# Backend-tailored quantum feature maps optimization through genetic algorithms in neutrino physics

**Roberto Moretti**
roberto.moretti@mib.infn.it
Milano Bicocca University
INFN – Milano Bicocca

*A. Giachero – UniMiB, INFN*
*M. Grossi – CERN QTI*
*V. Radescu – IBM Quantum*

*Second workshop on*
*Quantum Computing @ INFN*

# LArTPC potential at few-MeV

**LArTPC: Liquid Argon Time Projection Chamber**
**Broad physics reach:**

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

**Low-Energy sector**
- Supernova neutrinos
- Solar neutrinos
- WIMPs
- **$0\nu\beta\beta$**

*proposals*

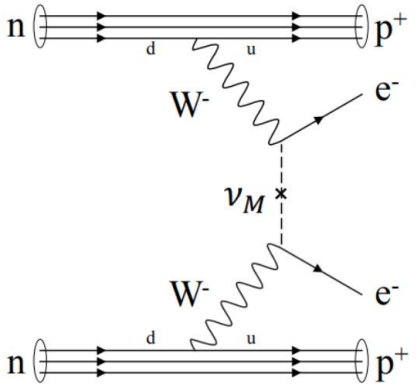
$$\Gamma^{0\nu}_{\beta\beta} = \frac{1}{T^{0\nu}_{\beta\beta}} = G|M^{0\nu}|^2 m^2_{\beta\beta}$$

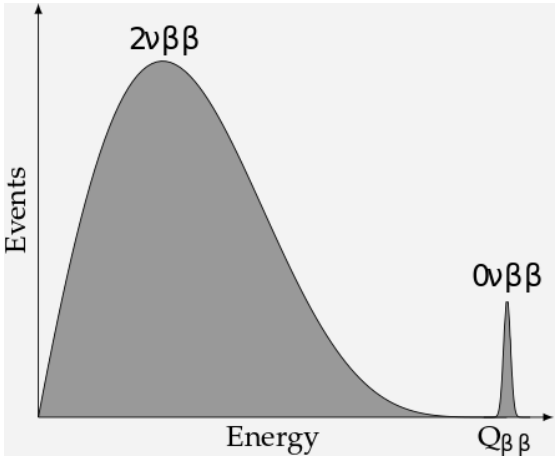
*phase space* — *nuclear matrix element*

*Majorana mass*

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

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




**Candidate:**

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

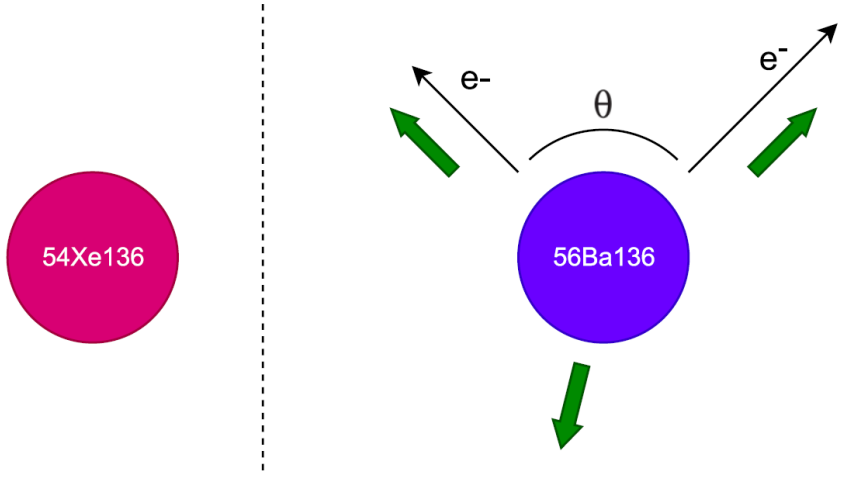
$$Q^{136Xe}_{\beta\beta} = 2.458\ MeV$$

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

# Topology-based classification

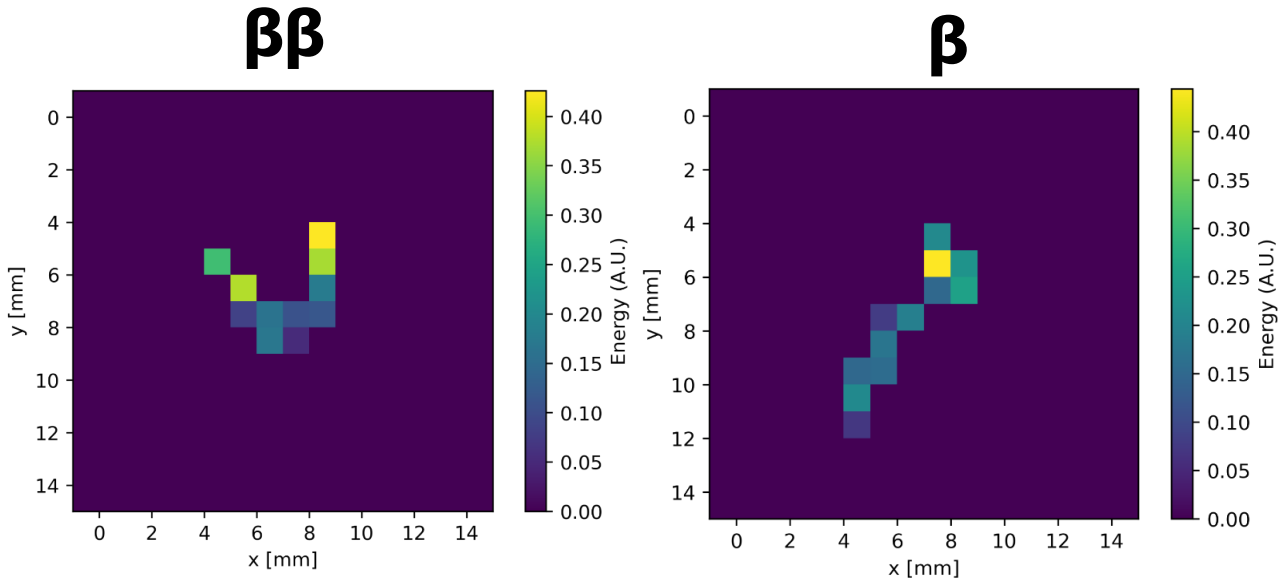
***ββ* topologies (signal):** two electrons originating from the same position in space.

***β* topology (background):** one electron with an energy close to $Q_{\beta\beta}^{^{136}Xe} = 2.458\ \mathrm{MeV}$




Energy-angle distribution for $\beta\beta$:

$$\frac{d\Gamma_{\beta\beta}^{0\nu}}{dE_1 d\cos\theta} \propto \frac{1}{16\pi^5} F(E_1, Z) F(E_2, Z) dE_1 d\cos\theta E_1 E_2 p_1 p_2 \left(1 - \frac{\vec{p}_1 \cdot \vec{p}_2}{E_1 E_2}\right)$$

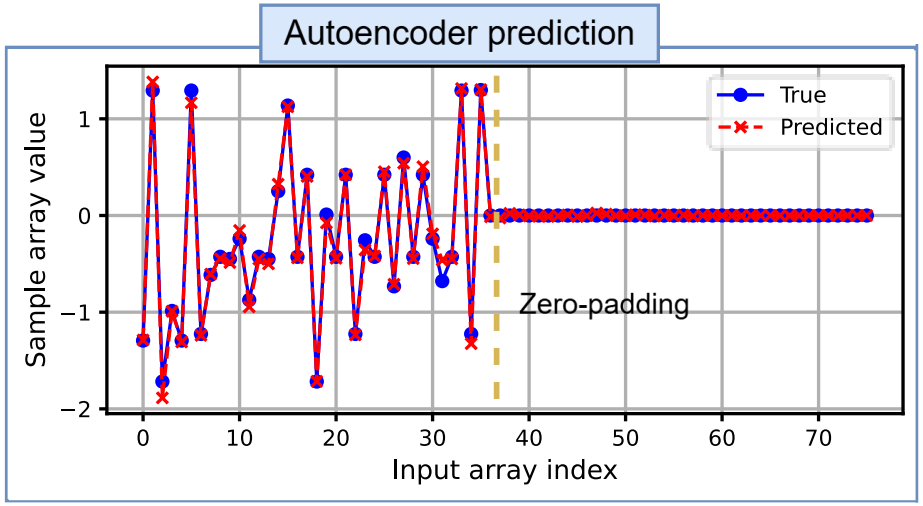


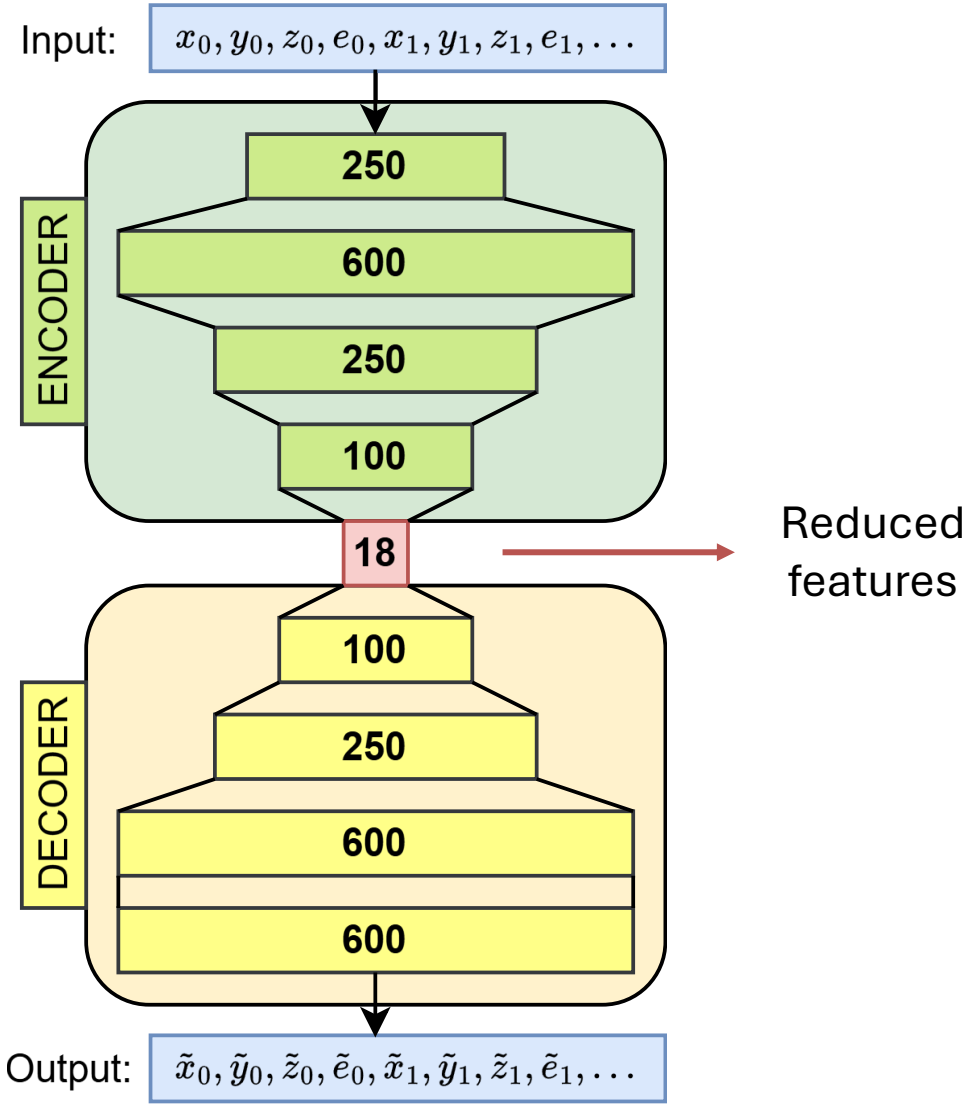
# Autoencoder feature extractor

We used a label-agnostic feature reduction algorithm, i.e. **autoencoder:**

- Stack of **feed-forward layers** divided into an Encoding and a Decoding part.
- Input and output should match as closely as possible.
- The hidden layer that produces the reduced feature distribution is called «Bottleneck».
- Training cost function minimizes the information loss (Mean/Absolute Square Error).

Reduction
$76 \rightarrow 18$
features



# Feature distributions

Features are not well-separated for $\beta/\beta\beta$ classes and not all of them have «gaussian-like» distributions.

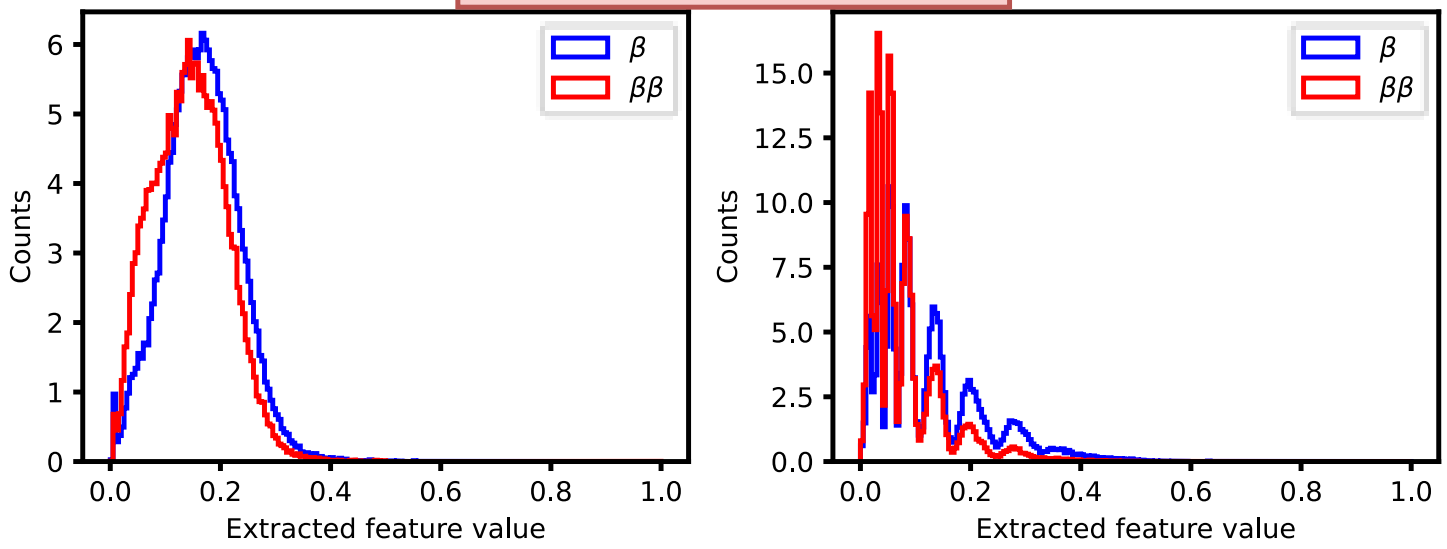
We expect low accuracy overall, but it is a viable testbed for testing quantum advantage.

Best accuracy on this dataset: **76%** with $> 10^5$ training samples.

*10.1140/epjp/s13360-024-05287-9*

*(R. Moretti et al, 2024).*

Latent space features **(Examples)**




- Some features are multi-modal.
- Big overlap between distributions.

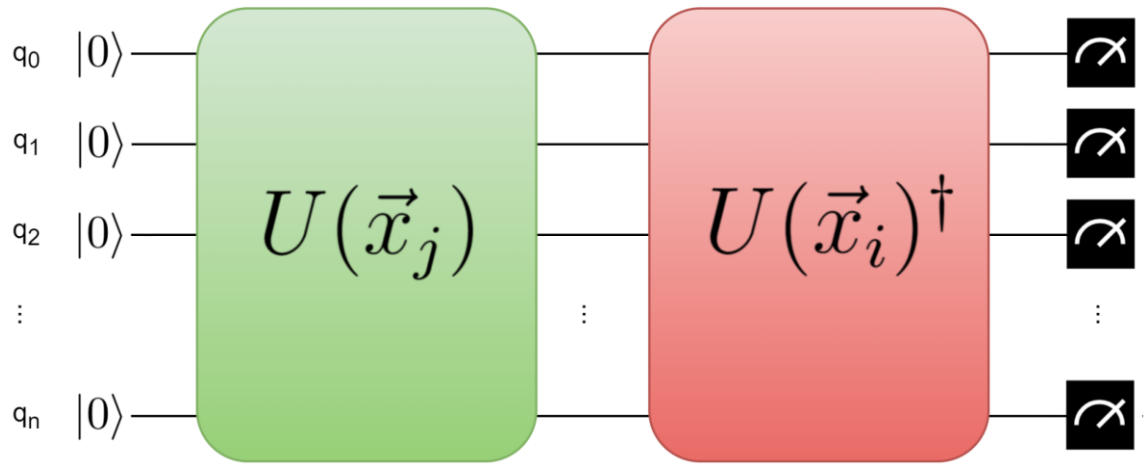
# Quantum Support Vector Machine

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.
- The ernel function is 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.**

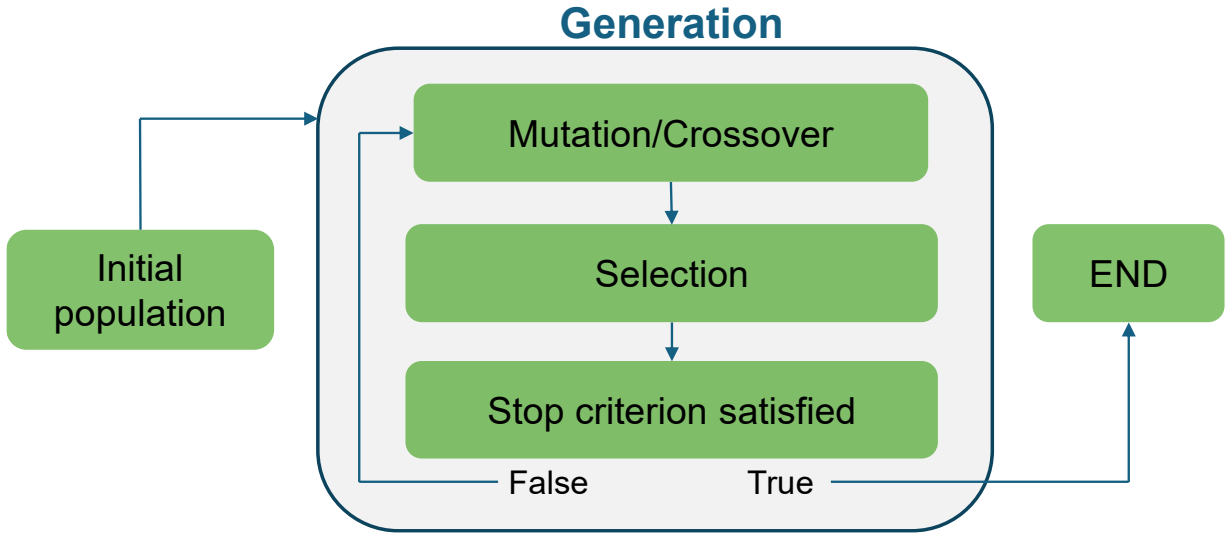
# Genetic algorithm for quantum kernels

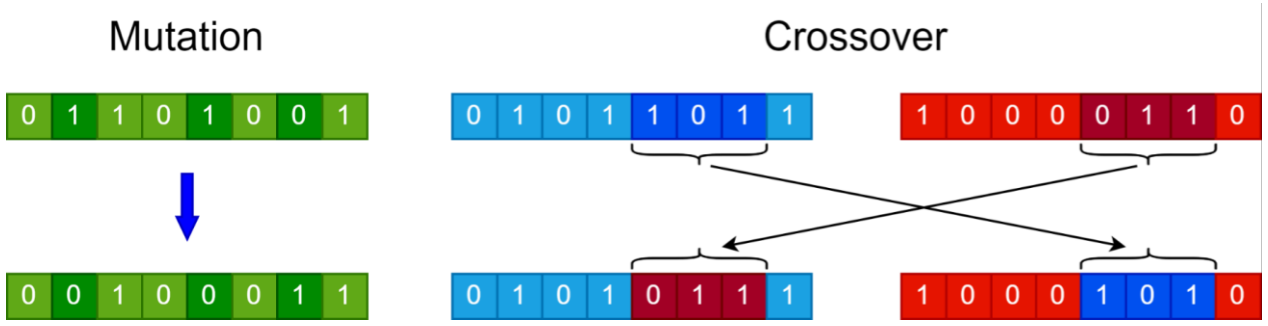
**Meta-heuristic approach – Genetic algorithm**

- *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 and **hardware** in use.

- Identifying useful metrics.
- Exploiting hardware knowledge.
- Fitness function engineering.
- Efficient gene encoding.

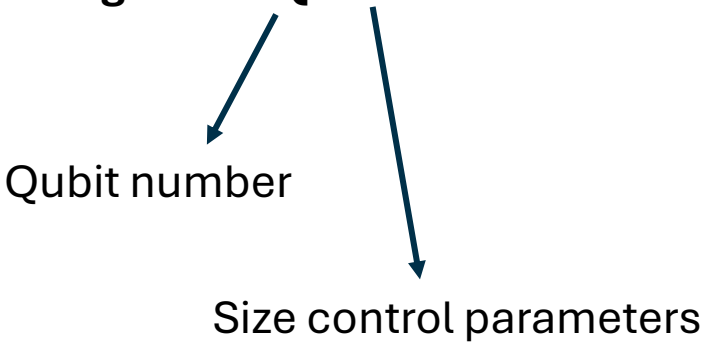


# Featuremap automatization

6 integer genes encoding the gate type and rotation angle.

**Tot genes = Q x S x 6**




**Q**: From 4 to 12 qubits tested.
**S**: 8-10.
Gates uniformly distributed on all qubits.

| | Min | Max | Description |
|---|---|---|---|
| **Gate-type** | 0 | custom | I, H, X, RX, RY, RZ, CX, S, CRX, … |
| **Feature index** | 0 | custom | Feature to use as a primary gate argument. |
| **Second feature index** | 0 | custom | Feature to use as a secondary argument. |
| **Featuremap type** | 0 | 2 | Linear, quadratic, trigonometric, … |
| **Multi-feature per gate** | 0 | 1 | Use one or two features in a gate rotation angle. |
| **Target qubit index** | 0 | custom | Target qubit index. |

| | 0 – single feature | 1 – two features |
|---|---|---|
| **0 – «linear»** | $2\pi(x_i - 0.5)$ | $2\pi x_i(1 - x_j) - \pi$ |
| **1 – «square»** | $2\pi(x_i - 0.5)(1 - \mathrm{x_i}) - \pi$ | $\big(2(x_i - 0.5)(1 - x_j) - 1\big) \times (2\pi(x_j - 0.5)(1 - x_i) - \pi)$ |
| **2 – «trig»** | $2\arcsin(2x_i - 1) - \pi$ | $2\arcsin\big((2x_i - 1)(2x_j - 1)\big) - \pi$ |

# Fitness function engineering

- **Mono-objective:**

$$f = a + \eta\sigma$$

Fitness function maximization.
Tuning of $\eta$ required.

- **Multi-objective:**

$$f = \begin{cases} a \\ -d \end{cases}$$

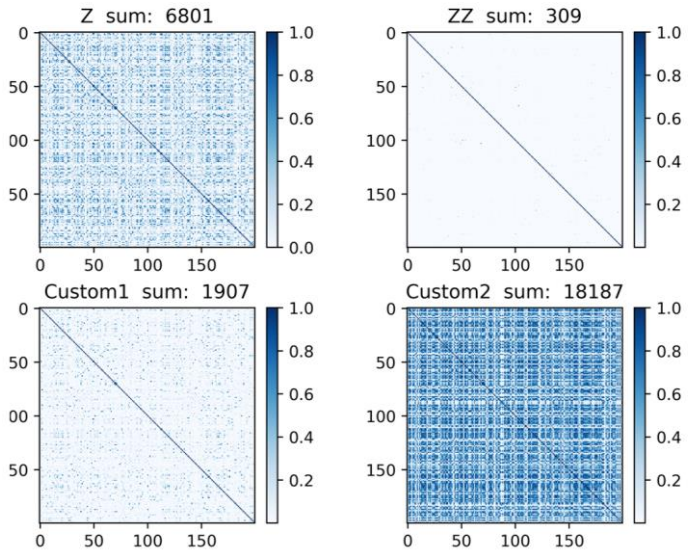
Pareto optimization.

**Metrics:**

- Classification accuracy $a$.
- Transpiled depth $d$.
- Off-diagonal kernel matrix standard deviation $\sigma$.

Quantum kernels matrices are prone to **exponential concentration – harder to estimate on real backends.**

$\sigma$ grows with qubit number and $d$, but ultimately depends on the individual feature map.



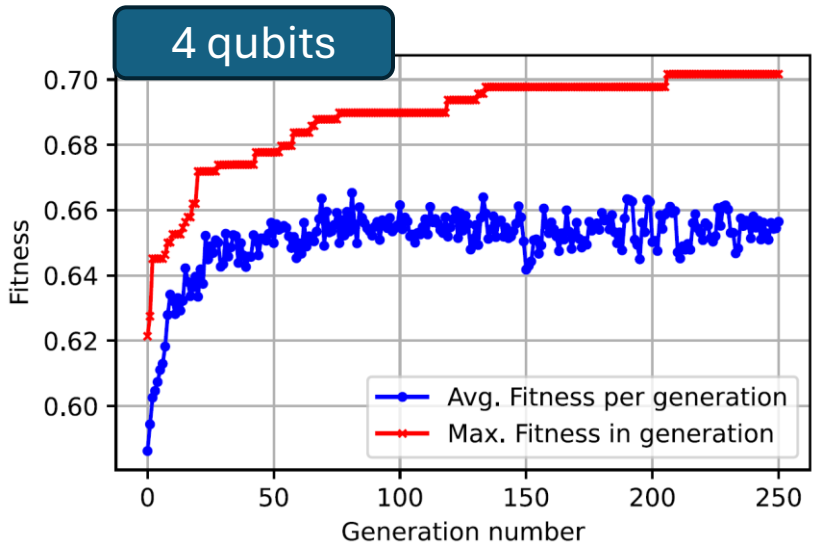
# Mono-objective genetic run

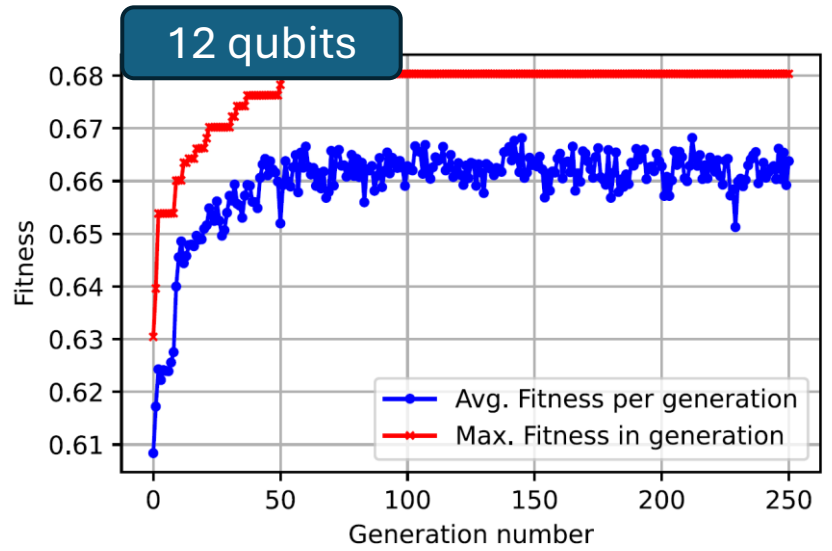
**250** feature map generations
Each generation contains **50** feature maps.

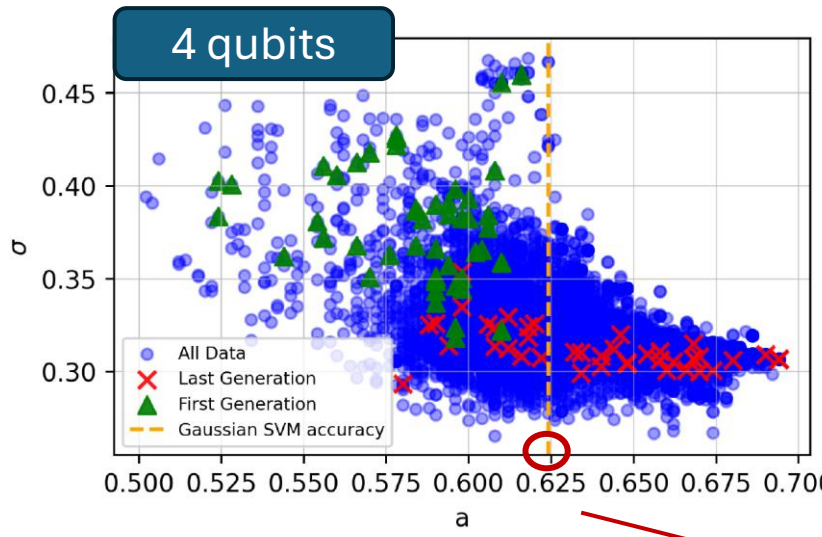
Elitism strategy preventing the loss of the 5 best kernels in each generation.

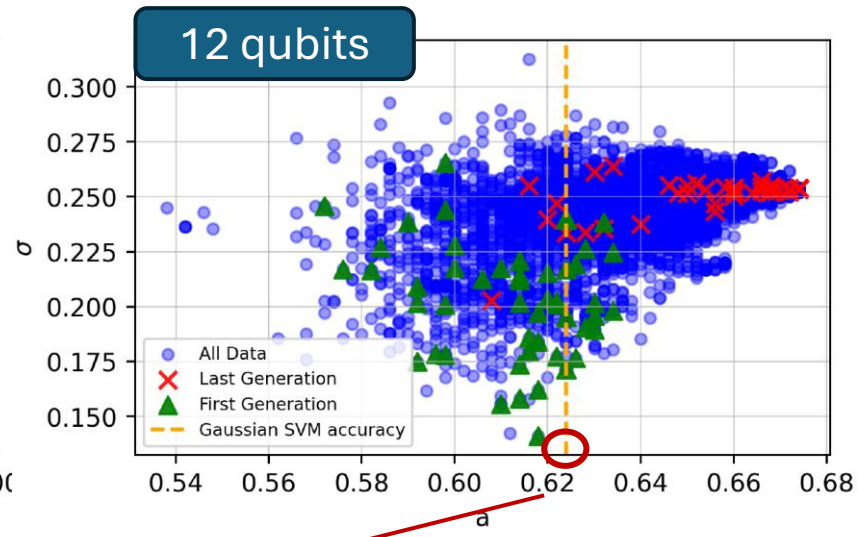
$$\textbf{fitness} = a + \eta\sigma$$
$$\eta = 0.25$$

Total kernels tested: **11300**
Dataset size: **500** samples.




**Best classical SVM result**

# Multi-objective genetic run

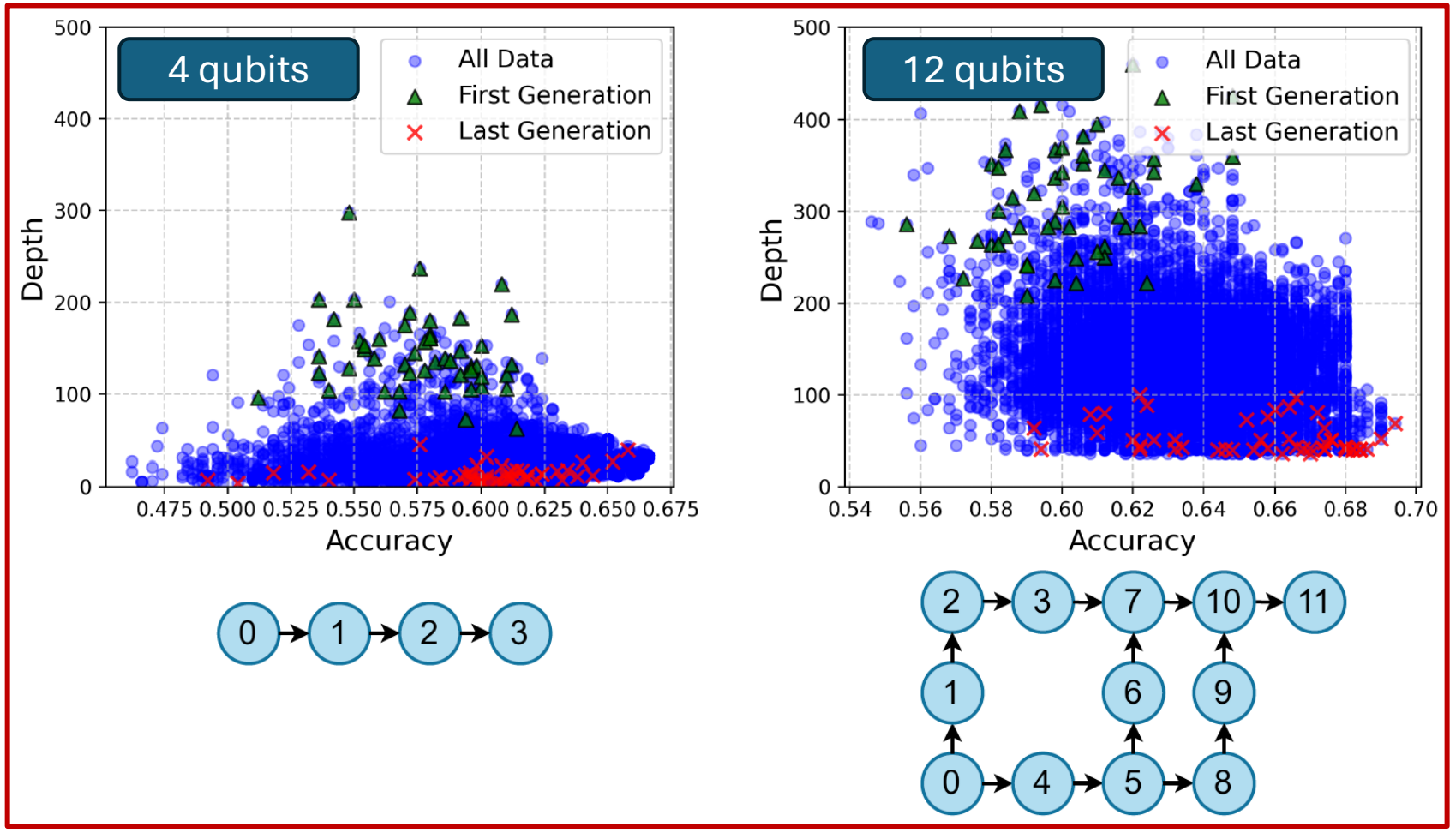
**250** feature map generations
Each generation contains **50** feature maps.

Allowed gates: *I, H, X, SX, RX, RY, RZ, CX, ECR CRX, CRY, CRZ*.

Basis gates: *I, X, SX, RZ, ECR*.

$$\textbf{fitness} = \begin{cases} \boldsymbol{a} \\ -\boldsymbol{d} \end{cases}$$

Total kernels tested: **11300**
Dataset size: **500** samples.



# Speedup through backend parallelization

**21 disjoint 4-qubit** sites that are equivalent in term of topology and entanglement.
Kernel evaluated simultaneusly within the same large quantum circuit.
→ parallelization attempt for a ~ **20 fold speedup**.

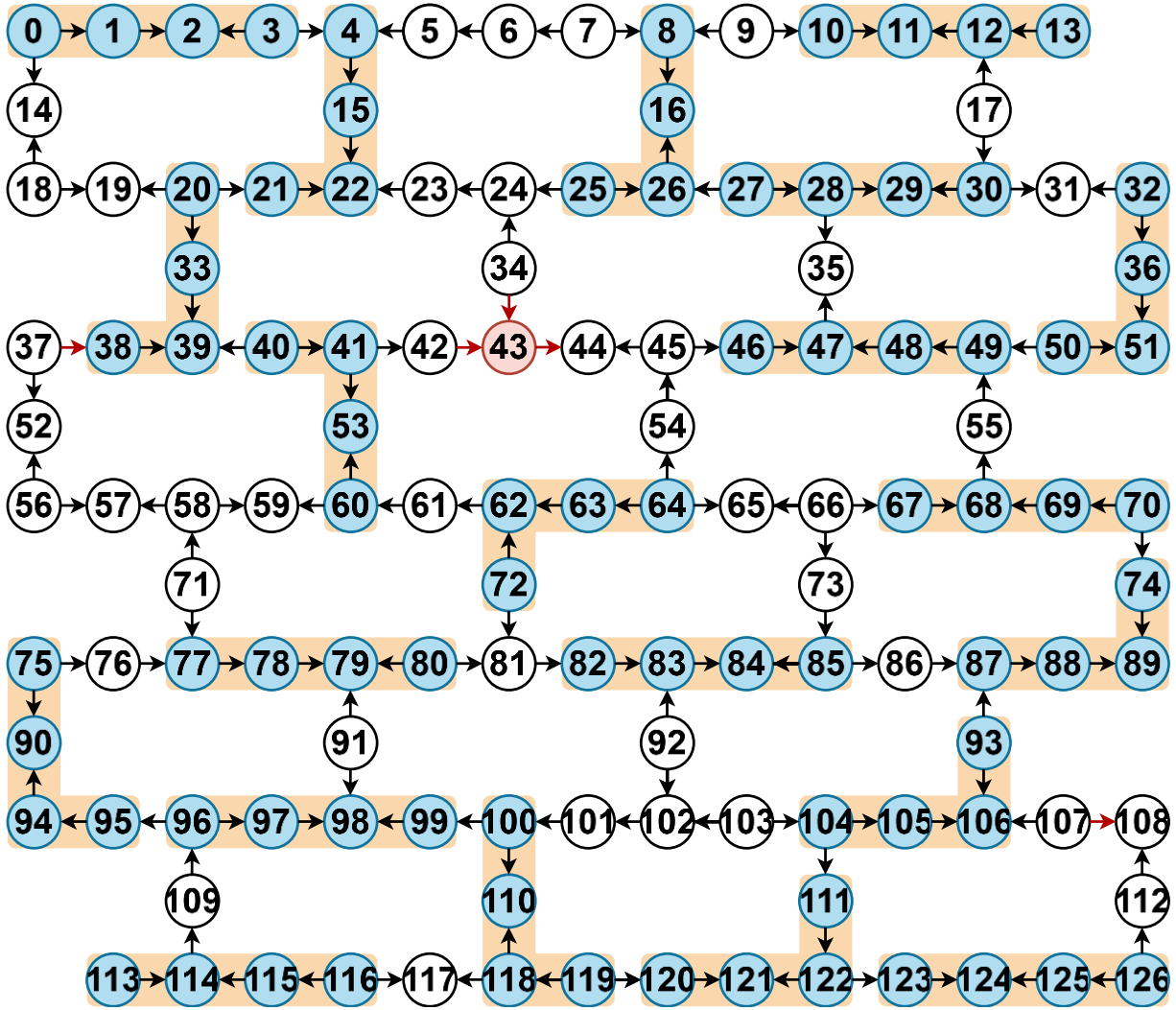
3 backends tested:

- IBM Nazca (127 QB).
- IBM Brisbane (127 QB).
- IBM Torino (133 QB).

IBM Nazca and IBM Brisbane have the same basis gates: I, RZ, SX, X, **ECR**.
All partitions were selected to have the same ECR directionality:

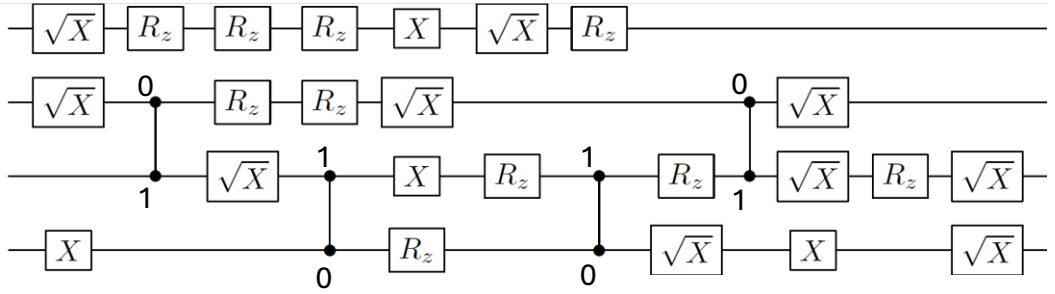
$$\mathrm{qb0} \rightarrow \mathrm{qb1} \rightarrow \mathrm{qb2} \leftarrow \mathrm{qb3}$$

IBM Torino: I, RZ, SX, X, **CZ**.

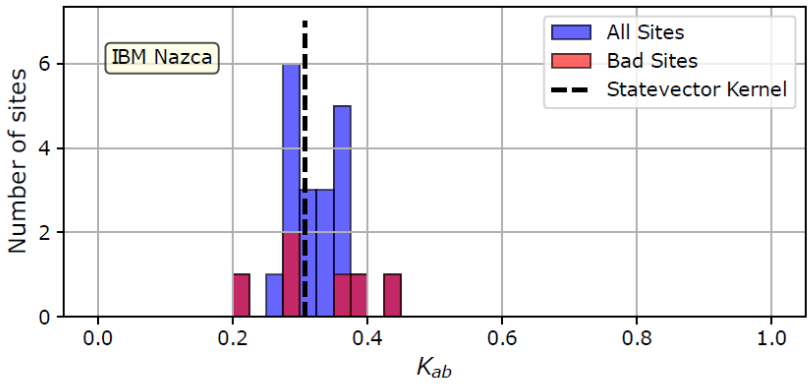


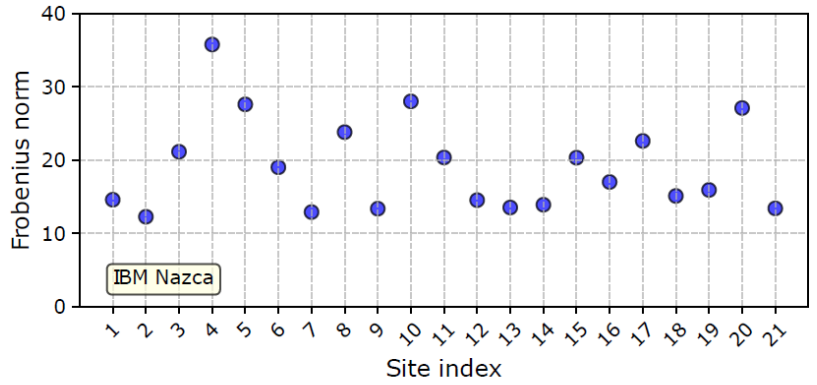
# Speedup through backend parallelization

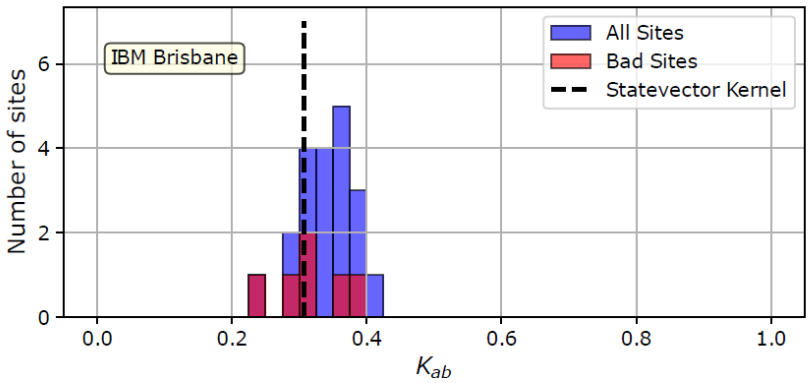
We can retrieve the kernel entries for each site and estimate the **output spread**.

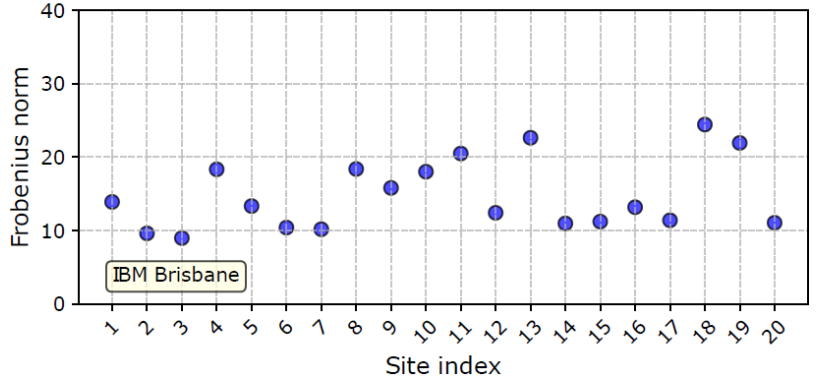


It turns out that **some sites are less performant than other**. We can discard them based on how much they differ from the statevector matrix.



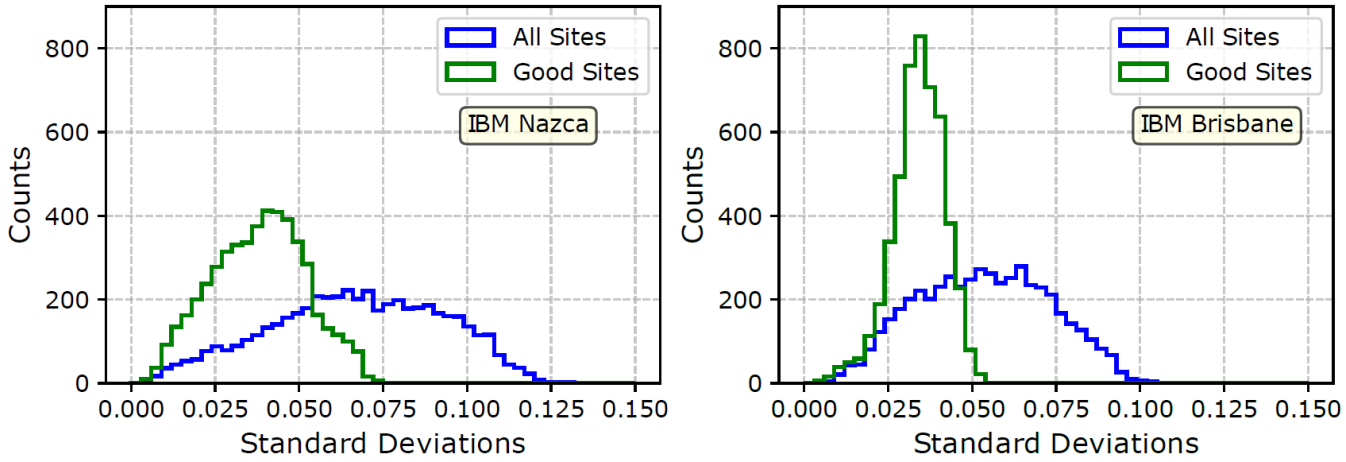



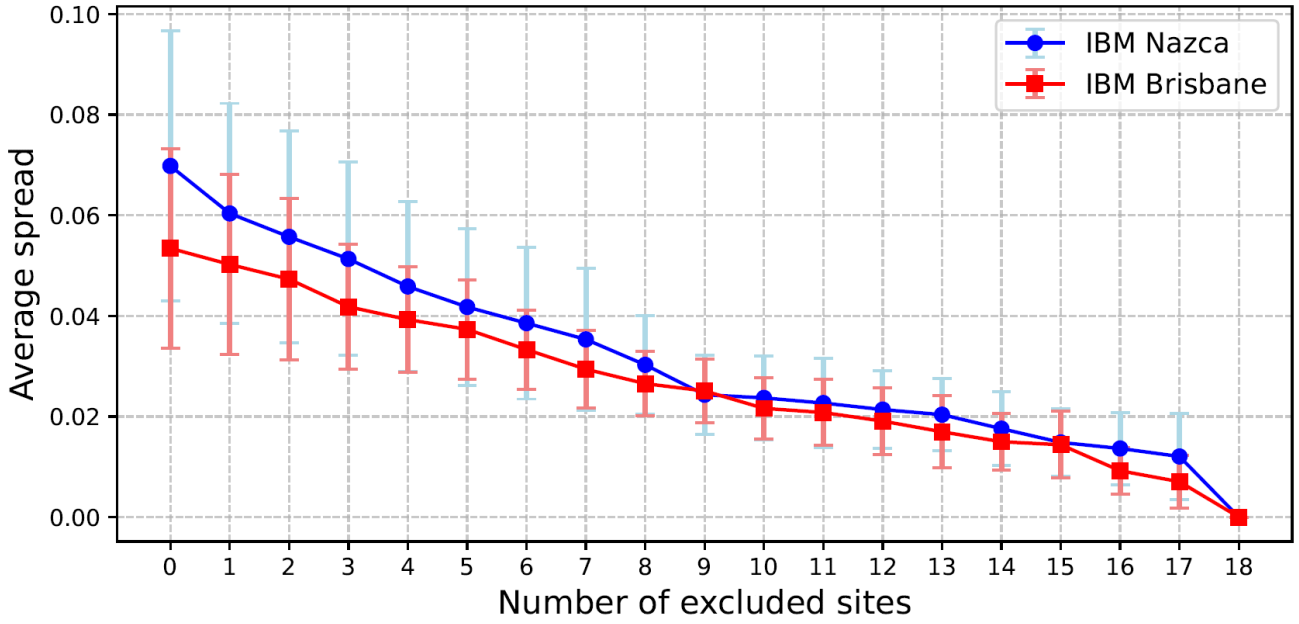



# Output spread distribution

Direct comparison of IBM Nazca and IBM Brisbane site spreads.



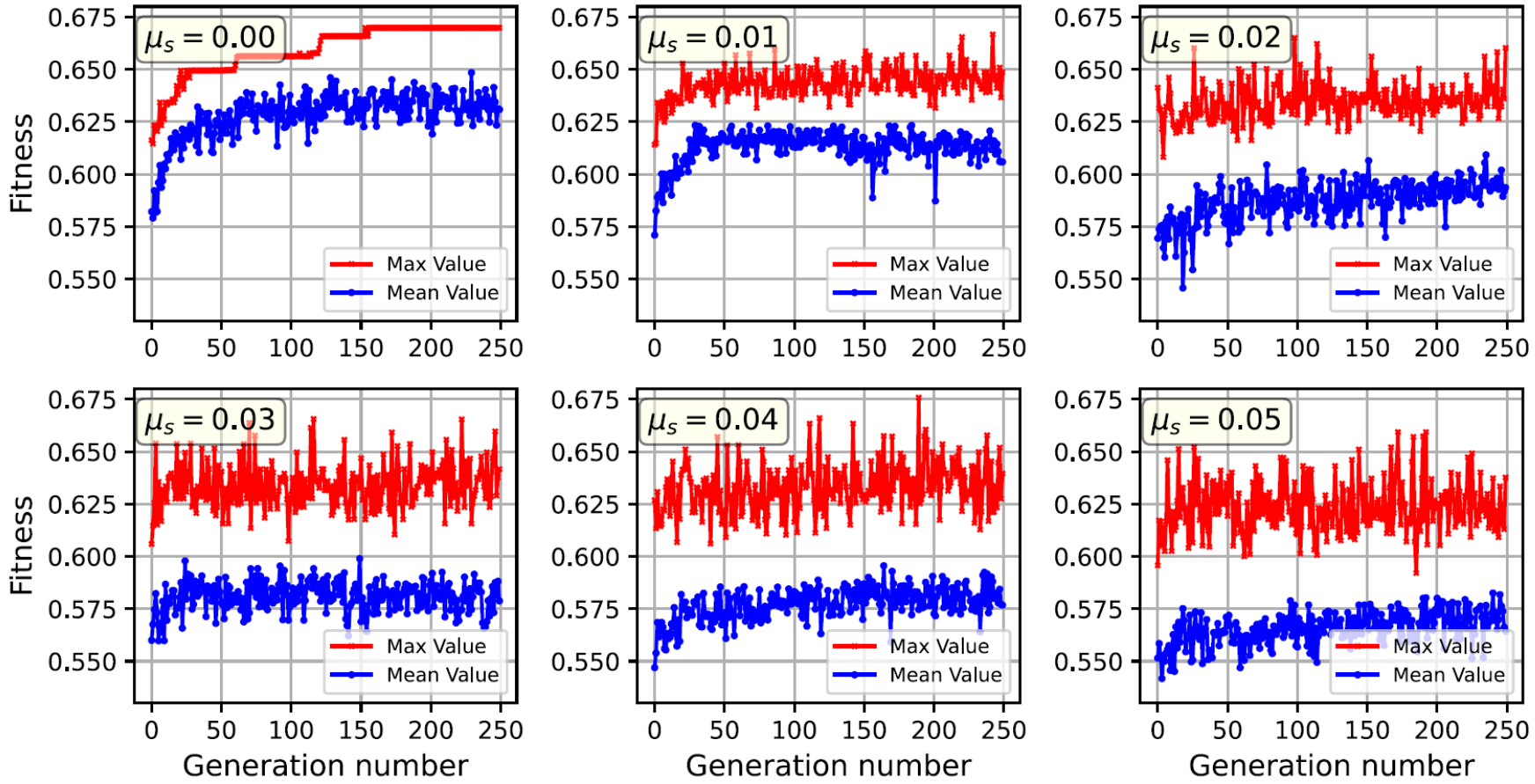

- The kernel matrix off-diagonal std (statevector) is **0.241.**
- When discarding enough sites, we can lower the noise spread to $\boldsymbol{\mu_s < 0.04}$.
- With real hardware we can «resolve», on average, one kernel entry from another.



# Data-driven spread effect on genetic runs

We simulated a genetic search with gaussian noise on the kernel matrix entries with increasing spread.
Up to $\boldsymbol{\mu_s = 0.02}$, the genetic optimization succeeds exhibiting a positive trend throughout generations.

→ The **best $20\%$** partitions could be used to carry out genetic searches for this classification task.



# Conclusions

## Neutrino Physics

- **Modest $\beta\beta$ vs $\beta$ classification accuracy overall** ($\sim 65\%$) for an ideal $2 \times 2\ \mathrm{mm}^2$ pixel-size LArTPC.
- **Detector changes are mandatory** to carry out neutrinoless double beta decay searches.

## Automatic optimization of quantum feature maps

- QSVMs have been thoroughly analysed for a small training set (500 samples) and compared with classical counterparts.
- Although Quantum Advantage can't be claimed, simulated, **genetic-optimized QSVMs exhibit promising performances.**
- **Parallelization attempts have been carried out** for three quantum backends with $> 100$ qubits.
- Commercially available NISQ hardware is close to be parallelizable for the application under consideration.
- More advanced partitioning schemes, transpiling, and error mitigation techniques could provide further benefit.

**For more information:** *"Enhancing quantum feature maps through quantum hardware parallelization"*, **in preparation.**