TiNiSi-type *vs*. ZrNiAl-type: One More Time with Interpretable ML and Experimental Validation

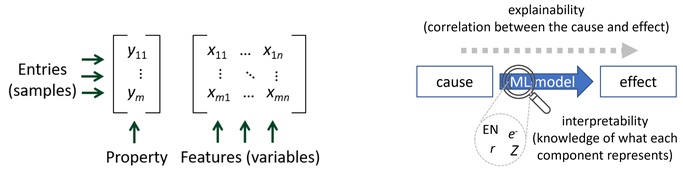
Anton Oliynyk <anton.oliynyk@hunter.cuny.edu>, Sangjoon (Bob) Lee, Nikhil Kumar Barua

Hunter College, Manhattan College, University of Alberta — *Columbia University, The Cooper Union* — *University of Waterloo*

Background

Machine-learning methods have exciting potential to aid materials discovery, but their wider adoption can be hindered by the opaqueness of many models. Even if these models are accurate, the inability to understand the basis for the predictions breeds skepticism. Thus, it is imperative to develop machine-learning models that are explainable and interpretable so that researchers can judge for themselves if the predictions are consistent with their own scientific understanding and chemical insight.

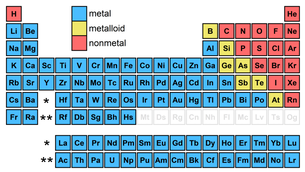
We search for complex correlations in order to predict desired property

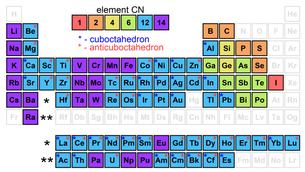


Data and structural variables

It is important to define the compositional playground. We work on intermetallics. What is an intermetallic? A compound that has a significant fraction of metallic bonding, but composition can include metalloids and small amounts of nonmetals.

In order to have a consistent dataset, we define that we consider metals only elements with CN over 12, which opens many possibilities for structural variables.

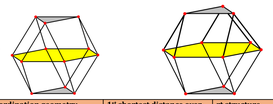




One structural variable that later will play a significant role is metal CN polyhedron distortion. To illustrate its importance, we have a short example below.

Are these equally highly symmetrical polyhedra?

Cuboctahedron Anticuboctahedron



	CN	Coordination geometry	1st shortest distance over 2nd shortest	rt structure type	Transition temperature
Sc	12	anticuboctahedron	0.984	Mg	1337 C
Co	12	anticuboctahedron	0.996	Mg	450 C
Y	12	anticuboctahedron	0.975	Mg	1478 C
Ru	12	anticuboctahedron	0.978	Mg	
La	12	Anticuboctahedron/ cuboctahedron	0.991	Nd	500 C
Ce	12	Anticuboctahedron/ cuboctahedron	0.990	Nd	800 C
Pr	12	Anticuboctahedron/ cuboctahedron	0.991	Nd	795 C
Nd	12	Anticuboctahedron/ cuboctahedron	0.992	Nd	863 C
Sm	12	Anticuboctahedron/ cuboctahedron	range	Sm	734 C
Gd	12	anticuboctahedron	0.983	Mg	1235 C
Tb	12	anticuboctahedron	0.978	Mg	1289 C
Dy	12	anticuboctahedron	0.976	Mg	1381 C
Ho	12	anticuboctahedron	0.975	Mg	
Er	12	anticuboctahedron	0.974	Mg	
Tm	12	anticuboctahedron	0.974	Mg	
Lu	12	anticuboctahedron	0.979	Mg	
Os	12	anticuboctahedron	0.978	Mg	

Fun fact: actually, in structures with anticuboctahedral polyhedron 6 atoms are at the shortest distance and the other 6 are at a slightly longer distances.

How important this is?
In no time ML identified that these tiny distortions are the factor that defines the phase transition temperature in elements (*e.g.*, Co).

Introducing structural variables in ML is essential!
Since structure defines property!

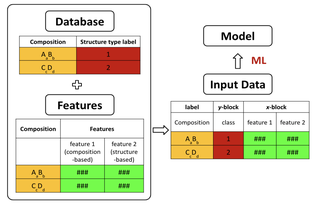
Extracting structural variables our way

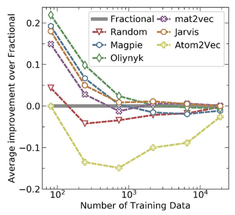
Reliable variables/descriptors/features are essential for ML. They matter more than the choice of the algorithms.

Basically, you want to convert chemical compositions and geometry of the structures into numbers to form a matrix, which becomes an input for machine learning.

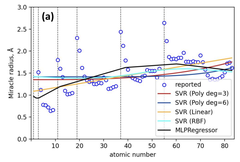
The most common featurizers are based on composition.
The most intuitive proto-ML features were introduced hundred years ago and are known to us as radius ratio and electronegativity difference. Besides size, electron configuration, physical properties of constituent elements there are infinite possibilities to combine them: ratio, sum, *etc*.

Composition-based features were extensively studied and tested, even with ML-predicted missing values in size scales

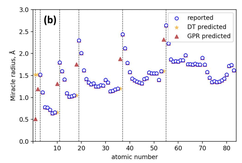




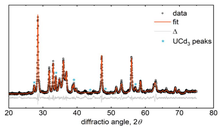
Unsuccessful prediction of missing property values



Successful prediction of missing property values

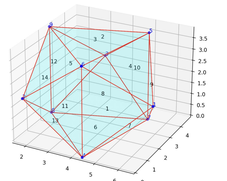


ML-generated composition features were experimentally validated with prediction of a new compound.

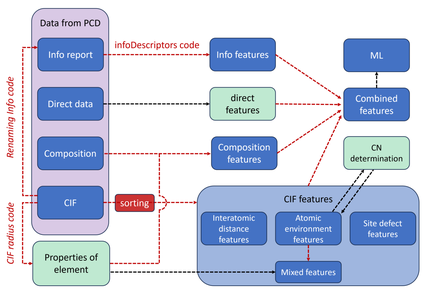


CIF-derived features
Structure governs property. We need to extract maximum of the information from structure (not just a space group number).

Hundreds of thousands standardized CIFs contain useful information which could be easily extracted with modern python packages and skilled undergraduate students.



Rh CN example. Finding a suitable coordination number (CN) is not an easy task, various methods are available, from d/d_{min} to Frank-Kasper .



We can extract useful data for structure analysis and ML:

- Various interatomic features: distances, refined atomic sizes, size scale fitness.
- Atomic environment features: symmetry, multiplicity of sites, CN, polyhedron distortions, packing efficiency.
- Site defect features: likelihood of site deficiency, the extent of site deficiency, polyhedron volume of site-deficient site.

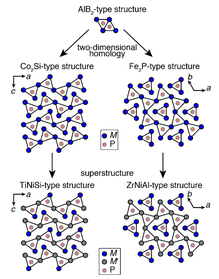
One of the challenges is selecting the method for determination of the coordination number. A user can pick the method they prefer, or the method that works for their systems, and this method will be used for calculation of the structural features, which advance any ML model.

URhIn CN determination example:

Ref atom	CN dist method	CN number	R atom count in CN	M atom count in CN	X atom count in CN	Volume of polyhedron, A^3	Dist from central atom to center of mass, A	Number of edges	Number of faces	Shortest distance to center of any face, A	Shortest distance to center of any edge, A	Volume of the inscribed sphere, A^3	Packing efficiency of inscribed sphere in polyhedron
In1	dist by shortest dist	14	6	4	4	86.72993	0.13697	36	24	2.43497	2.55678	60.47396	0.69727
In1	dist by CIF rad sum	14	6	4	4	86.72993	0.13697	36	24	2.43497	2.55678	60.47396	0.69727
In1	dist by CIF rad refined sum	12	6	4	2	71.61151	0.1598	30	20	2.03691	1.97172	35.4001	0.49434
In1	dist by Pauling rad sum	14	6	4	4	86.72993	0.13697	36	24	2.43497	2.55678	60.47396	0.69727
U1	dist by shortest dist	11	0	5	6	60.6529	0.32792	27	18	2.02892	1.94995	34.98527	0.57681
U1	dist by CIF rad sum	11	0	5	6	60.6529	0.32792	27	18	2.02892	1.94995	34.98527	0.57681
U1	dist by CIF rad refined sum	17	6	5	6	115.67168	0.03188	45	30	2.65258	2.66136	78.17957	0.67587
U1	dist by Pauling rad sum	17	6	5	6	115.67168	0.03188	45	30	2.65258	2.66136	78.17957	0.67587
Rh1	dist by shortest dist	9	6	0	3	52.07195	4.00E-05	21	14	1.9405	2.24708	30.60762	0.58779
Rh1	dist by CIF rad sum	9	6	0	3	52.07195	4.00E-05	21	14	1.9405	2.24708	30.60762	0.58779
Rh1	dist by CIF rad refined sum	9	6	0	3	52.07195	4.00E-05	21	14	1.9405	2.24708	30.60762	0.58779
Rh1	dist by Pauling rad sum	9	6	0	3	52.07195	4.00E-05	21	14	1.9405	2.24708	30.60762	0.58779
Rh2	dist by shortest dist	9	3	0	6	44.24485	0	21	14	1.9405	1.87273	30.60762	0.69178
Rh2	dist by CIF rad sum	9	3	0	6	44.24485	0	21	14	1.9405	1.87273	30.60762	0.69178
Rh2	dist by CIF rad refined sum	9	3	0	6	44.24485	0	21	14	1.9405	1.87273	30.60762	0.69178
Rh2	dist by Pauling rad sum	9	3	0	6	44.24485	0	21	14	1.9405	1.87273	30.60762	0.69178

Prediction and understanding TiNiSi/ZrNiAl-type changes in equiatomic structures

RE–M–X systems form equiatomic ternary compounds, most commonly in related TiNiSi- and ZrNiAl-type structures. The electron count plays a role, since by changing element *X* (In, Sn, Sb) structure changes from ZrNiAl-type to TiNiSi-type to no compound at all respectively. However, the gradual structure change happens with the *RE* substitution. Interpretable ML helps to identify a very specific factor that governs this structure change.



ZrNiAl-type | TiNiSi-type

rt – room-temperature polymorph, with the different structure existing at other temperatures.

	Ni			Pd			Pt	
	In	Sn		In	Sn		In	Sn
La								
Ce					other			
Pr								
Nd								
Sm								rt
Gd								rt
Tb								rt
Dy								
Ho					rt			
Er					both			
Tm					rt			
Lu								

The trend for the series, where In compound adopts ZrNiAl-type and the Sn compound adopts TiNiSi-type depends on the factors that could be identified with ML. The most important factors are relative compressibility of the elements and electron count divided by polyhedron volume (CIF-derived features).

For instance, Ni (bulk modulus 177 GPa) has this ZrNiAl-TiNiSi change for the entire series of *RE* compounds, Pd (187 GPa) has the series change around Ho (44.4 GPa), while Pt (276 GPa) changes the structure at Sm (37.8 GPa)

	La	Ce	Pr	Nd	Sm	Gd	Tb	Dy	Ho	Er	Tm	Lu
Bulk modulus, GPa	27.9	21.5	28.8	31.8	37.8	37.9	38.7	40.5	40.2	44.4	44.5	47.6

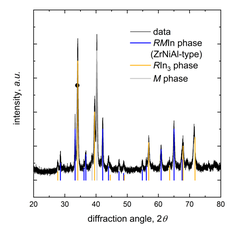
Experimental validation

Interpretable ML helps us to find specific factors that determine what structure forms. In other words, we extract chemical knowledge that allows us to predict what compositions can exhibit similar phenomenon, *e.g.*, structure change across *RE* series.

With the extracted knowledge, we screen chemical composition space, rank the compositions and synthesize them to confirm that with the factors identified with ML we will form *REM*In compound with ZrNiAl-type structure, *REM*Sn compound with TiNiSi-type structure, and no equiatomic ternary phase exists in *RE–M–*Sb system.

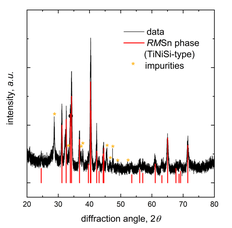
Experimental validation for a specific system is shown below.
(*RE* and *M* elements are not disclosed here, since the results are unpublished yet, and these new compounds are easy to make.)

*RE–M–*In



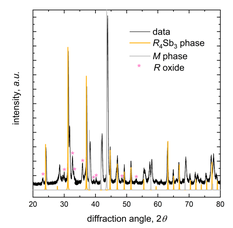
*REM*In is a new compound that adopts ZrNiAl-type structure, as predicted with ML

*RE–M–*Sn



*REM*Sn is a new compound that adopts TiNiSi-type structure, as predicted with ML

*RE–M–*Sb



*REM*Sn does not form, as predicted with ML

References:
Interpretable Machine Learning in Solid-State Chemistry, with Applications to Perovskites, Spinels, and Rare-Earth Intermetallics: Finding Descriptors Using Decision Trees, DOI: 10.1021/acs.inorgchem.3c01153
Revealing Hidden Patterns through Chemical Intuition and Interpretable Machine Learning: A Case Study of Binary Rare-Earth Intermetallics RX, DOI: 10.1021/acs.chemmater.2c02425
Machine learning descriptors in materials chemistry: prediction and experimental validation synthesis of novel intermetallic UCd_3, DOI: 10.26434/chemrxiv-2023-0nlzl
Disentangling Structural Confusion through Machine Learning: Structure Prediction and Polymorphism of Equiatomic Ternary Phases ABC, DOI: 10.1021/jacs.7b08460