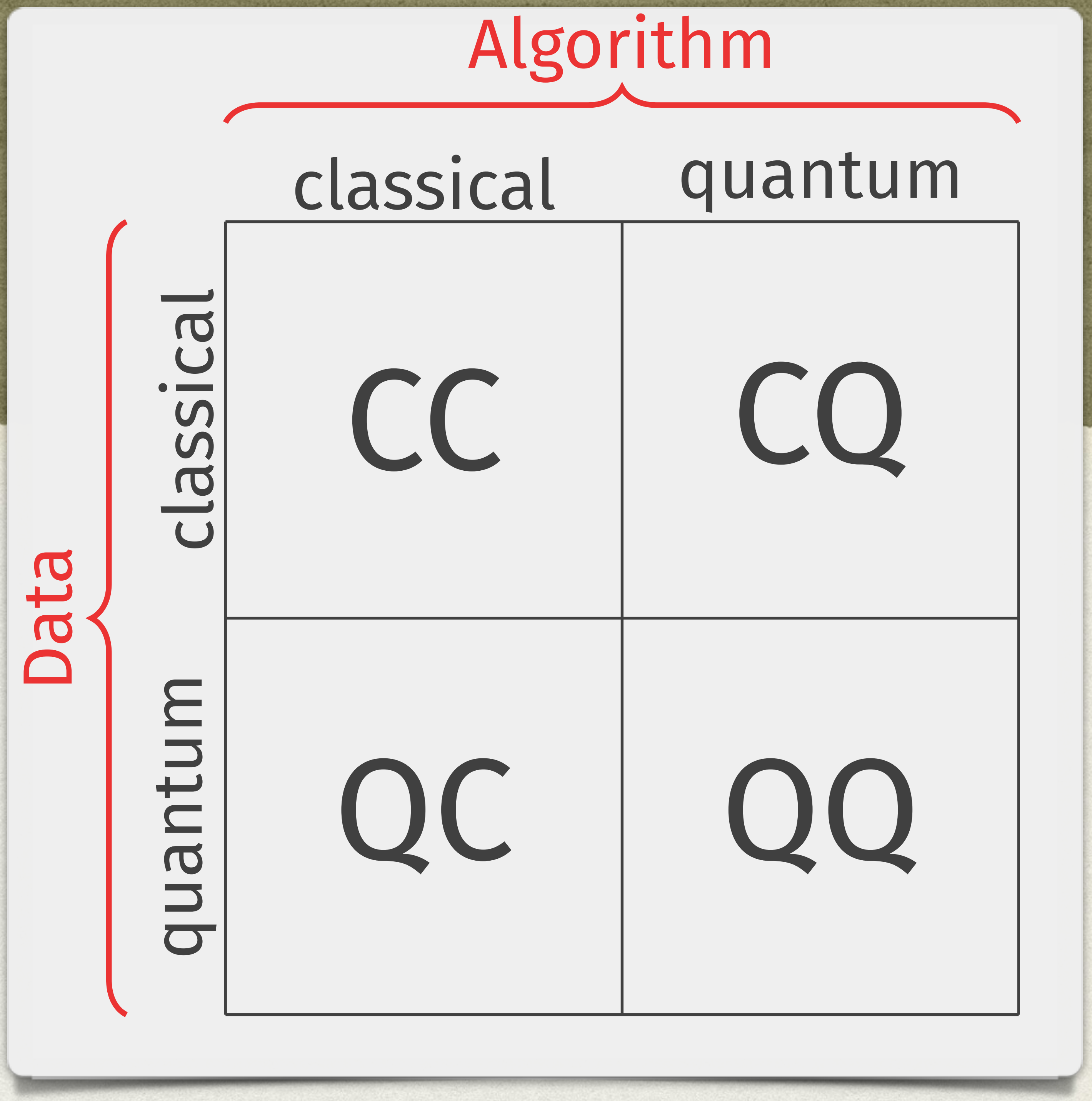
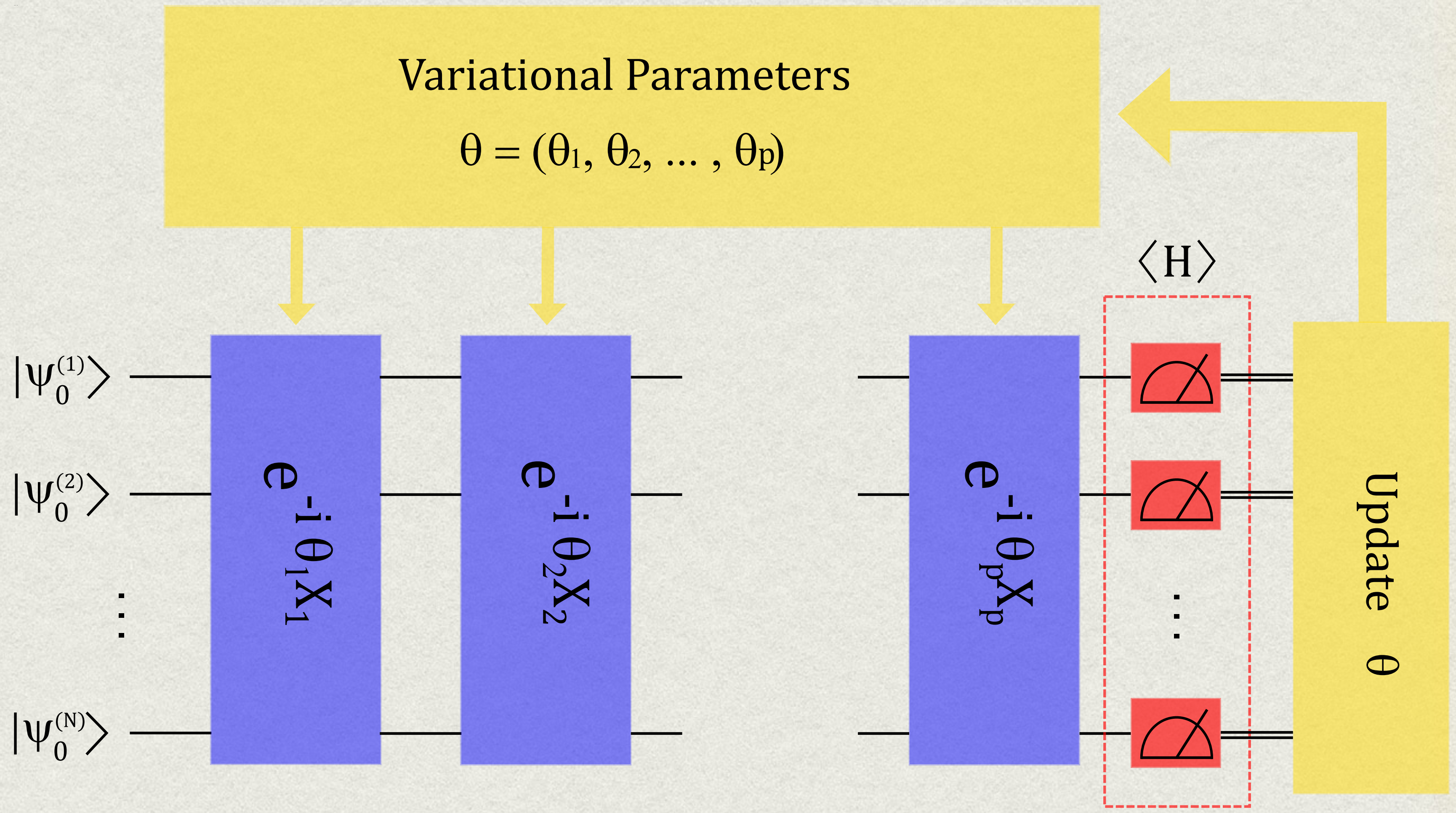
TRAINING AND TESTING:
A QUANTUM-INFO PERSPECTIVE

Leonardo Banchi — University of Florence & INFN Firenze

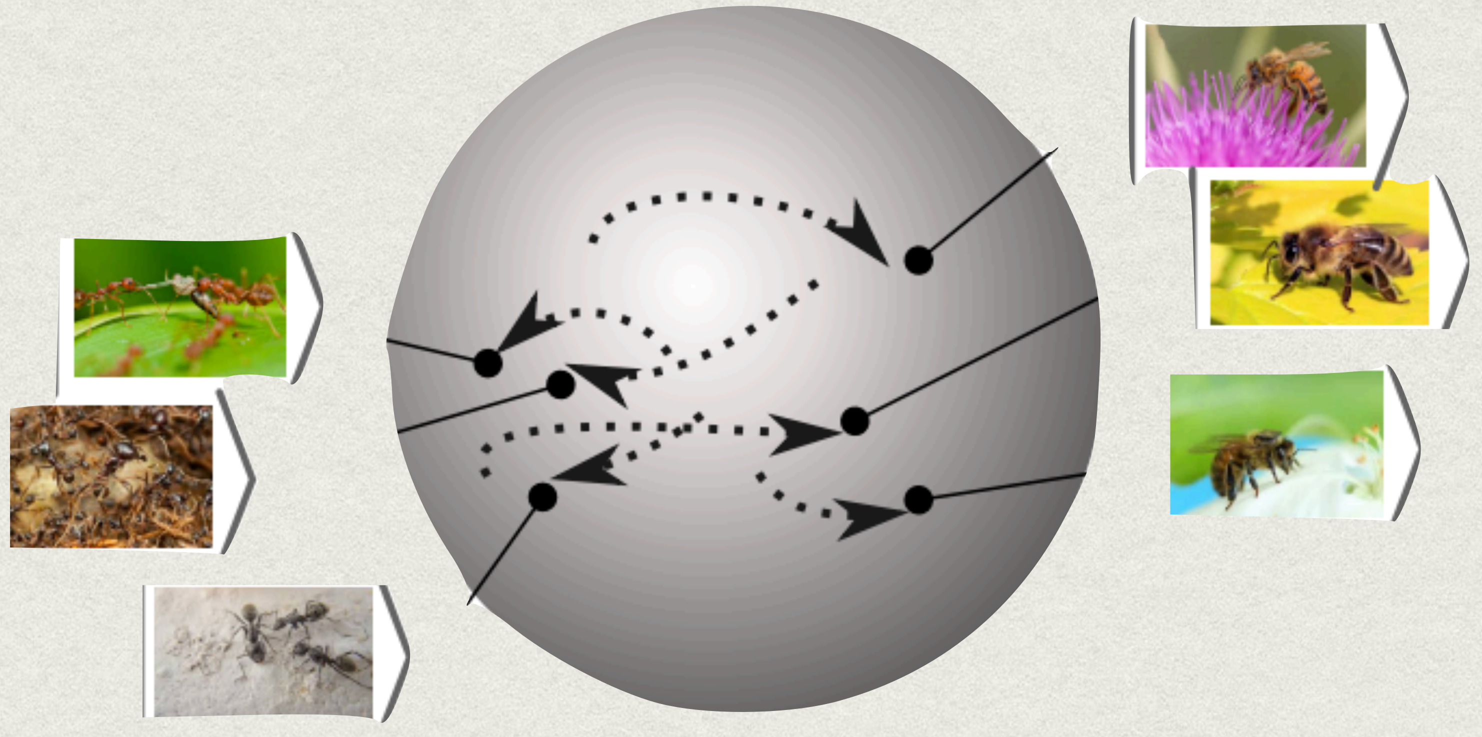
QML @ FLORENCE

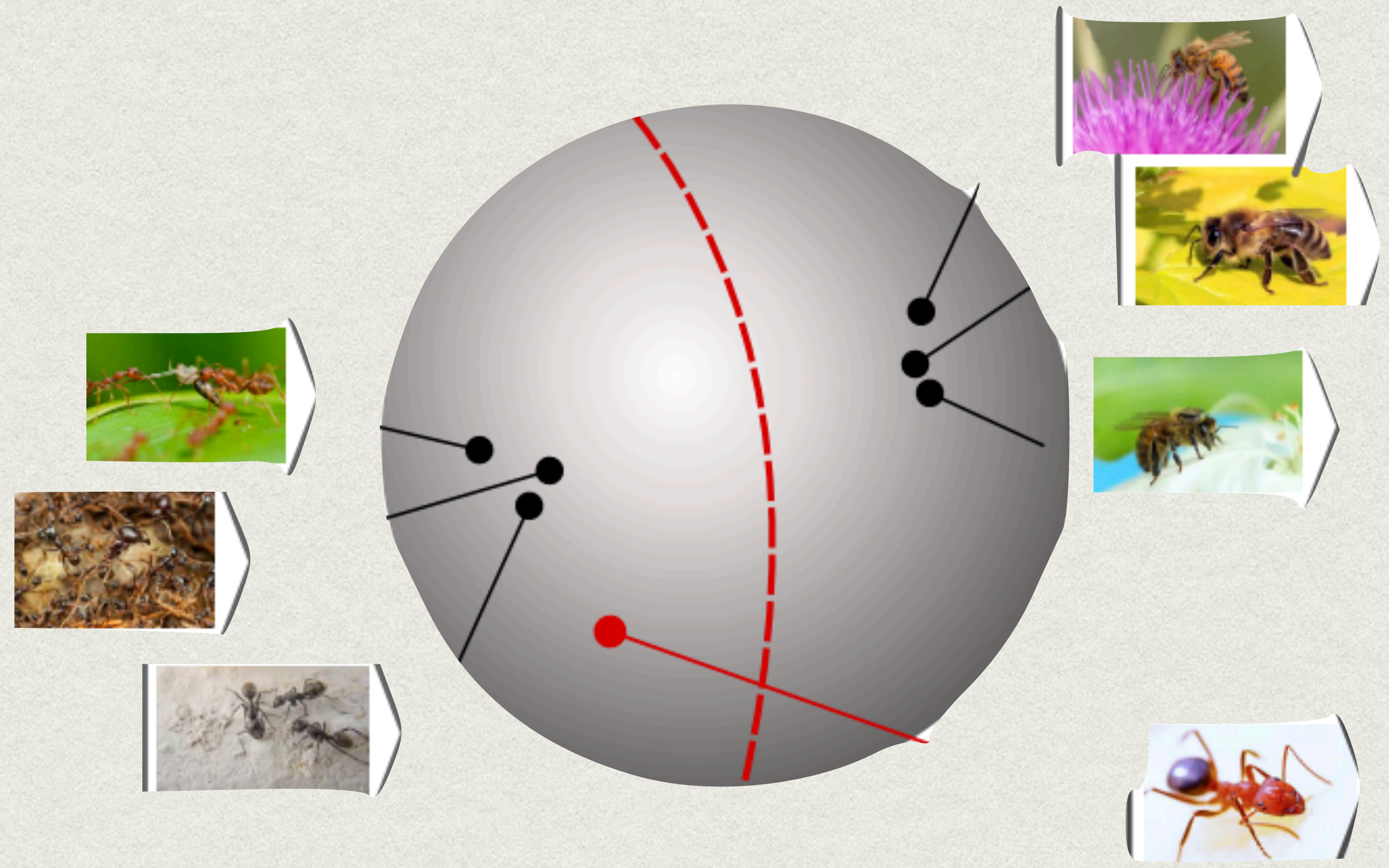


- Quantum Machine Learning in Florence
- Interest in Classical and Quantum data
- **Foundational aspects:** Generalisation
- Main focus: NISQ devices



- *Quantum-Enhanced Classifiers (CQ)*
  - M Schuld, N Killoran, Phys. rev. lett. 122 (4), 040504, (2019), V Havlicek, et al, Nature 567 (7747), 209, (2019)
  - L Banchi, et al., Phys. Rev. Applied 14, 064026 (2020), S Lloyd, et al, arXiv:2001.03622





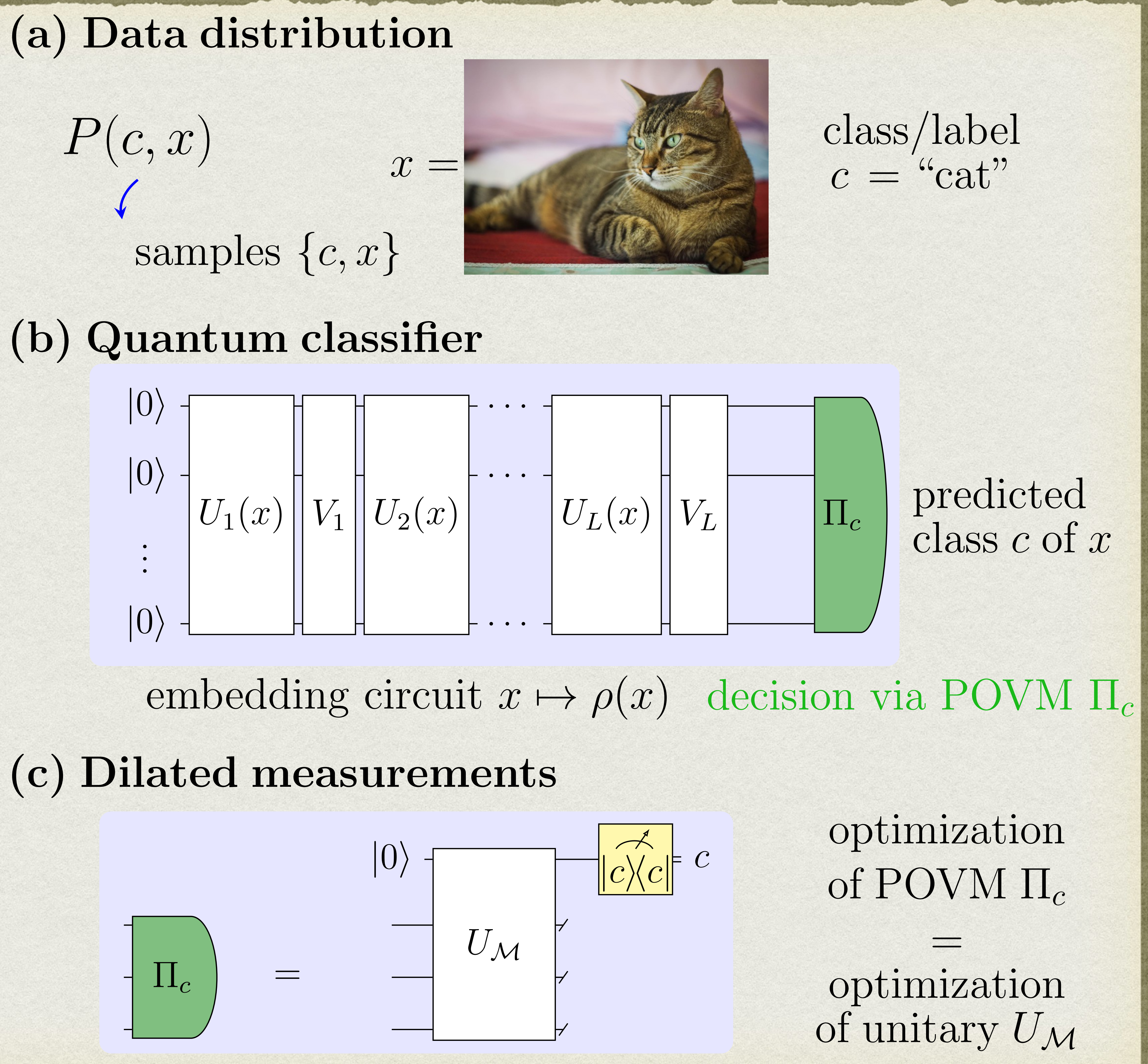
Main question: after training a quantum model using a few known examples, can the model accurately classify even unseen data?

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

- Study generalisation using tools from quantum information theory
- Study how the Hilbert space dimension, noise, pooling etc. affect generalisation
- Bias-Variance tradeoff and Information Bottleneck principle

QUANTUM EMBEDDINGS

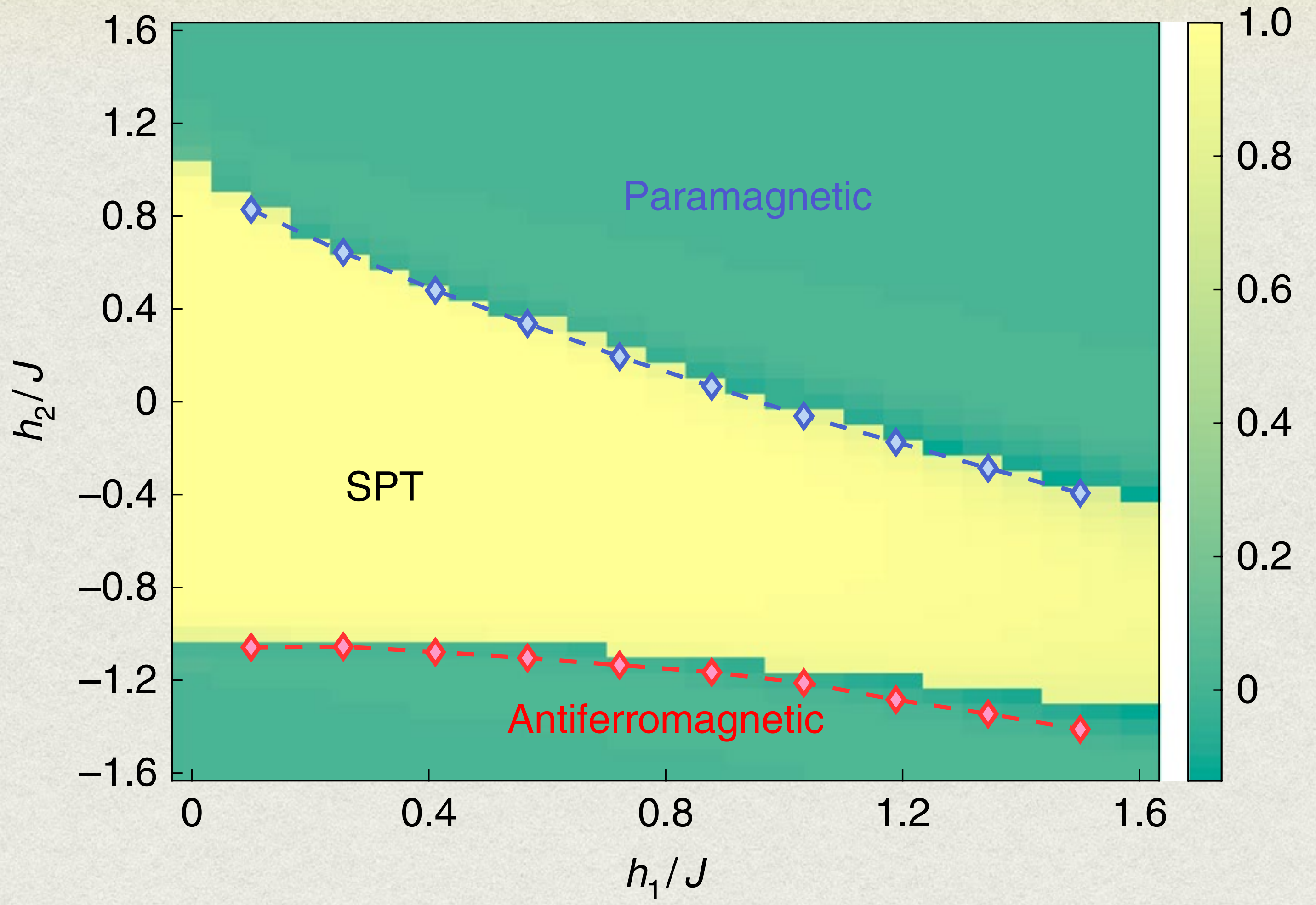
- Classify classical data (e.g. images)
- Embed images x onto a quantum state $x \mapsto \rho(x)$
- Decide the class from a quantum measurement $\{\Pi_c\}$



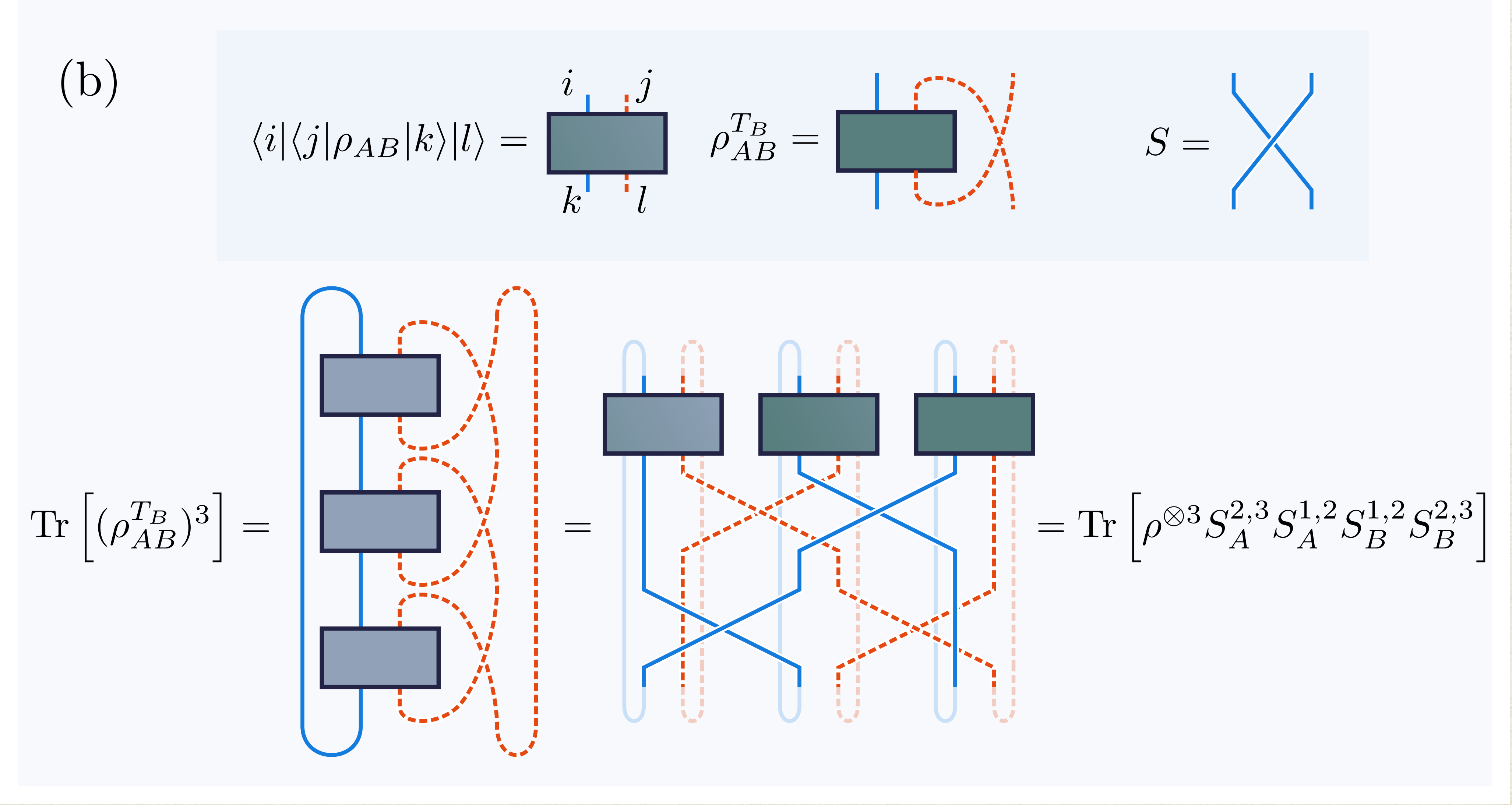
- M Schuld, N Killoran, Phys. rev. lett. 122 (4), 040504, (2019)
- V Havlicek, et al, Nature 567 (7747), 209, (2019)
- S Lloyd, et al, arXiv:2001.03622

MANY-BODY PHYSICS (QQ/QC)

- Quantum Phase Recognition
  - I Cong, S Choi, MD Lukin, Nature Physics 15, 1273 (2019)
  - L Banchi, J Pereira, S Pirandola, PRX Quantum 2, 040321 (2021)

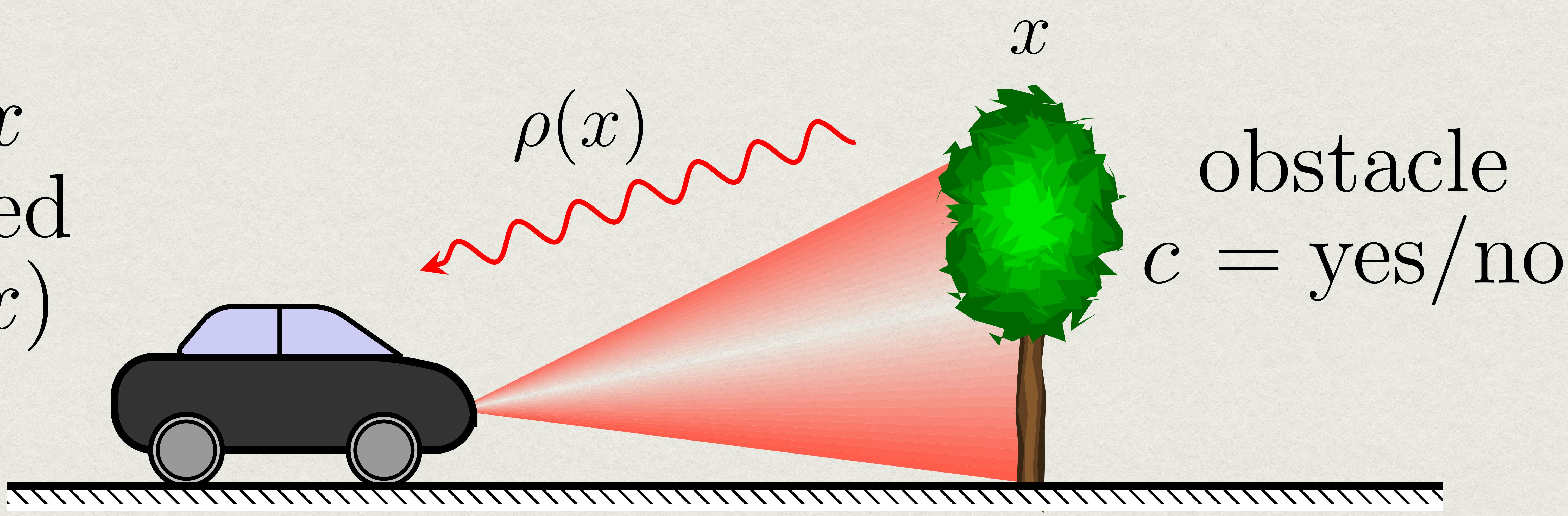


- Many-Body Entanglement Measurement from PPT-moments $\mathrm{Tr}\left[(\rho_{AB}^{T_B})^n\right]$
  - J Gray, L Banchi, A Bayat, S Bose, Phys. Rev. Lett. 121, 150503 (2018)



QUANTUM CHANNEL DISCRIMINATION

detect objects x
from the scattered
state of light $\rho(x)$



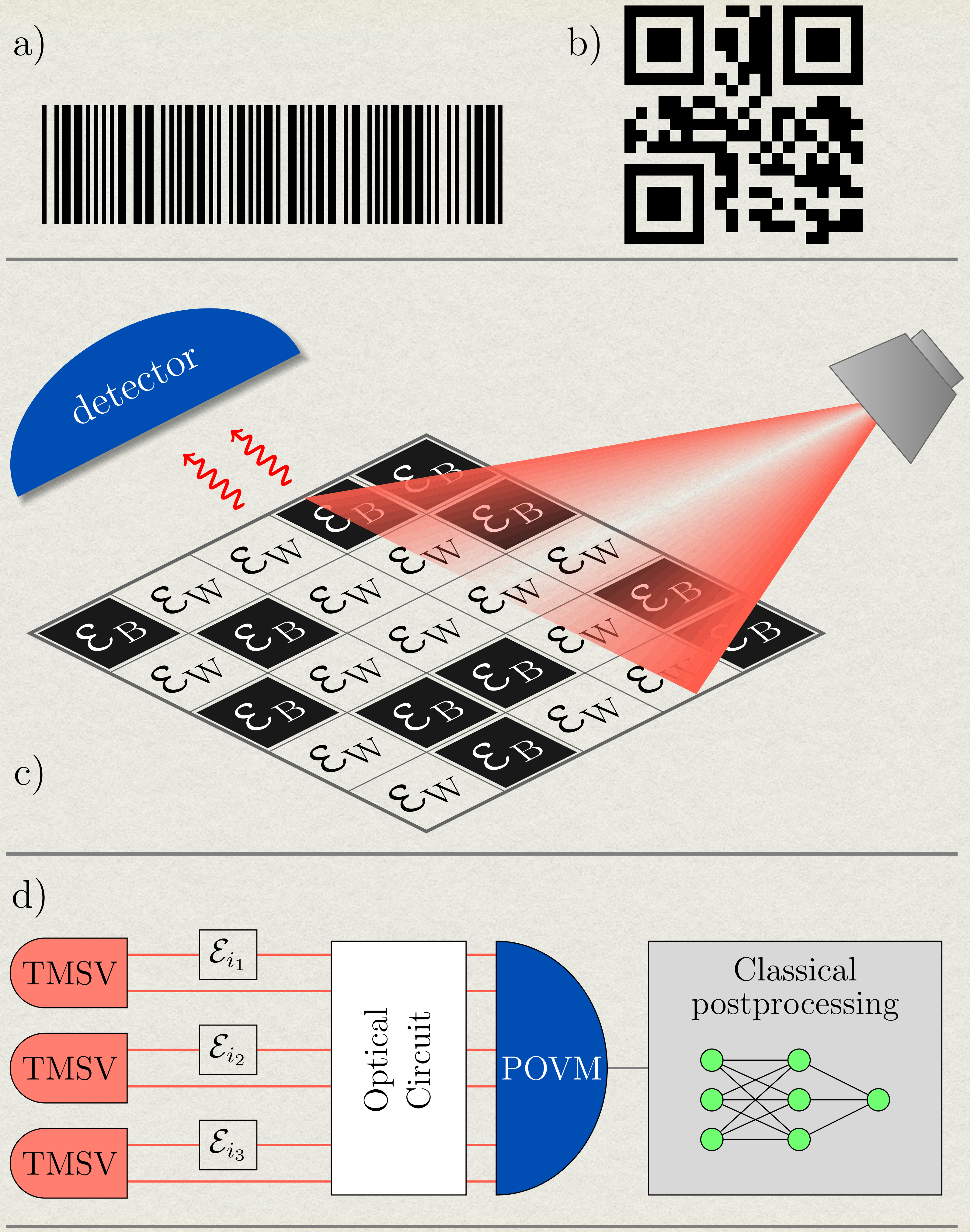
- Images live in the physical world
- Optimise over the (entangled) input probe state of light and over the detection POVM

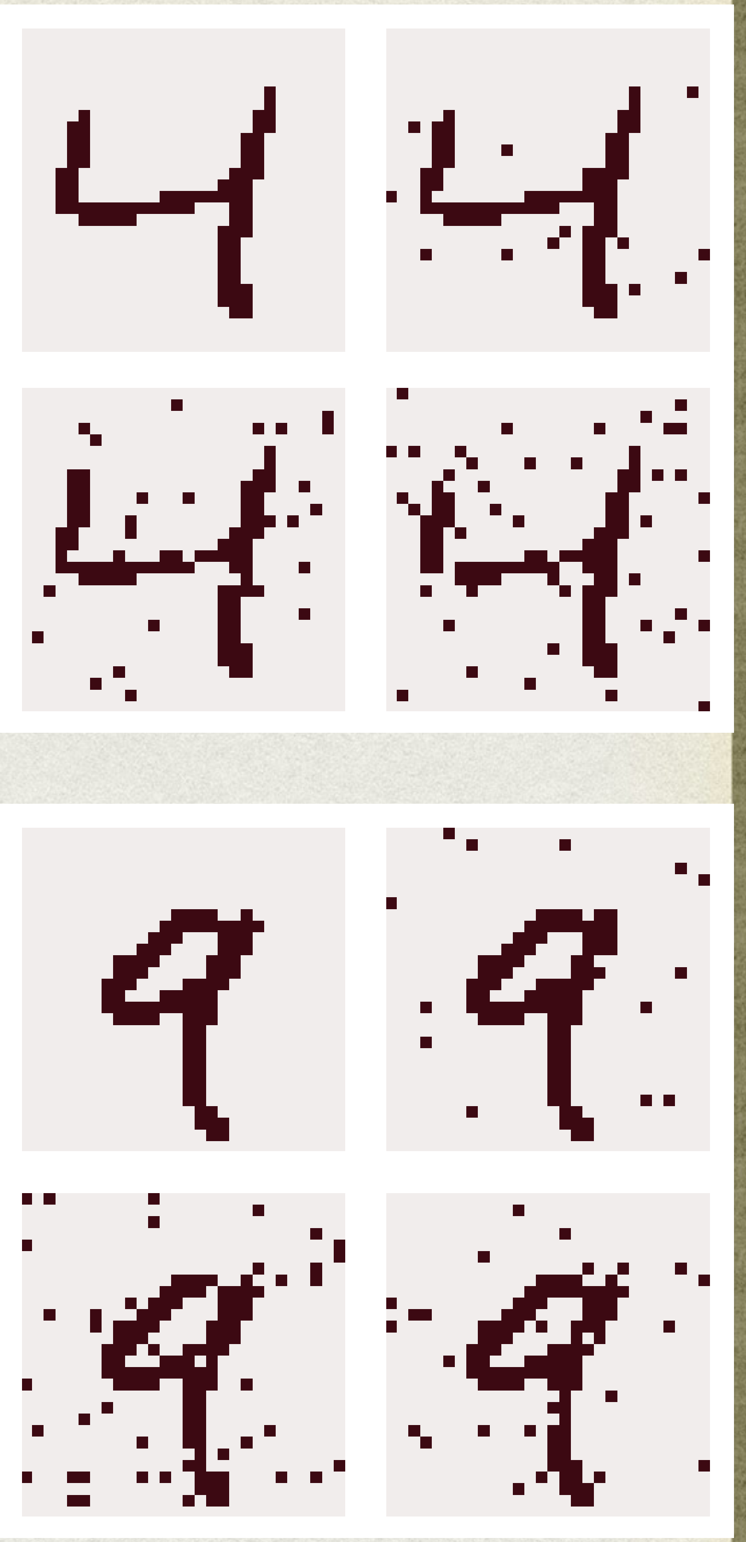
QUANTUM BARCODES AND PATTERN RECOGNITION

- 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)



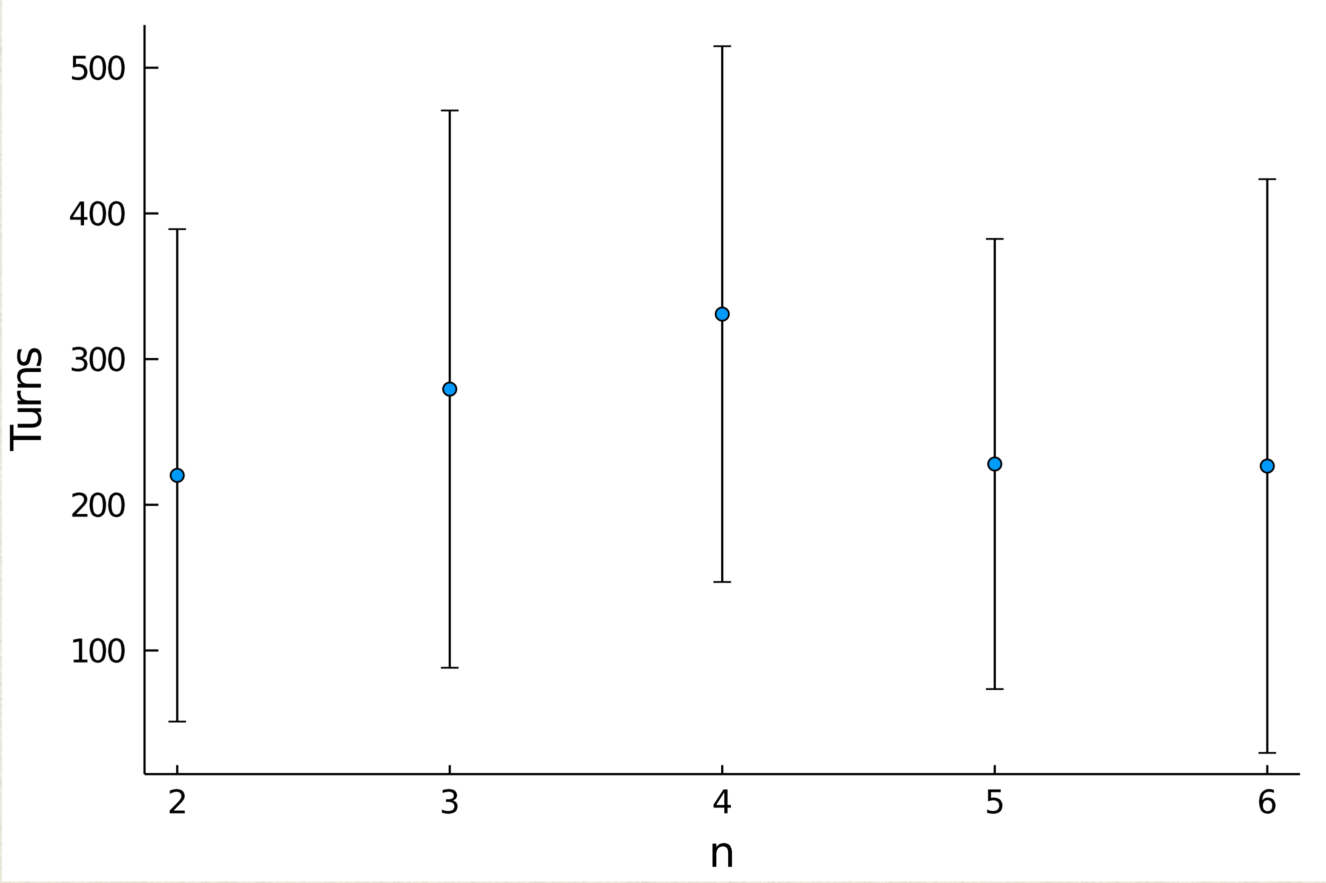


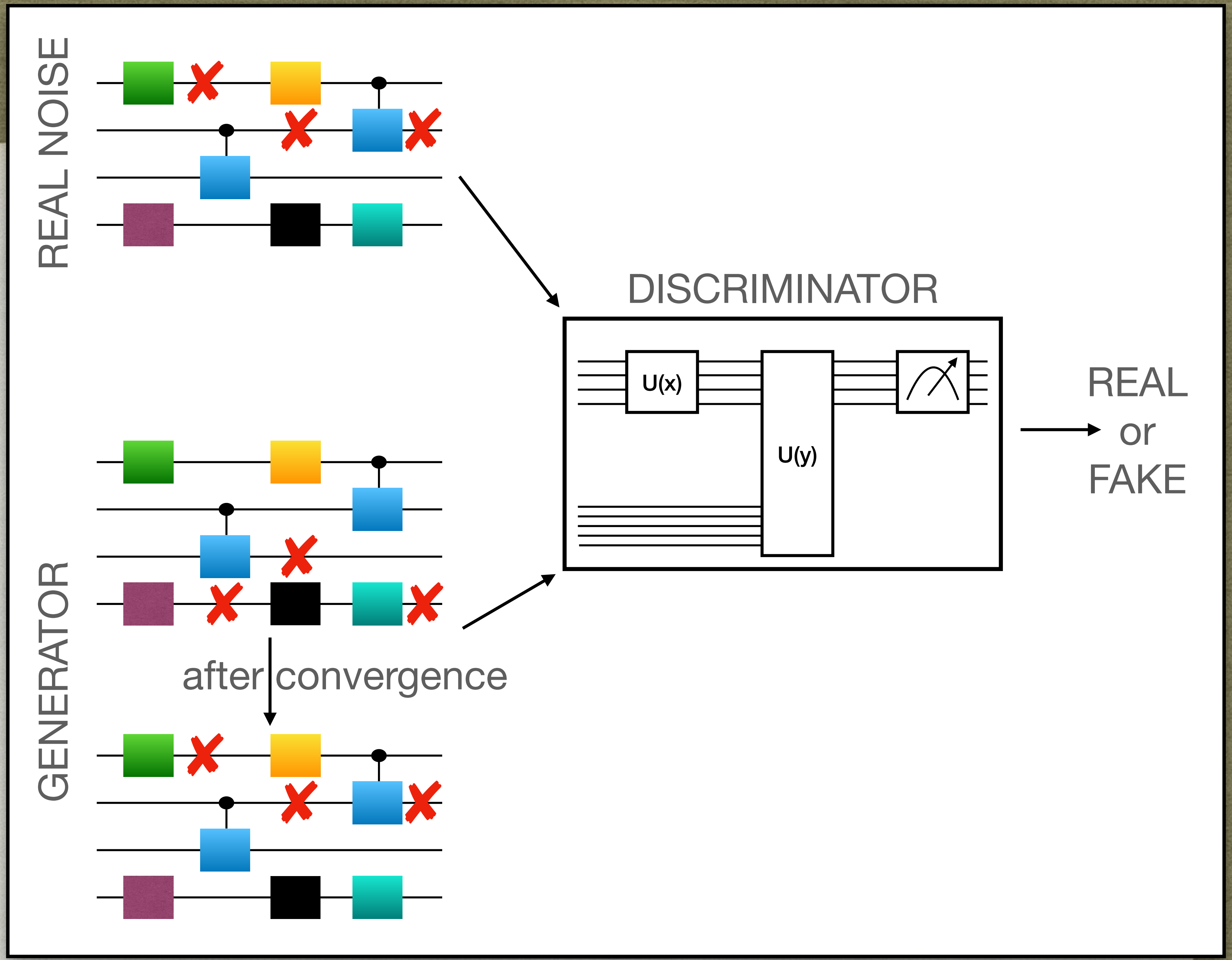
QUANTUM GANS FOR NOISE SENSING

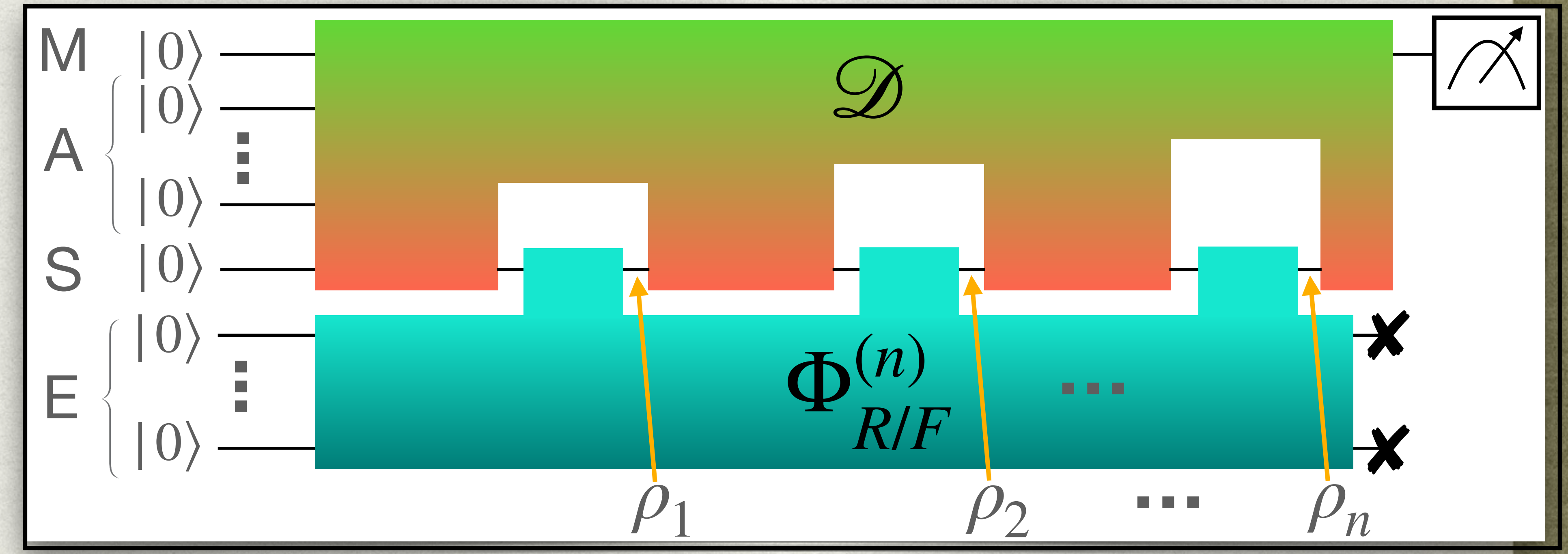
- SuperQGANs: Quantum Generative Adversarial Networks for learning Superoperators

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

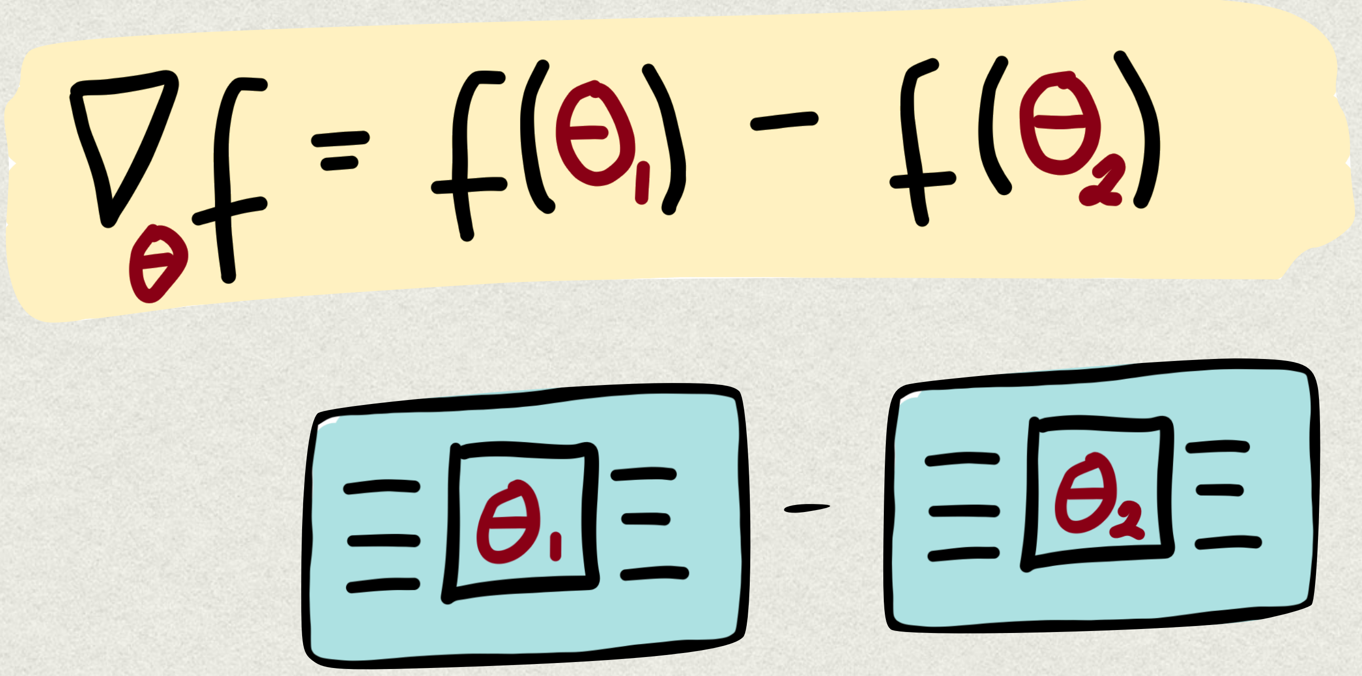
- Favourable Scaling



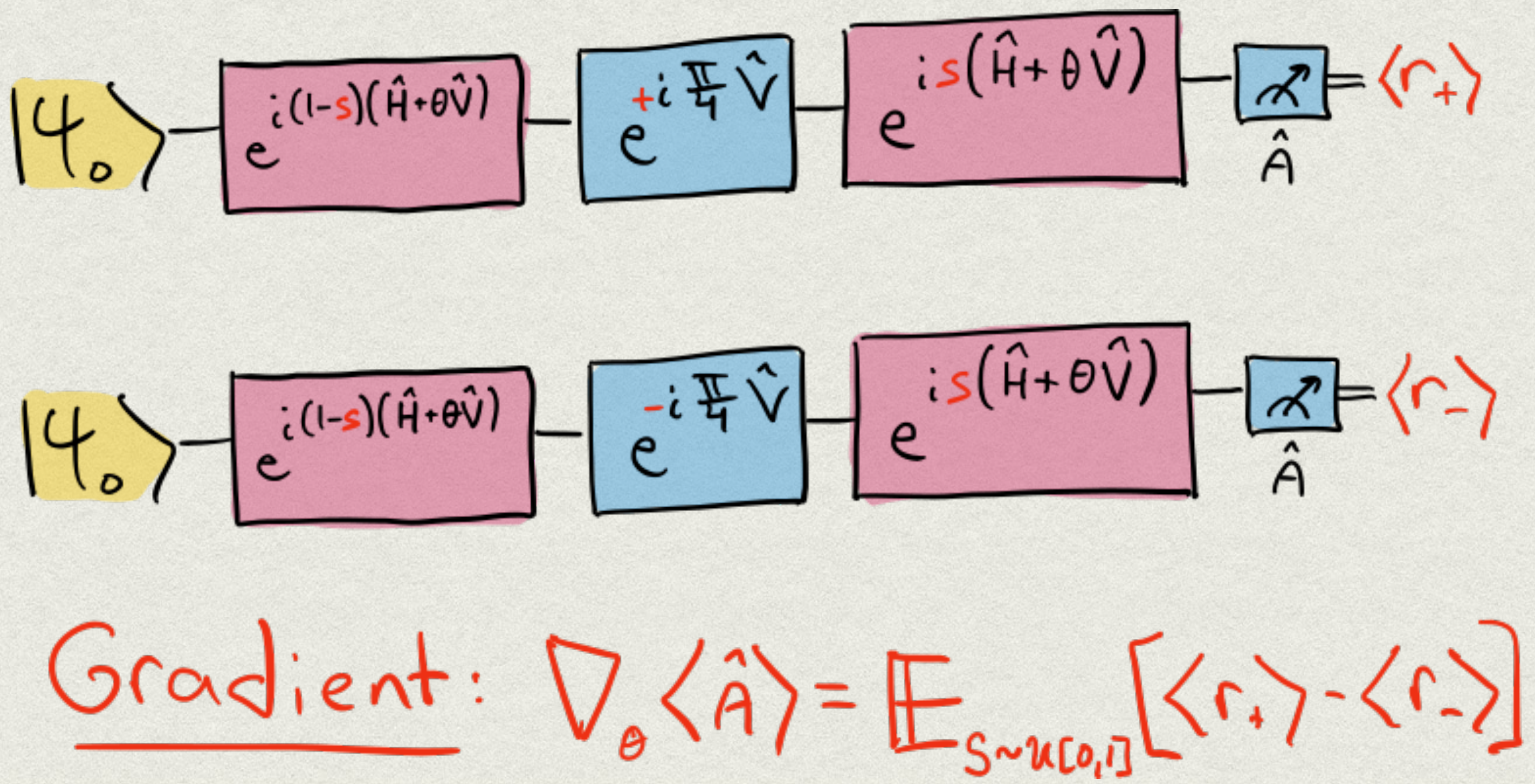




HOW DO WE OPTIMISE? GRADIENTS!



- Parameter Shift Rule / Hadamard test for $e^{i\theta\hat{o}}$ gates
  - K Mitarai, M Negoro, M Kitagawa, K Fujii - Physical Review A, 2018
  - M Schuld, V Bergholm, C Gogolin, J Izaac, N Killoran - Physical Review A, 2019
- Stochastic PSR for **general Hamiltonian evolution** $e^{i(\hat{H}+\theta\hat{V})}$
  - L Banchi, G Crooks, Quantum 5, 386 (2021)
- Continuous variable systems / GBS distribution
  - N Killoran, et al. Phys. Rev. Research (2019)
  - L Banchi, N Quesada, J M Arrazola, Phys. Rev. A 102, 012417 (2020)

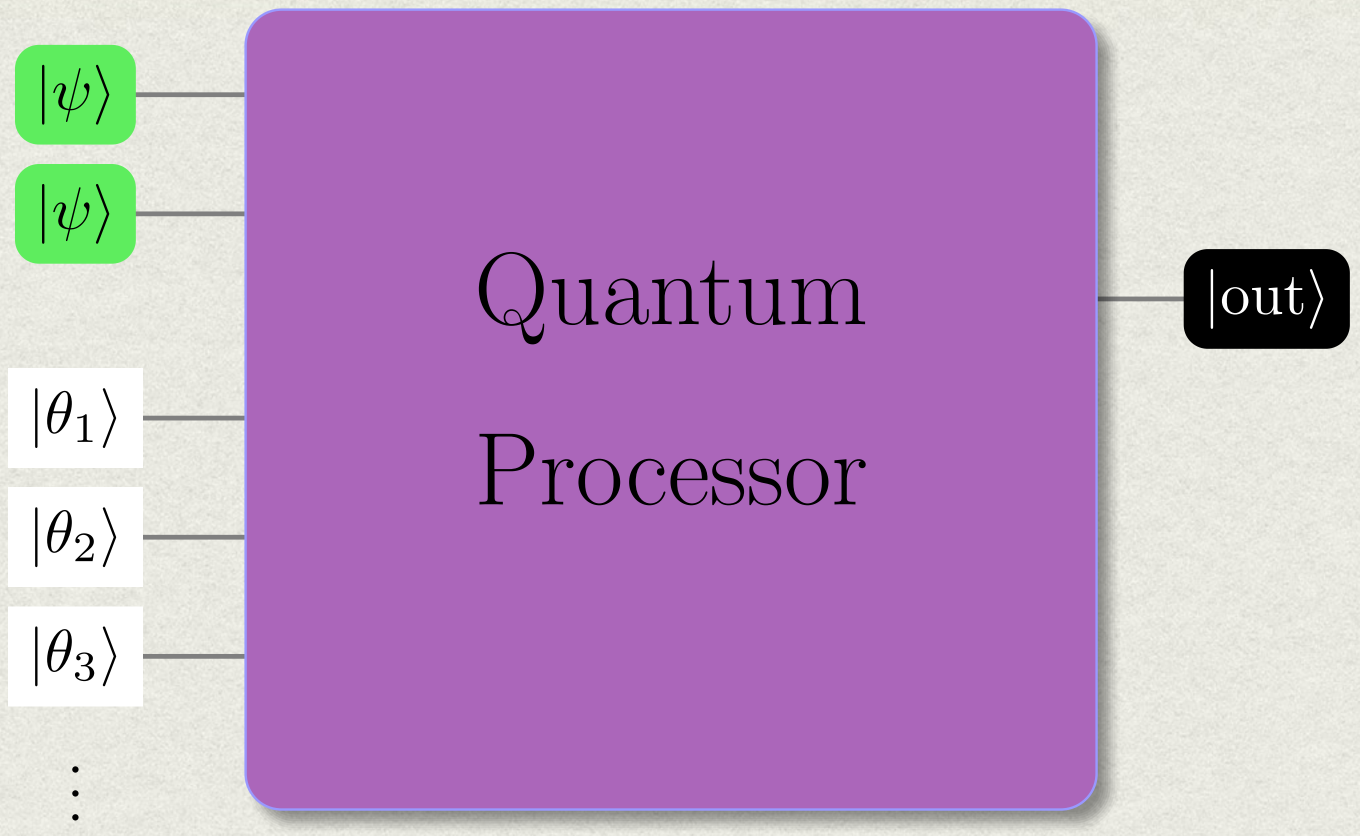


MORE GRADIENTS!

- Gradient-based optimisation of "Quantum Programs"

$$\rho_{\mathrm{output}} = \mathscr{E}[\rho_{\mathrm{input}} \otimes \theta]$$

  - L Banchi, J Pereira, S Lloyd, S Pirandola, npj Quantum Information 6 (1), 1-10, (2020)

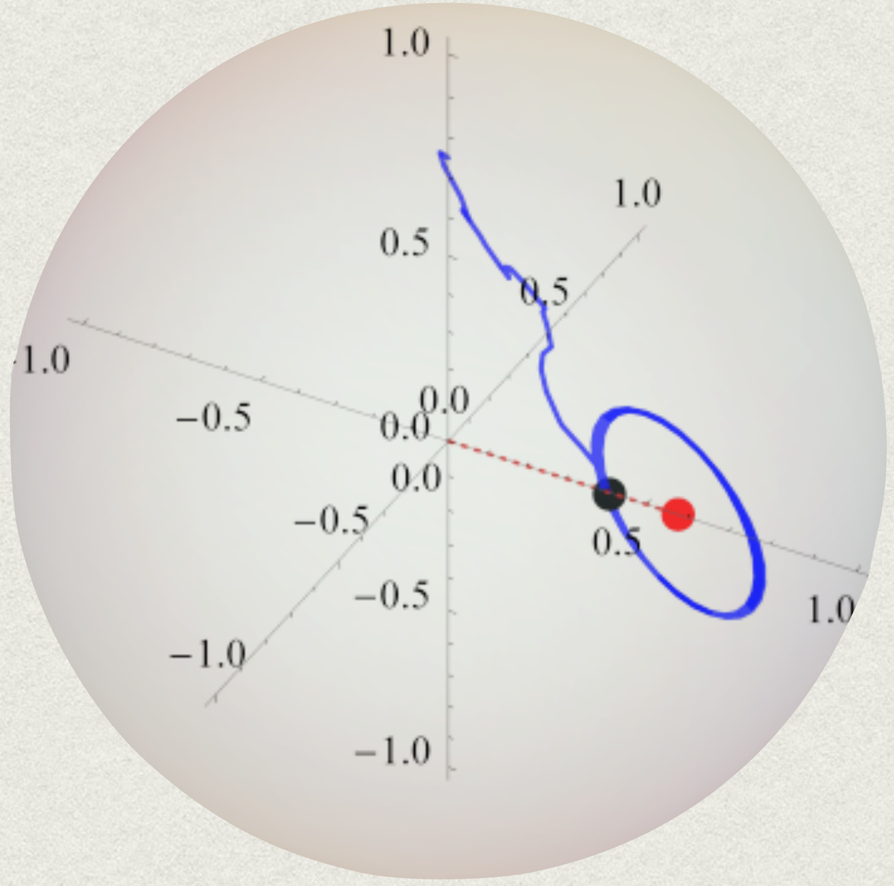


- "Optimism" for training QGANS without limit cycles

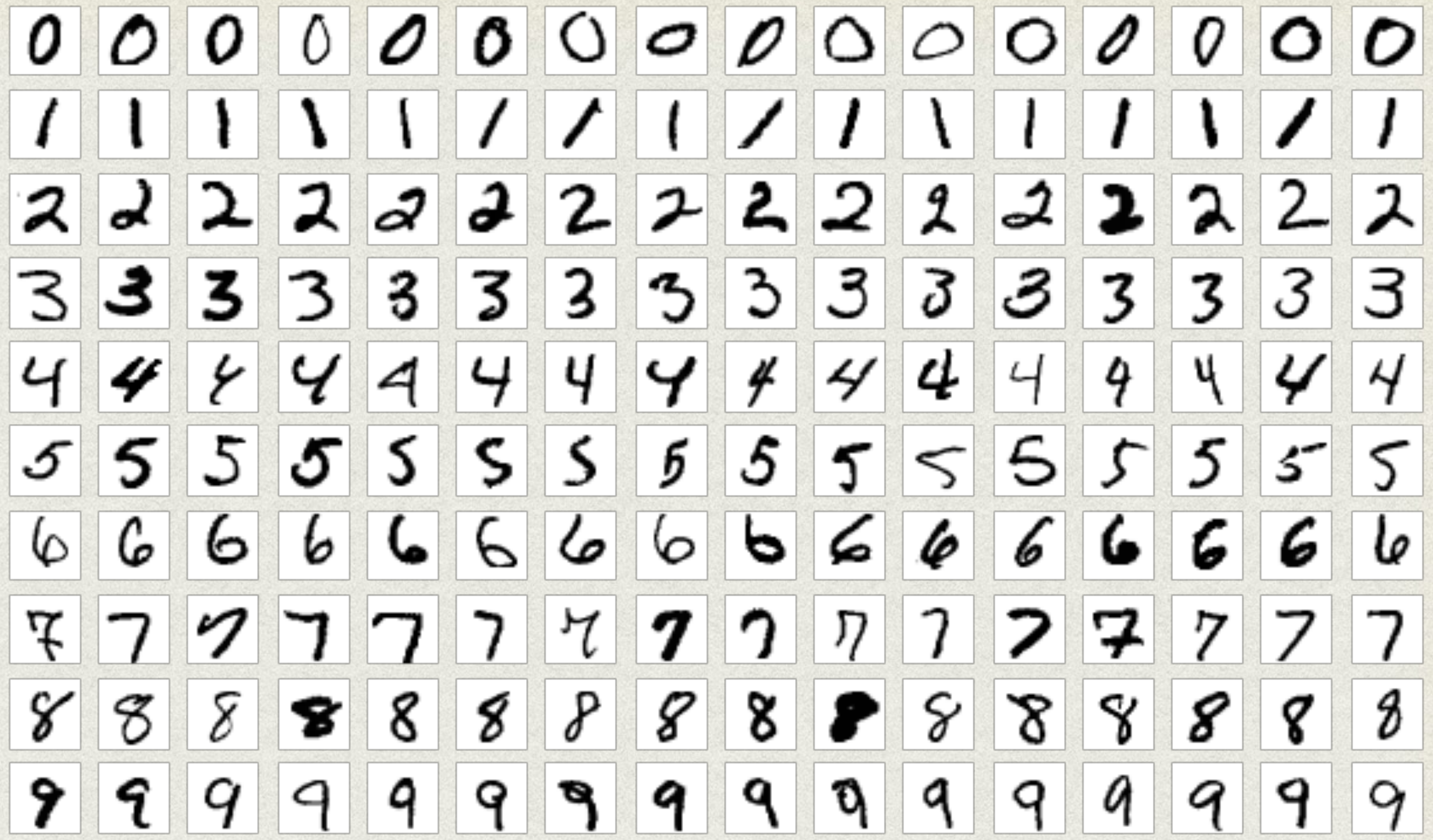
  - P Braccia, F Caruso, L Banchi, New J. Phys. 23 053024 (2021)

$$\boldsymbol{\theta}_D^{t+1} = \boldsymbol{\theta}_D^t + 2\eta_D \nabla_{\boldsymbol{\theta}_D} S(\boldsymbol{\theta}_t^D, \boldsymbol{\theta}_t^G) - \eta_D \nabla_{\boldsymbol{\theta}_D} S(\boldsymbol{\theta}_{t-1}^D, \boldsymbol{\theta}_{t-1}^G)$$

$$\boldsymbol{\theta}_G^{t+1} = \boldsymbol{\theta}_G^t - 2\eta_G \nabla_{\boldsymbol{\theta}_G} S(\boldsymbol{\theta}_{t+1}^D, \boldsymbol{\theta}_t^G) + \eta_G \nabla_{\boldsymbol{\theta}_G} S(\boldsymbol{\theta}_t^D, \boldsymbol{\theta}_{t-1}^G)$$

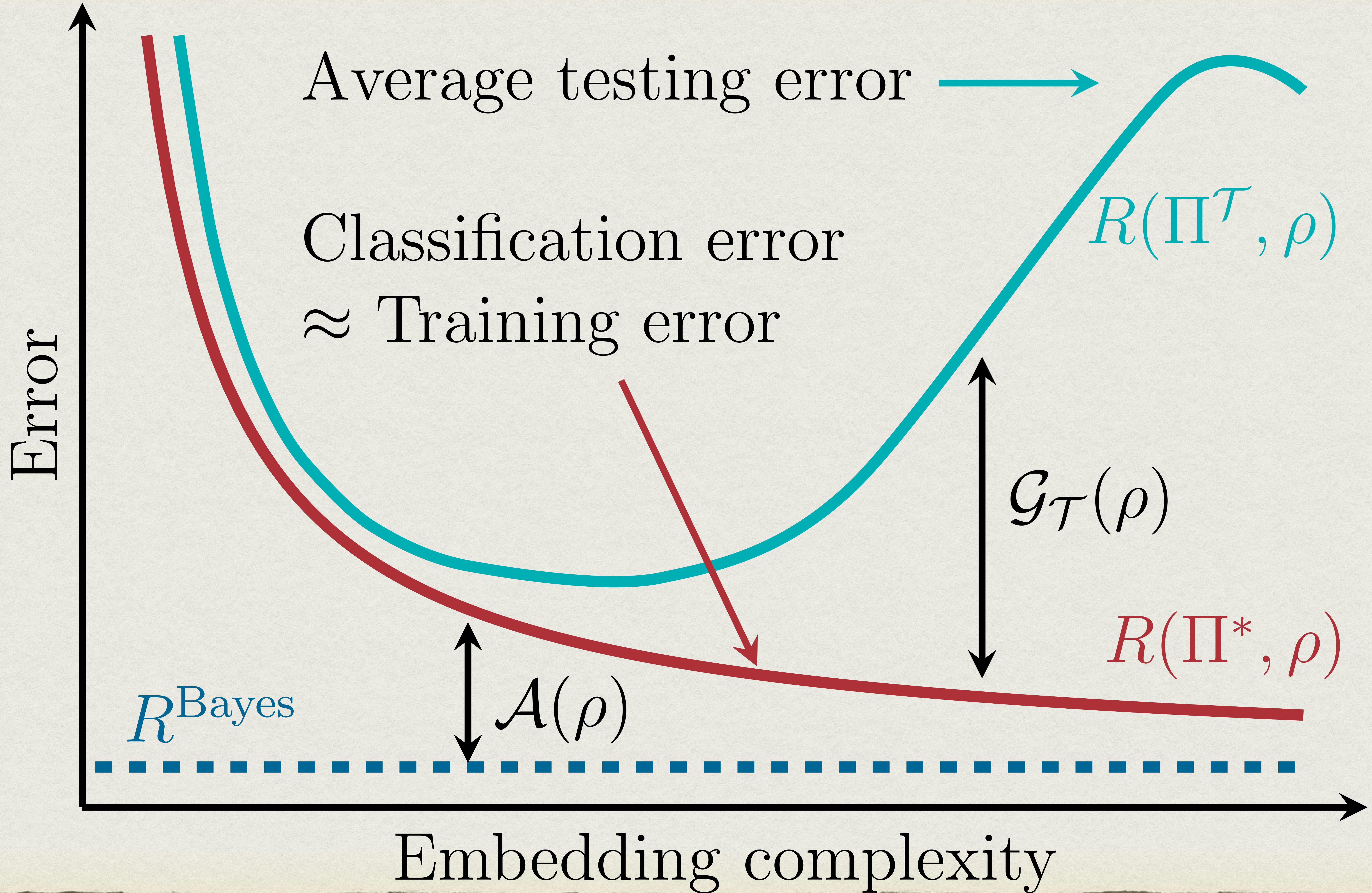


FOUNDATIONAL QUESTION: WHAT CAN WE EXPECT?



- We have a training set $\mathscr{T}$ made of T correctly classified images
- We can empirically check generalisation using a testing set $\mathscr{T}'$ with T' correctly classified images
- We choose a “good" quantum embedding $x \mapsto \rho(x)$ and optimal discrimination via POVM $\{\Pi_c\}$
- **What training error / testing error can we expect?**

"COMPLEXITY" OF QUANTUM EMBEDDINGS



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

QUANTUM BIAS-VARIANCE TRADEOFF

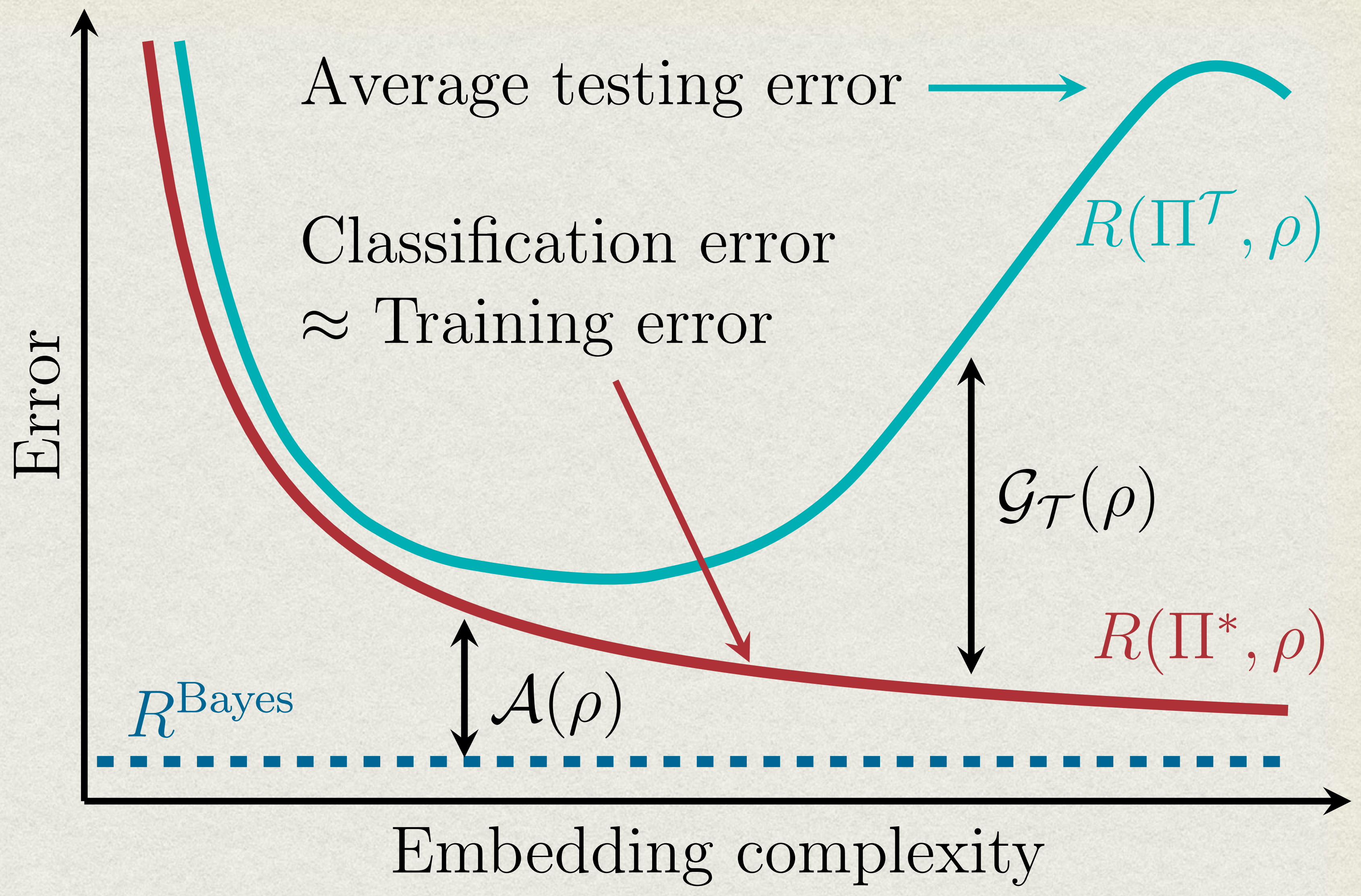
- Bound for the generalisation error

$$G_{\mathcal{T}} \leq 2\sqrt{\frac{\mathcal{B}}{T}} + \sqrt{\frac{2\log(1/\delta)}{T}}$$

- Bound for the approximation error

$$\mathcal{A} = \sum_x \frac{\left|P(x\,|\,0) - P(x\,|\,1)\right|}{2} - \frac{\|\rho_0 - \rho_1\|_1}{2}$$

where $\rho_c = \sum_x P(x\,|\,c)\rho(x)$

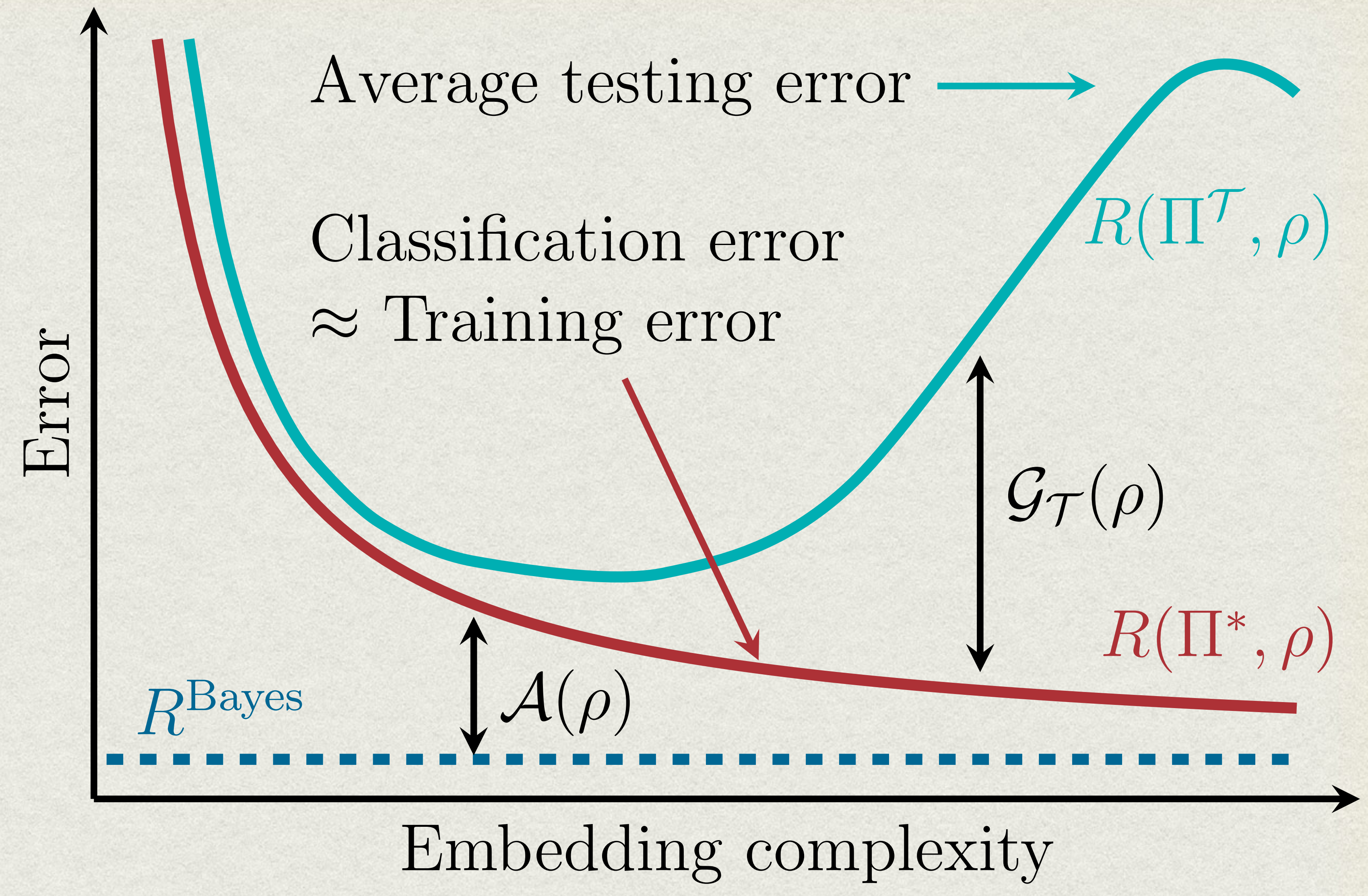


$$\mathcal{B} = 2^{I_2(X:Q)} \qquad \mathcal{A} \leq K - \frac{2^{I(C:Q)}}{N_C}$$

$$\rho_{CXQ} = \sum_{cx} P(c;x)\,|cx\rangle\langle cx| \otimes \rho(x)$$

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

Spoiler: Information Bottleneck!



*Two **extreme cases**:*

- Basis encoding: $x \mapsto \rho(x) = |x\rangle\langle x|$
  minimum $\mathscr{A}(\rho) = 0$, maximum $\mathscr{G}_{\mathscr{T}}(\rho)$

- Constant embedding : $x \mapsto \rho$
  maximum $\mathscr{A}(\rho)$, minimum $\mathscr{G}_{\mathscr{T}}(\rho) = 0$

$$\mathscr{B} = 2^{I_2(X:Q)}$$

$$\mathscr{A} \leq K - \frac{2^{I(C:Q)}}{N_C}$$

for MNIST with 8-bit colours
$28 \times 28 \times 8 \approx 6000$ qubits!

Bigger Hilbert spaces have lower approximation error, but larger generalisation error

In the numerical example we consider

$$x \mapsto |\psi(x)\rangle\langle\psi(x)|^{\otimes N}$$

$$|\psi(x)\rangle = \cos(x)|0\rangle + \sin(x)|1\rangle$$

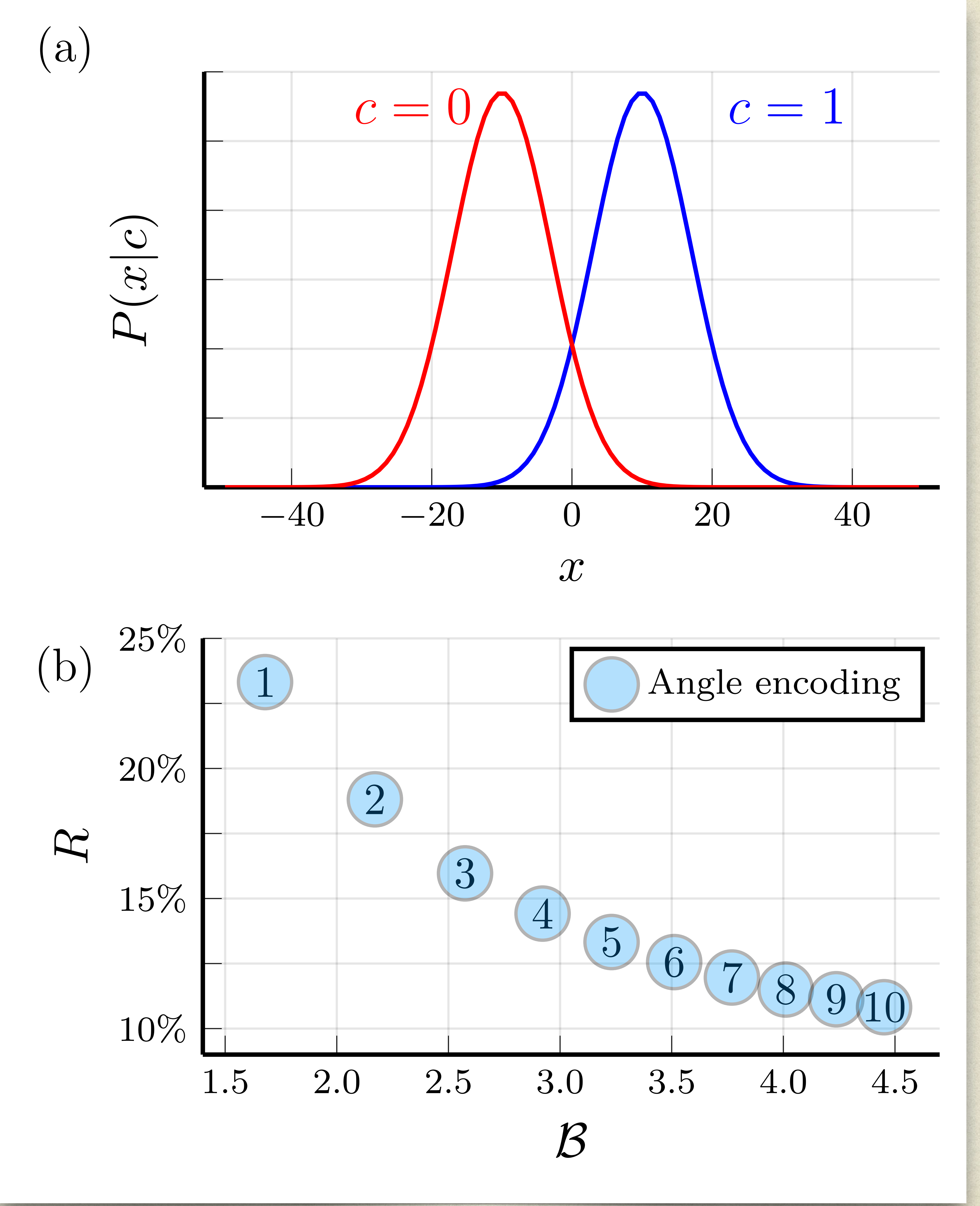
For large N

$$\mathscr{A}(\rho) \leq K F_{\max}^{N}$$

$$\mathscr{B} \approx \mathcal{O}(N)$$

where $F_{\max} = \max_{x \neq y} F(\rho(x), \rho(y))$

Proof from $\rho_{CXQ} = \sum_{cx} P(c,x)|cx\rangle\langle cx| \otimes \rho(x)^{\otimes N}$



QUANTUM KERNELS

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

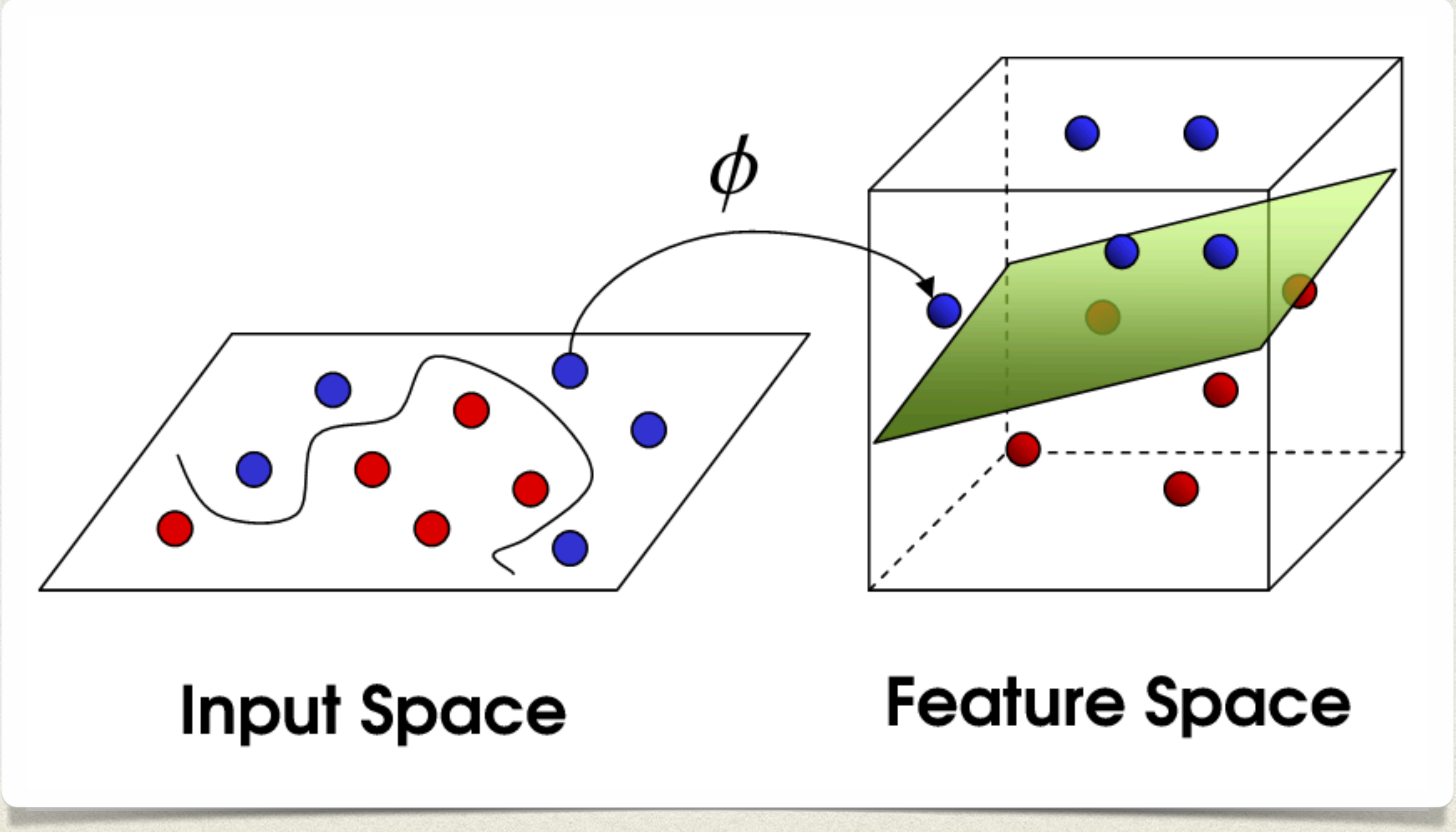
For pure state embeddings $\rho(x) = |\psi(x)\rangle\langle\psi(x)|$ we find

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

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 $K \propto$ identity (bad generalisation)

To favour generalisation the ***final*** *Hilbert space must be small, but the initial one can be big!*

We may iteratively discard information via ***pooling*** *layers (e.g. QCNN)*

Pooling favours generalisation but harms the accuracy (via data processing)

Take home message*: if low training error is achievable with pooling layers, then generalisation can only be better!*

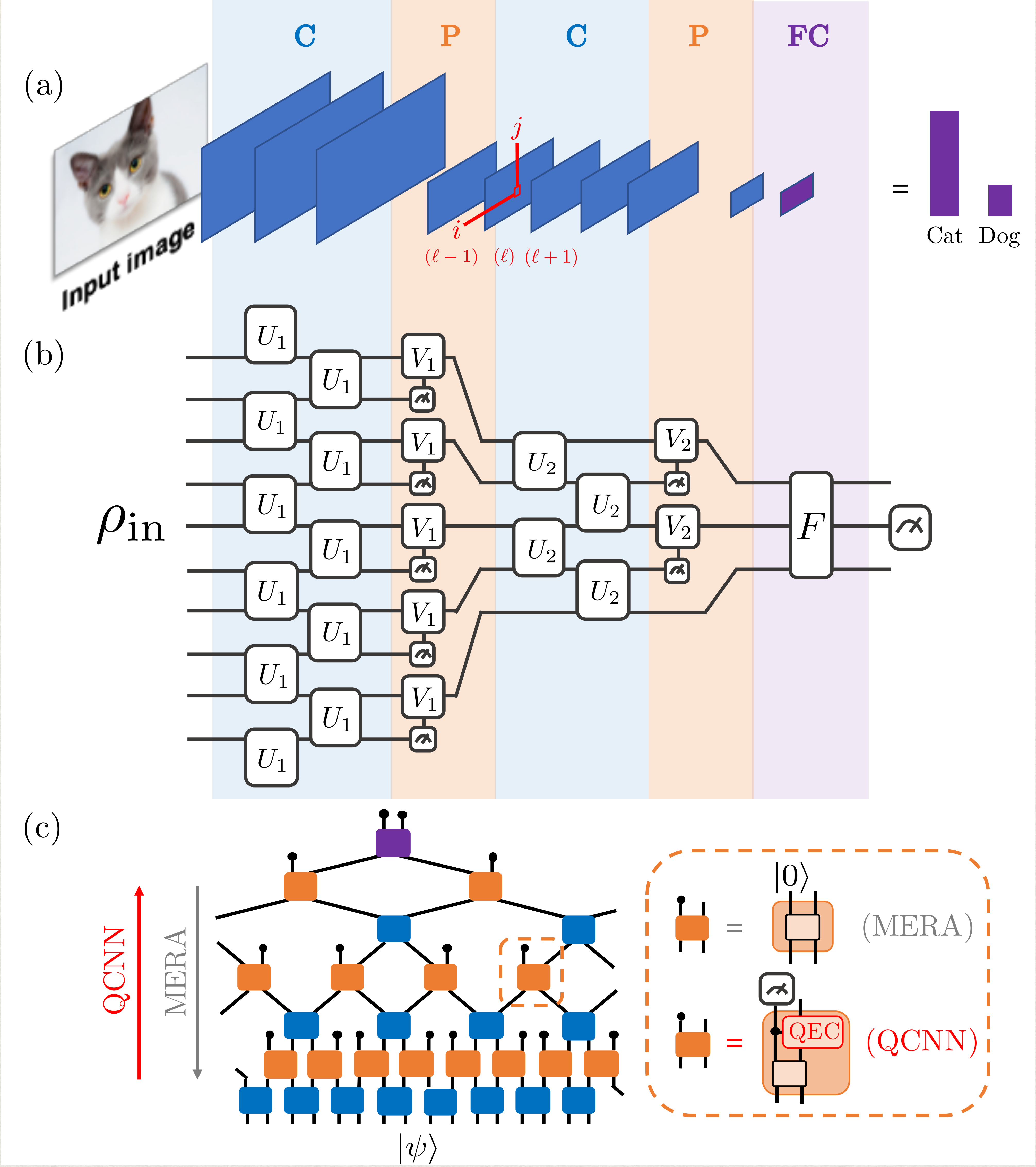


image from:
I Cong, et al. Nature Physics (2019)

INFORMATION BOTTLENECK FOR QUANTUM CLASSIFIERS

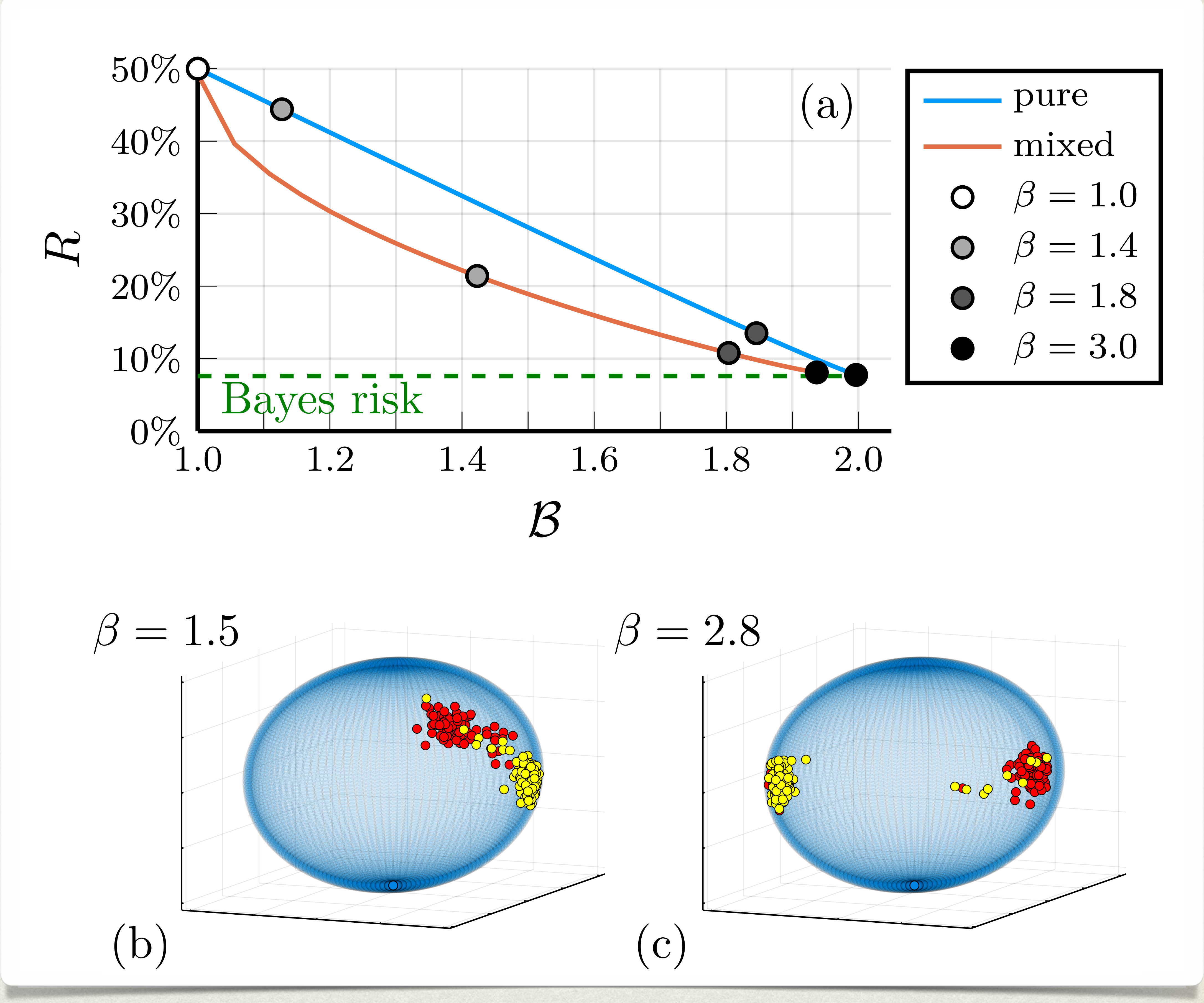
$\rho(x)$ as "bottleneck" that squeezes the relevant information that x provides about c

IB principle (loss independent): minimise

$$\mathscr{L}_{IB} = I(X{:}Q) - \beta I(C{:}Q)$$

Self-consistent solutions (similar for ρ)

$$\tilde{\lambda}_z |\psi(z)\rangle = e^{(1-\beta)\log\bar{\rho} + \beta\sum_c P(c|z)\log\rho_c} |\psi(z)\rangle$$



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

APPLICATIONS

QUANTUM PHASE RECOGNITION (QQ)

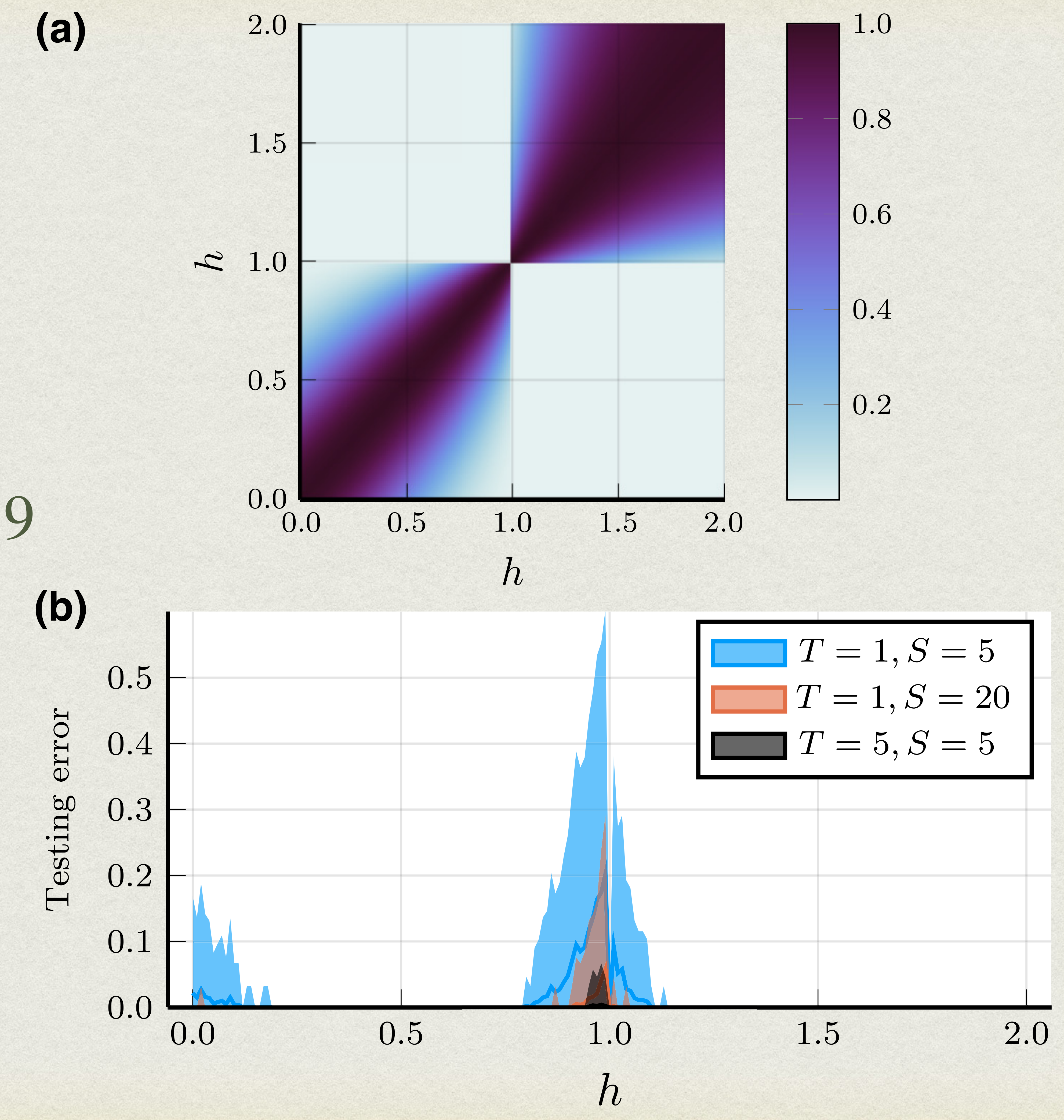
Task: recognize the phases of matter of a quantum many-body system by taking measurements on the quantum system itself

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

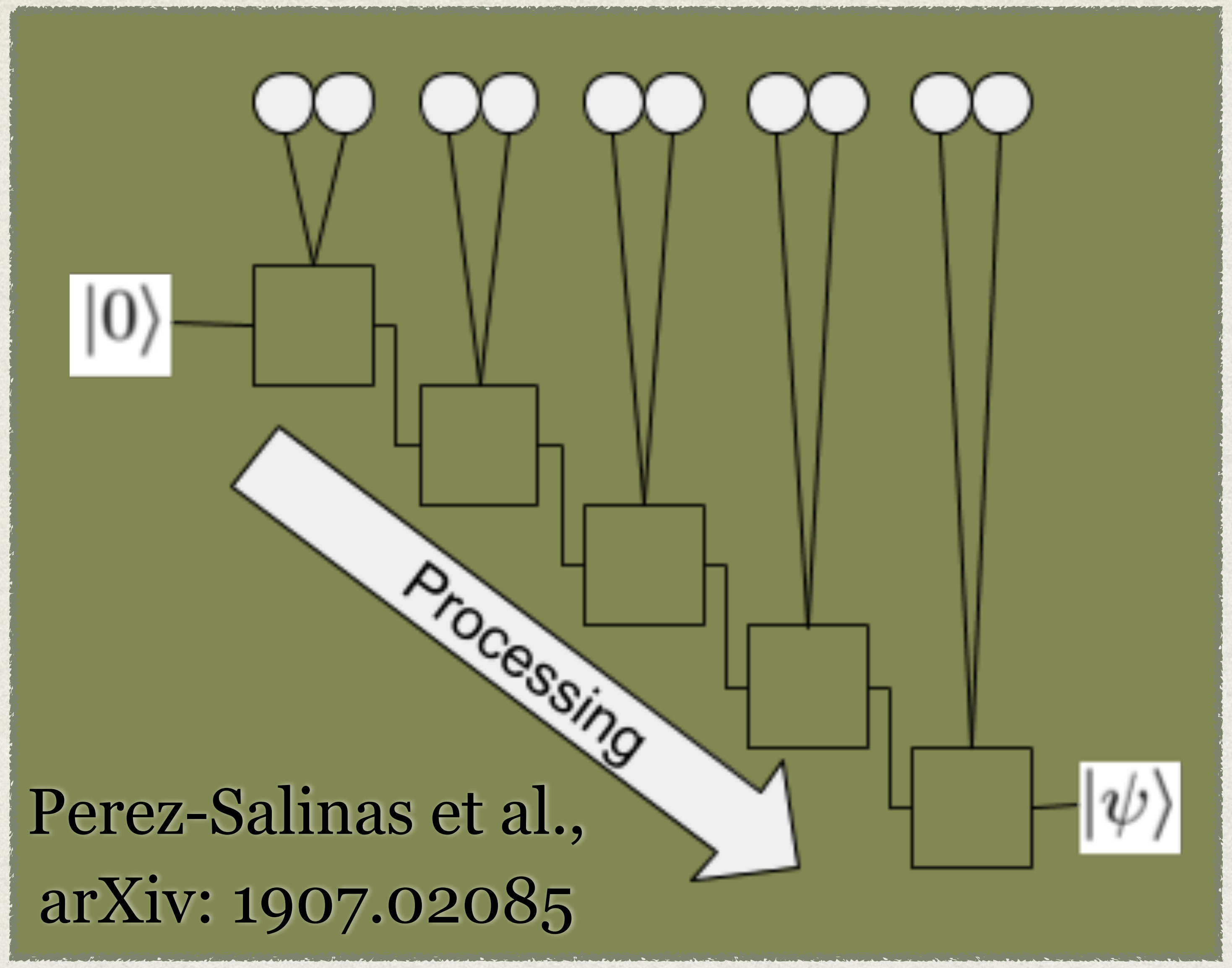
Ordered ($|h| < 1$) / disordered ($|h| > 1$) phases

T : number of training samples per class

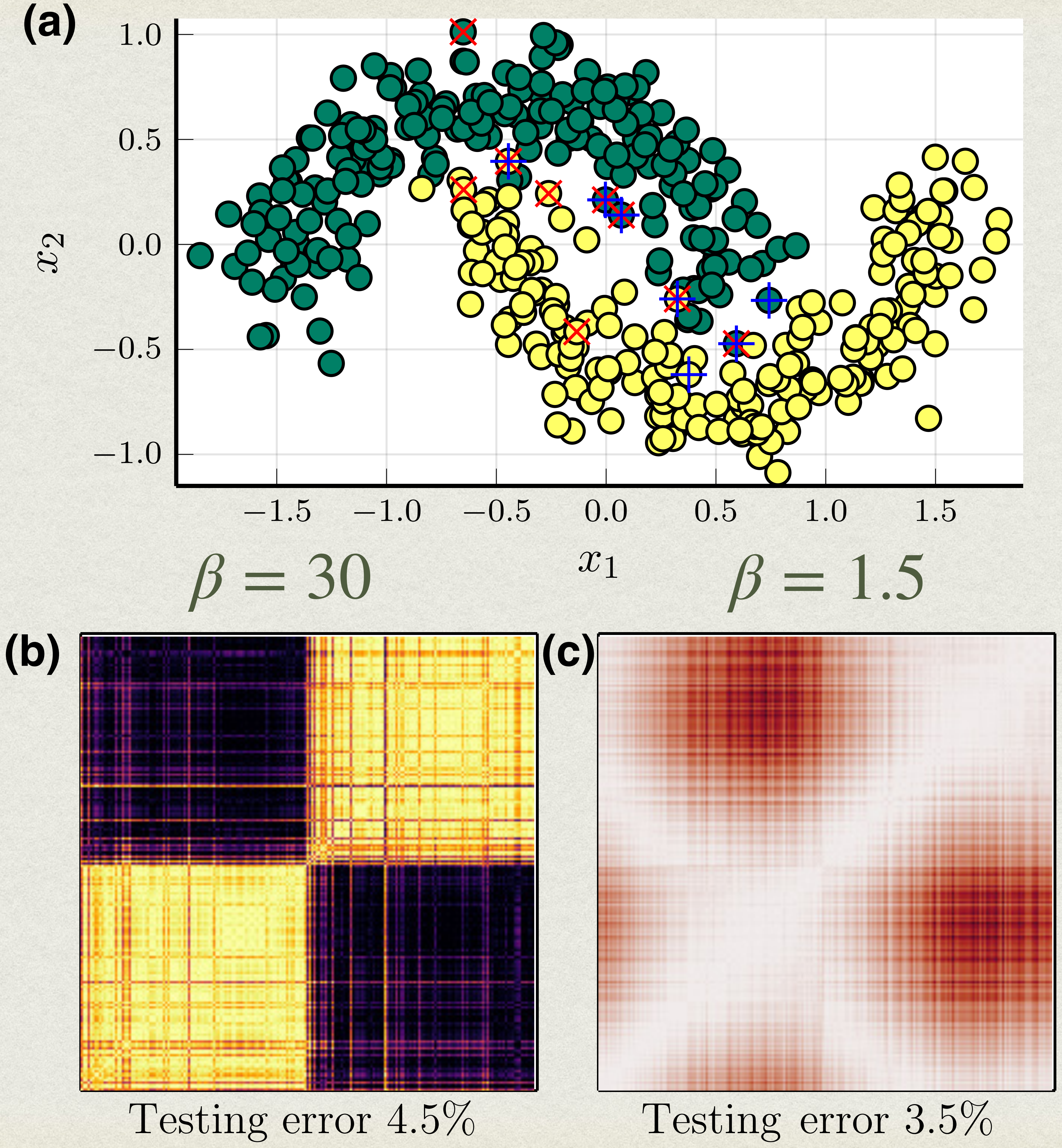
S : number of measurement shots



VARIATIONAL QUANTUM INFORMATION BOTTLENECK (CQ)



Single-qubit
"Data Re-uploading"
Classifier

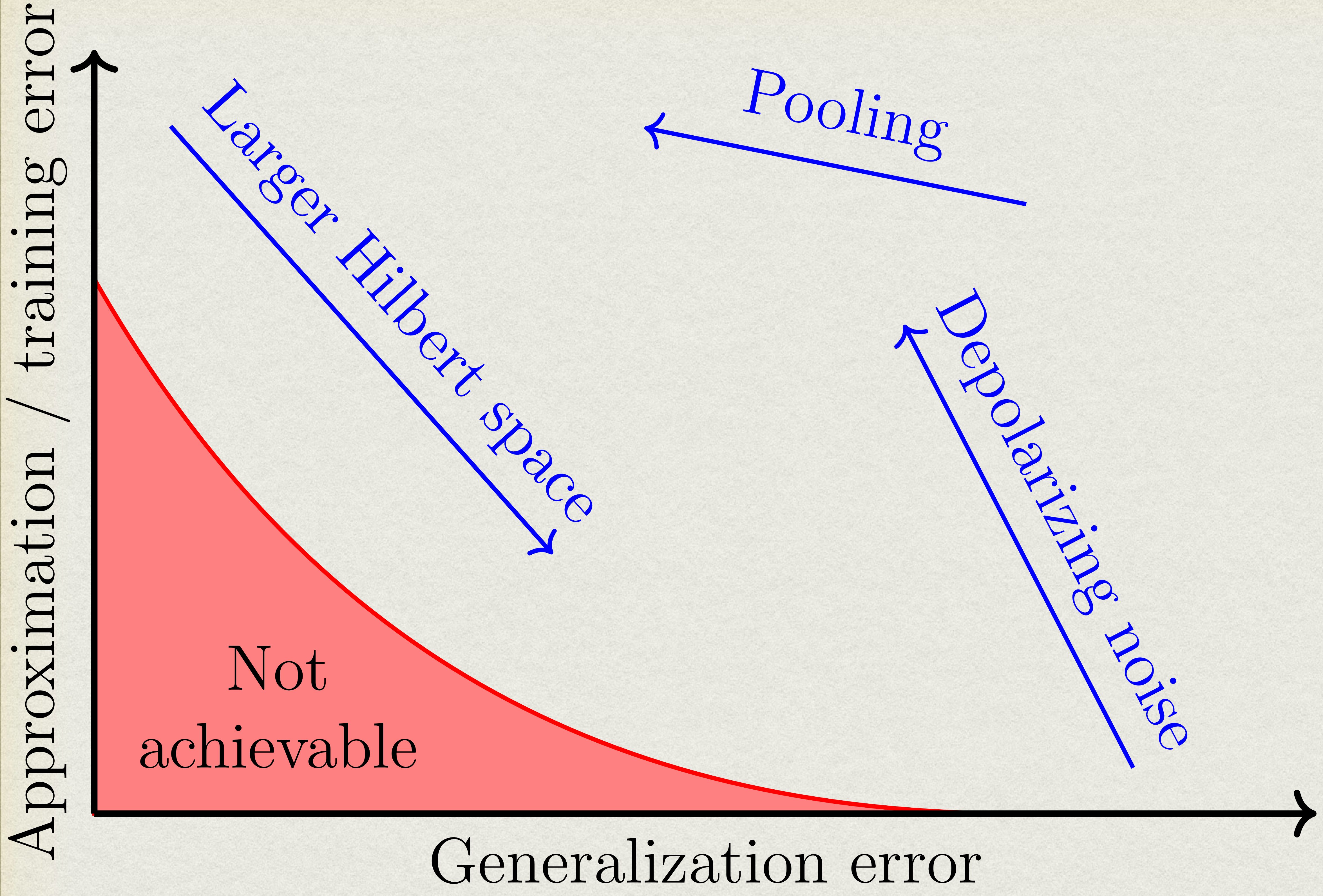


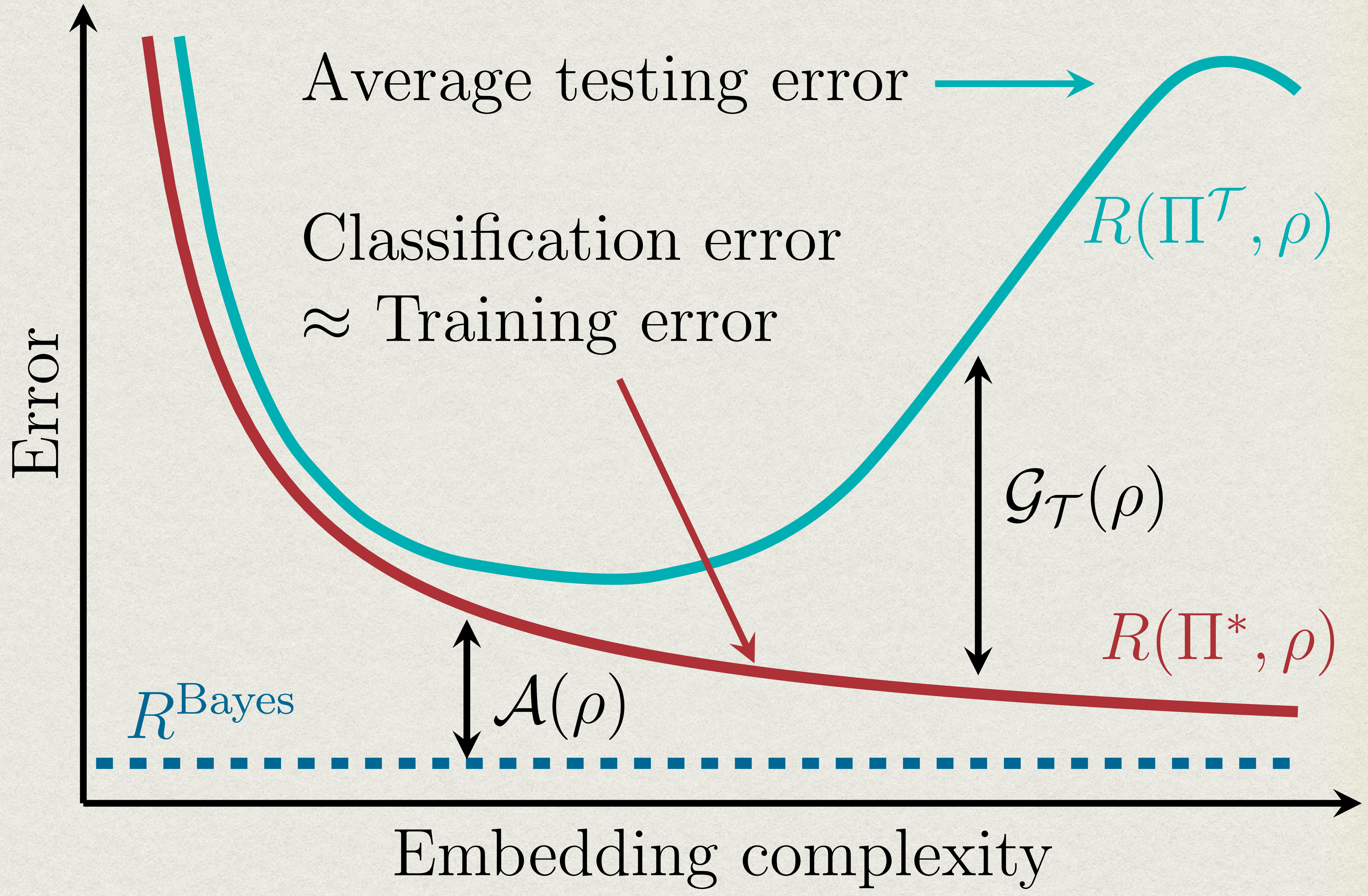
$$|\psi_w(x)\rangle = \prod_{\ell=1}^{L}[R^z(w^{z\ell}\cdot x + w_0^{z\ell})R^y(w^{y\ell}\cdot x + w_0^{y\ell})]|0\rangle,$$

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

SUMMARY

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





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

- Empirical loss / training error

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

- 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 $\Pi^{\mathcal{T}} = \mathrm{argmin}_{\Pi} R_{\mathcal{T}}(\Pi, \rho)$

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

- Testing error

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