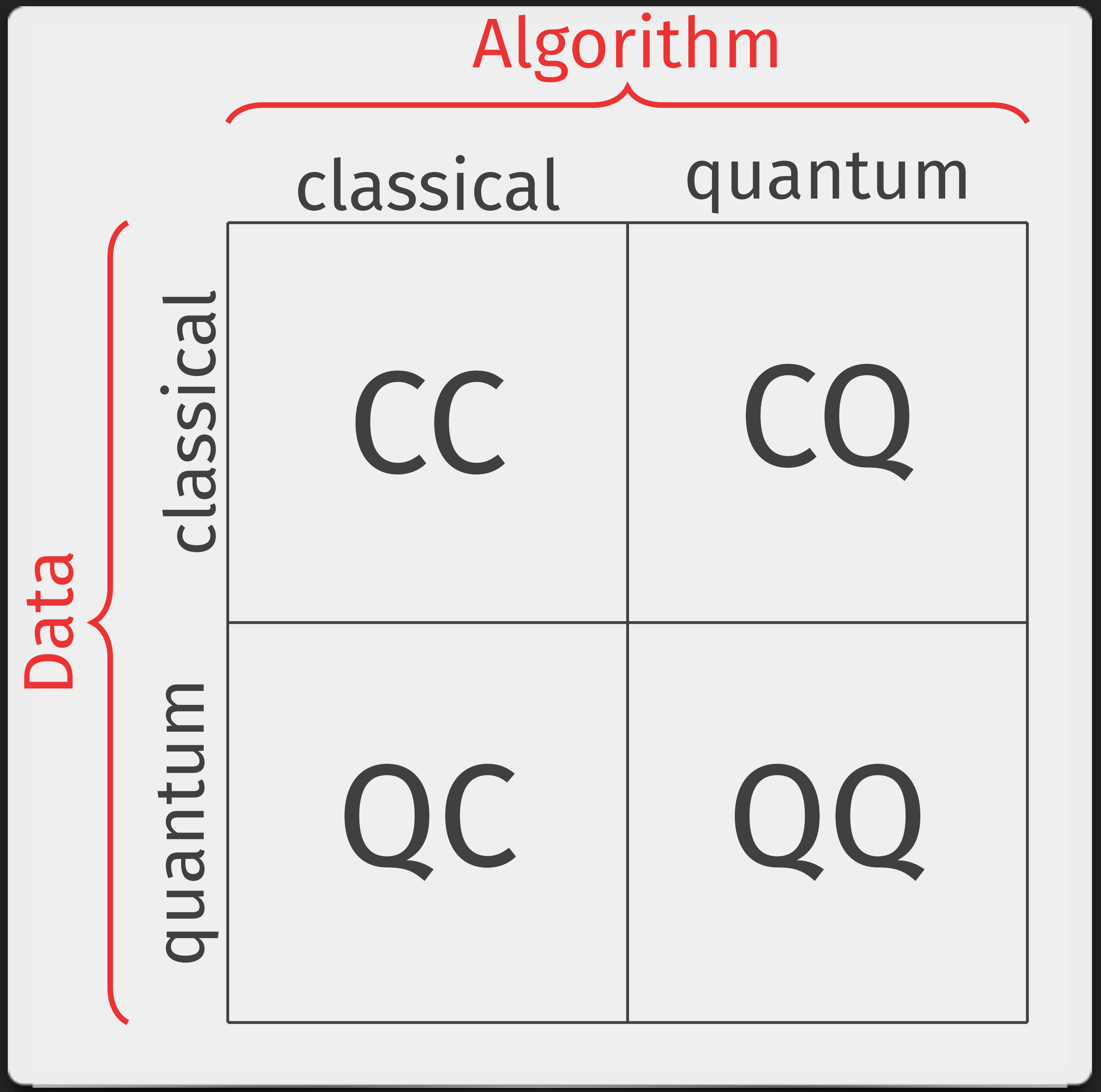
# OPENING THE BLACK BOX OF QUANTUM MACHINE LEARNING

LEONARDO BANCHI — UNIVERSITY OF FLORENCE & INFN FIRENZE

# OUTLINE OF QUANTUM MACHINE LEARNING

- Something has to be **quantum** (e.g. data, or algorithm)
- **Learning** problem
- We need to use a **machine**, the problem is too complex for pen and paper calculations
- **Data driven**: we learn from examples*, defining a mathematical model is often impossible
- **Current Personal interests: theoretical guarantees** (what kind of problems can be learnt easily? Why?)

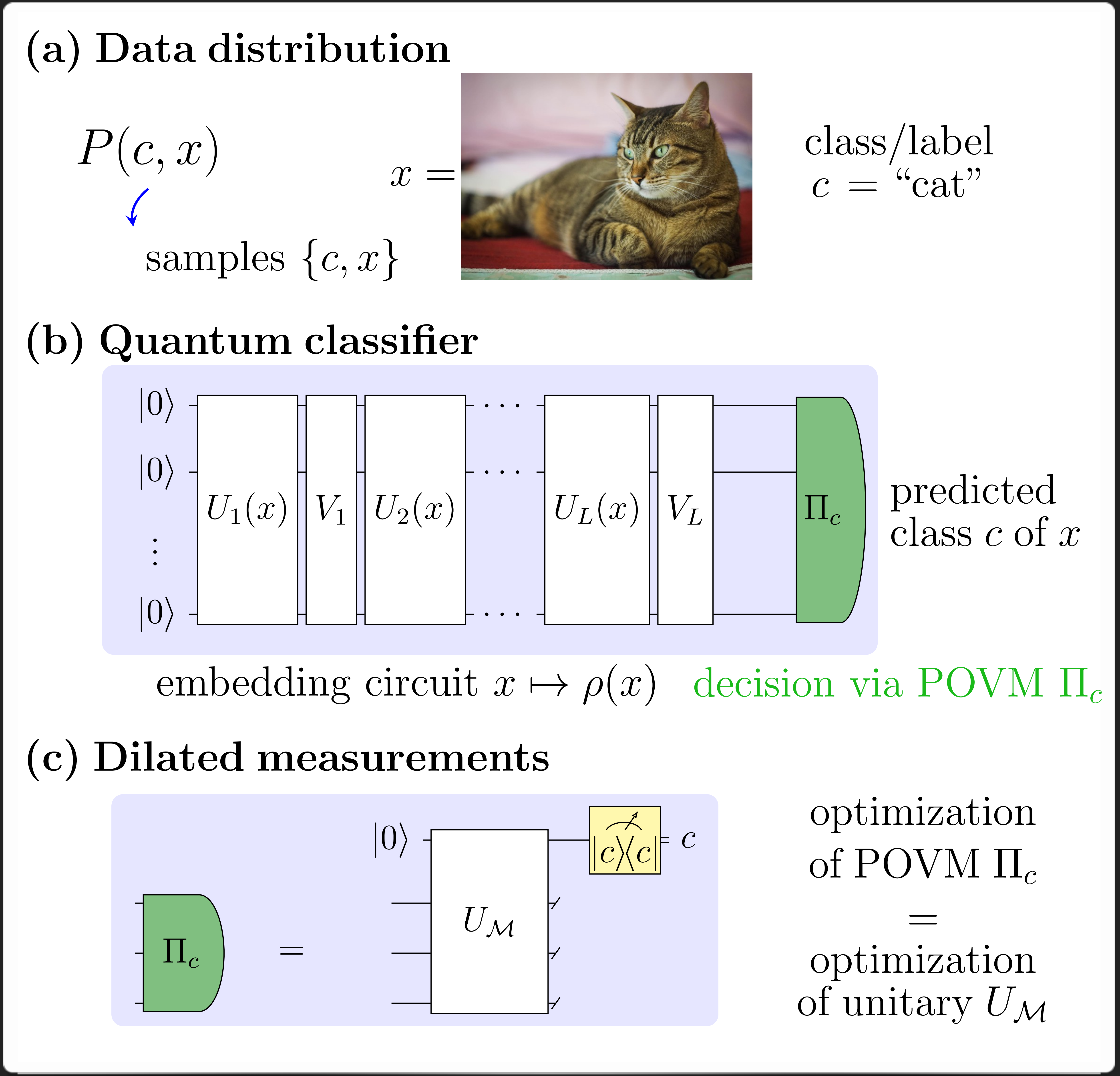


# CQ: QUANTUM LEARNING OF CLASSICAL DATA IN NISQ DEVICES

- Classical data is mapped onto a quantum state using a parametric quantum circuit, namely a composition of unitary gates that depend on classical variables.

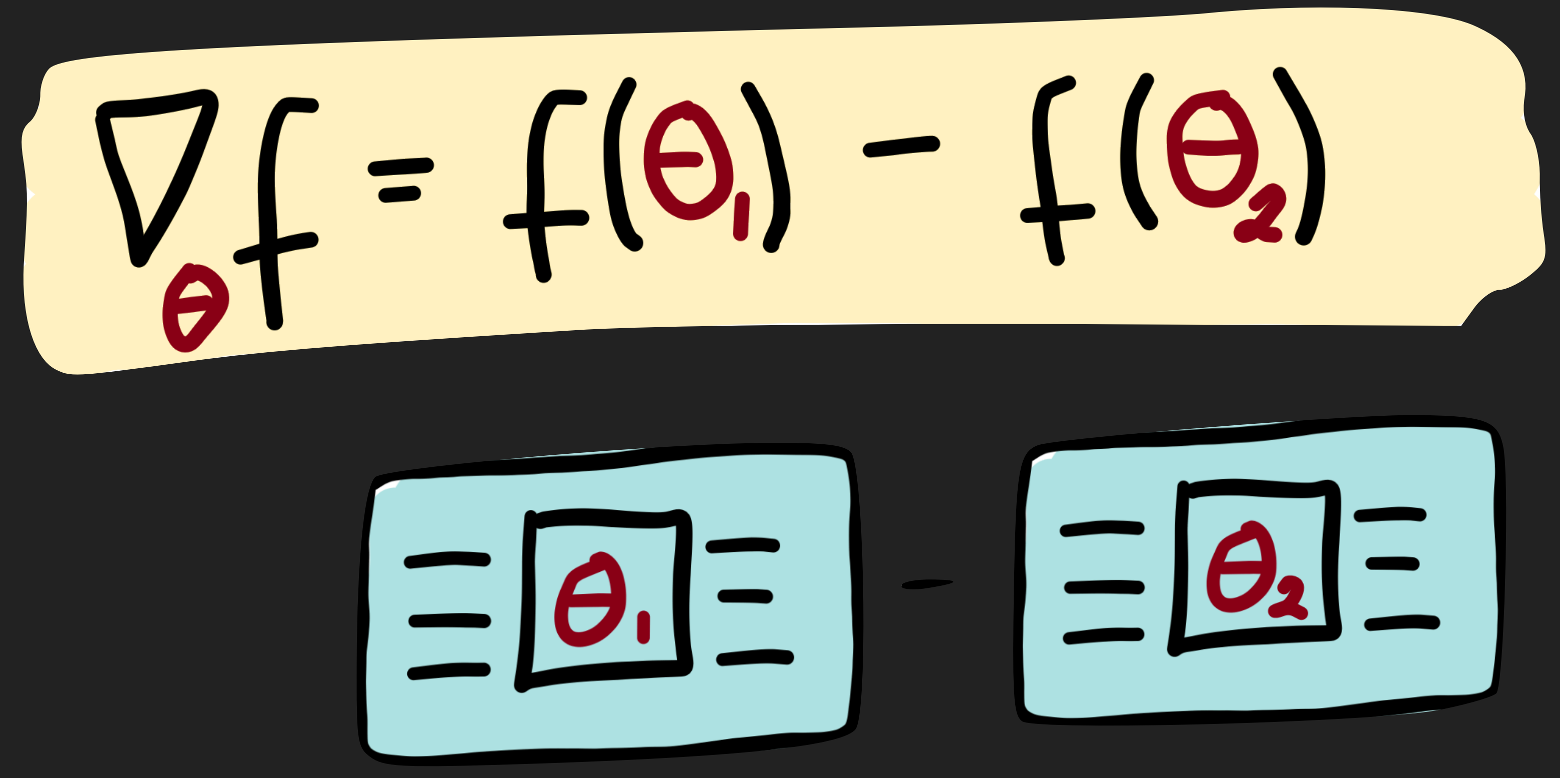
$$\hat{U}_j(x) = \exp\left(i\, x \cdot \hat{H}_j\right)$$

- Once data are loaded into a quantum register, we apply a classification unitary $\hat{U}_{\mathscr{M}}(\theta)$, with external parameters $\theta$
- A measurement in the computational basis (probabilistically) determines the class
- Parametric Quantum Circuits ≡ Quantum Neural Networks

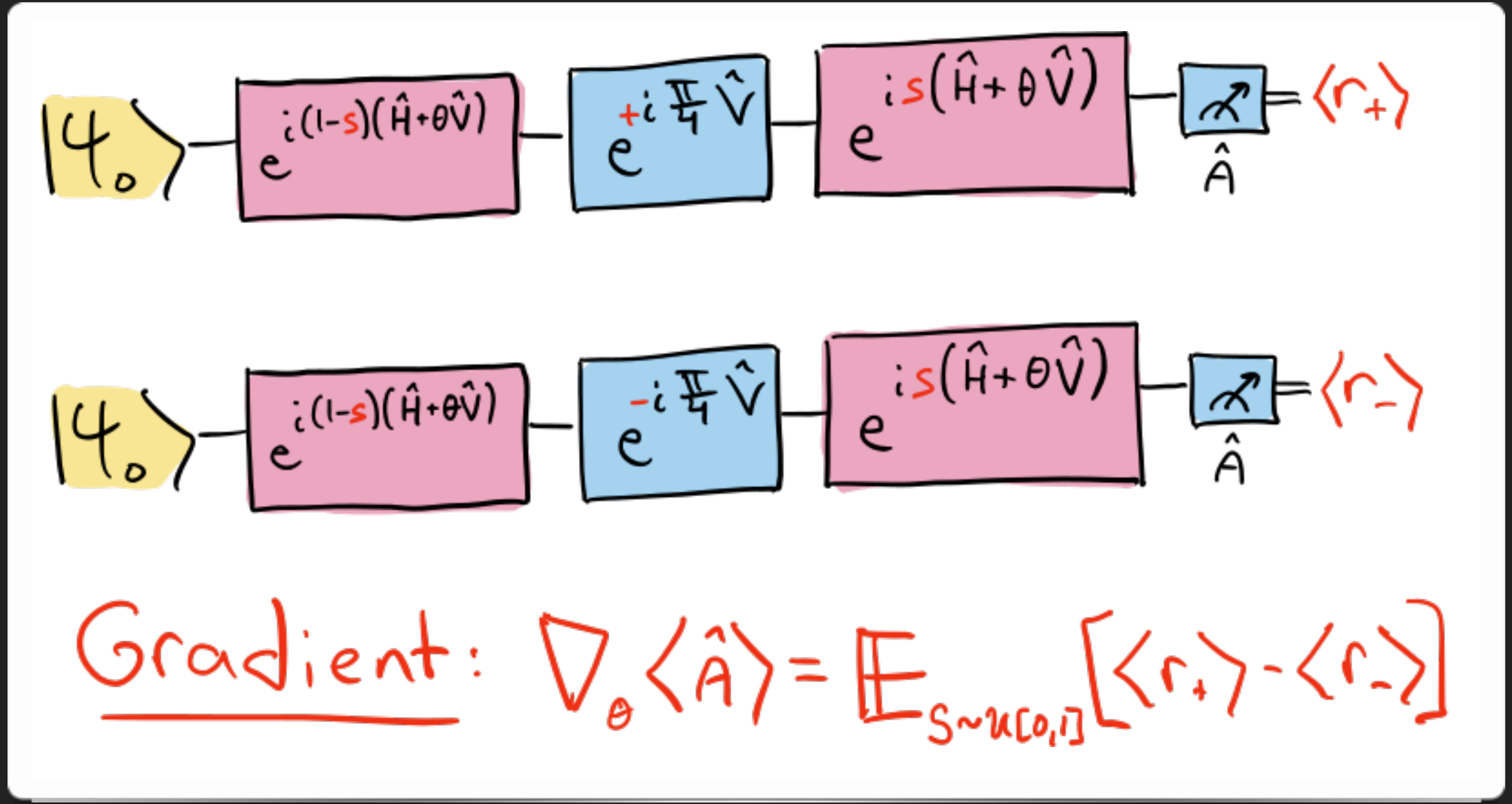


# HOW DO WE OPTIMISE QUANTUM NEURAL NETWORKS? GRADIENTS!

- Parameter Shift Rule / Hadamard test for $e^{i\theta\hat{\sigma}}$ gates
  - K Mitarai et al. - Physical Review A, 2018
  - M Schuld et al. - Physical Review A, 2019




- Stochastic PSR for **general Hamiltonian evolution** $e^{i(\hat{H}+\theta\hat{V})}$
  - L Banchi, GE Crooks, Quantum 5, 386 (2021)

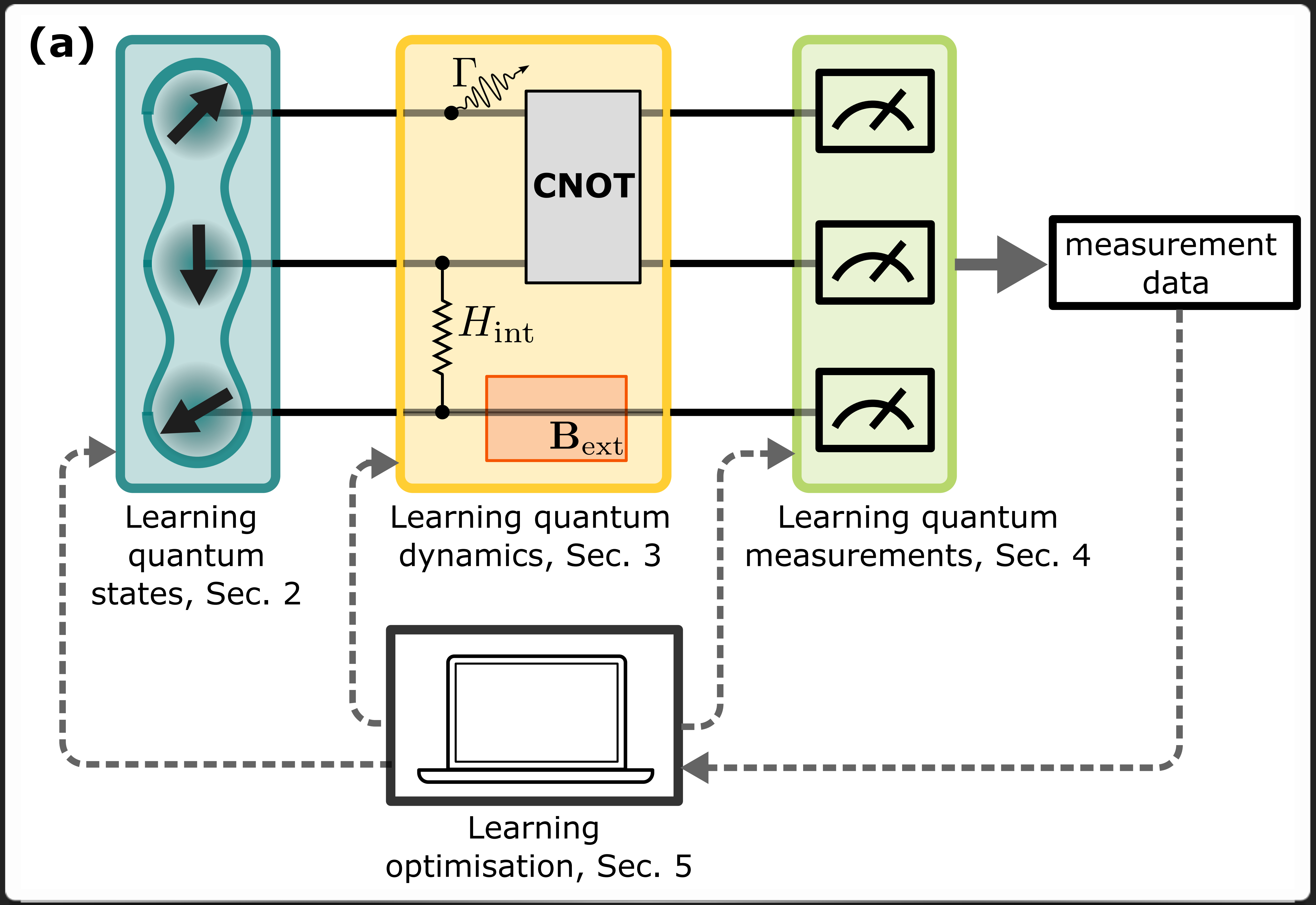



- Continuous variable systems / GBS distribution
  - N Killoran, et al. Phys. Rev. Research (2019)
  - L Banchi et al., Phys. Rev. A 102, 012417 (2020)

# QML WITH "QUANTUM DATA"

- For **classical data** we want to use a quantum computer, hoping to get a **computational advantage**
- Example problems with **quantum data**: quantum **sensing**, quantum state and process **tomography**

| Data \ Algorithm | classical | quantum |
|---|---|---|
| classical | CC | CQ |
| quantum | QC | QQ |

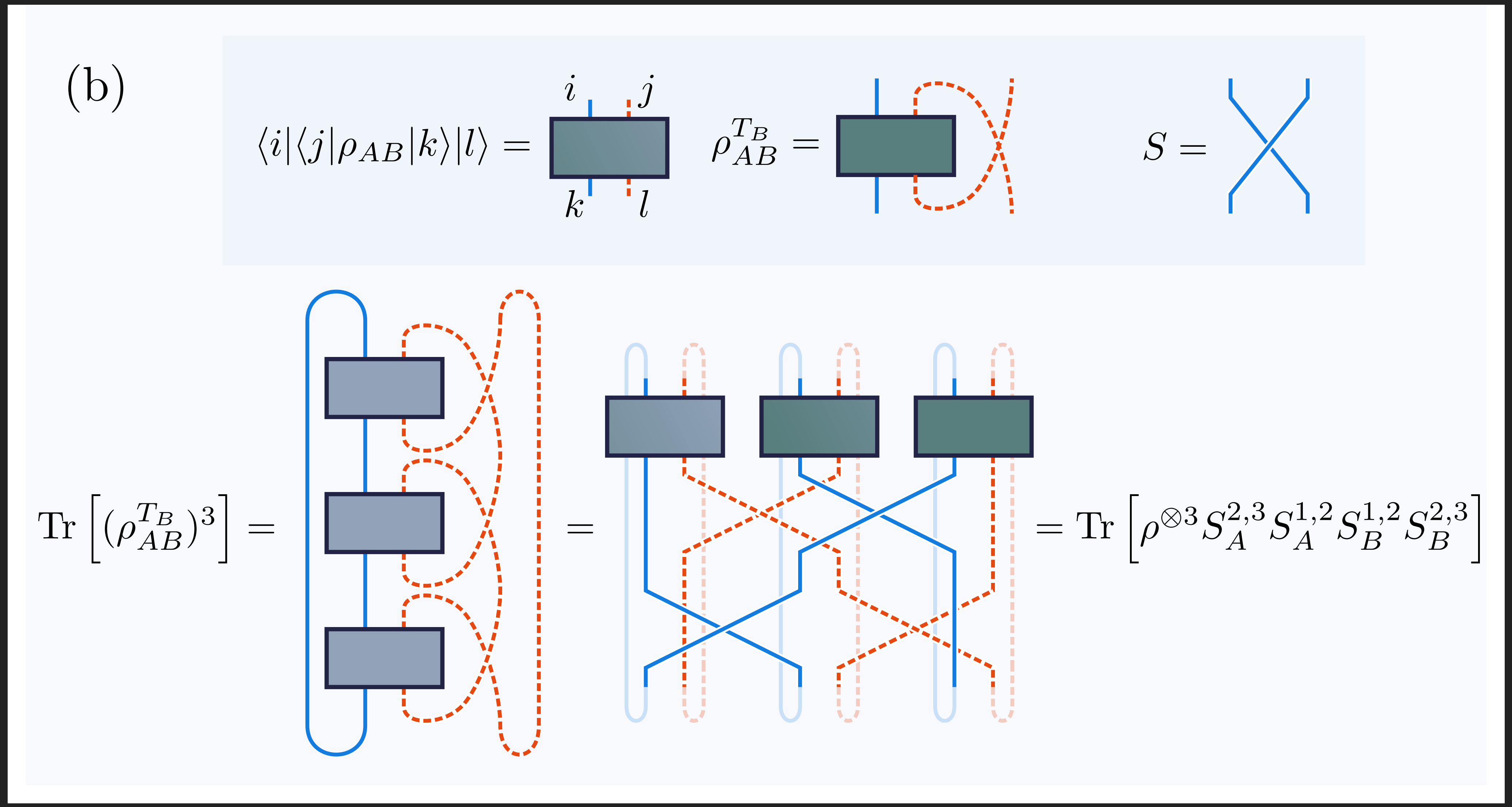



- Our review on QC

V Gebhart [...] L Banchi [...], arXiv:2207.00298
Nature Reviews Physics (accepted)

# MANY-BODY PHYSICS (QC/QQ)

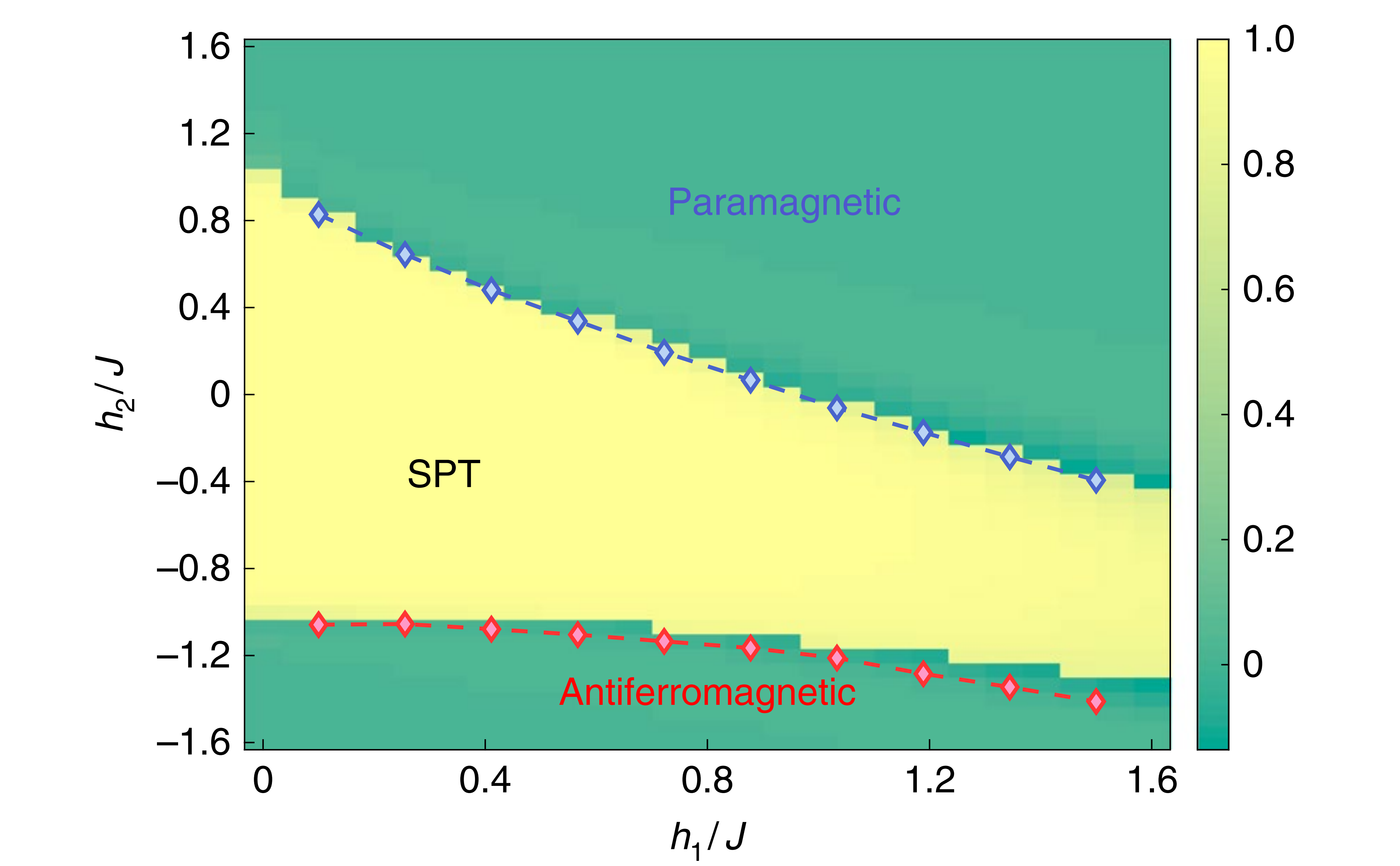
- Many-Body Entanglement Measurement from PPT-moments $\mathrm{Tr}\left[(\rho_{AB}^{T_B})^n\right]$
  - J Gray, L Banchi, A Bayat, S Bose, Phys. Rev. Lett. 121, 150503 (2018)



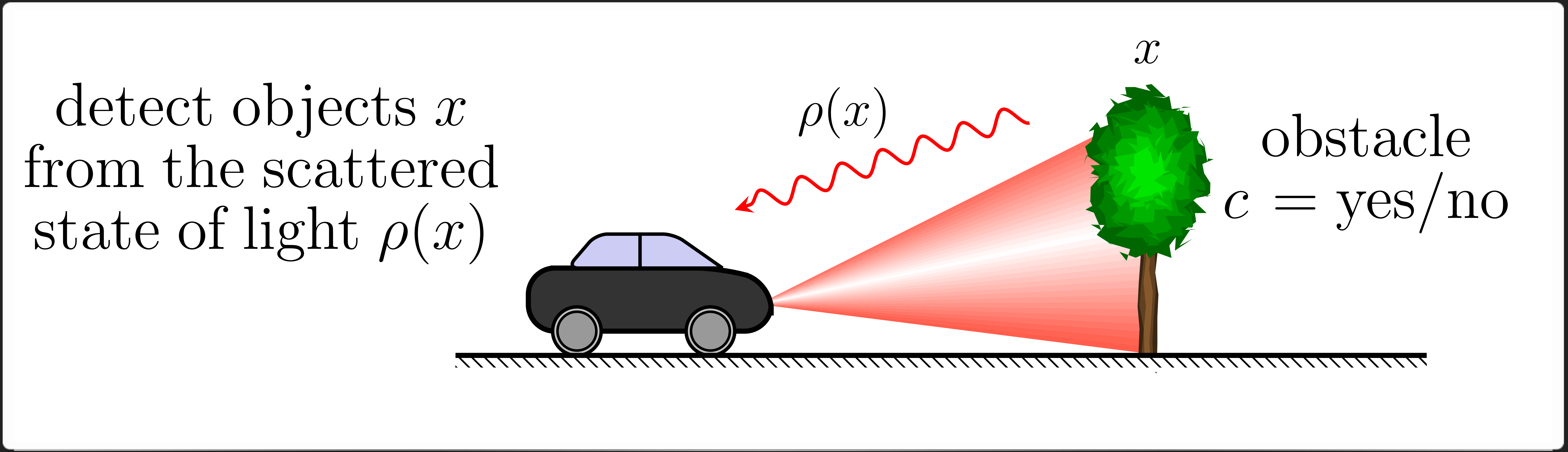

- Quantum Phase Recognition
  - I Cong, S Choi, MD Lukin, Nature Physics 15, 1273 (2019)

$$H = -J \sum_{i=1}^{N-2} Z_i X_{i+1} Z_{i+2} - h_1 \sum_{i=1}^{N} X_i - h_2 \sum_{i=1}^{N-1} X_i X_{i+1}$$

  - L Banchi, J Pereira, S Pirandola, PRX Quantum 2, 040321 (2021)



# QQ: QUANTUM CHANNEL DISCRIMINATION

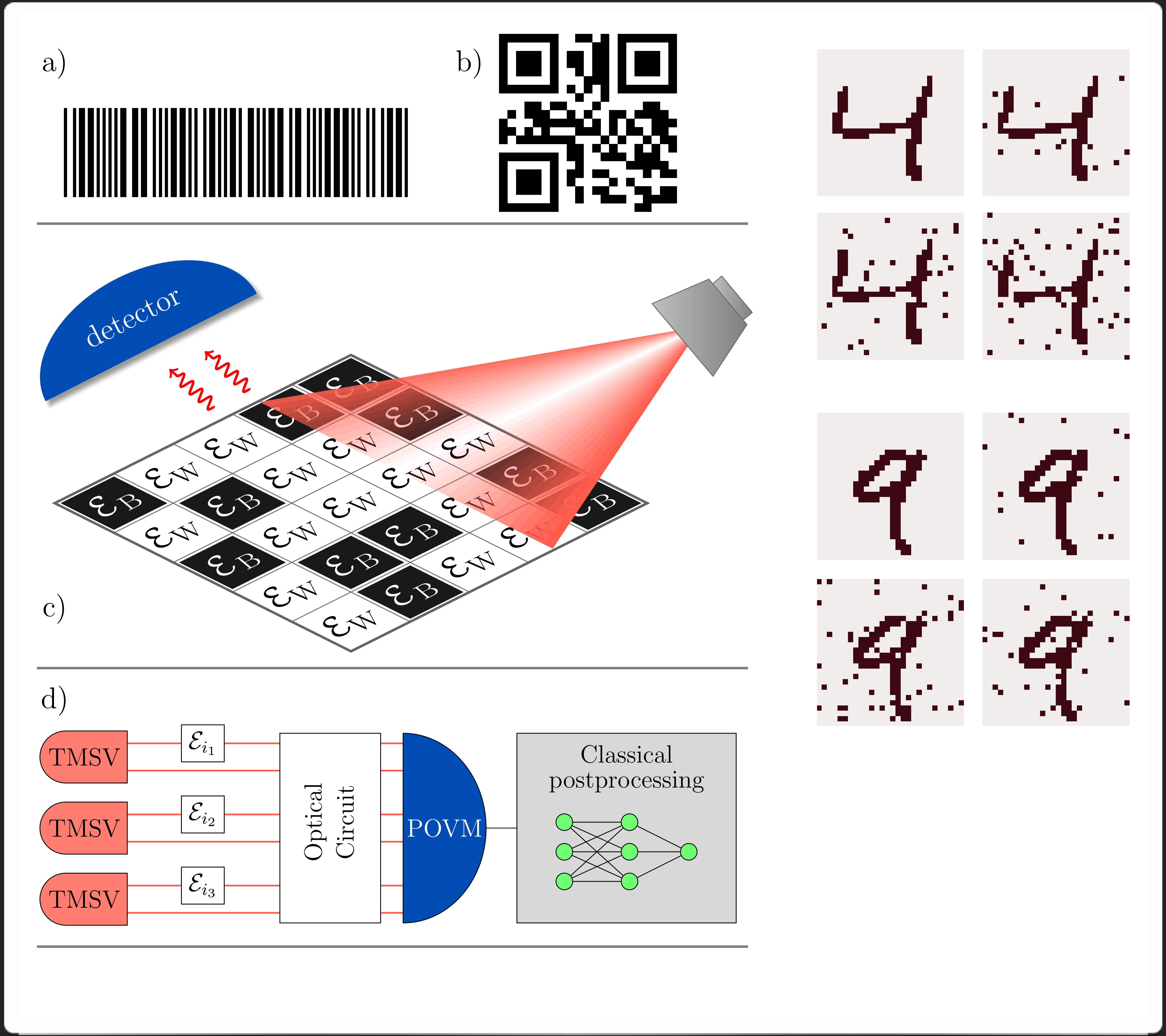



- Images live in the physical world
- Optimise the (entangled) input probe state of light and the detection POVM

# QUANTUM BARCODES AND PATTERN RECOGNITION

- Barcode classification must identify each pixel correctly
- Handwriting classification is easier as errors are tolerated!
- $$\text{error} \simeq F(\rho_{\text{black}}, \rho_{\text{white}})^{\text{Hamming}_{4\leftrightarrow 9}}$$


- L. Banchi, Q. Zhuang, S. Pirandola, Phys. Rev. Applied 14, 064026 (2020)
- C Harney, L Banchi, S Pirandola, Phys. Rev. A 103, 052406 (2021)
- JL Pereira, L Banchi, Q Zhuang, S Pirandola, Phys. Rev. A 103, 042614 (2021)

# QC: LEARNING QUANTUM SYSTEMS USING CLASSICAL ML

npj Quantum Information

*npj Quantum Information* (2016) **2,** 16019; doi:10.1038/npjqi.2016.19; published online 19 July 2016

www.nature.com/npjqi

ARTICLE OPEN

## Quantum gate learning in qubit networks: Toffoli gate without time-dependent control

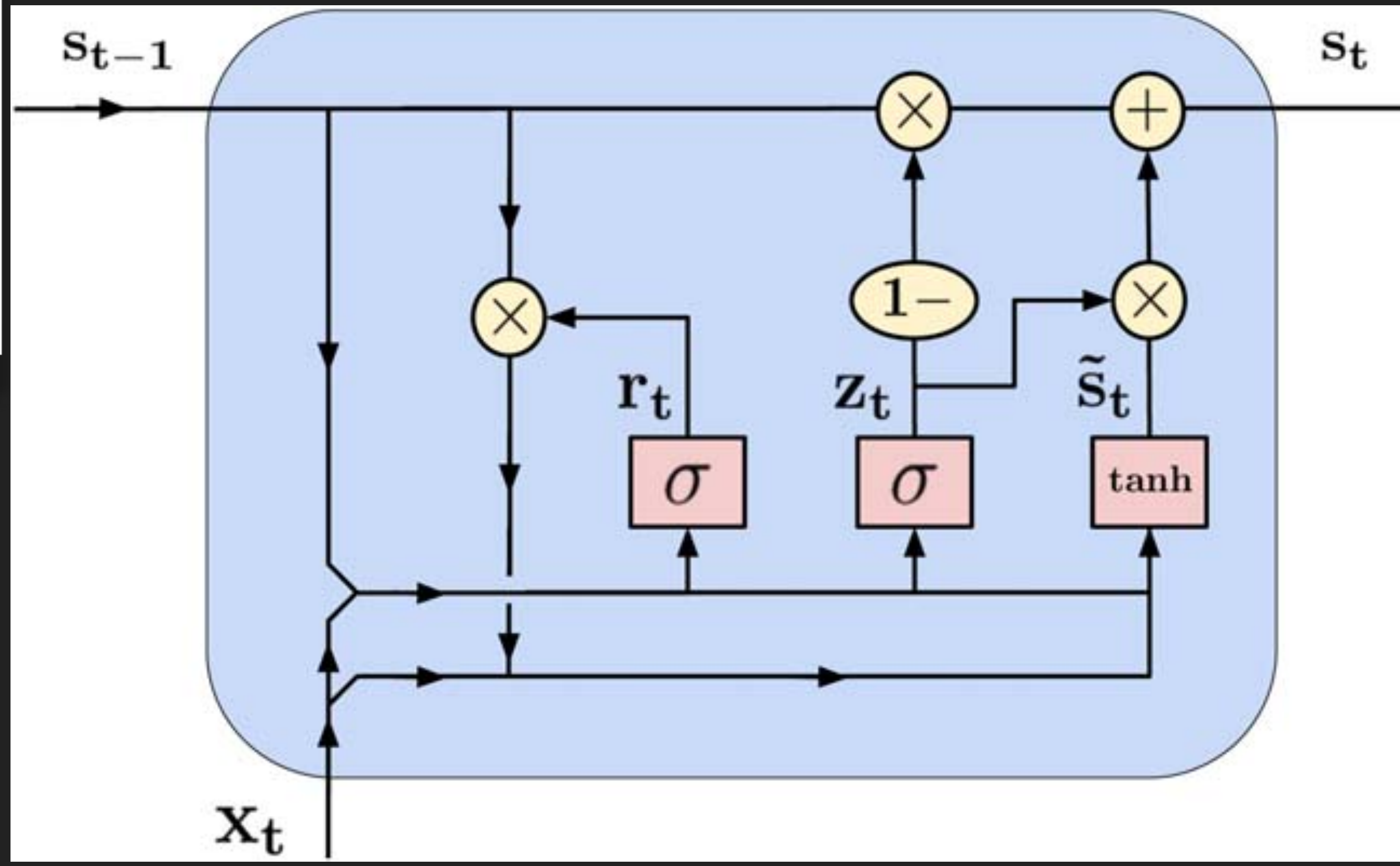
Leonardo Banchi[1], Nicola Pancotti[2,3] and Sougato Bose[1]

## Modelling non-markovian quantum processes with recurrent neural networks

IOP Publishing *New J. Phys.* **20** (2018) 123030

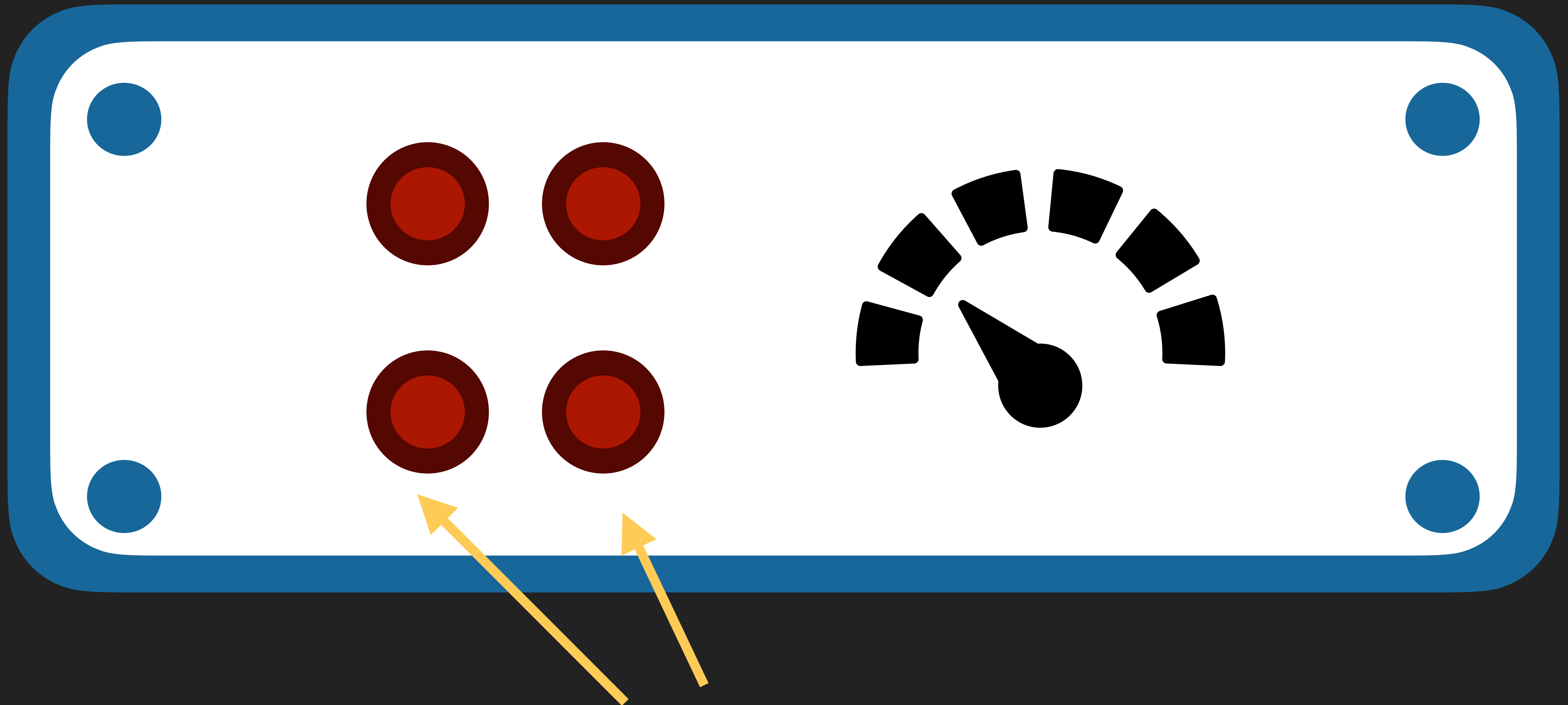
Leonardo Banchi[1,5], Edward Grant[2,4], Andrea Rocchetto[2,3] and Simone Severini[2]

$$\mathcal{L}_{\leqslant t}[\rho] = -\mathrm{i}[H + H^{LS}_{\leqslant t}, \rho] + \sum_{\mu}\left[L^{\mu}_{\leqslant t}\rho L^{\mu\dagger}_{\leqslant t} - \frac{1}{2}\{L^{\mu\dagger}_{\leqslant t}L^{\mu}_{\leqslant t}, \rho\}\right],$$



# DOES THE MODEL GENERALISE?

# FOUNDATIONAL QUESTION: WHAT CAN WE EXPECT?

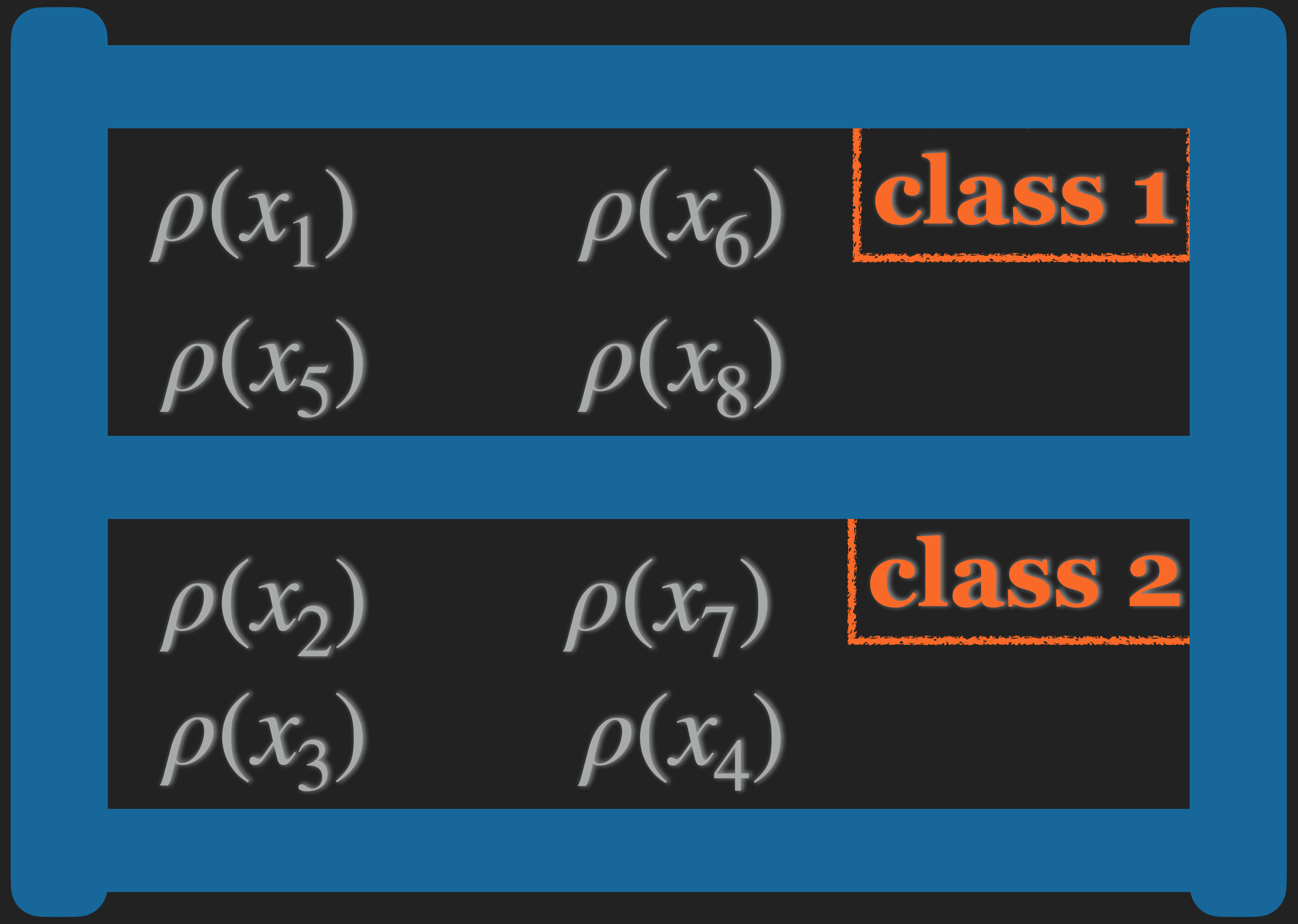


$x$

$\rho(x)$

$c = 1$ or $c = 2$ ?

- We have a training set $\mathcal{T}$ made of $T$ correctly classified states $\mathcal{T} = \{(\rho(x_t), c_t)$ for $t = 1,\ldots,T\}$
- We can empirically check generalisation using a testing set $\mathcal{T}'$ with $T'$ correctly classified states
- **How difficult is to define a reliable classier, given the available data?**



- Empirical loss / training error

$$R_{\mathscr{T}}(\Pi, \rho) = \frac{1}{T} \sum_{(c_k, x_k) \in \mathscr{T}} \sum_{c \neq c_k} \mathrm{Tr}\left[\Pi_c \rho(x_k)\right] = 1 - \frac{1}{T} \sum_{(c_k, x_k) \in \mathscr{T}} \mathrm{Tr}\left[\Pi_{c_k} \rho(x_k)\right]$$

- Abstract classification error

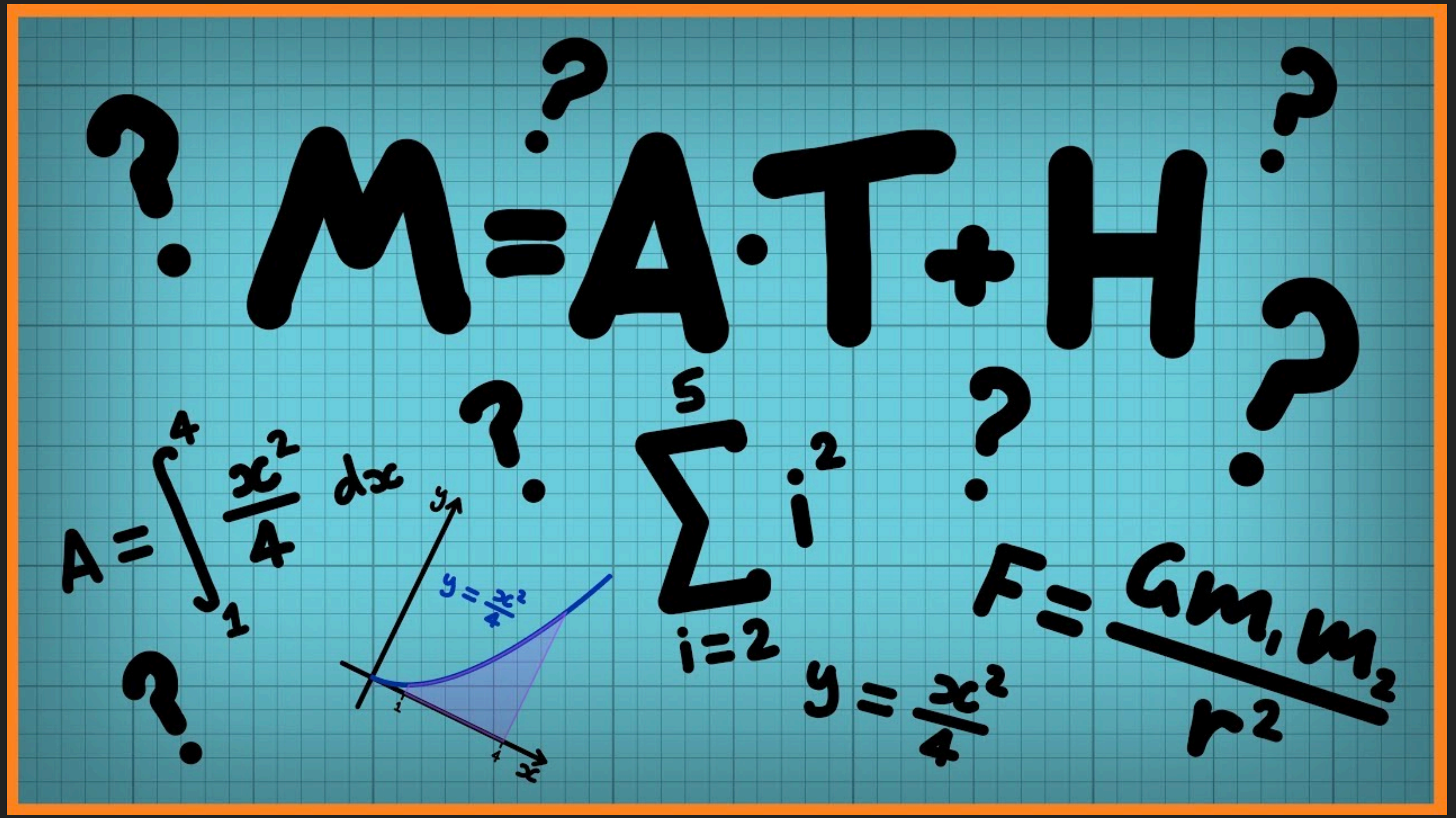
$$R(\Pi, \rho) = \mathbb{E}_{(c,x) \sim P(c,x)} \left( \sum_{c \neq \tilde{c}} \mathrm{Tr}\left[\Pi_{\tilde{c}} \rho(x)\right] \right) = 1 - \mathbb{E}_{(c,x) \sim P(c,x)} \mathrm{Tr}\left[\Pi_c \rho(x)\right]$$

- Optimal empirical measurement $\Pi^{\mathscr{T}} = \mathrm{argmin}_{\Pi} R_{\mathscr{T}}(\Pi, \rho)$

- Real optimal $\Pi^* = \mathrm{argmin}_{\Pi} R(\Pi, \rho)$

- Testing error

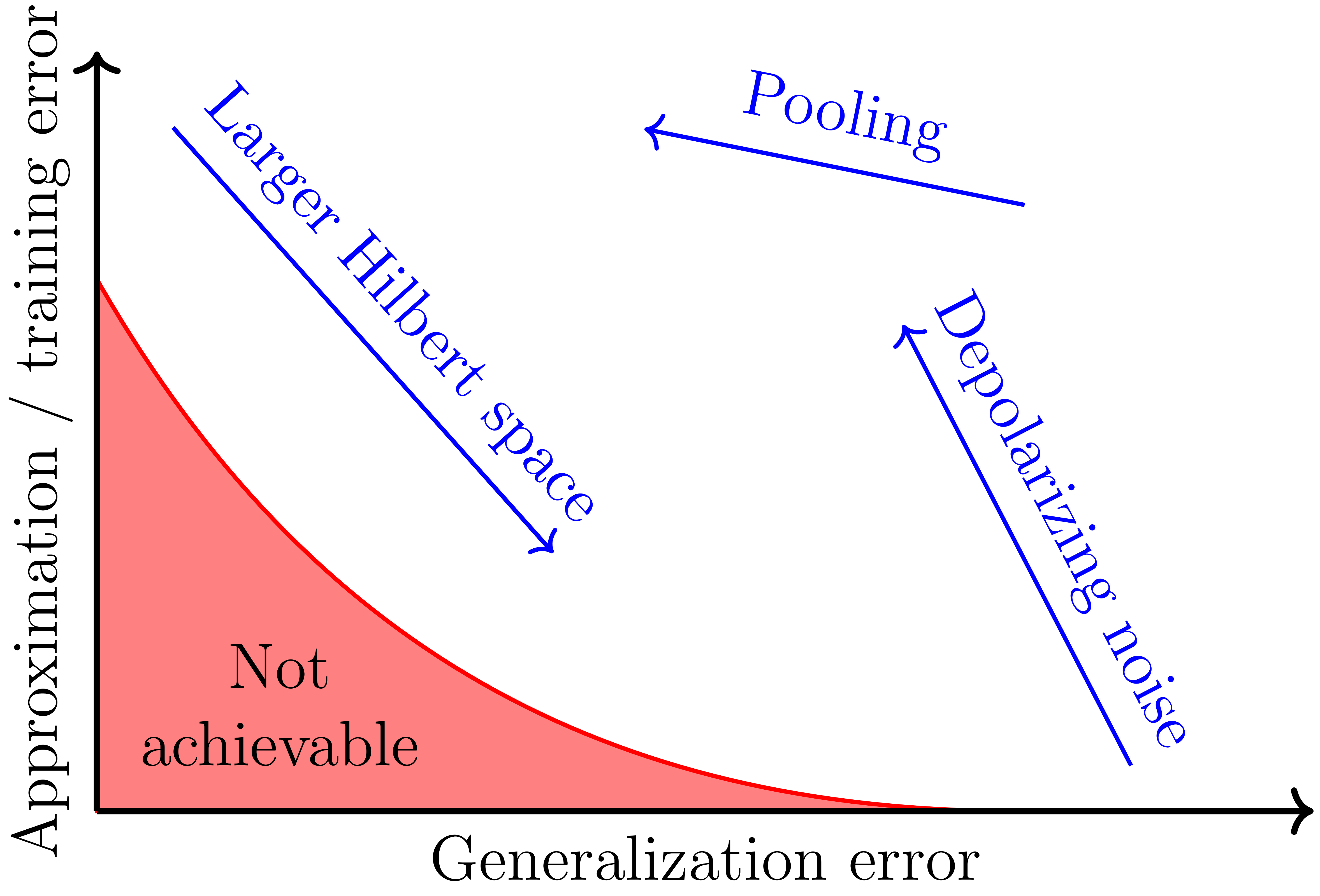
$$R_{\mathscr{T}'}(\Pi_{\mathscr{T}}, \rho)$$

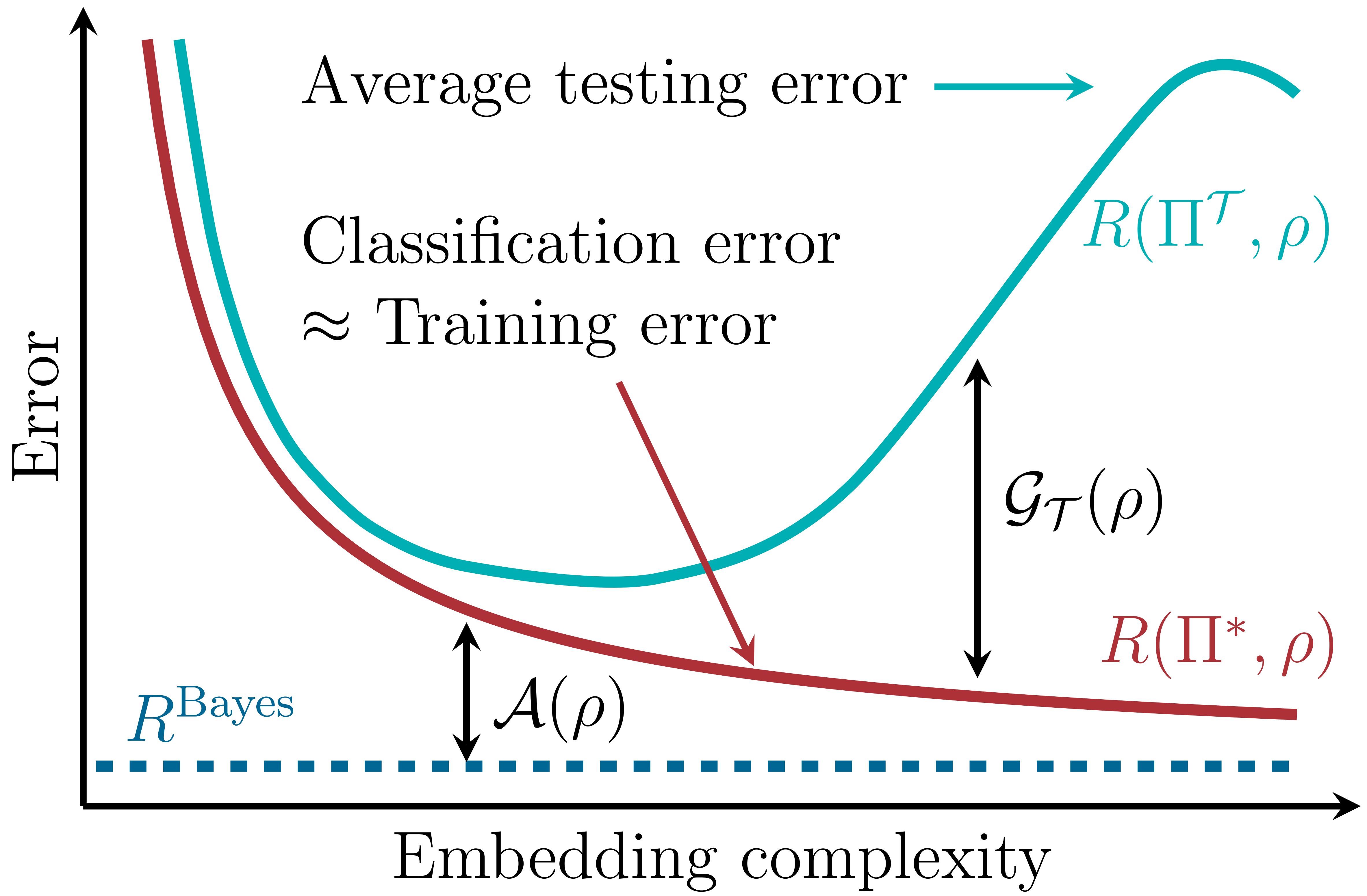


# LEARNING WITH QUANTUM STATES

L BANCHI, J PEREIRA, S PIRANDOLA
PRX QUANTUM 2, 040321 (2021)

$$\mathcal{G}_{\mathcal{T}} \approx \sqrt{\frac{2^{I_2(X:Q)}}{T}} \qquad \mathcal{A} = K - \frac{2^{I(C:Q)}}{N_c} \qquad \rho_{CXQ} = \sum_{cx} P(c,x)\,|cx\rangle\langle cx| \otimes \rho(x)$$



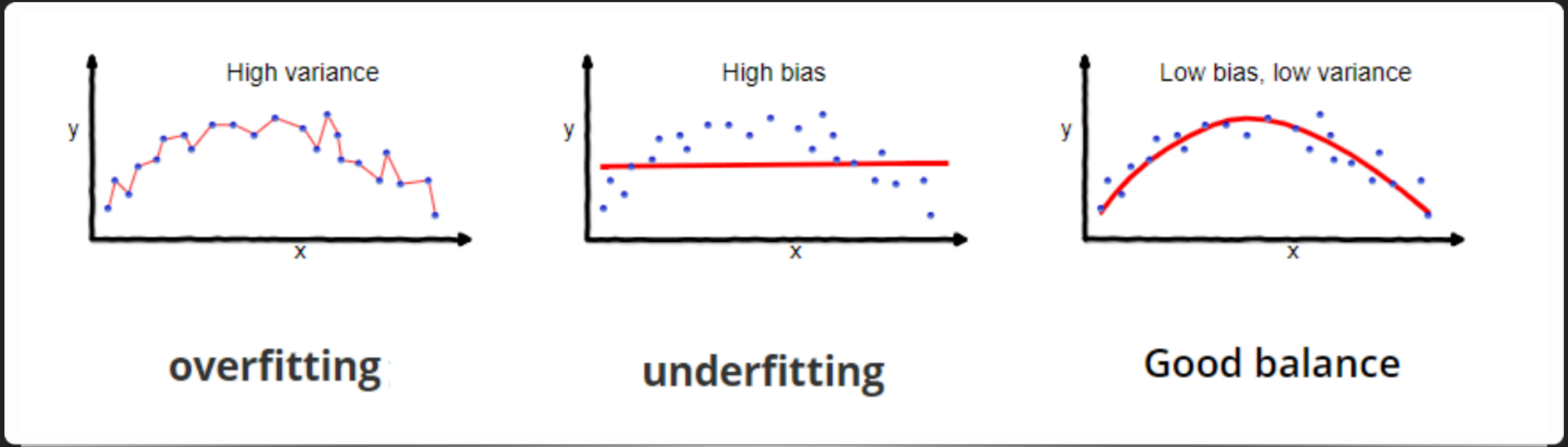



# STATISTICAL LEARNING THEORY

- The generalisation error $\mathscr{G}$ is the difference between the (average) training and testing error

- A fundamental result of statistical learning theory binds $\mathscr{G}$ using the Rademacher complexity

$$\mathcal{G} \leq 4\mathcal{C}_T(\mathcal{H}) + \sqrt{\frac{2\log(1/\delta)}{T}}, \qquad \mathcal{C}_T(\mathcal{H}) := \mathop{\mathbb{E}}_{\mathcal{T}\sim P^T} \mathop{\mathbb{E}}_{\boldsymbol{\sigma}} \left[ \sup_{h\in H} \frac{1}{T} \sum_{k=1}^{T} \sigma_k \ell_h(c_k, x_k) \right],$$



# TAKE HOME MESSAGE

- Computable upper bound via information theoretic quantities
- Generalisation favoured by **discarding information** about the input that is irrelevant for predicting the output. Quantum compression?
- Algorithm: variational information bottleneck
- Quantum convolutional neural networks iteratively discard information via **pooling** layers.
- Algorithmic-independent data-dependent bounds
- Greater physical insight (wait for QQ!)

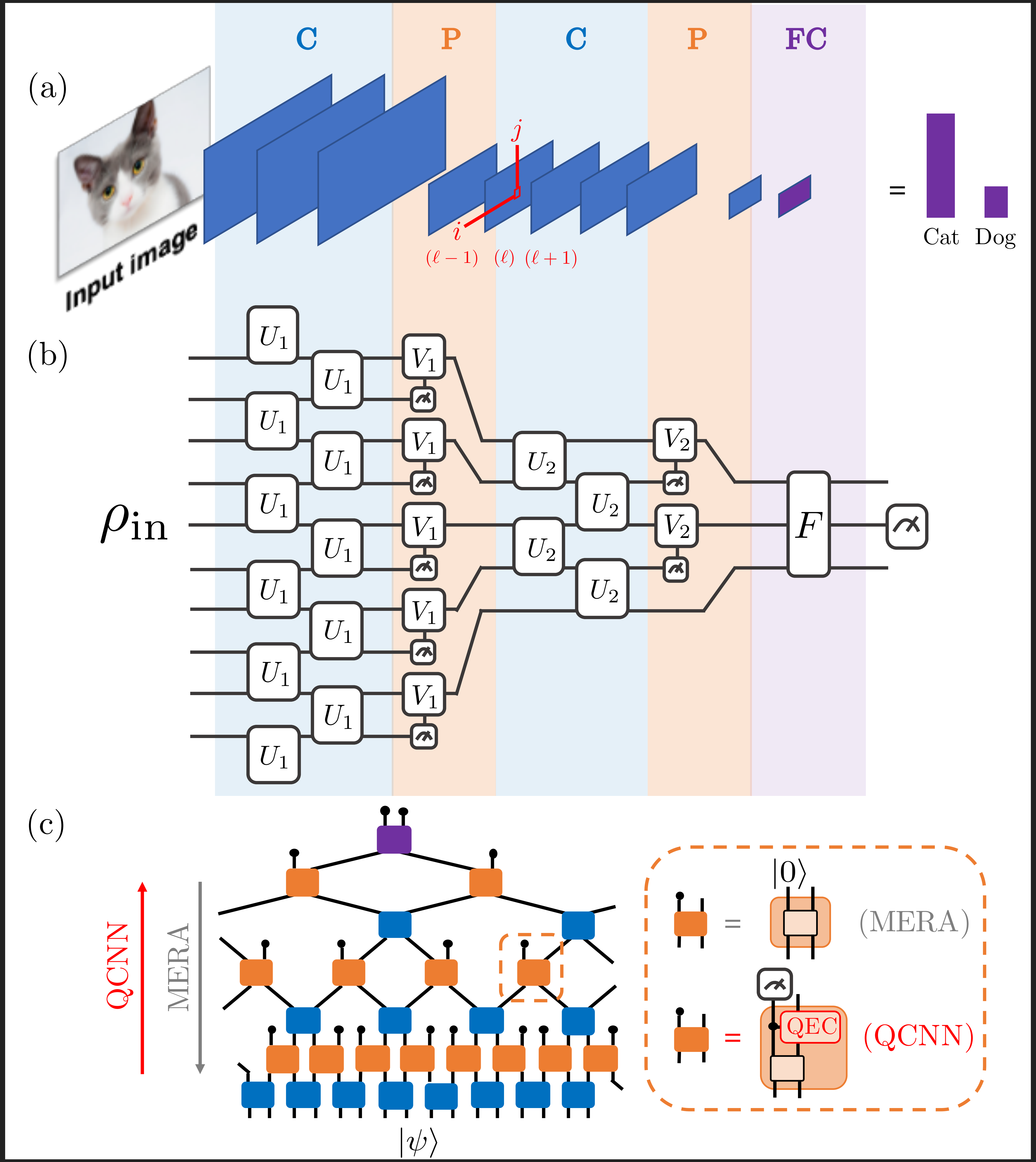


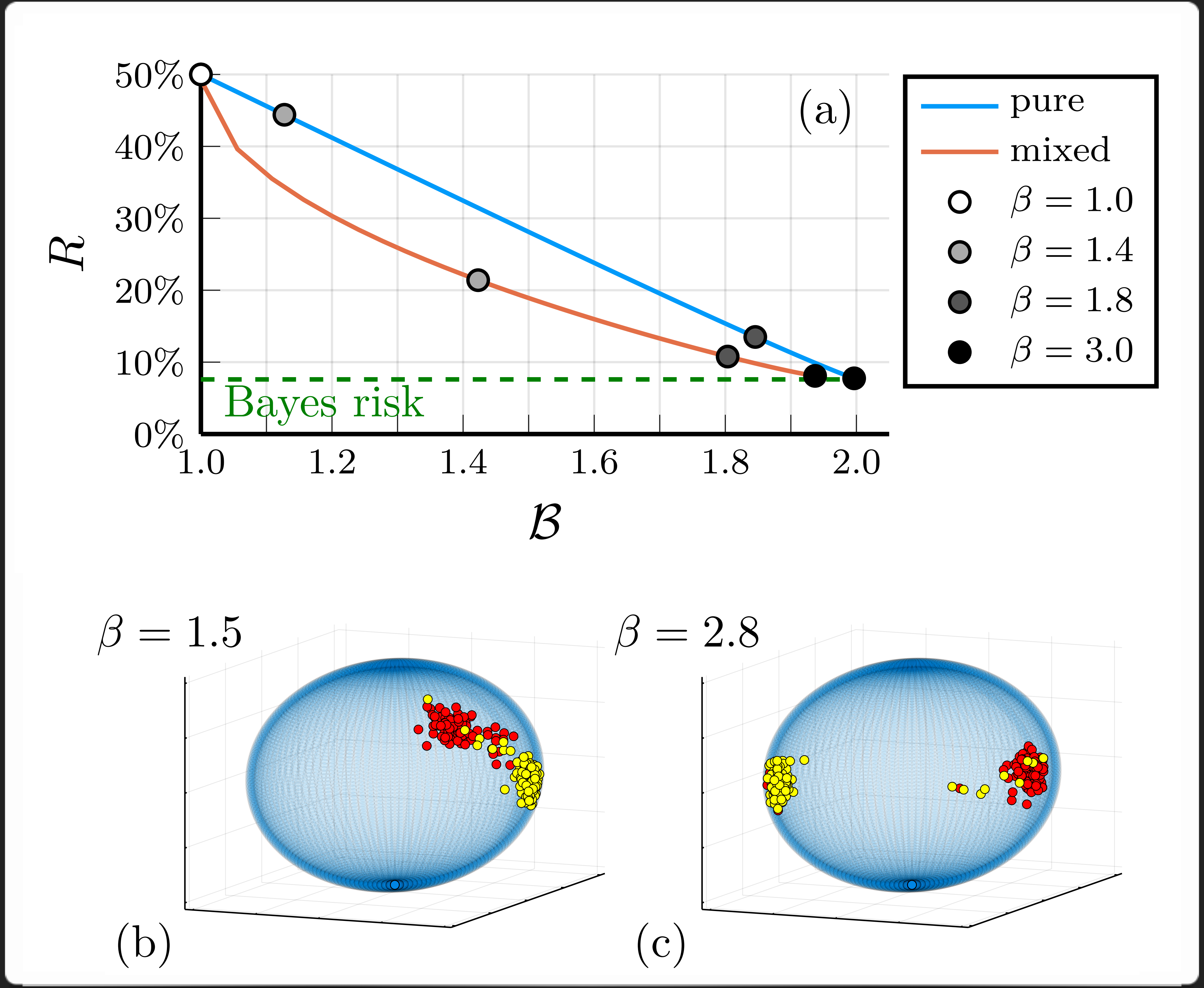

image from: I Cong et al. Nature Physics (2019)

# APPLICATIONS

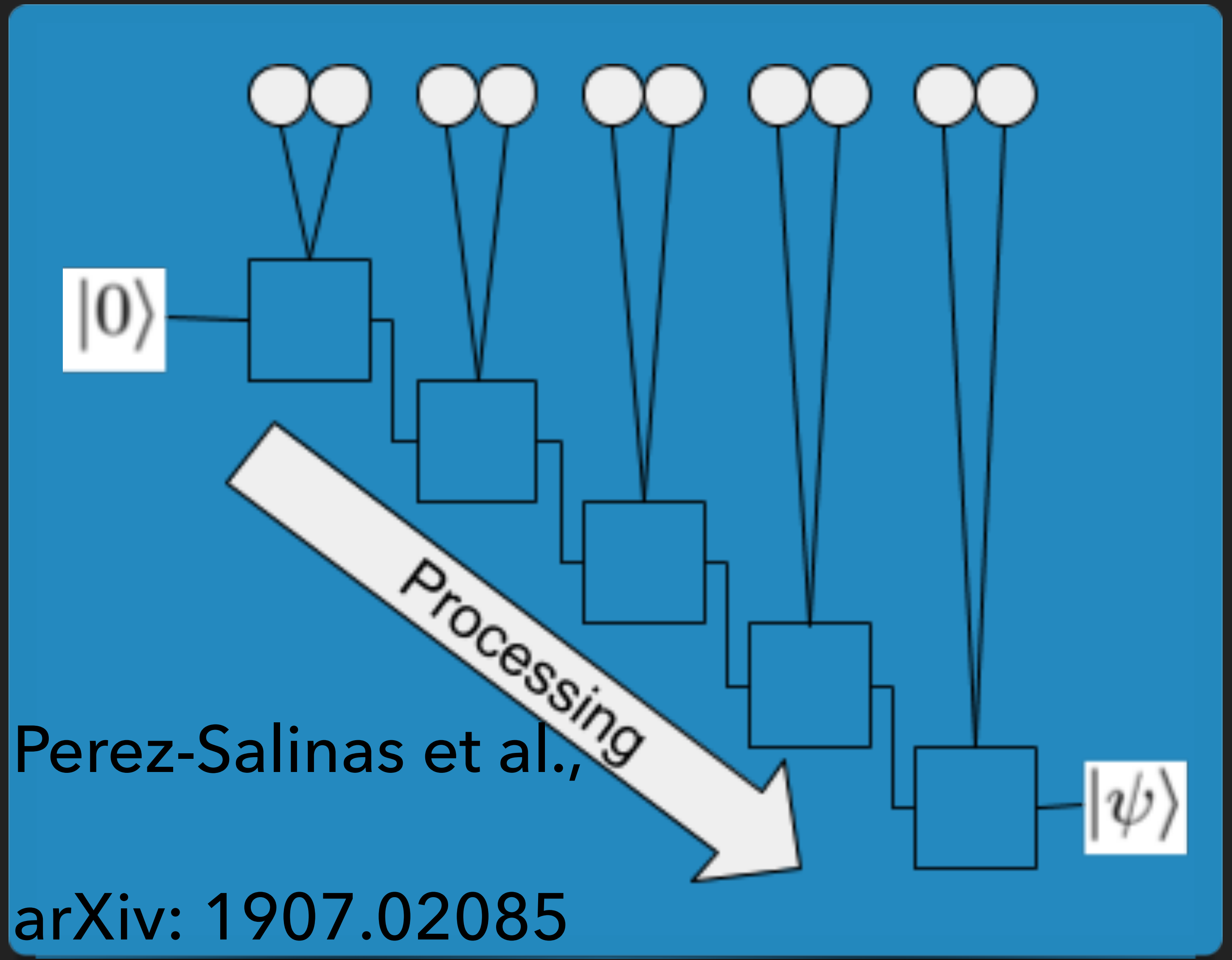
# INFORMATION BOTTLENECK FOR QUANTUM CLASSIFIERS

**L BANCHI, J PEREIRA, S PIRANDOLA**
**PRX QUANTUM 2, 040321 (2021)**

- $\rho(x)$ as "bottleneck" that squeezes the relevant information that $x$ provides about $c$. **Compression with side information**
- IB principle (loss independent): minimise
- $$\mathscr{L}_{IB} = I(X{:}Q) - \beta I(C{:}Q)$$
- Self-consistent solutions (similar for $\rho$)
- $$\tilde{\lambda}_z |\psi(z)\rangle = e^{(1-\beta)\log\bar{\rho} + \beta\sum_c P(c|z)\log\rho_c} |\psi(z)\rangle$$



# VARIATIONAL QUANTUM INFORMATION BOTTLENECK (CQ)




Perez-Salinas et al.,

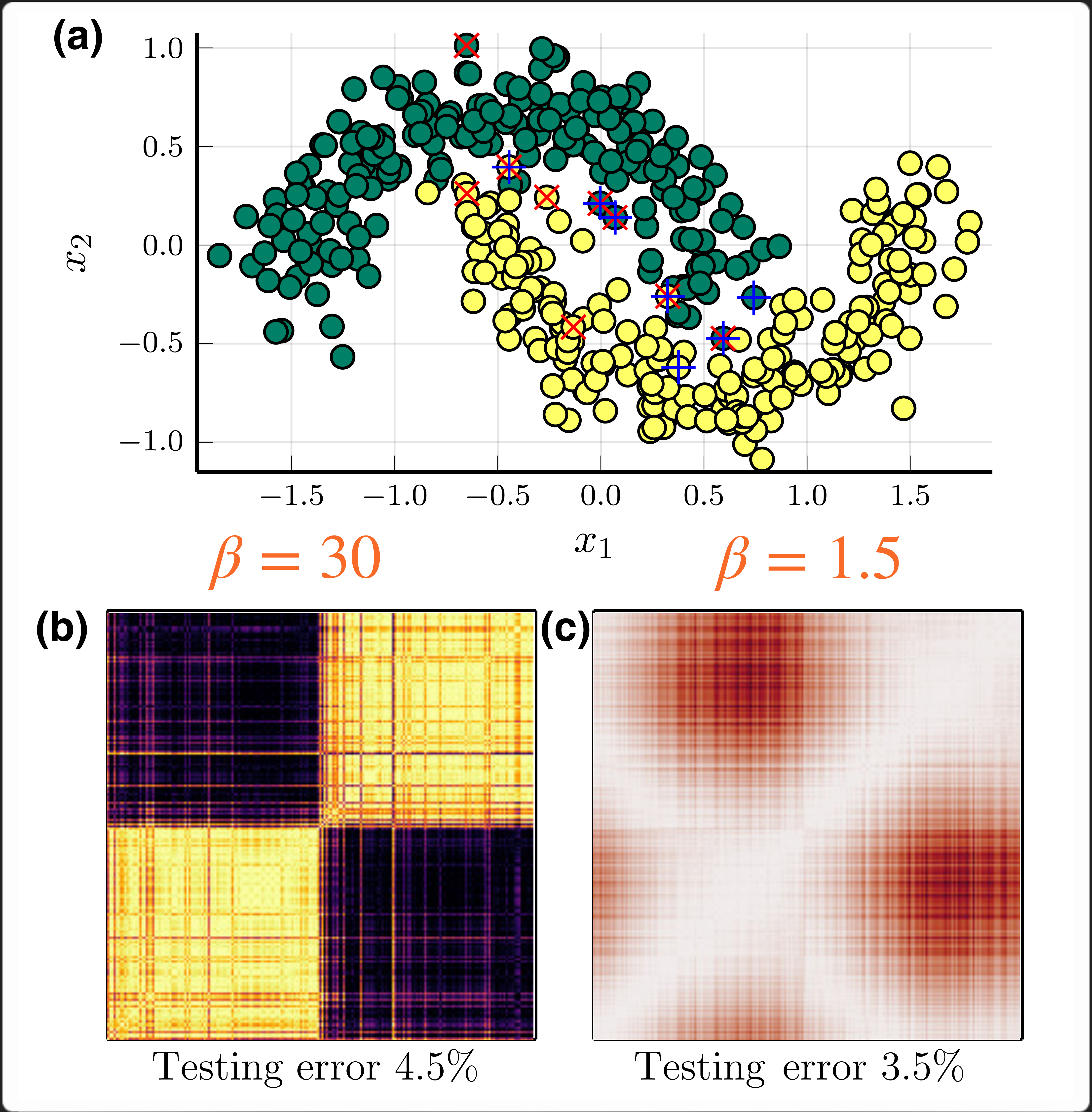
arXiv: 1907.02085

Single-qubit

“Data Re-uploading”

Classifier

$$|\psi_w(x)\rangle = \prod_{\ell=1}^{L}[R^z(w^{z\ell}\cdot x + w_0^{z\ell})R^y(w^{y\ell}\cdot x + w_0^{y\ell})]|0\rangle,$$



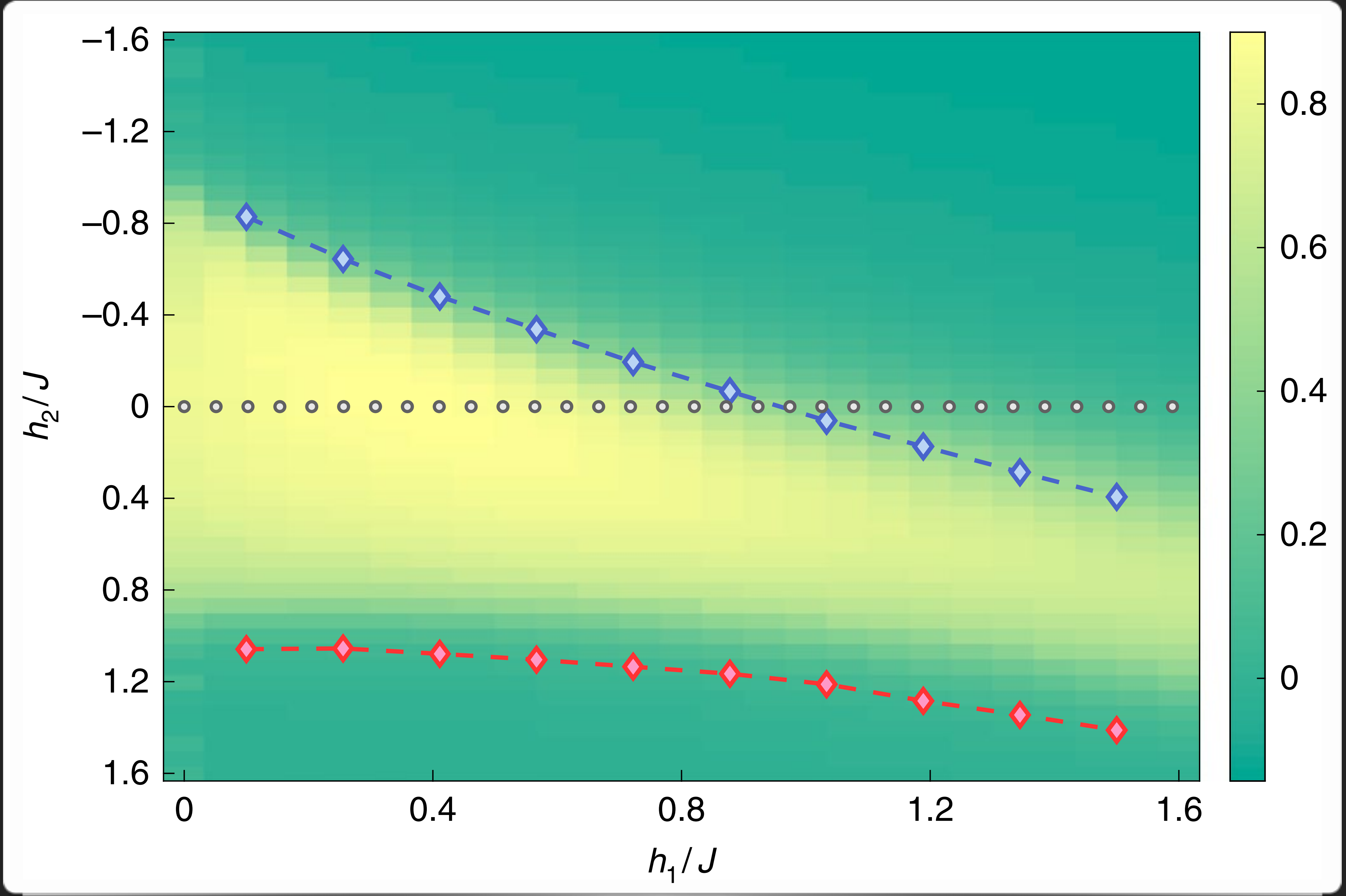

For two-qubit states the “re-upoading” embedding can be trained with an efficient variational minimisation of the IB Lagrangian

# QUANTUM PHASE RECOGNITION

- A Hamiltonian $\mathscr{H}(\theta)$ depends on external parameters $\theta = (\theta_1, \theta_2, \ldots)$, e.g. magnetic fields etc.
- Depending on $\theta$, the **ground state** of $\mathscr{H}(\theta)$ belongs to **different quantum phases**
- The order parameter is either unknown or too complex (e.g. non-local)
- For some $\theta$, the Hamiltonian is either **exactly solvable**, so and we can mathematically compute its phase, or **numerical approximations** work well
- ML perspective: we use the **computable points for training** and then check generalisation to arbitrary values of $\theta$

# QUANTUM PHASE RECOGNITION

$$H = -J \sum_{i=1}^{N-2} Z_i X_{i+1} Z_{i+2} - h_1 \sum_{i=1}^{N} X_i - h_2 \sum_{i=1}^{N-1} X_i X_{i+1}$$




Training along the curve $h_2 = 0$ where the model can be solved using a Jordan Wigner transformation

I Cong et al, Nature Physics 15, 1273 (2019)

**If** quantum phase recognition can be learnt by a QCNN, then generalisation is expected

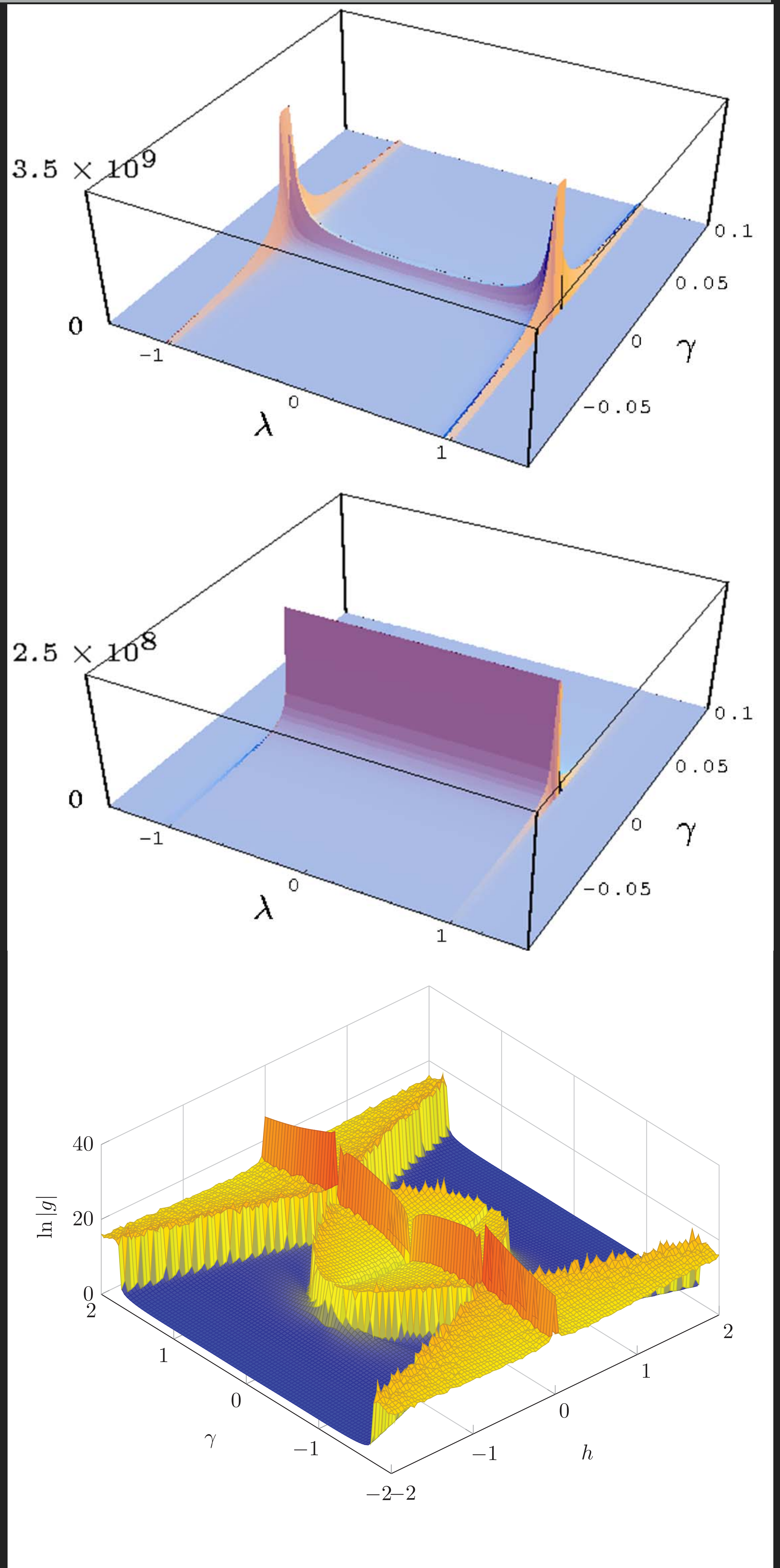
$$\mathscr{G} \approx \sqrt{\frac{N_{\text{parameters}}}{T}}$$

MC Caro et al, Nat. Comm. 13, 4919 (2022)

**What about the physics of this problem?**

# FIDELITY APPROACH TO QUANTUM PHASE TRANSITIONS

- Fidelity between two ground states
  $F(\theta, \phi) = |\langle \mathrm{gs}(\theta) | \mathrm{gs}(\phi) \rangle|^2$
- Metric in parameter space $1 - F(\theta, \theta + d\theta) = g_{\mu\nu} d\theta_\mu d\theta_\nu$
- Close to the phase transition $|g|$ diverges
- States belonging to different phases are quite **different**
- States belonging to the same phase are **clustered** (mind chaotic phases though!)



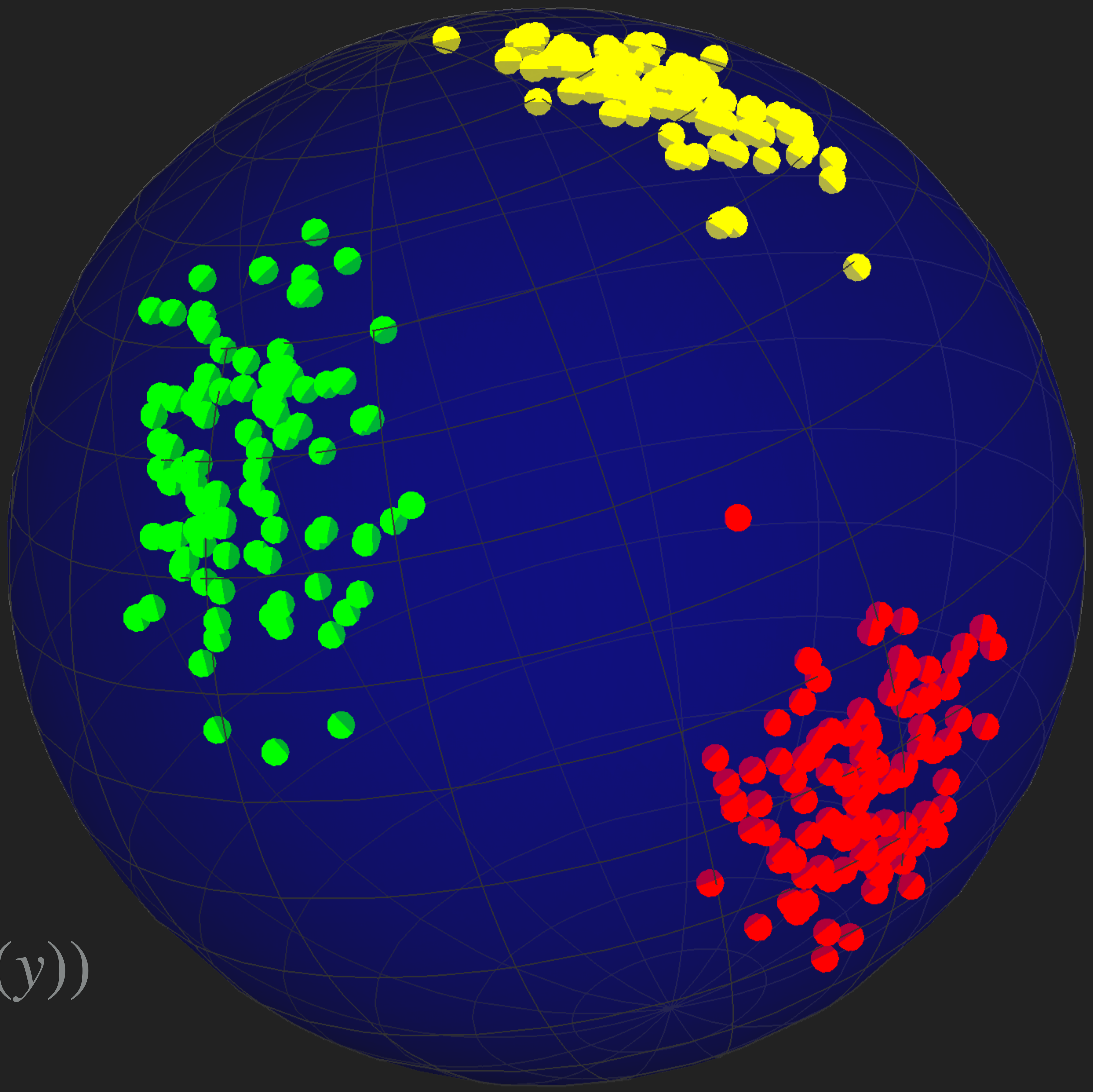

Zanardi et al. PRE 74,031123 (2006)

Banchi et al. PRE 89,022102 (2014)

# CLUSTERING IN HILBERT SPACES

- States are **clustered** in the (large!) Hilbert space
- There *should* be an efficient algorithm to classify the phases!
- **Generalisation depends on both data and algorithm**
- Data-dependent quantitative argument :

$$\mathcal{G} \approx \sqrt{\frac{\mathcal{B}}{T}} \qquad \mathcal{B}^2 \approx 1 - \sum P(x|c)P(y|c)F(\rho(x), \rho(y))$$



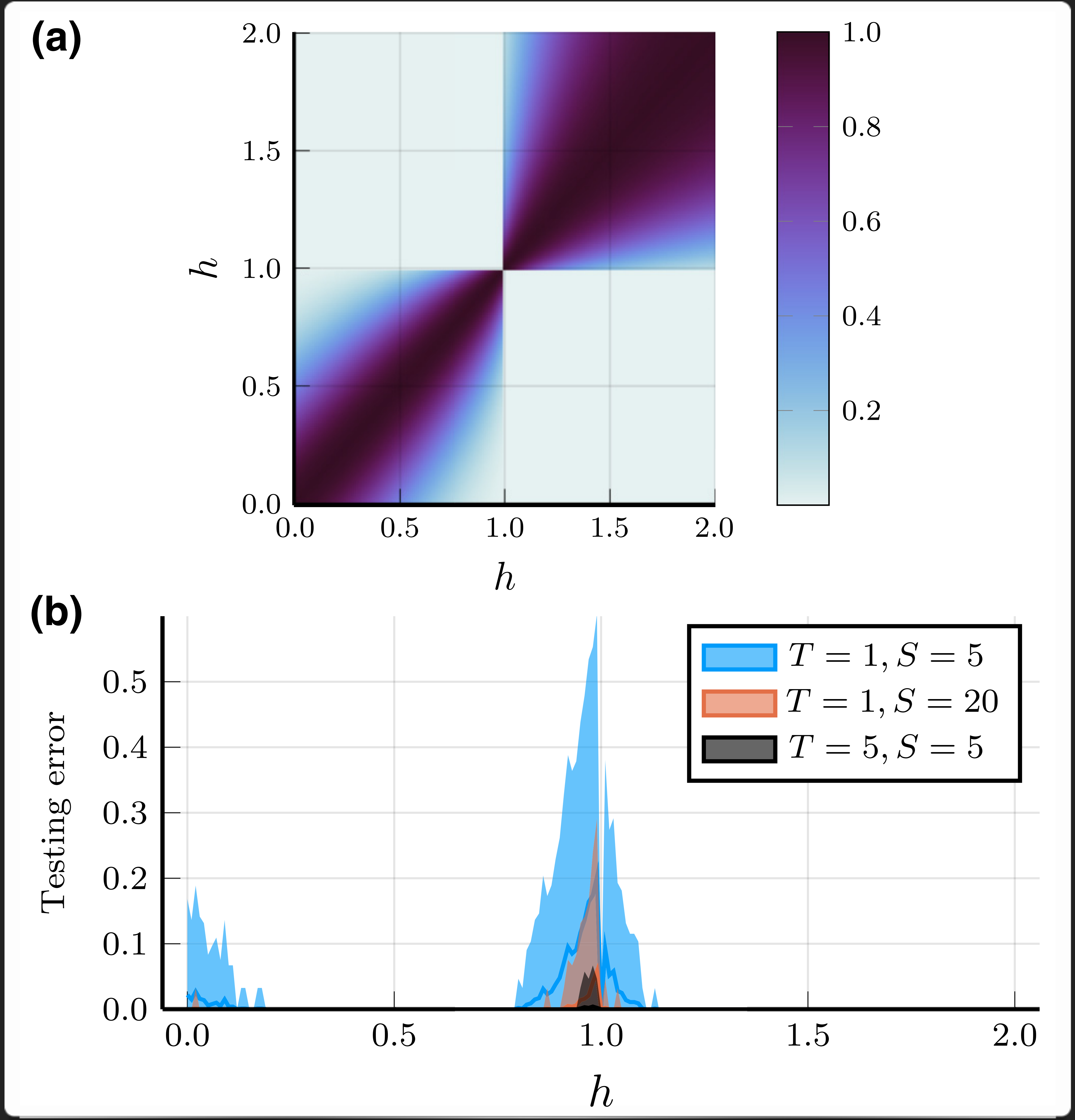
L BANCHI, J PEREIRA, S PIRANDOLA
PRX QUANTUM 2, 040321 (2021)

# QUANTUM PHASE RECOGNITION

L BANCHI, J PEREIRA, S PIRANDOLA
PRX QUANTUM 2, 040321 (2021)

Task: recognize the phases of matter of a quantum many-body system by taking measurements on the quantum system itself

$$H = -\sum_{i=1}^{L}(\sigma_i^x \sigma_{i+1}^x + h\sigma_i^z), \quad \Longrightarrow \quad \mathscr{B} \simeq 5.9$$




Ordered ($|h| < 1$) / disordered ($|h| > 1$) phases

T : number of training samples per class

S : number of measurement shots

# QUANTUM KERNELS

L BANCHI, J PEREIRA, S PIRANDOLA
PRX QUANTUM 2, 040321 (2021)

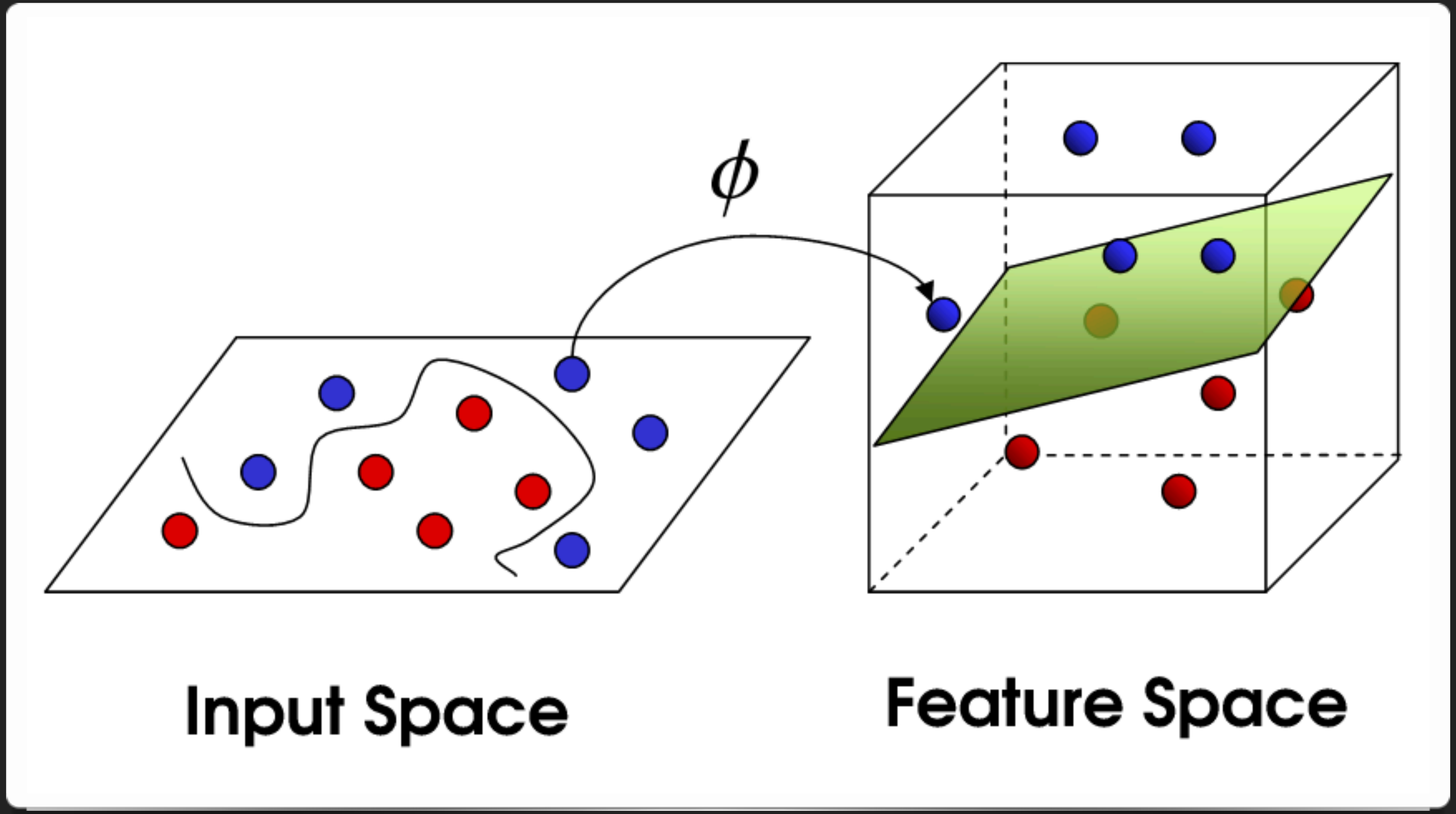
For pure state embeddings $\rho(x) = |\psi(x)\rangle\langle\psi(x)|$ we find

$$\mathscr{B} = \left[\mathrm{Tr}\sqrt{K}\right]^2$$

where $K_{xy} = \sqrt{p(x)p(y)}\,|\langle\psi(x)|\psi(y)\rangle|$ is a (normalised) **kernel** matrix. This makes the calculation $\mathscr{B}$ easier for large-dimensional embeddings.

Quantum kernels are used in

- Quantum support vector machines
- Quantum enhanced-feature space




**Take home message**: avoid $K \propto \mathrm{identity}$

(bad generalisation)

# CONCLUDING REMARKS

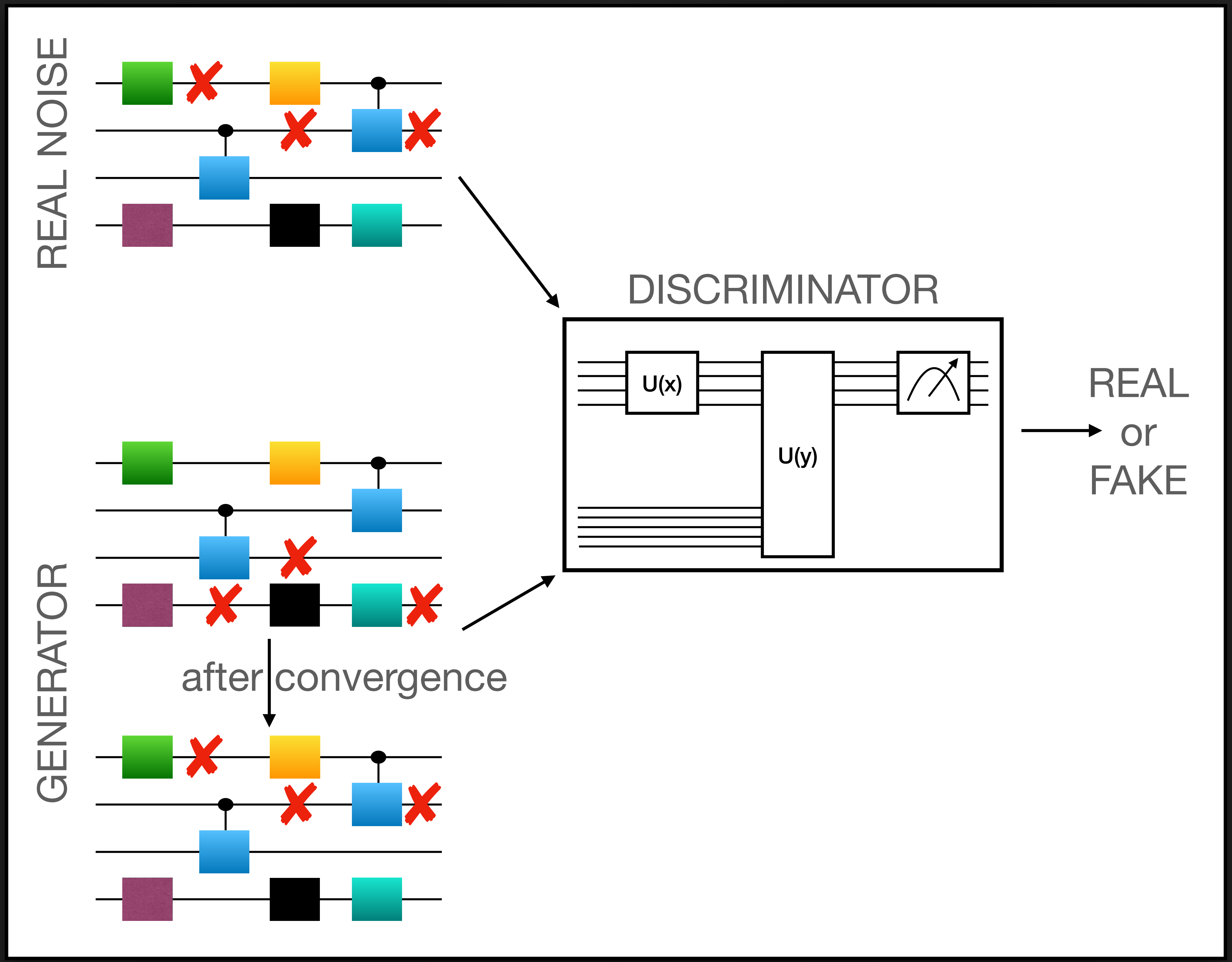
# CONCLUSIONS

- Quantum and Classical Algorithms to process either classical data (e.g. images) or quantum information encoded in quantum states
- Different applications:
  - quantum **pattern recognition** with entanglement-enhanced **quantum sensor**
  - Classification of **quantum states** and phases of matter
  - Quantum embeddings of classical data
- Foundational aspects: **generalisation & sample complexity**, information theoretic tools

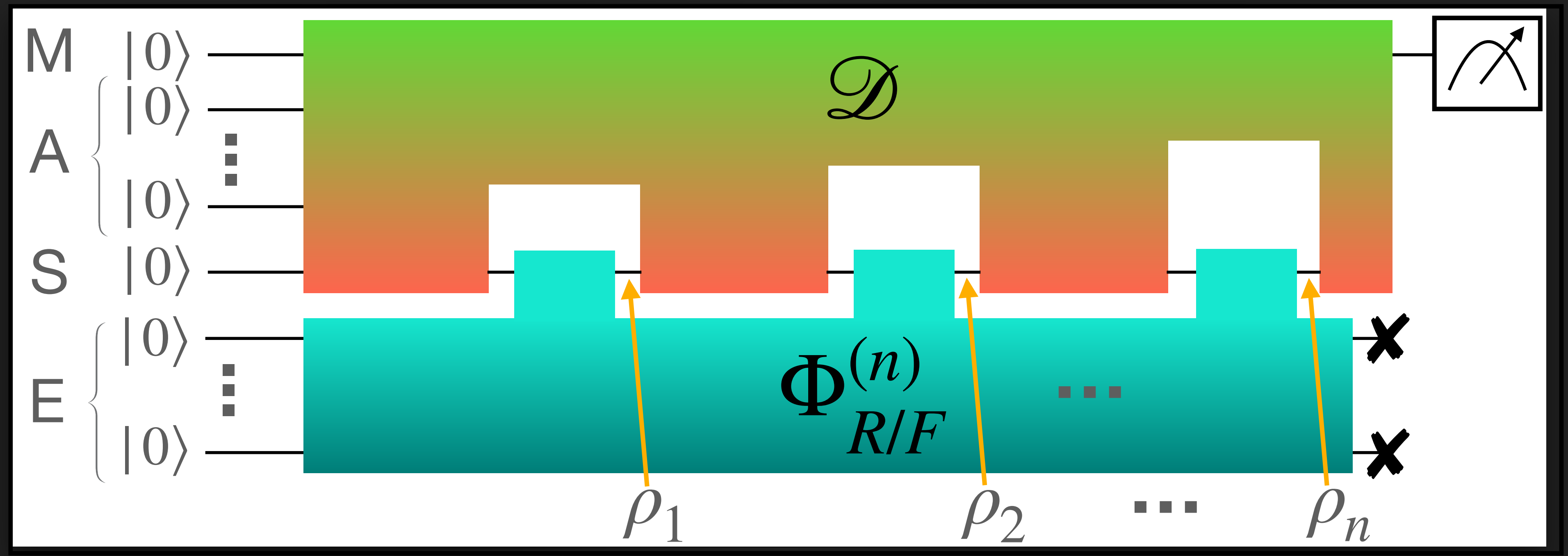
# EXTRA SLIDES

# QUANTUM GANS FOR NOISE SENSING ....

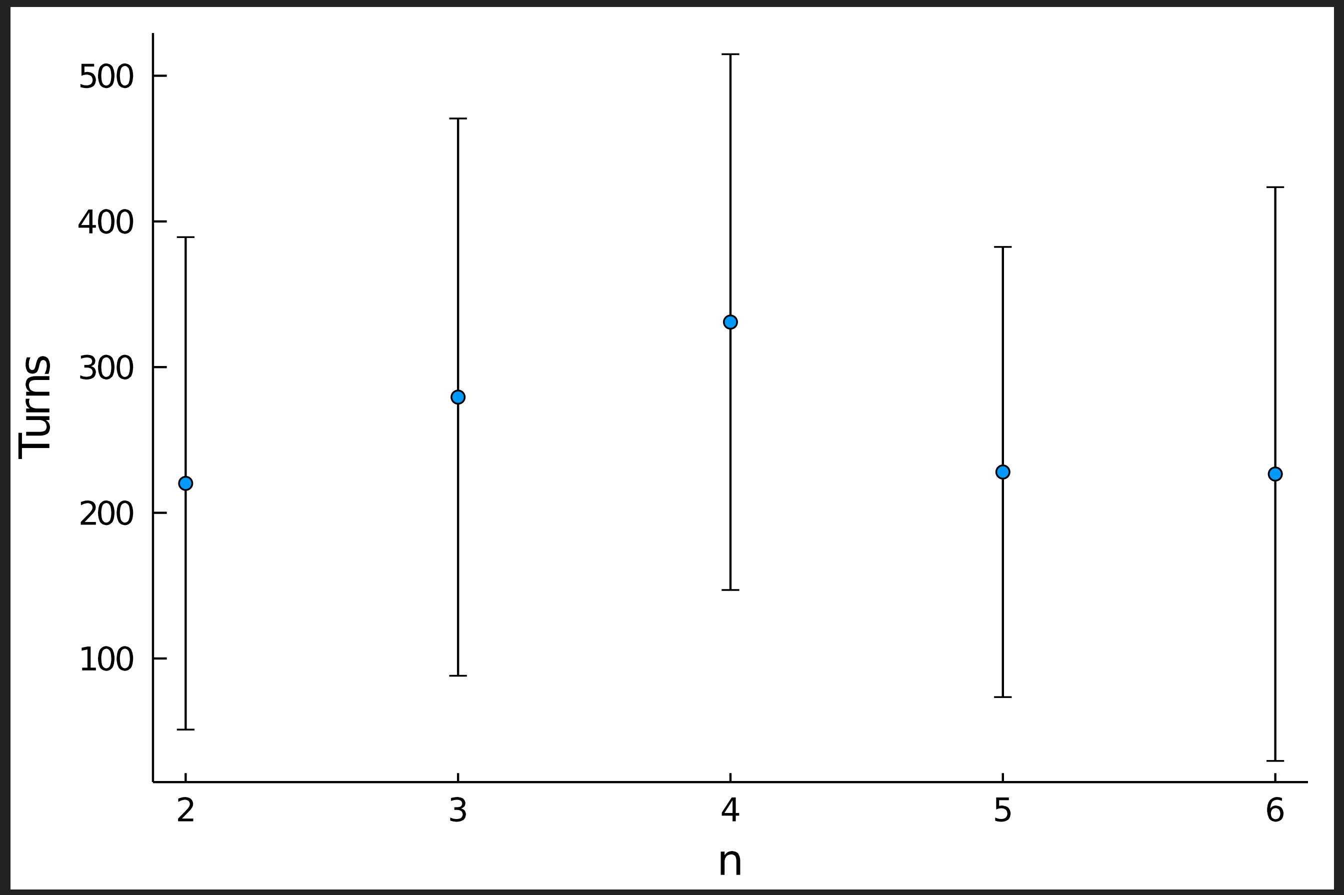
- SuperQGANs: Quantum Generative Adversarial Networks for learning Superoperators




P Braccia, L Banchi, F Caruso,
Phys. Rev. Applied 17, 024002 (2022)




Favourable Scaling



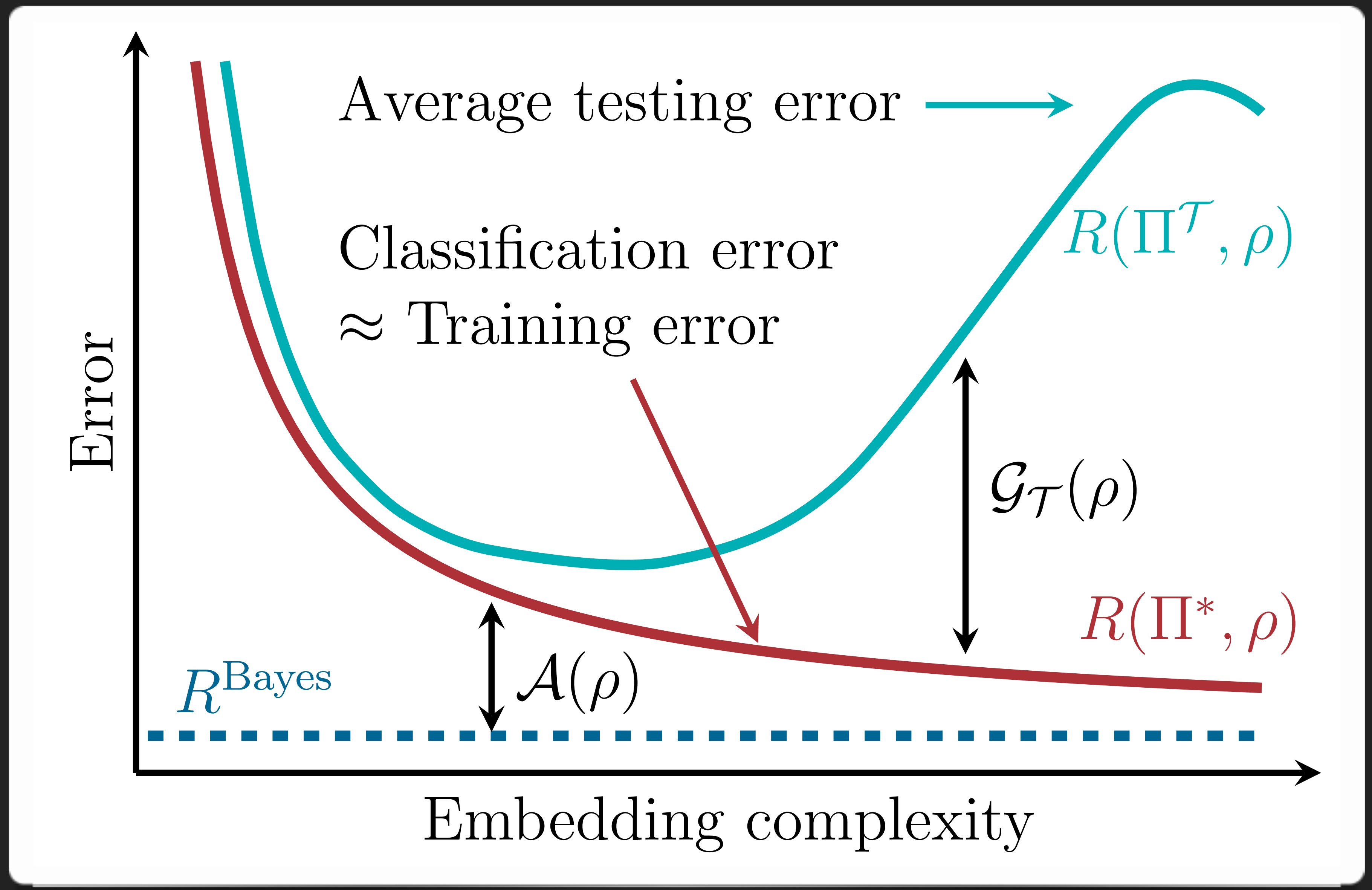
- Good embeddings should maximise $I(C{:}Q)$ and minimise $I_2(X{:}Q)$

$$\mathcal{G} \propto \sqrt{\frac{2^{I_2(X:Q)}}{T}} \qquad \mathscr{A} \leq K - \frac{2^{I(C:Q)}}{N_C}$$

- Spoiler: Information Bottleneck!
- Two **extreme cases**:

  - **BASIS ENCODING:** $x \mapsto \rho(x) = |x\rangle\langle x|$
    **MINIMUM** $\mathscr{A}(\rho) = 0$, **MAXIMUM** $\mathscr{G}_{\mathcal{T}}(\rho)$

  - **CONSTANT EMBEDDING :** $x \mapsto \rho$
    **MAXIMUM** $\mathscr{A}(\rho)$, **MINIMUM** $\mathscr{G}_{\mathcal{T}}(\rho) = 0$




for MNIST with 8-bit colours

$28 \times 28 \times 8 \approx 6000$ qubits!