

Machine Learning for Structural Analysis in FCC Crystal: Hands-on

Pankaj Rajak

Argonne National Laboratory

Ken-ichi Nomura, Nitish Baradwaj

Collaboratory for Advanced Computing & Simulations

University of Southern California



MAGICS

Materials Genome Innovation
for Computational Software



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Basic Energy Sciences

Outline of ML Classifier Development

- **Phase 1 : Atomic Coordinate to Feature Vector**
 - **Step 1: Extract feature vector from atomic coordinates for training data**
 - **Step 2: Extract feature vector from atomic coordinates for test data**
- **Phase 2 : Train and Validate ML model**
 - **Step 1: Train the ML classifier**
 - **Step 2: Check training error and accuracy**
 - **Step 3: Predict labels of test data**
 - **Step 4: Check test error and accuracy**
- **Phase 3 : Visualize predicted labels in OVITO**

Files in Ml_module in your staging

```
cd staging/Ml_module
```

```
ls Ml_module
```

Output

Makefile	Ni_test2.xyz	createfeature.c
Ni_ML.ipynb	Ni_train.xyz	createfeature.h
Ni_ML.py	atom_property.c	feature/
Ni_test1.xyz	atom_property.h	readinput.py

Compile C Code

Two C codes are used to extract feature vector from atomic coordinates

To build executables, **c_feature**, simply type “make”

make

```
gcc -c -Wall -std=c99 createfeature.c
gcc -c -Wall -std=c99 atom_property.c
gcc -o c_feature createfeature.o atom_property.o -lm
```

Phase 1: Atomic Coordinate to Feature Vector

Step 1: Extract feature vector from atomic coordinates
for training data

./c_feature inputfile outputfile

```
./c_feature Ni_train.xyz feature/train.txt
```

Total number of atoms	114376		
Box size	101.521004	130.103226	101.521004

```
ls feature
```

```
train.txt
```

Phase 1: Atomic Coordinate to Feature Vector

Step 2: Extract feature vector from atomic coordinates
for test data

./c_feature inputfile outputfile

```
./c_feature Ni_test1.xyz feature/test_1.txt
```

```
Total number of atoms    114376
Box size    101.521004    130.103226    101.521004
```

```
ls feature
```

```
test_1.txt  train.txt
```

Phase 2: Train and Validate ML Model

- **Step 1: Train the ML classifier**
- **Step 2: Check training error and accuracy**
- **Step 3: Predict labels of test data**
- **Step 4: Check test error and accuracy**

```
python SVM_model.py training_file testing_file
```

```
python SVM_model.py feature/train.txt  
feature/test_1.txt
```

MAKE SURE YOU TYPE THIS COMMAND IN A SINGLE LINE

Phase 2: Train and Validate ML model

```
python SVM_model.py  
feature/train.txt  
feature/test_1.txt
```

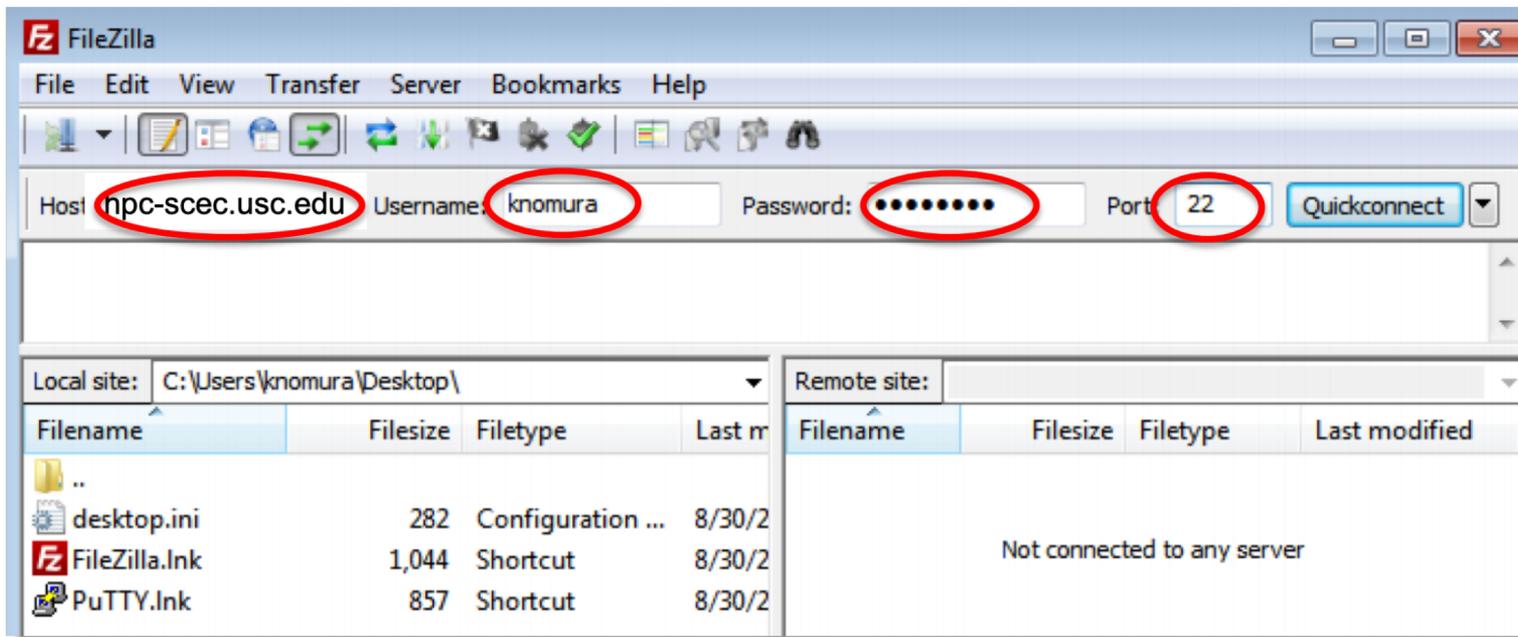
Output

```
Natoms and number of features per atom 114376 17  
Number of training examples: 18655  
training error: 3.2002144197266102  
training accuracy: 96.7997855802734  
Natoms and number of features per atom 114376 17  
Test error: 0.9232706162131898  
Test accuracy: 99.07672938378681
```

Phase 3 : Visualize predicted labels in OVITO

Phase 3: Transfer output.xyz to local machine using Filezilla

- There are four boxes you need to type in. 1-Hostname, 2-Username, 3-Password, and 4-Port.
- Port is always **22**.



Host: hpc-scec.usc.edu
Username: magicsXX

Phase 3: Transfer output.xyz to local machine using Filezilla

The screenshot shows the Filezilla interface with two panels. The left panel shows the local site at /Users/nbaradwaj/Desktop/ with a tree view of folders including Desktop, Documents, Downloads, Dropbox, Fall_2016, Google Drive, and HW3_NR. The right panel shows the remote site at /auto/rcf-40/magics60 with a tree view of folders including ., .pki, .ssh, .vim, ML_module, Pankaj_ML_module, magics, staging, authority, .bash_history, and .bash_logout. The 'staging' directory is selected and highlighted in blue. A red arrow points from the 'staging' directory to the text below.

Filename	Filesize	Filetype	Last modified
..			
APS Membership Application_files		Directory	10/26/2018 16:2...
CSCI_567_ML		Directory	10/02/2018 15:4...
DC_Workshop_Testimonials_raw		Directory	11/12/2018 18:0...
Day1_Handson		Directory	11/02/2018 17:3...
Jun 28		Directory	09/25/2018 01:4...
LECTURES on Machine Learning		Directory	09/19/2018 19:1...
Master		Directory	10/25/2018 17:0...
Pankaj_ML_module		Directory	11/11/2018 01:03...
Pankaj_ML_module 2		Directory	11/08/2018 19:3...
VESTA		Directory	03/27/2018 22:4...

Filename	Filesize	Filetype
..		Directory
.pki		Directory
.ssh		Directory
.vim		Directory
ML_module		Directory
Pankaj_ML_module		Directory
magics		Directory
staging		Directory
authority	377	File
.bash_history	13,247	File
.bash_logout	62	File

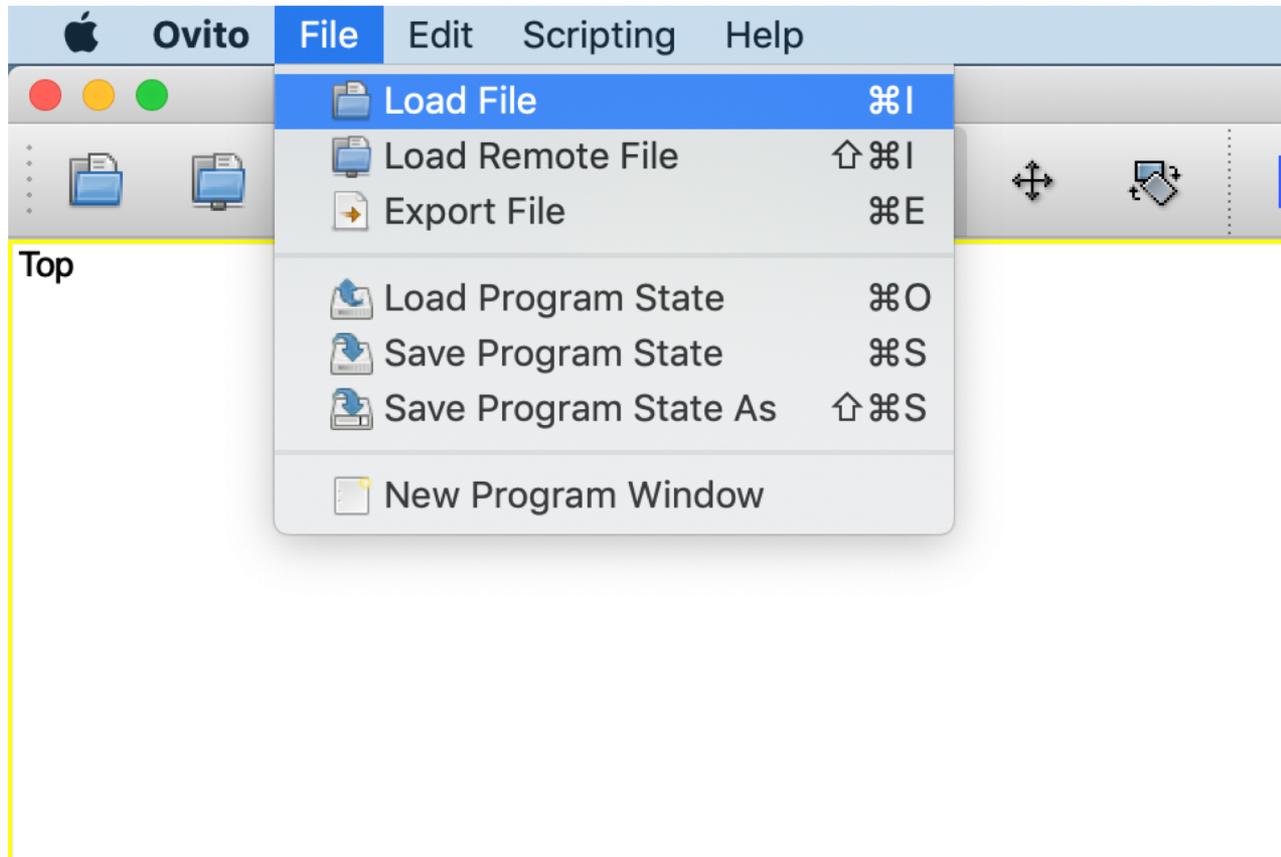
26 files and 10 directories. Total size: 16,856,250 bytes

Selected 1 directory.

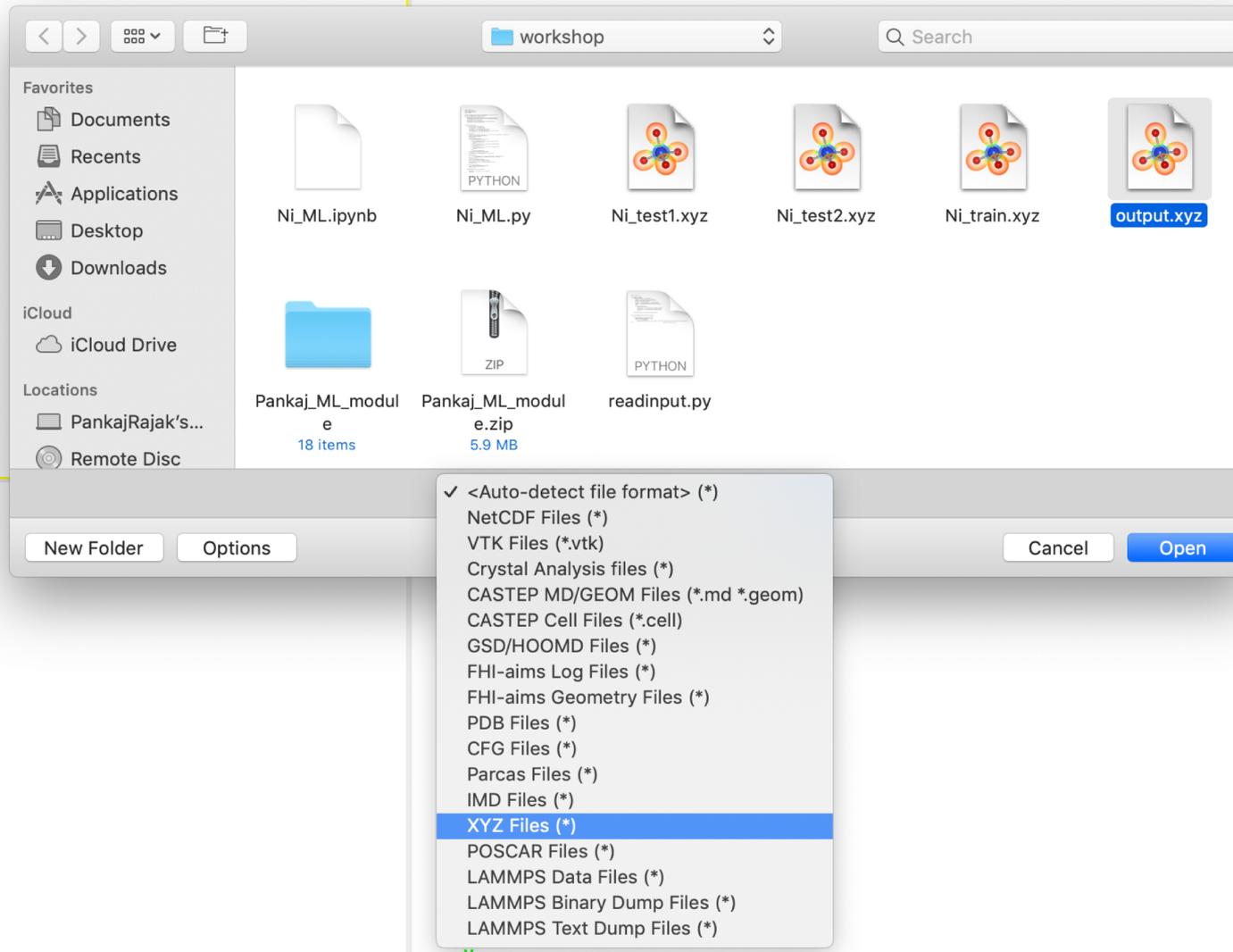
Double Click on staging. Navigate to ML_module directory and transfer output.xyz to your local machine

Load output.xyz in OVITO

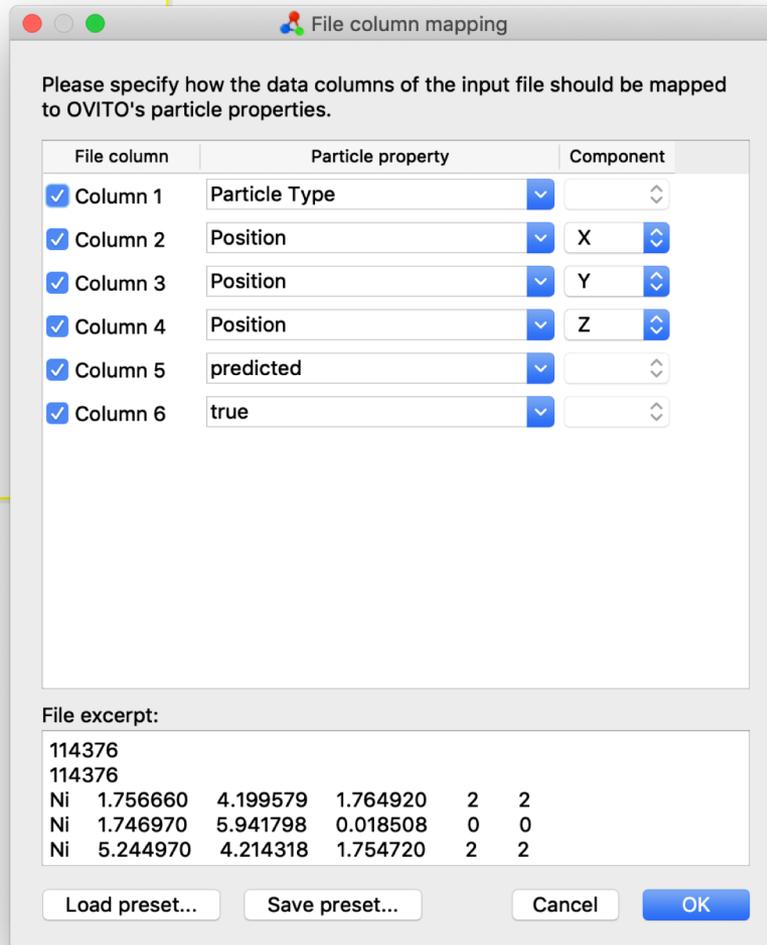
Open file **output.xyz** in OVITO



Load “output.xyz” in OVITO



Load “output.xyz” in OVITO



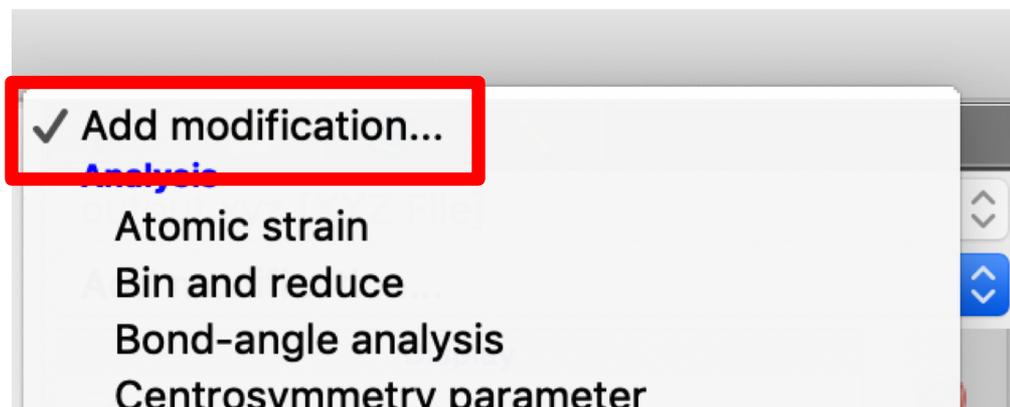
Load “output.xyz” in OVITO

File column	Particle property	Component
<input checked="" type="checkbox"/> Column 1	Particle Type	
<input checked="" type="checkbox"/> Column 2	Position	X
<input checked="" type="checkbox"/> Column 3	Position	Y
<input checked="" type="checkbox"/> Column 4	Position	Z
<input checked="" type="checkbox"/> Column 5	predicted	
<input checked="" type="checkbox"/> Column 6	true	

Type in these two columns explicitly

Select “Expression Select”

From “Add modification...”,
select “Expression Select”



Select Bulk atoms

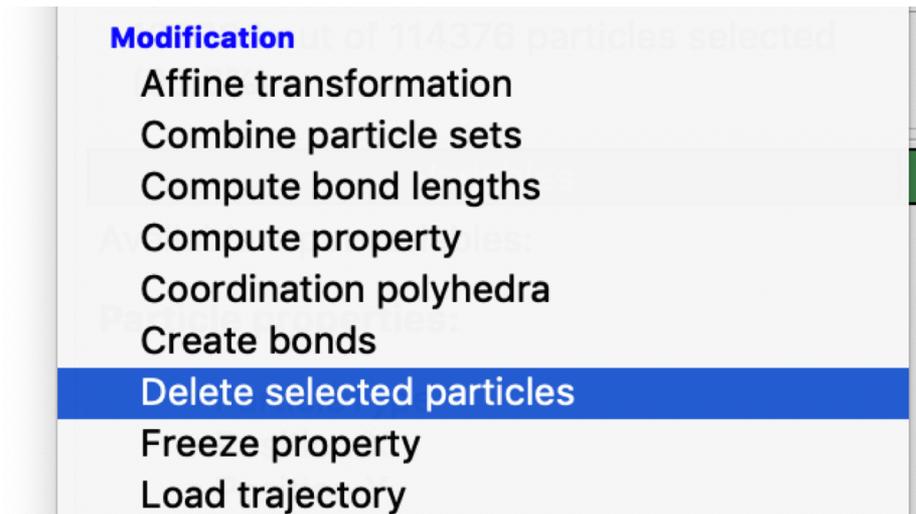
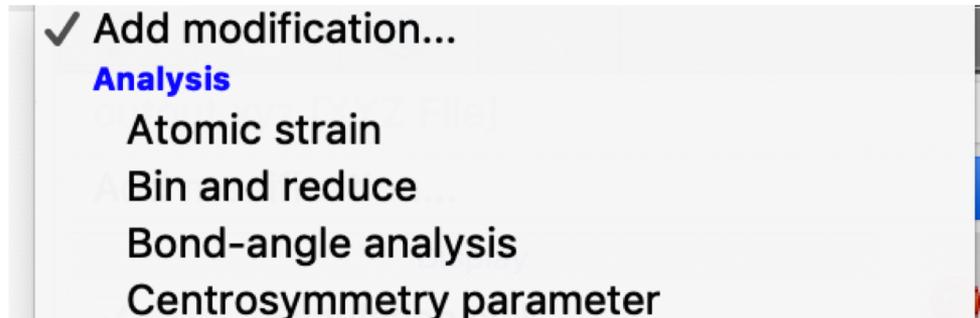
In Expression Select
write

predicted == 0

The screenshot shows a software interface for selecting atoms. At the top, a window title reads 'output.xyz [XYZ File]'. Below it is a section titled 'Add modification...' with a dropdown arrow. Underneath, there are three sections: 'Display', 'Modifications', and 'Input'. In the 'Display' section, 'Simulation cell' and 'Particles' are checked. In the 'Modifications' section, 'Expression select' is checked and highlighted with a red box. In the 'Input' section, 'output.xyz [XYZ File]', 'Simulation cell', and 'Particle types' are listed. To the right of the list are several icons: a red circle with a slash, a green up arrow, a green down arrow, a blue circular arrow, and a blue question mark. Below the list is a tab labeled 'Expression select' with a green question mark icon. A red box highlights the text 'Boolean expression: predicted == 0'. At the bottom, a status bar shows '109394 out of 114376 particles selected (95.6%)'.

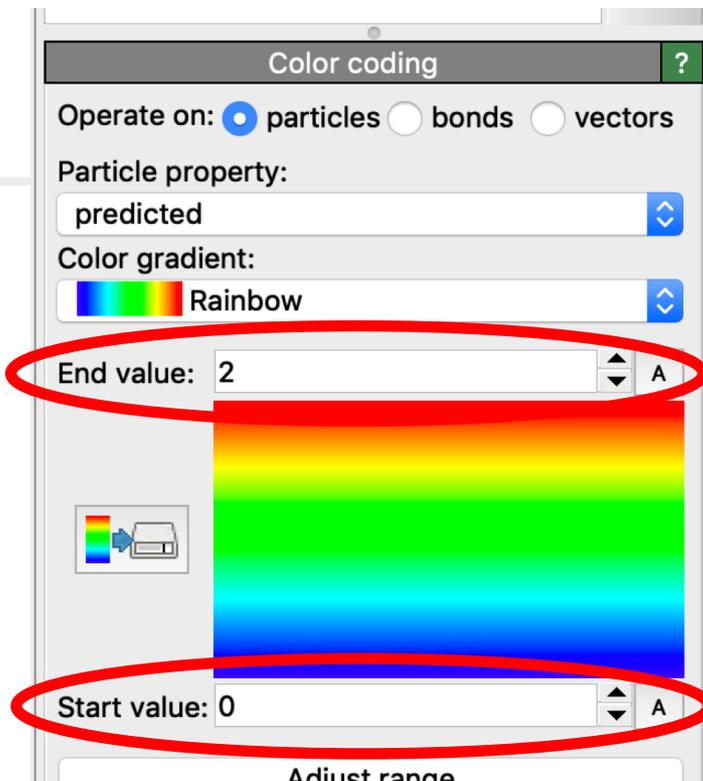
Delete Selected Particles

From "Add modification",
select "Delete selected particles"



Color Remaining Atoms

From “Add modification...”
Color coding



Type 0 in Start value
Type 2 in End value

- ✓ Add modification...
- Analysis**
- Atomic strain
- Bin and reduce
- Bond-angle analysis
- Centrosymmetry parameter
- Cluster analysis
- Common neighbor analysis
- Construct surface mesh
- Coordination analysis
- Correlation function
- Dislocation analysis (DXA)
- Displacement vectors
- Elastic strain calculation
- Histogram
- Identify diamond structure
- Polyhedral template matching
- Scatter plot
- Voronoi analysis
- VoroTop analysis
- Wigner-Seitz defect analysis
- Coloring**
- Ambient occlusion
- Assign color
- Color coding**
- Fields**
- Create isosurface
- Modification**

Step 5: Visualize predicted label

