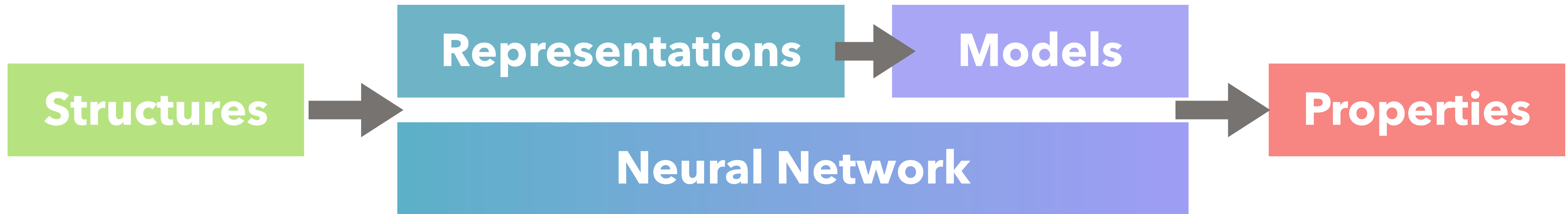
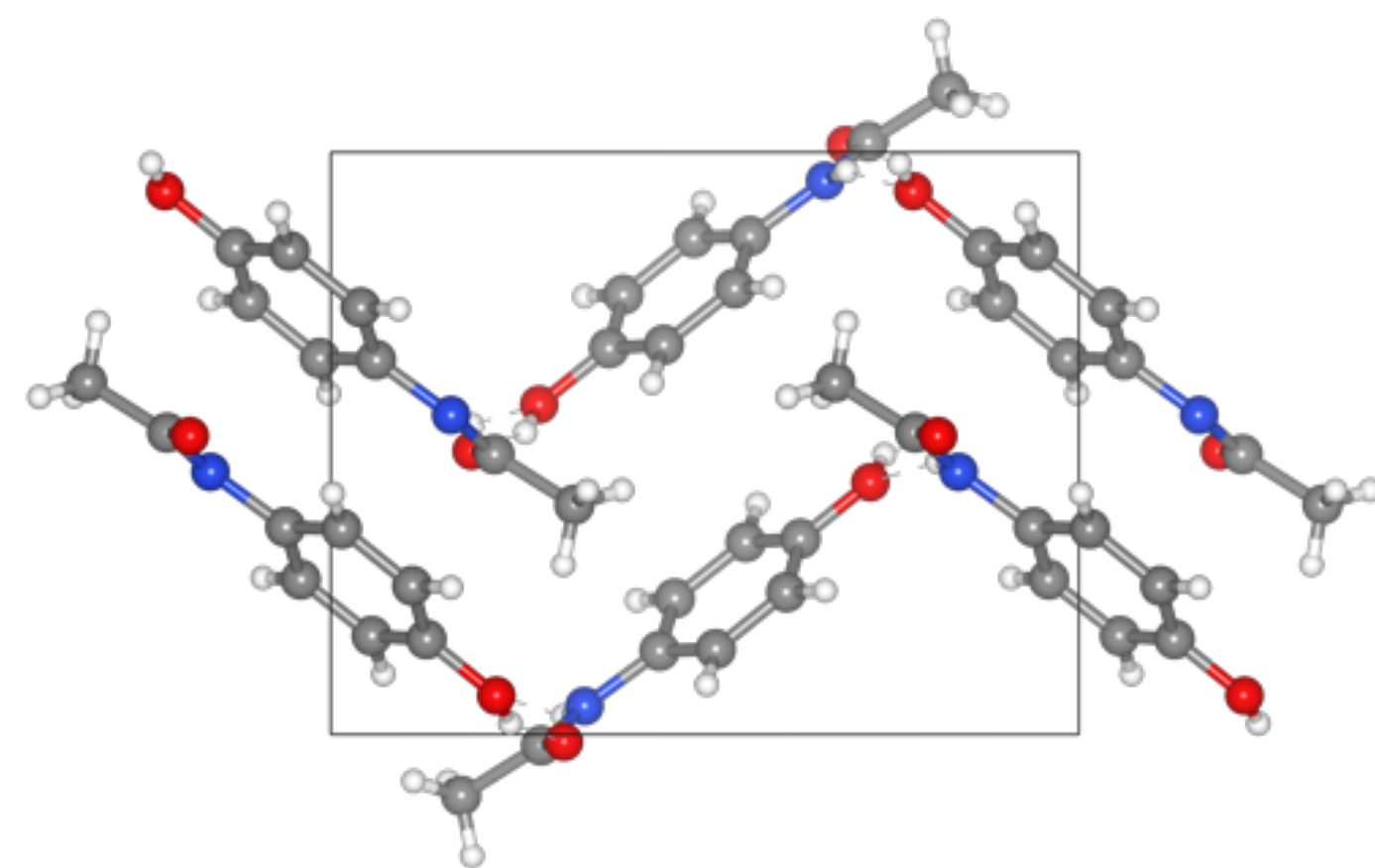
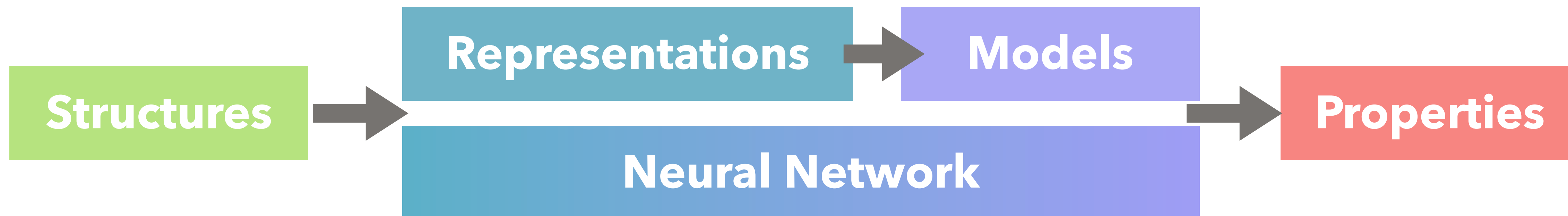


Exploring structure-property maps with kernel principal covariates regression and chemiscope





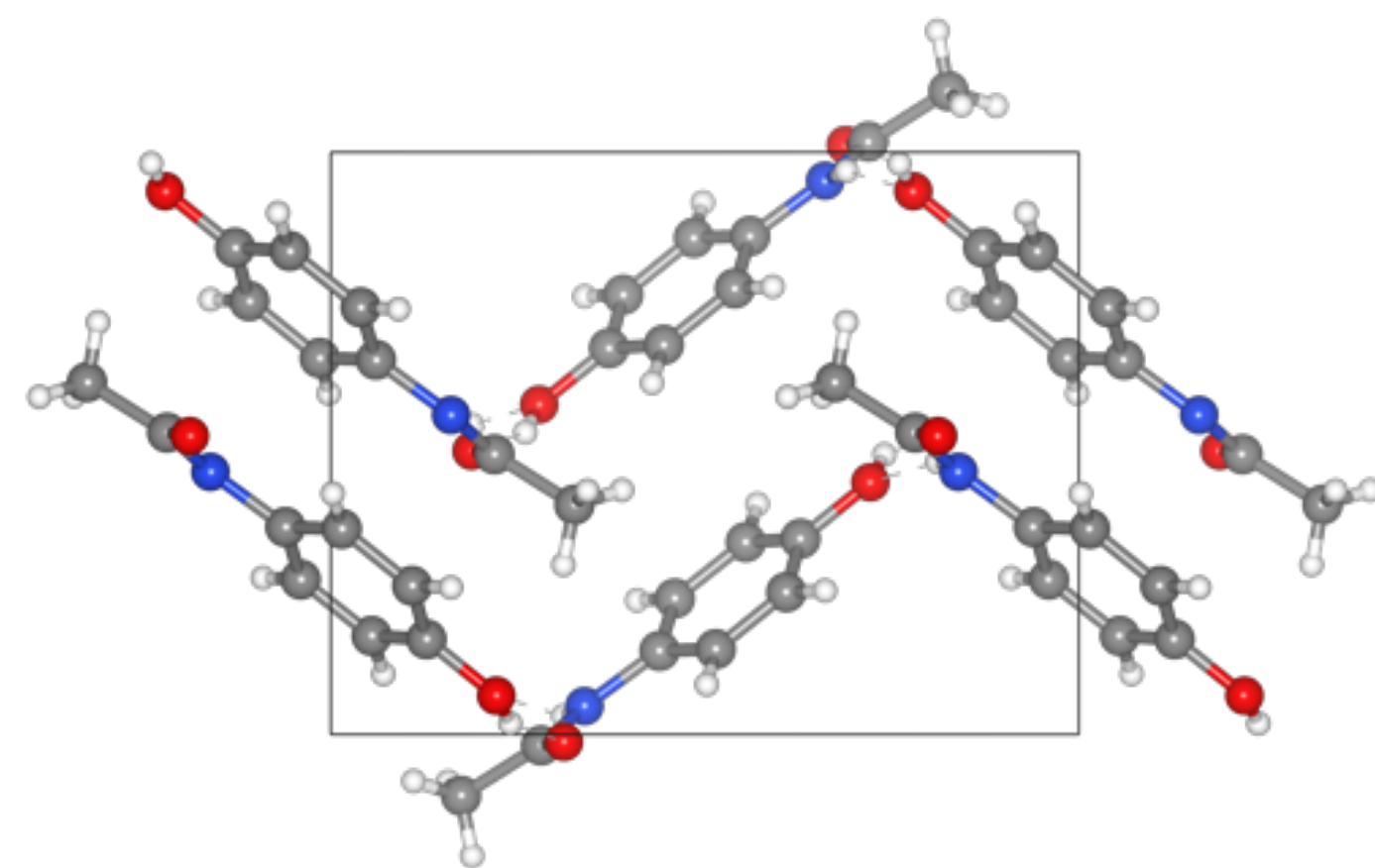
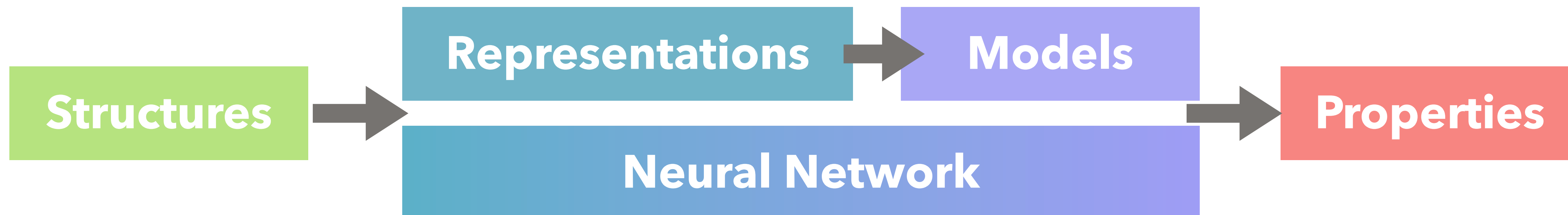
Structures space,
not ideal for ML



High dimensional
feature space



Property space



Structures space,
not ideal for ML



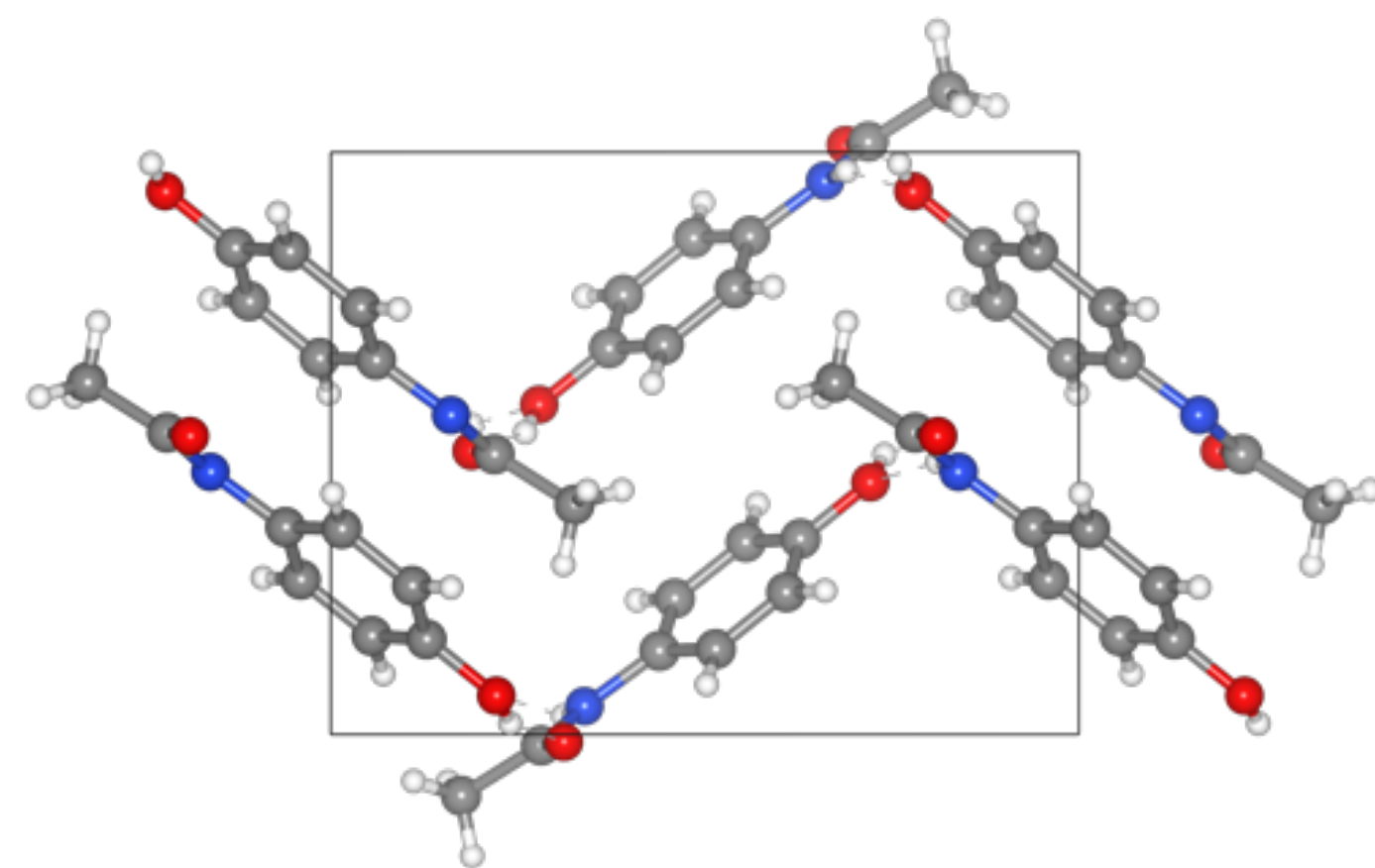
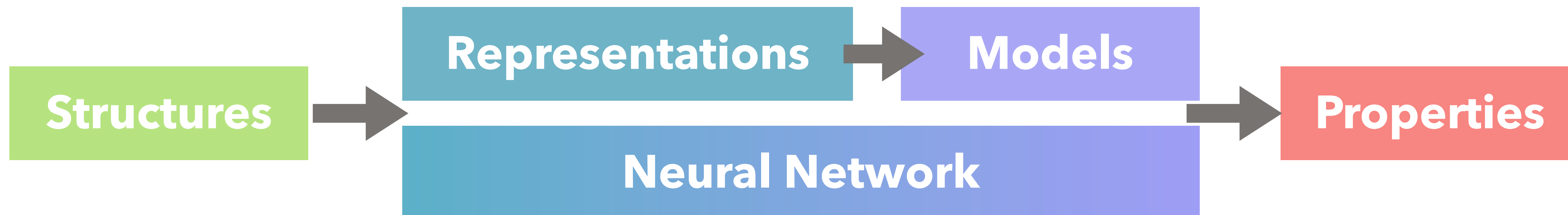
High dimensional
feature space



Low dimensional
latent space



Property space



Structures space,
not ideal for ML



High dimensional
feature space

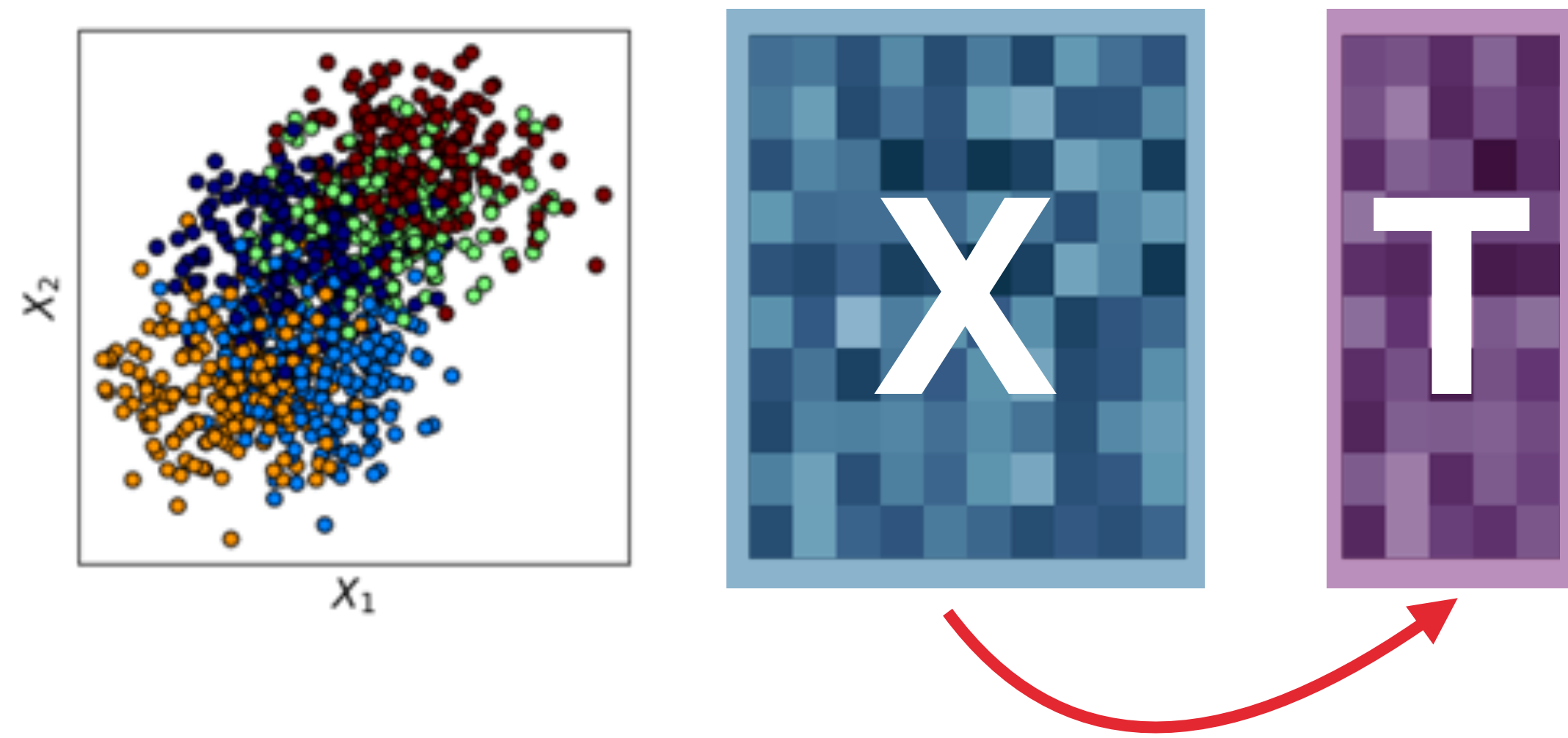


Low dimensional
latent space

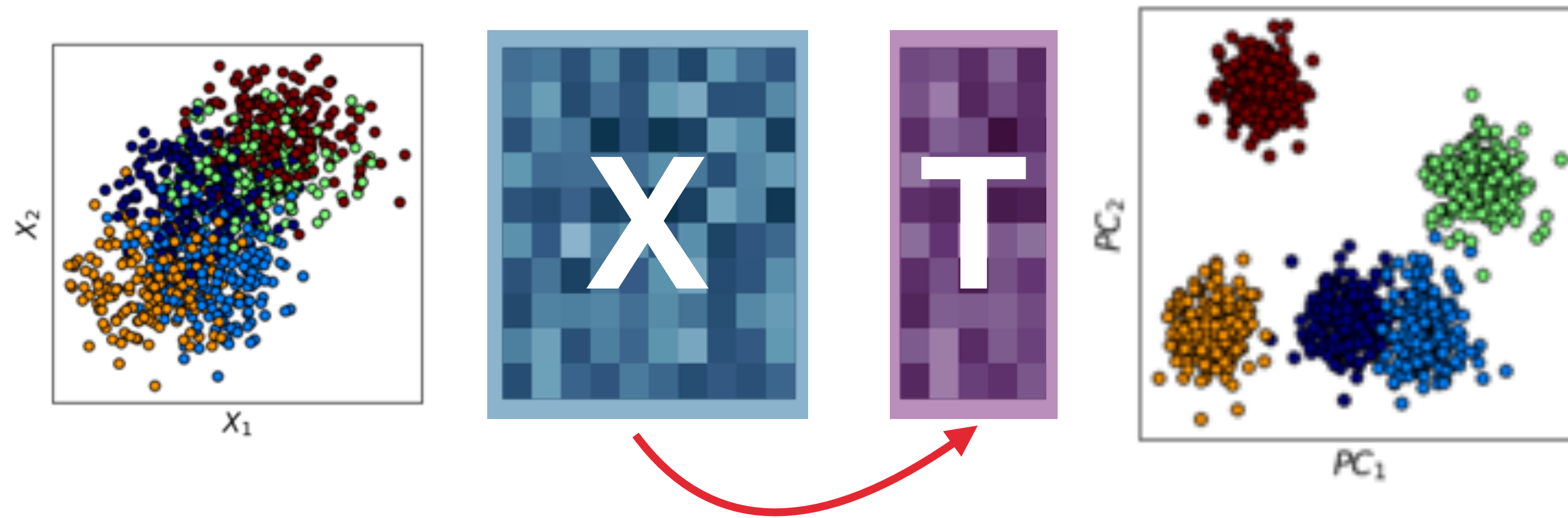


Property space

How can we understand the relationships between input space, representation and predictions?



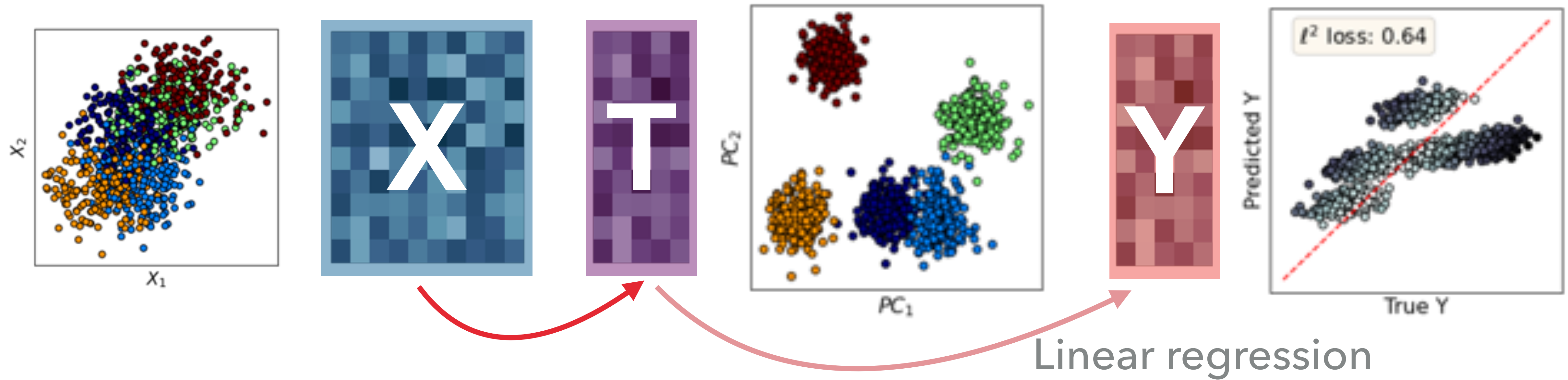
Principal Components Analysis (PCA)



Principal Components Analysis (PCA)

From PCA to Principal Covariates Regression

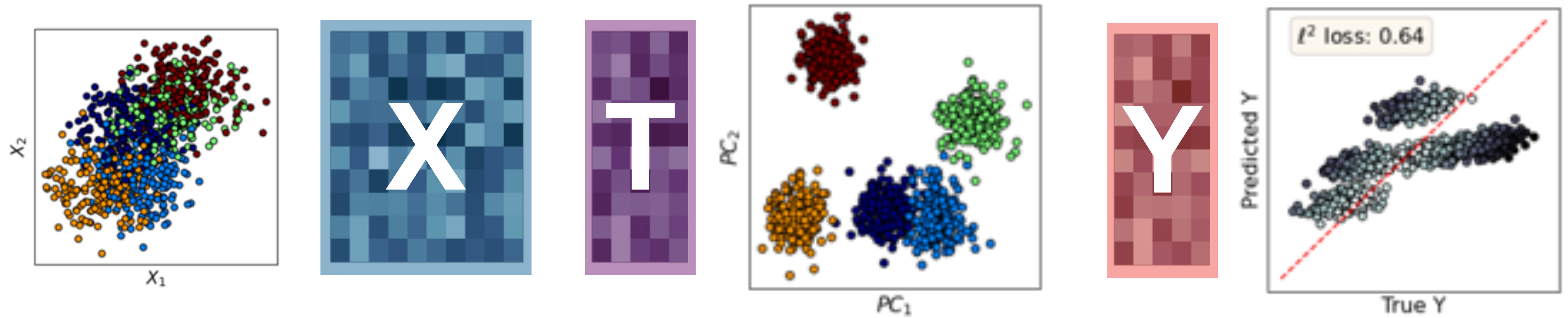
3



Principal Components Analysis (PCA)

From PCA to Principal Covariates Regression

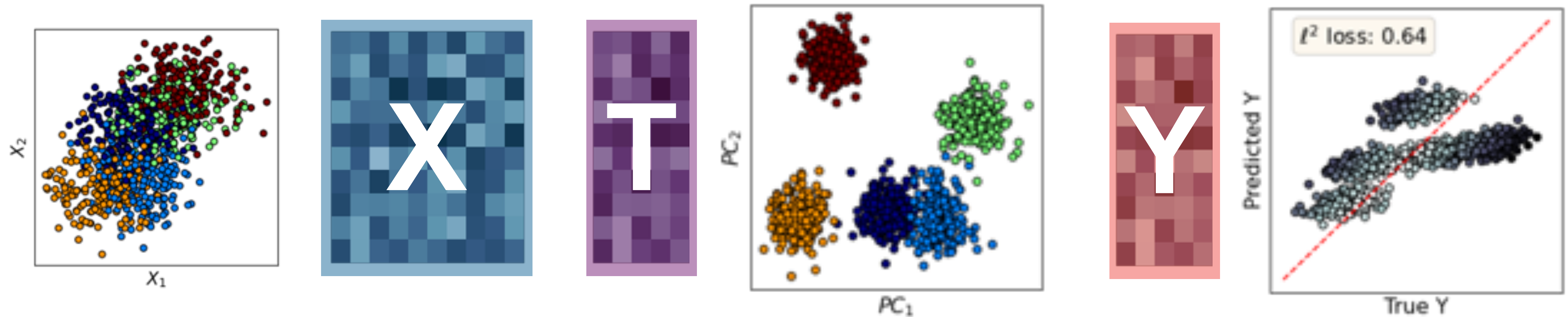
3



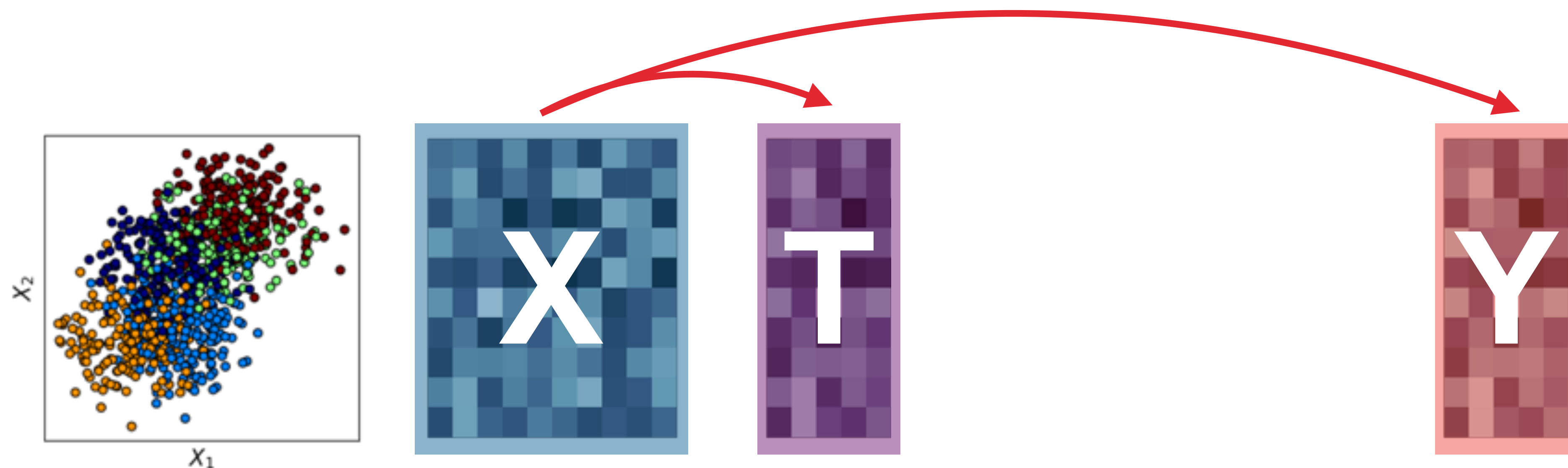
Principal Covariates Regression (PCovR)

From PCA to Principal Covariates Regression

3

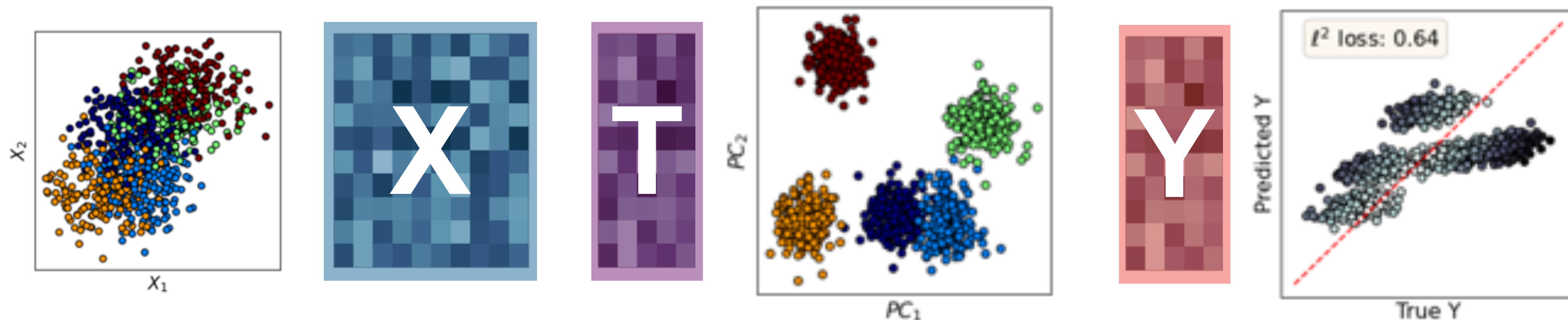


Principal Covariates Regression (PCovR)

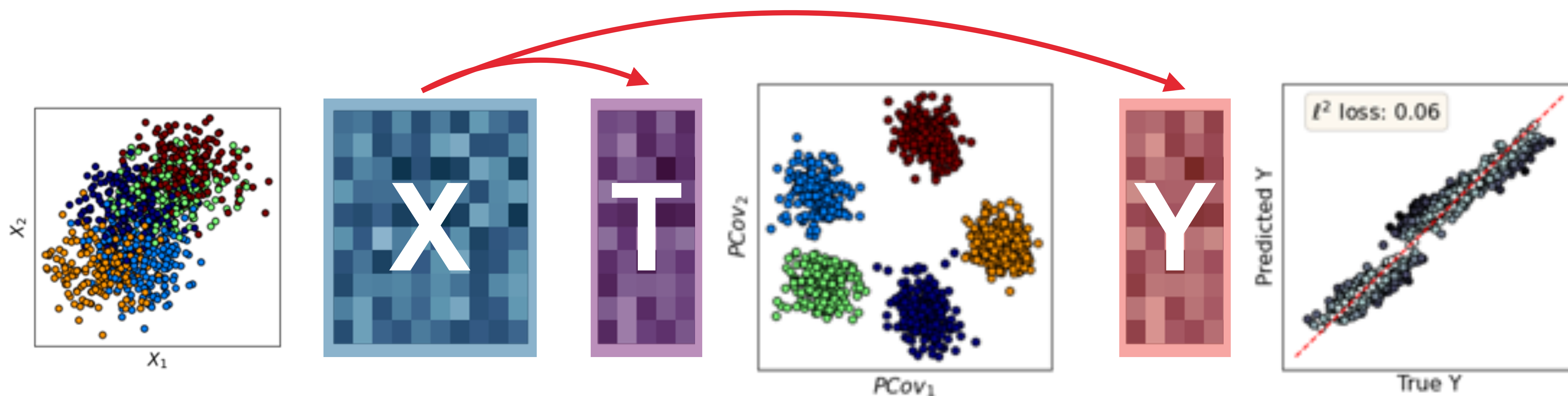


From PCA to Principal Covariates Regression

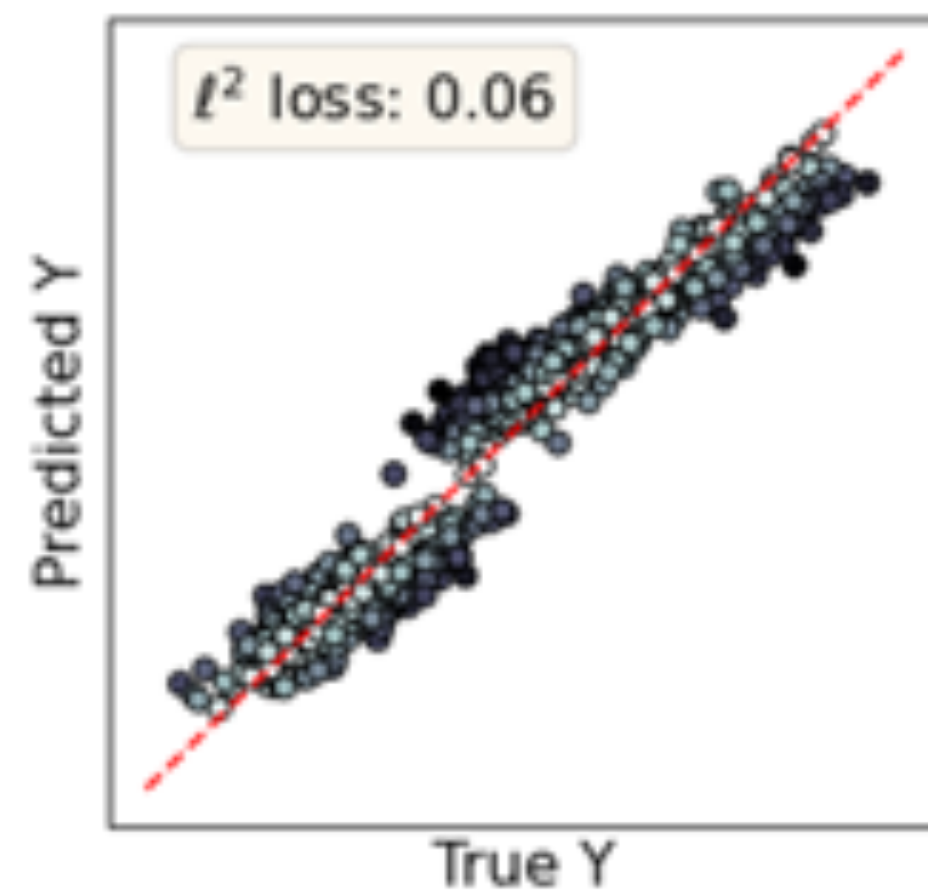
3



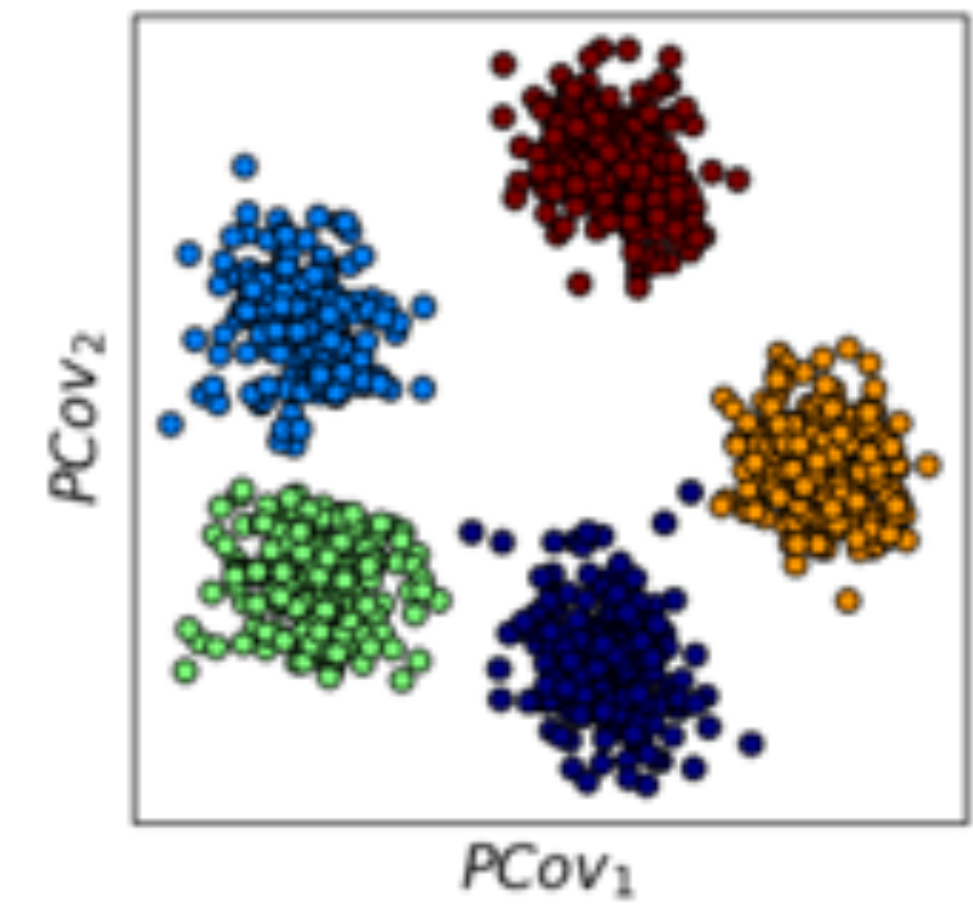
Principal Covariates Regression (PCovR)



$$\ell = (1 - \alpha) ||Y - XP_{XT}P_{TY}||^2 + \alpha ||X - XP_{XT}P_{TX}||^2$$

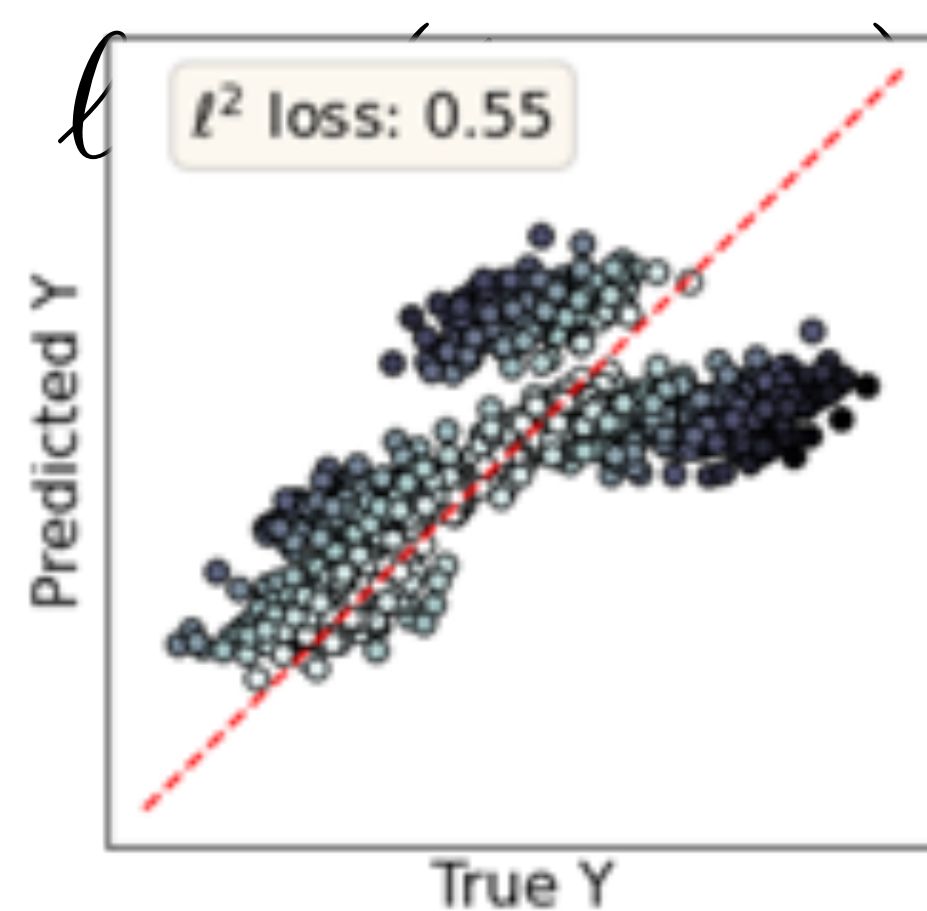


$$\alpha = 0.5$$



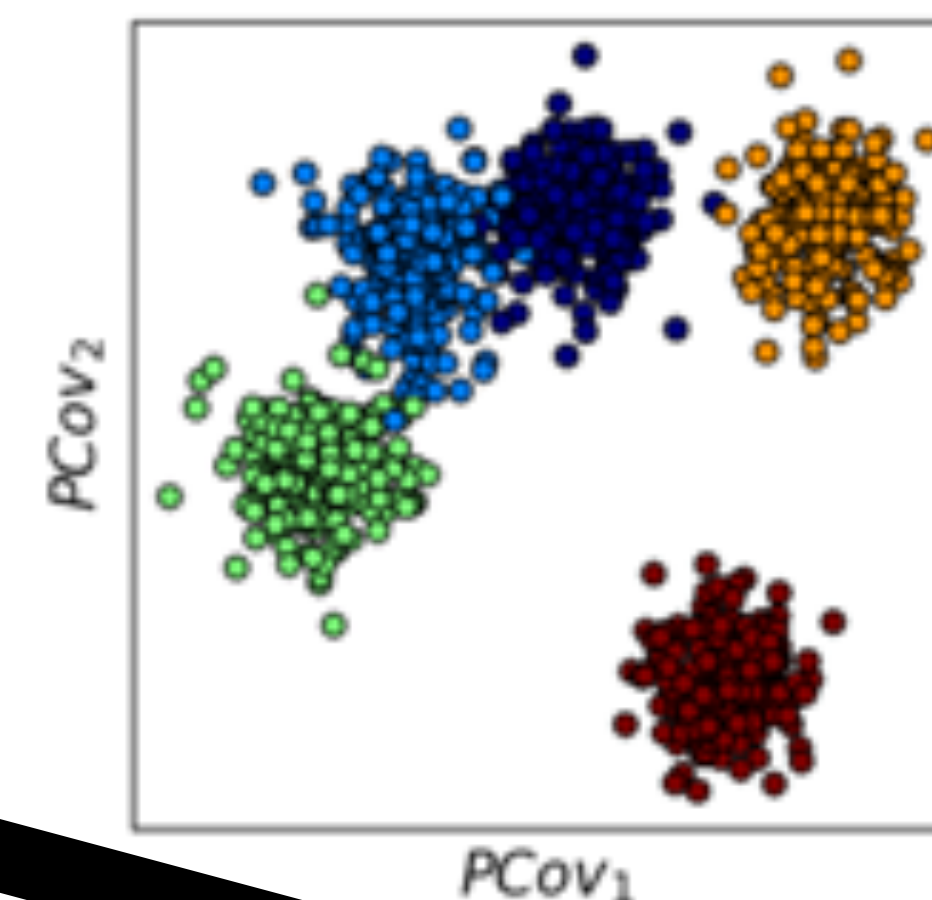
Linear Regression

PCA



$$||Y - XP_{XT}P_{TY}||^2 + \alpha ||X - XP_{XT}P_{TX}||^2$$

$$\alpha = 0.1$$



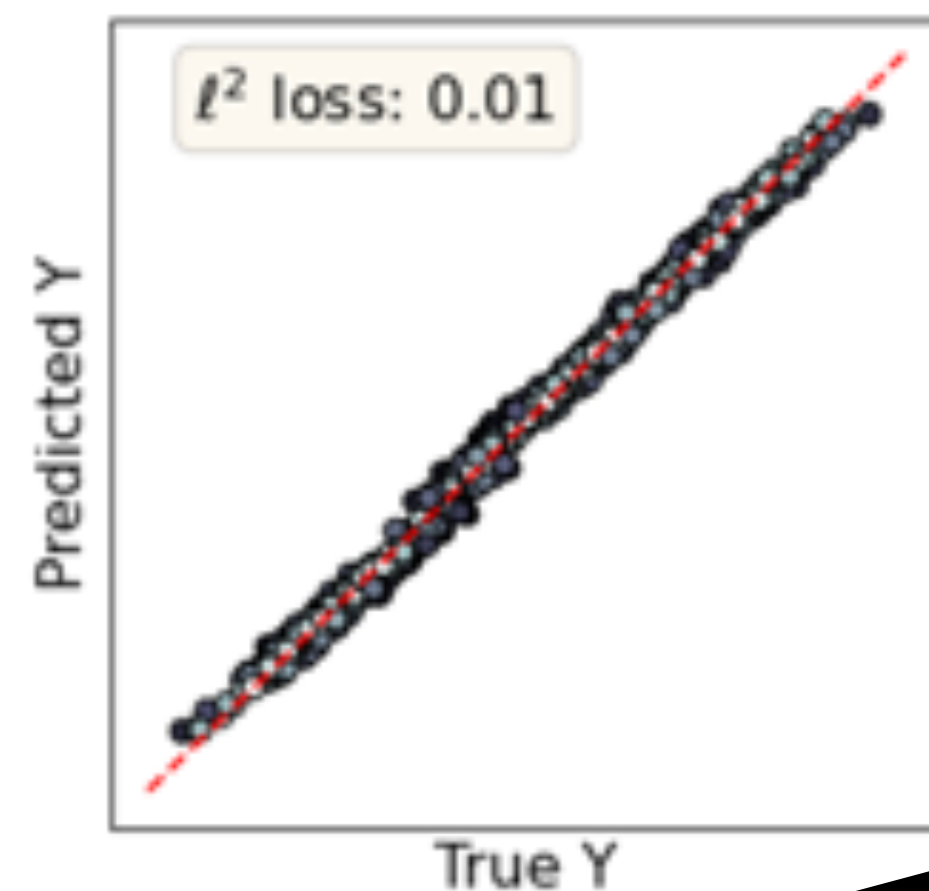
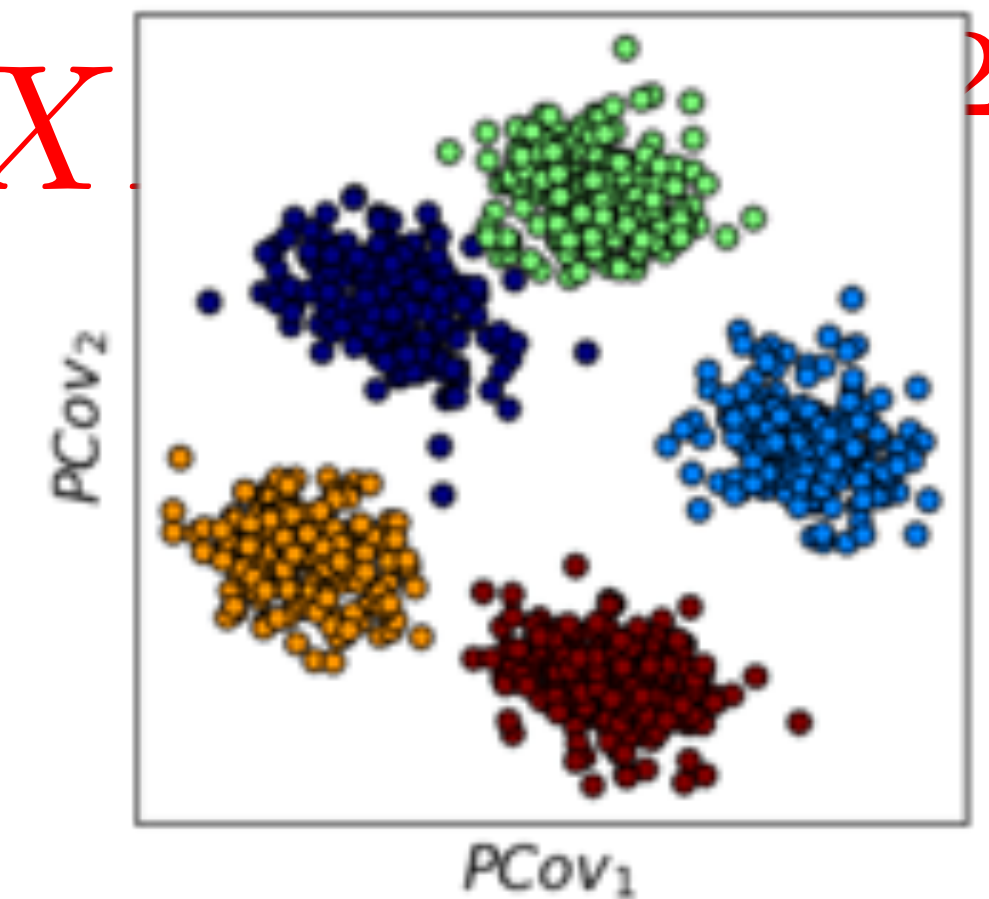
Linear Regression

PCA

Controlled mixing of PCA and linear regression

6

$$\ell = (1 - \alpha) ||Y - XP_X T P_{TY}||^2 + \alpha ||X - X$$



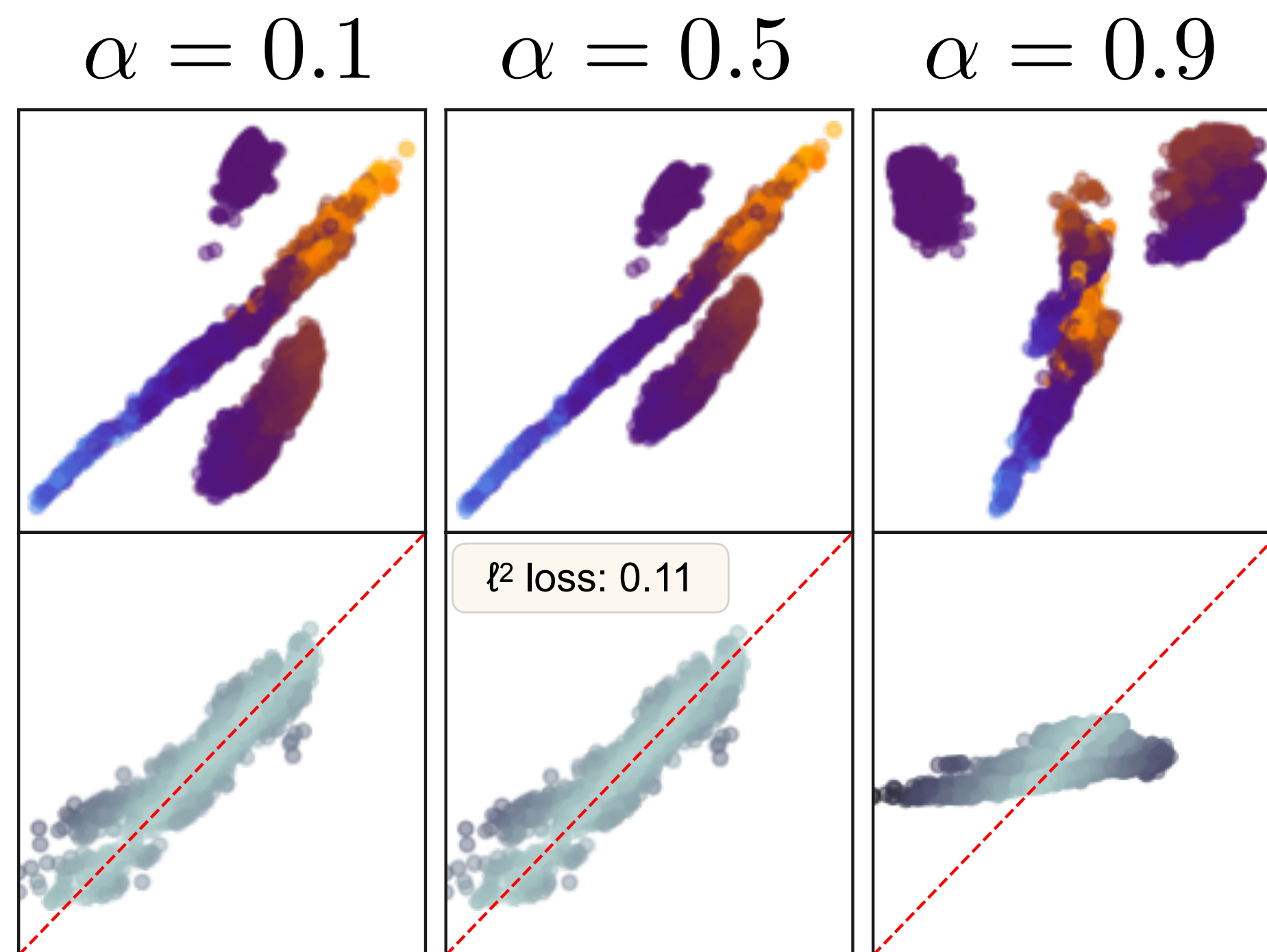
$$\alpha = 0.9$$

Linear Regression

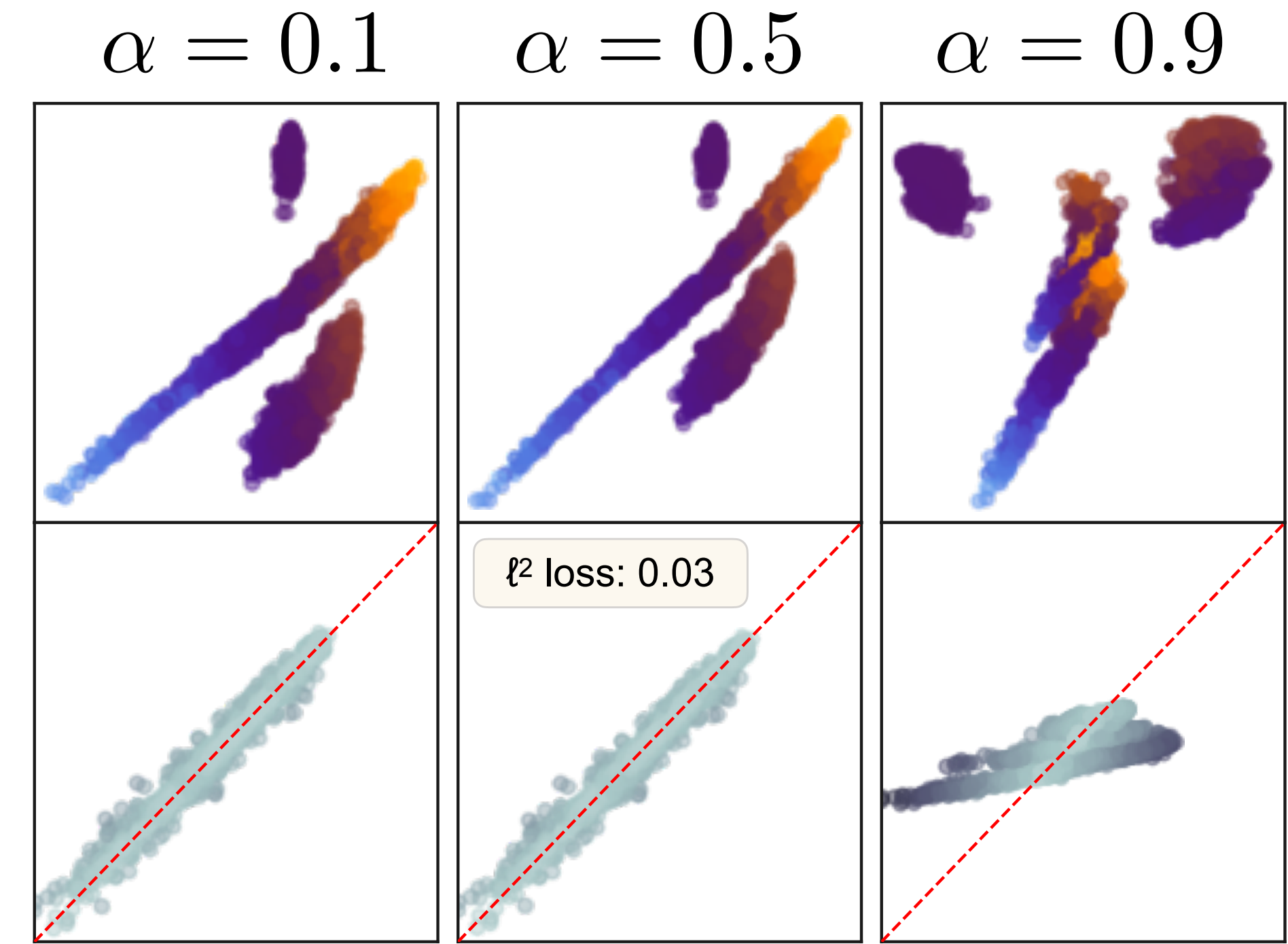
PCA

- ▶ What if we have non-linear correlations between features and properties?
- ▶ Use the "kernel trick": positive definite function define a reproducing kernel Hilbert space (RKHS) in which we can use linear operations

- ▶ What if we have non-linear correlations between features and properties?
- ▶ Use the "kernel trick": positive definite function define a reproducing kernel Hilbert space (RKHS) in which we can use linear operations



PCovR

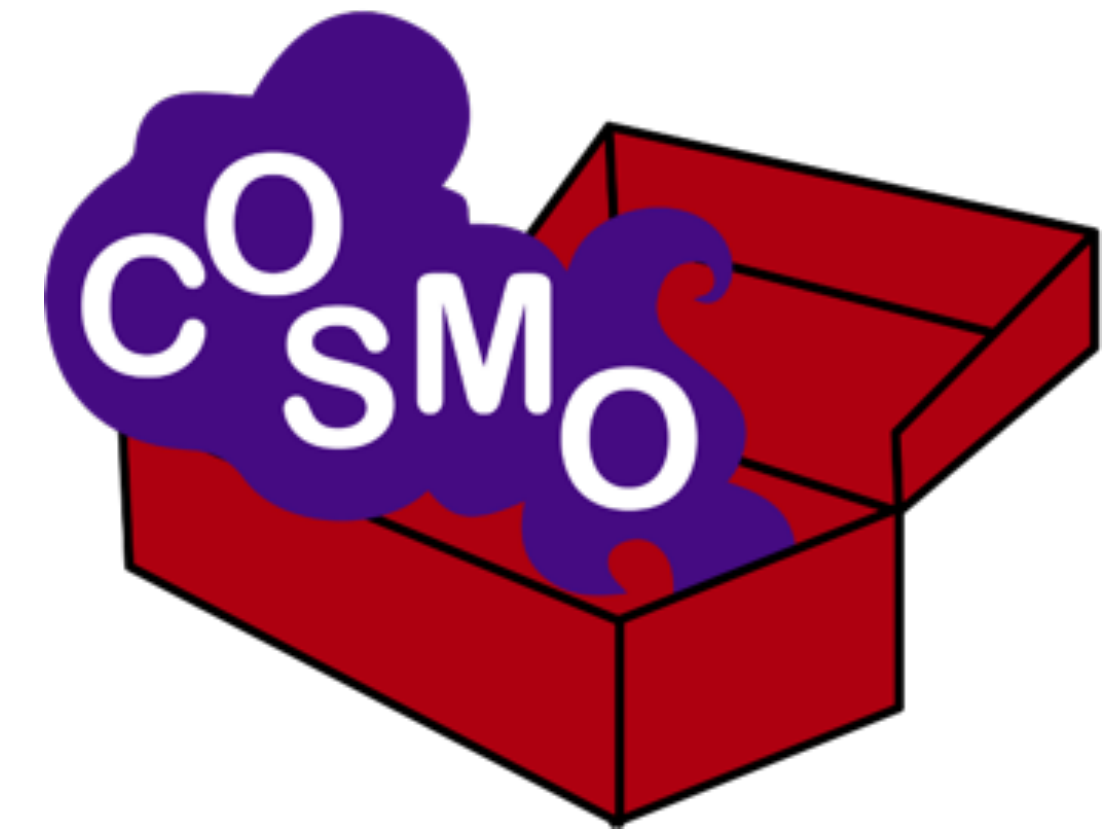


Kernel-PCovR

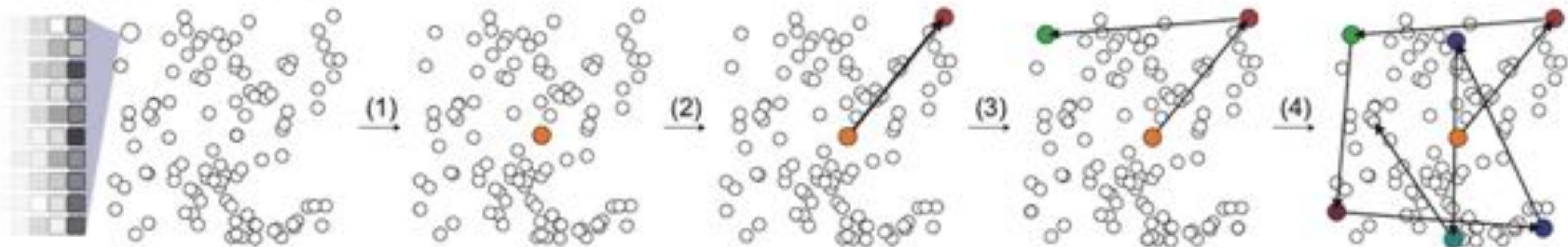
(K)-PCovR in practice with scikit-cosmo

8

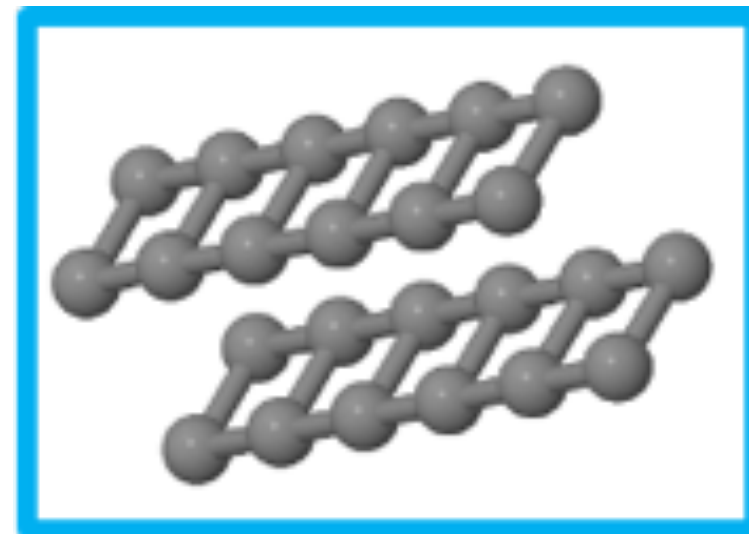
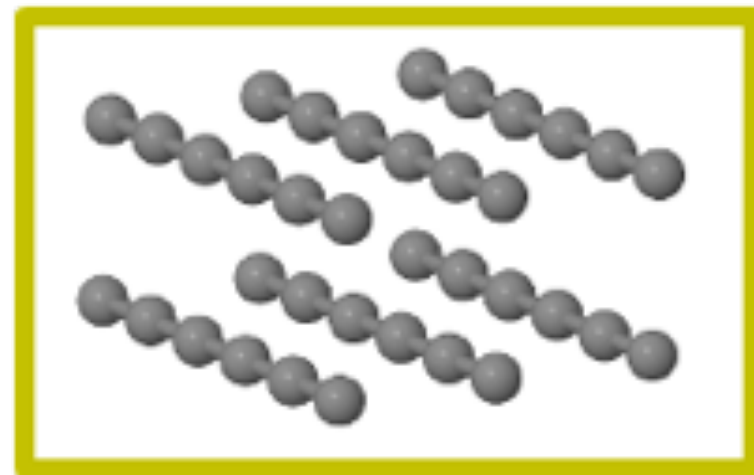
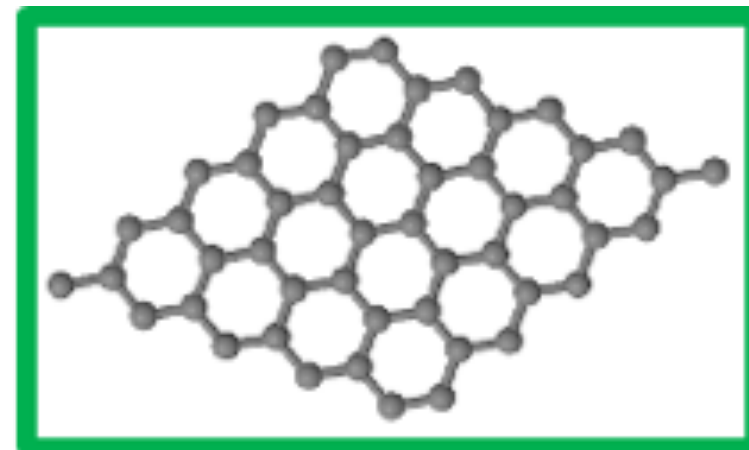
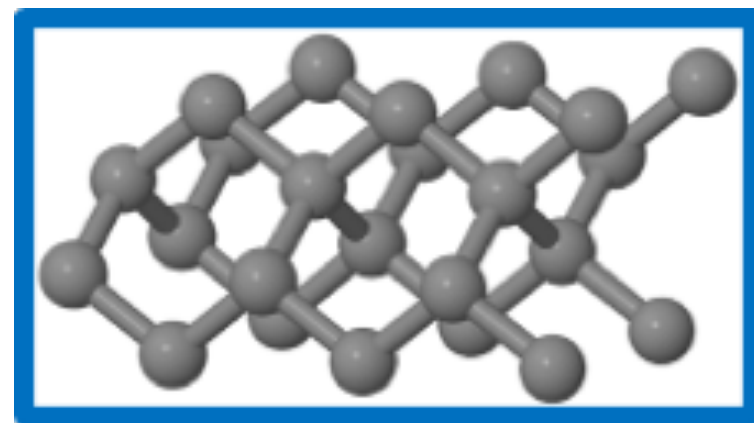
- ▶ Open source Python package implementing PCovR, Kernel-PCovR, and other methods developed in the COSMO lab
- ▶ Compatible with scikit-learn ecosystem
- ▶ <https://github.com/lab-cosmo/scikit-cosmo>
- ▶ Also: features and sample selections (FPS, CUR, PCov-FPS, PCov-CUR), feature space comparison (GFRE, LFRE, ...), and more machine learning tools!



scikit-cosmo



100k carbon structures at 10 GPa
generated with Ab-Initio Random
Structure Search

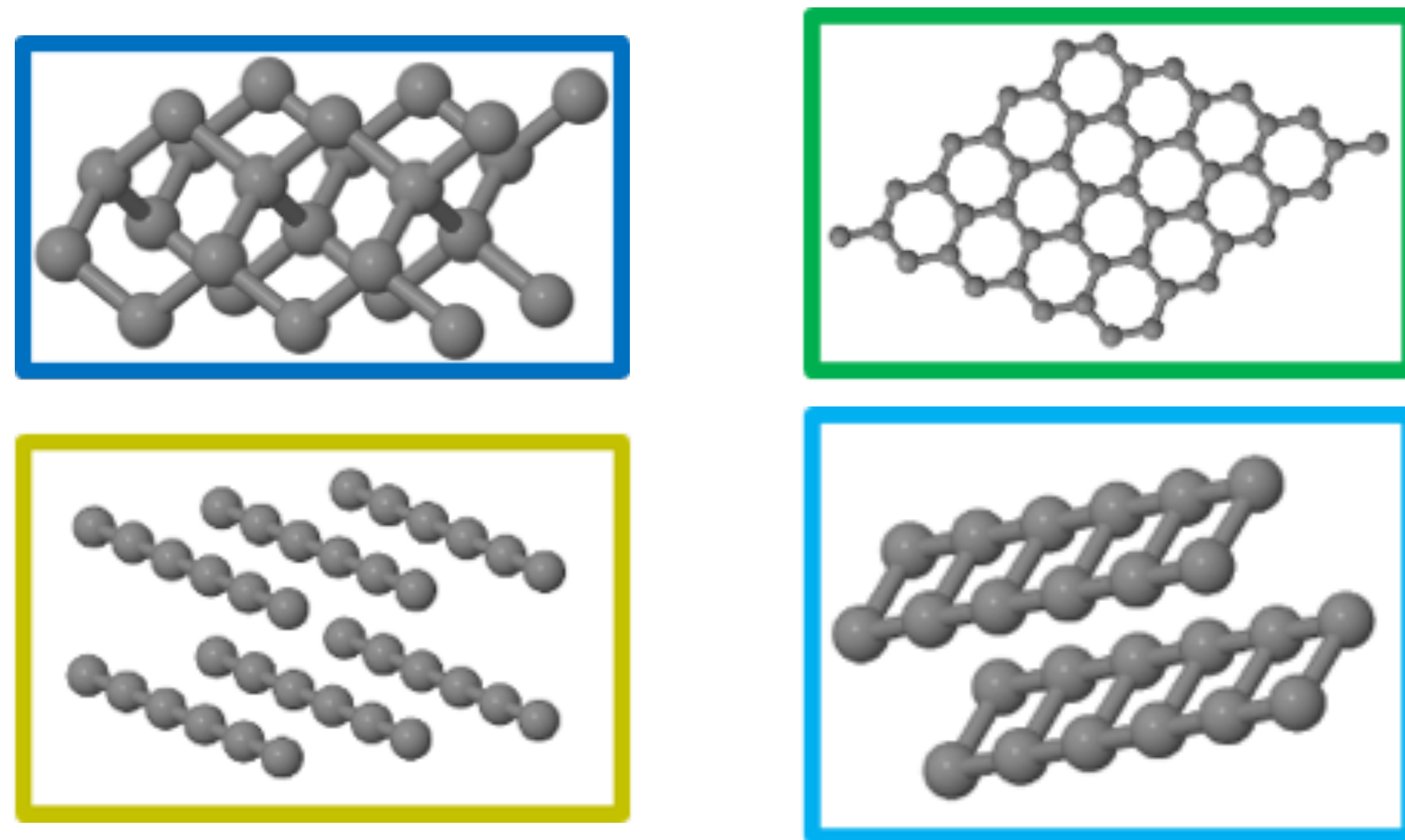


DFT energy computed with CASTEP

Chris J. Pickard

<https://doi.org/10.24435/materialscloud:2020.0026/v1>

100k carbon structures at 10 GPa
generated with Ab-Initio Random
Structure Search

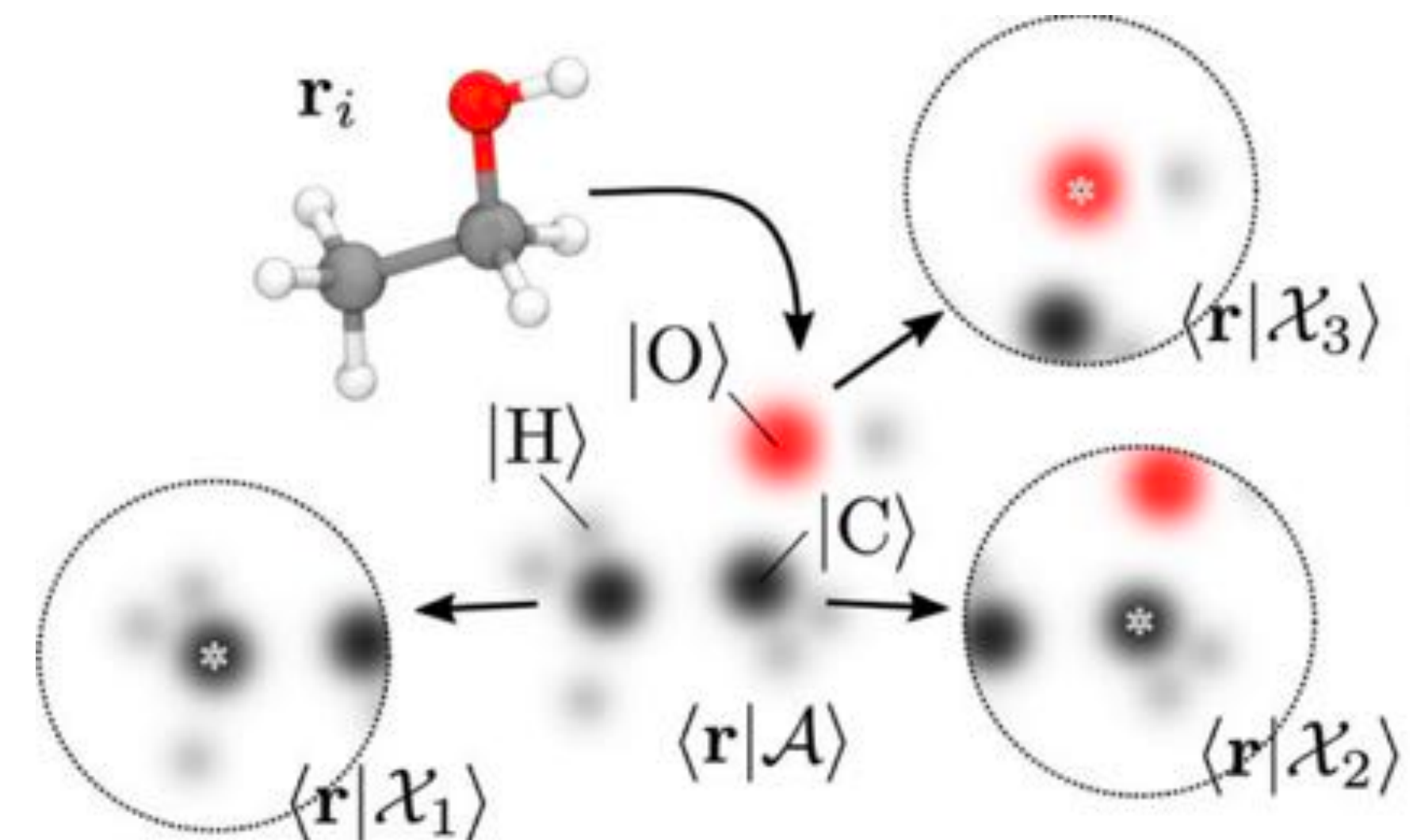


DFT energy computed with CASTEP

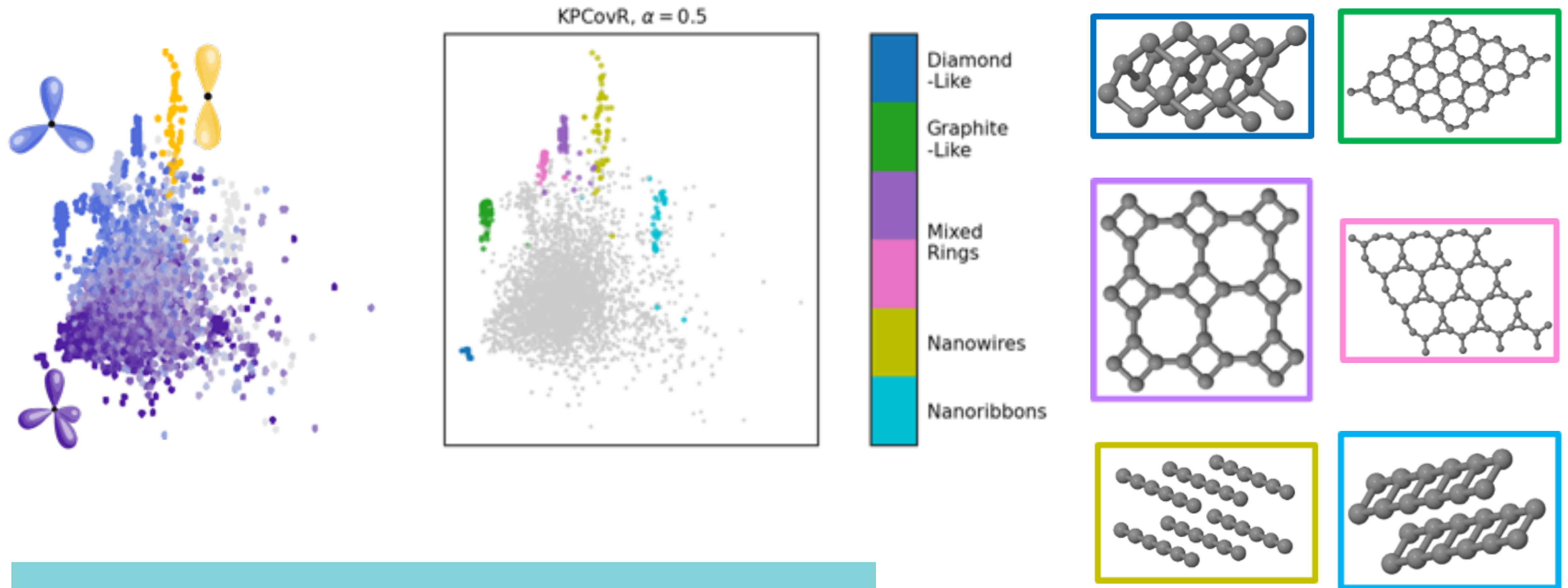
Chris J. Pickard

<https://doi.org/10.24435/materialscloud:2020.0026/v1>

SOAP power spectrum
structural representation



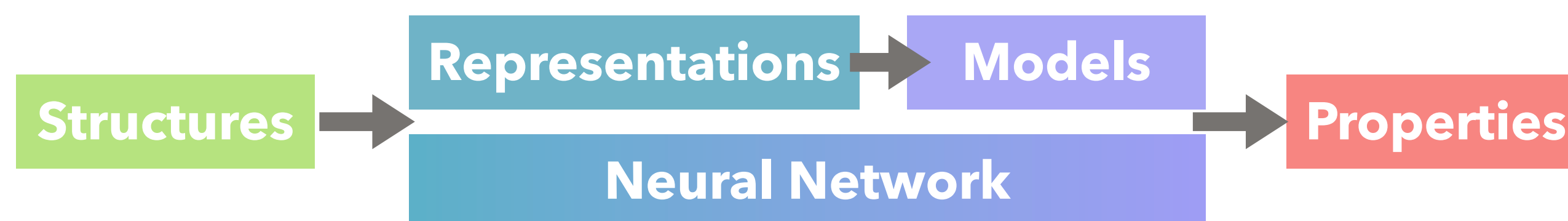
3-body correlations of the atomic
density around a given atomic center



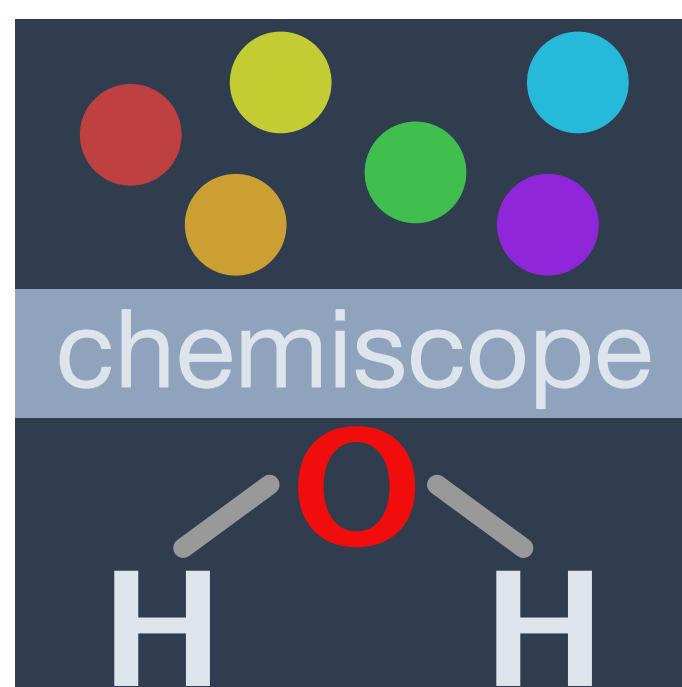
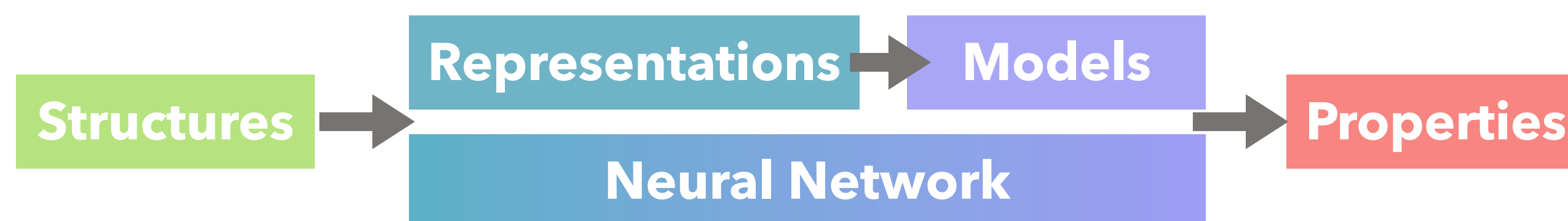
Clustering both by global and local chemical structure

B. Helfrecht, R. Cersonsky, G. Fraux, M. Ceriotti
Structure-property maps with Kernel principal covariates regression
Mach. Learn. Sci. Techno. **1**, p45021 (2020)

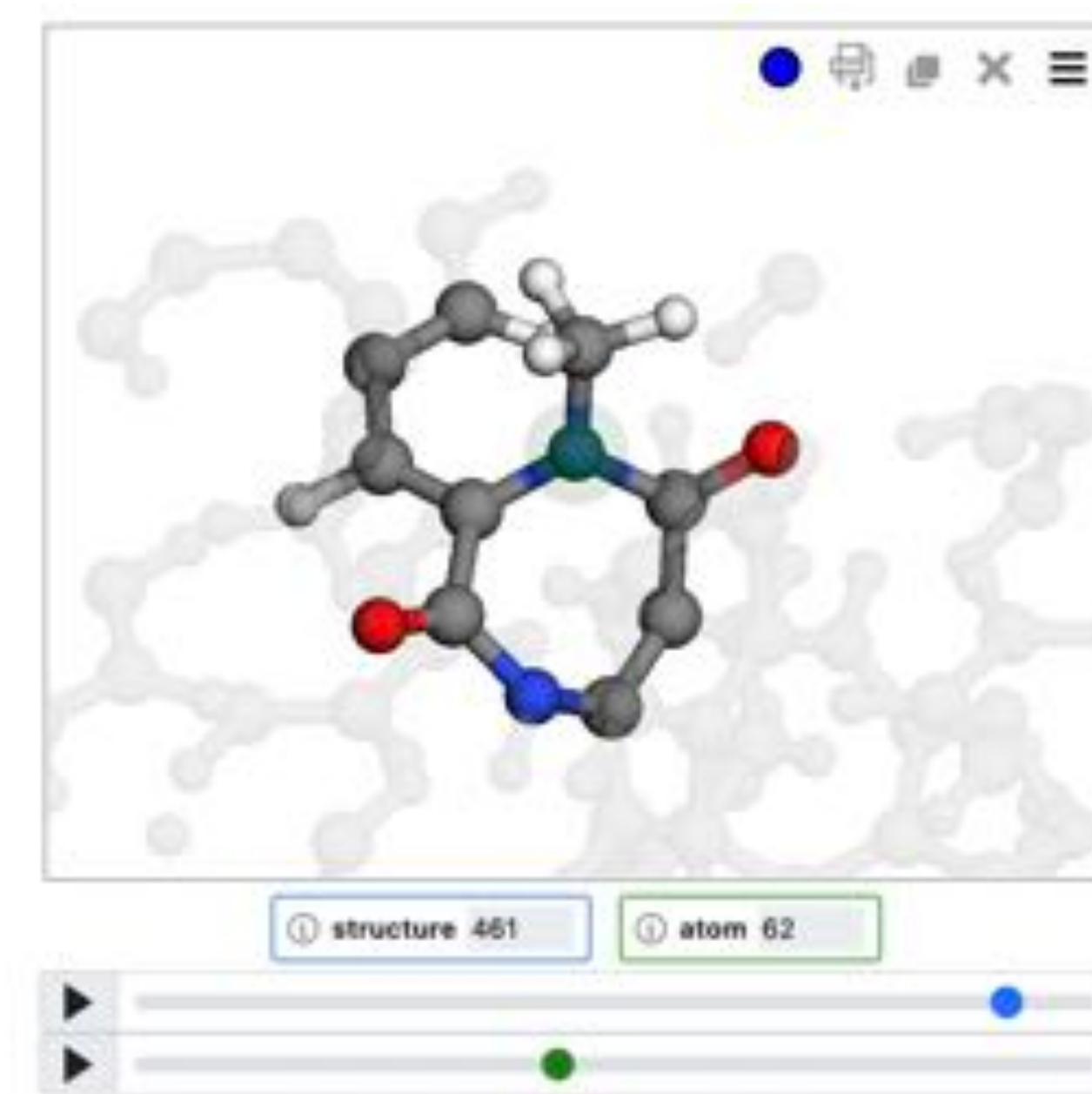
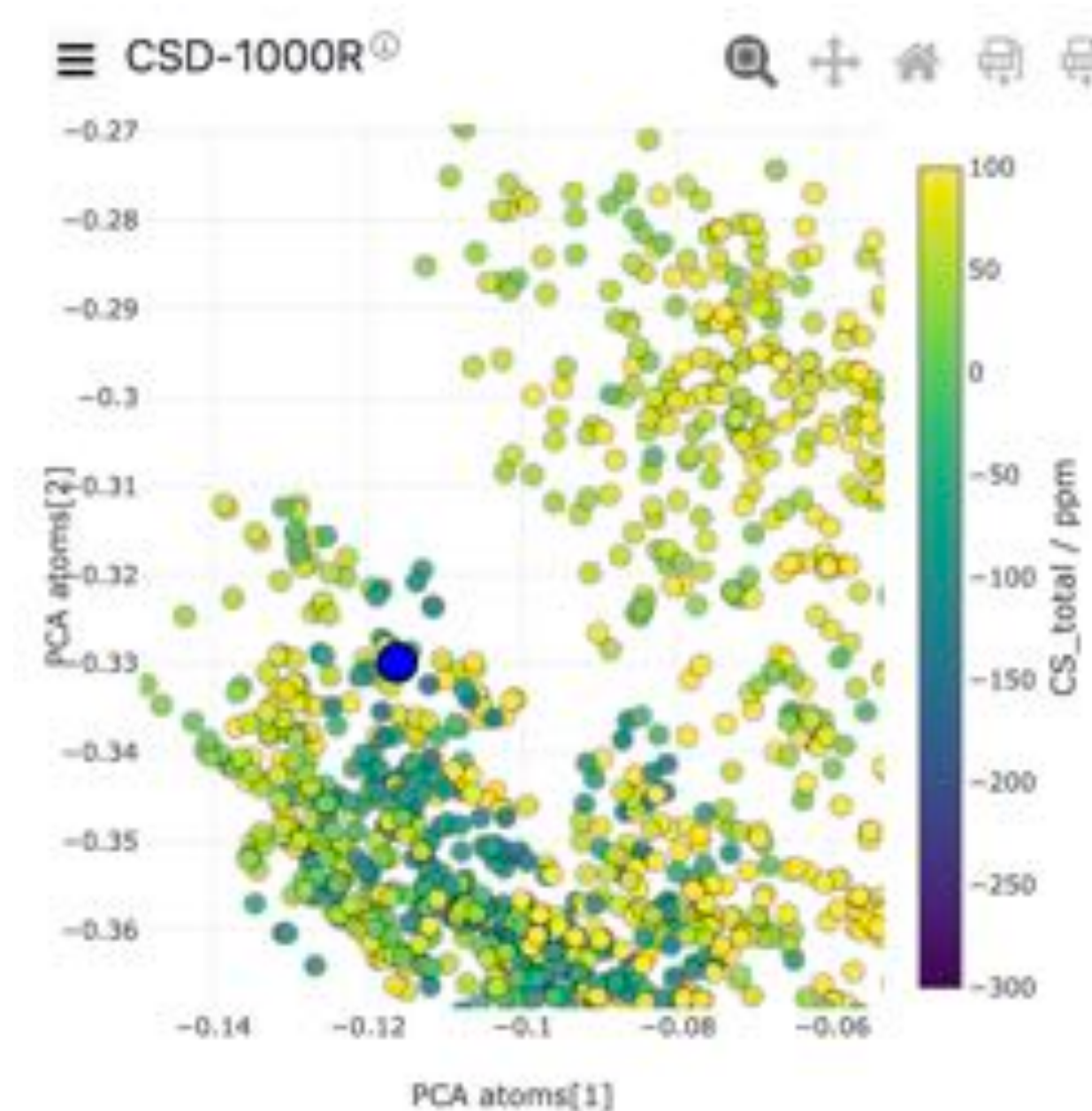
- ▶ Now that we have these large datasets, how can we rationalise the clustering behaviour, and structure-property correlations?
- ▶ Simultaneously explore the latent space and structure space



- ▶ Now that we have these large datasets, how can we rationalise the clustering behaviour, and structure-property correlations?
- ▶ Simultaneously explore the latent space and structure space



<https://chemiscope.org>



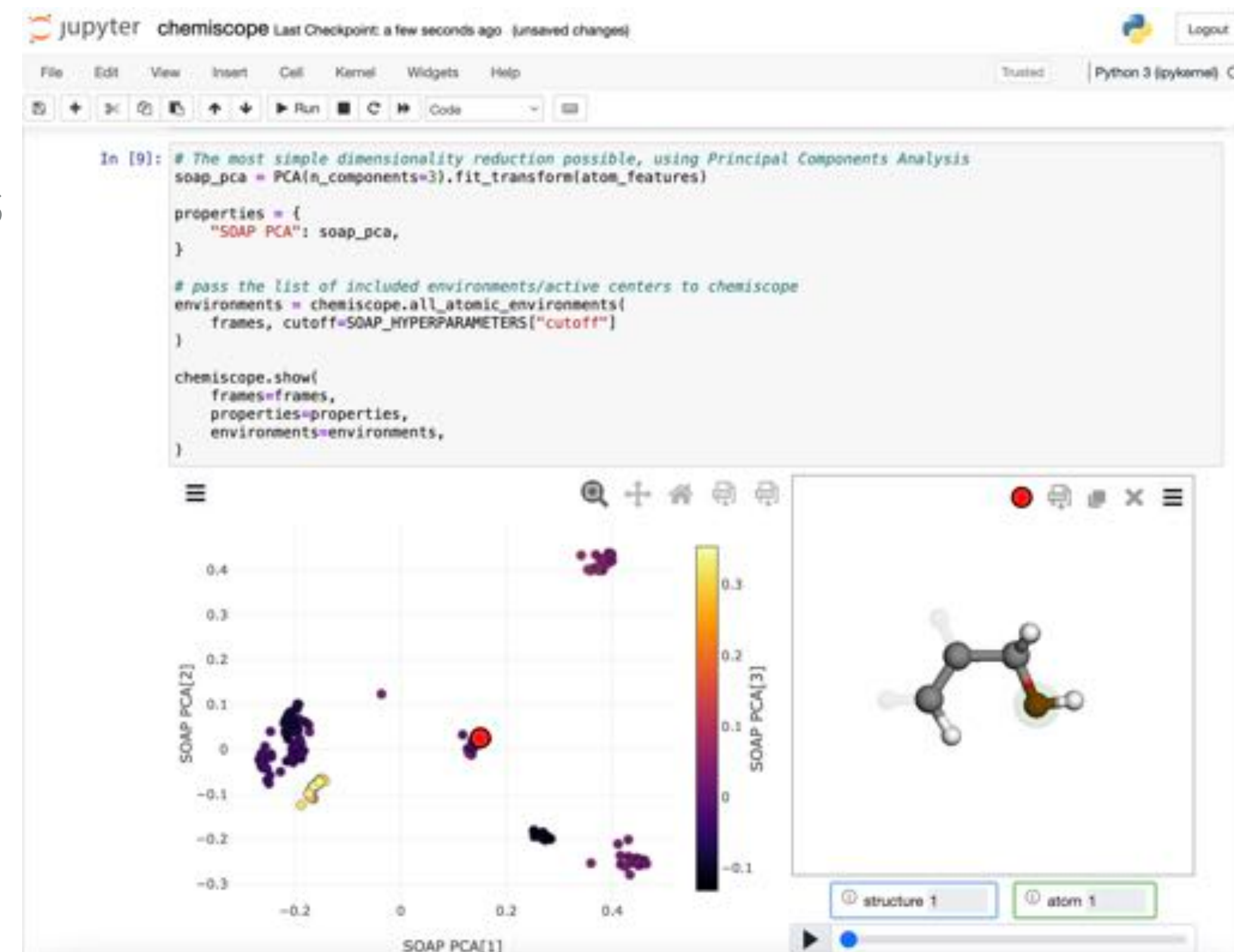
<https://chemiscope.org>

<https://chemiscope.org>

- ▶ Supports 100k data points without issues, up to 1M points in 2D mode
- ▶ Runs in web browser, no installation required, additional python tools to help generate input files
- ▶ Fully open source!
<https://github.com/lab-cosmo/chemiscope>

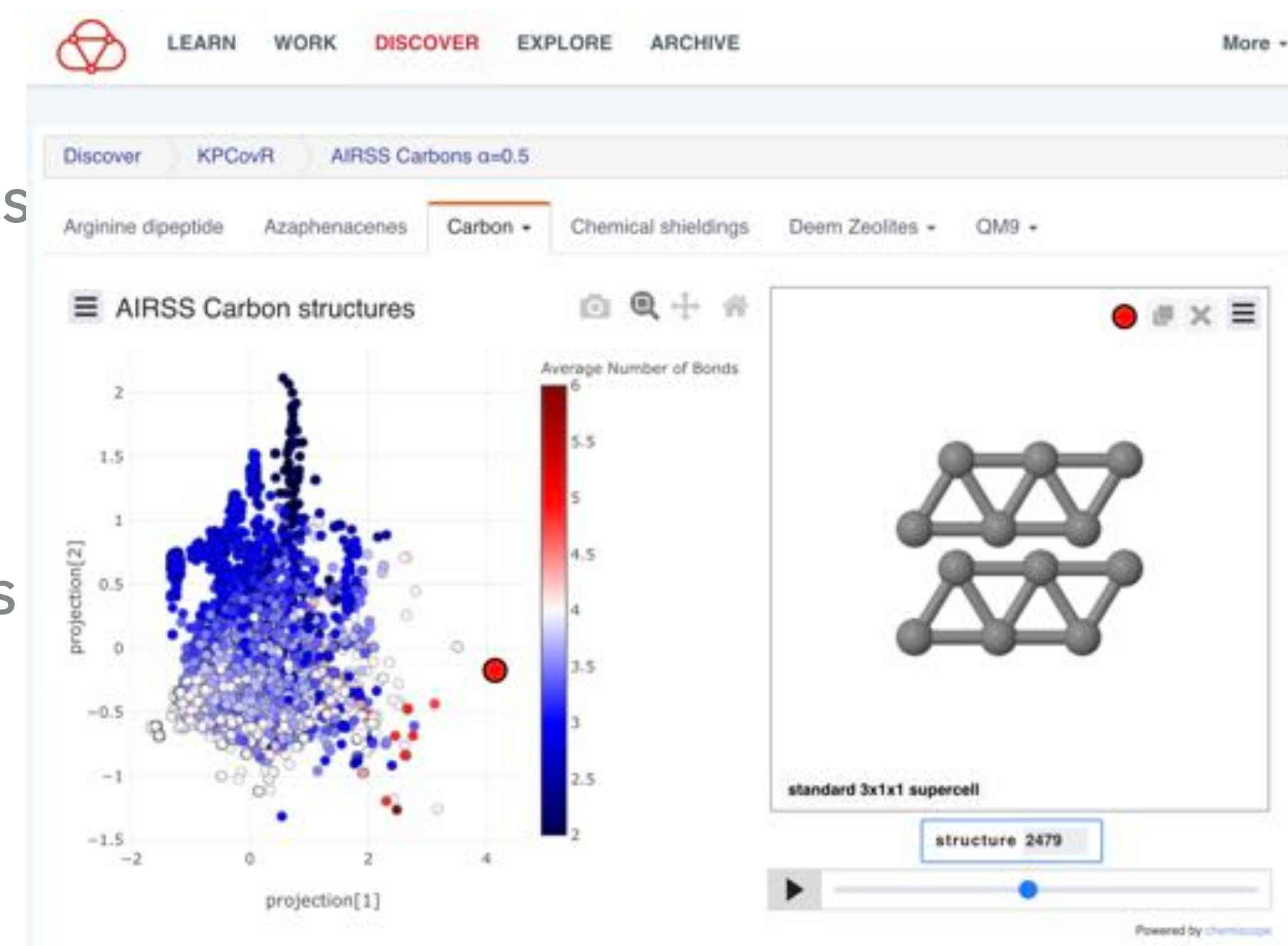
<https://chemiscope.org>

- ▶ Supports 100k data points without issues, up to 1M points in 2D mode
- ▶ Runs in web browser, no installation required, additional python tools to help generate input files
- ▶ Fully open source!
<https://github.com/lab-cosmo/chemiscope>
- ▶ Integration with Jupyter notebooks and workflows



<https://chemiscope.org>

- ▶ Supports 100k data points without issues, up to 1M points in 2D mode
- ▶ Runs in web browser, no installation required, additional python tools to help generate input files
- ▶ Fully open source!
<https://github.com/lab-cosmo/chemiscope>
- ▶ Integration with Jupyter notebooks and workflows
- ▶ Easy to share visualisations with collaborators (email files, host your own files, integrated with materialscloud)



Give it a try!

scikit-cosmo

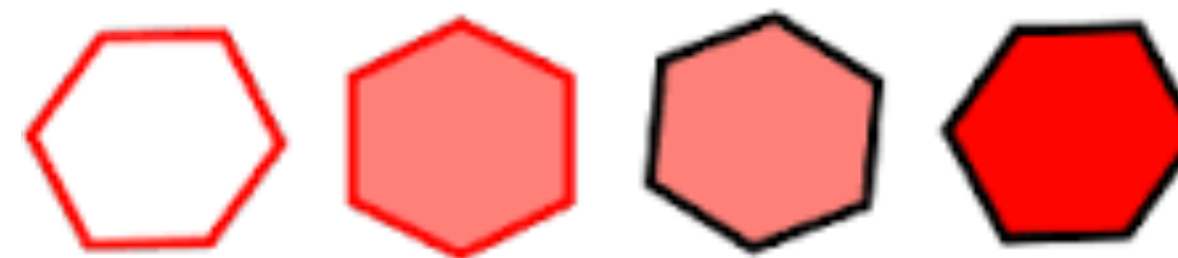
<https://chemiscope.org>



B7.26: Dr. Rose Cersonsky, *Identifying high-stability motifs of structural patterns in molecular crystals*

```
pip install skcosmo chemiscope
```

MARVEL



Come and see us at
the MARVEL booth!

- ▶ Tuesday 12:30
- ▶ Thursday 14:00

these slides: <https://tinyurl.com/psik22-fraux>

Kernel PCovR article: <https://tinyurl.com/kernel-pcovr>

Give it a try!

scikit-cosmo

<https://chemiscope.org>



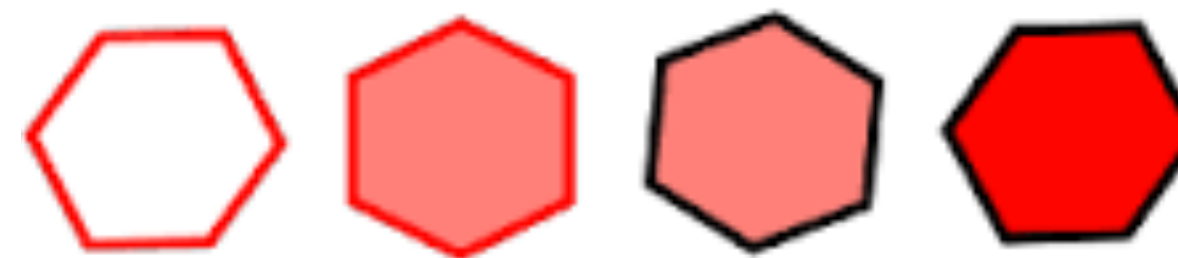
B7.26: Dr. Rose Cersonsky, *Identifying high-stability motifs of structural patterns in molecular crystals*

```
pip install skcosmo chemiscope
```

Thank you for your
attention!

MARVEL

Come and see us at
the MARVEL booth!



- ▶ Tuesday 12:30
- ▶ Thursday 14:00

these slides: <https://tinyurl.com/psik22-fraux>

Kernel PCovR article: <https://tinyurl.com/kernel-pcovr>