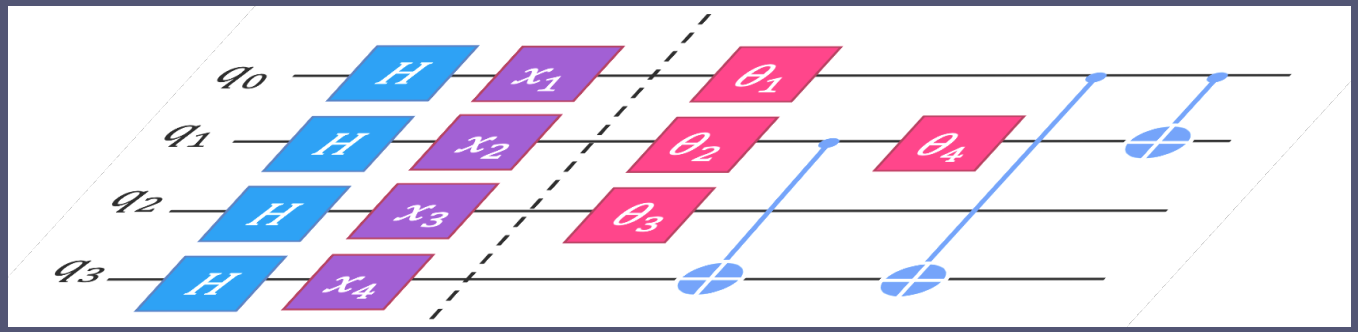
Can quantum computers enhance machine learning? If yes, then how?

Roman Krems University of British Columbia

Quantum machine learning

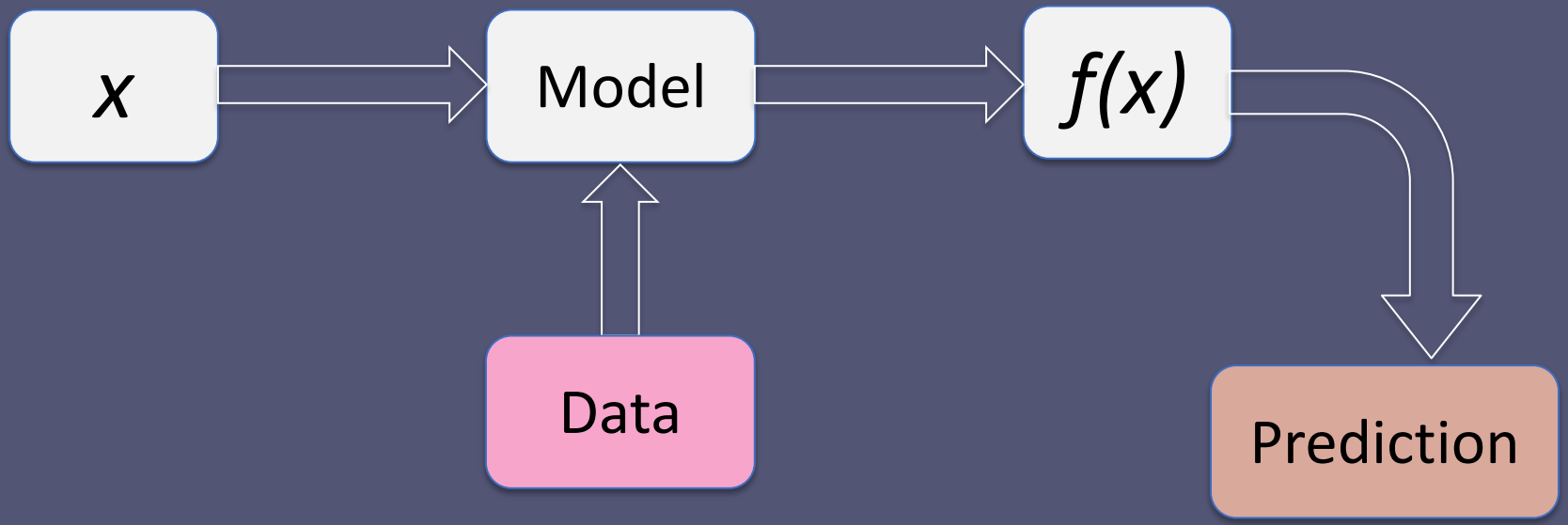
Uses quantum computers for any part of the machine learning algorithm



Goal

Perform computations that cannot be done with classical computers alone

Supervised machine learning: input – output pairs are used to make predictions at unseen parts of input space

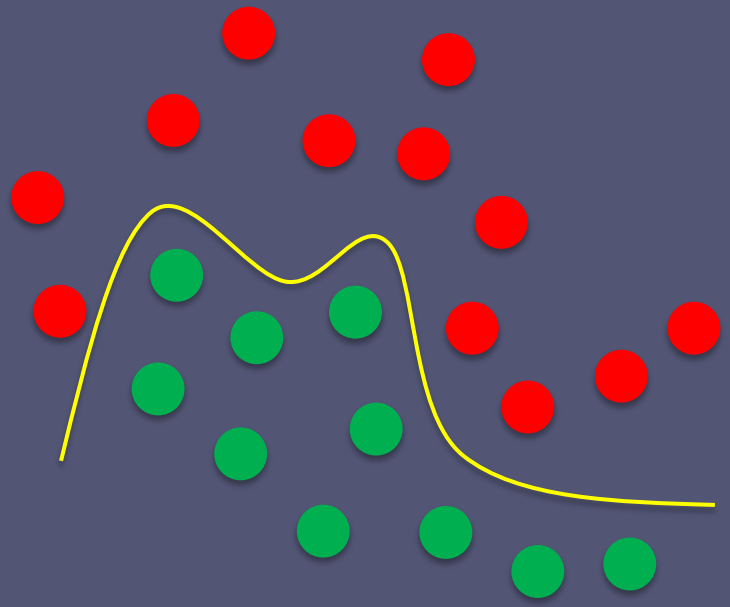


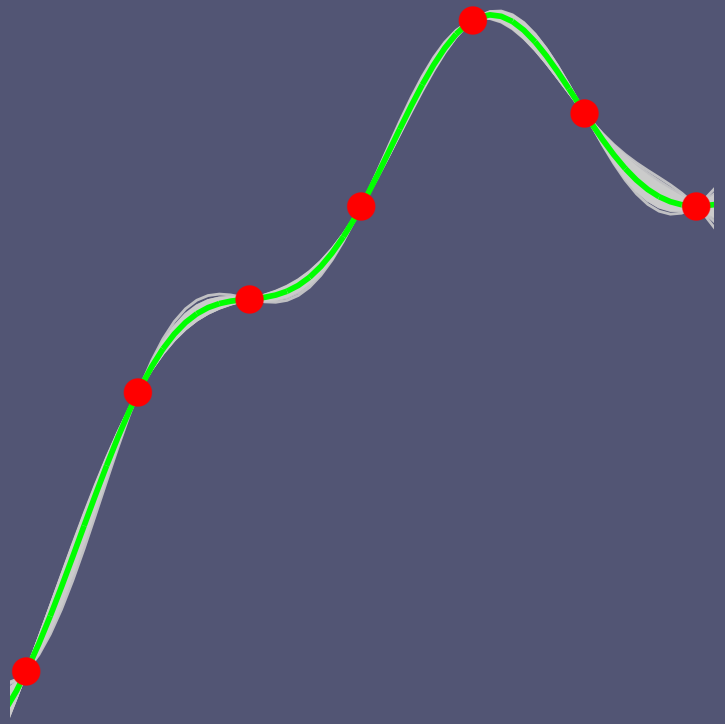
Inputs: variables collected into multi-dimensional vectors (e.g. parameters of experiment)

Outputs: quantity to predict (e.g. observable)

Supervised machine learning: input – output pairs are used to make predictions at unseen parts of input space

Classification (outputs: +/-)





Regression (continuous outputs)

Does quantum machine learning have the quantum advantage?

In principle, Yes! ⟸ Computational complexity theory

Universal expressiveness of variational quantum classifiers and quantum kernels for support vector machines

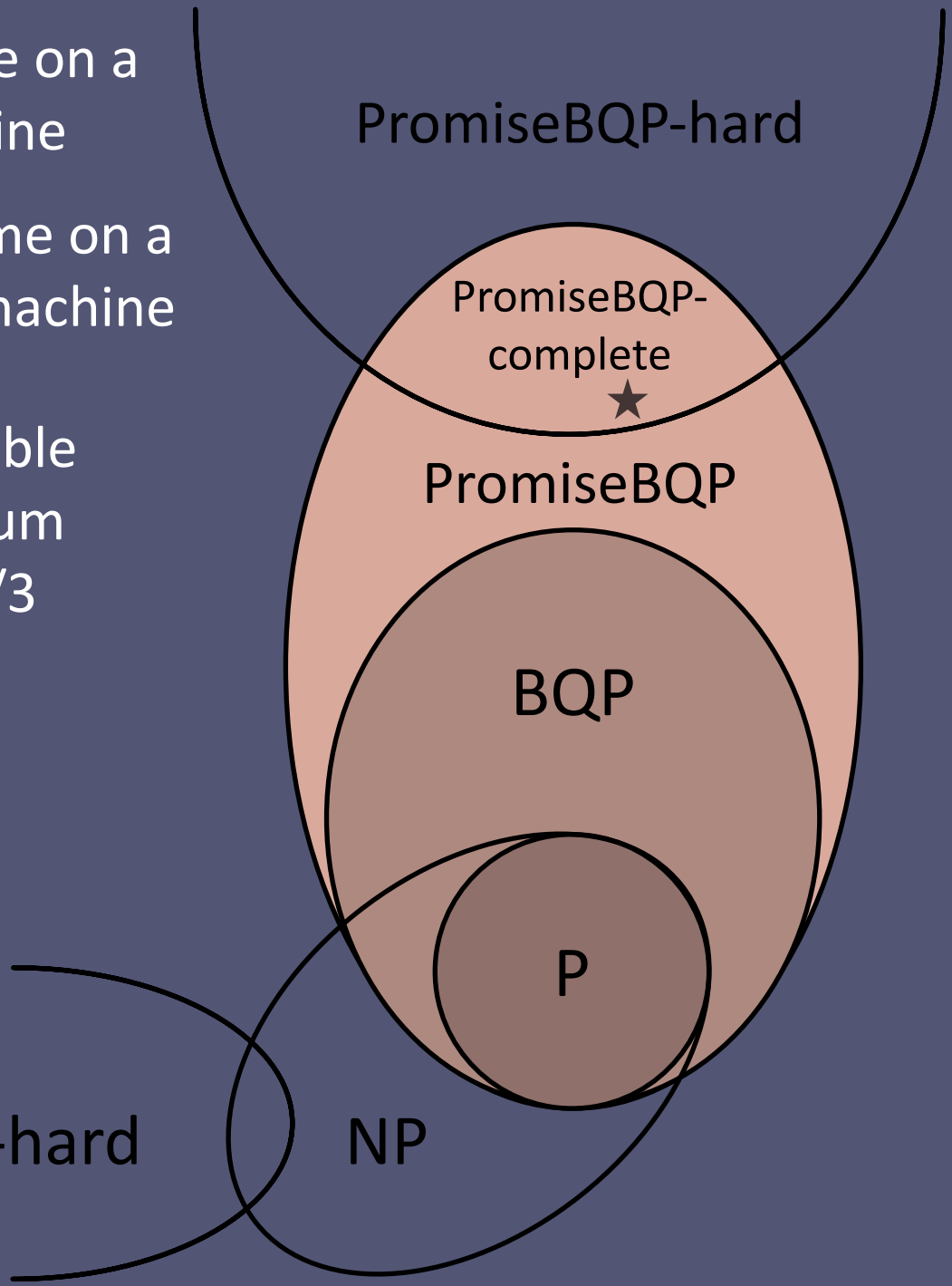
J Jäger, RV Krems

Nature Communications 14 (1), 576 (2023)

P = solvable in polynomial time on a deterministic Turing machine

NP = solvable in polynomial time on a non-deterministic Turing machine

BQP = decision problems solvable in polynomial time by a quantum computer with probability > 2/3



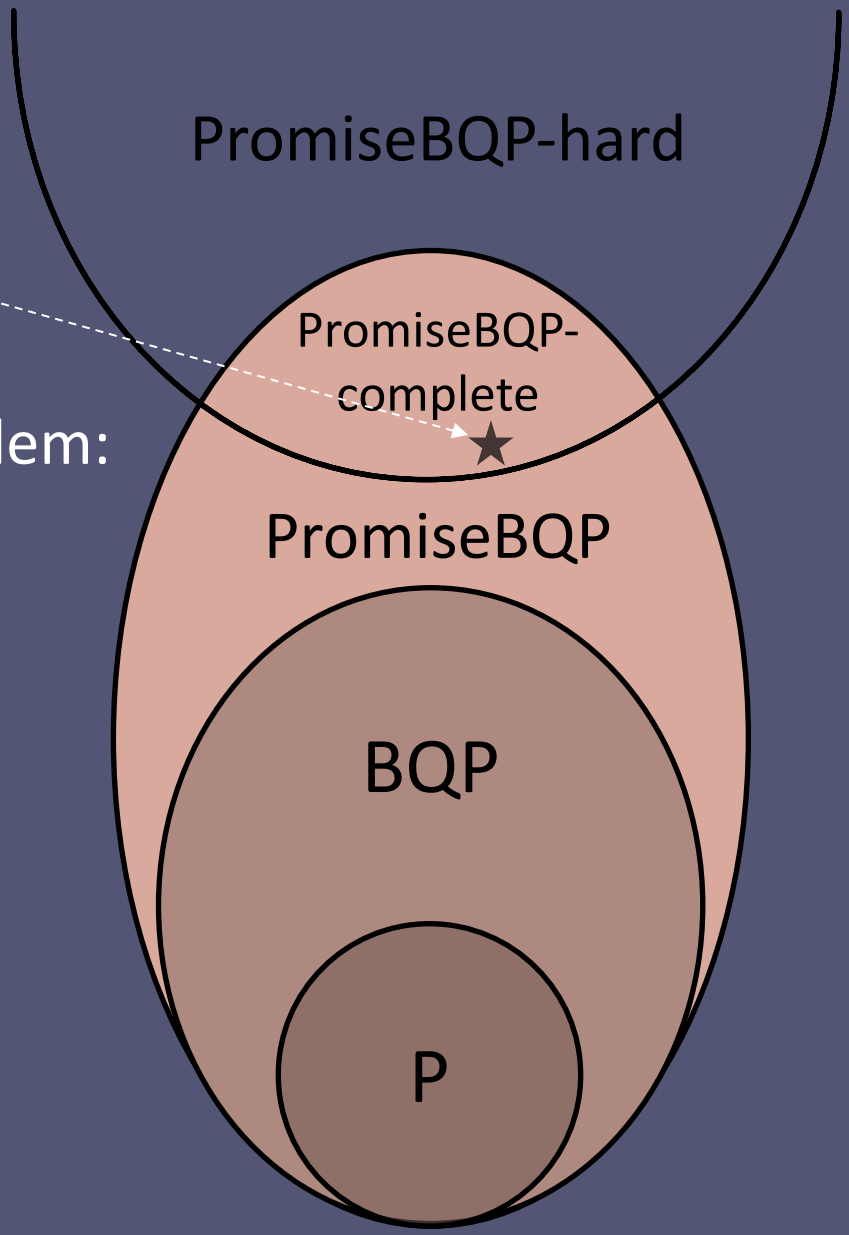
k-Forrelation problem

Generalization of the Forrelation problem:

Two Boolean functions *f* and *g*

Are f and g forrelated or not?

Forrelation = correlation between *f* and the Fourier transform of *g*

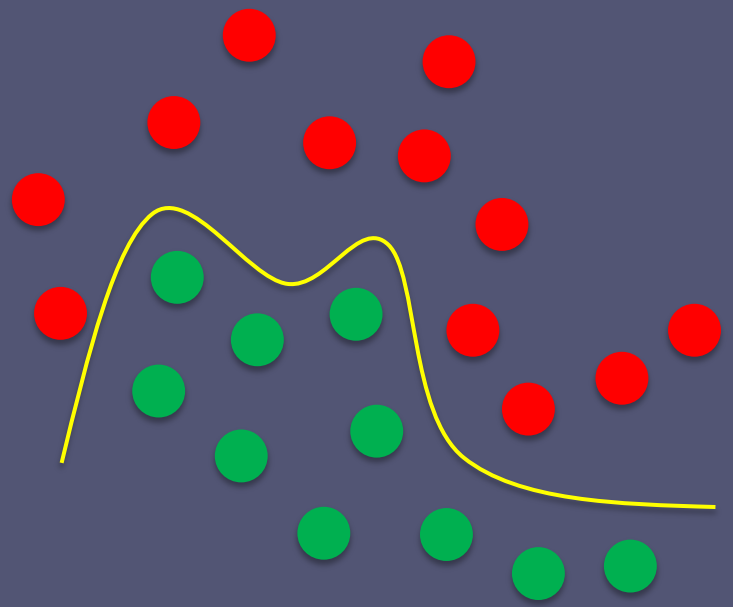


Aaronson, S. & Ambainis, A. Forrelation: A problem that optimally separates quantum from classical computing. SIAM Journal on Computing 47, 982–1038 (2018)

Universal expressiveness of variational quantum classifiers and quantum kernels for support vector machines
J Jäger, RV Krems
Nature Communications 14 (1), 576

Reformulated the k-Forrelation promise problem as a ML classification problem

k-Forrelation promise problem



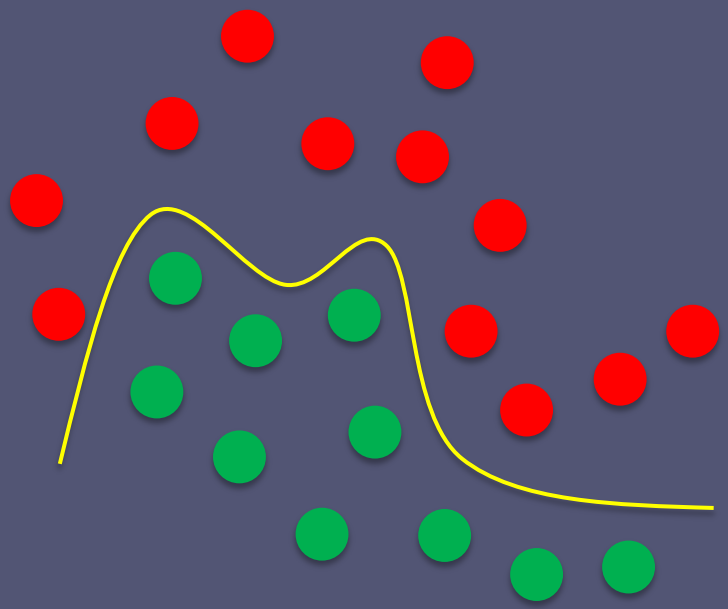
Separating boundary can be constructed efficiently using quantum circuits

Does quantum machine learning have a quantum advantage? … Yes!

Does quantum machine learning have the quantum advantage?

In principle, Yes! … Why “in pricniple”?

k-Forrelation quantum circuits solve the classification problem exactly…
What if the accuracy requirement is relaxed?



Is there an efficient classical algorithm that gives an approximate solution for the separating boundary?

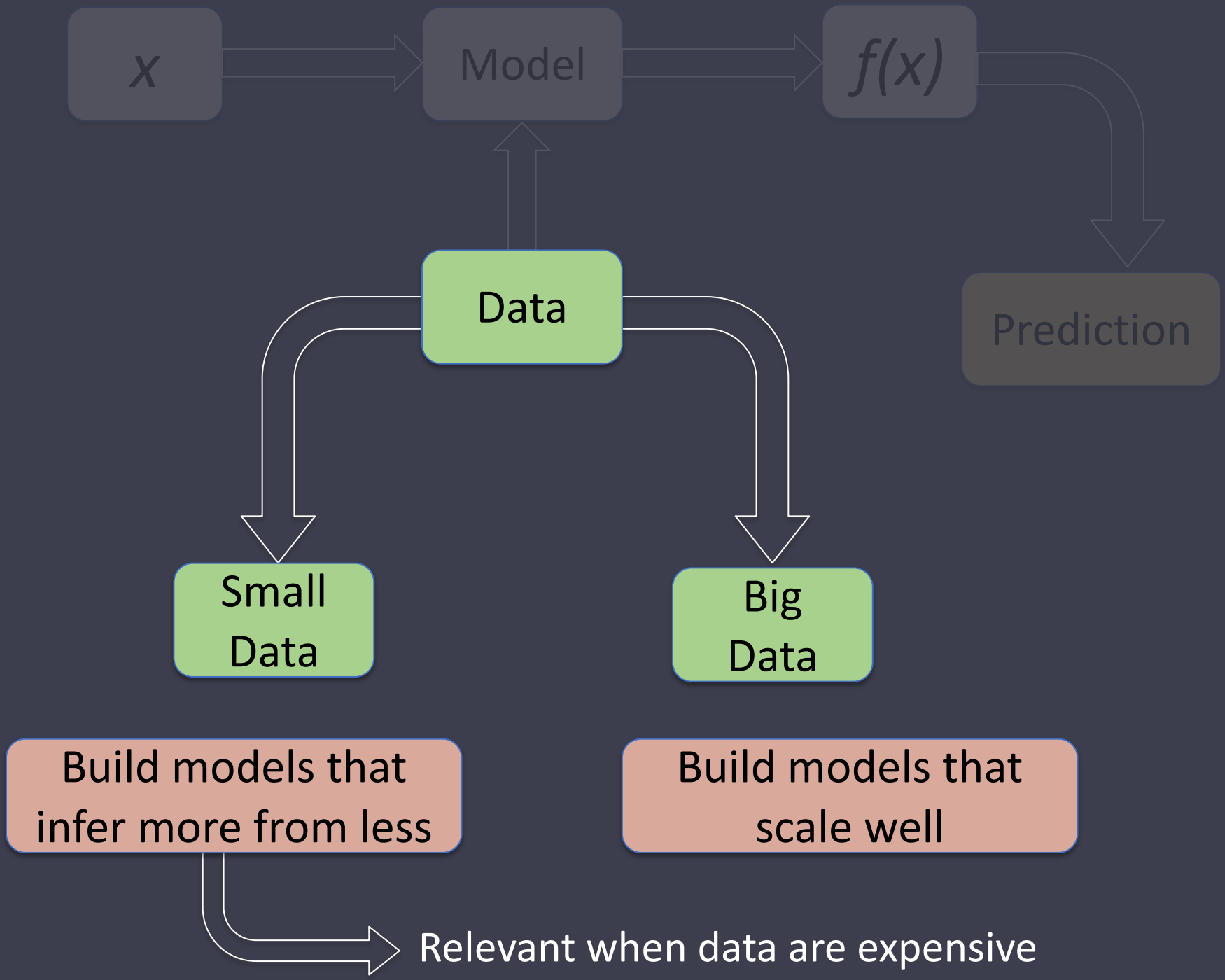
Does quantum machine learning have the quantum advantage?

In principle, Yes! … Why “in pricniple”?

We know how to bias the quantum models for the k-Forrelation data set…

However, we don’t necessarily know how to do this for arbitrary problems

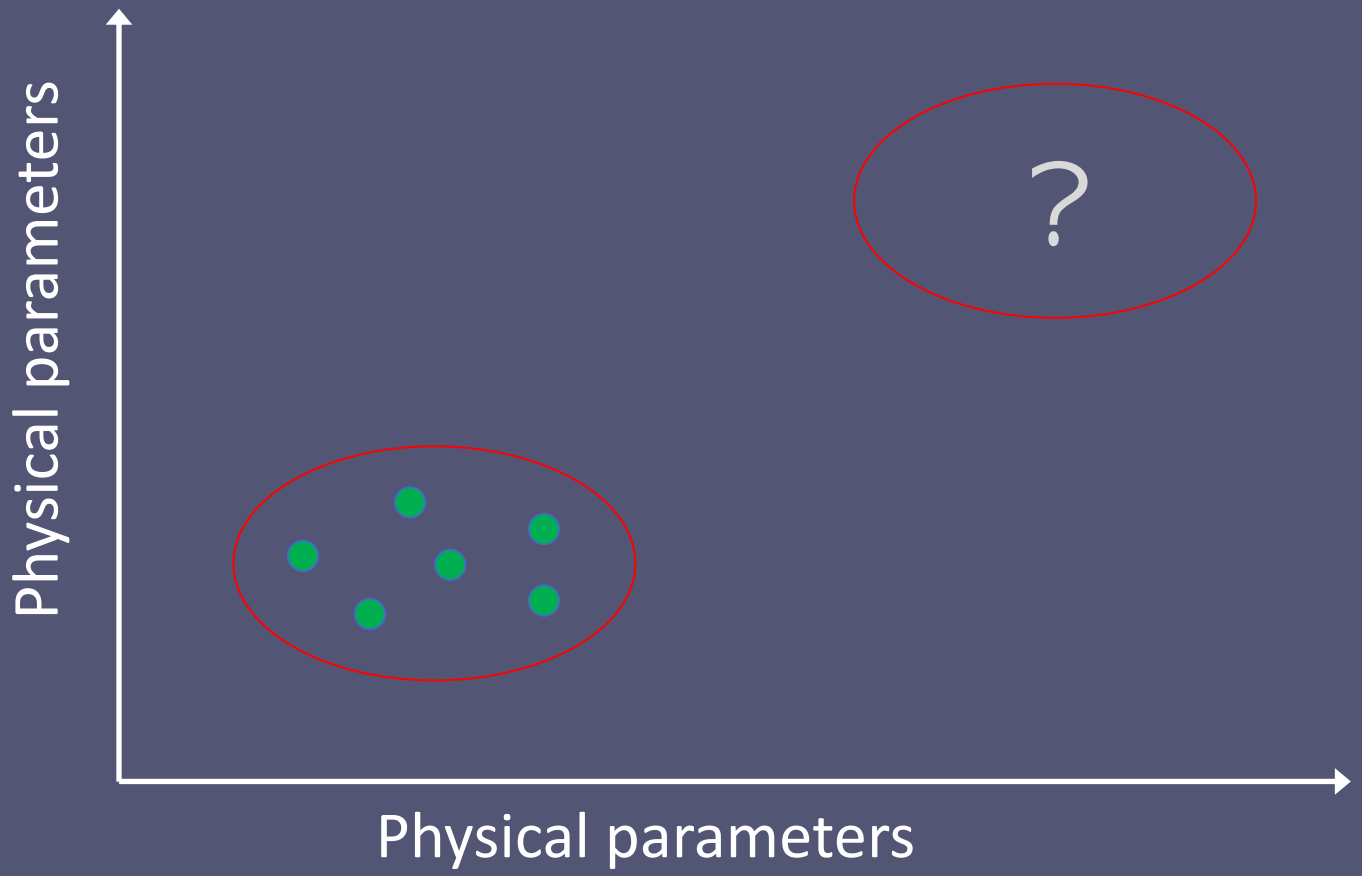
How to build quantum machine learning models that *outperform* classical machine learning for *practical* applications?


x
Model
f(x)
Prediction
Data
Small Data
Big Data
Build models that infer more from less
Build models that scale well
Relevant when data are expensive

Question:

Can one build physical ML models based on small data?

What is `physical'?



Goal:

Build models that infer more (physics) from less (data)

How?

Kernel models: mathematically grounded

Bayesian models: a way to select best kernels for limited data

Can quantum computing help?

Can quantum kernels offer better inference than classical kernels?

Summary of key points so far:

Quantum machine learning has a quantum advantage!

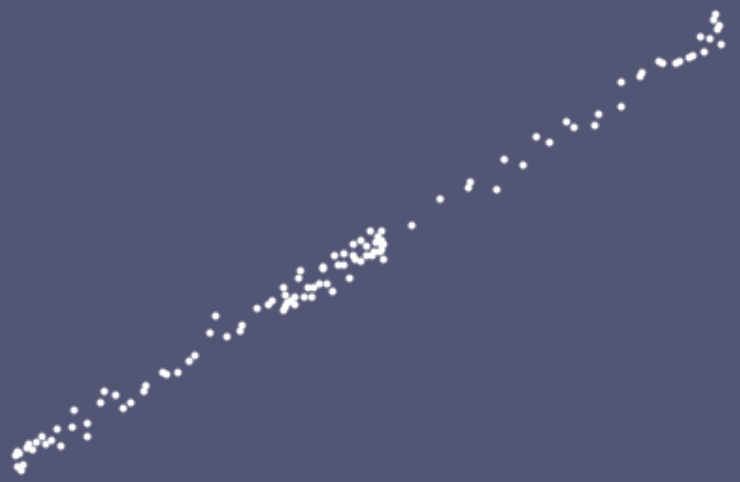
We know this because a QML algorithm can be designed to solve a PromiseBQP-complete problem.

We don't know: how to take advantage of the quantum advantage!

Let's aim to address this questions by constructing quantum ML models that "infer more from less" for low-dimensional problems.

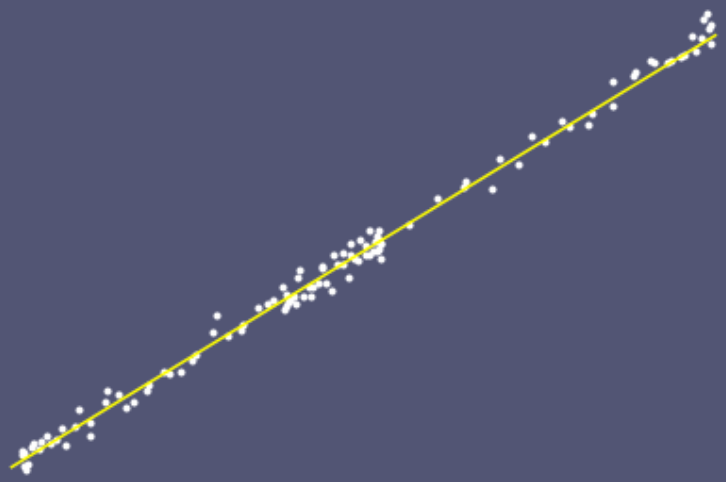
How does (kernel) machine learning work?

$$\sum_{i}^{n} \left(y_i - \boldsymbol{\beta}^\top \boldsymbol{x}_i \right)^2 + \gamma \boldsymbol{\beta}^\top \boldsymbol{\beta}$$

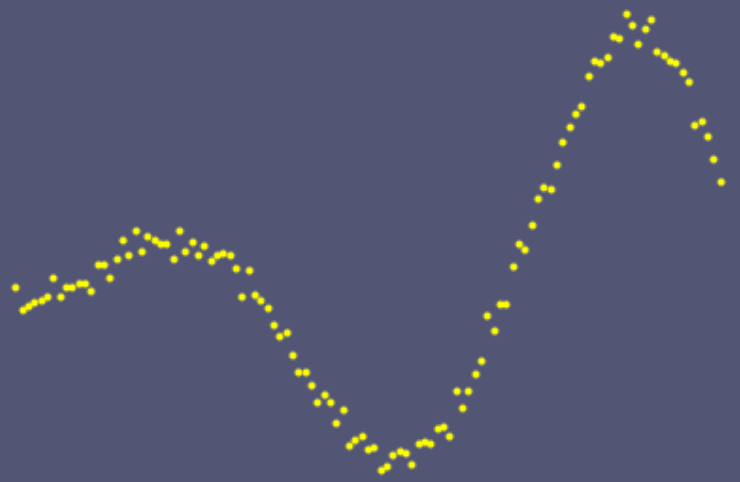


How does (kernel) machine learning work?

$$\sum_{i}^{n} \left(y_i - \boldsymbol{\beta}^\top \boldsymbol{x}_i \right)^2 + \gamma \boldsymbol{\beta}^\top \boldsymbol{\beta}$$

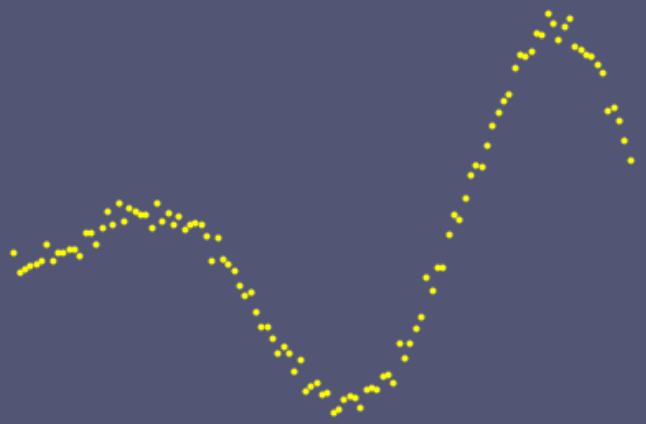


The reality is (most often) non-linear

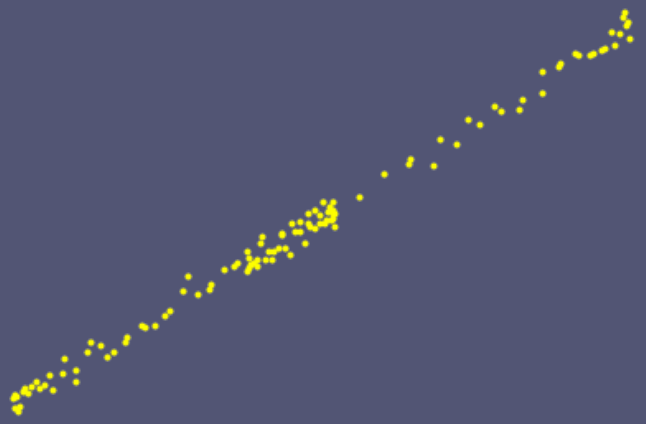


... but it can be made linear ...

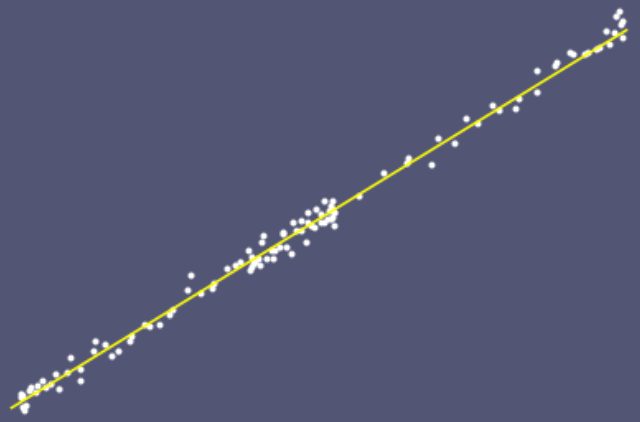
The change $x \to \Phi(x)$ with $\Phi(x) = x \sin(x)$, transforms



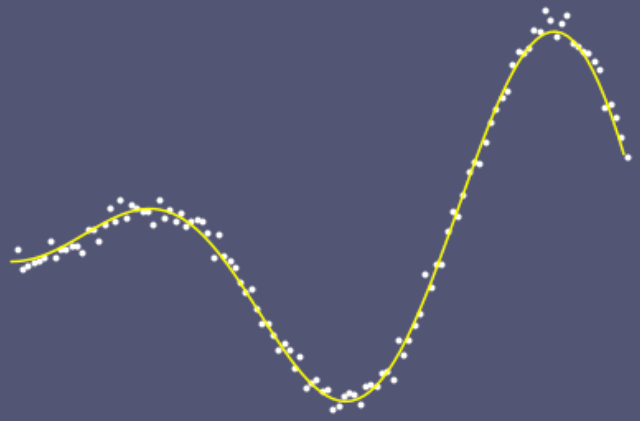
into



Performing regression

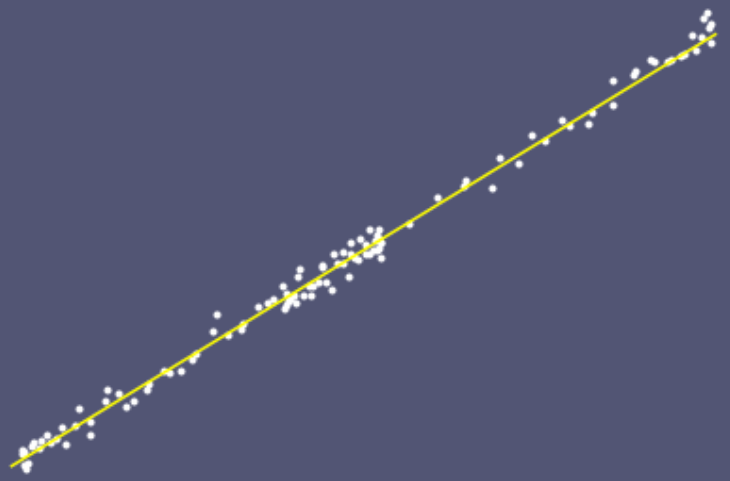


and transforming back, yields:

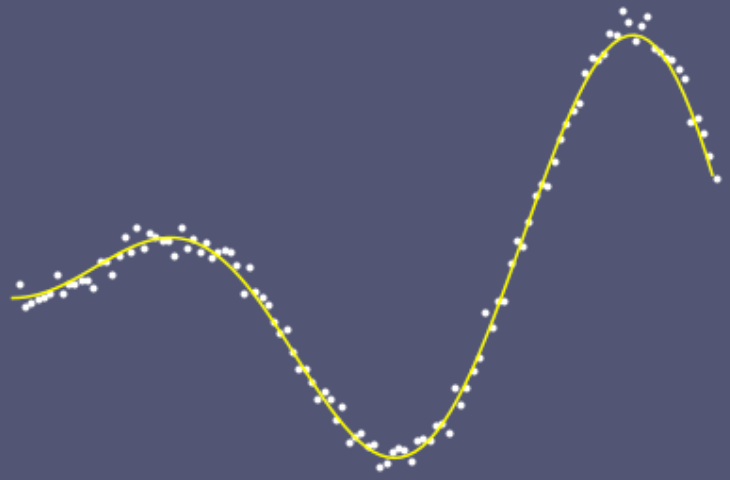


If only we knew $\Phi(x)$ for every problem ...

$$\sum_i^n \left(y_i - \boldsymbol{\beta}^\top \boldsymbol{x}_i\right)^2 + \gamma \boldsymbol{\beta}^\top \boldsymbol{\beta}$$



$$\sum_i^n \left(y_i - f(\boldsymbol{x}_i)\right)^2 + \gamma \langle f, f \rangle_{\mathcal{H}}$$



Inner product in a high-dimensional Hilbert space

Hilbert showed that the integrals over two real-valued, square-integrable functions

$$\text{inner product} \rightarrow \qquad \langle f, g \rangle = \int_a^b f(x) g(x) dx$$

have the same properties as a scalar product of two vectors in a Euclidean space.

For orthogonal functions, this means:

$$\langle f, g \rangle = 0$$

Consider an integral transform:

$$f(x) = \int_a^b K(x,y) f(y) dy$$

with the kernel $K(x,y)$ that is symmetric in x and y.

Hilbert's discovery implies that the kernel can be written as

$$K(x,y) = \sum_{n=1}^{\infty} \lambda_n \phi_n(x) \phi_n(y)$$

where $\phi_n(x)$ are orthogonal functions.

If $\phi_n(x)$ are ortho-normal, $\lambda_n = 1$.

$$K(x,y) = \sum_{n=1}^{\infty} \phi_n(x)\phi_n(y) \qquad \text{is a special function}$$

Let's index it by x and view it as a function of y:

$$K_x(y) = \sum_n a_{x,n}\phi_n(y)$$

We can then write:

$$f(x) = \int_a^b K(x,y)f(y)dy = \langle f, K_x \rangle$$

Thus, any function can be written as

$$f(x) = \langle f, K_x \rangle$$

Now, consider an arbitrary, positive-definite function $K(x, y)$ with the eigenvalue equation:

$$\int_a^b K(x,y)\phi_n(y)dy = \lambda_n \phi_n(x) \quad \text{with} \quad \lambda_n \geq 0$$

Can we still write: $f(x) = \langle f, K_x \rangle$?

The answer is Yes! – and this leads to the notion of RKHS = Reproducing Kernel Hilbert Space

The answer is yes, if we redefine the inner product of the Hilbert space as

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{n=1}^{\infty} \frac{\langle f, \phi_n \rangle \langle g, \phi_n \rangle}{\lambda_n}$$

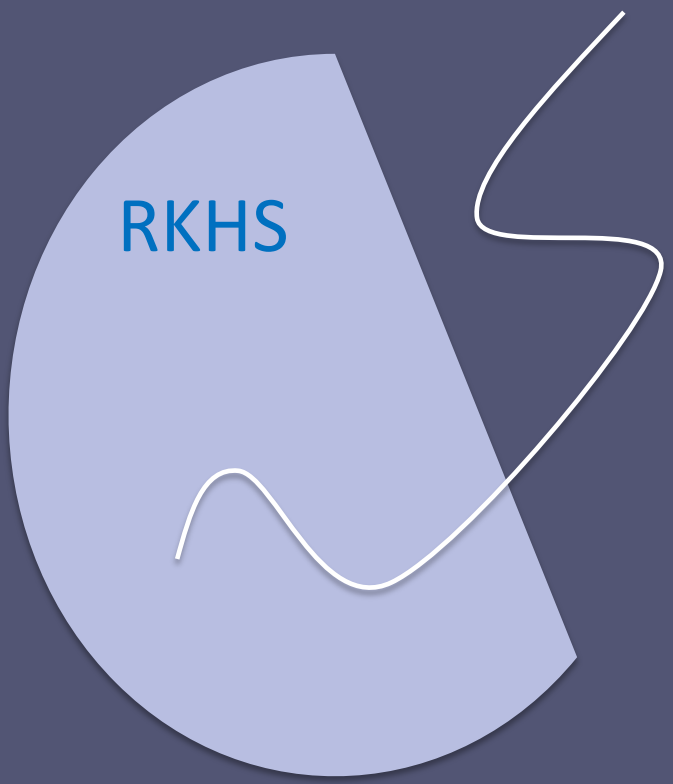
In this case,

$$\langle f, K_x \rangle_{\mathcal{H}} = \sum_{n=1}^{\infty} \frac{\langle f, \phi_n \rangle \langle K_x, \phi_n \rangle}{\lambda_n} = \sum_{n} \langle f, \phi_n \rangle \phi_n(x) = f(x)$$

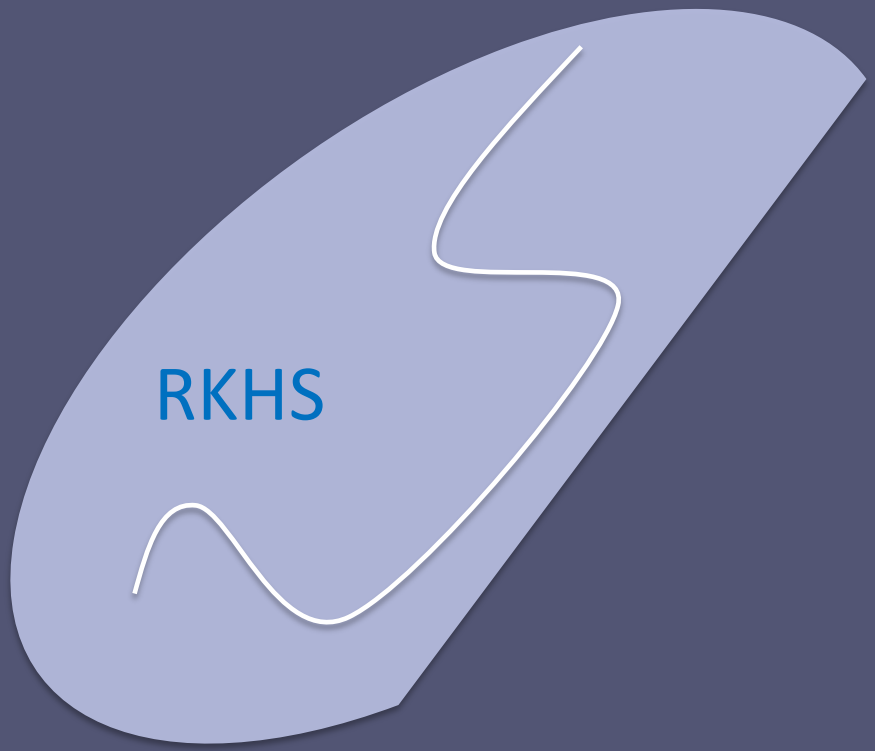
What if some of the eigenvalues λ_n are zero (or small)?

$$\int_a^b K(x, y) \phi_n(y) dy = \lambda_n \phi_n(x) \quad \text{with} \quad \lambda_n \geq 0$$

The choice of the kernel function determines the learning



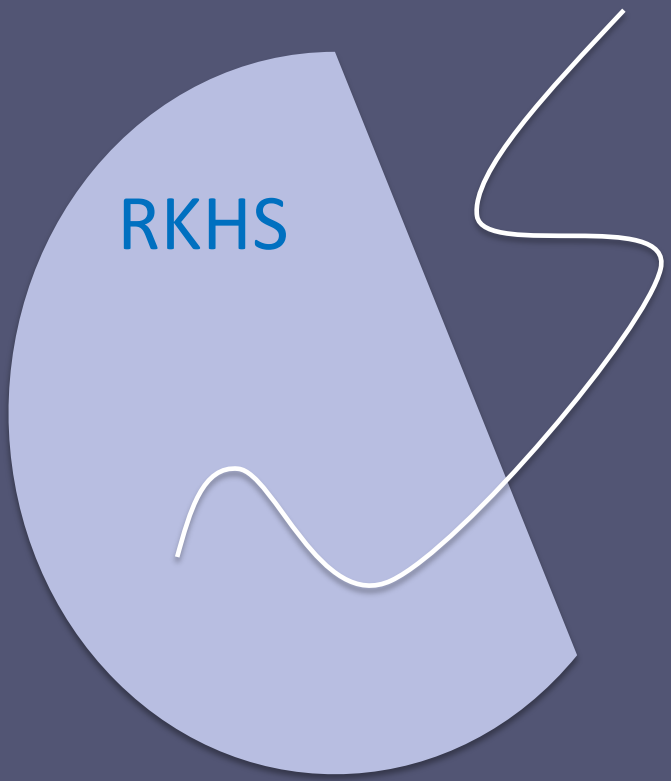
Bad kernel



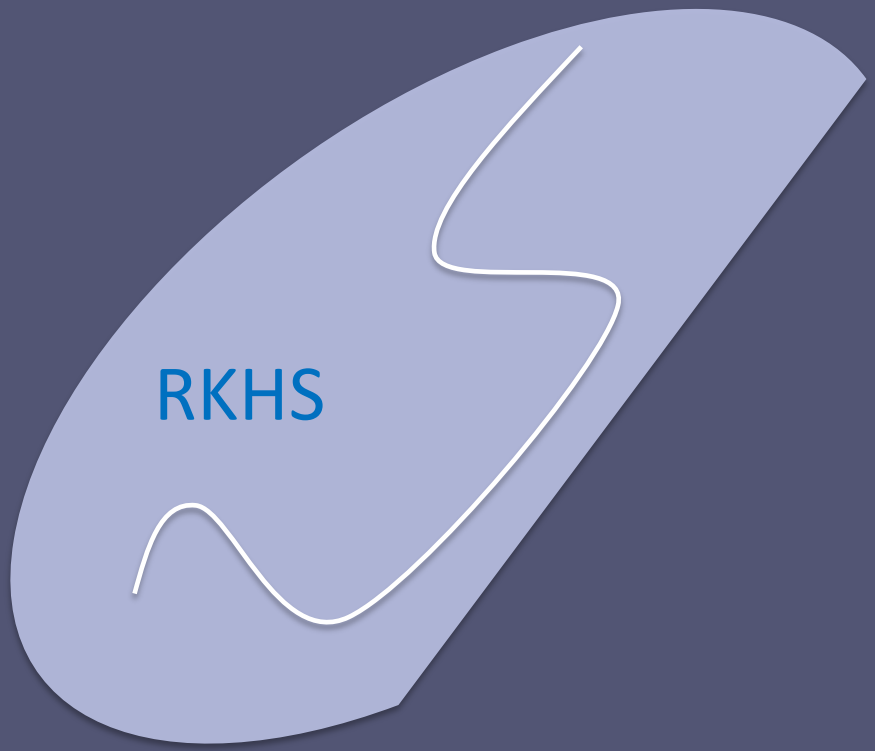
Good kernel

$$\int_a^b K(x,y)\phi_n(y)dy = \lambda_n\phi_n(x) \quad \text{with} \quad \lambda_n \geq 0$$

The choice of the kernel function determines the learning



Bad kernel



Good kernel

How to find the right kernel?

Given two kernels, how to tell which is better?

The Bayes' theorem: $$P(\mathcal{M}_i|\text{Data}) = \frac{P(\text{Data}|\mathcal{M}_i)P(\mathcal{M}_i)}{P(\text{Data})}$$

$$\frac{P(\mathcal{M}_1|\text{Data})}{P(\mathcal{M}_2|\text{Data})} = \frac{P(\text{Data}|\mathcal{M}_1)}{P(\text{Data}|\mathcal{M}_2)} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}$$

Assuming the same priors: $P(\mathcal{M}_1) = P(\mathcal{M}_2)$

The ratio of the posteriors: $$\frac{P(\mathcal{M}_1|\text{Data})}{P(\mathcal{M}_2|\text{Data})} = \frac{P(\text{Data}|\mathcal{M}_1)}{P(\text{Data}|\mathcal{M}_2)}$$

Marginal likelihood

Given two kernels, how to tell which is better?

We approximate marginal likelihood by Bayesian Information Criterion

The ratio of the posteriors:

$$\frac{P(\mathcal{M}_1|\text{Data})}{P(\mathcal{M}_2|\text{Data})} = \frac{P(\text{Data}|\mathcal{M}_1)}{P(\text{Data}|\mathcal{M}_2)}$$

Marginal likelihood

Start with several basic kernels, such as these ones:

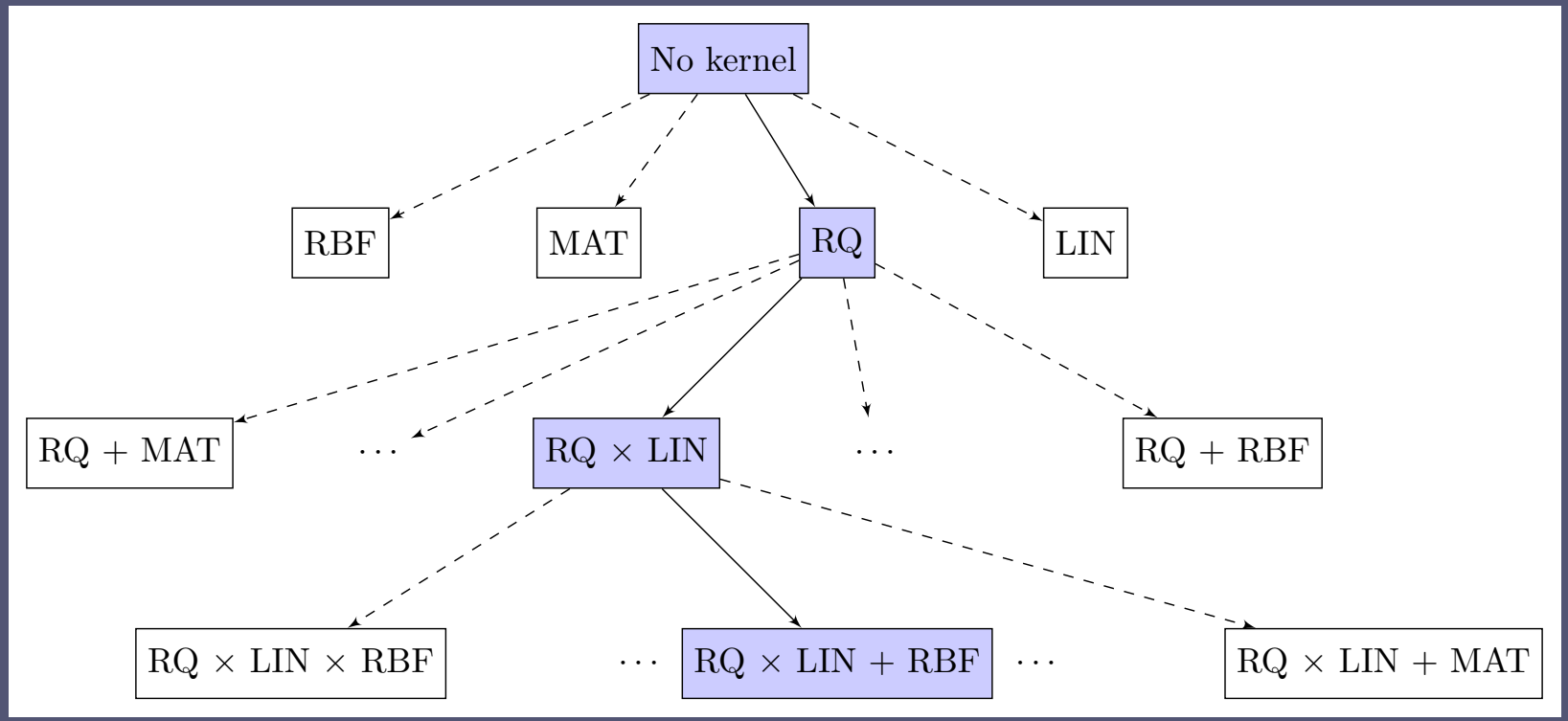
$$k_{\mathrm{LIN}}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$$

$$k_{\mathrm{RBF}}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{2} r^2(\mathbf{x}_i, \mathbf{x}_j)\right)$$

$$k_{\mathrm{MAT}}(\mathbf{x}_i, \mathbf{x}_j) = \left(1 + \sqrt{5} r^2(\mathbf{x}_i, \mathbf{x}_j) + \frac{5}{3} r^2(\mathbf{x}_i, \mathbf{x}_j)\right) \times \exp\left(-\sqrt{5} r^2(\mathbf{x}_i, \mathbf{x}_j)\right)$$

$$k_{\mathrm{RQ}}(\mathbf{x}_i, \mathbf{x}_j) = \left(1 + \frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{2\alpha\ell^2}\right)^{-\alpha}$$

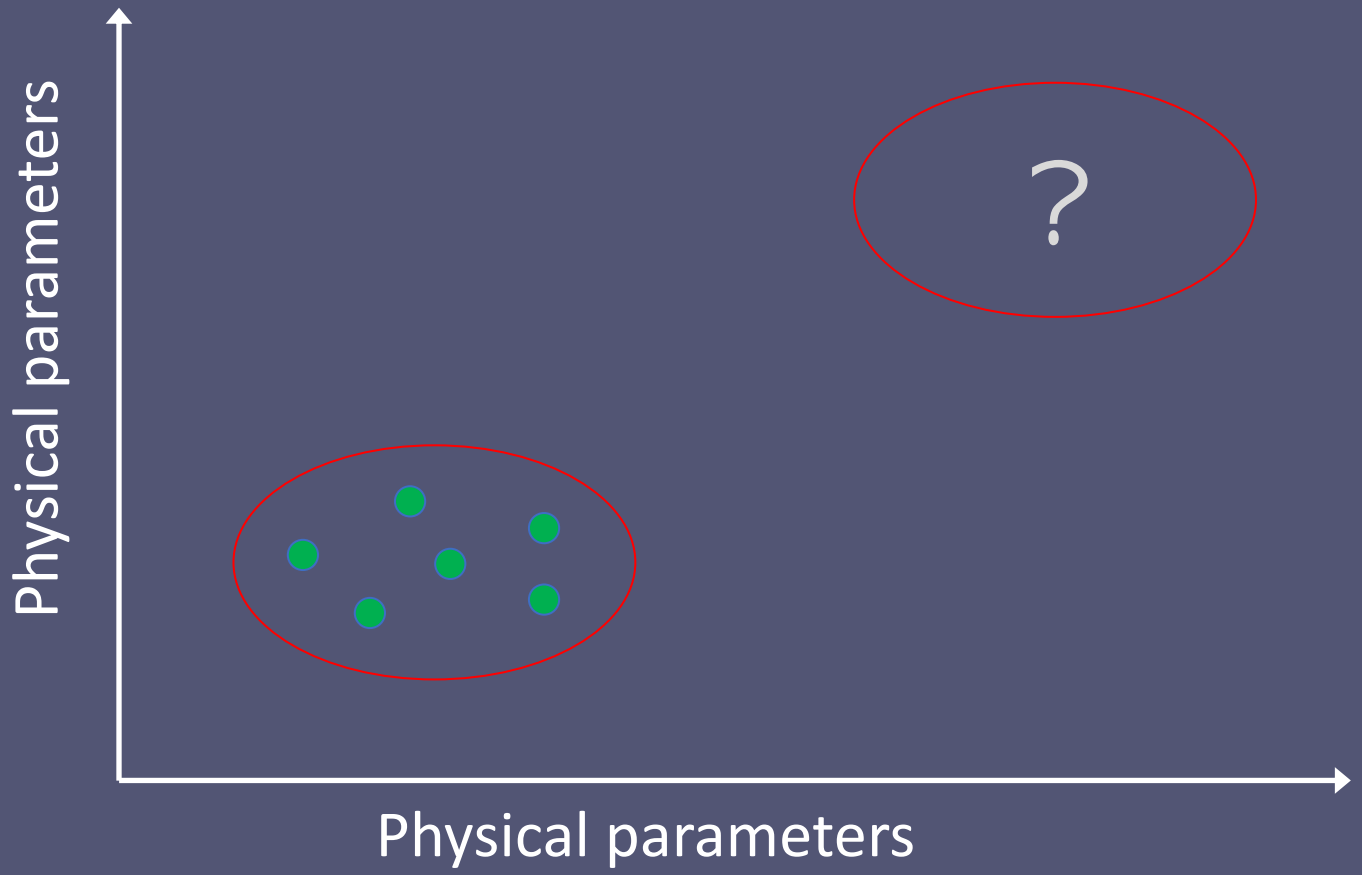
Combine them to increase complexity:



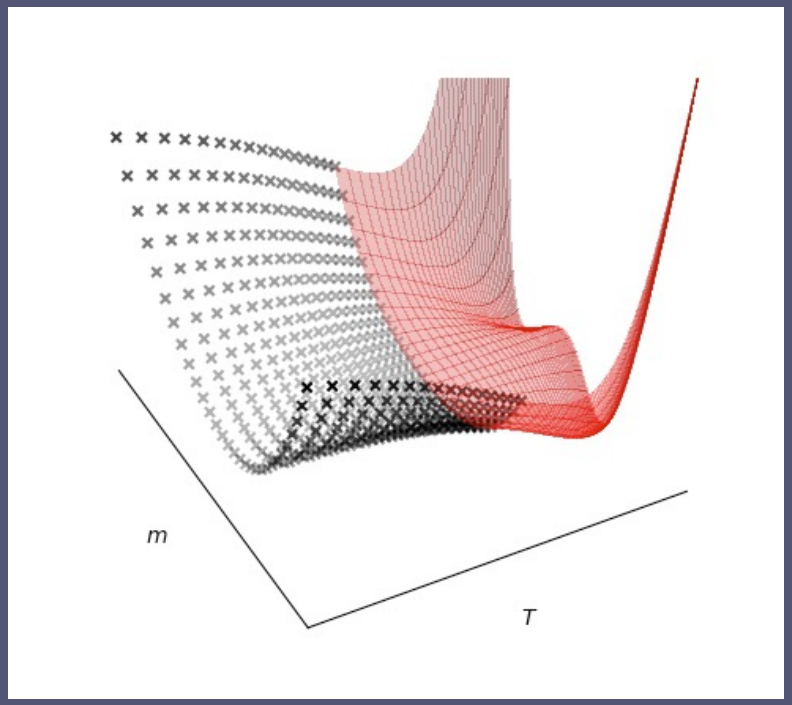
Bayesian Information Criterion: Balances maximum likelihood and model complexity

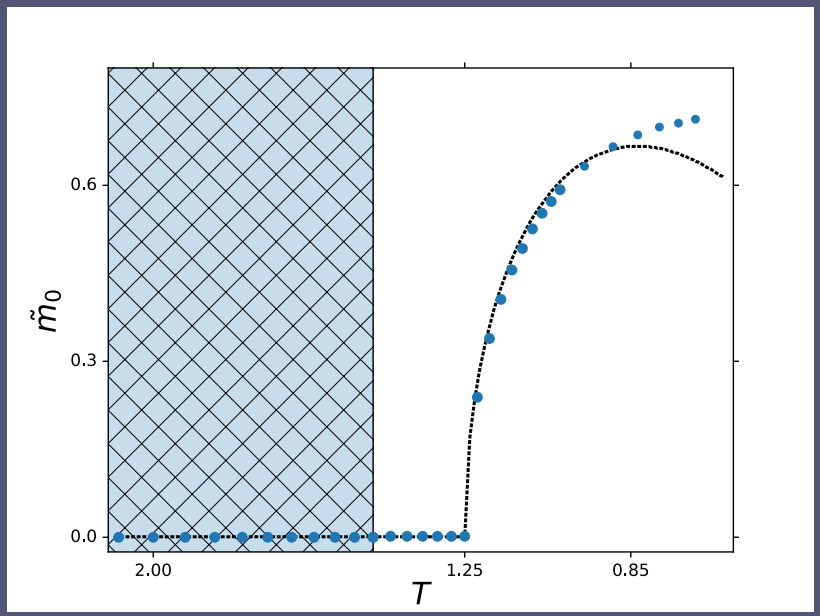
D. K. Duvenaud et al, Structure discovery in nonparametric regression through compositional kernel search, Proceedings of the 30th International Conference on Machine Learning Research 28, 1166 (2013).

This yields ML models that extrapolate!



Heisenberg spin model





Rodrigo Vargas, John Sous, Mona Berciu and R. V. Krems, **Phys. Rev. Lett. 121, 255702 (2018)**
Jun Dai and R. V. Krems**, J. Chem. Theory Comp. 16, 1386 (2020)**

Generalized polaron model

$$\mathcal{H} = \sum_k \epsilon_k c_k^\dagger c_k + \sum_q \omega_q b_q^\dagger b_q + V_{\mathrm{e-ph}}$$

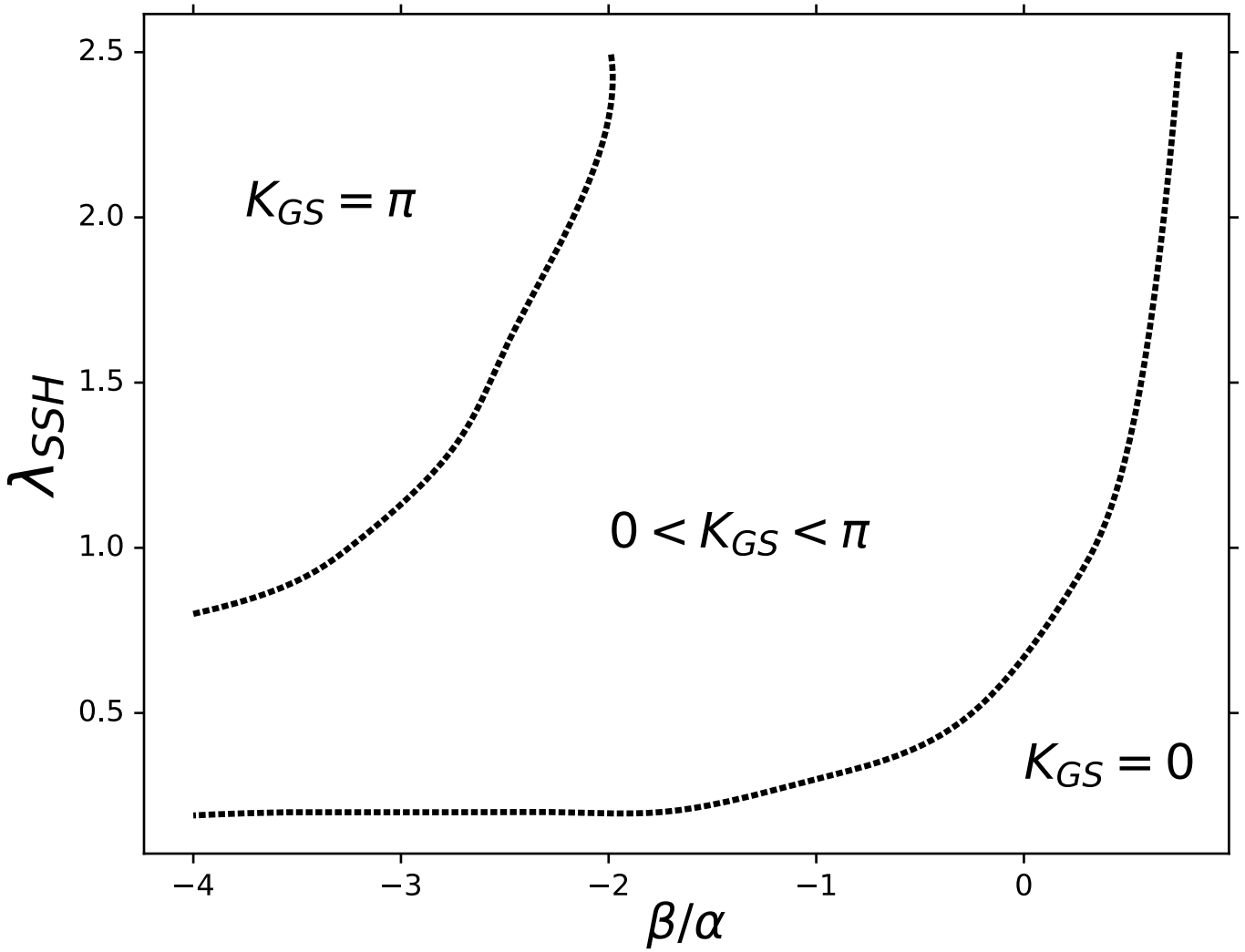
$$V_{\mathrm{e-ph}} = \alpha H_1 + \beta H_2$$

$$H_1 = \sum_{k,q} \frac{2i}{\sqrt{N}} \left[\sin(k+q) - \sin(k)\right] c_{k+q}^\dagger c_k \left(b_{-q}^\dagger + b_q\right)$$

$$H_2 = \sum_{k,q} \frac{2i}{\sqrt{N}} \sin(q) c_{k+q}^\dagger c_k \left(b_{-q}^\dagger + b_q\right)$$

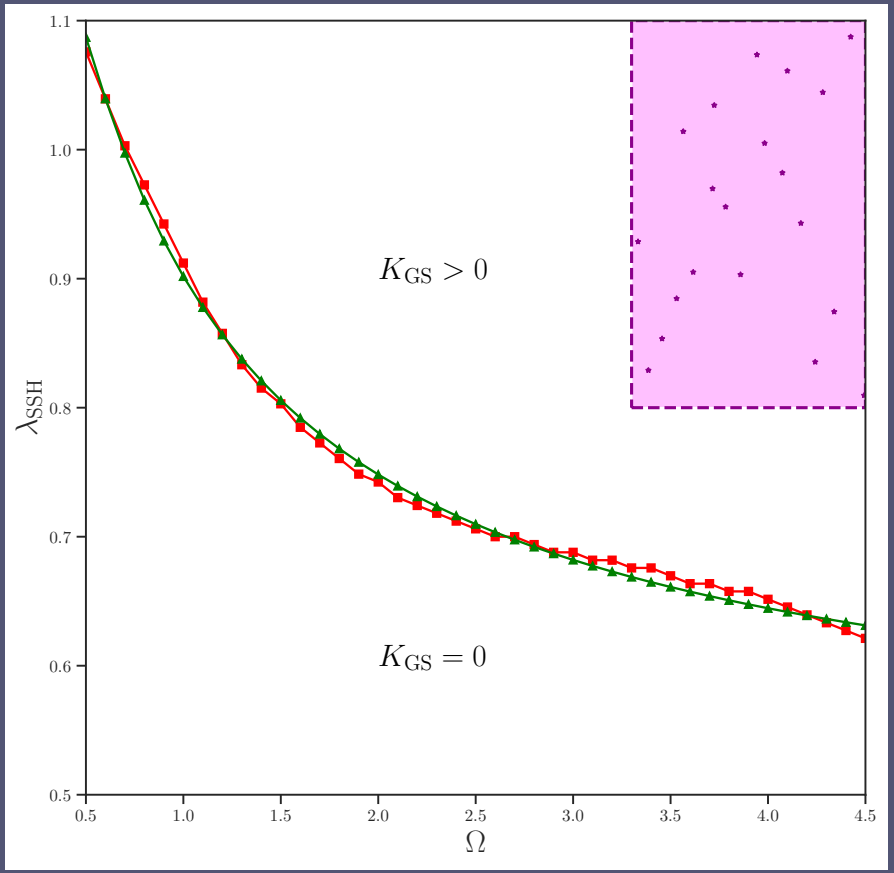
Felipe Herrera, Kirk Madison, RK, Mona Berciu, *Phys. Rev. Lett.* 110, 223002 (2013)

Generalized polaron model



Felipe Herrera, Kirk Madison, RK, Mona Berciu, *Phys. Rev. Lett. 110, 223002 (2013)*

Extrapolation from a smaller Hilbert space to a bigger Hilbert space:



Across phases

From high phonon frequencies to low phonon frequencies

Pranav Kairon, J. Sous, M. Berciu and RK, Phys Rev B (2024)

Summary of key points so far:

ML predictions can be improved

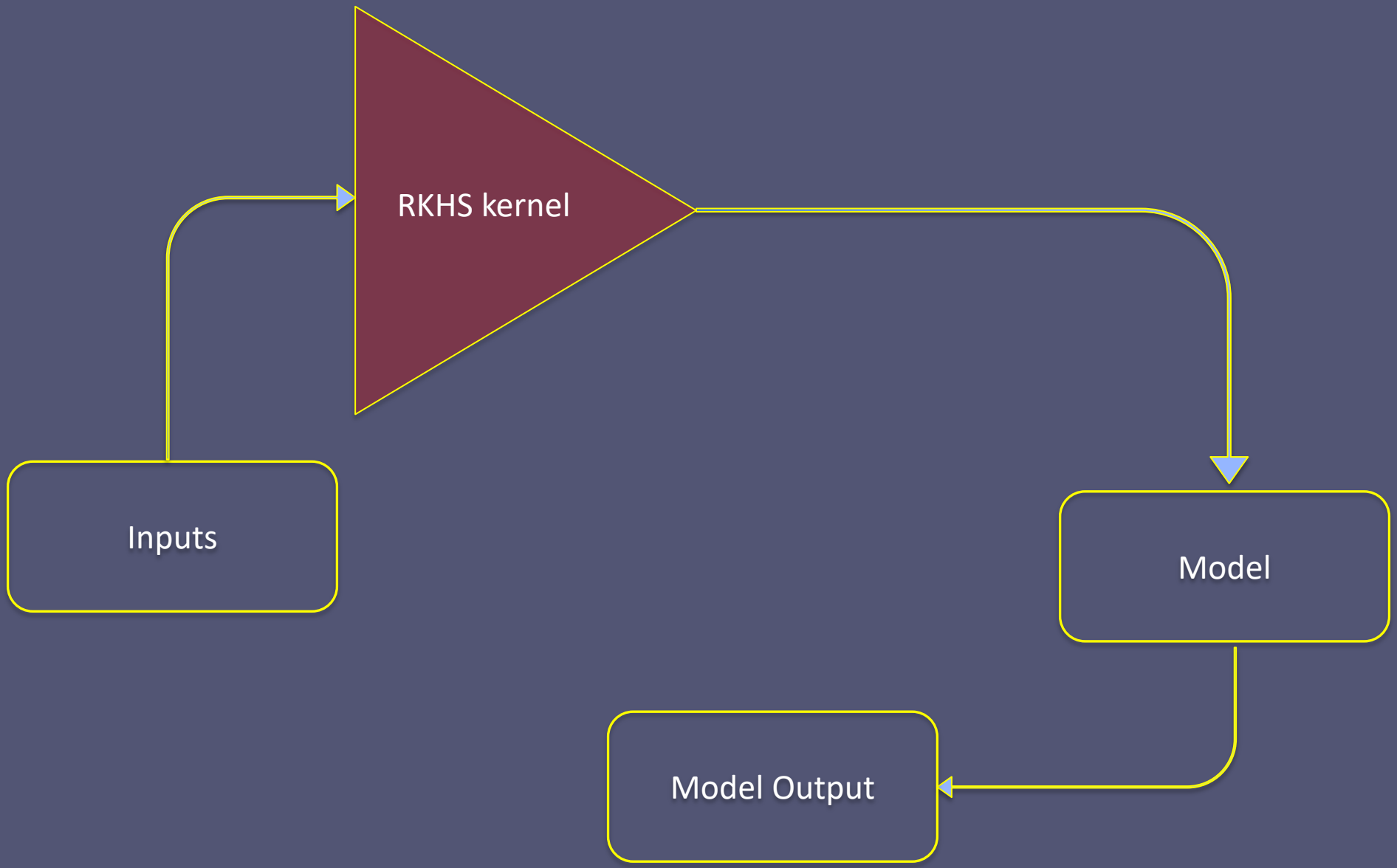
either by increasing the amount of training data

or by aligning RKHS kernels with fixed, limited data

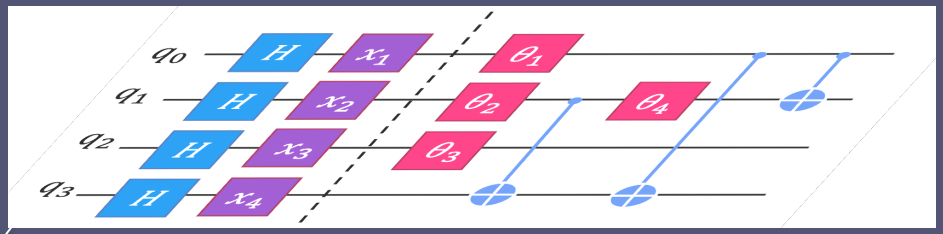
The Bayesian approach can improve kernels for machines to learn "more from less"

We will now use the same approach to build quantum kernels that outperform classical kernels

Kernel methods of machine learning



Quantum machine learning

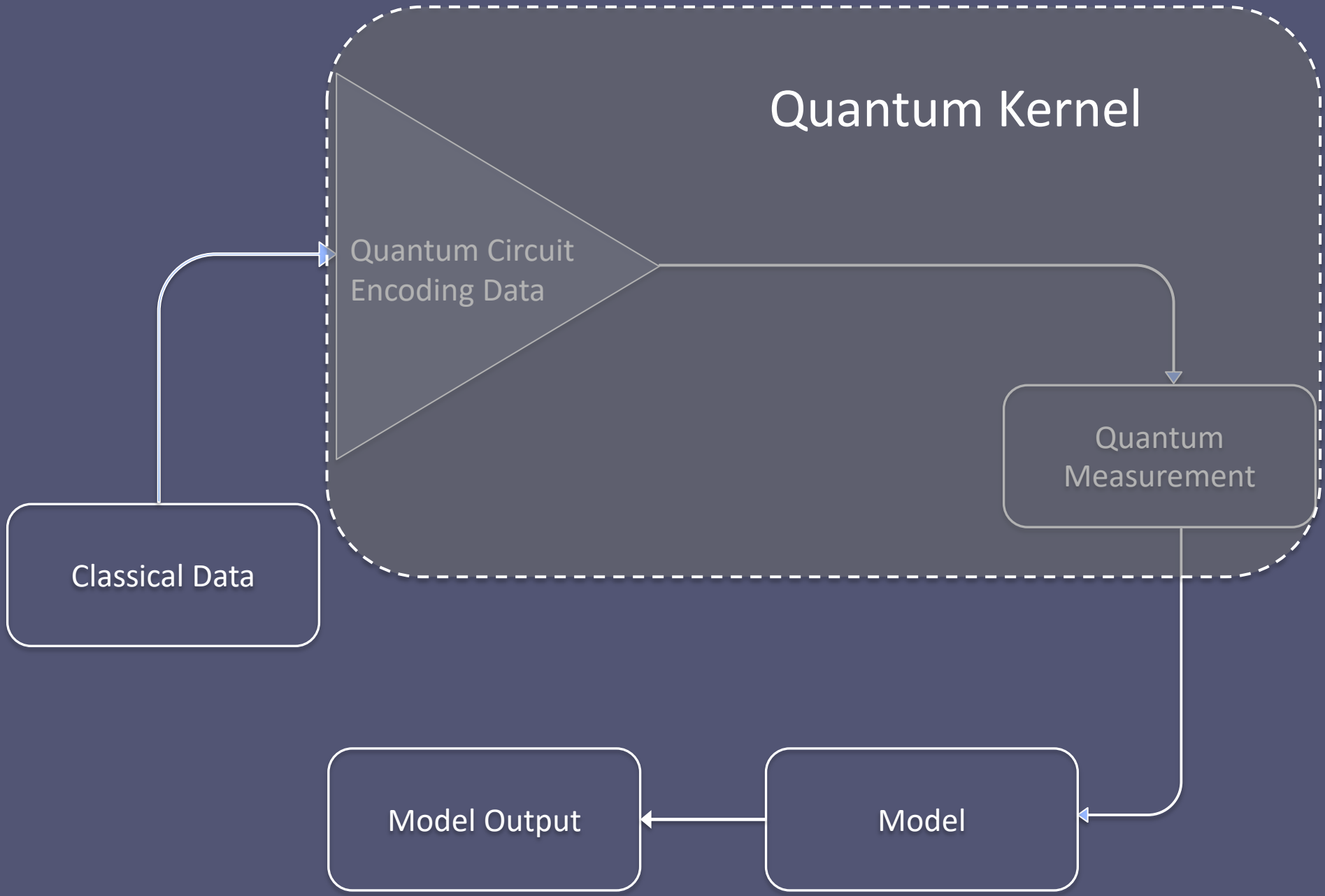


Quantum Circuit Encoding Data

Classical Inputs

Quantum Measurement

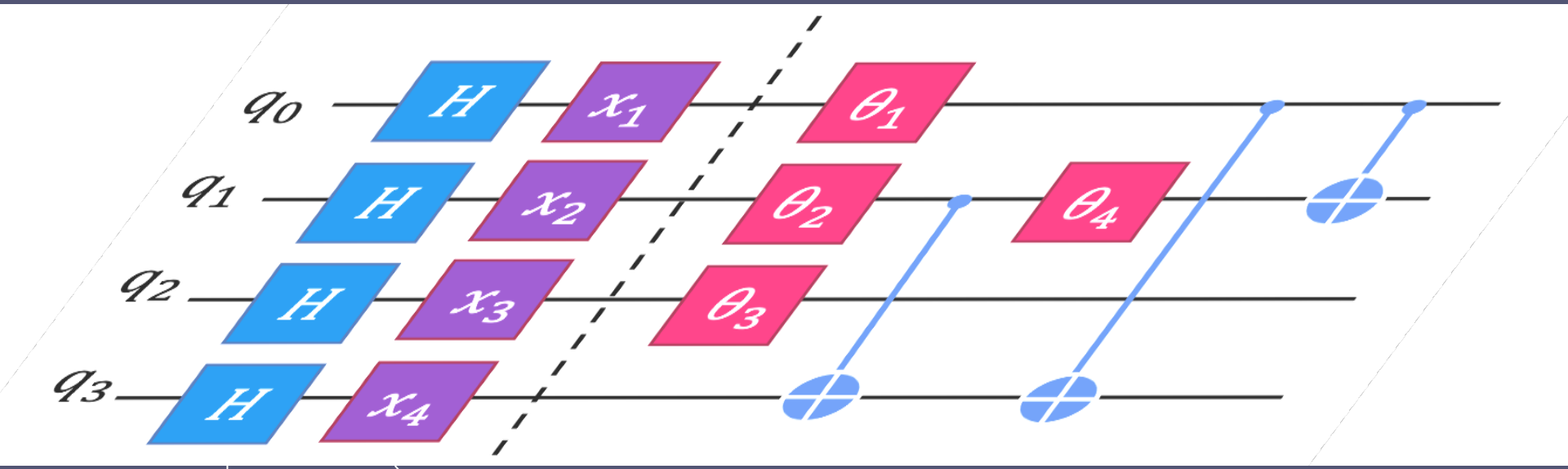
Output


Quantum Kernel
Quantum Circuit
Encoding Data
Quantum
Measurement
Classical Data
Model Output
Model

Consider a quantum computer with n qubits, initially in state $|0\rangle^{\otimes n}$.

Introduce a sequence of gates that produces a quantum state $\mathcal{U}(\boldsymbol{x})|0\rangle^{\otimes n}$

and another state $\mathcal{U}(\boldsymbol{x}')|0\rangle^{\otimes n}$



$$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \qquad \begin{pmatrix}e^{-i\boldsymbol{x}_i} & 0\\ 0 & e^{i\boldsymbol{x}_i}\end{pmatrix}$$

Consider a quantum computer with n qubits, initially in state $|0\rangle^{\otimes n}$.

Introduce a sequence of gates that produces a quantum state $\mathcal{U}(\boldsymbol{x})|0\rangle^{\otimes n}$

and another state $\mathcal{U}(\boldsymbol{x}')|0\rangle^{\otimes n}$

The measurable square of the inner product:

$$|\langle 0|^{\otimes n}\mathcal{U}^\dagger(\boldsymbol{x}')\mathcal{U}(\boldsymbol{x})|0\rangle^{\otimes n}|^2$$

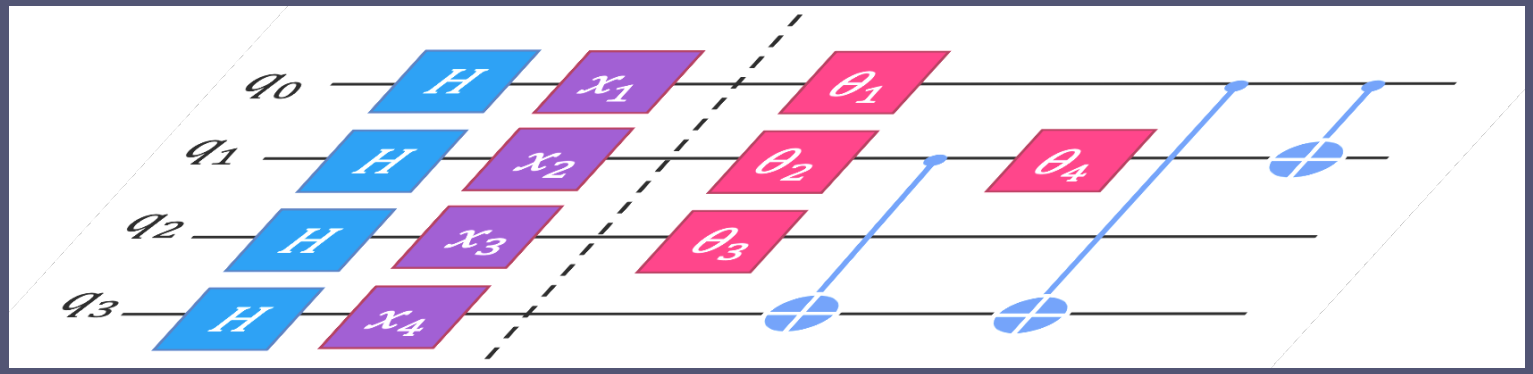
has all the properties of a kernel of an RKHS

Thus, projecting

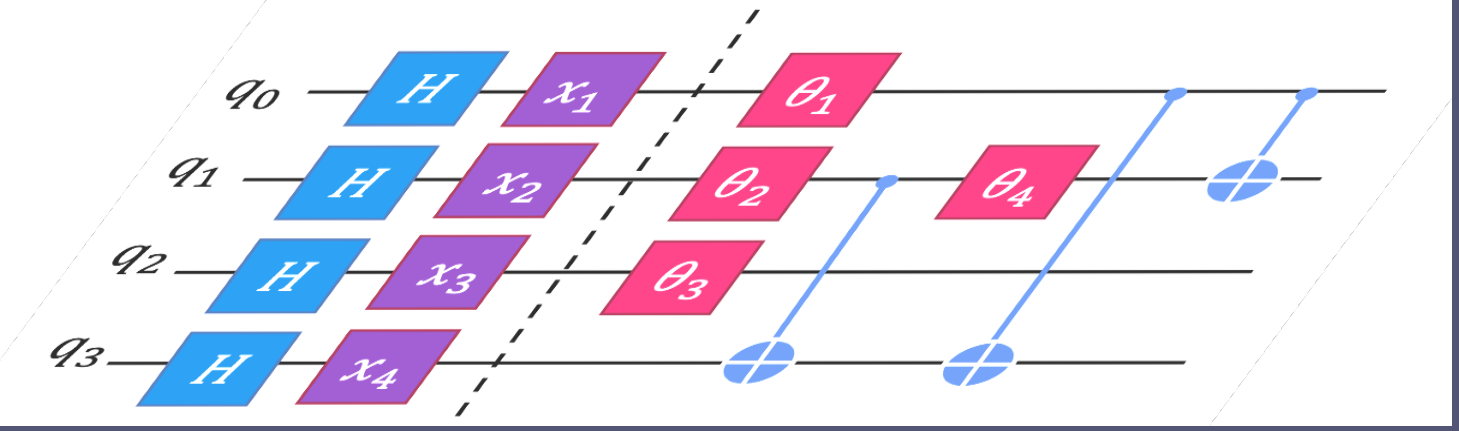
$$\mathcal{U}^\dagger(\boldsymbol{x}')\mathcal{U}(\boldsymbol{x})|0\rangle^{\otimes n} \quad \text{onto} \quad |0\rangle^{\otimes n}$$

can be another way of building kernels for kernel ML:

$$K(\boldsymbol{x}, \boldsymbol{x}') = |\langle 0|^{\otimes n}\mathcal{U}^\dagger(\boldsymbol{x}')\mathcal{U}(\boldsymbol{x})|0\rangle^{\otimes n}|^2$$



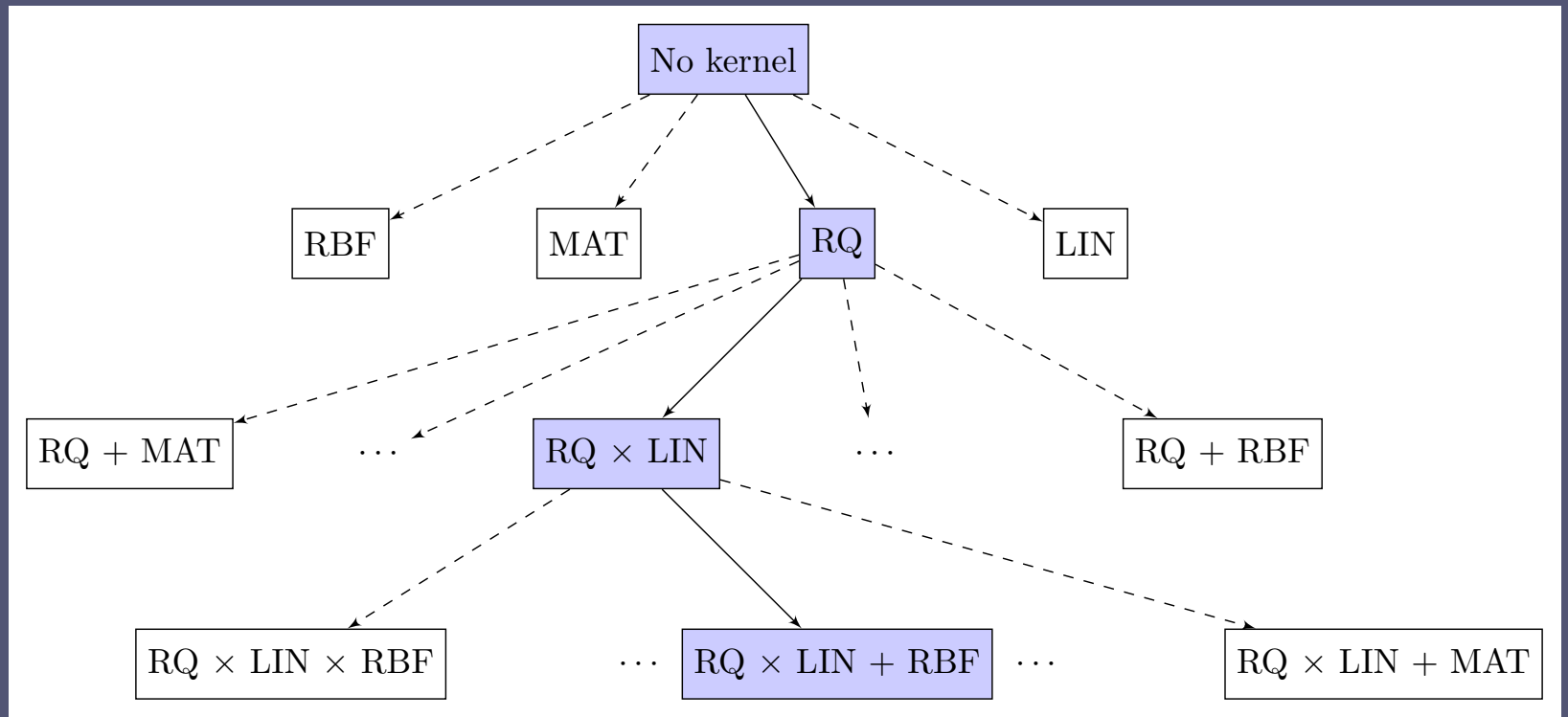
What is the best way to put the gates together to produce the most optimal kernel?

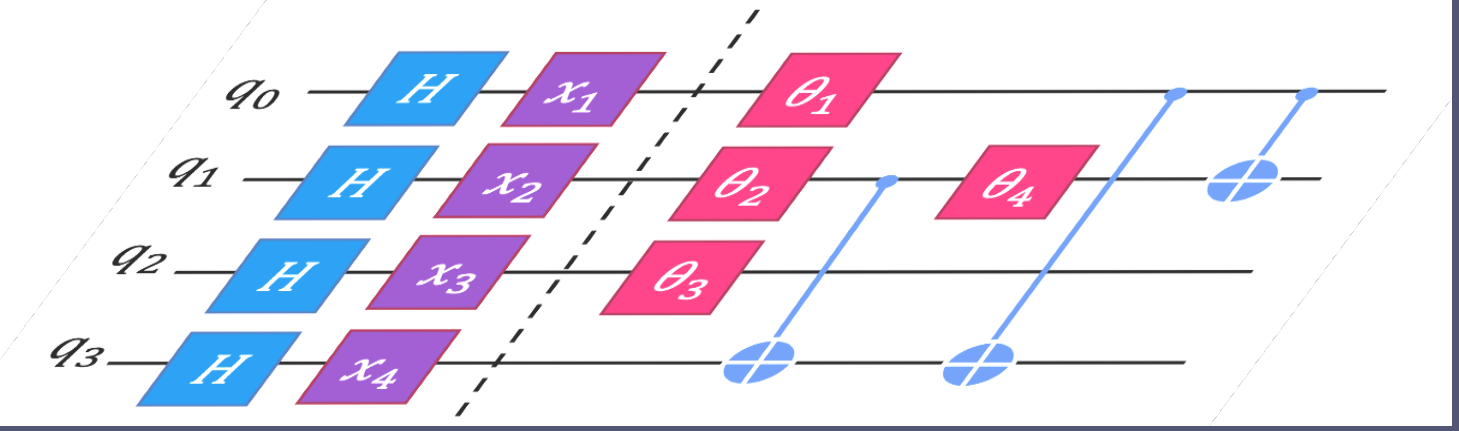


Compositional optimization of quantum circuits for quantum kernels of support vector machines

E Torabian, RV Krems

Physical Review Research 5 (1), 013211





Compositional optimization of quantum circuits for quantum kernels of support vector machines

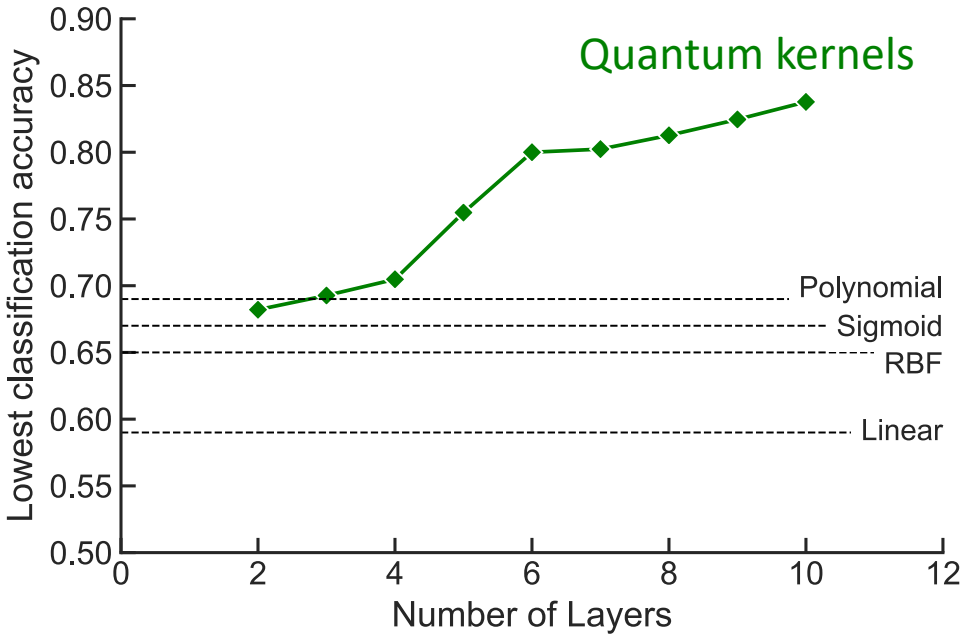
E Torabian, RV Krems

Physical Review Research 5 (1), 013211

Classification problem:

are perovskites A_2BBX_6 metals – Yes/No?

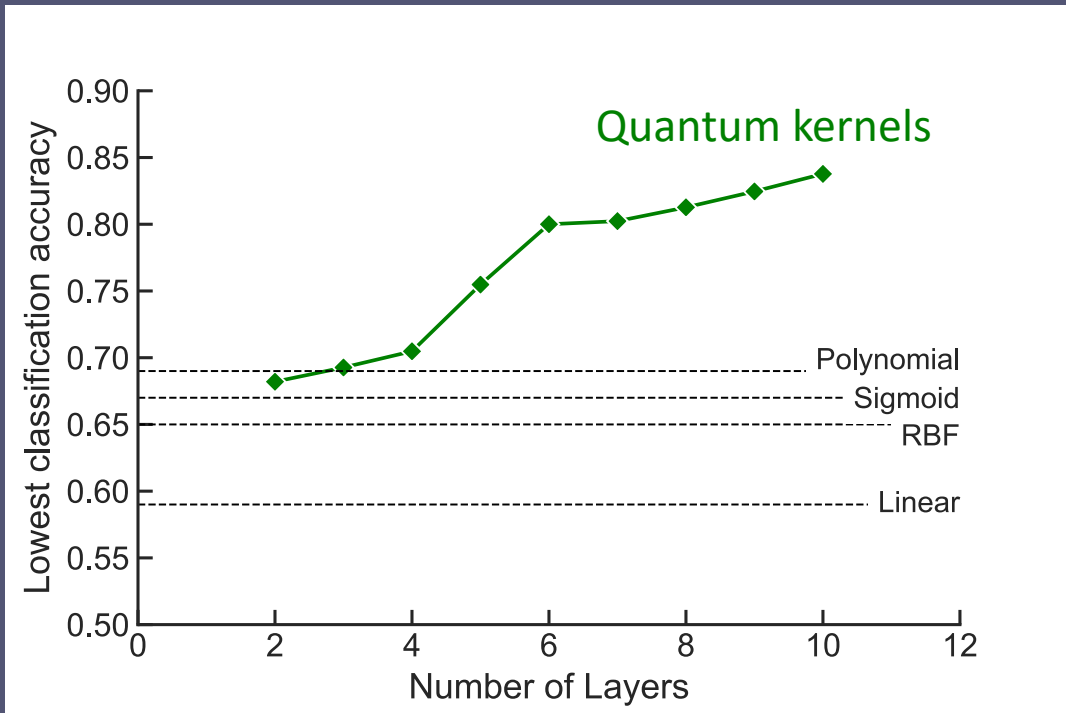
Qubits encode the size of the ions in a given perovskite



Quantum kernels can be constructed to offer powerful inference

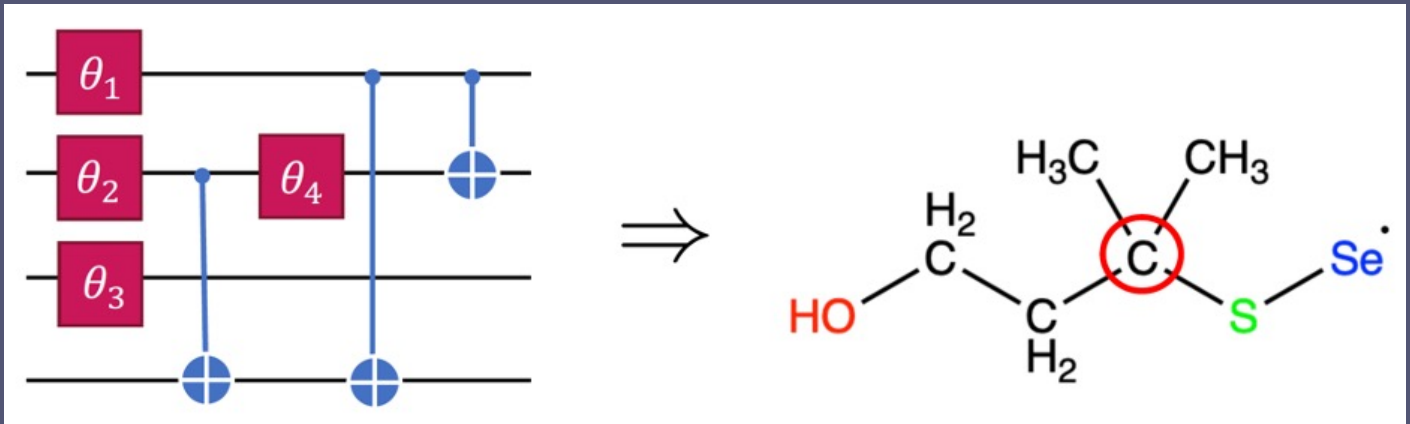
However… This algorithm is difficult to scale

Is there a another (better) way to build quantum circuits?



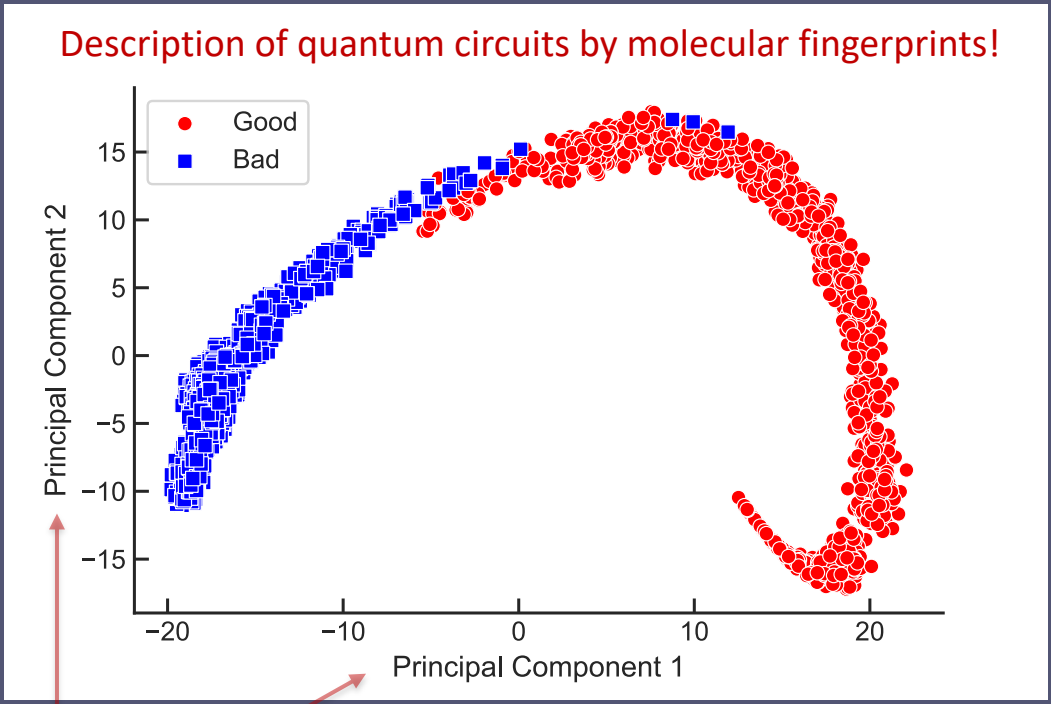
A radical idea:

Quantum circuits are isomorphic to polyatomic molecules!



A radical idea:

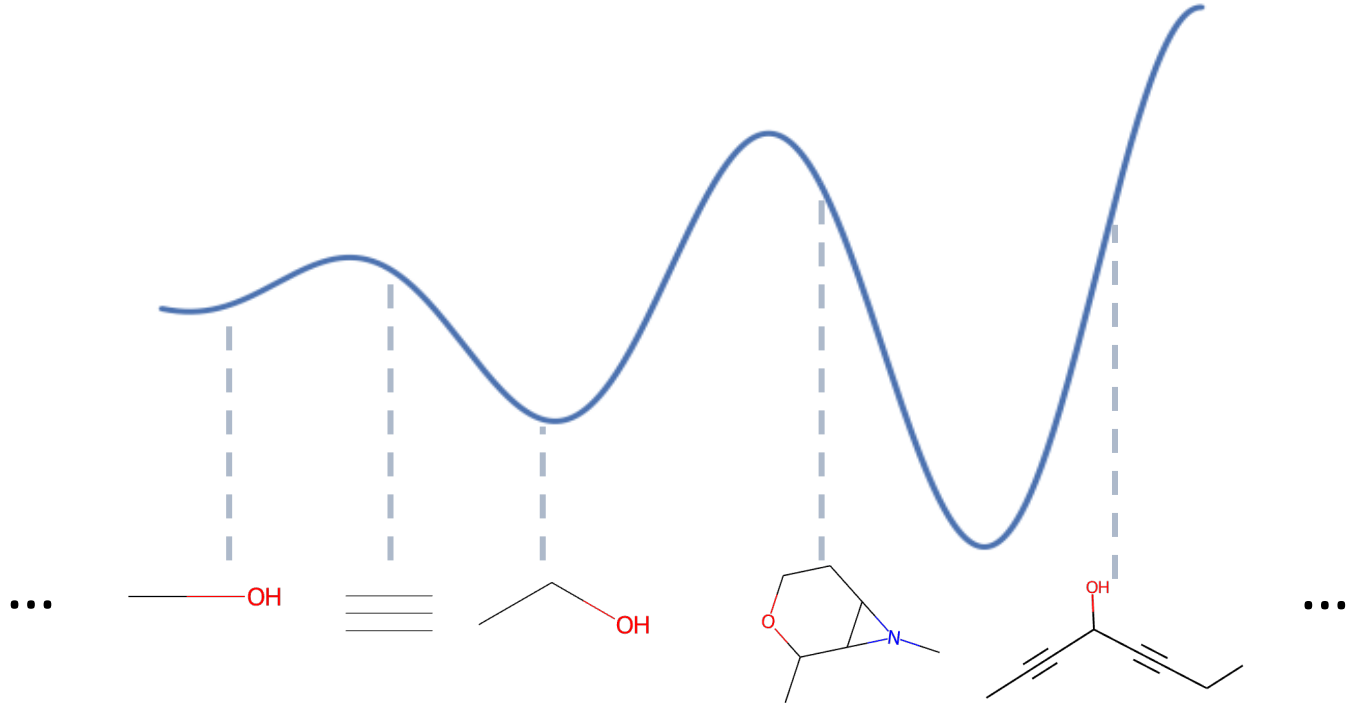
Quantum circuits are isomorphic to polyatomic molecules!



Principle components of molecular fingerprints from cheminformatics

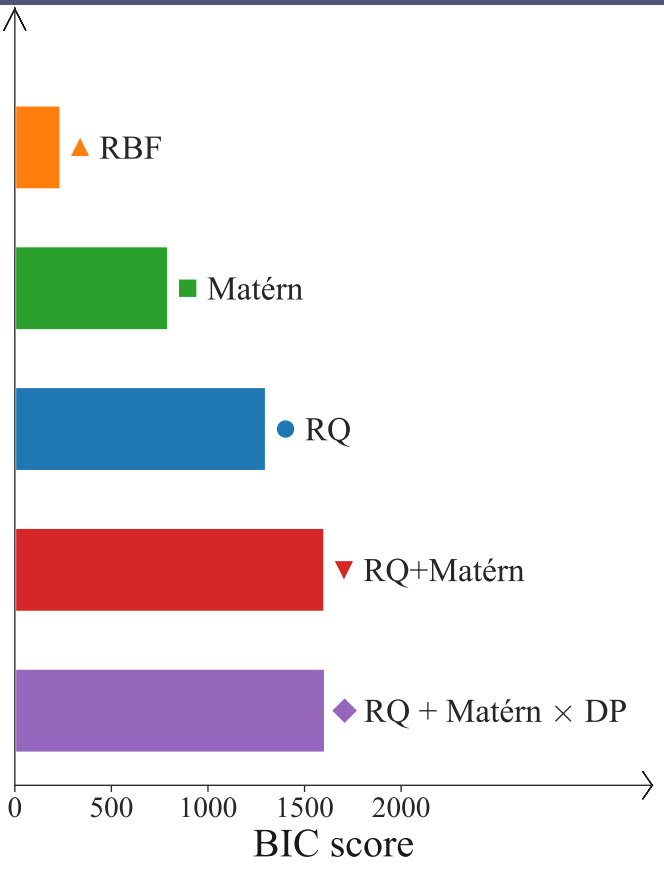
Why to simulate quantum circuits by molecules?

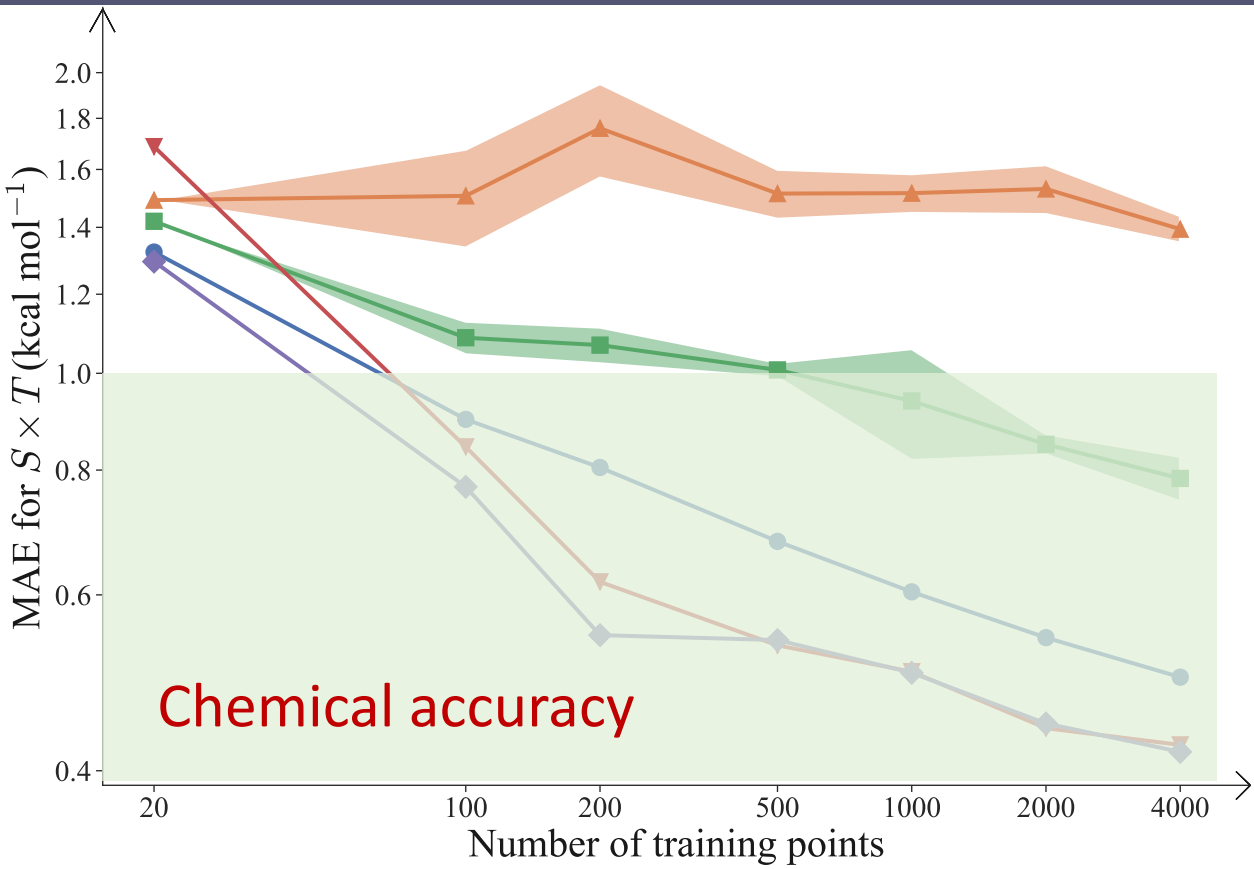
We know a lot about how to handle molecular compound spaces



Chemical subspace for drugs: **10^{23} to 10^{60}** molecules

Interpolation of entropy in the space of 133,000 molecules

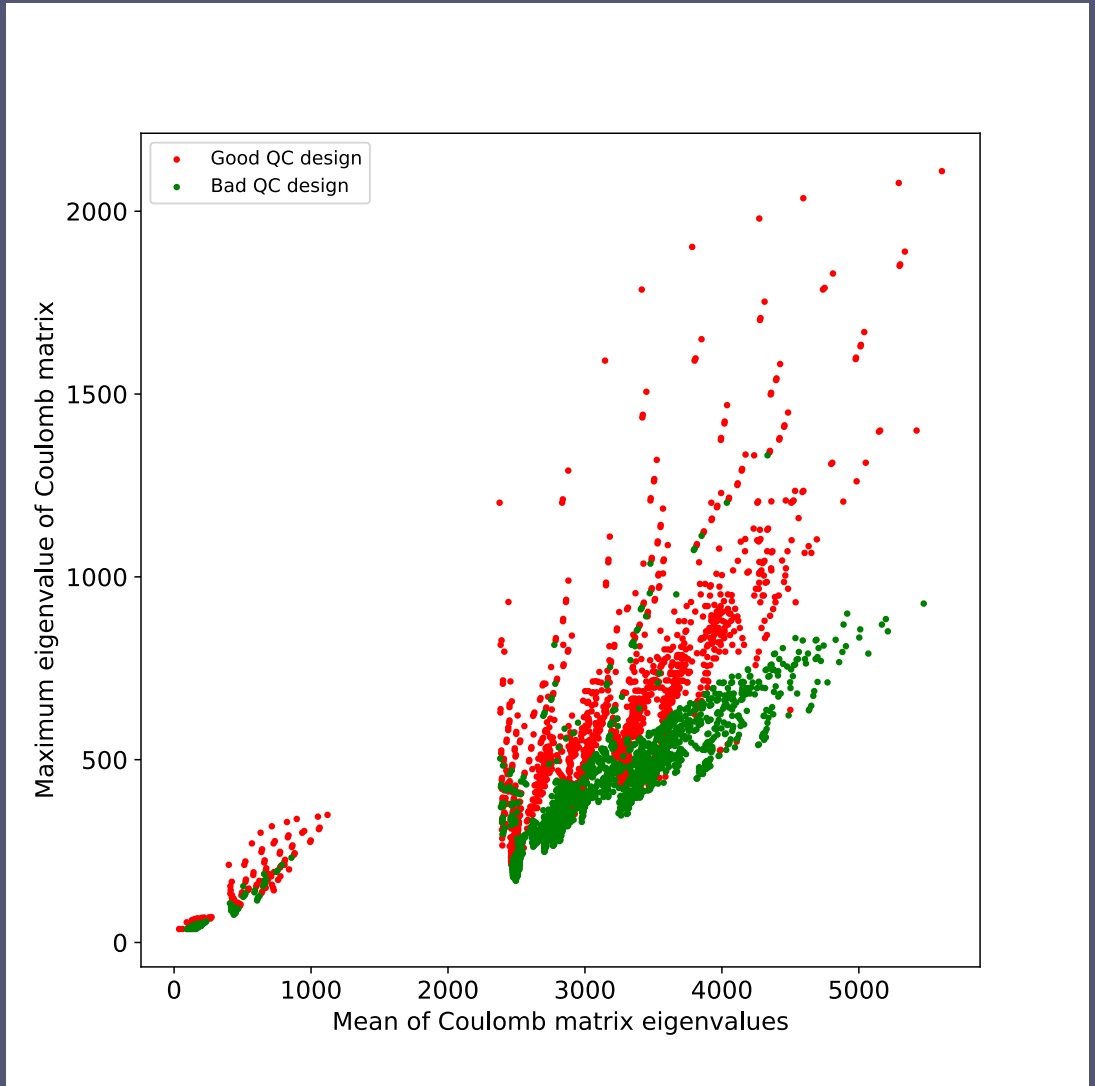




Dawn Mao and RK, Efficient interpolation of molecular properties across chemical compound space with low-dimensional descriptors, MLST (2024)

Molecules are physical objects

One can design effective physical descriptors of molecules



Compositional optimization of quantum circuits for quantum kernels of support vector machines

E Torabian, RV Krems

Physical Review Research 5 (1), 013211

Efficient interpolation of molecular properties across chemical compound space with low-dimensional descriptors

YW Mao, RV Krems

Machine Learning: Science and Technology 5 (1), 015059

Extrapolating quantum observables with machine learning: Inferring multiple phase transitions from properties of a single phase

RA Vargas-Hernández, J Sous, M Berciu, RV Krems

Physical review letters 121 (25), 255702

Extrapolation of polaron properties to low phonon frequencies by Bayesian machine learning

P Kairon, J Sous, M Berciu, RV Krems

Physical Review B 109 (14), 144523

Universal expressiveness of variational quantum classifiers and quantum kernels for support vector machines

J Jäger, RV Krems

Nature Communications 14 (1), 576