LEONARDO BANCHI

UNIVERSITY OF FLORENCE (ITALY)

QUANTIFYING THE COMPLEXITY OF LEARNING QUANTUM FEATURES

EASY VS COMPLEX QUANTUM LEARNING PROCESSES

$$x \longrightarrow \rho(x, y) \longrightarrow y$$

- Classical input x, classical feature y
- By varying x we can explore different states
- We want to extract the feature y e.g. via measurements

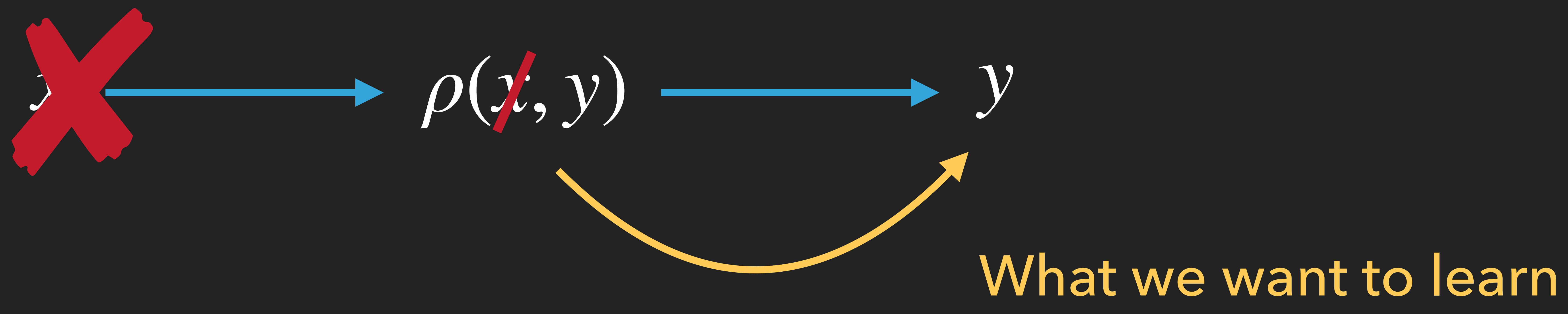
EXAMPLES

What we want to learn

- Classical machine learning or QML with classical data
- Sample complexity in supervised learning: number of training data N
- (Validation) Error

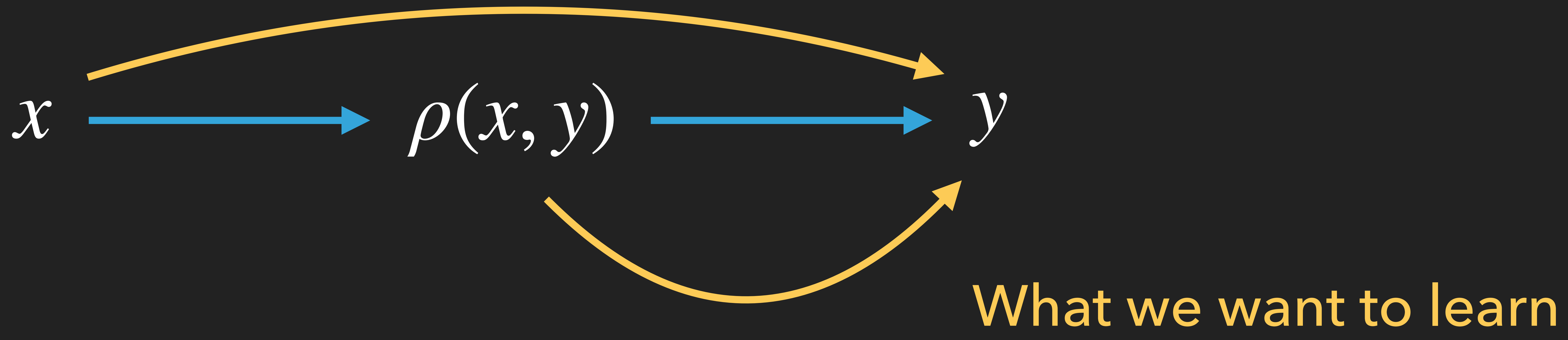
$$\epsilon = \mathcal{O}\left(\sqrt{\frac{B}{N}}\right)$$

EXAMPLES



- Quantum sensing / Quantum hypothesis testing
- Sample complexity is the number of copies of ρ (e.g. number of measurement shots) S
- Continuous y (metrology): mean square error $\epsilon = \mathcal{O}\left(S^{-1}\right)$ or $\epsilon = \mathcal{O}\left(S^{-2}\right)$
- Discrete y: probability of misclassification $\epsilon = 1 - 2^{-H_{\min}(Q|Y)} \approx e^{-\alpha S}$

OPEN QUESTION



- How many copies S (e.g. measurement shots) of each $\rho(x, y)$ do we need?
- How many values N of x does our learning strategy need to be able to extract y for all possible x? Generalisation
- Error: $\epsilon = f(N, S)$?

Open question

WORST CASE

$$\epsilon = \mathcal{O}\left((NS)^{-1/3}\right)$$

PERSPECTIVE *Adv. Quantum Technol.* **2024**, 2300311

ADVANCED QUANTUM TECHNOLOGIES
www.advquantumtech.com

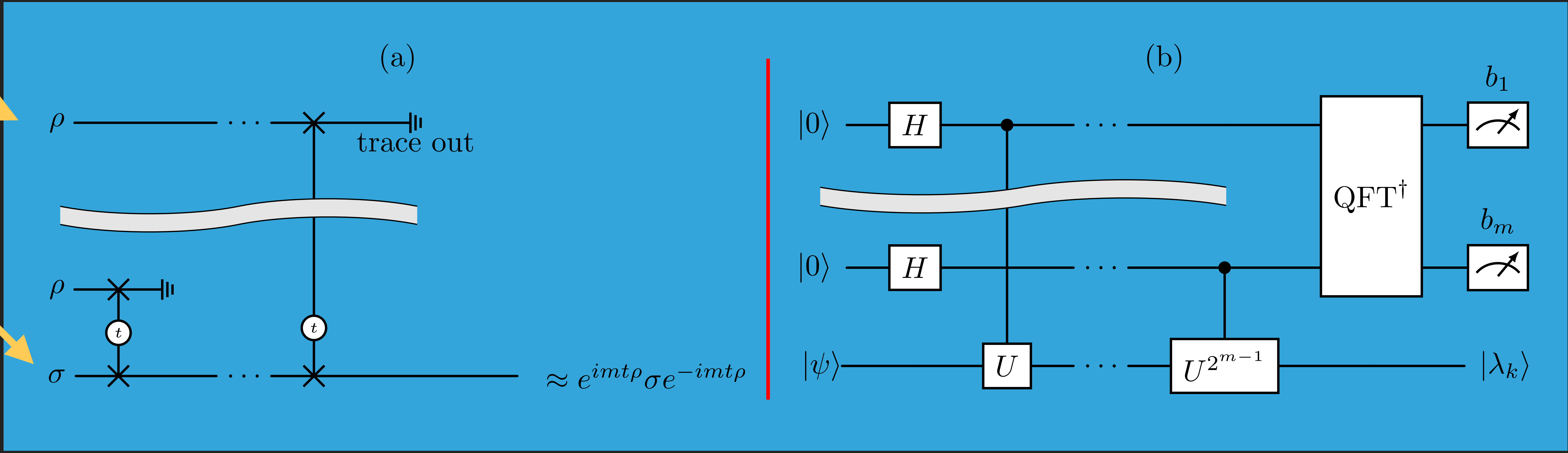
Statistical Complexity of Quantum Learning

Leonardo Banchi, Jason Luke Pereira, Sharu Theresa Jose, and Osvaldo Simeone*

$$\epsilon = \mathcal{O}\left(d\,(NS)^{-1/2}\right)$$ with tomography

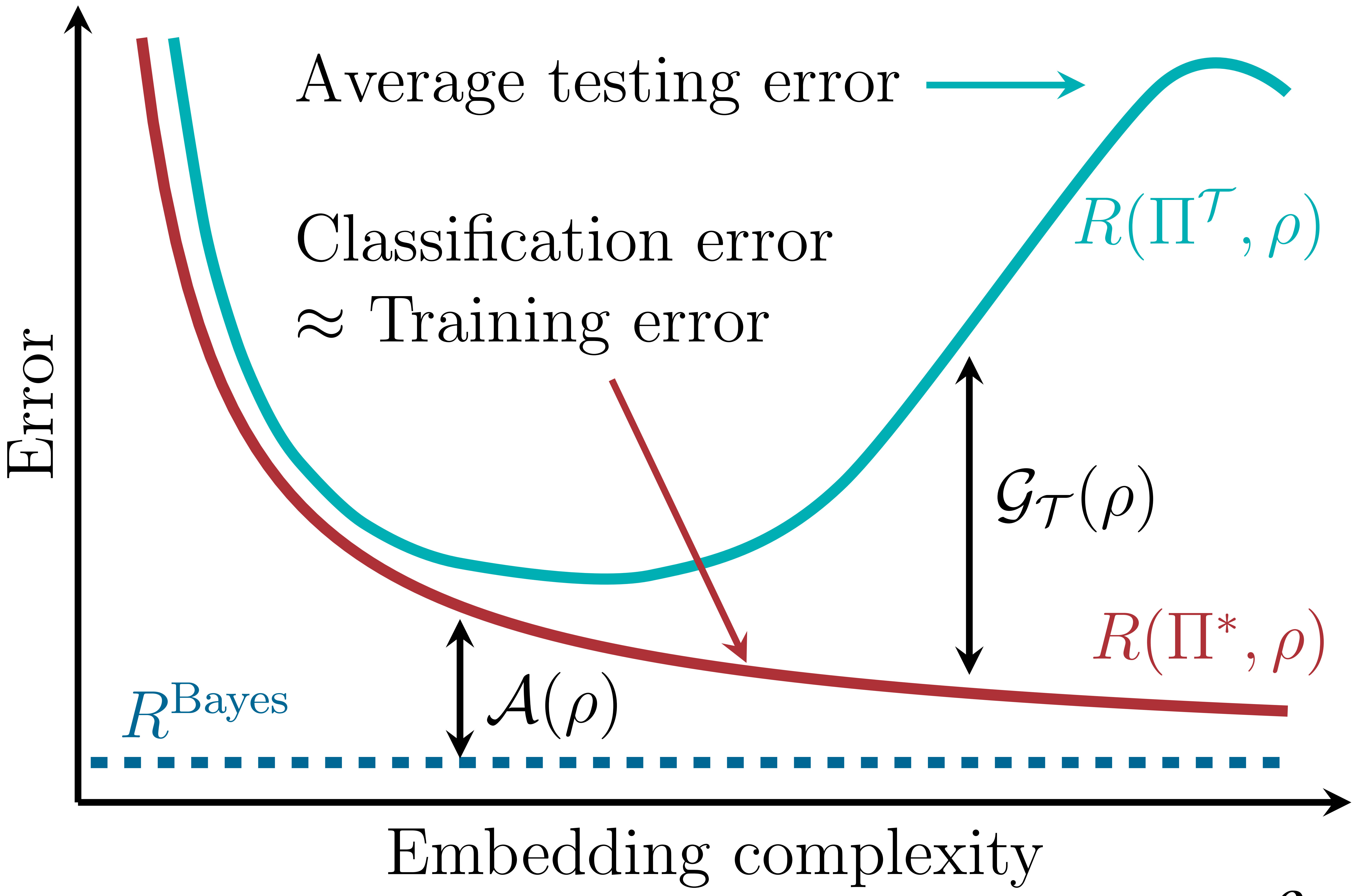
Training states

Test state



LEARNING WITH QUANTUM STATES

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



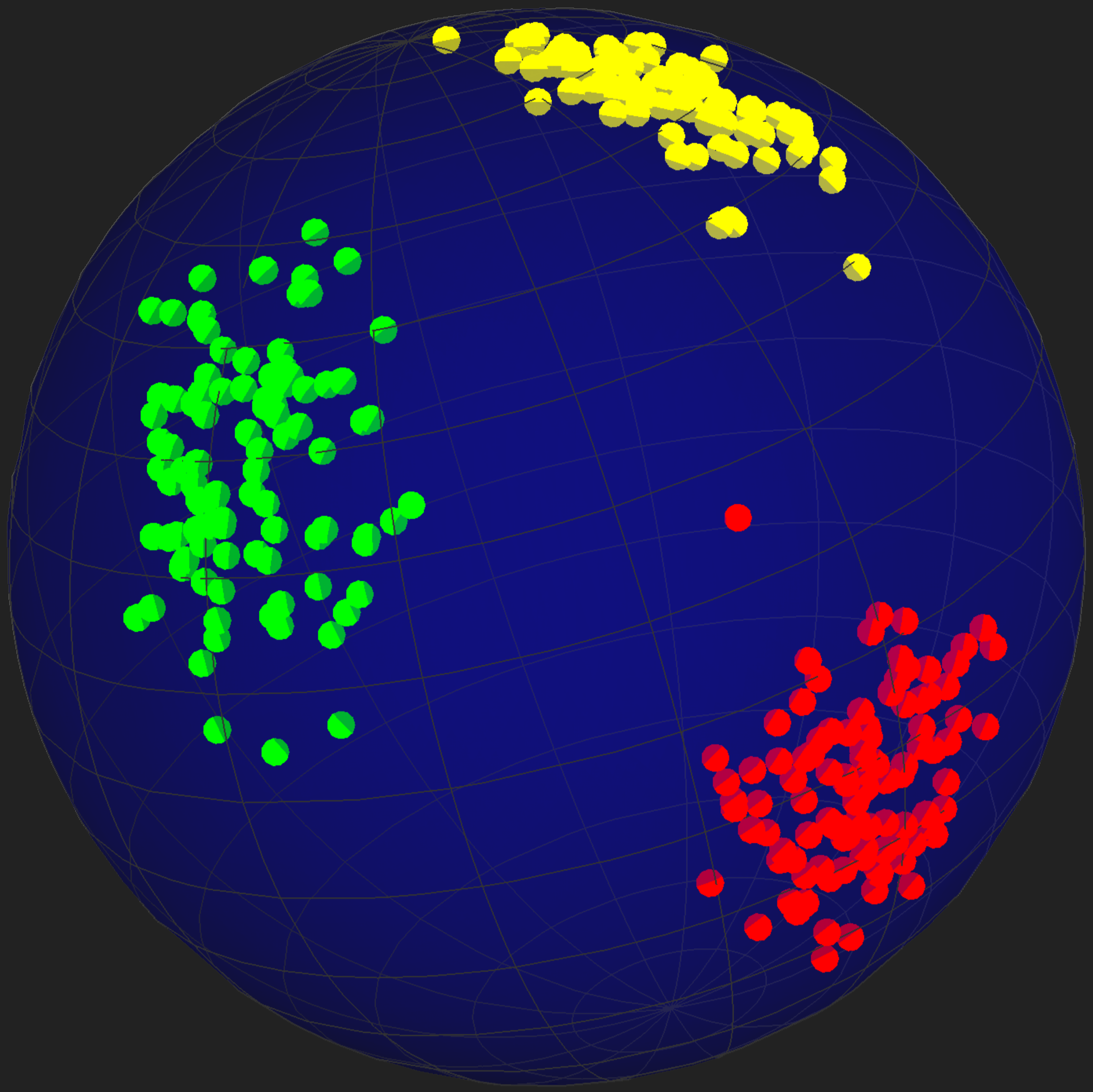
$$\mathscr{G}_{\mathscr{T}} \approx \sqrt{\frac{2^{I_2(X:Q)}}{N}}$$

$$\mathscr{A} = K - \frac{2^{I(Y:Q)}}{N_y}$$

$$\rho_{XYQ} = \frac{1}{N}\sum_{n=1}^{N} |x_n y_n\rangle\langle x_n y_n| \otimes \rho(x_n, y_n)$$

CLUSTERING IN HILBERT SPACES

- **Generalisation depends on both data and algorithm**
- Good generalisation when the states are **clustered** in the Hilbert space
- Large $I(Y : Q)$, small $I(X : Q)$



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

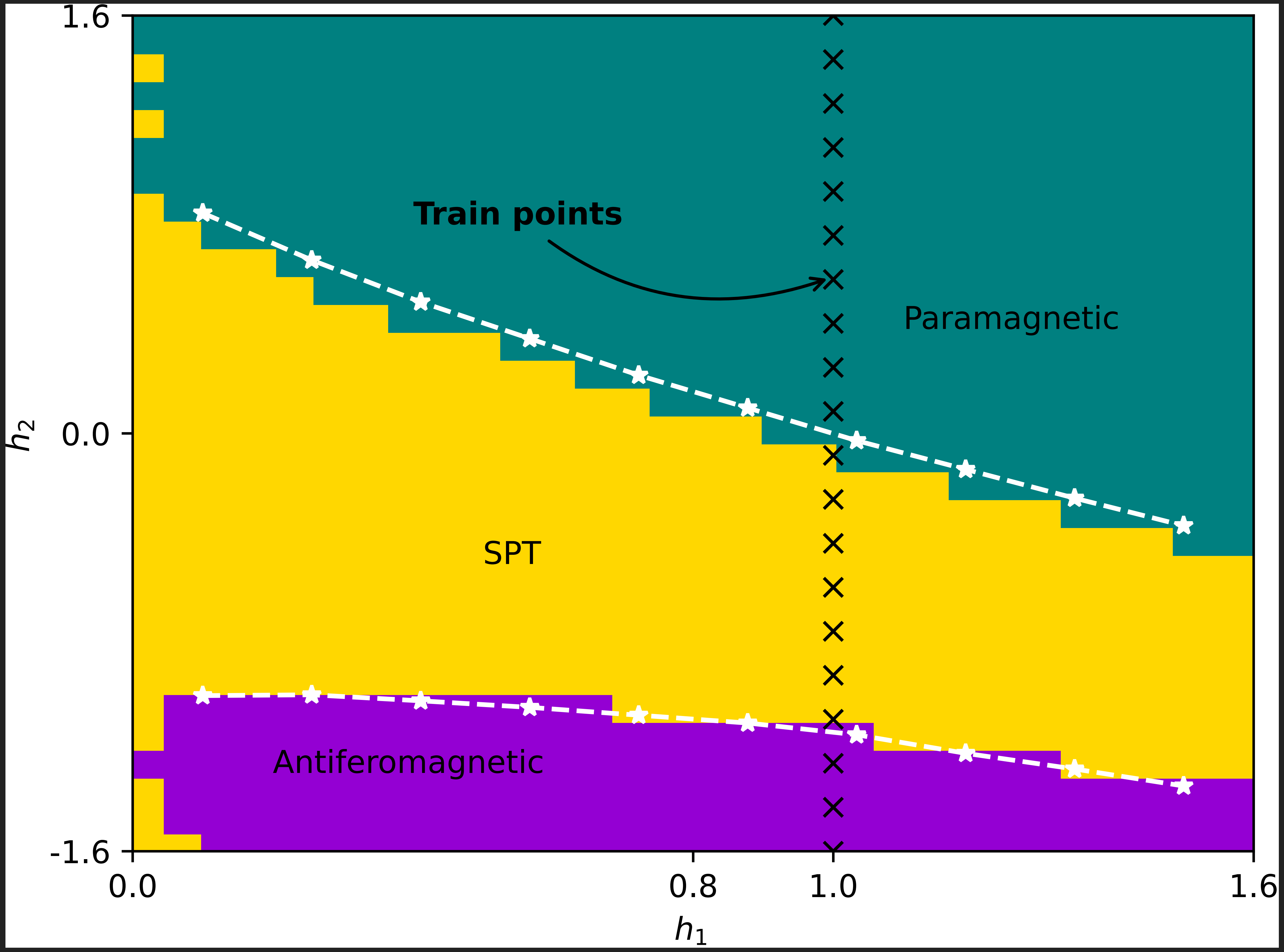
LEARNING TO CLASSIFY QUANTUM PHASES OF MATTER

RECOGNISING QUANTUM PHASES OF MATTER IS "EASY"

$$x \longrightarrow \rho(x, y) \longrightarrow y$$

- $x = (h_1, h_2)$ is the set of classical Hamiltonian parameters (magnetic fields, couplings)
- $\rho(x, y)$ is the ground state
- y is the phase (paramagnetic, topological…)

SEE MEHRAN'S TALKS



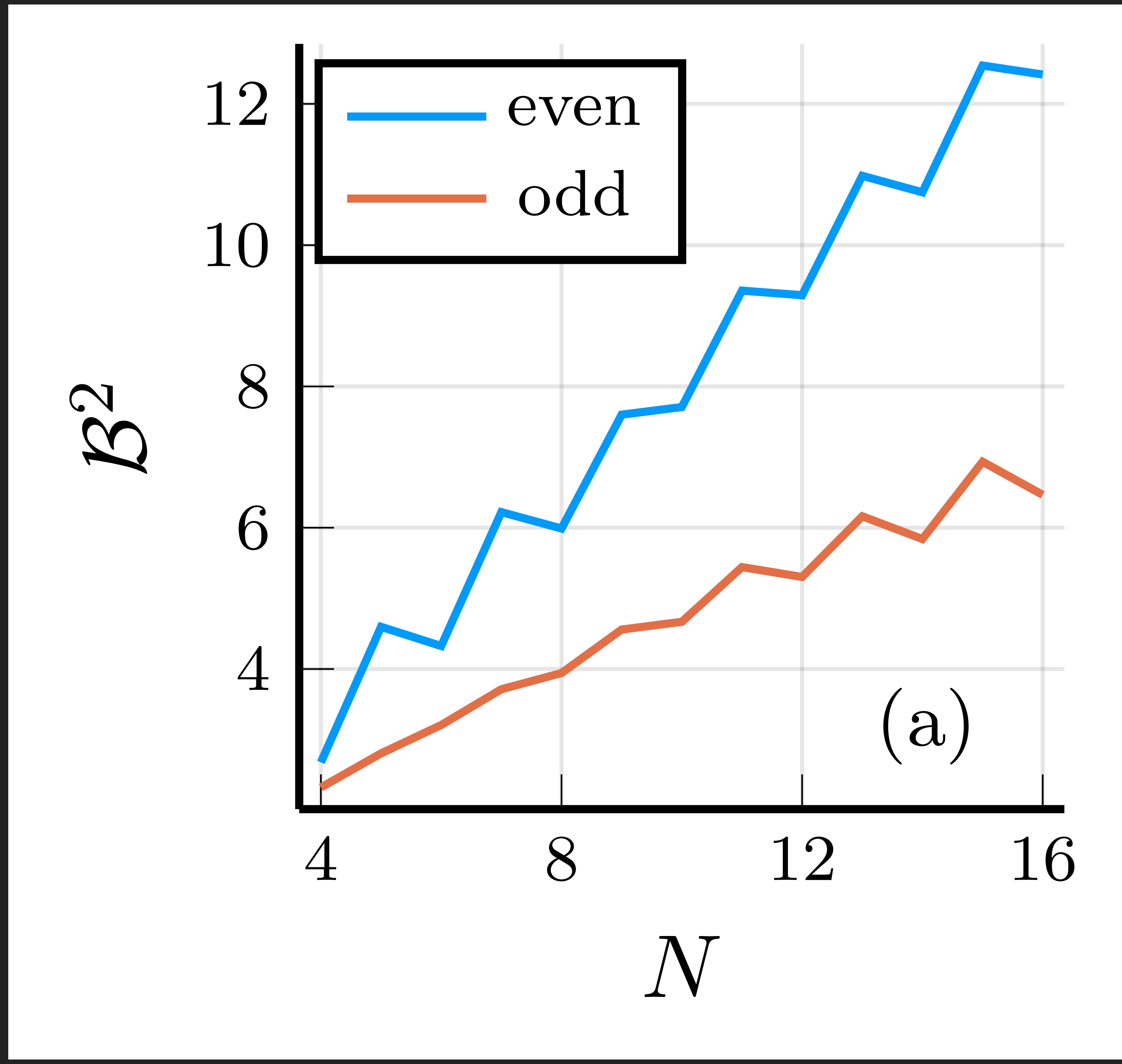
WHY LEARNING QUANTUM PHASES OF MATTER ?

$$x \longrightarrow \rho(x, y) \longrightarrow y$$

We learn the observable

Generalisation bound grows (linearly?) with the number of qubits

The learnt observable requires few measurements, e.g. single-qubit measurements



See arXiv:2409.05188

LEARNING TO CLASSIFY ENTANGLEMENT WITHOUT KNOWING ENTANGLEMENT THEORY

LEARNING TO CLASSIFY ENTANGLEMENT IS "HARD"

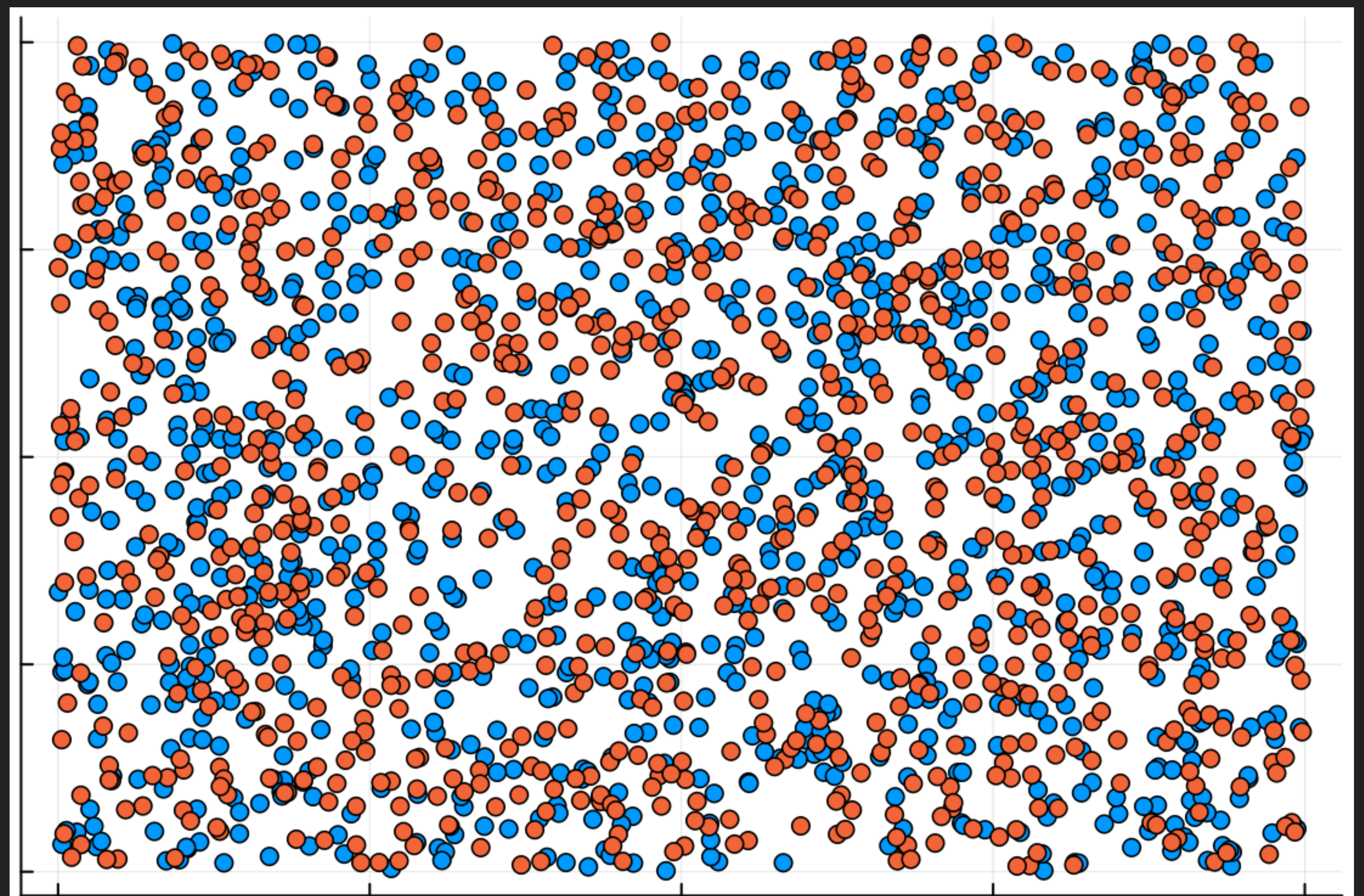
Toy problem: binary classification of separable (class 0) vs maximally entangled (class 1) states

$$|\psi_0(x)\rangle = U_A(x)|0\rangle \otimes U_B(x)|0\rangle \qquad\qquad |\psi_1(x)\rangle \propto U_A(x) \otimes U_B(x) \sum_{i=1}^{d} |i,i\rangle_{AB}$$

Simple analytical strategy: measure the purity (~ Renyi-2 entropy) of either A or B subsystems. Outcomes 1 vs $\mathcal{O}(d^{-1})$ variance $\left(1 - d^{-2}\right)/S$

The states from both classes uniformly populate the Hilbert space, the two classes are not clustered

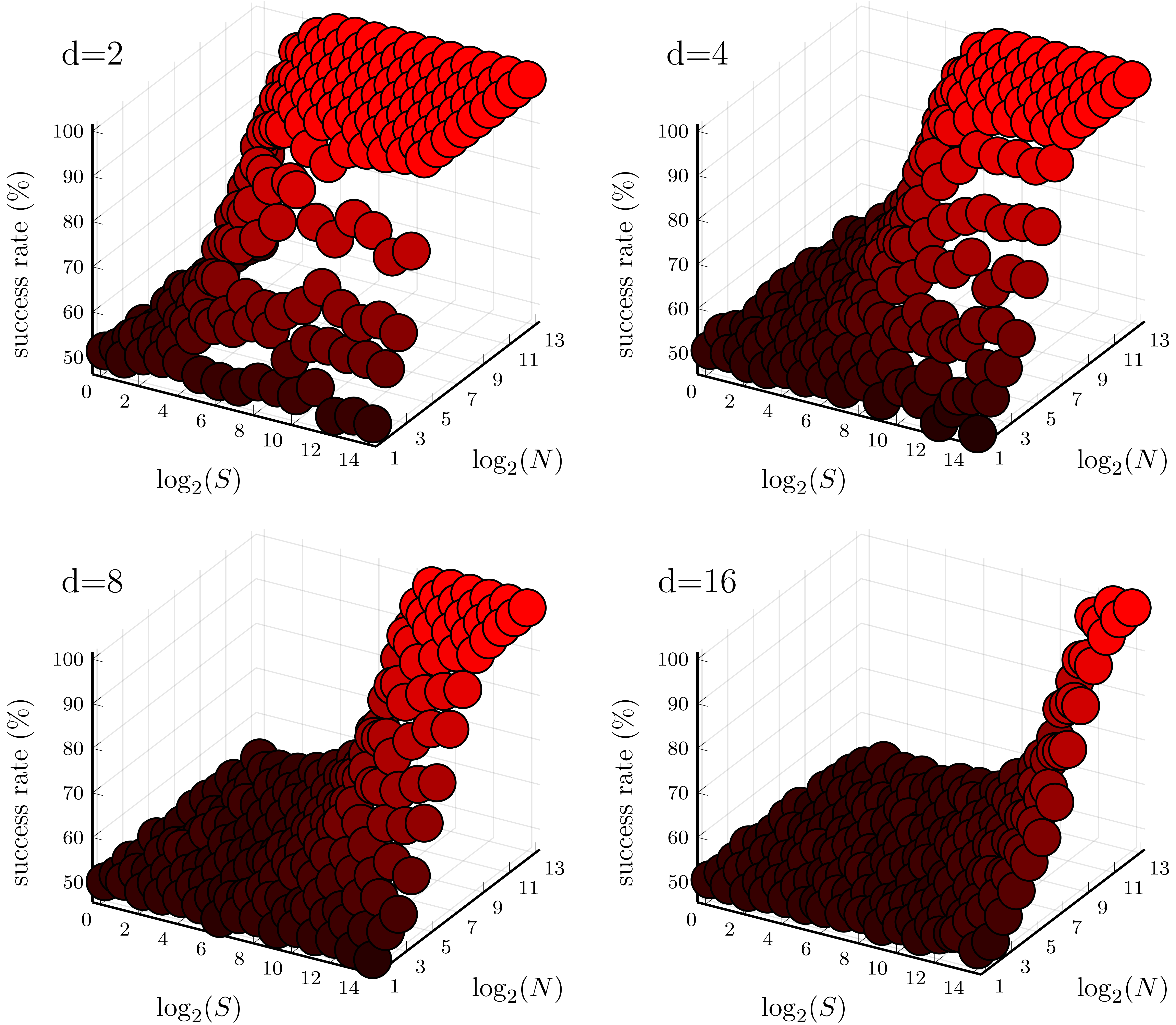
$$\int \mathrm{d}x \ |\psi_0(x)\rangle\langle\psi_0(x)| = \int \mathrm{d}x \ |\psi_1(x)\rangle\langle\psi_1(x)| = \frac{\mathbb{1} \otimes \mathbb{1}}{d^2}$$



Linear observables cannot discriminate the two states.

CLASSIFICATION WITH SUPPORT VECTOR MACHINES

Banchi et al. arXiv (soon)



$$K(x,x') = |\langle \psi(x) | \psi(x') \rangle|^4$$
$$= \mathrm{Tr}\left[\rho(x)^{\otimes 2} \rho(x')^{\otimes 2}\right]$$

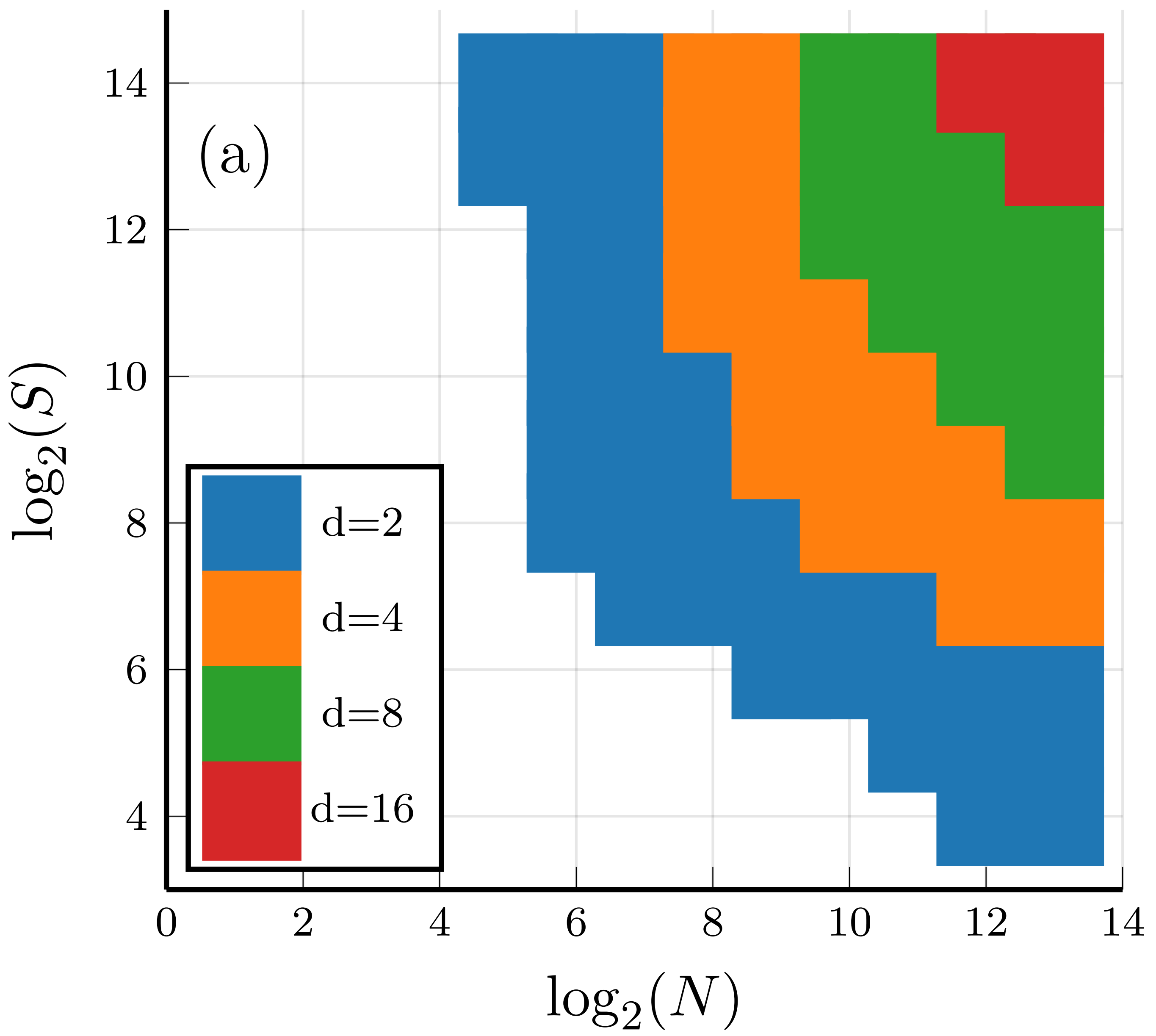
Feature space

Theorem: if we estimate kernels via the SWAP test, the optimal observable learnt from data converges to the purity measurement when $N \to \infty, S \to \infty$

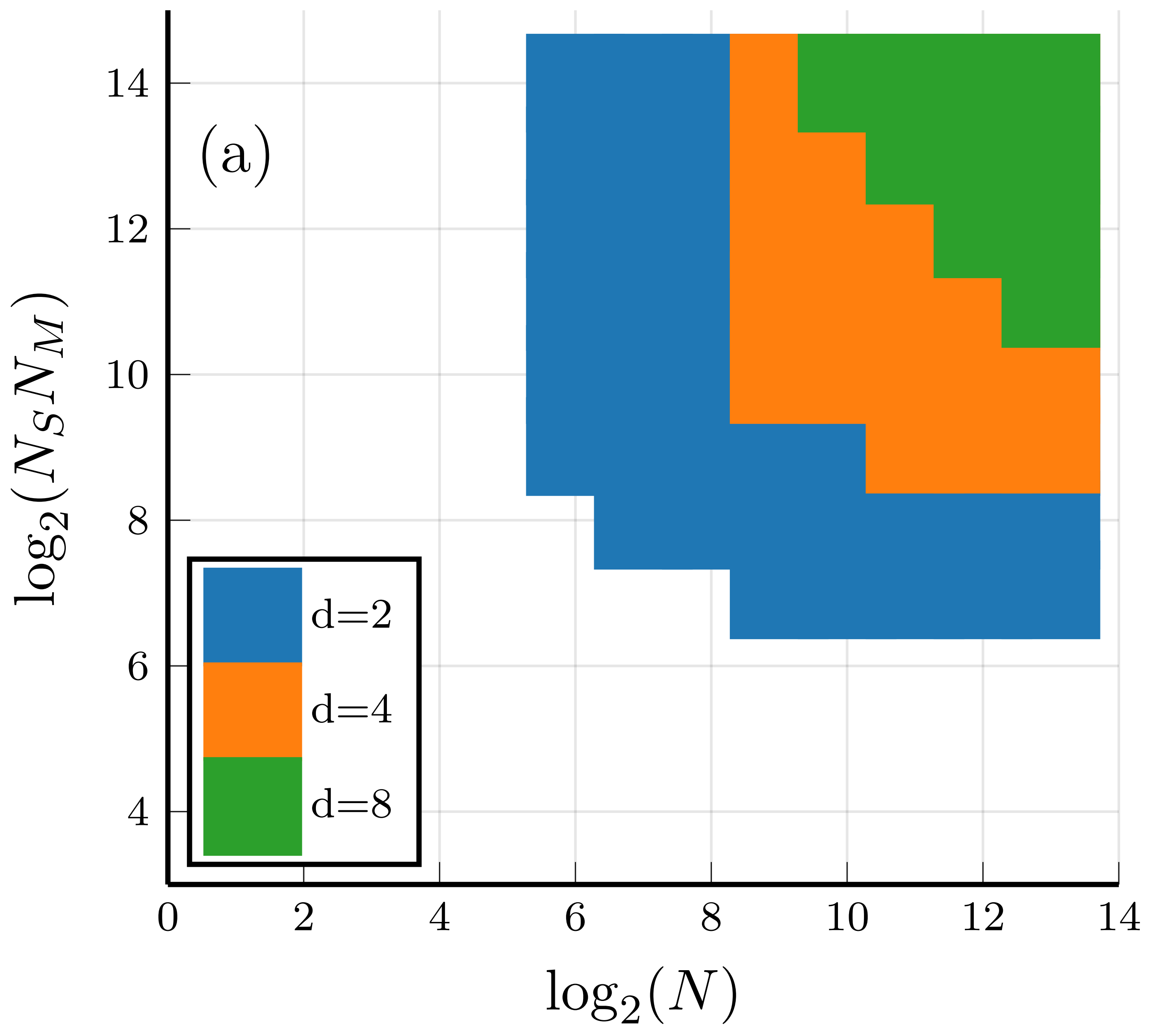
Error ϵ with $N = \frac{1}{\epsilon}\left(1 + \frac{d^4}{S}\right)^2$ data

PARAMETERS WITH ACCURACY HIGHER THAN 99%

Banchi et al. arXiv (soon)

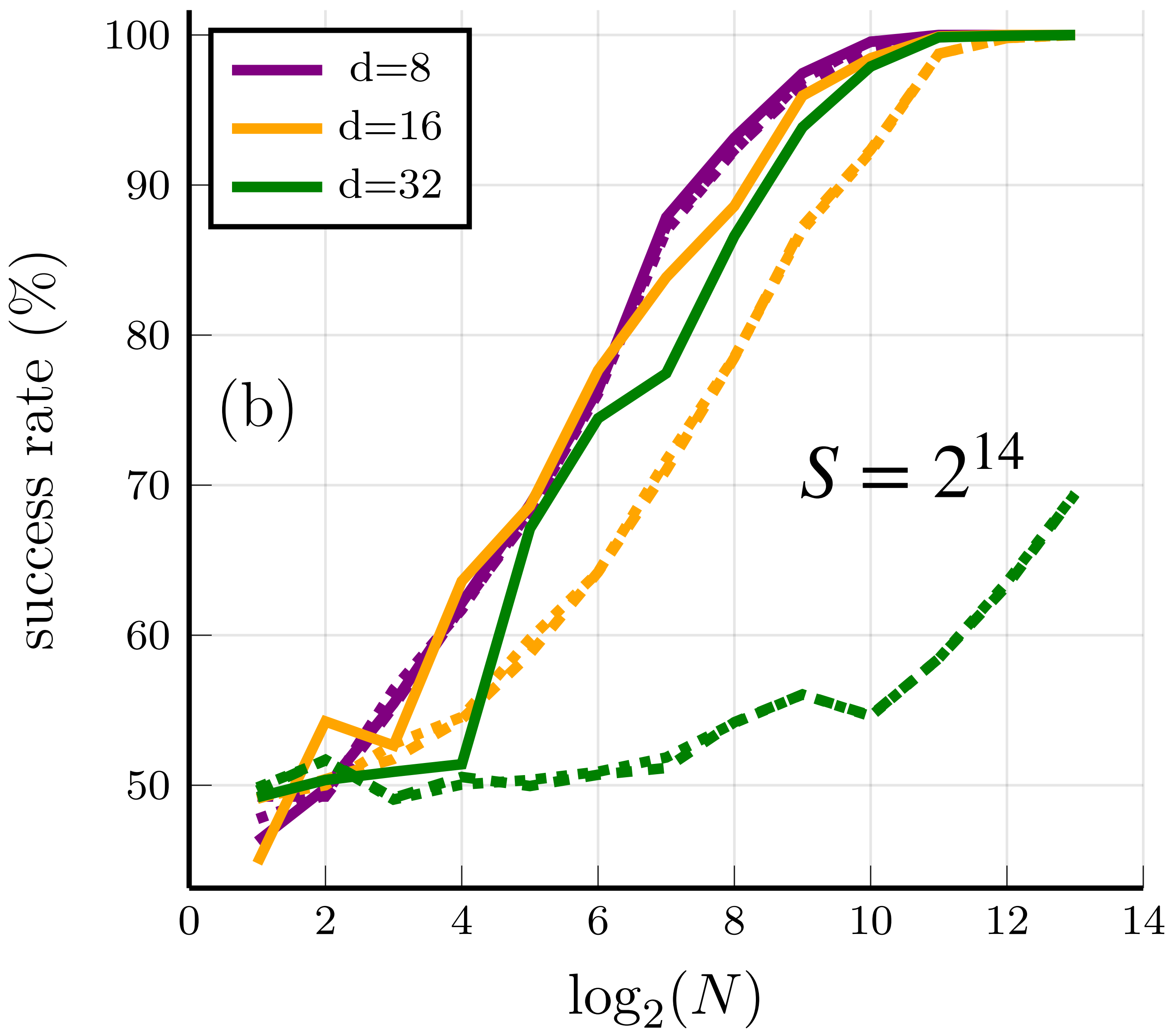


Error due to from few measurements may be the dominant source of generalisation error

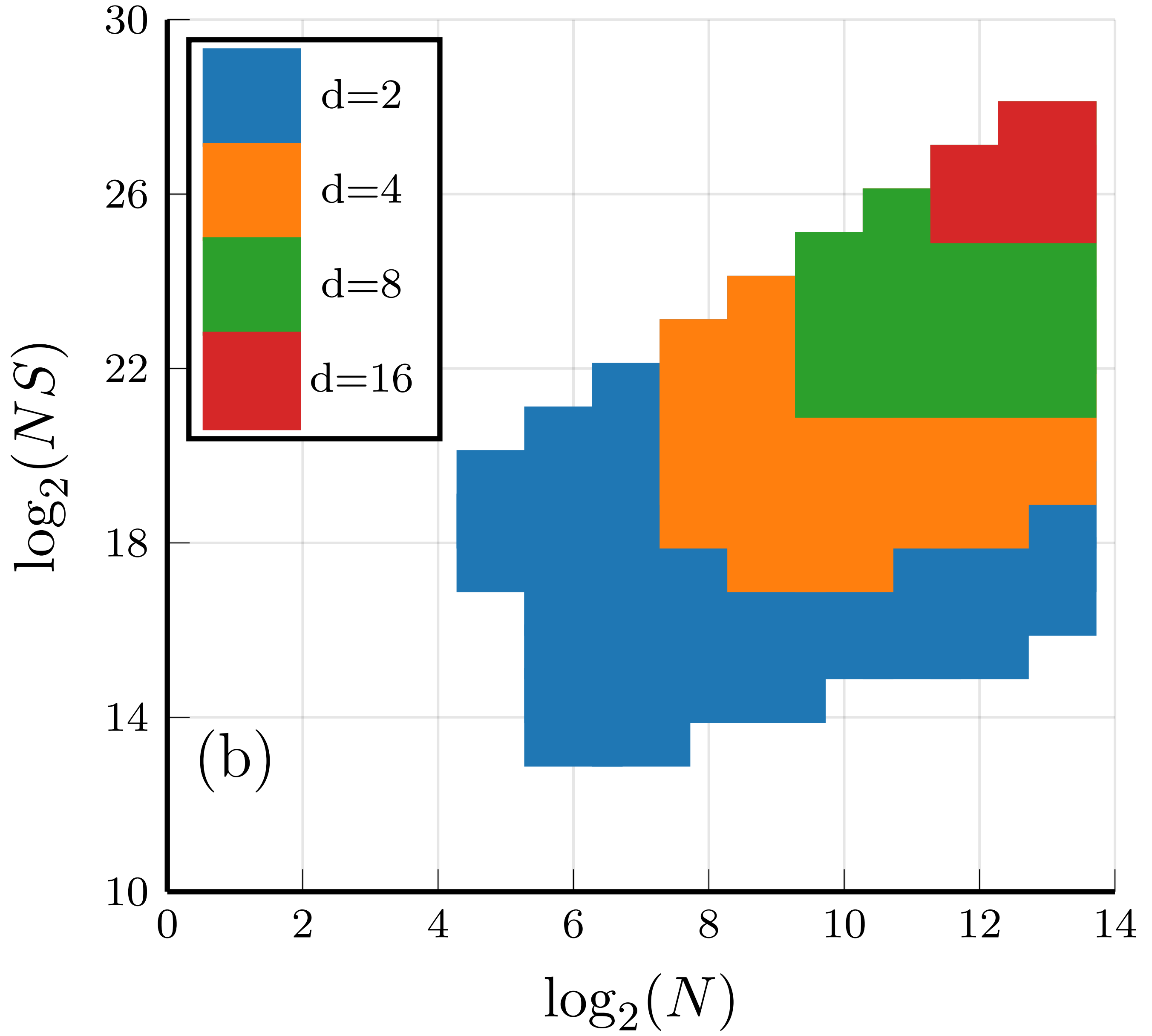


Shadow measurements

Swap test



In terms of overall copies, shadow fidelity estimator perform better than the swap test



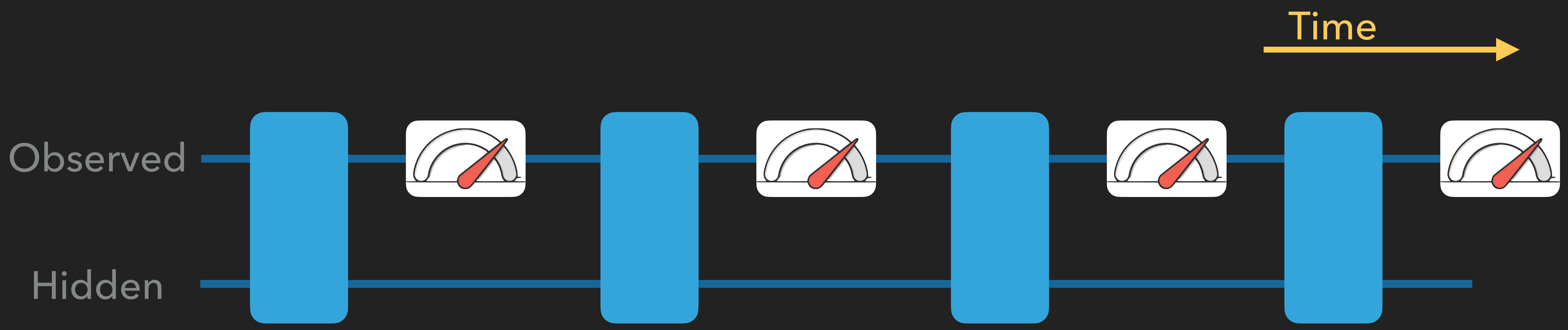
TEMPORAL PROCESS

TEMPORAL STOCHASTIC PROCESSES

- EUR vs 1 SGD
- What will happen in the future?

?

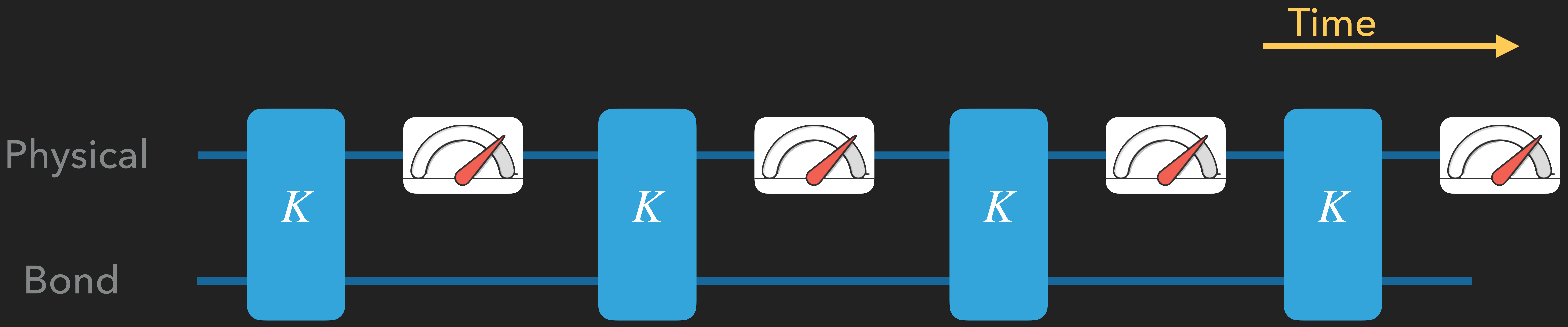
TEMPORAL STOCHASTIC PROCESSES



Not all hidden variables are relevant to predict future observations

Replace the huge hidden states with a finite memory, which contains all the relevant information to predict future outcomes

QUANTUM SIMULATORS



K are Kraus operators

Expressive power of tensor-network factorizations for probabilistic modeling

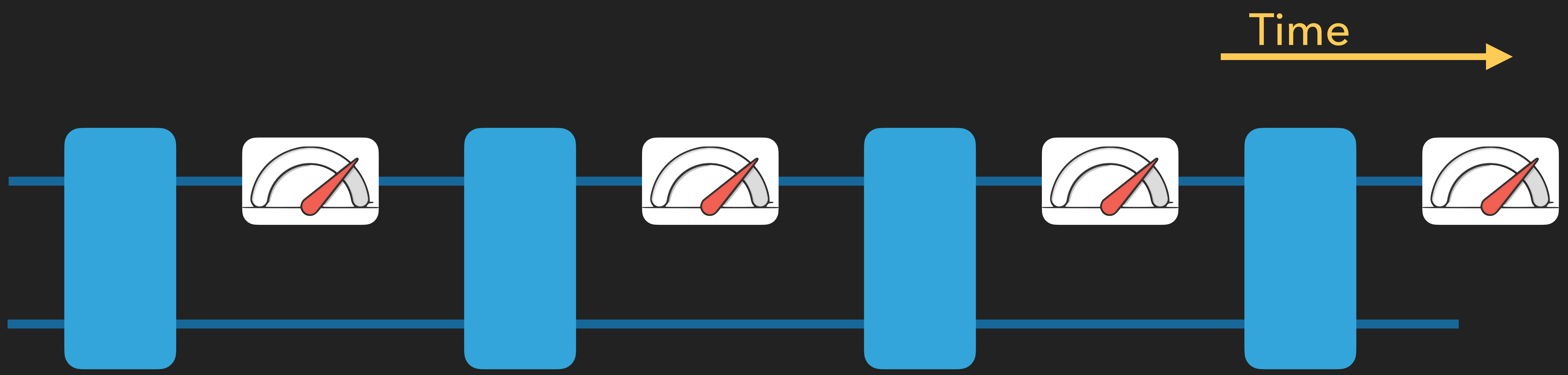
Ivan Glasser[1,2]*, Ryan Sweke[3], Nicola Pancotti[1,2], Jens Eisert[3,4], J. Ignacio Cirac[1,2]
[1]Max-Planck-Institut für Quantenoptik, D-85748 Garching
[2]Munich Center for Quantum Science and Technology (MCQST), D-80799 München
[3]Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, D-14195 Berlin
[4]Department of Mathematics and Computer Science, Freie Universität Berlin, D-14195 Berlin

PHYSICAL REVIEW LETTERS **121,** 260602 (2018)

Matrix Product States for Quantum Stochastic Modeling

Chengran Yang,[1,2,*] Felix C. Binder,[1,2,†] Varun Narasimhachar,[1,2] and Mile Gu[1,2,3,‡]
[1]*School of Physical and Mathematical Sciences, Nanyang Technological University, 637371 Singapore, Singapore*
[2]*Complexity institute, Nanyang Technological University, 639798 Singapore, Singapore*
[3]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543 Singapore, Singapore*

QUANTUM SIMULATORS

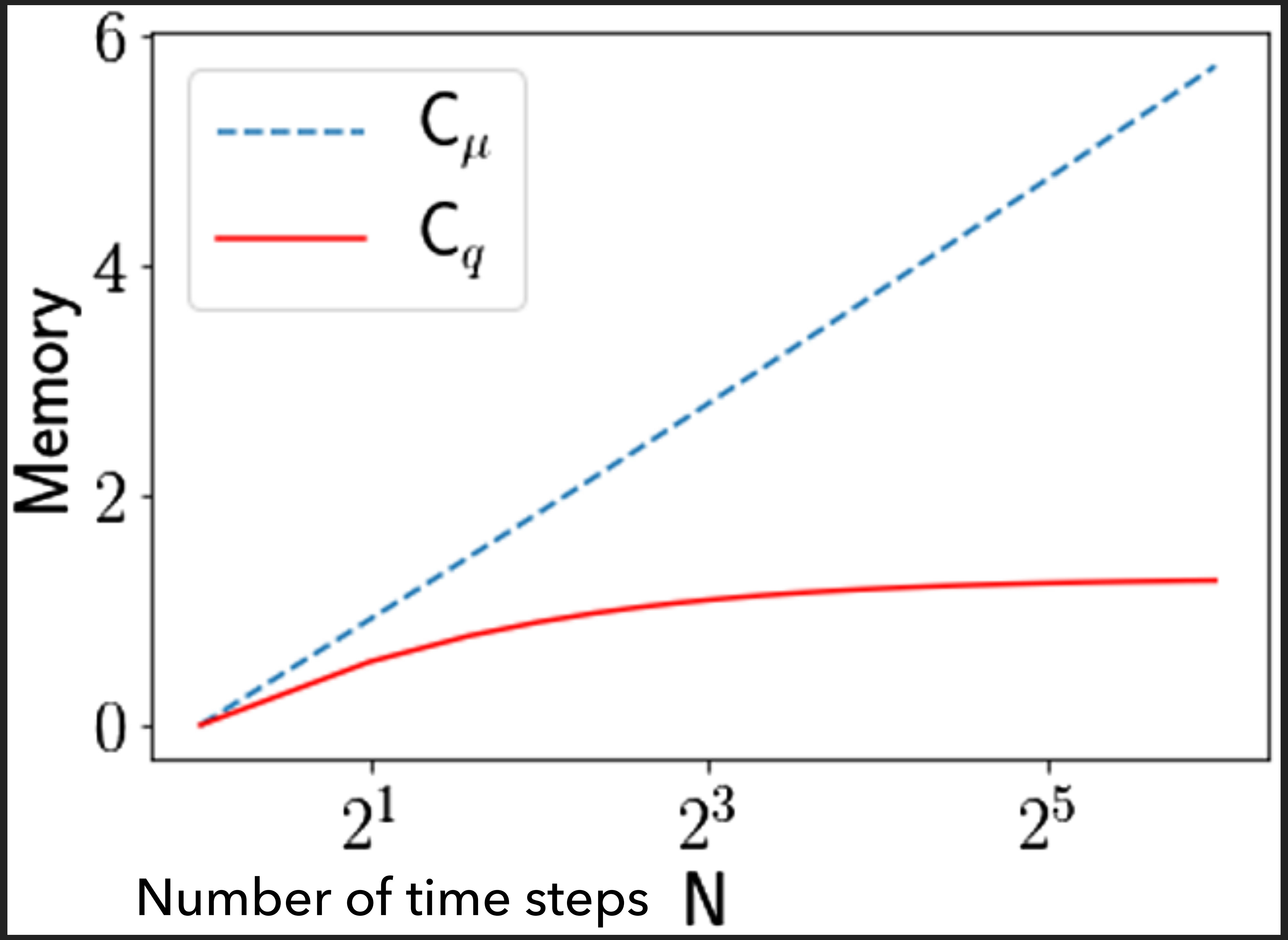


In general

Quantum: $P_K(x_1, \ldots, x_L) = \| K^{x_L} \ldots K^{x_1} | \phi \rangle \|^2$

Classical: $P_T(x_1, \ldots, x_L) = \sum_{\alpha} (T^{x_L} \cdots T^{x_1} p^{\mathrm{i}})_{\alpha}$

In some cases (ε-machines) the quantum model uses less memory



LEARNING FRAMEWORK

$$x \longrightarrow \rho(x, y) \longrightarrow y$$

Mach. Learn.: Sci. Technol. **5** (2024) 025036 https://doi.org/10.1088/2632-2153/ad444a

MACHINE LEARNING Science and Technology

IOP Publishing

OPEN ACCESS

PAPER

Accuracy vs memory advantage in the quantum simulation of stochastic processes

Leonardo Banchi

Department of Physics and Astronomy, University of Florence, via G. Sansone 1, I-50019 Sesto Fiorentino, (FI), Italy
INFN Sezione di Firenze, via G. Sansone 1, I-50019 Sesto Fiorentino, (FI), Italy

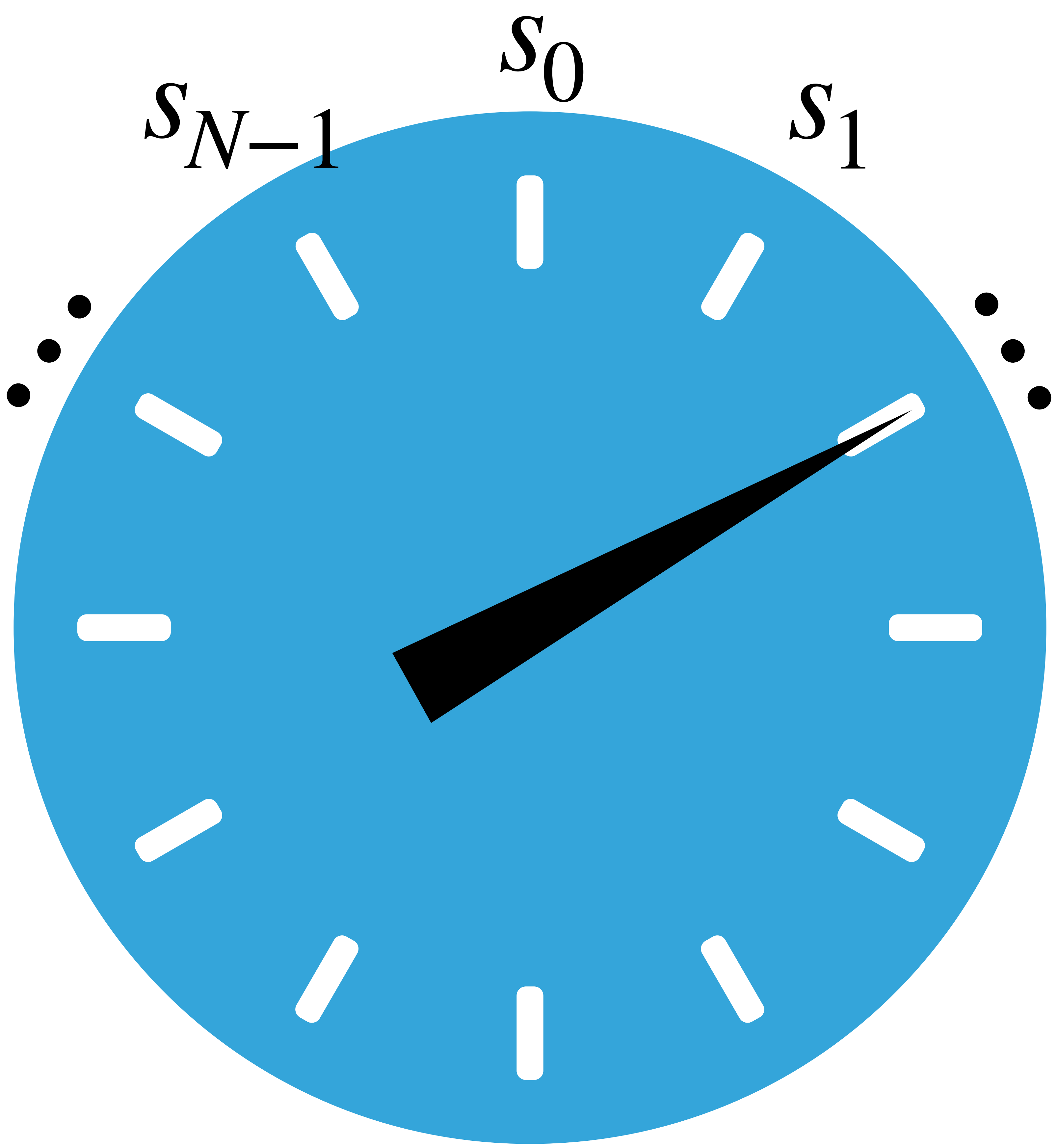
E-mail: leonardo.banchi@unifi.it

- x: vector of past observations
- y: vector of future observations

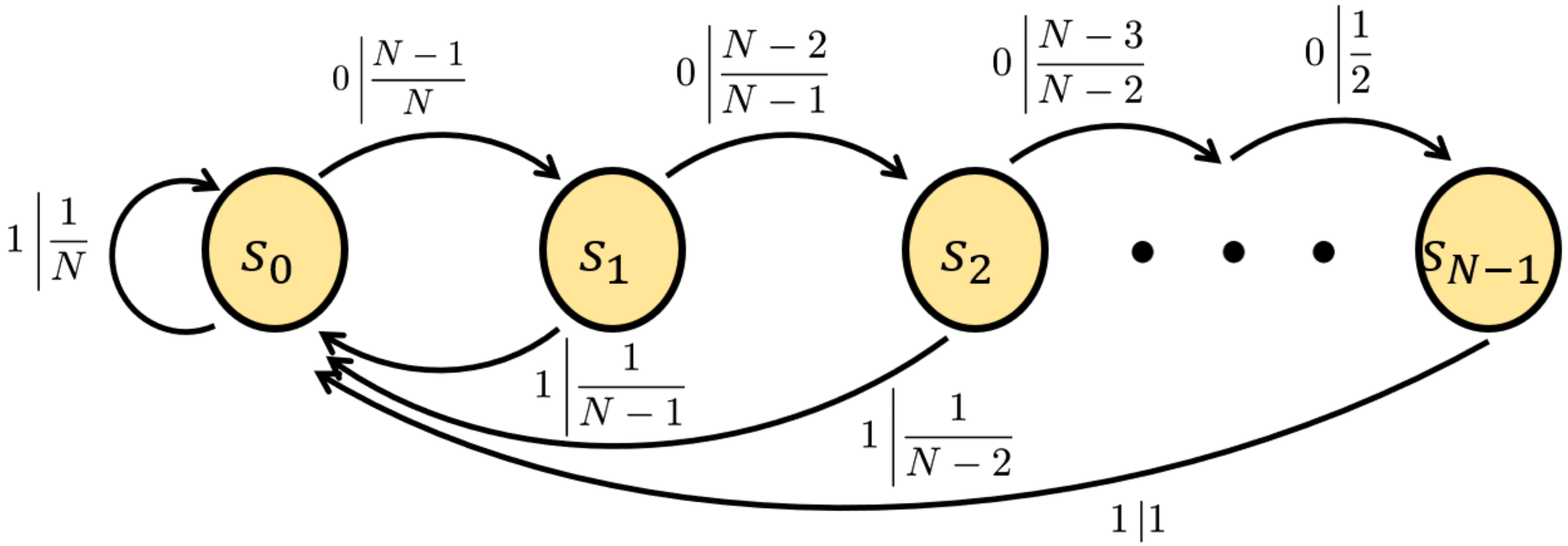
- The quantum model better compresses past observations into a quantum memory $$I(X : Q_{\text{quantum memory}}) \leq I(X : Q_{\text{classical memory}})$$
- Does the quantum model display an advantage in generalisation?

EXAMPLE: STOCHASTIC CLOCK

Yang et al. PRL 121, 260602 (2018)



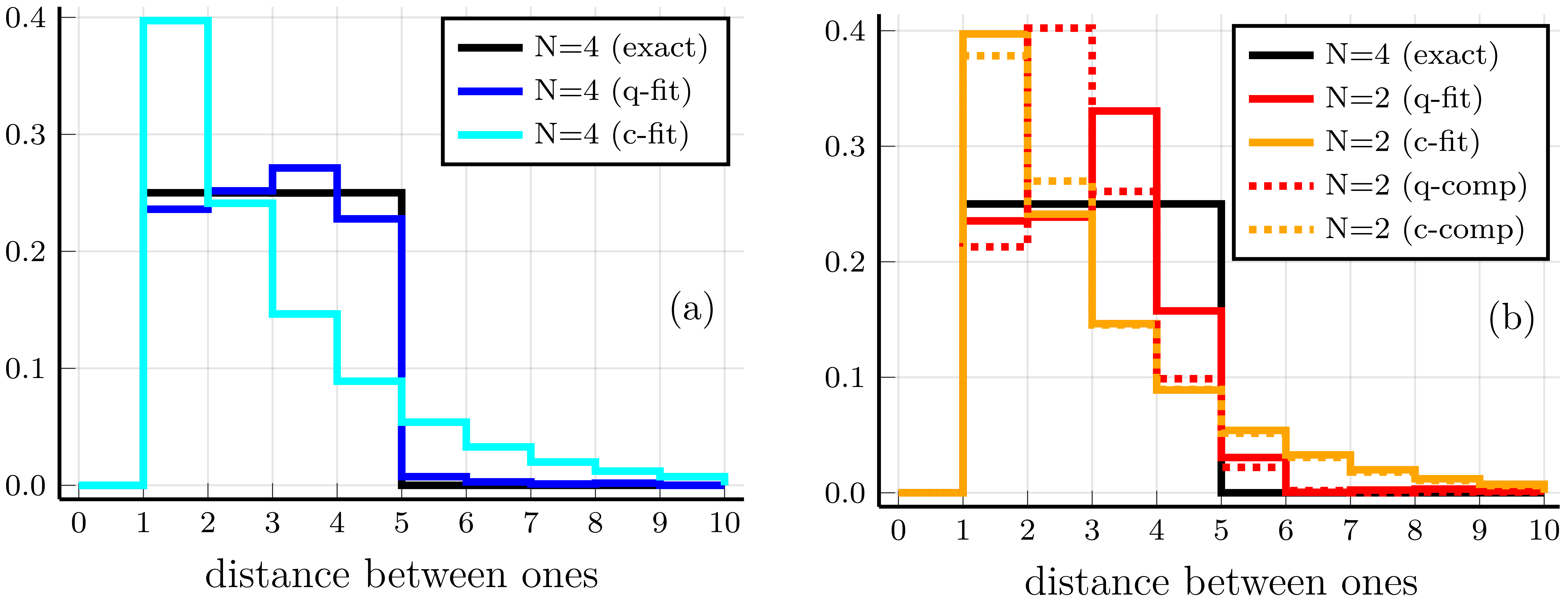
Increase clock and emit 0



Reset and emit 1

The distance between two ticks is uniformly distributed in $[0, N-1]$

EXAMPLE: STOCHASTIC CLOCK



EXAMPLE: REAL-WORLD DATA (LYMPHOGRAPHY)

Probability of observing the **last ten** outcomes in the Lymphography dataset, given the previous observations

Each entry shows the best value over 100 repetitions of the fitting procedure for random initial configurations, and the median in parenthesis.

D	Quantum	Classical
16	0.037 (0.025)	0.023 (0.0039)
32	0.134 (0.098)	0.051 (0.0002)
64	0.330 (0.261)	0.002 (0.0002)

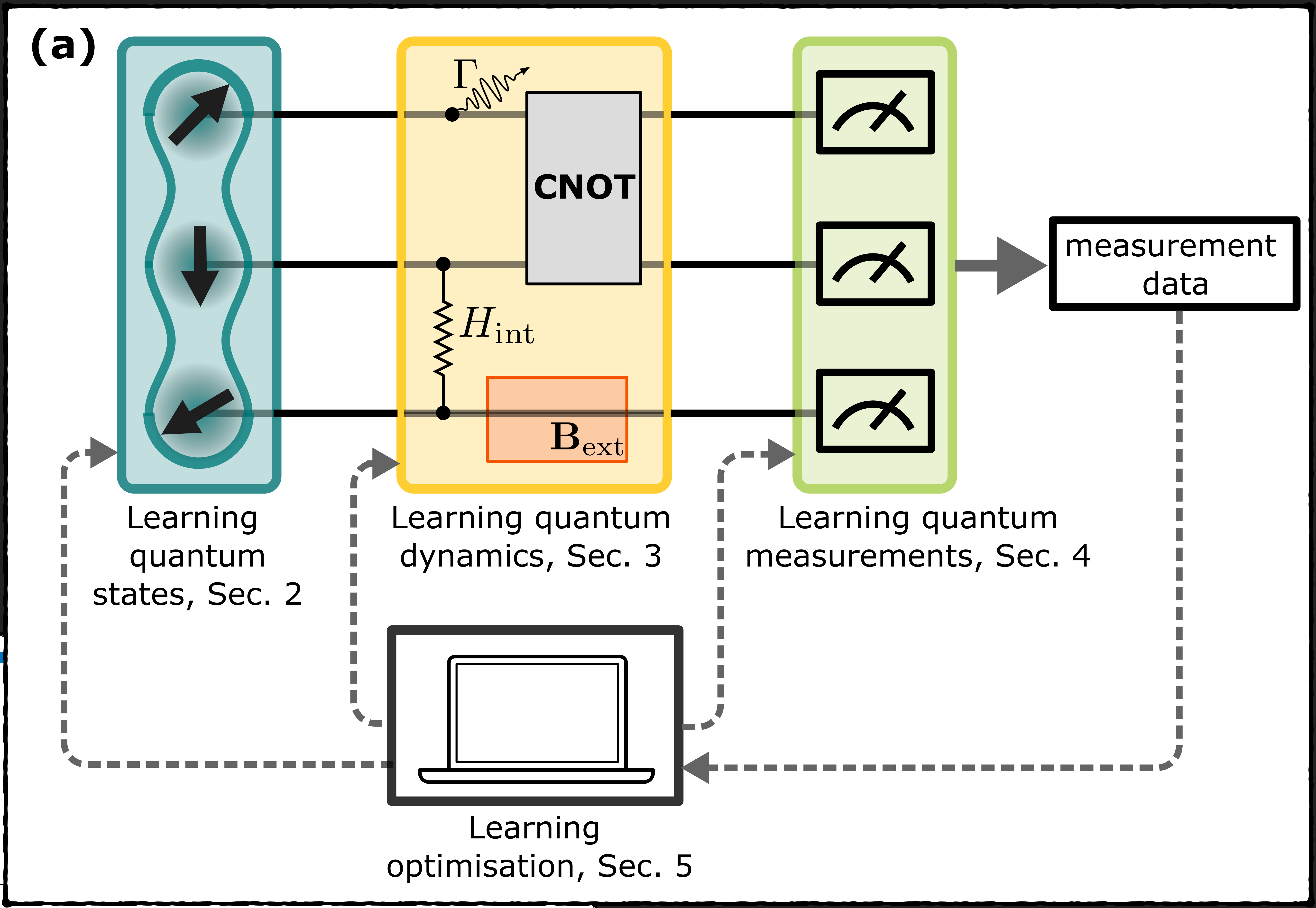
```
X = lymphography["data"]["features"]
```

	lymphatics	block of affere	bl. of lymph. c	bl. of lymph. s	by pass	extravasates
0	4	2	1	1	1	1
1	3	2	1	1	2	2
2	3	2	2	2	2	2
3	3	1	1	1	1	2
4	3	1	1	1	1	1
...	...	...	...	...	...	...
143	3	2	1	1	2	2
144	2	1	1	1	1	1
145	2	2	1	1	1	2
146	2	1	1	1	1	1
147	2	2	2	1	2	2

148 rows × 19 columns

OTHER APPLICATIONS

LEARNING WITH QUANTUM DATA



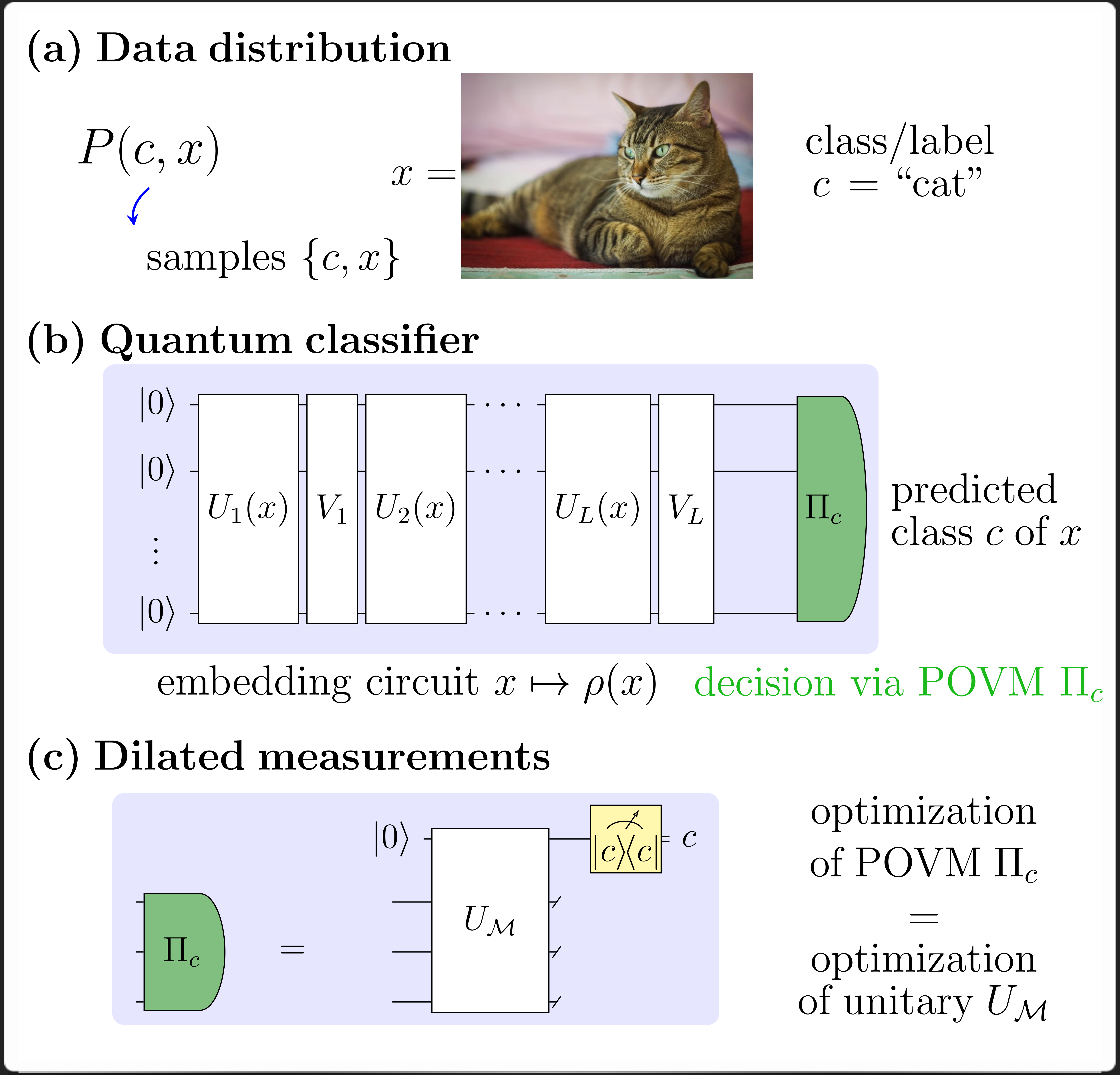
nature reviews physics

Nature Reviews Physics **5**, 141–156 (2023)

Learning quantum systems

Valentin Gebhart[1,2], Raffaele Santagati[3], Antonio Andrea Gentile[4], Erik M. Gauger[5], David Craig[6], Natalia Ares[7], Leonardo Banchi[8,9], Florian Marquardt[10,11], Luca Pezzè[1,2] & Cristian Bonato[5]

EXAMPLE: PARAMETRIC QUANTUM CIRCUIT EMBEDDING



The embedded input can be expressed as a Fourier-like series

$$|\psi(x)\rangle = \frac{1}{\sqrt{|\Omega|}} \sum_{\omega \in \Omega} e^{i\omega x} |\phi_\omega\rangle$$

Generalisation error from the Fourier matrix

$$F_{\omega\omega'} = \frac{1}{N\Omega} \sum_{n=1}^{N} e^{i x_n (\omega - \omega')}$$

with spectrum f

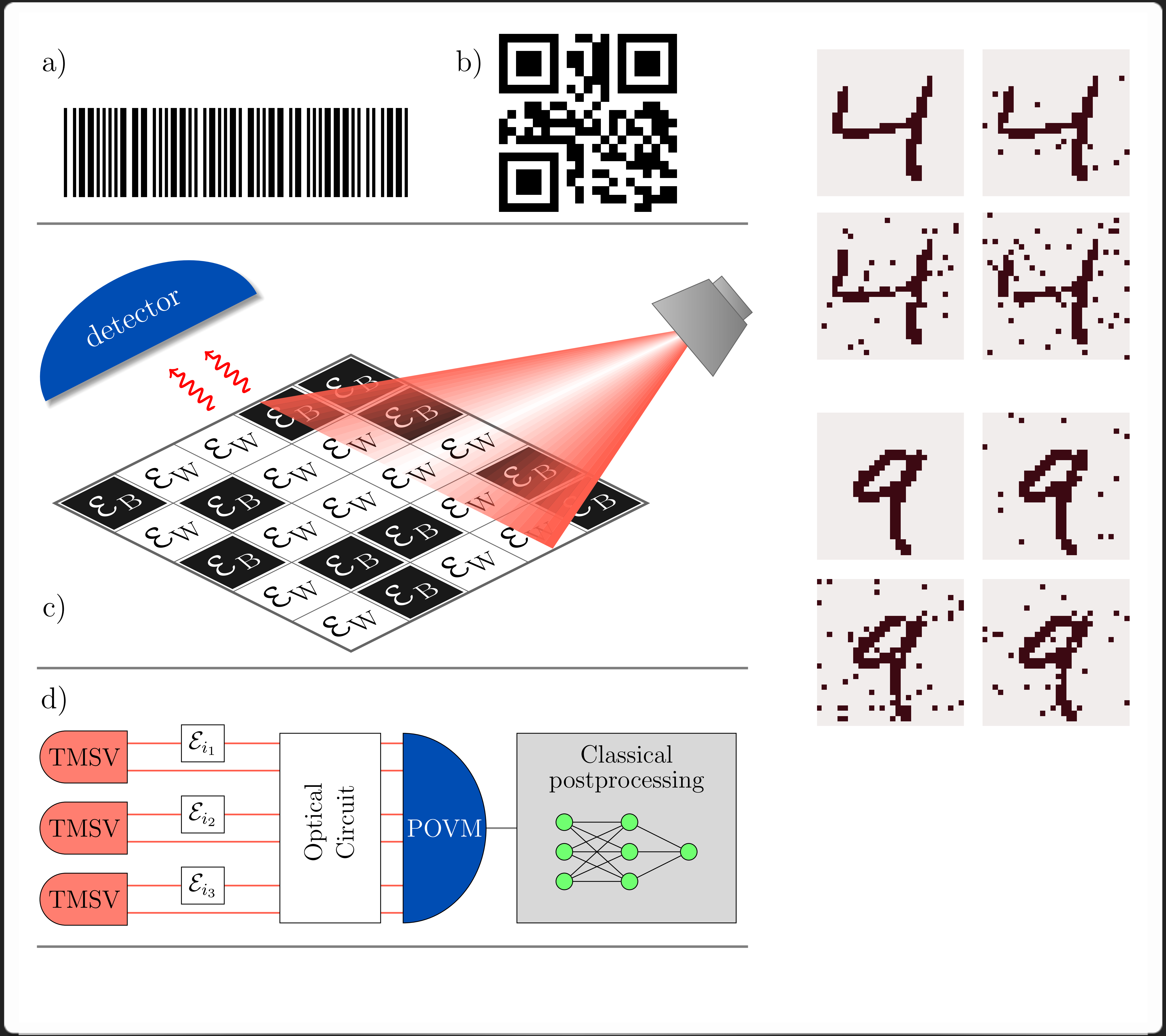
$$2^{I_{1/2}(X:Q)} \leq 2^{H_{1/2}(f)} \leq |\Omega|$$

See also Caro et al. Quantum (2021)

QUANTUM BARCODES AND PATTERN RECOGNITION

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

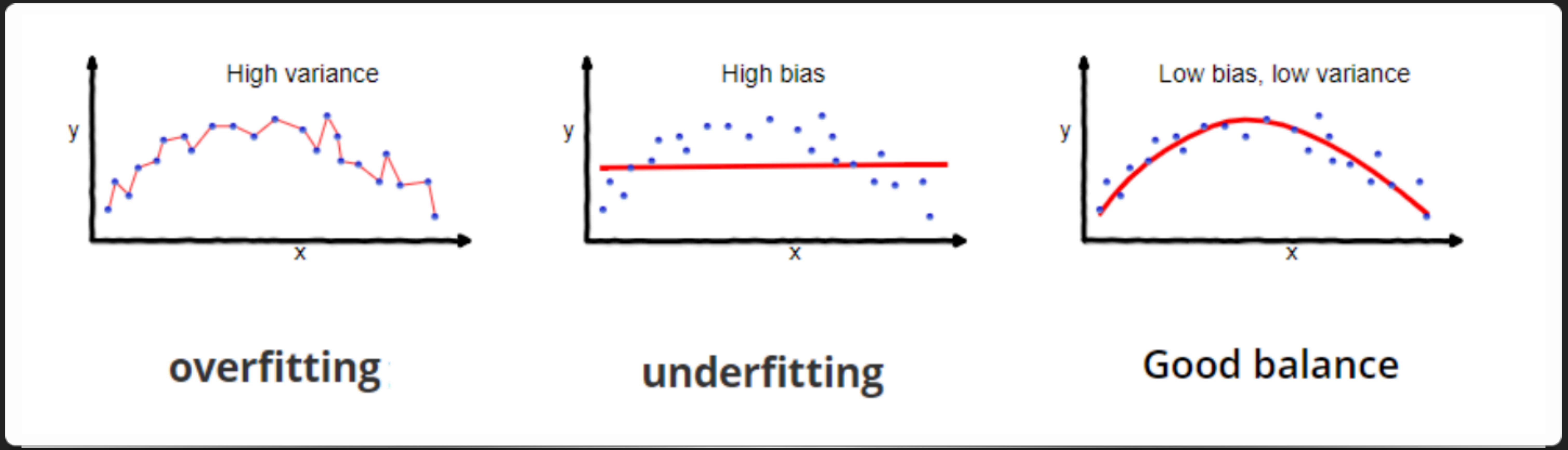
- L. Banchi, Q. Zhuang, S. Pirandola, Phys. Rev. Applied 14, 064026 (2020)
- C Harney, L Banchi, S Pirandola, Phys. Rev. A 103, 052406 (2021)
- JL Pereira, L Banchi, Q Zhuang, S Pirandola, Phys. Rev. A 103, 042614 (2021)

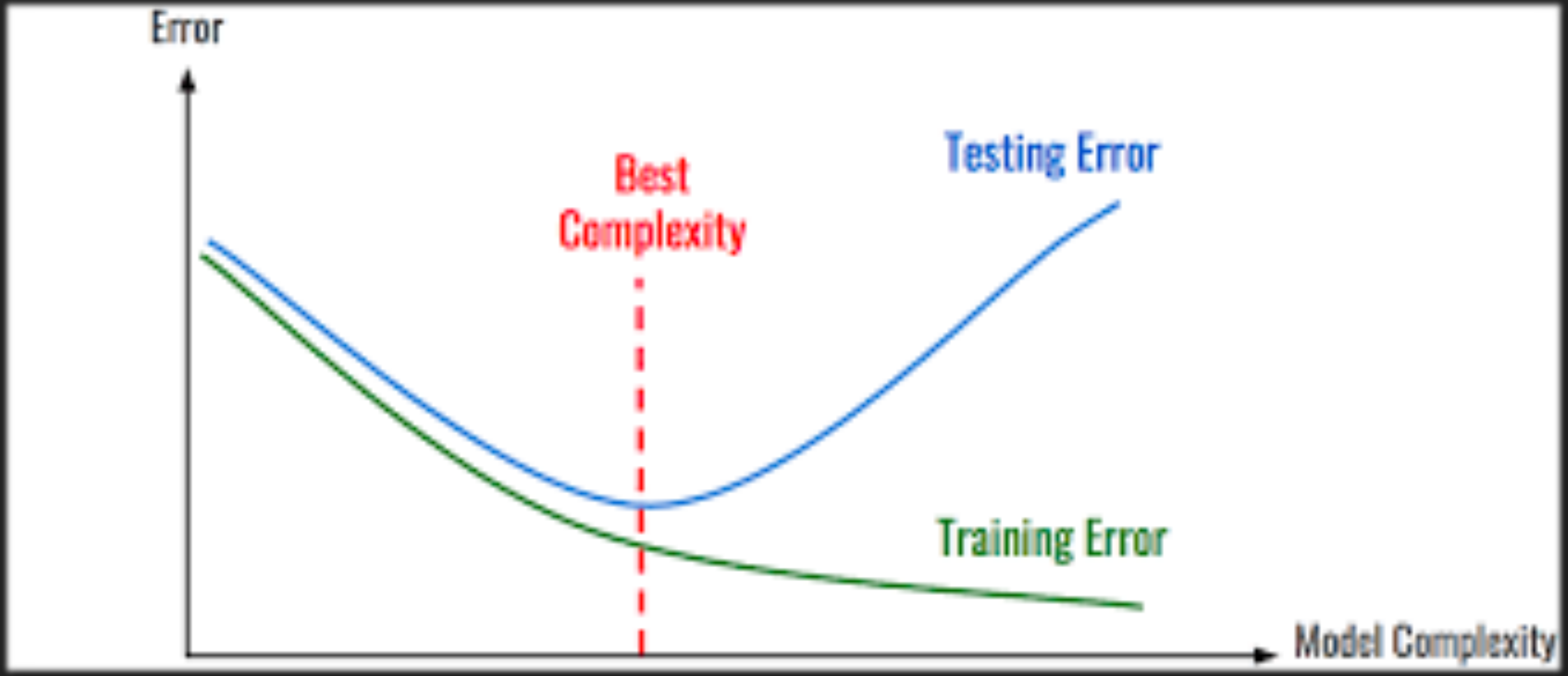


CONCLUSION

- Generalisation favoured by **discarding information** about the input that is irrelevant for predicting the output Quantum compression?
- **Few measurement shots** may be a dominant source of generalisation error
- Quantitative data-dependent bounds. Greater physical insight.
- Different applications:
  - Quantum pattern recognition with entanglement-enhanced quantum sensor
  - Classification of quantum states and phases of matter
  - Quantum machine learning of classical data

STATISTICAL LEARNING THEORY





?