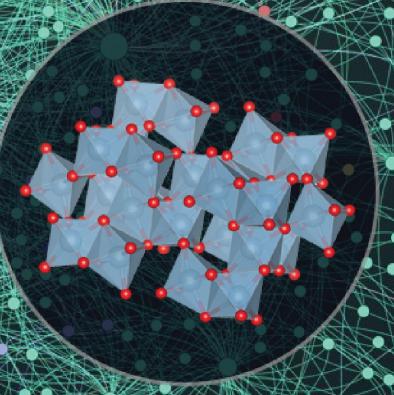
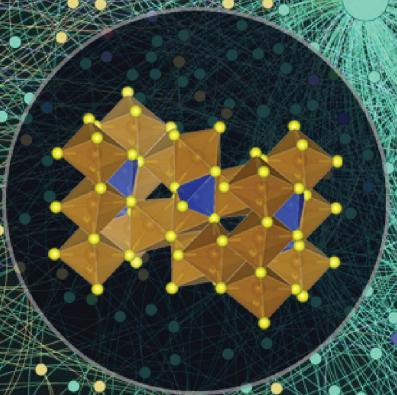
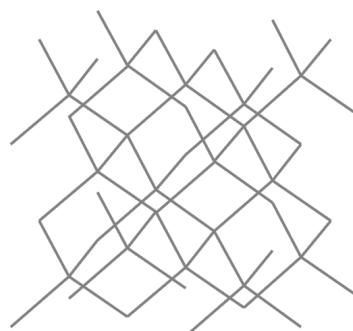


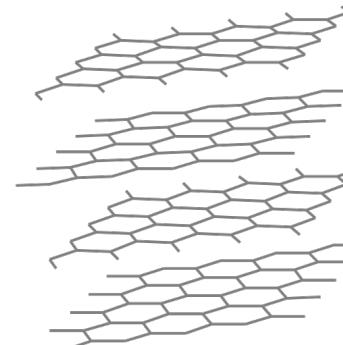
Structure-based feature vector



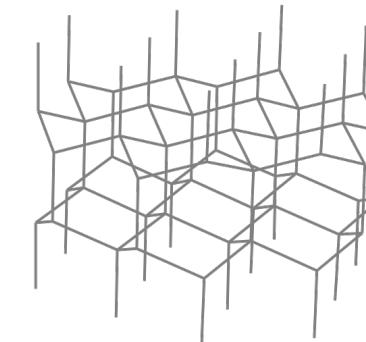
Carbon is a great example of why we can't simply rely on composition-based features



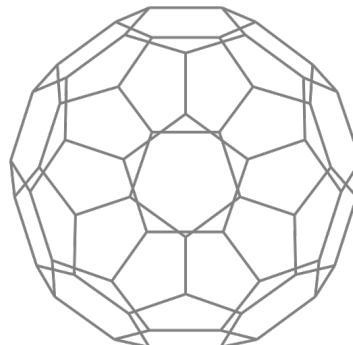
a



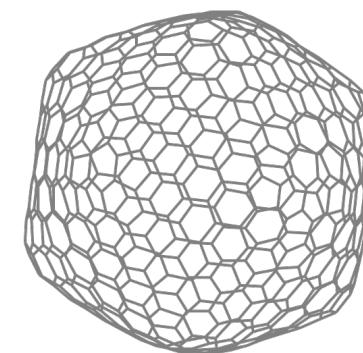
b



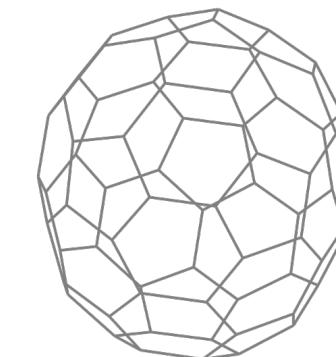
c



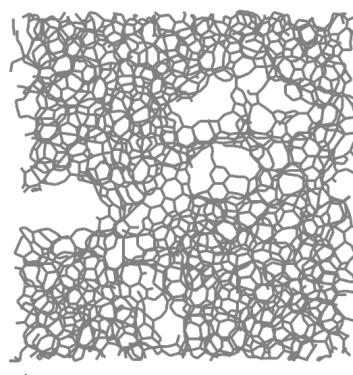
d



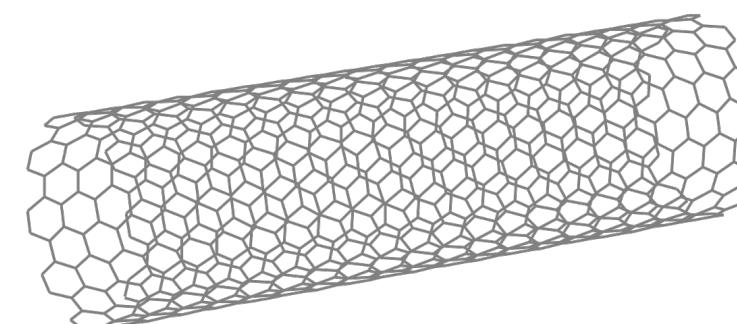
e



f



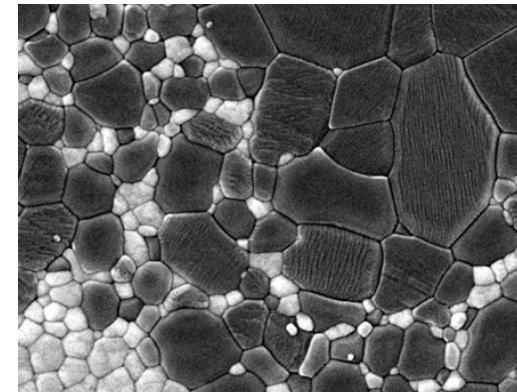
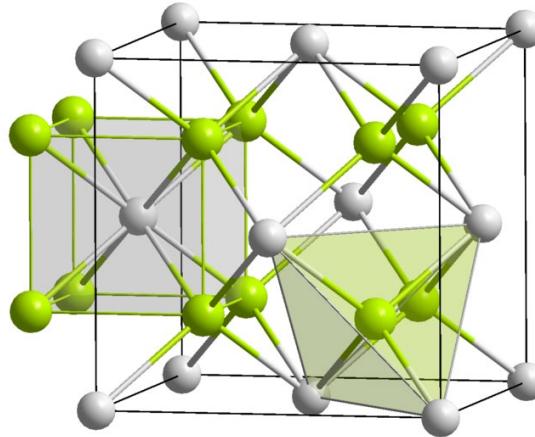
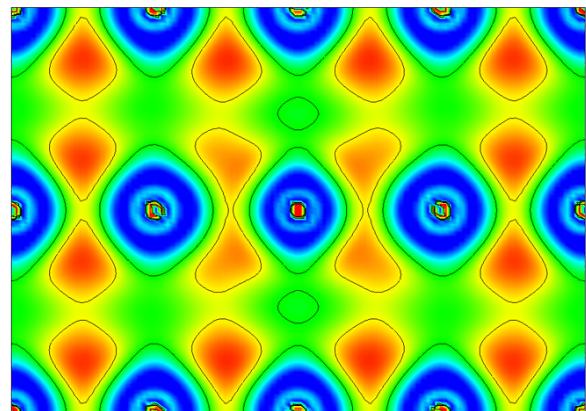
g



h

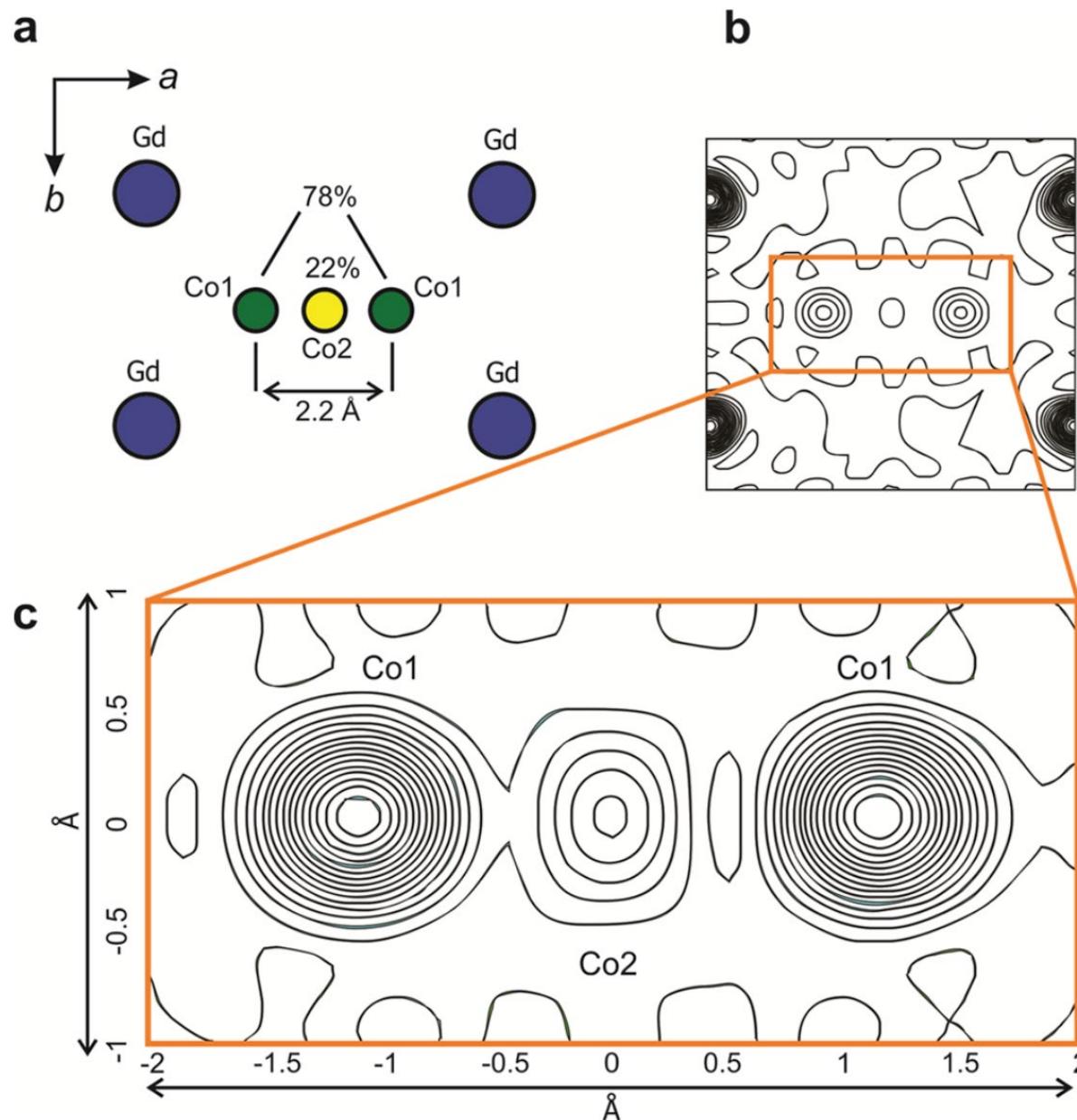
Which structure are we talking about for our structure-based feature vector?

Electronic → Atomic → Microstructure → Macroscale

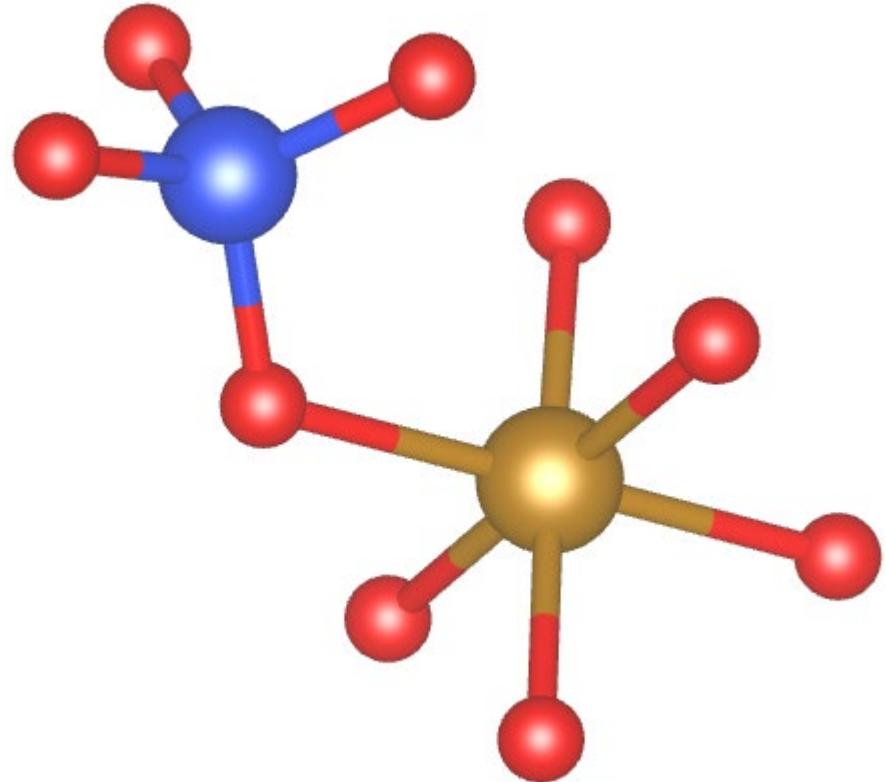


Atomic structure is described using atomic numbers

Atomic structure can also be approximated with electron density maps



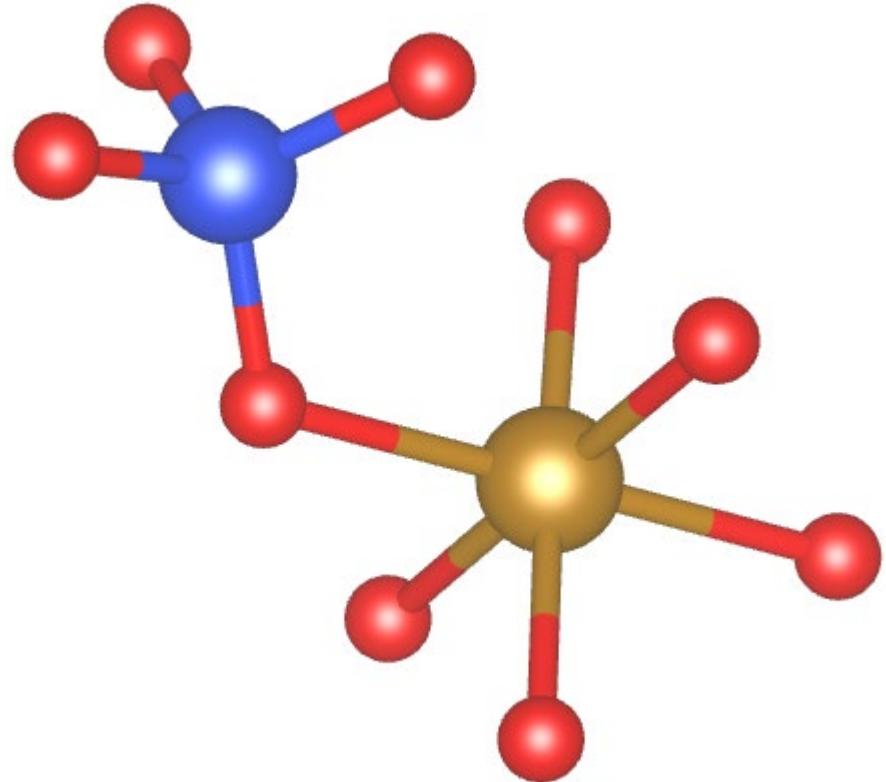
Crystal structure itself manifests unique features at different scales



Bonds

- Type
- Elements involved
- Length
- Distribution of different bonds

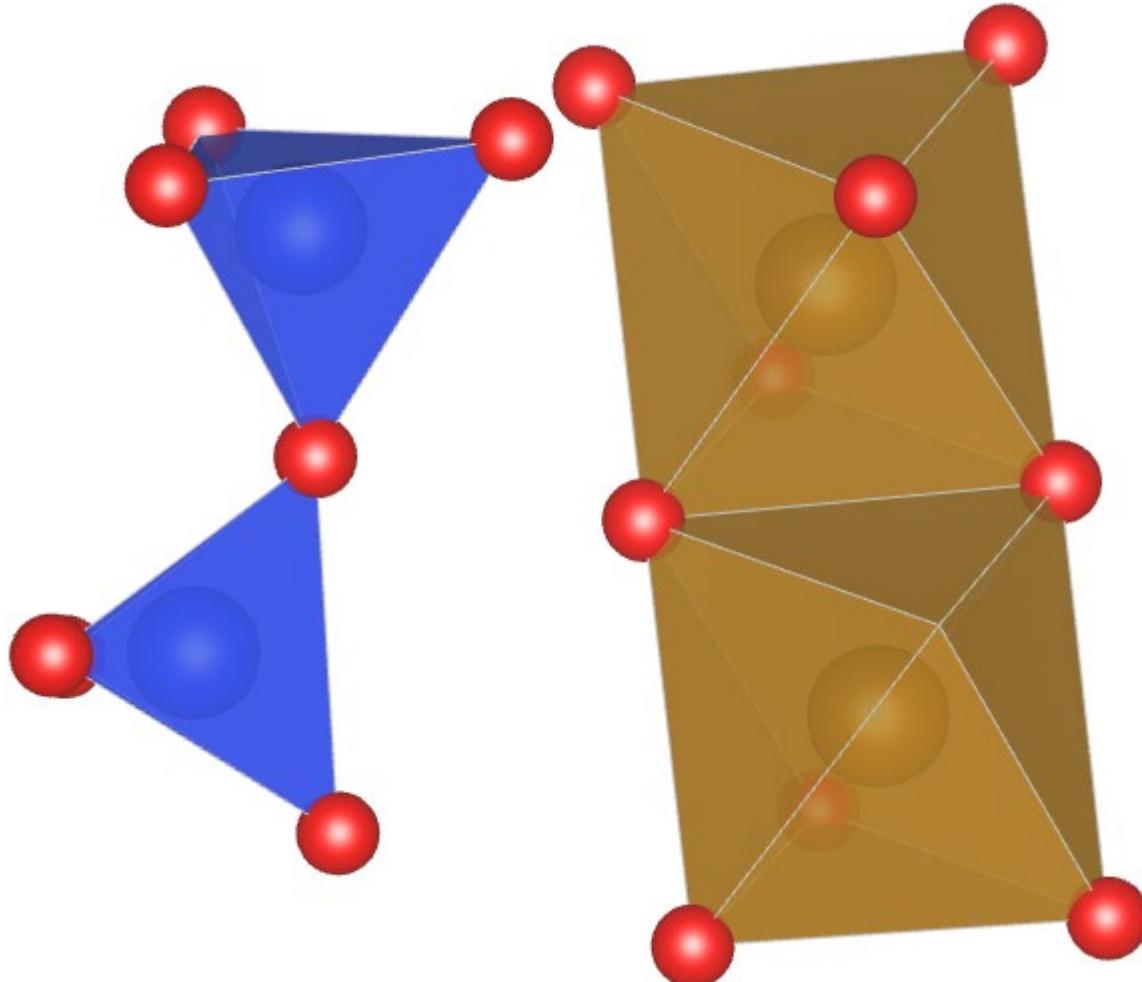
Crystal structure itself manifests unique features at different scales



Local environment

- Coordination number
- Angles
- Distortions (Jahn-Teller etc)

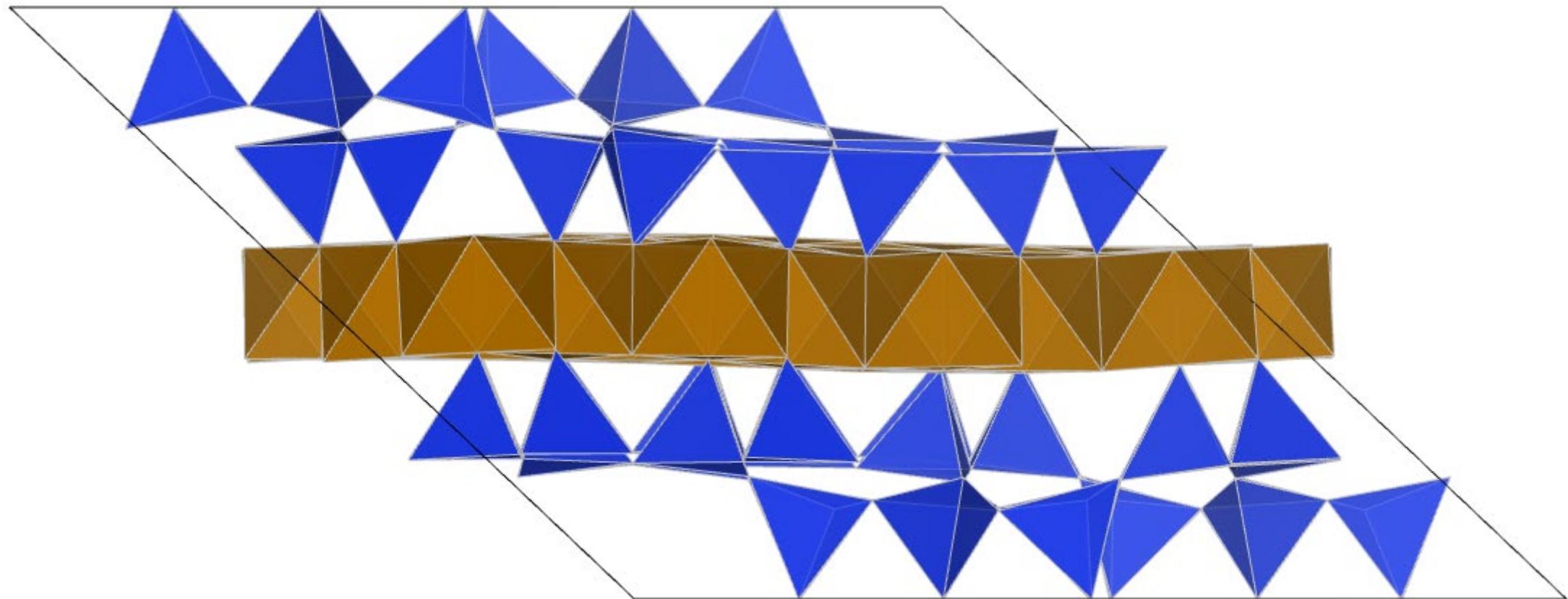
Crystal structure itself manifests unique features at different scales



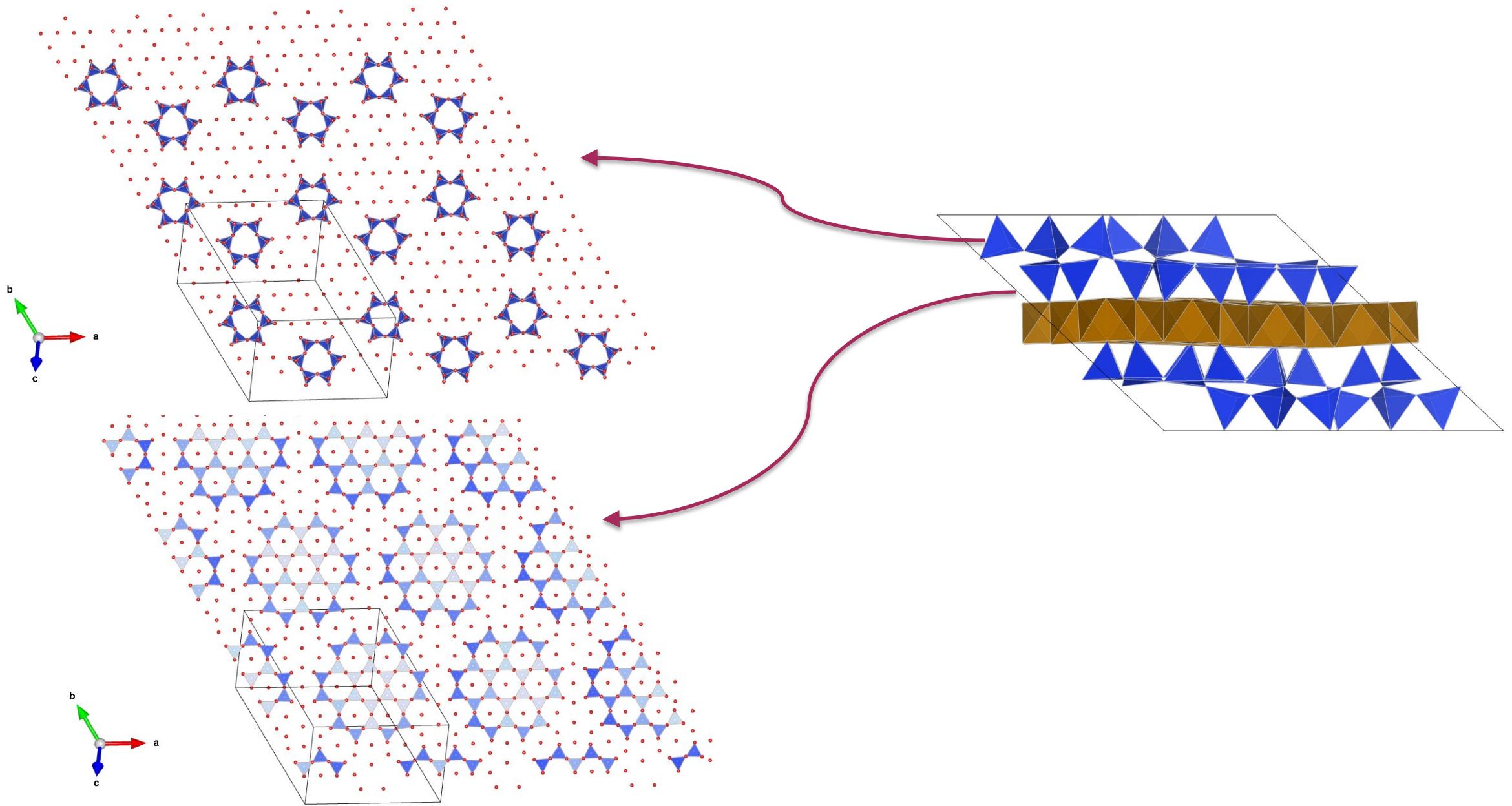
Polyhedra

- isolated vs corner vs edge vs face sharing
- multiple types
- angles
- distances

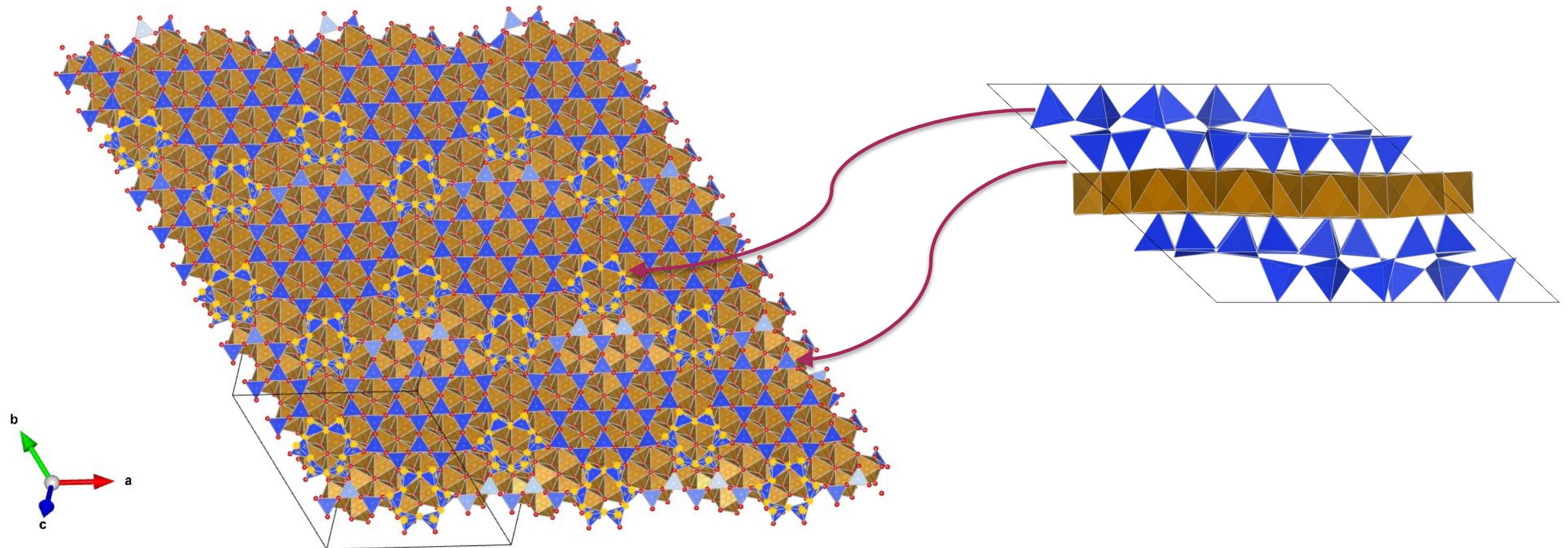
Structure can be atomic scale or unit cell scale



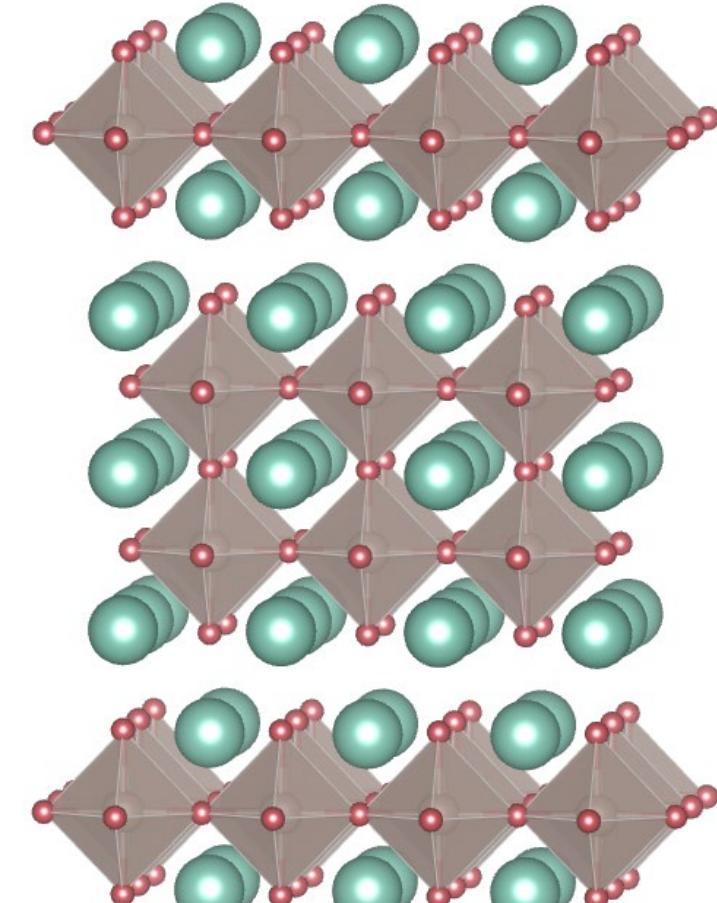
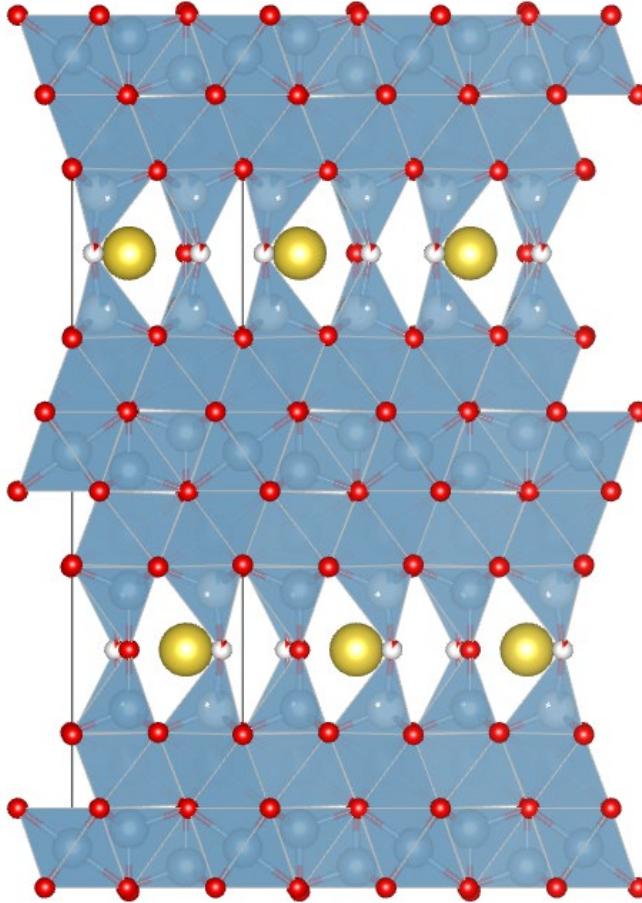
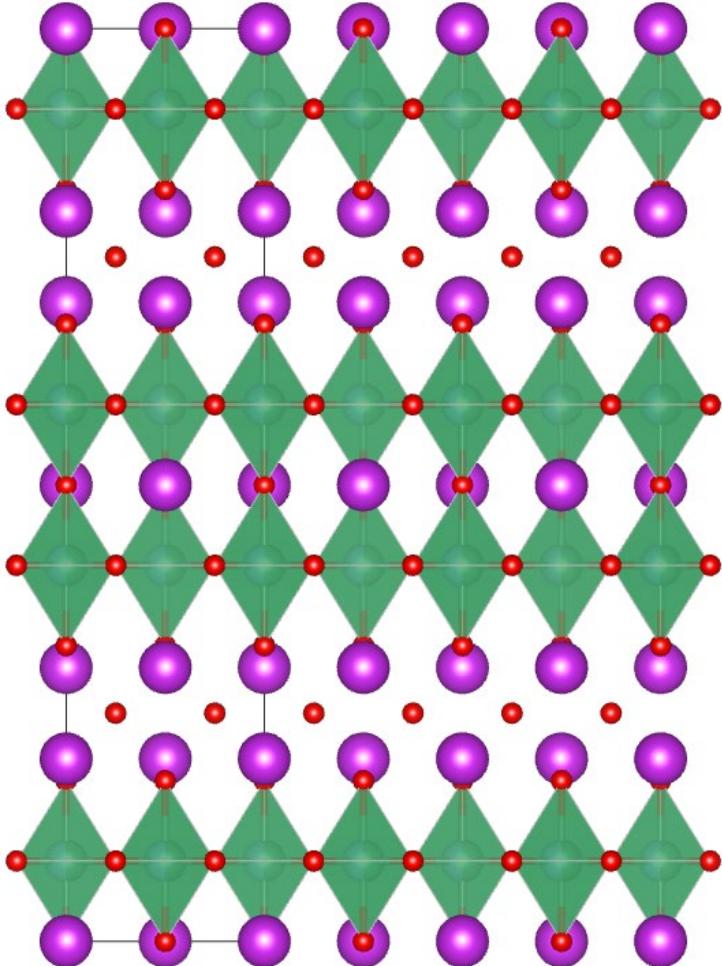
Structure can be atomic scale or unit cell scale



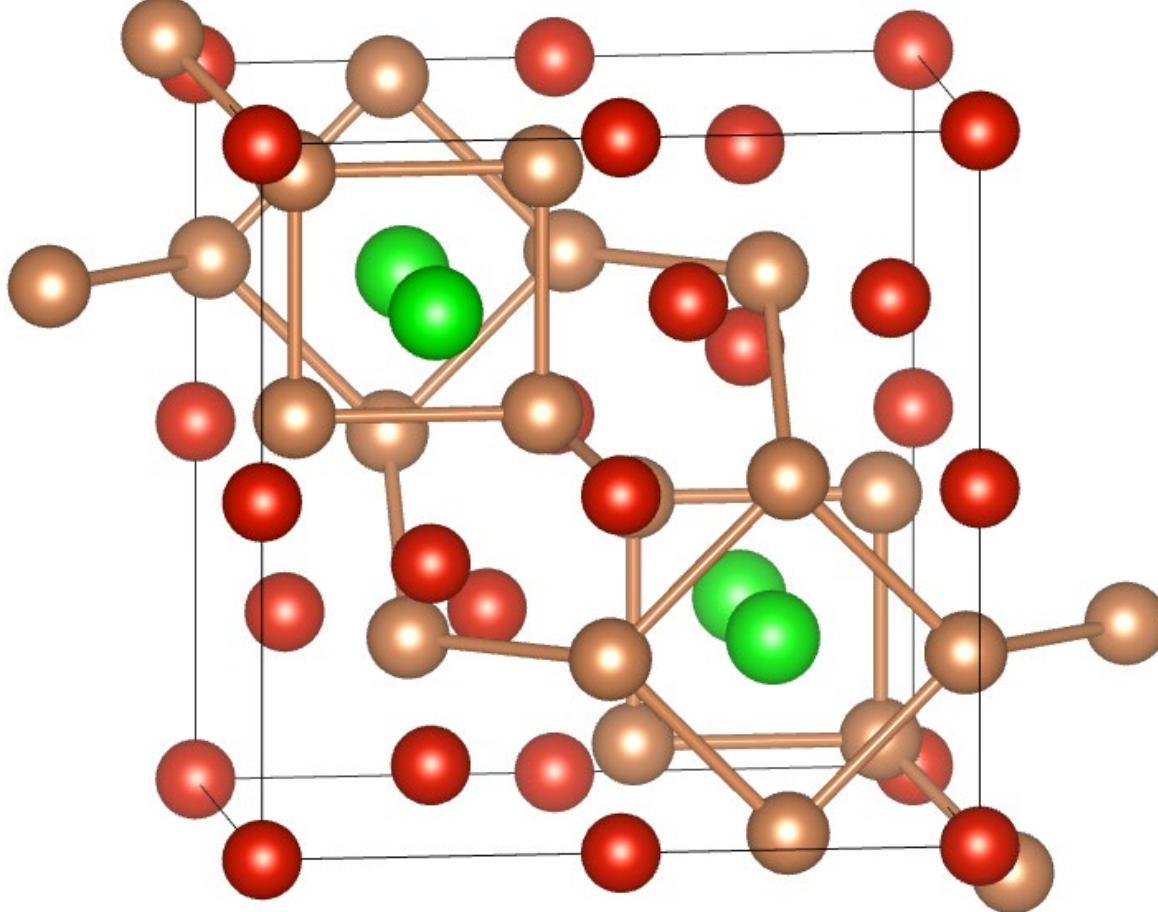
Structure can be atomic scale or unit cell scale



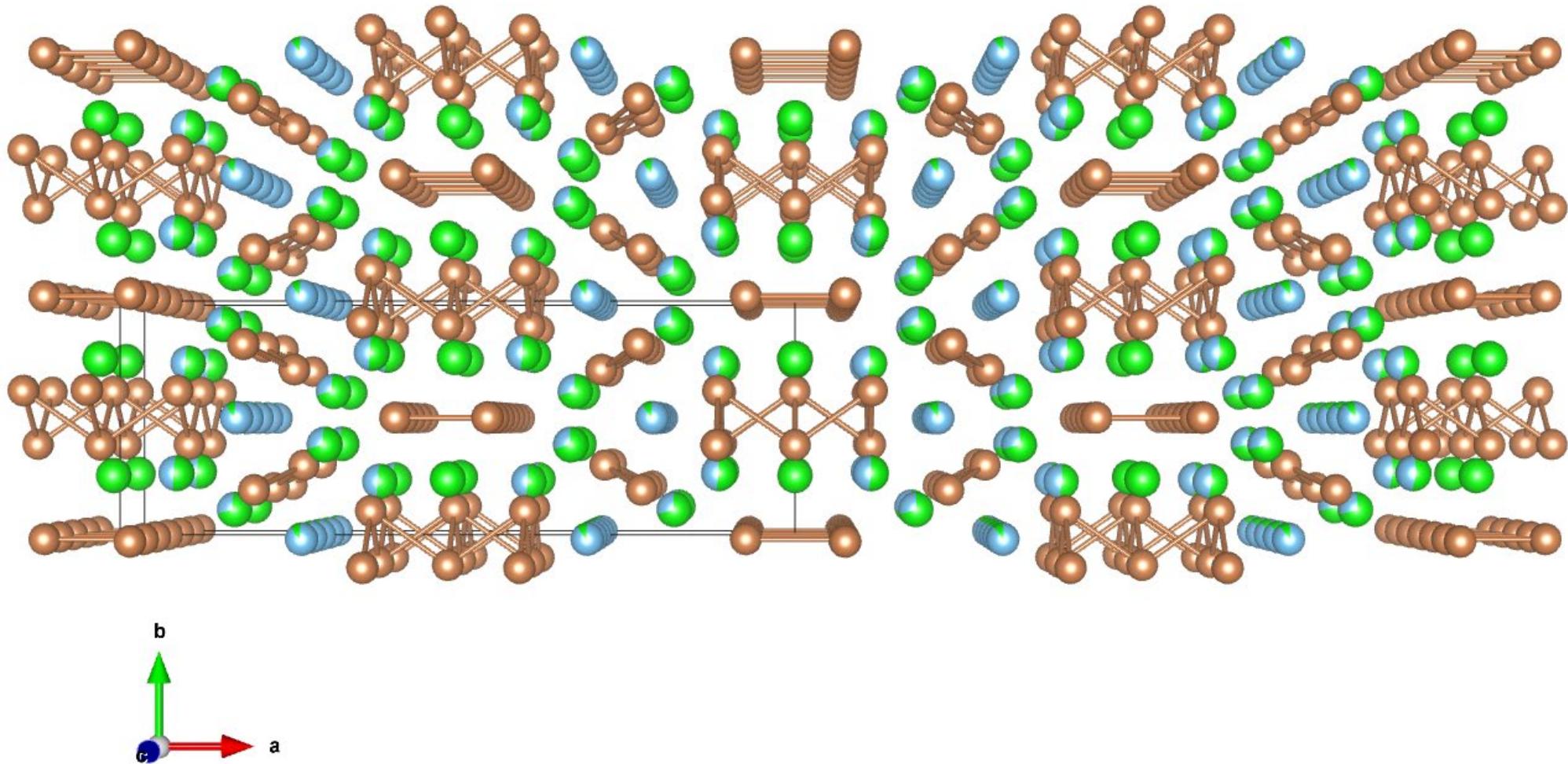
Many structures are made up of “slabs”



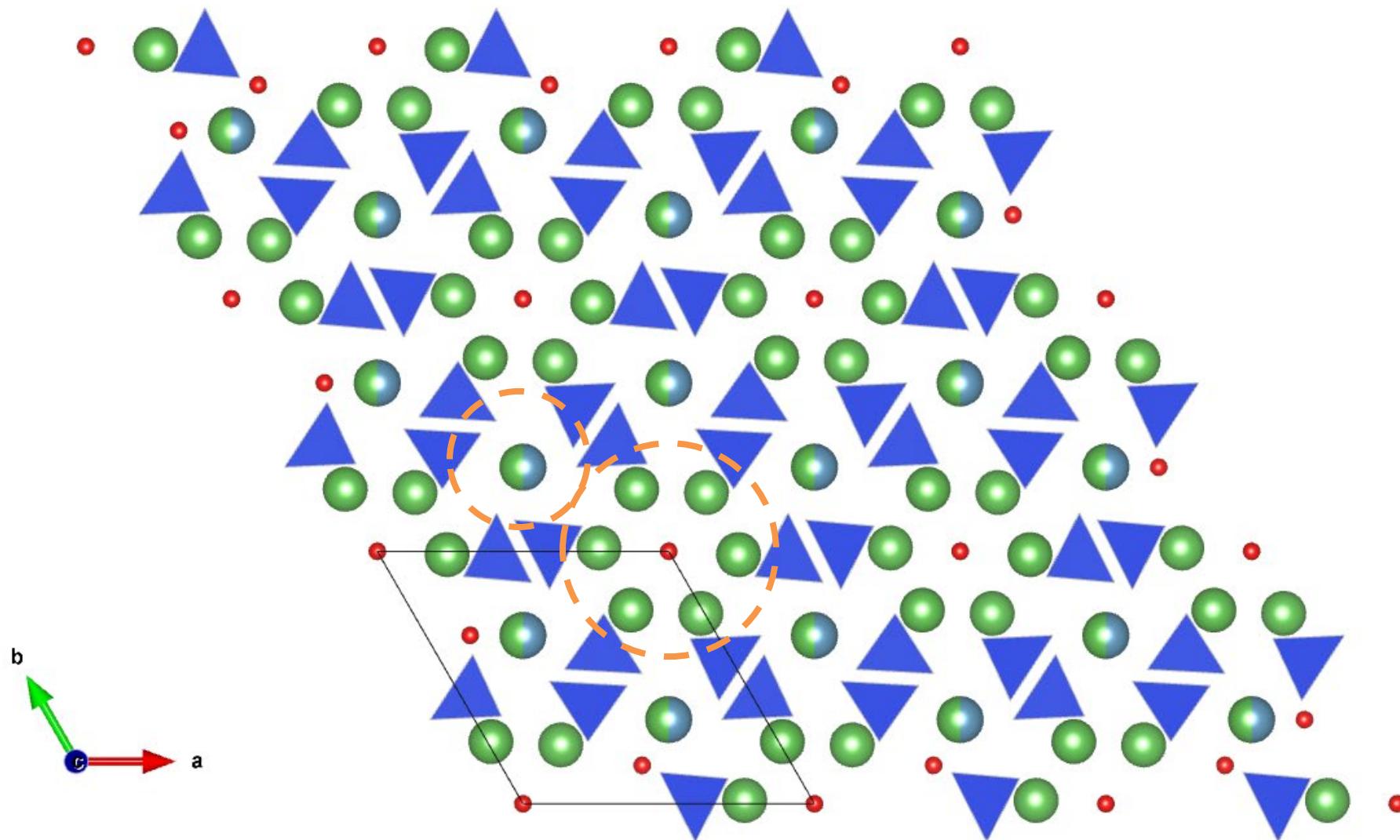
Chains and planes of atoms can define properties



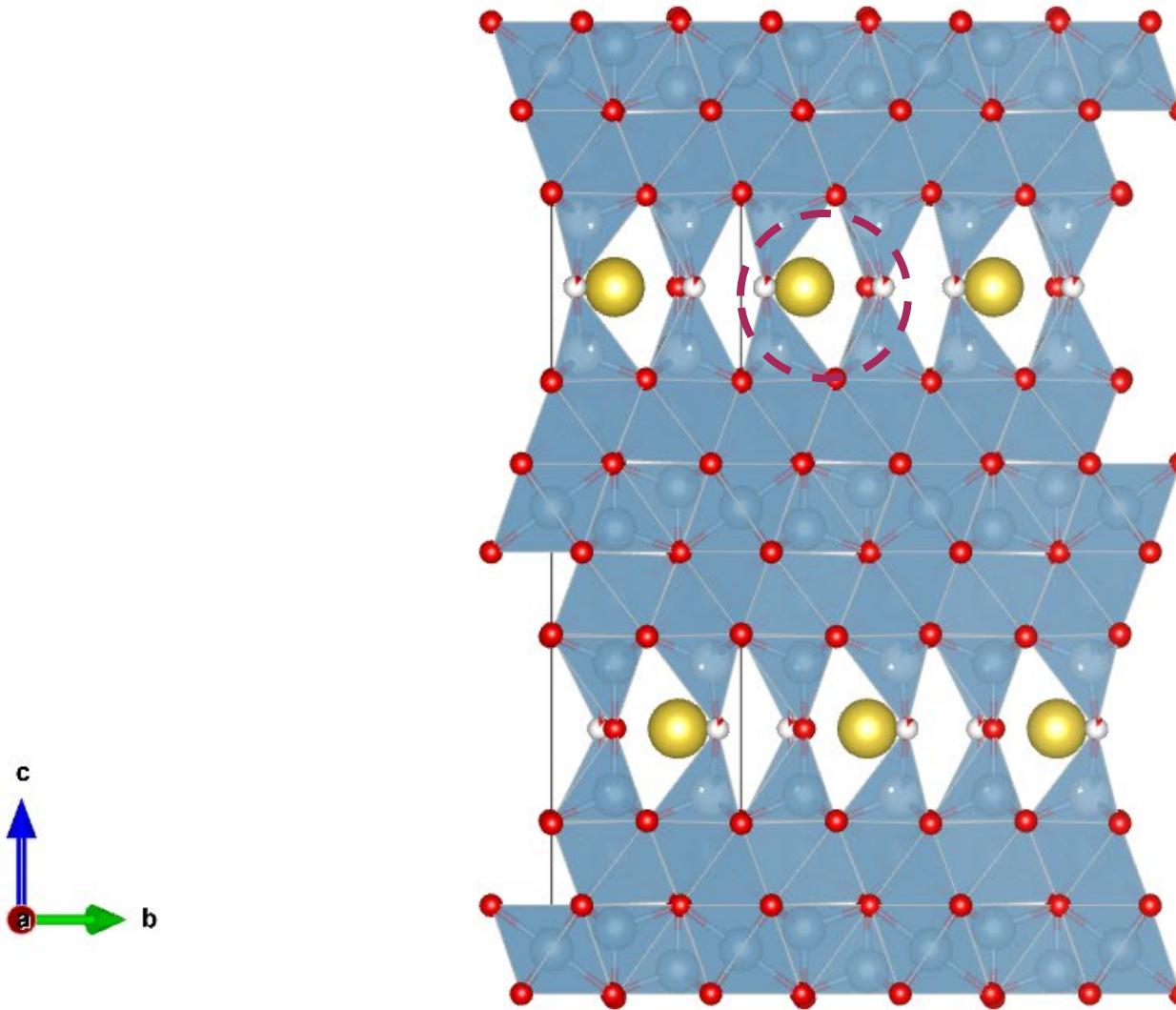
Chains and planes of atoms can define properties



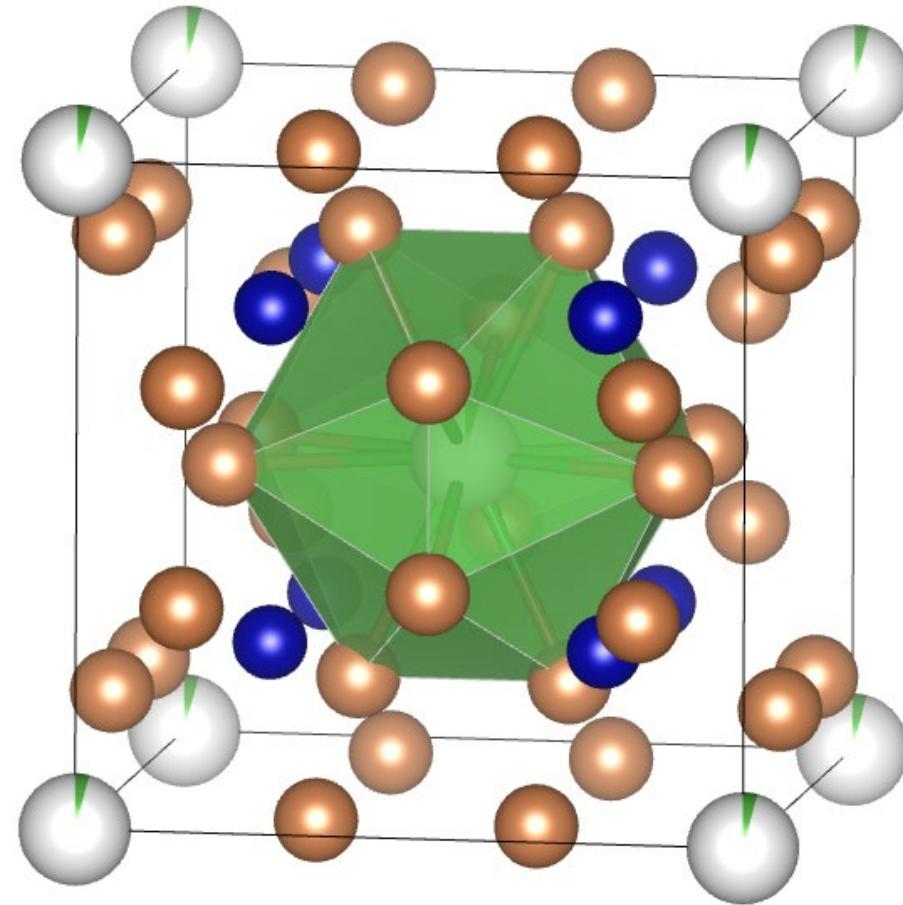
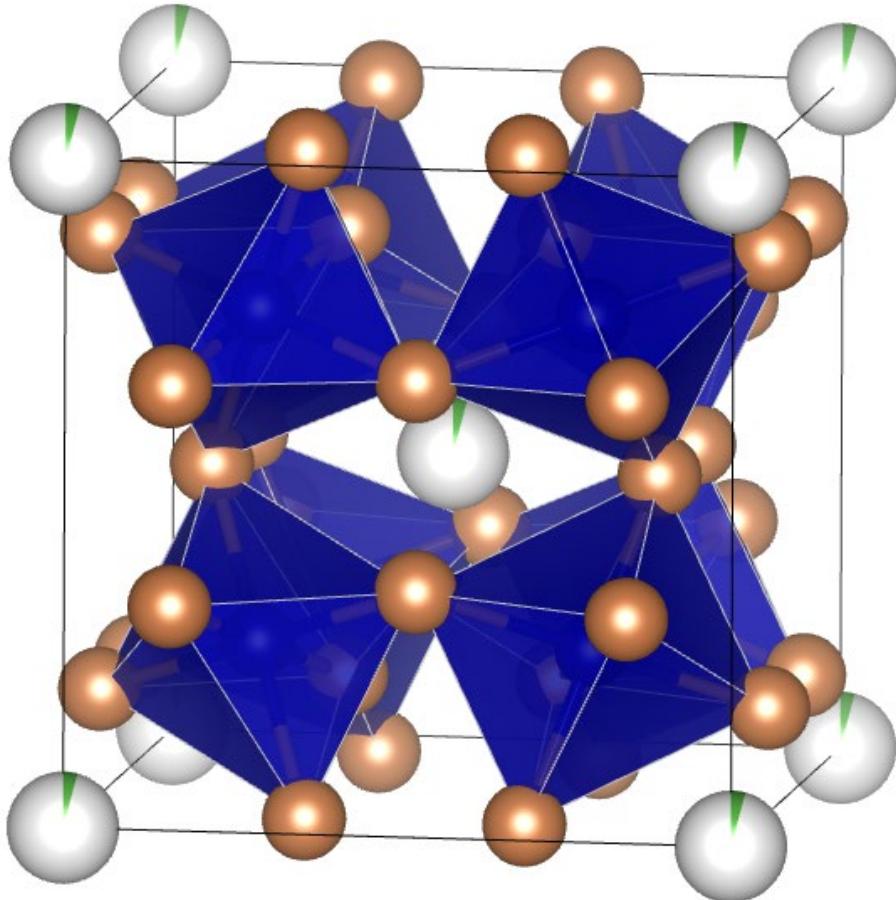
Negative space unoccupied by atoms can be defining feature



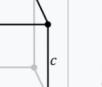
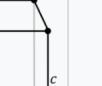
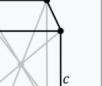
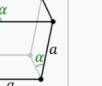
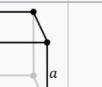
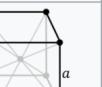
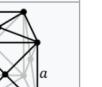
Negative space unoccupied by atoms can be defining feature



Negative space unoccupied by atoms can be defining feature



We could use symmetry parameters to define structure

Crystal family	Lattice system	Point group (Schönflies notation)	14 Bravais lattices			
			Primitive (P)	Base-centered (S)	Body-centered (I)	Face-centered (F)
Triclinic (a)	C _i					
Monoclinic (m)	C _{2h}					
Orthorhombic (o)	D _{2h}					
Tetragonal (t)	D _{4h}					
Rhombohedral	D _{3d}					
	Hexagonal (h)					
Cubic (c)	O _h					

7 crystal systems

14 bravais lattices

Lattice parameters (a,b,c)

Angles (α, β, γ)

Centering (primitive, base, body, face)

32 Schoenflies point groups

We could use symmetry parameters to define structure

230 The Space Group List Project



by Frank Hoffmann



More information at
crystalsymmetry.wordpress.com

Crystal structures drawn with VESTA
K. Momma and F. Izumi, J. Appl. Crystalogr. 2011, 44, 1272-1276

230 space groups for unique symmetries
Symmetry is applied to crystal basis

How do we encode all of these features?

Structural feature	Specific structure
Layered	Yes
Channels	No
RS blocks	Yes
Octahedra	Yes
Tetrahedra	No
...	
Corner-sharing	Yes
Edge-sharing	No

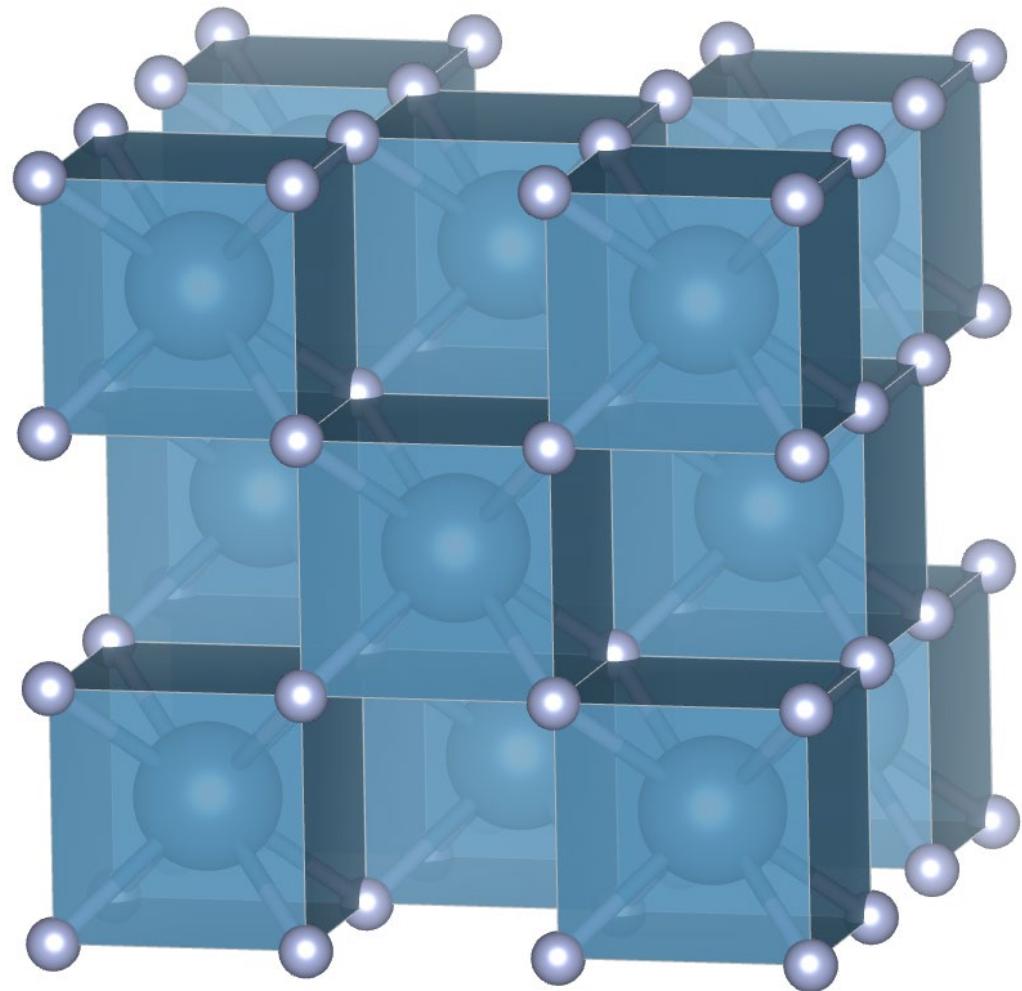
Maybe we could one-hot encode them?

How do we encode all of these features?

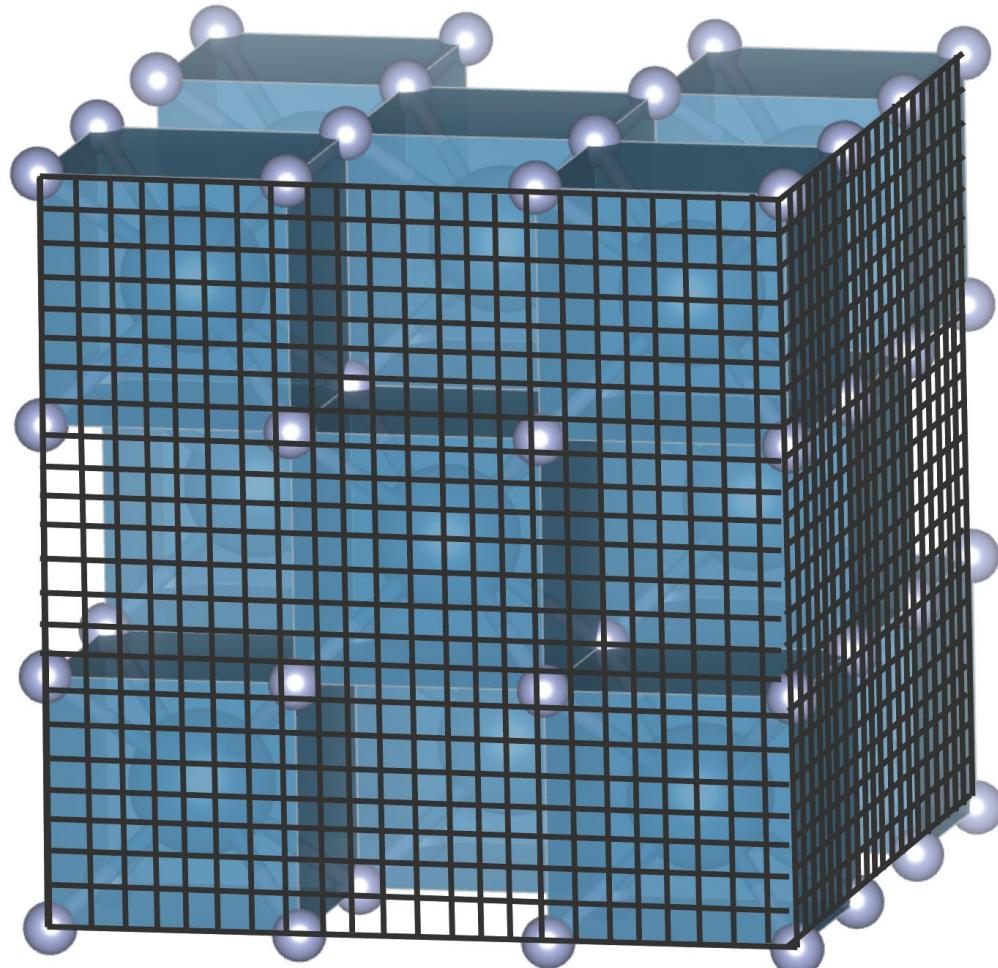
Structural feature	Ruddlesden-Popper	Aurivillius	Spinel	Rock Salt
Layered	1	1	0	0
Channels	0	0	1	0
RS blocks	1	0	0	1
Octahedra	1	1	1	1
Tetrahedra	0	0	1	0
...				
Corner-sharing	1	1	1	1
Edge-sharing	0	0	1	0

Maybe we could one-hot encode them?

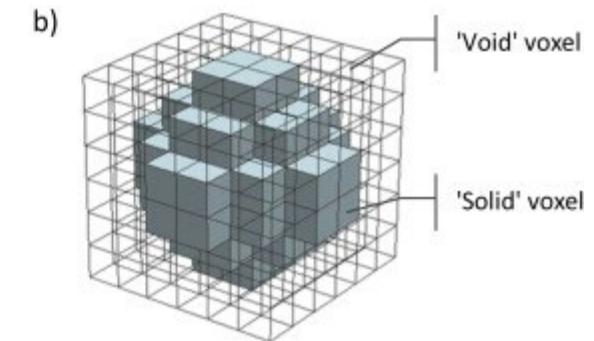
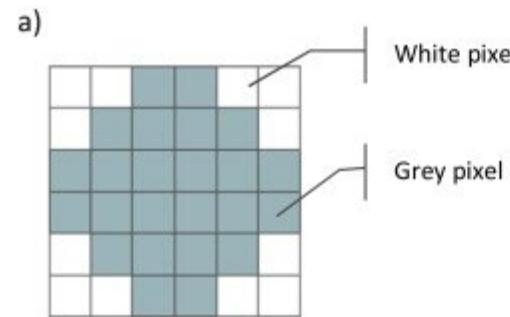
We could also represent the crystal as voxels occupied by different atoms



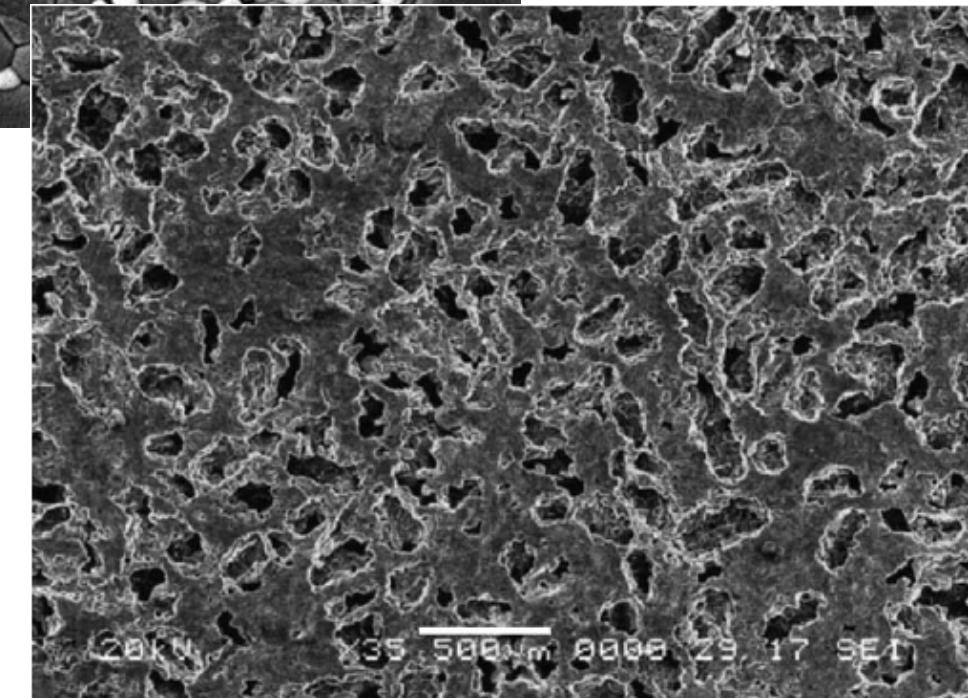
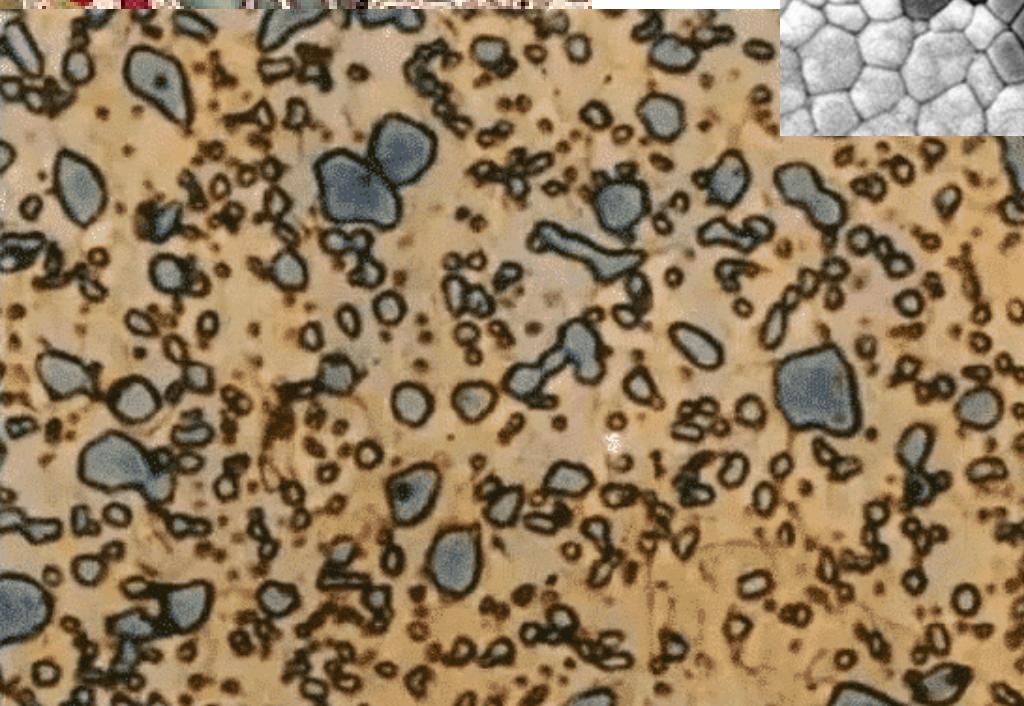
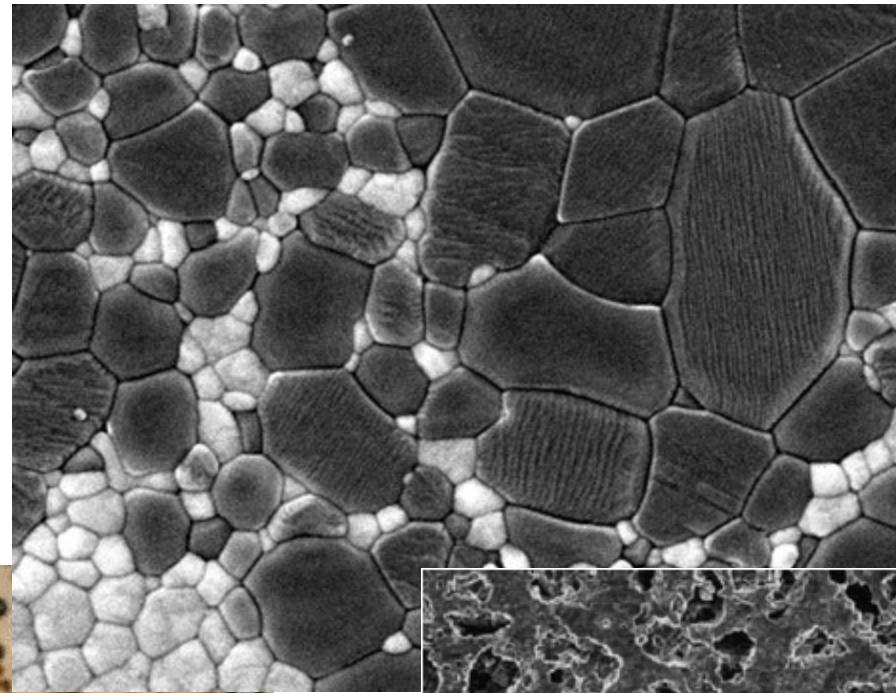
We could also represent the crystal as voxels occupied by different atoms



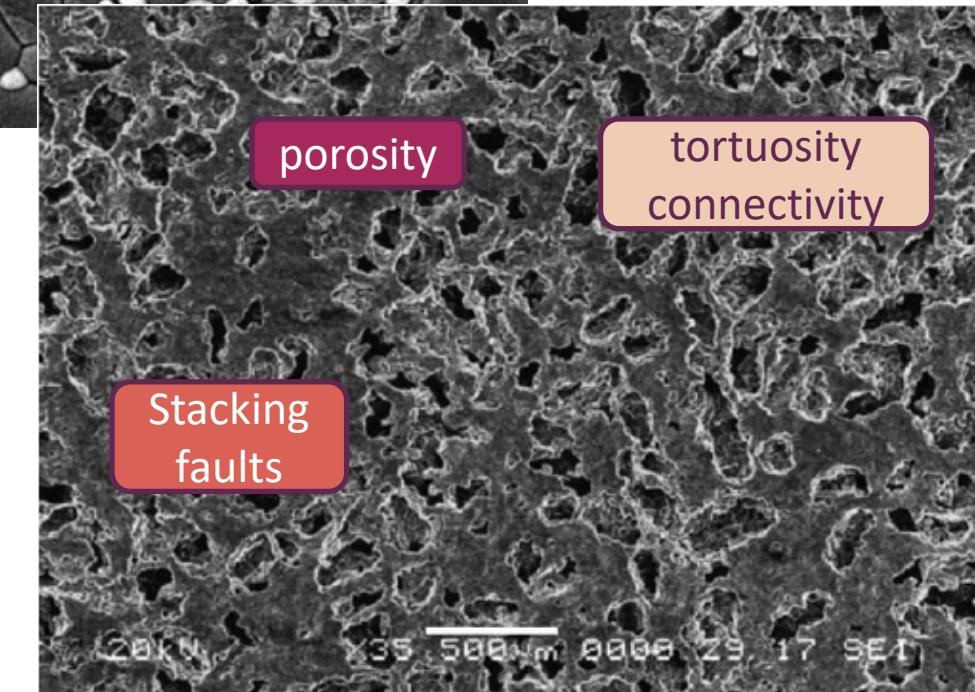
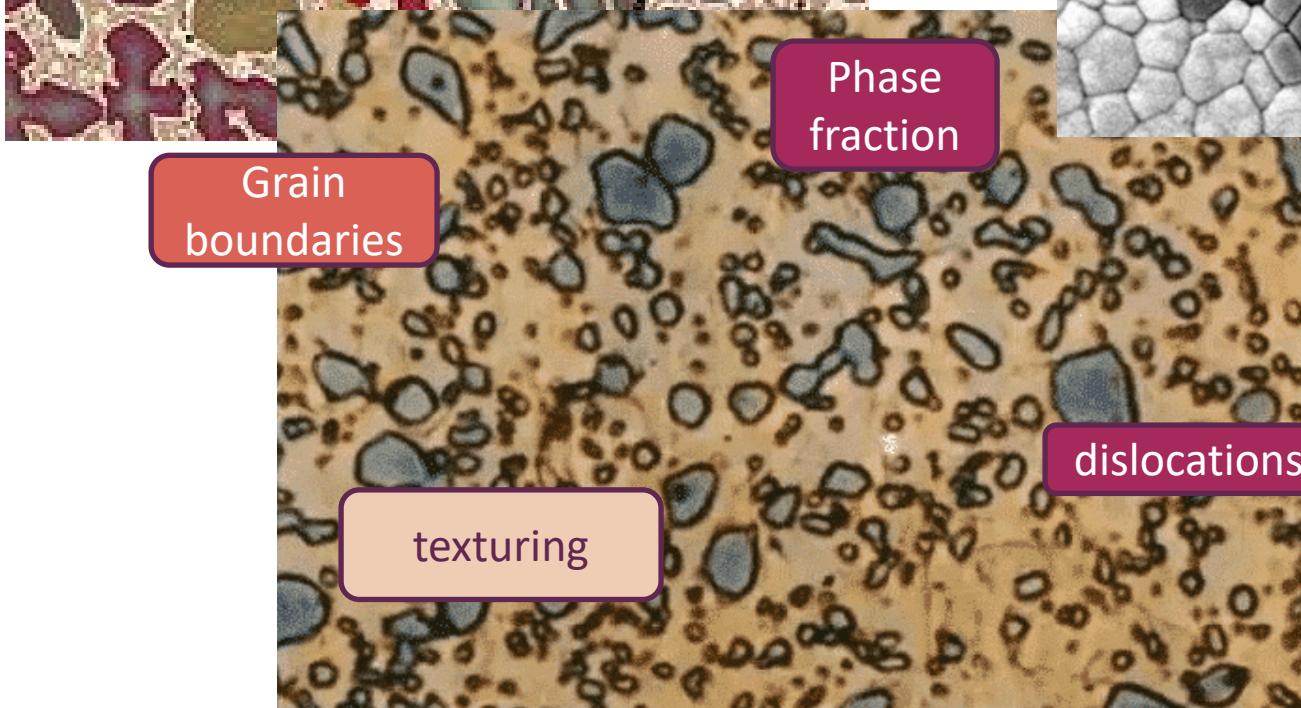
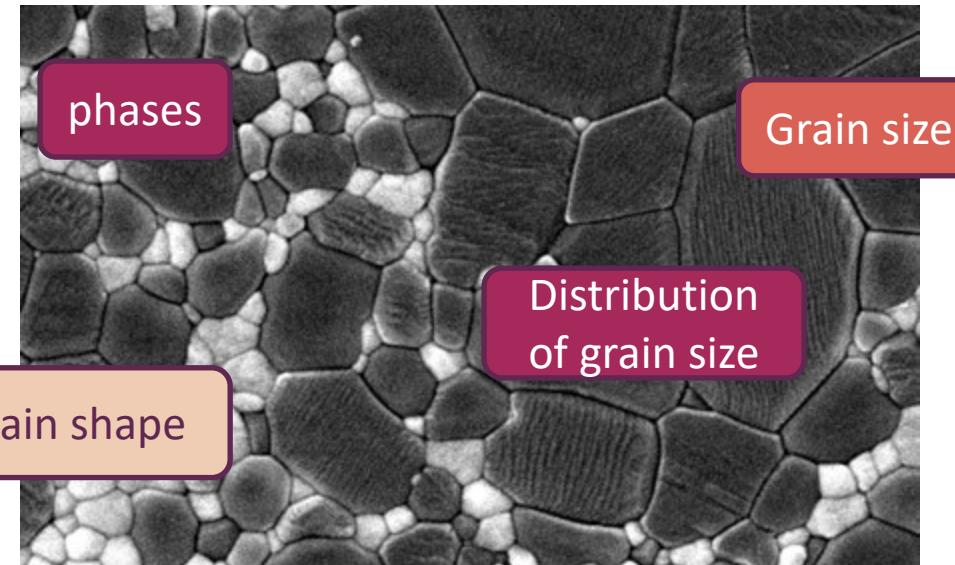
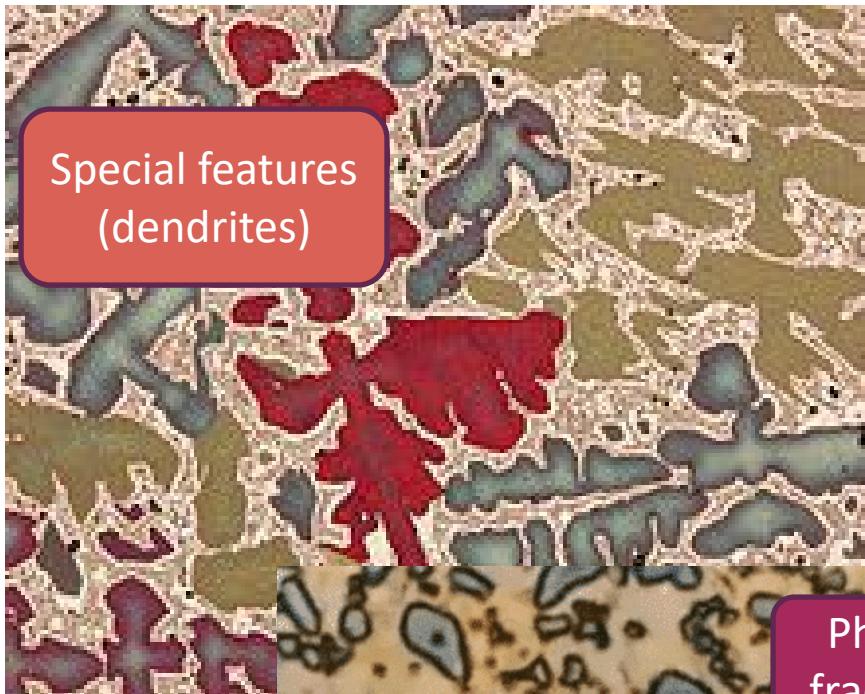
Do we put a point at atom center, or do we try to account for atom size?
How do we encode different atom chemistries?



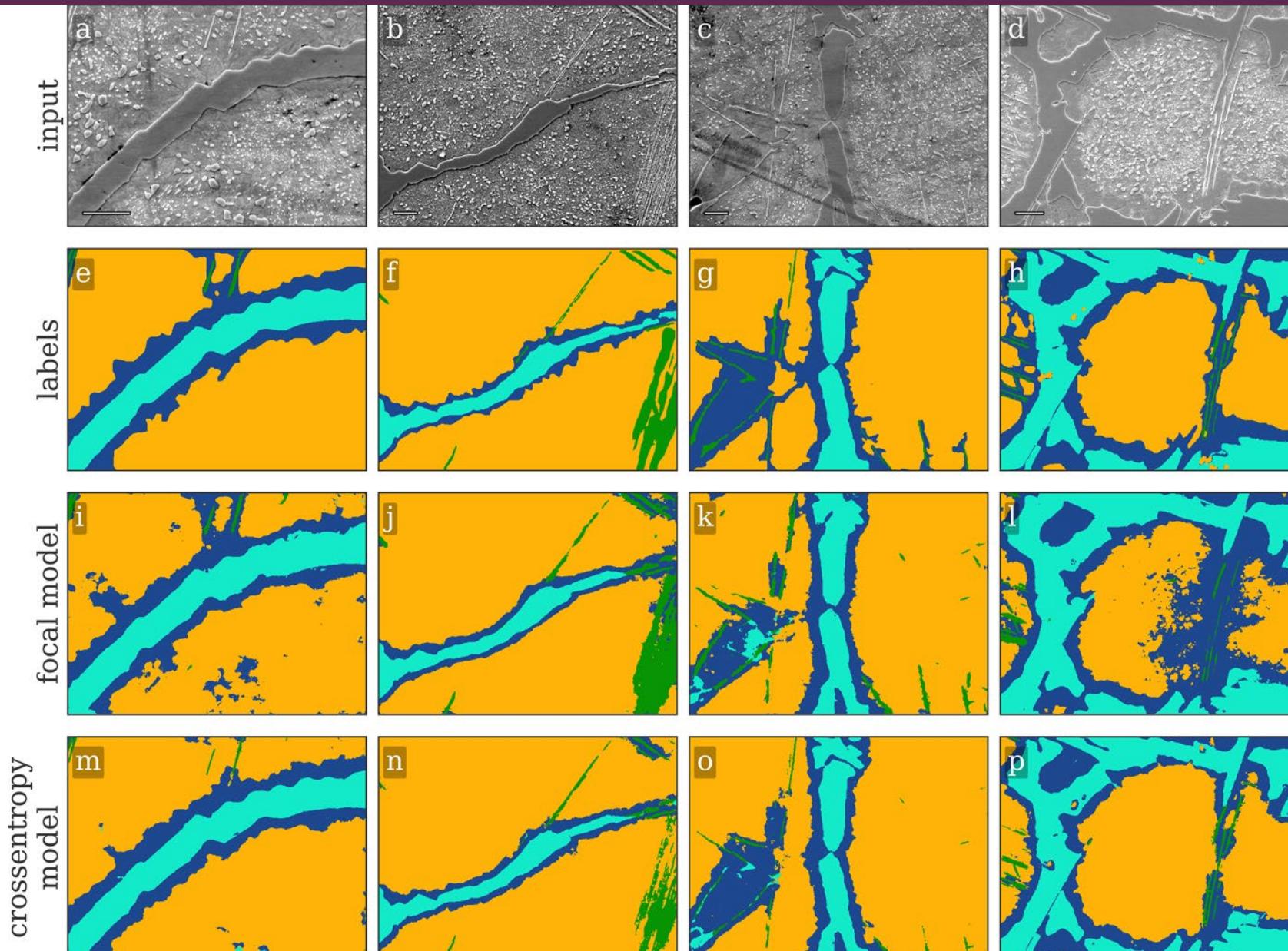
What information is present at the microstructure scale?



What information is present at the microstructure scale?

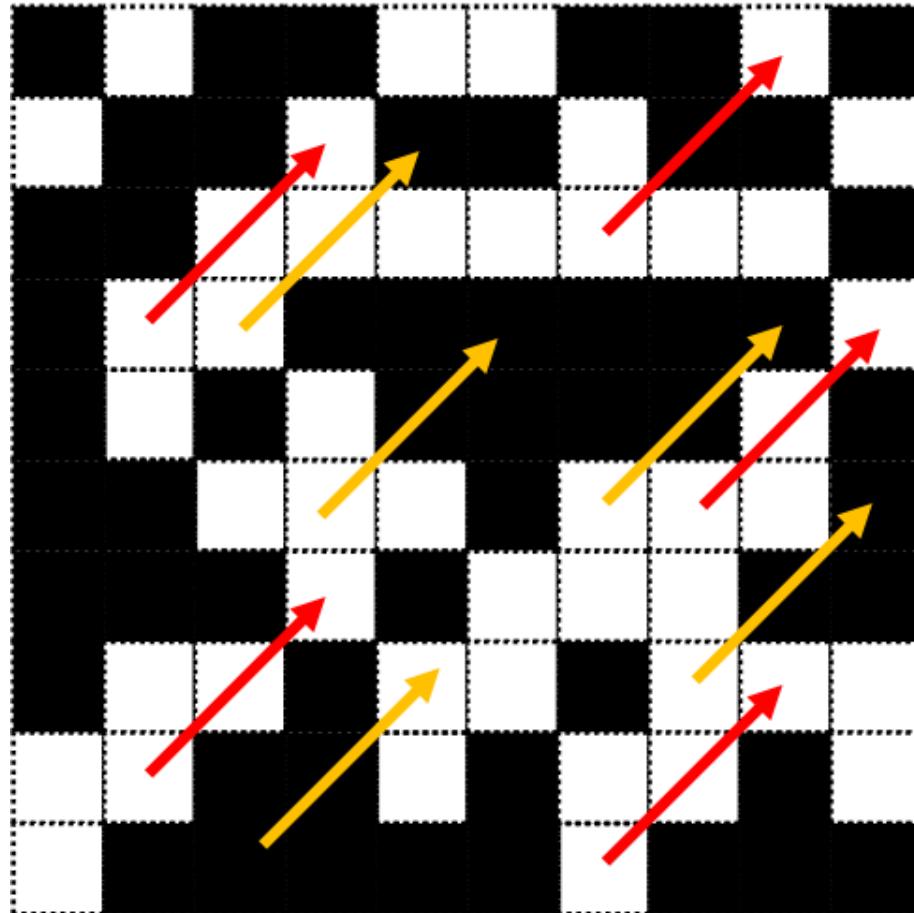


How do we identify and featurize these microstructural features?

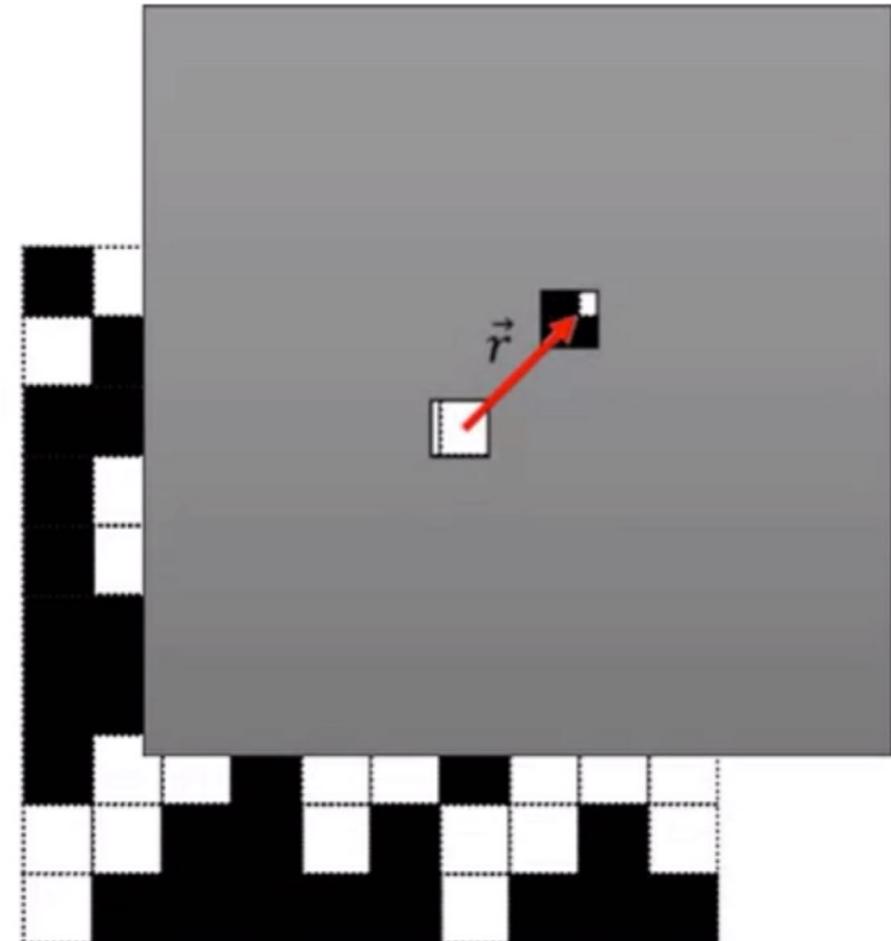


Two-point statistics can be a useful tool for quantifying microstructure

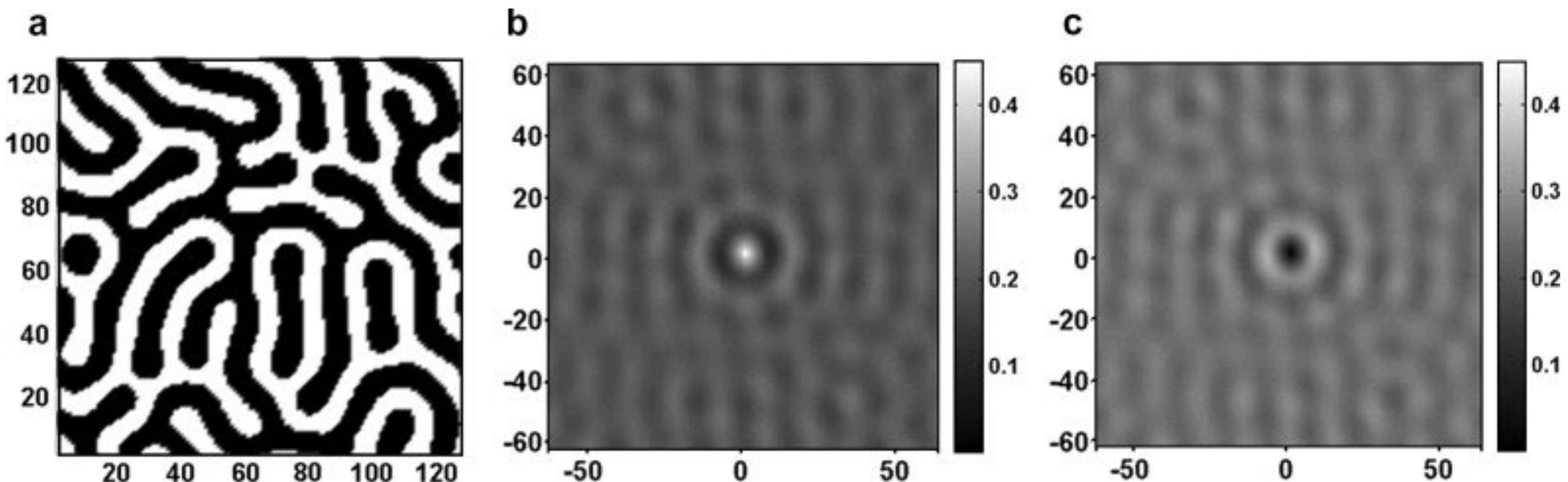
$$f_r^{hh'} = \frac{\# \text{Trials Successful}}{\# \text{Trials Attempted}}$$



$$P(A^h \cap A_{@\vec{r}}^{h'} | \vec{r}) = \frac{P(A^h \cap A_{@\vec{r}}^{h'} \cap \vec{r})}{P(\vec{r})}$$



Two-point statistics can be a useful tool for quantifying microstructure



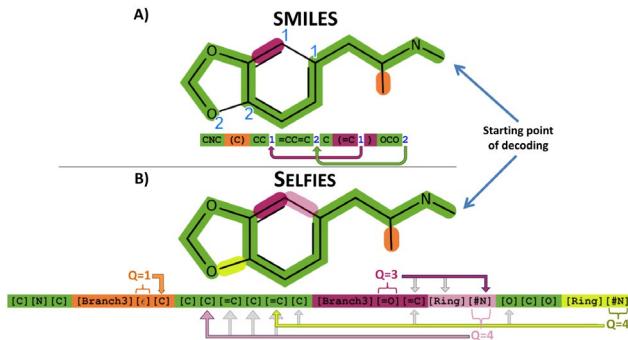
Fullwood, David T., Stephen R. Niezgoda, and Surya R. Kalidindi. *Acta Materialia* 56.5 (2008): 942-948.

A number of different tools exist for structural featurization

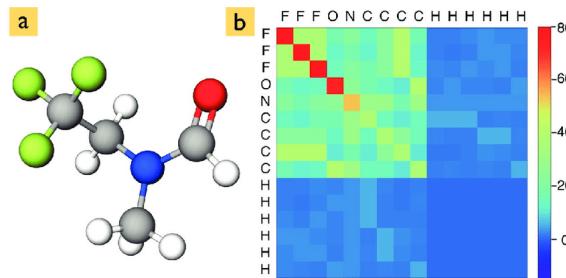
Microstructure segmentation/analysis



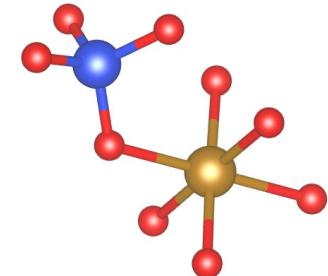
String representations of molecules



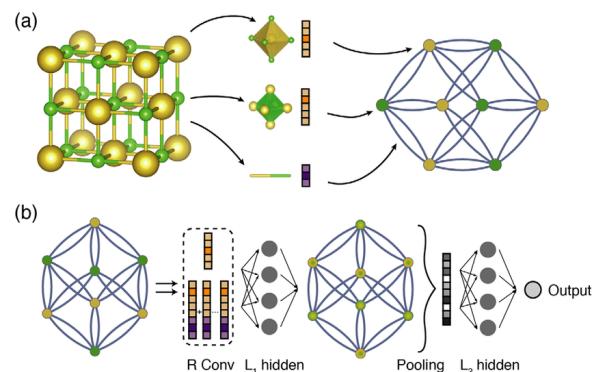
Coulomb matrix



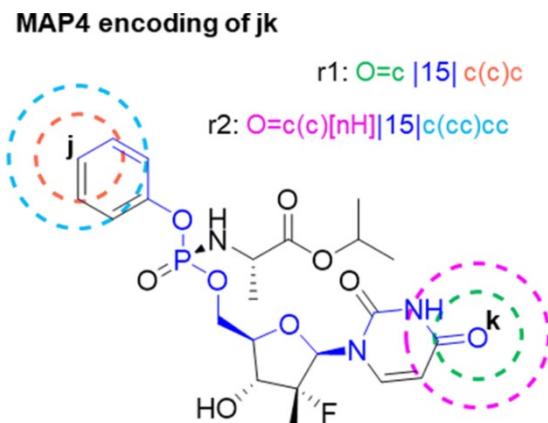
3D coordinates



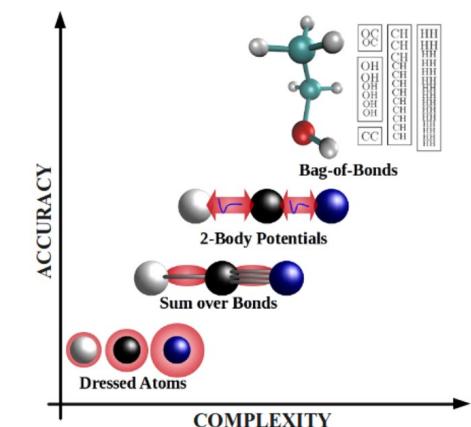
Crystal graphs



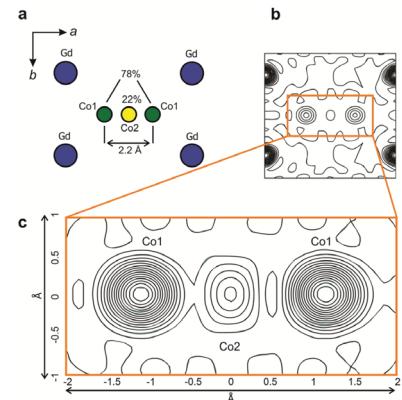
Fingerprints



Bag of Bonds / Fragments



Electron density



Crystal Structure graphs

