

Advances in Machine Learning in physics

```
data = source.fetch()  
if count > 100:  
if count > 100: pass
```

```
data = source.fetch()  
if count > 100:  
pass
```

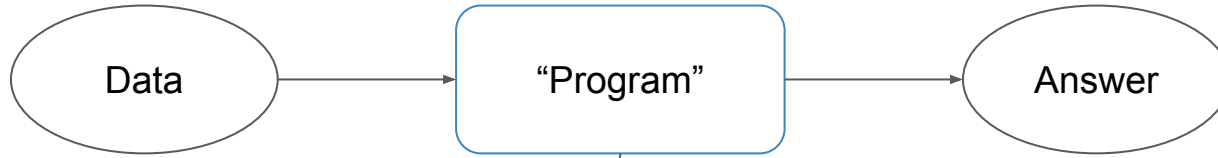
The work is supported by RSF grant 24-72-10056

→ Dear AI,

Please, make the Theory of Everything.
Here is all of the data in the world.

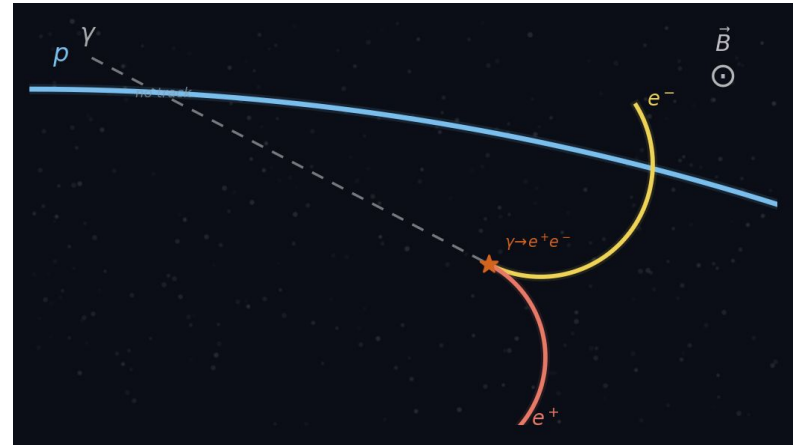
1. The idea of machine learning

What is machine learning

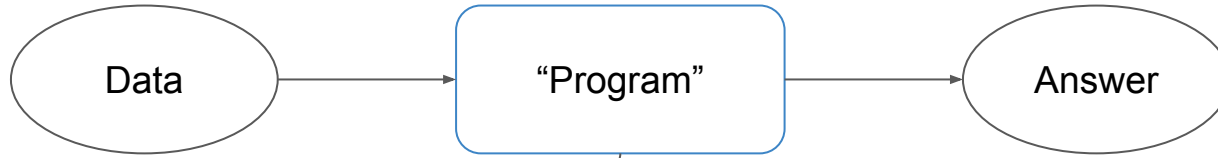


Standard algorithms:

“Program” - **fixed**, written by human **algorithm**.



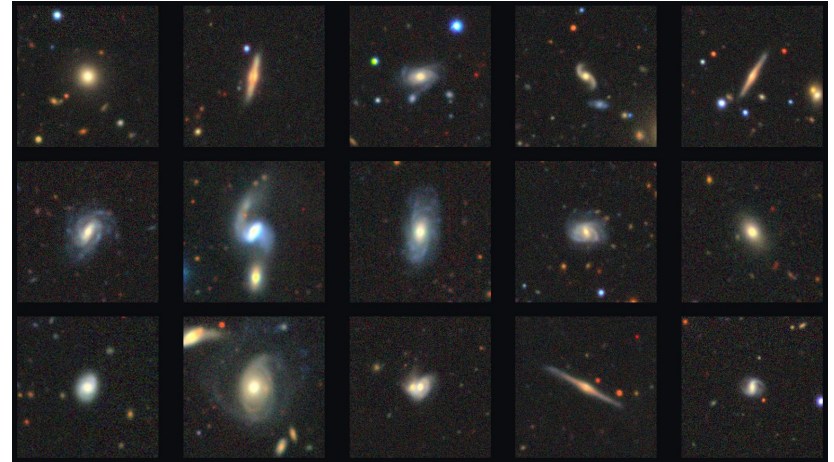
What is machine learning



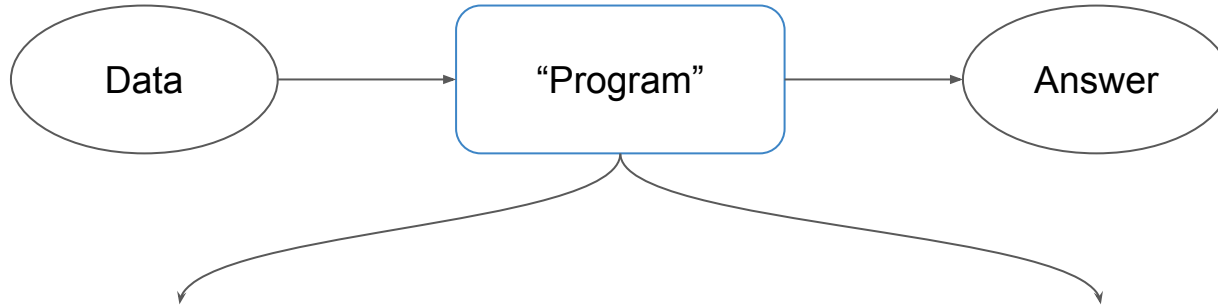
Standard algorithms:

“Program” - **fixed**, written by human **algorithm**.

How to determine and extract useful features?



What is machine learning



Standard algorithms:

“Program” - **fixed**, written by human **algorithm**.

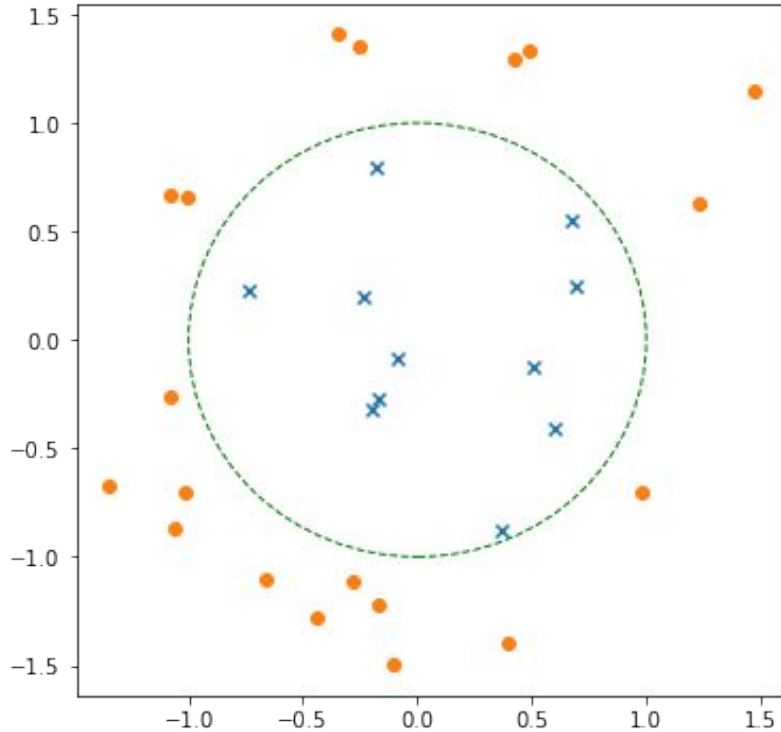
Manually created algorithms

Machine learning:

“Program” - **learning on examples** algorithm of extracting and parsing optimal features.

Programs, creating optimal algorithms.

Toy example

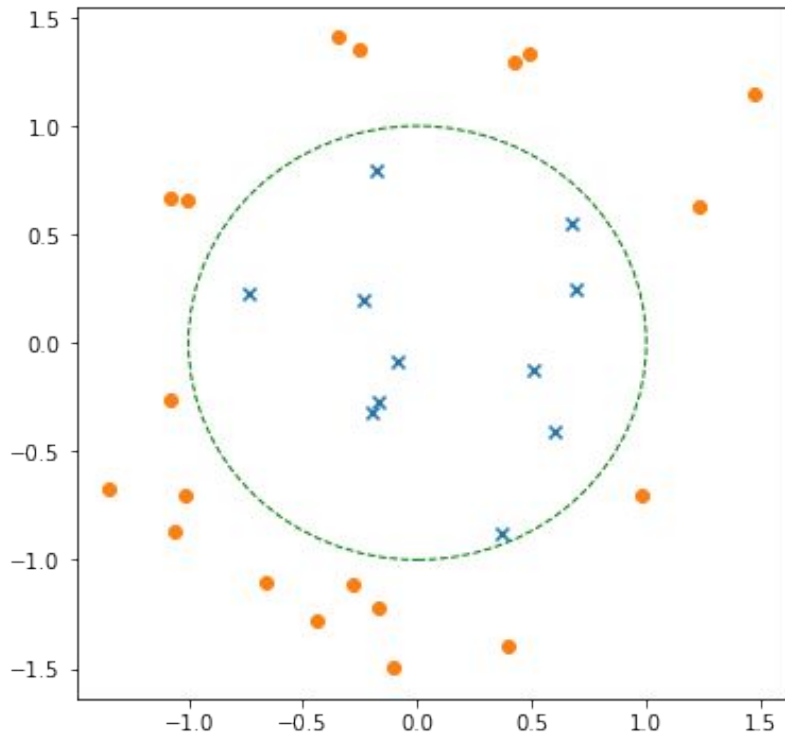


Data: points on 2D, (x_1, x_2)

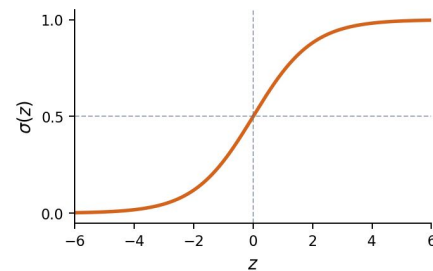
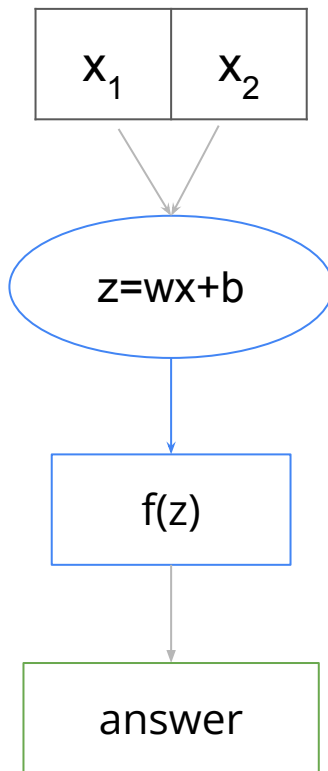
Answer: 0 - outside, 1 - inside

Program: how to automatically discover the rule from data?

Toy example



The Program

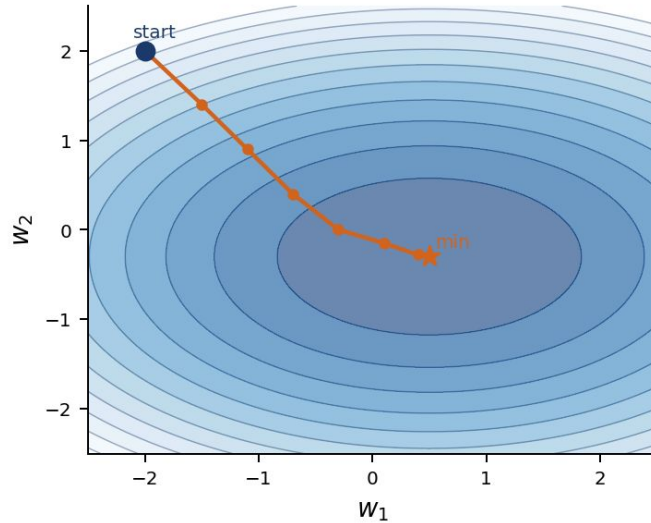


$P(\text{inside} | x)$

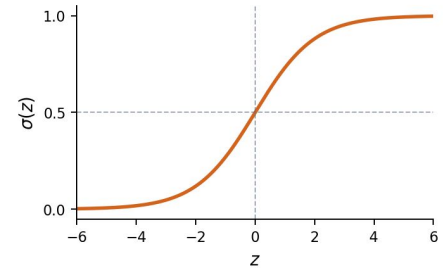
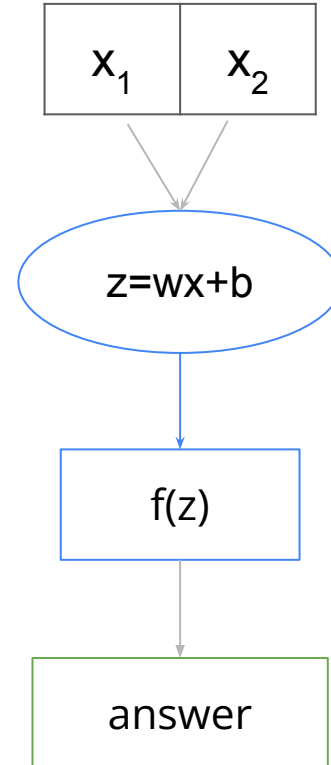
Toy example

Loss function:

$$\text{Loss} = | \text{prediction}(w,b) - \text{truth} | \geq 0$$

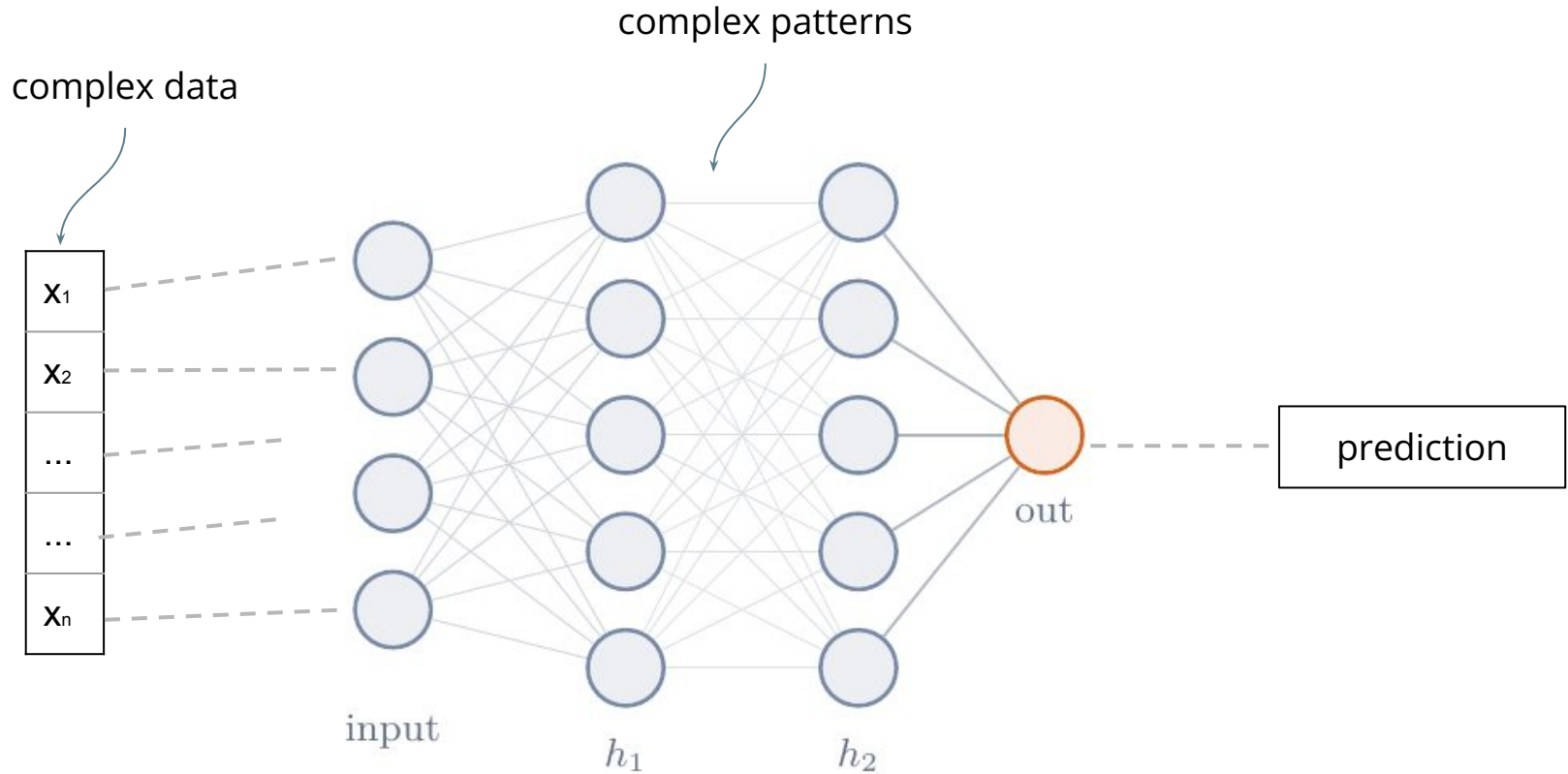


The Program



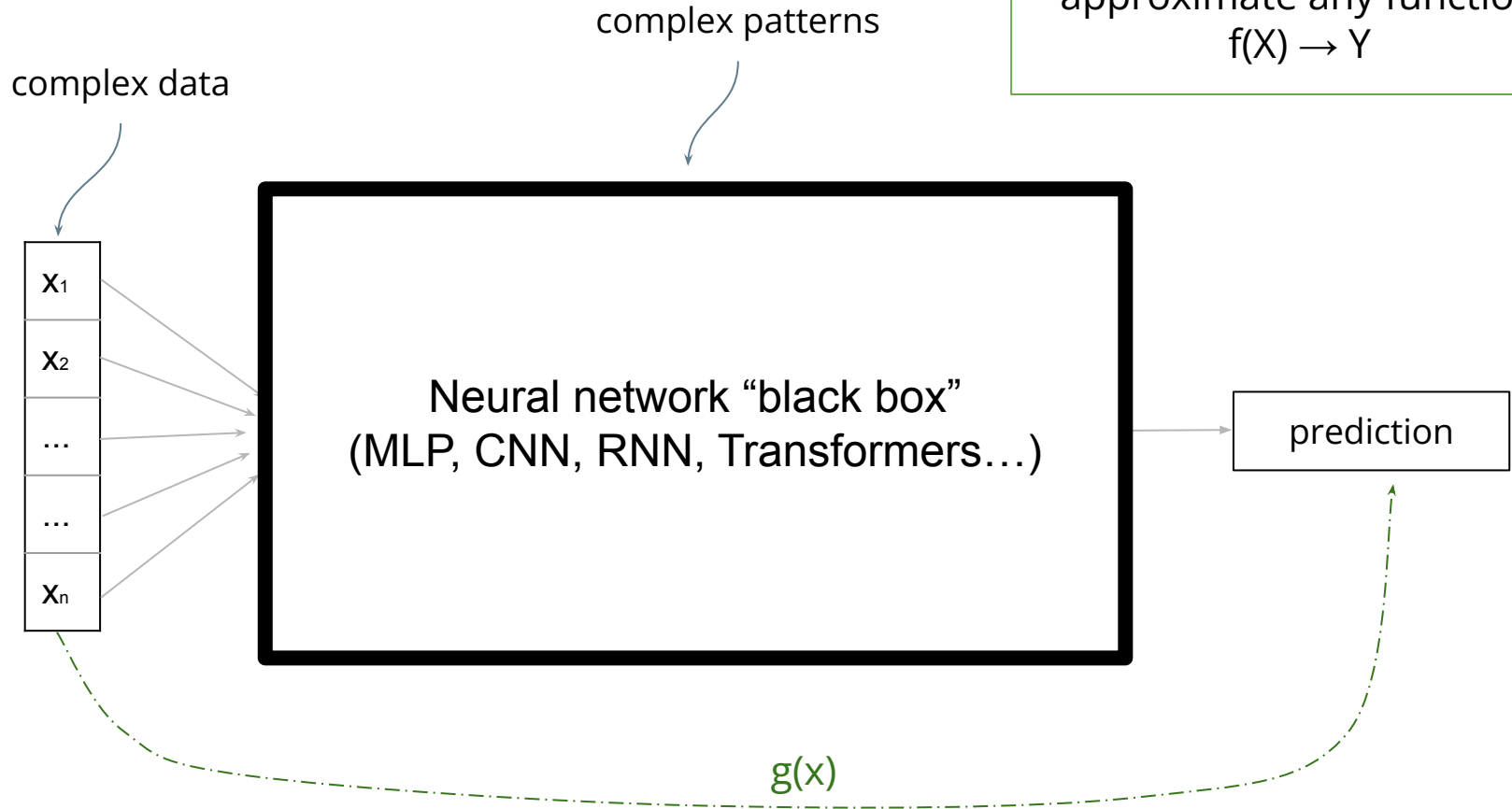
$P(\text{inside} | x)$

The power of machine learning



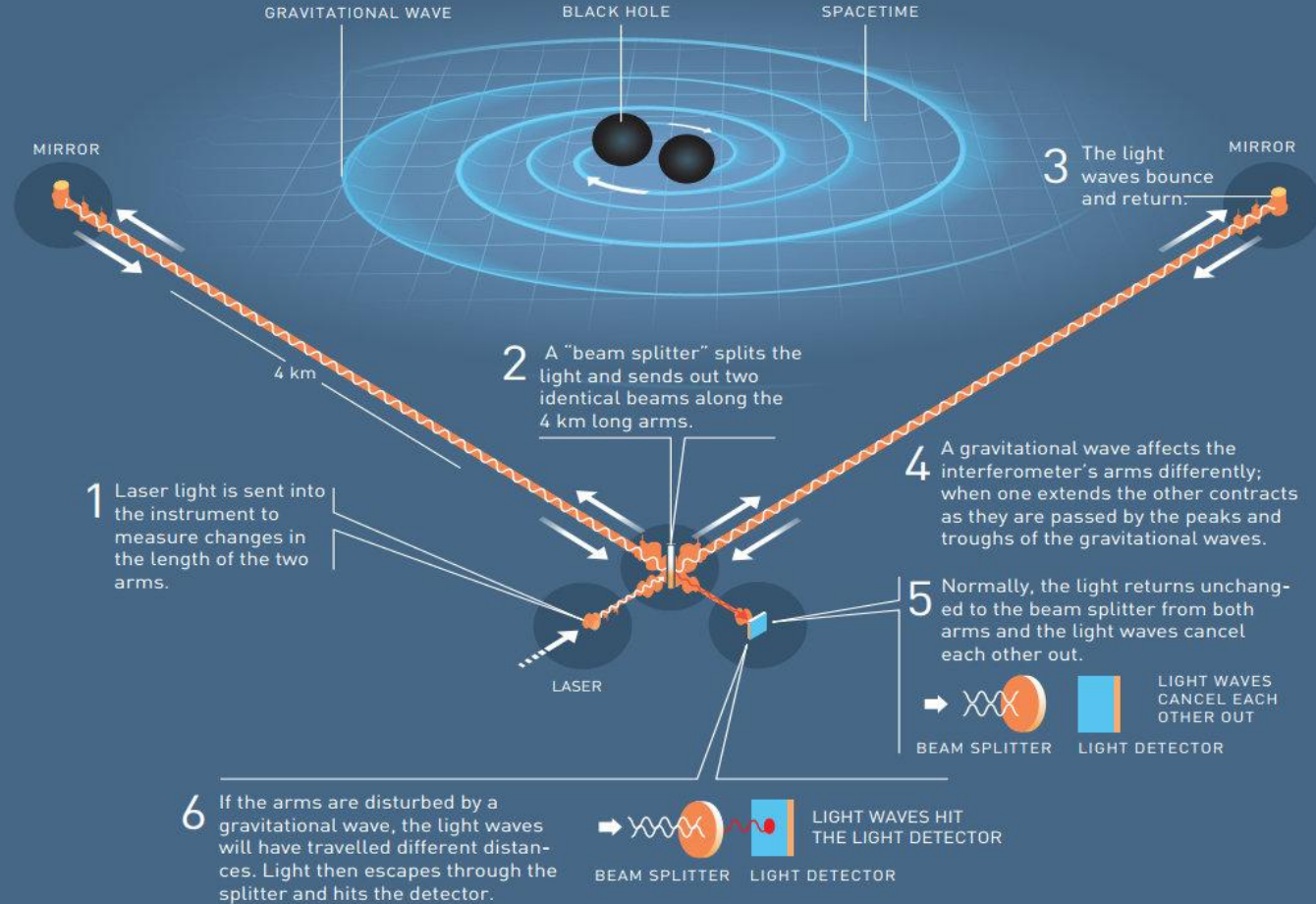
The power of machine learning

Neural network can approximate any function
 $f(X) \rightarrow Y$

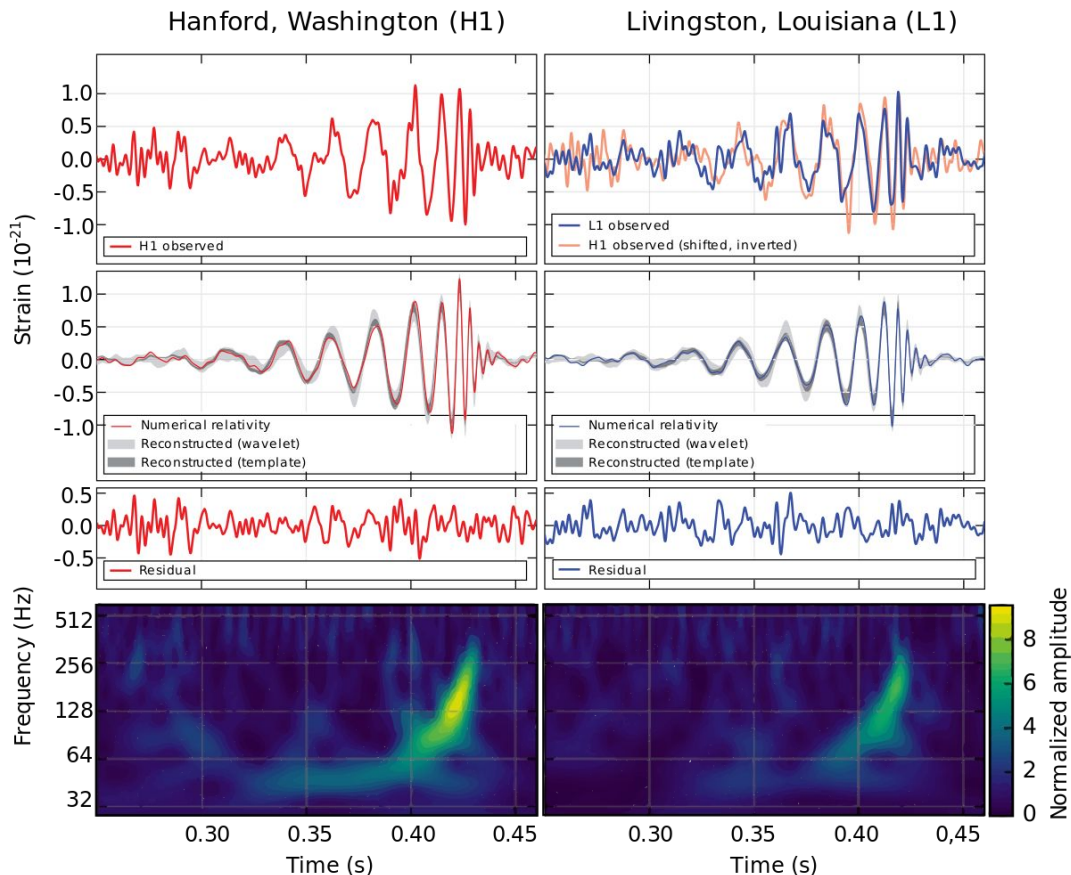


2. ML for automatization

LIGO – A GIGANTIC INTERFEROMETER



Search for gravitational waves



Processing registered signals
takes a while

2308.04545:

Neural network based analysis,
pre-merger alerts (-3.1 s)

2603.26405:

Full posterior estimate

Comparable or better accuracy,
Much faster data analysis

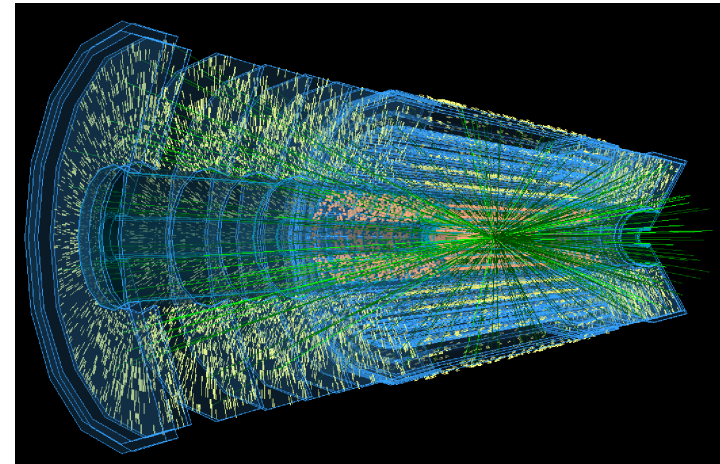
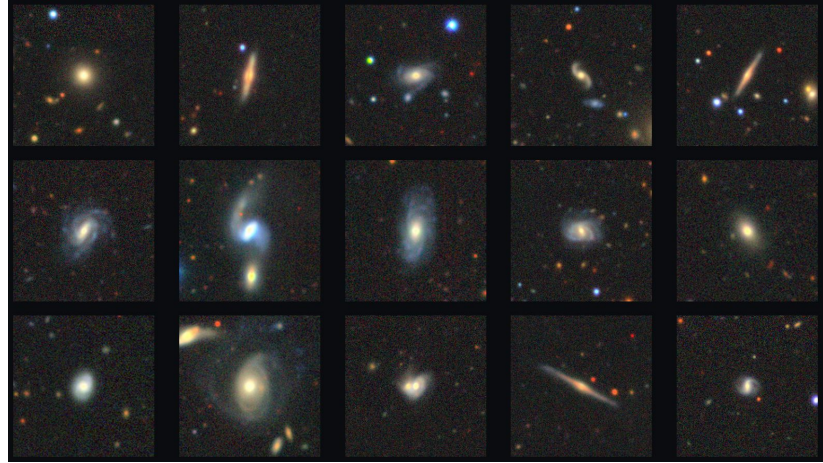
Big Data Era

Large amount of data:
Planck, JWST, Euclid, LHC

We need to:

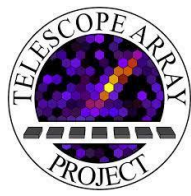
- Process the data
(identify and classify galaxies)
- Select candidates for details studies
(detailed spectrography)

Neural networks can **automate** this processing.
Moreover, they yield **better photometric** redshift estimation.

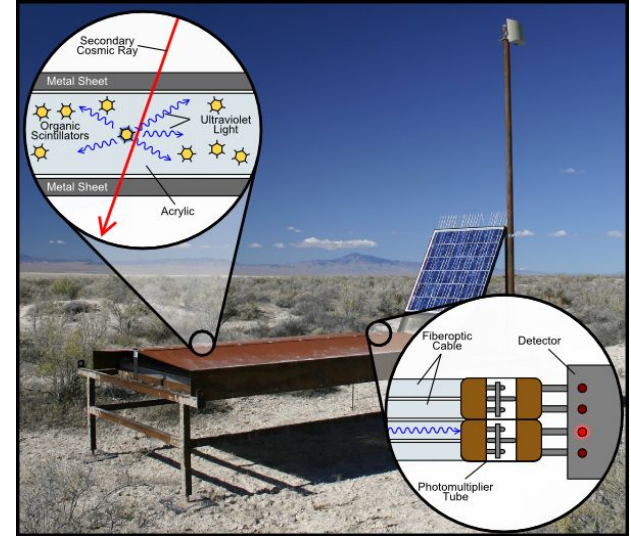
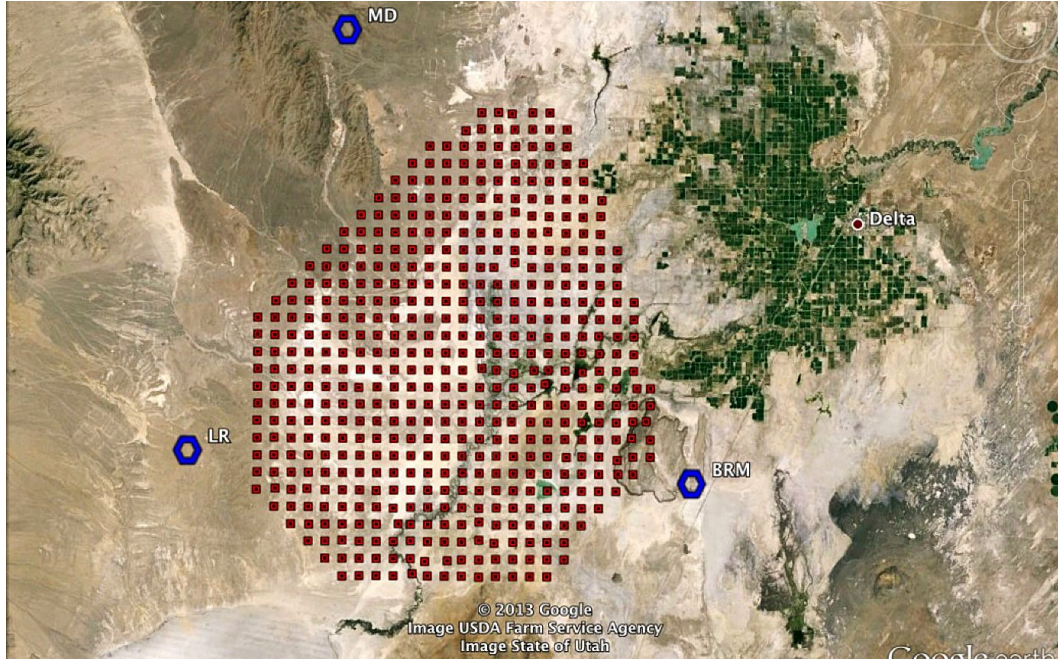


3. ML for accuracy

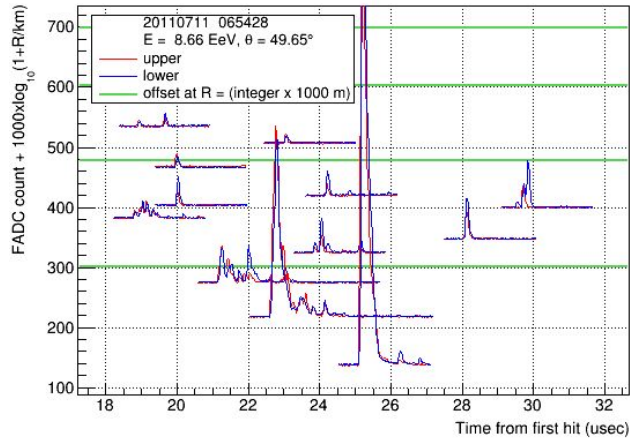
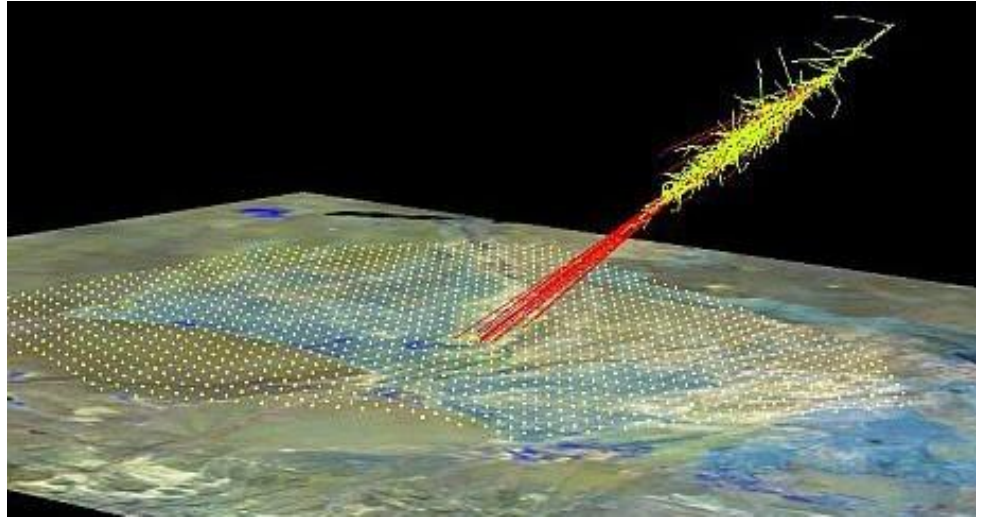
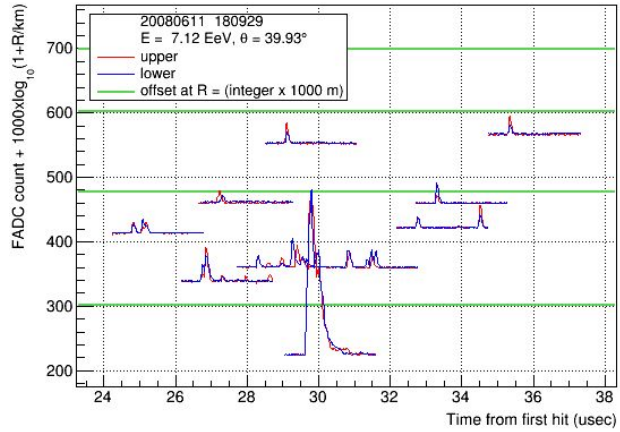
Telescope Array



Largest cosmic ray observatory in Northern hemisphere
(700 km², 507 surface detectors + 3 fluorescent detectors)

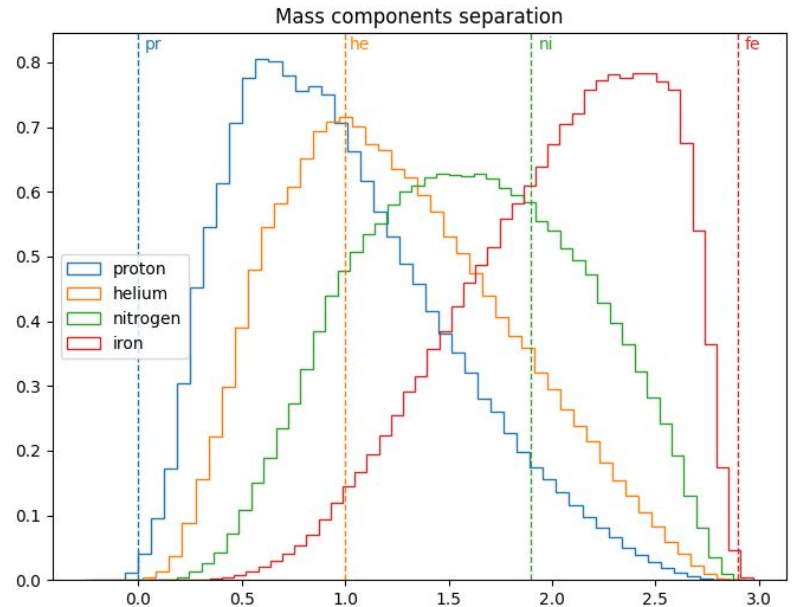
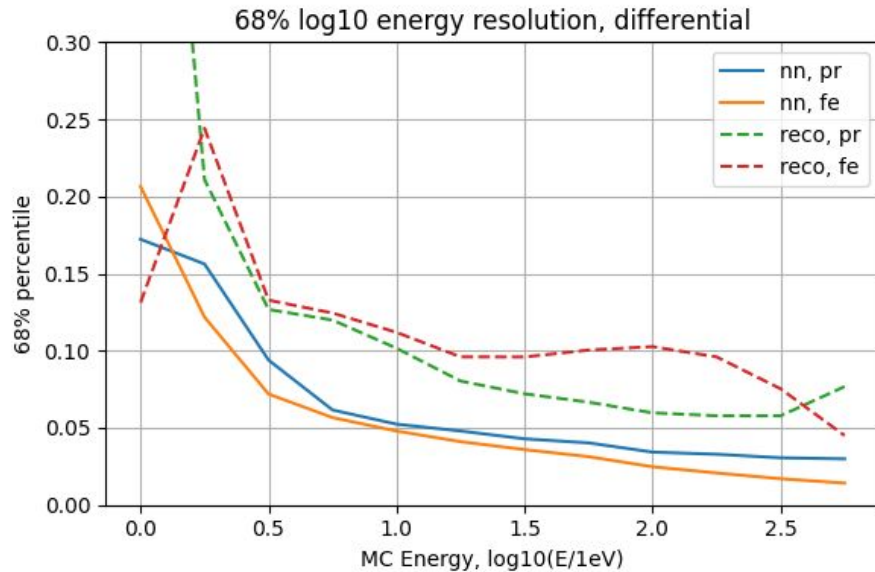


Telescope Array



Telescope Array

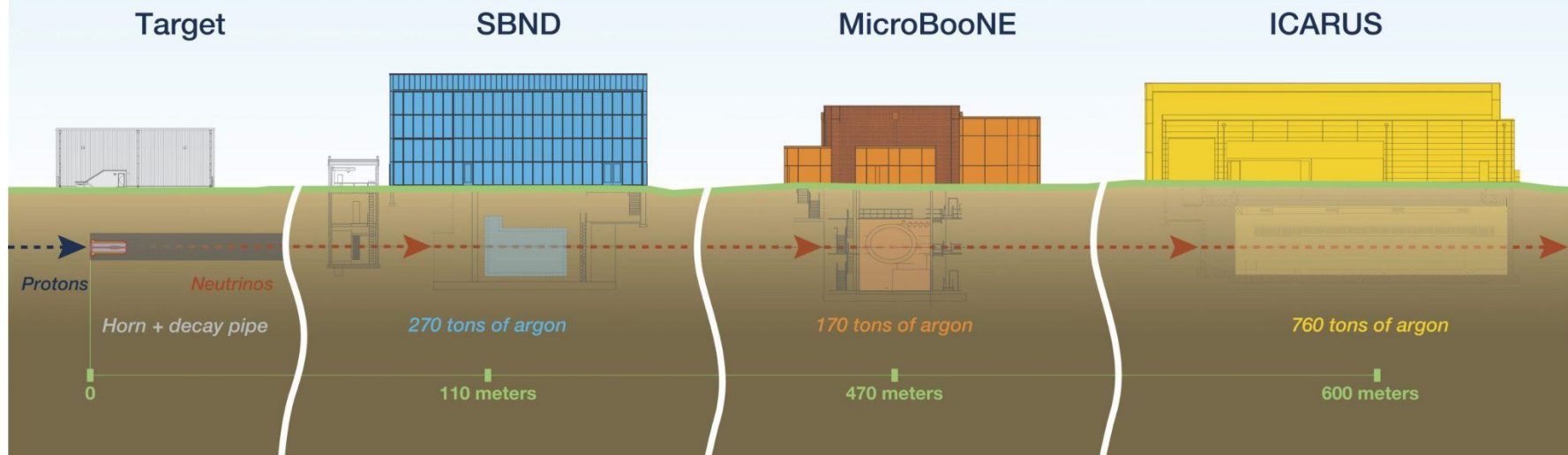
- Improved energy reconstruction
 - No bias with respect to primary EAS particle
- Makes possible mass spectrum analysis



4. ML as data simulator

ICARUS

Short-Baseline Neutrino Program at Fermilab



Fermilab

ICARUS

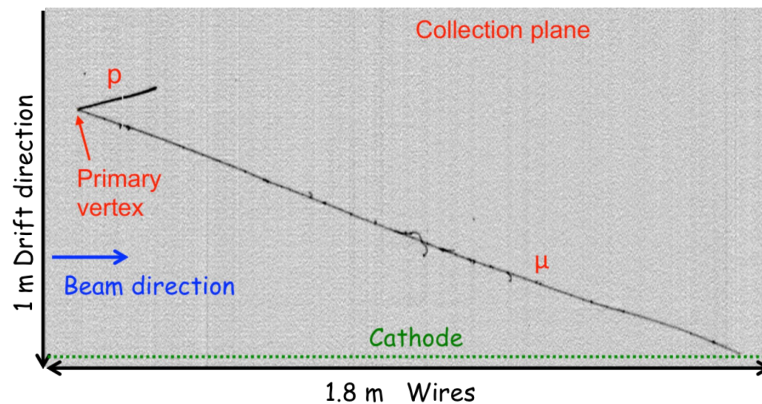
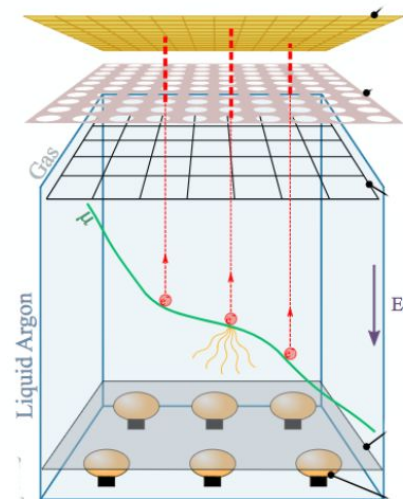
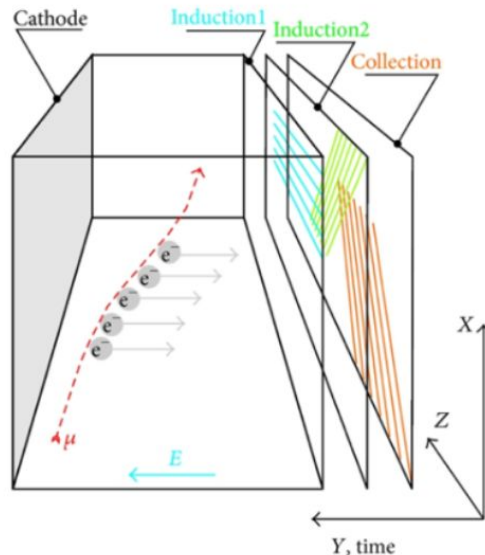
Simulation:

- Millions of voxels
- Complex photon reflections
- Baked-in simulation parameters
- Takes a while



Photon tables, $P(y_i | r)$:

- No spatial gradients
- Coarse interpolation



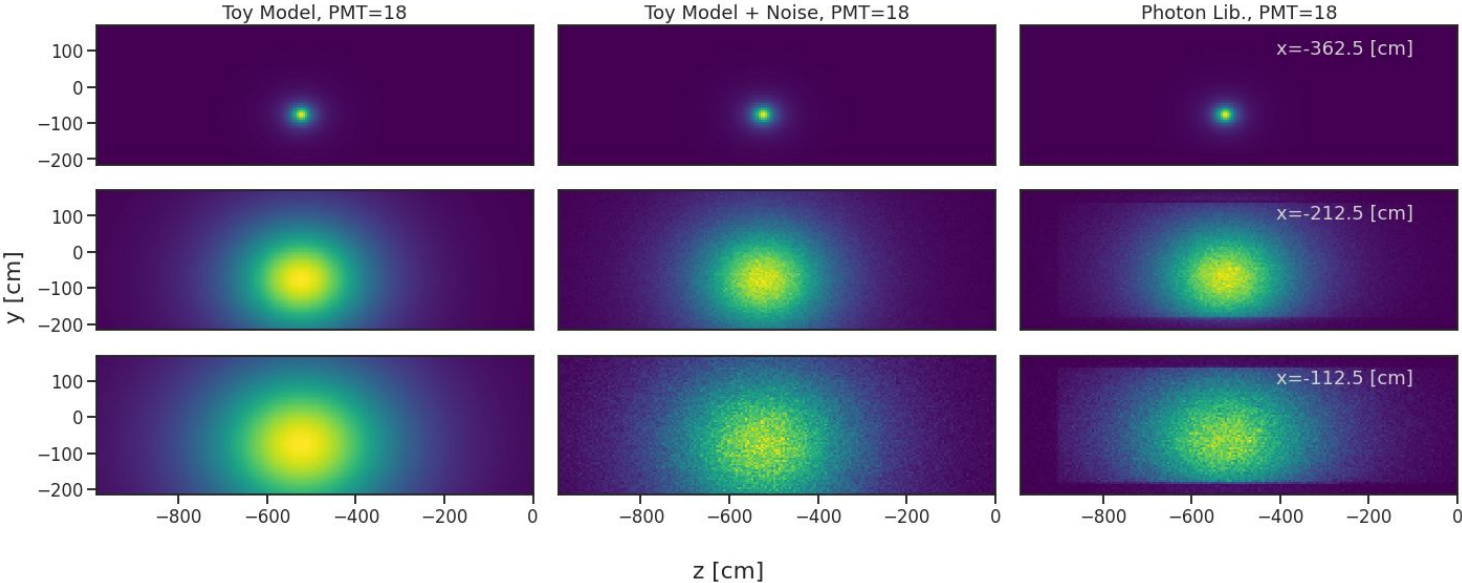
NN Surrogate Simulators

More compact than photon tables.

Calibration on real data

Train neural network to predict photon detection probability:

$$P(\gamma_i | r)$$



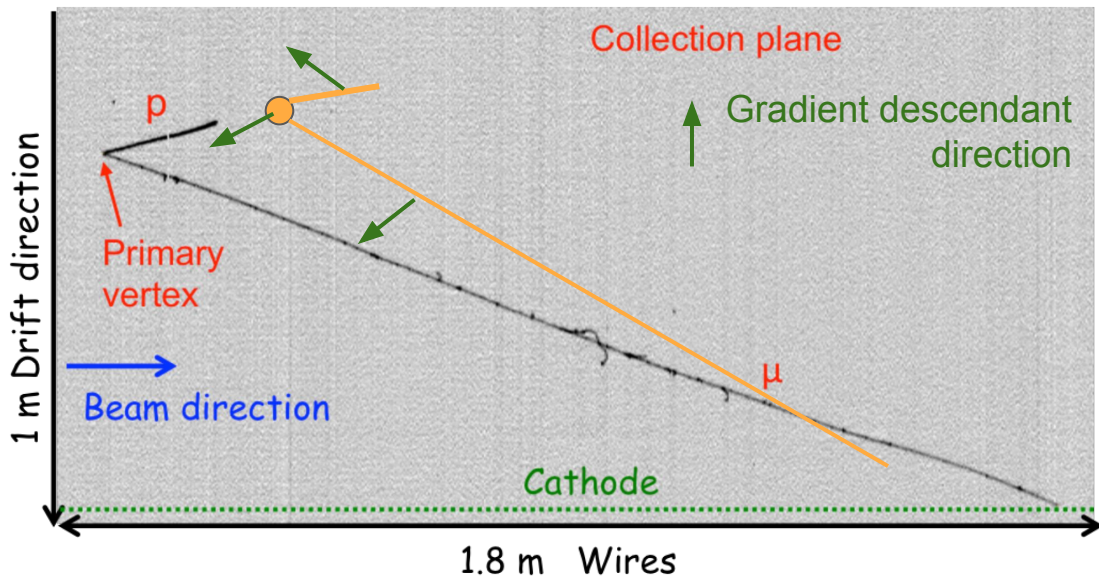
NN Surrogate Simulators

Train neural network to predict photon detection probability:

$$P(\gamma_i | r)$$

The mapping is differentiable!

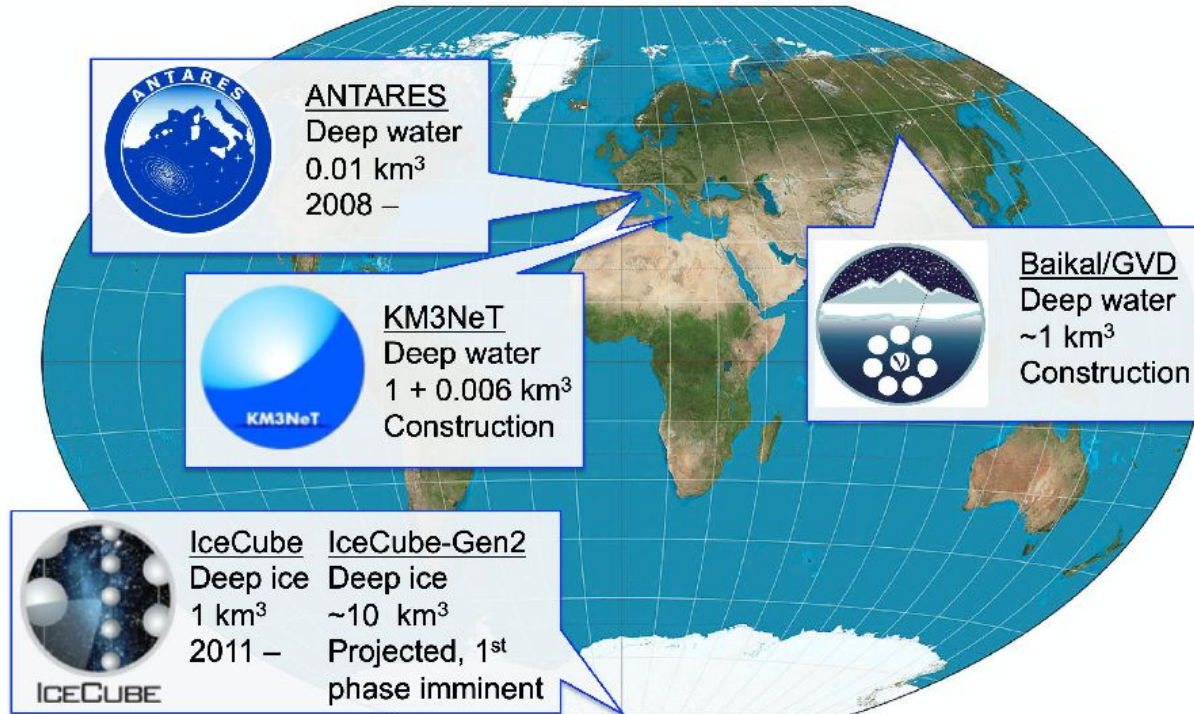
Direct likelihood optimization via gradient descendant



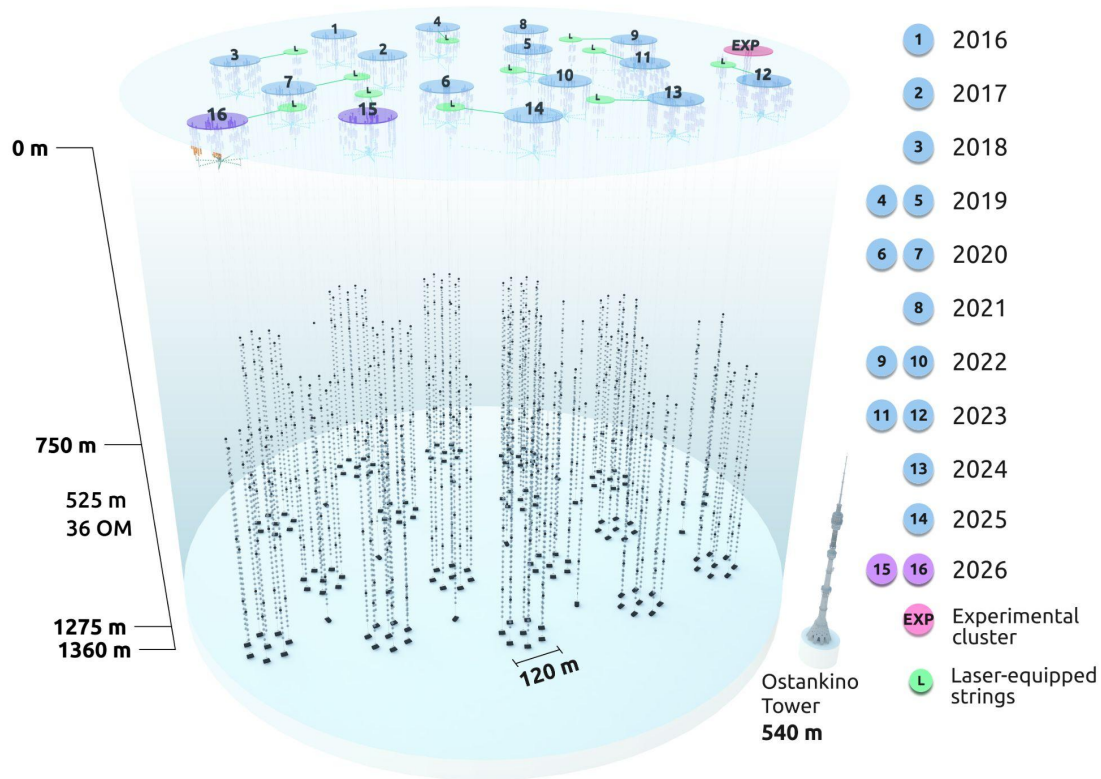
arXiv:2301.08643

5. ML for event selection

Baikal-GVD



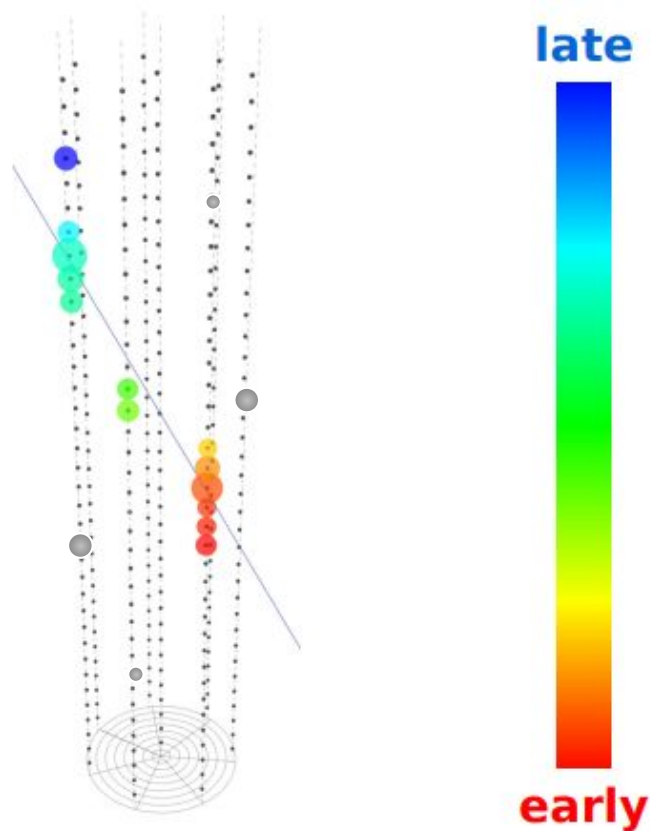
Baikal-GVD



Baikal-GVD

ML-based data analysis pipeline

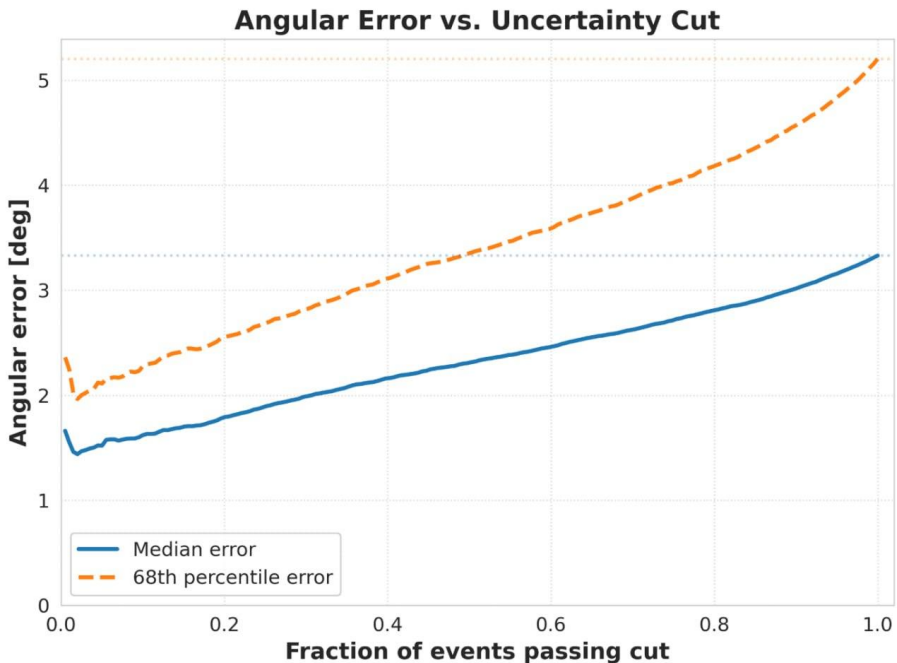
1. EAS suppression
 - Reduce background 10^4
2. Identification of signal hits
 - Surpass quality of algorithmic reconstruction
3. Estimation of energy and arrival direction
 - Predicted value and estimated error



Baikal-GVD

ML-based data analysis pipeline

1. EAS suppression
 - Reduce background 10^4
2. Identification of signal hits
 - Surpass quality of algorithmic reconstruction
3. Estimation of energy and arrival direction
 - Predicted value and estimated error



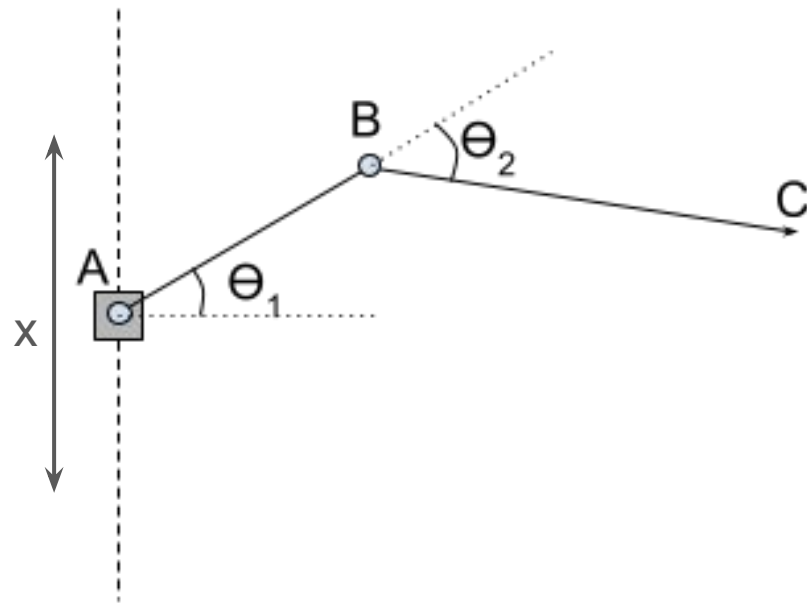
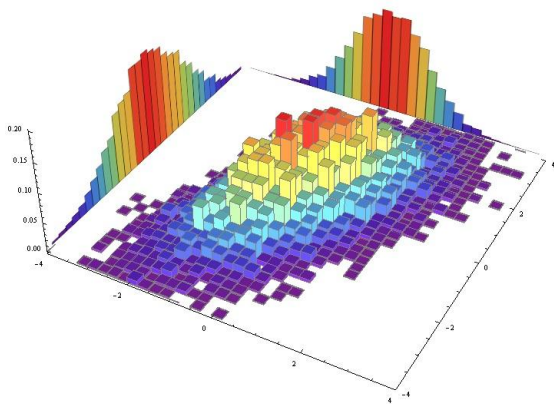
6. ML for posterior estimation

Non-unique problems

Given coordinates of point C,
define angles (and location of point A).

$\nexists f^{-1}!$

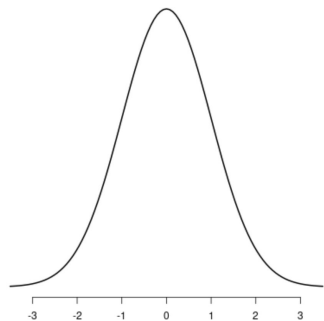
We need ot reconstruct probabilities!



Normalizing flow

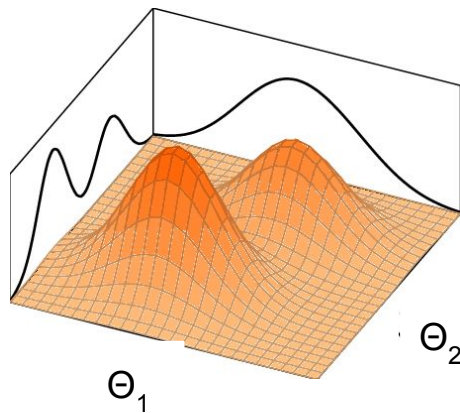
Arm
coordinate

$C \otimes$



- Choose some known distribution, like Gaussian.
- Learn mapping from this distribution to true posterior on data examples
- Neural network is trained to maximize likelihood
- At inference, sampling from Gaussian yields posterior.

$f^1(C,z)$

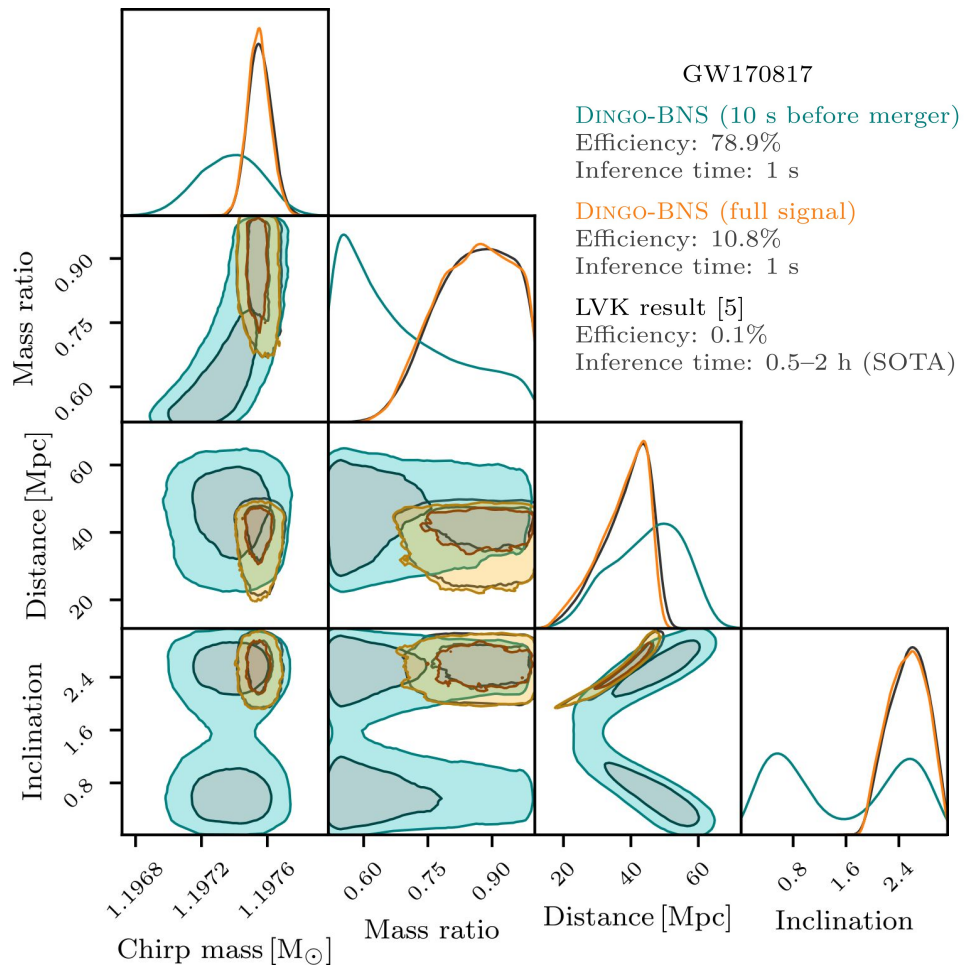


Simulation based inference

Simulators are exact,
but the likelihoods are intractable.

MCMC takes a **while**.

Neural network, once trained, **is fast**.



7. Why so long?

Applicability of trained neural networks

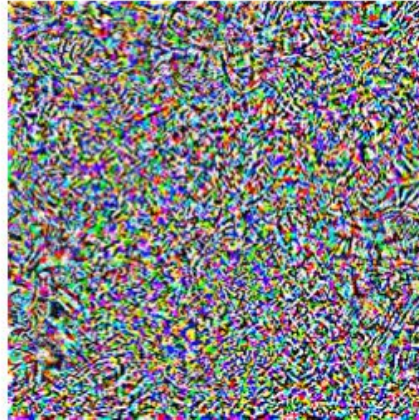
Neural network are trained on **specific data** with certain **data distribution** and are **very sensitive**.

Disagreement between training and to-be-applied-on data can be crucial:
systematic shifts, bad quality of MC data, ...

“pig”



+ 0.005 x



=

“airliner”



Domain adaptation as solution

Source (training) and target dataset might be different.
NN will specialize its data processing to source domain.



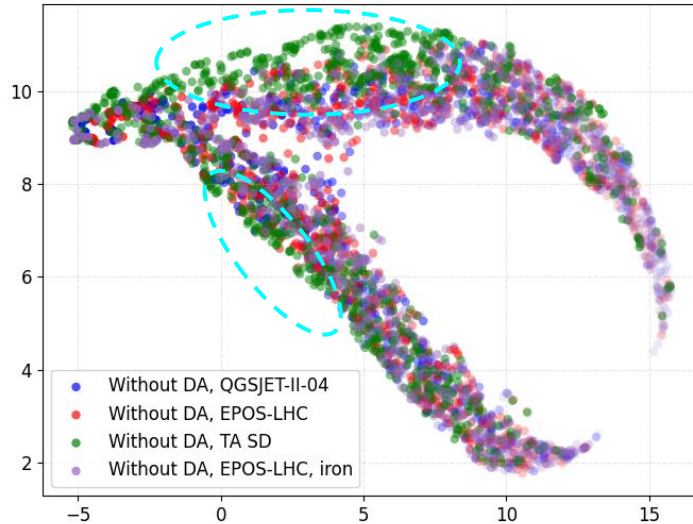
Force NN to learn domain-invariant features!

arXiv:1409.7495

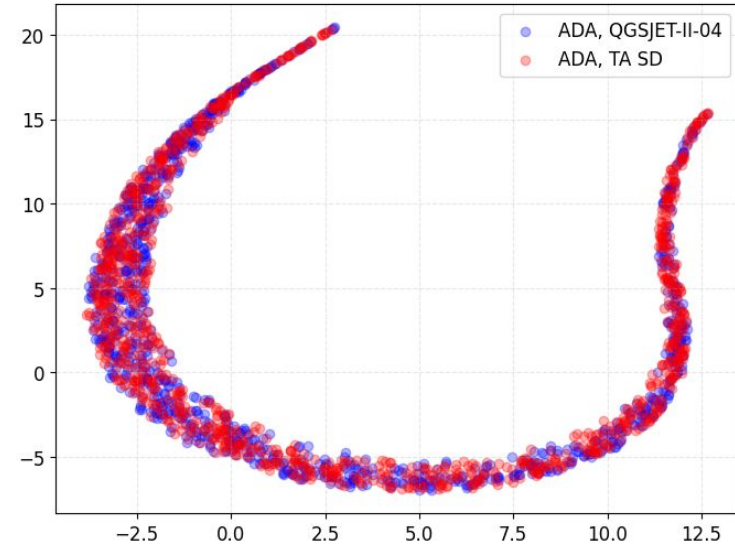
Domain adaptation effect

Comparison of feature distribution after the encoder

UMAP without DA



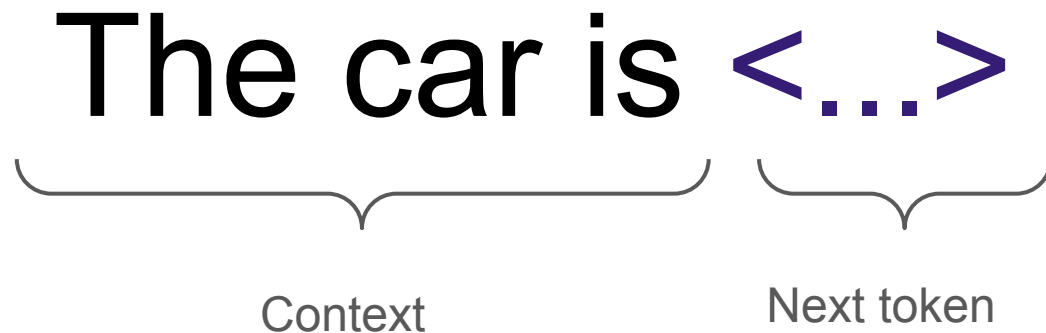
UMAP with DA



Domain Adaptation makes feature distributions close

8. Can LLMs help?


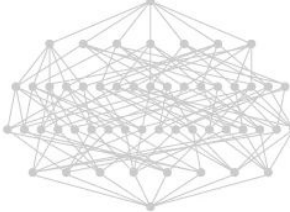
What is LLM



Maximize $P(\text{next token} \mid \text{context})$

AI for math and physics

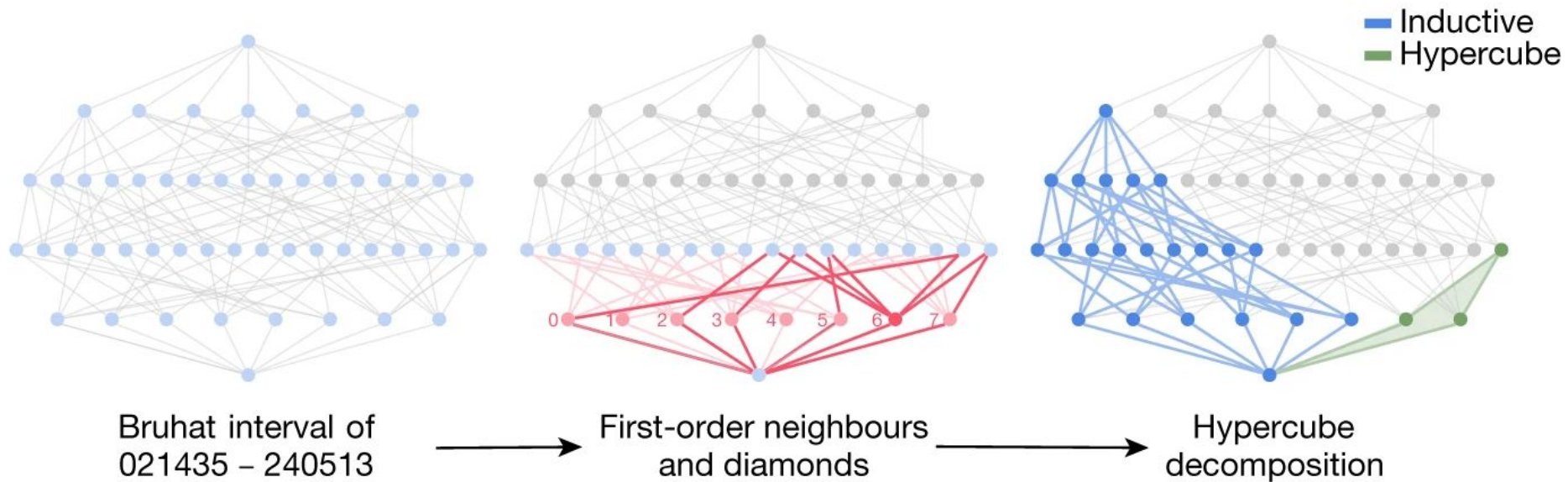
Nature: “Advancing mathematics by guiding human intuition with AI”

z: Pair of permutations	X(z): Unlabelled Bruhat interval	Y(z): KL polynomial
(03214), (34201)		$1 + q^2$
(021435), (240513)		$1 + 2q + q^2$

Can NN solve and find useful features?

AI for math and physics

c



Nature

Examine what NN uses and generalize!

Summary

- ML helps physics:
 - Automation, Precision, Posterior estimation
 - Tool of modern statistics and simulation
- Simulation-to-data gap is present but manageable
- Neural networks can assist in scientific discovery
 - Physics can guide ML