>CoNi<sub>2</sub>.

Space Group	Material	Formula Weighted Atomic Number	Number of Patterns Classified to Each Space Group						Accuracy	Precision	Recall
			221	223	225	227	229	230			
<b>221</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	FeNi <sub>3</sub>	27.5	<b>8,124</b>	0	1,881	0	13	1	81.1%	0.32	0.81
	Ni <sub>3</sub> Al	24.25	<b>8,524</b>	0	979	0	28	2	89.4%	0.33	0.89
	NiAl	20.5	<b>4,537</b>	0	2	0	0	1	99.9%	0.21	1
	FeAl	19.5	<b>647</b>	0	143	0	1,073	0	34.7%	0.04	0.35
<b>223</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Mo <sub>3</sub> Si	35	100	<b>10,373</b>	21	2	161	3	97.3%	0.95	0.97
	Cr <sub>3</sub> Si	21.5	249	<b>11,403</b>	280	2	654	0	90.6%	0.96	0.93
<b>225</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	TaC	39.5	577	90	<b>13,967</b>	77	562	10	91.4%	0.72	0.91
	Ni	28	362	1	<b>11,334</b>	0	8	9	96.8%	0.68	0.97
	NbC	23.5	1,462	261	<b>13,429</b>	122	3,290	36	72.2%	0.71	0.72
	TiC	14	9,899	8	<b>5,454</b>	0	231	0	35.0%	0.5	0.35
	Al	13	1,047	0	<b>33</b>	0	0	8	3.0%	0.01	0.03
<b>227</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Ge	32	141	34	0	<b>1,936</b>	127	2	86.4%	0.9	0.86
	Si	14	1,485	99	555	<b>6,387</b>	205	7	73.1%	0.97	0.73
<b>229</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	W	74	9	1	1,012	0	<b>41</b>	0	3.9%	0.01	0.04
	Ta	73	139	0	40	2	<b>8,613</b>	1	97.9%	0.57	0.98
	Fe	26	884	0	462	0	<b>8,170</b>	0	85.9%	0.56	0.86
<b>230</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Al <sub>4</sub> CoNi <sub>2</sub>	19.4	16	0	0	0	1	<b>1,812</b>	99.1%	0.96	0.99
	Al <sub>4</sub> Ni <sub>3</sub>	19.4	591	0	0	2	38	<b>1,161</b>	64.8%	0.94	0.65
Total Patterns	144,465		221	223	225	227	229	230	Average		
# Correct	115,945	Precision	0.56	0.98	0.89	0.98	0.72	0.97	0.83		
Accuracy (%)	80.26	Recall	0.84	0.94	0.71	0.76	0.87	0.82	0.80		

**Supplementary Table 16 Number of diffraction patterns classified to each space group.** The model was trained using a small subset of patterns from each of the available materials.

Space Group	Material	Formula Weighted Atomic Number	Number of Patterns Classified to Each Space Group						Accuracy	Precision	Recall
<b>221</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	FeNi3	27.5	9,370	2	551	5	88	3	93.5%	0.65	0.94
	Ni3Al	24.25	9,015	1	378	22	77	40	94.6%	0.64	0.95
	NiAl	20.5	4,381	0	41	16	11	91	96.5%	0.46	0.96
	FeAl	19.5	1,725	0	6	0	131	1	92.6%	0.25	0.93
<b>223</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Mo3Si	35	14	10,572	1	76	33	9	98.8%	0.98	0.99
	Cr3Si	21.5	6	12,544	10	21	7	0	99.7%	0.98	1
<b>225</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	TaC	39.5	115	132	13,601	695	719	21	89.0%	0.93	0.89
	Ni	28	3,160	1	8,107	32	381	33	69.2%	0.89	0.69
	NbC	23.5	272	19	16,404	1,304	553	48	88.2%	0.94	0.88
	TiC	14	166	34	15,230	52	110	0	97.7%	0.94	0.98
	Al	13	64	0	1,015	0	1	8	93.3%	0.5	0.93
<b>227</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Ge	32	5	1	0	2,222	6	6	99.2%	0.49	0.99
	Si	14	4	4	2	8,720	6	2	99.8%	0.79	1
<b>229</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	W	74	0	0	3	0	1,060	0	99.7%	0.33	1
	Ta	73	105	0	4	24	8,657	5	98.4%	0.8	0.98
	Fe	26	1,148	16	37	21	8,246	48	86.7%	0.79	0.87
<b>230</b>			<b>221</b>	<b>223</b>	<b>225</b>	<b>227</b>	<b>229</b>	<b>230</b>			
	Al4CoNi2	19.4	3	0	0	7	1	1,818	99.4%	0.85	0.99
	Al4Ni3	19.4	30	0	0	23	9	1,730	96.5%	0.85	0.97
Total Patterns	144,465		221	223	225	227	229	230	Average		
# Correct	134,417	Precision	0.83	0.99	0.98	0.83	0.89	0.92	0.93		
Accuracy (%)	93.04	Recall	0.94	0.99	0.87	1	0.93	0.98	0.92		

## **Appendix F Supplementary Information for Chapter 7**

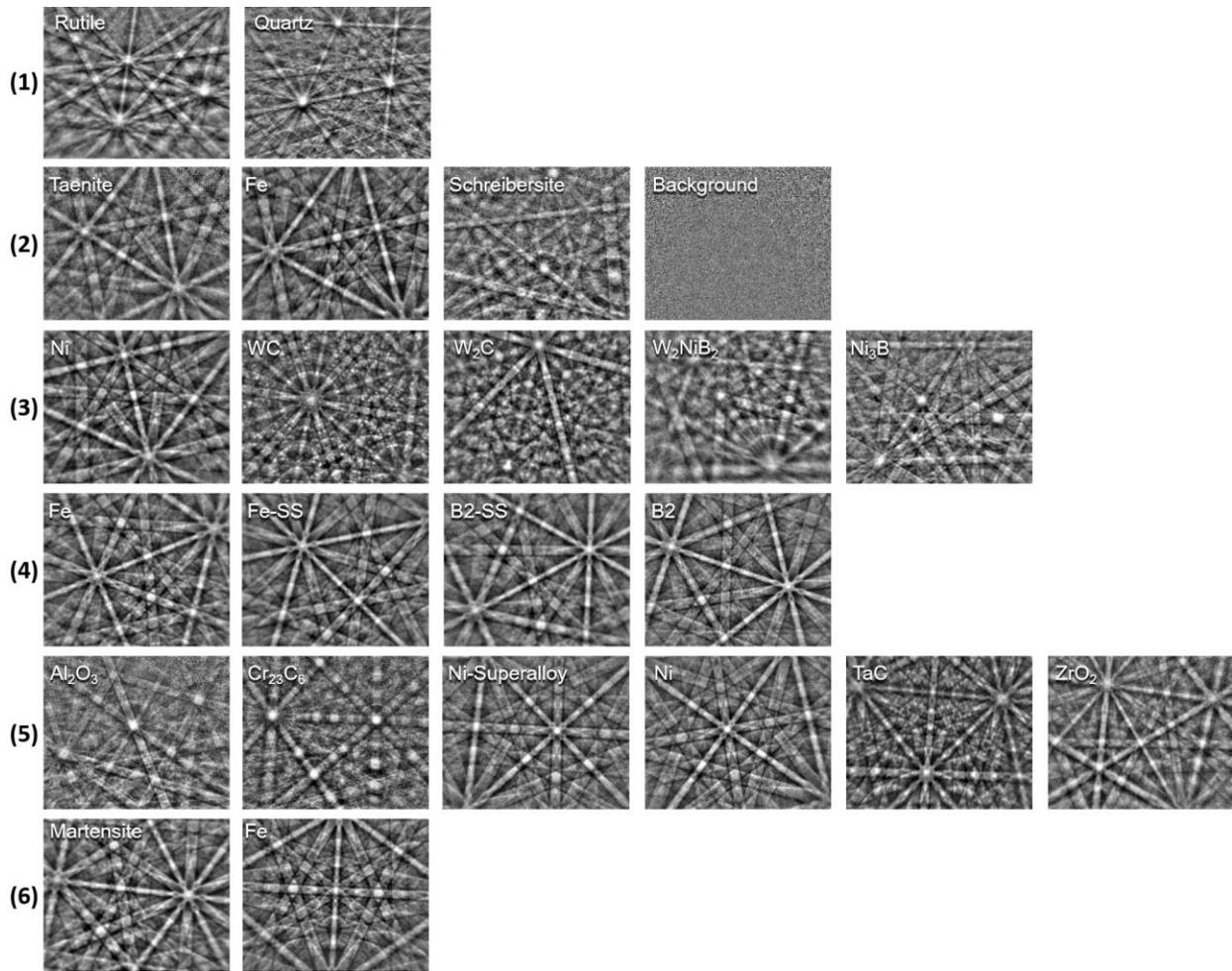
### Phase Mapping in EBSD using Convolutional Neural Networks

Kevin Kaufmann<sup>1</sup>, Chaoyi Zhu<sup>2</sup>, Alexander S. Rosengarten<sup>1</sup>, Daniel Maryanovsky<sup>3</sup>, Haoren Wang<sup>1</sup>, and Kenneth S. Vecchio<sup>1,2</sup>

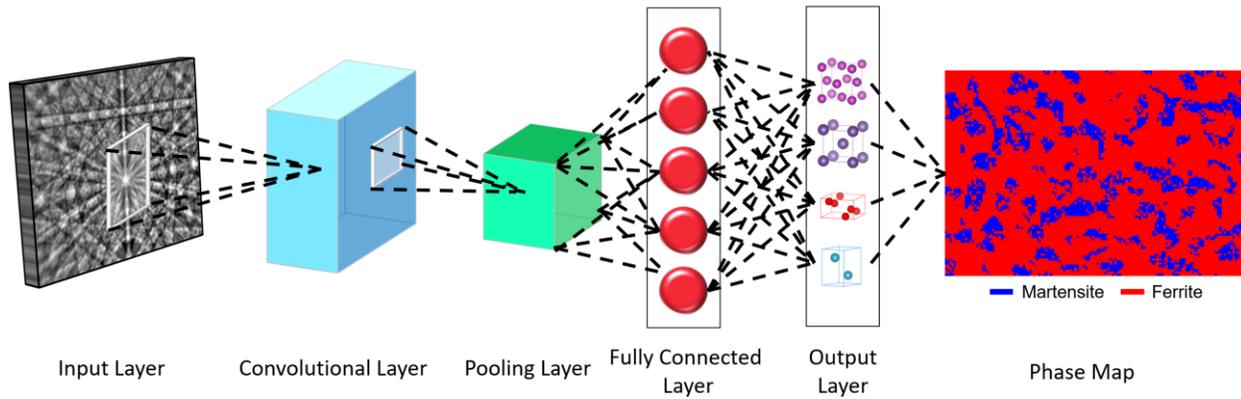
<sup>1</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA

<sup>2</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA

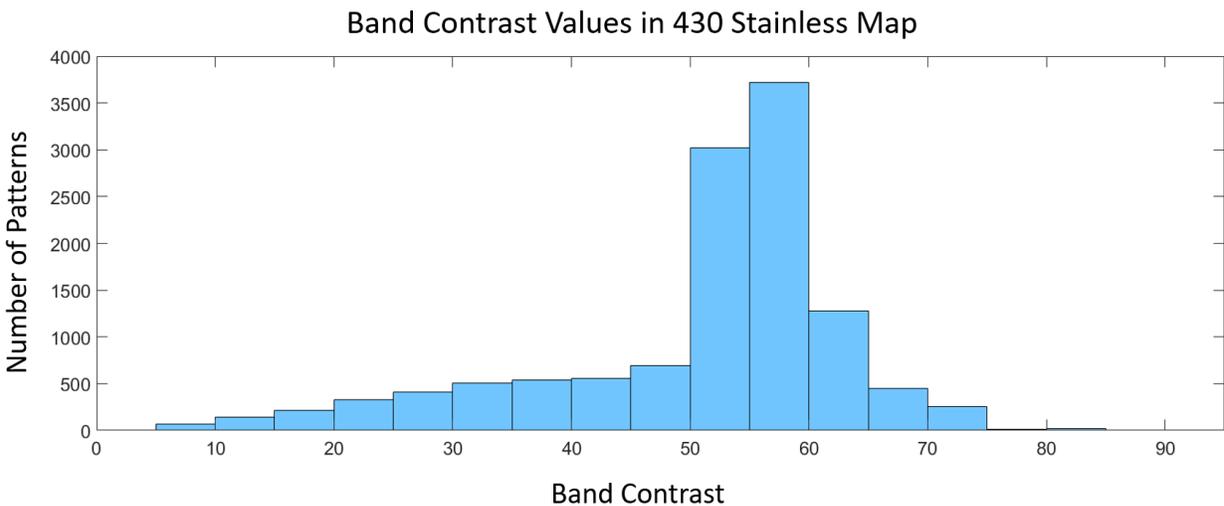
<sup>3</sup>Department of Cognitive Science, UC San Diego, La Jolla, CA 92093, USA



**Supplementary Figure 39 Representative diffraction patterns from each phase in the six materials studied.** All patterns shown are collected from the material studied. The sample number is denoted to the left of each group of patterns.



**Supplementary Figure 40 Schematic of the neural network.** In convolutional layers, a learnable filter is convolved across the image and the scalar product between the filter and the input at every position is computed to form a feature map. Pooling layers are placed after convolutional layers to down sample the feature maps and produce coarse grain representations and spatial information about the features in the data. A traditional dense neural network is placed as the last layer, where the probability that the input diffraction pattern belongs to a given class is computed. These outputs are used to construct a phase map.



**Supplementary Figure 41 Histogram of band contrast values for the 430 stainless steel map.** Band contrast values are binned in groups in steps of 5, starting from 0. The left edge value is included in the count for each bin, but not the right edge value.

## **Appendix G Supplementary Information for Chapter 8**

Efficient *few-shot* machine learning for classification of EBSD patterns

Kevin Kaufmann<sup>1</sup>, Hobson Lane<sup>2,3</sup>, Xiao Liu<sup>4</sup>, and Kenneth S. Vecchio<sup>1,4\*</sup>

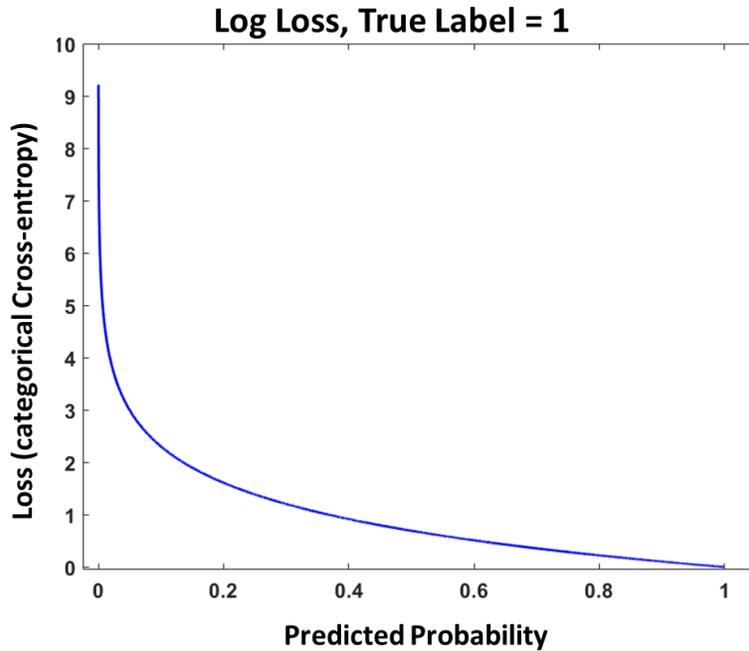
<sup>1</sup>Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA

<sup>2</sup>Tangible AI LLC, San Diego, CA 92037, USA

<sup>3</sup>Department of Healthcare Research and Policy, UC San Diego – Extension, San Diego, CA 92037, USA

<sup>4</sup>Materials Science and Engineering Program, UC San Diego, La Jolla, CA 92093, USA

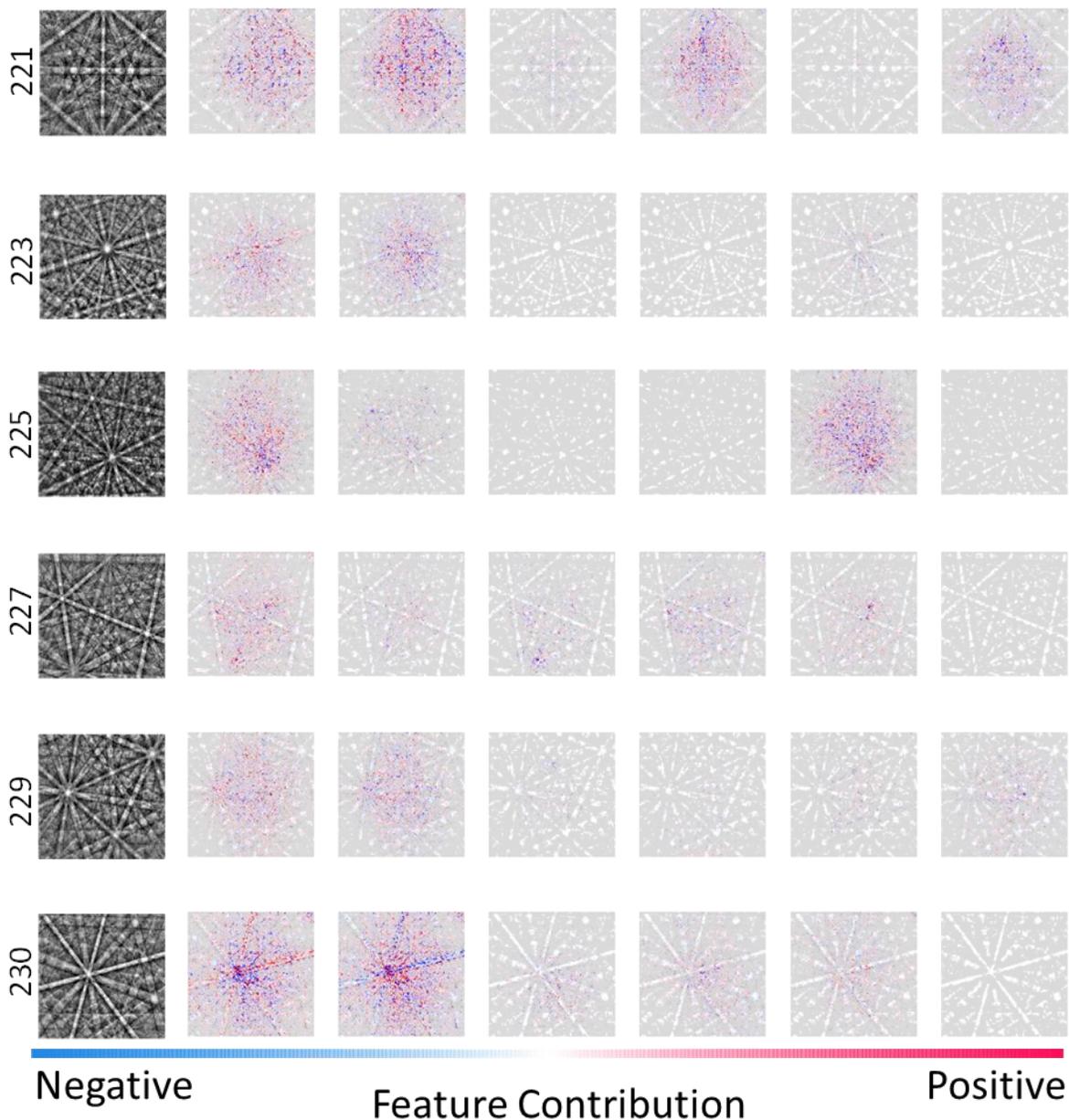
\*Corresponding author



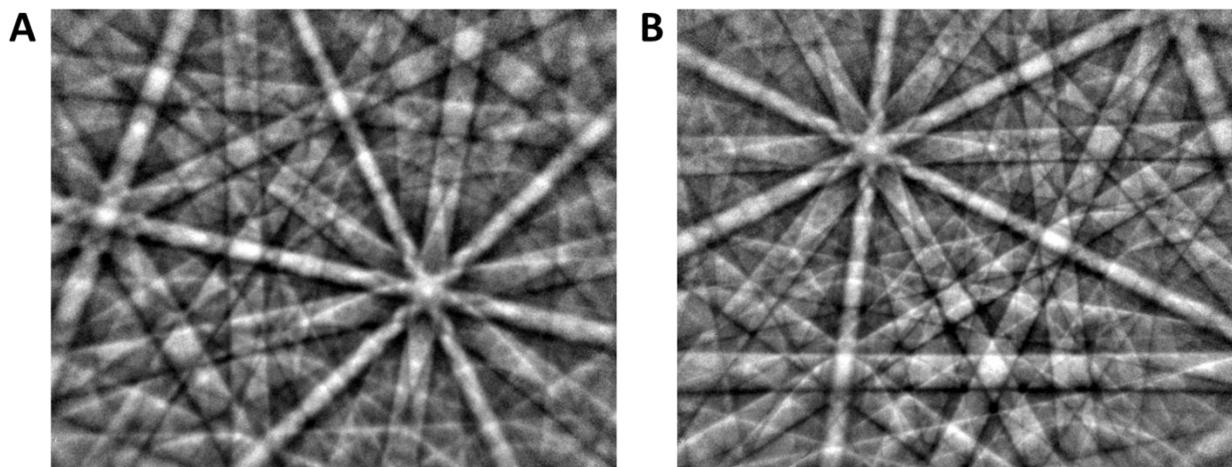
**Supplementary Figure 42 The categorical cross-entropy loss function.** Categorical cross-entropy increases as the predicted probability diverges from the actual label.



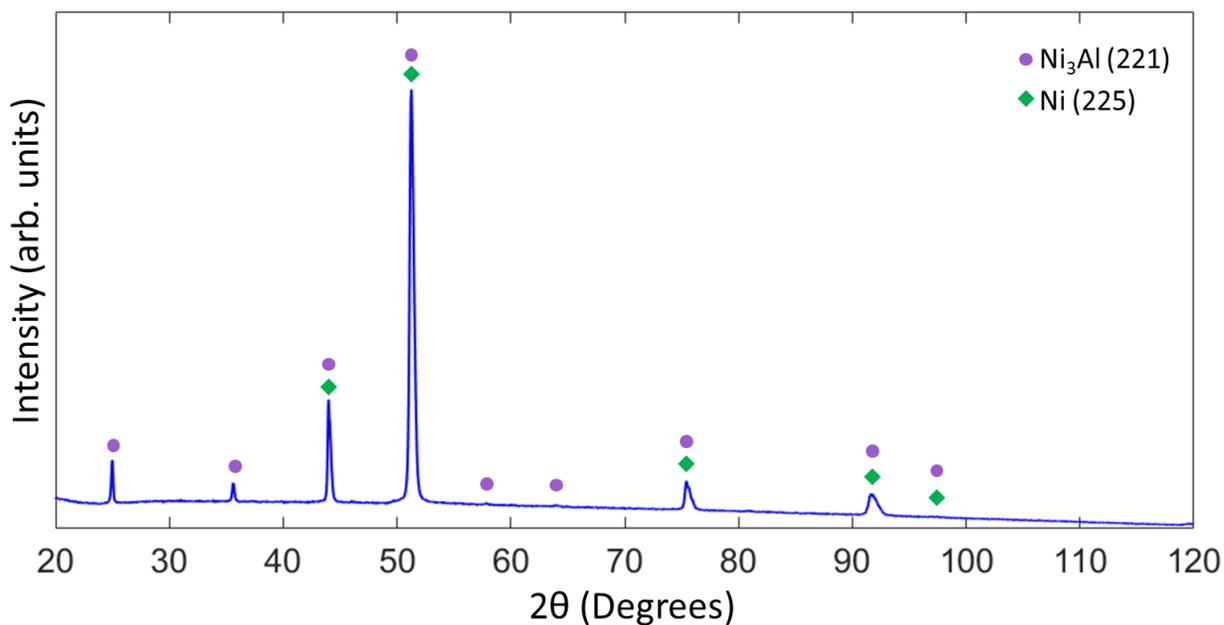
**Supplementary Figure 43 Shapley value analysis for handwritten digits.** Shapley values are computed for several new images to gauge the importance of features (present or not) in predicting the handwritten number. The first column is the raw input image. The ensuing columns correspond to the Shapley values for each possible class in 0-9 order.



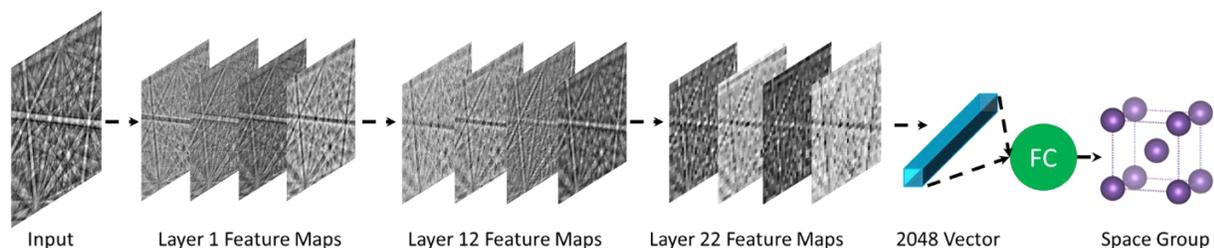
**Supplementary Figure 44 Visual explanation of feature contributions.** Shapley values are computed for each input image to gauge the importance of features in the EBSPs. The first column is the raw input image. The second column corresponds to the Shapley values for the correct prediction. Columns three through seven correspond to incorrect classifications in softmax order.



**Supplementary Figure 45 Example EBSPs from the  $\text{Ni}_{90}\text{Al}_{10}$  sample.** An EBSD from (A) the Ni-rich matrix and (B) the  $\text{Ni}_3\text{Al}$  precipitates are shown as examples.



**Supplementary Figure 46 XRD Pattern for the  $\text{Ni}_{90}\text{Al}_{10}$  sample.** The peaks for  $\text{Ni}_3\text{Al}$  (space group 221) and Ni (space group 225) are labeled with purple circles and green diamonds, respectively.



**Supplementary Figure 47 Schematic of Neural Network Operating on an EBSD.** Individual EBSDs are input to the neural network wherein a series of mathematical operations extract features learned during the training or fine-tuning process. Selected feature maps obtained from several layers are shown. Eventually, the input image is reduced to a 2048-dimensional vector that is passed into a series of fully connected layers (FC) followed by multi-class logistic regression.

**Supplementary Table 17 Classification metrics by space group for the transfer learning model.**

*Precision* is the number of patterns correctly identified to a class divided by the total number of patterns identified as that class. *Recall* is the percentage of patterns in a space group that were correctly identified. F1-score is the weighted harmonic mean of *Precision* and *Recall*. Test EBSDs is the total number of diffraction patterns in the test set for the given class.

Space Group	<i>Precision</i>	<i>Recall</i>	F1-score	Test EBSDs
<b>221</b>	0.89	0.98	0.93	25,955
<b>223</b>	1.00	1.00	1.00	23,248
<b>225</b>	0.99	0.94	0.96	62,277
<b>227</b>	0.95	1.00	0.97	10,978
<b>229</b>	0.98	0.96	0.97	19,374
<b>230</b>	0.87	0.99	0.93	3,621

**Supplementary Table 18 Classification metrics by space group for the model trained from scratch.** Precision is the number of patterns correctly identified to a class divided by the total number of patterns identified as that class. Recall is the percentage of patterns in a space group that were correctly identified. F1-score is the weighted harmonic mean of *Precision* and *Recall*. Support is the total number of images in the test set.

<b>Space Group</b>	<i>Precision</i>	<i>Recall</i>	<b>F1-score</b>	<b>Support</b>
<b>221</b>	0.83	0.94	0.88	25,955
<b>223</b>	0.99	0.99	0.99	23,248
<b>225</b>	0.98	0.87	0.92	62,277
<b>227</b>	0.83	1.00	0.91	10,978
<b>229</b>	0.89	0.93	0.91	19,374
<b>230</b>	0.92	0.98	0.95	3,621

**Supplementary Table 19 Comparison of the symmetry elements that describe each space group.** For each space group in this work, the Bravais lattice and primary, secondary, and tertiary symmetry operations are detailed. Within the ( $4/m\bar{3}2/m$ ) point group, the secondary symmetry element is always a 3-fold rotary inversion, while the primary and tertiary symmetry operations vary.

Space Group Number	Space Group Name	Bravais Lattice	Symmetry and [Direction]		
			Primary [100]/[010]/[001]	Secondary [111]	Tertiary [110]
221	$Pm\bar{3}m$	Primitive	Crystal is mapped onto itself by reflecting across a mirror plane perpendicular to [100]/[010]/[001]	Crystal is mapped back onto itself by a 3-fold rotary inversion in the [111]	Crystal is mapped onto itself by reflecting across a mirror plane perpendicular to [110]
223	$Pm\bar{3}n$	Primitive	Crystal is mapped onto itself by reflecting across a mirror plane perpendicular to [100]/[010]/[001]	Crystal is mapped back onto itself by a 3-fold rotary inversion in the [111]	Crystal is mapped back onto itself by a diagonal glide in the [110]
225	$Fm\bar{3}m$	Face Centered	Crystal is mapped onto itself by reflecting across a mirror plane perpendicular to [100]/[010]/[001]	Crystal is mapped back onto itself by a 3-fold rotary inversion in the [111]	Crystal is mapped onto itself by reflecting across a mirror plane perpendicular to [110]
227	$Fd\bar{3}m$	Face Centered	Crystal is mapped back onto itself by a diamond glide in the [100]/[010]/[001] direction	Crystal is mapped back onto itself by a 3-fold rotary inversion in the [111]	Crystal is mapped onto itself by reflecting across a mirror plane perpendicular to [110]
229	$Im\bar{3}m$	Body Centered	Crystal is mapped onto itself by reflecting across a mirror plane perpendicular to [100]/[010]/[001]	Crystal is mapped back onto itself by a 3-fold rotary inversion in the [111]	Crystal is mapped onto itself by reflecting across a mirror plane perpendicular to [110]
230	$Ia\bar{3}d$	Body Centered	Crystal is mapped onto itself by gliding half the lattice vector in the [100]/a direction	Crystal is mapped back onto itself by a 3-fold rotary inversion in the [111]	Crystal is mapped back onto itself by a diamond glide in the [110]

**Appendix H Supplementary Information for Chapter 9**

**Chapter 1 An Acquisition Parameter Study for Machine-Learning-Enabled**

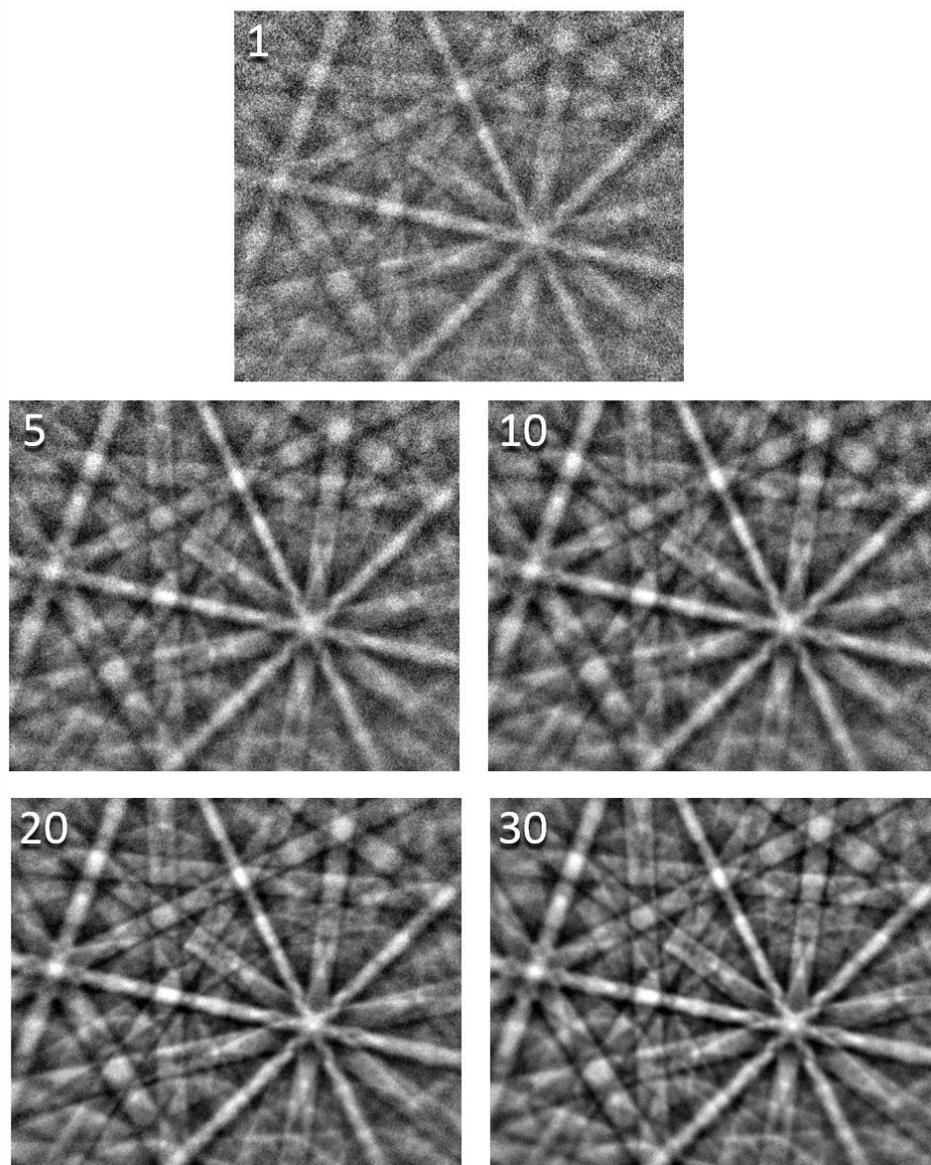
**Electron Backscatter Diffraction**

Kevin Kaufmann and Kenneth S. Vecchio\*

Department of NanoEngineering, UC San Diego, La Jolla, CA 92093, USA

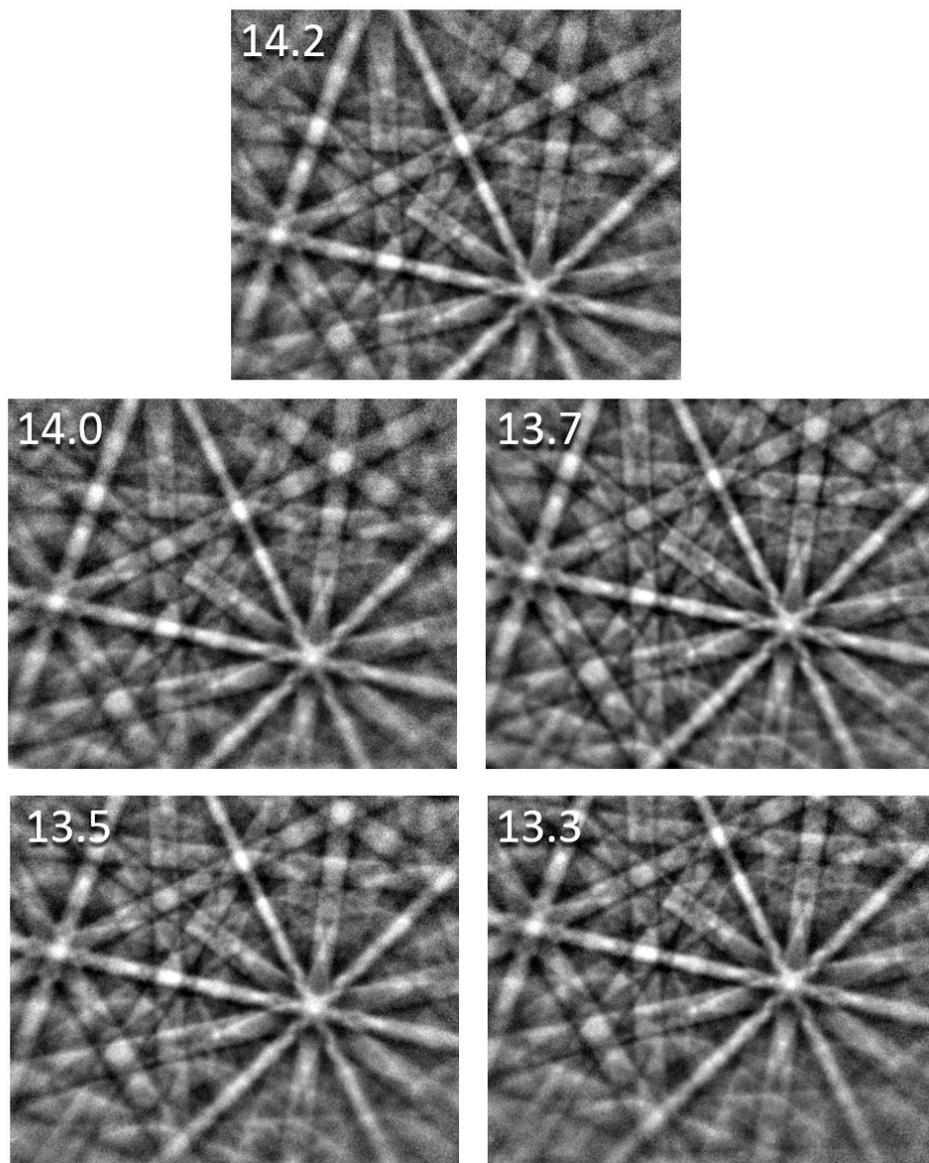
\* Corresponding Author

## Frame Averaging

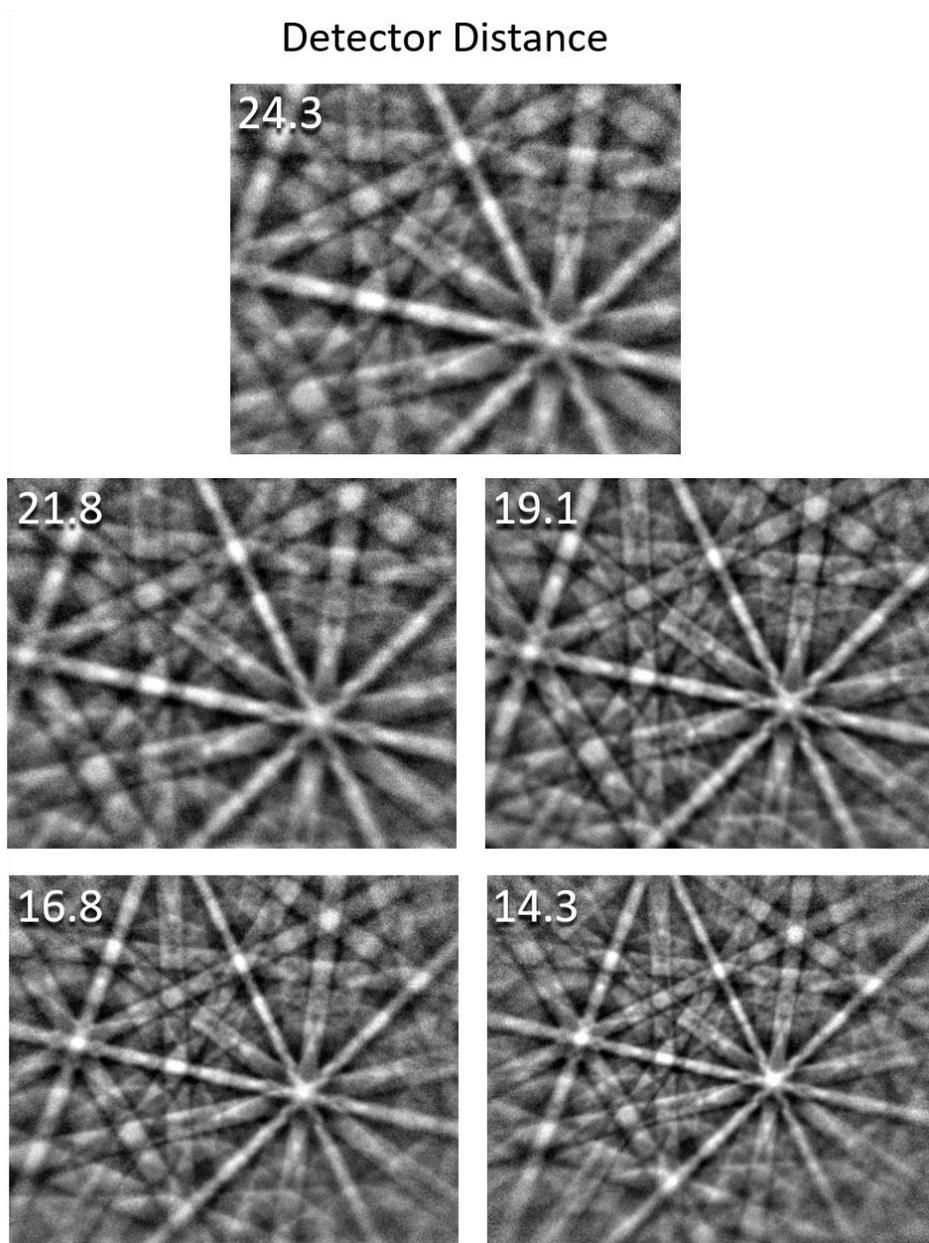


**Supplementary Figure 48 Diffraction pattern with increasing frame averaging.** A visual explanation of the observed changes based on the number of frames averaged during the capture of each diffraction pattern.

## Detector Tilt

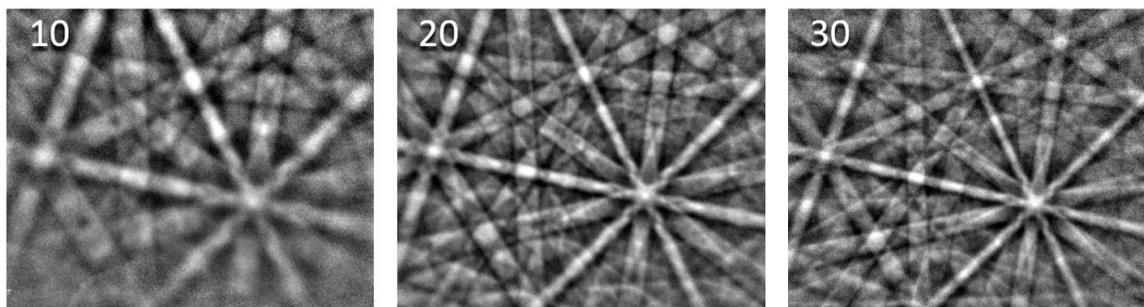


**Supplementary Figure 49 Diffraction pattern with different detector tilt.** A visual explanation of the observed changes based on the tilt of the EBSD detector above the horizontal.



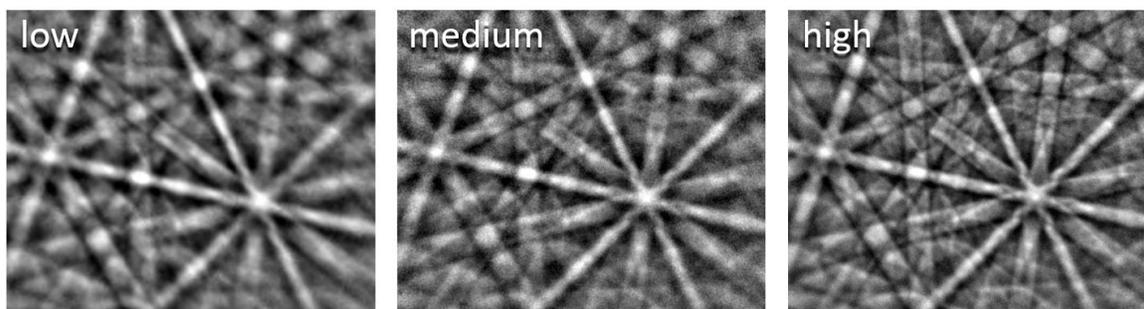
**Supplementary Figure 50 Diffraction pattern with decreasing sample-to-detector distance.** A visual explanation of the observed changes based on the proximity of the EBSD detector to the sample. Larger detector distances are further from the sample.

### Accelerating Voltage

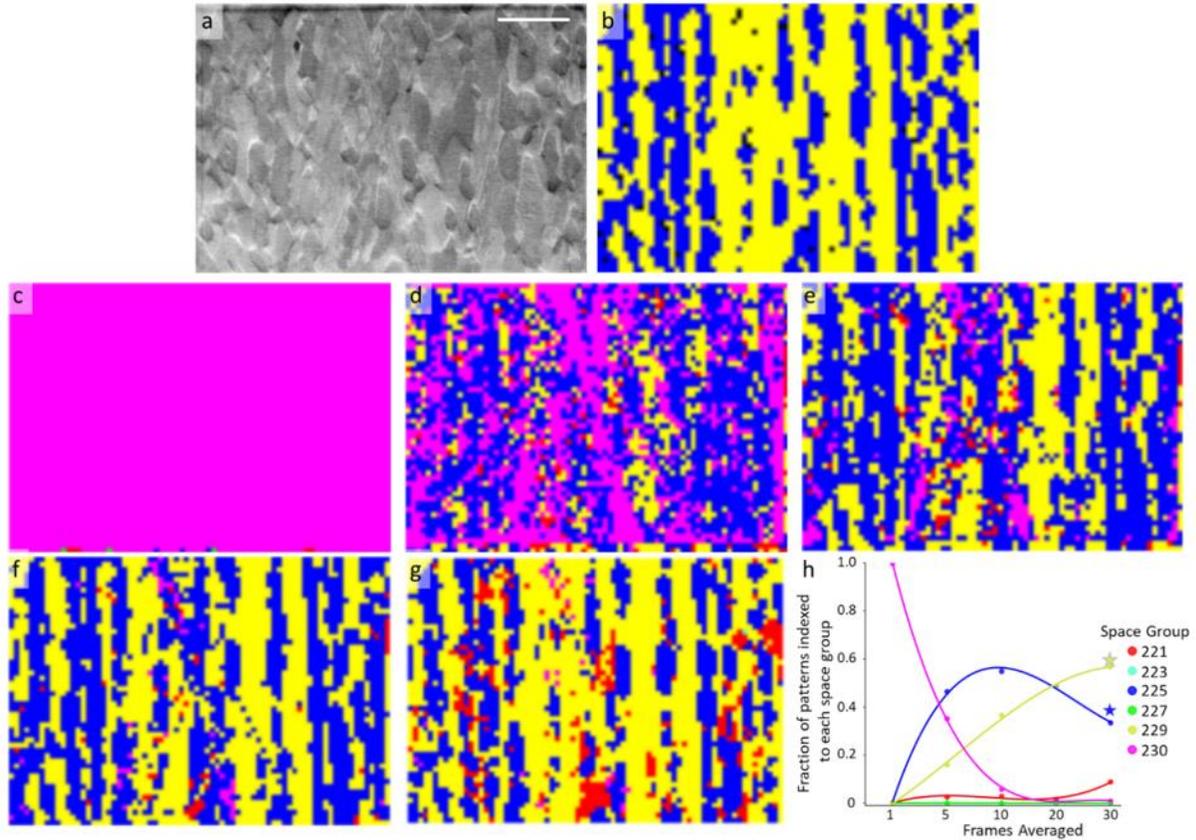


**Supplementary Figure 51 Diffraction pattern with variable accelerating voltage.** A visual explanation of the observed changes based on the accelerating voltage applied to the incoming electrons.

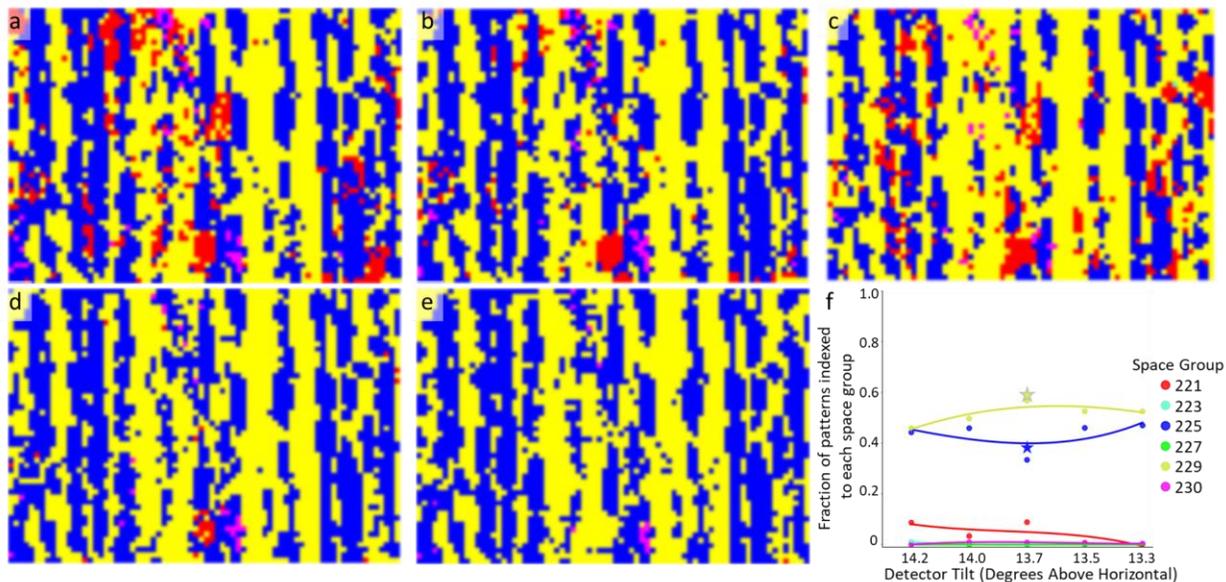
### Pattern Resolution



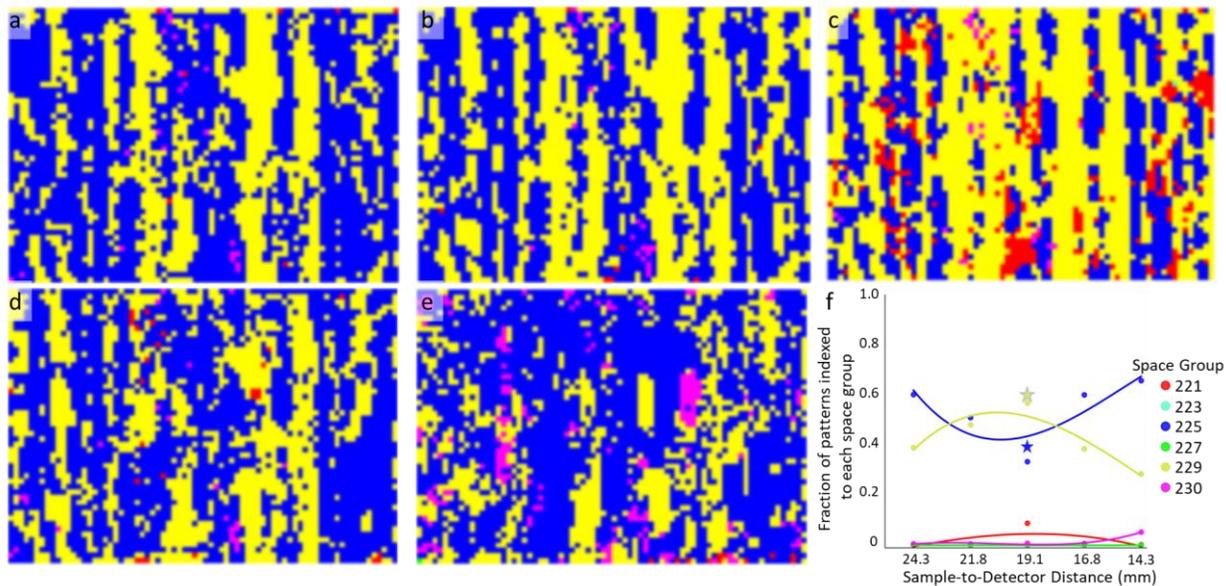
**Supplementary Figure 52 A visual explanation of the observed changes based on the resolution of the EBSD detector.**



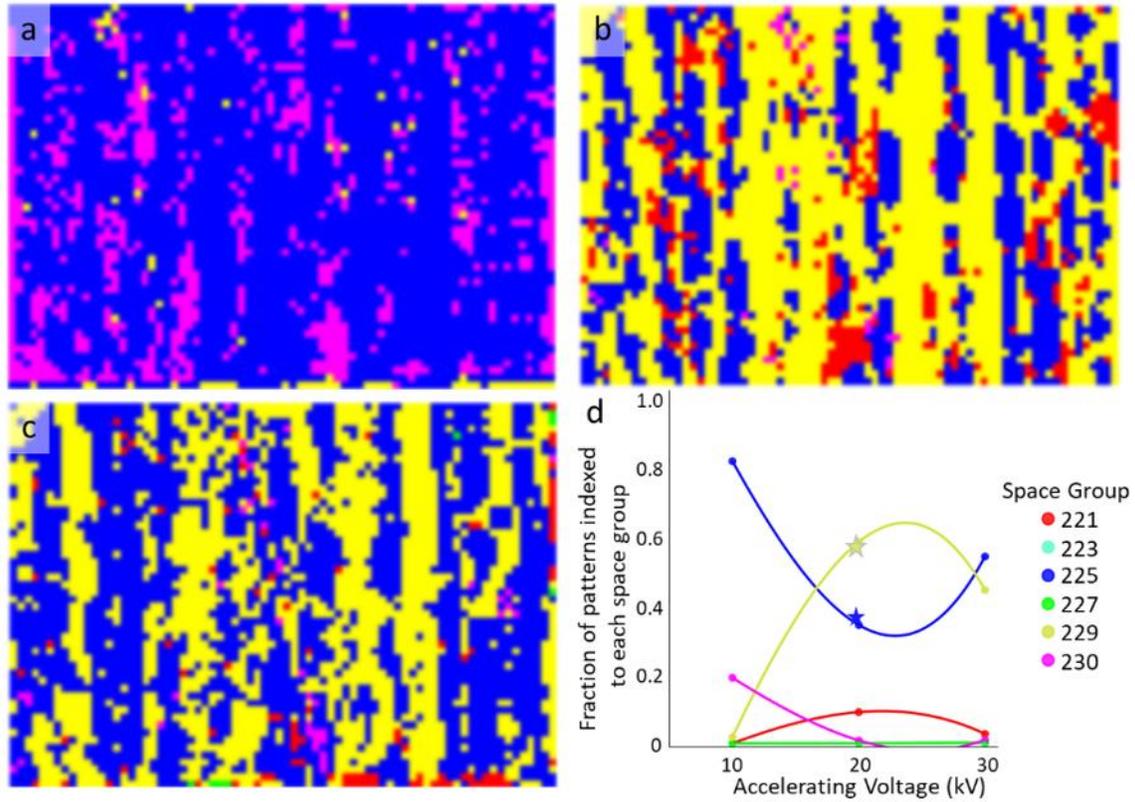
**Supplementary Figure 53 Visual overview of frame averaging on CNN classification for duplex steel.** (a) electron image of the region of dual-phase 2205 duplex steel. (b) Hough-based EBSD phase map of the fcc (225) austenite (blue) and bcc (229) ferrite (yellow). (c) phase map generated from EBSD patterns collected with no frame averaging applied (i.e. one frame). (d) phase map generated from EBSD patterns collected with five frame averaging applied. (e) phase map generated from EBSD patterns collected with ten frame averaging applied. (f) phase map generated from EBSD patterns collected with twenty frame averaging applied. (g) phase map generated from EBSD patterns collected with thirty frame averaging applied. (h) Plot showing the fraction of patterns indexed to each space group as a function of frame averaging. Thirty frame averaging is the default parameter and is designated as such by the blue star for space group 225 and a yellow star for space group 229. Trend lines are fit with a 3<sup>rd</sup> order polynomial. Scale bar 25 $\mu$ m. There are 3,848 diffraction patterns (pixels) in each phase map.



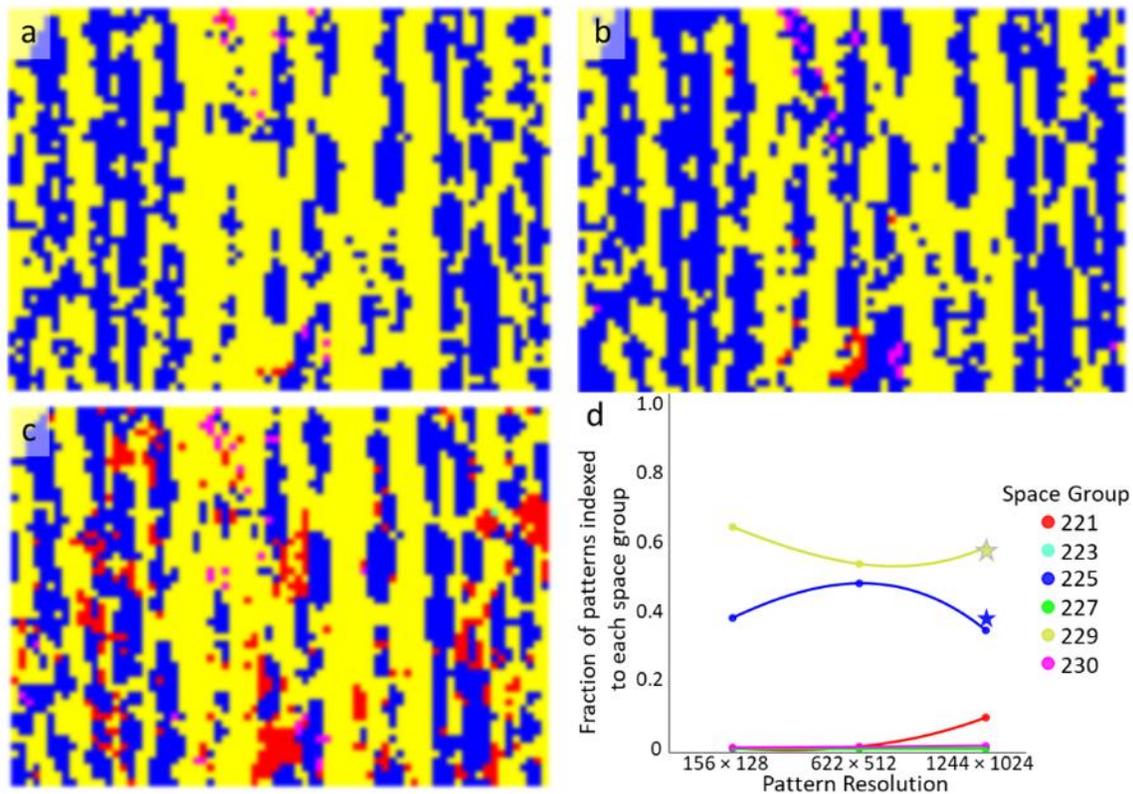
**Supplementary Figure 54 Visual overview of detector tilt on CNN classification for duplex steel.** (a) phase map generated from EBSD patterns collected with a detector tilt of 14.2 degrees. (b) phase map generated from EBSD patterns collected with a detector tilt of 14.0 degrees. (c) phase map generated from EBSD patterns collected with a detector tilt of 13.7 degrees. (d) phase map generated from EBSD patterns collected with a detector tilt of 13.5 degrees. (e) phase map generated from EBSD patterns collected with a detector tilt of 13.3 degrees. (f) Plot showing the fraction of patterns indexed to each space group as a function of detector tilt. A detector tilt of 13.7 degrees above horizontal is the default parameter and is designated as such by the blue star for space group 225 and a yellow star for space group 229. Trend lines are fit with a 3<sup>rd</sup> order polynomial. There are 3,848 diffraction patterns (pixels) in each phase map.



**Supplementary Figure 55 Visual overview of detector distance on CNN classification for duplex steel.** (a) phase map generated from EBSD patterns collected with a sample-to-detector distance of 24.3 mm. (b) phase map generated from EBSD patterns collected with a detector distance of 21.8 mm. (c) phase map generated from EBSD patterns collected with a detector distance of 19.1 mm. (d) phase map generated from EBSD patterns collected with a detector distance of 16.8 mm. (e) phase map generated from EBSD patterns collected with a detector distance of 14.3 mm. (f) Plot showing the fraction of patterns indexed to each space group as a function of sample-to-detector distance. A detector distance of 19.1 mm is the default parameter and is designated as such by the blue star for space group 225 and a yellow star for space group 229. Trend lines are fit with a 3<sup>rd</sup> order polynomial. There are 3,848 diffraction patterns (pixels) in each phase map.



**Supplementary Figure 56 Visual overview of accelerating voltage on CNN classification for duplex steel.** (a) phase map generated from EBSD patterns collected with an electron accelerating voltage of 10kV. (b) phase map generated from EBSD patterns collected with an electron accelerating voltage of 20kV. (c) phase map generated from EBSD patterns collected with an electron accelerating voltage of 30kV. (d) Plot showing the fraction of patterns indexed to each space group as a function of accelerating voltage. An electron accelerating voltage of 20kV is the default parameter and is designated as such by the blue star for space group 225 and a yellow star for space group 229. Trend lines are fit with a 3<sup>rd</sup> order polynomial. There are 3,848 diffraction patterns (pixels) in each phase map.



**Supplementary Figure 57 Visual overview of pattern resolution on CNN classification for duplex steel.** (a) phase map generated from EBSD patterns collected with a detector resolution of  $156 \times 128$  (low). (b) phase map generated from EBSD patterns collected with a detector resolution of  $622 \times 512$  (medium). (c) phase map generated from EBSD patterns collected with a detector resolution of  $1244 \times 1024$  (high). (d) Plot showing the fraction of patterns indexed to each space group as a function of EBSD pattern resolution. The default pattern resolution is  $1244 \times 1024$  and is designated as such by the blue star for space group 225 and a yellow star for space group 229. Trend lines are fit with a 3rd order polynomial. There are 3,848 diffraction patterns (pixels) in each phase map.

**Supplementary Table 20 Pattern acquisition rates.** A summary of the acquisition rates compared to the default settings used in this work. Default conditions: 30 frame averaging, high resolution, 20kV accelerating voltage, 13.7 degree detector tilt, and 19.1 mm detector-to-sample distance. Units: Patterns/second (Hz).

Setting	Default	20 Frames	10 Frames	5 Frames	1 Frame	Medium Resolution	Low Resolution	30 kV	10 kV
Rate (Hz)	5.7	8.6	17	34	171	35	40	9	1.8
Setting	14.2 Degrees	14.0 Degrees	13.5 Degrees	13.3 Degrees	24.3 mm	21.8 mm	16.8 mm	14.3 mm	
Rate (Hz)	5.7	5.7	5.7	5.7	3.5	4.6	7.1	9.5	