

Machine Learning crash course

Ahmed Hammad

High Energy Accelerator Research Organization (KEK).
IPMU, Tokyo U.

3rd summer school at CTP, BUE

Outline

In this series we are going to discuss the basics of classical and quantum ML as the following :

- Overview about recent classical and quantum machine learning algorithms in HEP.
- Introduction to ML, linear and non-linear regression models.
- Ensemble learning, decision trees, boosted decision tress and random forest.
- Introduction to deep learning and feed forward deep neural network.
- Convolution based neural network for image recognition.
- Auto-Encoders and variational auto-encoders for anomaly detection.
- Introduction to quantum machine learning, quantum gates, quantum feature map and classical data encoding.
- Variational quantum circuit for non-linear separable data analysis.
- Hybrid classical-quantum models for image classification.

References

Classical ML:

- 1- Hands-on machine learning with Sikit-Learn and Tensorflow. By Aurélien Géron
- 2- Probabilistic machine learning. By Kevin P.Murphy
- 3- Machine learning with Pytorch and scikit-Learn. By Sebastian Raschka, etal

Quantum ML:

- 1- Machine learning with quantum computers. By Maria Schuld
- 2- Quantum computers and quantum information. By Micheal A. Nielsen & Issac L.Chuang

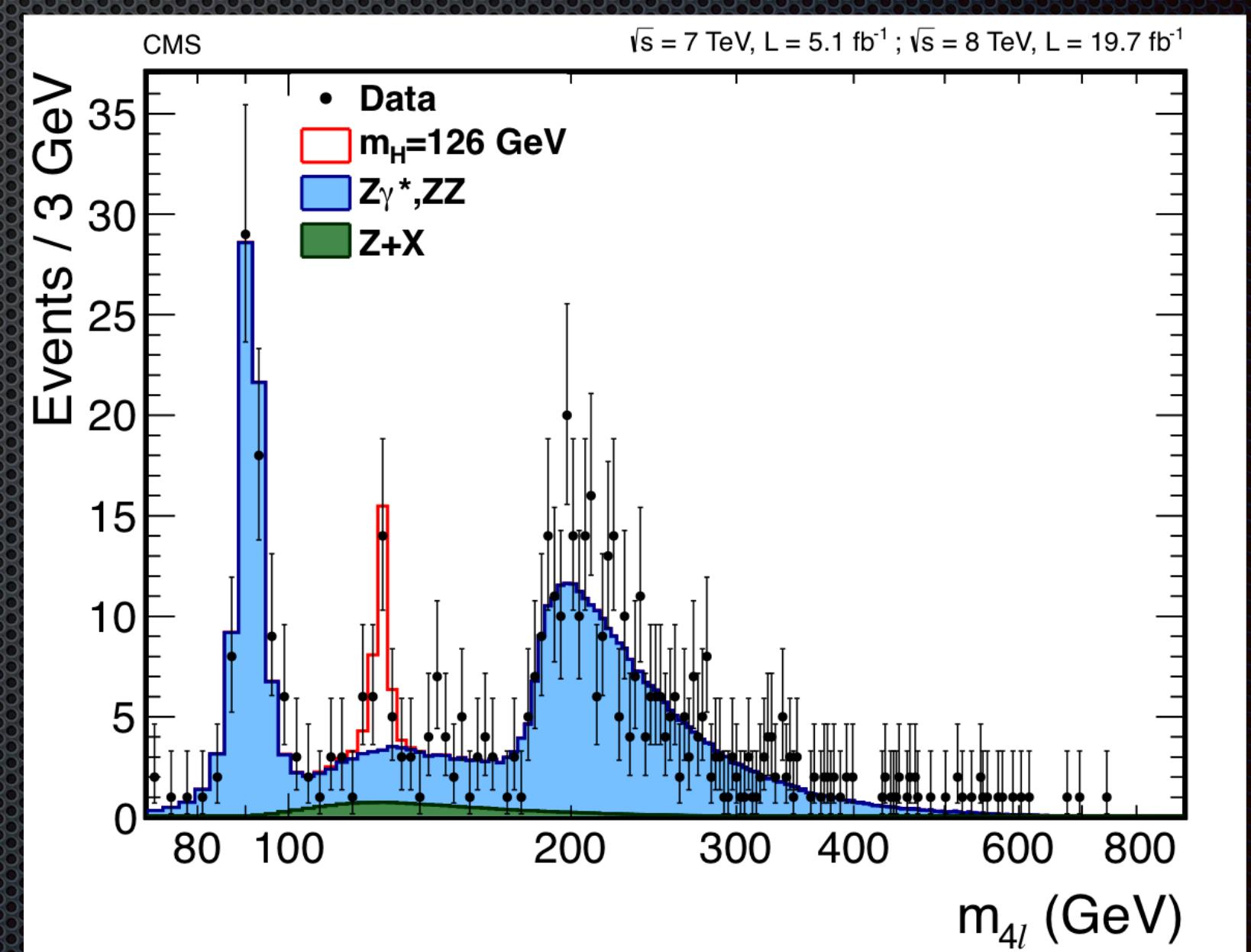
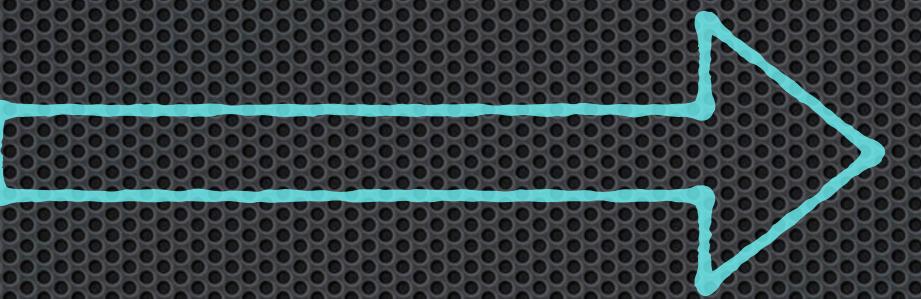
Machine Learning crash course

(Part-1)

Classical & quantum machine learning methods in HEP

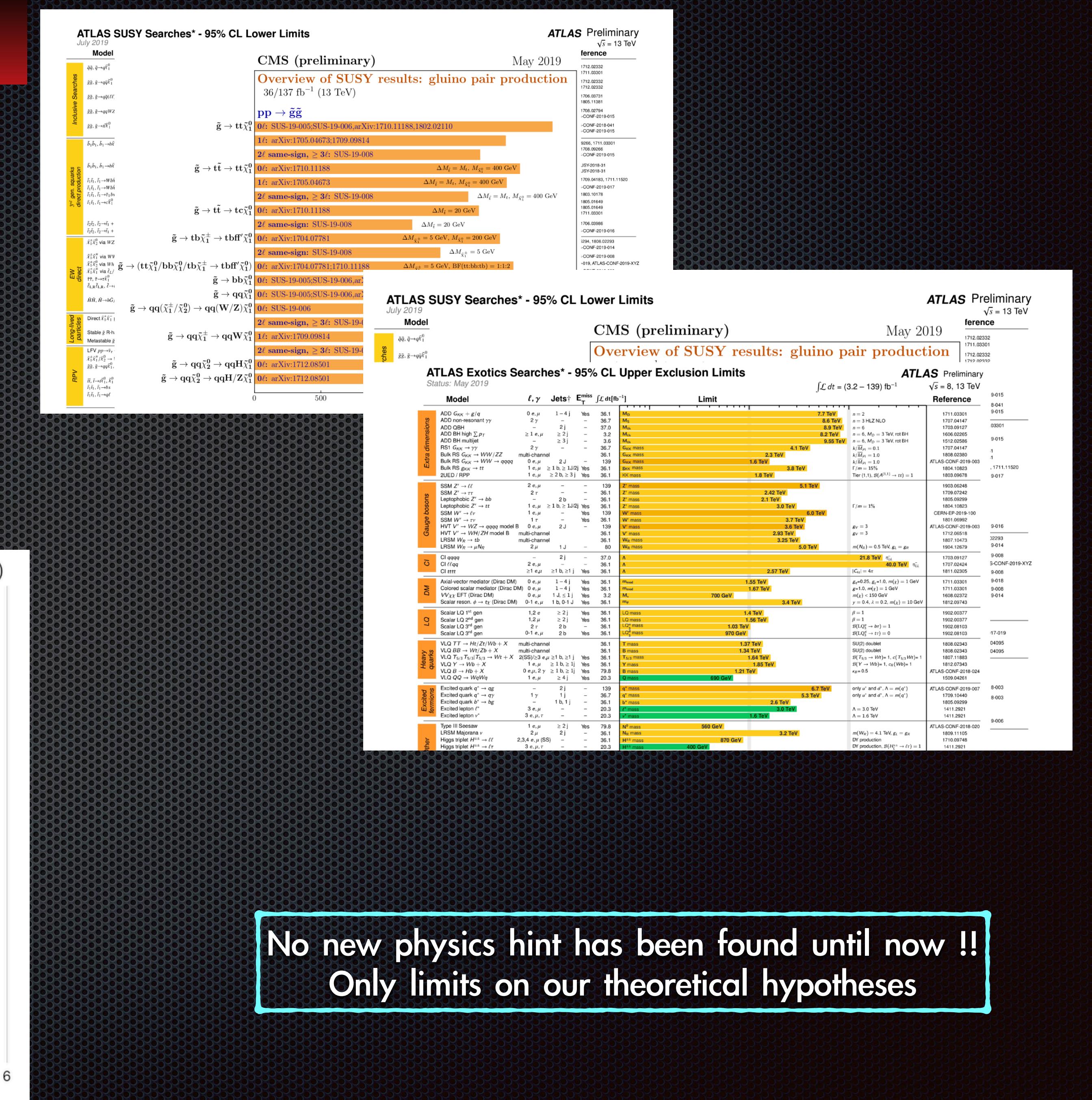
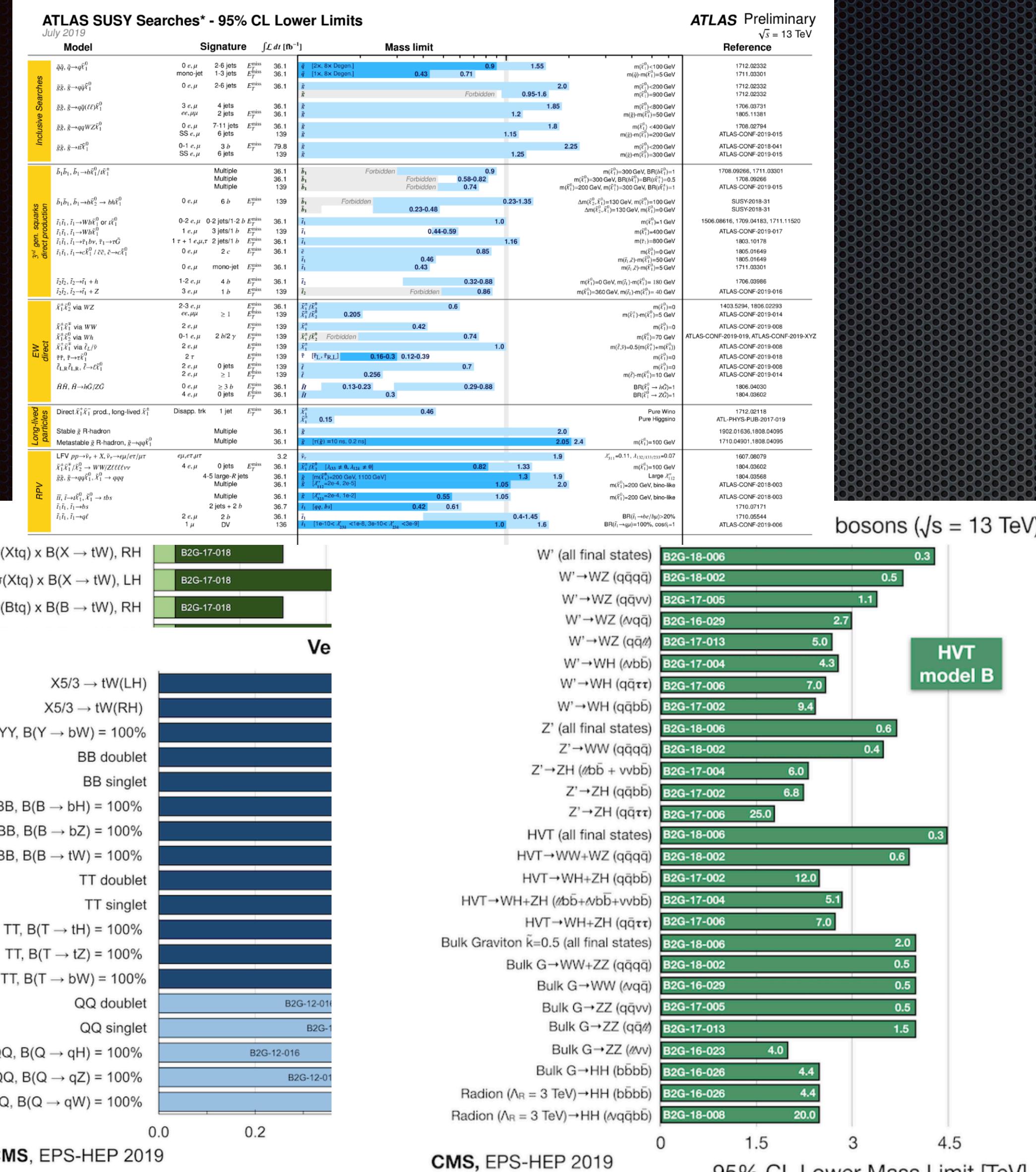
Introduction

In 2012 the Higgs boson has been discovered and considered as the last missing piece in the SM



Since then the community was optimistic to find new physics after the Higgs discovery.
They pursued their search using **cut and count analysis**

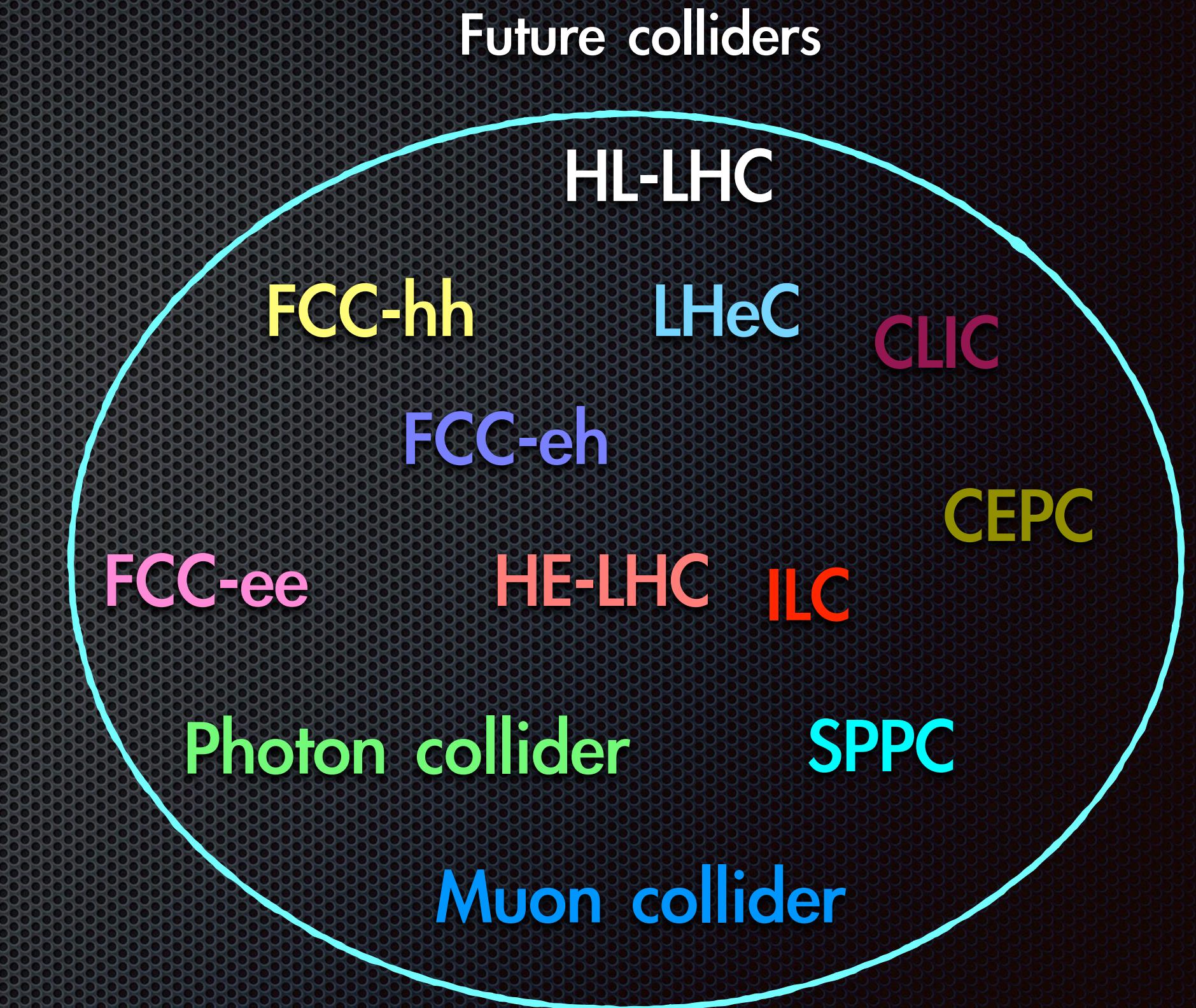
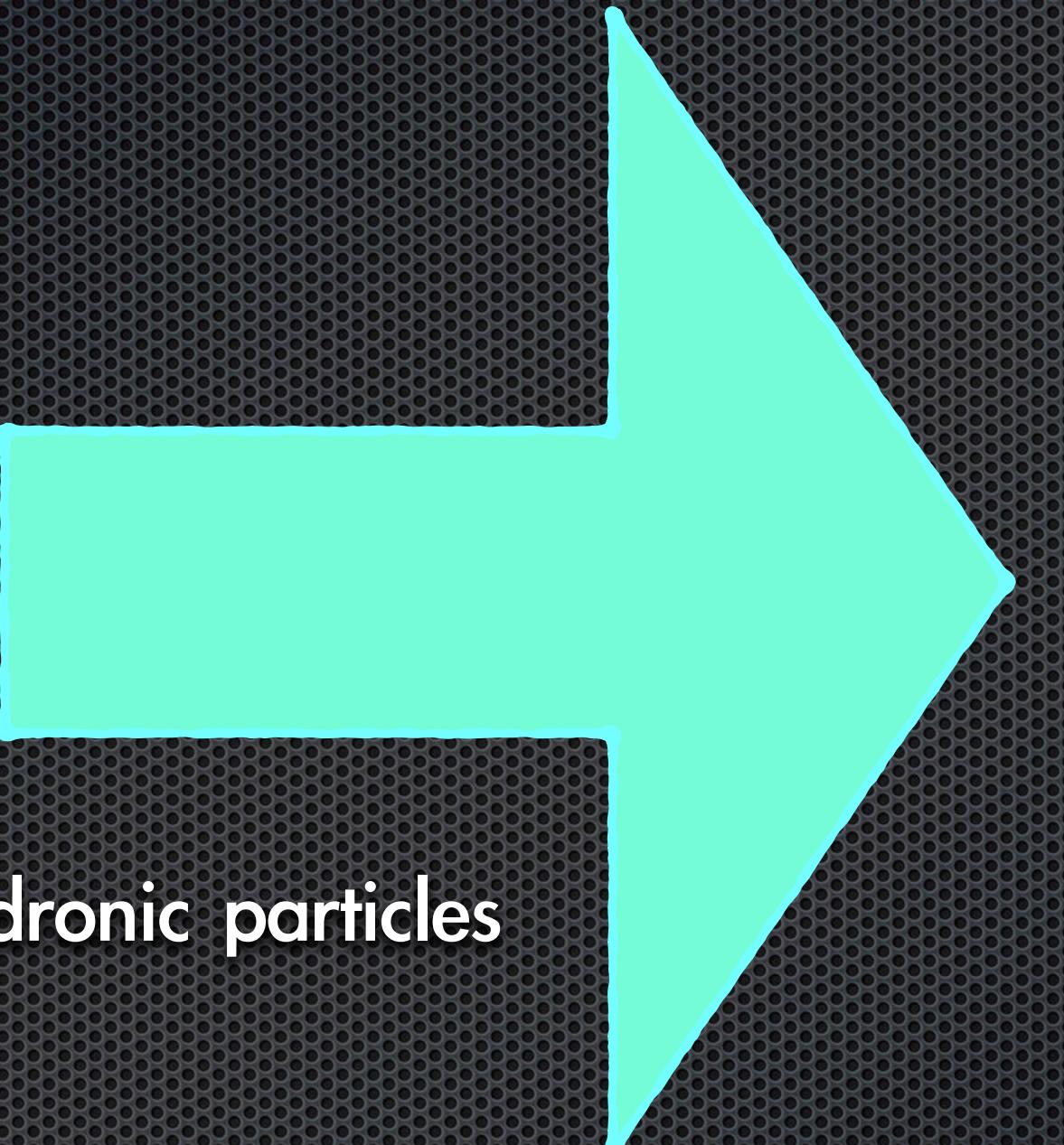
Introduction



Introduction

Suggestions for new physics discovery:

- Increase the center of mass energy
- Higher integrated Luminosity
- Reduce the QCD contamination
- Improve tracking efficiency
- Asymmetric detector to separate the boosted hadronic particles
- Leptonic collisions for precise measurements
- Lepton-proton collisions for leptoquarks signatures, etc..



Introduction

Suggestions for new physics discovery:

- Increase the center of mass energy
- Higher integrated Luminosity
- Reduce the QCD contamination
- Improve tracking efficiency
- Asymmetric detectors
- Leptonic collision
- Lepton-proton



Future colliders

HL-LHC

FCC-hh

LHeC

CLIC

CEPC

HE-LHC ILC

FCC-ee

SPPC

Photon collider

Muon collider

Introduction



If you torture the data long enough,
it will confess.

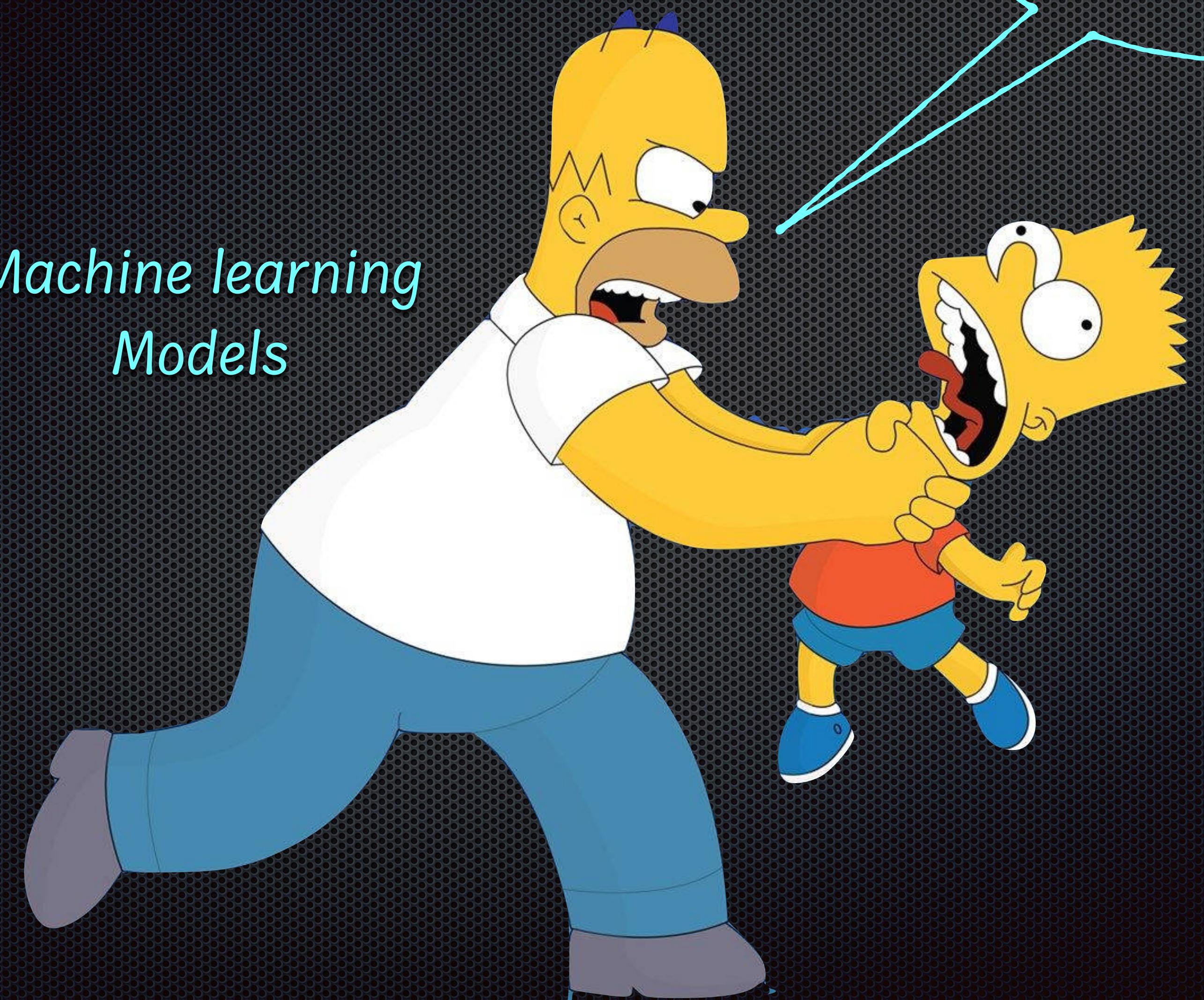
— Ronald Coase —

Introduction

Machine learning
Models

LHC data
(Current)

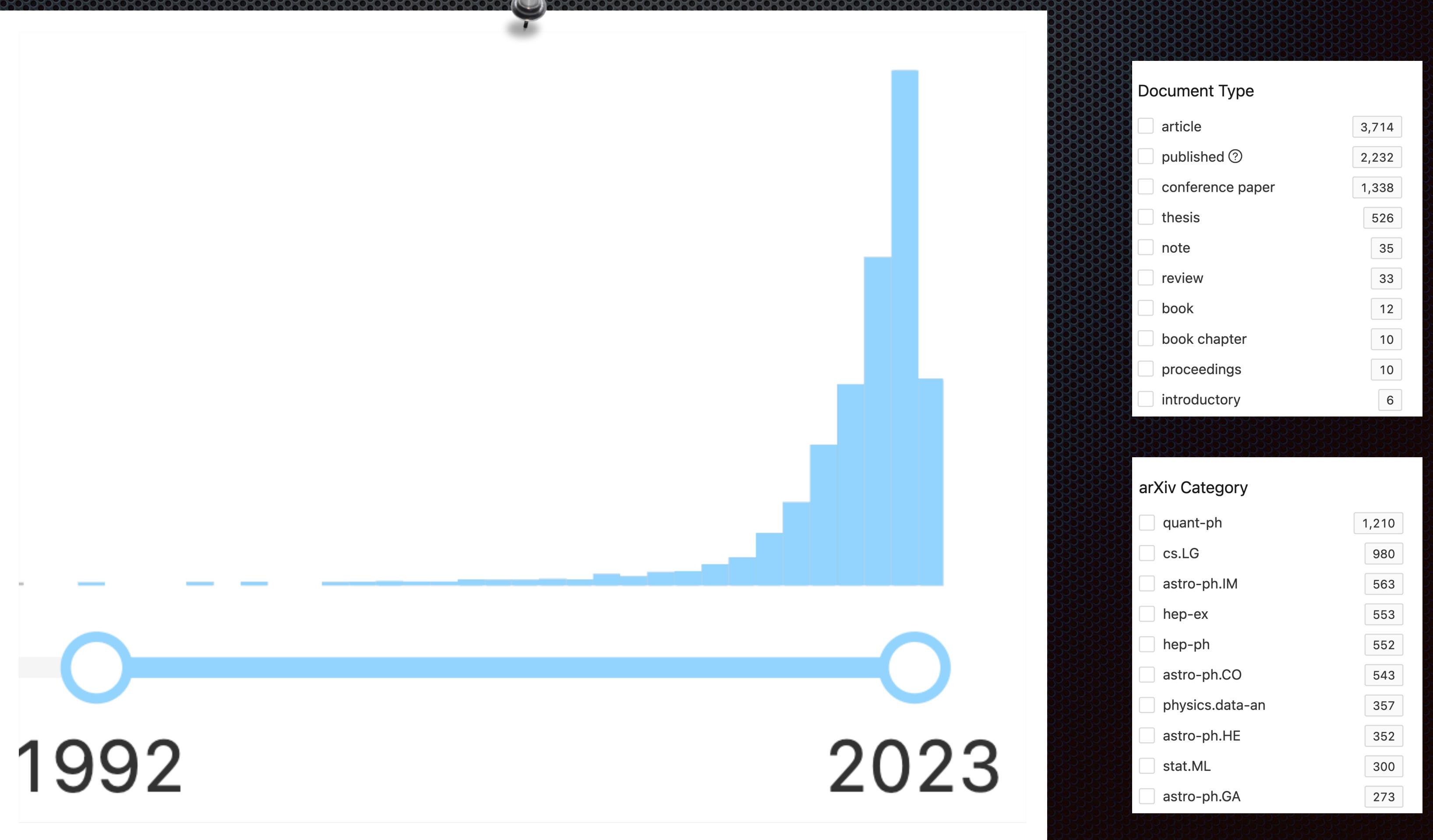
Where is the hidden
new physics ???



Introduction

*Rapid growing of ML
searches in HEP!!*

Search result for “machine learning” on INSPIRE



Introduction

~ 1000 paper with codes

A Living Review of Machine Learning for Particle Physics

Modern machine learning techniques, including deep learning, is rapidly being applied, adapted, and developed for high energy physics. The goal of this document is to provide a nearly comprehensive list of citations for those developing and applying these approaches to experimental, phenomenological, or theoretical analyses. As a living document, it will be updated as often as possible to incorporate the latest developments. A list of proper (unchanging) reviews can be found within. Papers are grouped into a small set of topics to be as useful as possible. Suggestions are most welcome.

[download](#) [review](#) [GitHub](#)

The purpose of this note is to collect references for modern machine learning as applied to particle physics. A minimal number of categories is chosen in order to be as useful as possible. Note that papers may be referenced in more than one category. The fact that a paper is listed in this document does not endorse or validate its content - that is for the community (and for peer-review) to decide. Furthermore, the classification here is a best attempt and may have flaws - please let us know if (a) we have missed a paper you think should be included, (b) a paper has been misclassified, or (c) a citation for a paper is not correct or if the journal information is now available. In order to be as useful as possible, this document will continue to evolve so please check back before you write your next paper. If you find this review helpful, please consider citing it using \cite{hepmllivingreview} in `HEPML.bib`.

This review was built with the help of the HEP-ML community, the [INSPIRE REST API](#), and the moderators Benjamin Nachman, Matthew Feickert, Claudio Krause, and Ramon Winterhalder.

Reviews

Modern reviews

<https://github.com/iml-wg/HEPML-LivingReview>

Topics:

- ⦿ Modern reviews
- ⦿ Parametrized classifiers
- ⦿ Jet images
- ⦿ Event images
- ⦿ Sequances
- ⦿ Trees
- ⦿ Graphs
- ⦿ Point clouds
- ⦿ Physics inspired basis
- ⦿ W/Z tagging
- ⦿ H-> bb
- ⦿ Quarks and gluons
- ⦿ Top quark tagging
- ⦿ Strange jets
- ⦿ Bottom quark tagging
- ⦿ Flavor physics
- ⦿ BSM phsyics
- ⦿ Particle identification
- ⦿ Neutrino detector
- ⦿ Dark matter detectors
- ⦿ Cosmology
- ⦿ Astro particle
- ⦿ Tracking
- ⦿ Heavy ions
- ⦿ Semi supervision
- ⦿ Quantum machine learning
- ⦿ Attention models
- ⦿ Parameter estimation
- ⦿ Lattice gauge theory
- ⦿ And more in the list

So, what is machine learning ?

Introduction

So, what is machine learning ?

ChatGPT says:

A Hi GPT

 Hello again! How can I assist you today?

A What is machine learning algorithm in one sentence?

 A machine learning algorithm is a set of mathematical instructions that enables a computer to learn patterns from data and make predictions or take actions based on those patterns.

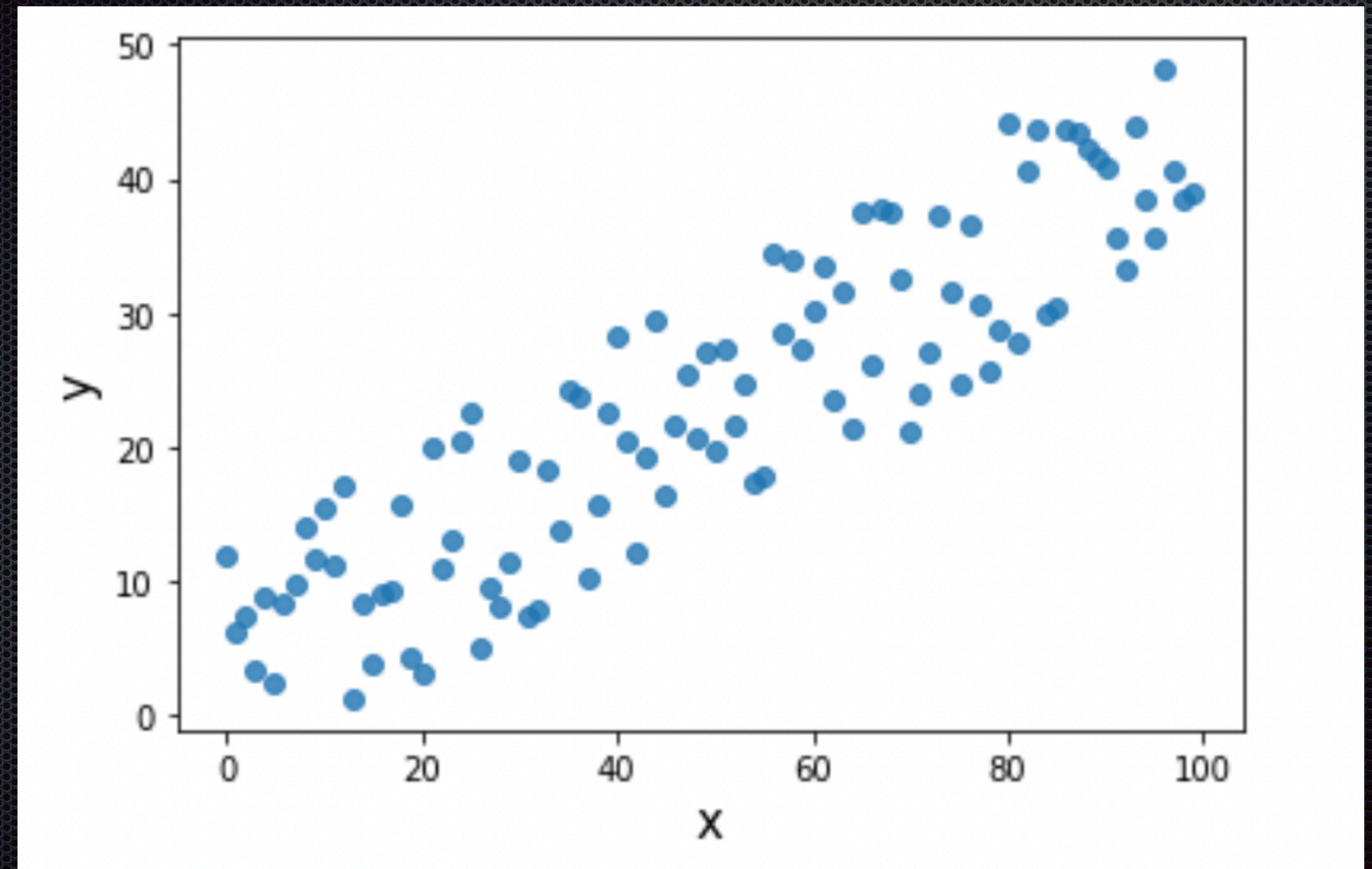
So, what is machine learning ?

$$Y = F(X)$$

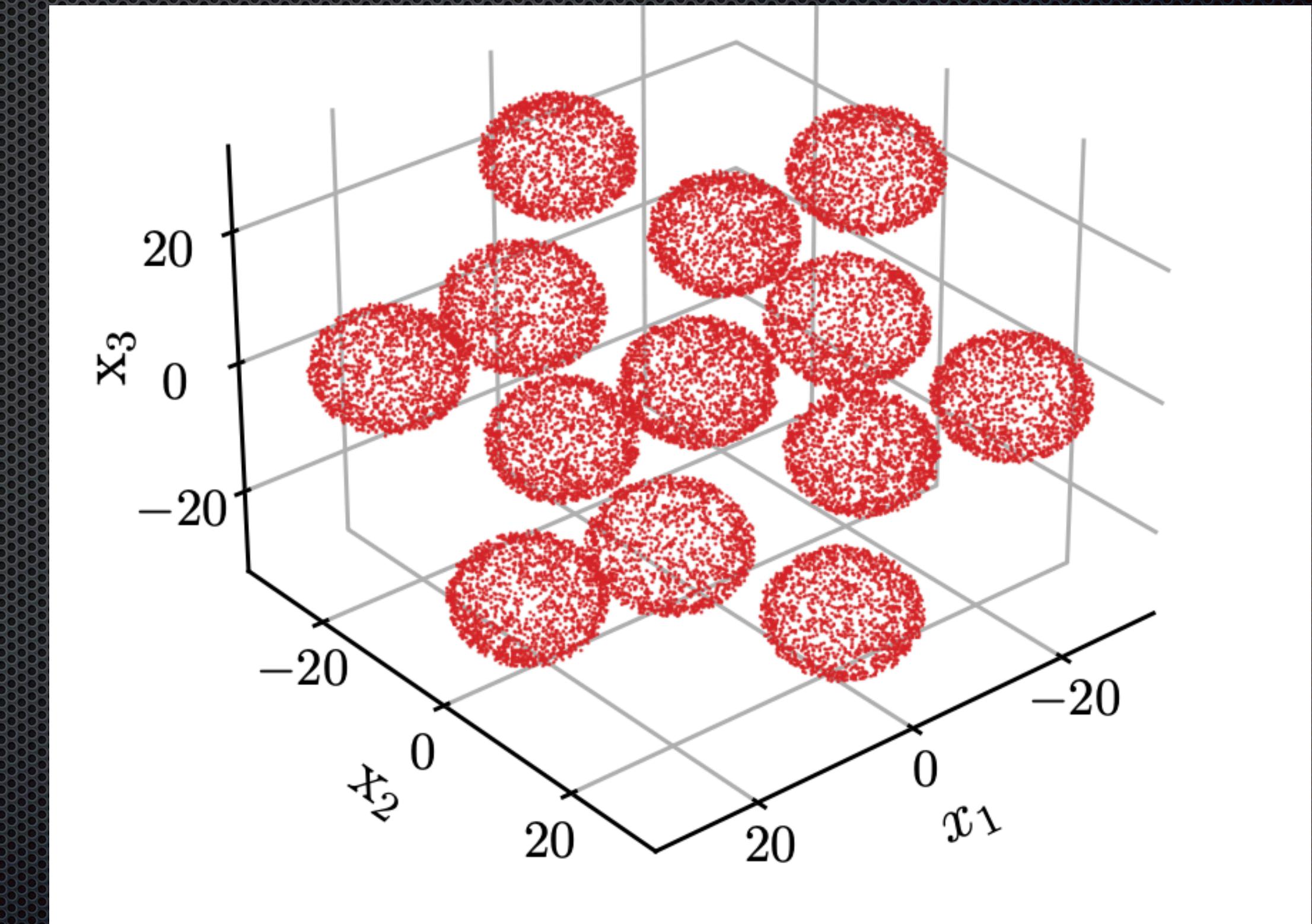
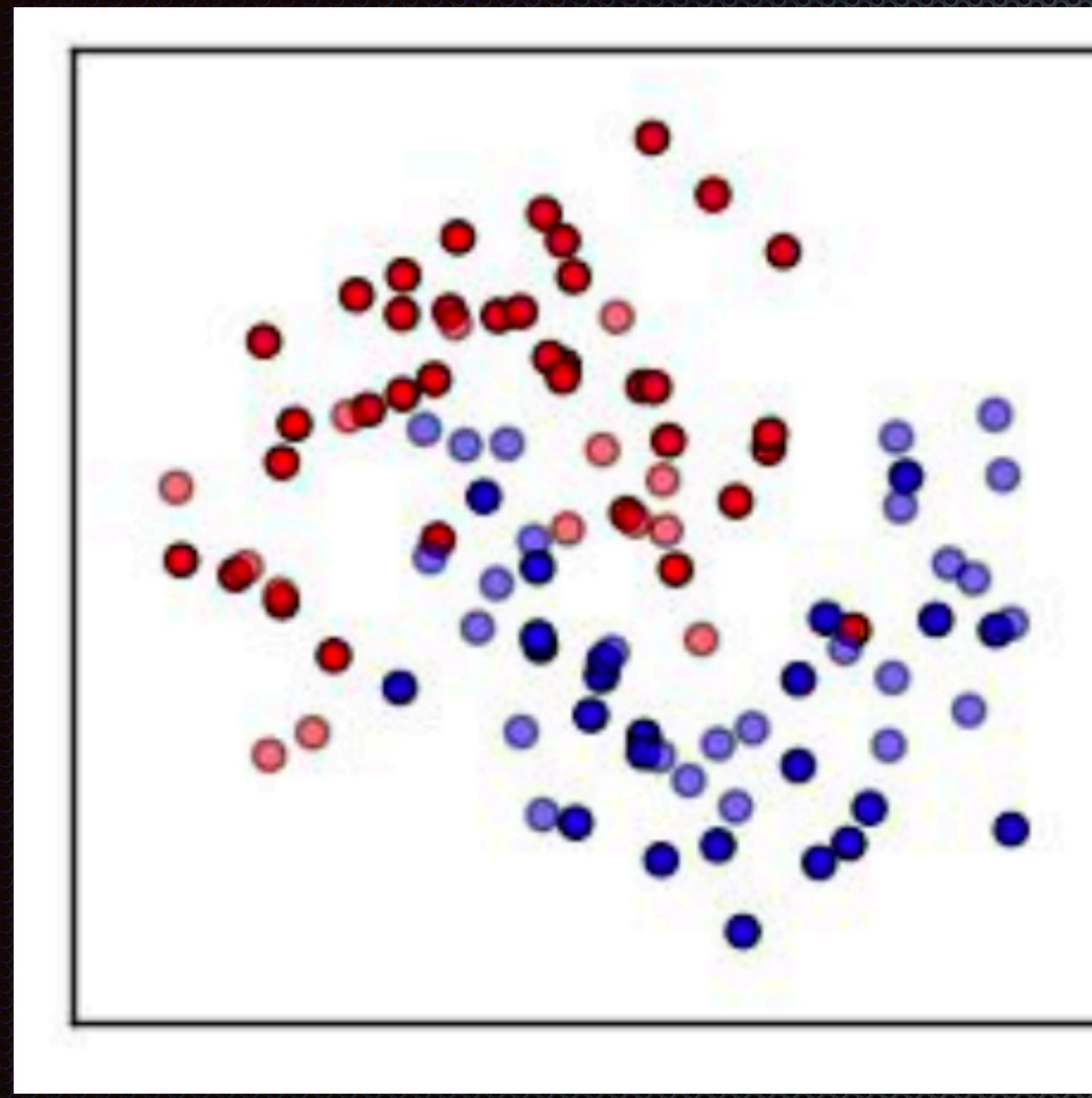
Machine learning models approximate the function F for any type of data X

Introduction

Approximated function can be very simple: One dimension-Linear function



Low dimensions - Non-linear function



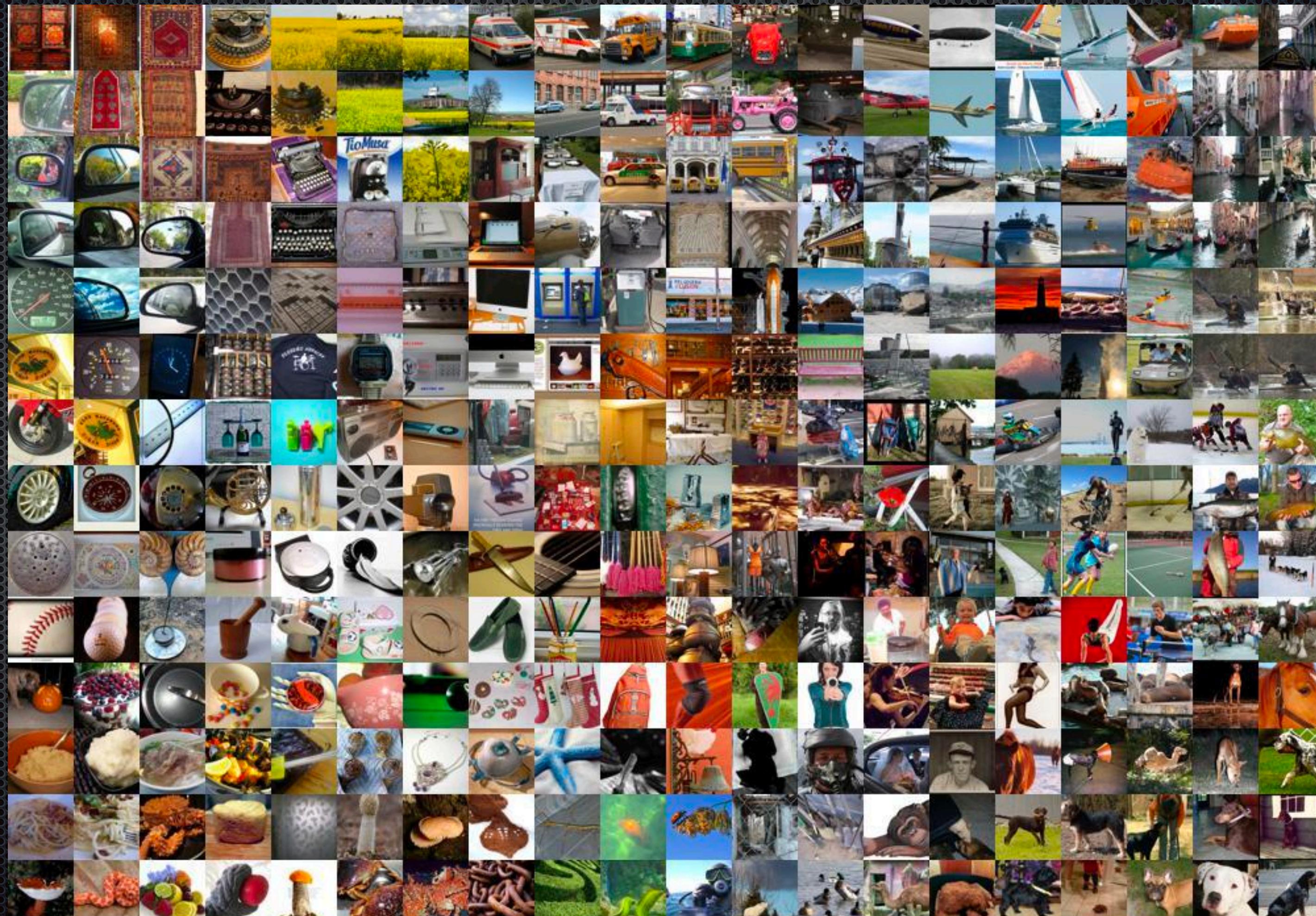
Introduction

ImageNet dataset

Size: 167.6 GB

High dimensions - Non-linear function

$$d = (100000, 225, 225, 3)$$



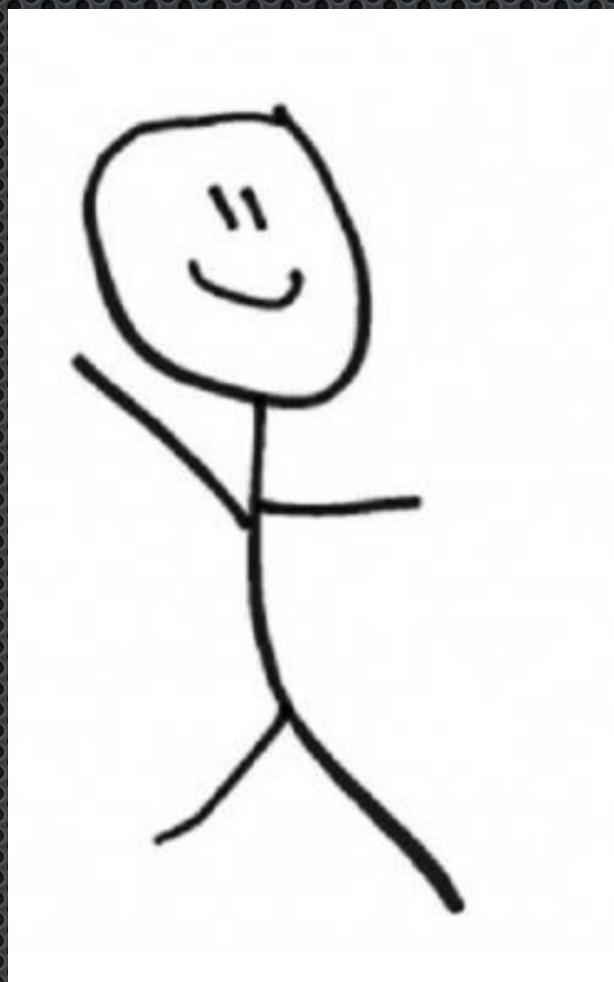
Introduction

How does machine learning approximate F ?

Inputs



Data scientist



0
10
20
30

Bunch of numbers

3	1	4	1
5	9	2	6
5	3	5	8
9	7	9	3
2	3	8	4
6	2	6	4

2	1	8	8	1
2	8	5	0	5
2	4	9	4	5
2	3	6	2	8
7	7	1	5	2

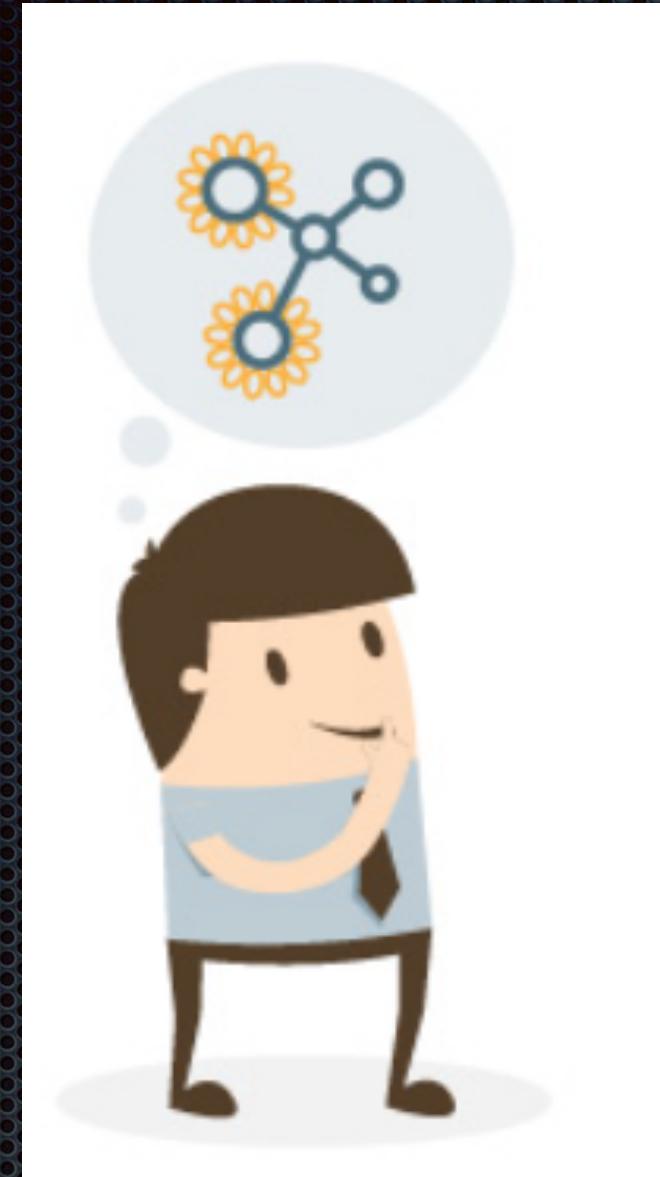
Find a pattern



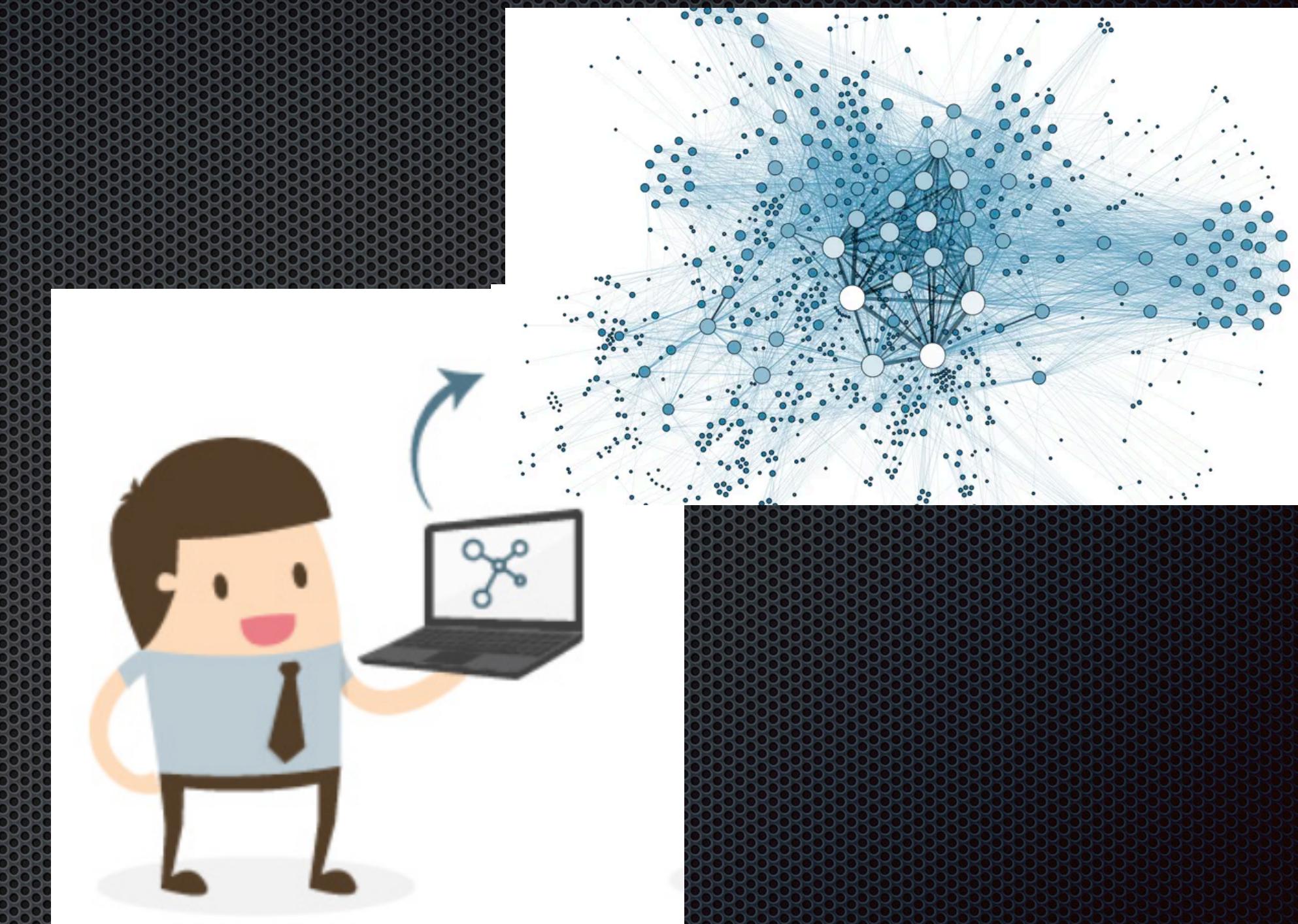
New data

Prediction

Introduction



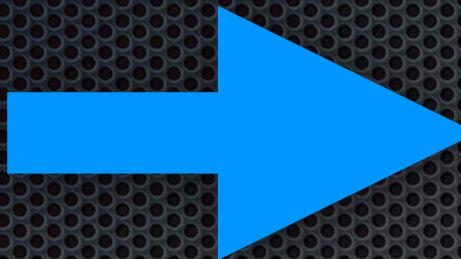
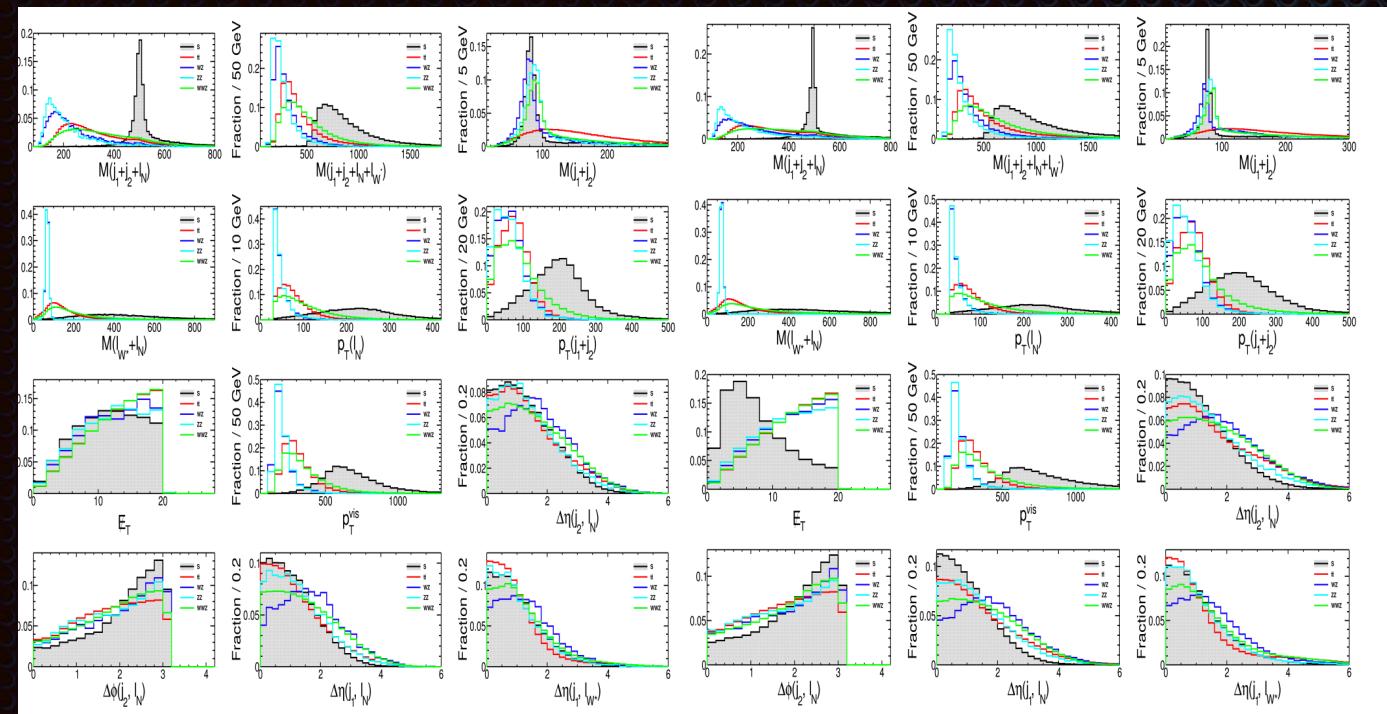
Pattern in our mind



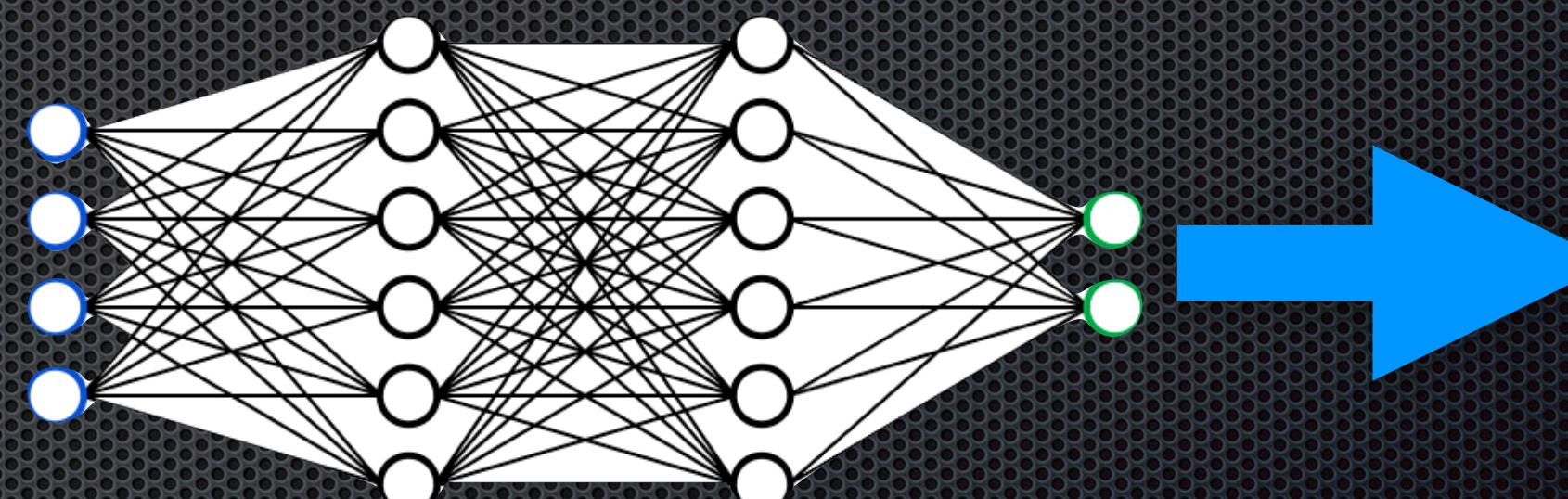
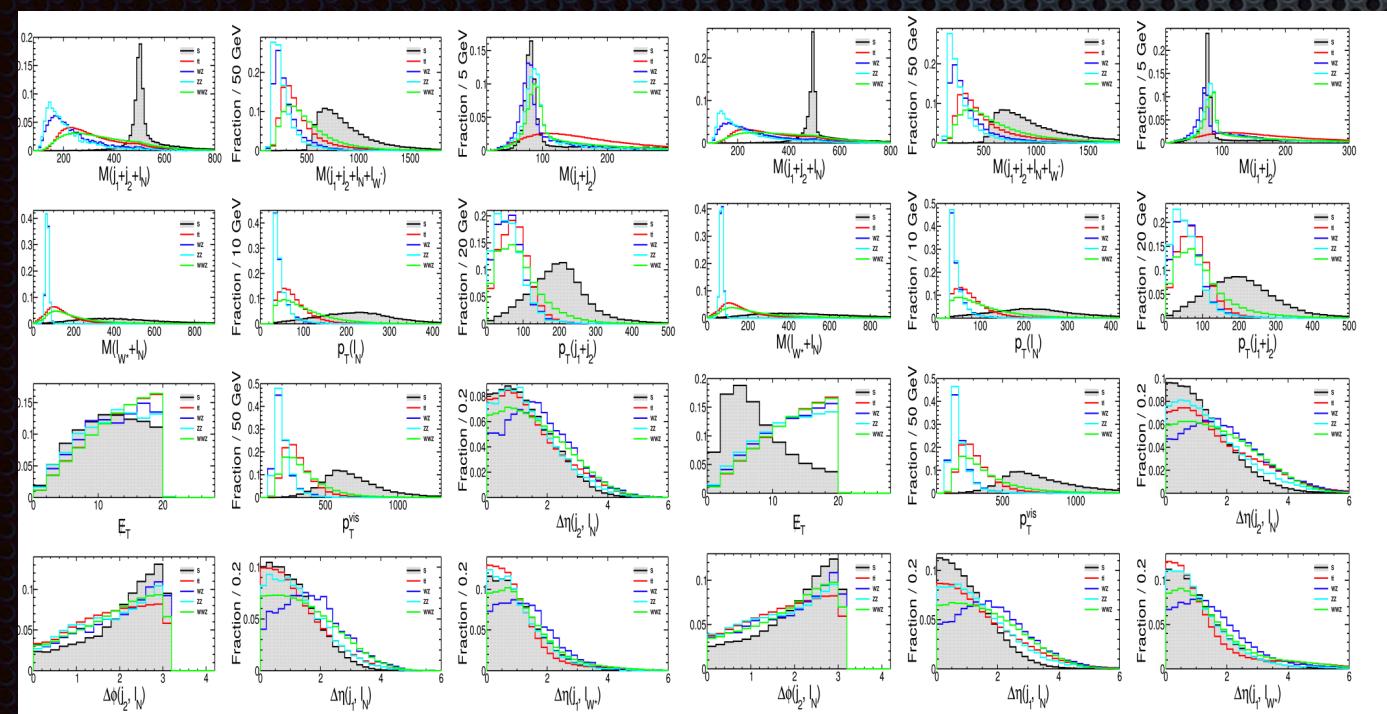
Patterns by ML

Our brains or analytic calculations are unable to create patterns between large correlated data

Multi-Layers perceptron (MLP)



Different cuts on
different distributions
to suppress the bkg



ML learns the different
patterns for signal
and bkg

Our brains can understand the difference between two distributions or more, but the machine can understand the difference between each bin in the different distributions

Convolution Neural Network (CNN)

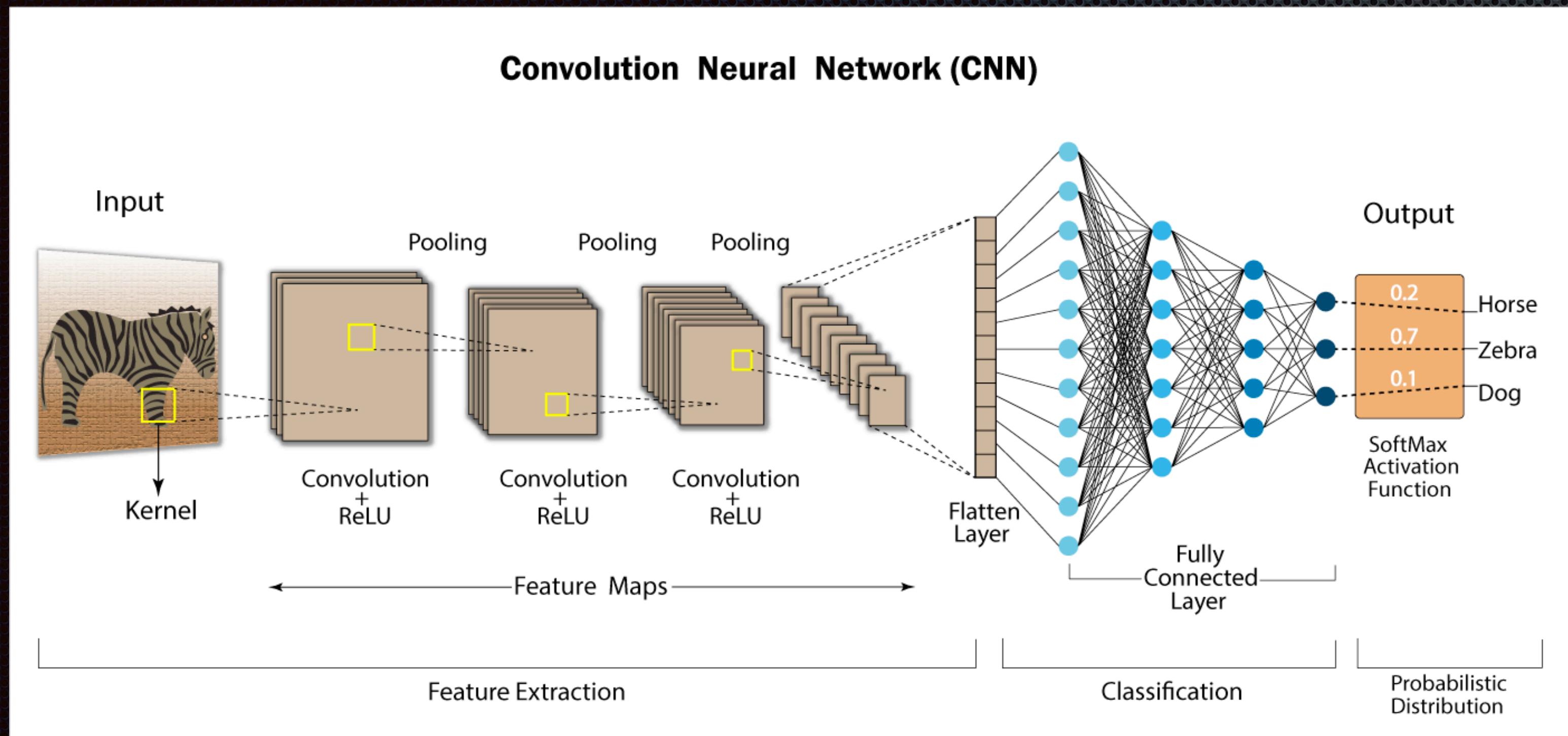


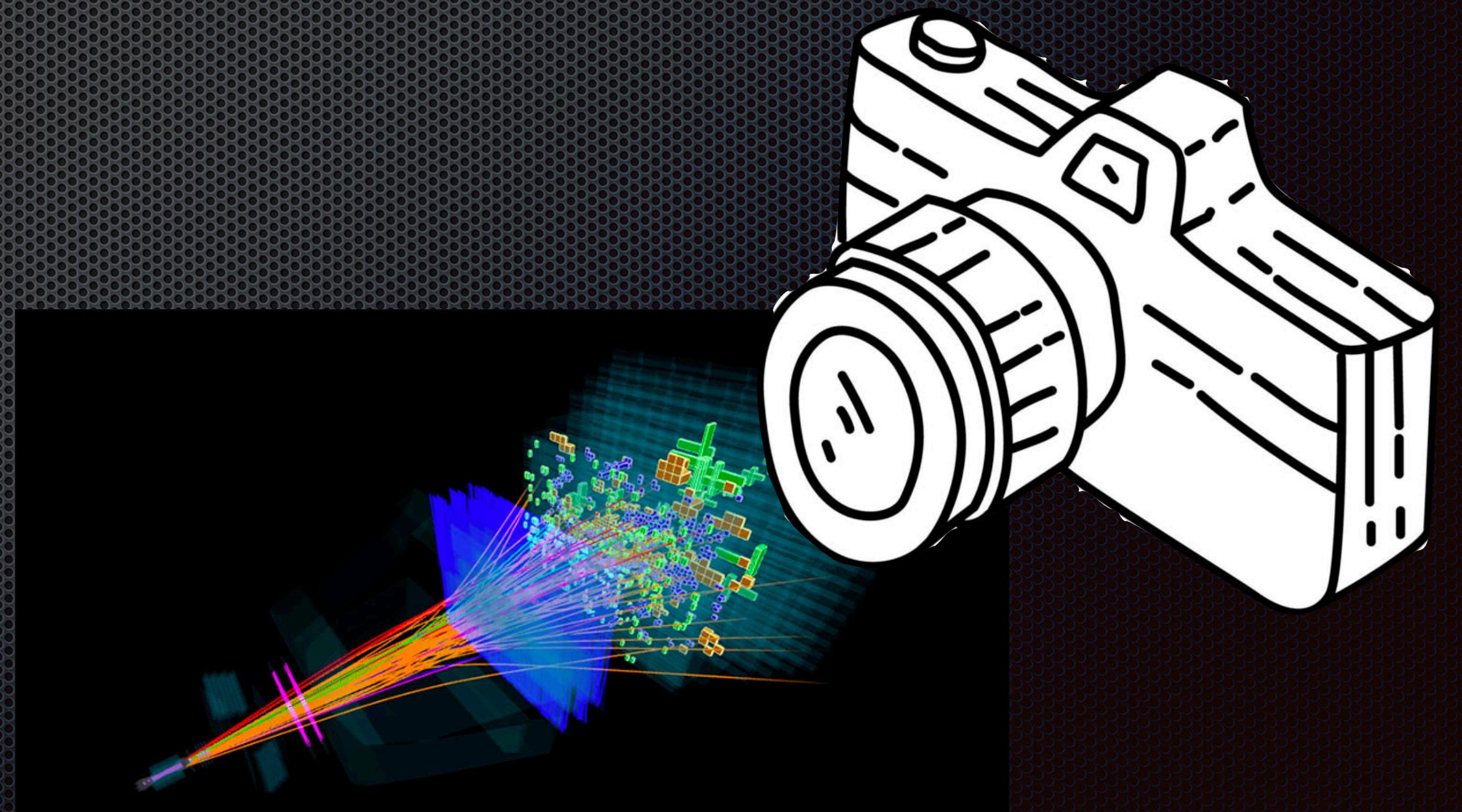
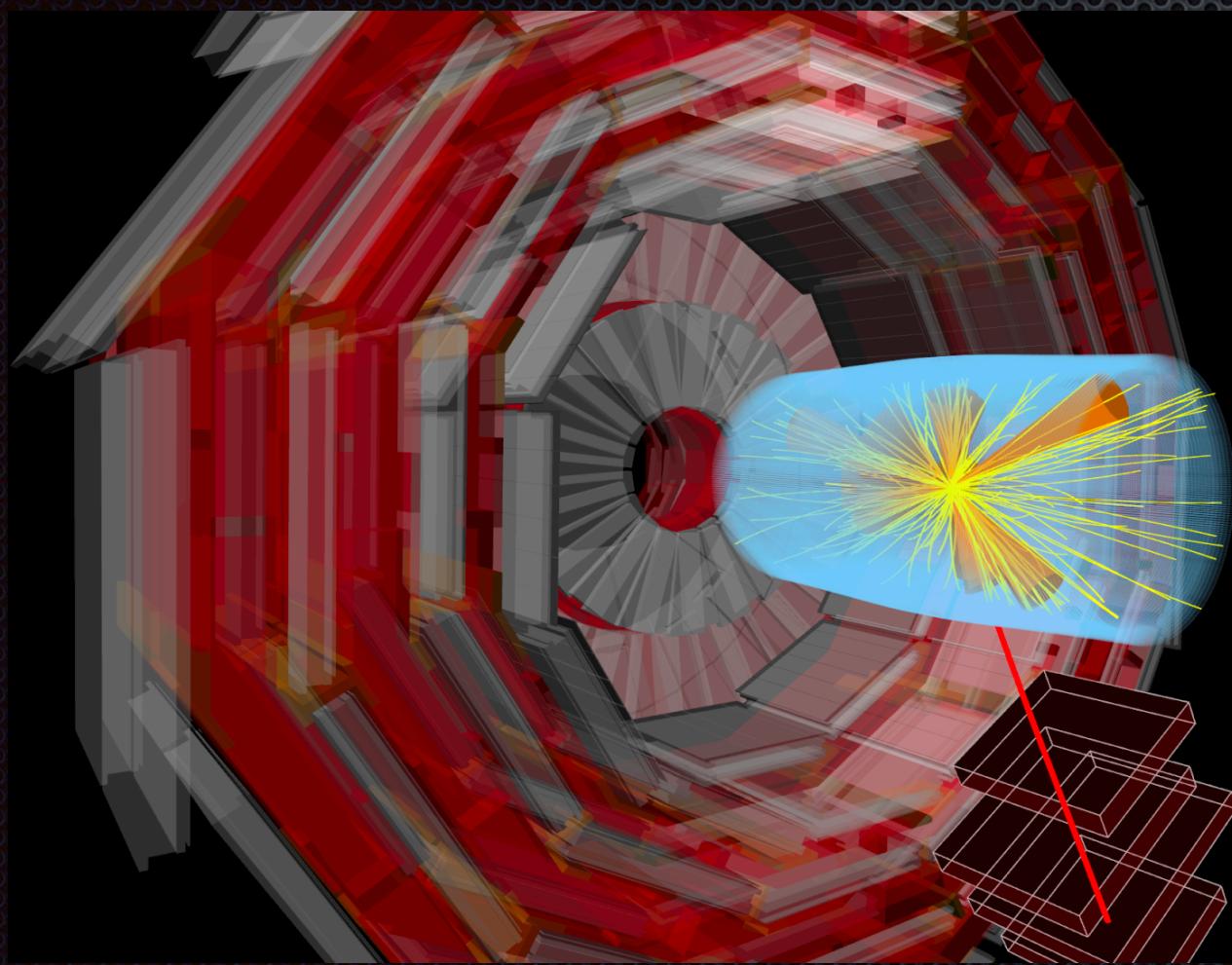
Image pixels can be transformed to matrix with probability of each pixel from 0 to 1 according to the pixel brightness, colors, etc

0	0	0	0	0	0	0
0	1	0	0	0	1	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	1	0	0	0	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0

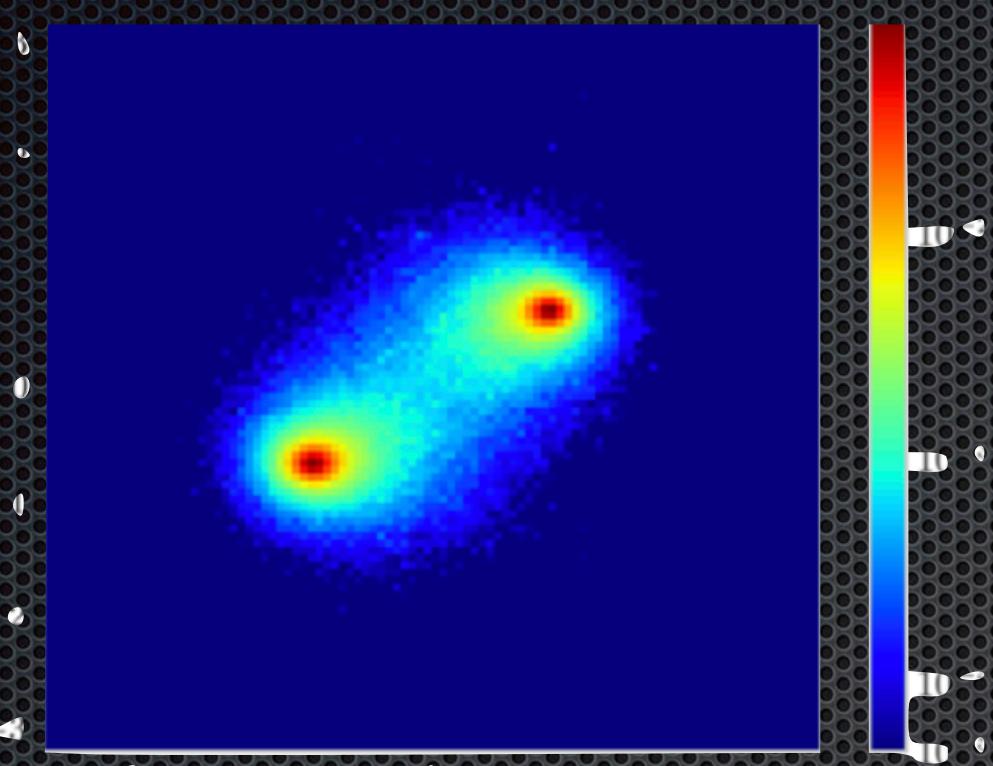
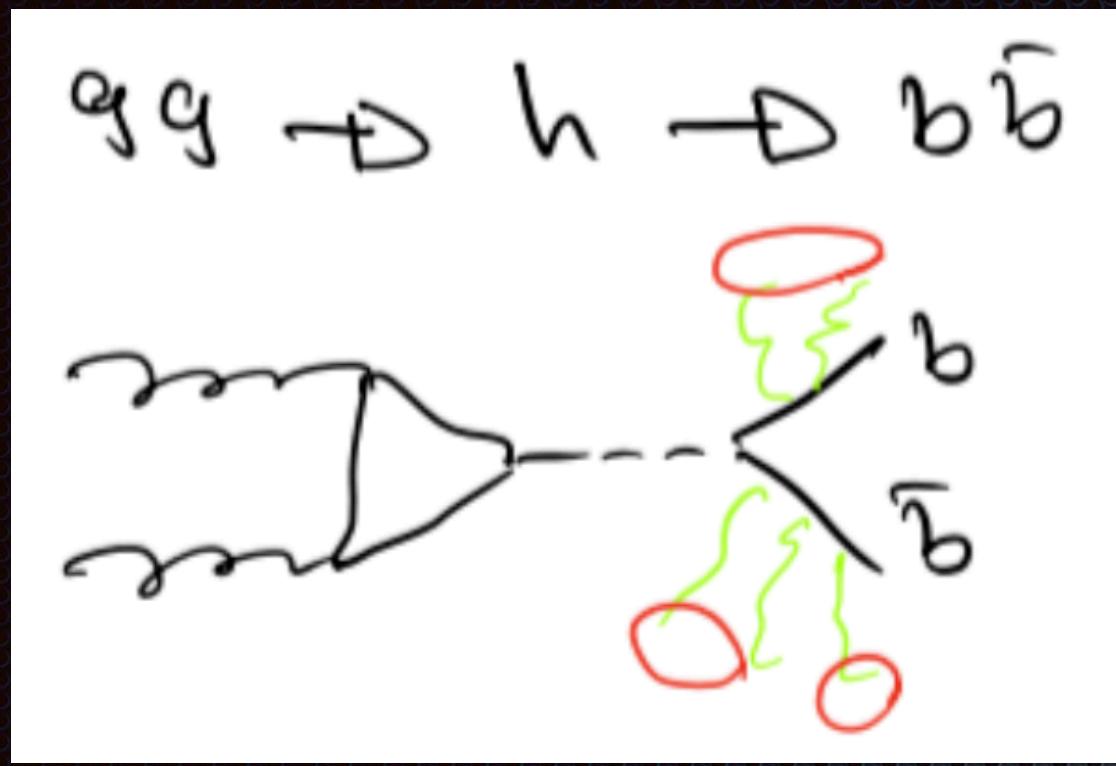
Input Image

Convolution Neural Network (CNN)

Consider the LHC detector as a giant live camera that picture the events in the eta-phi plane.
This way we can construct images with pixel intensity equals the detector strips and each image pixel can be weighted by the energy deposit of each particle.



QCD color flow

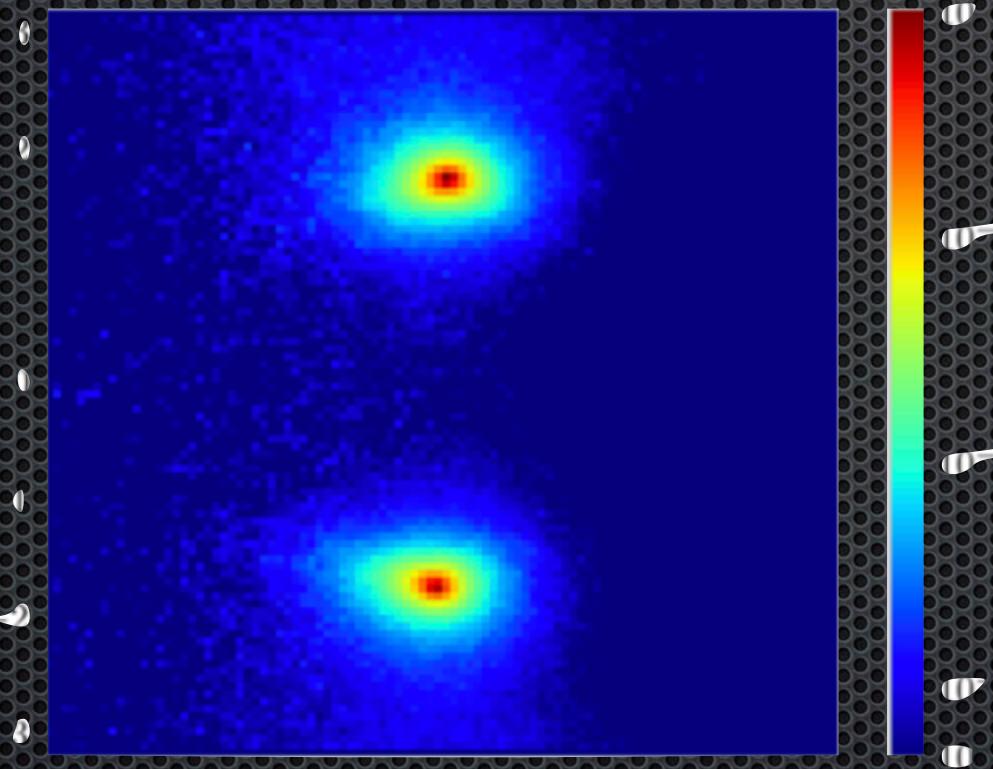
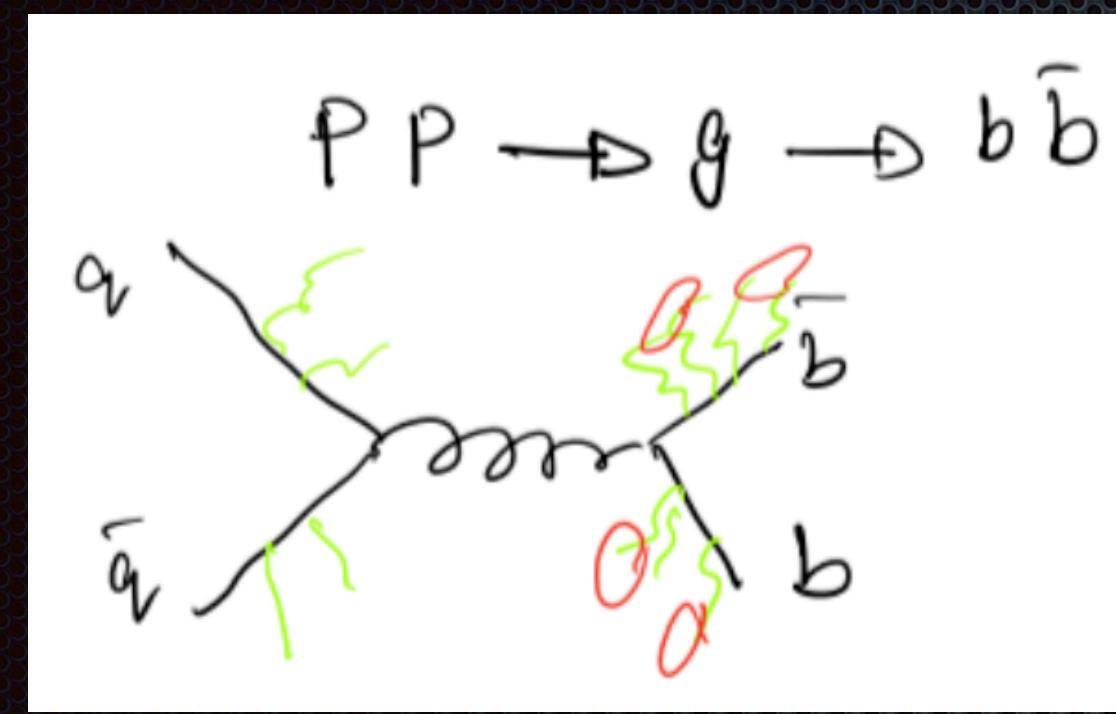


$$\alpha \text{Tr}[\tau^A \tau^C] \text{Tr}[\tau^B \tau^D]$$

$$\sim \delta^{Ac} \delta^{BD}$$

$$\alpha \text{Tr}[\tau^A \tau^B] \text{Tr}[\tau^C \tau^D]$$

$$\sim \delta^{AB} \delta^{CD}$$

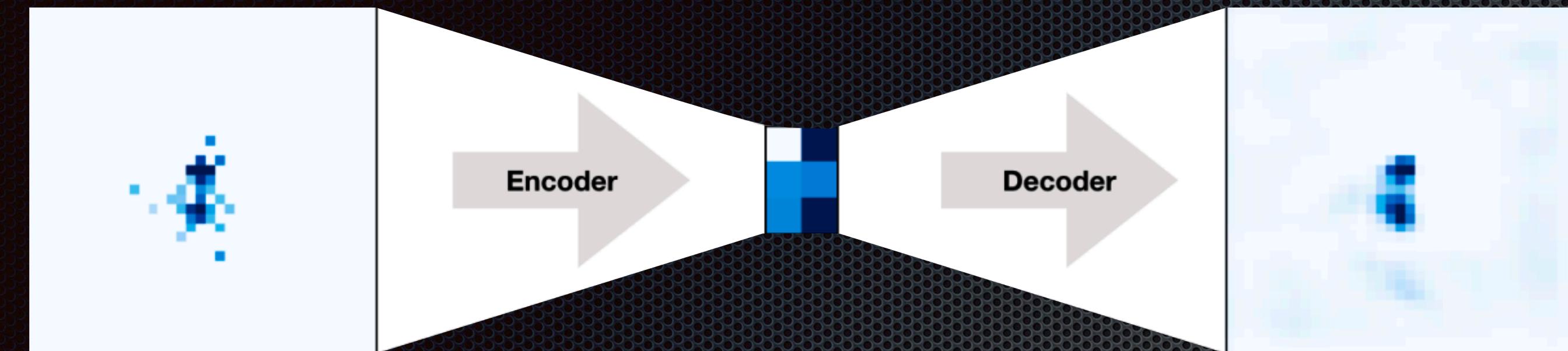


$$\frac{1}{\sqrt{2}} \delta_i^{l'} \delta_{k'}^j \quad \frac{1}{\sqrt{2}} \delta_{l'}^l \delta_k^{k'}$$

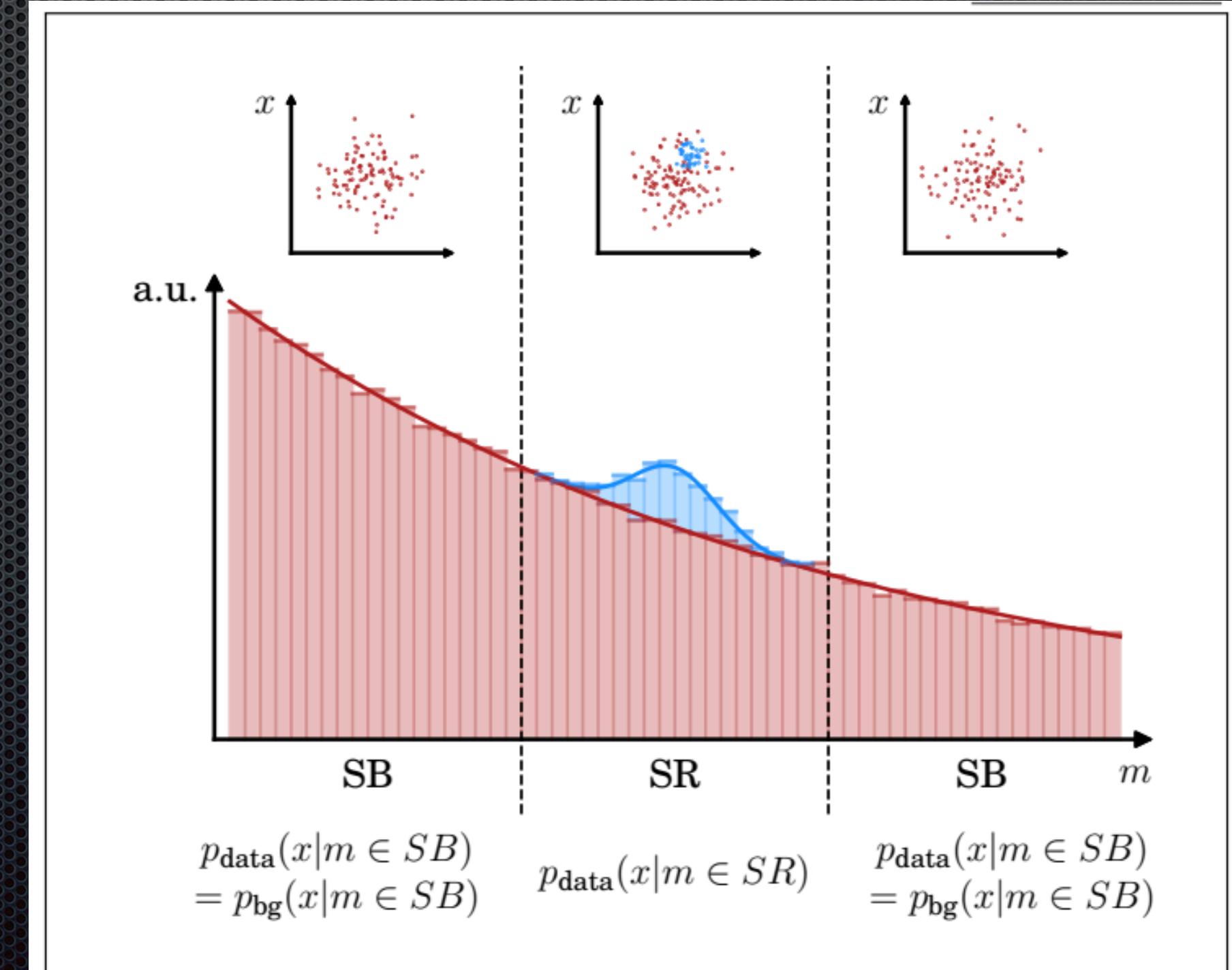
$$\frac{-1}{\sqrt{2}} \delta_i^j \quad \left(\frac{-1}{N}\right) \quad \frac{-1}{\sqrt{2}} \delta_k^l$$

Auto-Encoders for anomaly detection

Arxiv:1808.08992



For anomaly detection we train the ML model
to learn the features of the background events only.



Graph neural network (GNN)

Mostly the constructed jet images are sparse !!
(random zeros every where in the image)

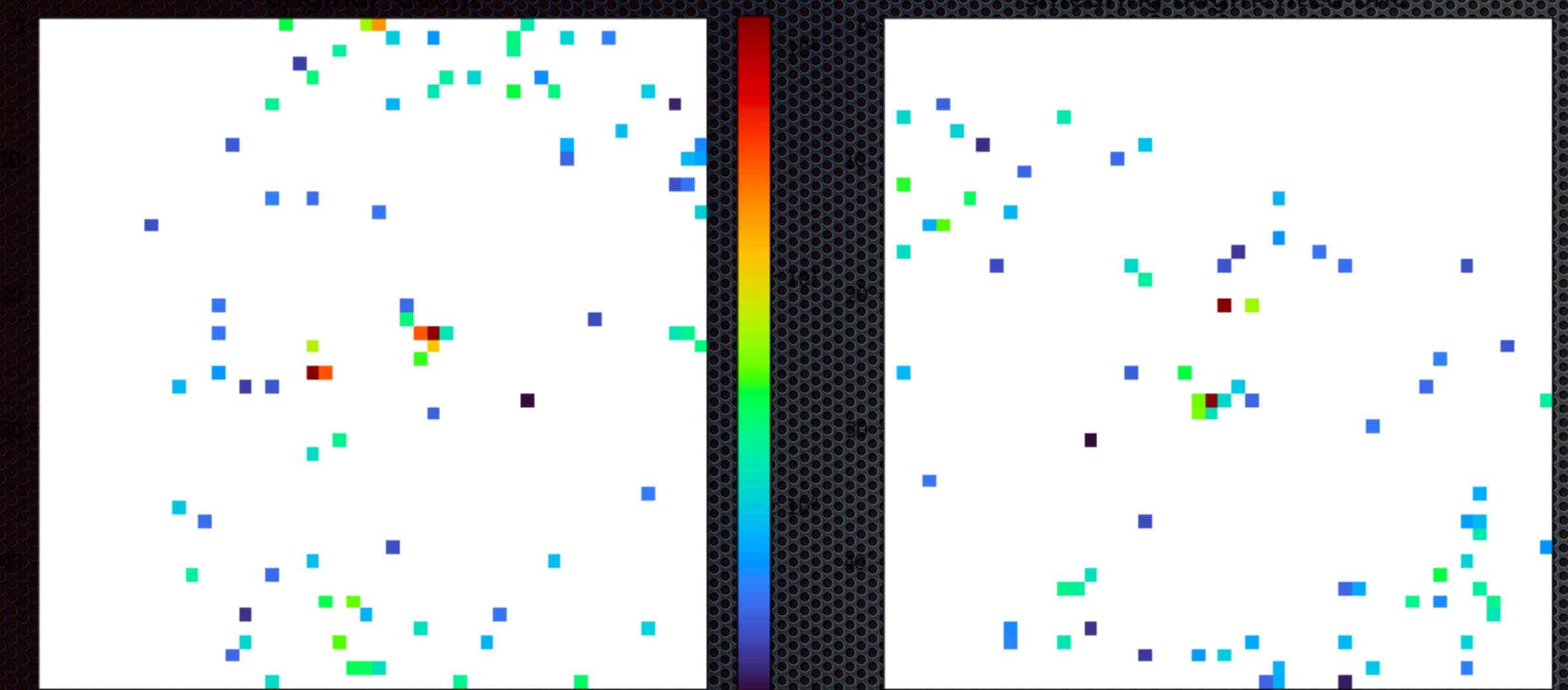
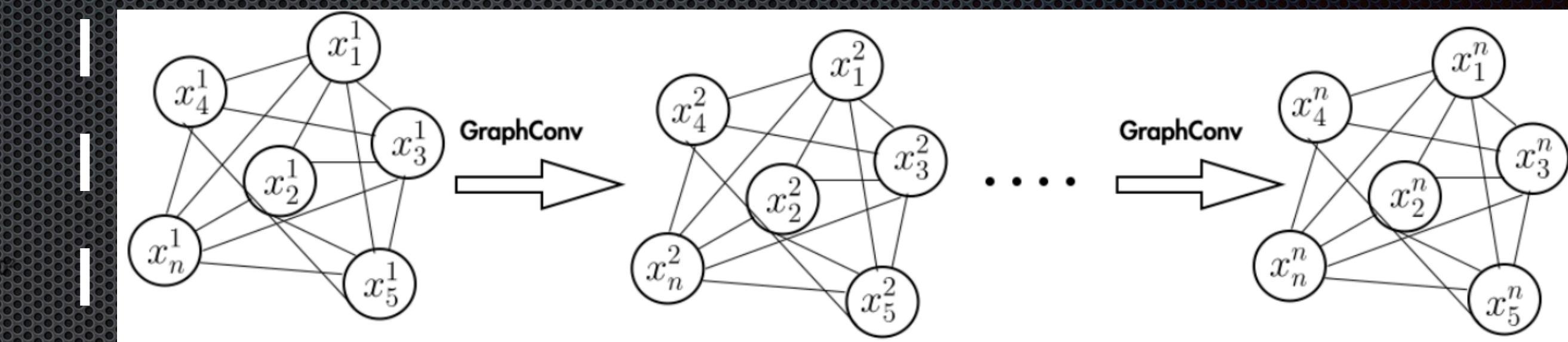


Image dimension: (50x50X1)

2500 pixel, few are filled !!

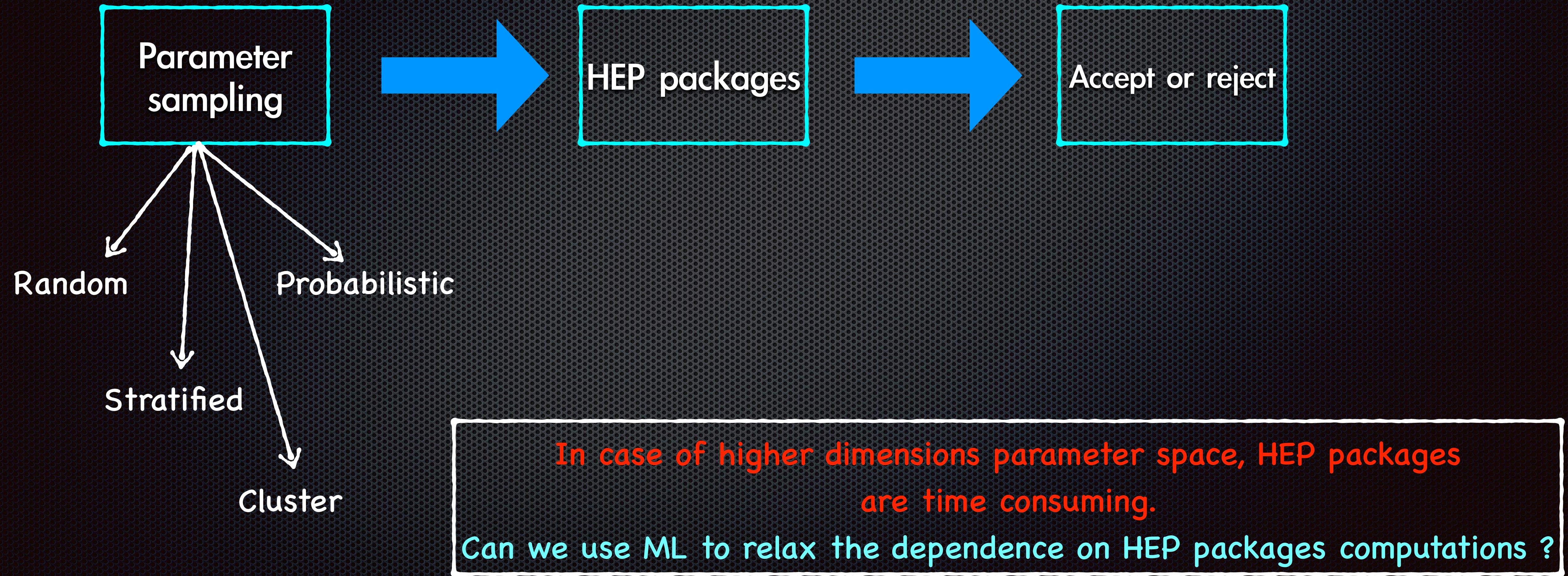


Graph nodes: particle four momenta

Graph edges: angular distance between particles

Parameter space sampling

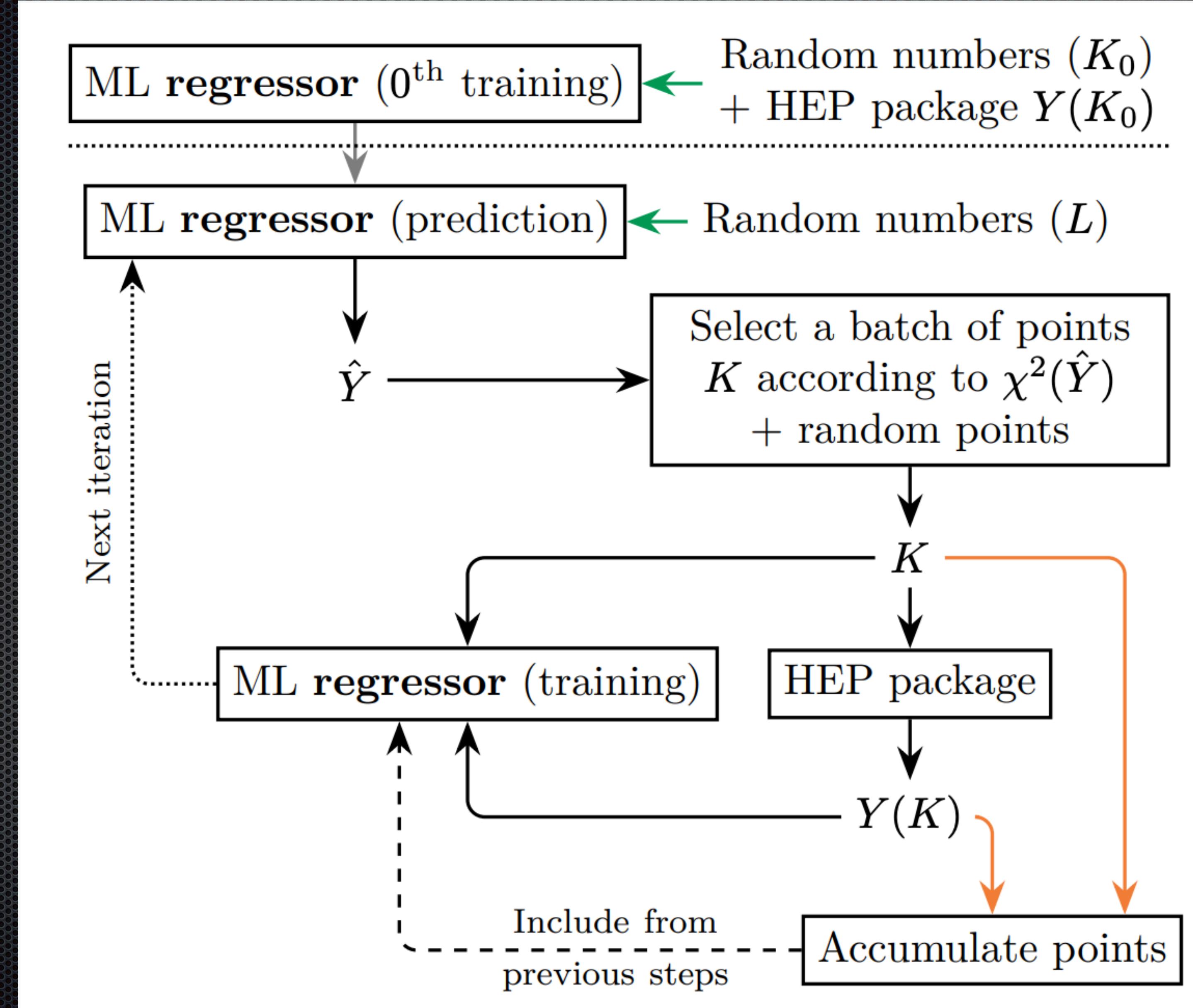
Find the values of Lagrangian free parameters that satisfy some measurements



Parameter space sampling

Arxiv:2207.09959

Code:
<https://github.com/AHamamdi150/MLscanner>

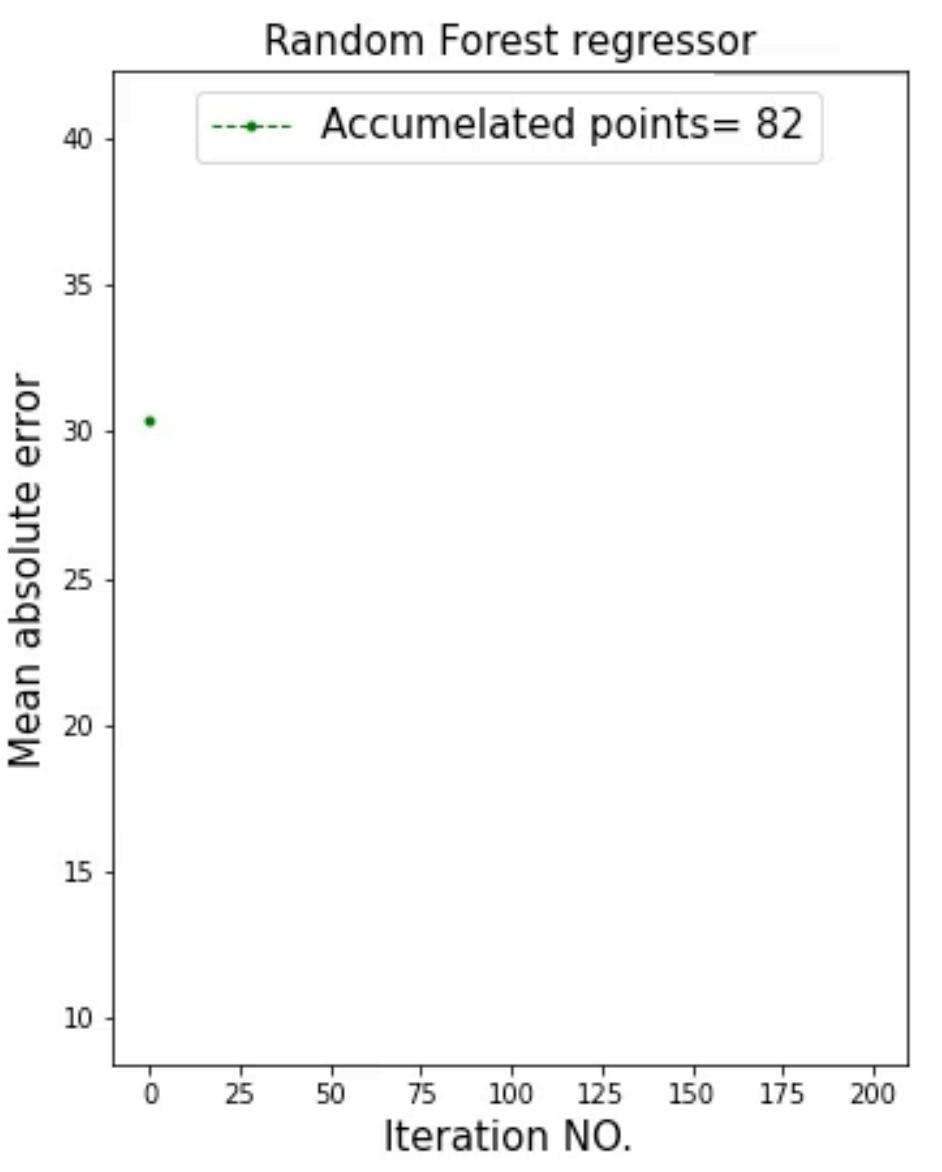
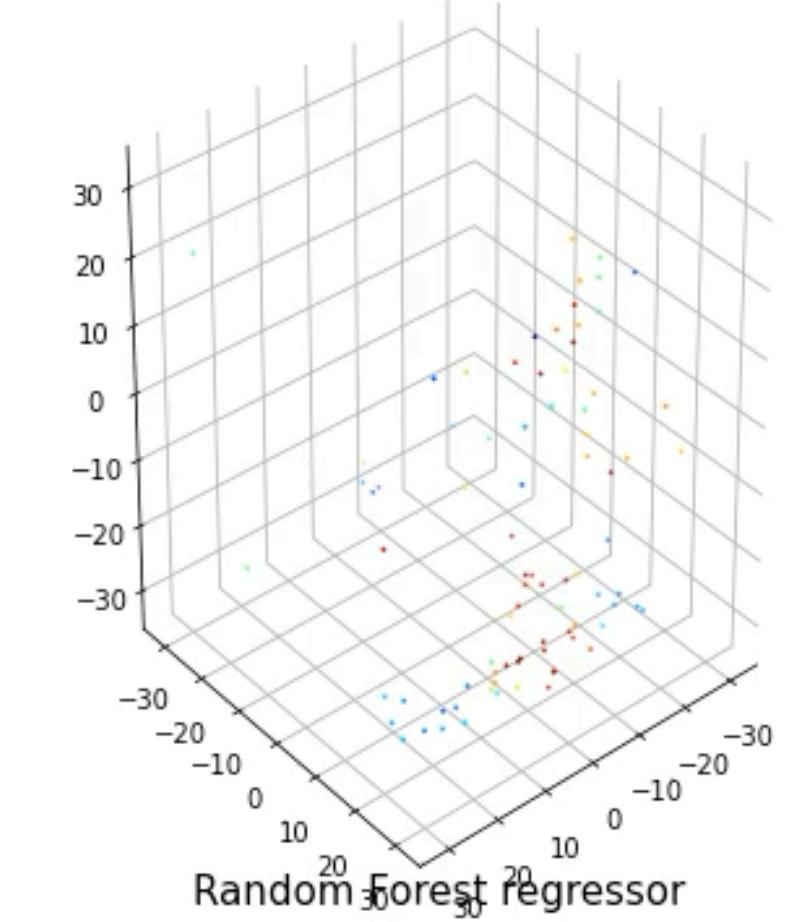
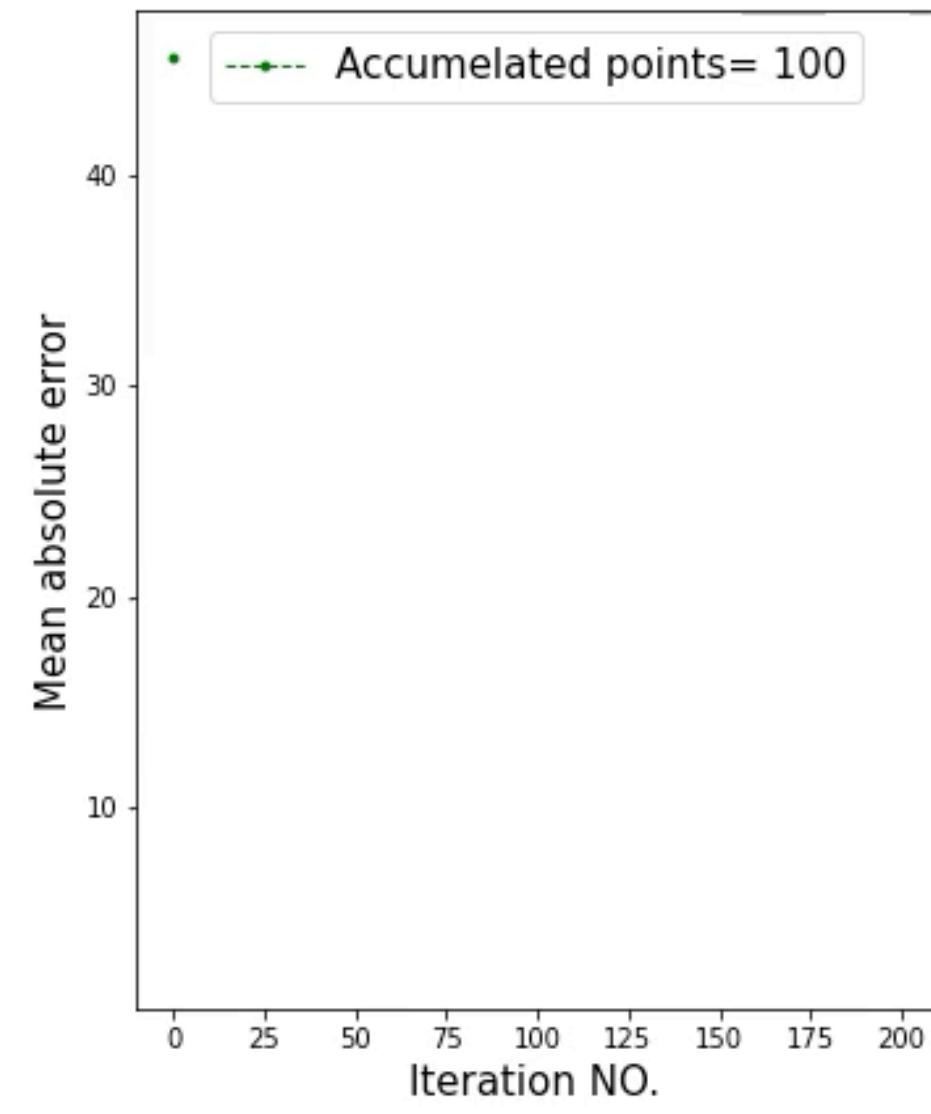
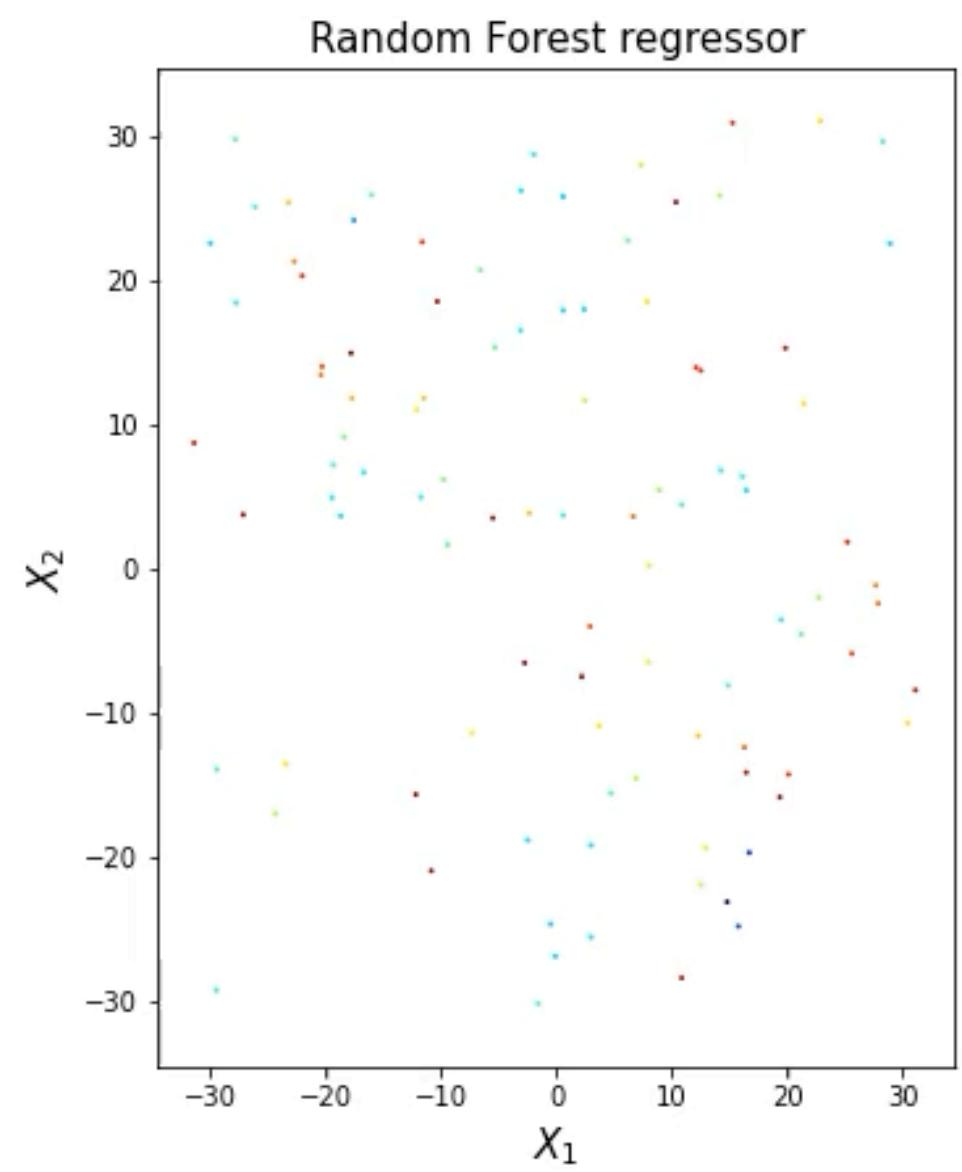


Parameter space sampling

Toy Examples:

$$\left[2 + \cos\left(\frac{x_1}{5}\right) \cos\left(\frac{x_2}{7}\right)\right]^5 = 100 \pm 5$$

$$\left[2 + \cos\left(\frac{x_1}{7}\right) \cos\left(\frac{x_2}{7}\right) \cos\left(\frac{x_3}{7}\right)\right]^5 = 100 \pm 5$$



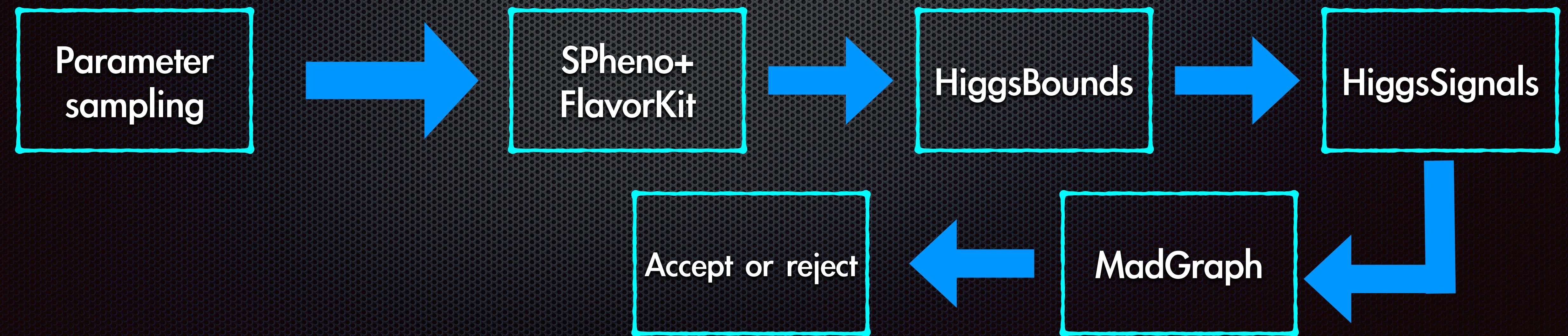
Parameter space sampling

THDM Example:

$$V_\phi = m_{11}^2(\phi_1^\dagger \phi_1) + m_{22}^2(\phi_2^\dagger \phi_2) - [m_{12}^2(\phi_1^\dagger \phi_2) + \text{h.c.}] + \lambda_1(\phi_1^\dagger \phi_1)^2 + \lambda_2(\phi_2^\dagger \phi_2)^2 \\ + \lambda_3(\phi_1^\dagger \phi_1)(\phi_2^\dagger \phi_2) + \lambda_4(\phi_1^\dagger \phi_2)(\phi_2^\dagger \phi_1) + \frac{1}{2} [\lambda_5(\phi_1^\dagger \phi_2)^2 + \text{H.c.}] ,$$

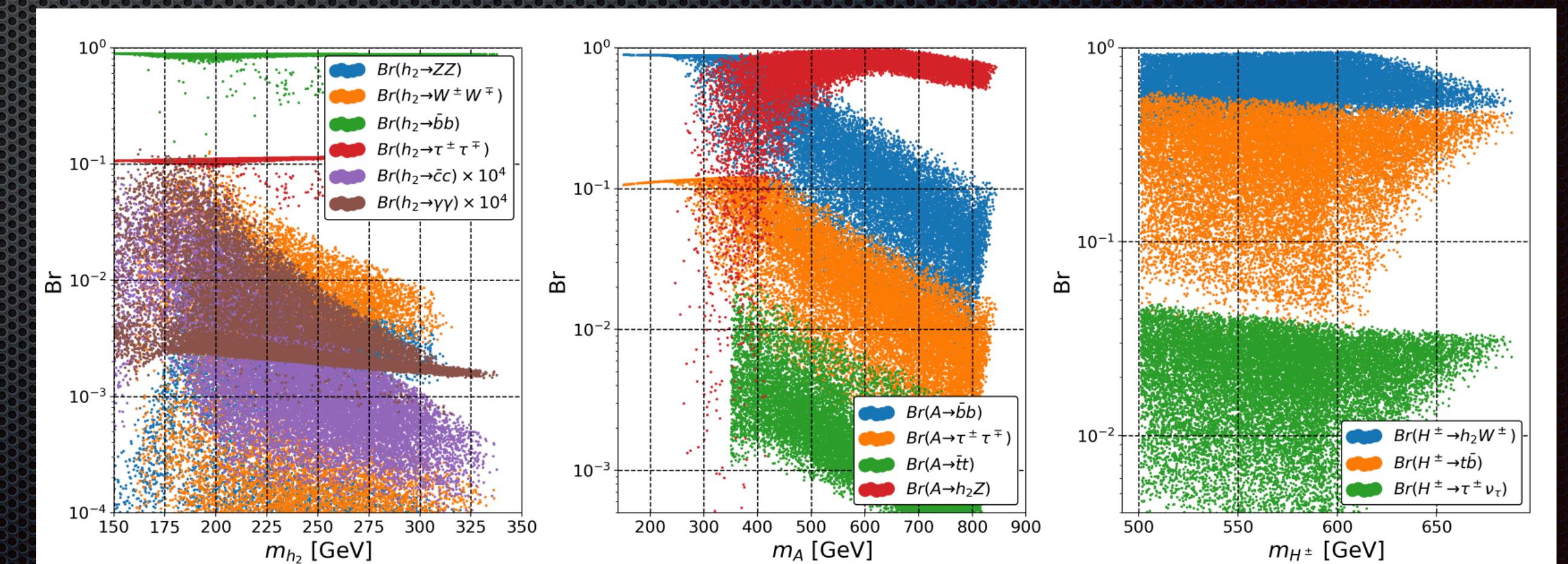
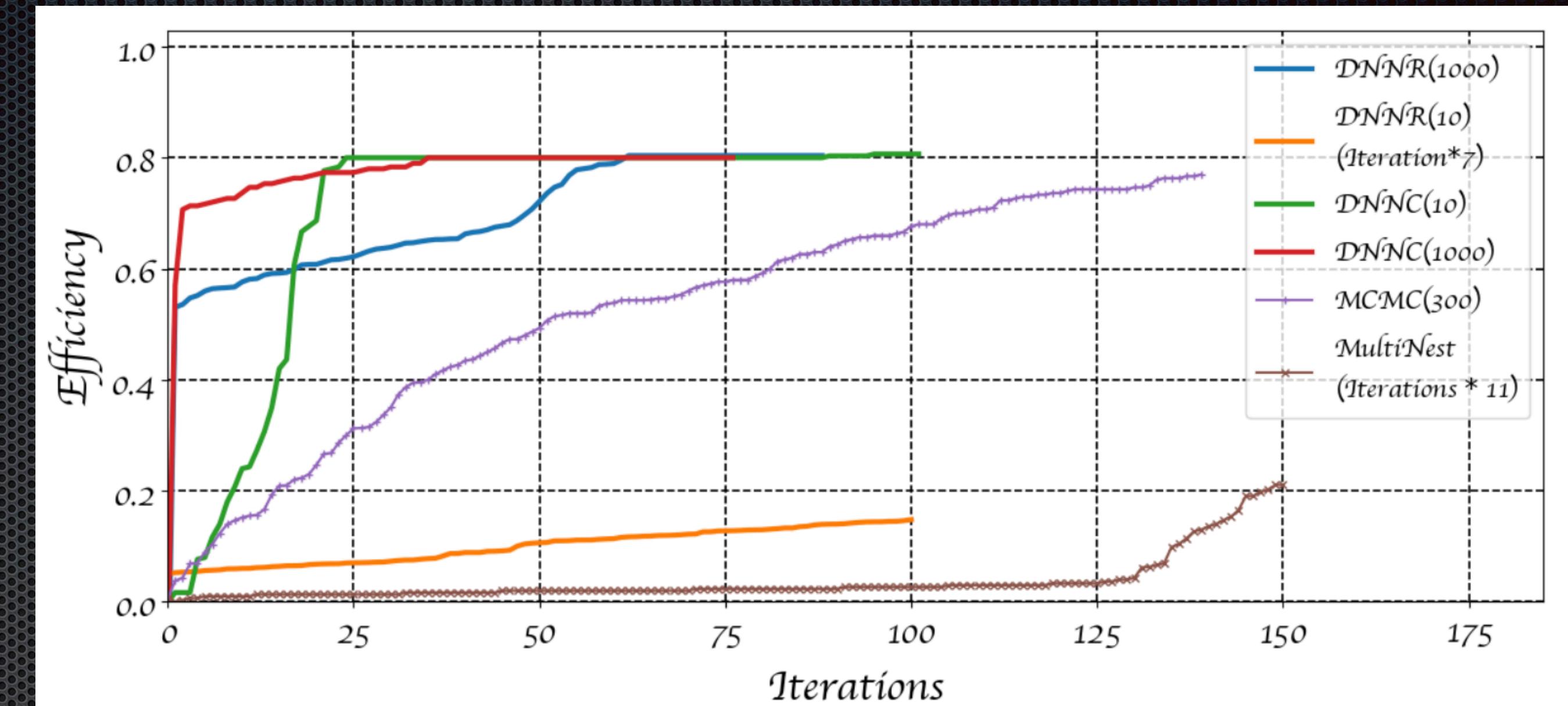
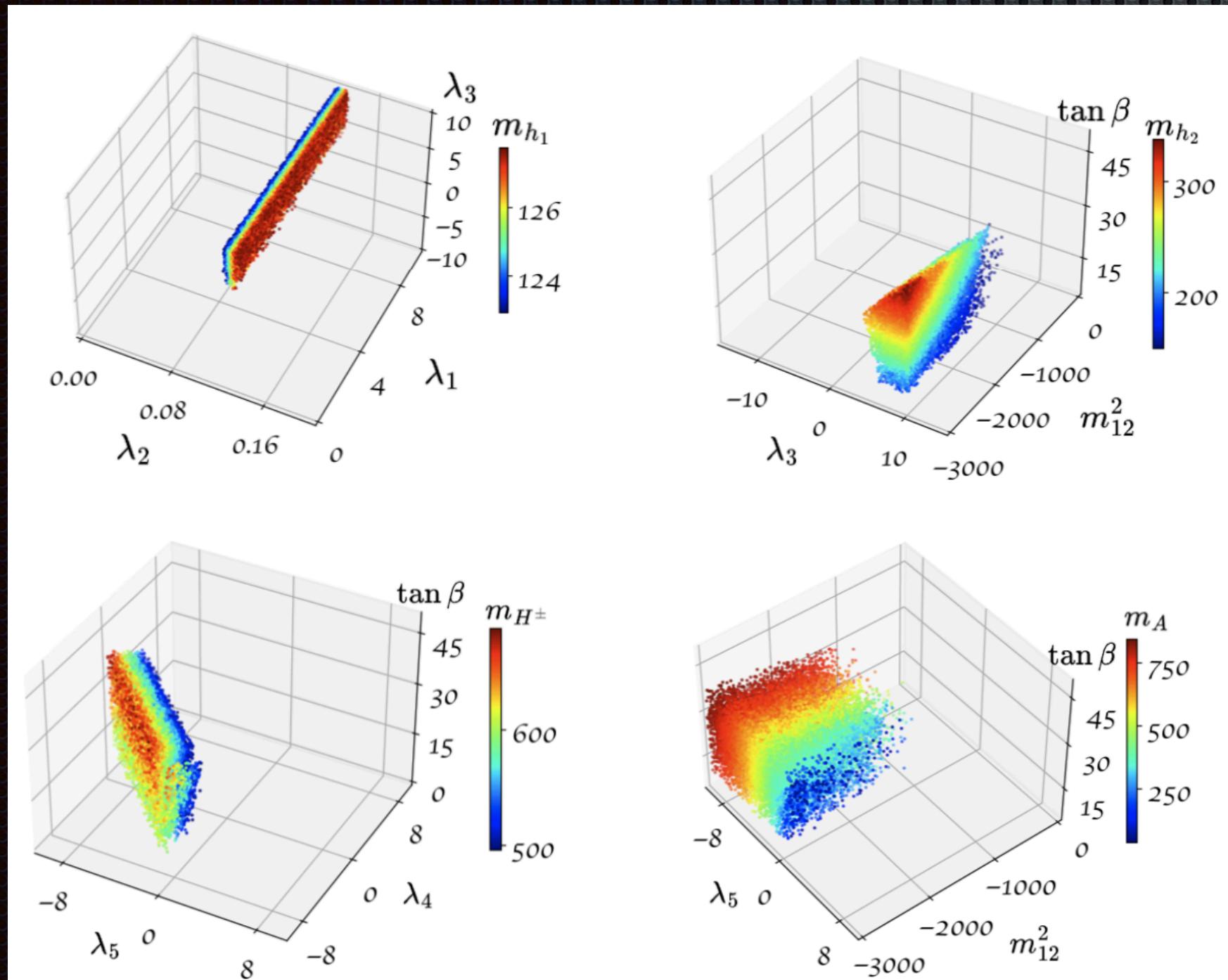
Parameter space of dimensions 7

$$0 \leq \lambda_1 \leq 10, \quad 0 \leq \lambda_2 \leq 0.2, \quad -10 \leq \lambda_3 \leq 10, \quad -10 \leq \lambda_4 \leq 10, \\ -10 \leq \lambda_5 \leq 10, \quad 5 \leq \tan \beta \leq 45, \quad -3000 \text{ GeV}^2 \leq m_{12}^2 \leq 0 \text{ GeV}^2 ,$$



Parameter space sampling

THDM Example:

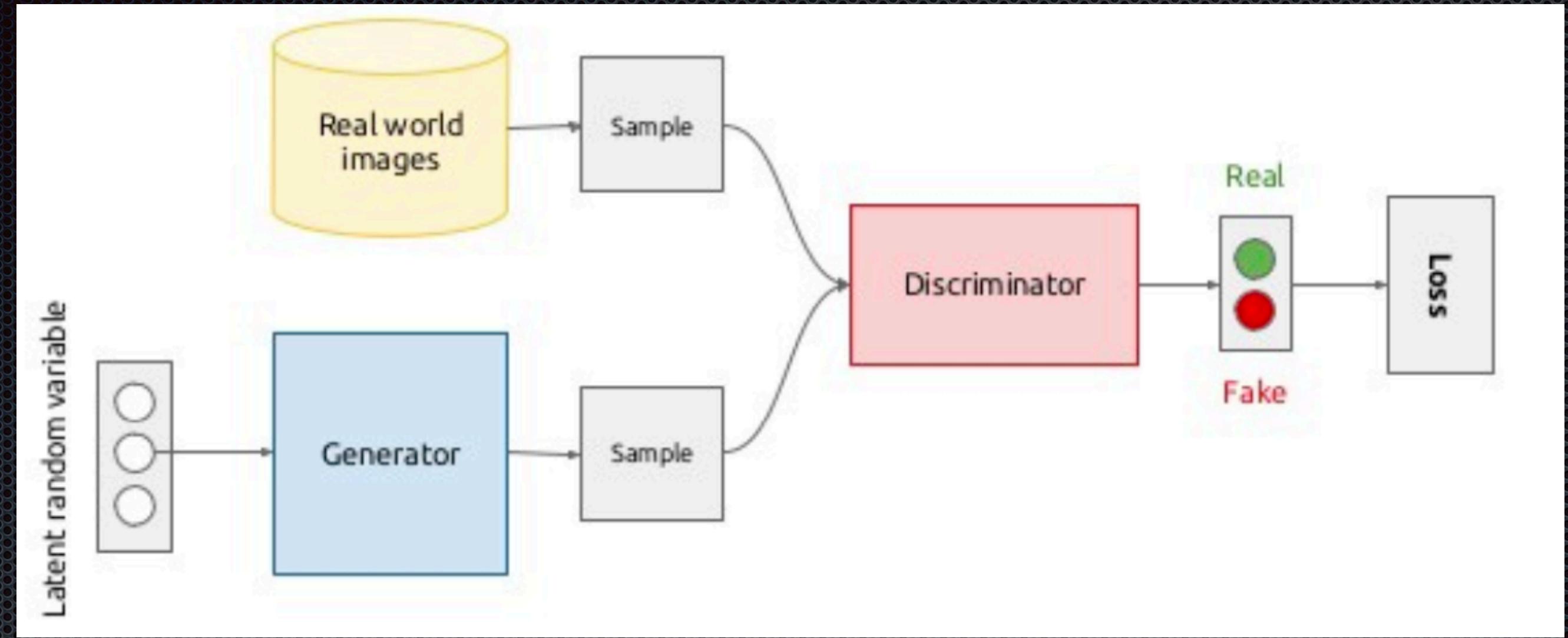


Generative ML models

These people do not exist in real !!



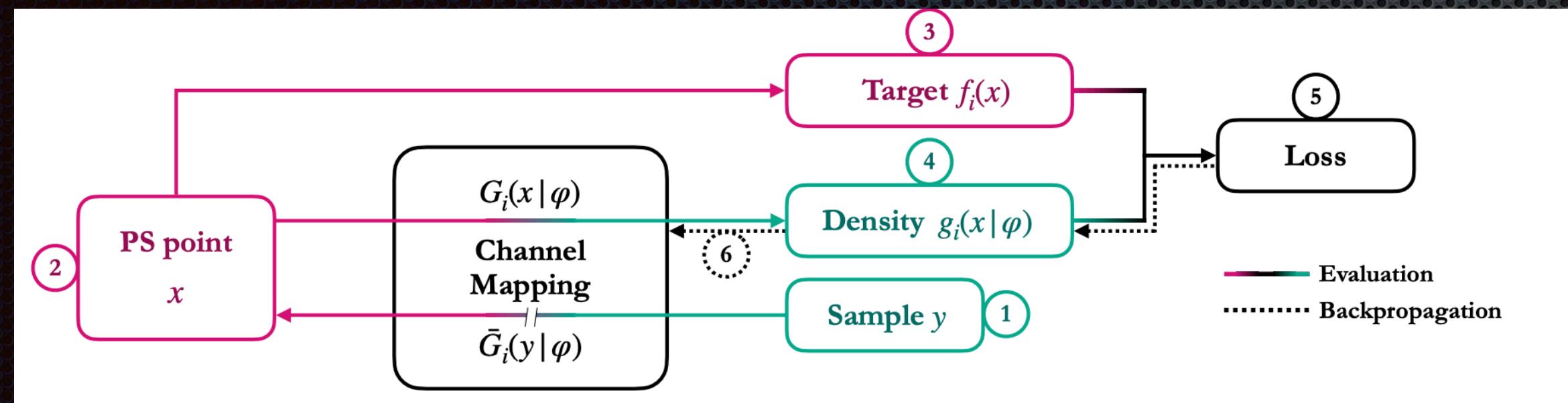
Generative ML models



Generative ML models

GANs are used for phase space integration and events generation

Used network



SciPost Physics

Submission

IRMP-CP3-22-56, MCNET-22-22, FERMILAB-PUB-22-915-T

MadNIS – Neural Multi-Channel Importance Sampling

Theo Heimel¹, Ramon Winterhalder²,
Anja Butter^{1,3}, Joshua Isaacson⁴, Claudio Krause¹,
Fabio Maltoni^{2,5}, Olivier Mattelaer², and Tilman Plehn¹

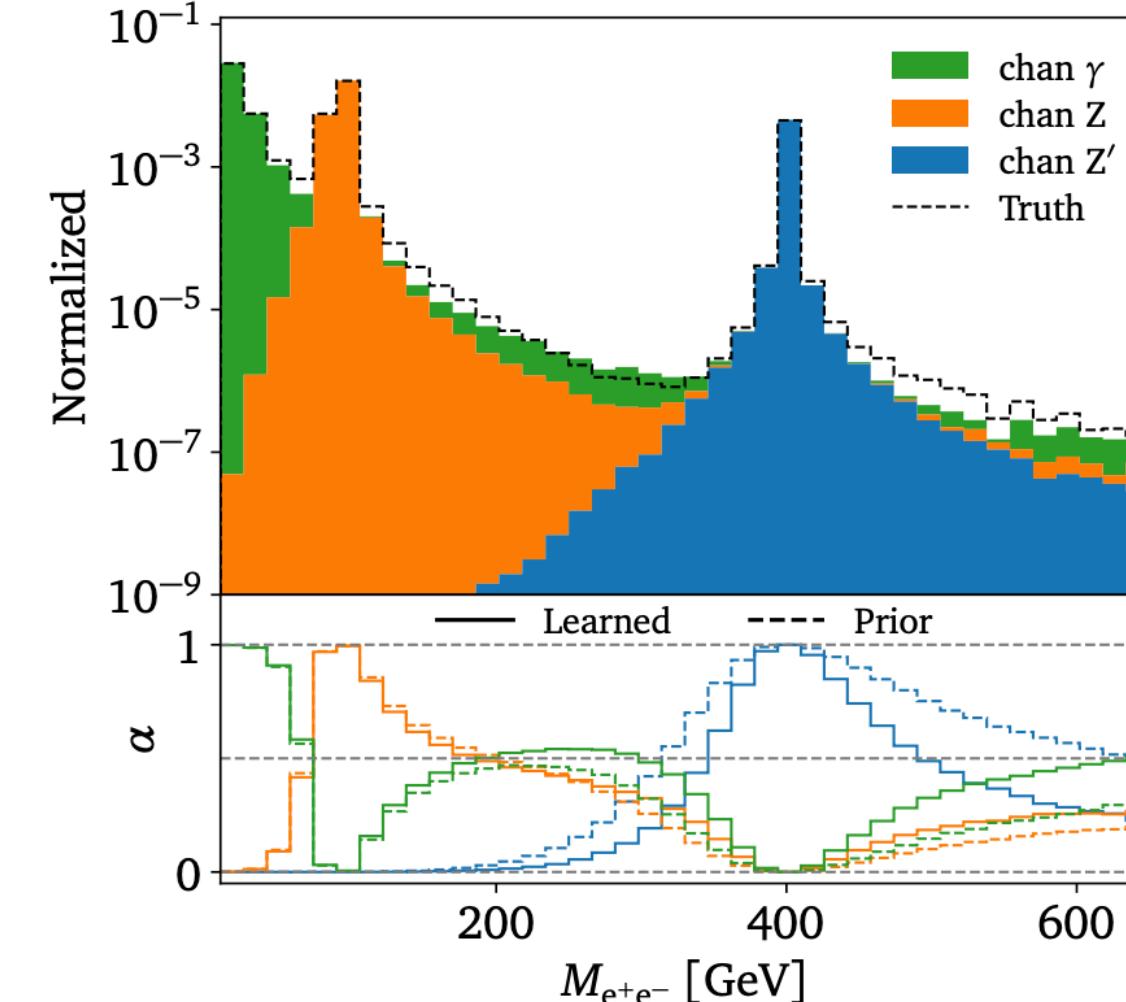
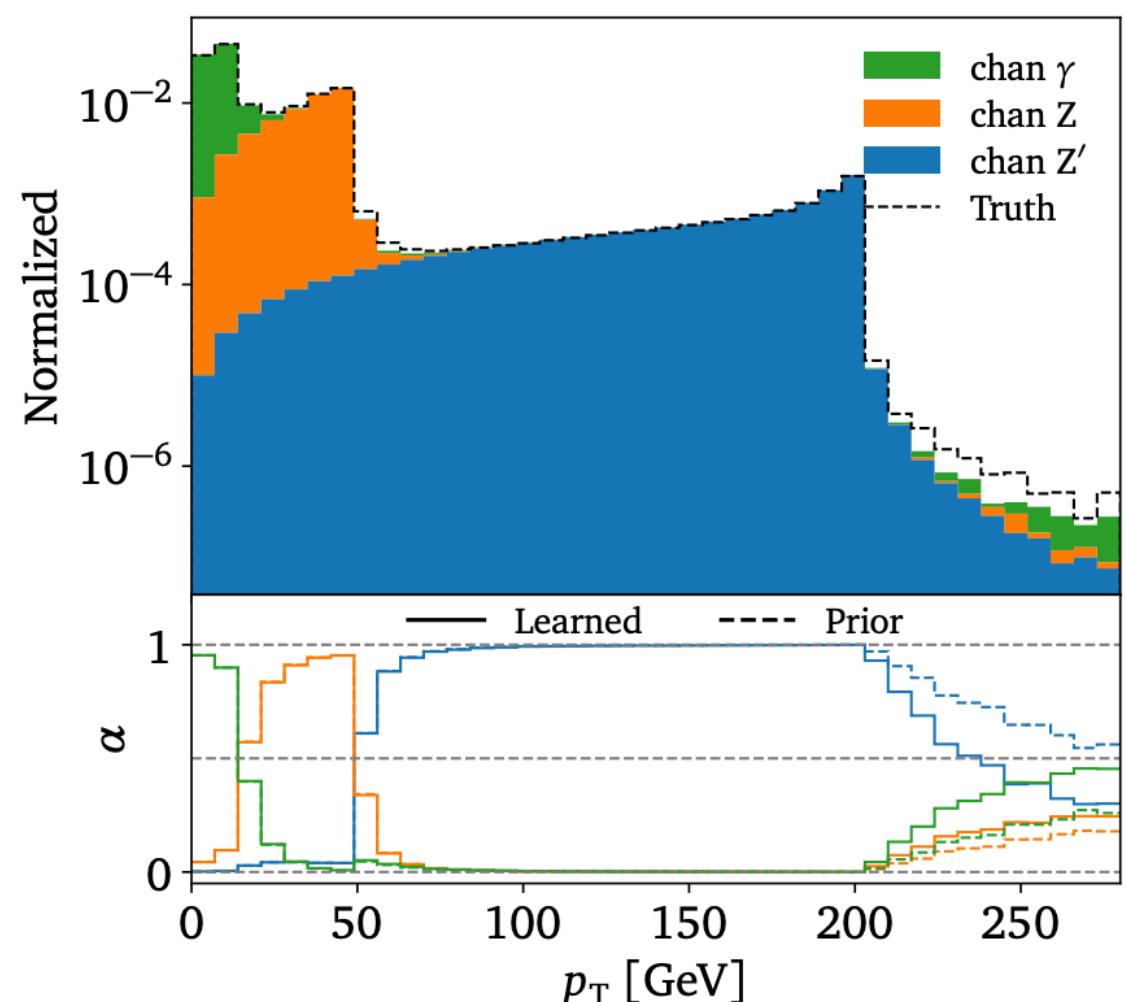
1 Institut für Theoretische Physik, Universität Heidelberg, Germany

2 CP3, Université catholique de Louvain, Louvain-la-Neuve, Belgium

3 LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, France

4 Theoretical Physics Division, Fermi National Accelerator Laboratory, Batavia, IL, USA

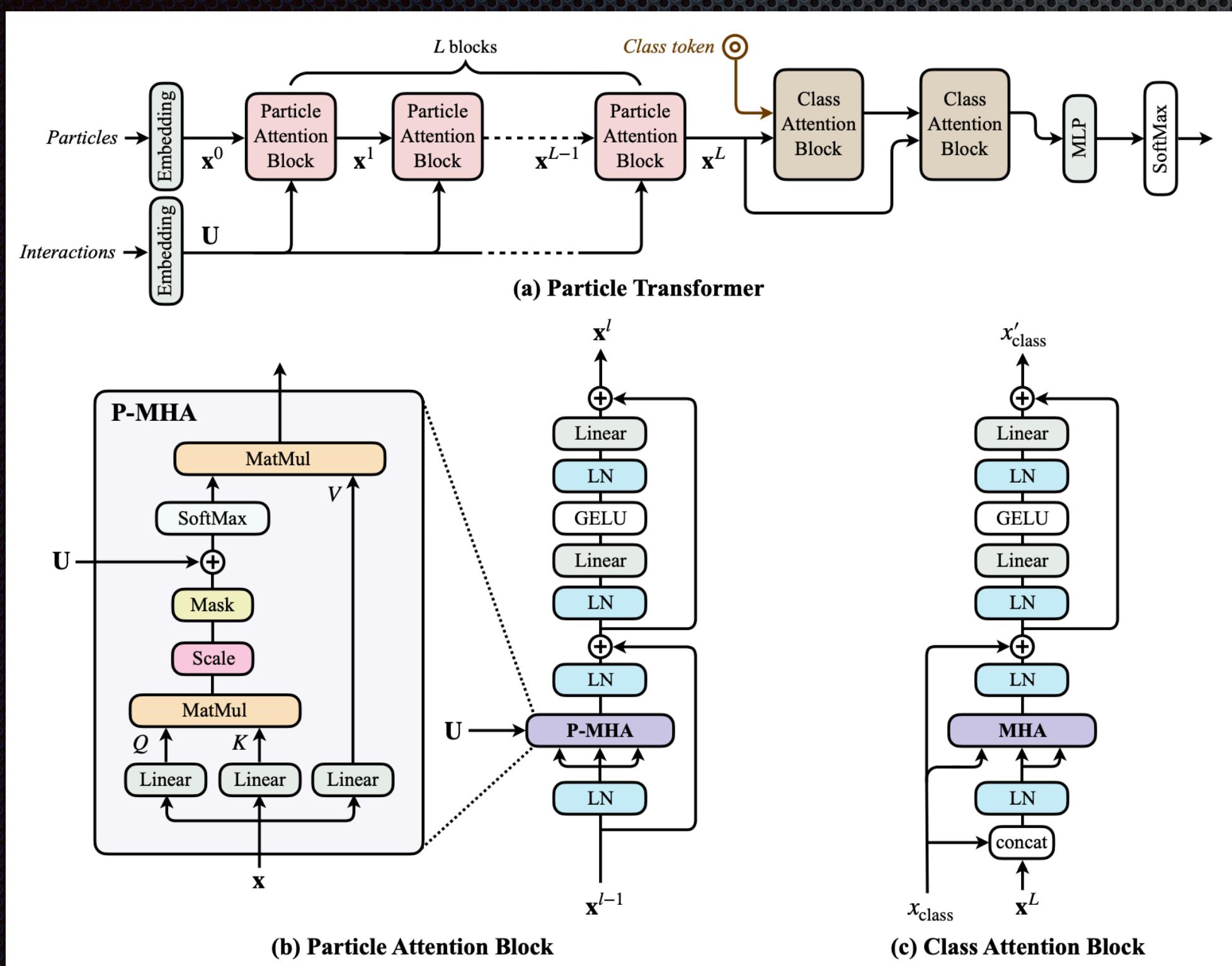
5 Dipartimento di Fisica e Astronomia, Università di Bologna, Italy



ramon.winterhalder@uclouvain.be

Transformer models

Particle transformer acts soon particle clouds and can achieve Higher accuracy

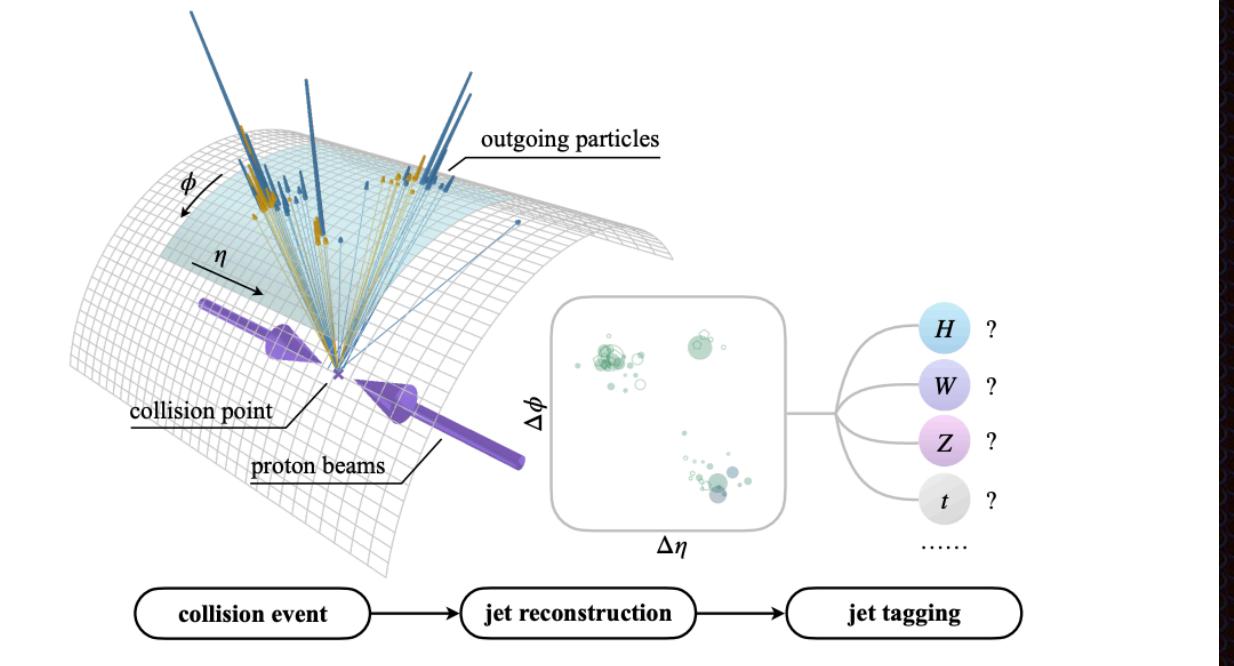


Particle Transformer for Jet Tagging

Huilin Qu¹ Congqiao Li² Sitian Qian²

Abstract

Jet tagging is a critical yet challenging classification task in particle physics. While deep learning has transformed jet tagging and significantly improved performance, the lack of a large-scale public dataset impedes further enhancement. In this work, we present JETCLASS, a new comprehensive dataset for jet tagging. The JETCLASS dataset consists of 100 M jets, about two orders of magnitude larger than existing public datasets. A total of 10 types of jets are simulated, including several types unexplored for



Quantum machine learning

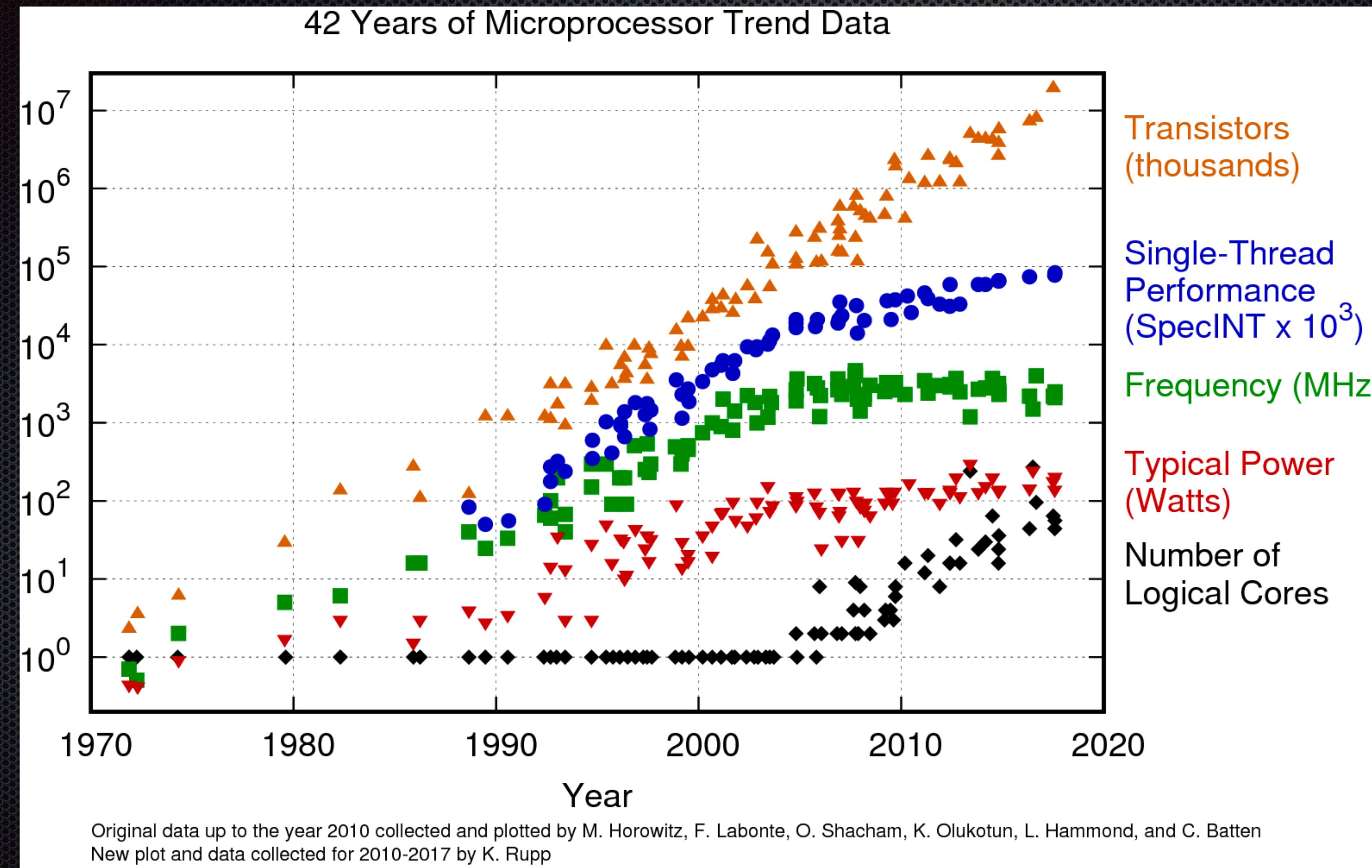
For the undergrad students, please don't panic.
You just need to get your hands dirty



What the hell this
guy is talking about !!

Quantum machine learning

Why do we need quantum computation



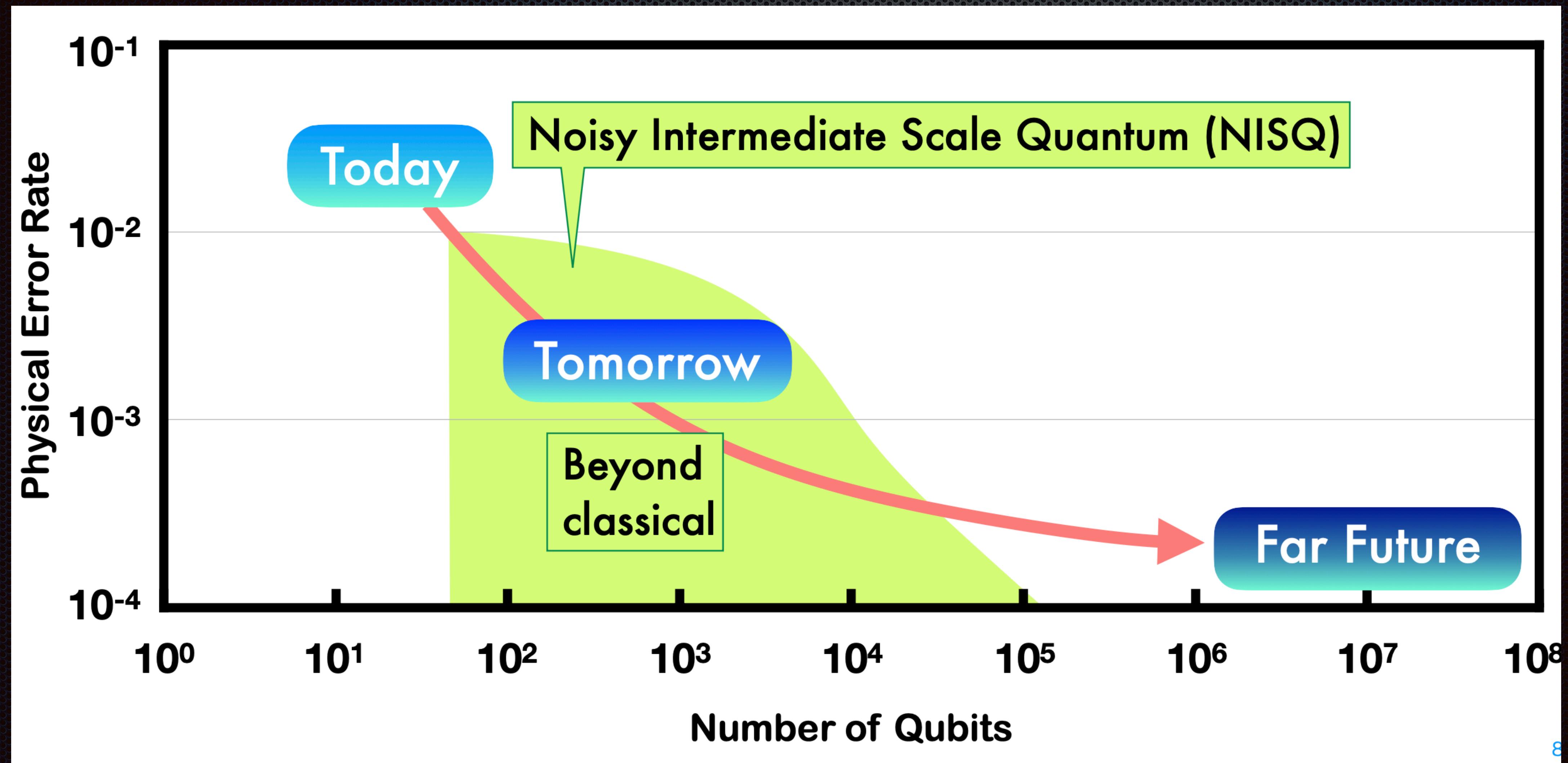
Quantum machine learning

Noisy intermediate quantum computers



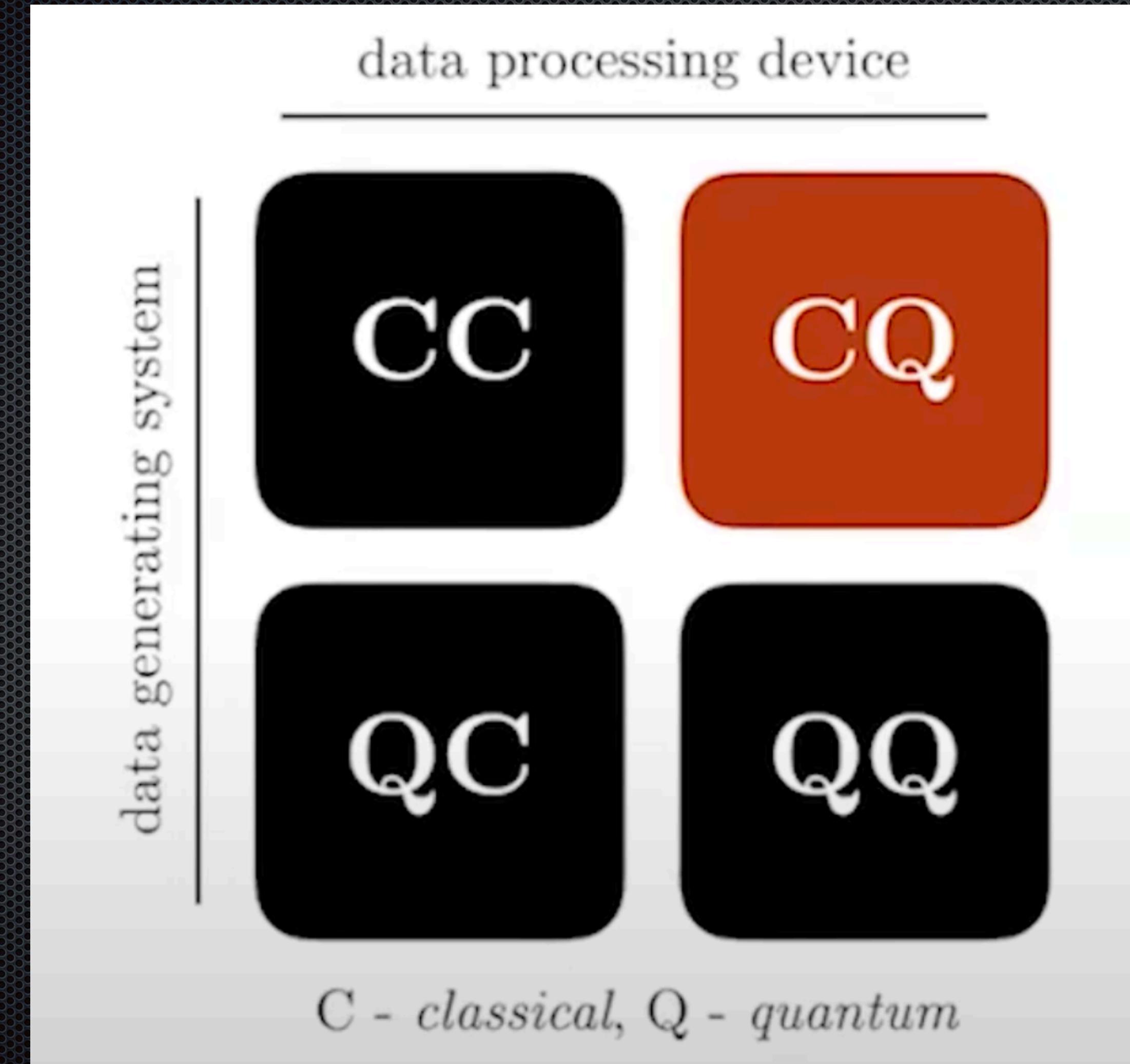
Quantum machine learning

From Daniel Park talk



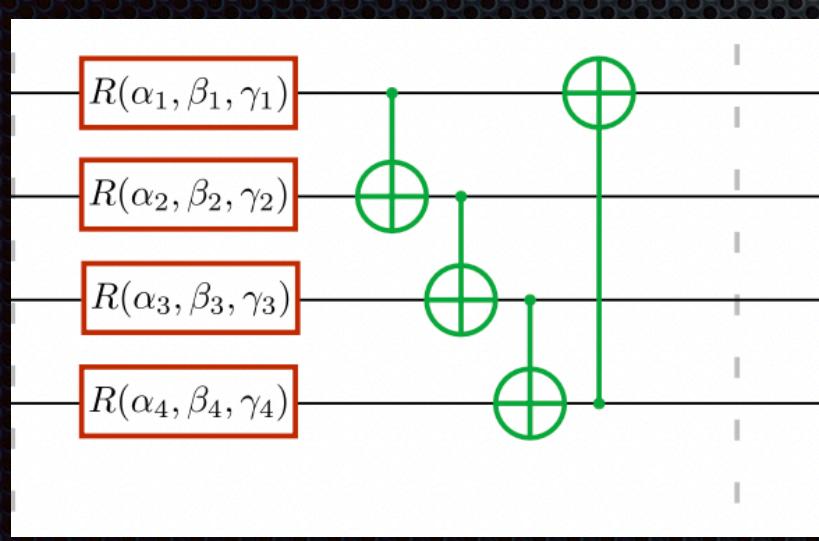
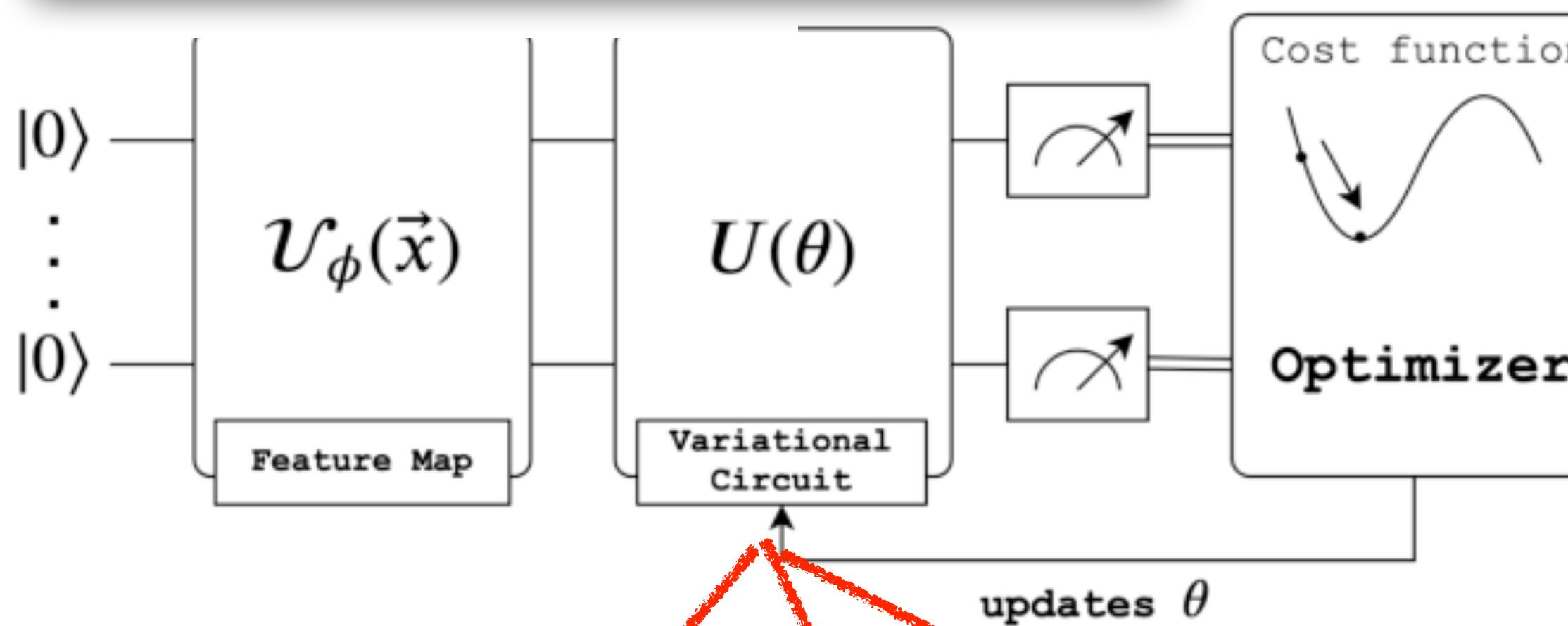
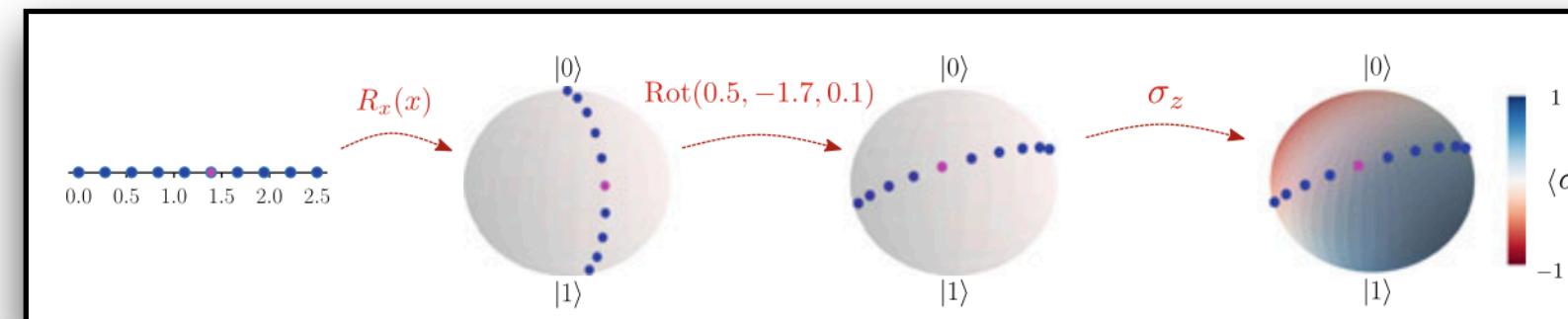
Quantum machine learning

What do we mean by quantum machine learning ?

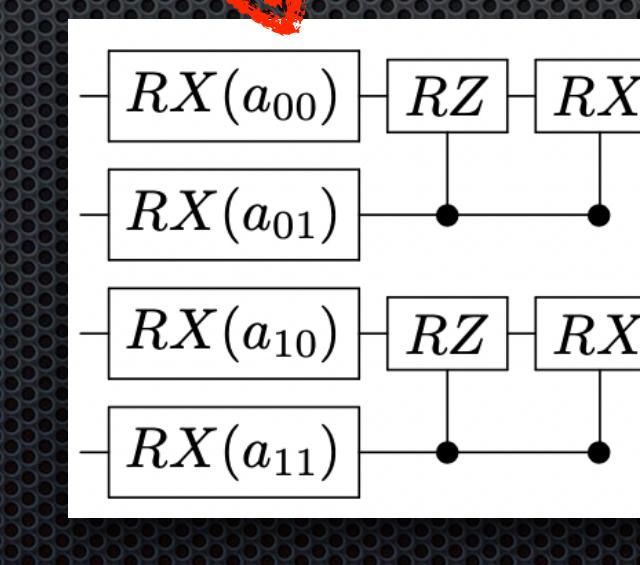


Quantum machine learning

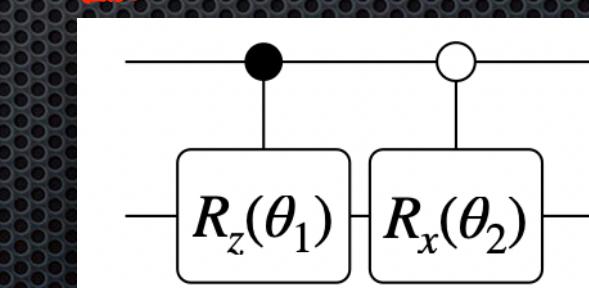
Quantum computers as ML models



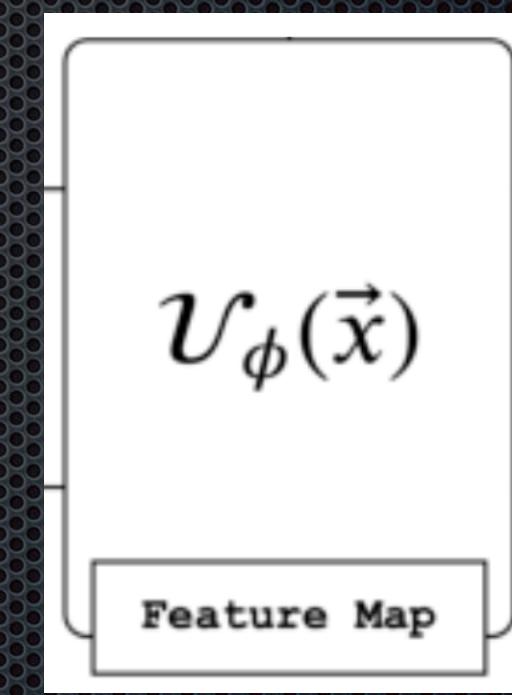
Fully connected



Conv circuit



Pooling circuit

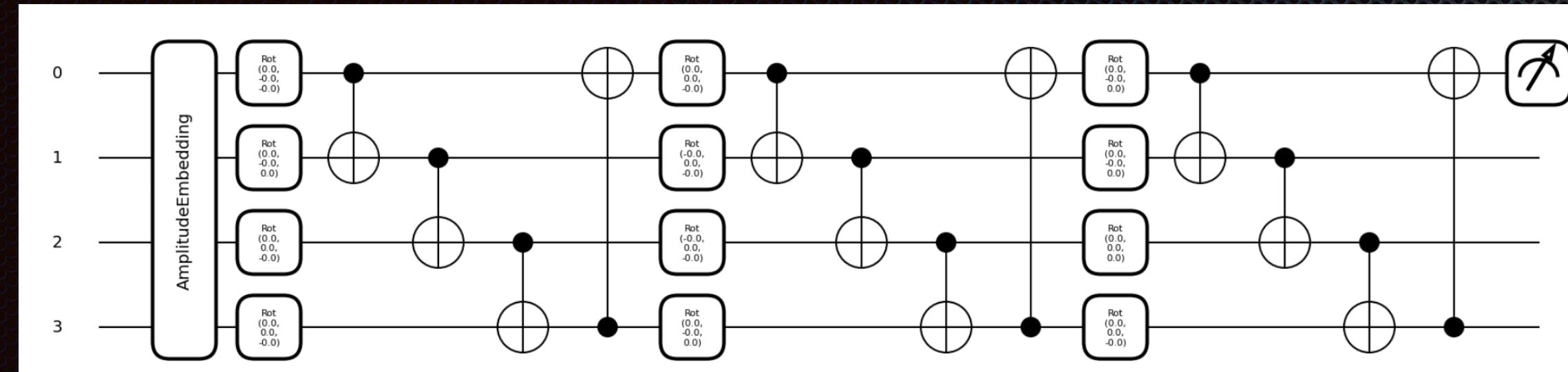


- Fixed
1- Basis embedding
2- Amplitude embedding
3- Angle embedding

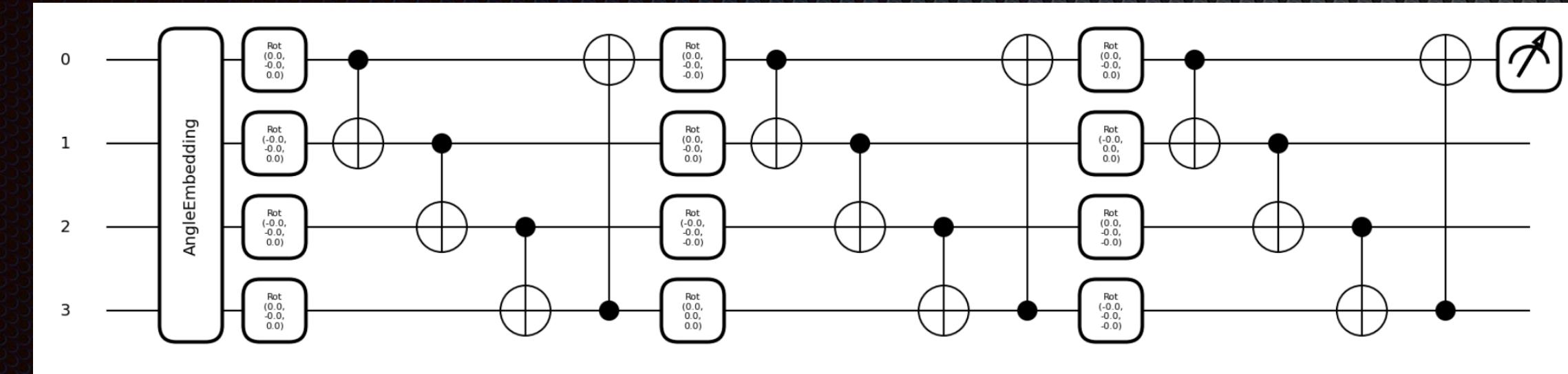
Variational
embedding

Quantum machine learning

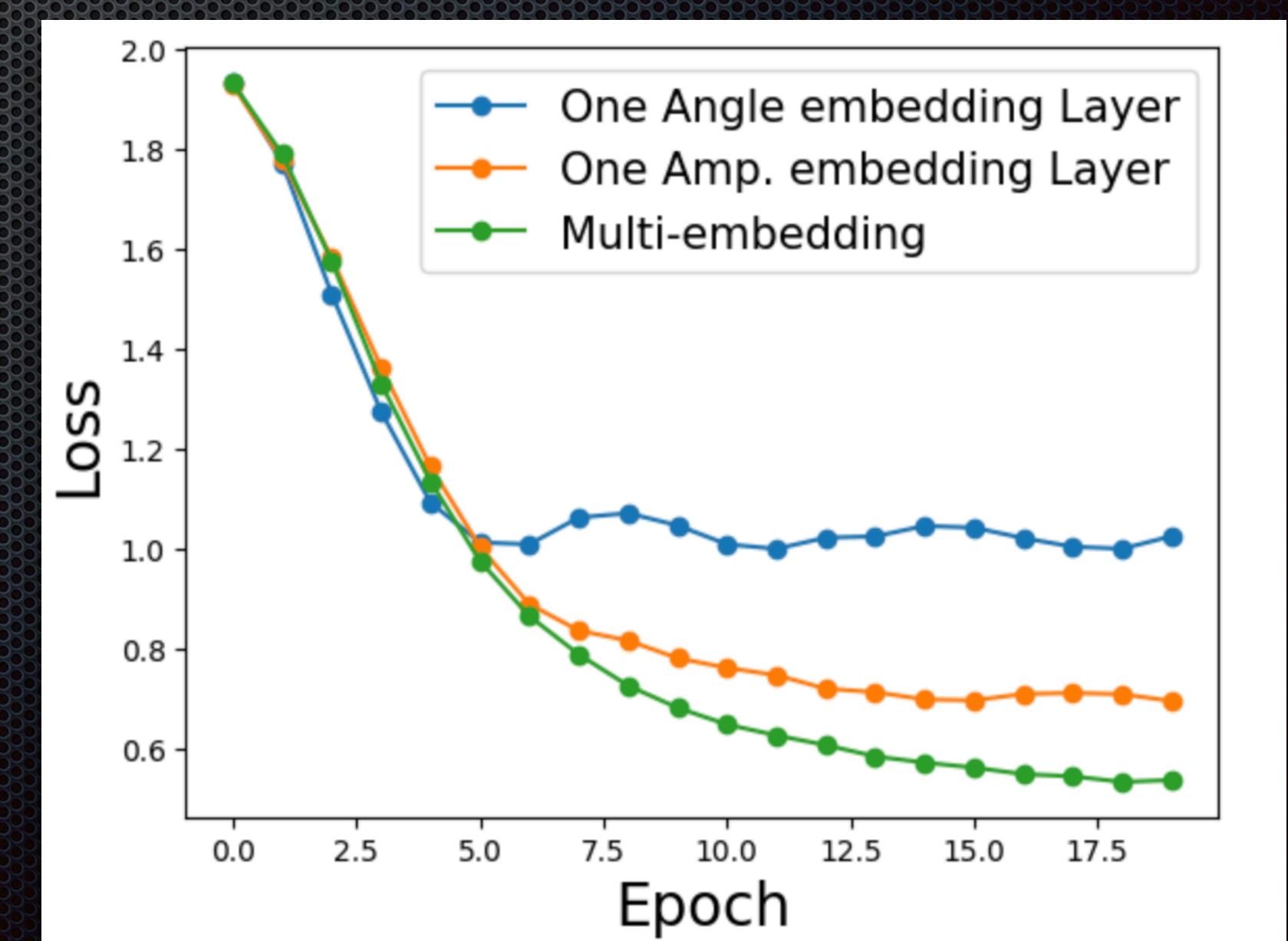
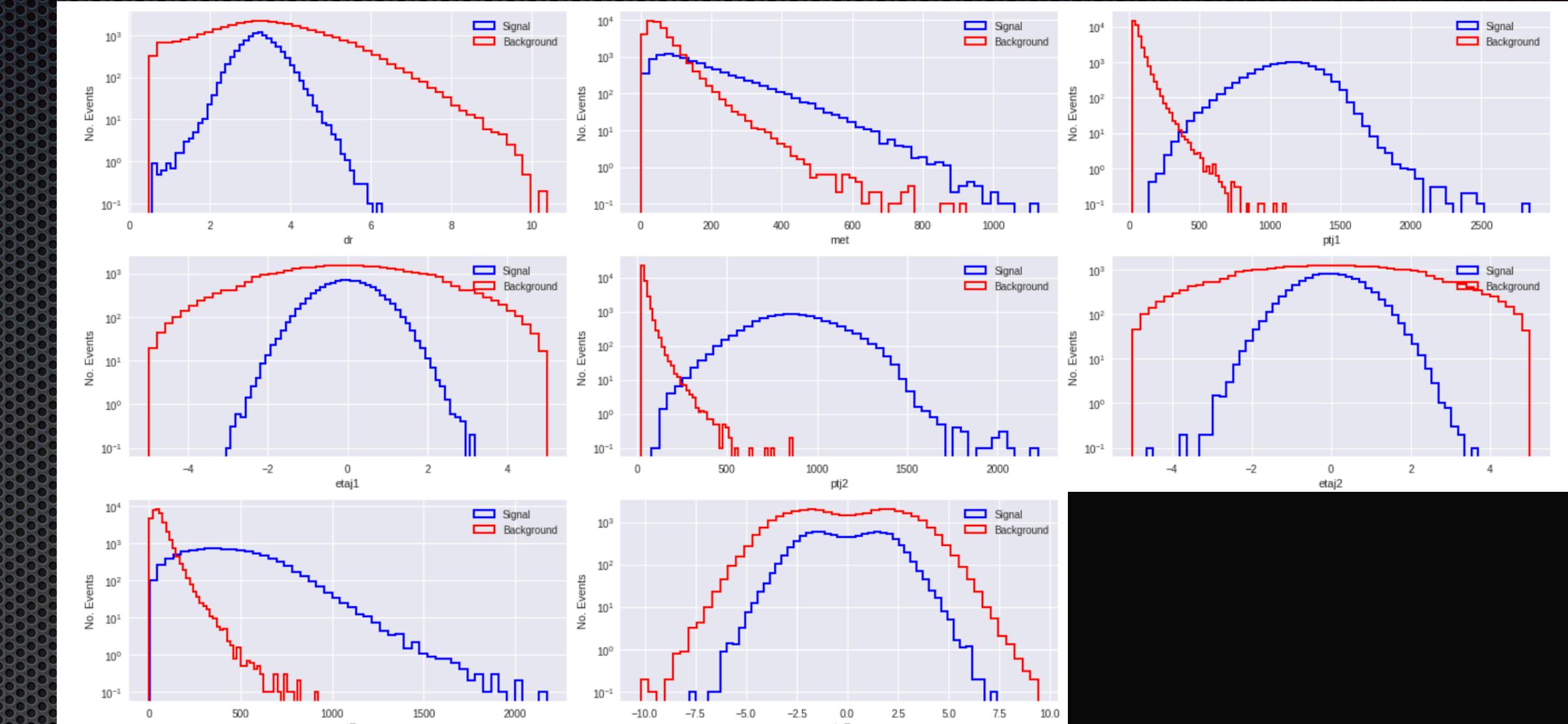
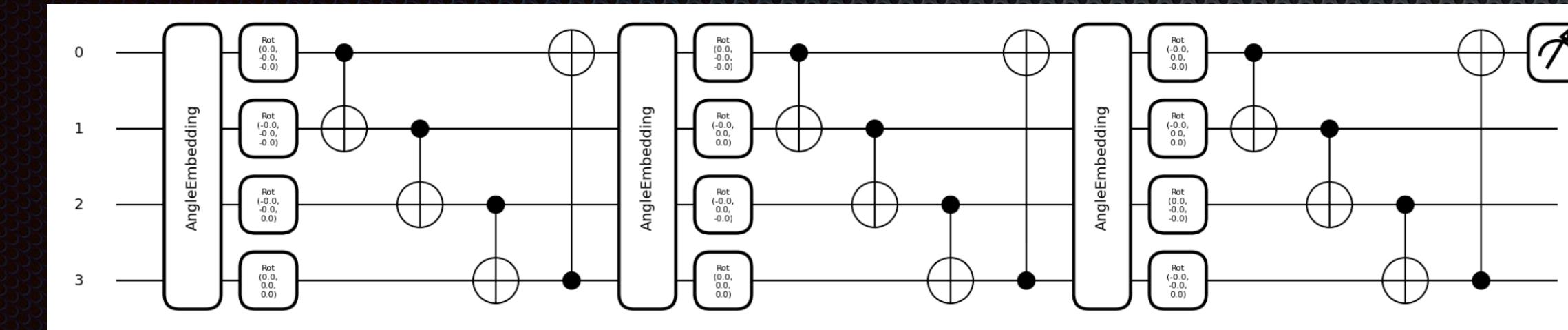
Amplitude embedding + 3 basic entangled layers



Angle embedding + 3 basic entangled layers



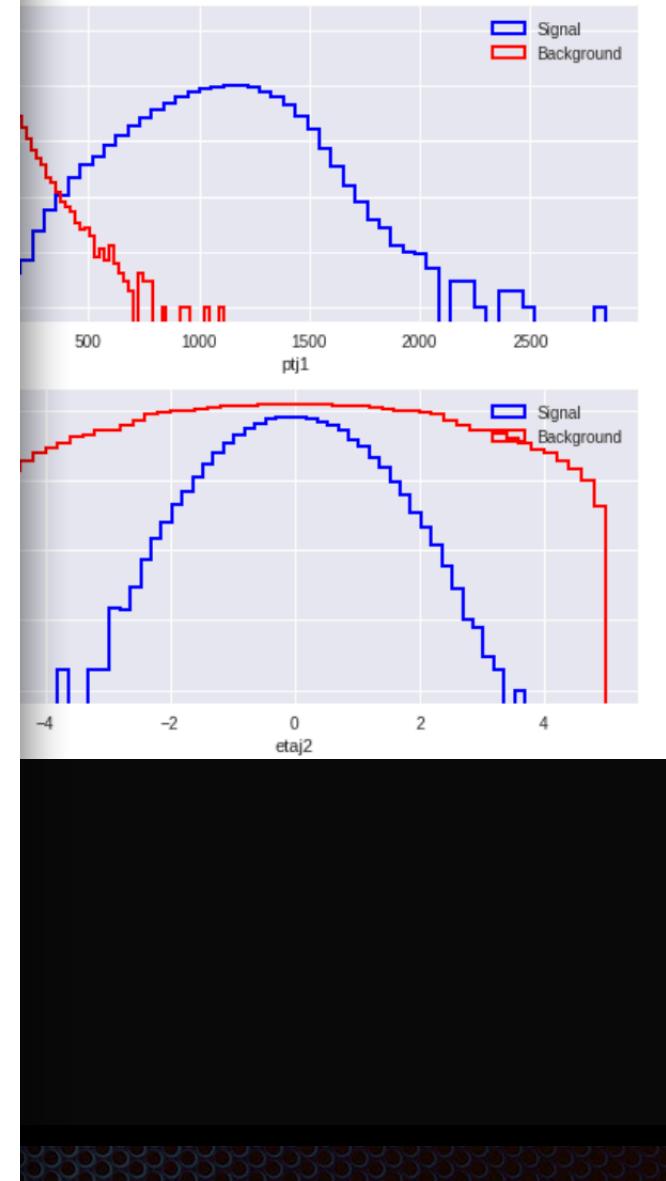
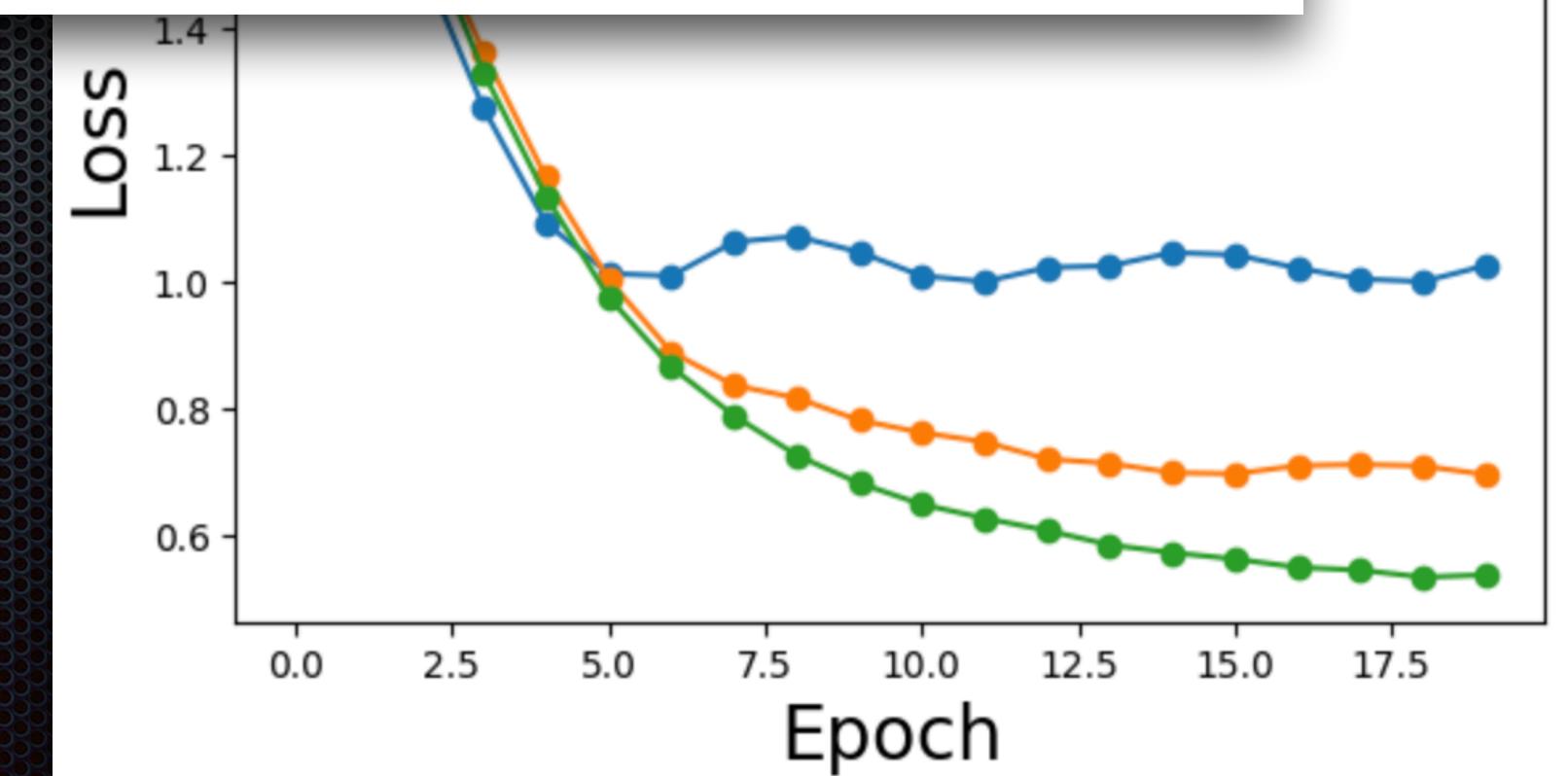
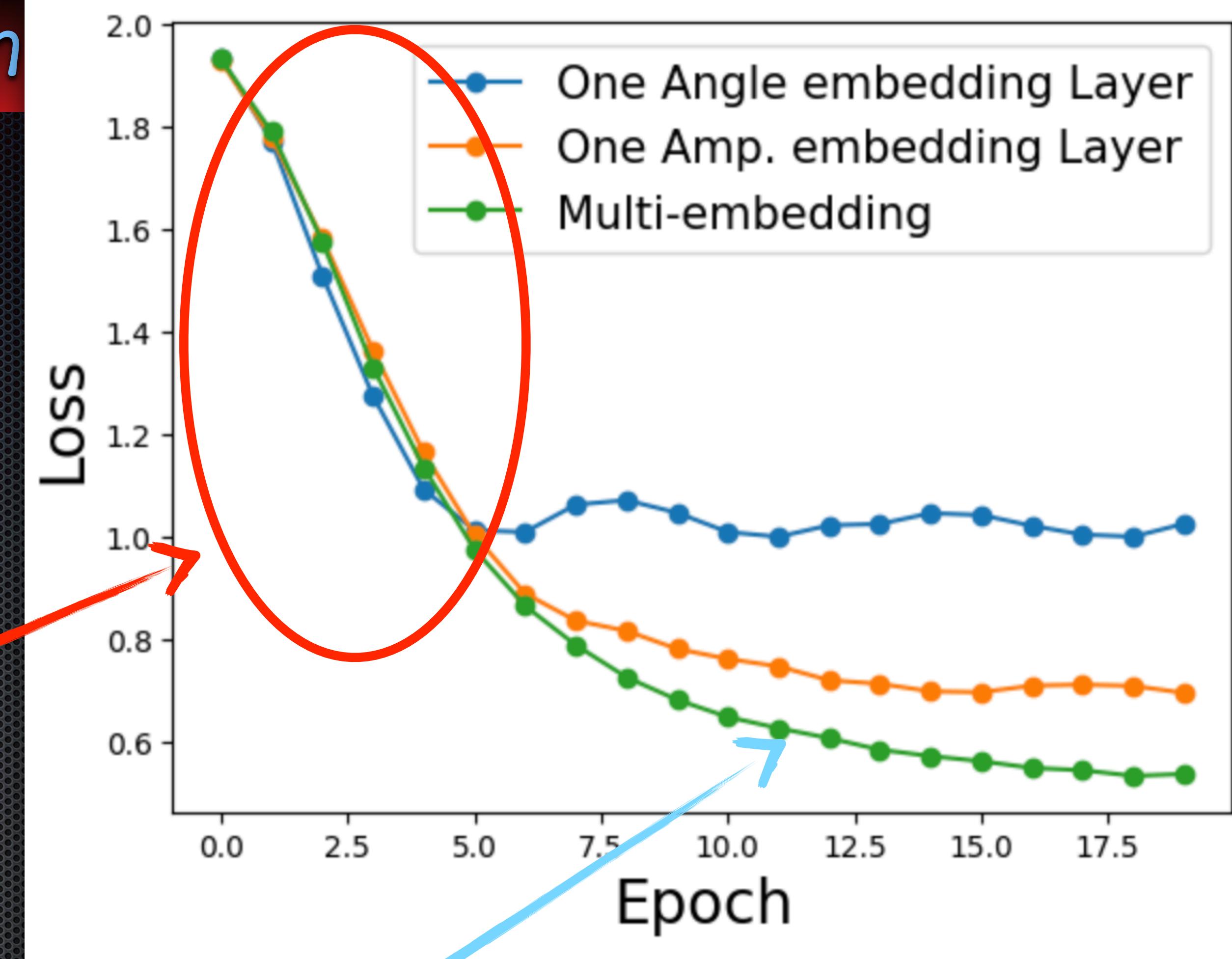
Repeated angle embedding + 3 basic entangled layers



Quantum machine learn

High performance for
small training data/few steps
(Quantum kernel mapping)

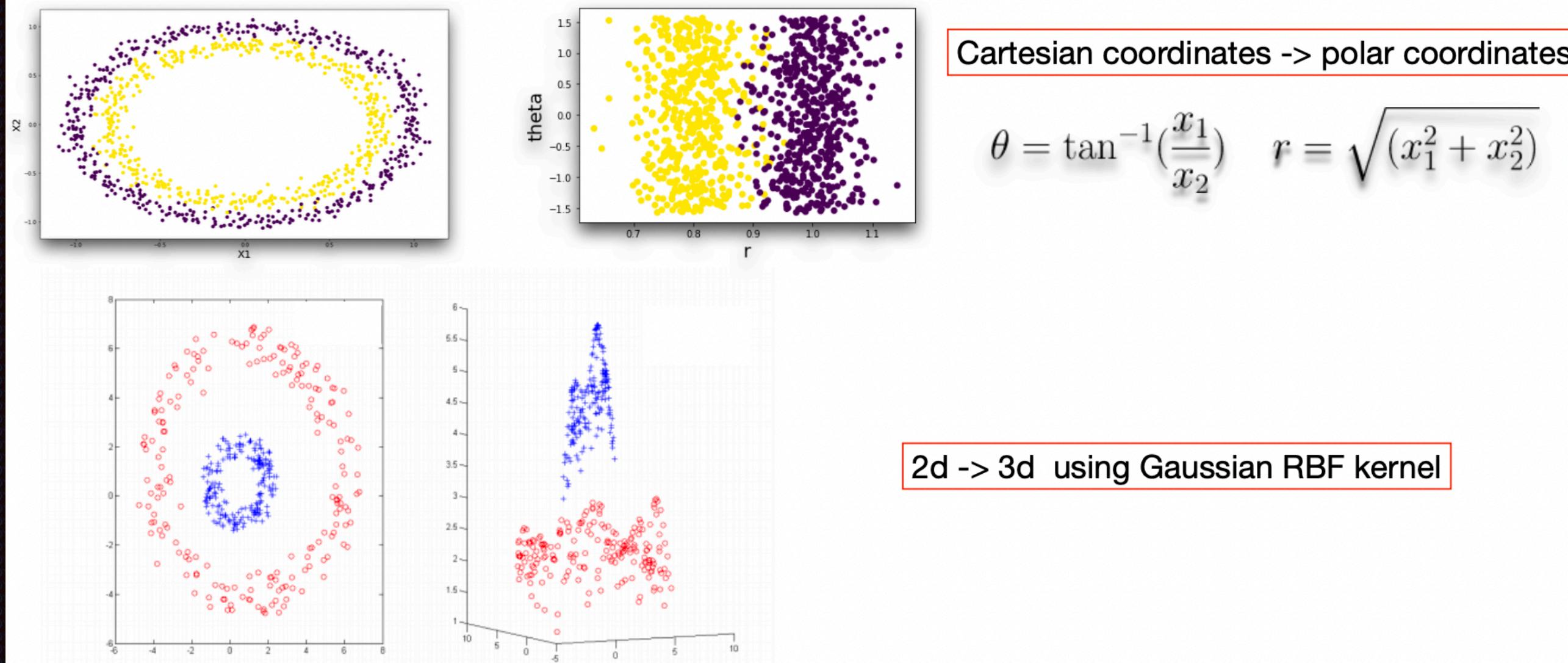
Repeated embedding can
achieve higher performance



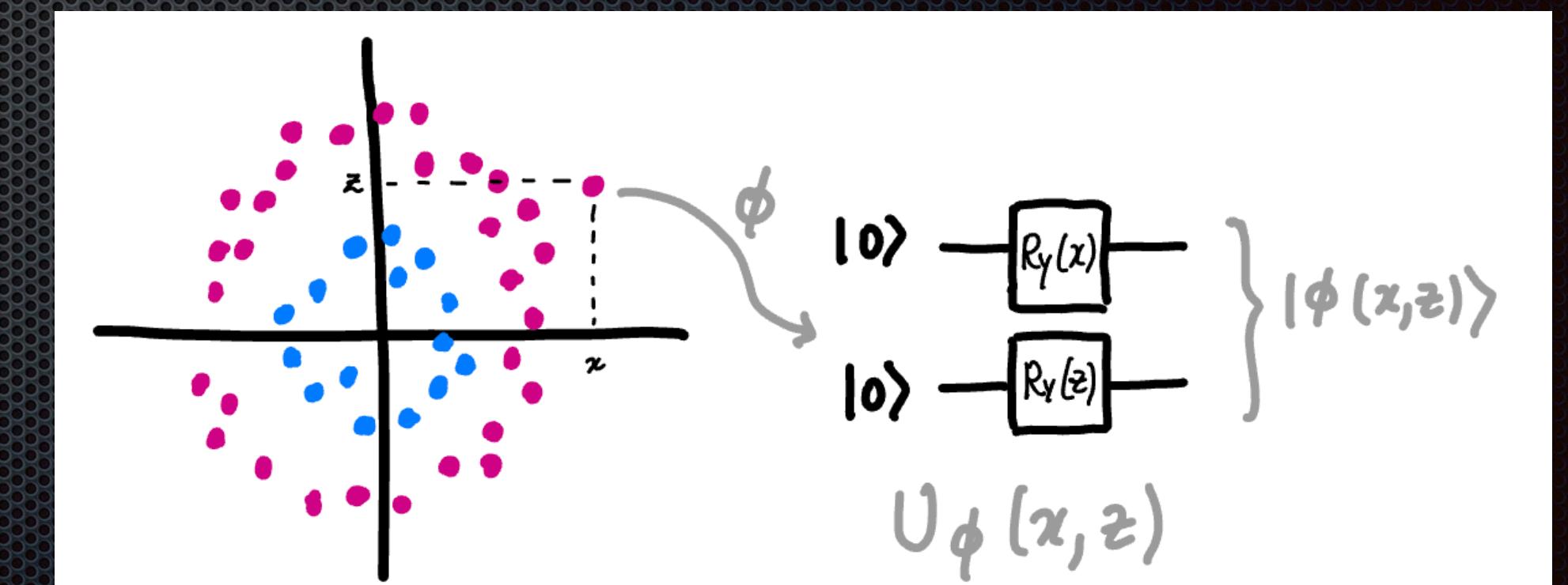
Quantum machine learning

Mapping non-linear separable data from low dimensional space to other coordinates by using specific kernel, one can find a hyper-plane that can easily separate between the data

Classical mapping

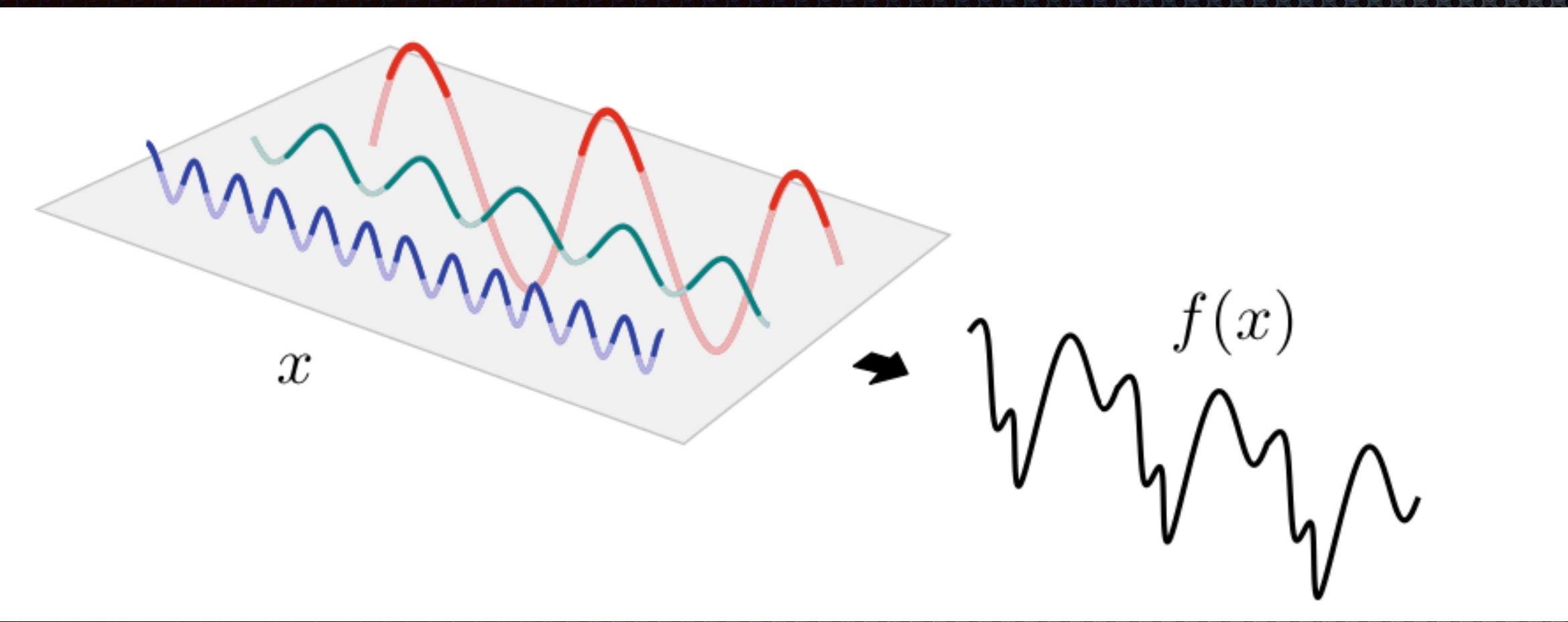


Quantum mapping



https://pennylane.ai/qml/glossary/quantum_feature_map.html

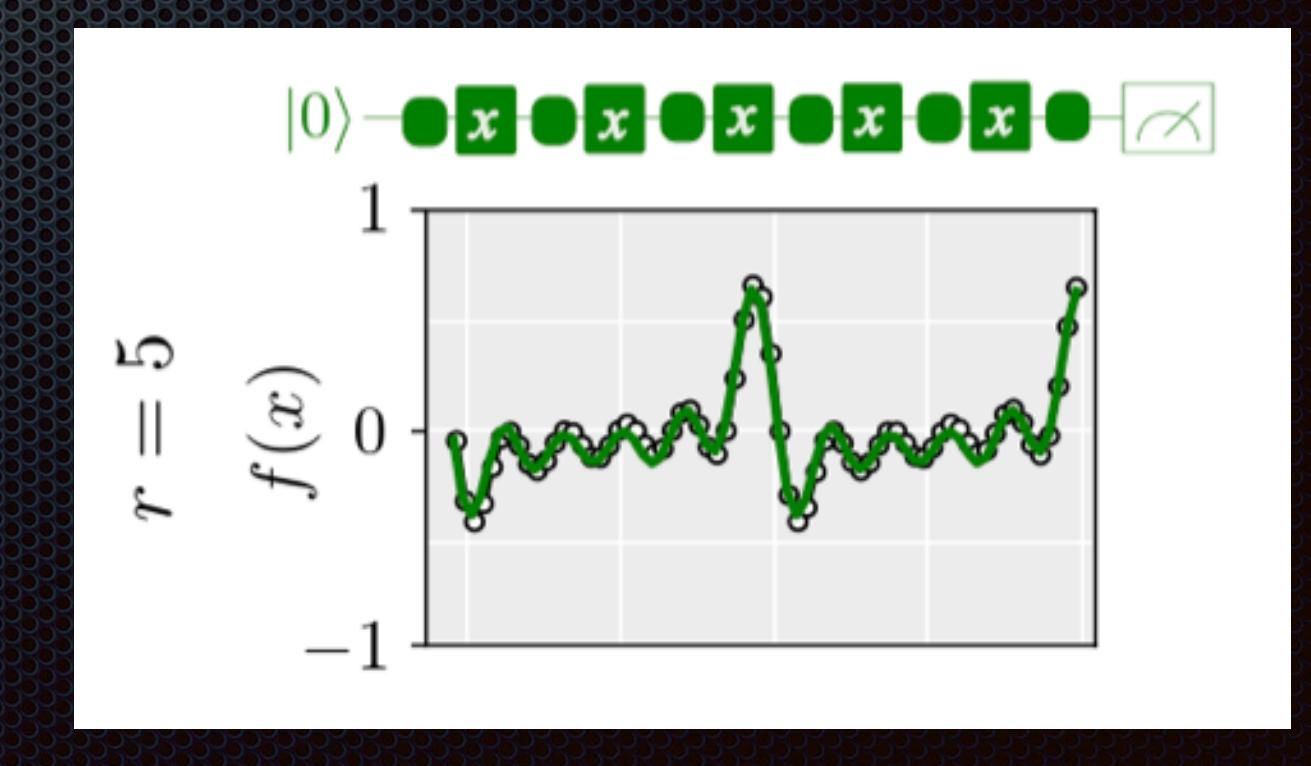
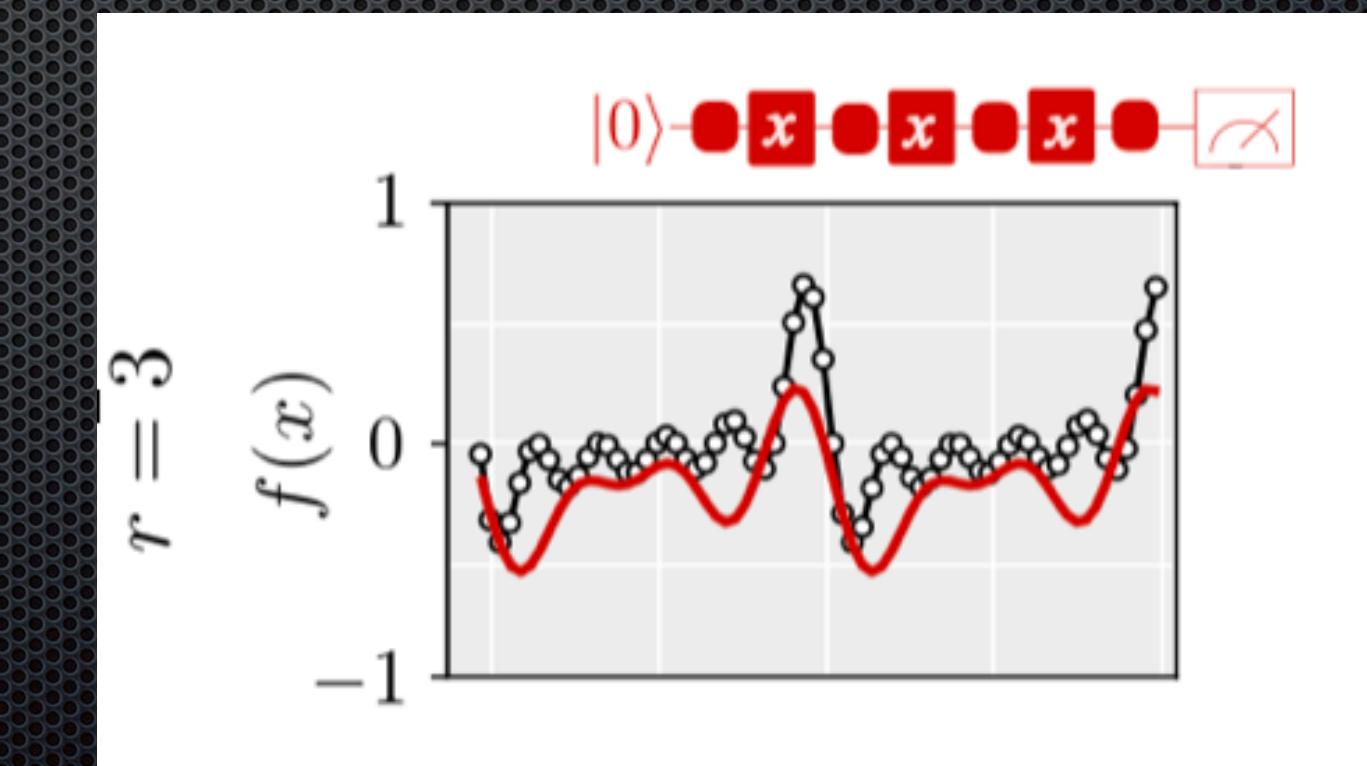
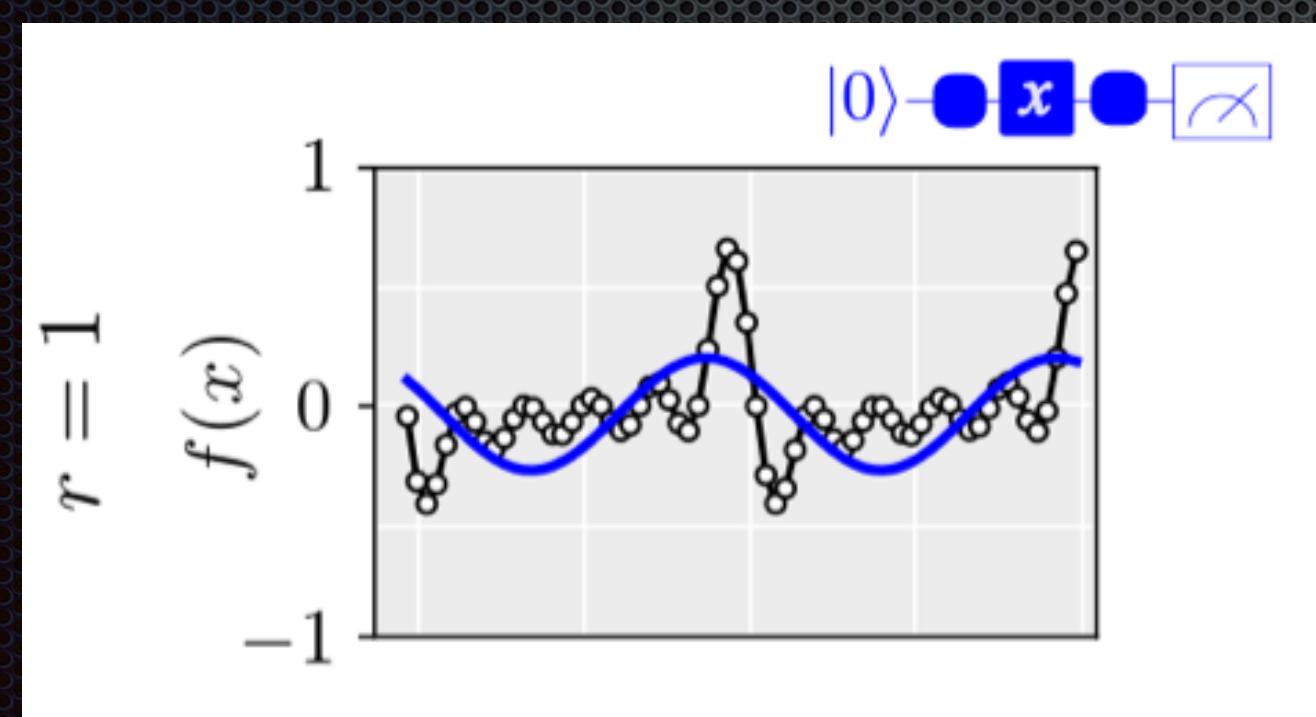
Quantum machine learning



Quantum model with data re-uploading
is a sum of a multi-dimensional
partial Fourier series

$$f_0(\omega) = \sum_{\omega} c_{\omega}(\theta) e^{i\omega \cdot \theta}$$

arXiv:2008.08605v2 [quant-ph]

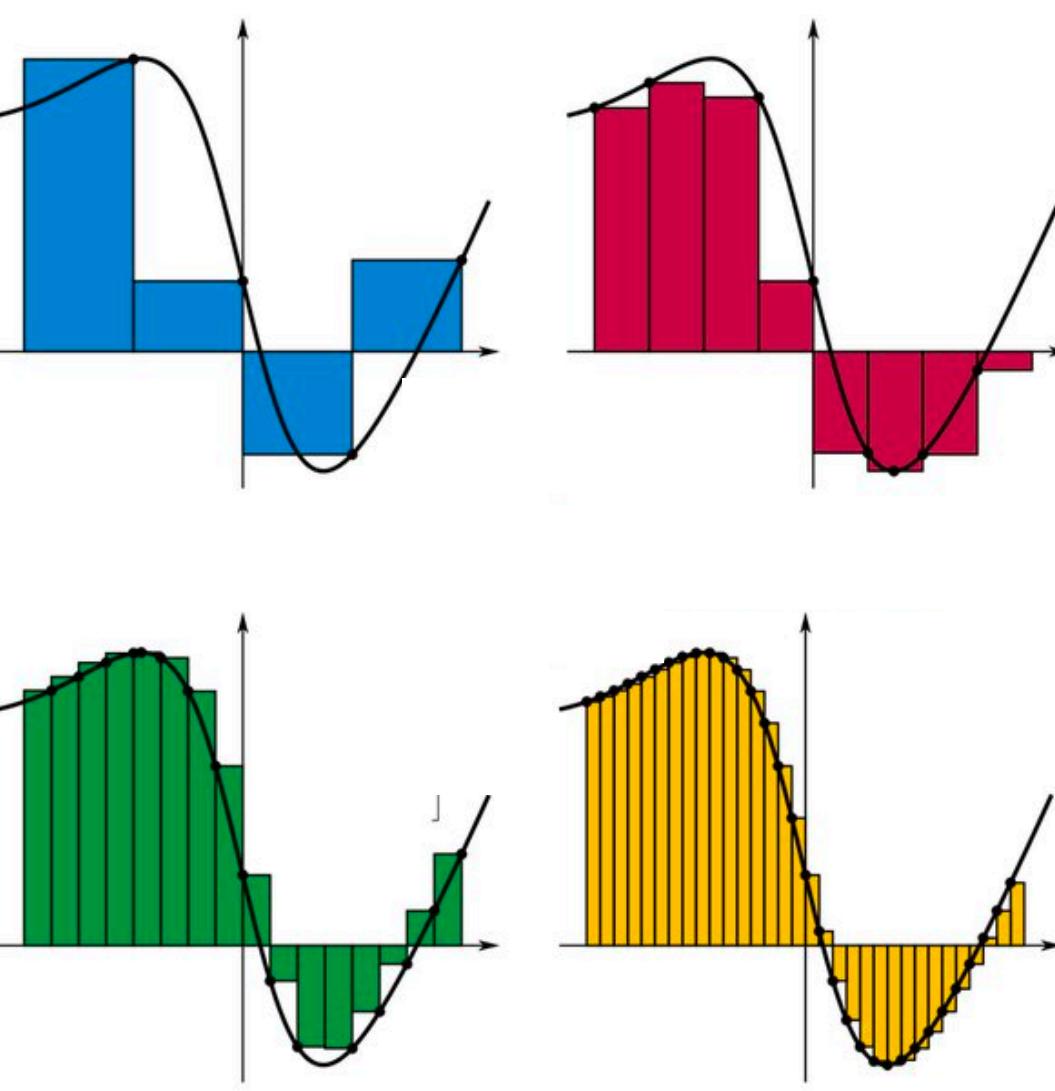
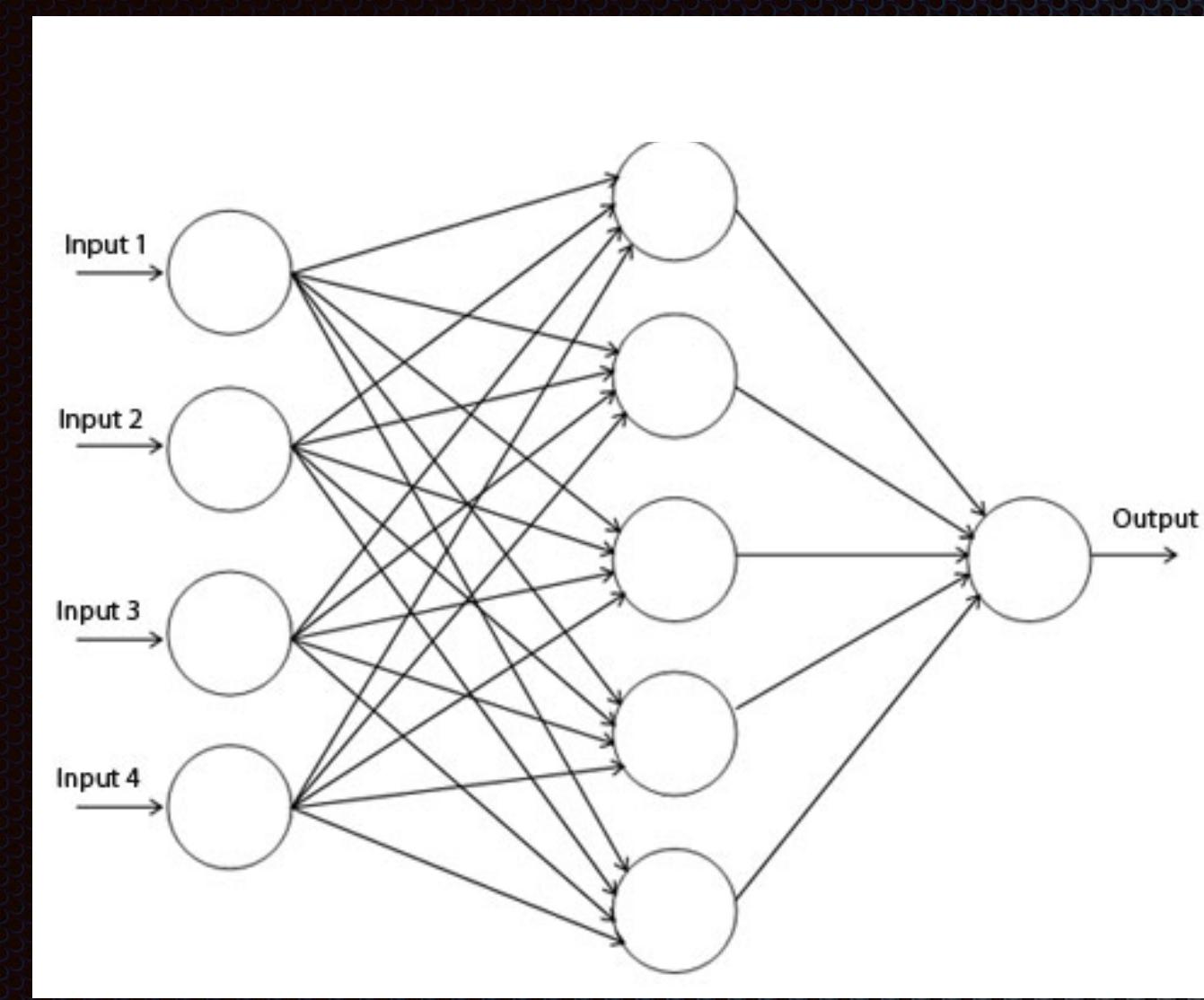


Question:

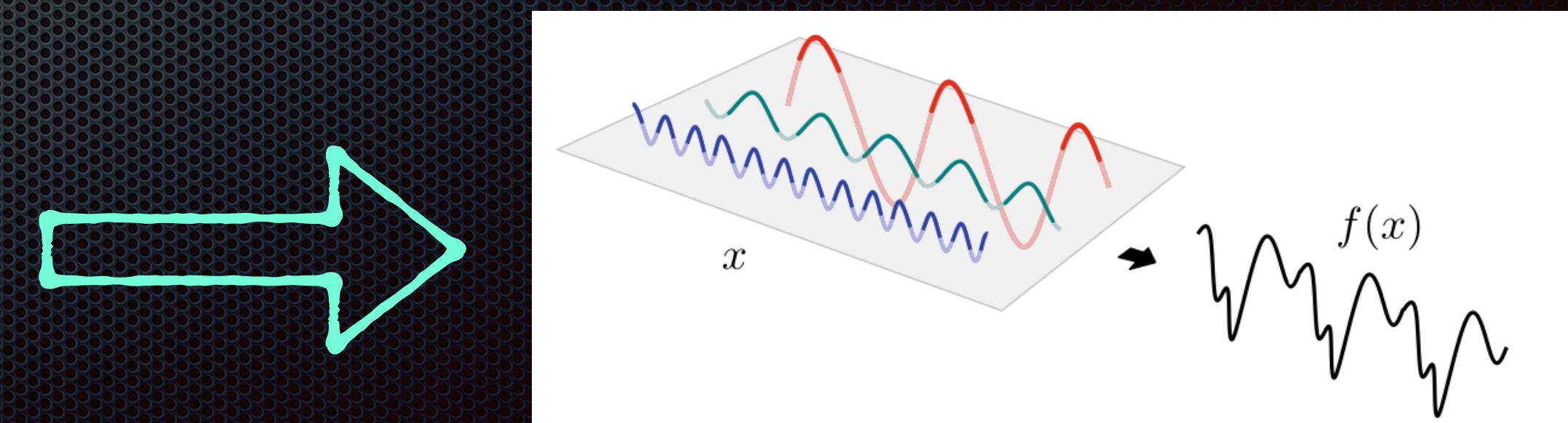
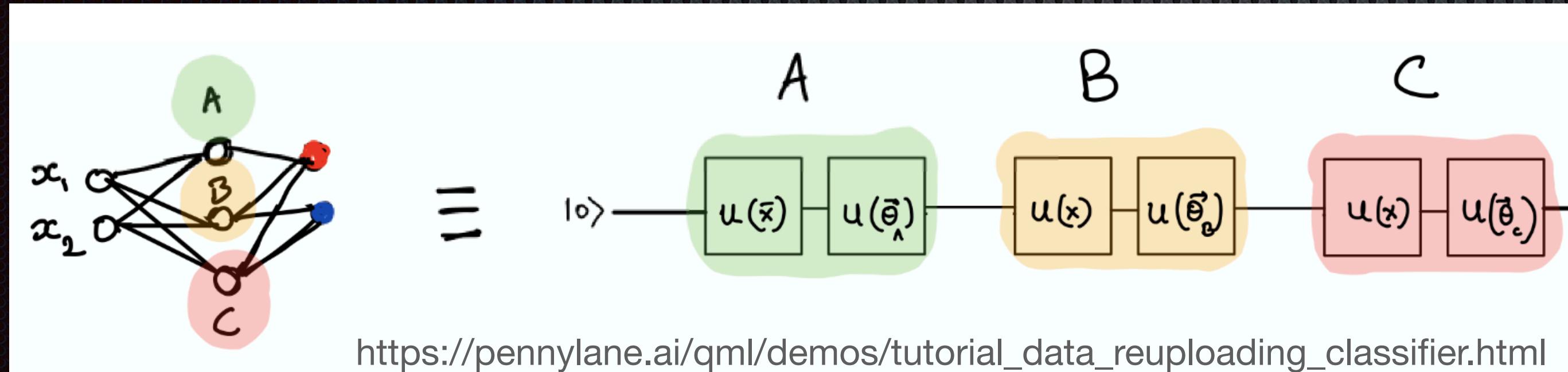
Can we construct a variational quantum classifier with only one qubit ?

Single qubit classifier

Given any continuous function, no matter how complicated it is, there is always exist a network that can approximate this function

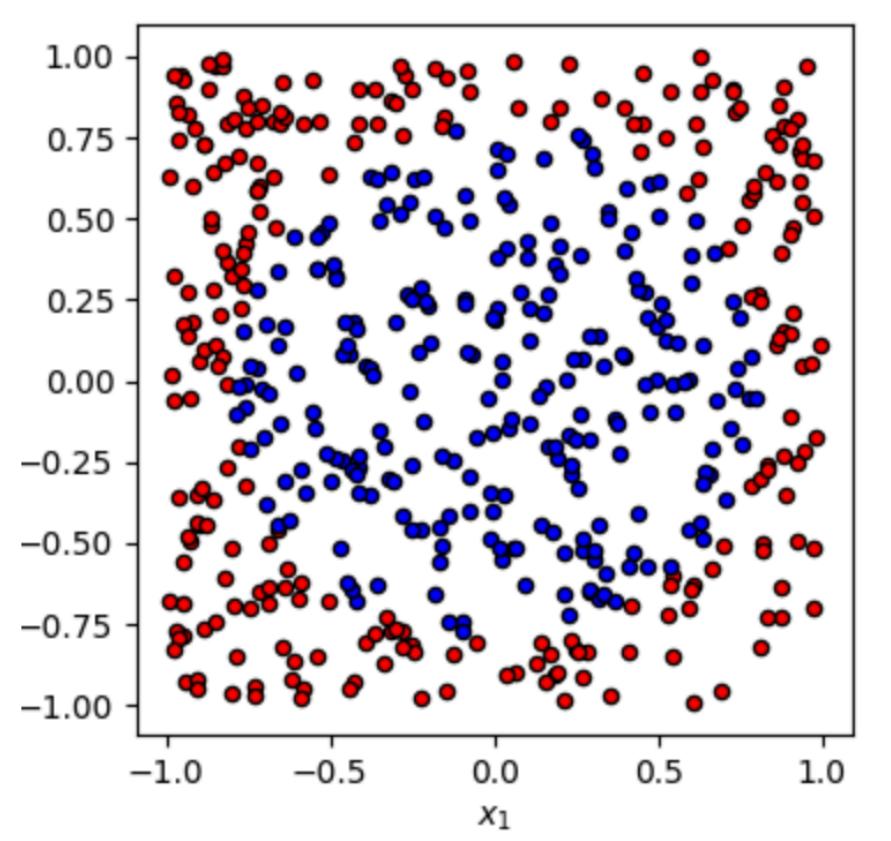


Universal approximation theorem

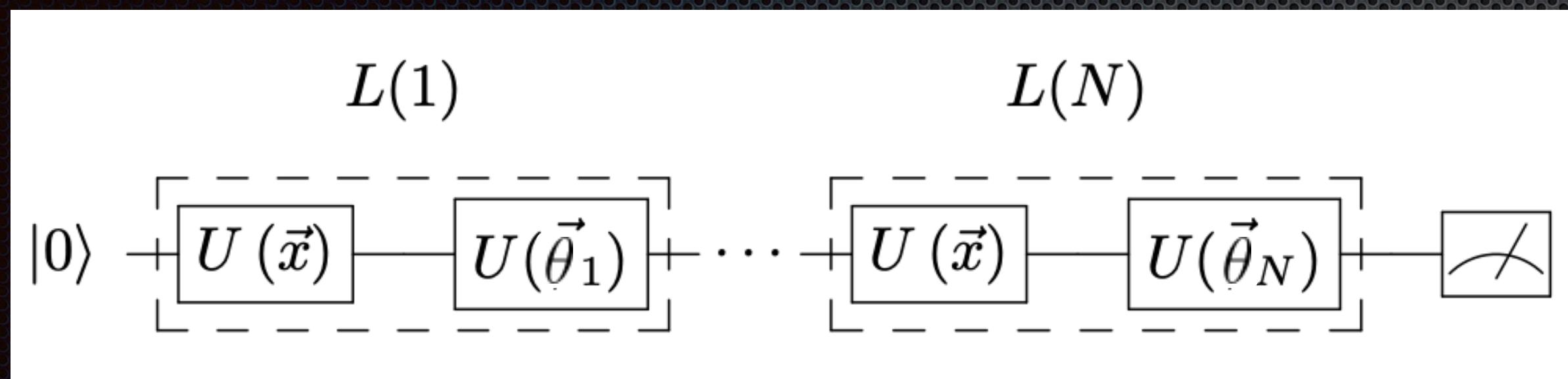


Single qubit classifier

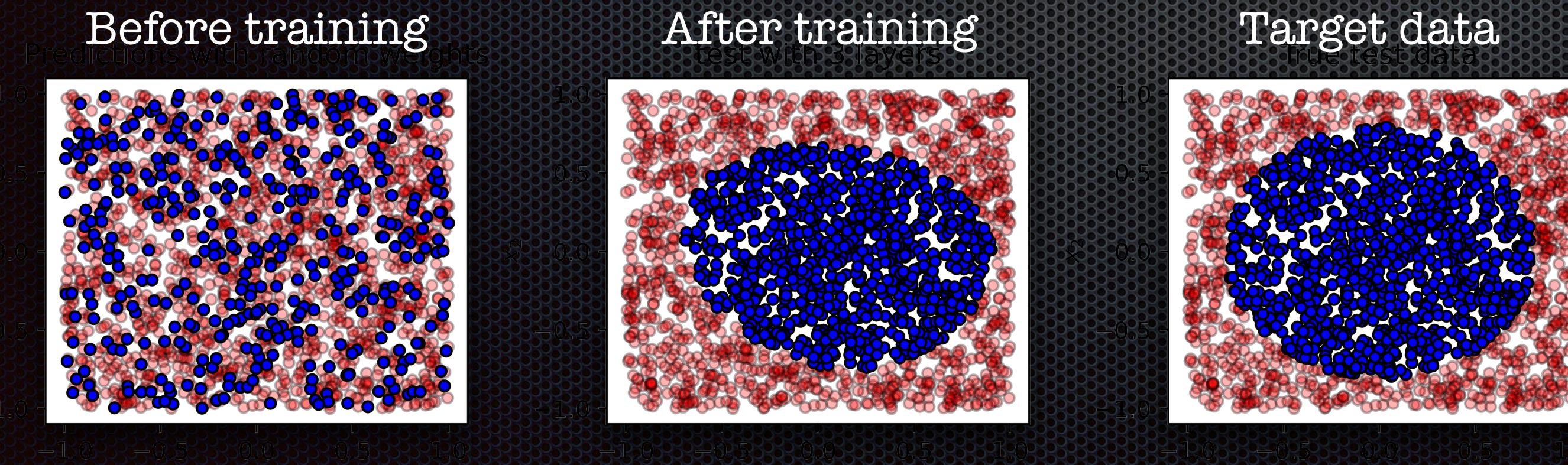
https://pennylane.ai/qml/demos/tutorial_data_reuploading_classifier.html



Task: classify non-linear two dimensional data using one qubit with data re-uploading



$$\text{Cost} = \sum_{\text{data points}} (1 - \text{fidelity}(\psi_{\text{output}}(\vec{x}, \vec{\theta}), \psi_{\text{label}}))$$



Training 200 points with 10 epochs

Single qubit classifier

$$\text{Cost} = \sum_{\text{data points}} (1 - \text{fidelity}(\psi_{\text{output}}(\vec{x}, \vec{\theta}), \psi_{\text{label}}))$$

What does this cost function mean ?

Fidelity

$$F(|\psi\rangle, |\phi\rangle) = |\langle\psi, \phi\rangle|^2$$

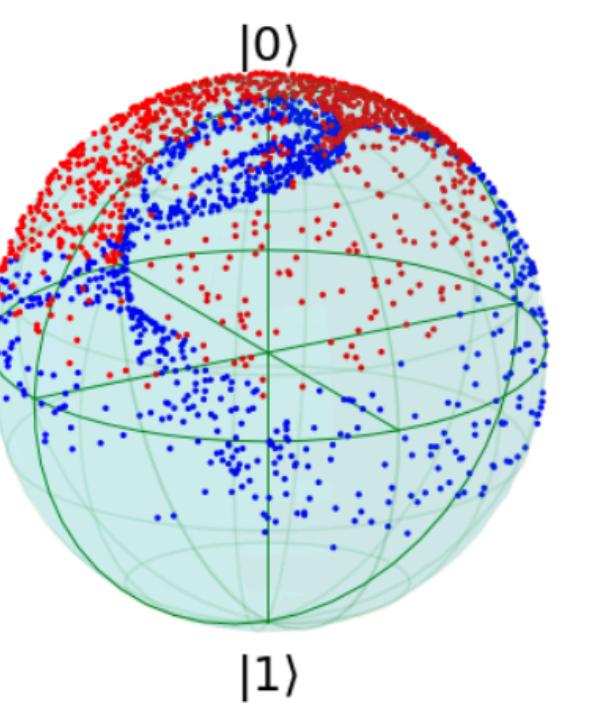
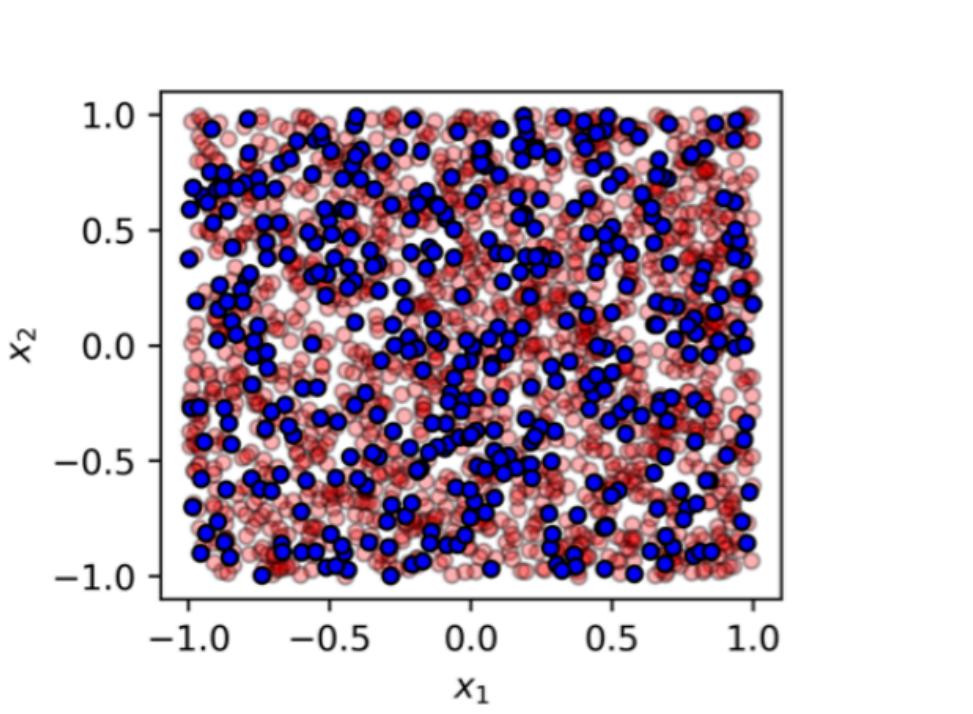
Fidelity is the distance (similarity measure) between two quantum states

Minimizing the cost function means we increase
the purity of quantum states

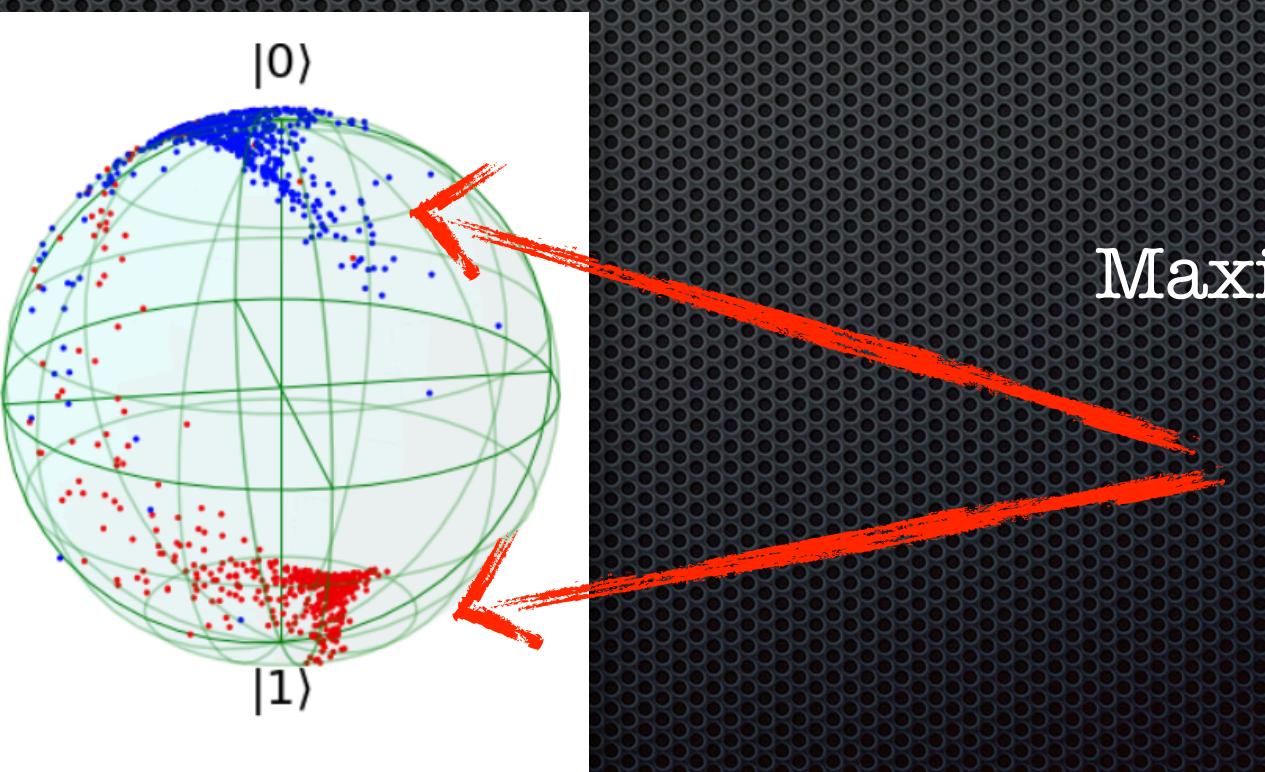
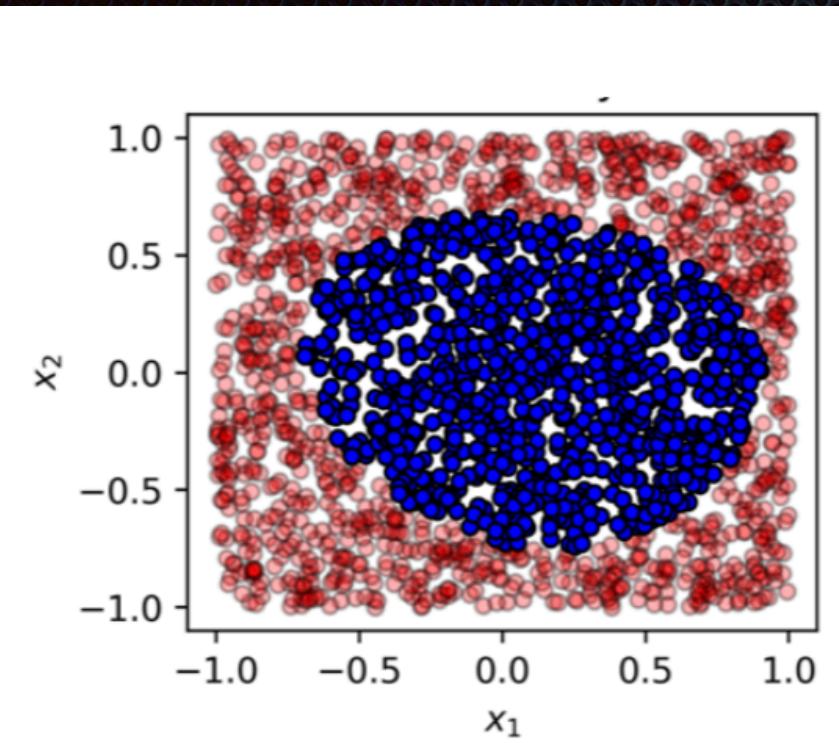
Single qubit classifier

Bloch sphere representation of the embedded data on the single qubit

Before training



After training



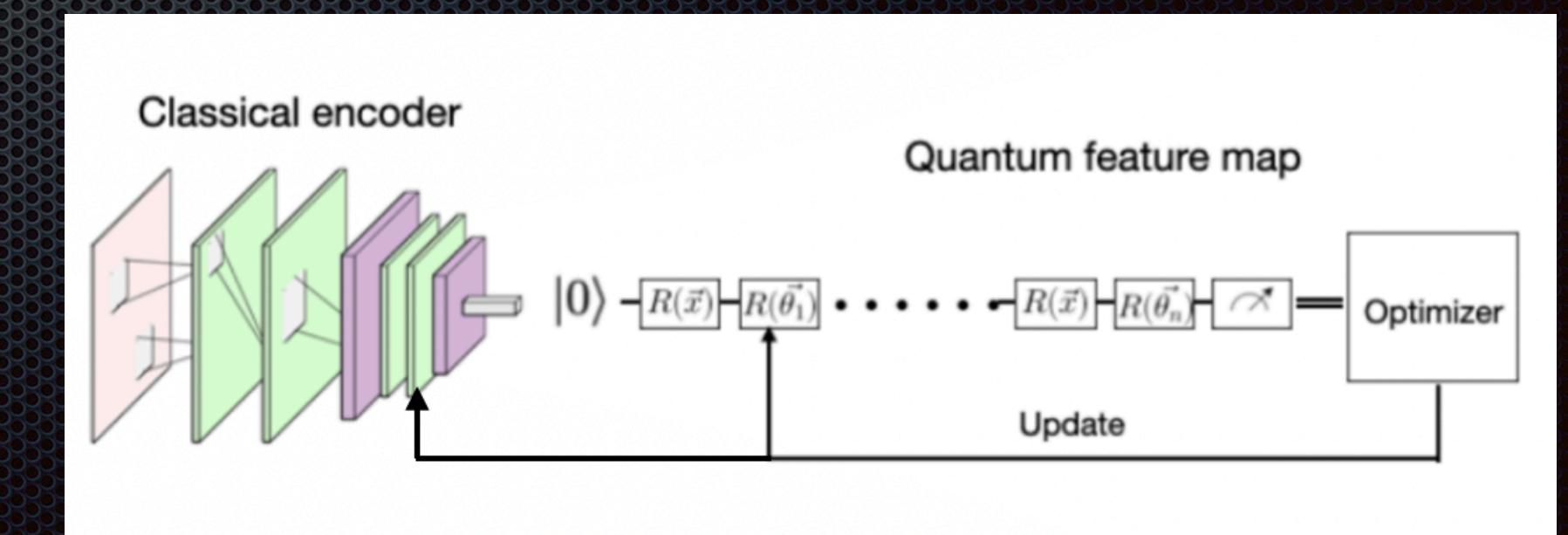
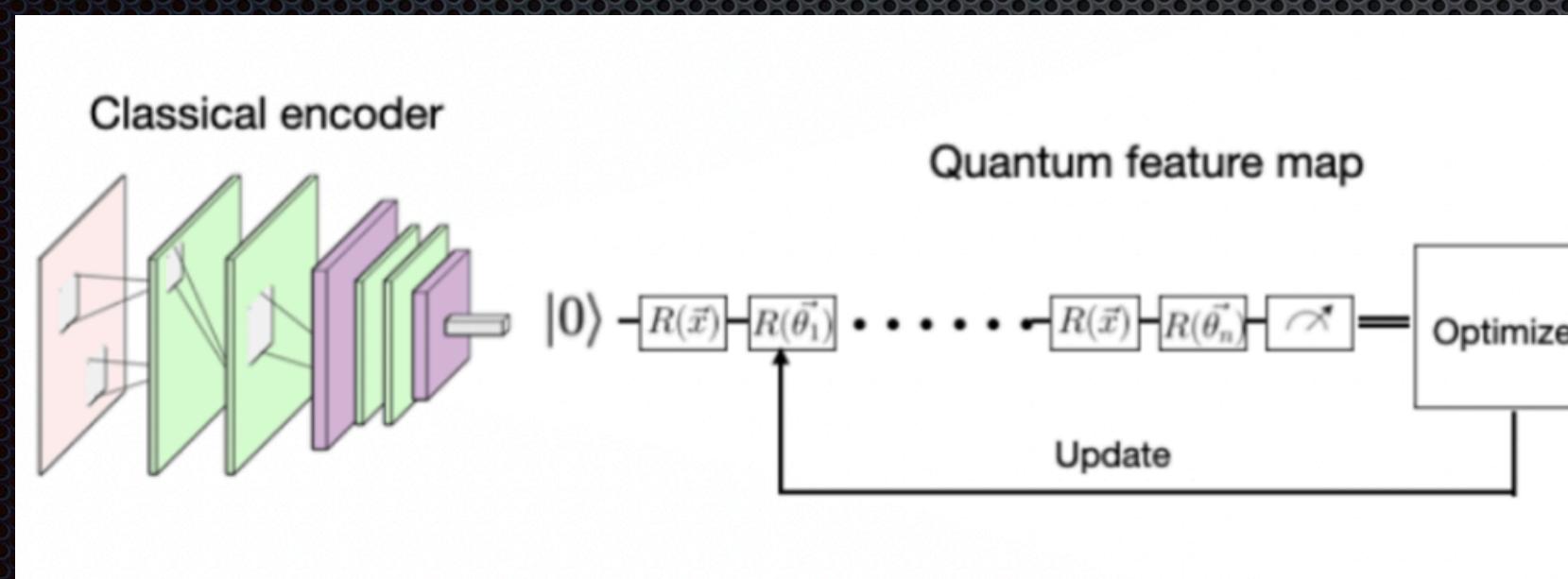
Maximize the quantum fidelity = maximize the distance between the measured quantum states on the qubit

Quantum model for image classification

Hybrid classical-quantum model for image classification

Classical pre-trained Encoder
+
Trainable quantum layers

Classical trainable Encoder
+
Trainable quantum layers



Quantum model for image classification



Training images: 200

Test images: 6000

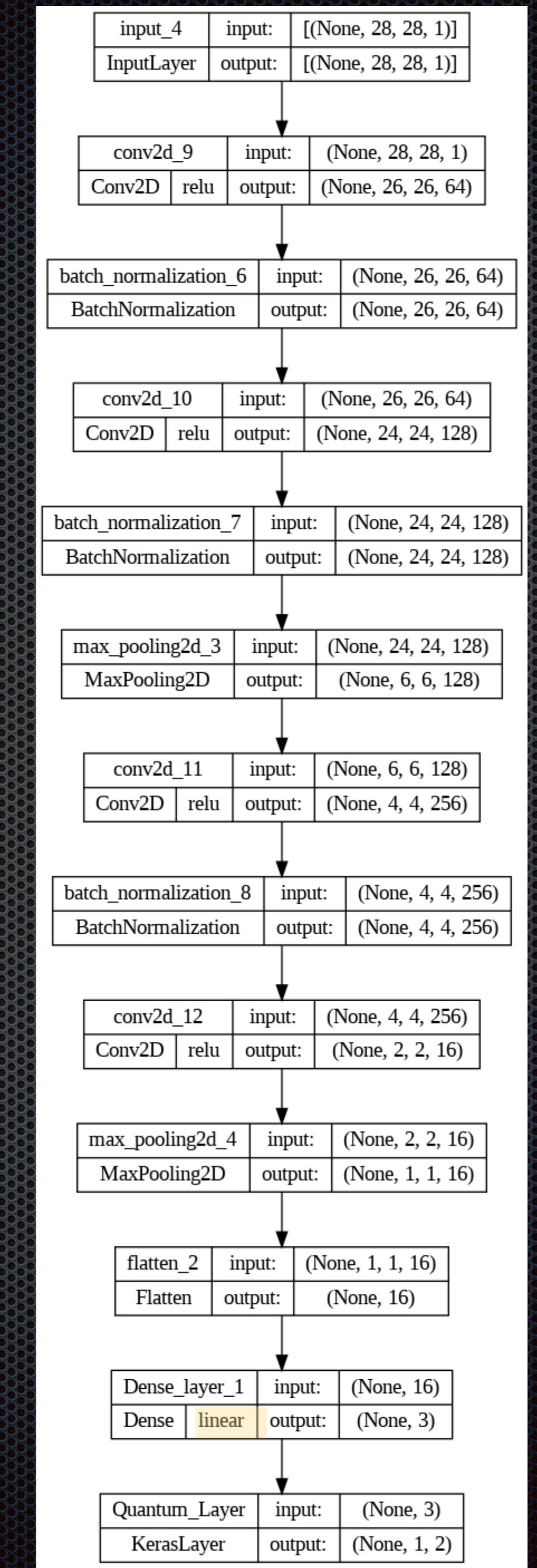
Validation images : 100

Optimizer: Adam (lr= 0.001)

Epochs: 10

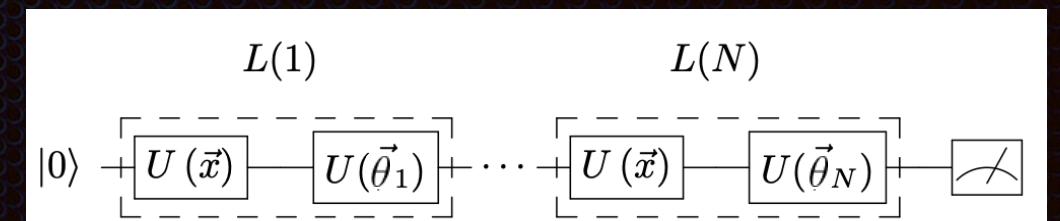
Batch size: 5

Quantum block: 10



Classical Encoder

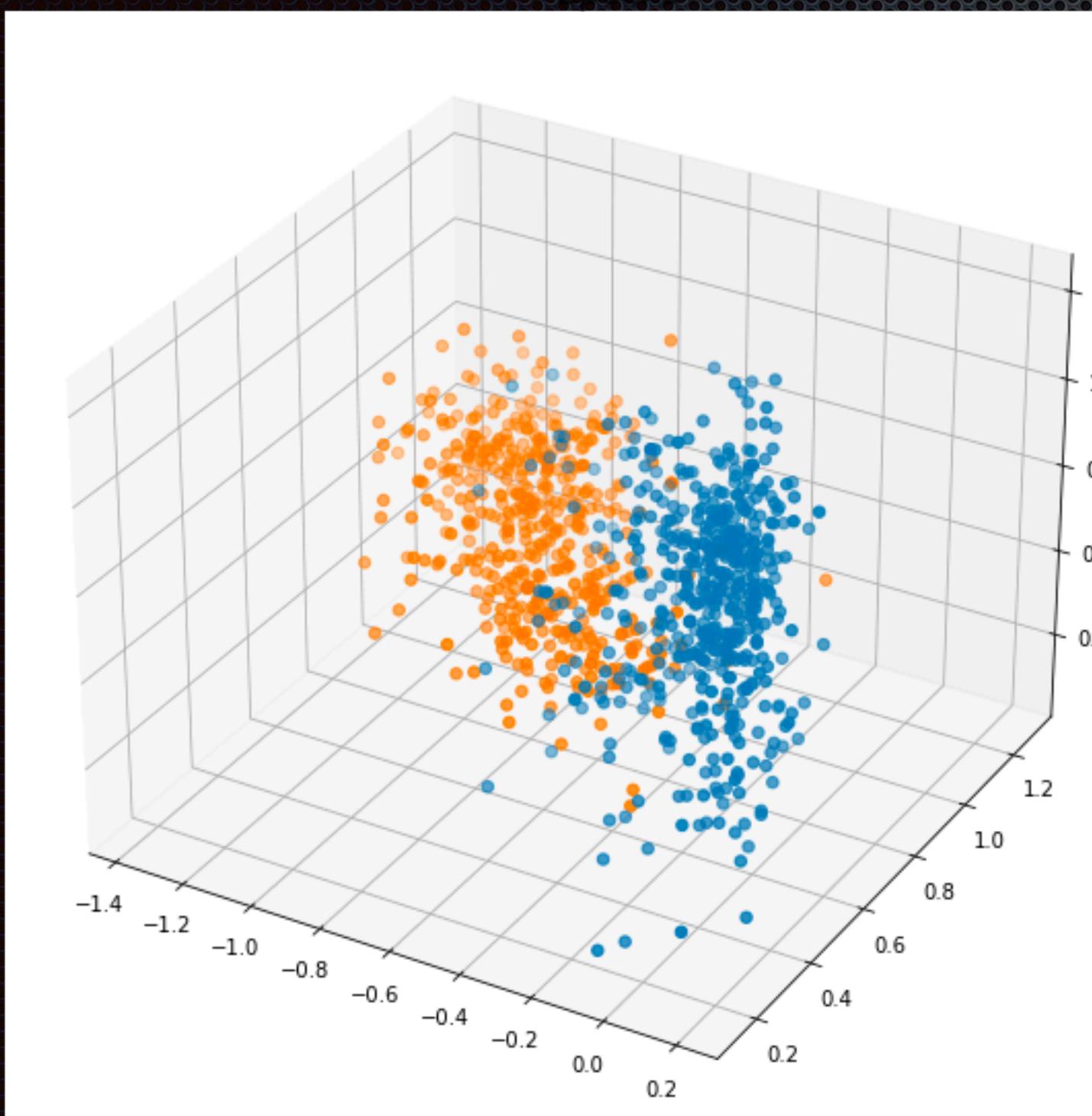
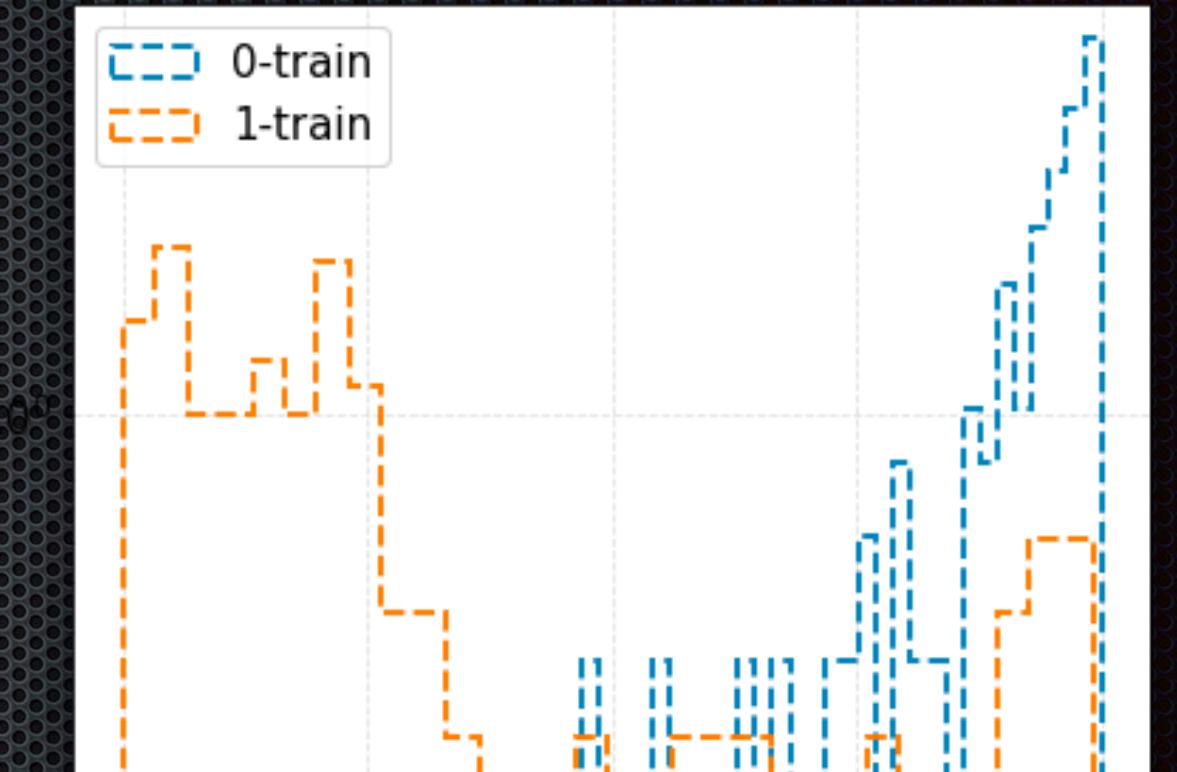
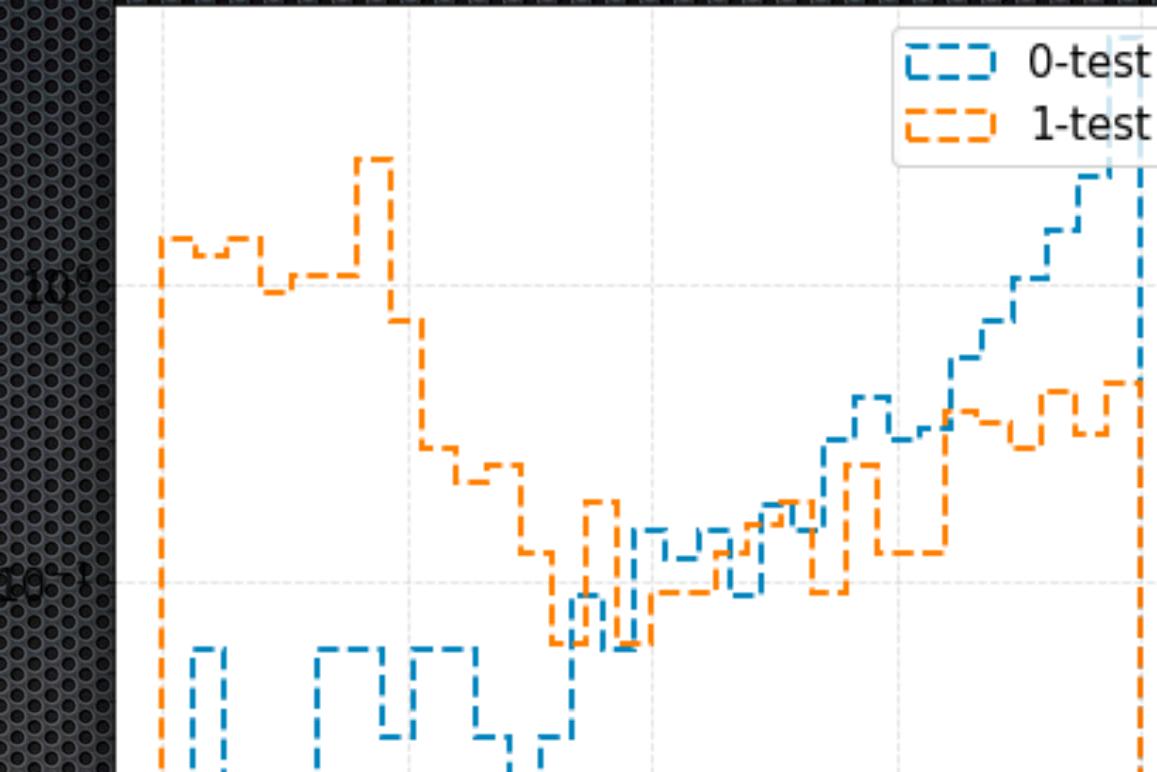
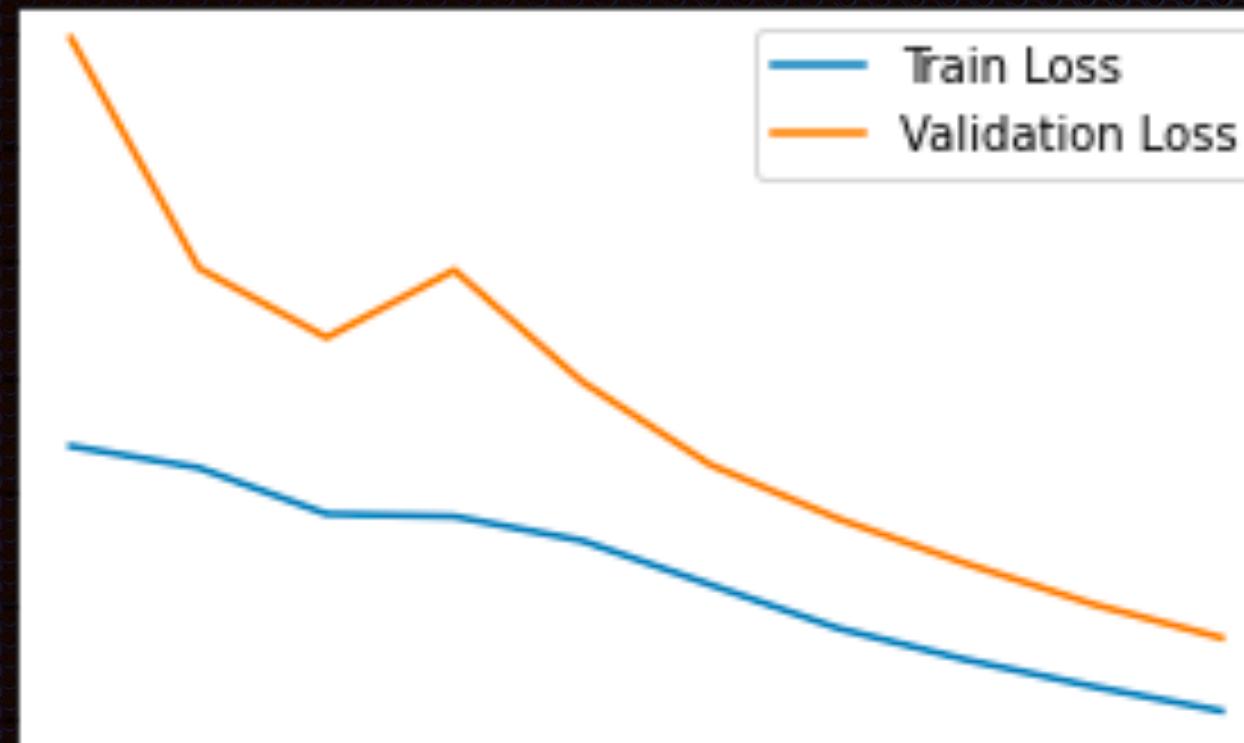
Quantum circuit



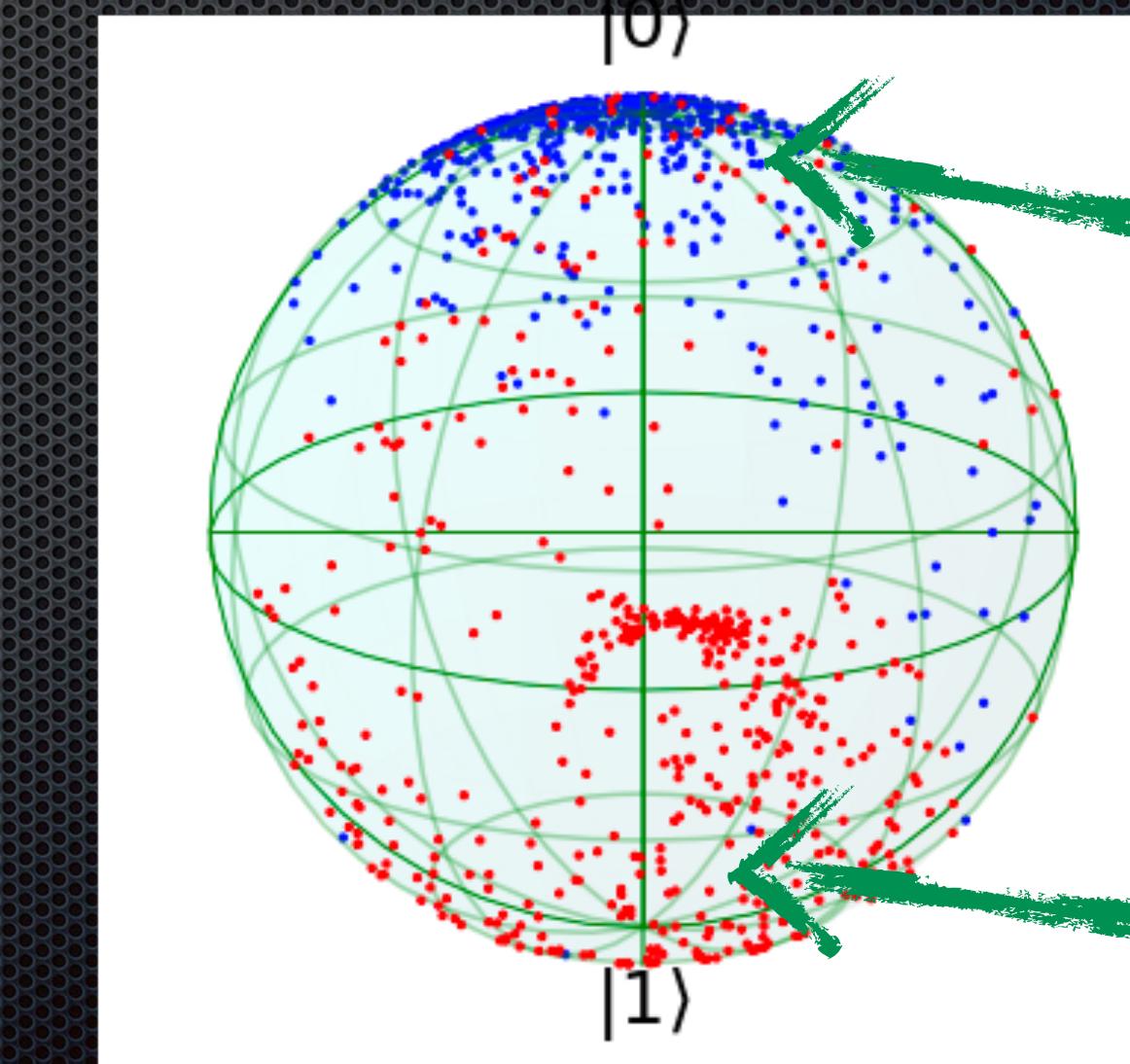
Quantum model for image classification

Test accuracy = 86.8%

Train accuracy= 93.5%



Latent space of
the classical encoder
Not-normalized

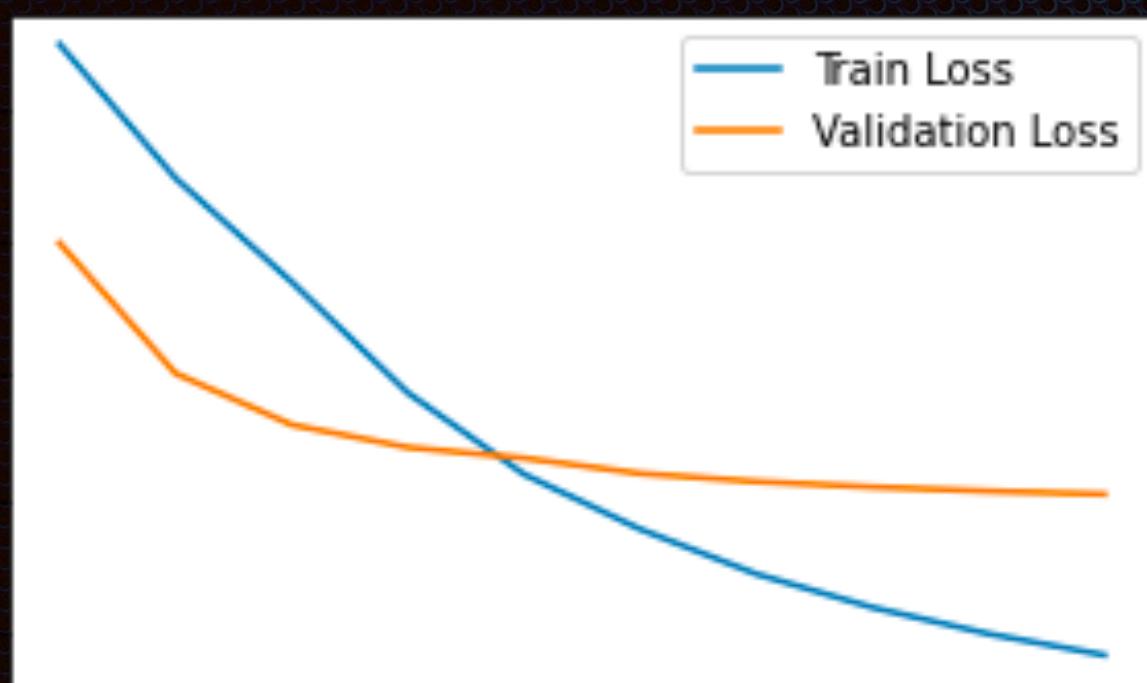


Shirt images

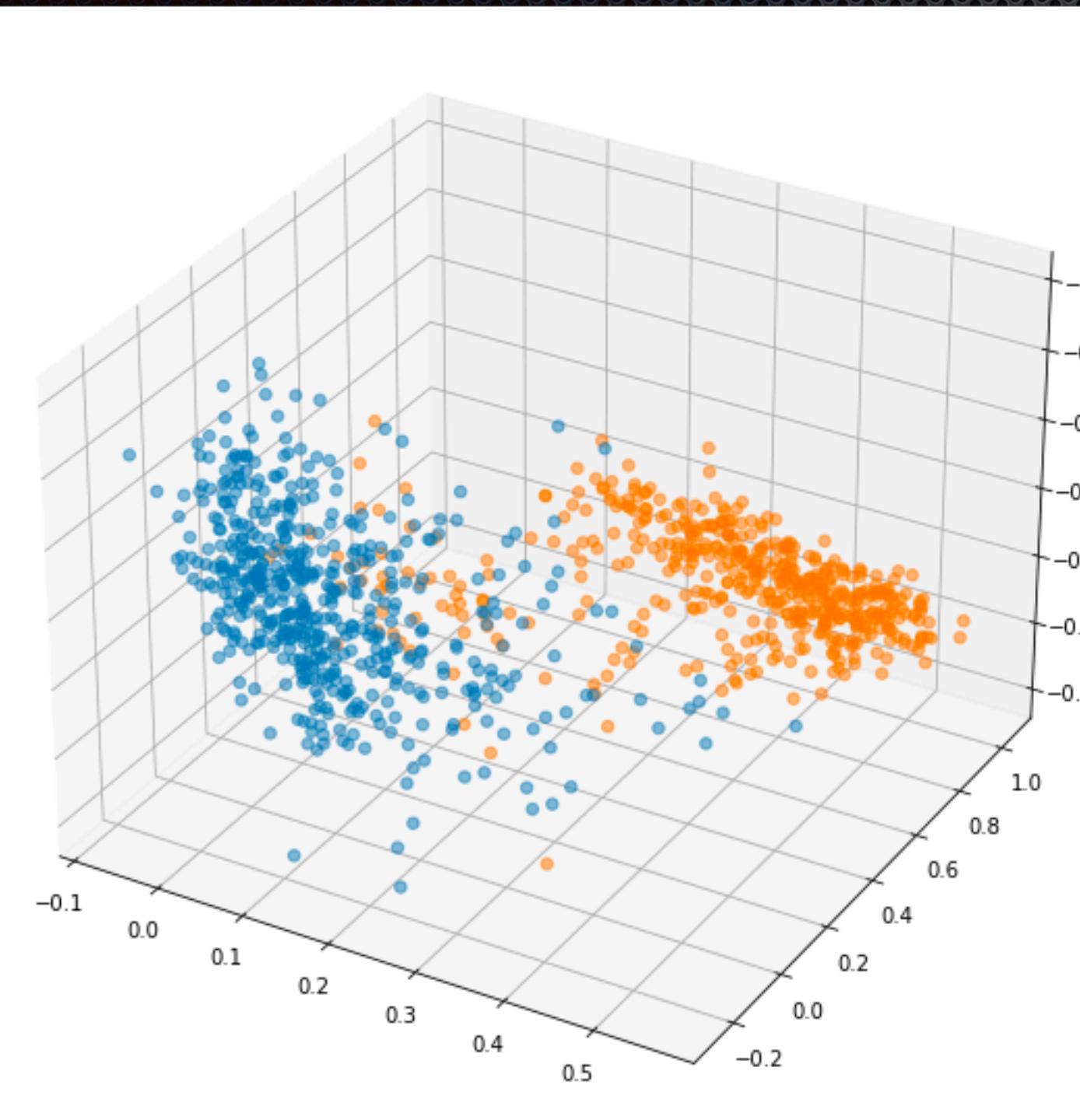
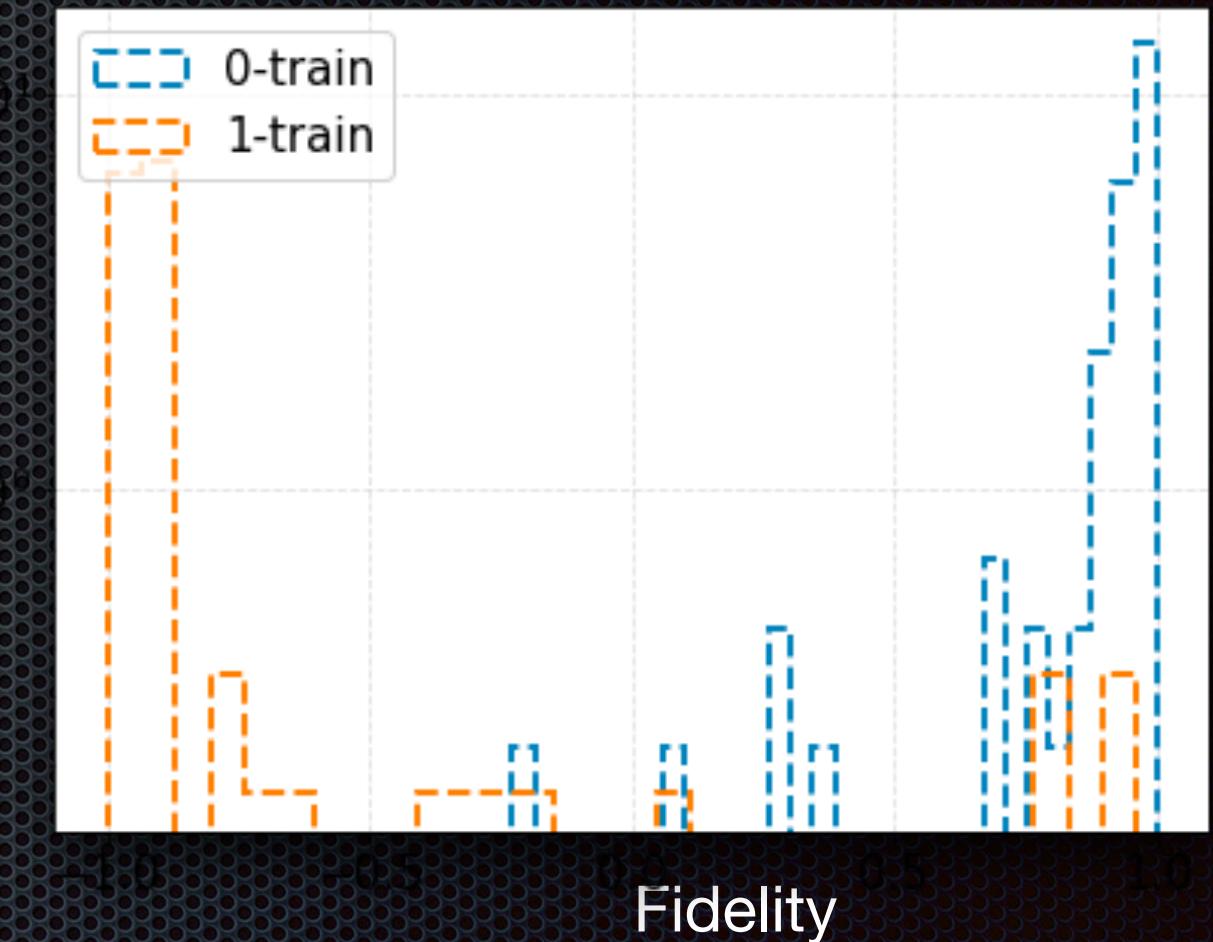
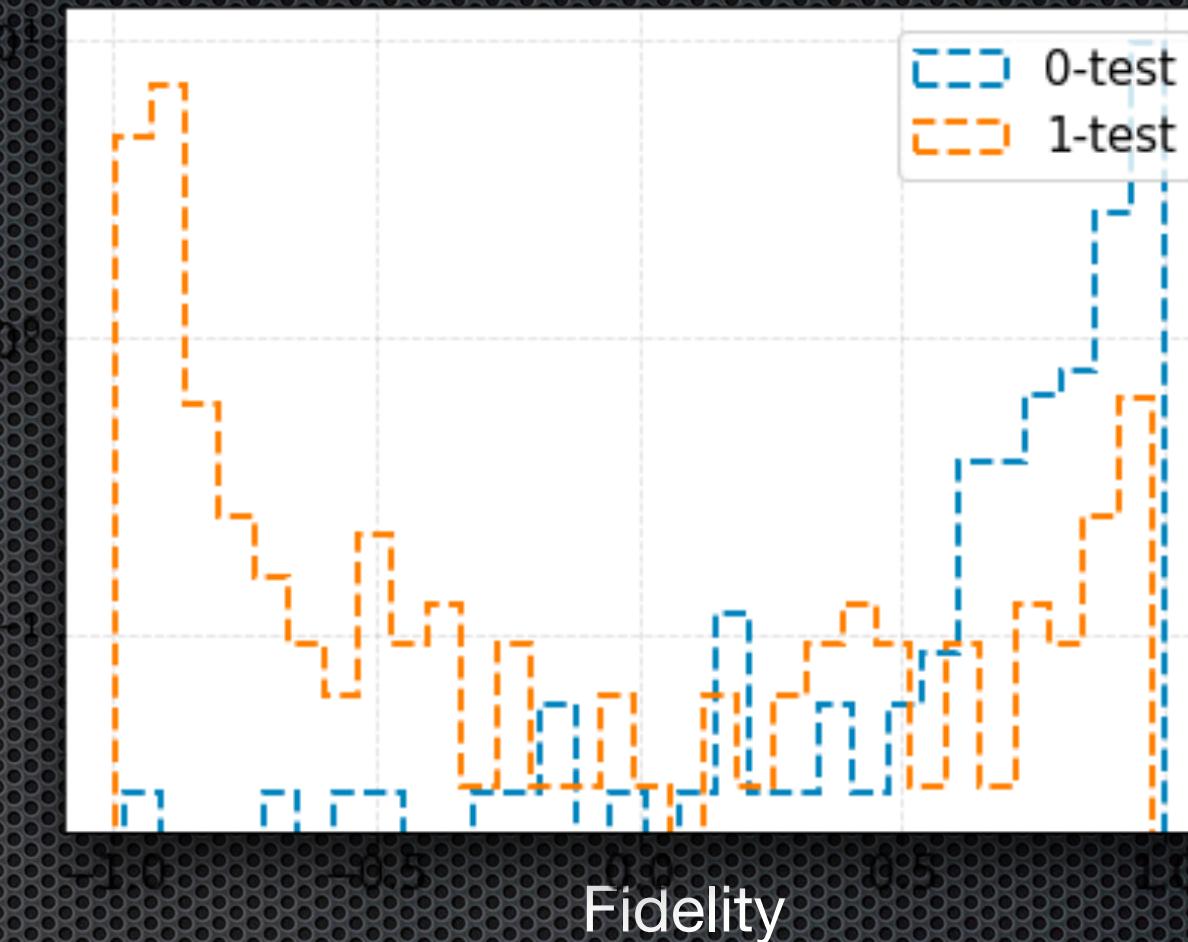
Trousers images

Quantum model for image classification

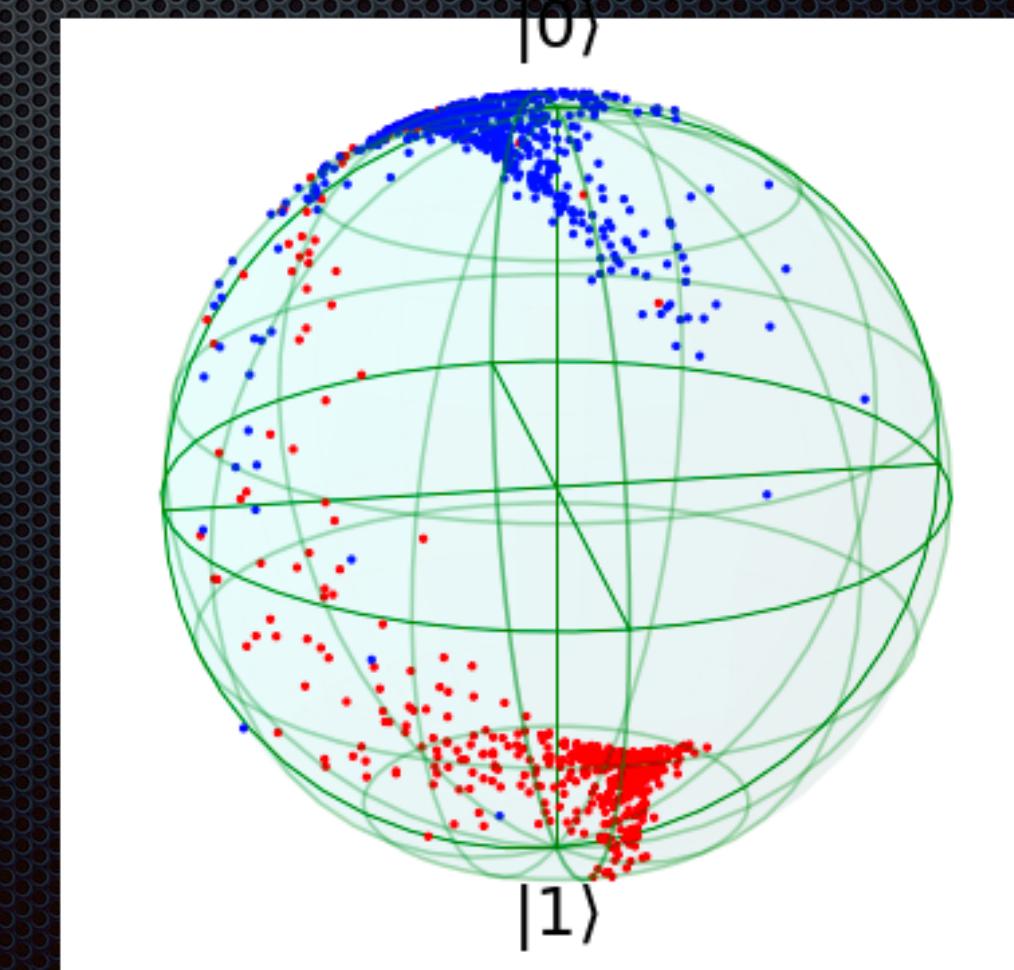
Normalizing the latent space (UnitNorm weights constrain)



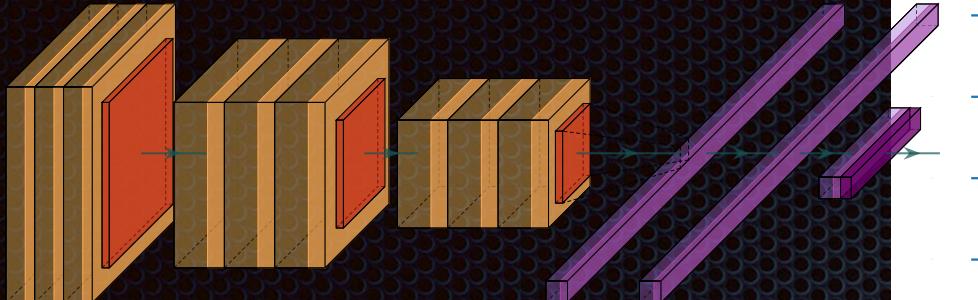
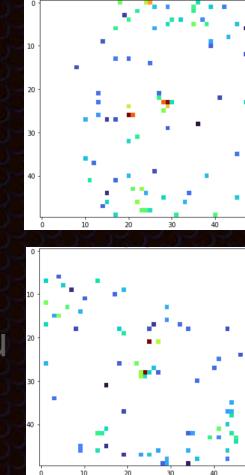
Test accuracy = 93.5% | Train accuracy= 97.0%



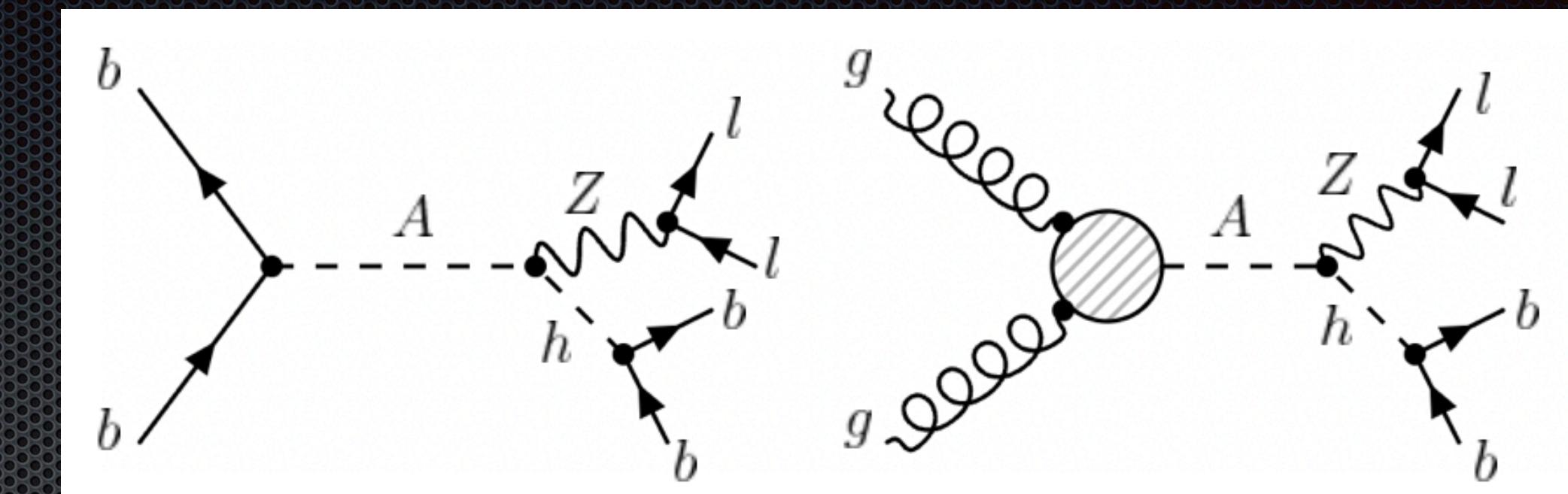
Latent space of
the classical encoder
Normalized



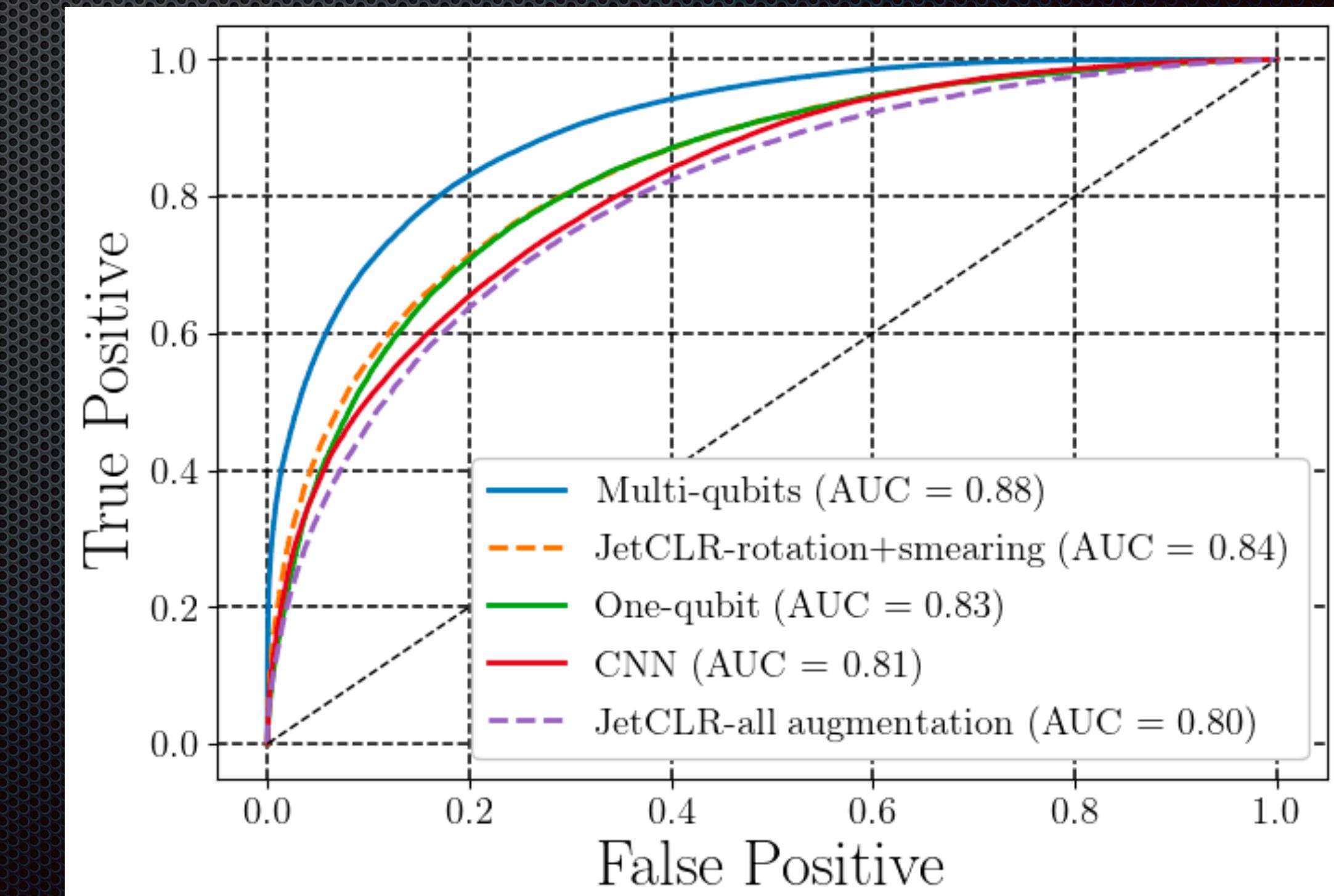
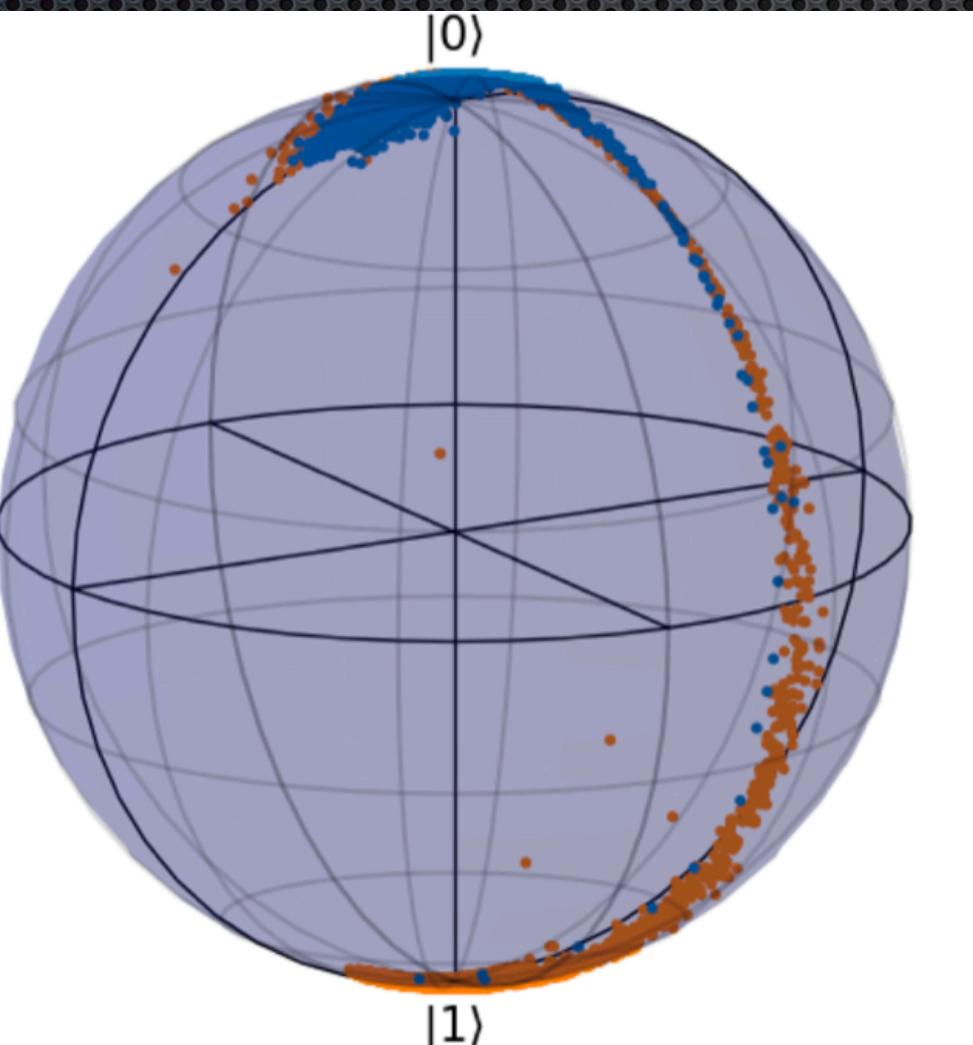
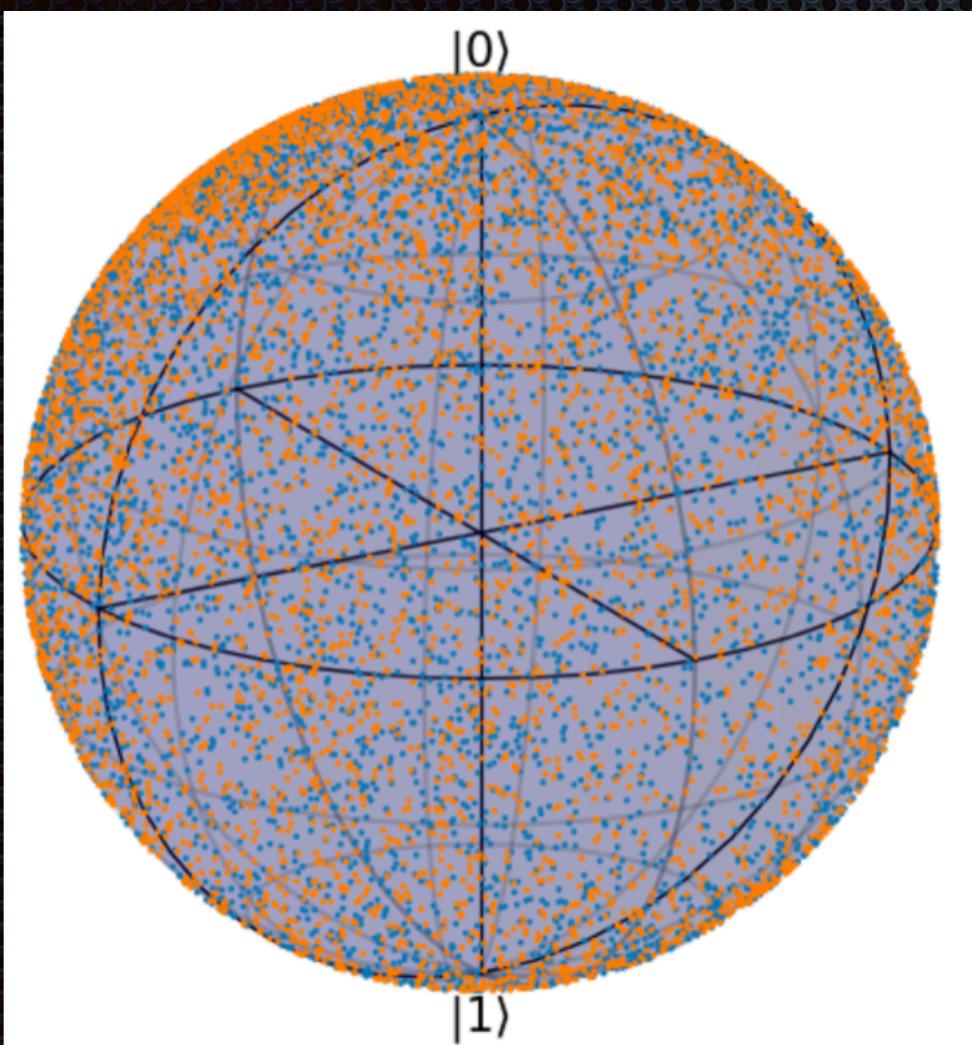
Quantum model for jet image classification



$\mathcal{L} = 6$



A.Hammad,etal



To be continued...