UNIVERSITÀ DEGLI STUDI DI MILANO - BICOCCA

Scuola di Scienze
Dipartimento di Fisica «Giuseppe Occhialini»
Corso di Laurea Magistrale in Fisica

Feature Extraction Neural Networks for Quantum Kernel classifiers: low-energy background rejection in Xenon-Doped LArTPC detectors

Relatore: dott. Andrea Giachero
Correlatore: dott. Daniele Guffanti
Relatore esterno: dott. Michele Grossi

Tesi di Laurea Magistrale
Roberto Moretti
Matr. 825617

Anno Accademico 2021/2022 – Sessione di Ottobre

The 0νββ decay in DUNE

DUNE: Deep Underground Neutrino Experiment
Several physics goals:

High-Energy sector
- Mass hierarchy
- CP violation
- Proton decay

Low-Energy sector
- Supernova neutrinos
- Solar neutrinos
- WIMPs
- $\boldsymbol{0\nu\beta\beta}$

proposals

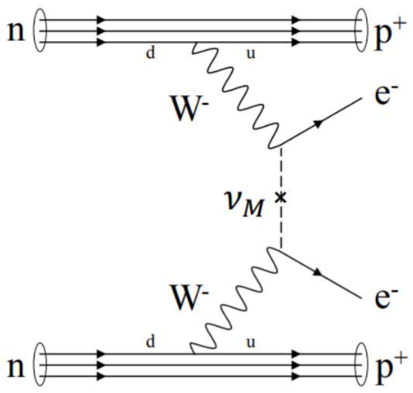
$$\Gamma^{0\nu}_{\beta\beta} = \frac{1}{T^{0\nu}_{\beta\beta}} = G|M^{0\nu}|^2 m^2_{\beta\beta}$$

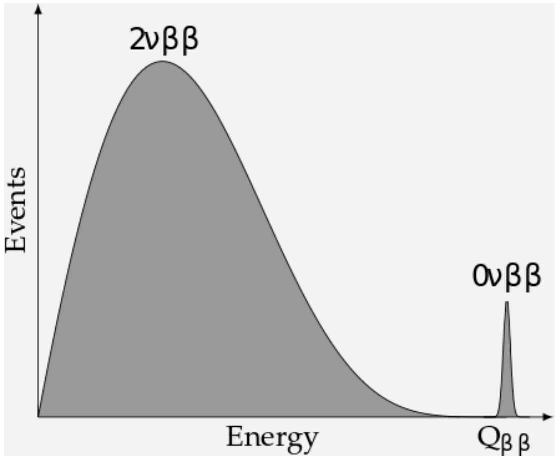
phase space (G) · *nuclear matrix element* ($M^{0\nu}$)

Majorana mass

$$m_{\beta\beta} = \sum_i U^2_{ei} m_{\nu i}$$

The Neutrinoless double beta decay ($0\nu\beta\beta$)
- Hypothetical BSM process
- Consequences:
  - Neutrinos are Majorana particles.
  - Lepton number does not conserve.





Candidate:

$$^{136}\mathrm{Xe}_{54} \rightarrow {}^{136}\mathrm{Ba}_{56} + 2e^- + \cancel{2\bar{\nu}_e}$$

$$Q^{136Xe}_{\beta\beta} = 2.458\ MeV$$

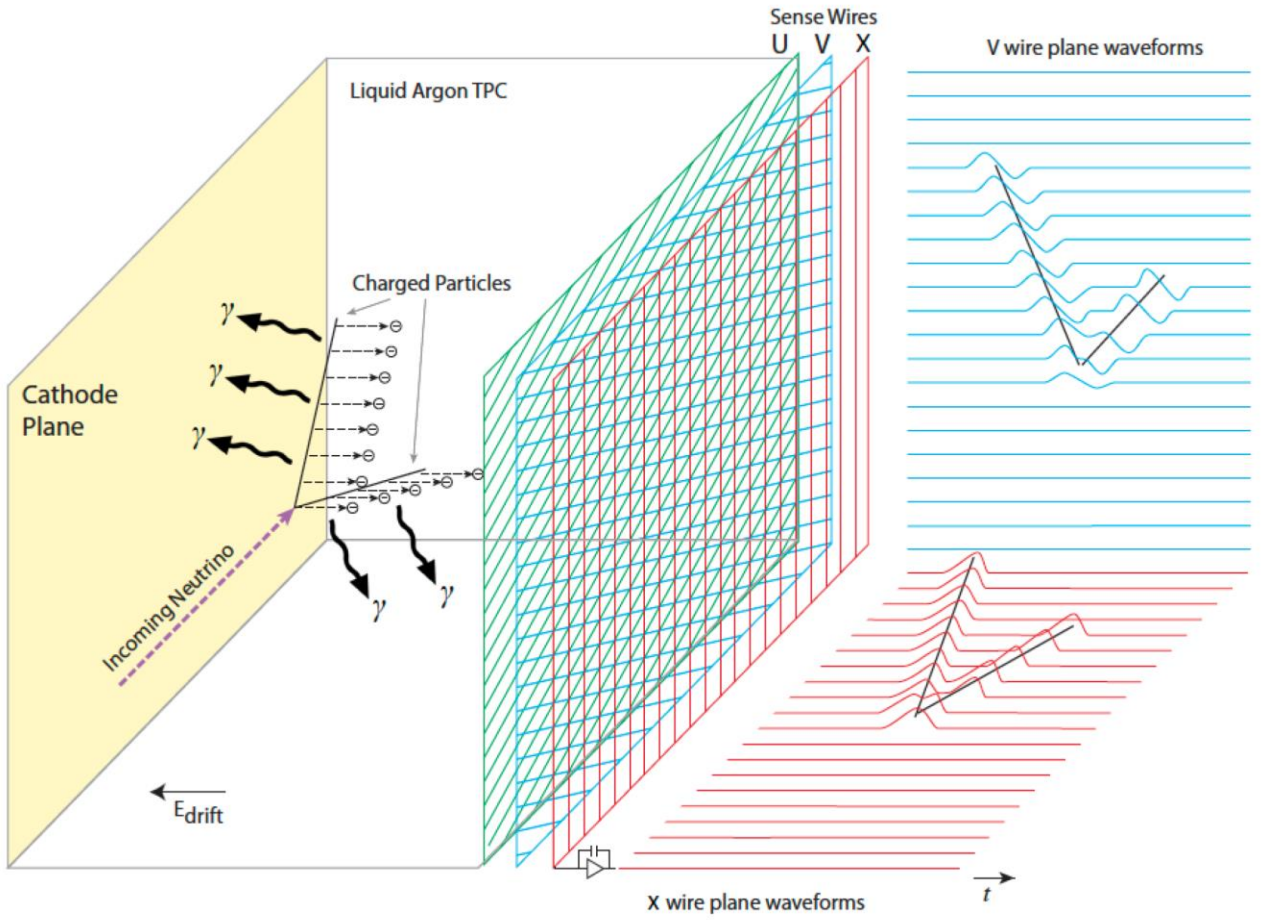
$T^{0\nu}_{\beta\beta} > 1.07 \cdot 10^{26} y$ at 90% $C.L.$

KamLAND-Zen, 2016, PhysRevLett.117.082503

DUNE LArTPCs and Quantum Machine Learning

DUNE is composed of a Near Detector (ND) and Far Detector (FD) facilities.

- FD: four modules of 17kton **Liquid Argon Time Projection Chambers (LArTPCs)**.
- **Proposal:** an «*opportunity*» module with argon doped with xenon at 2% concentration for the search of the $^{136}\mathrm{Xe}\ 0\nu\beta\beta$ decay.
- Careful background studies (β, n, solar ν, etc …)
  β from $^{42}\mathrm{Ar}$ dominates.



Thesis goal: leverage TPC tracking for background mitigation.
Challenging tasks at the MeV-scale in FD LArTPCs:

Opportunity to explore **Quantum Machine Learning** models (QML).

Qubits and Quantum Circuits

A qubit is a **2-level quantum system** described by the wavefunction:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

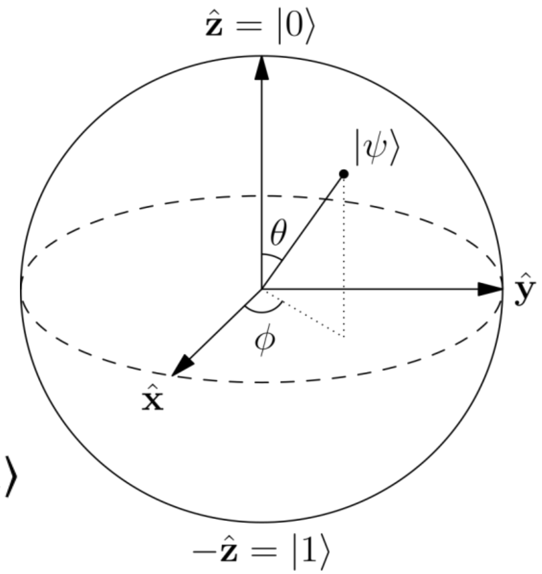
$$|\alpha|^2 + |\beta|^2 = 1 \qquad \alpha;\beta \in \mathbb{C}$$

- Fundamental unit of quantum computation.
- $|0\rangle$ and $|1\rangle$ are the two computational basis, in analogy with 0 and 1 of classical computing.

Qubit states can be visualized as points on a sphere's surface.

Bloch Sphere representation

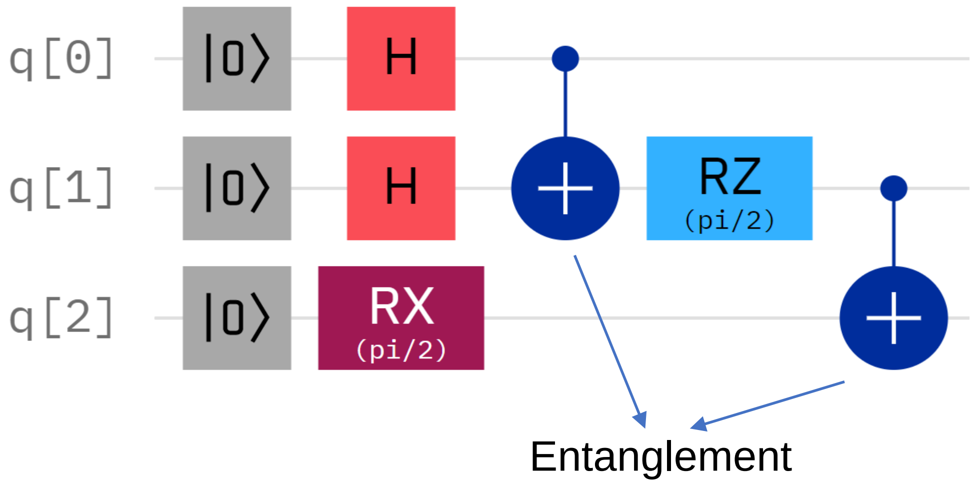
$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$



Qubits are controlled by unitary operators called **quantum gates**, organized in **quantum circuits**.

H	$H\|0\rangle = \frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle) \quad H\|1\rangle = \frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle)$	Hadamard
	$CX\|\psi_0\rangle\|\psi_1\rangle = \|\psi_0\rangle\|\psi_0 \oplus \psi_1\rangle$	CNOT
RZ RX RY	$R_i(\theta) = e^{-\frac{i\theta}{2}\vec{\sigma}_i}$	Pauli rotations

⋮

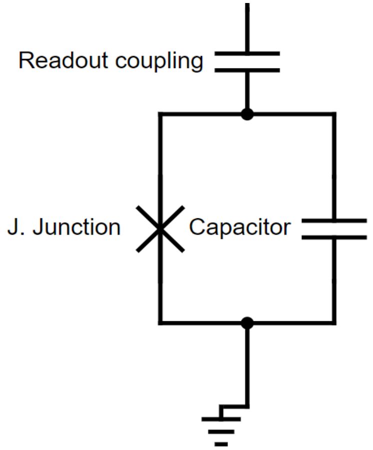


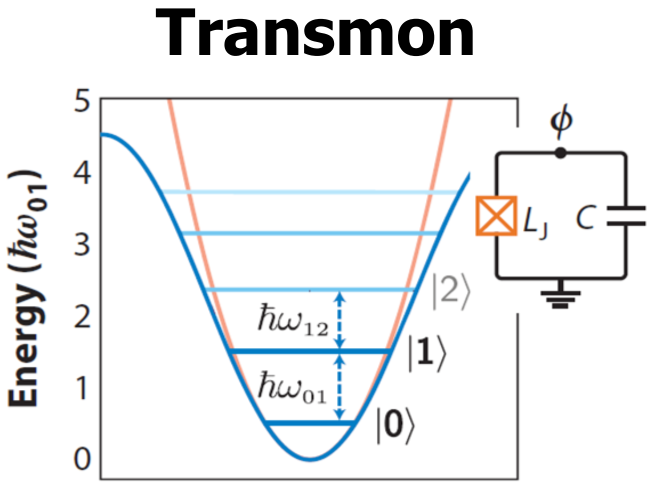
Superconductive qubit and NISQ era

Transmon: superconductive modified LC circuit.
The nonlinear inductance is called **Josephson junction** and allows for anharmonicity in the energy level separation:

↓

It's possible to isolate ground and excited states as $|0\rangle$ and $|1\rangle$ bases.



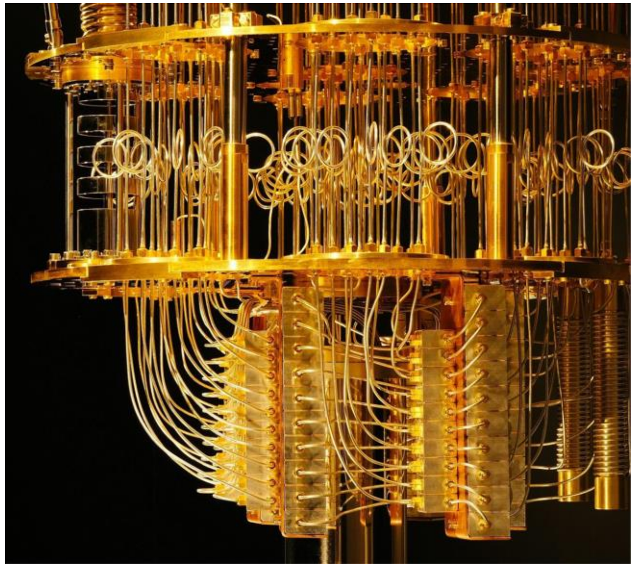


State of the art quantum processors are limited in:

- Noise: computing errors.
- Scalability: increasing qubit numbers.
- Entanglement: not all qubits can be entangled.

→ **Noise Intermediate Scale Quantum** (NISQ) era

Open line of research: achieving proof of quantum advantage with **NISQ** algorithms.
Example: **Quantum Support Vector Machine** (QSVM).



Eagle r1 quantum processor, the 127-qubits IBM quantum computer.

Support Vector Machine

- Well-known Machine Learning model suited for binary and multilabel classification.
- Useful for signal/background discrimination.

Task: binary classifications of feature vectors $\vec{\boldsymbol{x}} \in \mathbb{R}^{\boldsymbol{n}}$
i.e. predicting the class outcome $y \in \{-1;\ +1\}$.

Idea: given a **feature map** $\phi(\vec{x}),\ \phi(\vec{x}_i) \in M\colon \dim(M) = m > n$,
finding the best linear decision boundary $\vec{w}^{\boldsymbol{T}} \boldsymbol{\phi}(\vec{\boldsymbol{x}}) - \boldsymbol{b} = \boldsymbol{0}$
by maximizing:

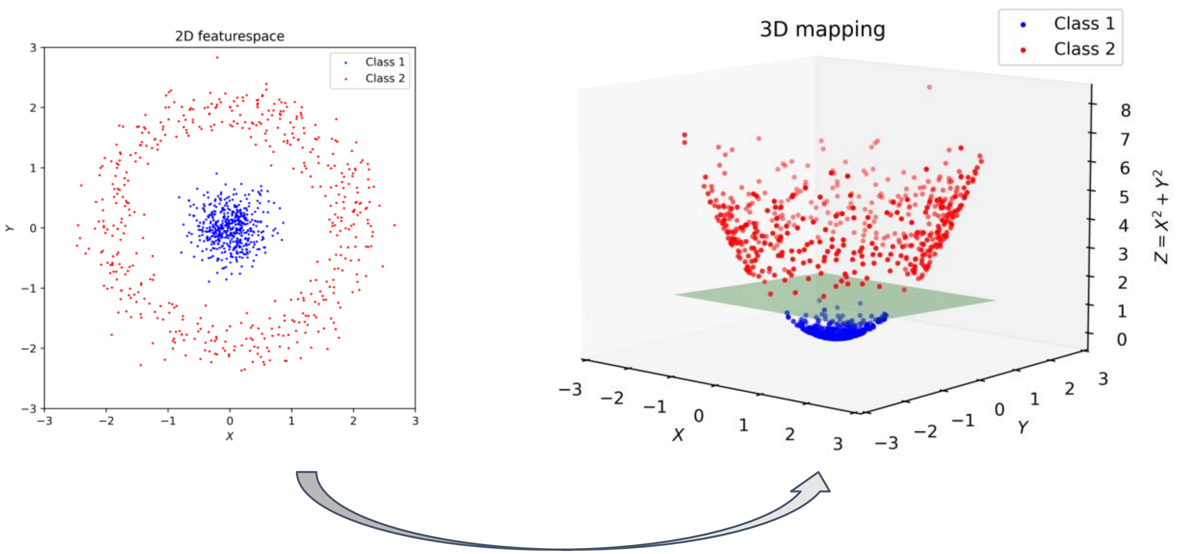
$$f(c_1, c_2, \dots, c_n) = \sum_i c_i - \frac{1}{2} \sum_{ij} y_i c_i y_j c_j \underbrace{\langle \phi(\vec{x}_i), \phi(\vec{x}_j) \rangle}_{\boldsymbol{k}(\vec{\boldsymbol{x}}_{\boldsymbol{i}}, \vec{\boldsymbol{x}}_{\boldsymbol{j}})}$$

with $\vec{w} = \sum_i \boldsymbol{c_i y_i \phi}(\vec{\boldsymbol{x}}_{\boldsymbol{i}})$.

$\downarrow$

Kernel function

When projecting on the original feature space, the decision boundary will be generally nonlinear.



Feature embedding

Common kernel choices:

Linear	$K(\vec{x}_i\ ,\ \vec{x}_j)\ =\ \vec{x}_i \cdot \vec{x}_j$
Polynomial	$K(\vec{x}_i, \vec{x}_j) = (\gamma\, \vec{x}_i \cdot \vec{x}_j + r)^d$
RBF	$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\gamma \lvert\lvert \vec{x}_i - \vec{x}_j \rvert\rvert^2 + C\right)$

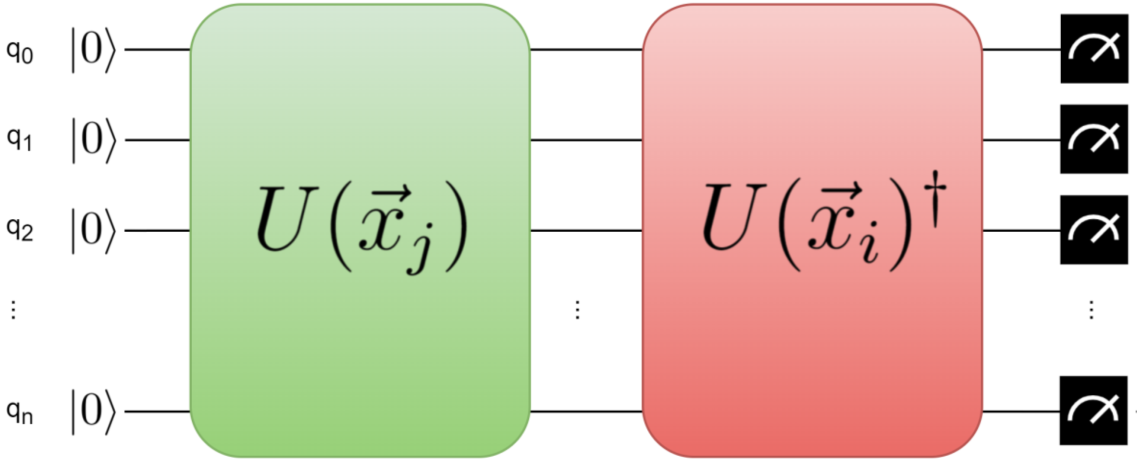
Quantum Kernel

Promoting the classical feature mapping to a quantum state:

$$\phi(\vec{x}) \rightarrow |\phi(\vec{x})\rangle\langle\phi(\vec{x})| =$$
$$= U(\vec{x})|0\rangle\langle 0|U(\vec{x})^\dagger$$

$$K(\vec{x}_i, \vec{x}_j) = \left|\langle 0|U(\vec{x}_i)^\dagger U(\vec{x}_j)|0\rangle\right|^2$$

- Feature maps are still implicitly defined.
- Kernel function is still a measure of similarity between different samples.



Quantum circuits of with this structure are suitable kernels.

Pros:

- Hilbert space grows rapidly with qubit's number
  - **Expressive classifiers.**
- Quantum kernels are generally hard to compute classically
  - **No classical counterpart.**
- Good results even with small sized circuits
  - **Is a NISQ-era algorithm.**

Cons:

- Lack of featuremap explainability
  - Unintuitive relation between circuit and outcome.
- Usually set arbitrarily
  - Problem of chosing a good Quantum Kernel.

Room for quantum advantage.

Toy model dataset

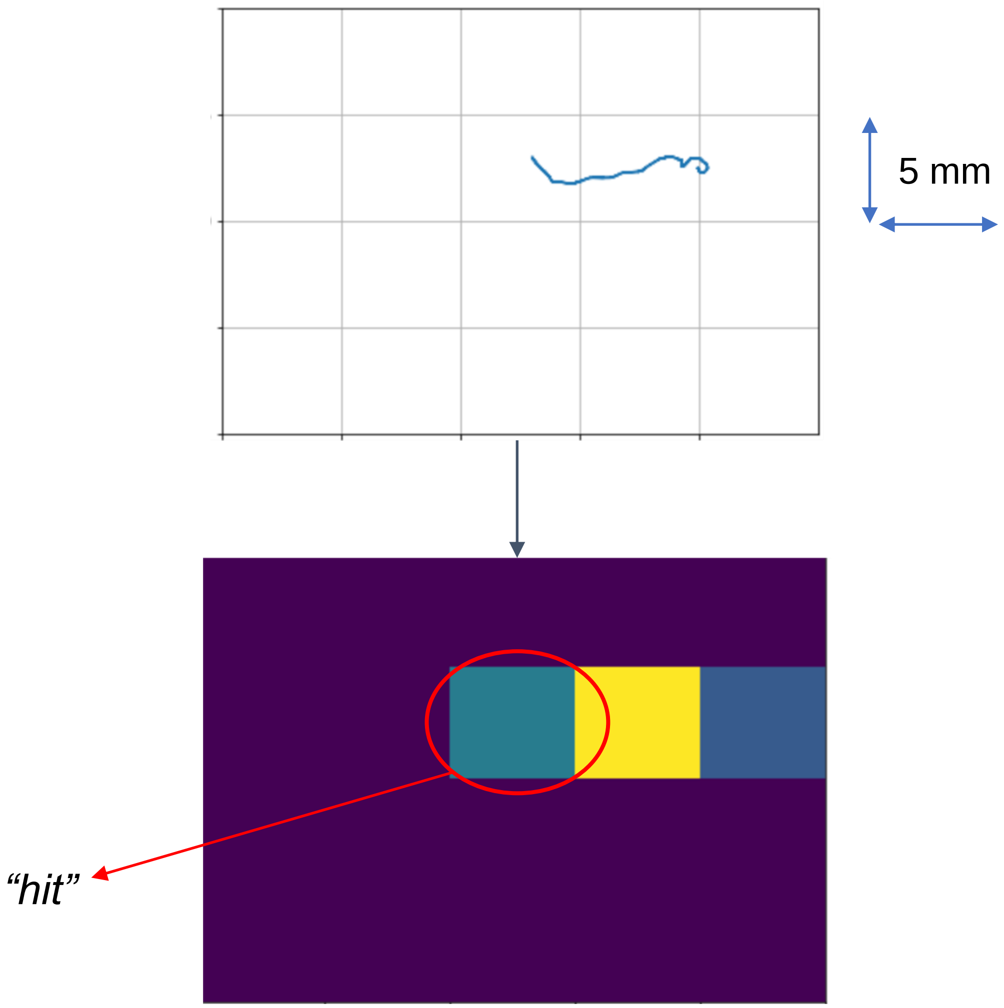
Datsaset:

- Geant4 simulated high-resolution **β** and **ββ** tracks in LAr at $E = Q_{\beta\beta}^{136Xe} = 2.458\ \mathrm{MeV}$.
- Tracks have been downsampled to **3D voxelized data** with resolution (bin-widths) of $[\mathbf{5} \times \mathbf{5} \times \mathbf{1}]$ **mm³**, in order to simulate a DUNE-like spatial resolution.
- Detector effects not accounted.

Issues:

- DUNE's energy resolution won't be excellent at MeV energies.
- Track lengths are similar to the spatial resolution ($R_{CSDA} \sim 1.2\ \mathrm{cm}$ for $2.5\ \mathrm{MeV}$ elecrons).

→ **Hard classification problem**.

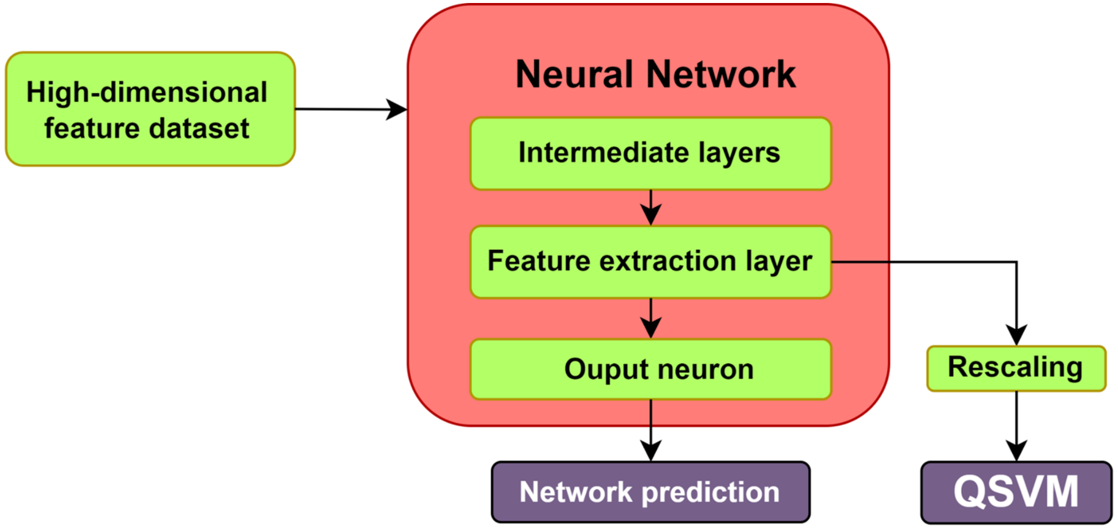


(XY) view (5 x 5) mm²

Feature extraction

For implementing the NISQ Quantum Support Vector Machine (QSVM) with LArTPC measurements, the input features must be reduced, while maintaining useful informative content.

Proposed approach: training Neural Networks as *standalone* classifiers, while defining specific *feature extraction layers* for the QSVM input.

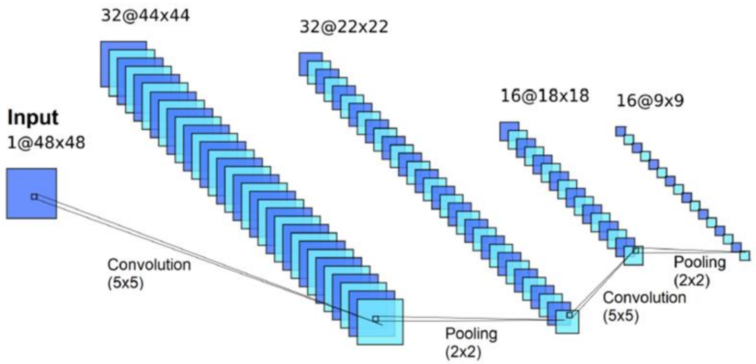


Models investigated:

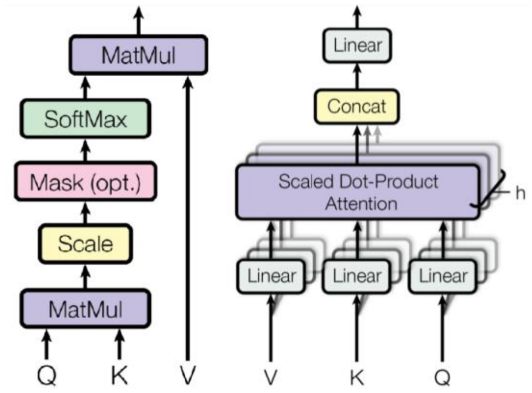
- **Convolutional Neural Network (CNN)**
  Tracks are elaborated visually and features are extracted making use of convolutional filters.

- **Transformer (or Attention Network) (ATN)**
  Tracks are treated as sequences of correlated hits, recurring to the mechanism of self-attention.

CNN



ATN

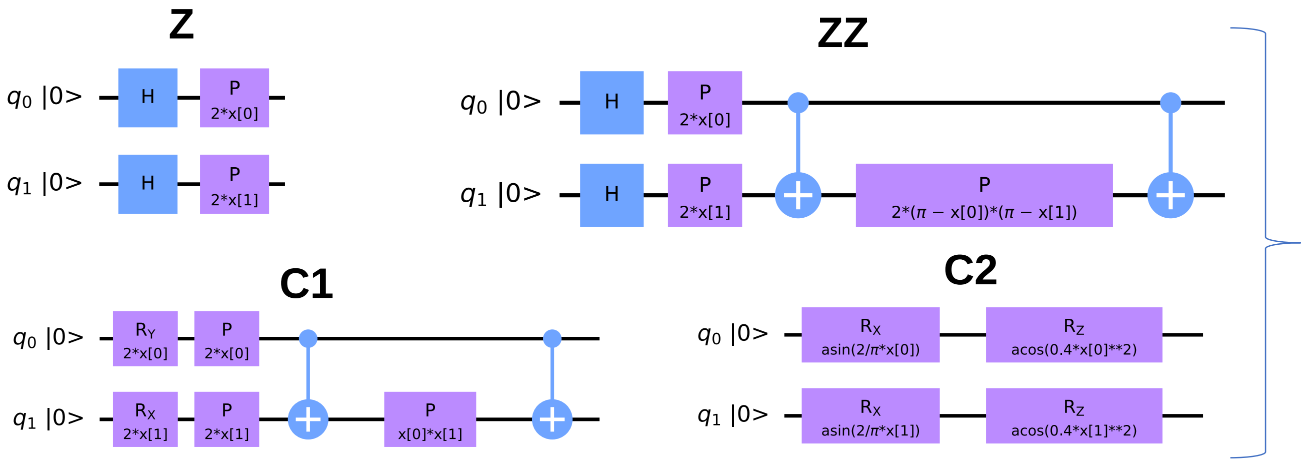


Quantum Kernel choices

Heuristic approach – arbitrary circuits

Creating with trial and error, or suggested by symmetries in the feature distribution.

- **Z, ZZ:** Pauli Feature Map class.
- **C1, C2**: Defined arbirarily.

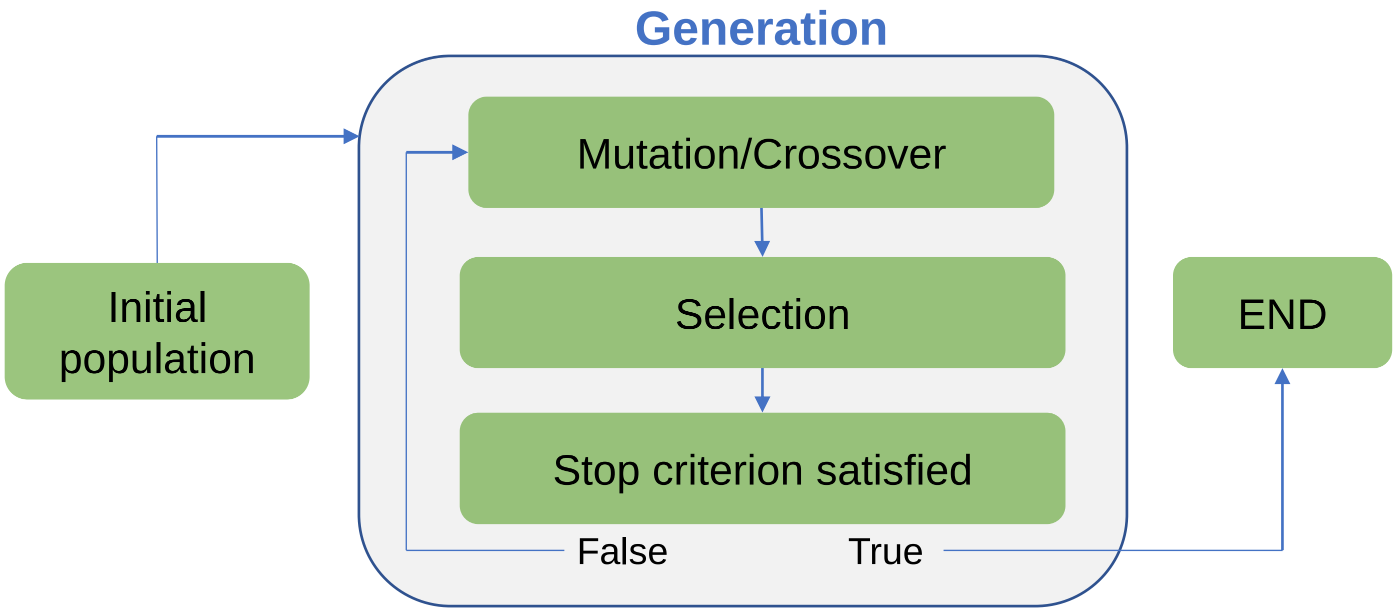


Two-qubits case

Meta-heuristic approach – Genetic algorithm

- *Fitness function* – quantifies the goodness of a kernel.
- *Mutation* and *Crossover* operators – introduce variability through generations.
- A *parent/offspring* selection criteria.
- Initial population – Generation zero.

Goal: specialize the kernel population for the given classification task.



Two-qubits QSVM

Accuracy comparison between SVM and QSVM – two feature case.

Classical:
Linear
Polynomial
Rbf

Quantum:
Z
ZZ (*ent*)
C1 (*ent*)
C2

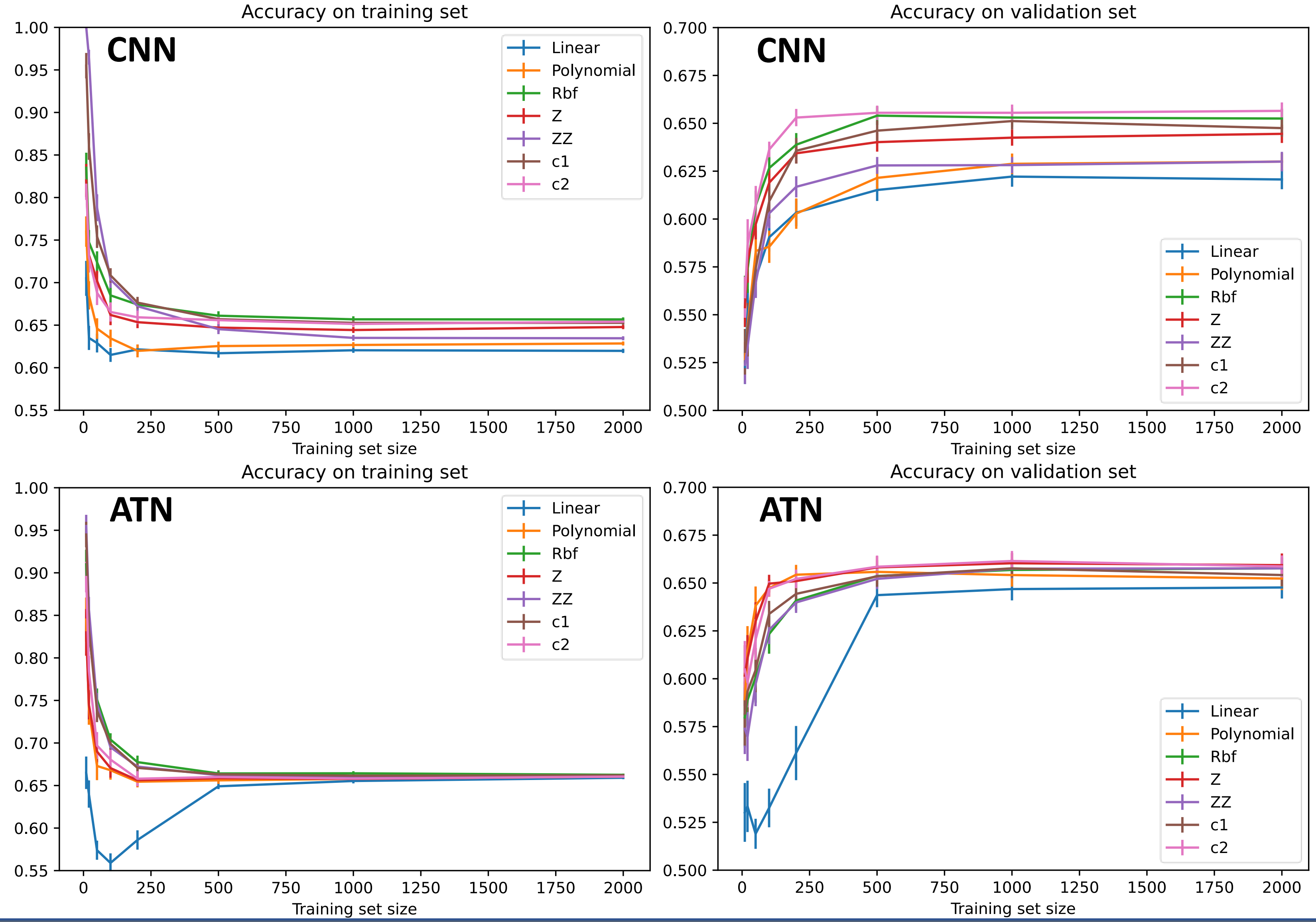
CNN

- Best kernels: C2 and RBF.
- C2 has the steepest validation curve.
- Accuracy ~ 66%.

ATN

- All kernels share similar behaviour.
- Accuracy ~ 64%.

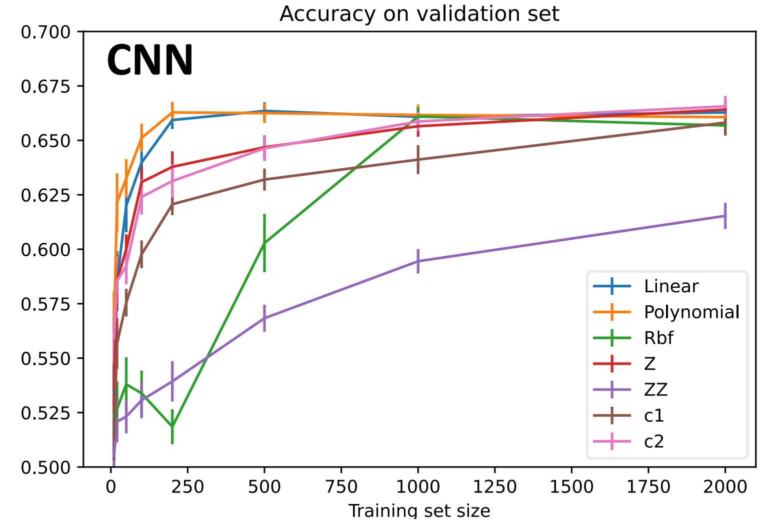
Ideal QSVMs have been simulated **classically**



More qubits and the role of entanglement

- More qubits and features have been added to explore larger embedding spaces.
- Some Quantum Kernels tends to **lose generalizability** when increasing the circuit size.

10-qubits:

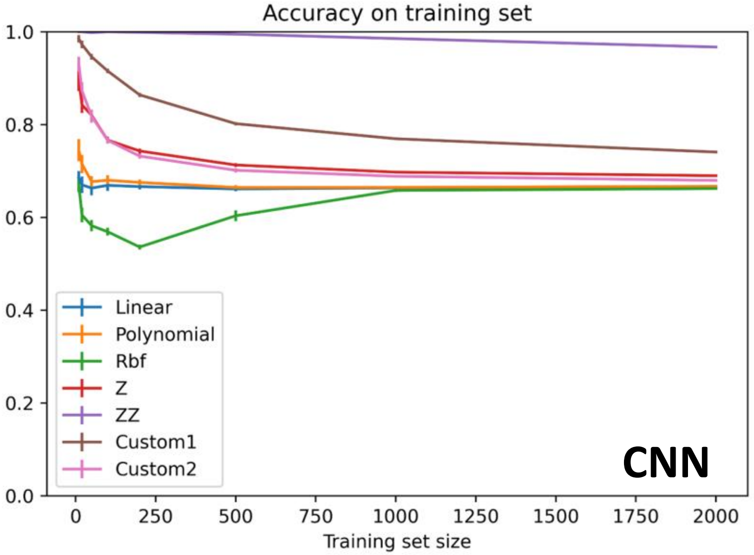


The accuracy on a training set is higher for Quantum Kernels, especially the ones with entanglements.
→ **Quantum Kernels with entanglement are very expressive, at the risk of overfitting.**

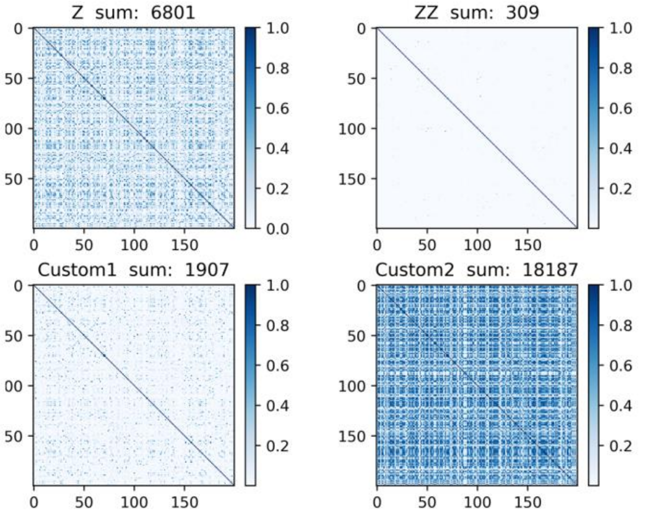
For such kernels, the Gram matrix:

$$G_{ij} = K(\vec{x}_i, \vec{x}_j) = \langle \phi(\vec{x}_i), \phi(\vec{x}_j) \rangle$$

approximates the Identity.



Training accuracies with CNN-extracted features.



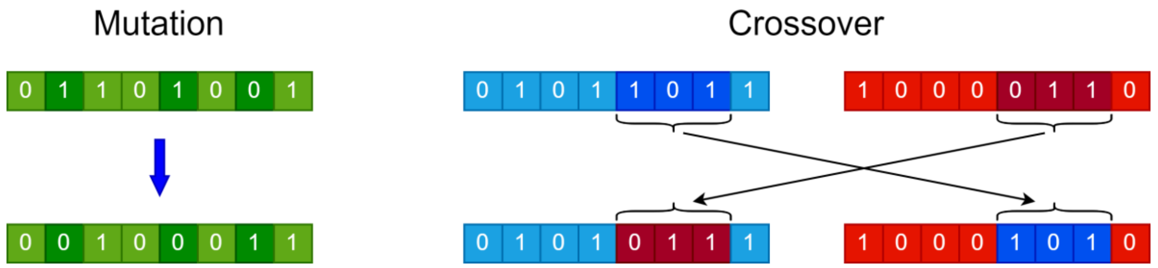
Kernel's Gram matrices.

Genetic Quantum Kernels

- Recurring to a **binary representation** of the quantum feature map circuit, for applying Mutation and Crossover.
- Each gate has been described by **6 bits sequences**, encoding the gate type and the respective rotation angle.
- Rigid constraint: two qubits and three gates per line.

First 3 bits	*Gate type*
000	I
001	H
010	$\oplus$
011	RZ
100	RX

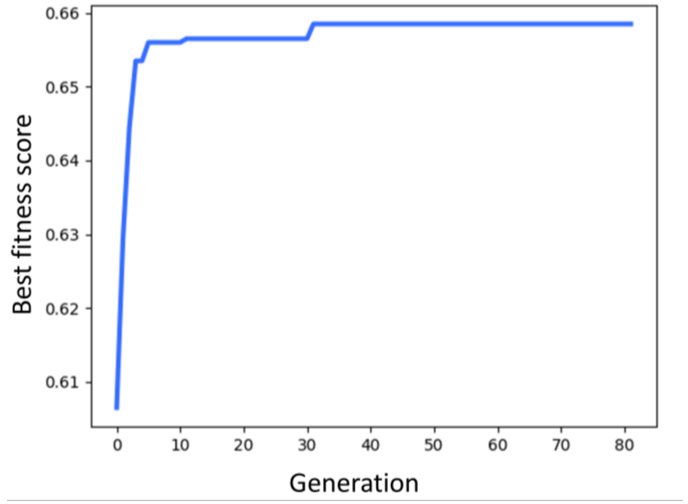
Last 3 bits	*Rotation angle*
000	$2x_0$
001	$2x_1$
010	$\text{asin}\left(\frac{2}{\pi}x_0\right)$
011	$\text{acos}\left(\frac{4}{\pi^2}x_0\right)$
100	$(\pi - 2x_0)(\pi - 2x_1)$



$$Fitness\ function = Valid.Accuracy - 10^{-4} * Gate\ Number$$

→ **Maximizing accuracy, preferring smaller circuits for the same classification performance.**

Fitness function vs. generation number



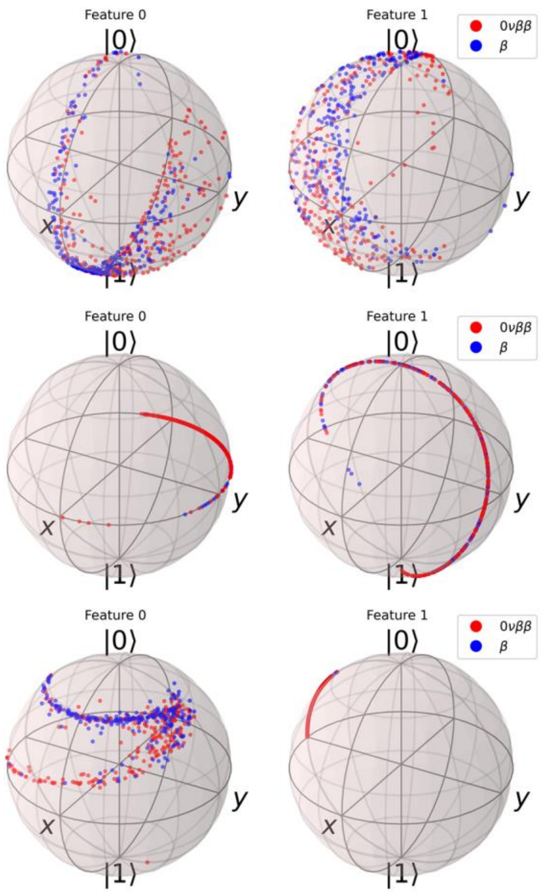
Genetic outcome and performances

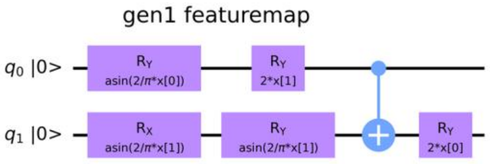
The Genetic Algorithm produced satisfactory quantum feature maps, giving new insights on the datasets.

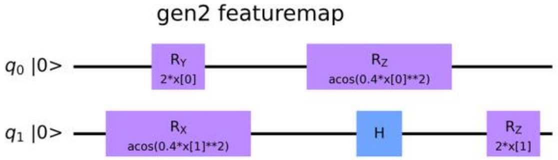
CNN

CNN

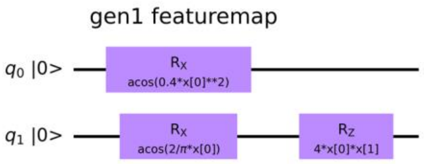
ATN



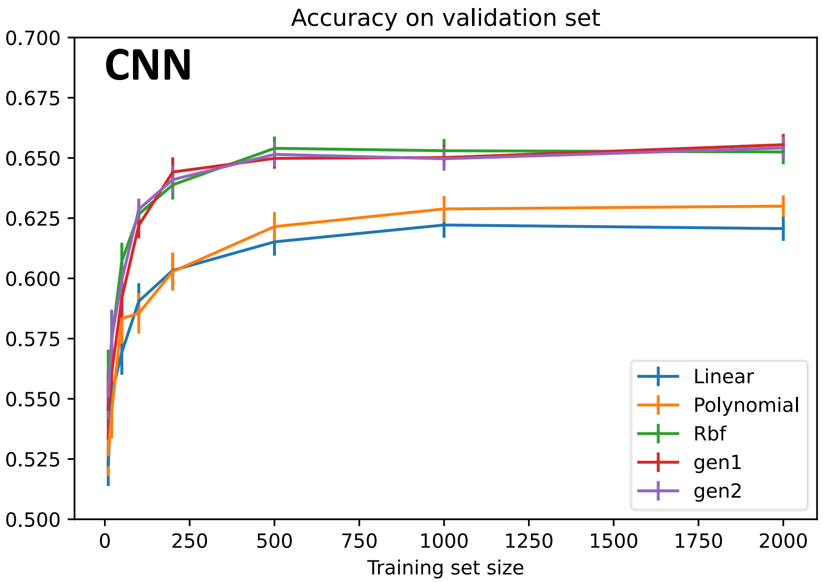


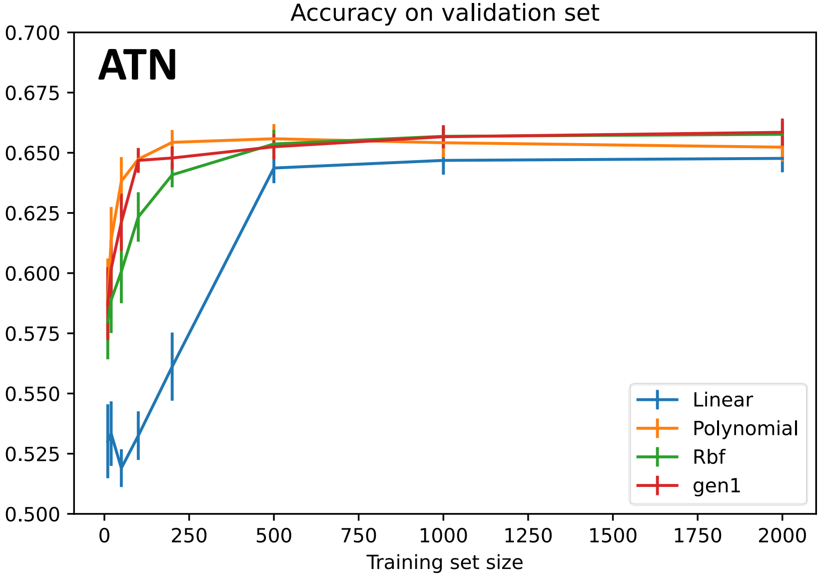


Also entanglement has been premiated



Only three-gates feature map



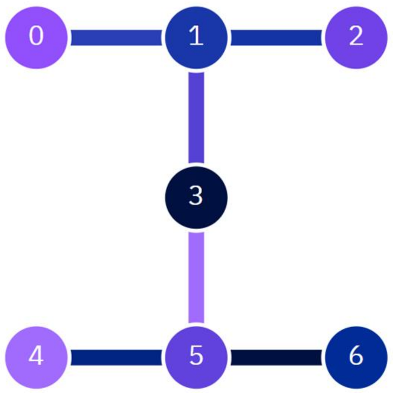


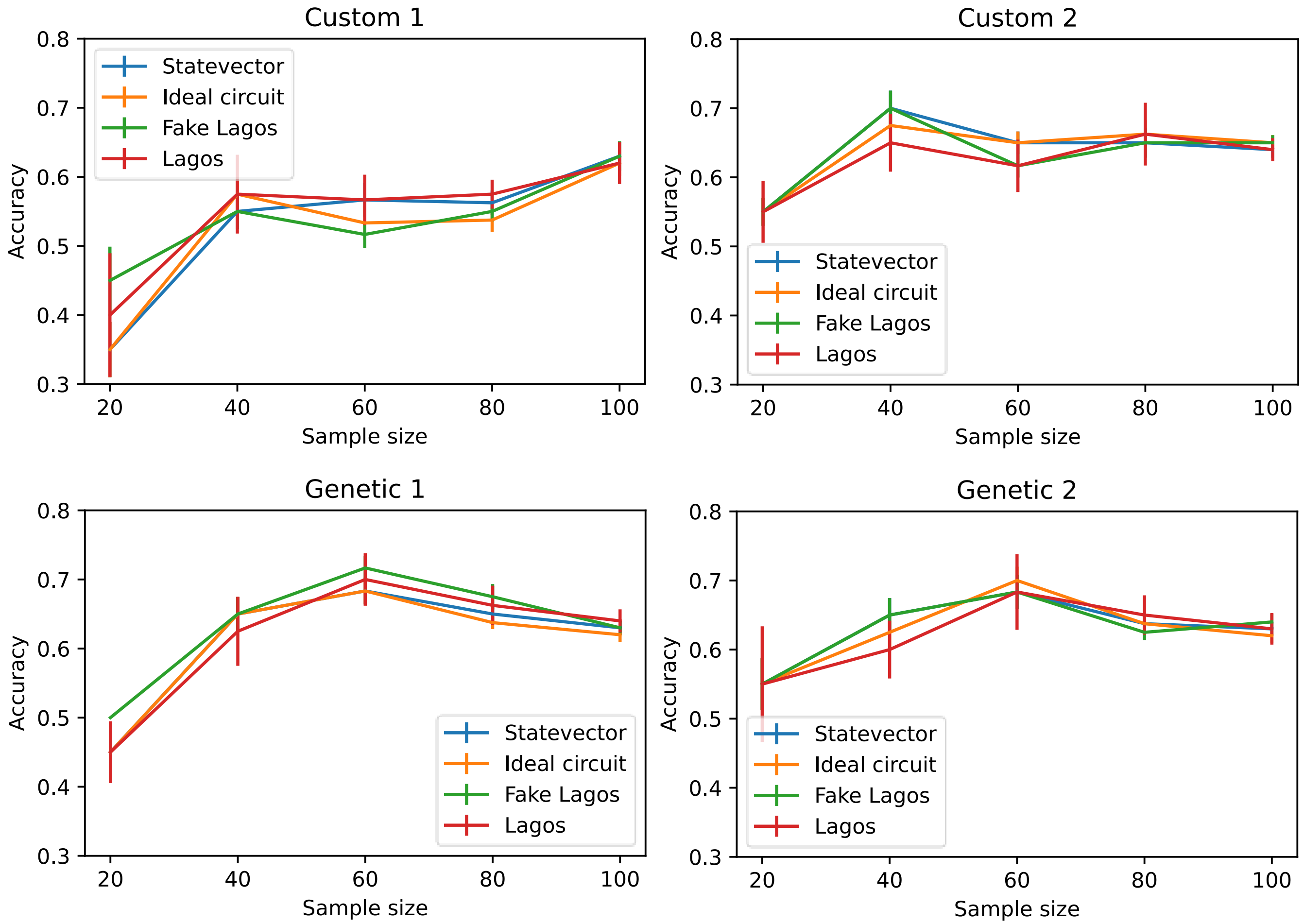
QSVM test on IBM hardware

QSVMs were trained with different *backends:*

- ***Statevector***
  Access to probability amplitudes.
- ***Qasm simulator***
  Simulating an ideal circuit behaviour.
- **Simulated noise**
  Simulating the noise scheme of a real device.
- **Real hardware**
  Remotely running on one of IBM’s devices.

Tested hardware:
ibm_lagos





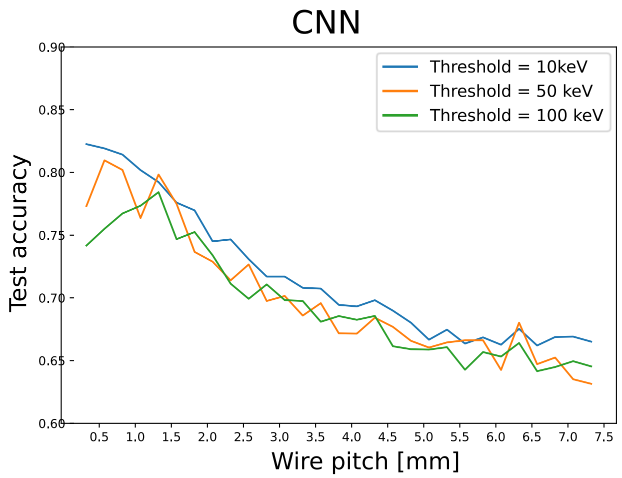
No evidence of performance loss due to noise.

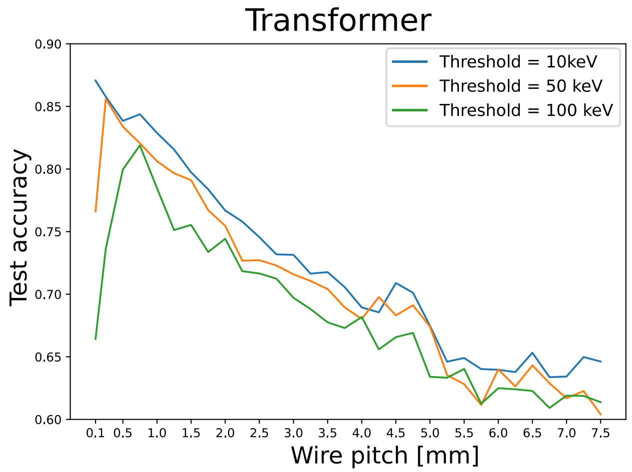
Enhancing the track resolution

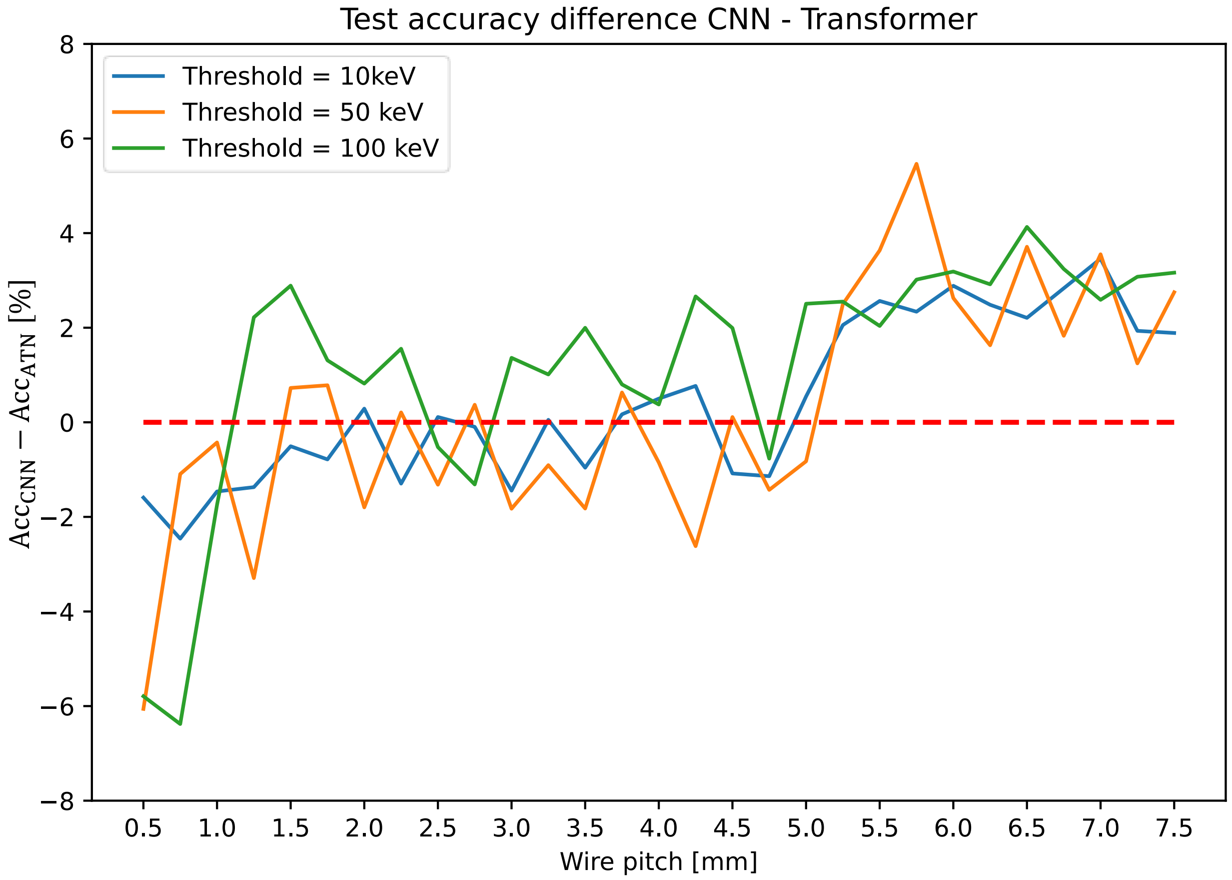
By changing the wire pitch on the toy model dataset, it was possible to highlight three regimes:

- > 5 mm: CNN outperforms ATN.
- Between 5 mm and 1 mm: Transition region.
- < 1 mm: ATN outperforms CNN.

→ Few hits available: **CNN**
→ Many hits available, ratio hit number / picture size becomes small ($\lesssim 8\%$): **ATN**



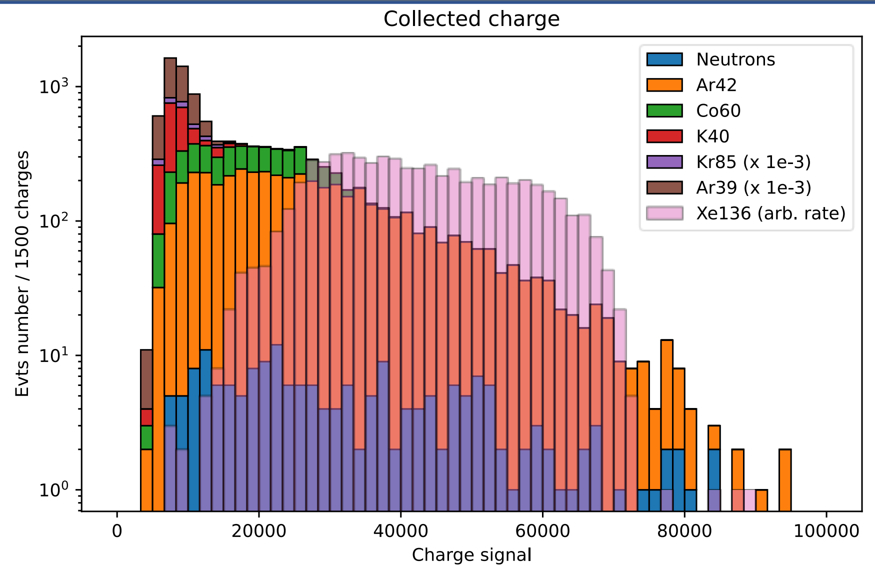




DUNE FD Horizontal Drift dataset

Simulation of radiological samples and $0\nu\beta\beta$ events collected at the DUNE Far Detector Horizontal Drift module, **accounting for detector effects**.

- **Dominant** bkg: 42Ar (β topology).
- **Subdominant** bkg: neutron capture (n topology).

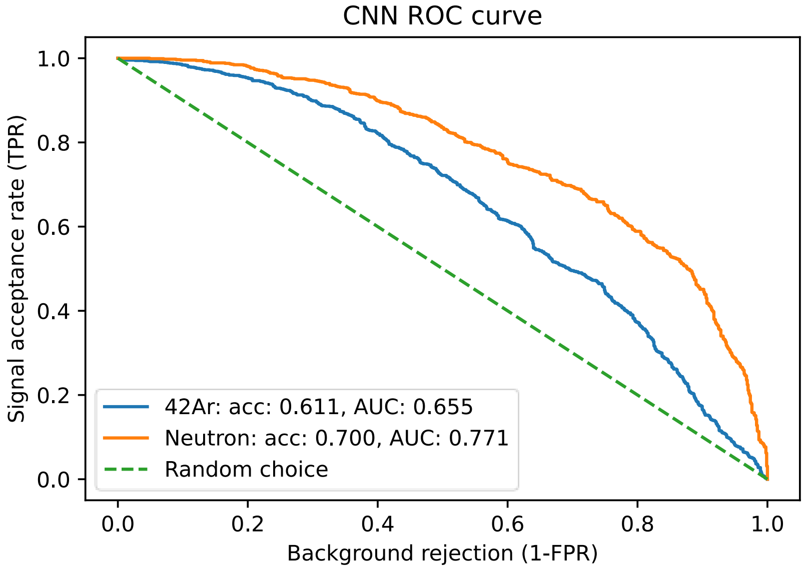


Charges signal at the LArTPC collection plane for radiological background and $0\nu\beta\beta$ events.

CNN has been trained as a binary classifier:

- $\mathbf{0\nu\beta\beta - \beta}$: accuracy $\sim 61\%$.
- $\mathbf{0\nu\beta\beta - n}$: accuracy $\sim 70\%$.

→**Neutron rejection is easier to achieve.**



Tradeoff efficiency/purity.

Conclusions

Physics

- **Modest $\beta\beta$ vs β classification accuracy** (66%) for an ideal 5mm pitch LArTPC.
- Slight decrease in DUNE FD Horizontal Drift LArTPC (61%)
  $\rightarrow$ **Depleted argon and better spatial resolution are mandatory.**
- Better performance in neutron discrimination (70%).
- Interesting technique for **other low energy physics channels in DUNE.**

Preprocessing – Neural Networks

- ATN outperforms CNN with high-resolution tracking (pitch $< 1\ \text{mm}$).
- CNN outperforms ATN with low-resolution tracking (pitch $> 5\ \text{mm}$).

Quantum classifier – QSVM

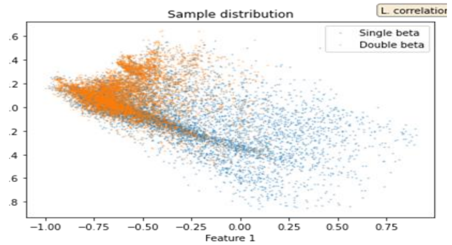
- Kernel function evaluated with a quantum circuit.
- A genetic algorithm provides suitable quantum kernels.
- QSVMs have proven to be **noise-robust on a 7-qubit IBM processor.**
- **QSVMs matches** the classical **SVM** performances.

Backup

Workflow outline

Binned tracks

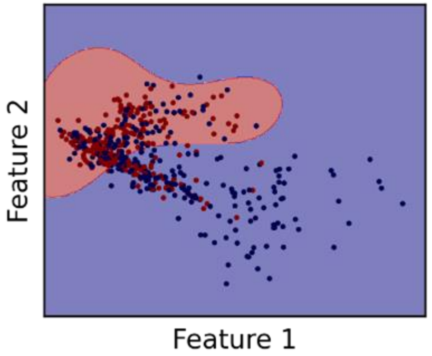
Feature extraction (CNN/ATN)



Distributions of NN-extracted features (2D case).

Rescaling and outlier removal

Classical or quantum classifier (SVM/QSVM)



Classifier's **decision boundary.**

Prediction

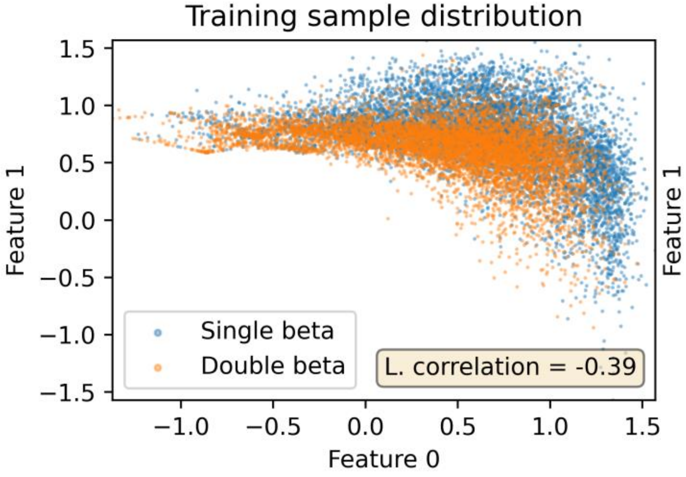
$\beta\beta$ (signal) or β (background).

CNN and ATN standalone predictions

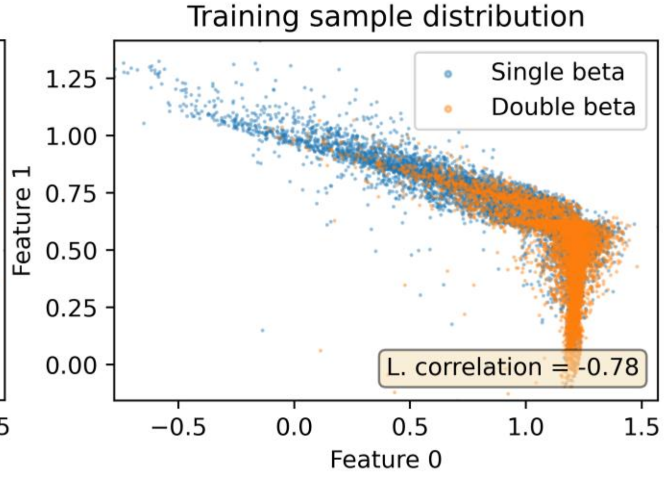
Both models have been trained for different number of neurons in the feature layers.
In both cases the accuracies on a validation set don't depend on the number of extracted features.

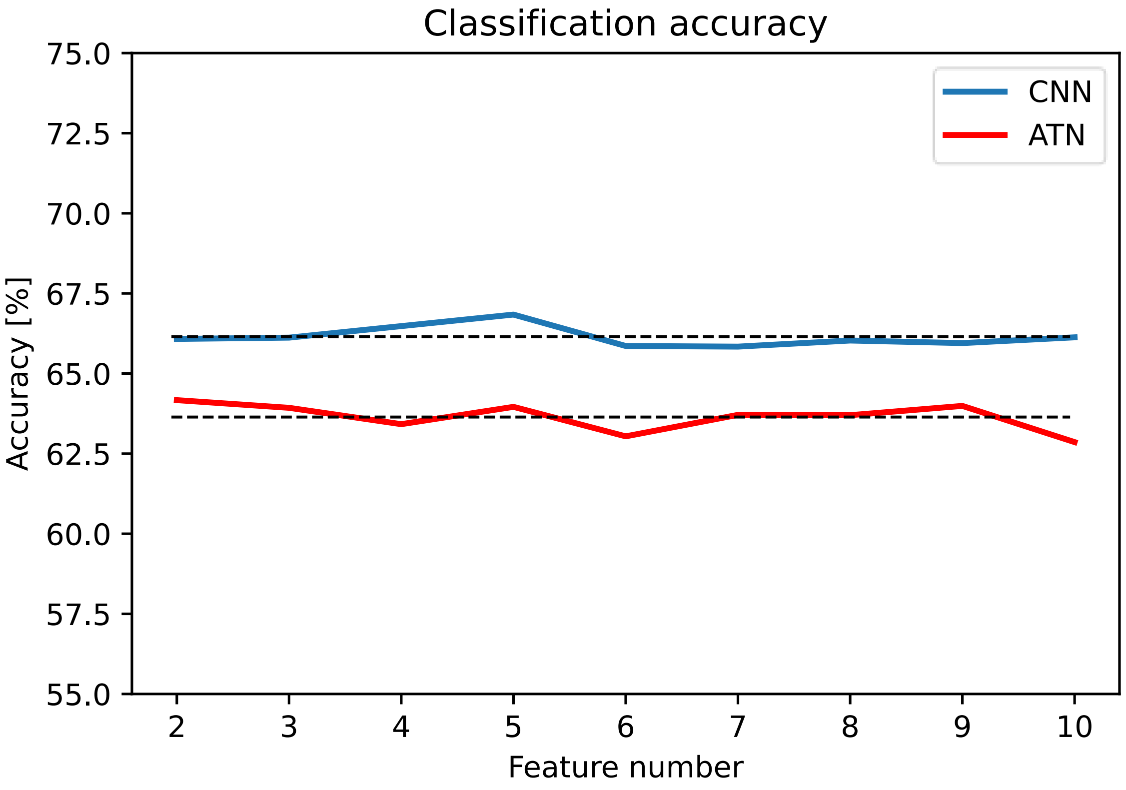
It's possible to give in input to SVM and QSVM just two features without compromising the overall performance → suitable for small quantum circuits

CNN



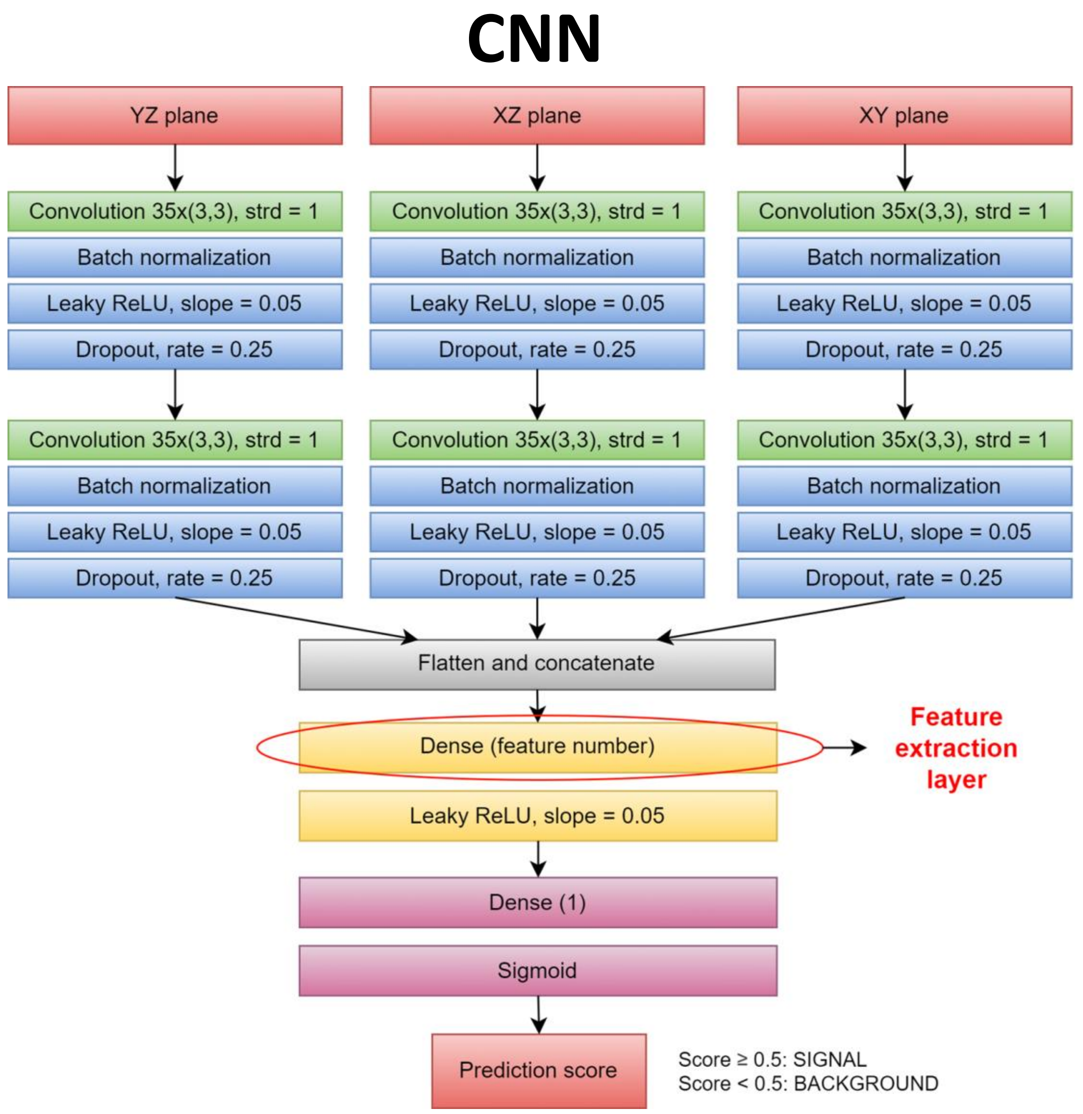
ATN

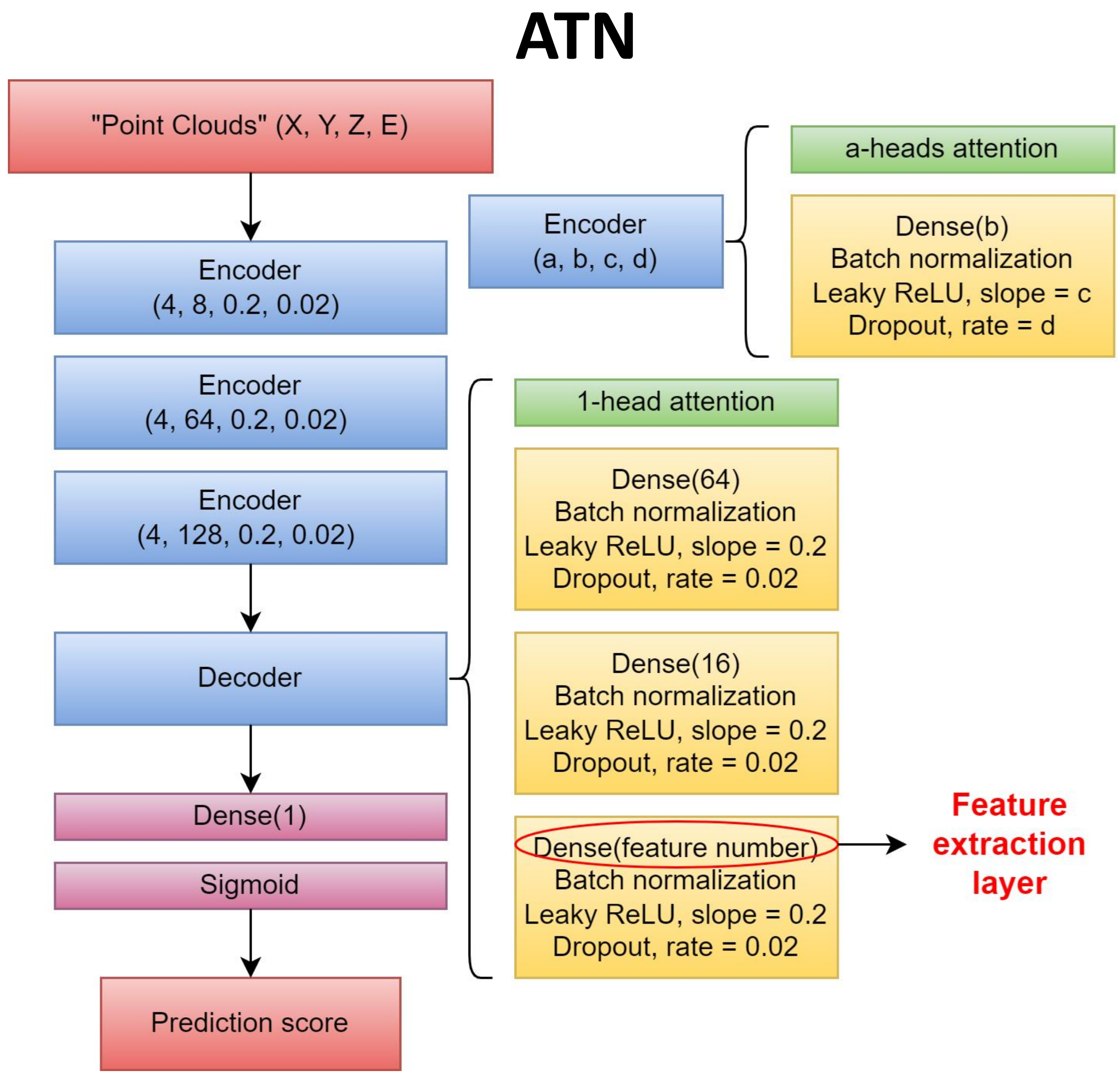




CNN accuracy: ~ 66%
ATN accuracy: ~ 64%

CNN and ATN architectures

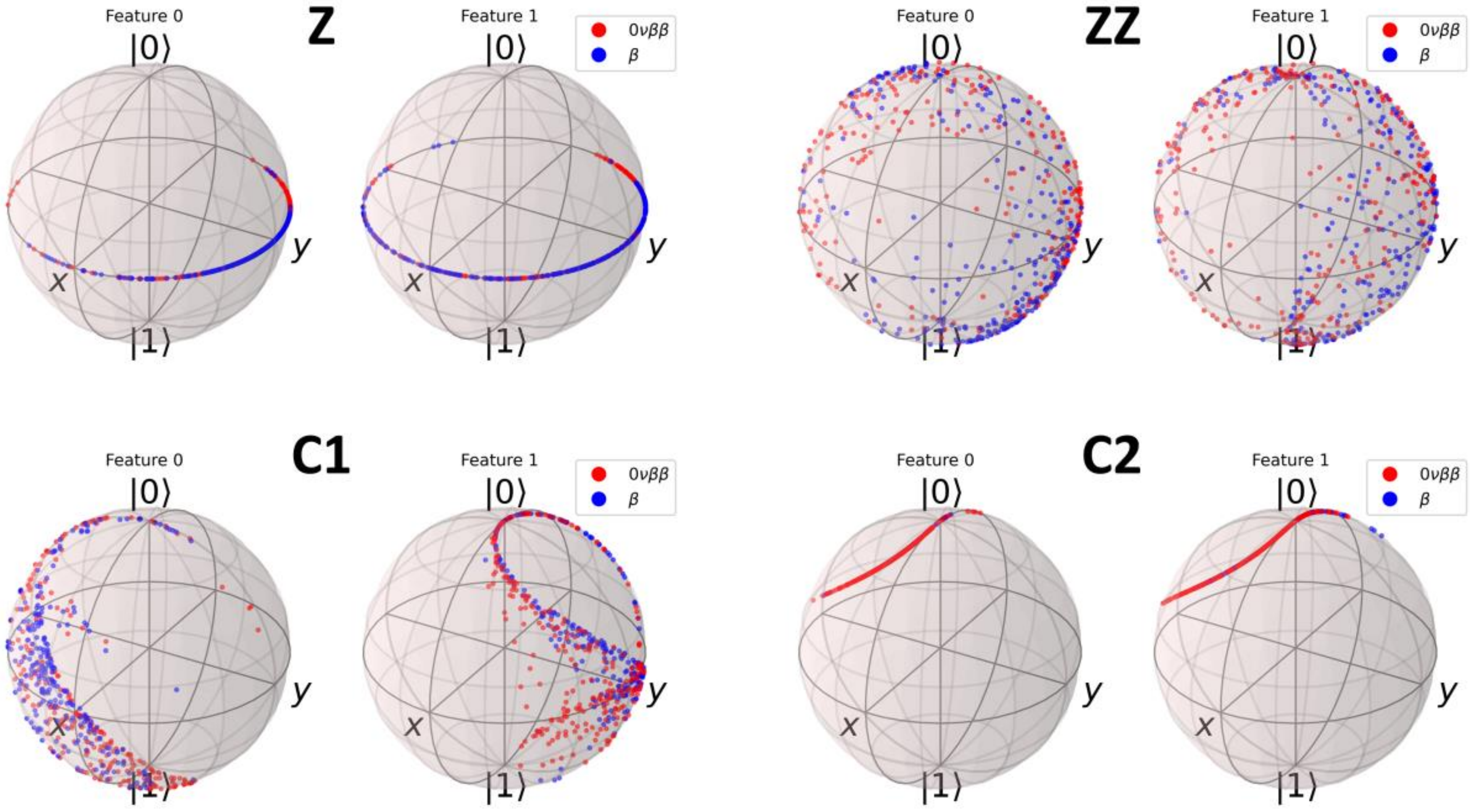




Bloch sphere visualization

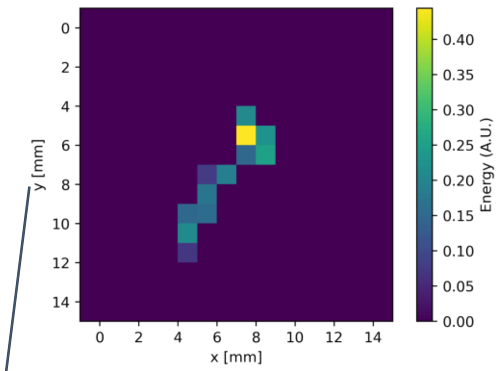
Quantum-encoded feature distribution of heuristic feature maps.

Bloch spheres allows to visualize the degree of class separation.

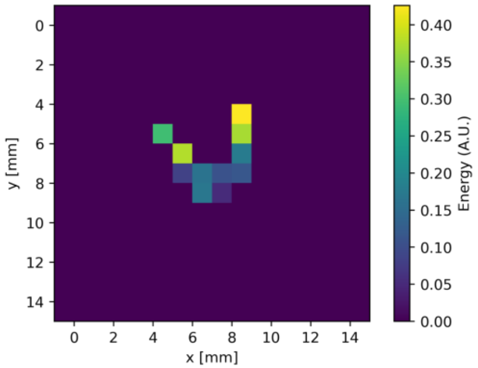


QSVM at $[1 \times 1 \times 1]\ \mathrm{mm}^3$

β

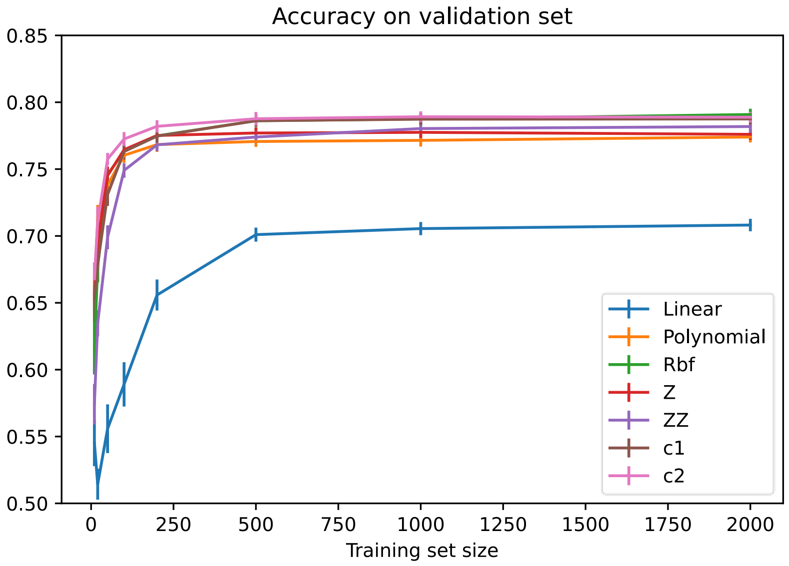


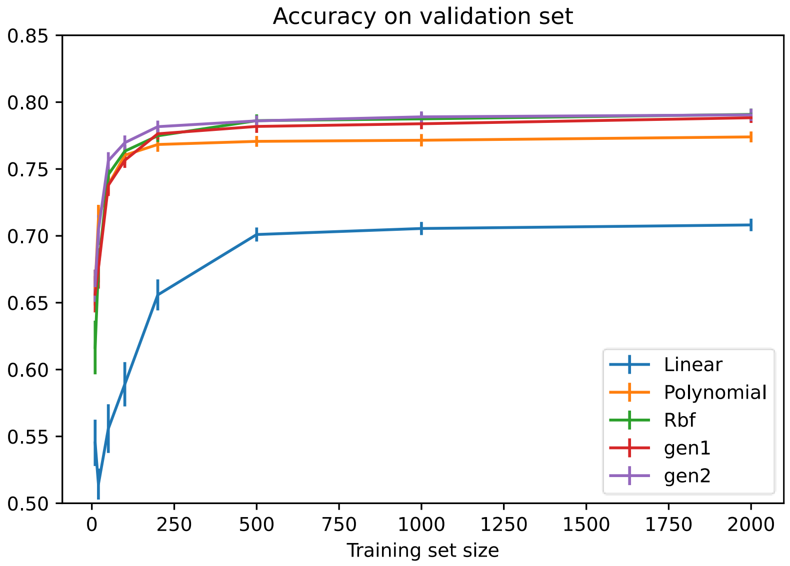
ββ



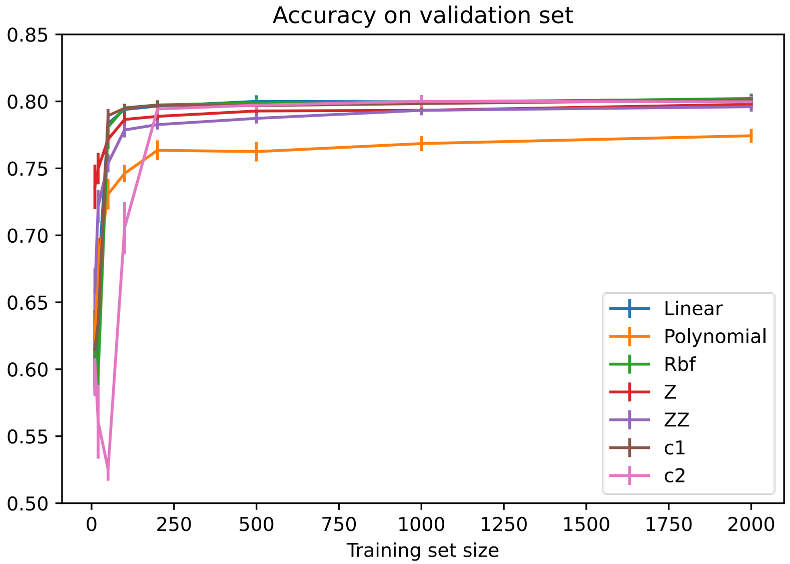
Topologies start being visible by the naked eye

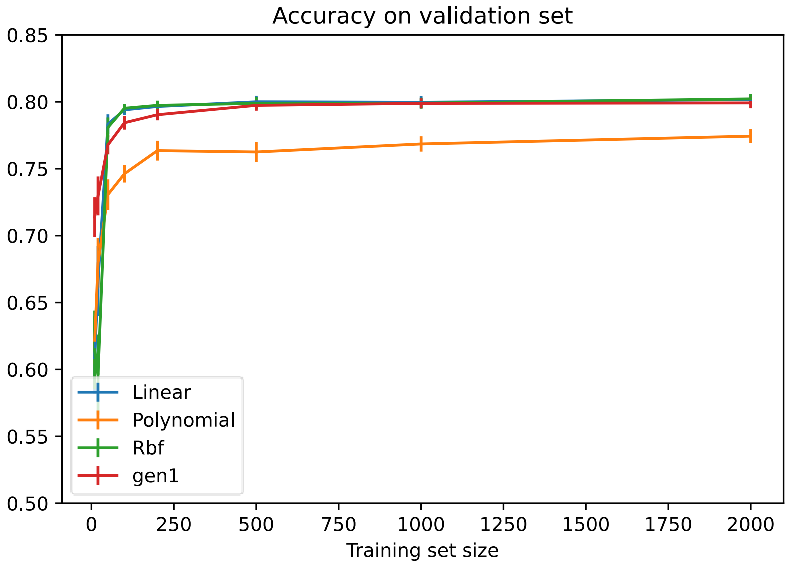
CNN
2 qubits
2 features



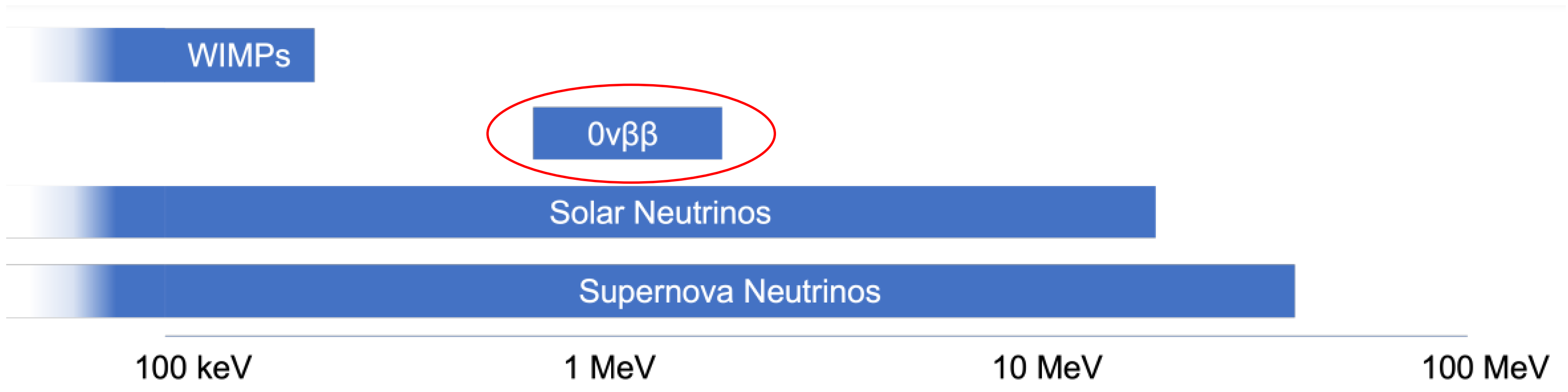


ATN
2 qubits
2 features

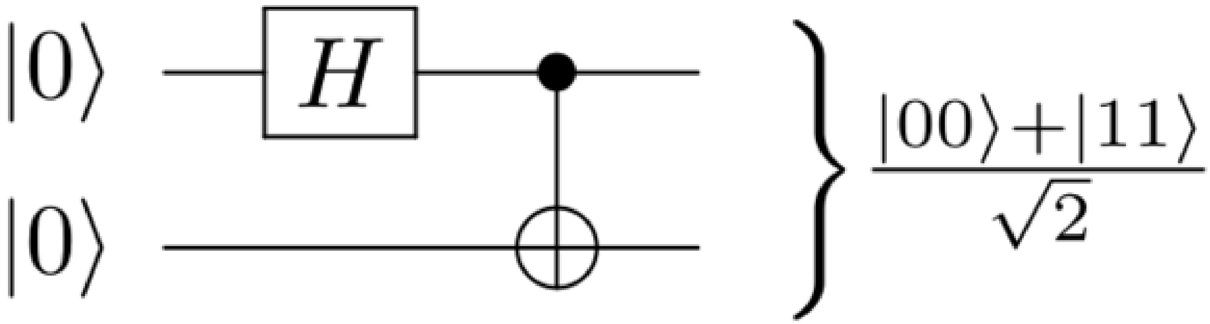




DUNE low-energy physics



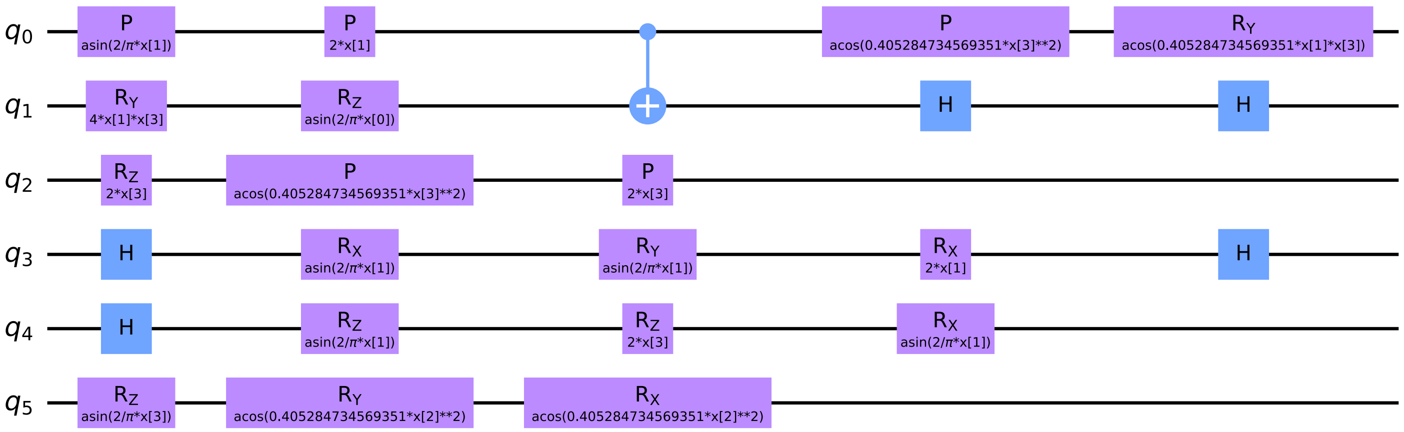
Bell's pair:

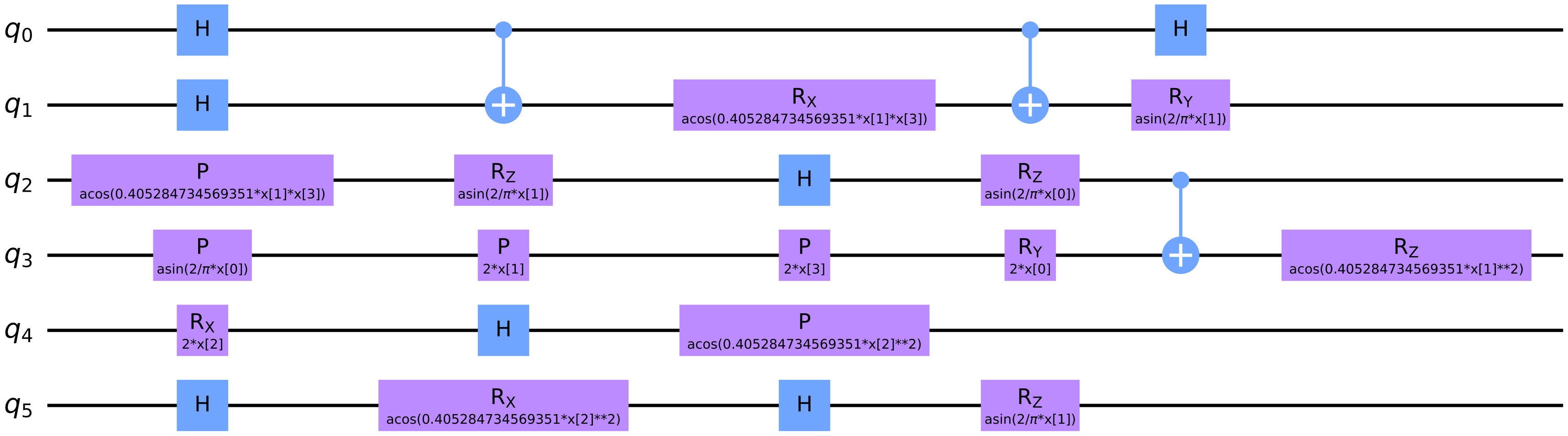


$$\mathrm{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

More genetic kernels

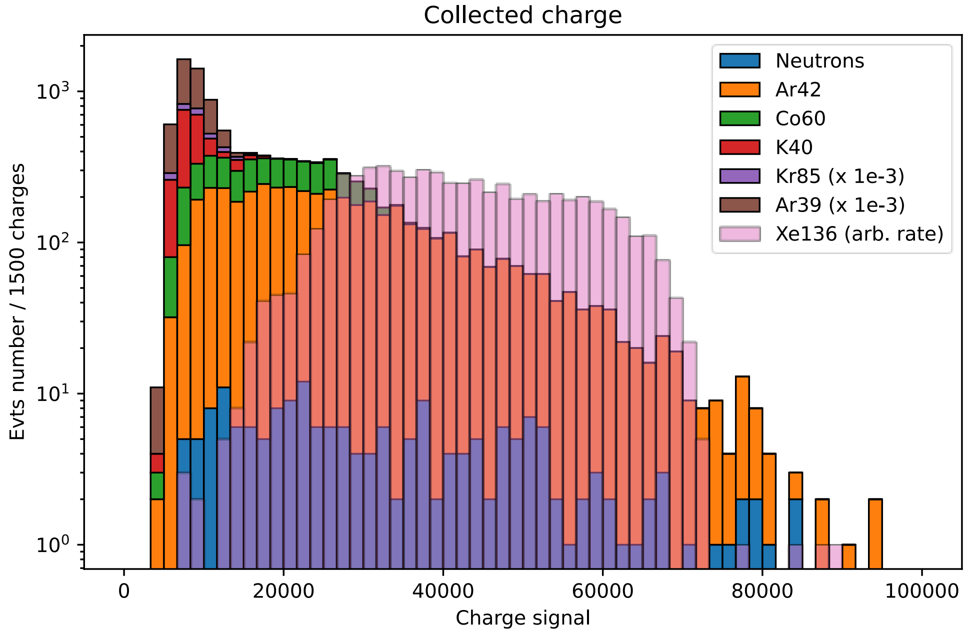
6 qubits – 4 features (2 from CNN, 2 from ATN):





DUNE FD Horizontal Drift dataset

Simulation of radiological samples and $0\nu\beta\beta$ events collected at the DUNE Far Detector Horizontal Drift module.
Dominant background: 42Ar (β topology).
Subdominant background: neutron capture.



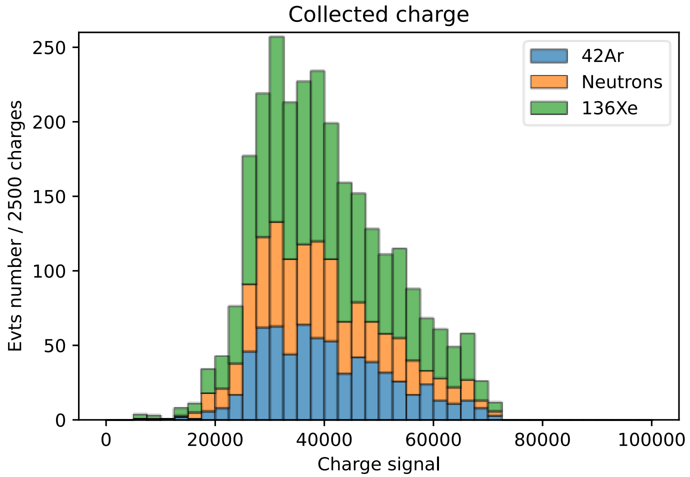
Charges signal at the LArTPC collection plane for radiological background and $0\nu\beta\beta$ events.

The CNN have been trained with a balanced dataset for classifying signal ($0\nu\beta\beta$) and background (β +neutrons).

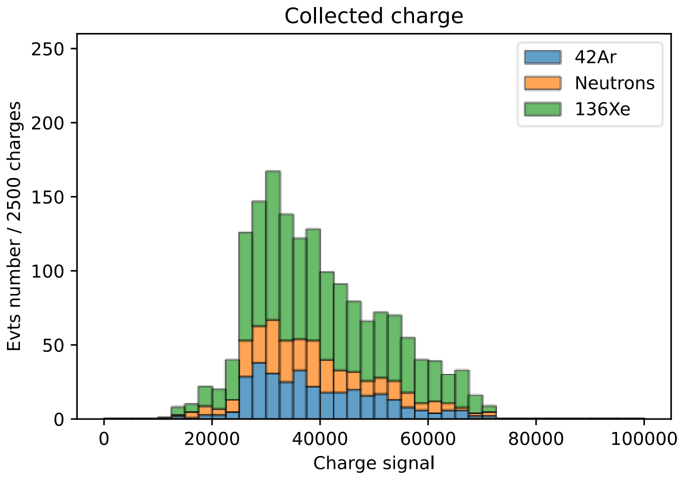
Accuracy: $\sim 63\%$

Corresponding to
Efficiency:$\sim 73\%$ Purity: $\sim 55\%$

Before CNN



After CNN



β and neutron background topologies

CNN has been trained to distinguish $0\nu\beta\beta$ from β and from neutrons separately.
Significantly higher performances in neutron rejection.

→Neutron topology is easier to identify than β.