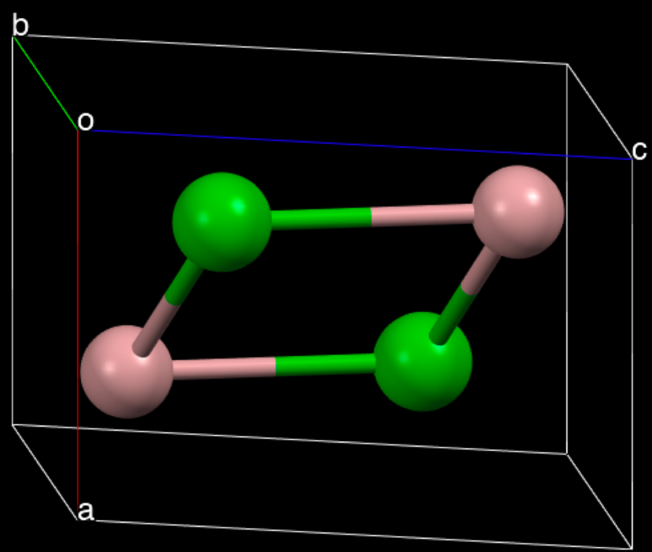
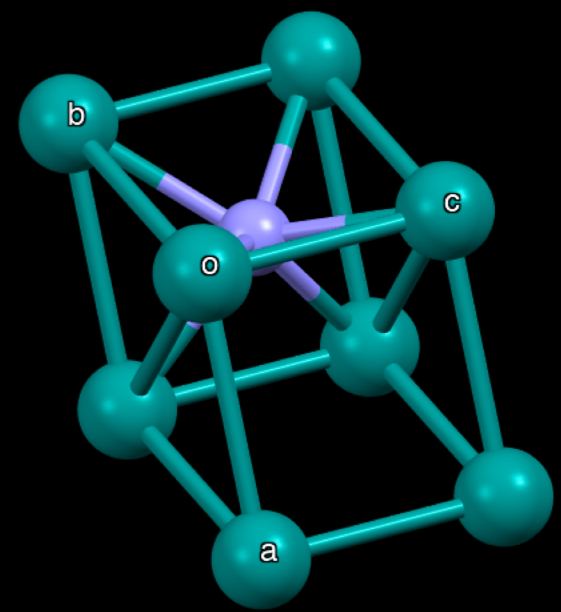
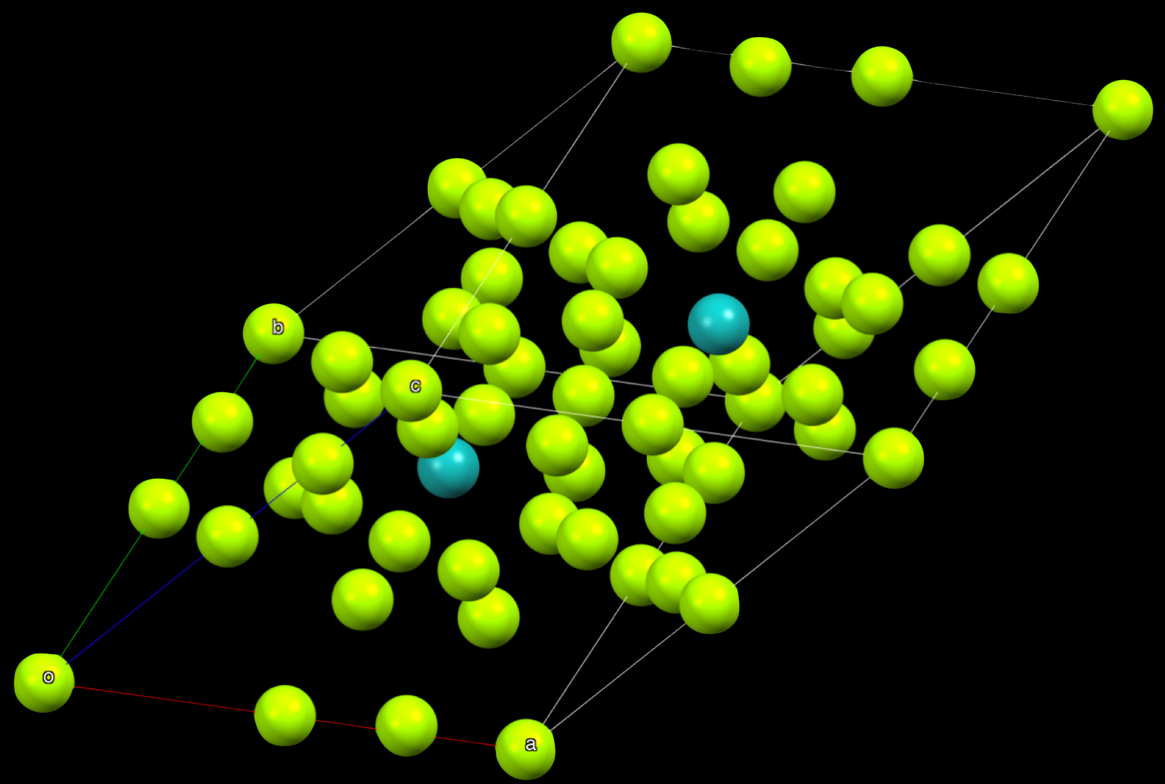
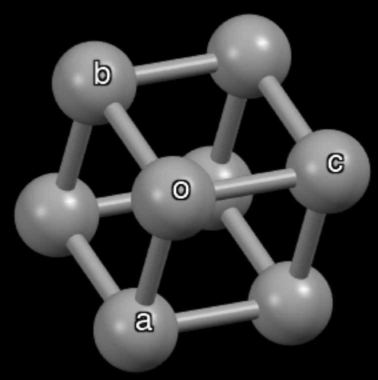
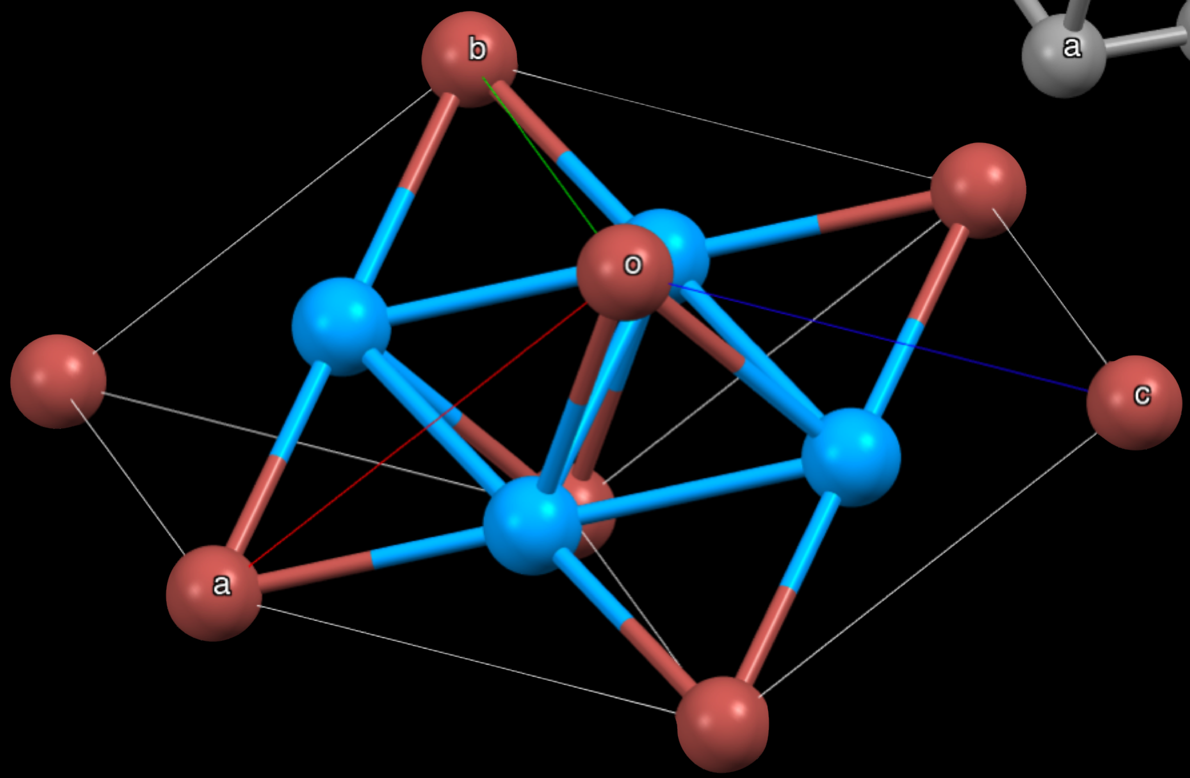
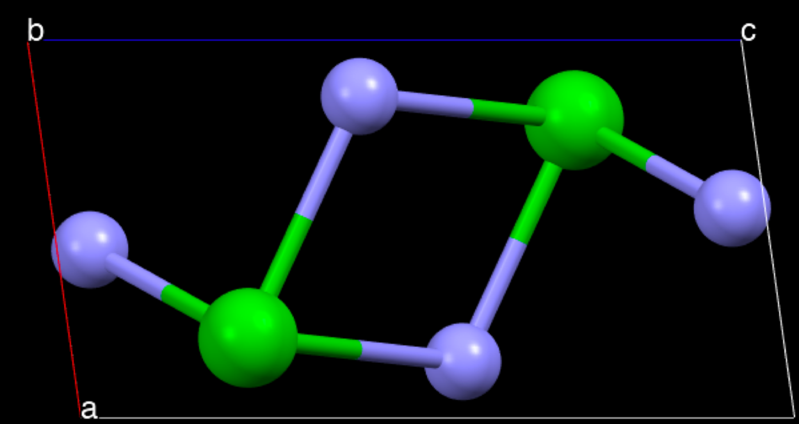

b
o
c
a

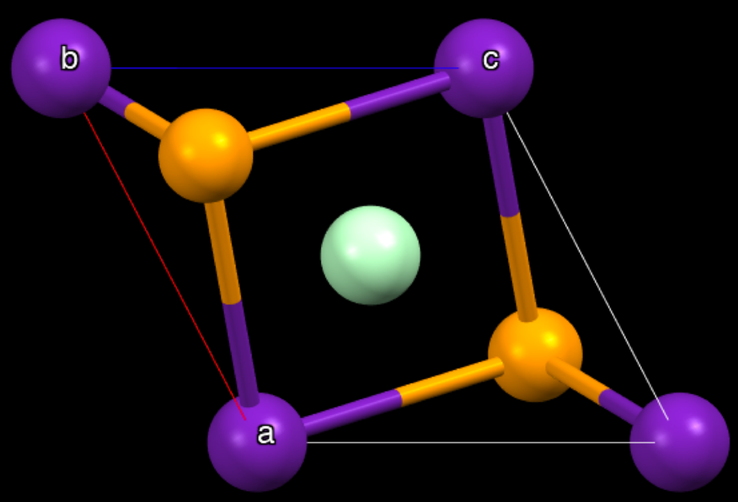

b
o
c
a


b
c
o
a


b
o
c
a


b
o
c
a


b
c
a


b
c
a

Encoding and Decoding 3-D Crystals

Jordan Hoffmann • Louis Maestrati • Yoshihide Sawada
Jian Tang • Jean Michel Sellier • Yoshua Bengio

Applications

Specific compounds of interest:

- batteries
- solar cells

Applications

Specific compounds of interest:

- batteries
- solar cells

Targeted

(Now) Works well for drug discovery.

Applications

Specific compounds of interest:

- batteries
- solar cells

Targeted

(Now) Works well for drug discovery.

Random structure searching

Abstract

It is essential to know the arrangement of the atoms in a material in order to compute and understand its properties. Searching for stable structures of materials using first-principles electronic structure methods, such as density-functional-theory (DFT), is a rapidly growing field. Here we describe our simple, elegant and powerful approach to searching for structures with DFT, which we call *ab initio* random structure searching (AIRSS). Applications to discovering the structures of solids, point defects, surfaces, and clusters are reviewed. New results for iron clusters on graphene, silicon clusters, polymeric nitrogen, hydrogen-rich lithium hydrides, and boron are presented.

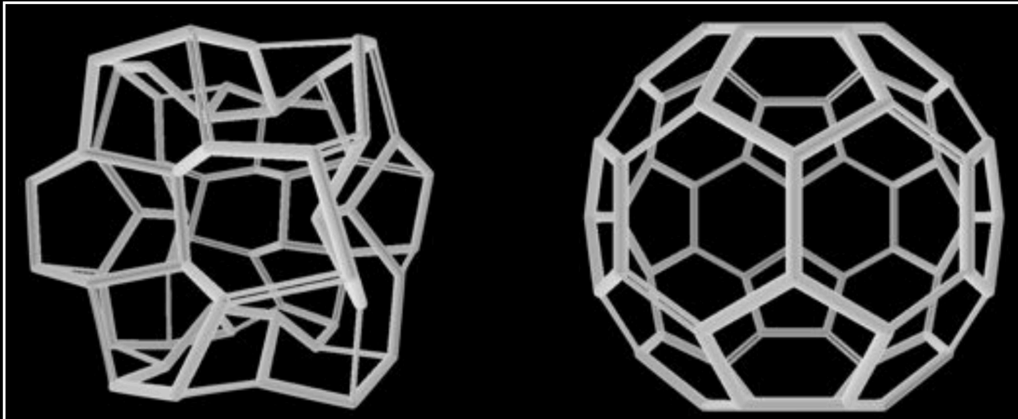


Figure 3. Left: A structure built by placing carbon atoms randomly within a small sub-box, subject to symmetry constraints. Random structures were generated and then screened to determine whether the atoms were three-fold coordinated. If not, the structure was rejected and another one was generated. Right: relaxation of this structure within DFT gave the well-known C_{60} "buckyball".

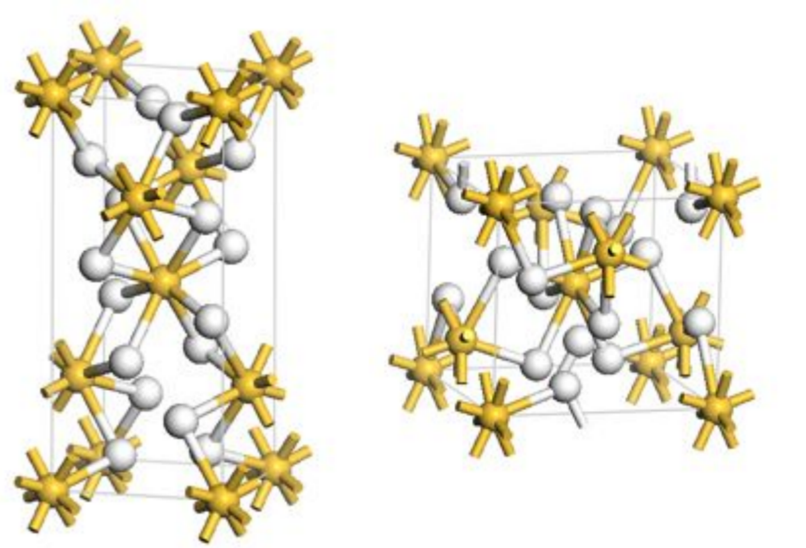


Figure 8. The $I4_1/a$ structure of silane (left) and the slightly less stable $I\bar{4}2d$ structure (right). Silicon atoms are shown in gold and hydrogen atoms are in white. All of the bonds in $I4_1/a$ and $I\bar{4}2d$ are of the Si–H–Si type. Both phases were subsequently found experimentally.

High-Pressure Phases of Silane

Chris J. Pickard and R. J. Needs

Phys. Rev. Lett. **97**, 045504 – Published 27 July 2006

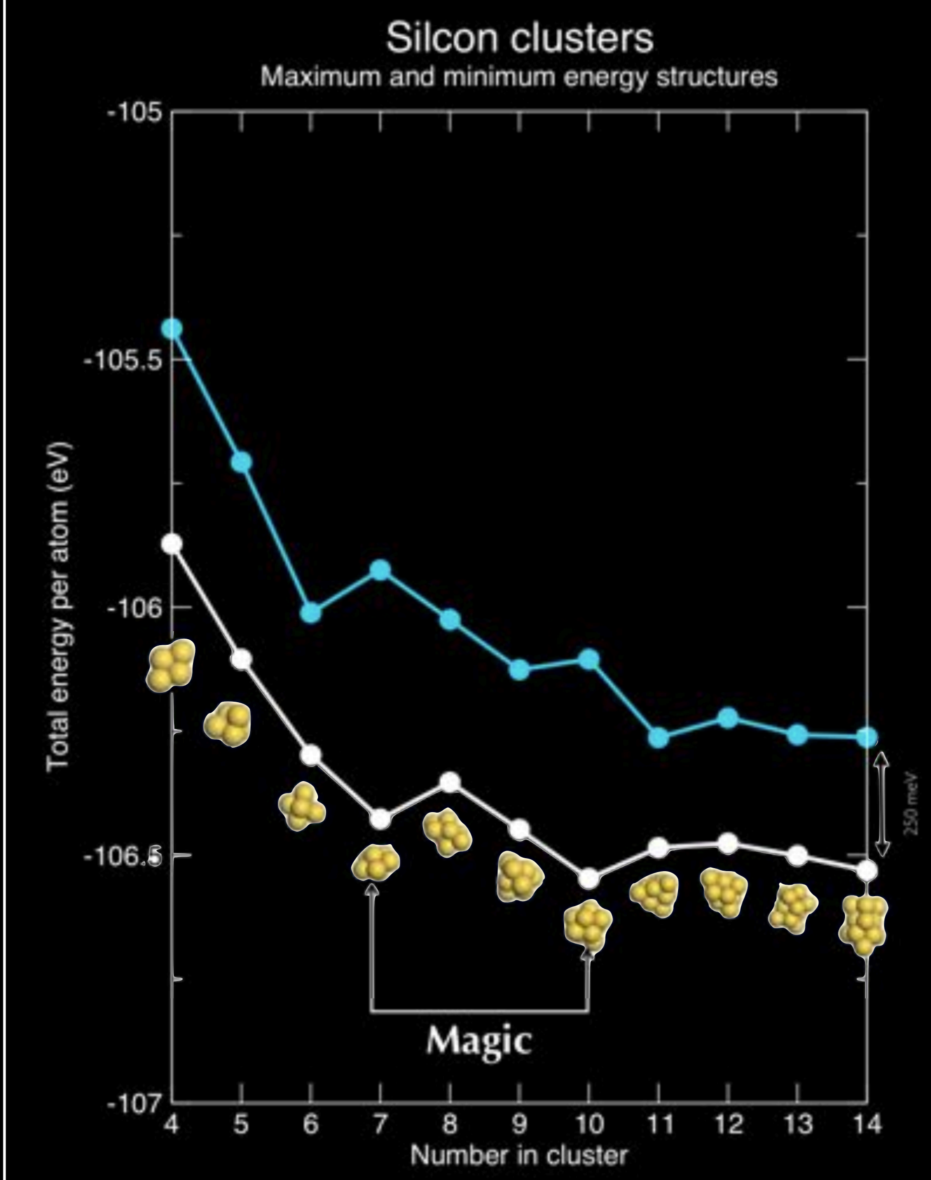


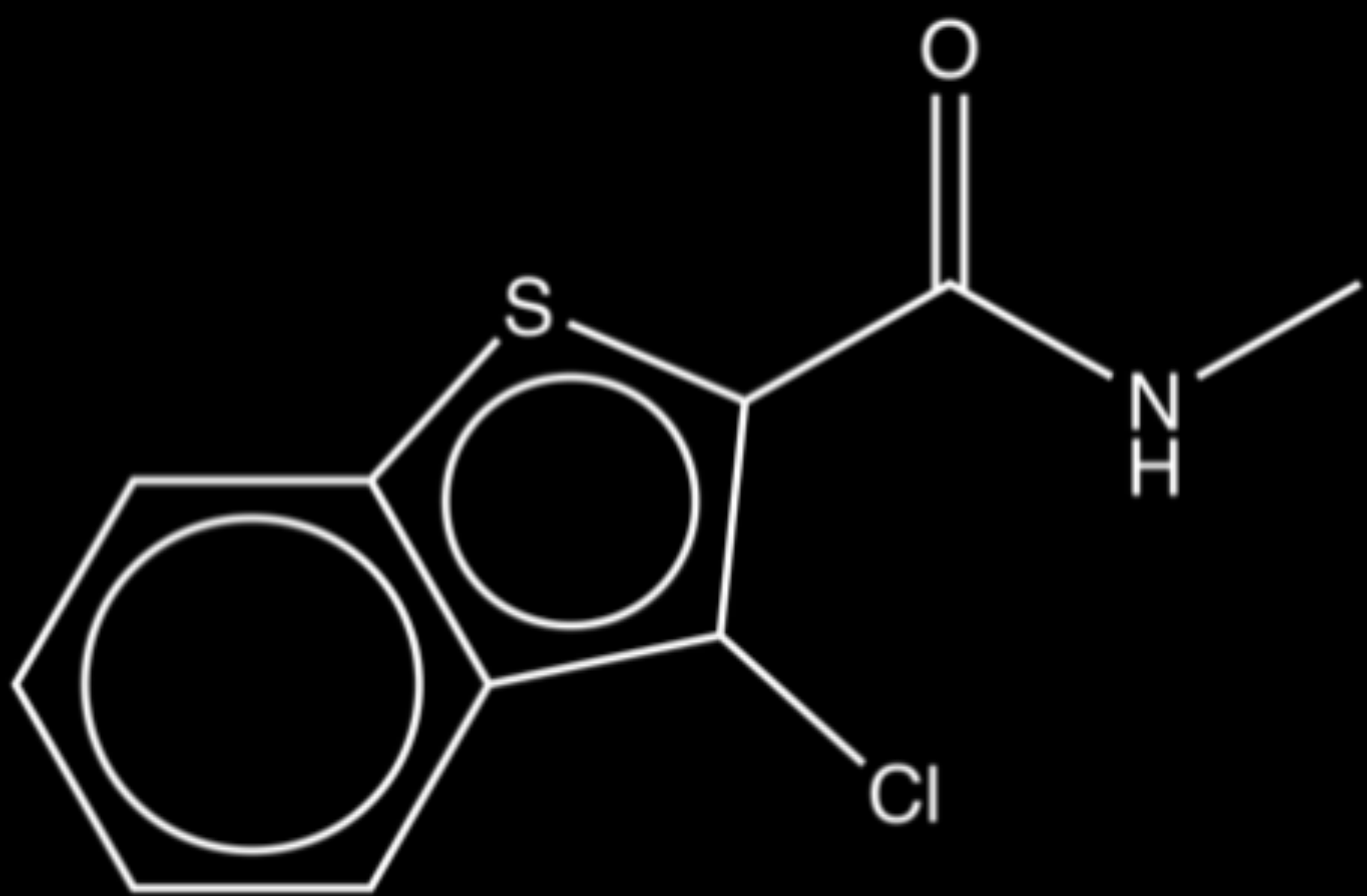
Figure 2. Silicon clusters were generated by placing atoms randomly within a small box inside a large unit cell and relaxing within DFT. The algorithm generated the same lowest-energy structures obtained in previous DFT studies [22], including the two "magic" number clusters with seven and ten atoms. We also found many local minima. The highest-energy minimum for each cluster size is only about 0.25 eV per atom higher in energy than the minimum energy structure.

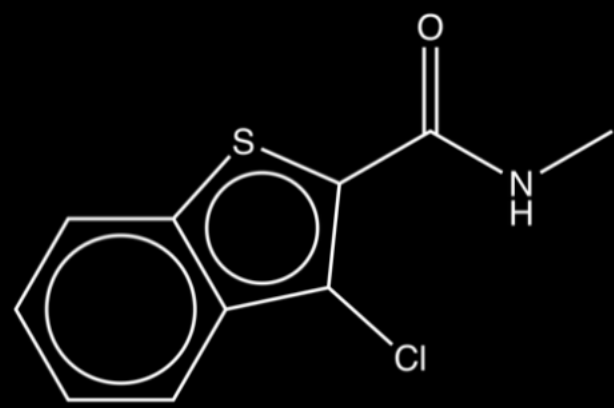
Ab initio random structure searching

Chris J Pickard[1] and R J Needs[2]

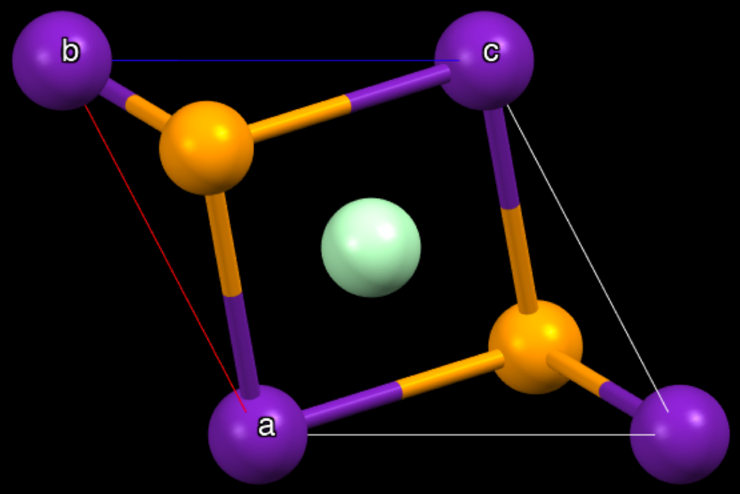
Published 5 January 2011 • IOP Publishing Ltd

Journal of Physics: Condensed Matter, Volume 23, Number 5


O
S
NH
Cl



Drug-like Molecule



Crystal Structure

SMILES string `Cn1cnc2n(C)c(=O)n(C)c(=O)c12`

Graph representation

Tend to use 5-20 total atoms

Mostly C,H,O- less than 10 total species

Not applicable

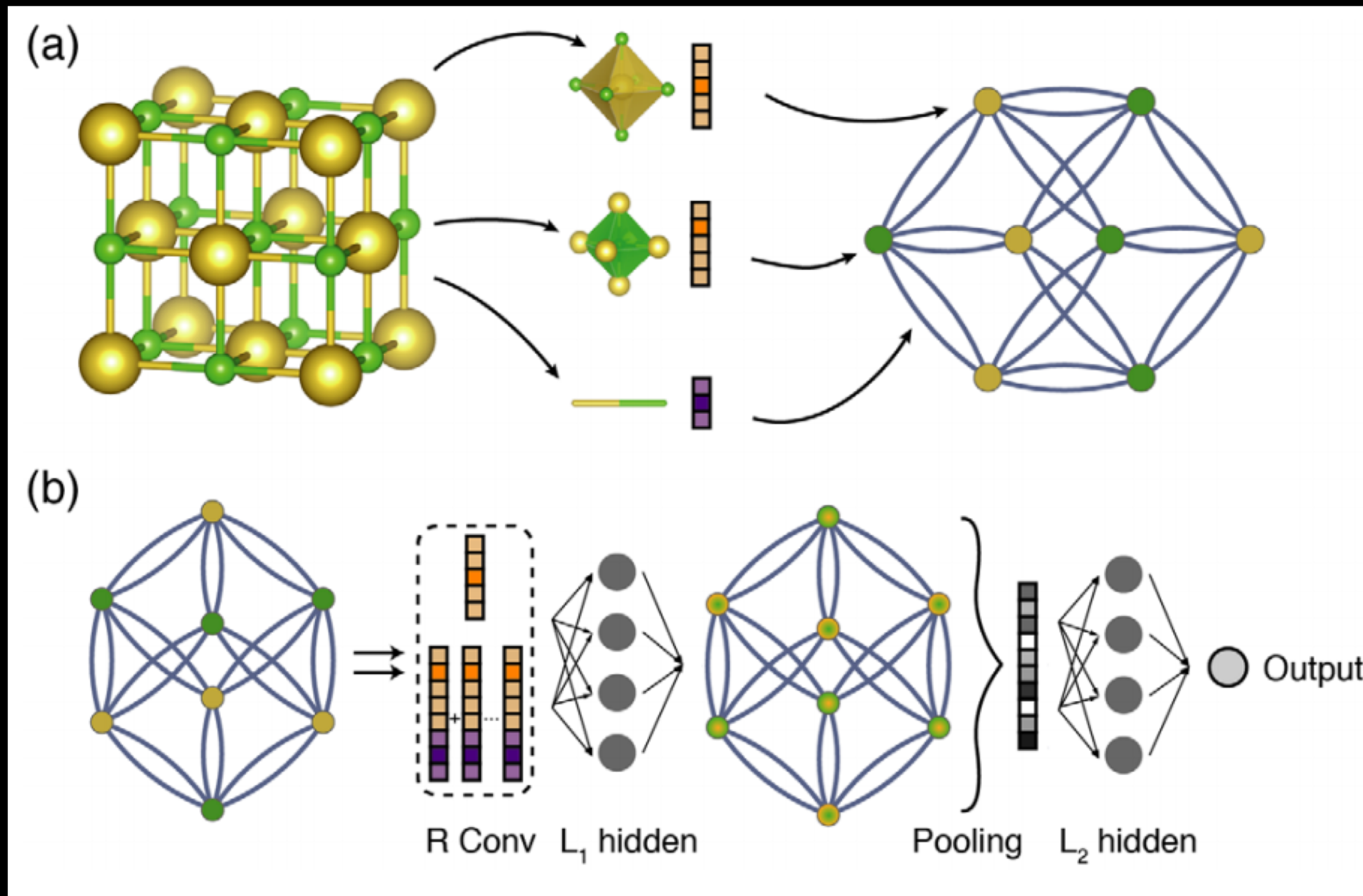
Need location information

Tend to use 1-100 total atoms

Entire periodic table

Less data exists
slower to compute

6 additional degrees of freedom
are needed



Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties

Tian Xie and Jeffrey C. Grossman
Phys. Rev. Lett. **120**, 145301 – Published 6 April 2018

Input format

```
# generated using pymatgen
data_Ti2MnIn
_symmetry_space_group_name_H-M   'P 1'
_cell_length_a   4.48903651
_cell_length_b   4.48903651
_cell_length_c   4.48903651
_cell_angle_alpha   60.00000000
_cell_angle_beta   60.00000000
_cell_angle_gamma   60.00000000
_symmetry_Int_Tables_number   1
_chemical_formula_structural   Ti2MnIn
_chemical_formula_sum   'Ti2 Mn1 In1'
_cell_volume   63.96529632
_cell_formula_units_Z   1
loop_
 _symmetry_equiv_pos_site_id
 _symmetry_equiv_pos_as_xyz
  1  'x, y, z'
loop_
 _atom_site_type_symbol
 _atom_site_label
 _atom_site_symmetry_multiplicity
 _atom_site_fract_x
 _atom_site_fract_y
 _atom_site_fract_z
 _atom_site_occupancy
  Ti  Ti1  1  -0.000000  0.000000  0.000000  1
  Ti  Ti2  1  0.750000  0.750000  0.750000  1
  Mn  Mn3  1  0.500000  0.500000  0.500000  1
  In  In4  1  0.250000  0.250000  0.250000  1
```

We need:

1.) Locations

2.) Species

$$\boldsymbol{M}_{i,j,k} = \frac{1}{\sigma^3(2\pi)^{3/2}} \sum_m Z_m \exp\left(-\frac{d(\vec{Z}_m,(i,j,k))^2}{2\sigma^2}\right)$$

```
# generated using pymatgen
data_Ti2MnIn
_symmetry_space_group_name_H-M   'P 1'
_cell_length_a   4.48903651
_cell_length_b   4.48903651
_cell_length_c   4.48903651
_cell_angle_alpha   60.00000000
_cell_angle_beta   60.00000000
_cell_angle_gamma   60.00000000
_symmetry_Int_Tables_number   1
_chemical_formula_structural   Ti2MnIn
_chemical_formula_sum   'Ti2 Mn1 In1'
_cell_volume   63.96529632
_cell_formula_units_Z   1
loop_
 _symmetry_equiv_pos_site_id
 _symmetry_equiv_pos_as_xyz
  1  'x, y, z'
loop_
 _atom_site_type_symbol
 _atom_site_label
 _atom_site_symmetry_multiplicity
 _atom_site_fract_x
 _atom_site_fract_y
 _atom_site_fract_z
 _atom_site_occupancy
  Ti  Ti1  1  -0.000000  0.000000  0.000000  1
  Ti  Ti2  1  0.750000  0.750000  0.750000  1
  Mn  Mn3  1  0.500000  0.500000  0.500000  1
  In  In4  1  0.250000  0.250000  0.250000  1
```

We need:

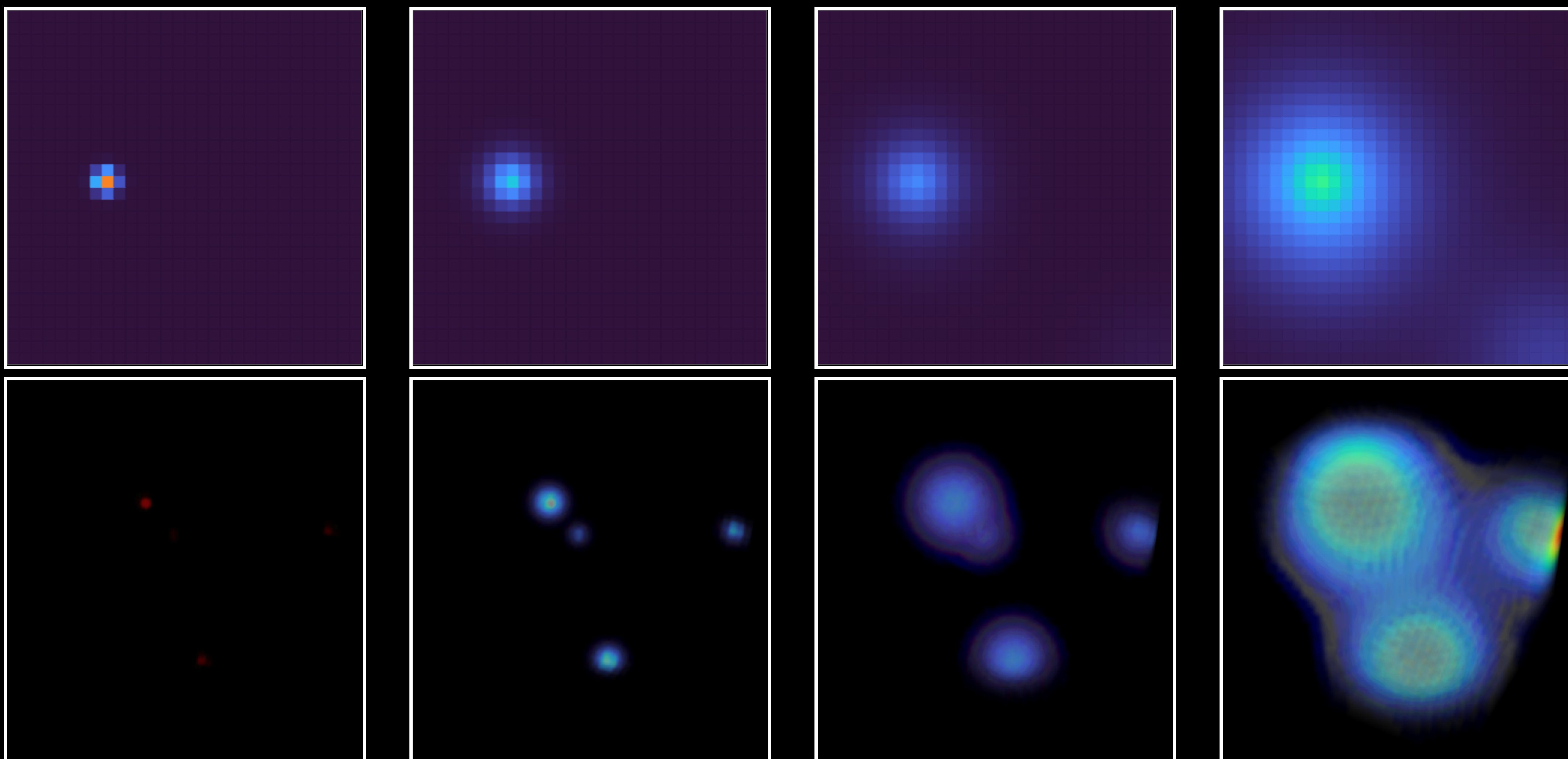
1.) Locations

2.) Species

$$\boldsymbol{M}_{i,j,k} = \frac{1}{\sigma^3(2\pi)^{3/2}} \sum_m Z_m \mathrm{exp}\left(-\frac{d(\vec{Z}_m, (i,j,k))^2}{2\sigma^2}\right)$$

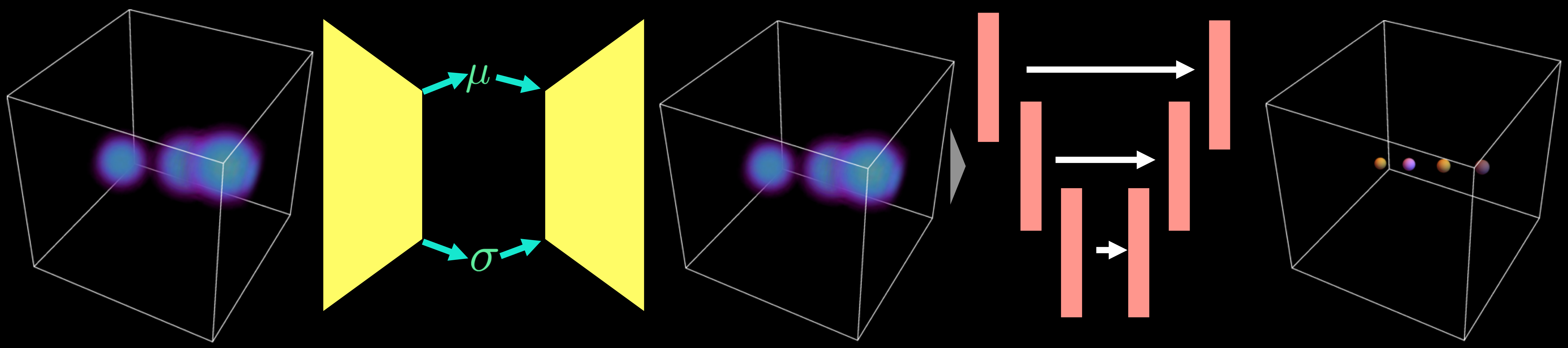
$$\boldsymbol{M}_{i,j,k} = \frac{1}{\sigma^3(2\pi)^{3/2}} \sum_m Z_m \exp\left(-\frac{d(\vec{Z}_m,(i,j,k))^2}{2\sigma^2}\right)$$

$$\sigma$$

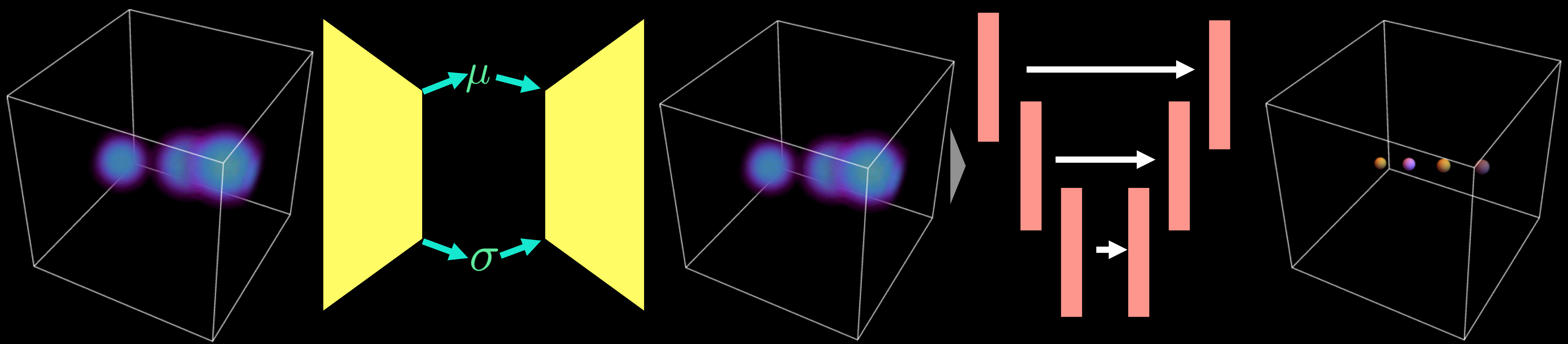


Variational Autoencoder

U-Net

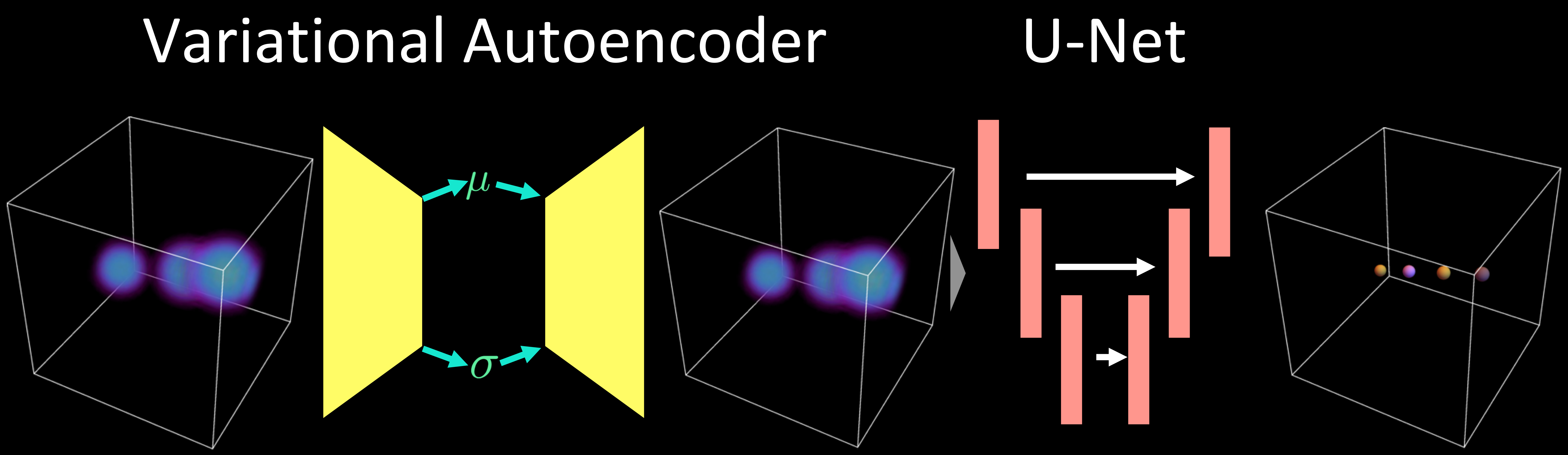


Variational Autoencoder U-Net



Encodes and decodes the density matrix.

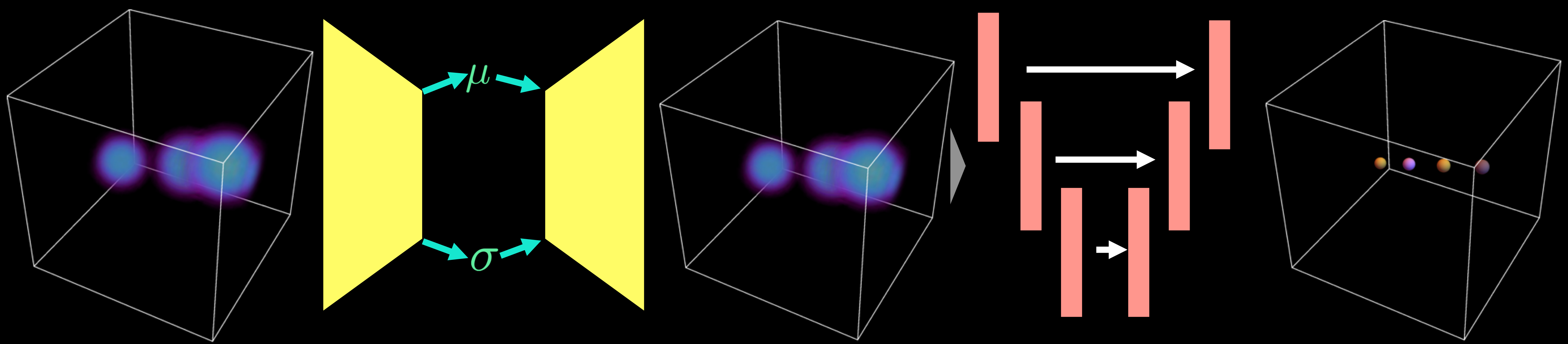
Solves the problem of converting a density field back to species. If sigma = 0, trivial task. In theory, this is just solving a complicated optimization problem.



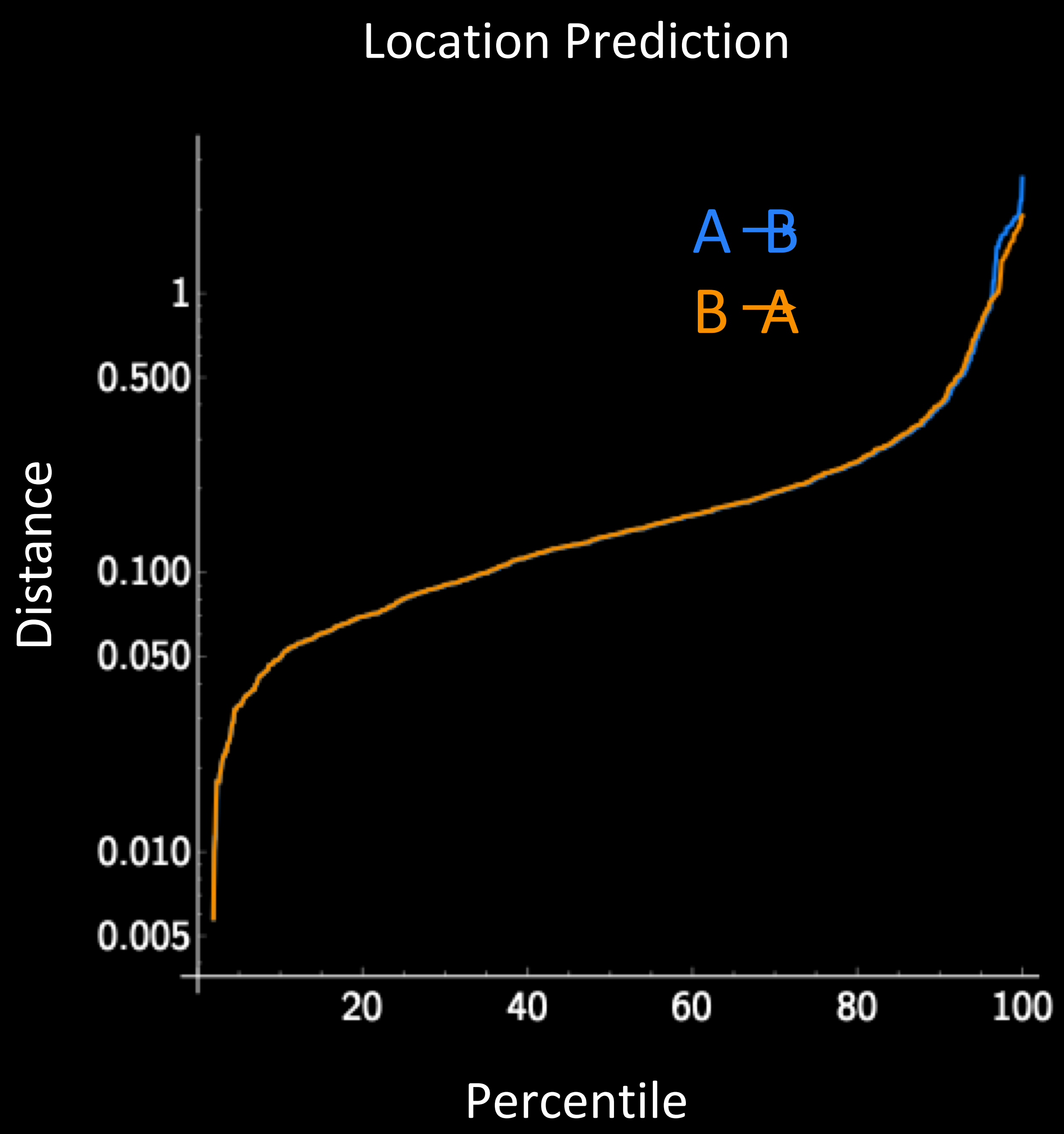
$$\mathcal{L}_{\text{VAE}} = L_{\text{RE}}(\hat{\boldsymbol{M}}, \boldsymbol{M}) + \beta(D_{\text{KL}}(q(\mathbf{z}|\boldsymbol{M})||p(\mathbf{z}))) + \gamma L_{\text{BCE}}(\hat{\boldsymbol{S}}, \boldsymbol{S})$$

$$\mathcal{L}_{\text{U-Net}} = L_{\text{BCE}}(\hat{\boldsymbol{S}}, \boldsymbol{S})$$

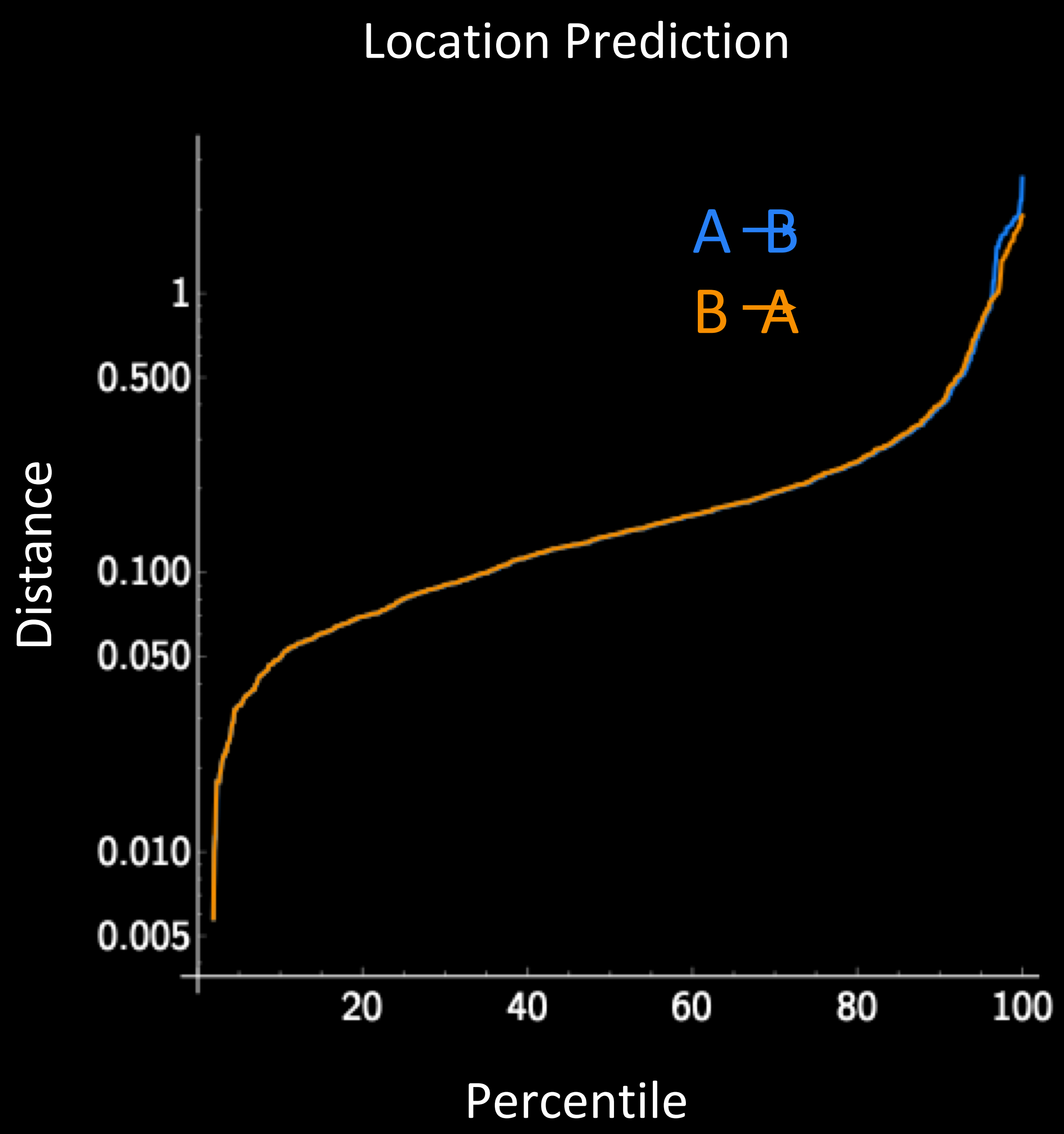
Variational Autoencoder U-Net

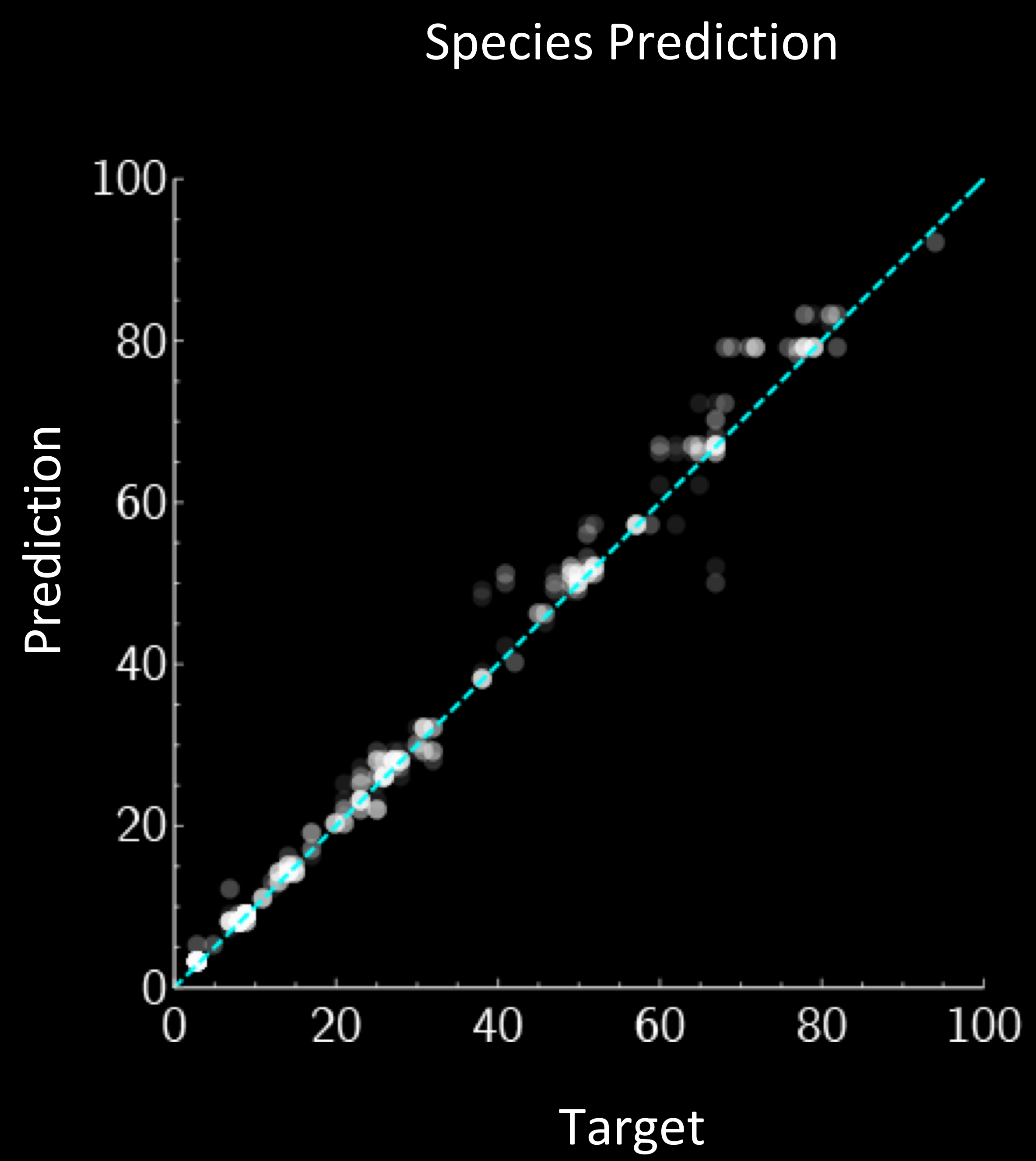


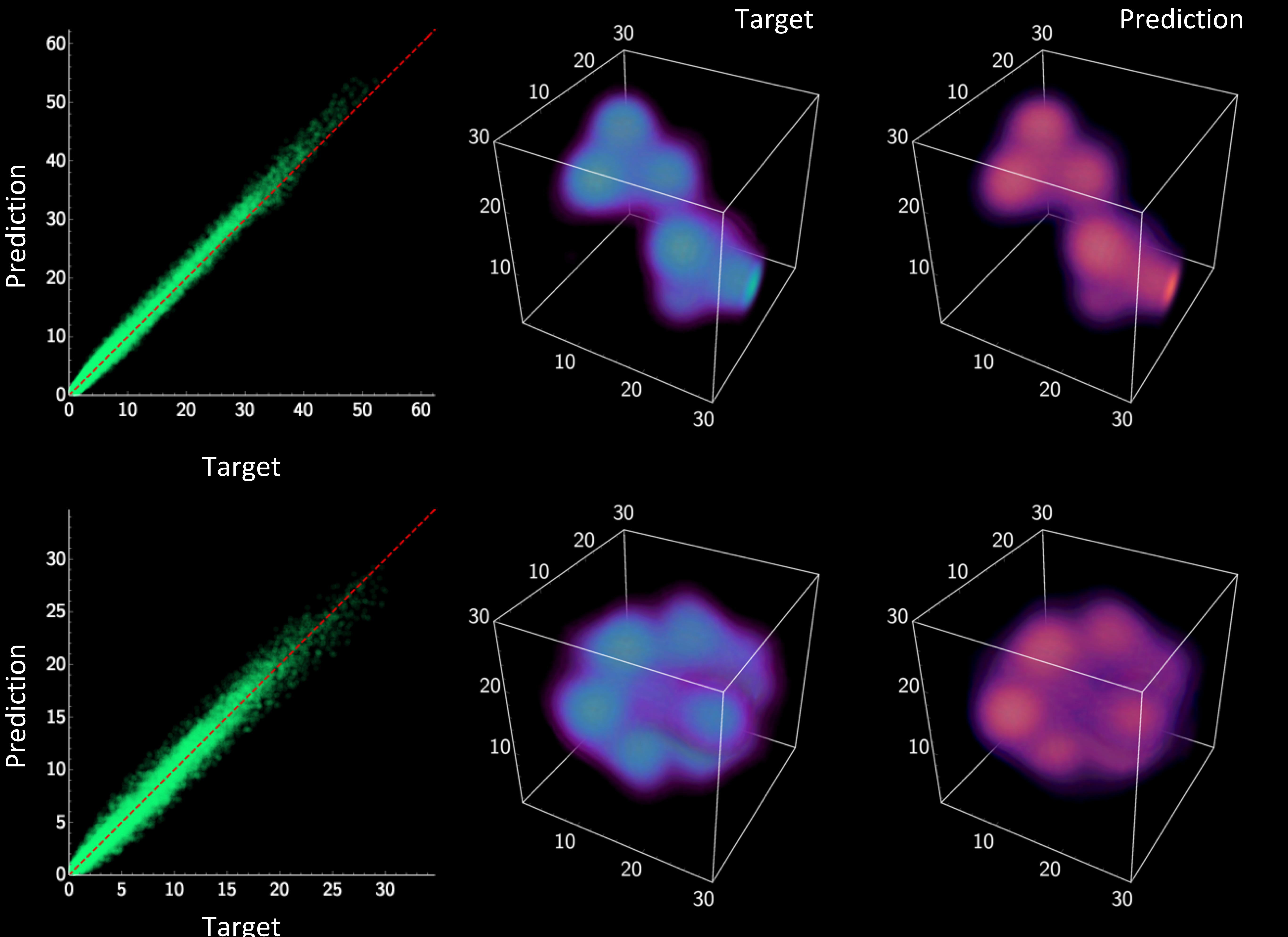
- 10 Angstrom on a side, used 30 and 60 pixels across
- 46,000 crystal structures, 35,000 considered
- Always centered, but we can rotate them randomly



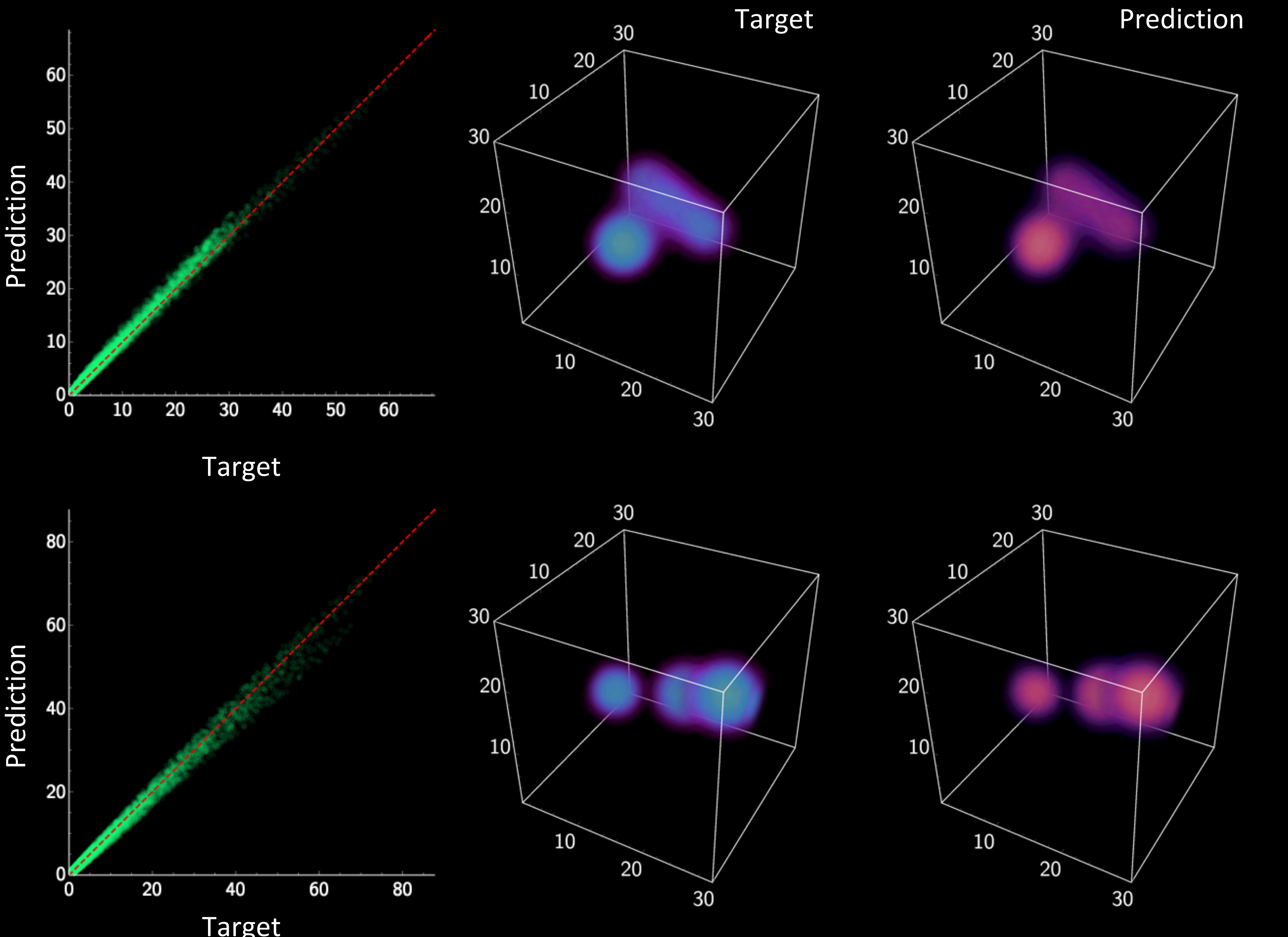
Location Prediction
A → B
B → A
1
0.500
0.100
0.050
0.010
0.005
Distance
20
40
60
80
100
Percentile


Location Prediction
A → B
B → A
Distance
1
0.500
0.100
0.050
0.010
0.005
20
40
60
80
100
Percentile

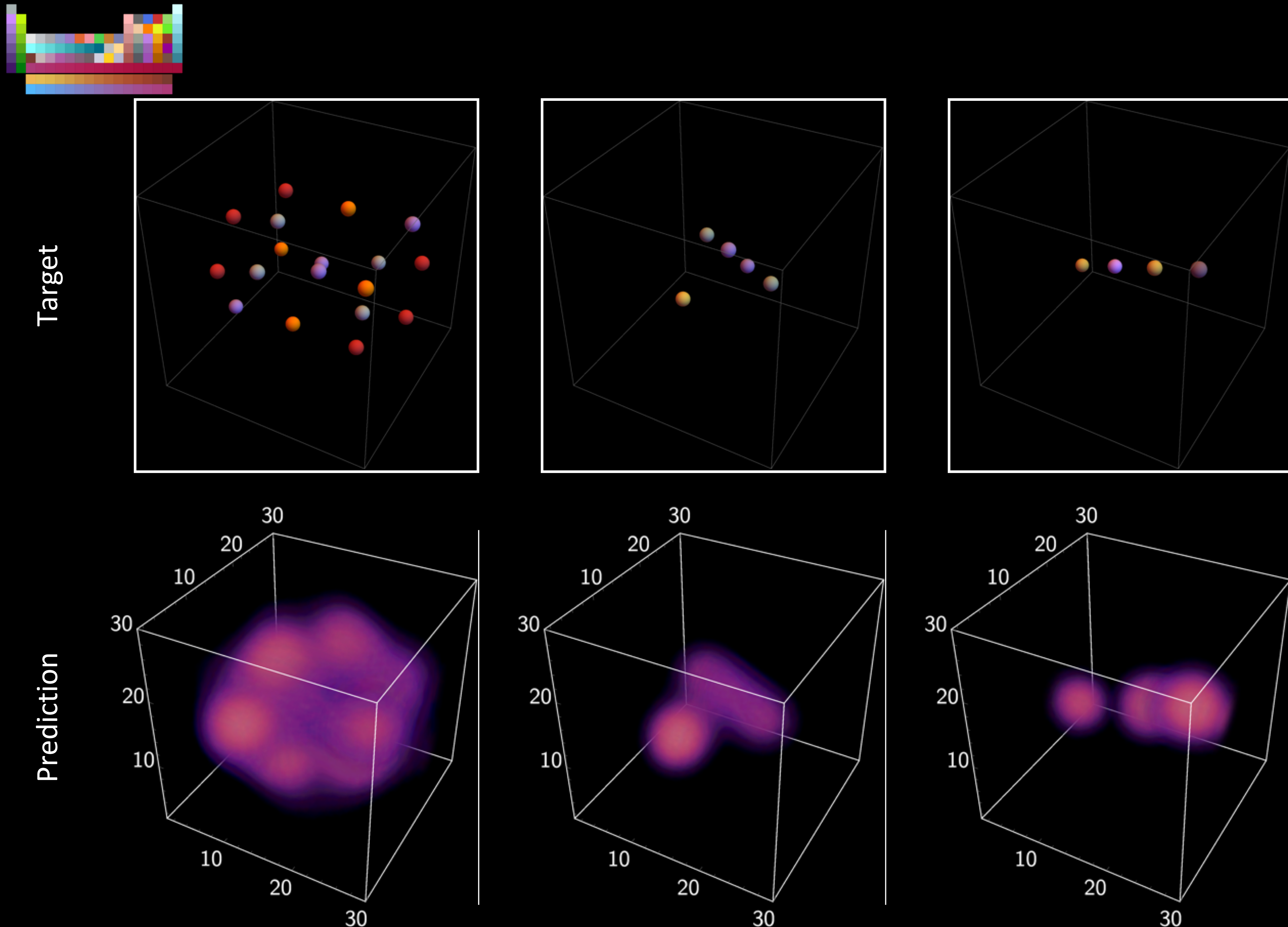

Species Prediction
100
80
60
40
20
0
Prediction
0
20
40
60
80
100
Target



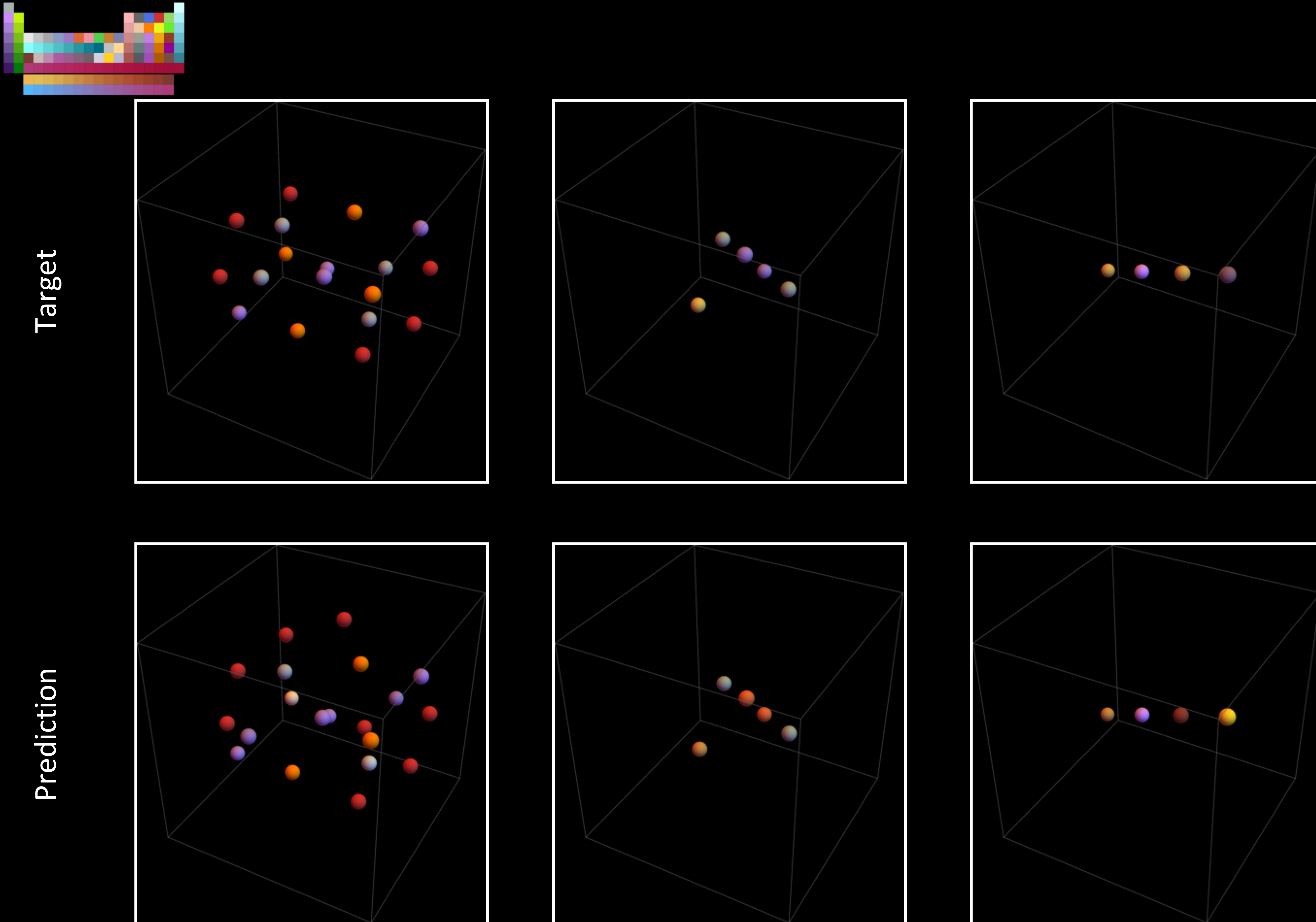
Target
Prediction
Prediction
Target
Prediction
Target

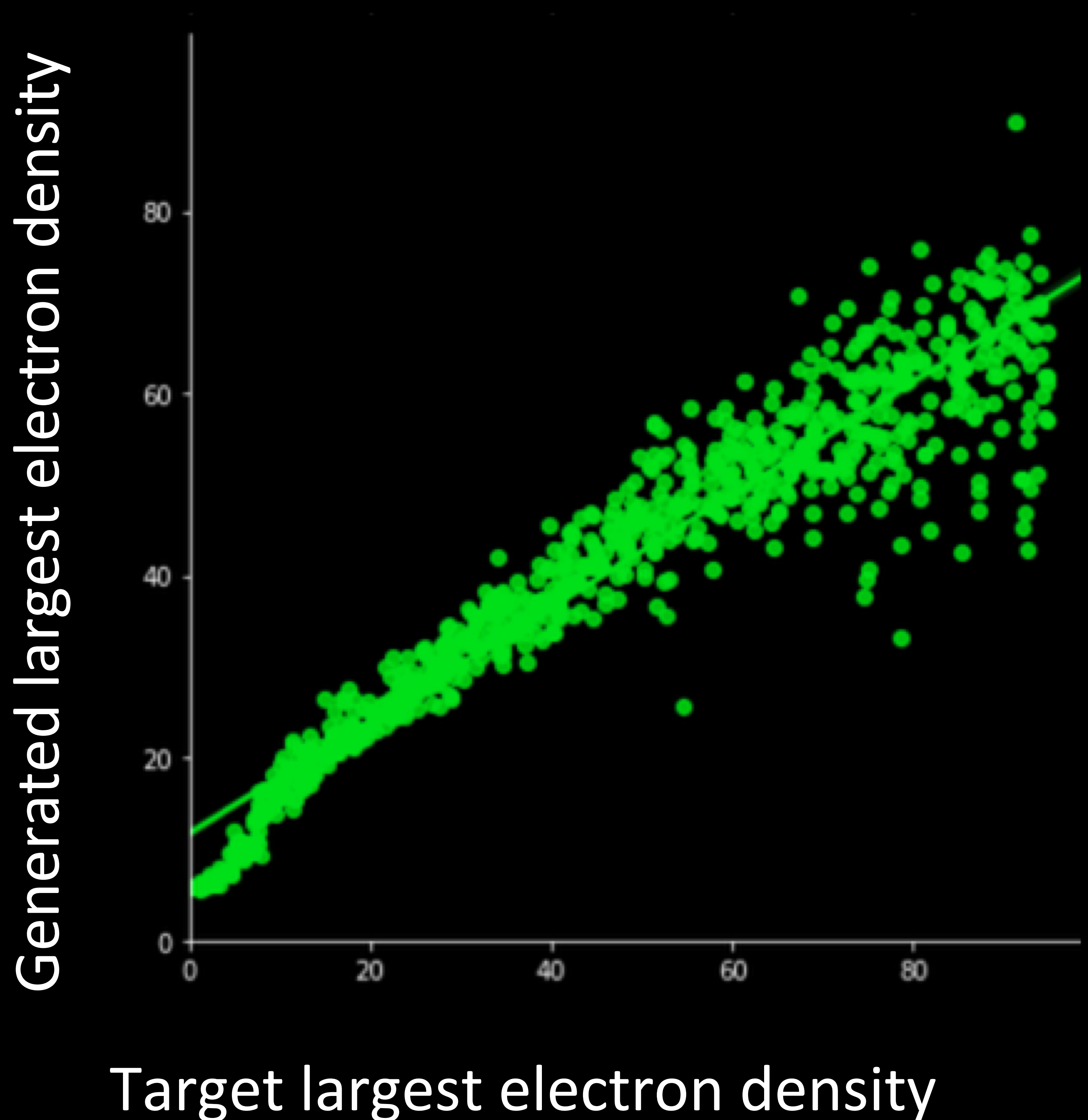


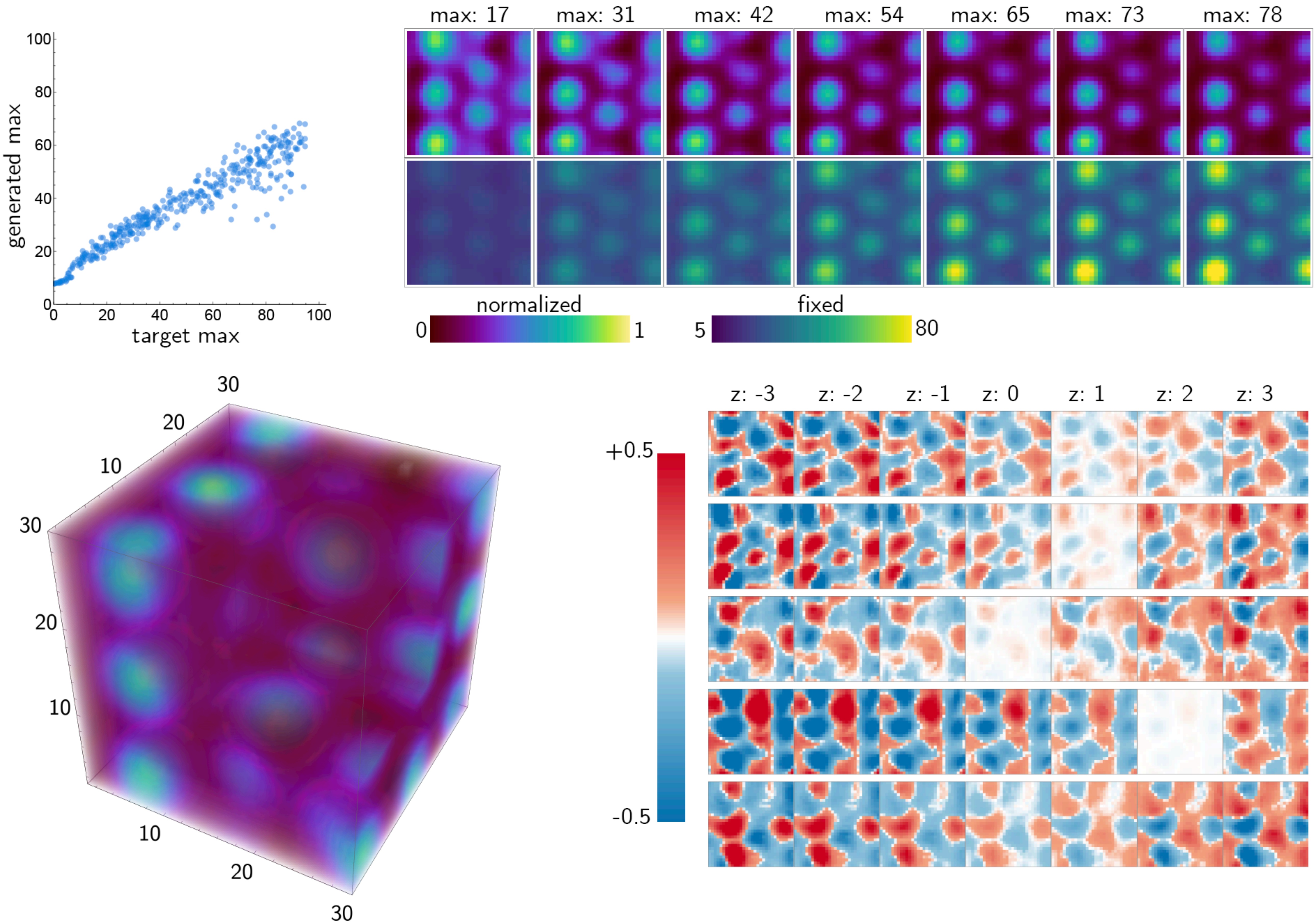
Target
Prediction
Prediction
Target
Prediction
Target
0
10
20
30
40
50
60
0
20
40
60
80
10
20
30



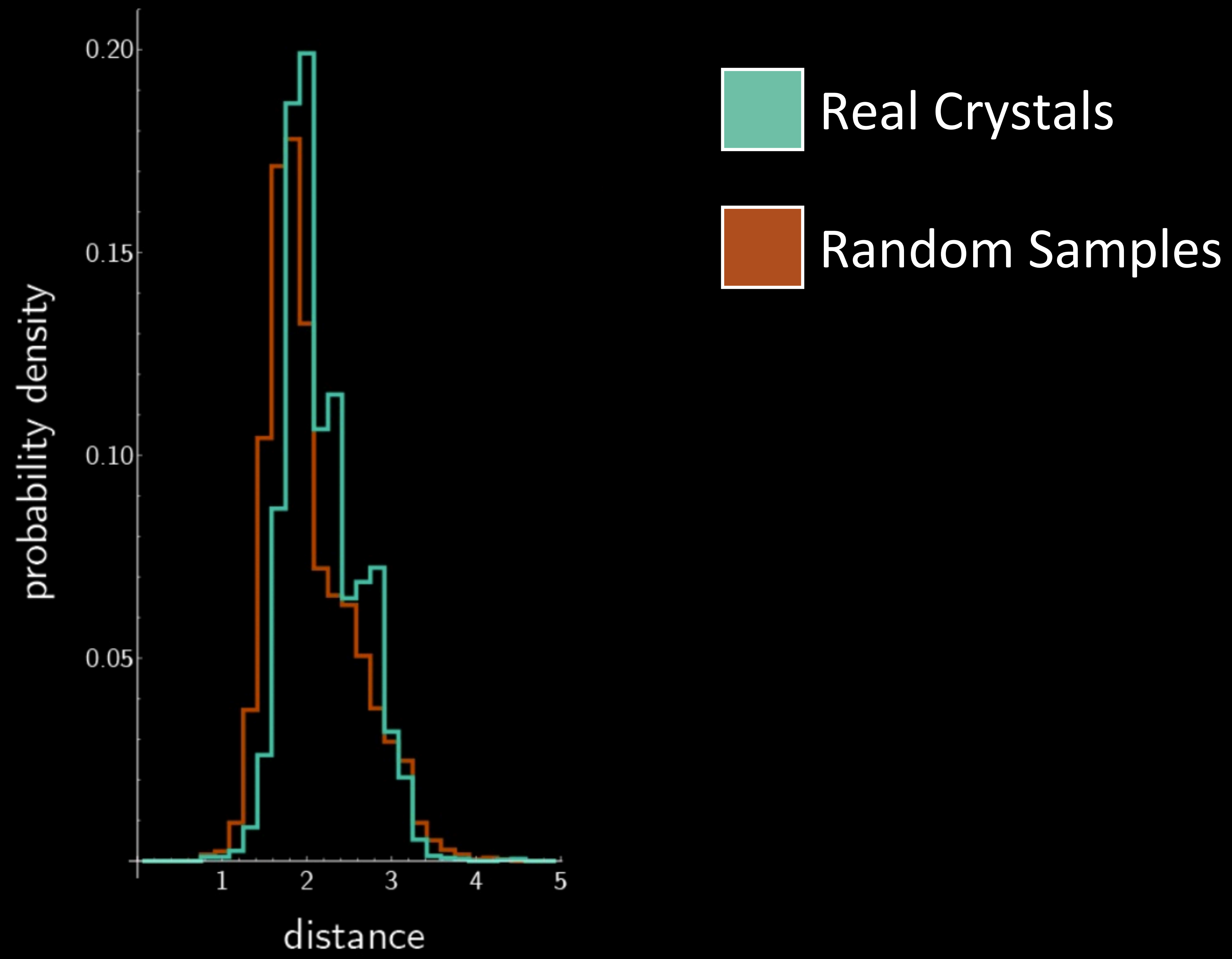
Target
Prediction
30
20
10
30
20
10
10
20
30


Target
Prediction

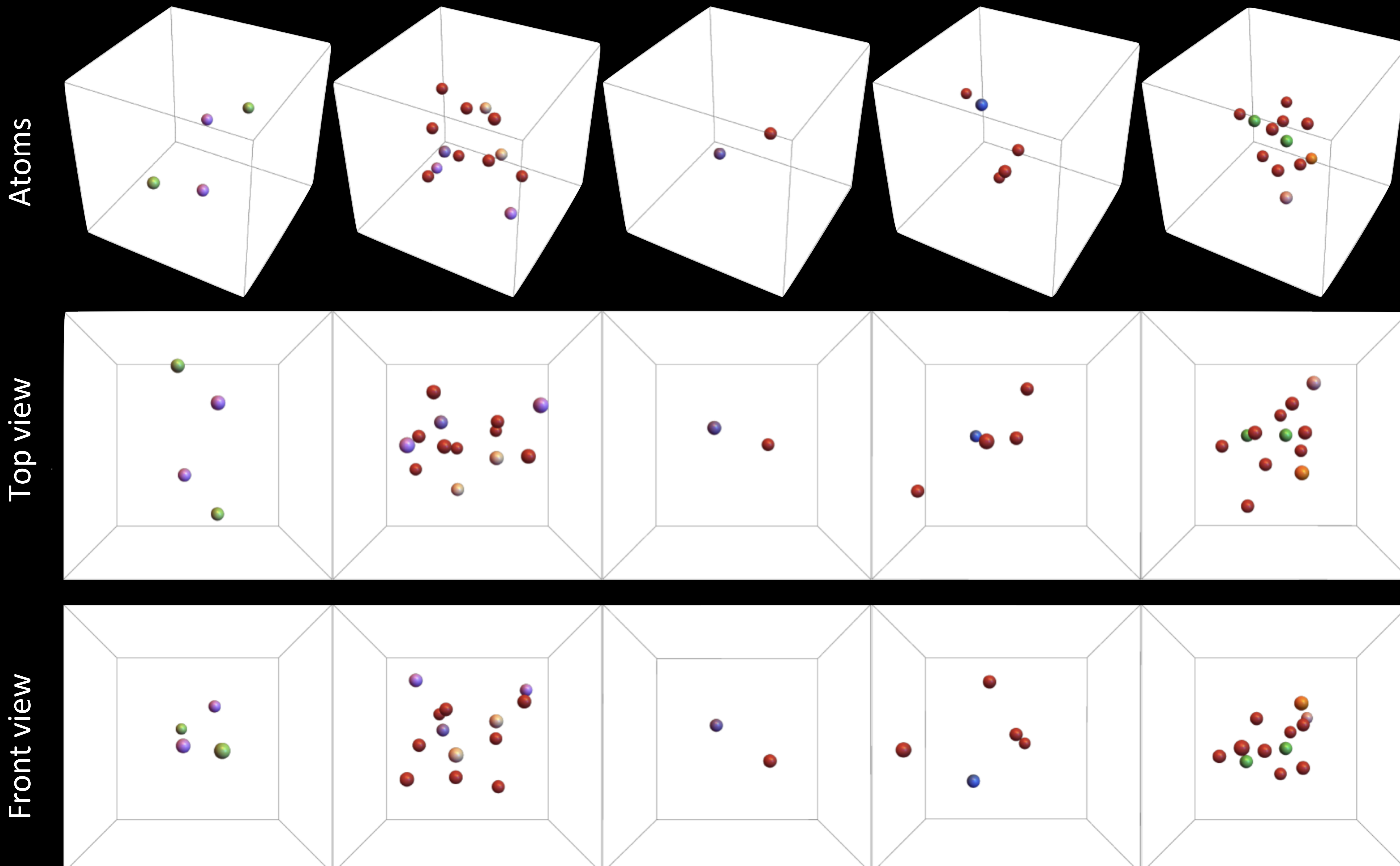

Generated largest electron density
0
20
40
60
80
Target largest electron density

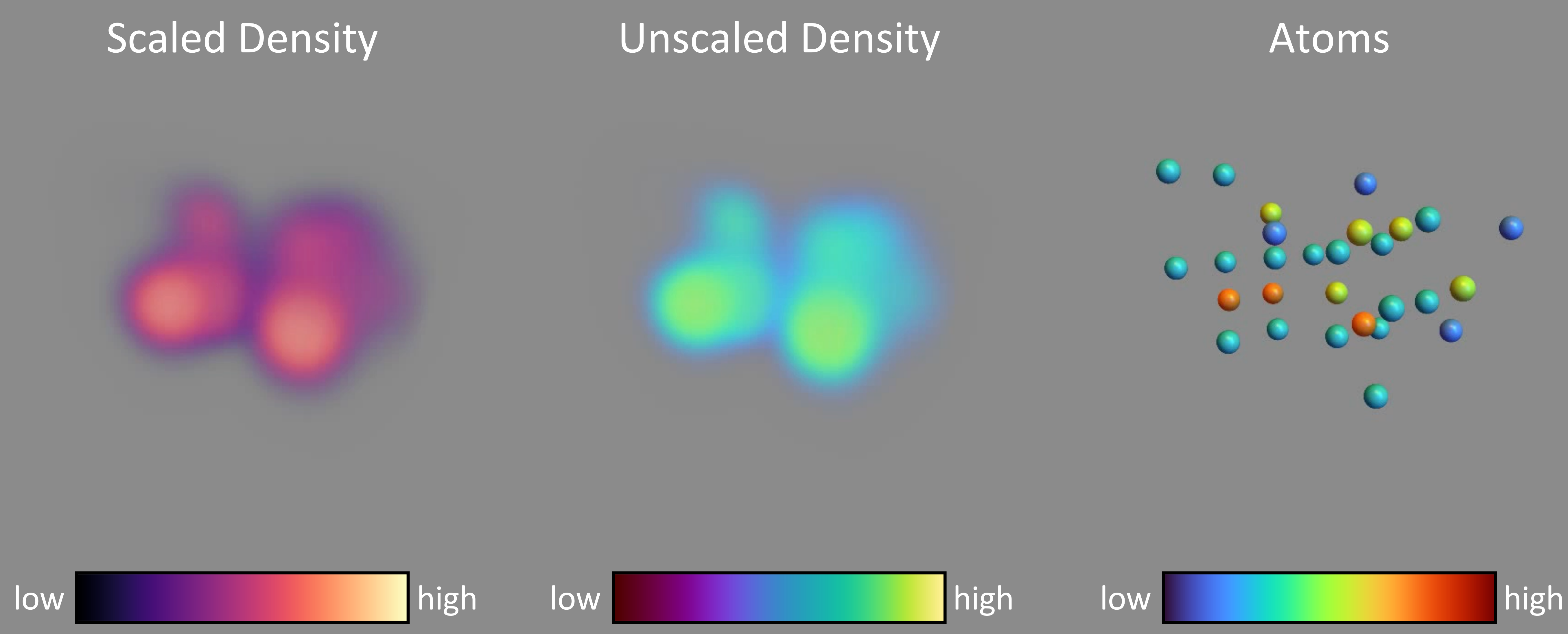


generated max
target max
0
20
40
60
80
100
max: 17
max: 31
max: 42
max: 54
max: 65
max: 73
max: 78
normalized
0
1
fixed
5
80
30
20
10
30
20
10
10
20
30
+0.5
-0.5
z: -3
z: -2
z: -1
z: 0
z: 1
z: 2
z: 3



0.20
0.15
0.10
0.05
probability density
1
2
3
4
5
distance
Real Crystals
Random Samples


Atoms
Top view
Front view


Scaled Density
Unscaled Density
Atoms
low
high
low
high
low
high