MATTEO ROSATI - ROMA TRE UNIVERSITY, ITALY

Learning at the Interface Between Quantum Hardware and AI:

From Data-Driven Quantum Design to Quantum-Native Models

Outline

ML-ENHANCED
DESIGN AND CONTROL
OF QUANTUM DEVICES

LEARNING THEORY & ALGORITHMS
FOR QUANTUM SYSTEMS

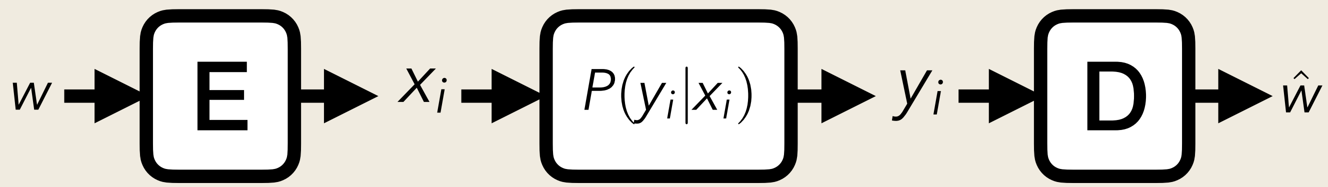
QUANTUM/AI
INTERFACE
TECHNOLOGIES

QUANTUM-ENHANCED
AI PRIMITIVES

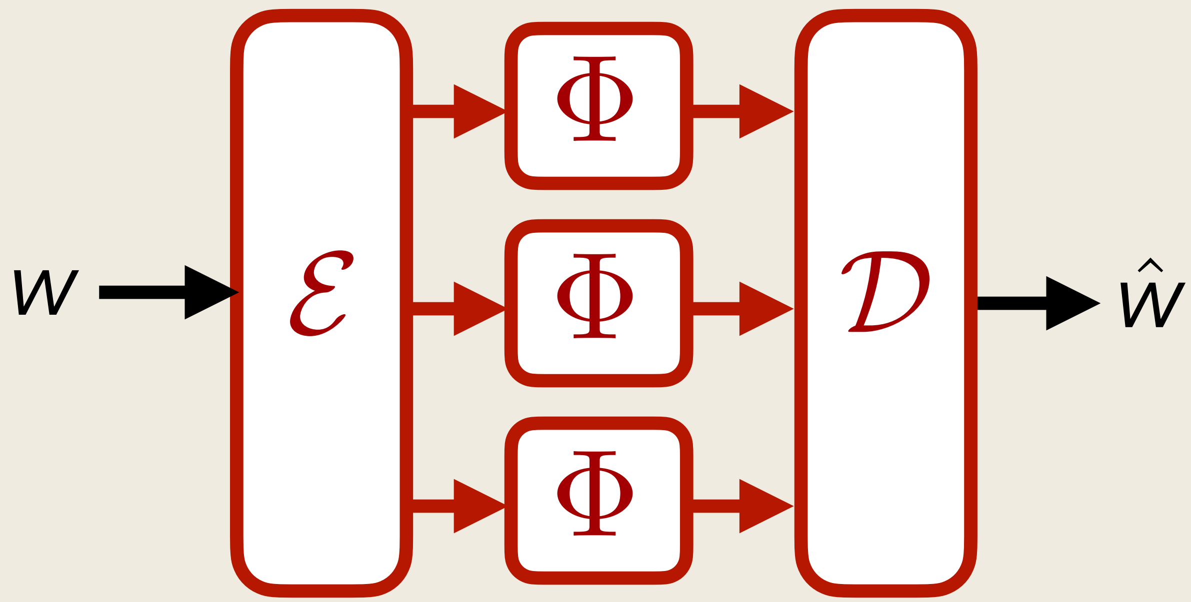
ML-enhanced control and design of quantum devices

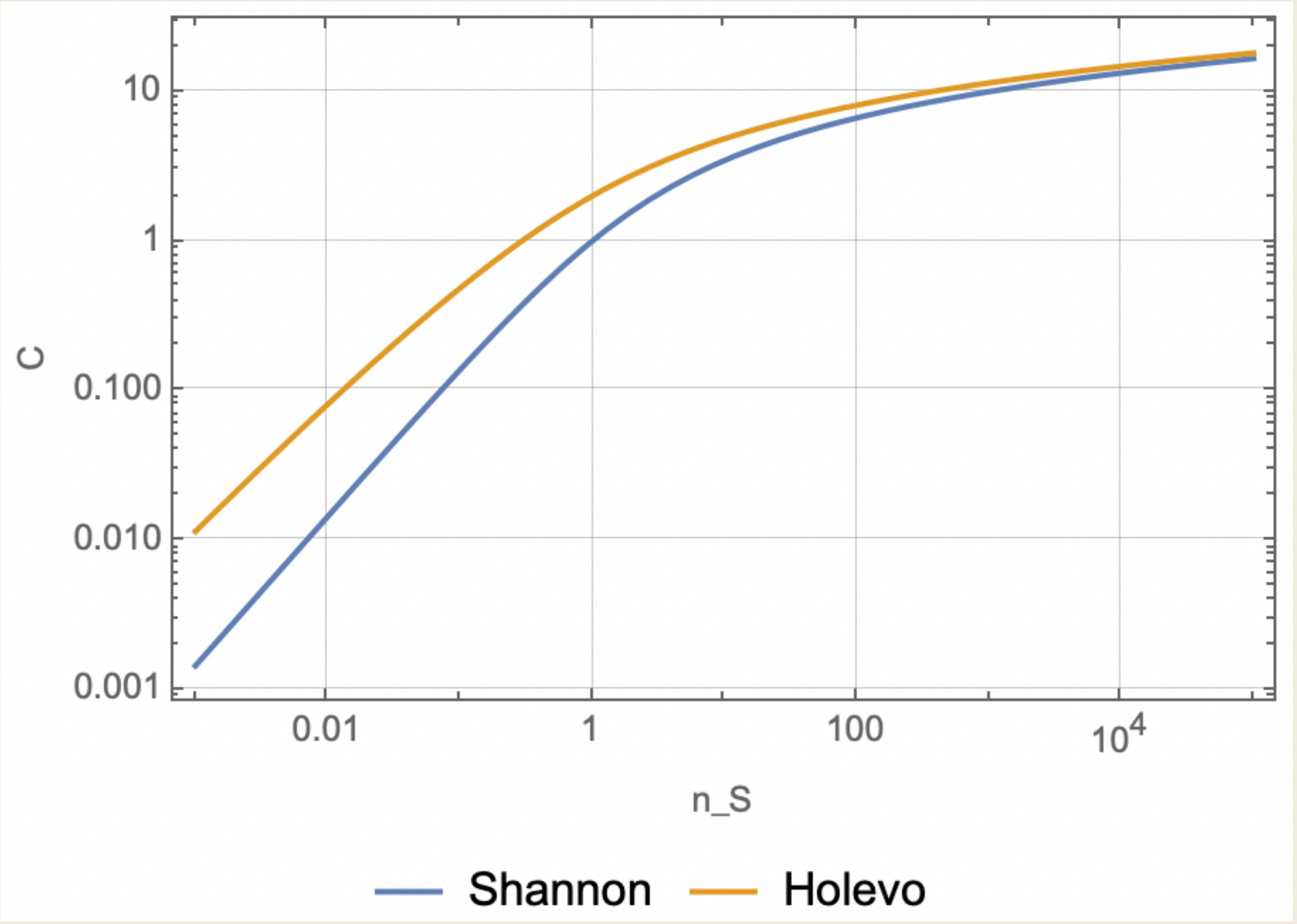
Bit transmission using quantum states

Classical Shannon system



Quantum Holevo system



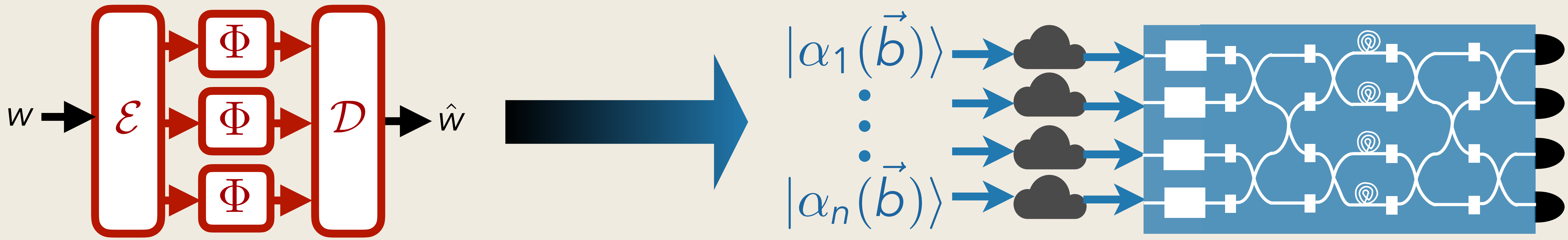


Join detection: use cases

- Short-distance: constant advantage
- Long-distance: log(1/S) advantage (no amplification)
- Entanglement-assisted: infinite-fold enhancement at large noise (microwave frequencies)
- Medium-distance: ~57% energy reduction at 200Km (with amplification)

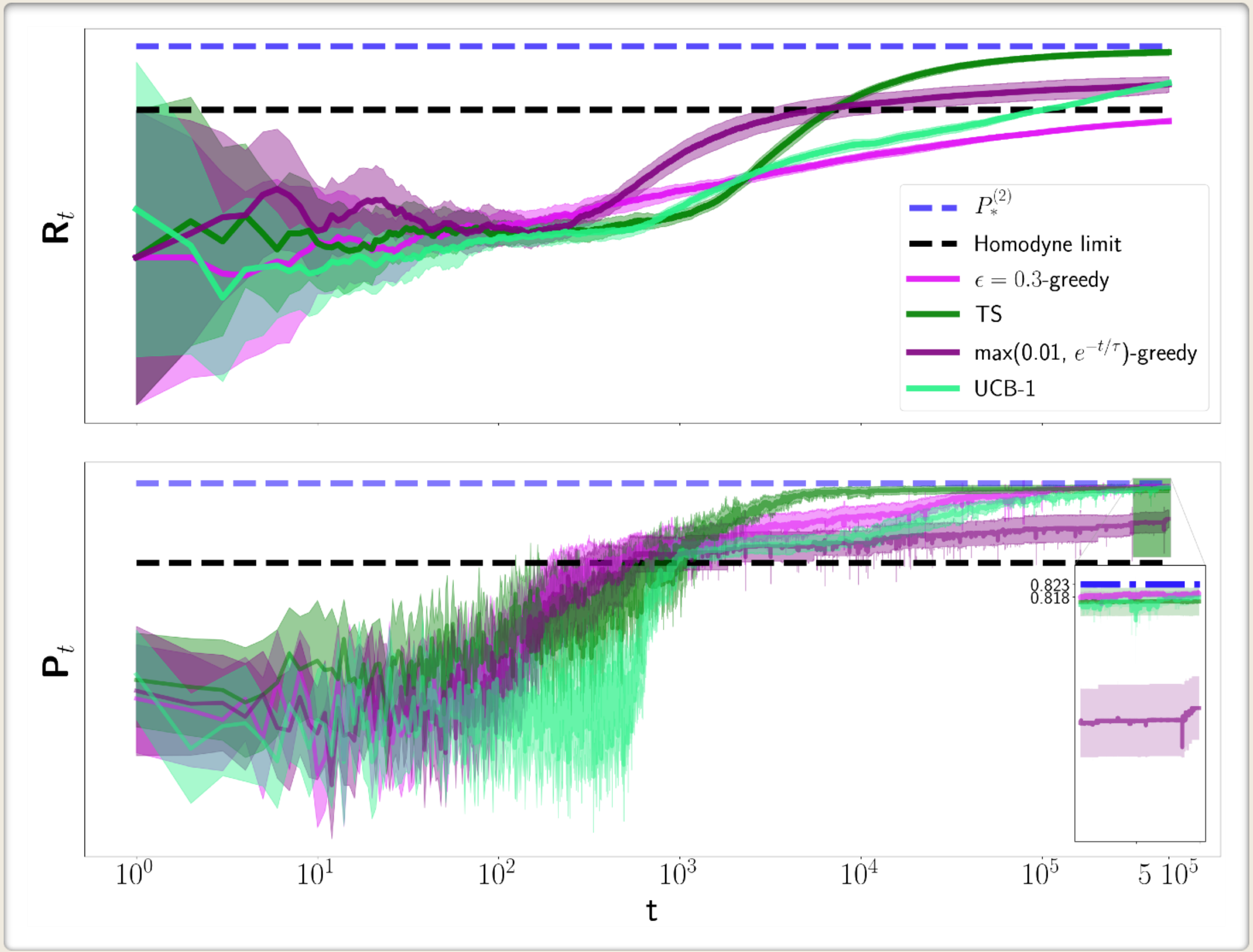
J. Nötzel and MR, *J. Light. Technol.*, 1 (2023); MR, *JOSA B* 42, 454 (2025);
S. Guha, Q. Zhuang, and B. A. Bash, *2020 IEEE ISIT*, 1835 (2020).

The quest for an optimal joint detection receiver (JDR)



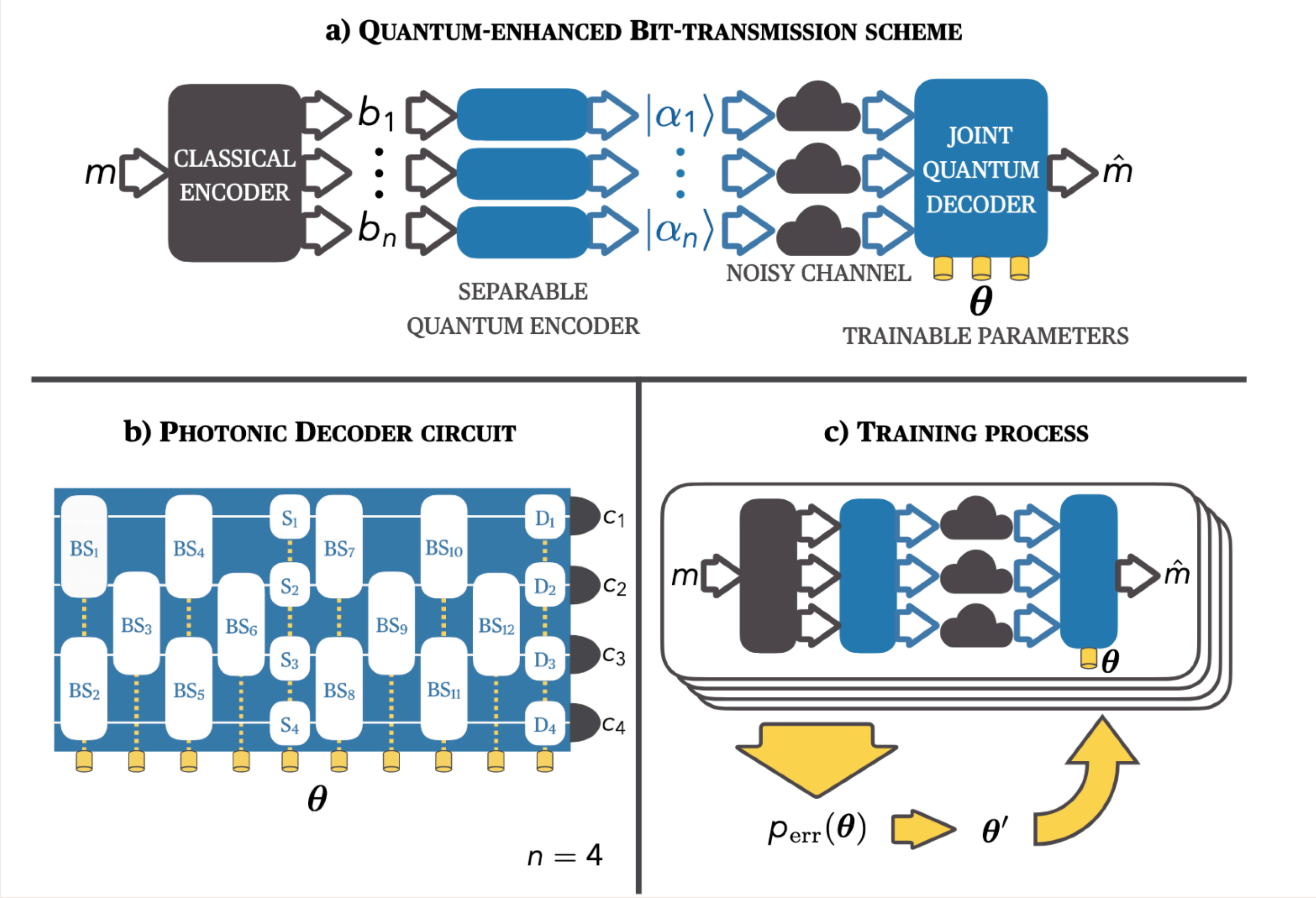
- Are there optical JDR designs for other codes, e.g., for q. advantage at higher signal power?
- Is there a systematic way of studying them?
- What is their performance and scalability?
- Can optical squeezing or ancillae help?

Reinforcement-learning a single-symbol receiver



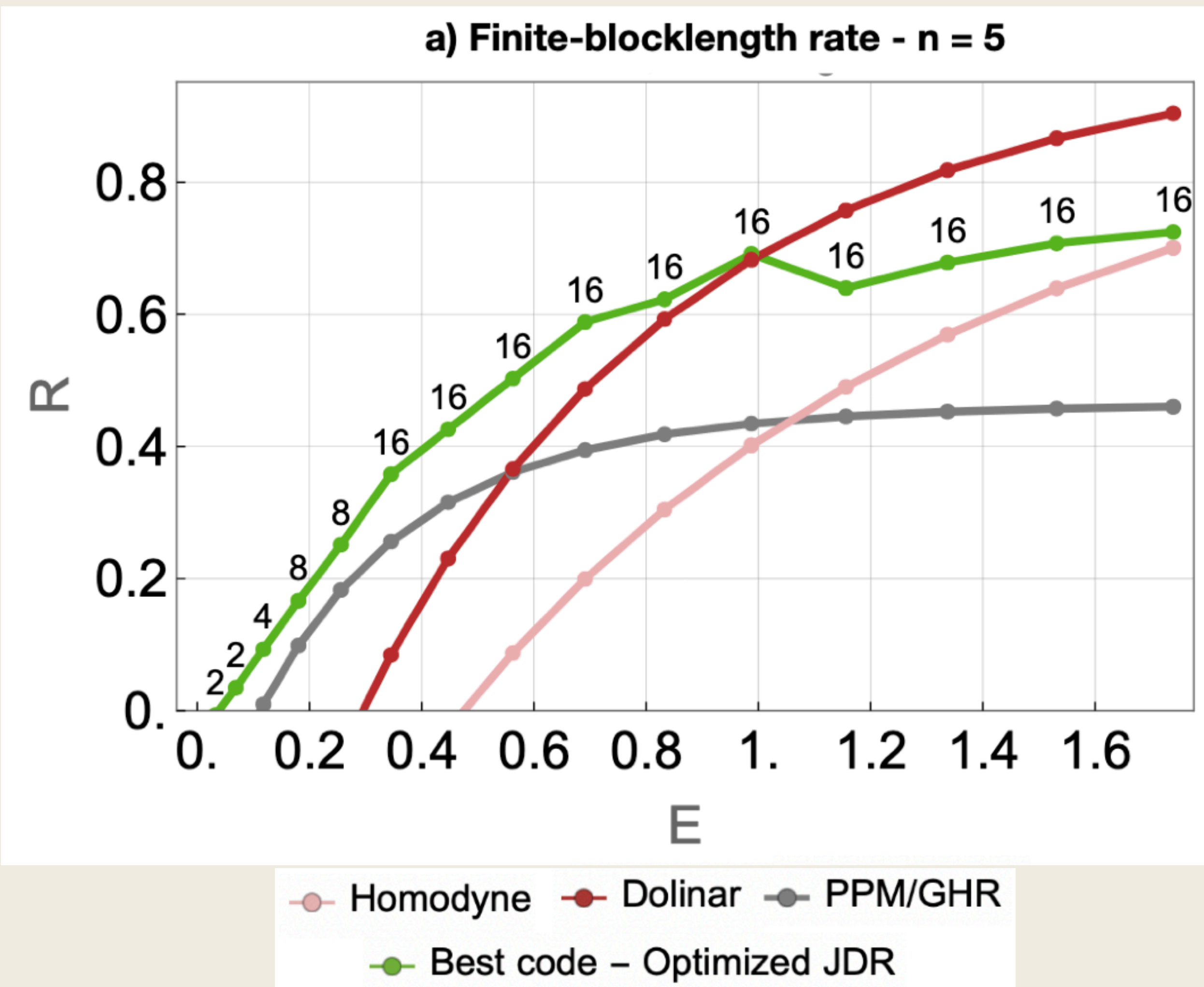
M. Bilkis, MR et al., Phys. Rev. Res. 2, 033295 (2020).

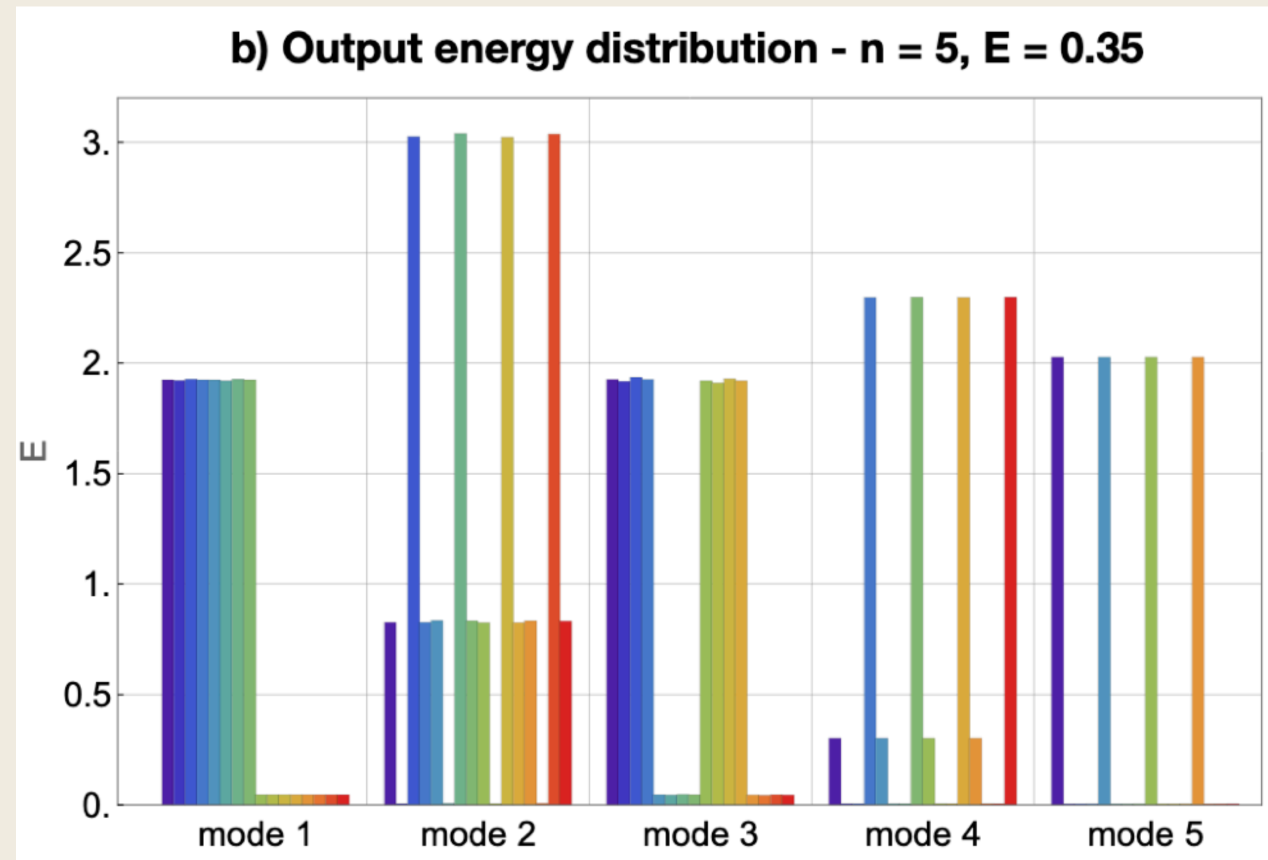
Optical JDR designs: a supervised learning approach



MR and A. Solana, *Opt. Quantum* 2, 390 (2024).

Results: linear codes

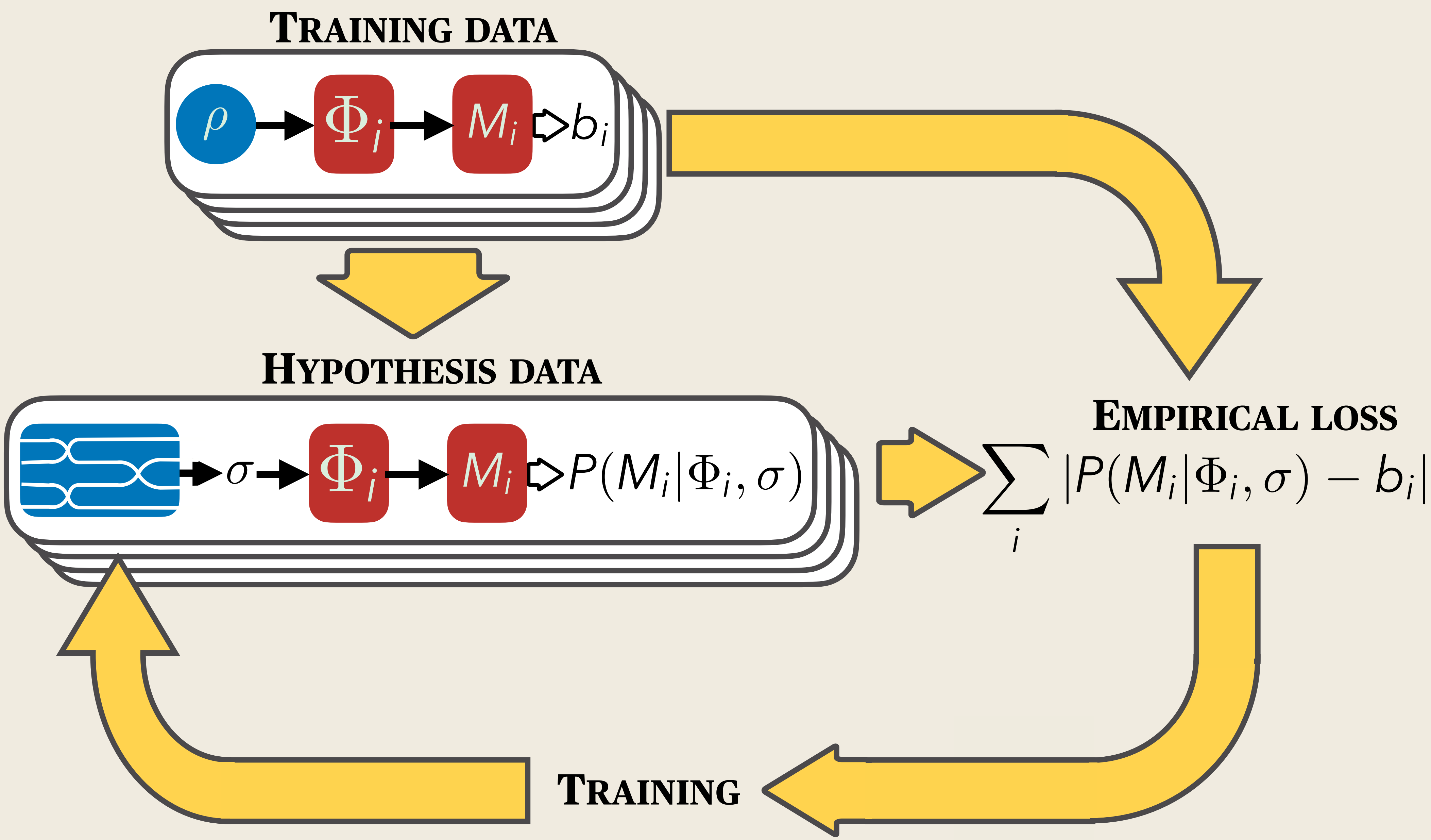




MR and A. Solana, *Opt. Quantum* 2, 390 (2024); MR, *New J. Phys.* 27, 013010 (2025).

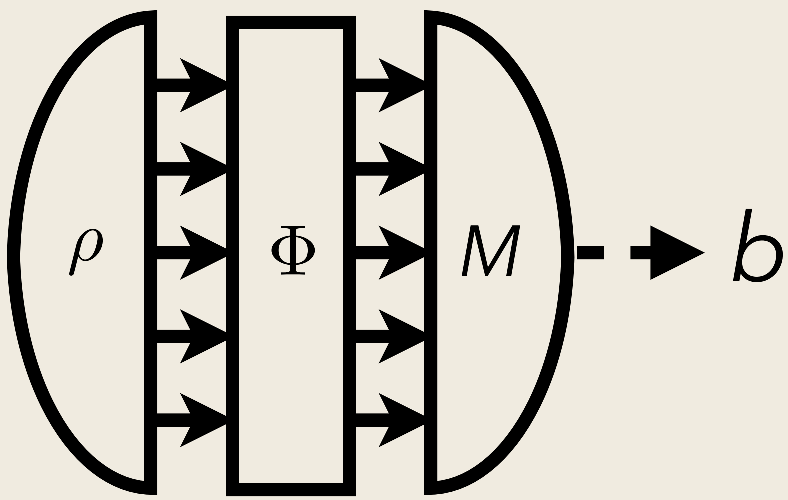
Learning theory & algorithms for quantum systems

Learning probability distributions generated by bosonic circuits



MR, *Quantum* 8, 1433 (2024); S. Aaronson, *Proc. R. Soc. A Math. Phys. Eng. Sci.* 463, 3089 (2006).

Gaussian and non-Gaussian circuits



$b = 1$ "accept"

with probability

$$P(M|\Phi, \rho) = \mathrm{Tr}[M \cdot \Phi(\rho)]$$

Gaussian

$$P(M|\Phi, \rho) = \mathcal{G}_{\mathbf{m}_{\mathrm{out}}, V_{\mathrm{out}}+V'}(\mathbf{m}')$$

Gaussian plus photodetection

$$P(M|\Phi, \rho) = \mathcal{G}_{\mathbf{m}_{\mathrm{out}}, V_{\mathrm{out}}+V'}(\mathbf{m}') \cdot H_{\mathbf{k}}^{(\tilde{V})}(\tilde{\mathbf{m}})$$

Non-Gaussian (complex sum of complex-valued Gaussian Wigner f.)

$$P(M|\Phi, \rho) = \sum_i c_i \mathcal{G}_{\mathbf{m}_{i,\mathrm{out}}, V_{i,\mathrm{out}}+V'_i}(\mathbf{m}'_i)$$

MR, *Quantum* 8, 1433 (2024); J. E. Bourassa *et al.*, *PRX Quantum* 2, 040315 (2021).

Results: sample complexity

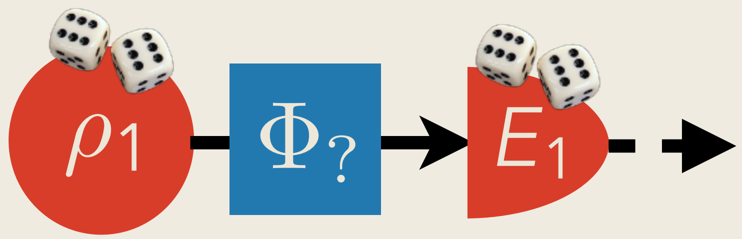
We obtain efficient learnability for k-mode circuits:

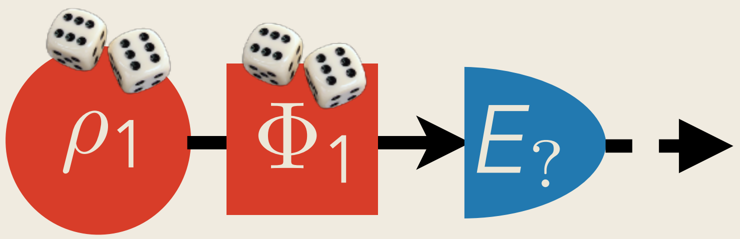
Gaussian $n = O(\epsilon^{-2} k^2)$

Gaussian plus photodetection $n = O(\epsilon^{-2} k^2 \log N)$

Non-Gaussian $n = O(\epsilon^{-4} k^2 B^2)$

Similar scaling: learning channels /measurements



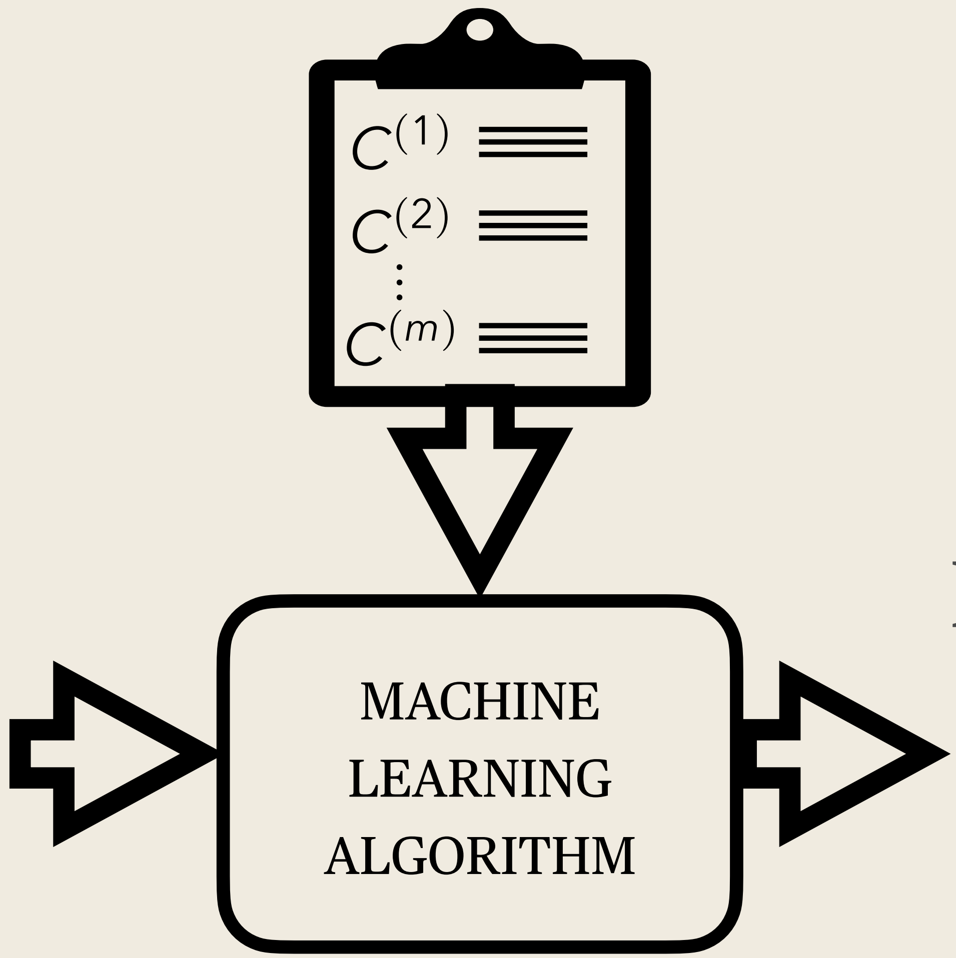


MR, *Quantum* 8, 1433 (2024).

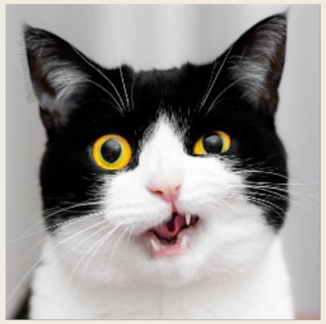
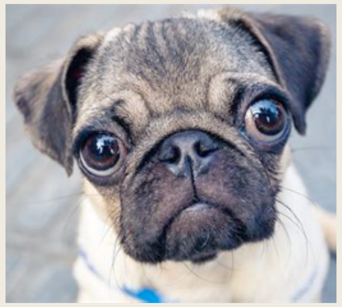
Classical Statistical Learning Theory

POSSIBLE EXPLANATIONS

theoretical model



OBSERVATIONS

classical "input" x	classical "label" y
	cat
	dog

Learning a deterministic function

$$y = f(x)$$

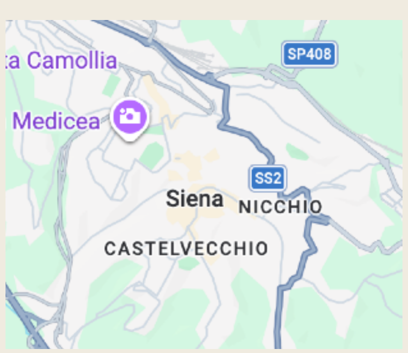
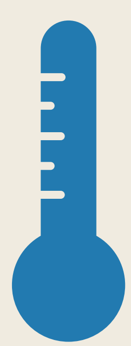
BEST EXPLANATION

$$C_{op}$$

M. Fanizza, Y. Quek, and MR, *PRX Quantum* 5, 020367 (2024).

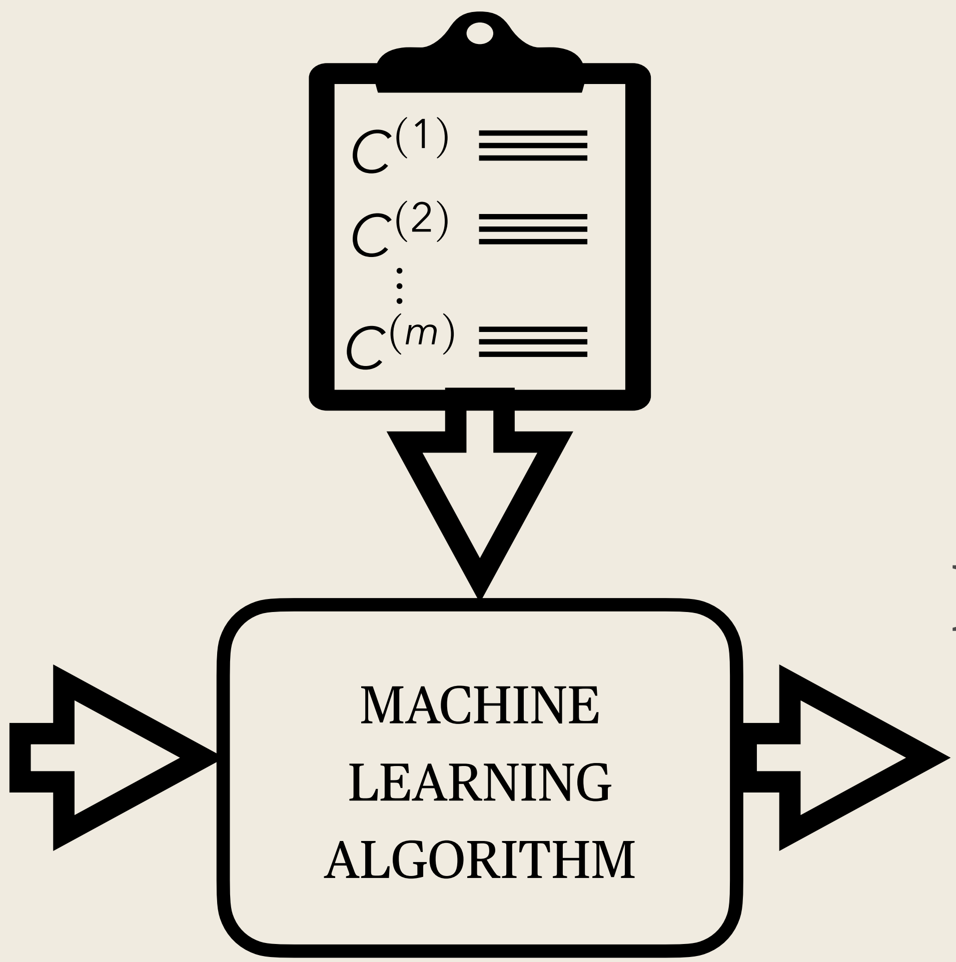
Classical Statistical Learning Theory

OBSERVATIONS

classical “context” x	classical “signal” y
	
	

POSSIBLE EXPLANATIONS

theoretical model



Learning a probabilistic relation

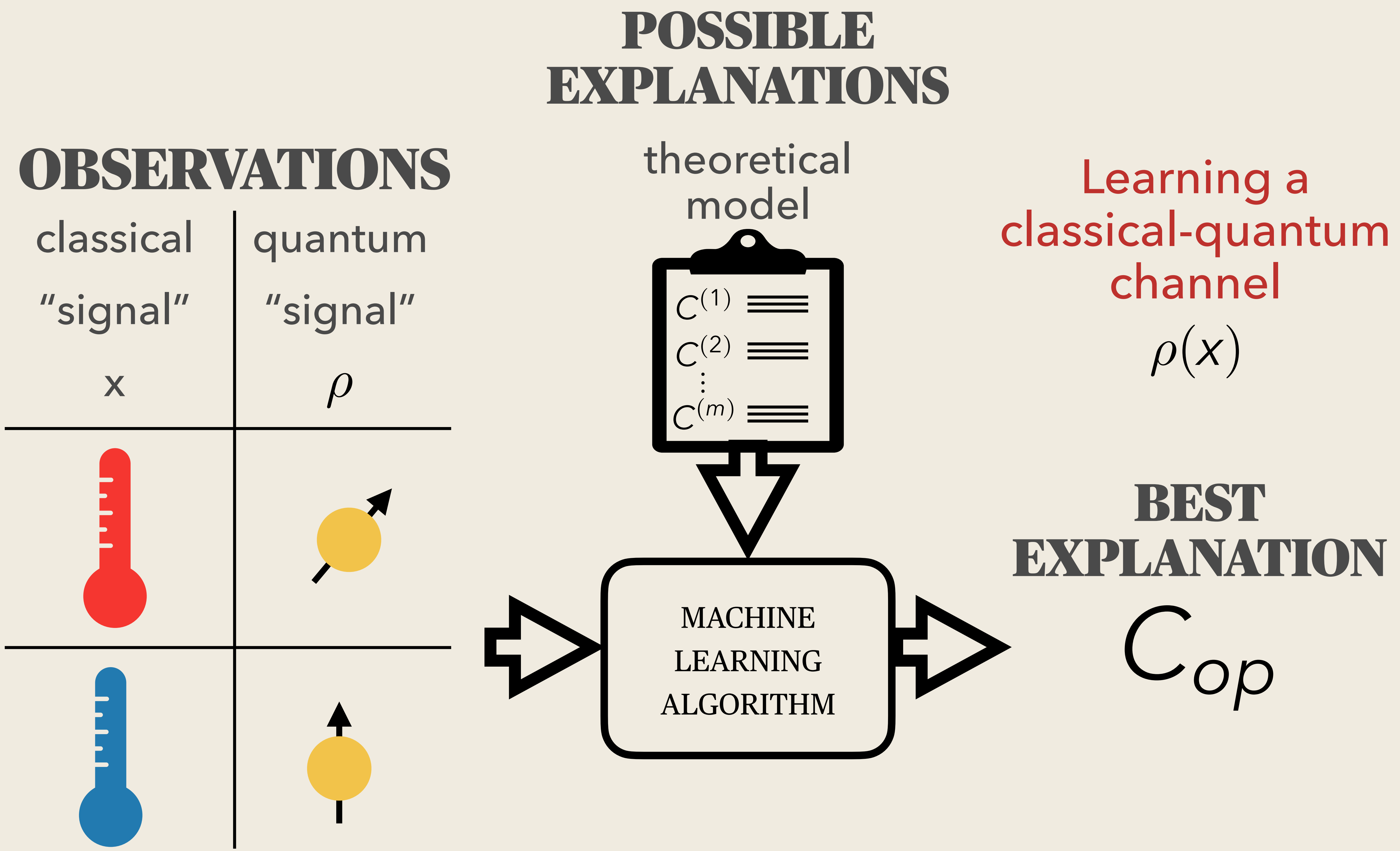
$p(y|x)$

BEST EXPLANATION

C_{op}

M. Fanizza, Y. Quek, and MR, *PRX Quantum* 5, 020367 (2024).

Quantum Statistical Learning Theory



M. Fanizza, Y. Quek, and MR, *PRX Quantum* 5, 020367 (2024).

Classical Statistical Learning Theory

- Vapnik: $f_? : \mathcal{X} \to \mathcal{Y}$ $\mathcal{C} = \{g : \mathcal{X} \to \mathcal{Y}\}$

  Target Hypothesis class

given samples $(x_i, f_?(x_i))$ $i = 1, \cdots, n$ $x_i \sim \mathcal{D}$

find a good candidate $g \in \mathcal{C}$

ideally, minimize $R(g) = \mathbb{E}_{x\sim\mathcal{D}} L(f_?(x), g(x))$

in practice $\hat{R}(g) = \frac{1}{n} \sum_{i=1}^{n} L(f_?(x_i), g(x_i))$

- Solution: a law of large numbers for the risk

$\hat{R}(g) \lim_{n\to\infty} \inf_{g'\in\mathcal{C}} R(g')$ with probability $1 - \gamma(n, \epsilon, \mathcal{C}) e^{-n\epsilon^2}$

V. N. Vapnik, *IEEE Trans. Neural Networks* 10, 988 (1999).

Learning Classical-Quantum Channels

Target (c-q function) $\rho_? : \mathcal{X} \to \mathcal{D}(\mathcal{H})$

Hypothesis class $\mathcal{C} = \{\sigma : \mathcal{X} \to \mathcal{D}(\mathcal{H})\}$

given samples $(x_i, \rho_?(x_i)) \quad i = 1, \cdots, n \quad x_i \sim \mathcal{D}$

find a good candidate

$$O \in \mathcal{C} : \hat{R}(O) - \eta \inf_{O' \in \mathcal{C}} R(O') \leq \epsilon \text{ w.h.p.}$$

- Theorem: a sufficient condition for learnability of classical-quantum processes is

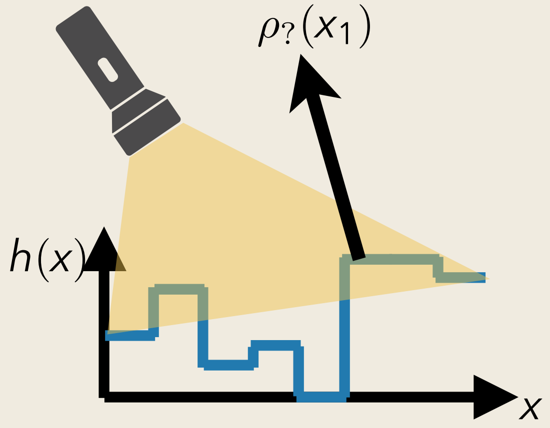
$$\lim_{n \to \infty} \frac{\log^2 \gamma(n, \epsilon, \mathcal{C})}{n} = 0$$

M. Fanizza, Y. Quek, and MR, *PRX Quantum* 5, 020367 (2024).

Applications

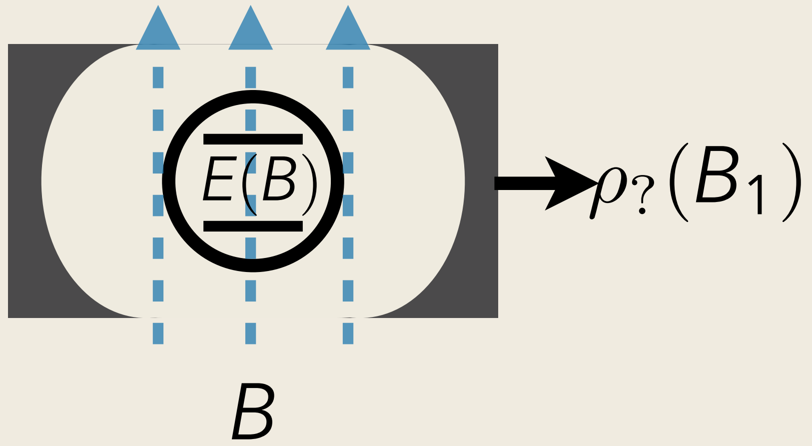
Approximate an unknown quantum process from "noisy" or "random" samples:

- Quantum imaging



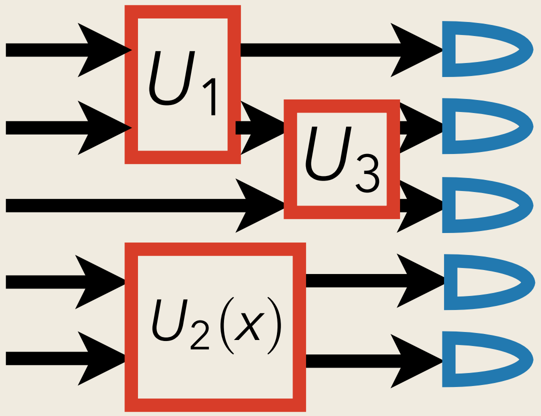
Learn the optical depth as a function of the position

- Hamiltonian learning



Learn the excited state energy as a function of the magnetic field

- Circuit compilation



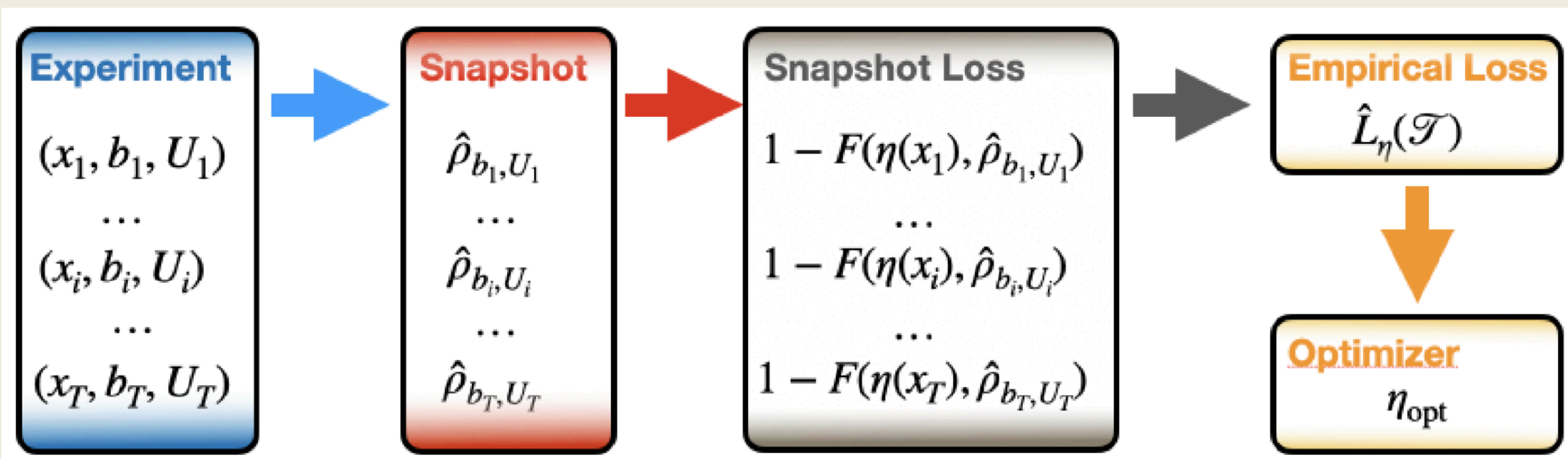
Learn circuit as a function of the classical encoding parameters

M. Fanizza, Y. Quek, and MR, *PRX Quantum* 5, 020367 (2024).

Photon-starved polarimetry with functional classical shadows

Reconstruct a a polarimetric profile, i.e., qubit angles as a function of wavelength, from standard measurements

$$\mathcal{C} = \left\{ |\eta(x)\rangle = \cos\frac{\theta(x)}{2}\,|H\rangle + e^{i\phi(x)}\sin\frac{\theta(x)}{2}\,|V\rangle \right\}$$

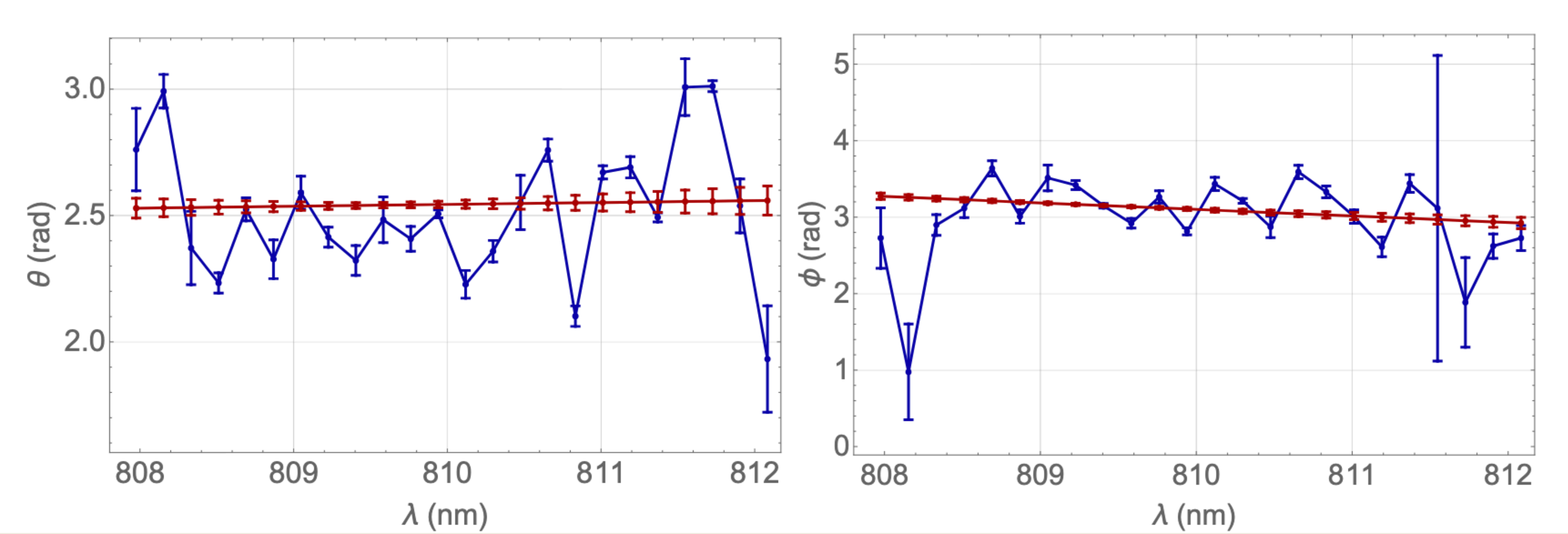


MR et al., arXiv:2509.19547, accepted for publication in AVS Quantum (2025)

Photon-starved polarimetry with functional classical shadows

Reconstruct a a polarimetric profile, i.e., qubit angles as a function of wavelength, from standard measurements

$$\mathcal{C} = \left\{ |\eta(x)\rangle = \cos\frac{\theta(x)}{2}\,|H\rangle + e^{i\phi(x)}\sin\frac{\theta(x)}{2}\,|V\rangle \right\}$$

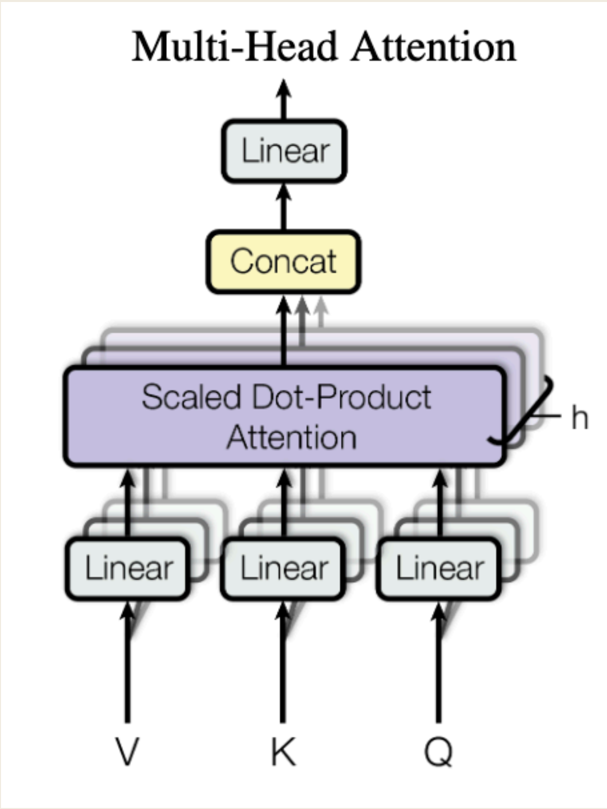


MR et al., arXiv:2509.19547, accepted for publication in AVS Quantum (2025)

Quantum-enhanced AI primitives

A Quantum Transformer Architecture

- Classical self-attention: non-linear processing of inputs



$$\mathbf{x}_1, \cdots, \mathbf{x}_j$$

$$\mathbf{z}_j = \sum_{i=1}^{j} \mathrm{softmax}\left(\frac{\mathbf{q}_j \cdot \mathbf{k}_i}{\sqrt{d_K}}\right) \mathbf{v}_i,$$

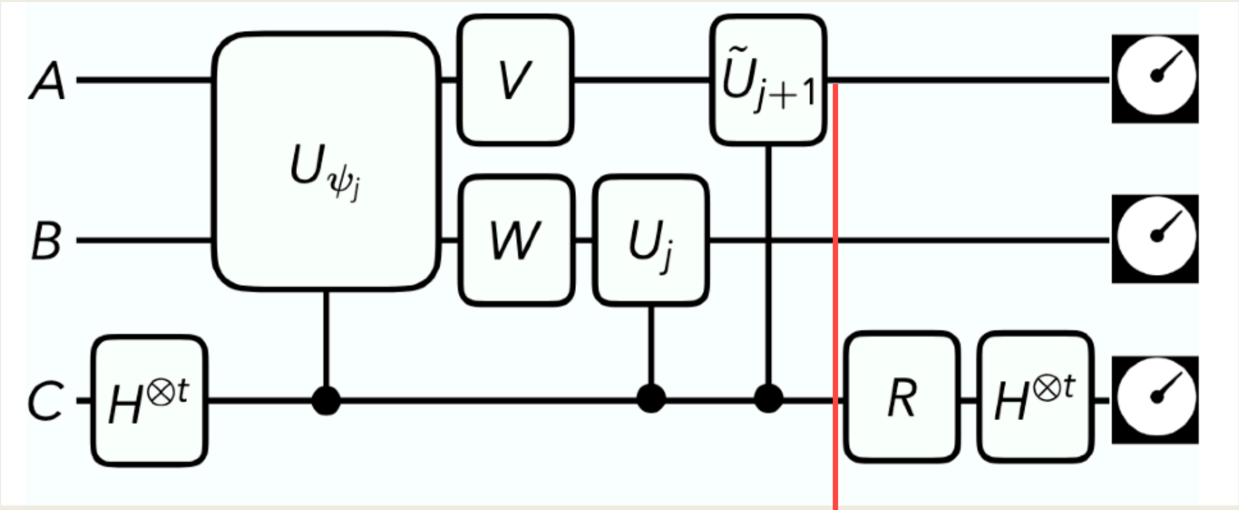
$$\tilde{\mathbf{z}}_j \propto \sum_{i=1}^{j} (\mathbf{x}_j Q^T K \mathbf{x}_i) V \mathbf{x}_i,$$

A. Vaswani *et al.*, "Attention Is All You Need," arXiv:1706.03762 (2017).

A Variational Quantum Transformer Architecture

- Quantum self-attention:
  - inputs data three times
  - variational unitary action
  - processes multiple words in coherent superposition
  - outputs Renyi cross-entropy loss as a Pauli observable's expectation

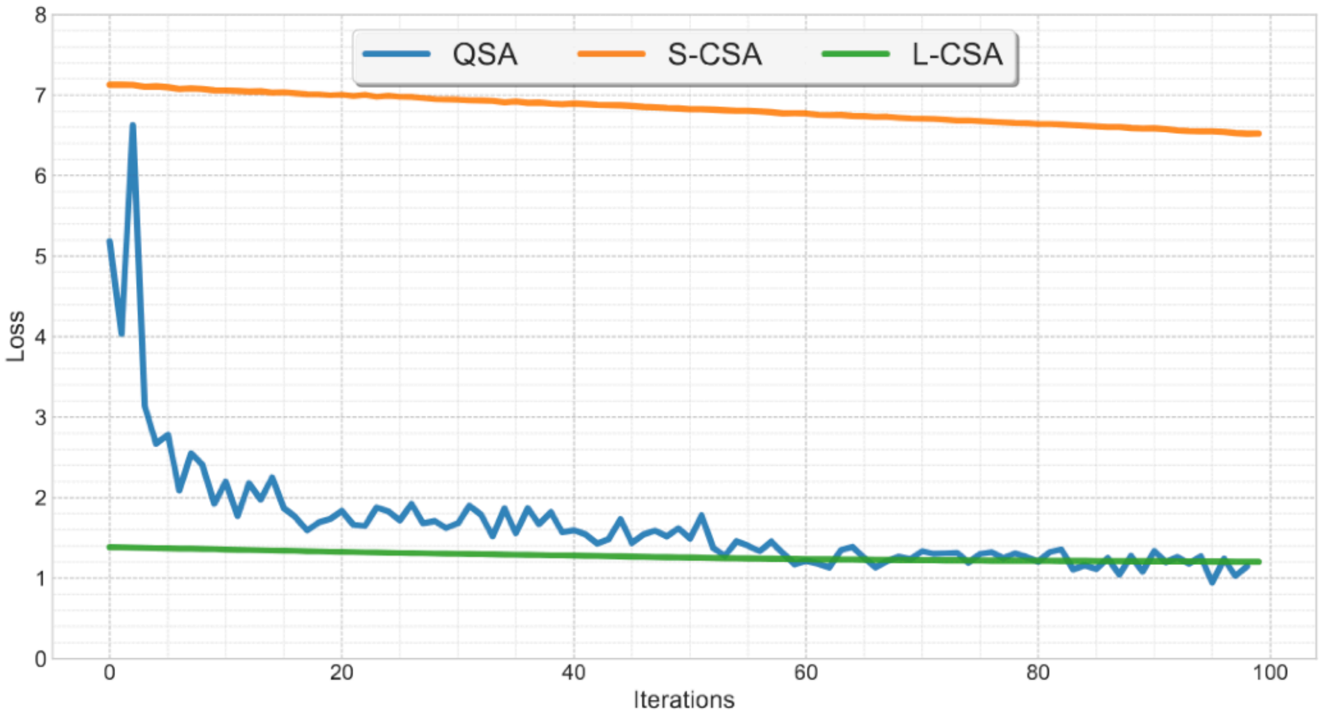
$$\mathcal{L}_{\frac{1}{2}}(p) = -2\log\left|\sum_{j=1}^{T}\sqrt{p_{j+1}}\right|$$

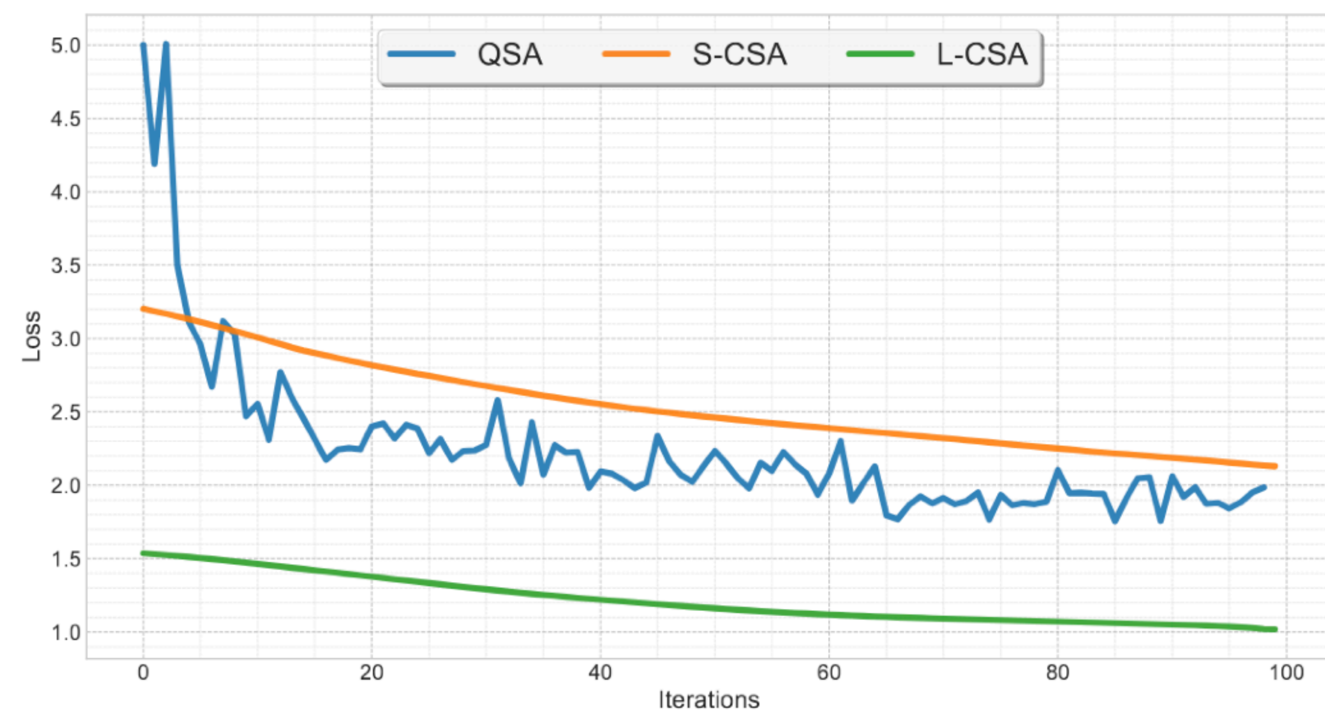


$$\propto \sum_{j=1}^{T} \langle \tilde{\mathbf{x}}_{j+1} | \tilde{\mathbf{z}}_j \rangle \, |j\rangle_C \, .$$

A. Pecilli and MR, arXiv:2602.06699, (2026).

A Quantum Transformer Architecture





Classical sequence prediction

$$(my, dog, is, black)$$

Quantum sequence prediction

$$\left(|\psi_0\rangle, e^{-iH}|\psi_0\rangle, e^{-i2H}|\psi_0\rangle, e^{-i3H}|\psi_0\rangle\right)$$

$$H_{TFIM} = \sum_i h_i X_i + \sum_{i<j} J_{ij} Z_i Z_j$$

- Complexity: QSA beats CSA for long sequences $\tilde{O}(Td^2)$ vs. $O(T^2 d)$

A. Pecilli and MR, arXiv:2602.06699, (2026).

Conclusions

- ML can calibrate and optimize quantum devices
- A learning approach replaces tomography with sample-efficient estimation of targeted quantum system properties
- Efficient mid-term Quantum LLM design