QVO VADIS QVANTVM MACHINE LEARNING

JENS EISERT, FU BERLIN

QTML, MELBOURNE, 2024

https://www.physik.fu-berlin.de/en/einrichtungen/ag/ag-eisert

- **Quantum computing** was a purely conceptual idea

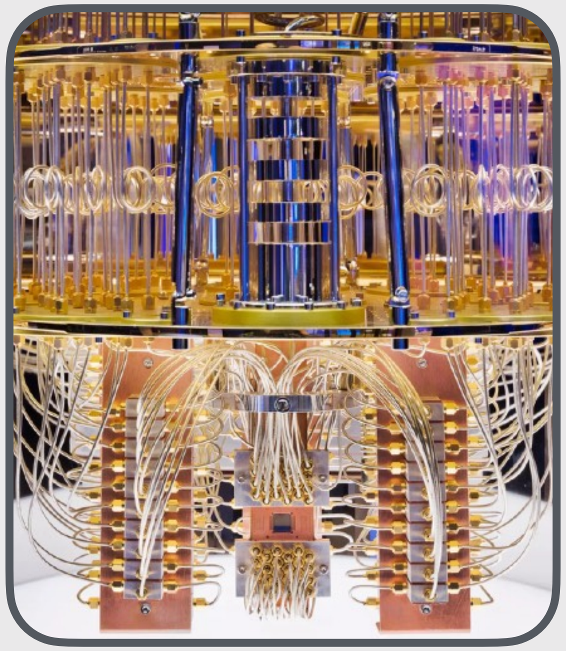
Feynman

Benioff, Deutsch

Shor

QUANTUM COMPUTERS

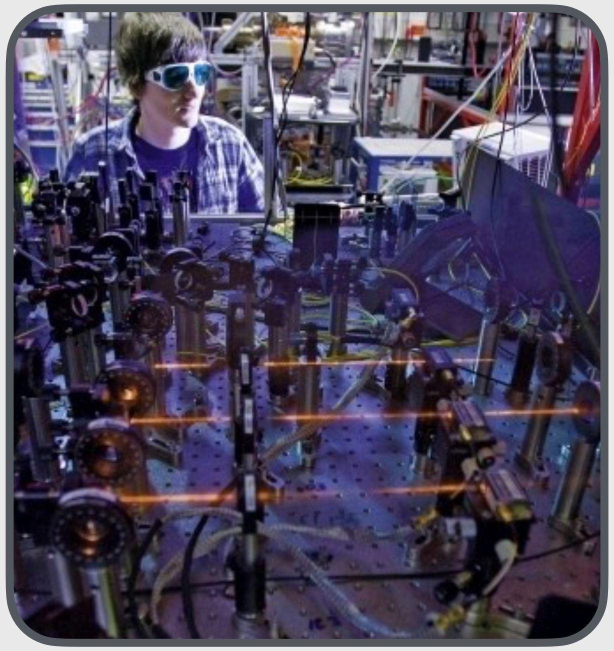
- Seems to be moving closer to **reality**



IBM

Google

QuERA

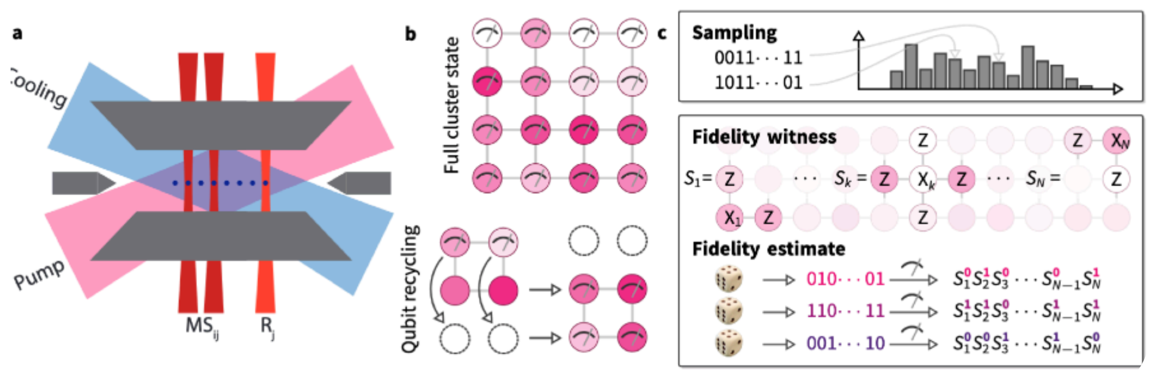


European efforts: BMBF-funded efforts, planqc, Pasqal, etc

- Can we reasonably hope for realistic quantum devices to provide a **speedup over classical computers**?

- **Quantum advantage** for paradigmatic sampling problems
- Sampling up to a constant error in $||.||_{l_1}$ distance is **classically hard**

  Aaronson, Arkhipov, Th Comp 9, 143 (2013)
  Arute, Arya, ..., Martinis, Nature 574, 505 (2019)
  Wang, Qin, Ding, Chen, Chen, You, He, Jiang, Wang, You, Renema, Hoefling, Lu, Pan, Phys Rev Lett 123, 250503 (2019)



- Efficient **verification**

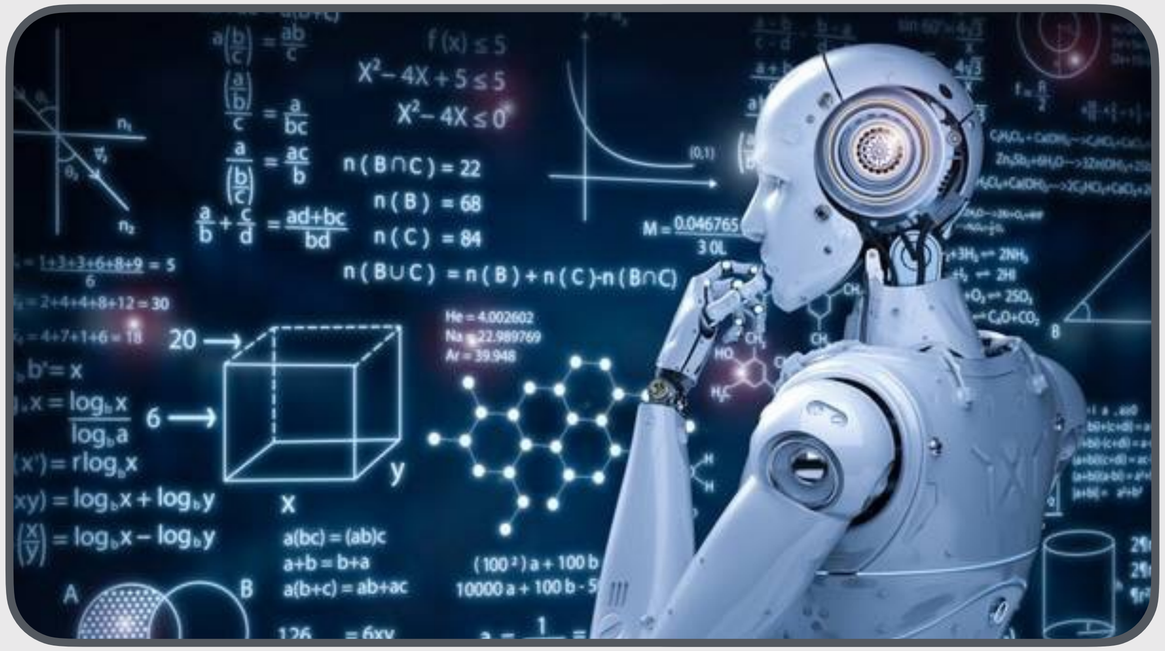
  Ringbauer, Hinsche, Feldker, Faehrmann, Bermejo-Vega, Edmunds, Stricker, Marciniak, Meth, Pogorelov, Postler, Blatt, Schindler, Eisert, Monz, Hangleiter, Nature Communications, in press (2024)

ded

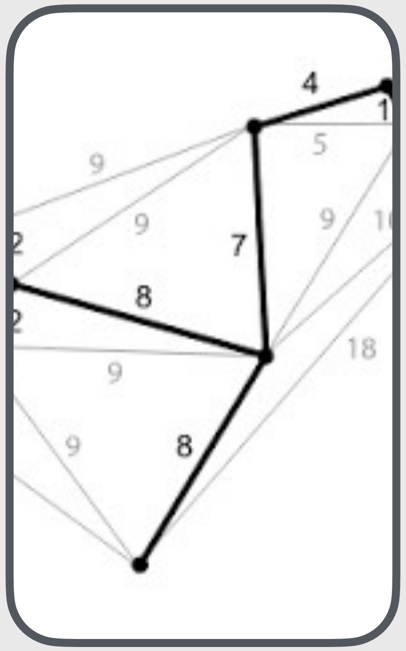
- Can we reasonably hope for realistic quantum devices to provide a **speedup over classical computers**?

QUANTUM ADVANTAGES

- **Practically** minded applications



Machine learning



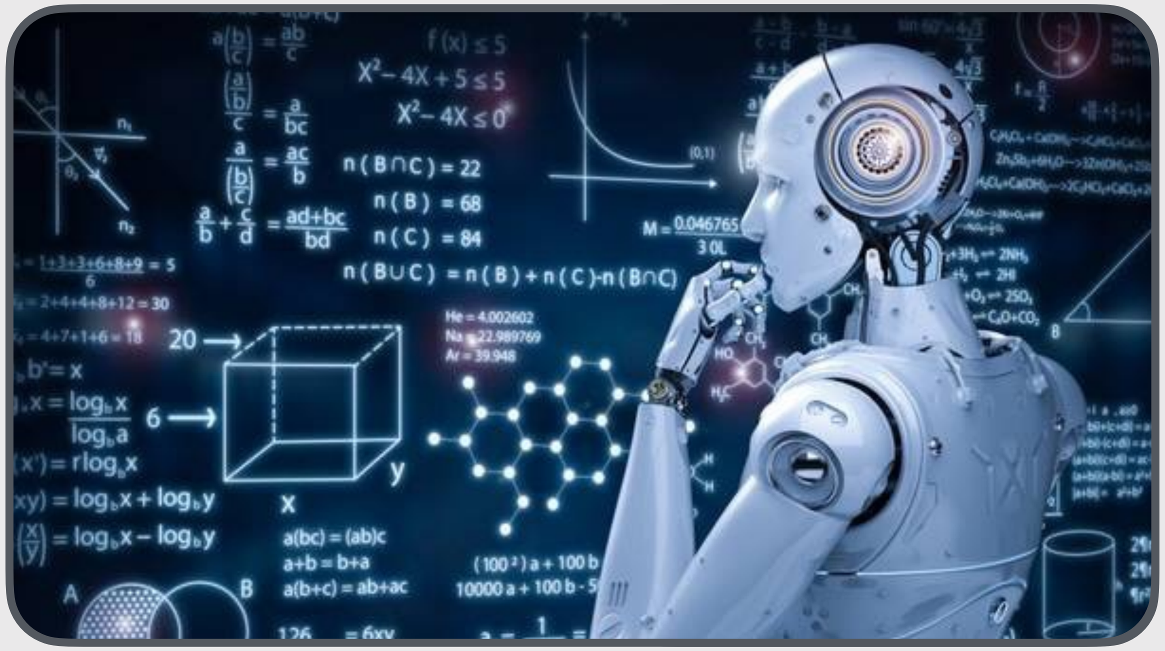
Optimization

Quantum simulation

- Can we reasonably hope for realistic quantum devices to provide a **speedup over classical computers**?

QUANTUM ADVANTAGES

- **Machine learning** has changed the world



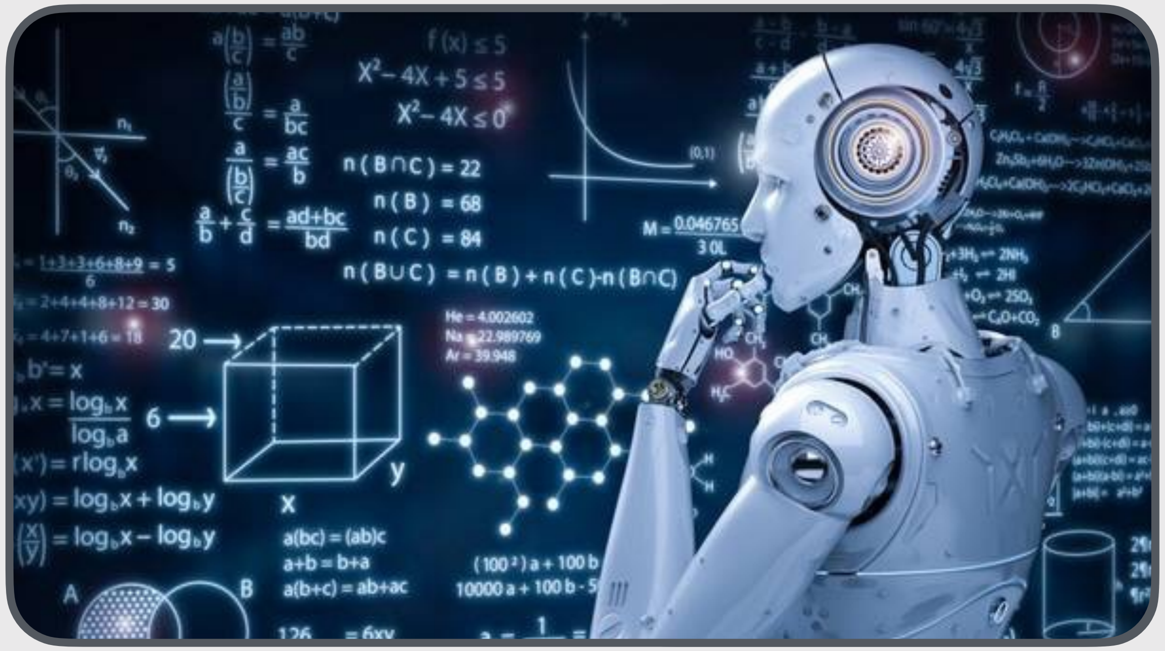
Machine learning

Optimization

Quantum simulation

- Can quantum computers assist in meaningful **machine learning tasks**?

- **Machine learning** has changed the world



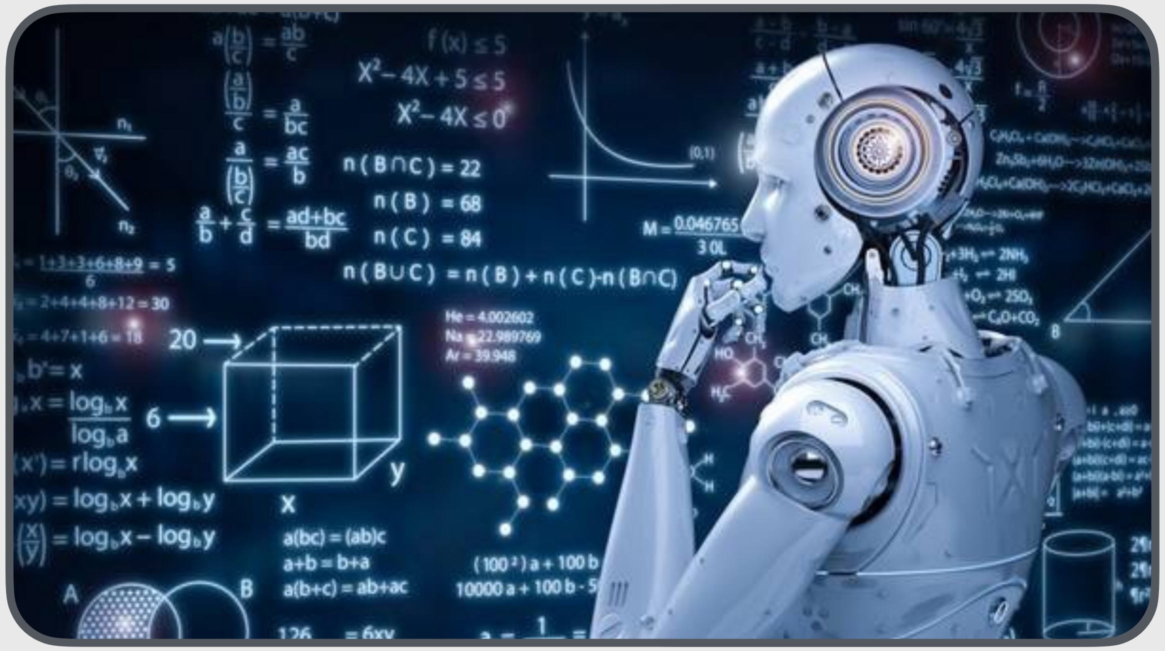
Machine learning

Optimization

Quantum simulation

- **The good:** Steps towards showing quantum advantages in **machine learning tasks**

- **Machine learning** has changed the world

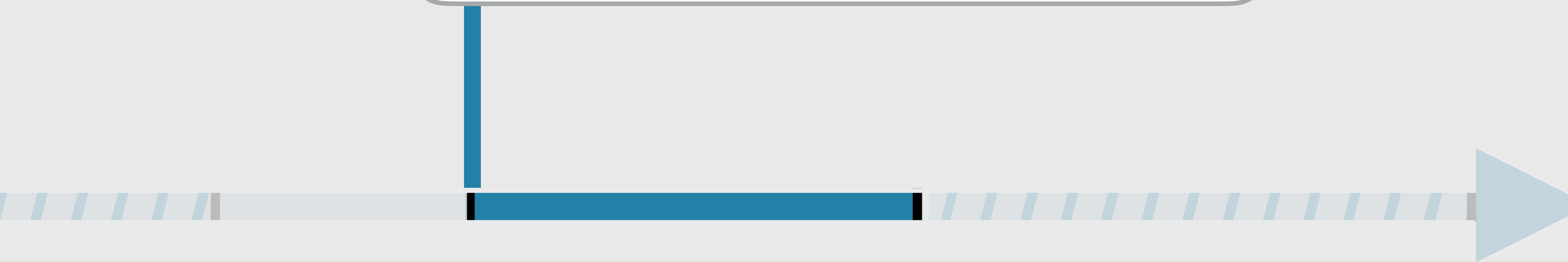


Machine learning

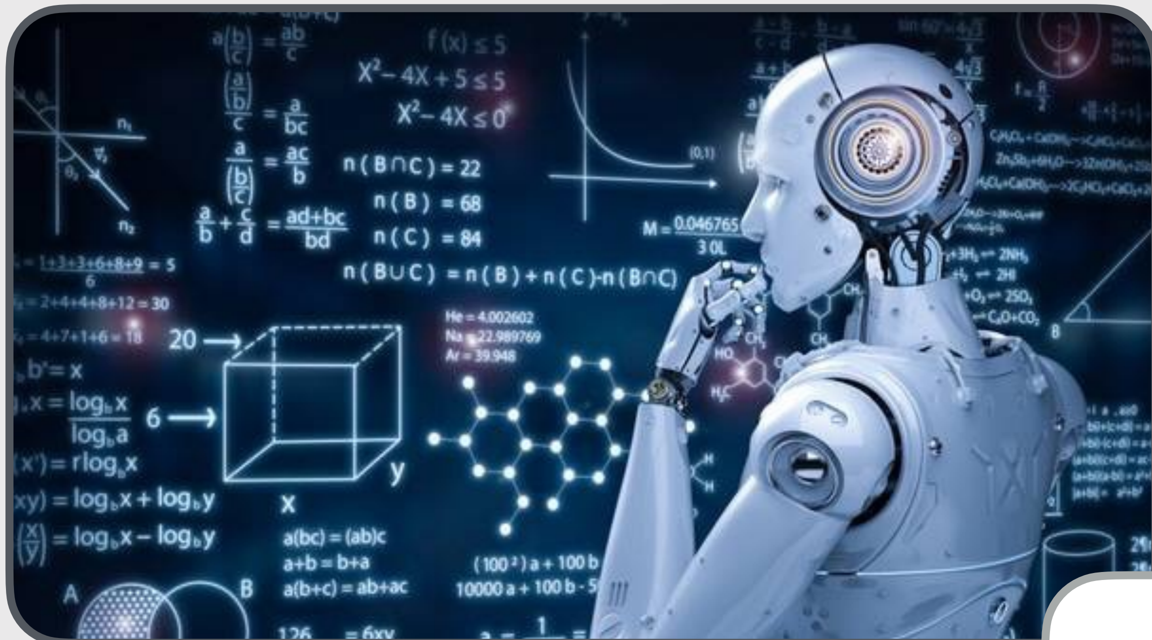
Optimization

Quantum simulation

- **The bad:** What **limitations** do we find?



- **Machine learning** has changed the world



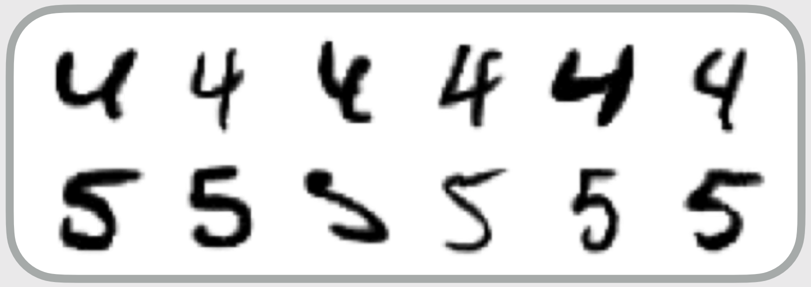
Machine learning

- **The good, the bad, and the ugly:** What are **perspectives**?

THE GOOD: QUANTUM ADVANTAGES IN MACHINE LEARNING

Pirnay, Ulitzsch, Wilde, Eisert, Seifert, Science Advances 10, eadj5170 (2024)
Liu, Liu, Liu, Ye, Alexeev, Eisert, Liang, Nature Communications 15, 434 (2024)
Pirnay, Jerbi, Seifert, Eisert, in preparation (2024)
Pirnay, Sweke, Eisert, Seifert, Phys Rev A 107, 042416 (2023)
Sweke, Seifert, Hangleiter, Eisert, Quantum 5, 417 (2021)

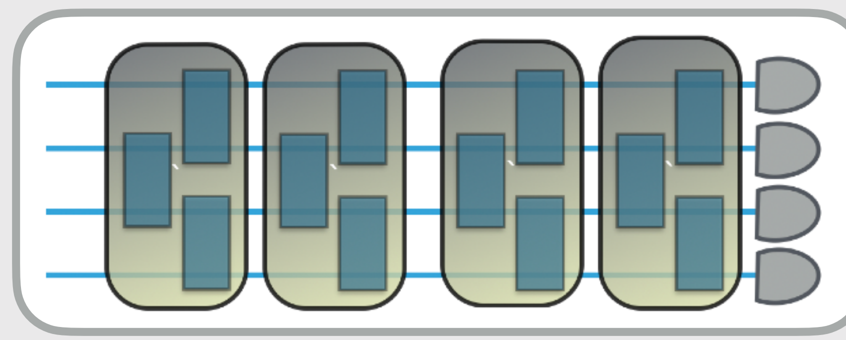
- We would like to have **rigorous guarantees** for **state-of-the-art learning** algorithms applied on **real-world datasets**



- **Data set**
  - Pictures of cats and dogs
  - Stock market data
  - Protein configurations

- **Sample complexity, computational complexity, generalisation bounds**

- **Algorithm**
  - Stochastic gradient descent
  - Expectation maximization

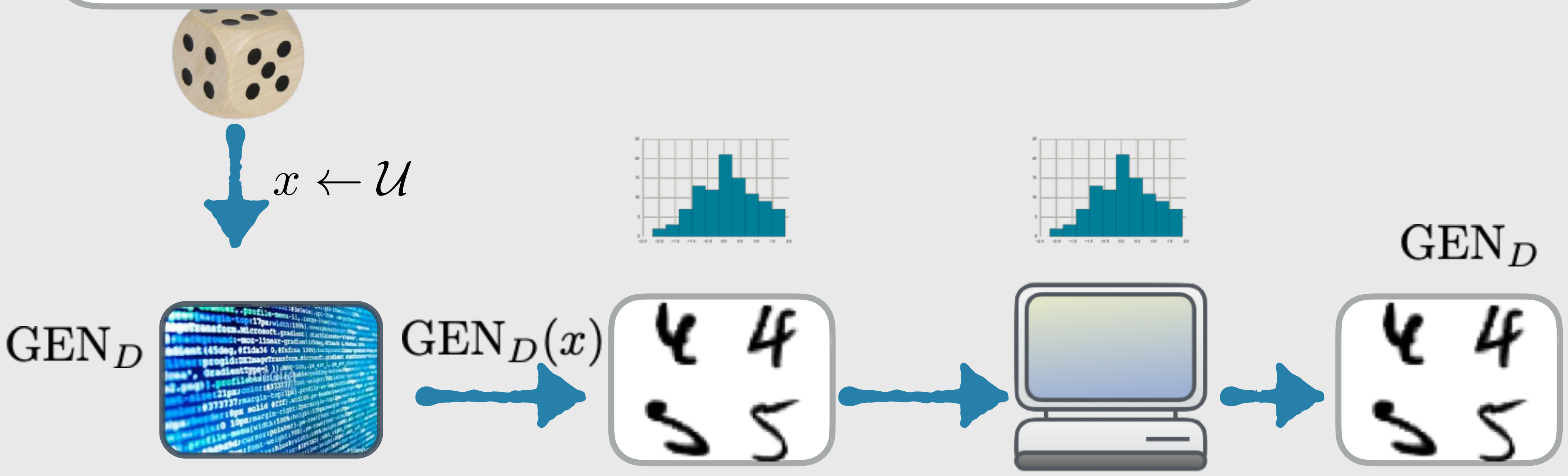


- **Model**
  - Neural networks
  - Parameterised quantum circuits

Dunjko, Briegel, Rep Prog Phys 81, 074001 (2018)
Biamonte, Wittek, Pancotti, Rebentrost, Wiebe, Lloyd, Nature 549, 195 (2017)
Arunachalam, de Wolf, arXiv:1701.06806 (2017)

LEARNING TASKS

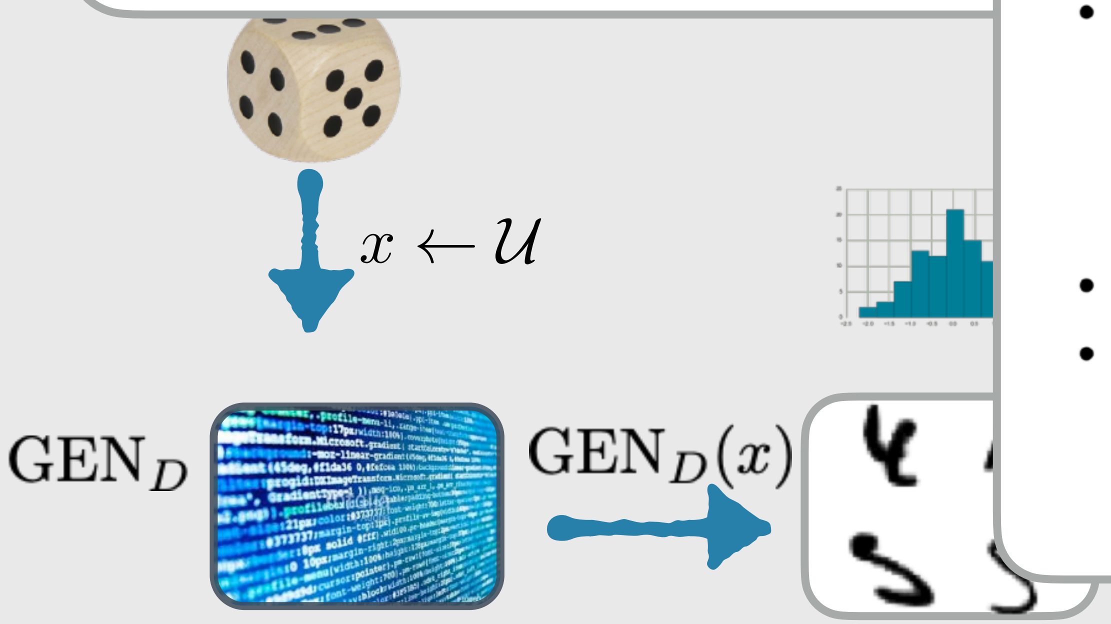
- **Theorem 1** (informal): There is a quantum advantage in PAC distribution learning



Sweke, Seifert, Hangleiter, Eisert, Quantum 5, 417 (2021)
Pirnay, Sweke, Eisert, Seifert, Phys Rev A 107, 042416 (2023)

LEARNING TASKS

- **Theorem 1** (informal): There is a quantum advantage in PAC distri



- **"Probably approximately correct" learning of distribution classes**
  - A distribution class $\mathcal{C}$ is efficiently PAC learnable w.r.t. distance d if there is an algorithm $\mathcal{A}$ which for every $D \in \mathcal{C}$ and every $\epsilon, \delta > 0$ given access to an oracle $O(D)$, outputs in time $\mathrm{poly}(|D|, 1/\epsilon, 1/\delta)$
  - with probability at least $1 - \delta$ ("probably") a generator $\mathrm{GEN}_{D'}$
  - of a distribution D' such that

$$d(D, D') < \epsilon$$

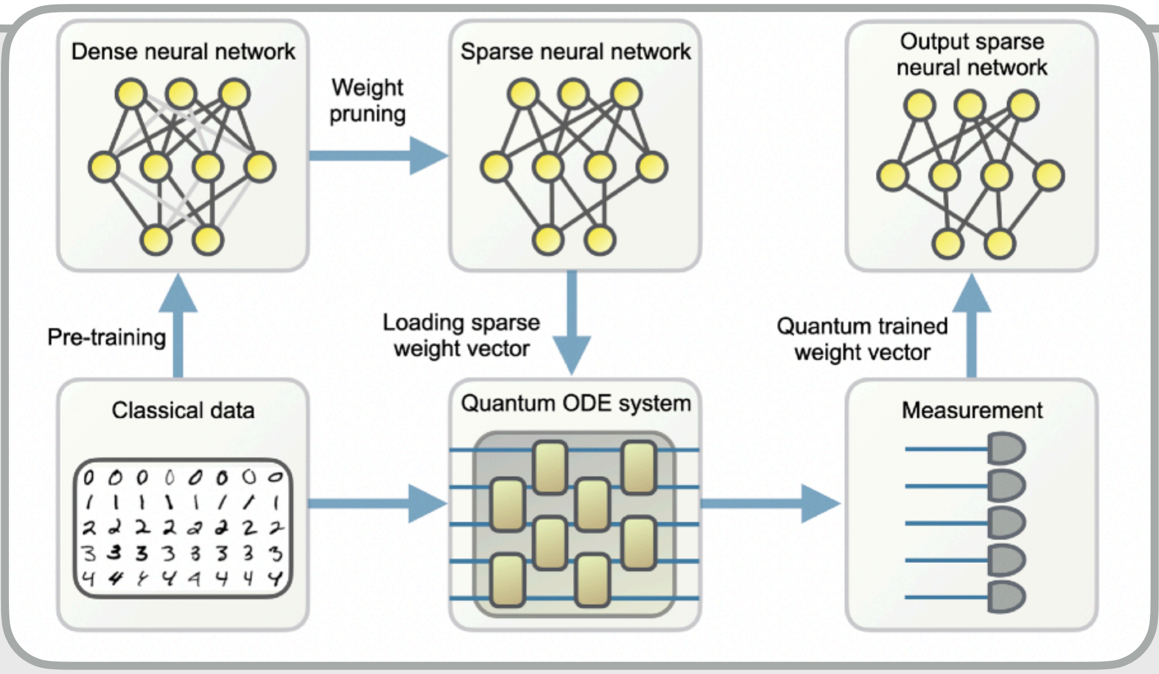
("approximately correct")

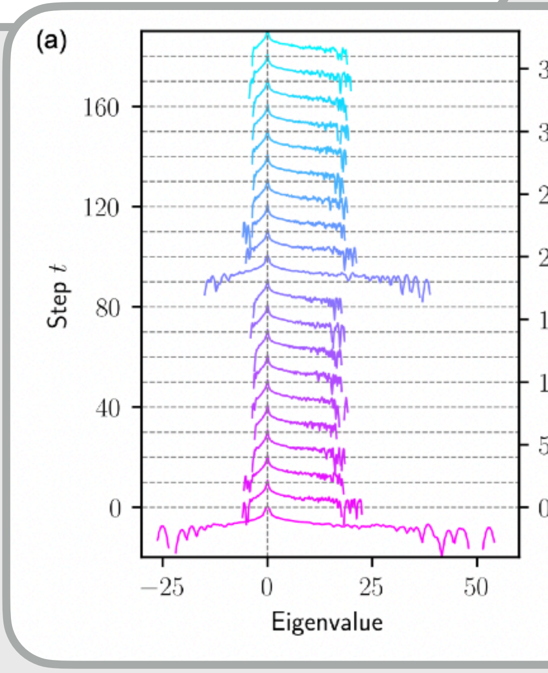
- **Proof techniques**
  - Hard to learn distributions using pseudorandom functions
  - Goldreich-Goldwasser-Micali trees
- **Features**
  - Superpolynomial advantage for learning task
  - Classical but artificial data
  - Fault tolerant quantum computer

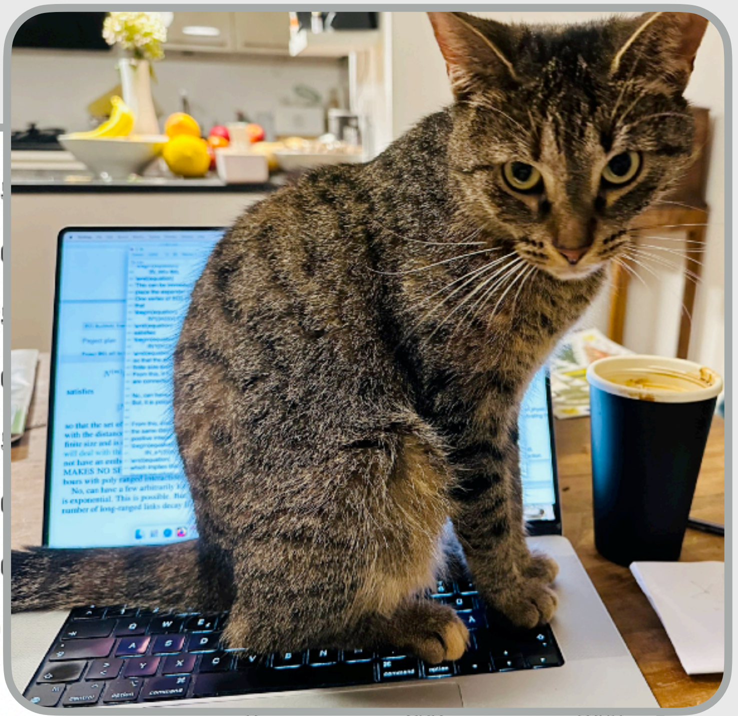
Sweke, Seifert, Hangleiter, Eisert, Quantum 5, 417 (2021)
Pirnay, Sweke, Eisert, Seifert, Phys Rev A 107, 042416 (2023)
Kearns, Mansour, Sellie, STOC (1994)
Compare Liu, Arunachalam, Temme, Nature Phys 17, 1013 (2021)

LEARNING TASKS

- **Theorem 2** (informal): There is a quantum advantage in **training** pruned classical networks





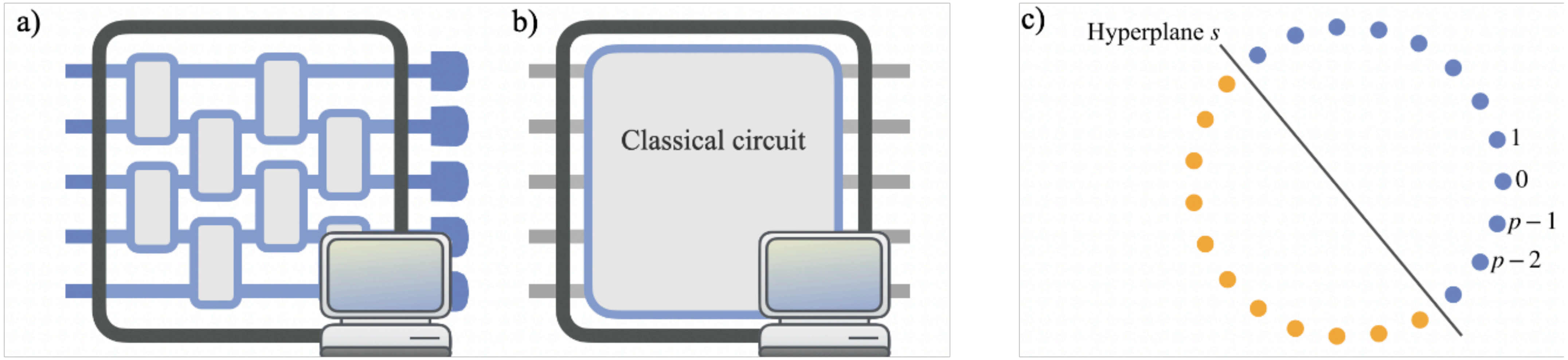


- **Proof techniques**
  - New variant of HHL algorithm for training pruned sparse classical networks
- **Features**
  - Superpolynomial advantage in training over gradient descent
  - Classical and natural data
  - Fault tolerant quantum computer

Liu, Liu, Liu, Ye, Alexeev, Eisert, Liang, Nature Comm 15, 434 (2024)

LEARNING TASKS

- **Theorem 3** (informal): There is a quantum advantage for **shallow quantum circuits**



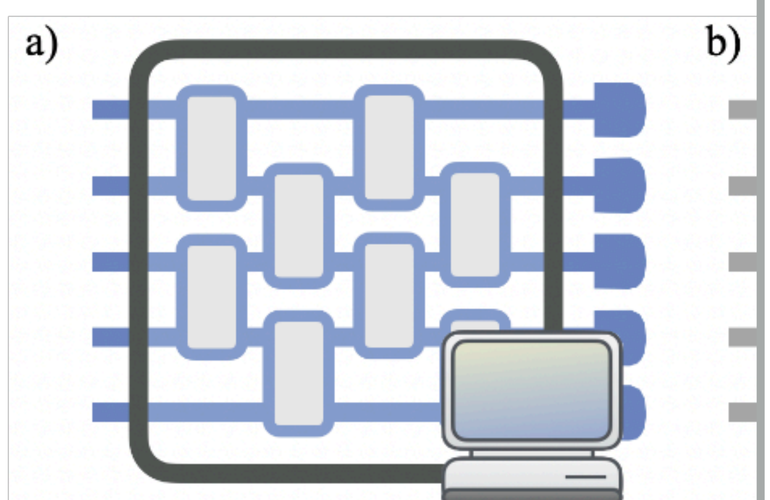
- **Proof techniques**
  - Construct concept classes from quantum advantages from shallow circuits

- **Features**
  - Constant-depth quantum vs $\omega(\log\log(n))$ (NC^0) circuits
  - Classical and artificial data
  - Near-term quantum computer

Pirnay, Jerbi, Seifert, Eisert, today (2024)

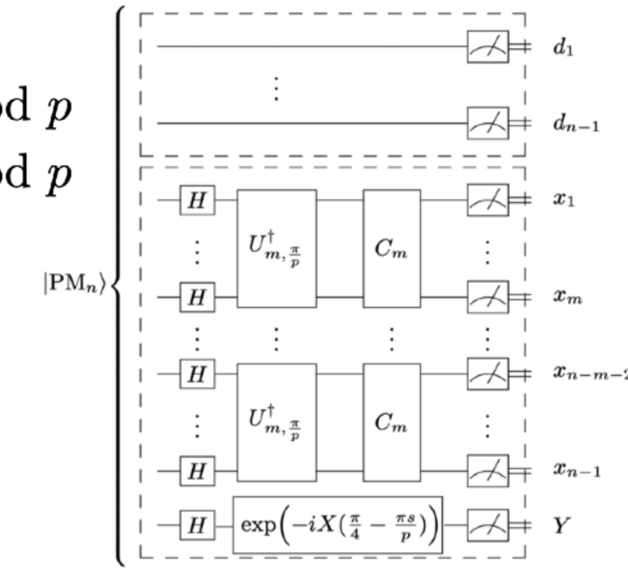
Compare Huang, Liu, Broughton, Kim, Anshu, Landau, McClean, arXiv:2401.10095 (2024)

- **Theorem 3** (informal): Th…
  advantage for shallo…



- Core idea: Devise a PAC generator learning advantage from an unconditional sampling advantage of QNC^0 over NC^0
- Encode hyperplane learning problem into the "majority mod p" function

$$\text{majmod}_{p,s}(x) = \begin{cases} 0, \text{ if } |x| + s < p/2 & \mod p \\ 1, \text{ if } |x| + s > p/2 & \mod p \end{cases}$$



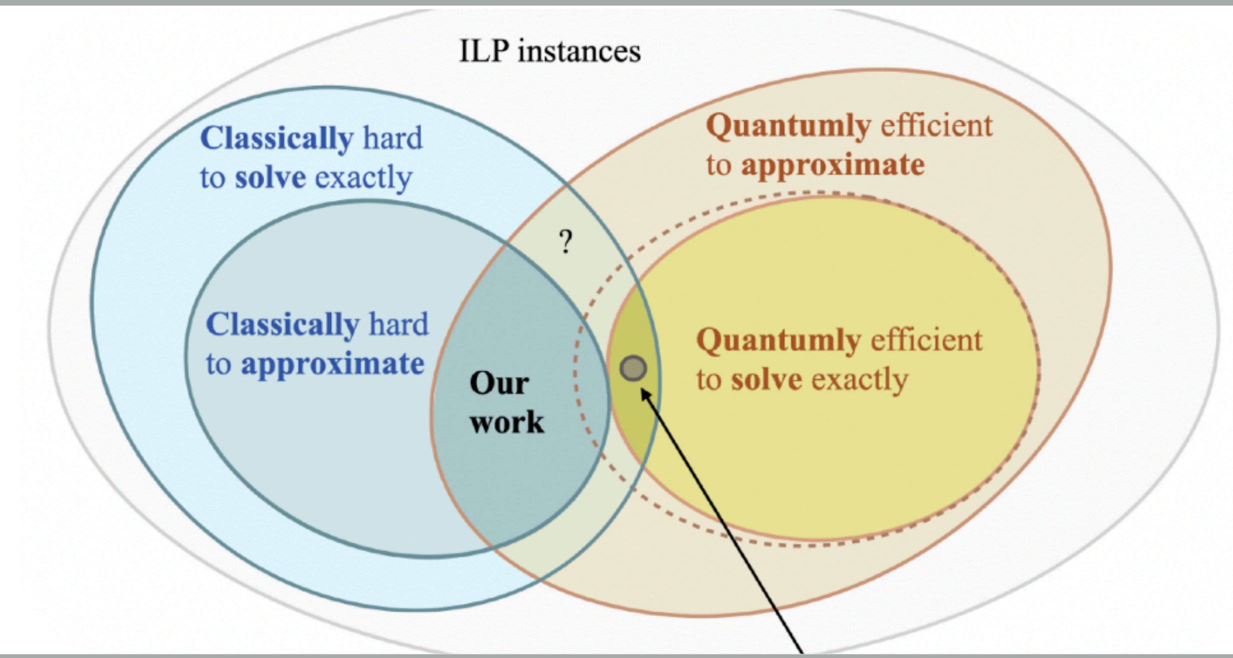
Building on Bene Watts, Parham, arXiv:2301.00995 (2013)
Bravyi, Gosset, König, Science 362, 308 (2018)

- **Proof techniques**
  - Construct concept classes from quantum advantages from shallow circuits
  - Constant-depth quantum vs $\omega(\log\log(n))$ (NC^0) circuits
  - Classical and artificial data
  - Near-term quantum computer

Pirnay, Jerbi, Seifert, Eisert, today (2024)
Compare Huang, Liu, Broughton, Kim, Anshu, Landau, McClean, arXiv:2401.10095 (2024)

STEPS TOWARDS COMBINATORIAL OPTIMISATON

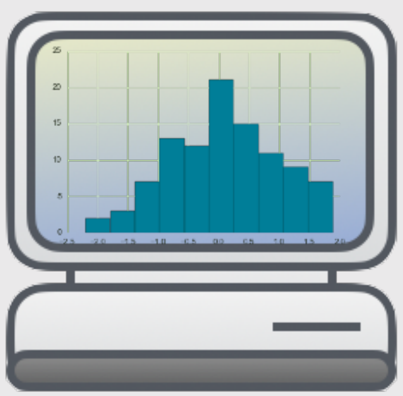
- **Theorem 4** (informal): There is a quantum advantage in **integer programming**



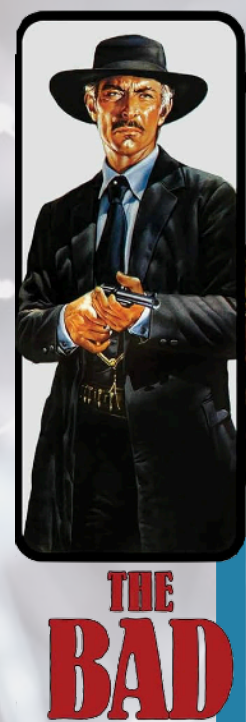
- **Proof techniques**
  - Occam's razor, computational learning theory applied to formula coloring problem
  - Hardness of inverting RSA encryption
- **Features**
  - Approximating hard classically hard instances
  - Artificial instances
  - Fault tolerant quantum computer

Pirnay, Ulitzsch, Wilde, Eisert, Seifert, Science Advances 10, eadj5170 (2024)
Szegedy, arXiv:2212.12572 (2022)

- Proven **quantum advantages** in interesting learning (and some optimization) problems



<

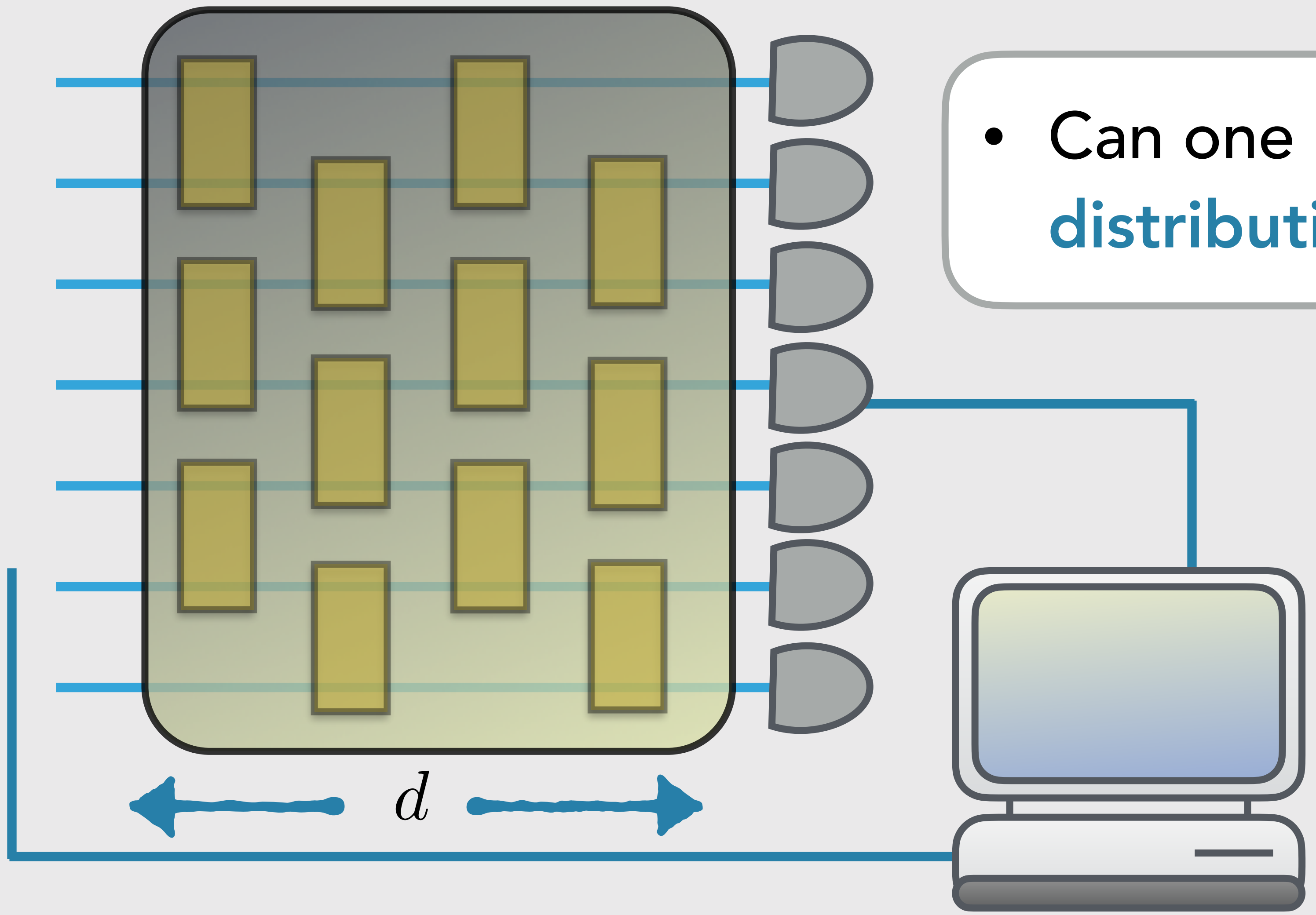


THE BAD: LIMITATIONS ALREADY FOR SHALLOW CIRCUITS

Quek, Stilck França, Khatri, Meyer, Eisert, Nature Physics 20, 1648 (2024)
Mele, Angrisani, Ghosh, Khatri, Eisert, Stilck Franca, Quek, arXiv:2403.13927 (2024)
Hinsche, Ioannou, Nietner, Haferkamp, Quek, Hangleiter, Seifert, Eisert, Sweke, Phys Rev Lett 130, 240602 (2023)

LIMITATIONS FOR SHORT REALISTIC CIRCUITS?

- **Question:** What **limitations** do we find?

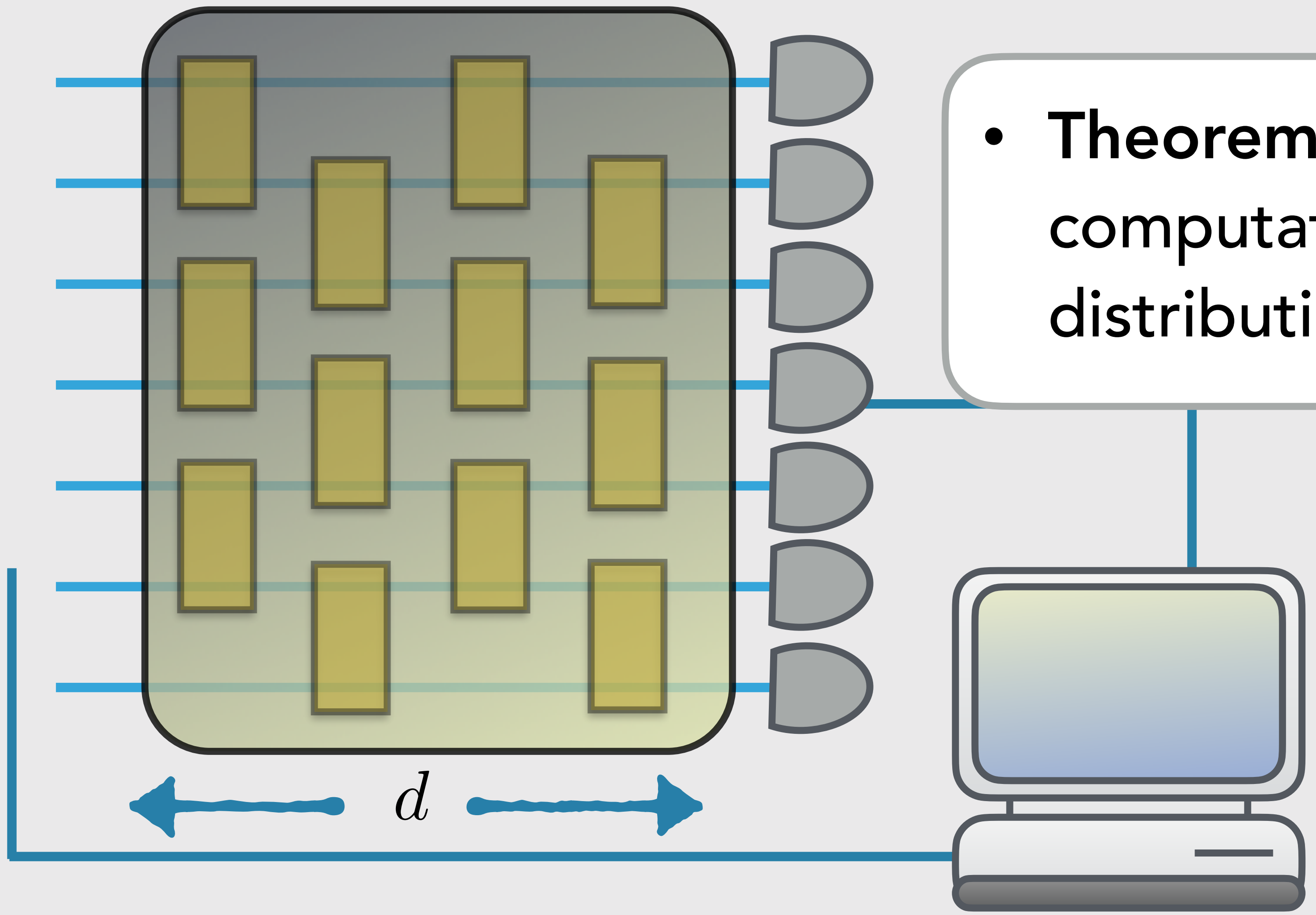


- Can one (PAC) learn the **output distribution** of quantum circuits?

Hinsche, Ioannou, Nietner, Haferkamp, Quek, Hangleiter, Seifert, Eisert, Sweke, Phys Rev Lett 130, 240602 (2023)

LIMITATIONS FOR SHORT REALISTIC CIRCUITS?

- **Question:** What limitations do we find?



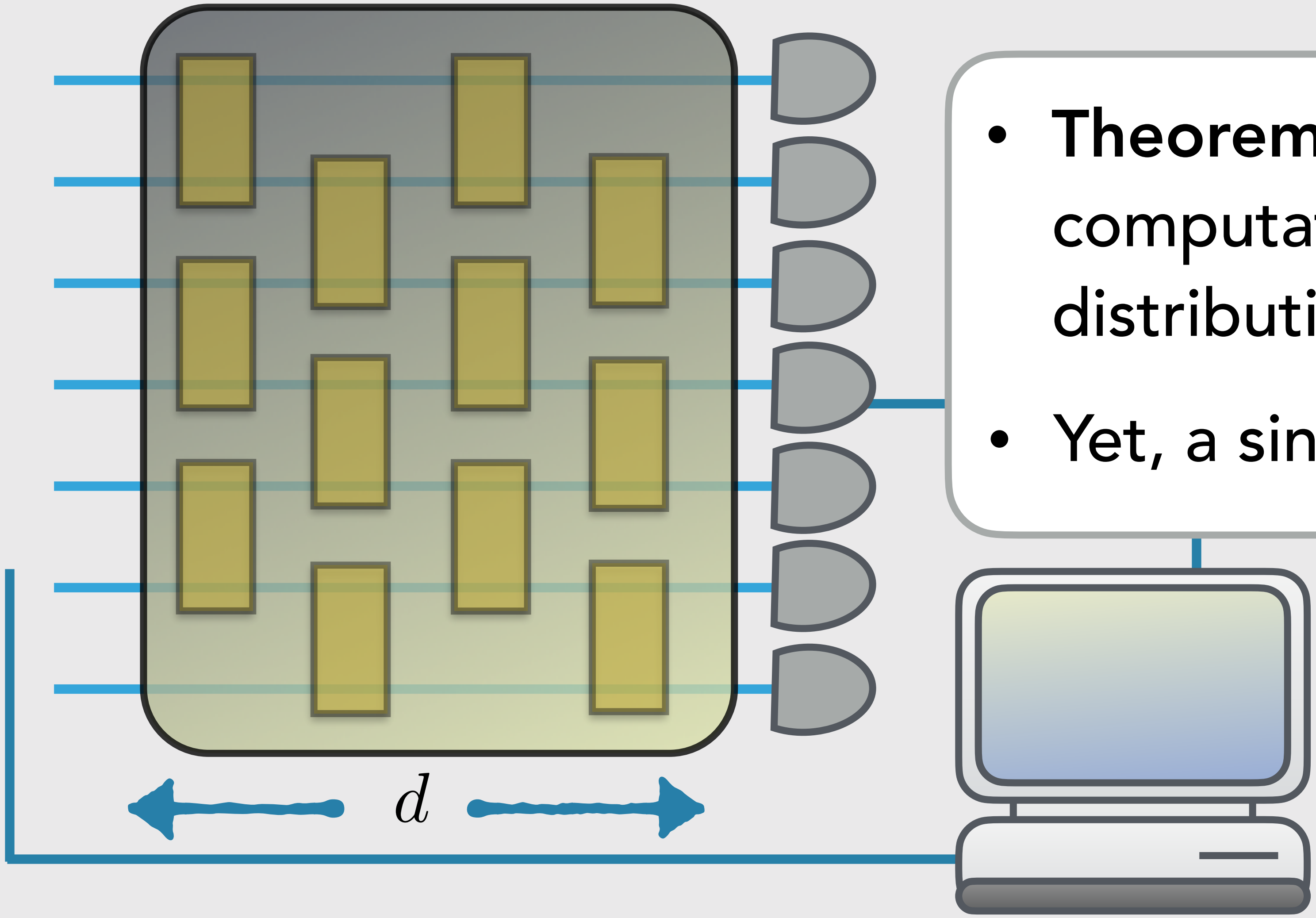
- **Theorem 5:** Can efficiently (sample and computationally) learn the output distribution of Clifford circuits

- Proof technique
  - Clifford-circuit output distributions uniform over affine subspaces of finite-dimensional vector space $\mathbb{F}_2^n$

Hinsche, Ioannou, Nietner, Haferkamp, Quek, Hangleiter, Seifert, Eisert, Sweke, Phys Rev Lett 130, 240602 (2023)

LIMITATIONS FOR SHORT REALISTIC CIRCUITS?

- **Question:** What limitations do we find?



- **Theorem 5:** Can efficiently (sample and computationally) learn the output distribution of Clifford circuits
- Yet, a single T-gate renders learning hard

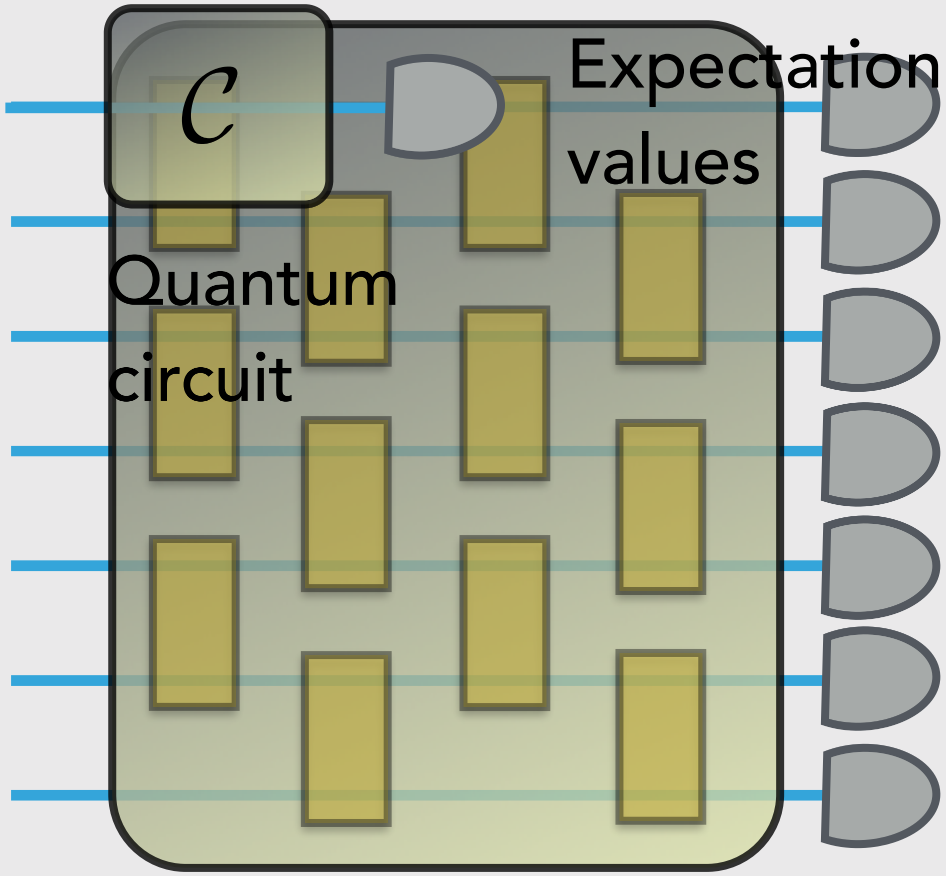
- **Proof technique**
  - Learning parity with noise
  - Remains average-case hard

Nietner, Ioannou, Sweke, Kueng, Eisert, Hinsche, Haferkamp, arXiv:2305.05765 (2023)

Hinsche, Ioannou, Nietner, Haferkamp, Quek, Hangleiter, Seifert, Eisert, Sweke, Phys Rev Lett 130, 240602 (2023)

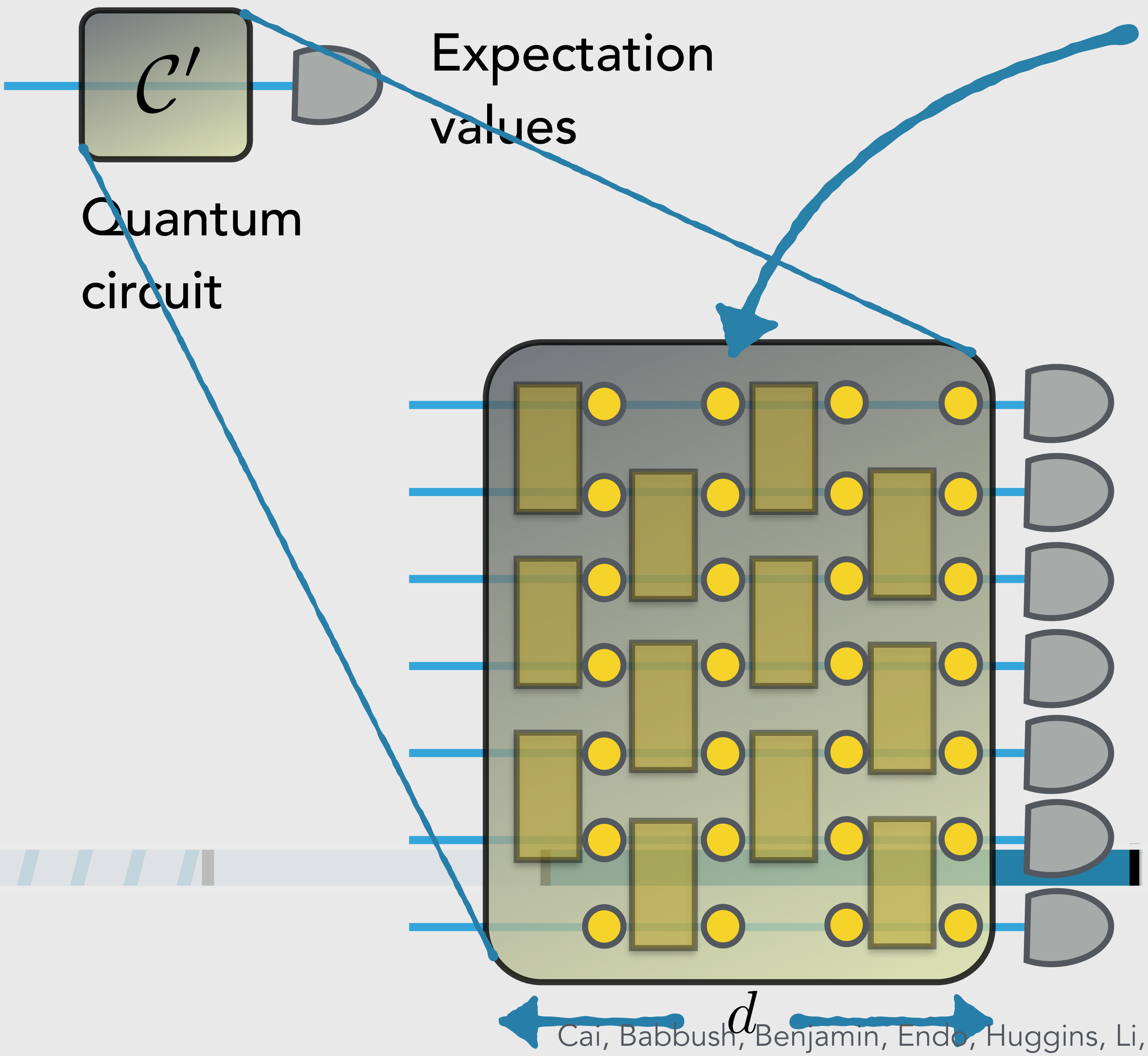
NOISY CIRCUITS

- Actual circuits are **noisy**



Temme, Bravyi, Gambetta, Phys Rev Lett 119, 180509 (2017)
Cai, Babbush, Benjamin, Endo, Huggins, Li, McClean, O'Brien, Cai et al, Rev Mod Phys 95, 045005 (2022)

- Actual circuits are **noisy**



- Noise and decoherence, for now **depolarizing**

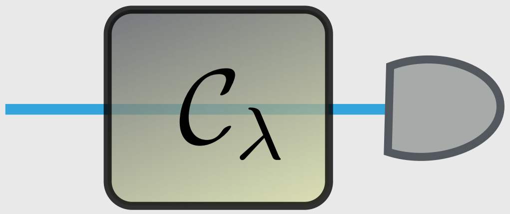
$$\mathcal{D}_p(M) = pM + (1-p)\mathrm{tr}(M)\frac{\mathbb{I}}{2}$$

but can also be non-unital

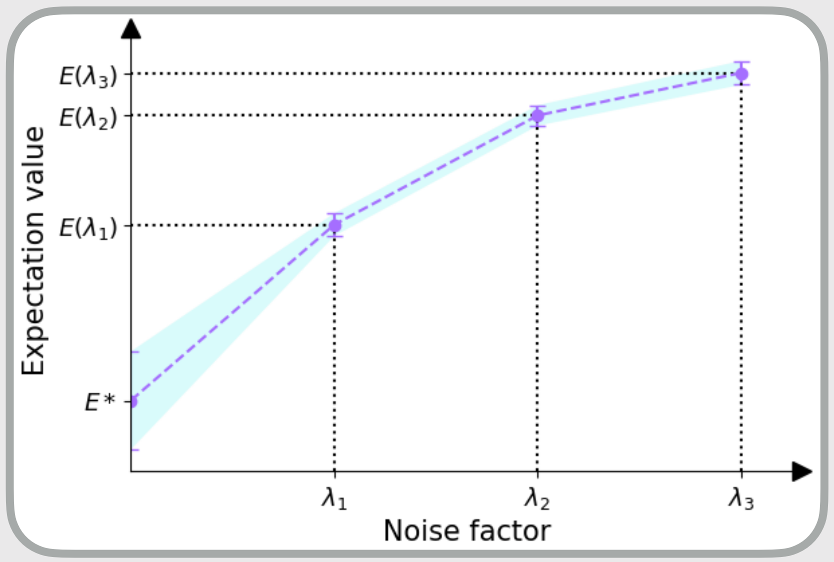
Temme, Bravyi, Gambetta, Phys Rev Lett 119, 180509 (2017)
Cai, Babbush, Benjamin, Endo, Huggins, Li, McClean, O'Brien, Cai et al, Rev Mod Phys 95, 045005 (2022)

EXAMPLE OF AN ERROR MITIGATION PROTOCOL

- Zero-noise extrapolation



- Add noise levels to $\mathcal{C}_\lambda$
- Measure $E(\lambda) = \mathrm{tr}(C_\lambda(\rho_{\mathrm{in}})O)$
- Extrapolate to $\lambda = 0$



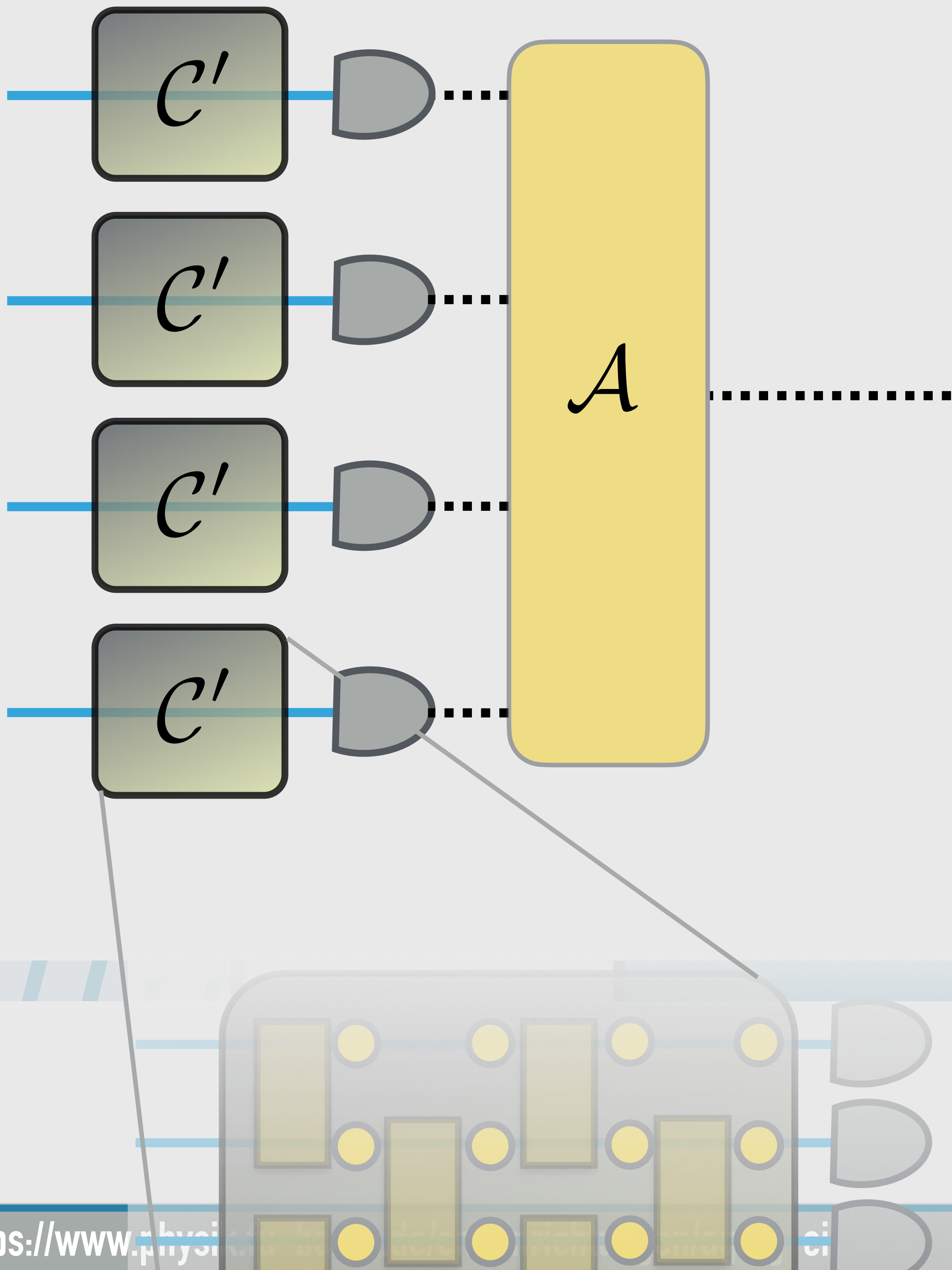
Majumdar, Rivero, Metz, Hasan, Wang, 2023 IEEE Int Conf Quant Comp Eng (2023)

- Or, probabilistic error cancellation

Temme, Bravyi, Gambetta, Phys Rev Lett 119, 180509 (2017)
Cai, Babbush, Benjamin, Endo, Huggins, Li, McClean, O'Brien, Cai et al, Rev Mod Phys 95, 045005 (2022)

OUR ABSTRACTION

- Captures large classes of protocols

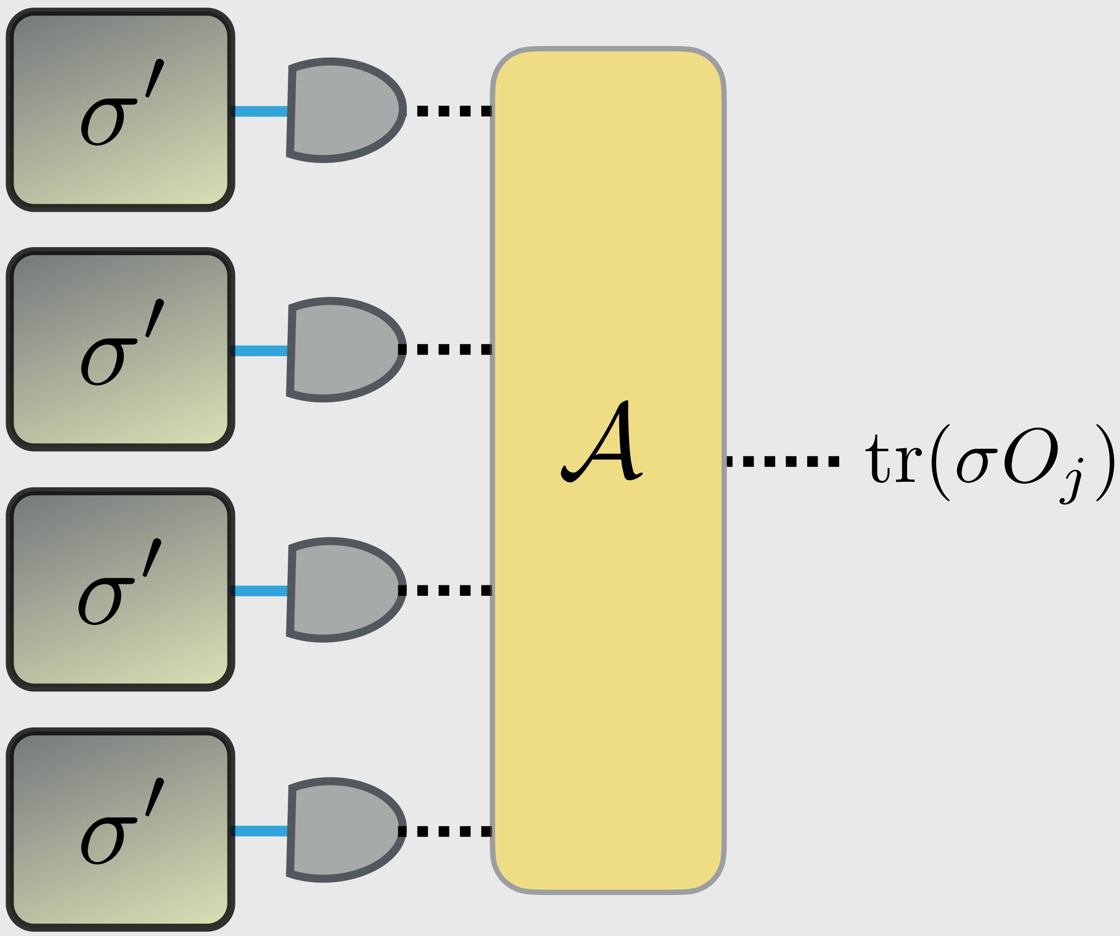


Expectation values

$\mathrm{tr}(\sigma O_j), O_j \in \mathcal{M}$

Quek, Stilck França, Khatri, Meyer, Eisert, Nature Phys 20, 1648 (2024)

- How many copies of σ' are needed to estimate $\mathrm{tr}(\sigma O_j)$ for $O_j \in \mathcal{M}$ to precision ε and probability $1-\delta$?



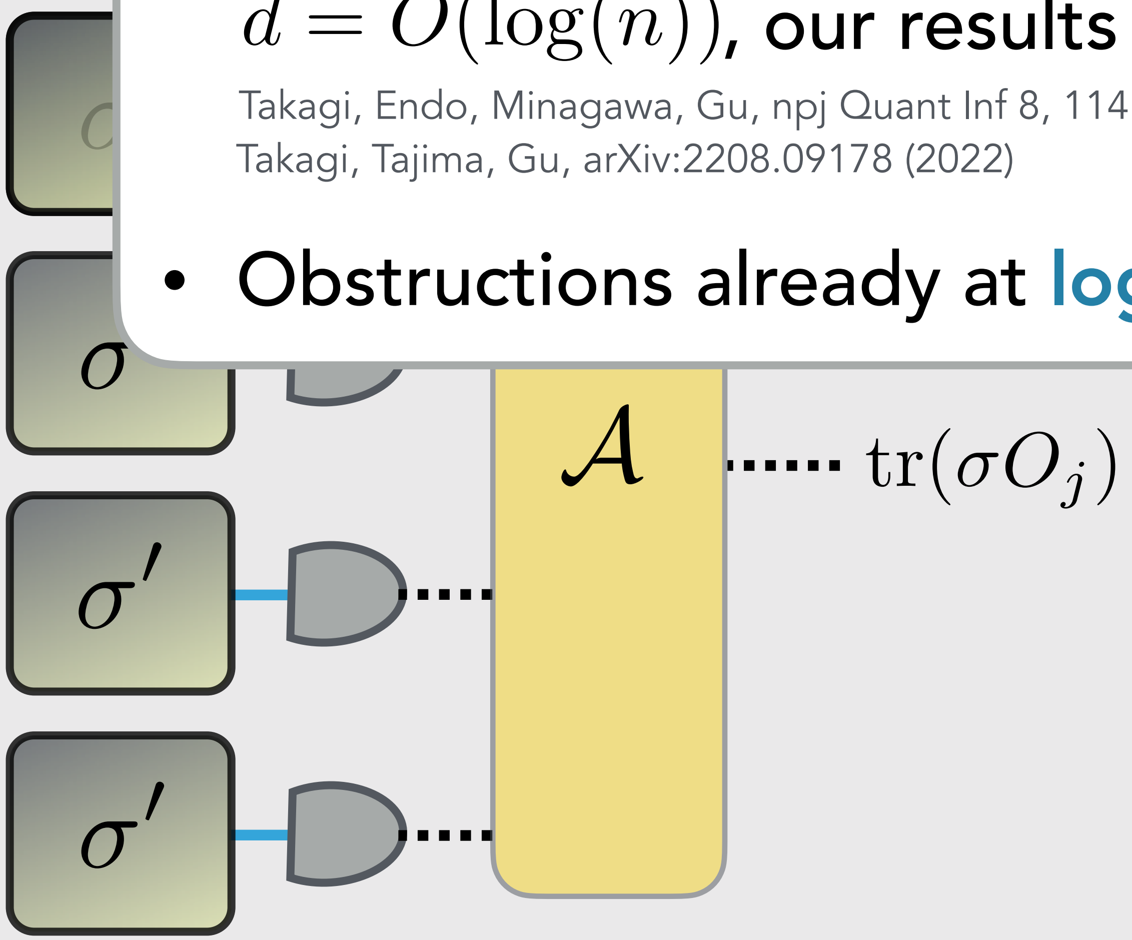
- **Statistical inference problem**
  - in terms of circuit depth d
  - circuit width n
  - (depolarizing) noise strength p

Quek, Stilck França, Khatri, Meyer, Eisert, Nature Phys 20, 1648 (2024)

RESULTS

- **Theorem 6** (informal): One needs $\exp(\Omega(nd))$ **many samples**
- Previously thought: $\exp(\Omega(d))$, and since in NISQ regime $d = O(\log(n))$, our results are **exponentially** stronger

  Takagi, Endo, Minagawa, Gu, npj Quant Inf 8, 114 (2022)
  Takagi, Tajima, Gu, arXiv:2208.09178 (2022)
- Obstructions already at **log-log depth**

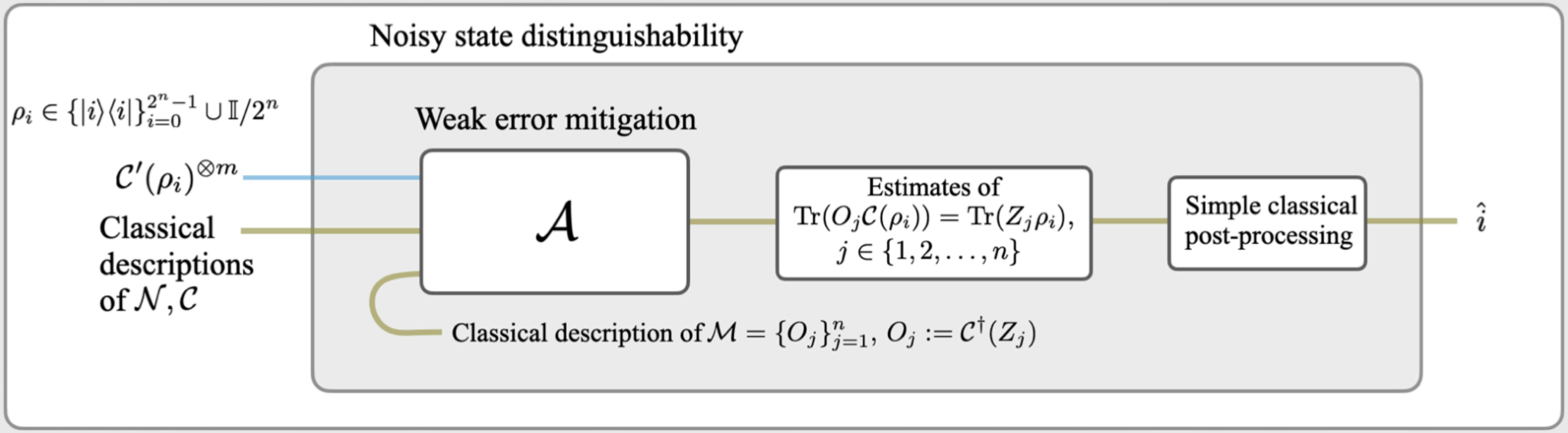


- Both for depolarizing
- and general local, including non-unital, noise

Quek, Stilck França, Khatri, Meyer, Eisert, Nature Phys 20, 1648 (2024)

THREE PROOF IDEAS

- Set up a **learning problem** that can be solved by error mitigation



- Discriminate outputs of noisy circuits of computational **basis states** from **maximally mixed** state, $D(\mathcal{C}'(|i\rangle\langle i|)||\mathcal{C}'(\mathbb{I}/2))$

Quek, Stilck França, Khatri, Meyer, Eisert, Nature Phys 20, 1648 (2024)

THREE PROOF IDEAS

- Set up a **learning problem** that can be solved by error mitigation
- Lower bound the **sample complexity**

- Use **Fano's Lemma**: Any single-sample test distinguishing $N+1$ distributions $P_0, \ldots, P_N$ must fail with probability at least $1-\alpha$,

$$\frac{1}{\log(N)} \frac{1}{N+1} \sum_{k=0}^{N} D(P_k || P_N) \leq \alpha$$

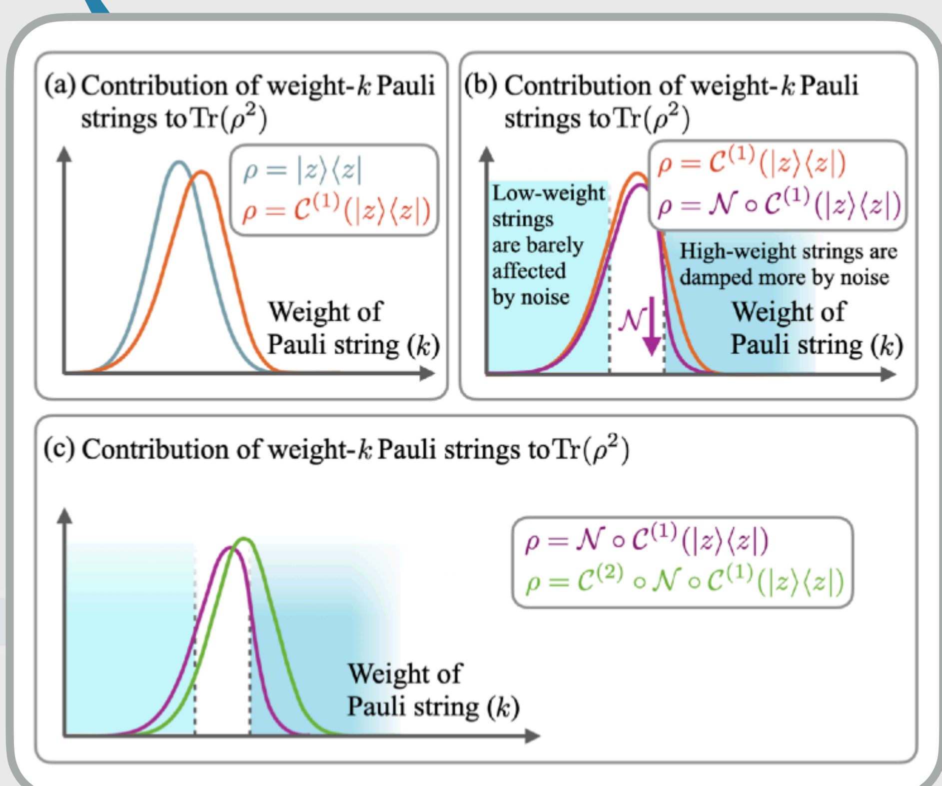
- Apply this to **noisy state distinguisher** for computational basis state or maximally mixed inputs

Gives right scaling

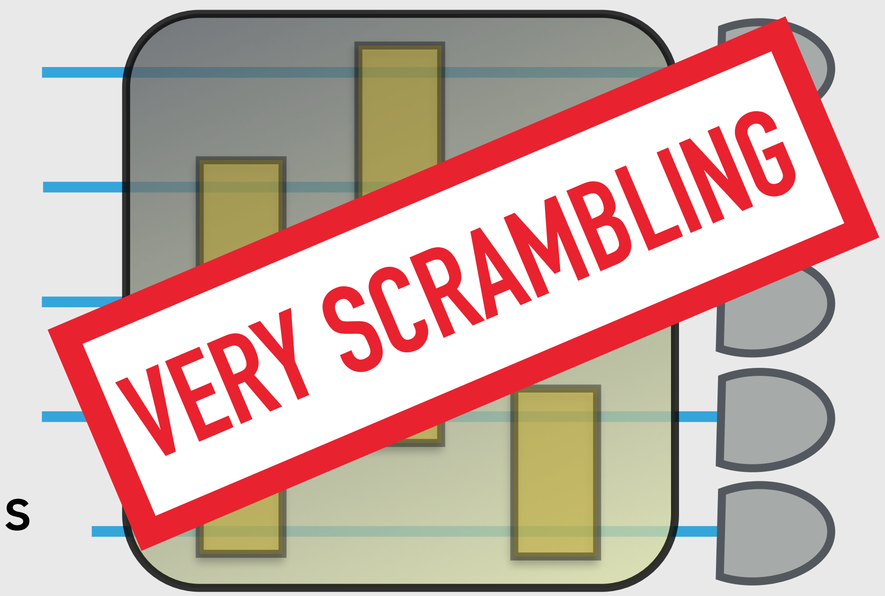
Quek, Stilck França, Khatri, Meyer, Eisert, Nature Phys 20, 1648 (2024)

- Set up a learning problem that can be solved by error mitigation
- Lower bound the sample complexity
- Construct circuits for which bound $D(\mathcal{C}'(|i\rangle\langle i|)||\mathcal{C}'(\mathbb{I}/2))$ is huge
  - Good enough to make the purity $\mathrm{tr}(\mathcal{C}'(|i\rangle\langle i|)^2)$ small
  - Clifford circuits for 2-designs in depth $\log^2(n)$

Cleve, Leung, Liu, Wang, Quant Inf Comp 16, 0721 (2016)



- Random circuits shift 'purity contribution' to larger weight Pauli words



Quek, Stilck França, Khatri, Meyer, Eisert, Nature Phys 20, 1648 (2024)

LESSON

- There are **strong obstructions against error mitigation** already at log-log depth

- Does not constrain **quantum error correction**
- New **intermediate** schemes?

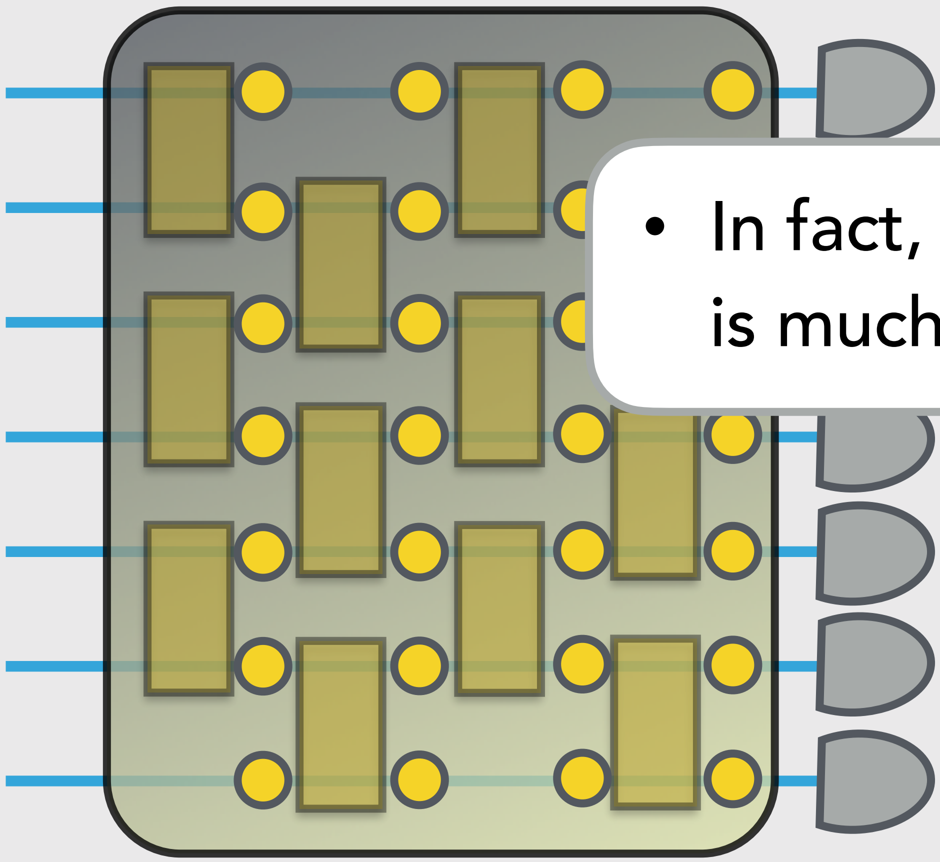
  Onorati, Kitzinger, Helsen, Ioannou, Werner, Roth, Eisert, arXiv:2403.04751 (2024)
  Seif, Cian, Zhou, Chen, Jiang, arXiv:2203.07309 (2022)
  Koczor, Phys Rev X 11, 031057 (2021)
  Huggins et al Phys Rev X 11, 041036 (2021)

- Does not **scale** and not work well **for all circuits**

  Compare also Gonzales-Garcia, Trivedi, Cirac, arXiv:2203.15632 (2022)

Quek, Stilck França, Khatri, Meyer, Eisert, Nature Phys 20, 1648 (2024)
Compare also Nietner, arXiv:2310.17716 (2023)

NON-UNITAL NOISE



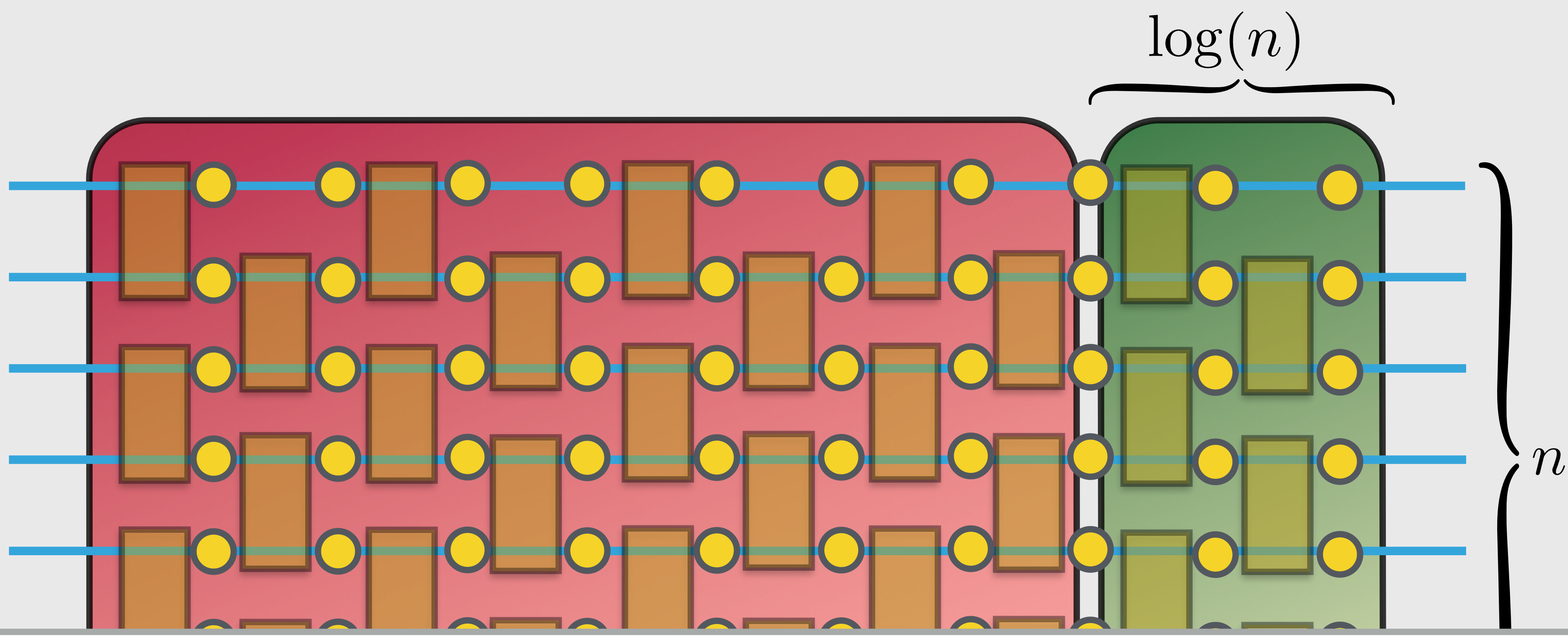
- In fact, the impact of **non-unital noise** is much more drastic than anticipated

Ben-Or, Gottesman, Hassidim, arXiv:1301.1995

Mele, Angrisani, Ghosh, Khatri, Eisert, Stilck França, Quek, arXiv:2403.13927 (2024)

Compare Fefferman, Ghosh, Gullans, Kuroiwa, Sharma, arXiv:2306.16659 (2023)

EFFECTIVE LOGARITHMIC DEPTH

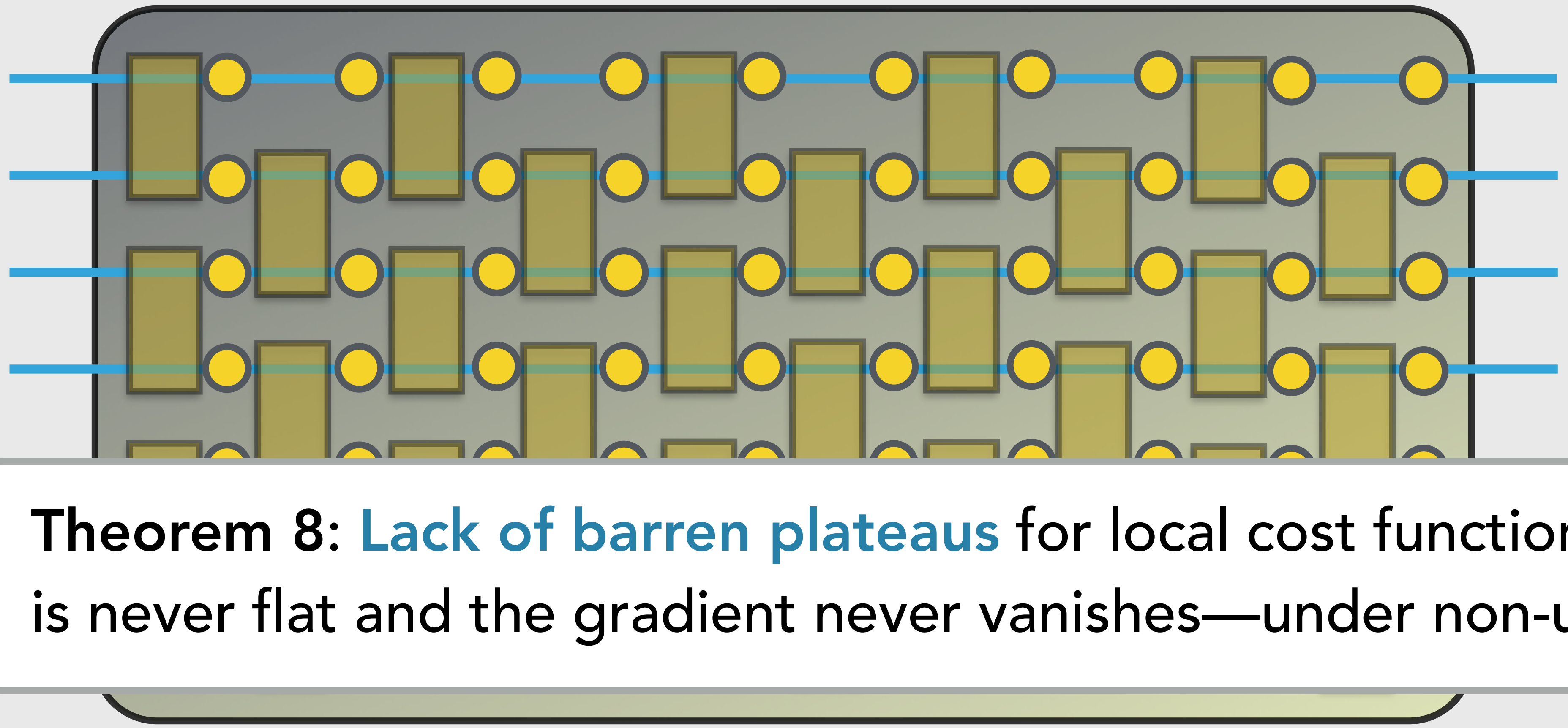


- **Theorem 7**: Deep random quantum circuits, under any uncorrected, possibly non-unital, noise, effectively get "truncated"

Mele, Angrisani, Ghosh, Khatri, Eisert, Stilck França, Quek, arXiv:2403.13927 (2024)
Compare also Fawzi, Mueller-Hermes, Shayeghi, ITCS (2022)

LACK OF BARREN PLATEAUS FOR NON-UNITAL NOISE

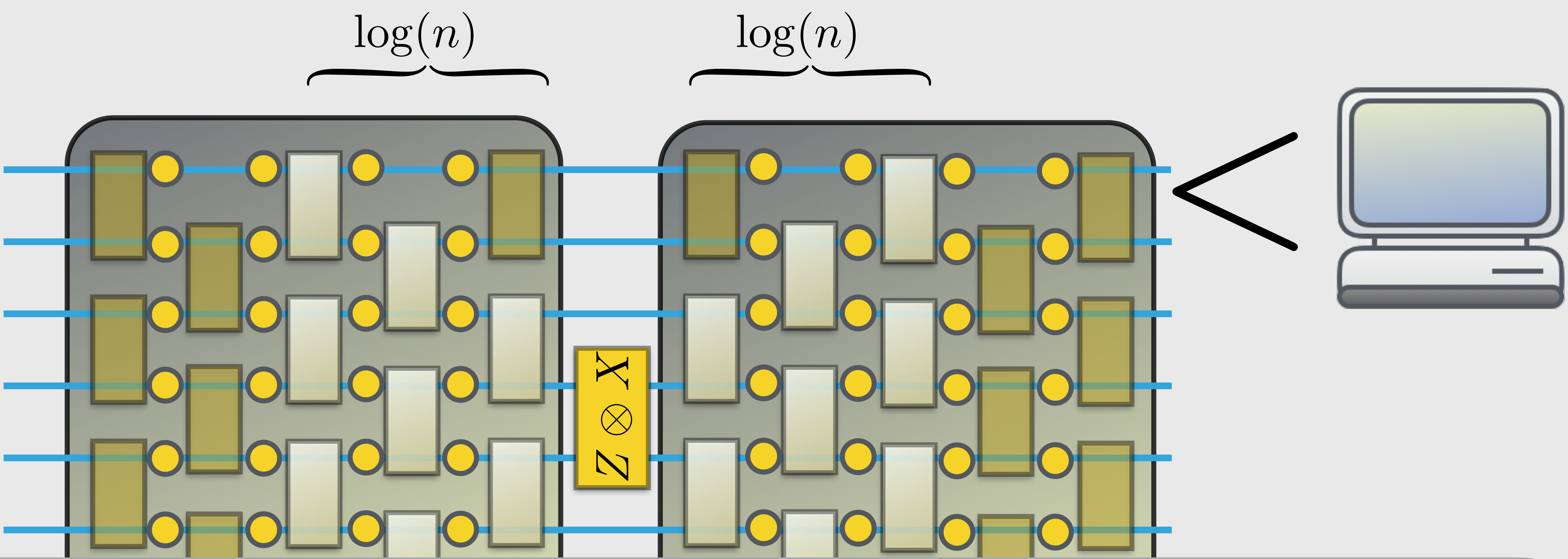
$$\min_{U_j} \mathrm{tr}(H\rho(U_1,\dots,U_m))$$
$$\mathrm{var}_{U_1,\dots,U_m}(C) = O(\exp(-n))$$



- **Theorem 8:** Lack of barren plateaus for local cost functions—cost landscape is never flat and the gradient never vanishes—under non-unital noise

Mele, Angrisani, Ghosh, Khatri, Eisert, Stilck França, Quek, arXiv:2403.13927 (2024)

EFFICIENT CLASSICAL SIMULATION

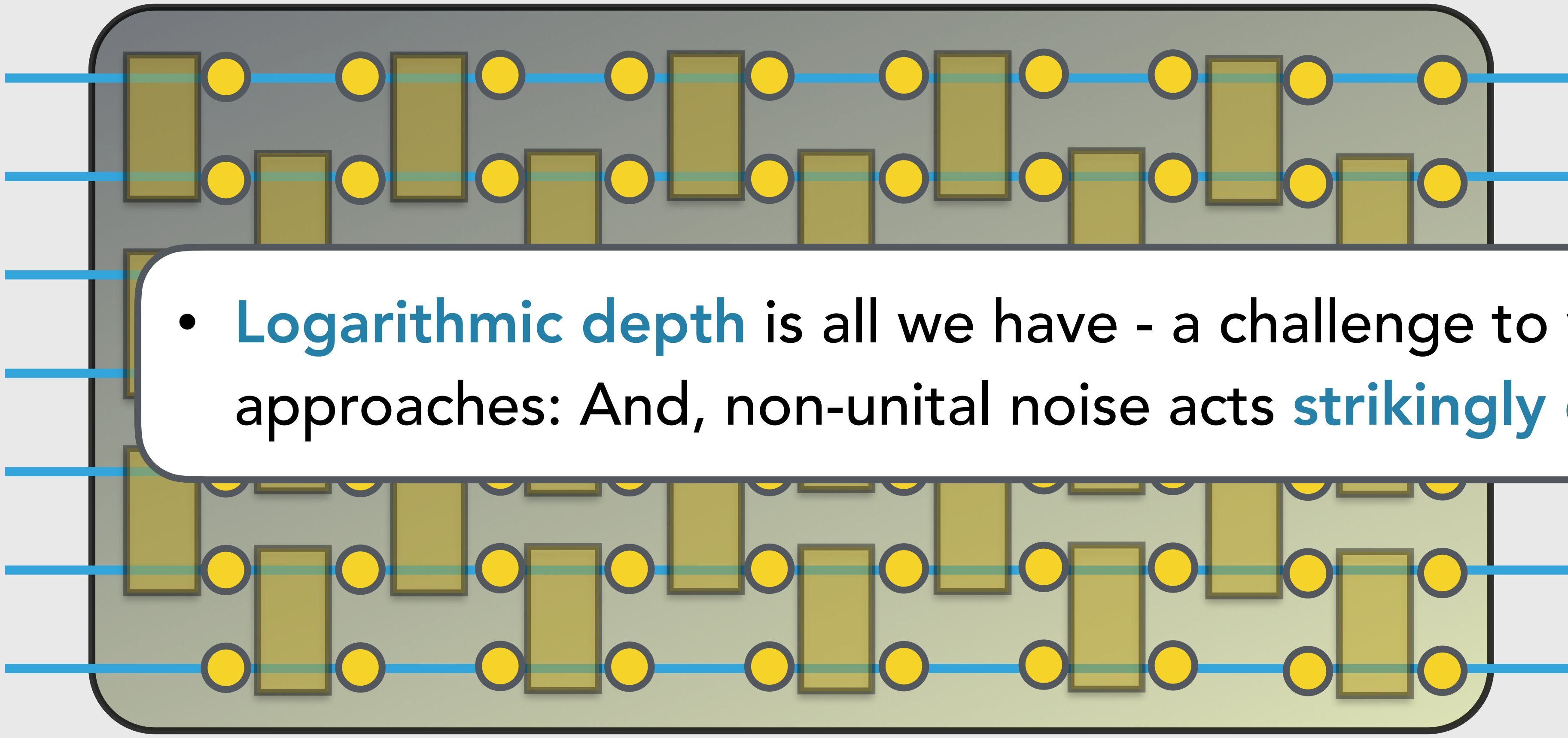


- **Theorem 9**: We can classically simulate on average expectation values of any observable to additive precision, with probability $1-\delta$, at any depth, with

$$O(\exp\left(\log^D(\delta^{-1}\varepsilon^{-2})\right)$$

Mele, Angrisani, Ghosh, Khatri, Eisert, Stilck França, Quek, arXiv:2403.13927 (2024)

AGAIN, LESSON



- **Logarithmic depth** is all we have - a challenge to variational QML approaches: And, non-unital noise acts **strikingly differently**

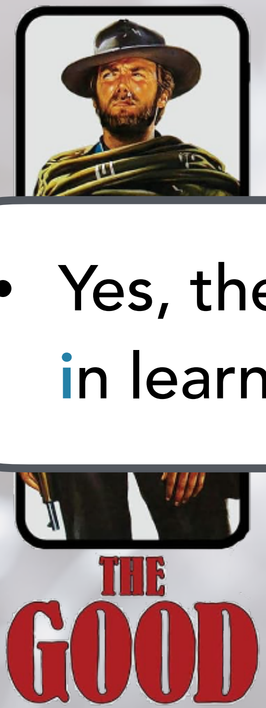
Mele, Angrisani, Ghosh, Khatri, Eisert, Stilck França, Quek, arXiv:2403.13927 (2024)

THE GOOD, THE BAD AND THE UGLY

Gil-Fuster, Eisert, Bravo-Prieto, Nature Comm 15,1 (2024)
Hangleiter, Roth, Fuksa, Eisert, Roushan, Nature Comm 15, 9595 (2024)
Recio-Armengol, Eisert, Meyer, arXiv:2406.13812 (2024)
Schreiber, Eisert, Meyer, Phys Rev Lett 131, 100803 (2023)
Sweke, Recio, Jerbi, Gil-Fuster, Fuller, Eisert, Meyer, Quantum (2024)
And others

THE GOOD, THE BAD AND THE UGLY

- Can quantum comput
meaningful **machine le**



- Yes, there are proven **separations in** learning and training

- How mu

Property	Problems studied in quantum computing	Problems solved by machine learning
classical performance	**low** – problems are carefully selected to be provably difficult for classical computers	**high** – machine learning is applied on an industrial scale and many algorithms run in linear time in practice
size of inputs	**small** – near-term algorithms are limited by small qubit numbers, while fault-tolerant algorithms usually take short bit strings	**very large** – may be millions of tensors with millions of entries each
problem structure	**very structured** – often exhibiting a periodic structure that can be exploited by interference	**"messy"** – problems are derived from the human or "real-world" domain and naturally complex to state and analyse
theoretical accessibility	**high** – there is a large bias towards problems about which we can theoretically reason	**shifting** – theory is currently been re-built around the empirical success of deep learning
evaluating performance	**computational complexity** – the dominant measure to assess the performance of an algorithm is asymptotic runtime scaling	**practical benchmarks** – machine learning research puts a strong emphasis on empirical comparisons between methods

Schuld, Killoran, arXiv:2203.01340 (2022)

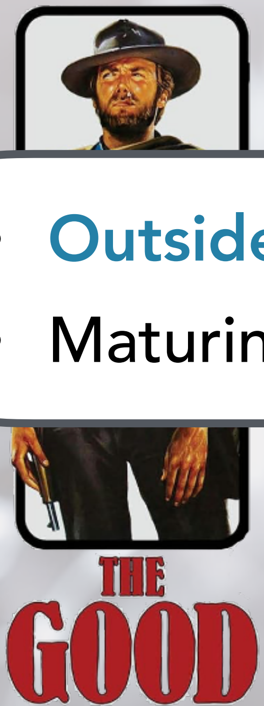
- Is **quantum advantage** the right aim?

- But there are **constant depth** advantages

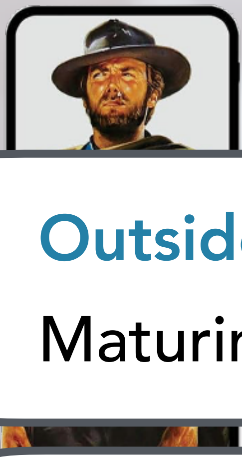
THE GOOD, THE BAD AND THE UGLY

- Can quantum computers assist in meaningful **machine learning tasks**?

- **Outside the box** considerations
- Maturing from perspective of classical AI



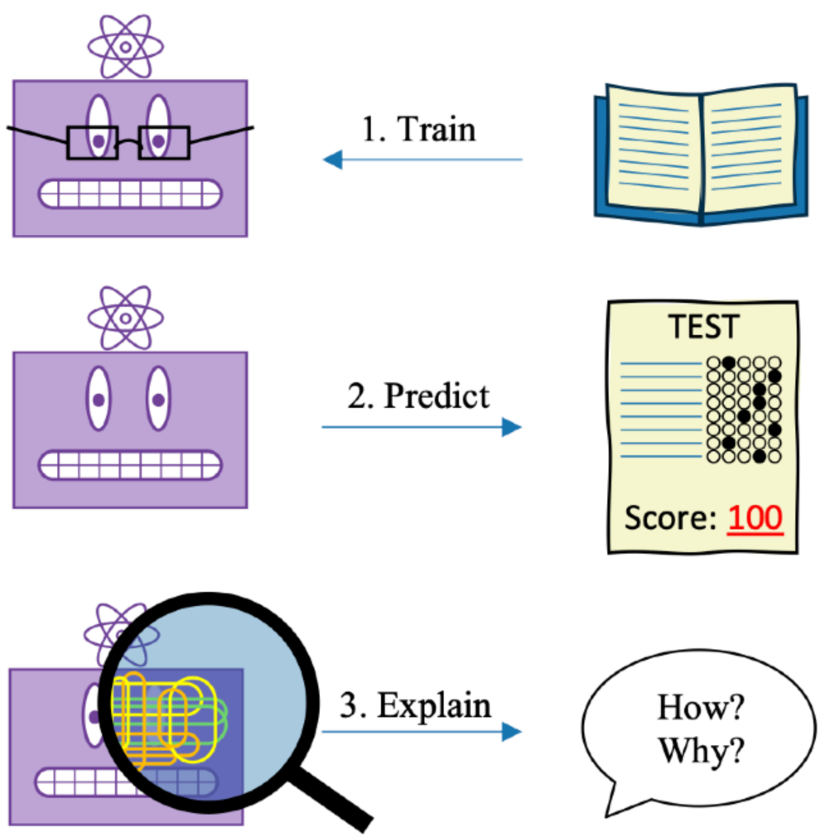
THE GOOD, THE BAD AND THE UGLY



- Can quantum computers assist in meaningful **machine learning tasks**?

- **Outside the box** considerations
- Maturing from perspective of classical AI

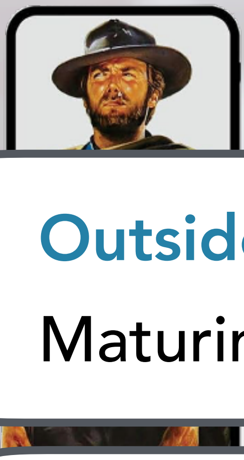
- **Explainable quantum AI:** "What is the role of each of the parameters in the model for classification?"



→ Jonas' talk

Gil-Fuster, Naujoks, Montavon, Wiegand, Samek, Eisert, next week (2024)

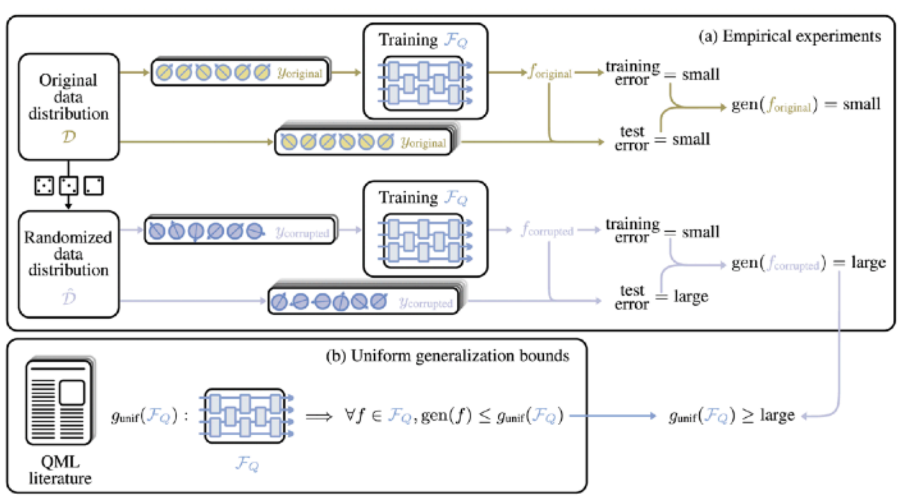
THE GOOD, THE BAD AND THE UGLY



- Can quantum computers assist in meaningful **machine learning tasks**?

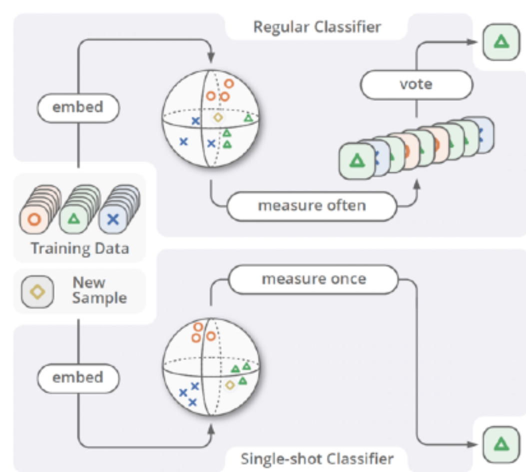
- **Outside the box** considerations
- Maturing from perspective of classical AI

- **Generalization:** Traditional approaches to generalization fail to explain the behavior QML models



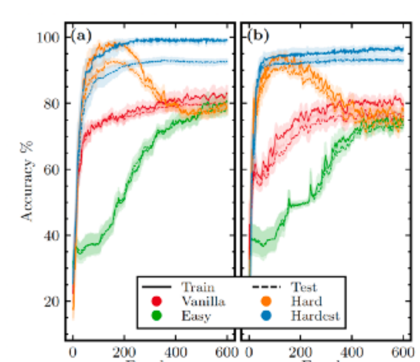
Gil-Fuster, Eisert, Bravo-Prieto, Nature Comm 15,1 (2024)

- **Single-shot** QML



Recio-Armengol, Eisert, Meyer, arXiv:2406.13812 (2024)

- Learning complexity **gradually:** Favorable inductive bias through curriculum learning and hard example mining



Recio-Armengol, Schreiber, Eisert, Bravo-Prieto, arXiv:2411.11954 (2024)

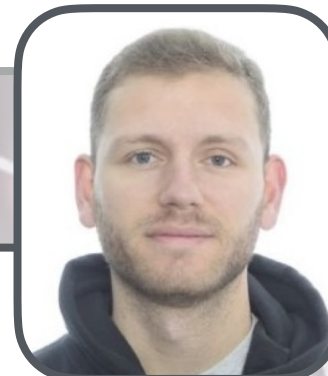
THE GOOD, THE BAD AND THE UGLY

- Can quantum computers assist in meaningful **machine learning tasks**?

- **Log depth** is all we have

THE GOOD, THE BAD AND THE UGLY

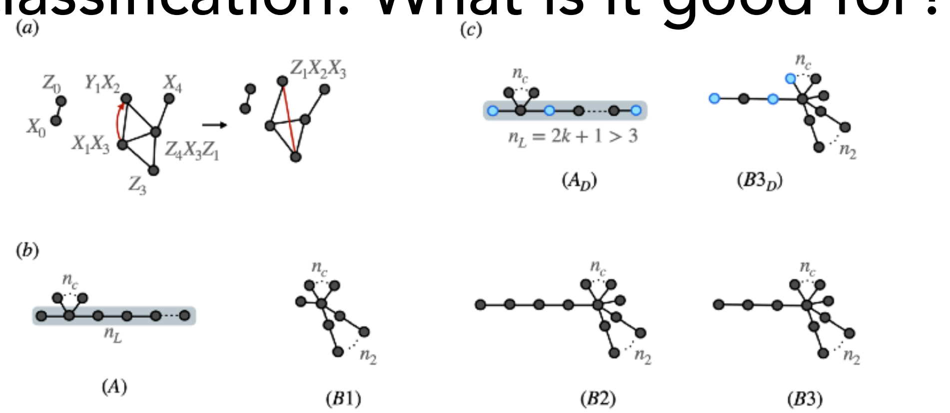
- Can quantum computers assist in meaningful **machine learning tasks**?



- **Expressivity:** How expressive are quantum kernels? → Elies' talk
- Can all quantum kernels be expressed as inner products of quantum feature states?

Gil-Fuster, Eisert, Dunjko, Mach Learn Sc Tech 5, 025003 (2024)

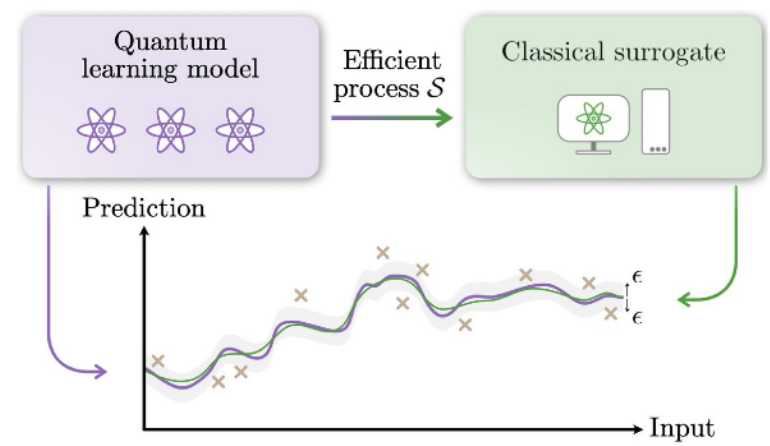
- **Pauli Lie algebras:** We have a full classification: What is it good for?



Aguilar, Cichy, Eisert, Bittel, arXiv:2408.00081 (2024)

- **Quantum data**

- **Classical surrogates** and dequantization



- **Trainability** vs **simulatability**

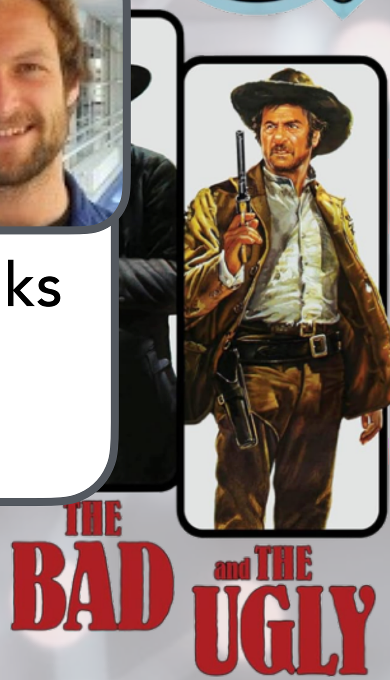
Schreiber, Eisert, Meyer, Phys Rev Lett 131, 100803 (2023)
Sweke, Recio, Jerbi, Gil-Fuster, Fuller, Eisert, Meyer, Quantum (2024)
Gil-Fuster, Gyurik, Pérez-Salinas, Dunjko, arXiv:2406.07072 (2024)

THE GOOD, THE BAD AND THE UGLY

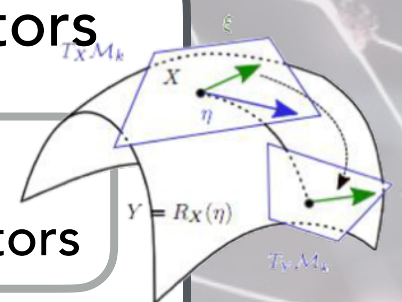
- Highly fruitful: **Learning theory/property testing**

Chen, Eisert, Phys Rev Lett 132, 220201 (2024)
Bertoni, Haferkamp, Hinsche, Ioannou, Eisert, Pashayan, Phys Rev Lett 133, 020602 (2024)
Denzler, Mele, Derbyshire, Guaita, Eisert, Phys Rev Lett 133 (2024)
Raza, Caro, Eisert, Khatri, arXiv:2406.04250 (2024)
Bittel, Mele, Eisert, Leone, arXiv:2409.17953 (2024)
Mele, Mele, Bittel, Eisert, Giovannetti, Lami, Leone, Oliviero, arXiv:2404.03585 (2024)
Bittel, Mele, Eisert, Leone, arXiv:2405.01431 (2024)
Teng, Samajdar, Van Kirk, Wilde, Sachdev, Eisert, Sweke, Najafi, arXiv:2406.00193 (2024)
Caro, Eisert, Hinsche, Ioannou, Nietner, Sweke, arXiv:2410.23969 (2024)

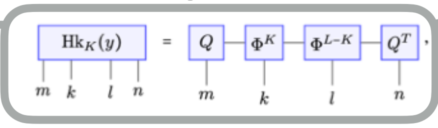
→ Sumeet's and Alex' talks



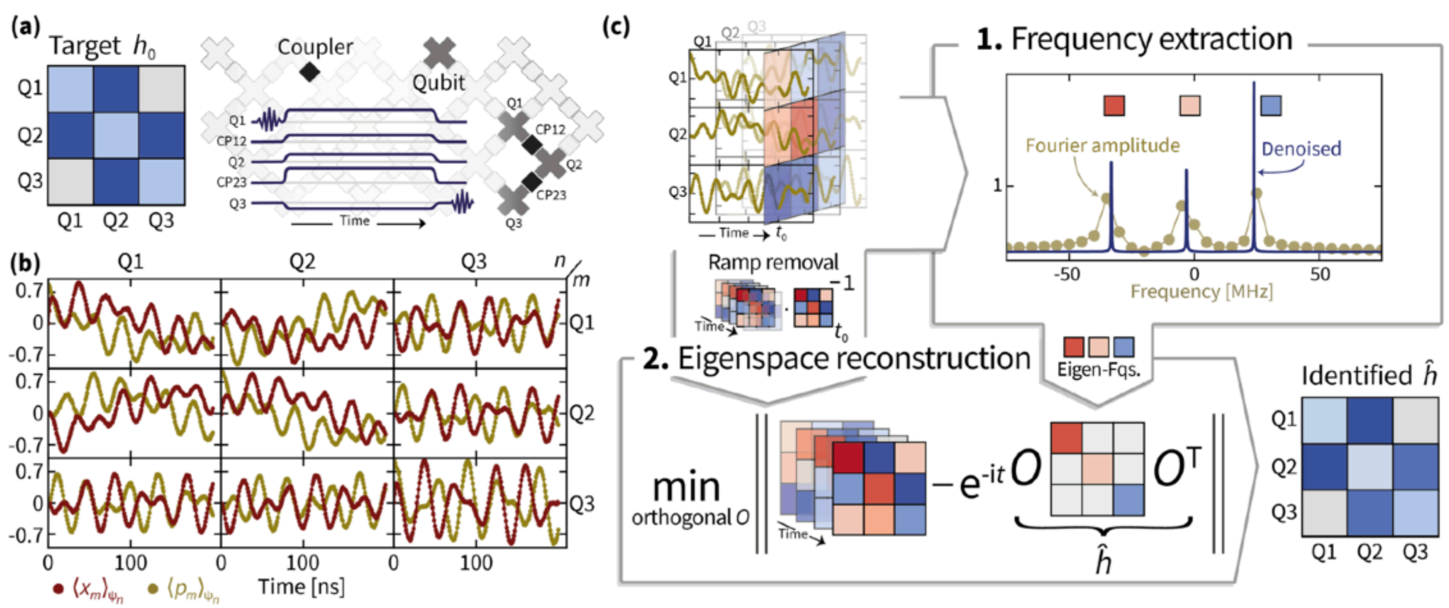
- **Hamiltonian learning** for analog quantum simulators



- New superresolution method tensorESPRIT for eigenvalues



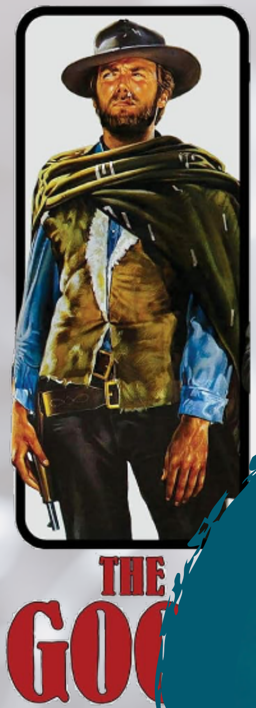
- Manifold optimization over $O(n)$ for eigenvectors



Hangleiter, Roth, Fuksa, Eisert, Roushan, Nature Comm 15, 9595 (2024)

THE GOOD, THE BAD AND THE UGLY

- Can quantum computers assist in meaningful **machine learning tasks**?



Definitely Maybe

THANKS FOR YOUR ATTENTION

erc Advanced Grant

POSTDOCS AVAILABLE