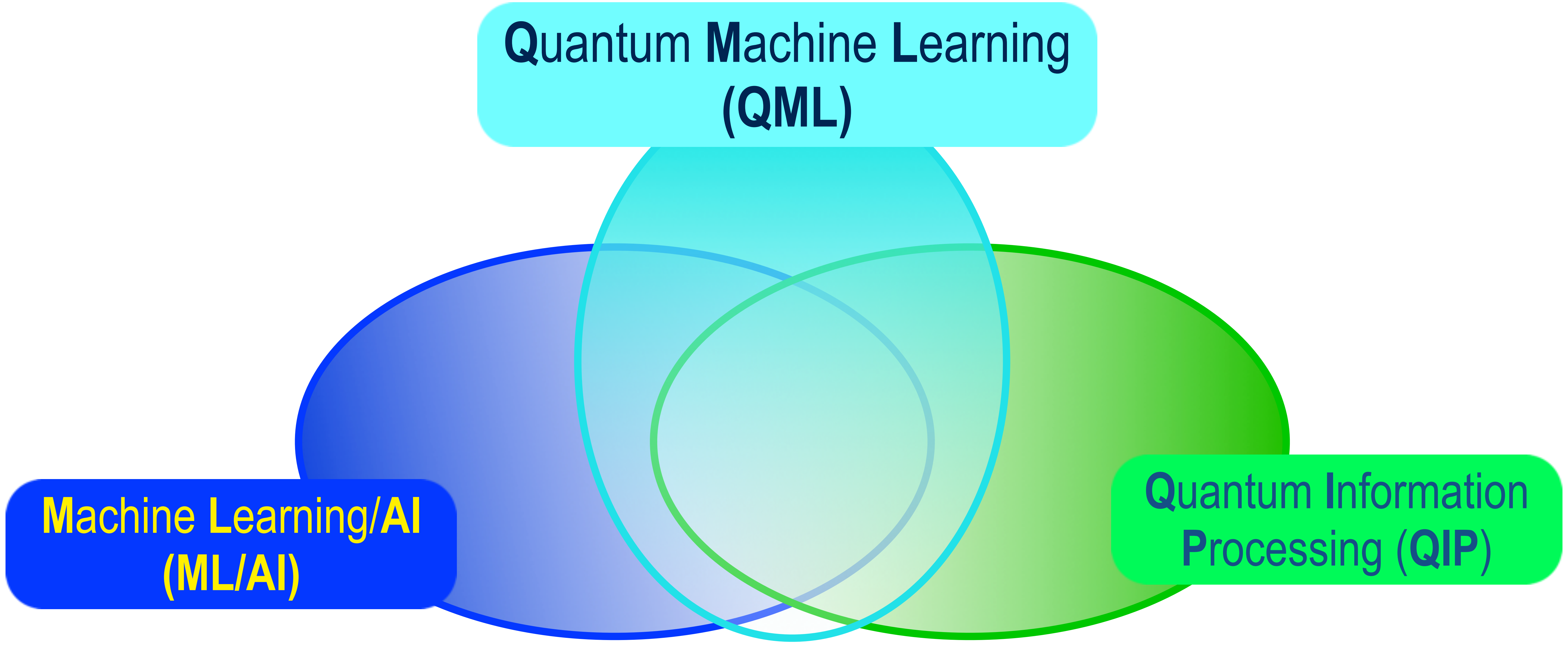
AI goes MAD

QML beyond kernels with smaller devices

Vedran Dunjko
applied Quantum algorithms Leiden
v.dunjko@liacs.leidenuniv.nl

- Very broad intro
- Two results

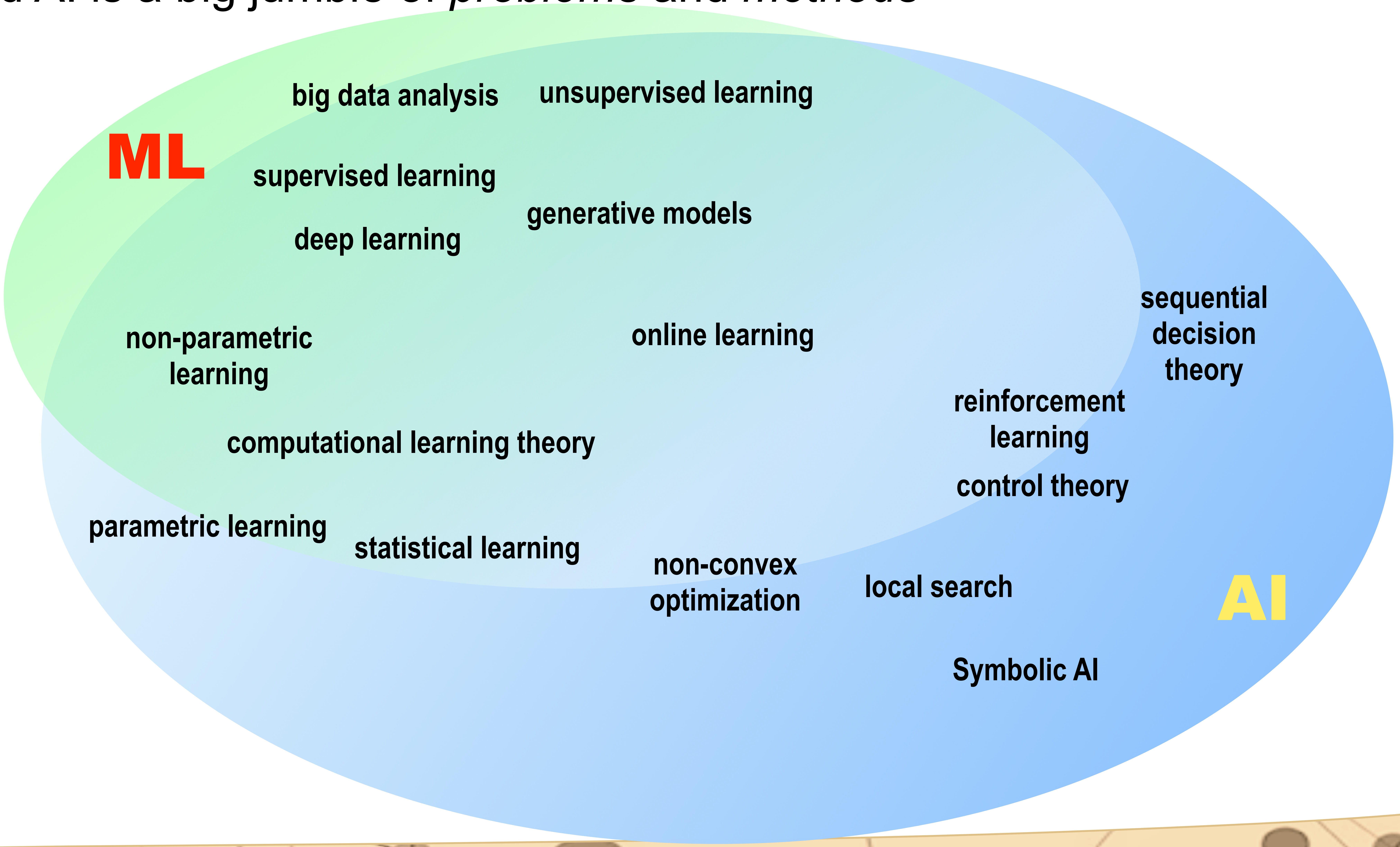


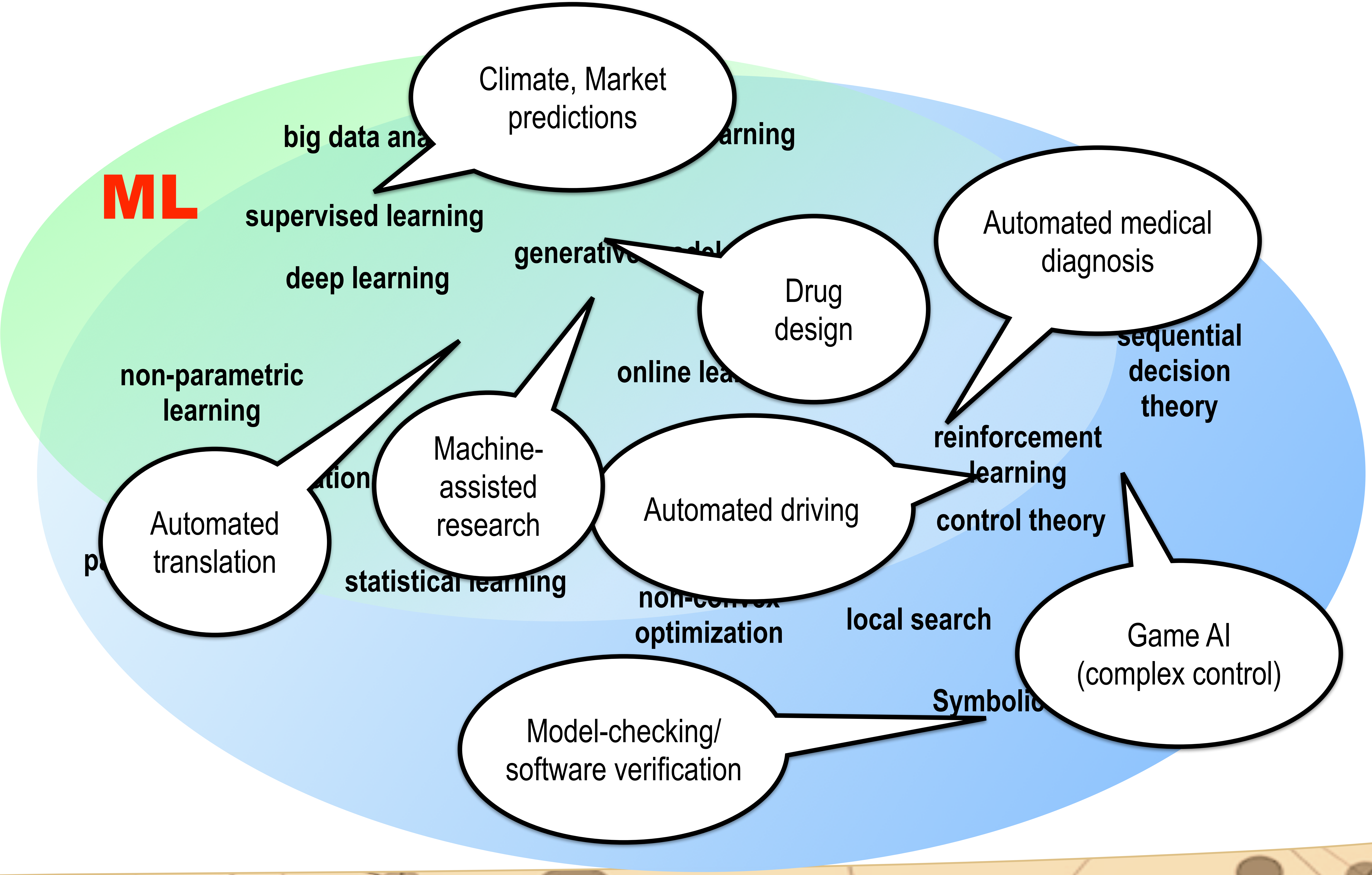
- **QIP→ML** (quantum-enhanced ML) ['94]
- **ML→QIP** (ML applied to QIP problems) ['74]
- **QIP↔ML** (generalizations of learning concepts) ['00]

- ML-inspired QM/QIP
- Physics inspired ML/AI
- *beyond (Q. AI)?*

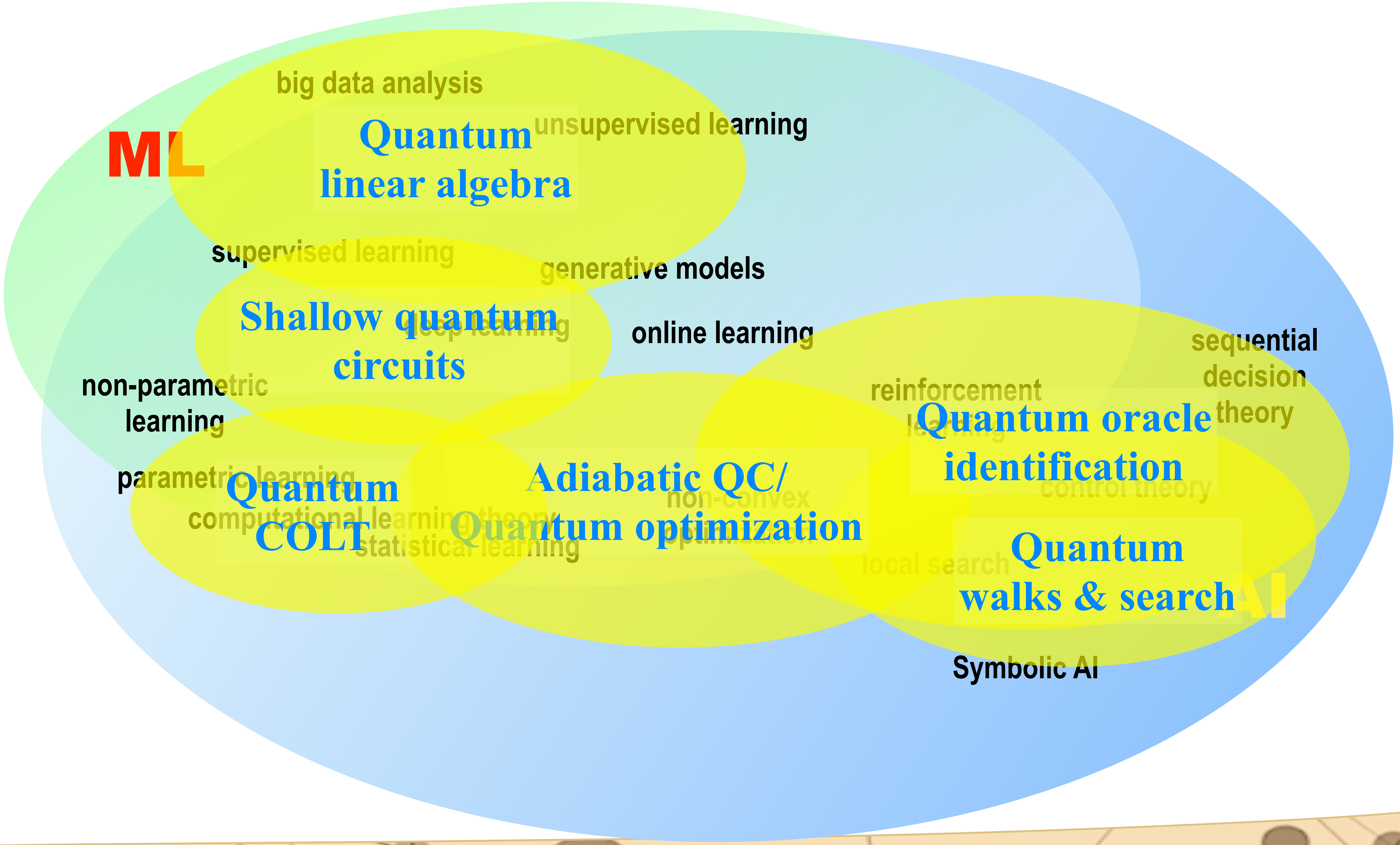
⟨aQa⟩

ML and AI is a big jumble of *problems* and *methods*




ML
Climate, Market predictions
supervised learning
deep learning
non-parametric learning
Drug design
Automated medical diagnosis
sequential decision theory
reinforcement learning
control theory
Machine-assisted research
Automated translation
Automated driving
statistical learning
optimization
local search
Game AI (complex control)
Model-checking/ software verification

Large playground to find new exciting QC applications

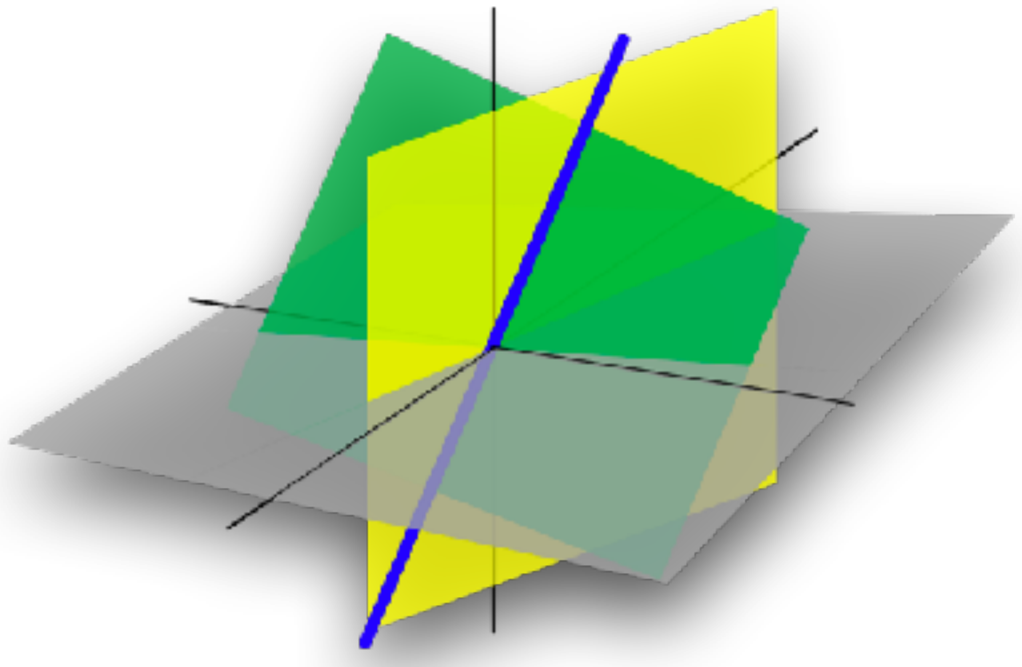


Why QML? Many ways quantum can help

- Speed-up training (optimization bottlenecks)

$$(\dot{\boldsymbol{w}}, \dot{b})^* = \arg\min_{\dot{\boldsymbol{w}},\dot{b}} \left\{ \frac{1}{S}\sum_{s=1}^{S} L_{\text{square}}\left(y_s(\dot{\boldsymbol{w}}^T \boldsymbol{x}_s + \dot{b})\right) + \lambda\|\dot{\boldsymbol{w}}\|_0 \right\}$$

- Linear-algebraic, big data (big data bottleneck)



https://www.nature.com/articles/nature24047

Near-term quantum computing: potential and limitations

The world this week

News in focus

The Sycamore chip is composed of 54 qubits, each made of superconducting loops.

GOOGLE PUBLISHES LANDMARK QUANTUM SUPREMACY CLAIM

The company says that its quantum computer is the first to perform a calculation that would be practically impossible for a classical machine.

By Elizabeth Gibney

Scientists at Google say that they have achieved quantum supremacy, a long-awaited milestone in quantum computing. The announcement, published

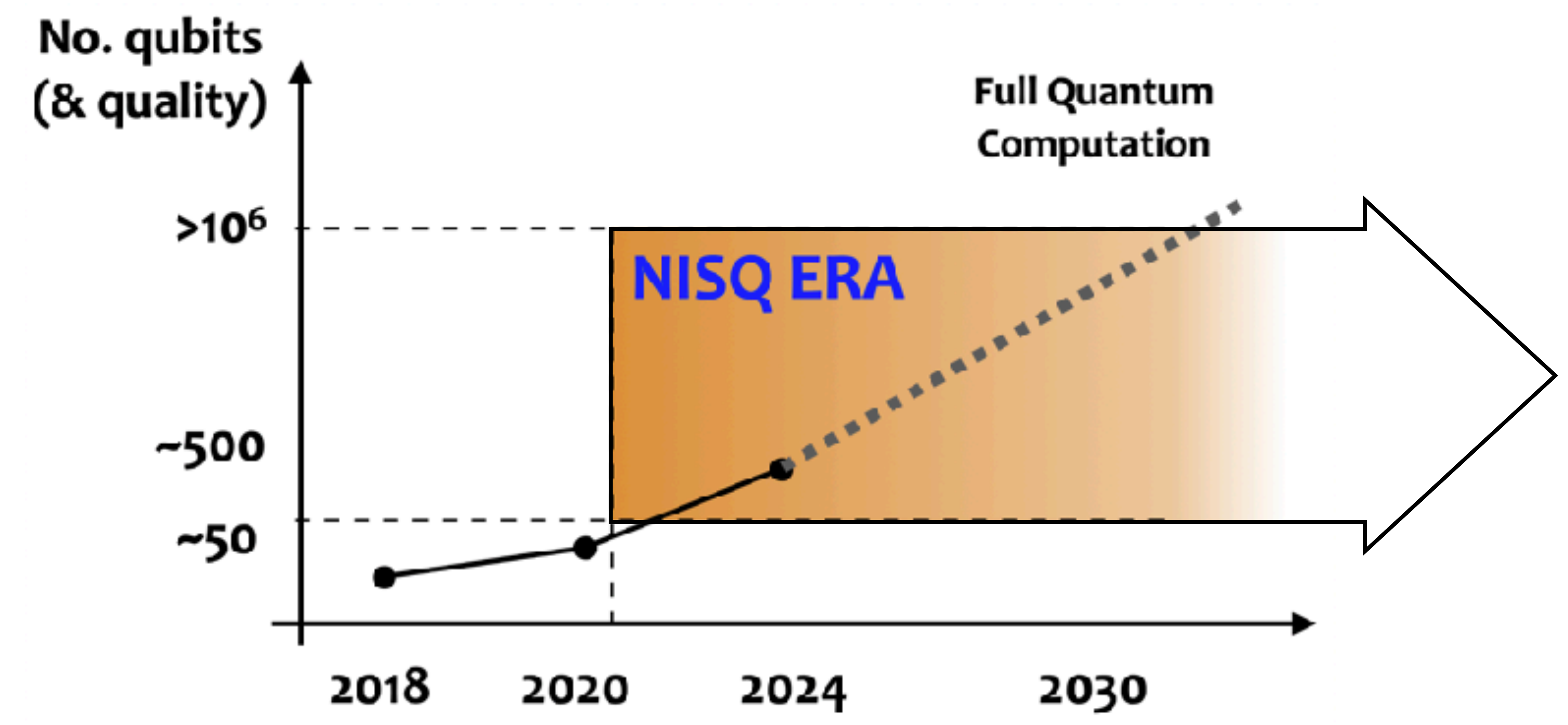
classical machines (F. Arute *et al. Nature* **574**, 505–510; 2019). The same calculation would take even the best classical supercomputer 10,000 years to complete, Google estimates.

Quantum supremacy has long been seen as a milestone because it proves that quantum

Michelle Simmons, a quantum physicist at the University of New South Wales in Sydney, Australia.

The feat was first reported in September by the *Financial Times* and other outlets, after an early version of the paper was leaked on

Noisy Intermediate-Scale Quantum (NISQ) Era



- Size-limited
- Architecture limited
- Noisy (error-prone)

➡ Hard to “program”; no useful applications

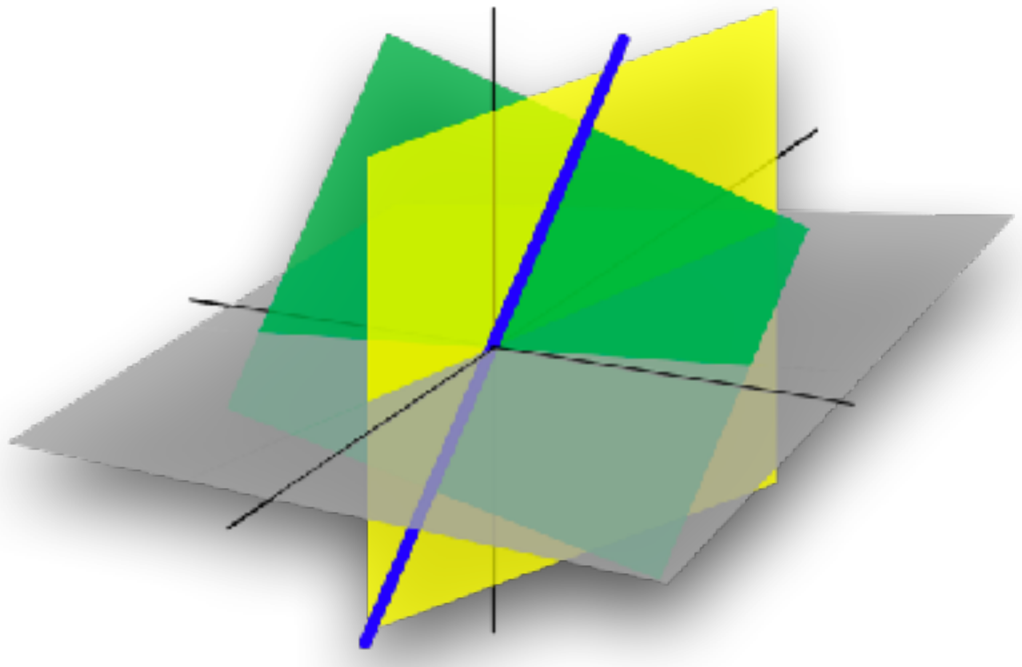
Yet powerful!

Why QML? Many ways quantum can help

- Speed-up training (optimization bottlenecks)

$$(\dot{\boldsymbol{w}}, \dot{b})^* = \arg\min_{\dot{\boldsymbol{w}},\dot{b}} \left\{ \frac{1}{S}\sum_{s=1}^{S} L_{\text{square}}\left(y_s(\dot{\boldsymbol{w}}^T \boldsymbol{x}_s + \dot{b})\right) + \lambda\|\dot{\boldsymbol{w}}\|_0 \right\}$$

- Linear-algebraic, big data (big data bottleneck)



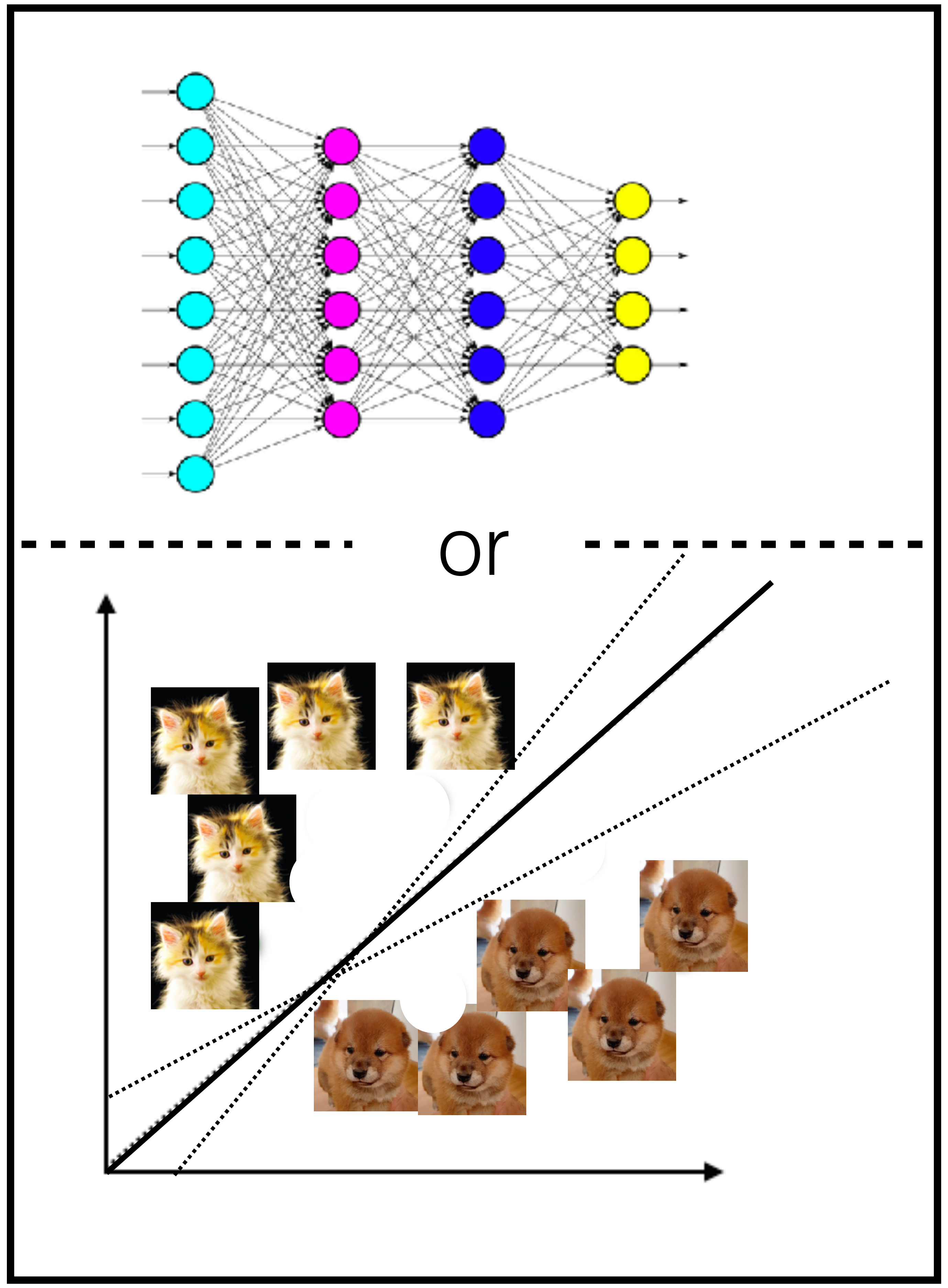
- “Genuinely quantum models”, or
  Near-term-QC-motivated approaches
  (not speed-ups… better *quality*)

https://www.nature.com/articles/nature24047

Models in machine learning?

Supervised learning

or

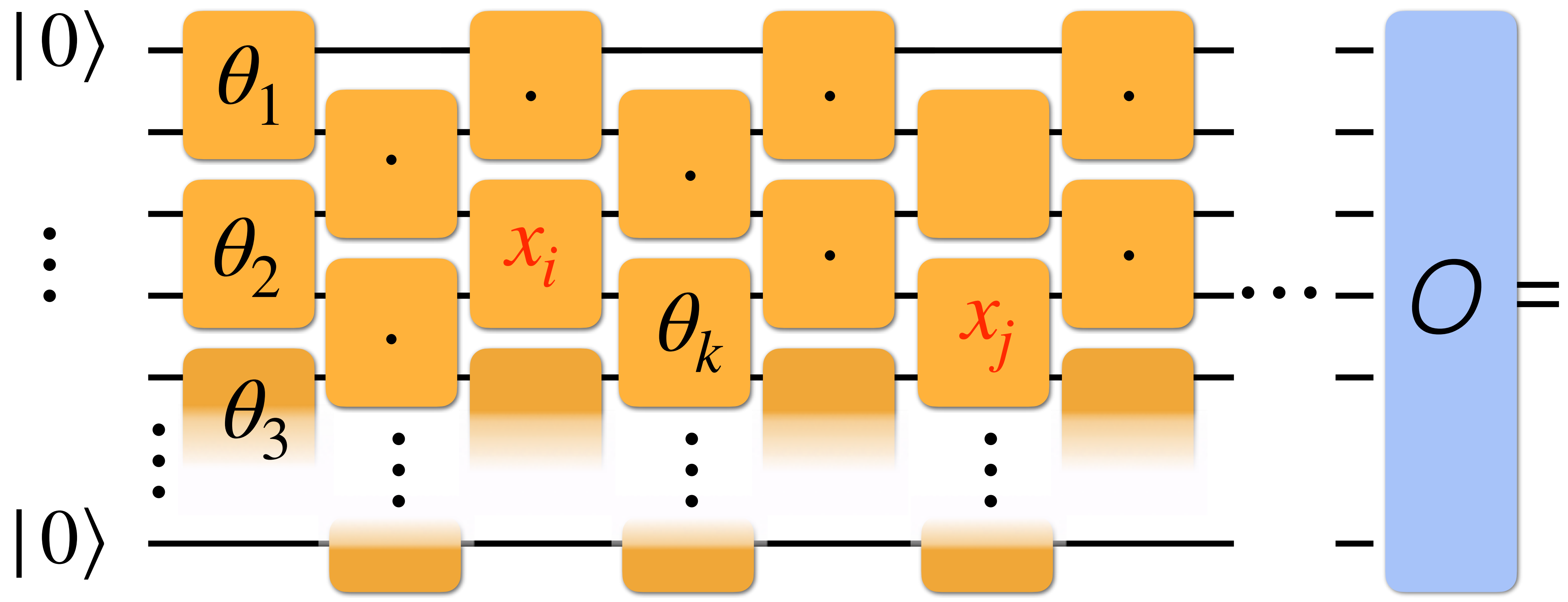


cat!

Model = hypothesis family or family of functions/distributions

Parametrized quantum circuits as machine learning models

(variational quantum circuits)



$$\Rightarrow \quad f_{\vec{\theta}}(\vec{x}) = \mathrm{Tr}[\rho(\vec{\theta}, \vec{x})O(\cdot)]$$

Typically:

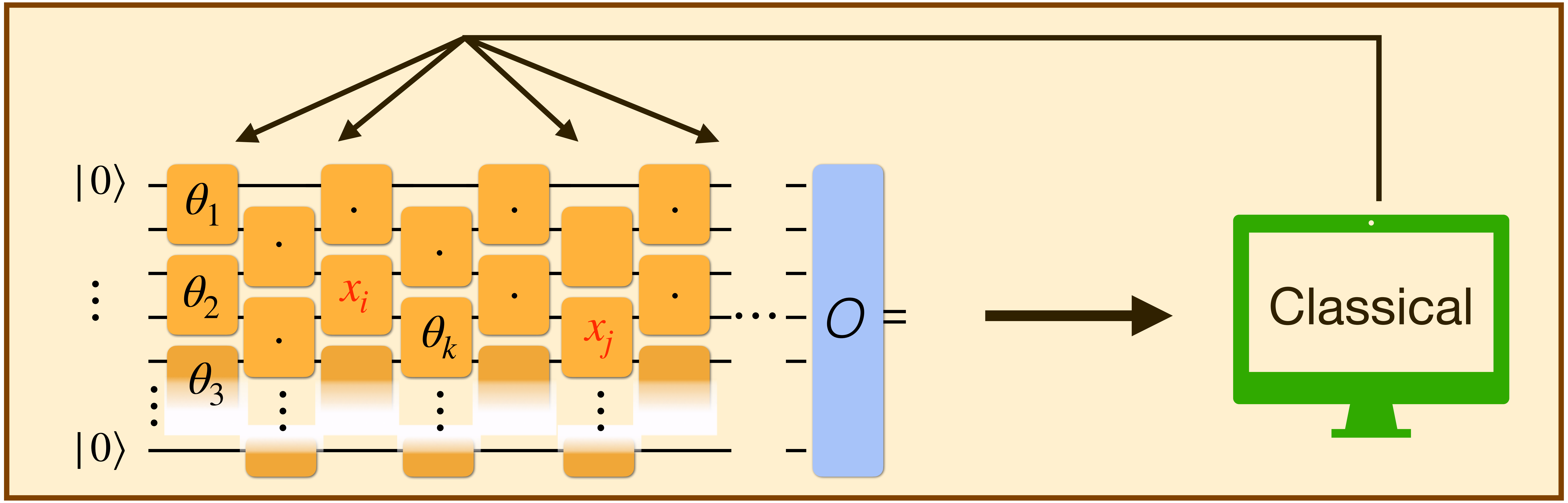
input = sets the parameters of some gates

trainable parameters = settings of other gates

output = expectation values of observable / measurement output

Parametrized function; hypothesis class; machine learning ***model***;

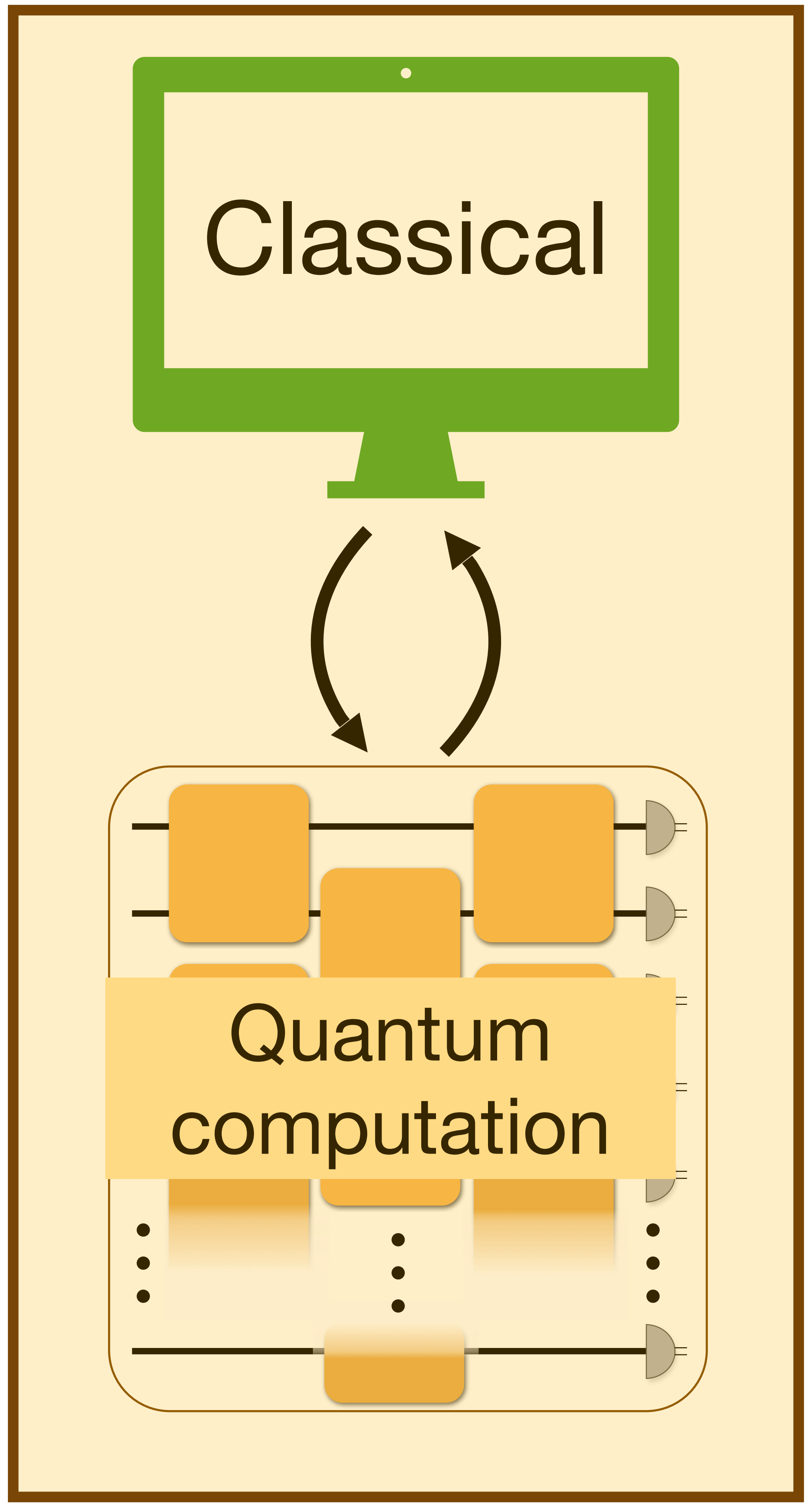
Parametrized quantum circuits as machine learning models



tune $\vec{\theta}s$, *test x, as to minimize error on training set*

⟨aQa⟩

Not just quantum cats v. quantum dogs



ML: supervised, generative, reinforcement learning

Combinatorial optimization (QAOA)

Quantum chemistry & many-body (variational ground states)

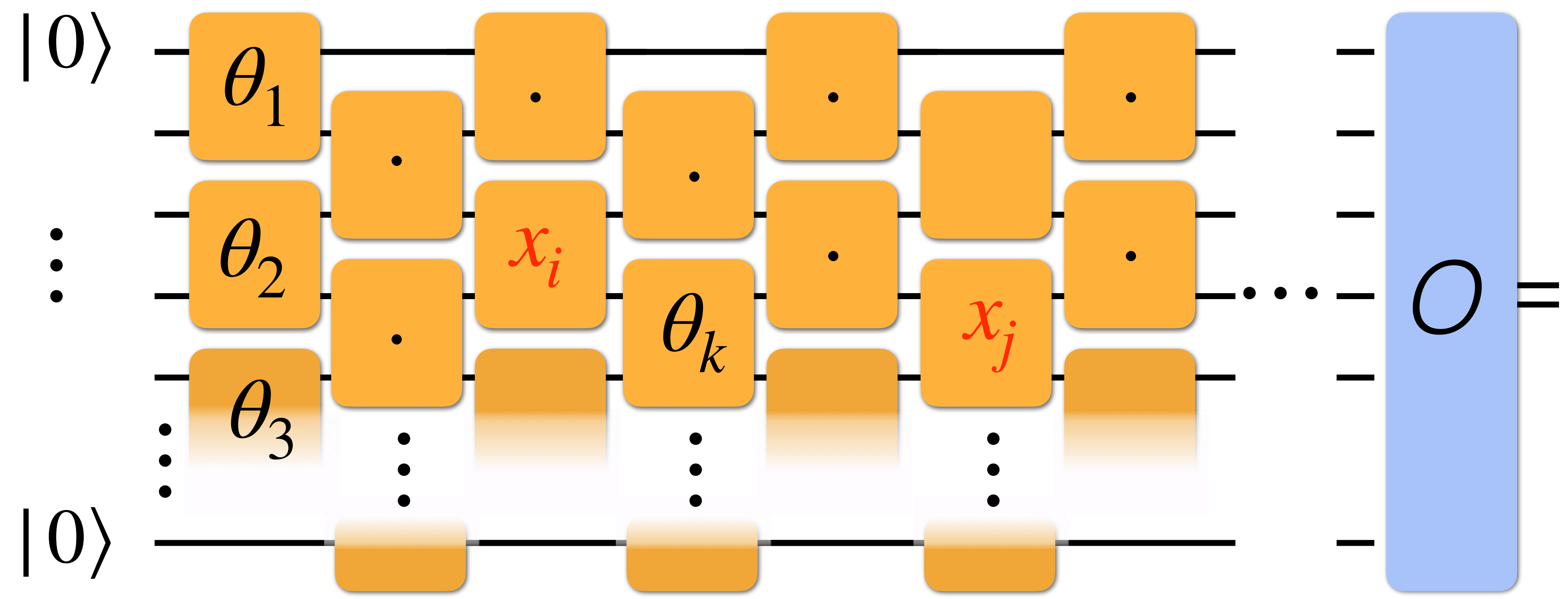
Differential equations (numerical methods, variational) & finance

Linear systems

…

Parametrized quantum circuits as machine learning models

(variational quantum circuits)



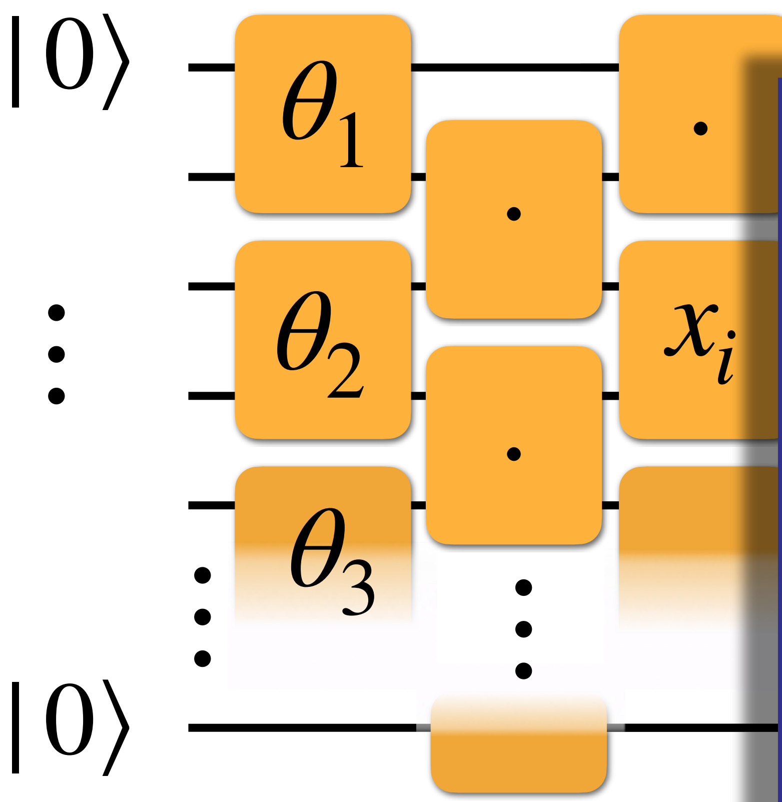
$$\Rightarrow \quad f_{\vec{\theta}}(\vec{x}) = \mathrm{Tr}[\rho(\vec{\theta}, \vec{x}) O(\cdot)]$$

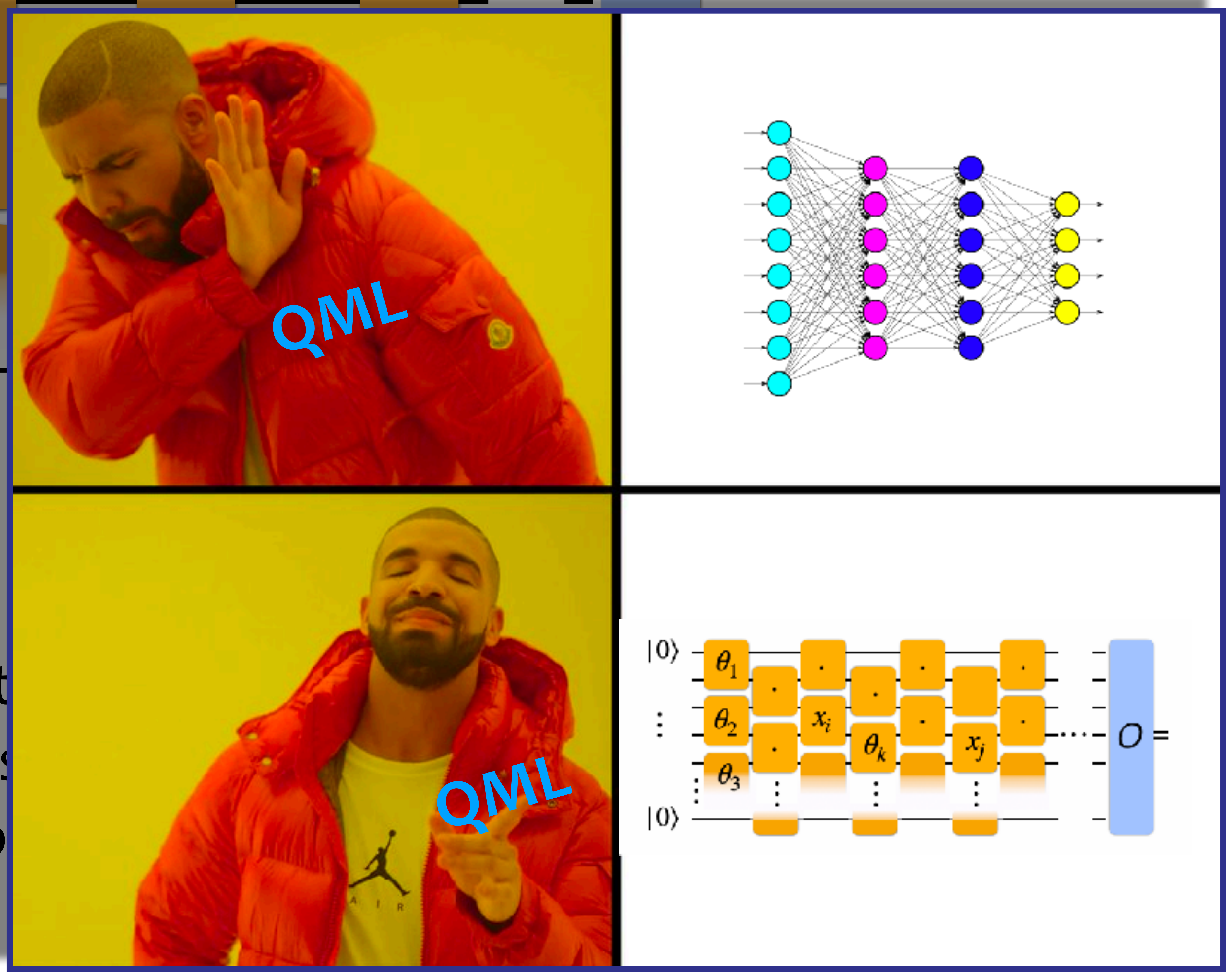
Basic concepts

- Supervised learning: classification (e.g. cats v. dogs)
- *Training error/empirical risk:* regularized error on training set: $RegRisk$ = Train_error(f,D) + Reg(f,D)
- *Total error:* on all data including unseen
- *Generalization performance:* (rate of decrease of) gap between training and total error

Parametrized quantum circuits as machine learning models

(variational quantum circuits)





$$= \mathrm{Tr}[\rho(\vec{\theta}, \vec{x})O(\cdot)]$$

Typically:

input = sets paramet

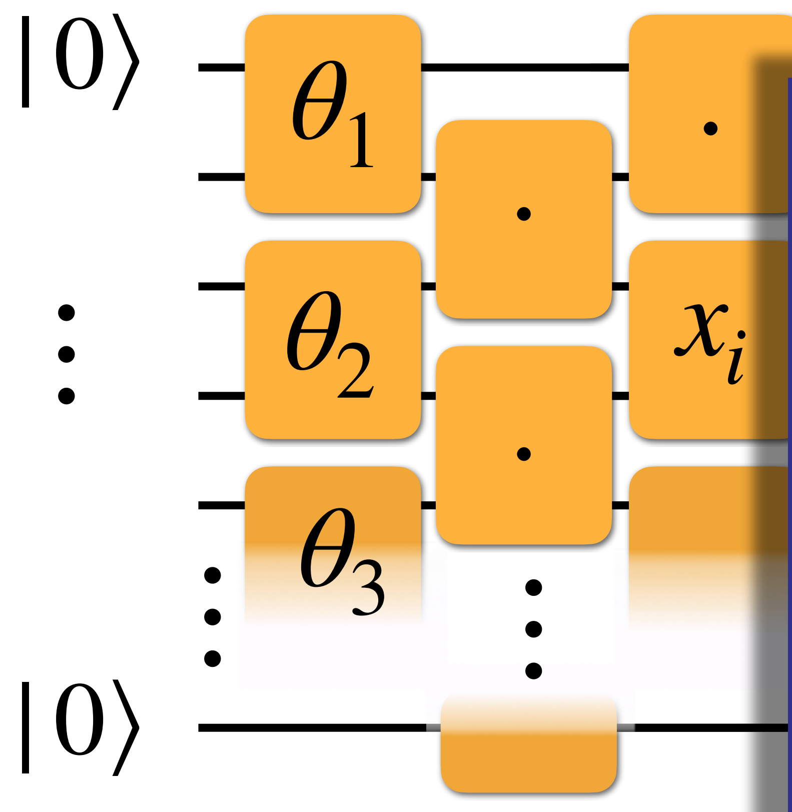
trainable parameter

output = expectatio

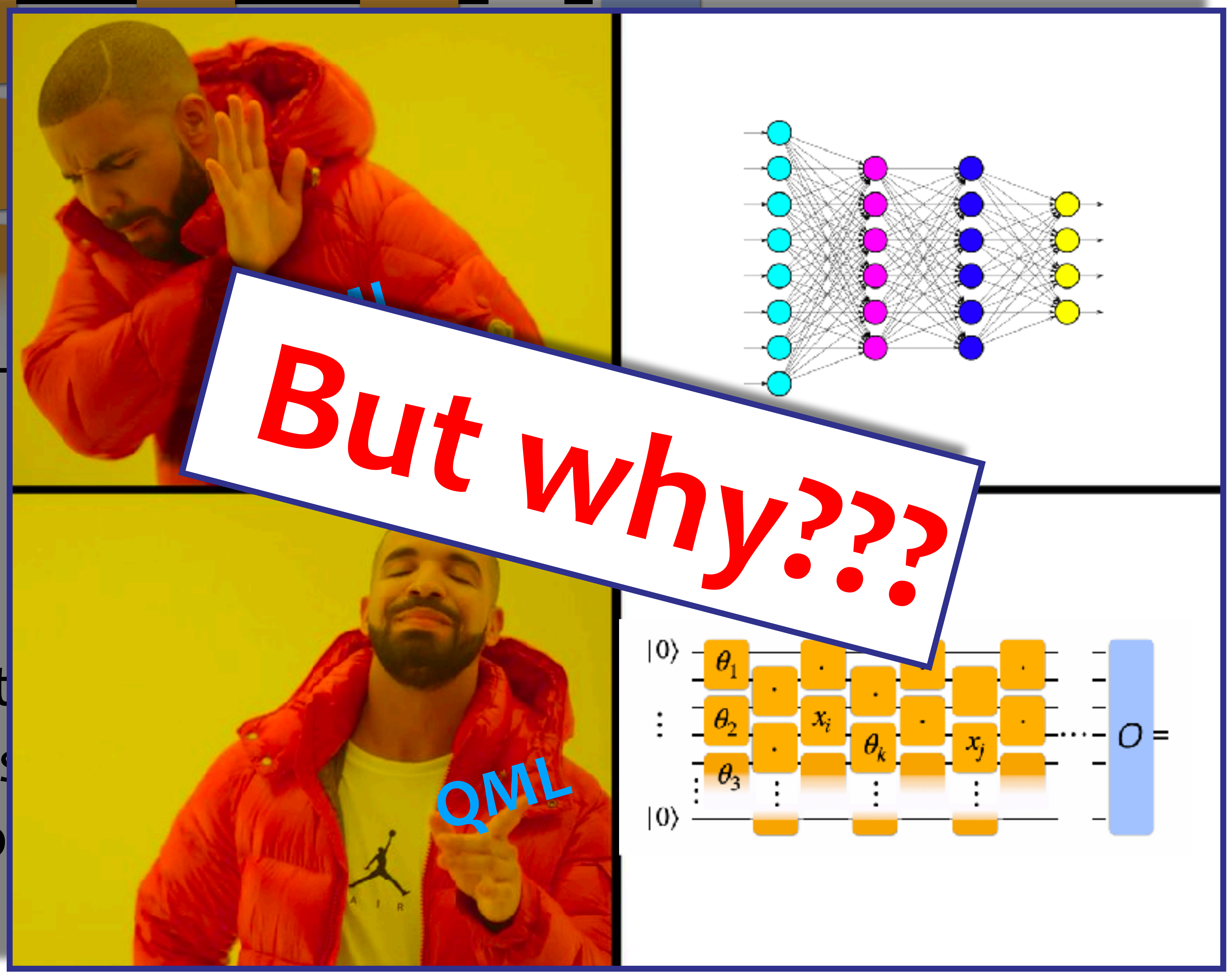
Parametrized function; hypothesis class; machine learning ***model***;

Parametrized quantum circuits as machine learning models

(variational quantum circuits)



$$= \mathrm{Tr}[\rho(\vec{\theta}, \vec{x})O(\cdot)]$$



Typically:

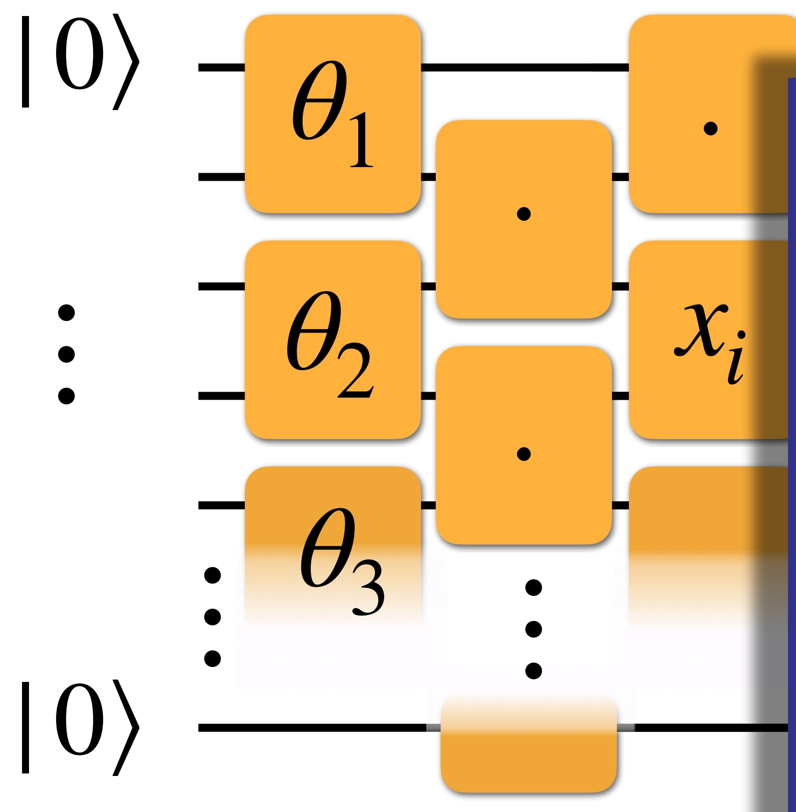
input = sets paramet

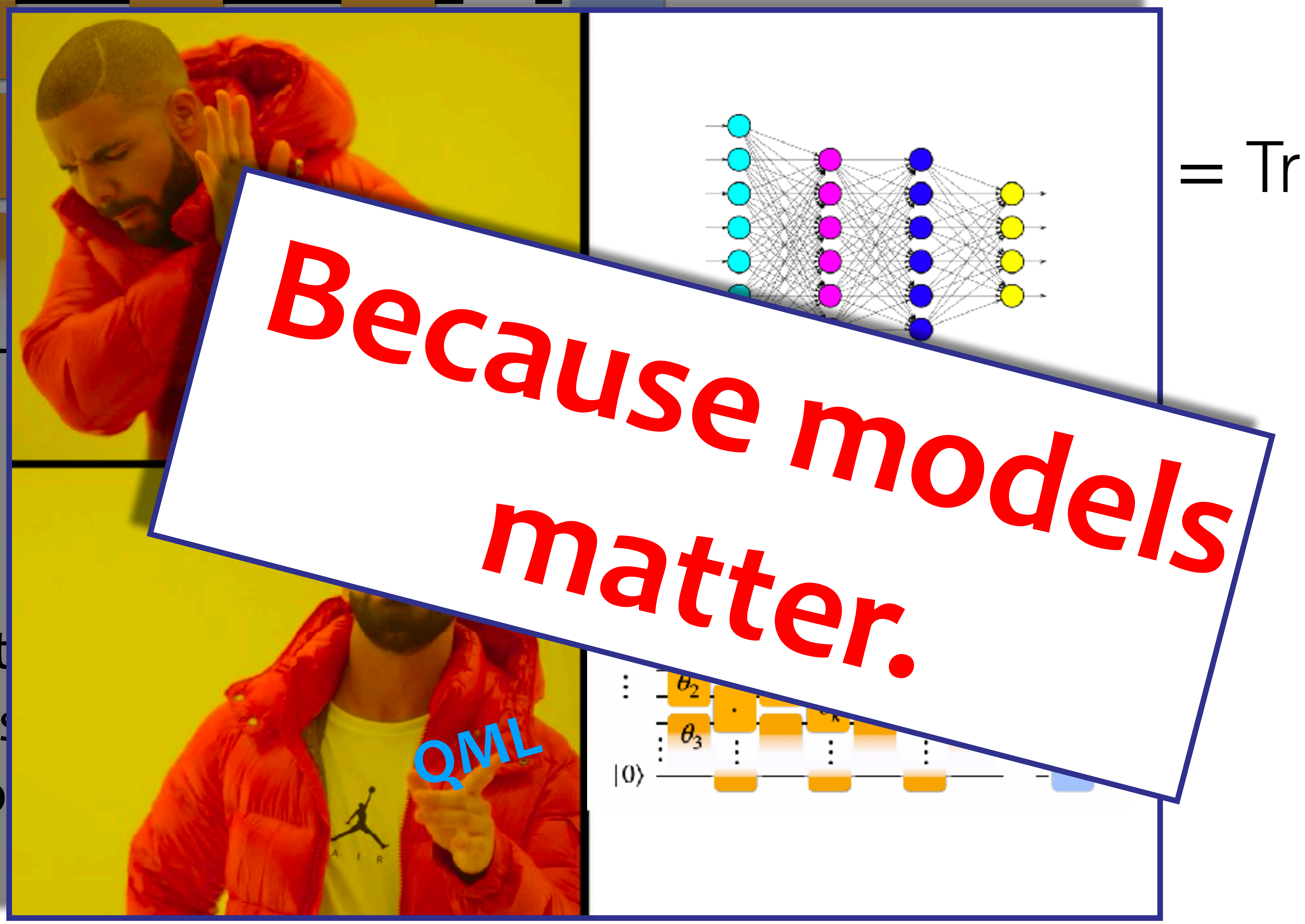
trainable parameters

output = expectatio

Parametrized quantum circuits as machine learning models

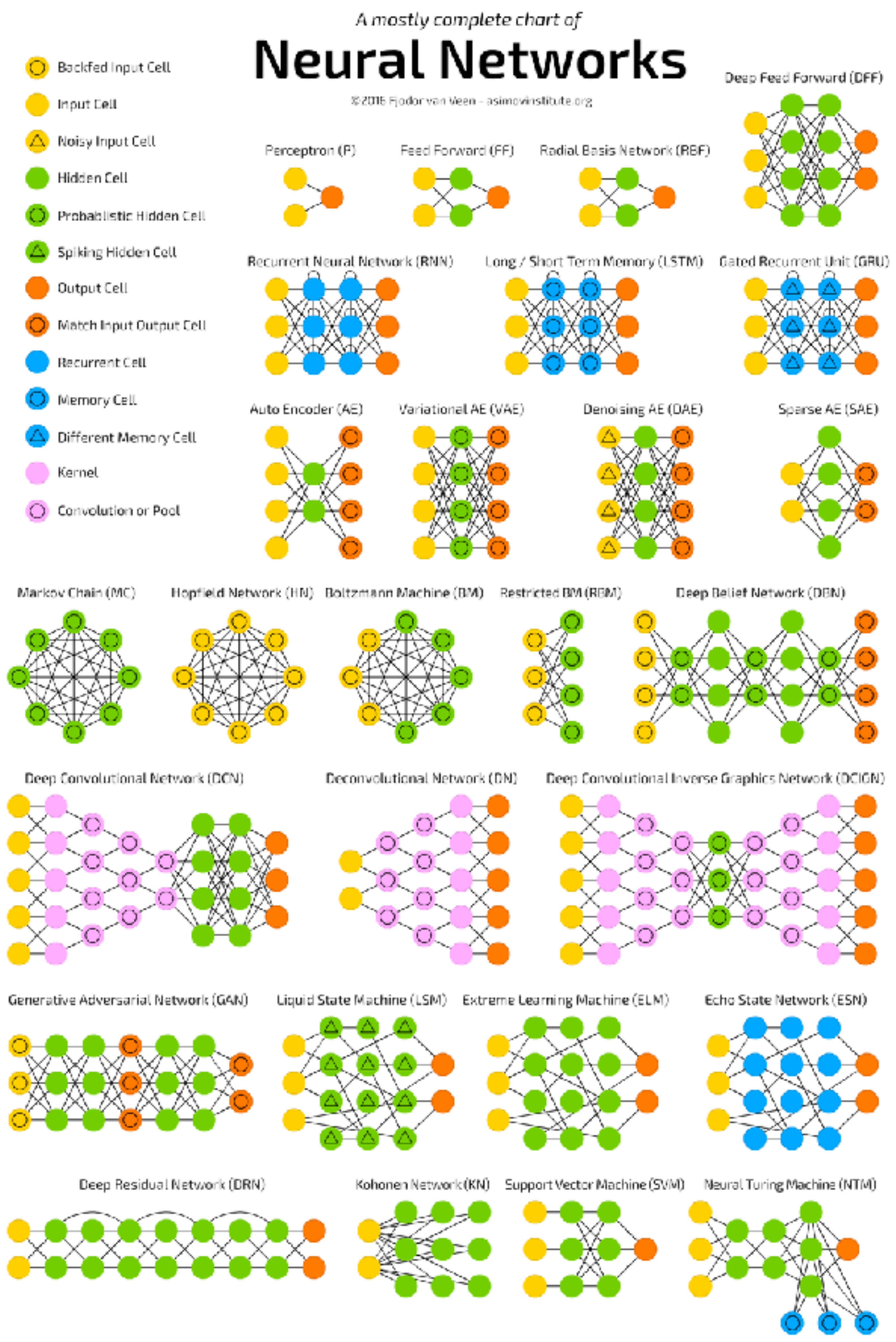
(variational quantum circuits)





$= \mathrm{Tr}[\rho(\vec{\theta}, \vec{x})O(\cdot)]$

Typically:
input = sets paramet
trainable parameter
output = expectatio



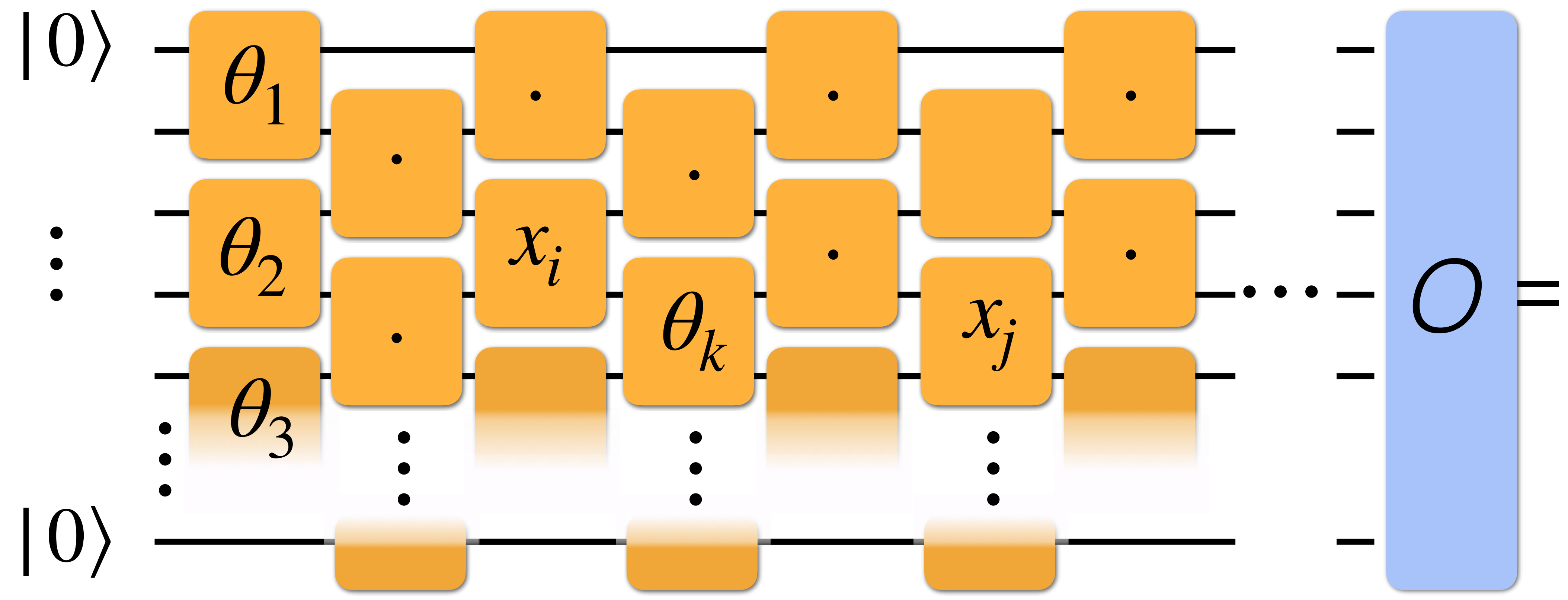
Good models?

- expressive
- flexible (various NNs, e.g. convolutional nets)
- easy to regularize well
- easy to optimize
- can match important real-world distributions
- can compute them!

https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464

Parametrized quantum circuits as machine learning models

(variational quantum circuits)



$$\Rightarrow \quad f_{\vec{\theta}}(\vec{x}) = \mathrm{Tr}[\rho(\vec{\theta}, \vec{x})O(\,\cdot\,)]$$

Q1: Can we get it to work?
(robustness to noise, training features, match with devices)

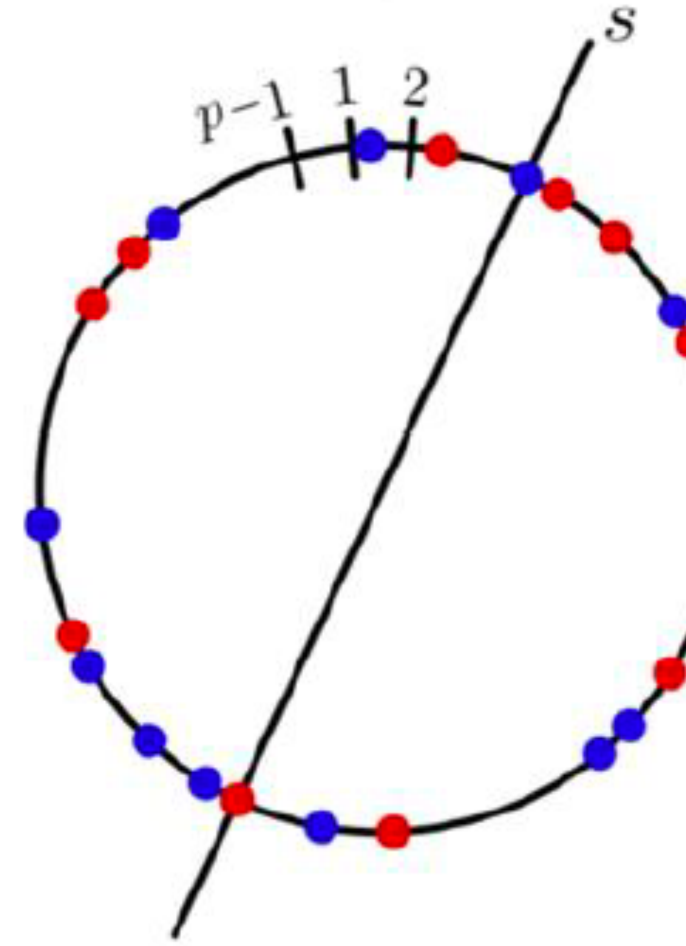
Q2: Should we get it to work?
(Expressivity/"type"? Generalization bounds? Capacity for q. advantage?)

What we to know about QC ML models (complexity separations)

General QC ML models can do *more*:

- classically intractable model families (“supremacy” for generative)
- there are classical/quantum *learning separations*

$$f_s(x) = \begin{cases} +1, & \text{if } \log_g x \in \left[s, s + \frac{p-3}{2}\right] \\ -1, & \text{otherwise} \end{cases}$$



There exist (contrived) supervised/generative learning problems
(even with *classical, classically efficiently generatable data*!)

which require exponential compute time classically, poly quantumly
(unlesss discrete log is easy)

Basis: Cryptographic function used to instill
classically hard, quantum easy structure in data-label correlations

arxiv.org/abs/2010.02174
arxiv.org/abs/2007.14451

What we know about PQC ML models (learning 2)

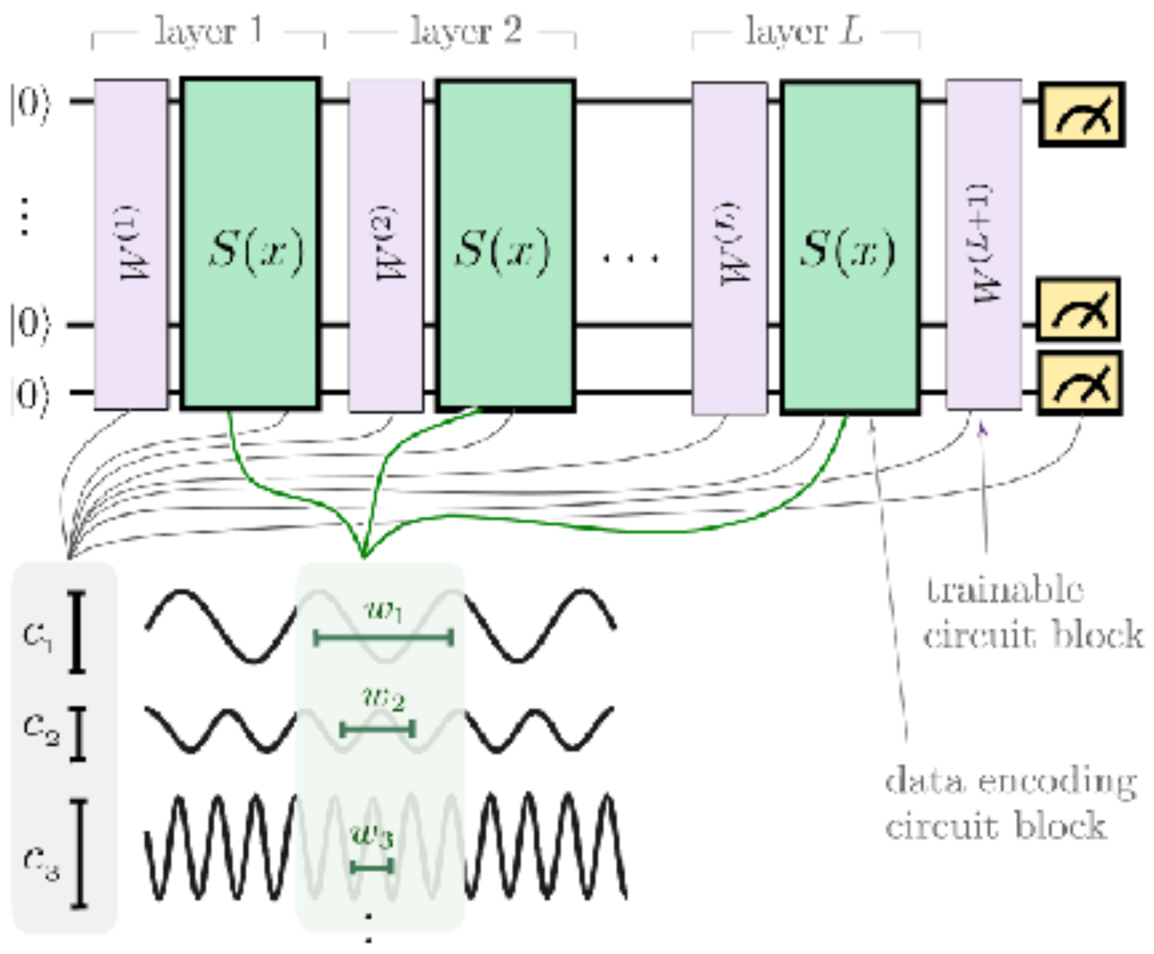
- correspond to (big!) generalized trigonometric polynomials
- very different from NNs so likely different applications…

- increasing understanding of generalization performance and regularization (10+ papers), e.g. in explicit map

$$x \rightarrow Tr[\rho(x)O(\theta)]$$

rank and Frobenious norm of O directly influence the VC, and fat shattering dimensions

- some metrics on data indicating a performance advantage may be achievable



arxiv.org/pdf/2008.08605
arxiv.org/abs/2106.03880
arxiv.org/abs/2105.05566
arxiv.org/abs/2011.01938

In essence

- a peculiar family that we are starting to understand
- different than NNs:
  so promising where NNs struggle
- not clear what it should be used for (esp. in ML)

WHAT IS THE RIGHT NAIL FOR OUR HAMMER?

Challenges:

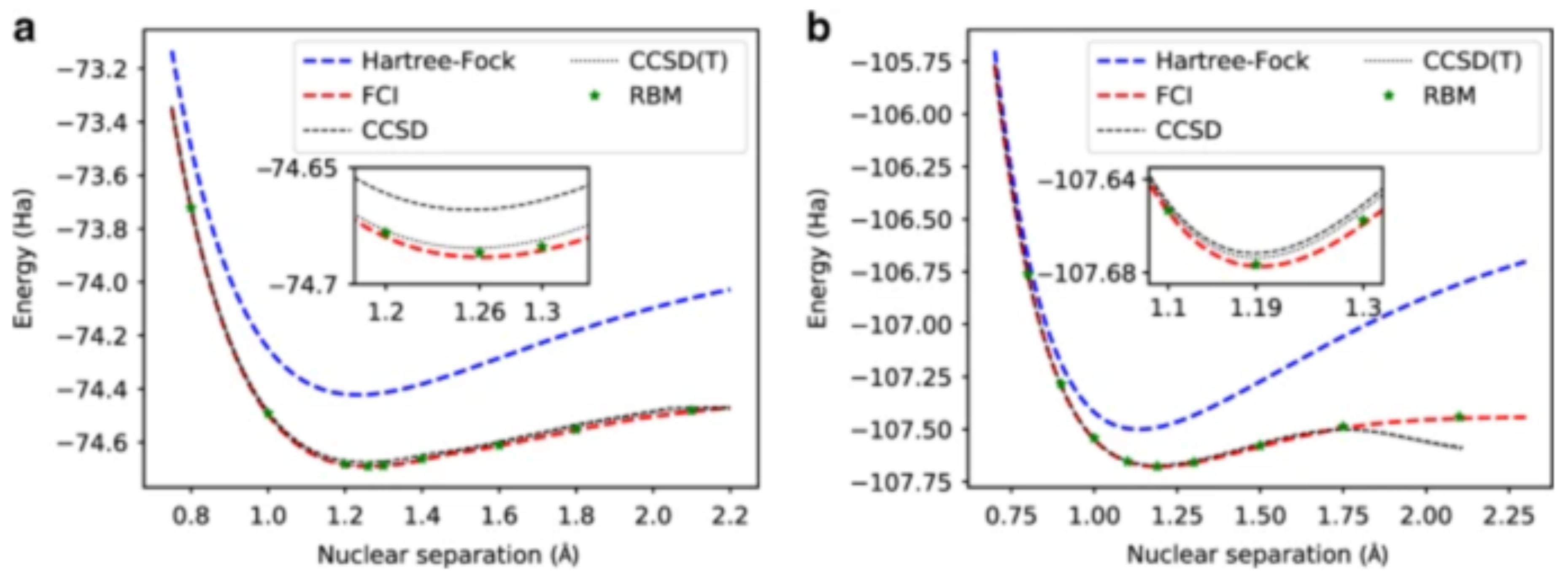
- **more understanding**
- **empirical advantage: need real experiments (guided by theory above)**
- **theoretically supported separations for relevant problems**

What is the right application?

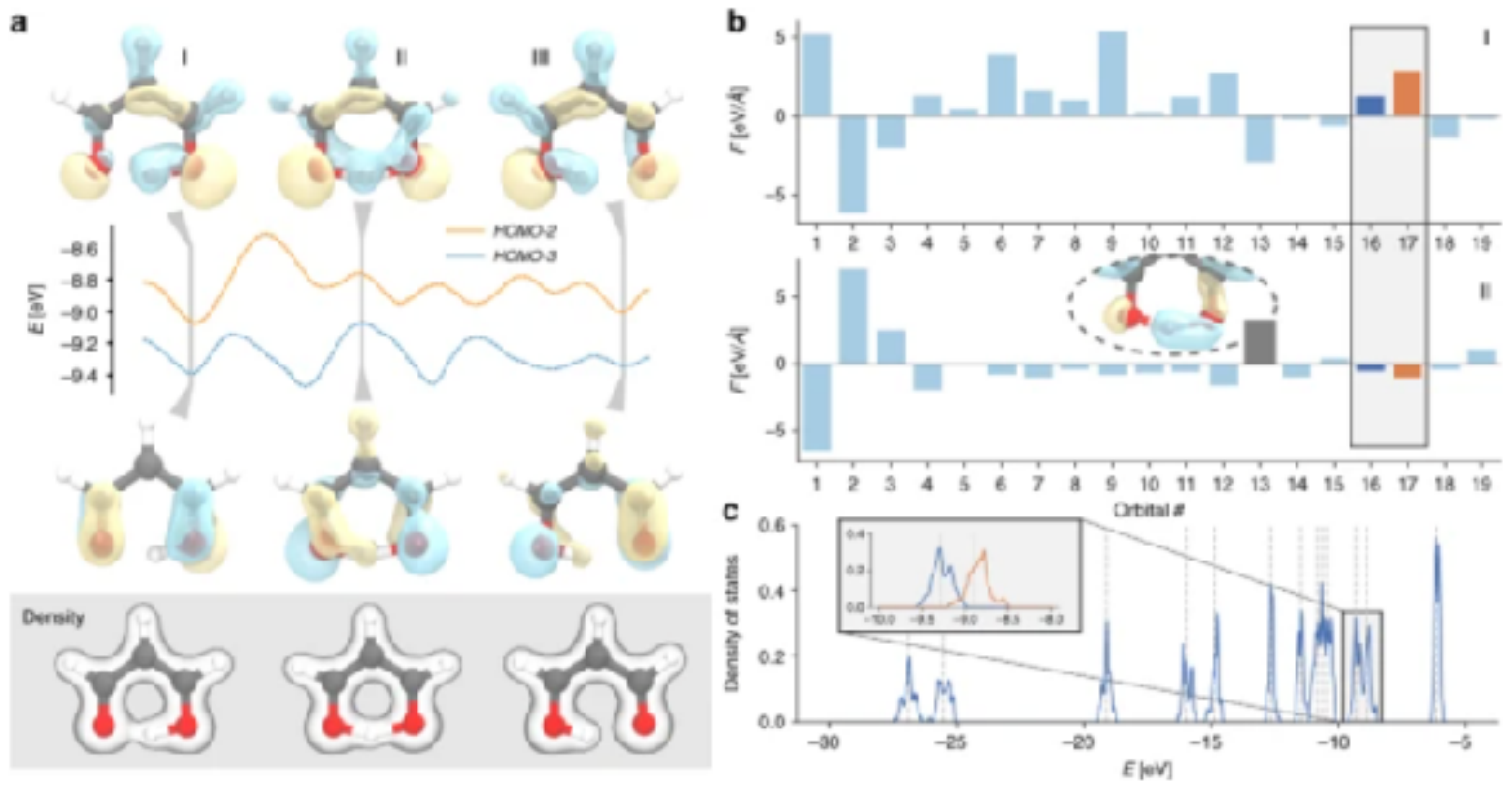
Learning from data generated by highly interacting systems: chemistry, condensed matter

cryptographic structure ↔ quantum mechanical nature of ground truth/underlying distribution

But can apply to real-world data now!



https://www.nature.com/articles/s41467-020-15724-9



Proton transfer in malondialdehyde

https://www.nature.com/articles/s41467-019-12875-2

Applications in HEP: arXiv:2005.08582

Applications in astronomy:
npj Quantum Information 7, 161 (2021) (arXiv:2101.09581)
Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC)

But for now capacities to help very limited…

This is not the end…
…end of the beginning

In essence

WHAT IS THE RIGHT NAIL FOR OUR HAMMER?

Challenges:

- more understanding
- empirical advantage: need real experiments (guided by theory above)
- theoretically supported separations for relevant problems

What we (want) to know about QC ML models (performance parameters)

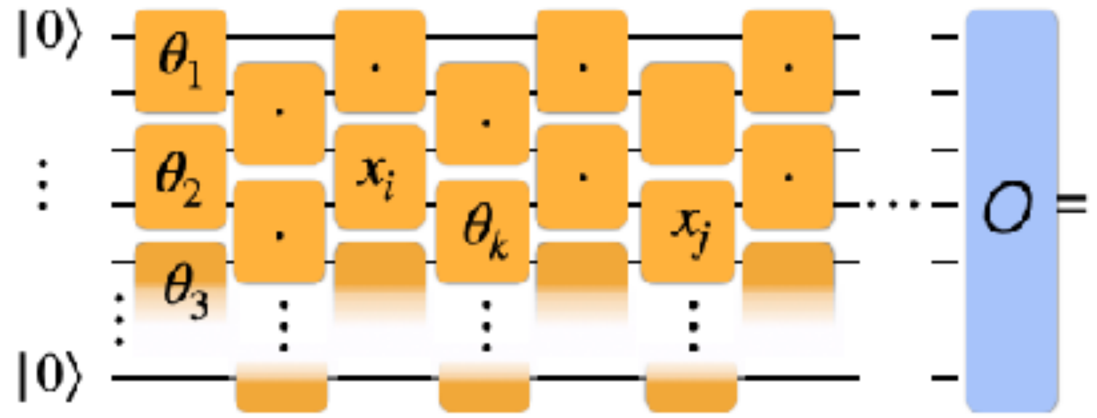
Generalization performance bounds

Want: $R(h) = \int_{\mathcal{X}\times\mathcal{Y}} \ell(h(x), y)dP(x, y).$ error *everywhere*

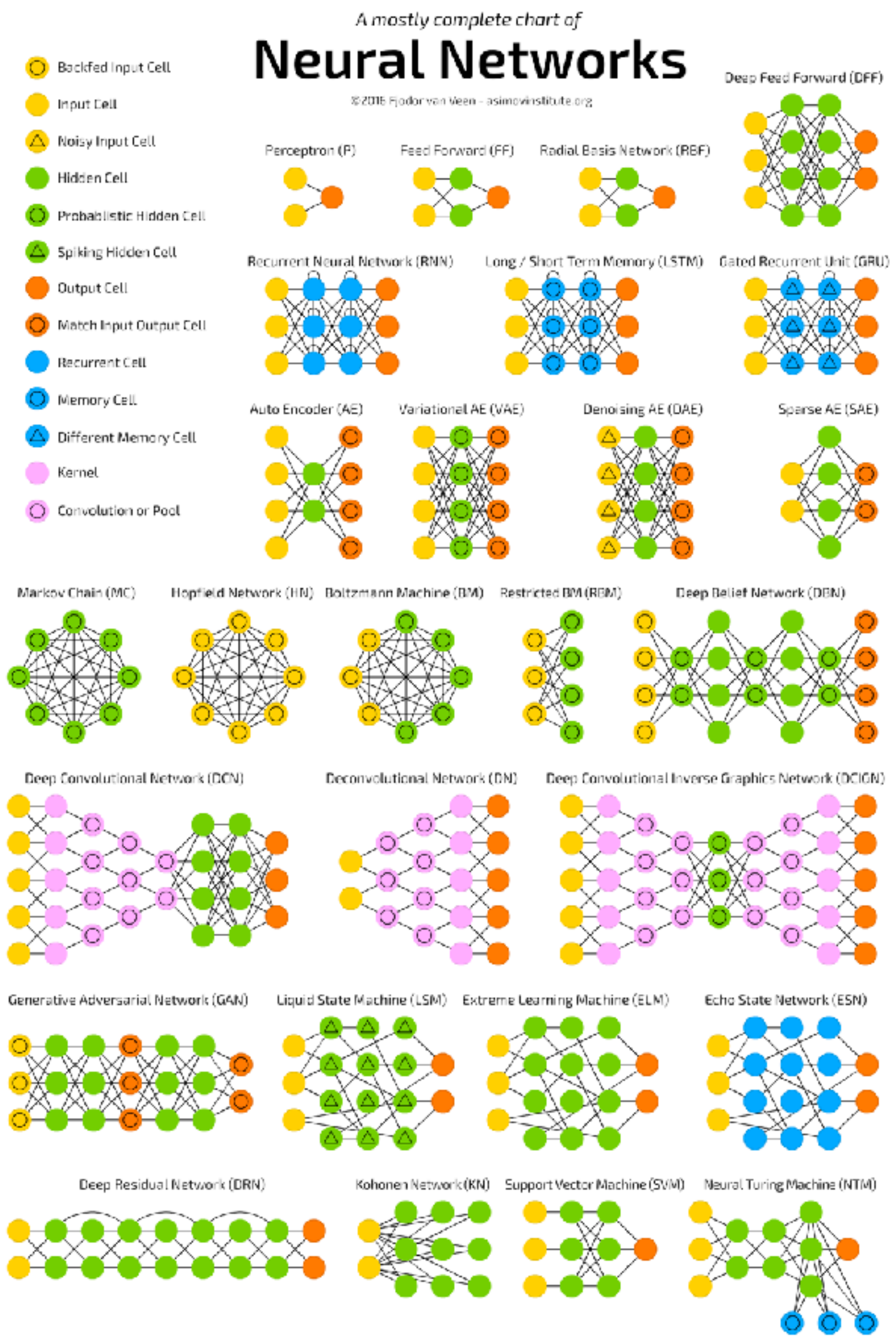
Have: $\hat{R}_S(h) = \frac{1}{|S|} \sum_{(x_i,y_i)\in S} \ell(h(x_i), y_i)$ error *on training set*

Can prove: $P\left(R(h) < \hat{R}_S(h) + g(\mathcal{F}, m, \delta)\right) > 1 - \delta$

P - true distribution/correlation data-label
ℓ-error function e.g abs-value
h - a classifier
S - dataset of size m
$\mathcal{F}$ *- function family*

$\mathcal{F} =$ 

How this behaves is critical!!!



Good models?

- expressive
- flexible (various NNs, e.g. convolutional nets)
- easy to regularize well
- easy to optimize
- can match important real-world distributions
- can compute them!

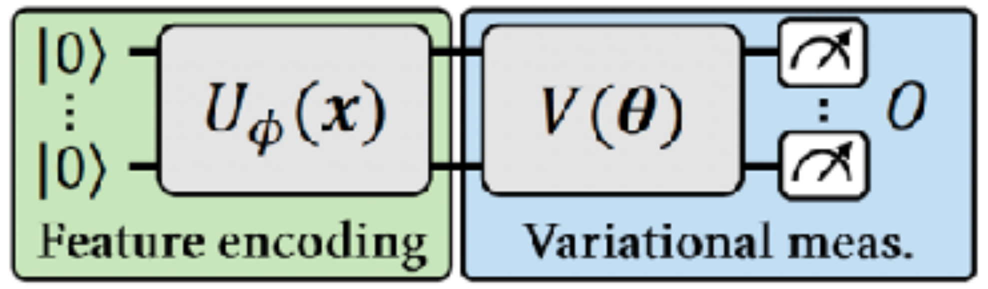
https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464

Parametrized quantum circuits as machine learning models

(variational quantum circuits)

Various types

Explicit



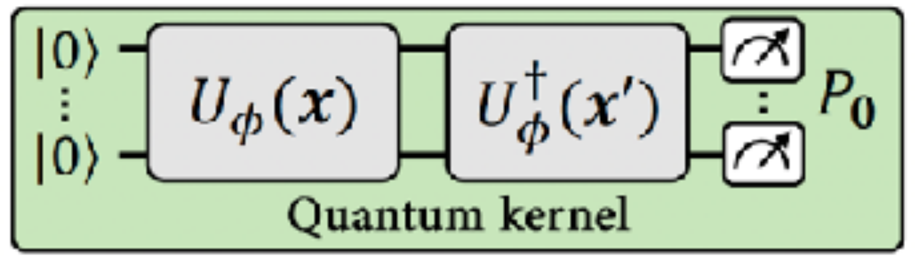
$\rho(x)$ $\quad\quad$ $O(\theta)$

$x \rightarrow Tr[\rho(x)O(\theta)]$

First ideas

linear models (and SVMs) [11]

Implicit/Kernels

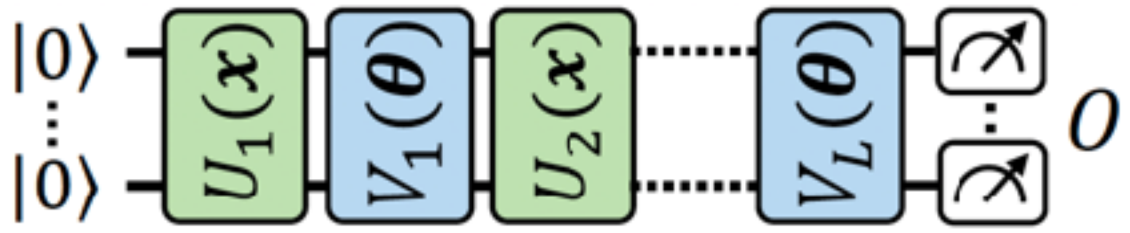


$K(x, x') = Tr[\rho(\vec{x})\rho(\vec{x}')]$

Learning separations [2]

Famous connection to kernel models [6,7]:

Data re-uploading



$x \rightarrow Tr[\rho(x,\theta)O(\theta)]$

Universal approximators [1]

$f(\boldsymbol{x}) = \sum_{\omega \in \Omega(H)} c_\omega \exp(-i\boldsymbol{\omega x})$ *[3];*

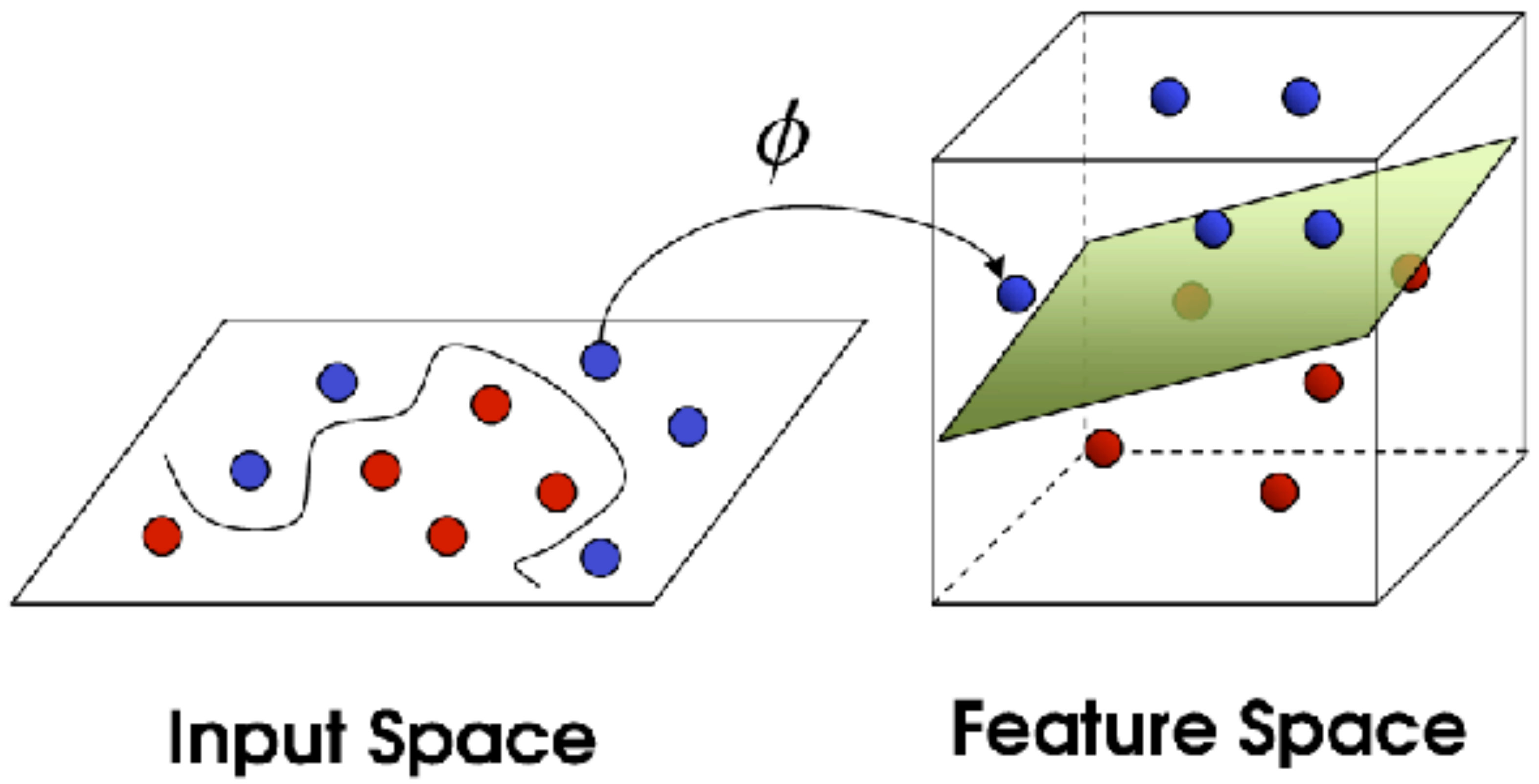
(for all) various bounds (Rademacher, Pseudodimension, VC, Fat-shattering) [3,4,5,g]

[g] C Gyurik, D van Vreumingen, VD, "Structural risk minimization for quantum linear classifiers", arXiv:2105.05566 (2021).

Parametrized quantum circuits as machine learning models

(variational quantum circuits)

Linear classifiers & feature spaces



Cover's theorem

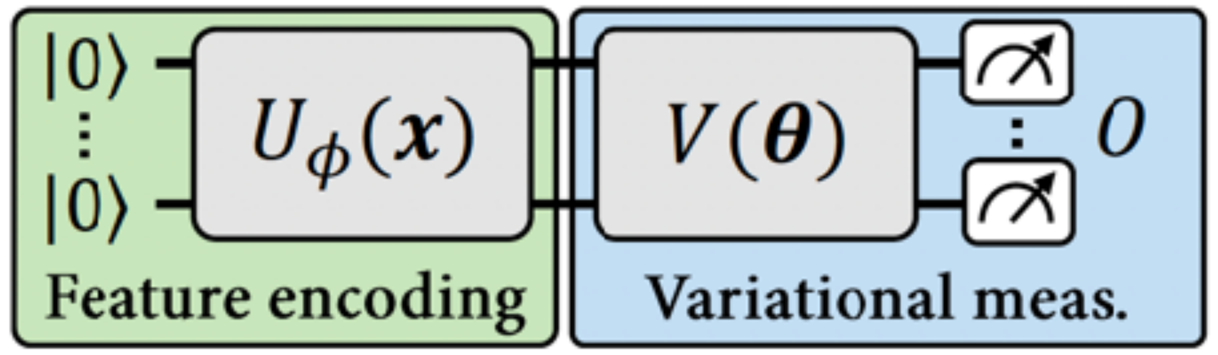
Support vector machine: maximum margin classifier

Classifier: inner product of normal vector and mapped data

Linear classifiers: feature map is data-independent.

"Quantum feature spaces"

Explicit:



$\rho(x)$ $O(\theta)$

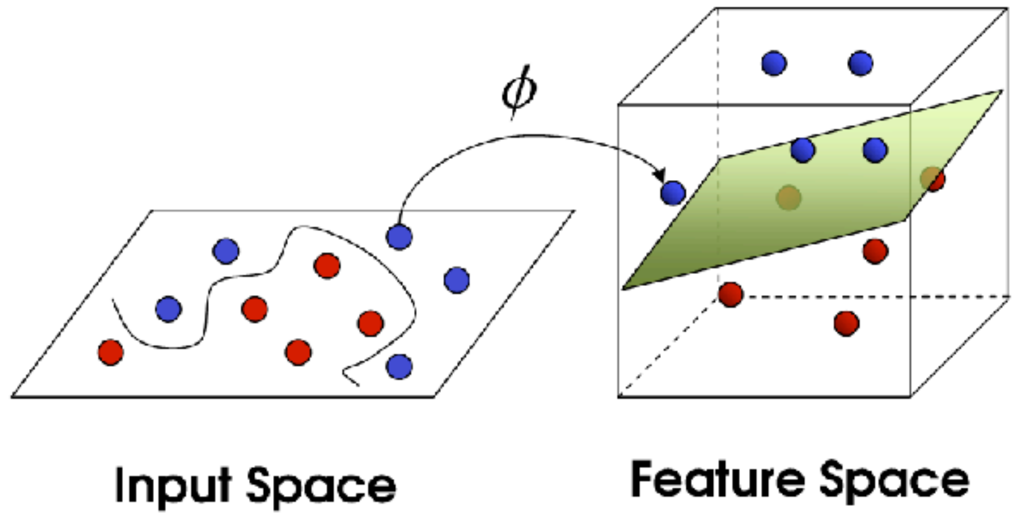
$x \to Tr[\rho(x)O(\theta)]$

Frobenius inner product in feature space $\mathbb{R}^{4n}$*;* ***restricted.***

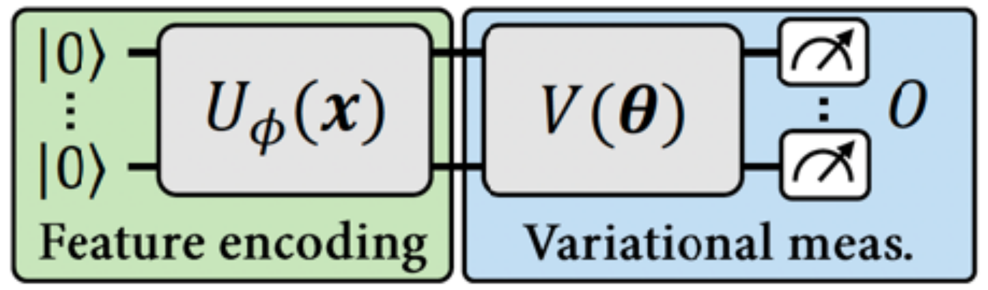
Parametrized quantum circuits as machine learning models

(variational quantum circuits)

"Quantum feature spaces"



Explicit



$\rho(x)$ $O(\theta)$

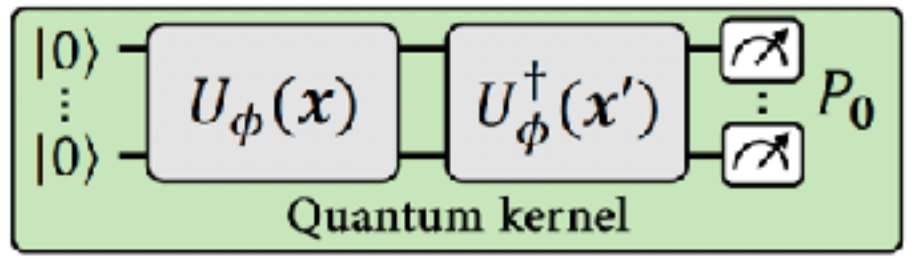
$x \rightarrow Tr[\rho(x)O(\theta)]$

Kernel methods and the representer theorem

$$f_{\boldsymbol{\alpha},\mathcal{D}}(\boldsymbol{x}) = \sum_{m=1}^{M} \alpha_m k(\boldsymbol{x}, \boldsymbol{x}^{(m)}).$$

$$k(\vec{x}, \vec{y}) = \mathrm{Tr}[\rho(\vec{x})\rho(\vec{y})]$$

Implicit/Kernels



$$K(x, x') = Tr[\rho(\vec{x})\rho(\vec{x}')]$$

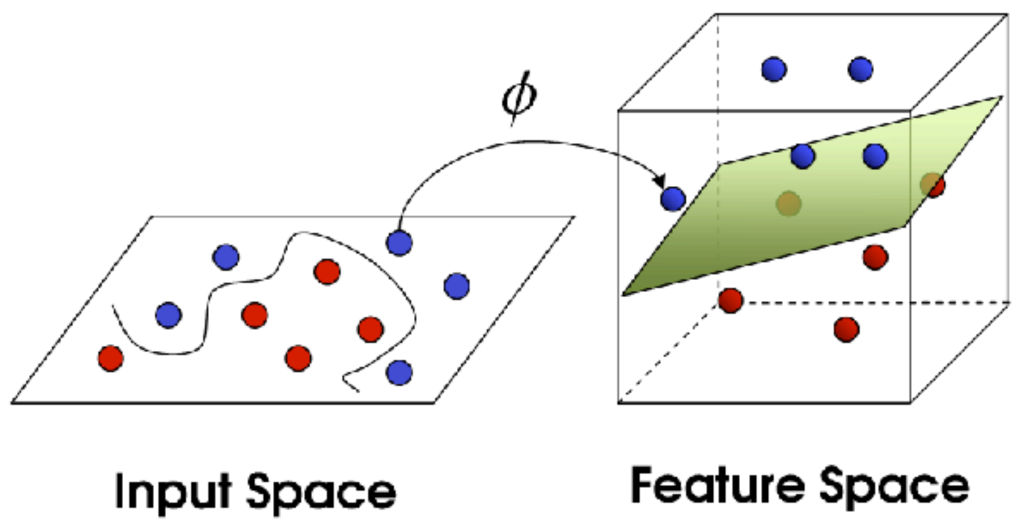
Representer theorem: reguralized loss is minimized by an implicit model $\widehat{\mathcal{L}}_\lambda(f) = \widehat{\mathcal{L}}(f) + \lambda \|O\|_{\mathcal{F}}^2$

Kernels are powerful; if they all there is, is this bad news?

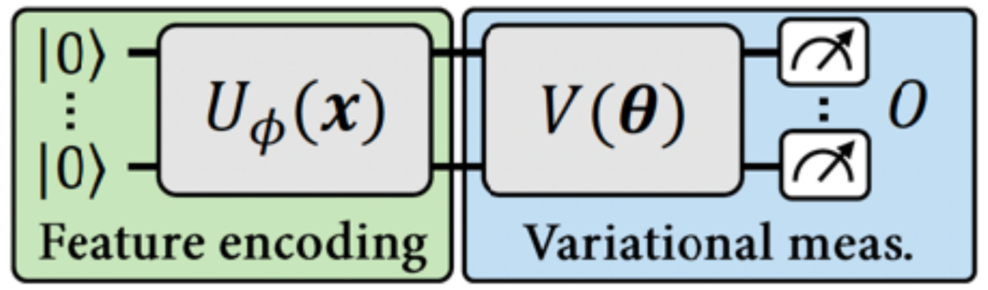
Parametrized quantum circuits as machine learning models

(variational quantum circuits)

“Quantum feature spaces”



Explicit



$\rho(x)$ $\quad O(\theta)$

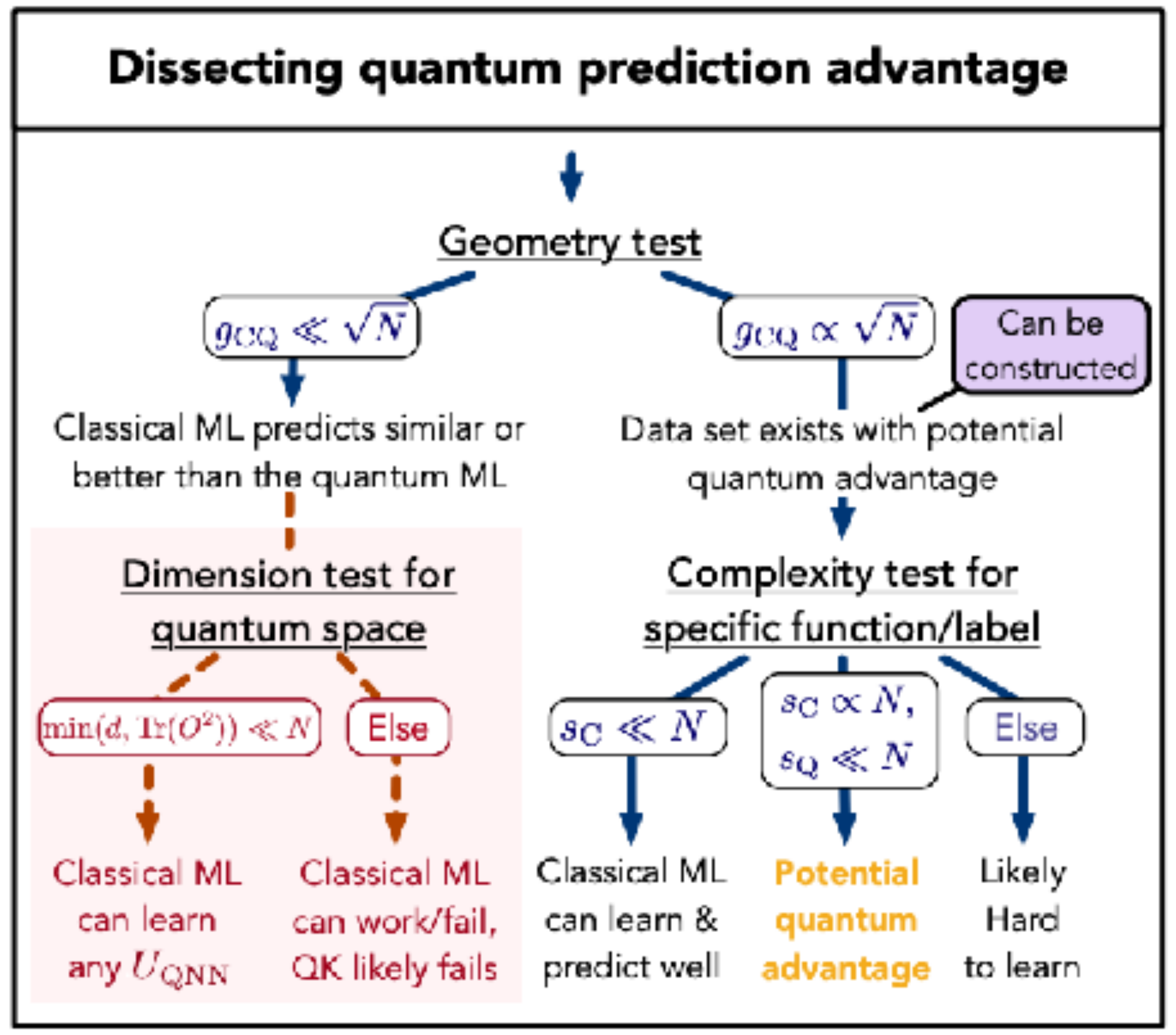
$x \to Tr[\rho(x)O(\theta)]$

Kernel methods and the representer theorem

$$f_{\boldsymbol{\alpha},\mathcal{D}}(\boldsymbol{x}) = \sum_{m=1}^{M} \alpha_m k(\boldsymbol{x}, \boldsymbol{x}^{(m)}).$$

$$k(\vec{x}, \vec{y}) = \mathrm{Tr}[\rho(\vec{x})\rho(\vec{y})]$$

Representer theorem: reguralized loss is minimized by an implicit model $\widehat{\mathcal{L}}_\lambda(f) = \widehat{\mathcal{L}}(f) + \lambda \|O\|_{\mathcal{F}}^2$



From:
Power of data in quantum machine learning

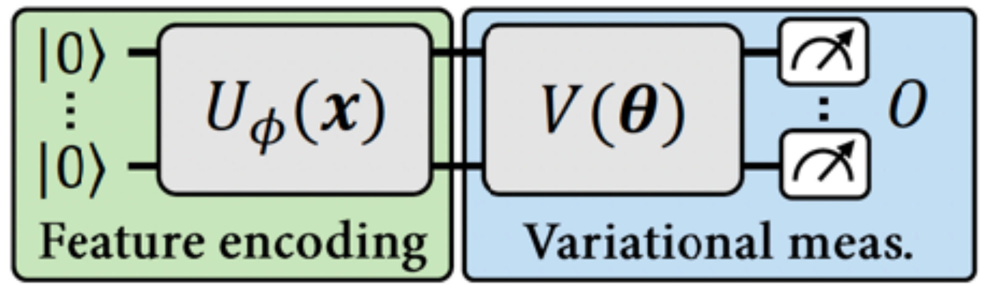
=Kernels are powerful; if they all there is, is this bad news?

Parametrized quantum circuits as machine learning models

(variational quantum circuits)

Questions: is this the full picture? What about data re-uploading (not kernel)?

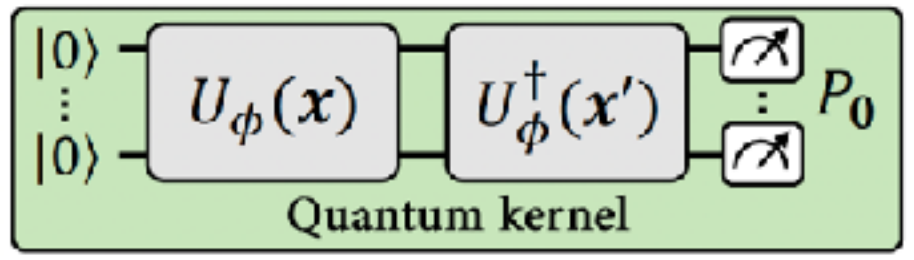
Explicit



$\rho(x)$ $\qquad O(\theta)$

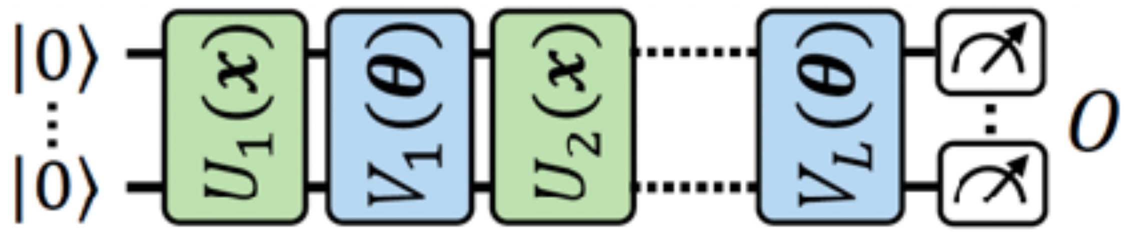
$x \rightarrow Tr[\rho(x)O(\theta)]$

Implicit/Kernels



$K(x, x') = Tr[\rho(\vec{x})\rho(\vec{x'})]$

Data re-uploading



$x \rightarrow Tr[\rho(x, \theta)O(\theta)]$

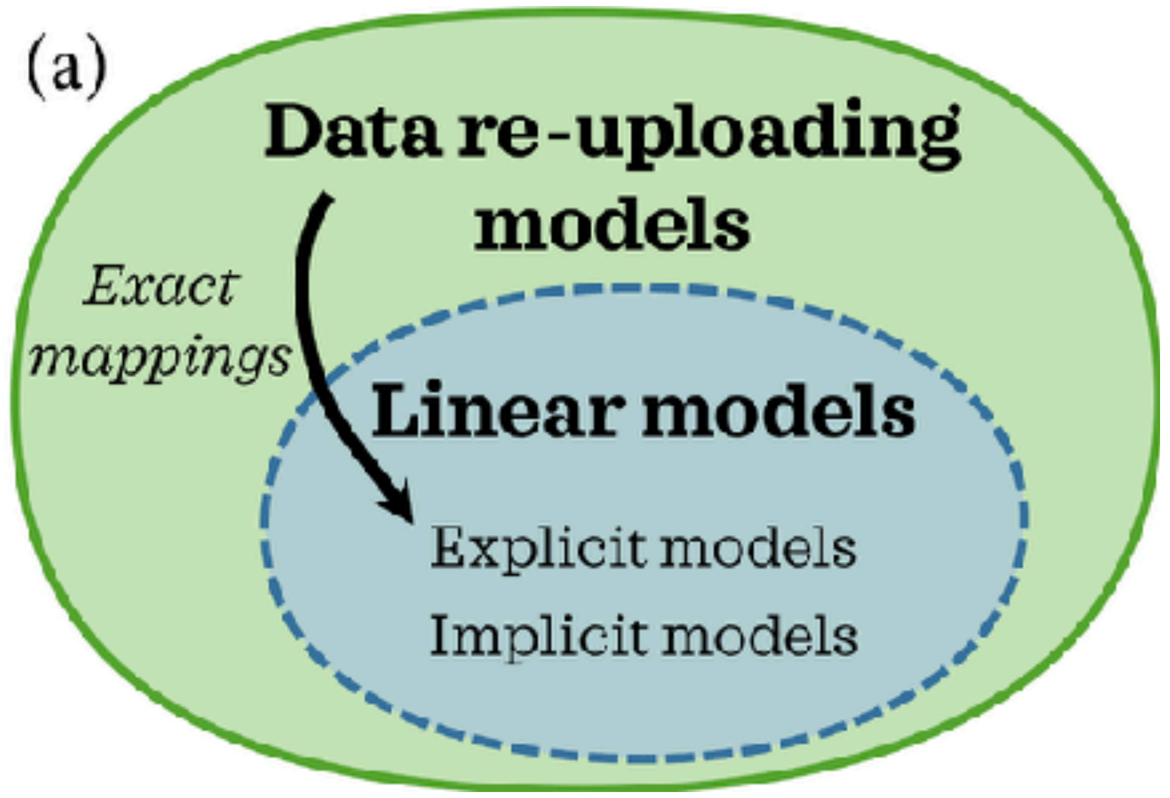
[h] S Jerbi, LJ Fiderer, HP Nautrup, JM Kübler, HJ Briegel, VD, "Quantum machine learning beyond kernel methods",arXiv:2110.13162 (2021)

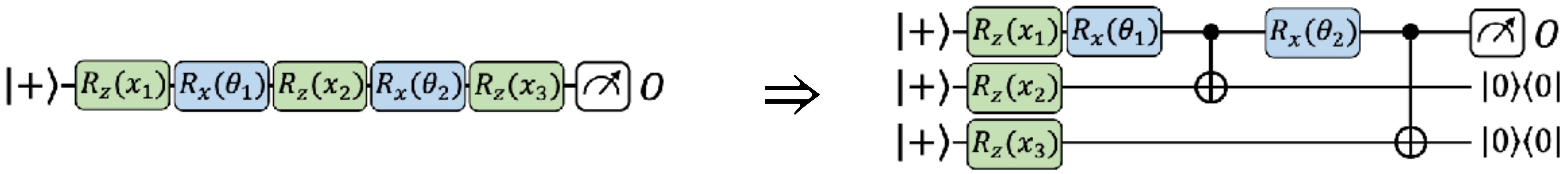
Parametrized quantum circuits as machine learning models

(variational quantum circuits)

Summary of results [h]:

a) Poly-sized data-re-uploading models are exactly (restricted) poly-sized explicit models (and are *not* poly-sized implicit models)





- via MBQC or gate teleportation
- but there is a penalty for *implementation due to postselection* ($O \Rightarrow O \otimes |0\rangle\langle 0|^{\otimes k}$)

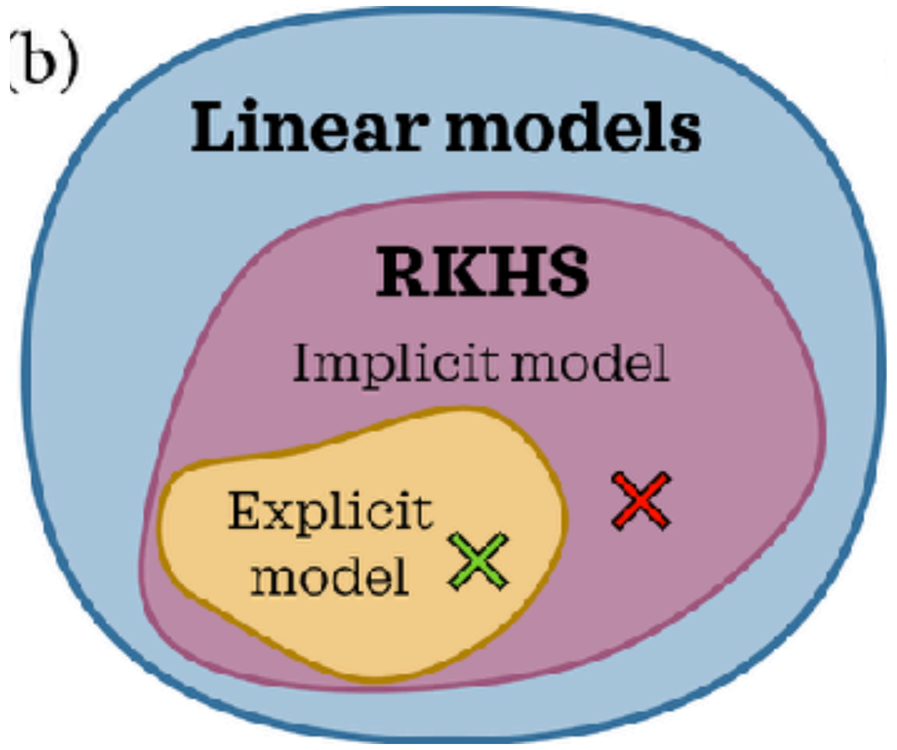
[h] S Jerbi, LJ Fiderer, HP Nautrup, JM Kübler, HJ Briegel, VD, "Quantum machine learning beyond kernel methods",arXiv:2110.13162 (2021)

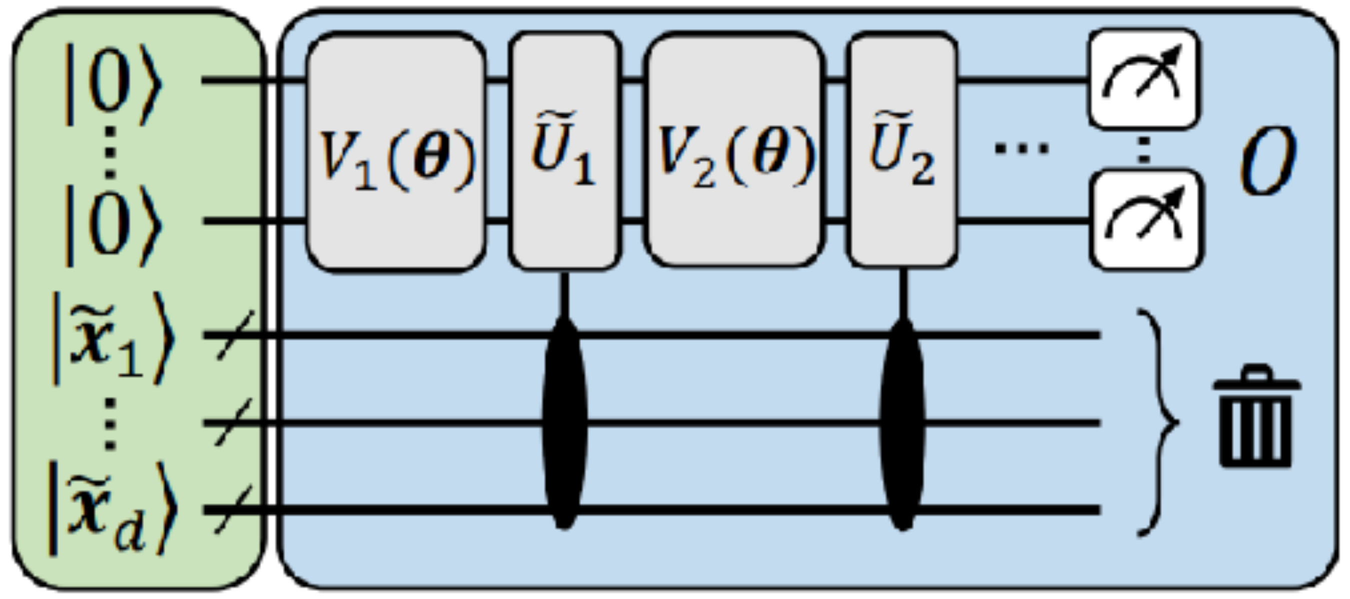
Parametrized quantum circuits as machine learning models

(variational quantum circuits)

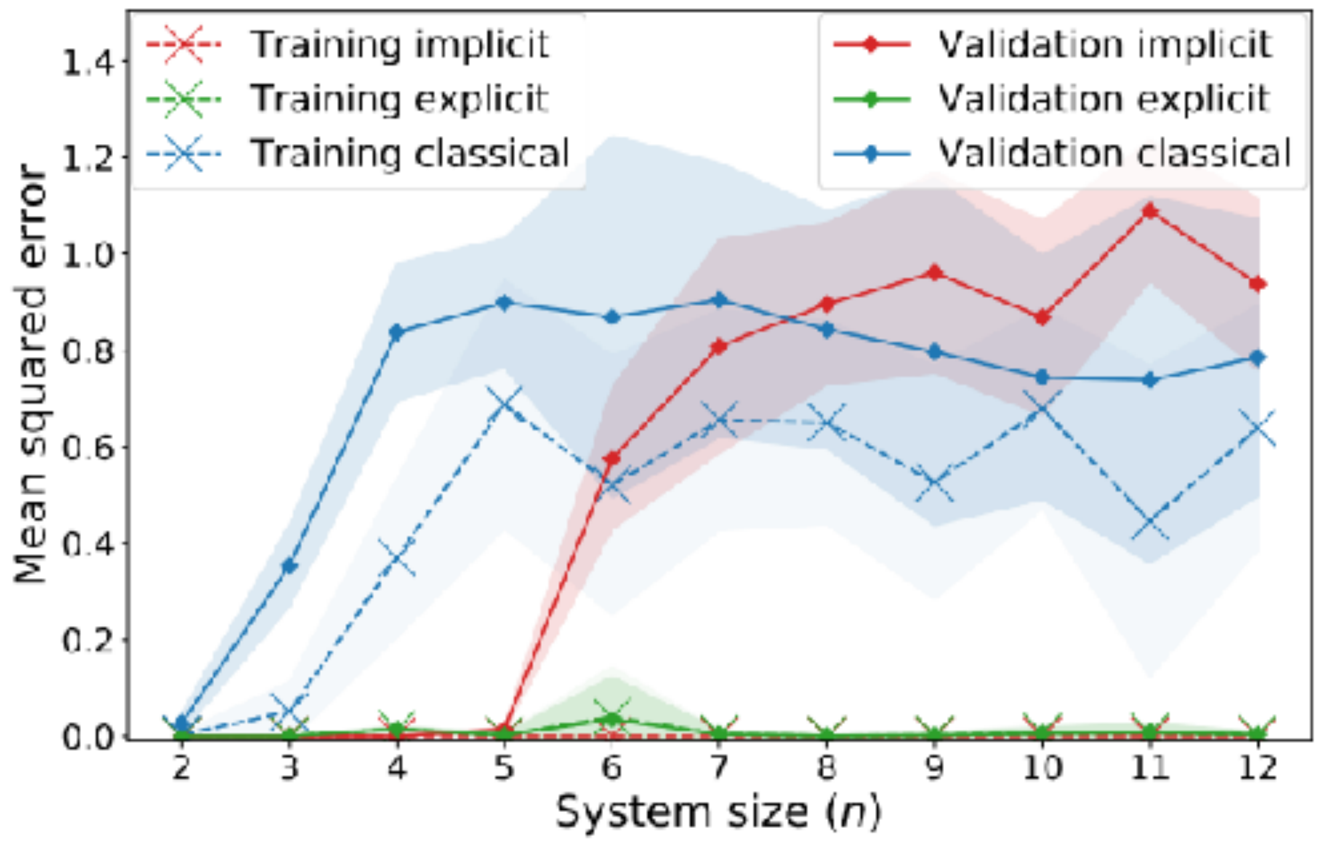
Summary of results [h]:

b) Quantum kernel models can dramatically overfit. Also gives different result to [8]





Kernel diagonal… yet circuit computes an effective kernel



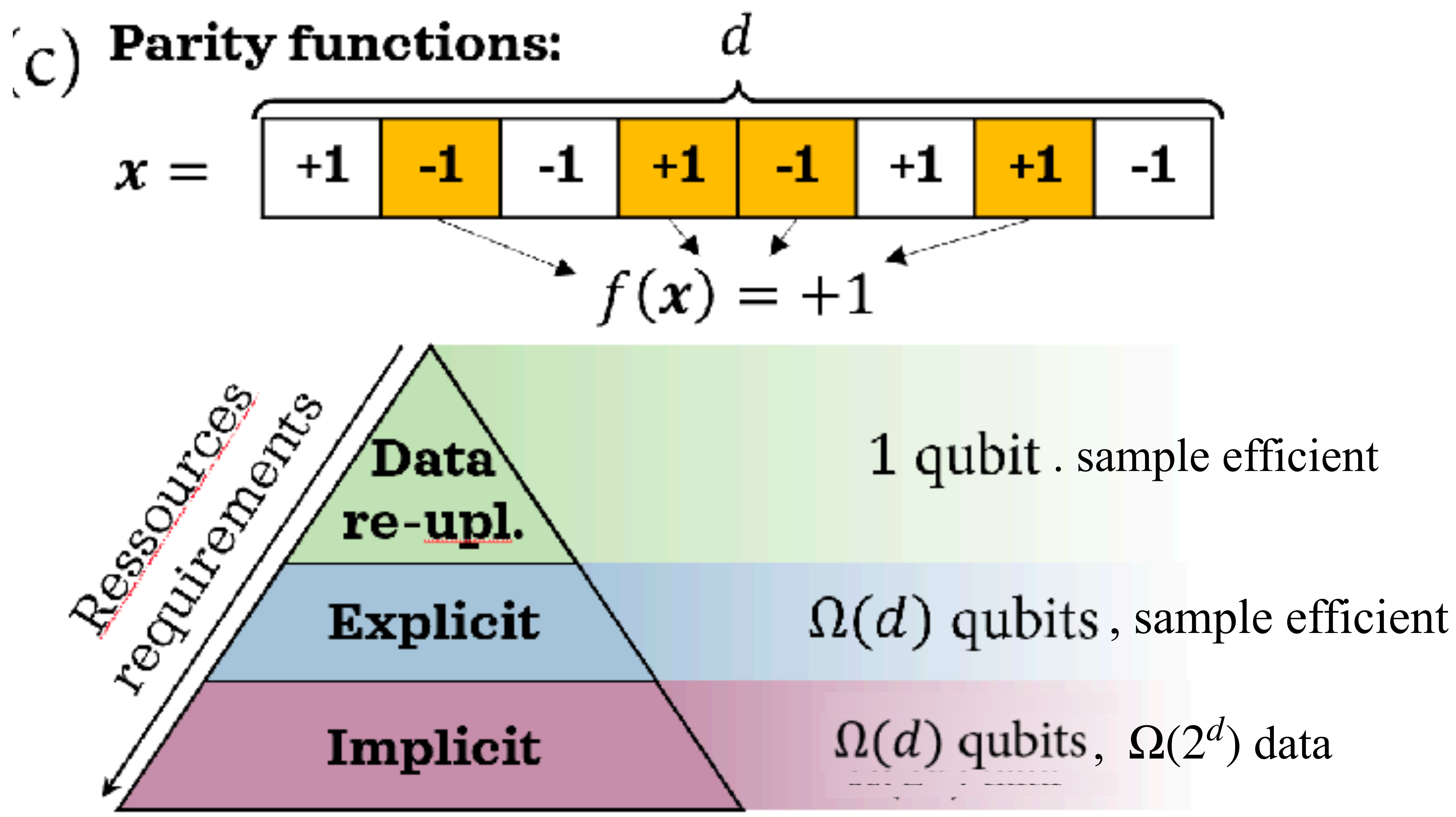
[h] S Jerbi, LJ Fiderer, HP Nautrup, JM Kübler, HJ Briegel, VD, "Quantum machine learning beyond kernel methods",arXiv:2110.13162 (2021)

Parametrized quantum circuits as machine learning models

(variational quantum circuits)

Summary of results [h]:

c) Fine-grained distinctions and sample complexity differences.
There exist learning problem involving parities such that…



qubit no. also matters from implementation

[h] S Jerbi, LJ Fiderer, HP Nautrup, JM Kübler, HJ Briegel, VD, "Quantum machine learning beyond kernel methods",arXiv:2110.13162 (2021)

Parametrized quantum circuits as machine learning models

Mid- way summary

- For theory: a broad class of QML models fit in one paradigm: explicit models
  - Not all, not in a useful sense?
  - What does this imply for analysis of properties? New quantum-specific tools needed.

- Practice, situation more contrived. Trade-offs;
  There will be no one-size-fits all solutions…

Challenges:

When does QML make sense?

- **empirical advantage: need real experiments (guided by theory)**
- theoretically supported separations for relevant problems

Divide-and-quantum hybrid models

Divide-and-quantum hybrid models

"When does QML make sense?"

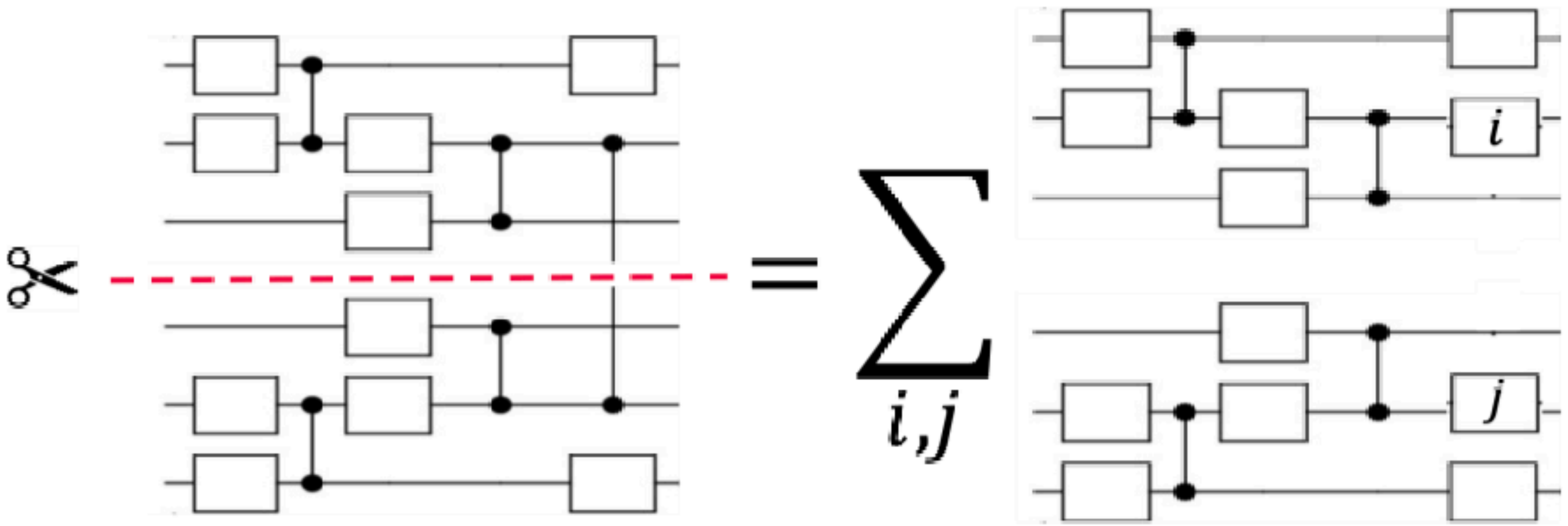
Easy ML is too easy for advantages.
Hard ML is too big for devices.

Can we have our cake and eat it too?

Idea: "mimic" *QML with large quantum circuits using smaller QCs*

Divide-and-quantum hybrid models

the "circuit chop" [9,10]



$$\boxed{U^{(1,2)} = \sum_{i,j} \alpha_{i,j} U_i^{(1)} \otimes U_j^{(2)}}$$

$$\tilde{f}_{\boldsymbol{\theta}}(x) = \sum_{i=1}^{T} c_i \prod_{k=1}^{K} \langle 0| U'^{i,k\dagger}(\boldsymbol{\theta}, x) M_k U^{i,k}(\boldsymbol{\theta}, x) |0\rangle$$

Exponential in # entangling gates cut

(think of these as hypotheses families/models again)

Can we approximate behaviour with just few terms? Concentration?

Even if yes… which terms? Combinatorial problem…

Divide-and-quantum **hybrid models**

Approach:

Combinatorially many sub-circuits…
but differ only in the cut gates.

Make those variational.
Pre-define number of terms.

Relaxation of combinatorial problem.

Relaxations still all valid solutions (!)
… but no longer truncated large circuits.

➡ *Big(er)-quantum-inspired hybrid models*

$$\tilde{f}_{\boldsymbol{\theta}}(x) = \sum_{i=1}^{T} c_i \prod_{k=1}^{K} \langle 0| U'^{i,k\dagger}(\boldsymbol{\theta}, x) M_k U^{i,k}(\boldsymbol{\theta}, x)|0\rangle$$

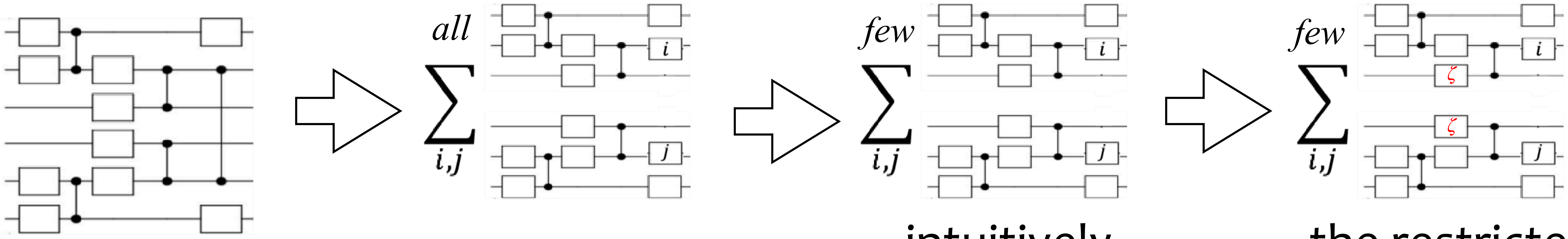
$$\bar{f}_{\boldsymbol{\theta},\boldsymbol{\varsigma},\boldsymbol{\lambda}}(x) = \sum_{i\in[L]} \lambda_i \prod_{k\in[K]} \langle 0|U^k(\boldsymbol{\theta}, x, \zeta_{i,k}) M_k U^k(\boldsymbol{\theta}, x, \zeta_{i,k+K})|0\rangle$$

Divide-and-quantum hybrid models

Findings:

Theory:

Generalization performance: whole-circuit bounds from literature apply.



same family

intuitively
restricted
(simpler)

the restricted
generalized
(more complex)

Divide-and-quantum hybrid models

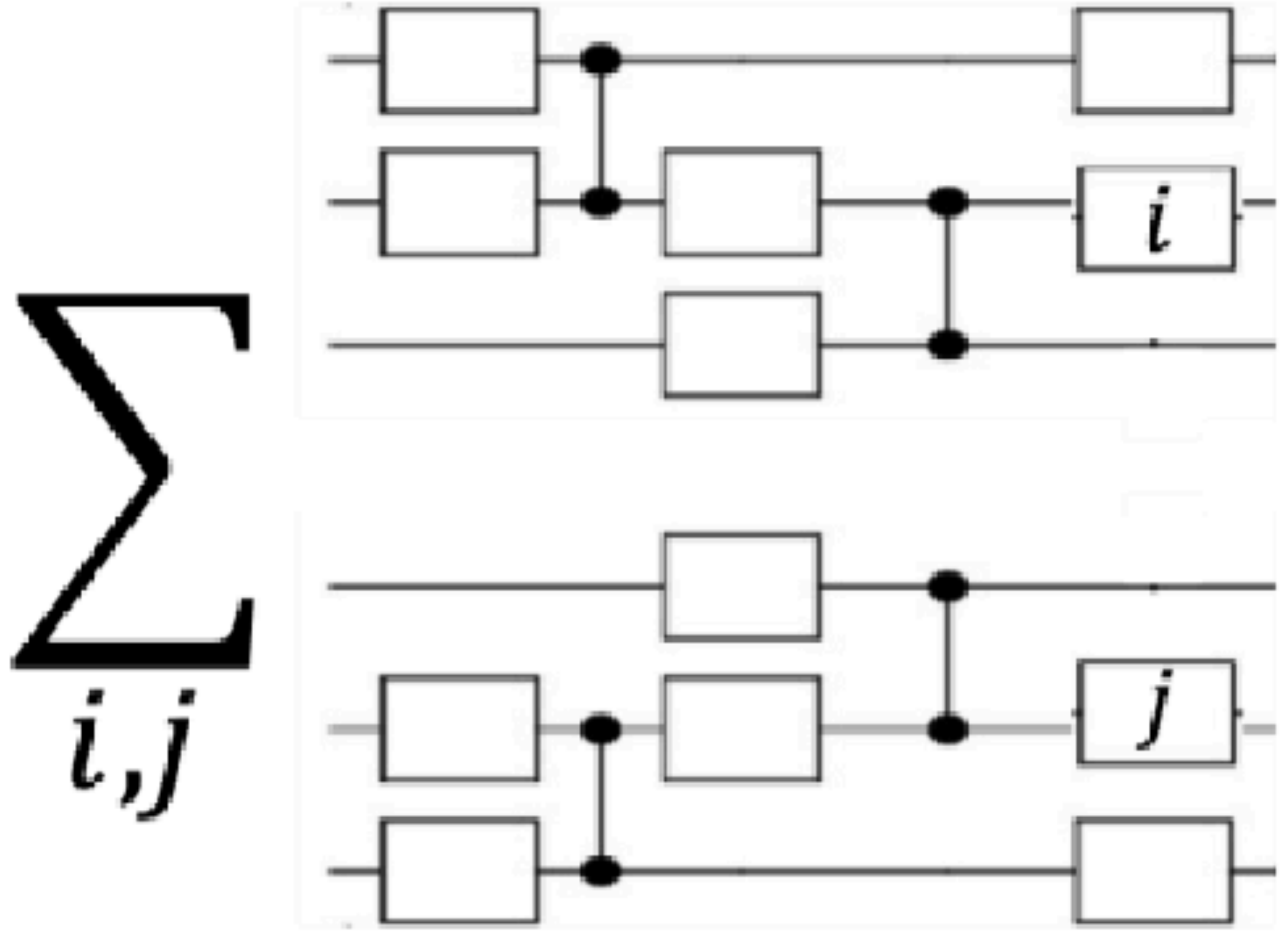
Findings:

Theory:

Generalization performance:
We need new quantum-specific approaches for g. bounds;
Some are possible through rewriting!

Training:
May have **advantages w.r.t. barren plateaus** in training.
Possible error levels can be taken into account

Applications:
A convenient framework for further hybridization
(e.g. localized Fourier functions)



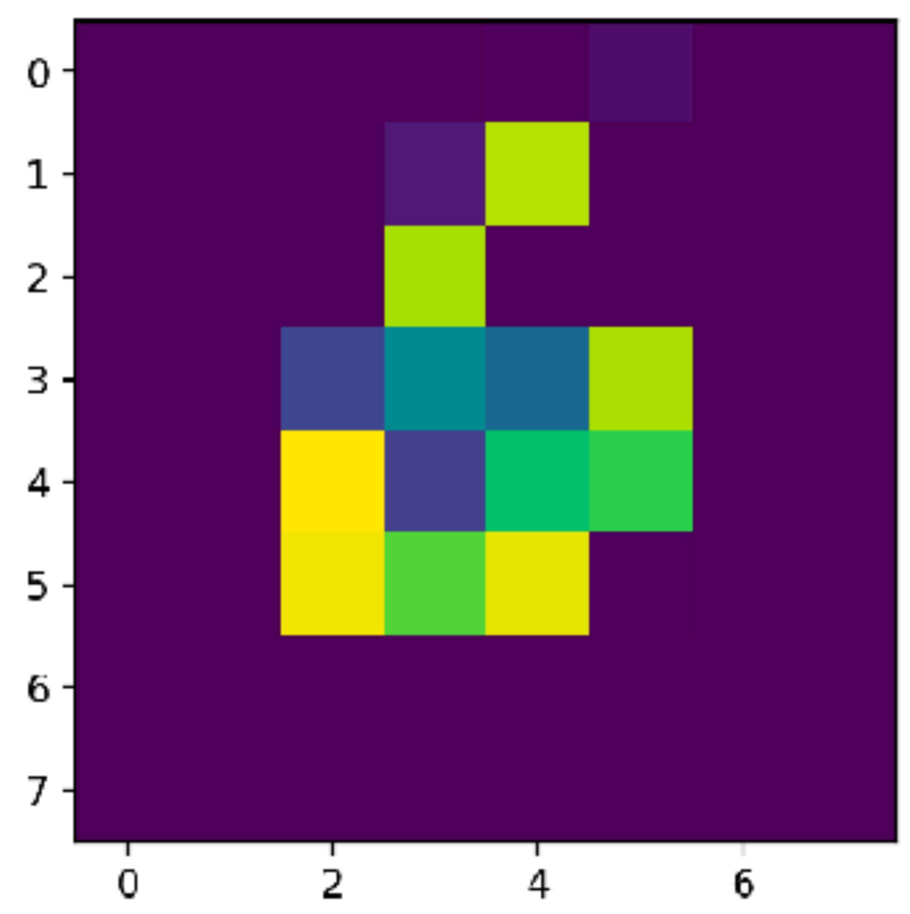
$$\mathcal{R}(\mathcal{F}+\mathcal{G}) \leq \mathcal{R}(\mathcal{F}) + \mathcal{R}(\mathcal{G}).$$

Divide-and-quantum hybrid models

Findings:

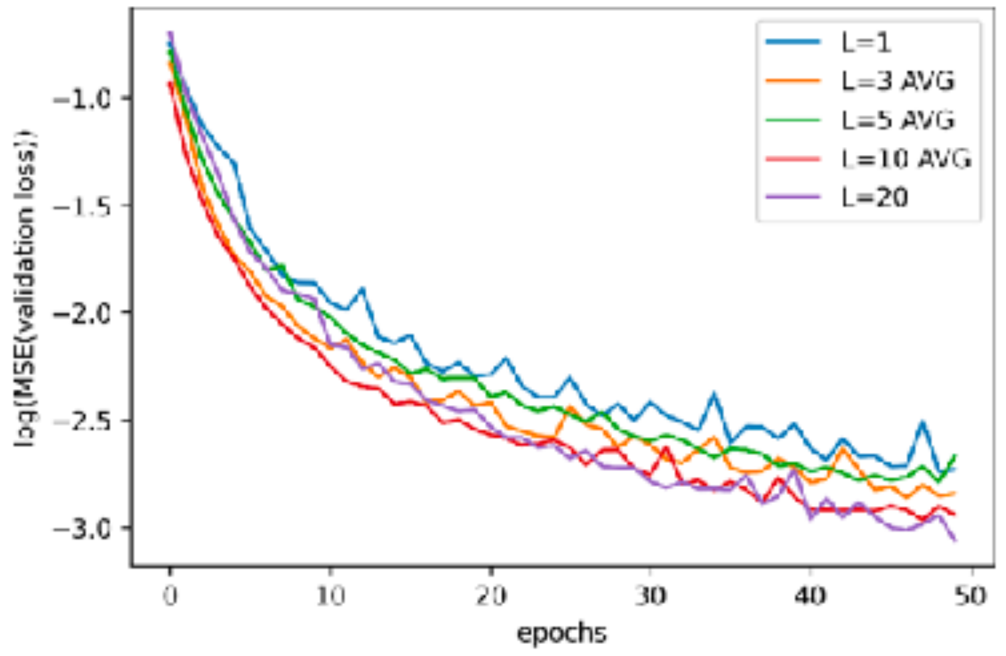
Empirical:

Numerics show what one would expect.
For quantum problems more terms improve
Classical not as much

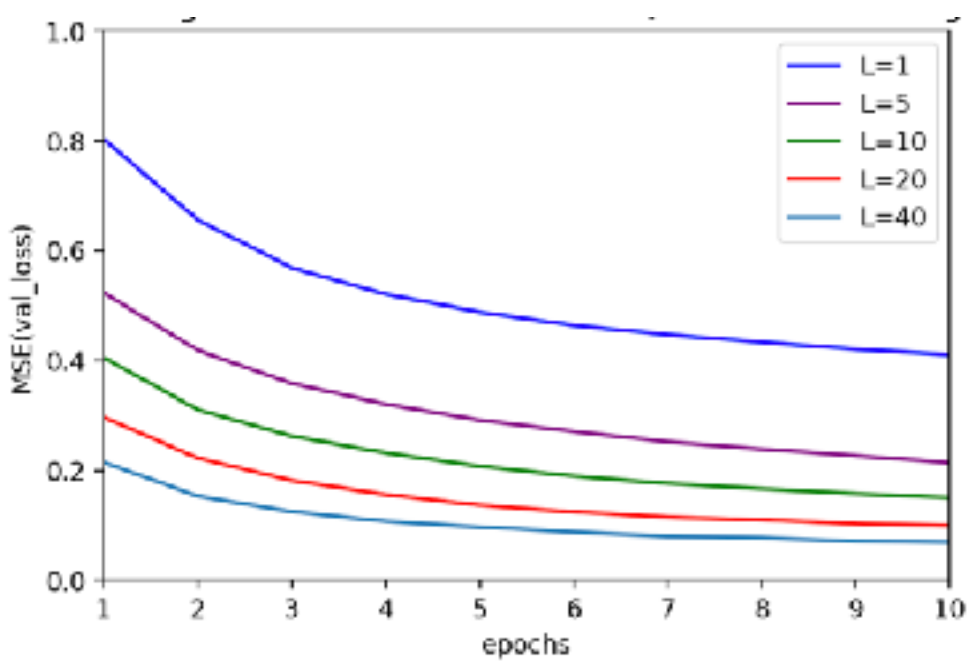


64* qubit case

MNIST



Synthetic



* 64 > 53, just saying

A take home

All models can be understood via explicit picture.
But each types of QML model has advantages in certain aspects. No one to rule all.

Slicing circuit is a promising starting point to devise hybrid models.

QC has been mostly driven by theory. Real-world ML by empirical success; heuristic. Domain specific.

Theory can be extremely practical, but QML is unlikely to show all it can show using theory alone.

Need case studies *and tools to do them*. Domain analyses…

Quantum Machine Learning Beyond Kernel Methods, arxiv:2110.13162
High Dimensional Quantum Machine Learning With Small Quantum Computers, arXiv:2203.13739

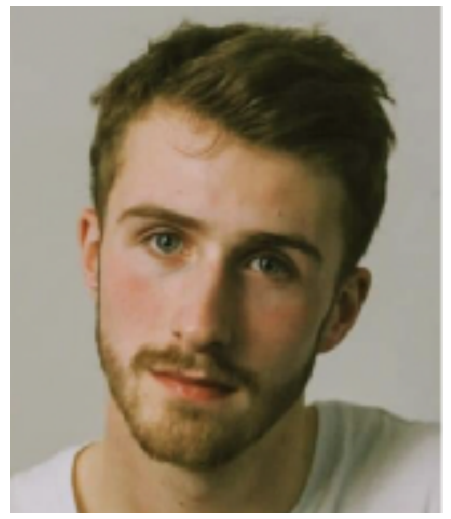
https://www.tensorflow.org/quantum/tutorials/quantum_reinforcement_learning

With:

Sofiené Jerbi

Andrea Skolik

Casper Gyurik

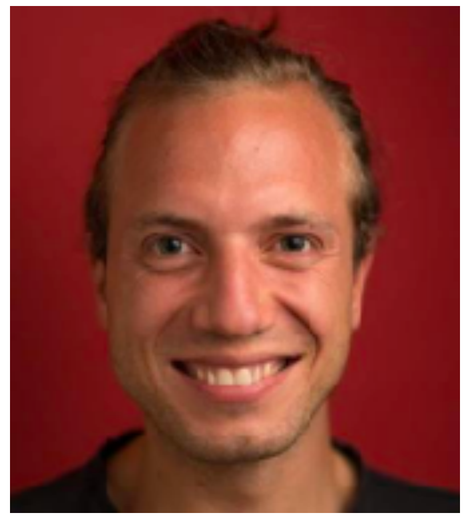


Simon Marshall

Hans Briegel
(Innsbruck)

Hendrik Poulsen Nautrup
(Innsbruck)

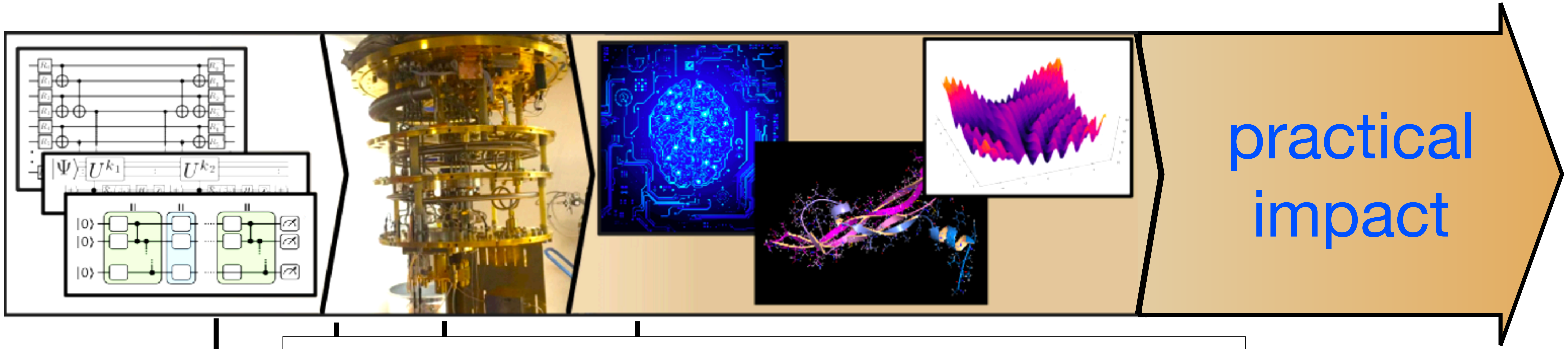
Lukas Fiderer
(Innsbruck)



Jonas Kübler
(MPI for Intelligent Systems)

aQa: Open PhD and PostDoc positions - drop me an email!

aQa Leiden



Development of novel quantum algorithms

Hardware-aware designs and optimizations

AI and ML for quantum information

with experimental groups (Delft)

Application-specific aspects

with industrial partners

References:

[1] 1907.02085
[2] 2010.02174
[3] 2106.03880
[4] 2002.01490
[5] 2103.03139
[6] 1803.07128
[7] 2101.11020
[8] 2011.01938
[9] 1506.01396
[10] 1904.00102
[11] 1804.11326