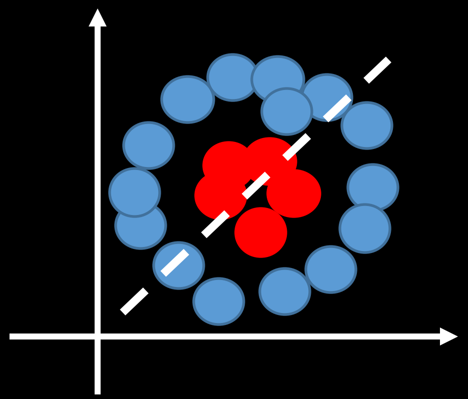
Automatic and effective discovery of quantum kernels

Michele Grossi, PhD (CERN QTI)

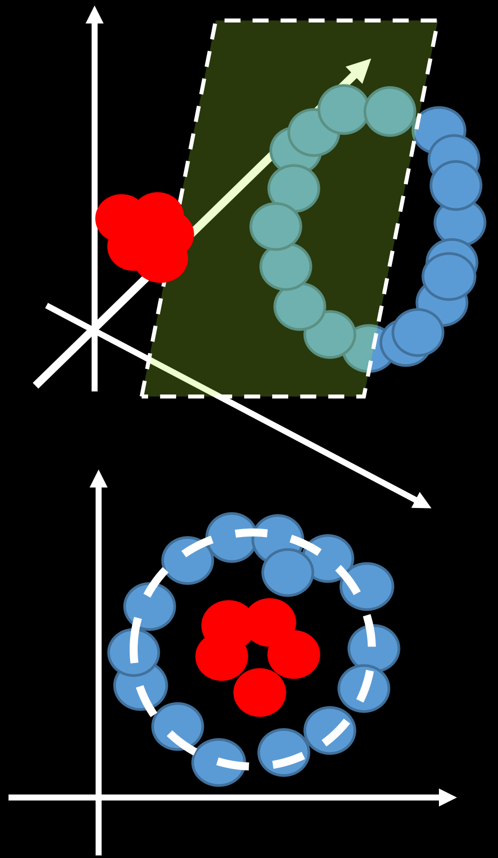
Hybrid Quantum Computing Infrastructures and Algorithms Coordinator

*Incudini M, Lizzio Bosco D, Martini F, **MG**, Serra G, Di Pierro A*
'Automatic and effective discovery of quantum kernels'.
Accepted in IEEE Transactions on Emerging Topics in Computational Intelligence (2024)

KERNEL METHODS



EMBEDDING
(FEATURE MAP)



Generalization of the conventional Euclidean product by using as a kernel any arbitrary positive **semi-definite function**.

KERNEL METHODS

- Many different kernel exists, and they all can be written in the form $\kappa(x, x') = \langle \phi(x), \phi(x') \rangle$

- Different kernels adapt to different assumptions on the data
  - $\kappa(x, x') = (\sum_i x_i x_i' + c)^d$ for polynomial dependencies
  - $\kappa(x, x') = \exp(-\|x - x'\|^2/\sigma)$ for linearly separable data
  - $\kappa(x, x') = \exp(-d(x, x')/\sigma)$ if you have the distance function
  - $\kappa(G, G') = \exp(-\mathrm{GED}(G, G')/\sigma)$ for graphs
  - $\kappa(x, x') = \int_\gamma \nabla_\theta f(x; \theta) \cdot \nabla_\theta f(x'; \theta)\, d\theta$ (path kernel)

M. Incudini, M. Grossi, et al. "The Quantum Path Kernel: a Generalized Quantum Neural Tangent Kernel for Deep Quantum Machine Learning." IEEE Transactions on Quantum Engineering - 10.1109/TQE.2023.3287736}

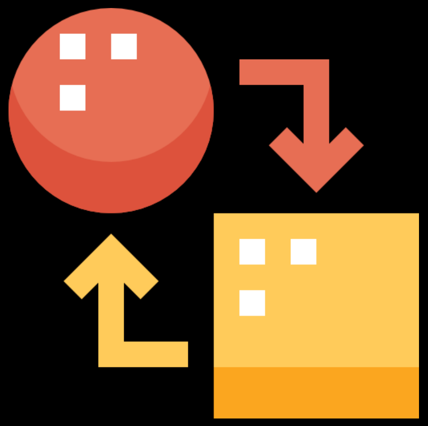
PROPERTIES OF KERNELS

- If κ is positive definite, then it is a kernel (and $\exists \phi. \kappa(x, x') = \phi(x)\phi(x')$)

- Kernel trick: you don't need to know ϕ to use a kernel

- Each kernel admit spectral decomposition $\kappa(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x)\phi_i(x')$ due to Mercer's theorem with λ_i real, non-negative
- Flexible to perform supervised (regression, classification) and unsupervised (dimensionality reduction, clustering) on more exotic data structures than real vectors.

QUANTUM KERNELS

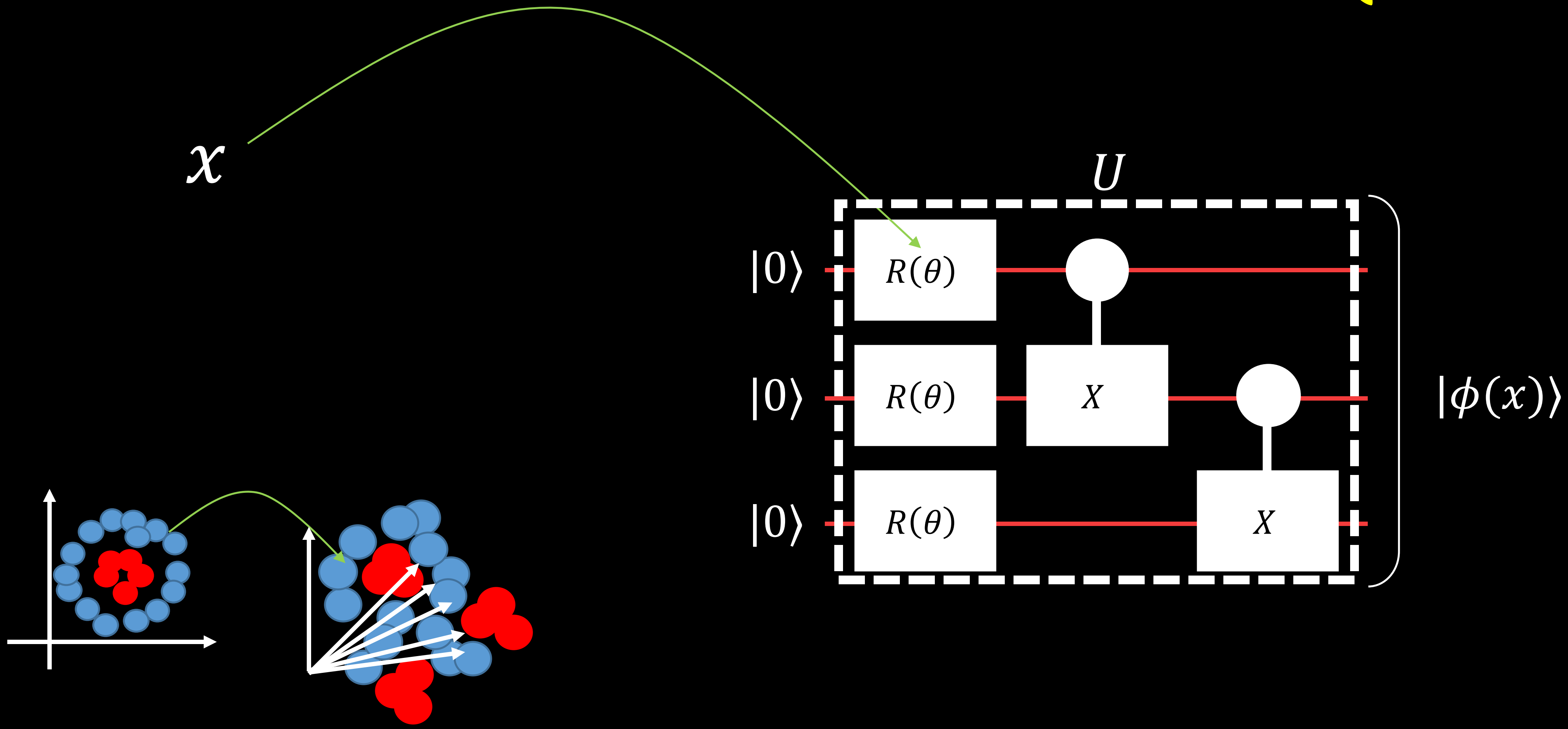


- Common aspect: Hilbert space
- Common aspect: efficient calculation of the inner product

Havlíček et al. "Supervised learning with quantum-enhanced feature spaces." *Nature* 567.7747 (2019).
Schuld and Killoran. "Quantum machine learning in feature Hilbert spaces." *Physical review letters* 122.4 (2019).

QUANTUM KERNELS

FTQC vs NISQ



General Problem

- We usually do not know much about the nature of the task
- We usually do not know whenever there exists a quantum feature map solving the task any better than some classical feature map

General Question

Is it possible to design an ansatz that is tailored for the task at hand, has favorable statistical properties, and retains a computational advantage over classical devices?

AUTOMATIC CONSTRUCTION OF QUANTUM KERNELS

MOTIVATION

$$|\phi(x)\rangle = \exp\left(-i\sum_{\ell} f_\ell(x) \cdot \sigma_{\ell,1} \otimes \cdots \otimes \sigma_{\ell,n}\right)|0\rangle$$

angles
(scalar values)

generators

MOTIVATION

HOW CAN I CHOOSE THESE ONES?

$$|\phi(x)\rangle = \exp\left(-i \sum_{\ell} f_{\ell}(x) \cdot \sigma_{\ell,1} \otimes \cdots \otimes \sigma_{\ell,n}\right)|0\rangle$$

angles
(scalar values)

generators

OPEN QUESTION

What are the characteristics of a quantum kernel?

and… how can we construct a quantum kernel practically?

DESIDERATA

- **AUTOMATIC PROCEDURE**
- **CHOOSE** the **ALL THE COMPONENTS**
- As **FAST** as possible

INGREDIENTS

- A model for a quantum kernel
- One or more criteria to evaluate the quality of quantum kernels
- One or more optimization techniques to adaptively modify the quantum kernel

MODEL

- QK is represented as a *combinatorial* object
- n-qubits, m-gates QK as a *discrete* object
- vector encodes the QK over n-qubits, with a parameterized quantum circuit containing m gates
- a family of criteria identified to assess properties such as **expressivity**, efficiency of **classical simulability**, or **compatibility** with the task at hand.
- *heuristic algorithm* is employed to iteratively explore the space of quantum kernels → optimization

$$\boldsymbol{k}^* = \arg \min_{\boldsymbol{k} \in \mathcal{QK}_{n,m}} C(\boldsymbol{k})$$

NB: nonconvexity of the cost function → no guarantee of finding an optimal solution

CRITERIA I: CLASSICAL SIMULABILITY

- The unitary $U(x)$ should not be classically simulable
  - No need for a quantum computer!

- Avoid quantum circuit without entangling gates

- Check the dimensionality of the dynamical Lie algebra of your circuit
  - $U = \exp(-i\theta_k H_k) \cdot \ldots \cdot \exp(-i\theta_1 H_1)$
  - $d = \dim \mathrm{span}\{iH_1, \ldots, iH_k\} \in \exp(\#\ \mathrm{qubits})$

Somma et al. "Efficient solvability of Hamiltonians and limits on the power of some quantum computational models." Physical review letters 97.19 (2006): 190501.

CRITERIA II: EXPRESSIBILITY

- “Expressible” circuits are able to define an ensemble of states $\{U(\theta)|0\rangle \mid \theta \in \Theta\}$ that resemble the “uniform” distribution $\{U_{Haar}|0\rangle \mid U_{Haar} \text{ is Haar random}\}$.

- $A = \int_{Haar} |\phi\rangle\langle\phi|^{\otimes t}\, d\phi - \int_{\Theta} |\psi(\theta)\rangle\langle\psi(\theta)|\, d\theta$

- The dimension of the dynamical Lie algebra is an approximate measure of the expressibility

Larocca et al. "Diagnosing barren plateaus with tools from quantum optimal control." Quantum 6 (2022): 824.

CRITERIA II: EXPRESSIBILITY

- There is a compromise between expressibility and practical applicability

- Expressible circuit exploit the full, exponentially sized Hilbert space and that lead to “complicated” kernels
  - Unfavorable eigenvalue distributions
    - All eigenvalues exponentially small, needing an exponential amount of data to learn the corresponding component
  - Coefficients concentrated around some value
    - Need exponential amount of measurement to distinguish the coefficients

Kübler, Buchholz, and Schölkopf. "The inductive bias of quantum kernels." Advances in Neural Information Processing Systems 34 (2021): 12661-12673. // Thanasilp, Supanut, et al. "Exponential concentration and untrainability in quantum kernel methods." arXiv preprint arXiv:2208.11060 (2022).

EVALUATION I

- $\mathrm{KernelEvaluator}(K, \kappa, X, y)$
  - Target-Kernel Alignment
  - Centered Target-Kernel Alignment
  - Task-Model Alignment
  - Lie Rank evaluator
  - Covering numbers evaluator
  - Haar evaluator
  - Geometric difference

Does the kernel function fit a specific task?

Strong correlation to generalization error!

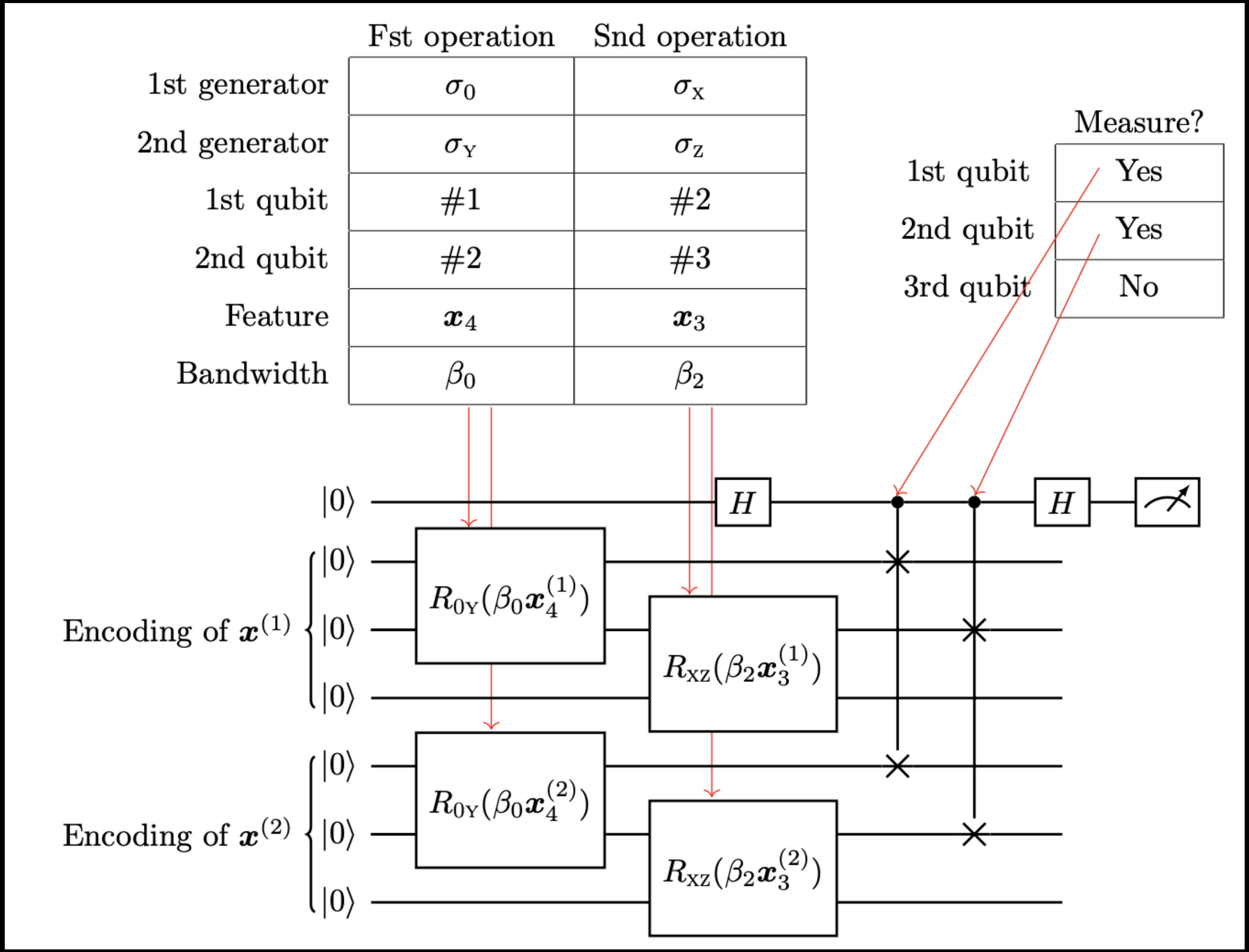
How expressive and classical simulable is the kernel?

How much of the power of the target function lies within the kernel eigenfunctions?

How is the kernel performance?

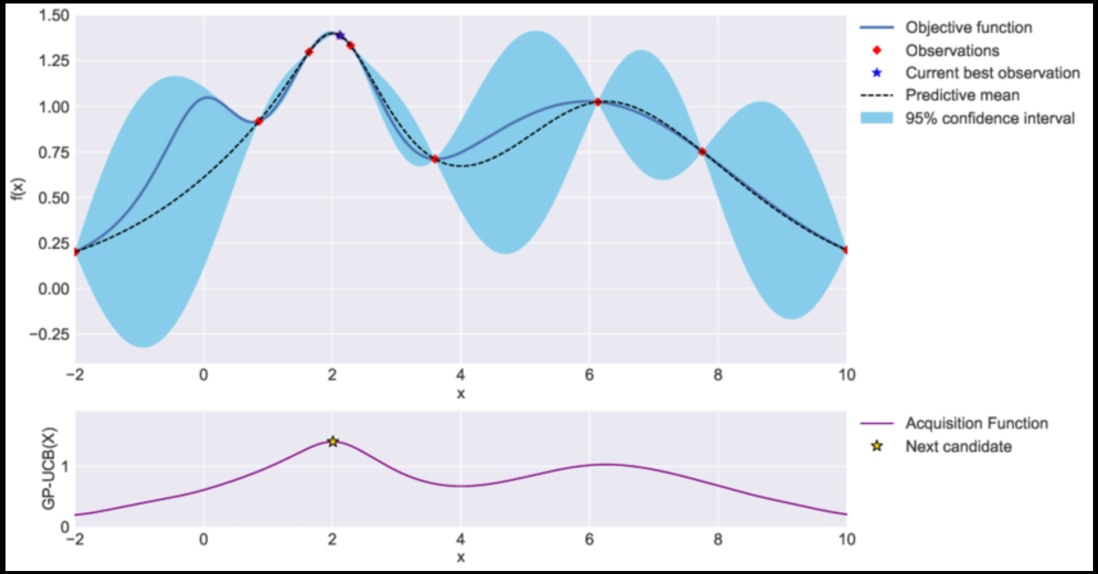
MODEL

$$\boldsymbol{k}^* = \arg \min_{\boldsymbol{k} \in \mathcal{QK}_{n,m}} C(\boldsymbol{k})$$



OPTIMIZATION I

- **Bayesian optimization**
- Useful for costly functions
- Usually struggles with high-dimensional optimization problems



OPTIMIZATION II

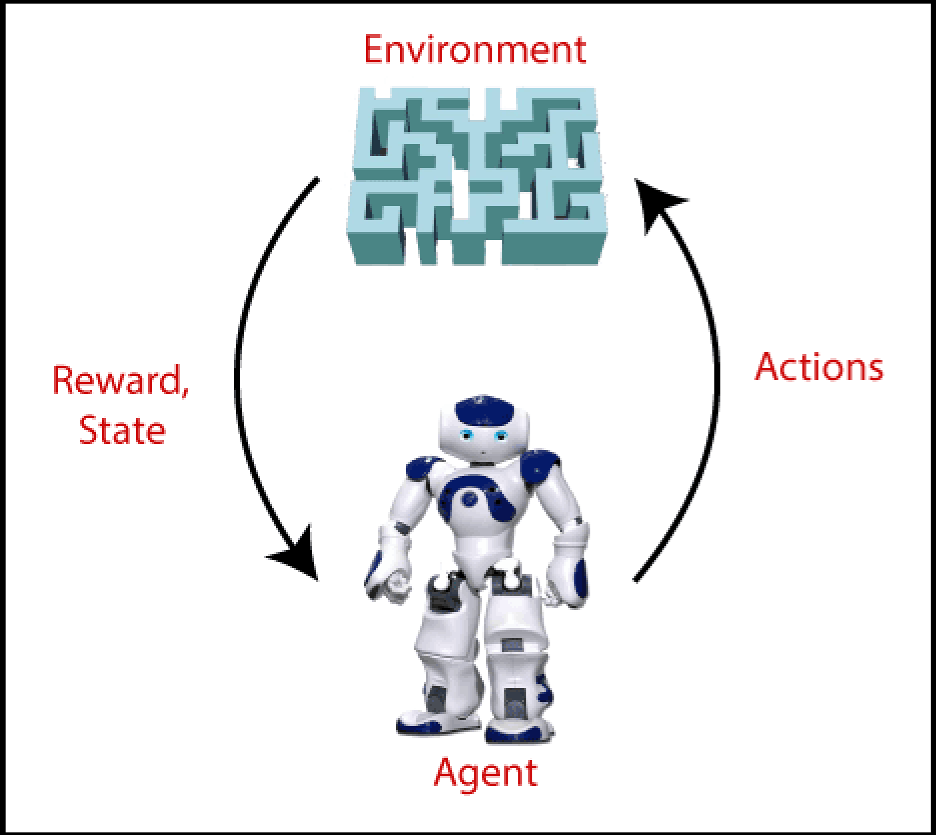
- **Evolutionary algorithms**

- Maintain a population of solutions and, for each iteration, perturb and select a subset of them

- Very expensive but can return high quality solutions

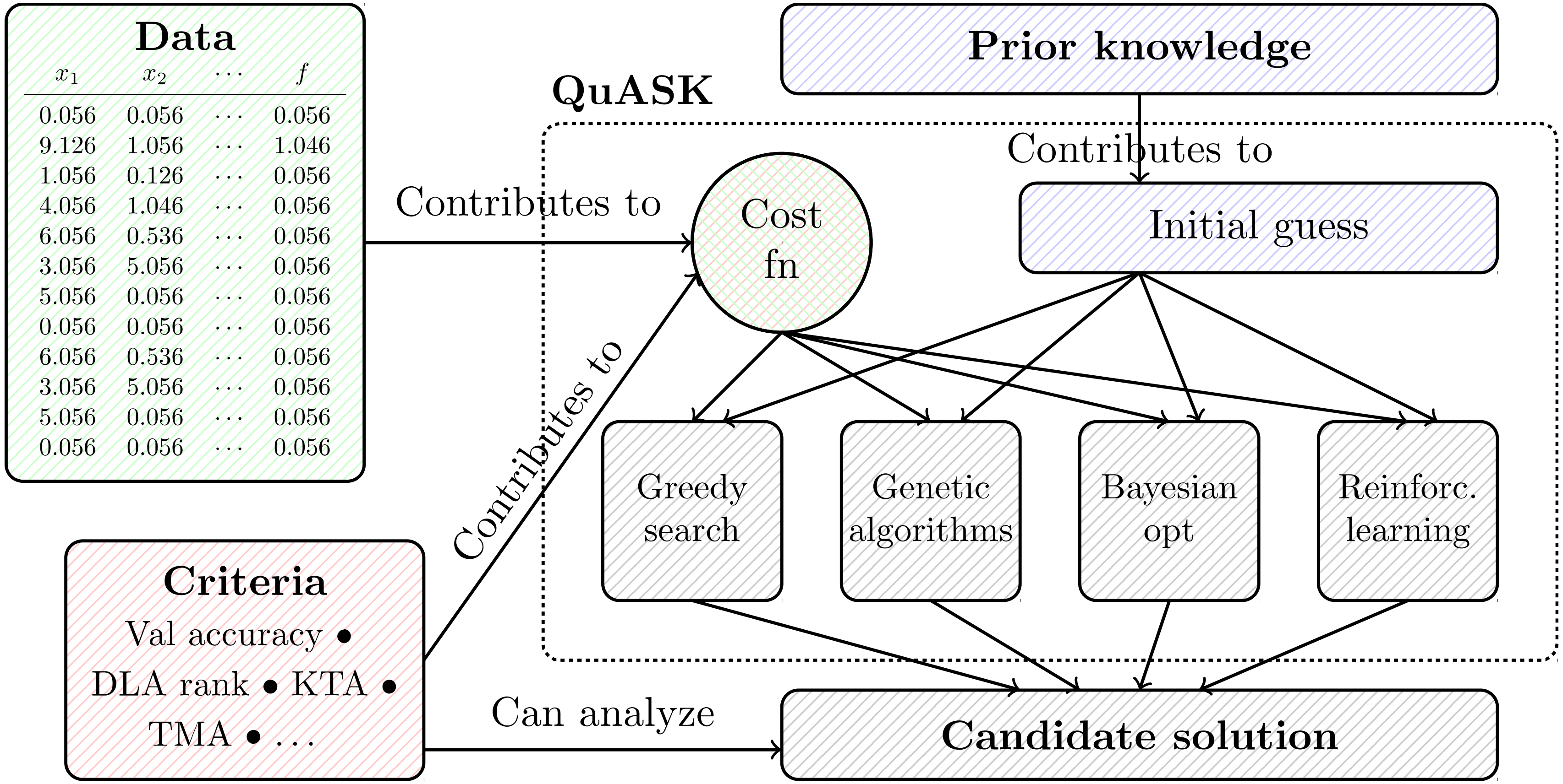
OPTIMIZATION III

- **Reinforcement learning** (SARSA-λ)
- Huge space of actions, short horizon
- Other schemes can be used



IMPLEMENTATION OF QUANTUM KERNELS

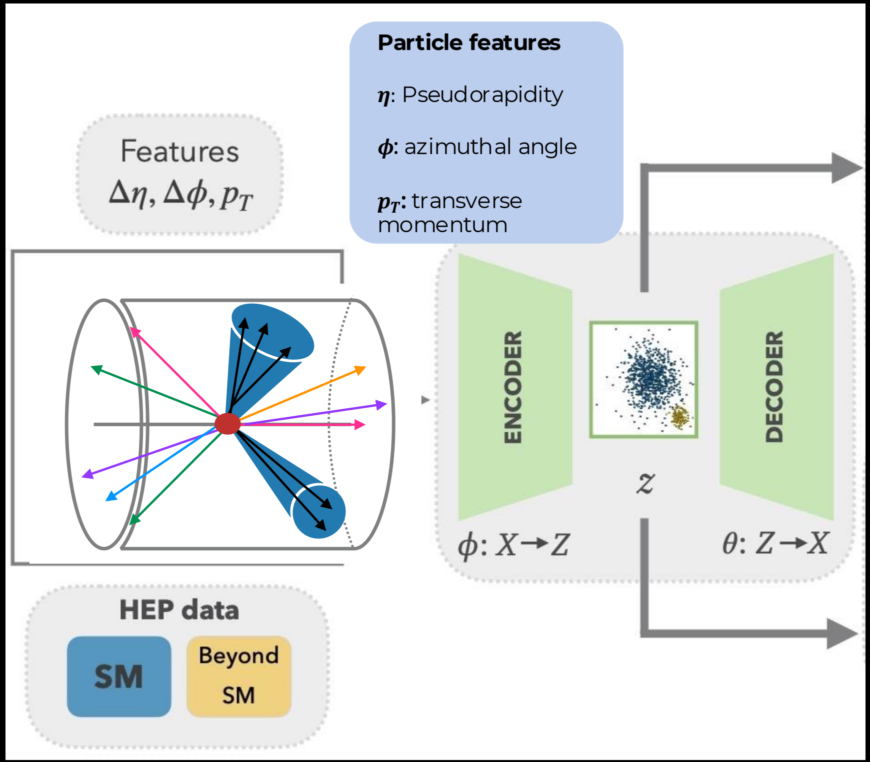
QUASK: QUANTUM ADVANTAGE SEEKER WITH KERNELS

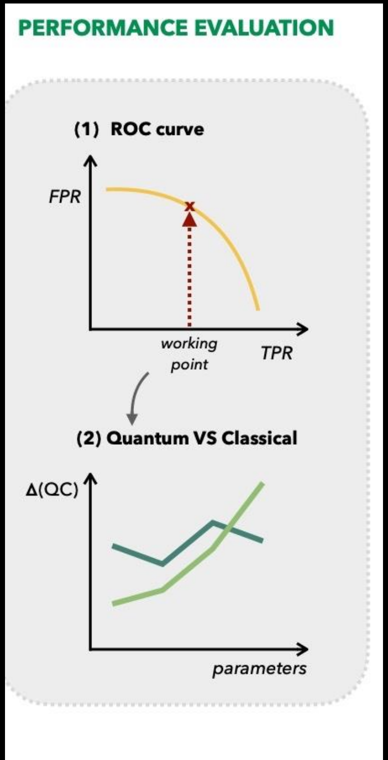


Di Marcantonio, **MG**, et al. "QuASK--Quantum Advantage Seeker with Kernels.»

Quantum Mach. Intell. 5, 20 (2023) - **https://quask.web.cern.ch/**

Anomaly detection of Beyond Standard Model events





Approach	Dataset Narrow G	Dataset $A \to HZ$	Dataset Broad G
Best classical kernel	99.54 ± 0.42	97.94 ± 0.69	50.90 ± 3.97
Best quantum kernel	99.65 ± 0.23	98.05 ± 0.58	55.20 ± 3.96
This work ($m = 8, t = 5$)	99.28 ± 0.58	97.66 ± 0.61	47.62 ± 4.17
This work ($m = 8, t = 10$)	99.28 ± 0.58	97.66 ± 0.61	$\mathbf{61.61 \pm 4.67}$
This work ($m = 12, t = 5$)	99.37 ± 0.46	94.57 ± 1.60	58.70 ± 4.38
This work ($m = 12, t = 10$)	$\mathbf{99.73 \pm 0.20}$	$\mathbf{98.34 \pm 0.37}$	58.70 ± 4.38

Background: QCD dijet events. 600 features per event ⟶ Too many for current hardware.

BSM anomalies: Graviton **&** New Scalar Boson

CONCLUDING REMARKS

TAKE-AWAY MESSAGE

Quantum kernels can be **useful**

but

we **lack information** to **choose** the **quantum transformation**

therefore

we use **combinatorial optimization** to **choose** a **quantum circuit**

we exploit **numerical properties** to **speedup** the **optimization**