Application-Motivated, Holistic Benchmarking of a Full Quantum Computing Stack

Dan Mills

Contents

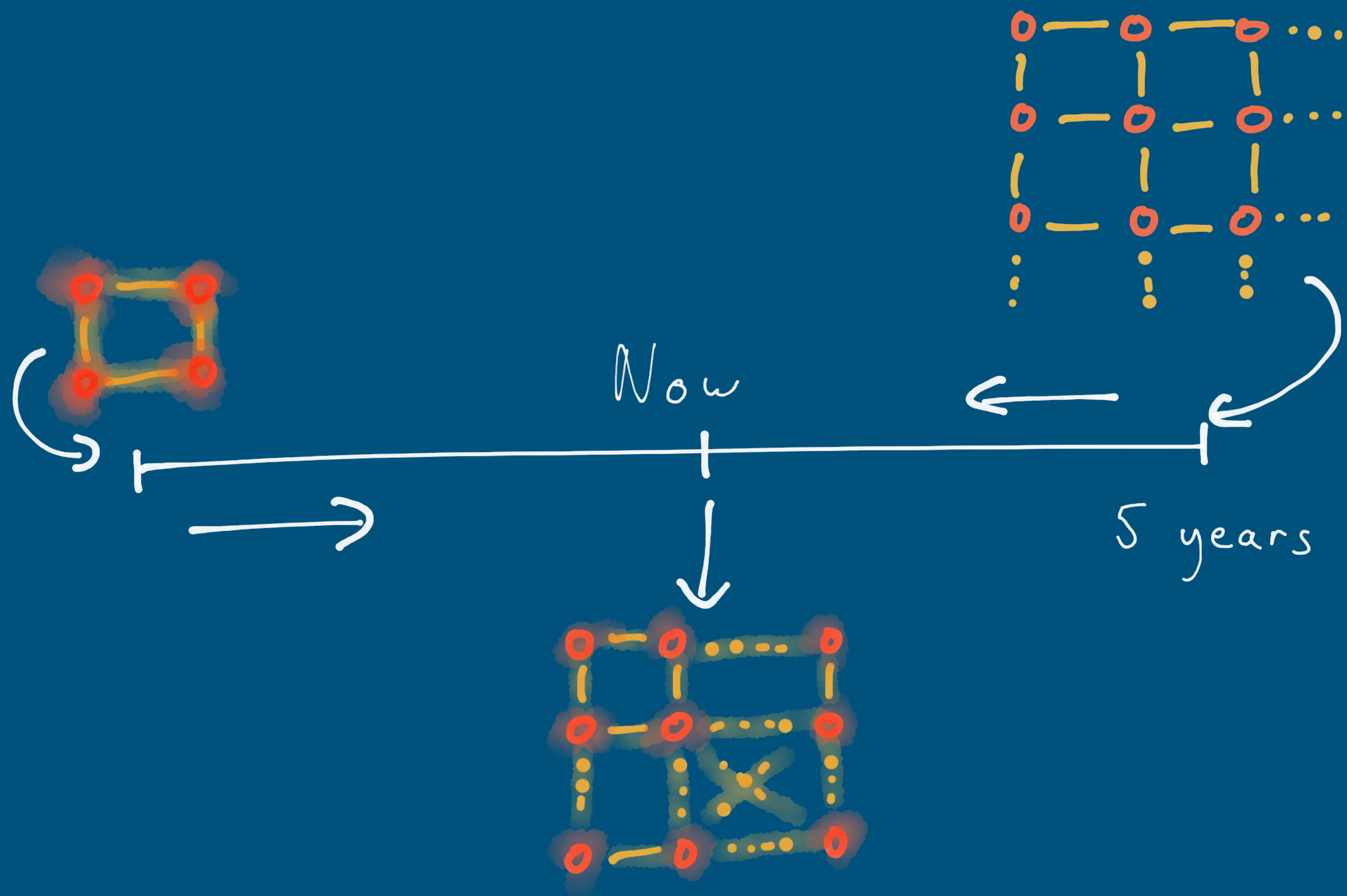
- The domain we will be benchmarking
  - Motivation
  - What's been done
- Benchmarking the full stack
  - Our philosophy
  - Some results
- An expansion of the stack
  - Error-mitigation

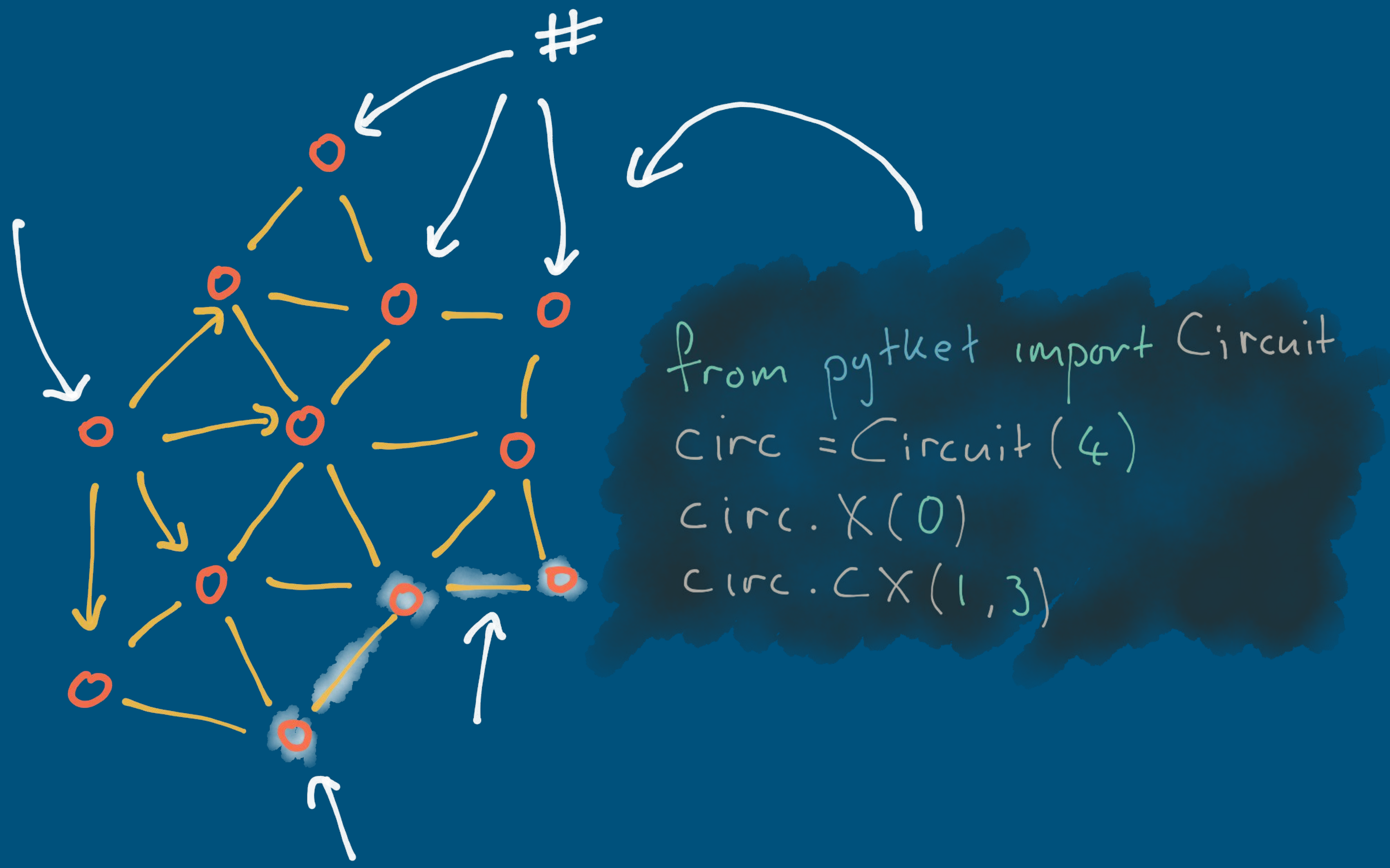
Benchmarking Near-Term Quantum Computers

A different beast

We will discuss:

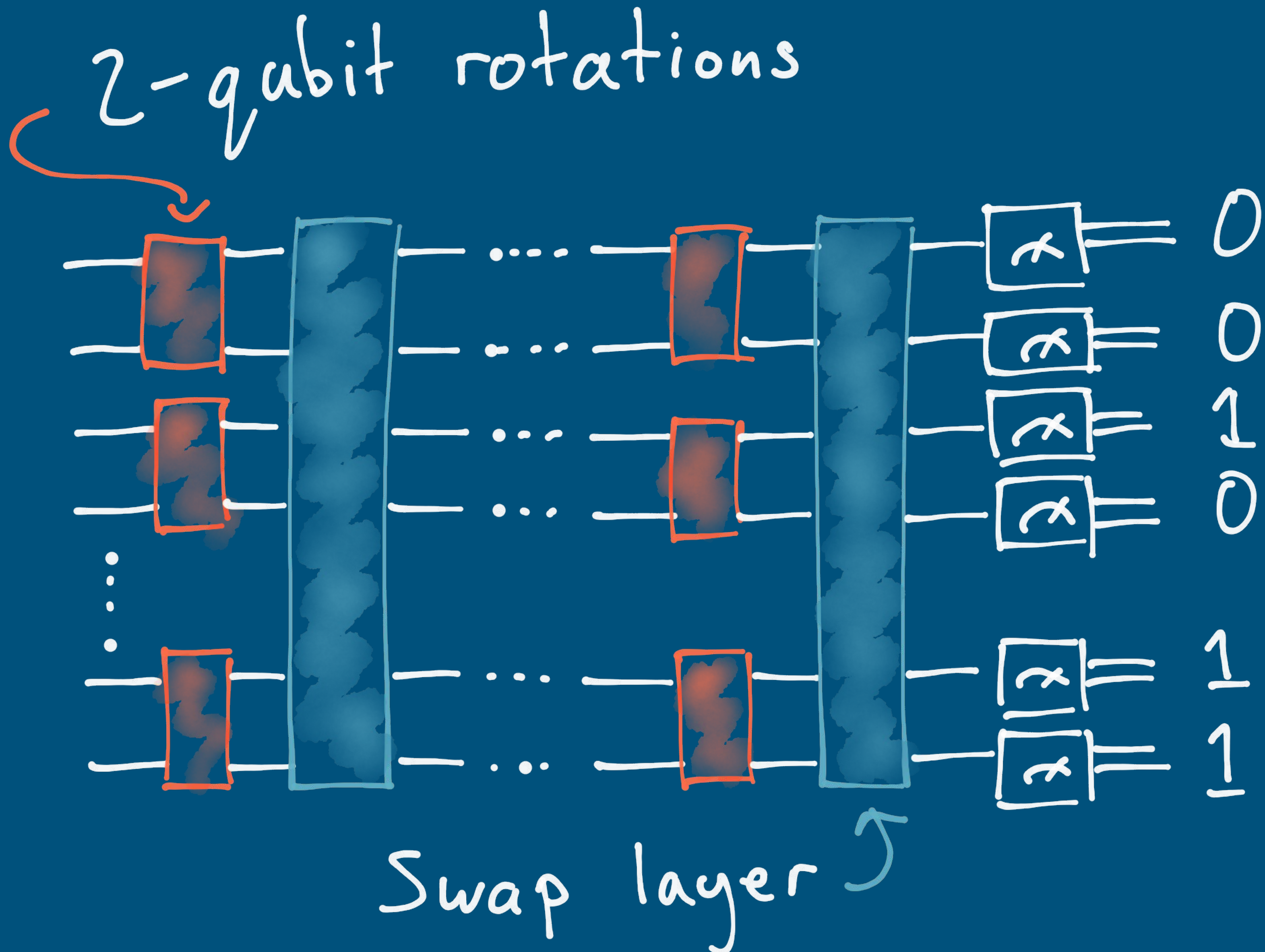
- What should benchmarking here do?
- Which schemes are available to do it now?

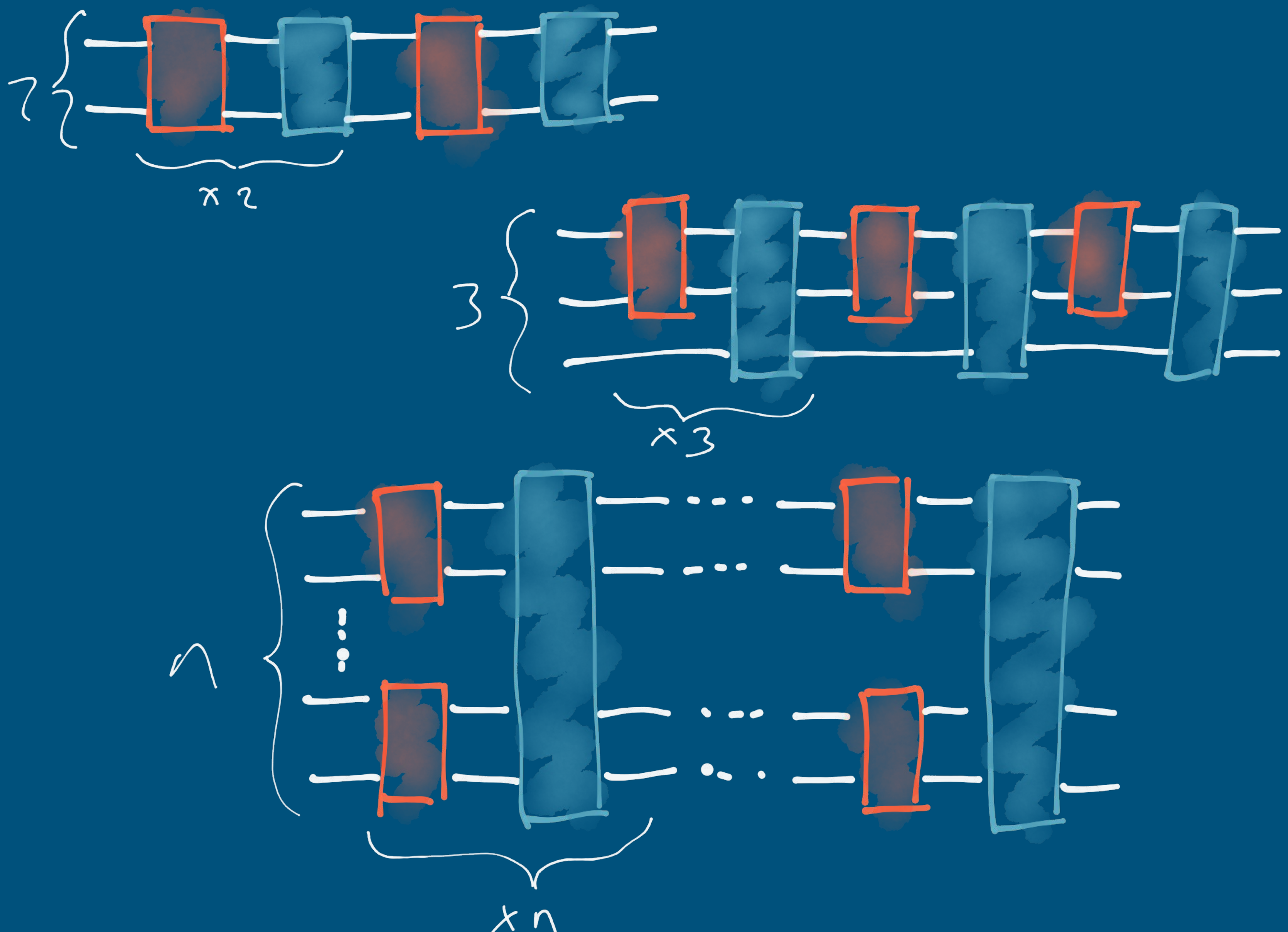

Now
5 years

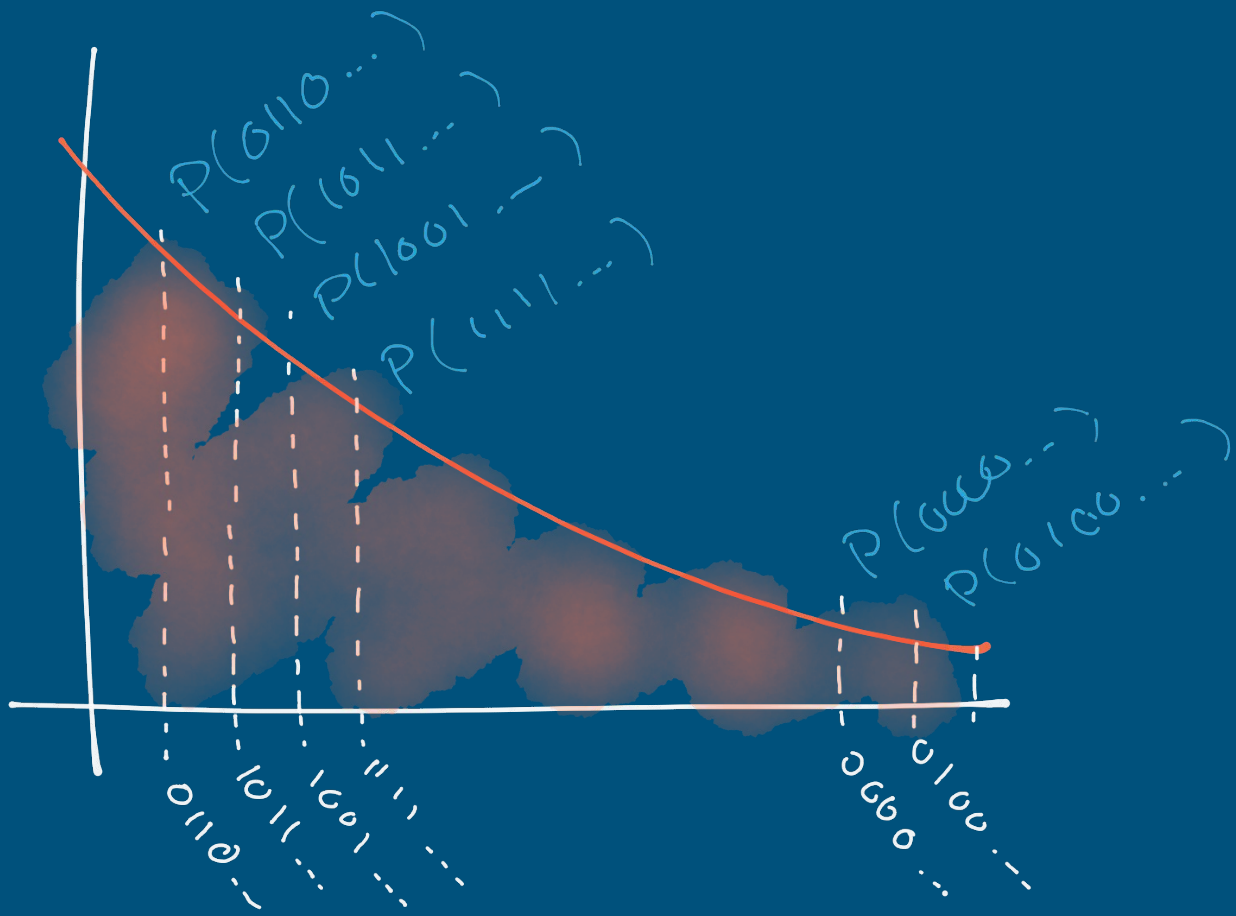

#
from pytket import Circuit
circ = Circuit(4)
circ.X(0)
circ.CX(1,3)

We wish our benchmarks to...

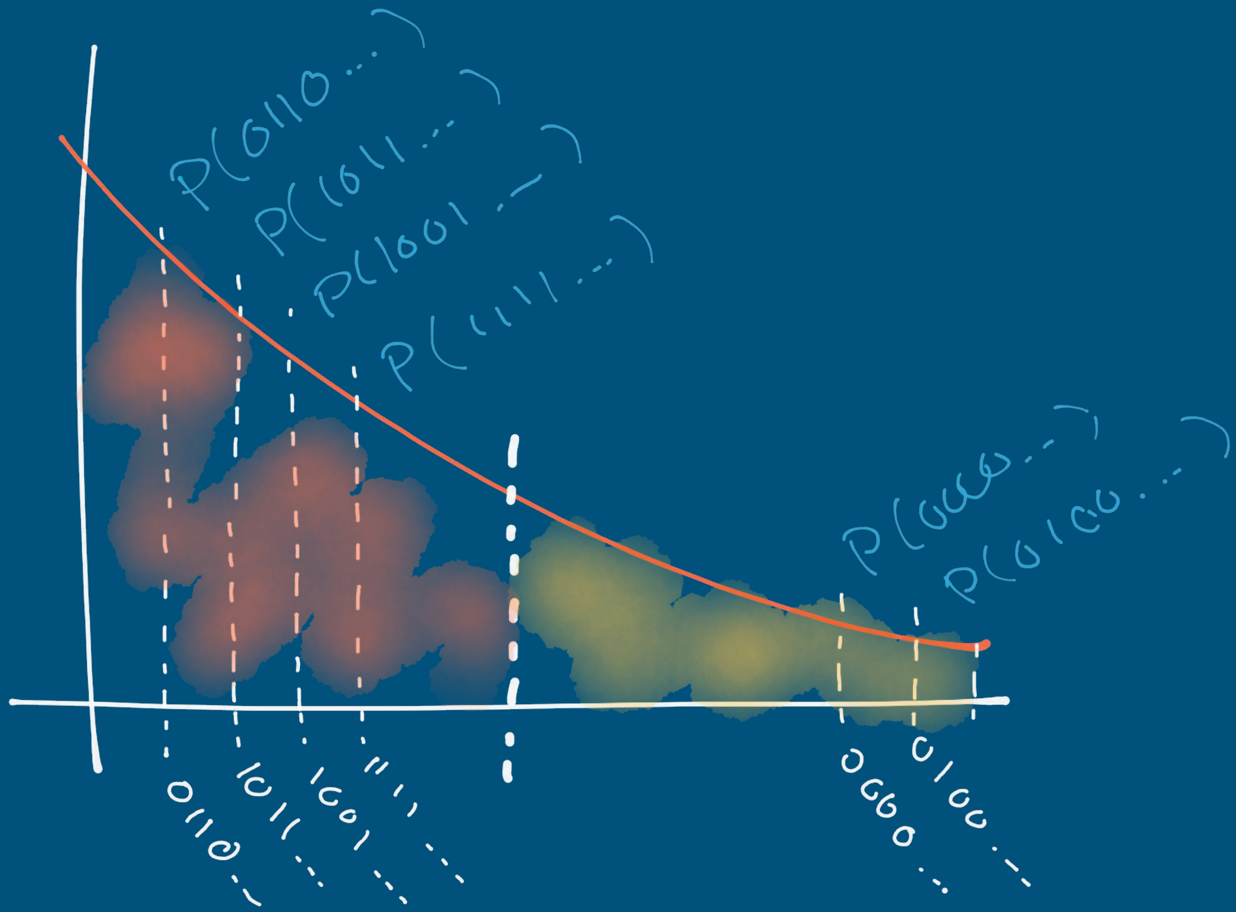
- **Holistic:** Measure the performance of the device in its entirety, rather than proxies of practical performance like gate fidelity.
- **Full-stack:** Include all contributions to the performance.
- **Application-motivated:** Give predictions of the performance of the stack in practice.

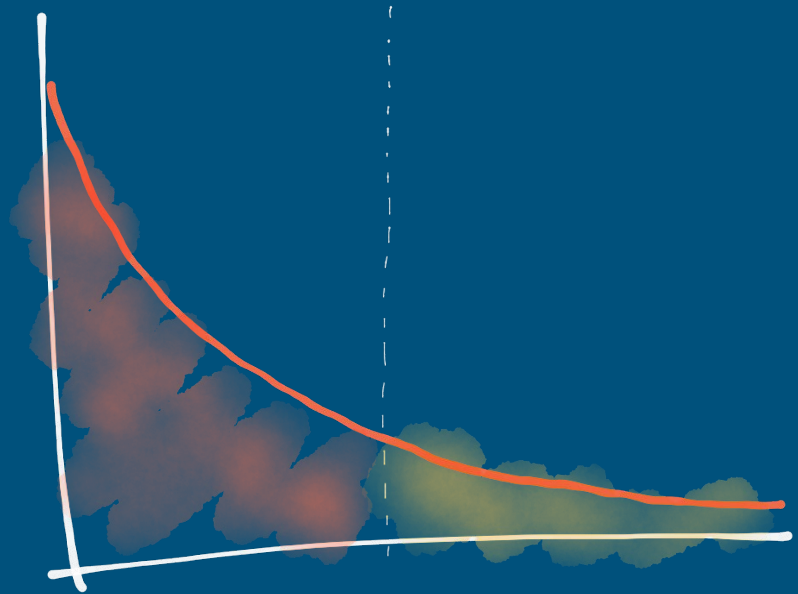

2-qubit rotations
0
0
1
0
1
1
Swap layer

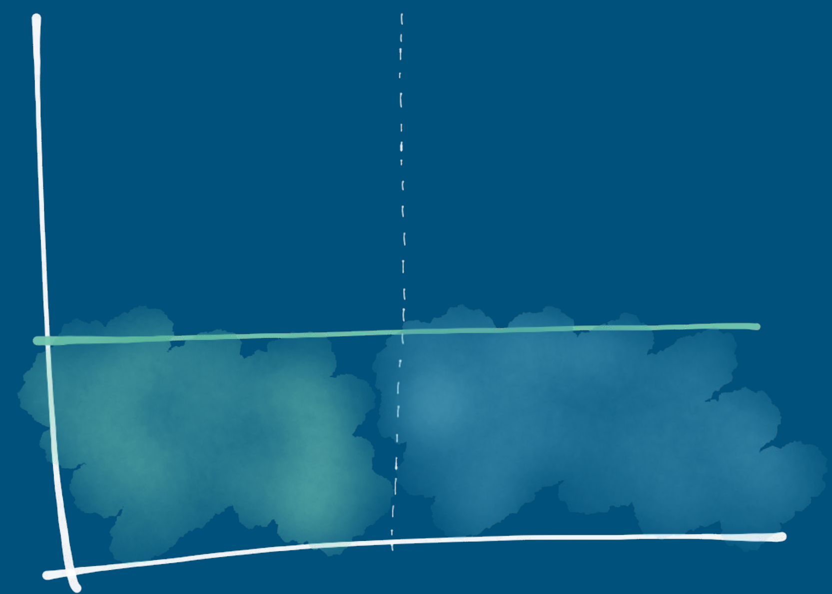

2
x 2
3
x 3
n
x n

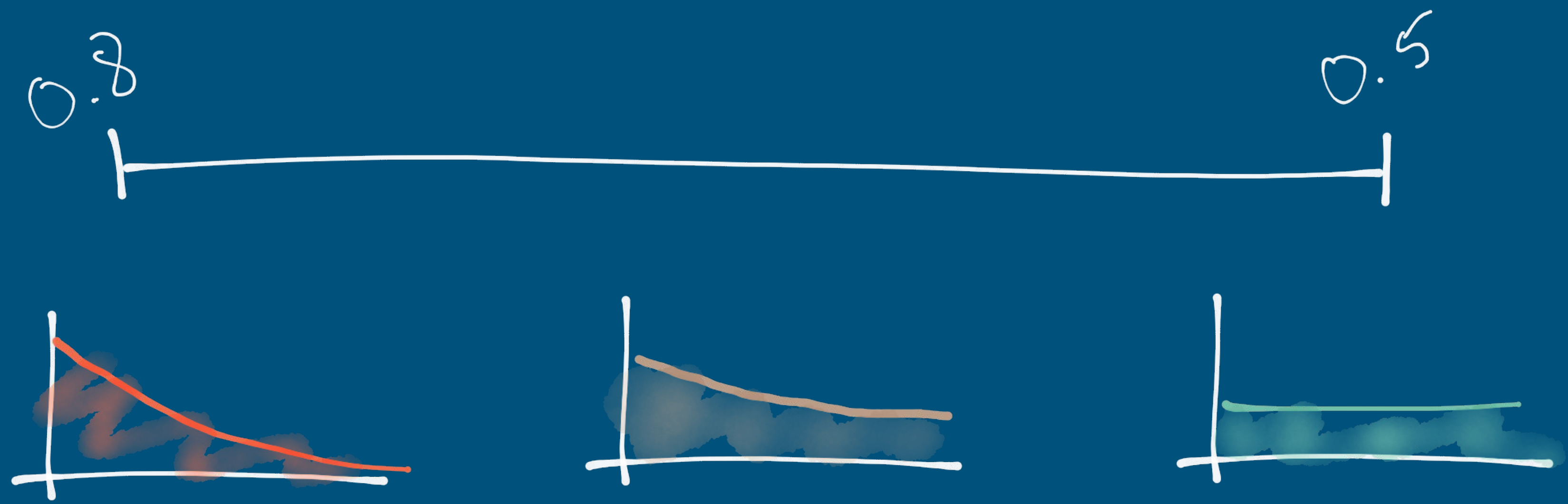


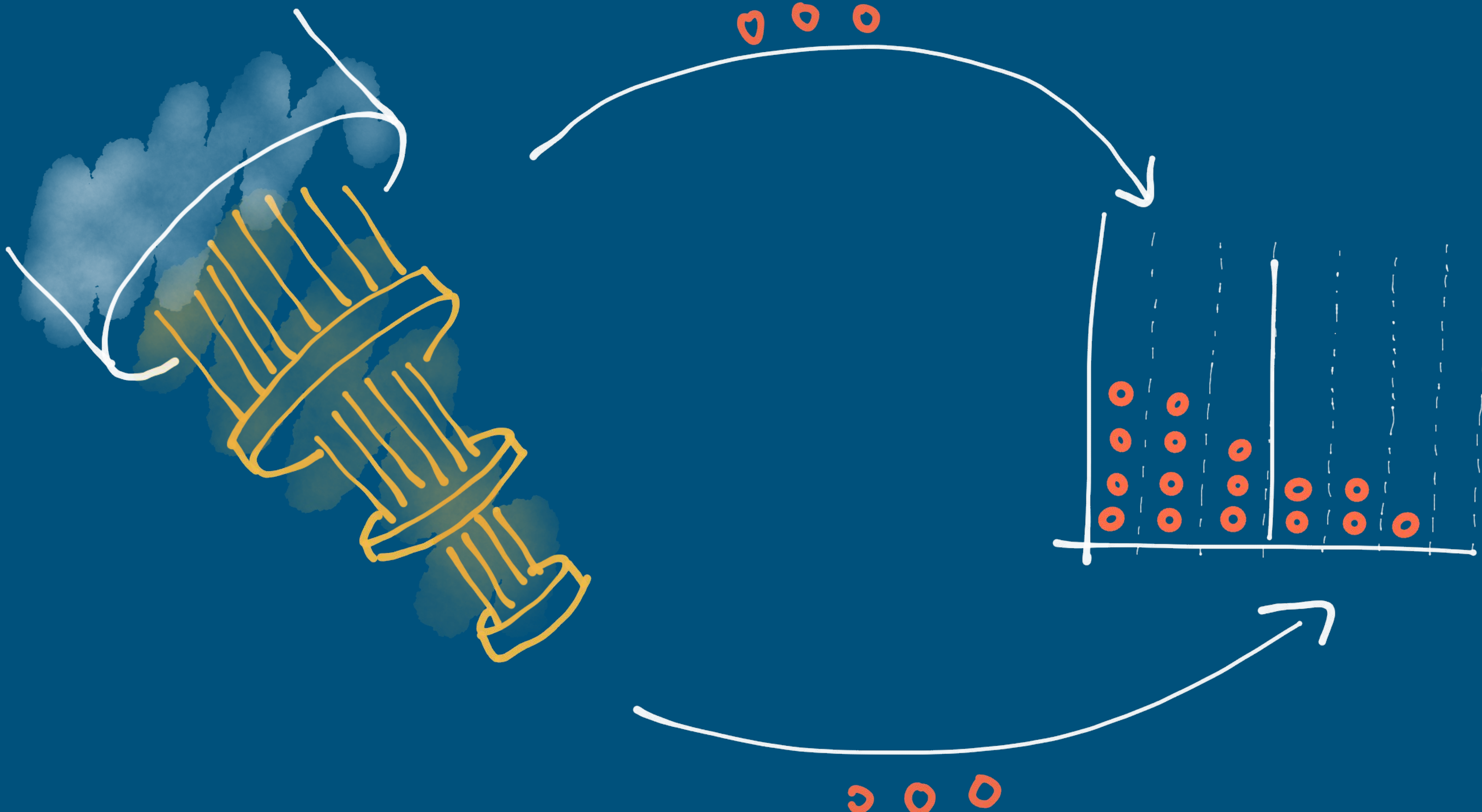
P(0110...)
P(1011...)
P(1001...)
P(1111...)
P(0000...)
P(0100...)
0110...
1011...
1001...
1111...
0000...
0100...

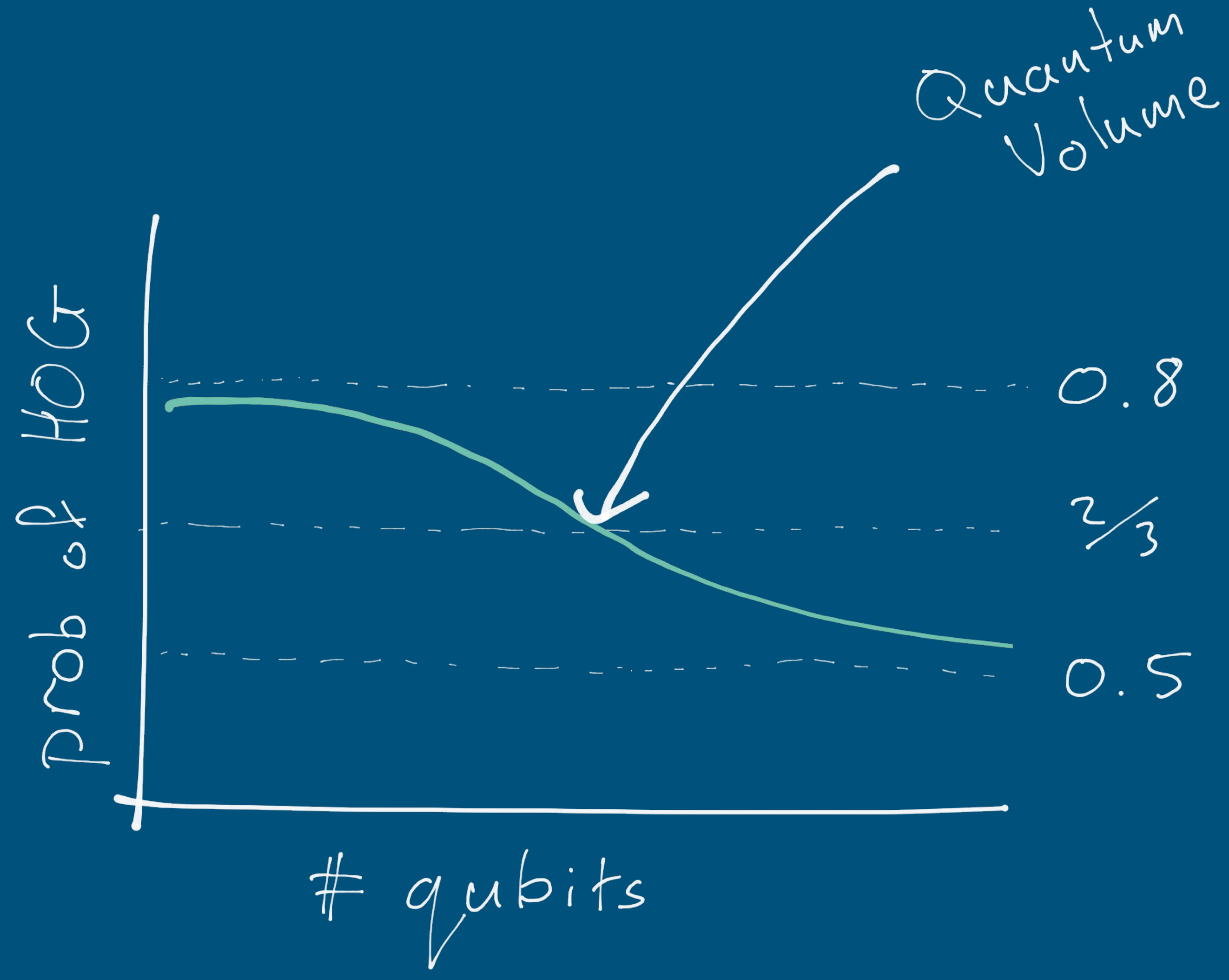

P(0110...)
P(1011...)
P(1001...)
P(1111...)
P(0000...)
P(0100...)
0110...
1011...
1001...
1111...
0000...
0100...



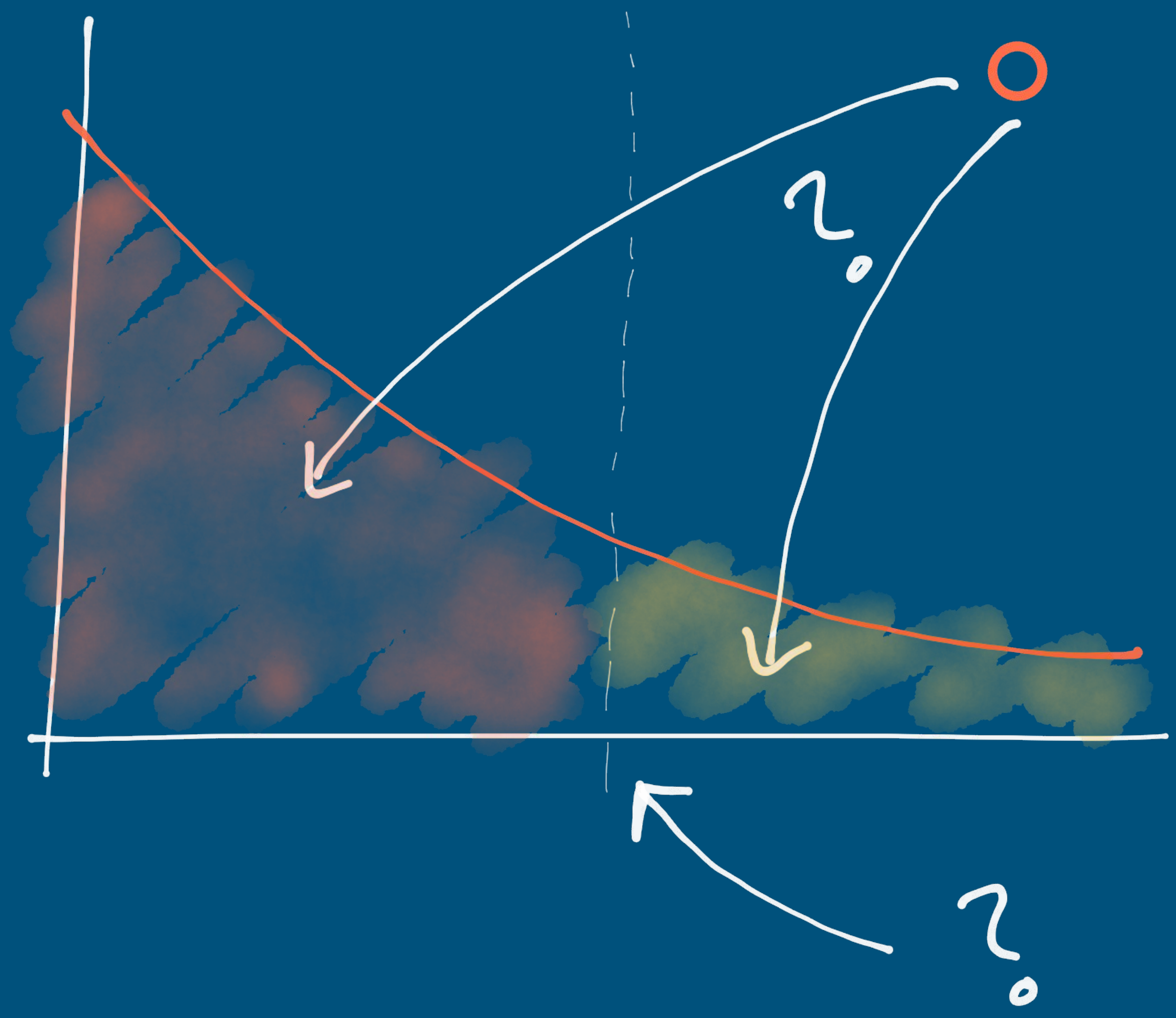



0.8
0.5





Quantum Volume
prob of HOG
0.8
2/3
0.5
qubits



Heavy Output Generation Probability

$$\mathrm{HOG}\left(D_C, p_C\right) = \sum_{x\in[0,1]^n} D_C\left(x\right)\delta_C\left(x\right)$$

- Cannot be used to bound the l1 distance.
- Only polynomially many single output probabilities are required, allowing the utilisation of Feynman simulators.
- Calculating probabilities takes exponential time.

Pros and Cons

Advantages:

- Gives one number to assess performance.
- Sample efficient.
- A test of general purpose, programmable quantum computers.
- Strong complexity theoretic foundations

Disadvantages:

- Does not give insights into performance in practice.
- Does not teach us how to make improvements at different stack layers.

Benchmark of Practical Performance

Application-Motivated, Holistic Benchmarking of a Full Quantum Computing Stack

We aim for our benchmarks to indicate

- The best complete stack to use
- The applications where the stack performs best

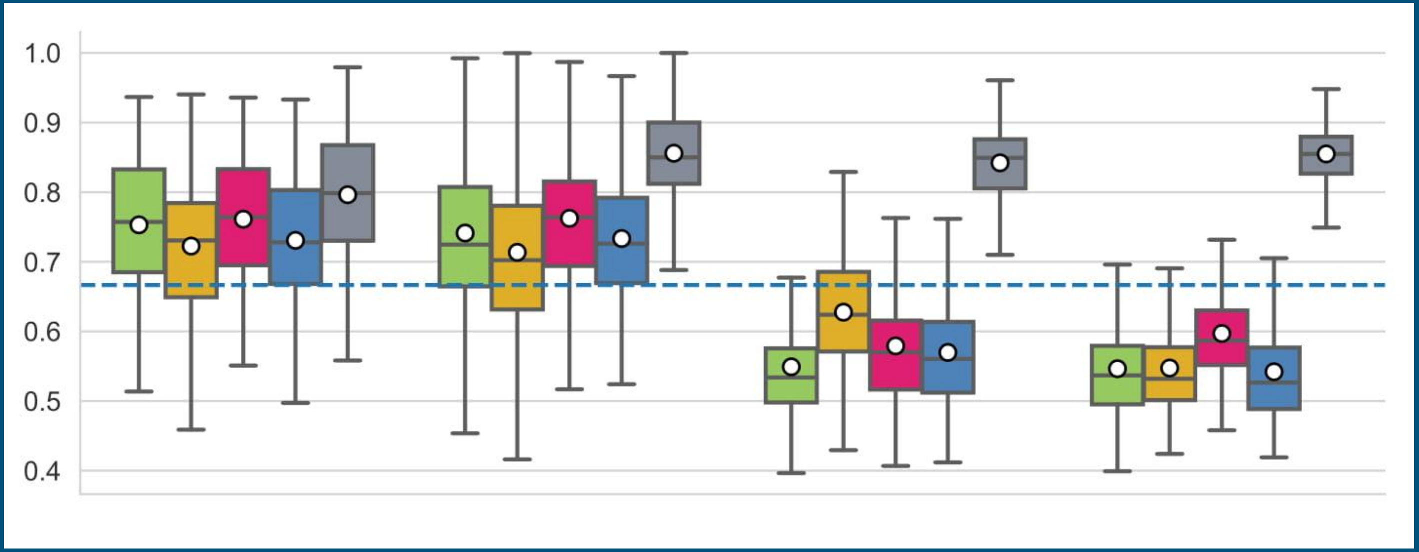
Circuits

- A minimal benchmark suite, rather than a collection of circuits (no lack of coverage, or unnecessary repeated coverage).
- Motivated by near term applications, but not particular instances of near term applications.
- Avoid bias against one architecture in particular.

Figures of Merit

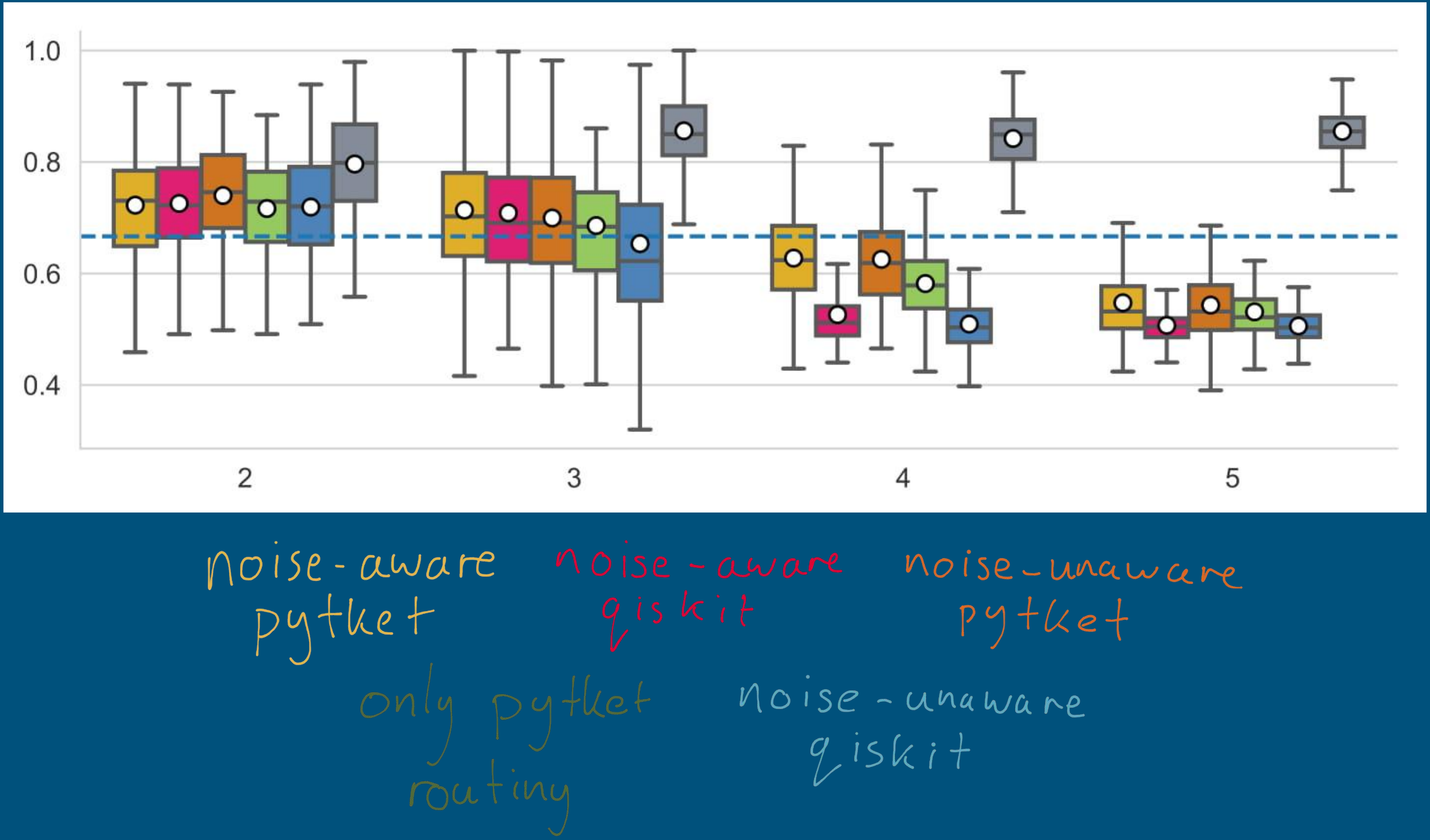
- Continuous figures of merit.
- Estimating figures of merit requires ideal outcome probabilities.
- Scaling to tens or hundreds of qubits will be challenging in general.
- Improvements in the time to perform benchmarks can be made if the circuits and figures of merit are developed jointly.

QV - Device Comparison

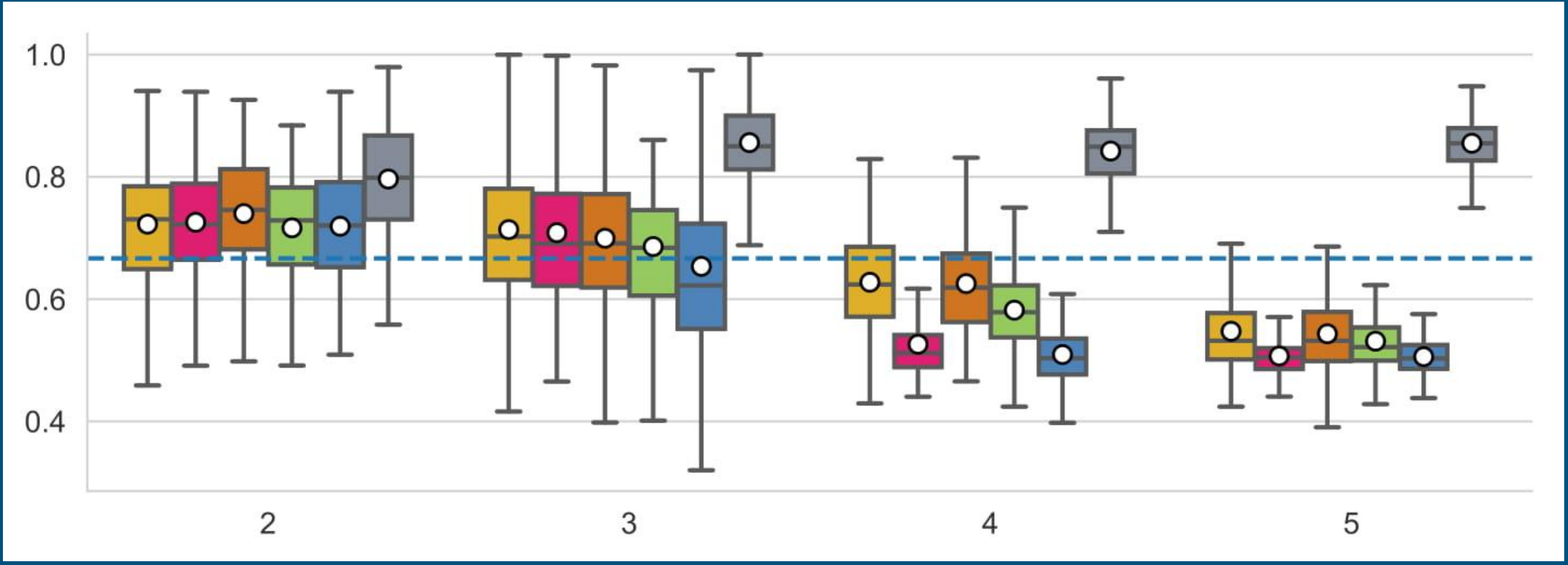


Singapore Yorktown Melbourne Ourense

QV - Melbourne, Compiler Comparison



QV - Melbourne, Compiler Comparison



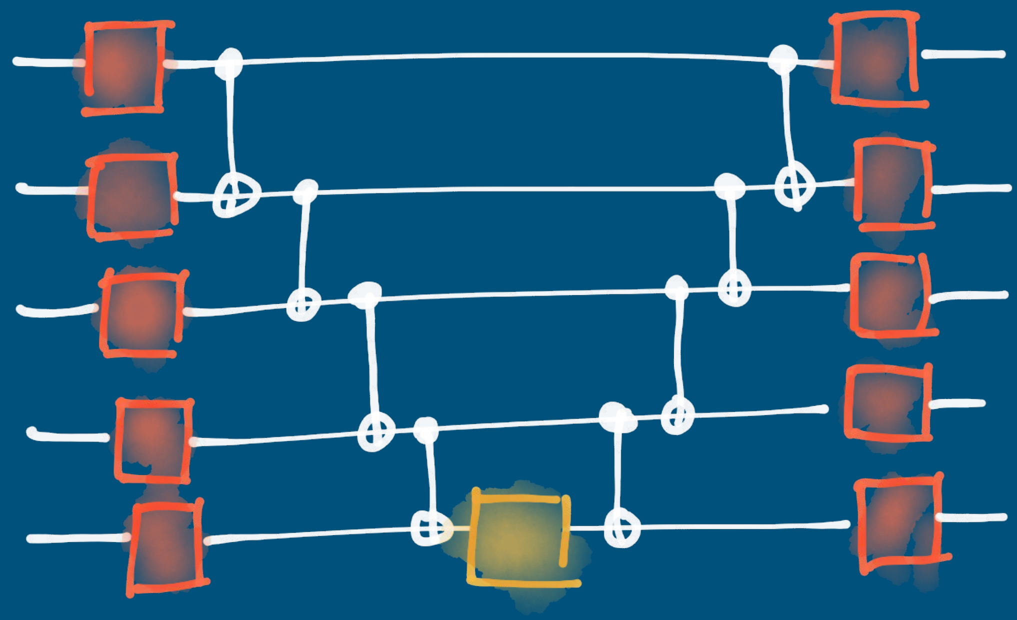
noise-aware pytket

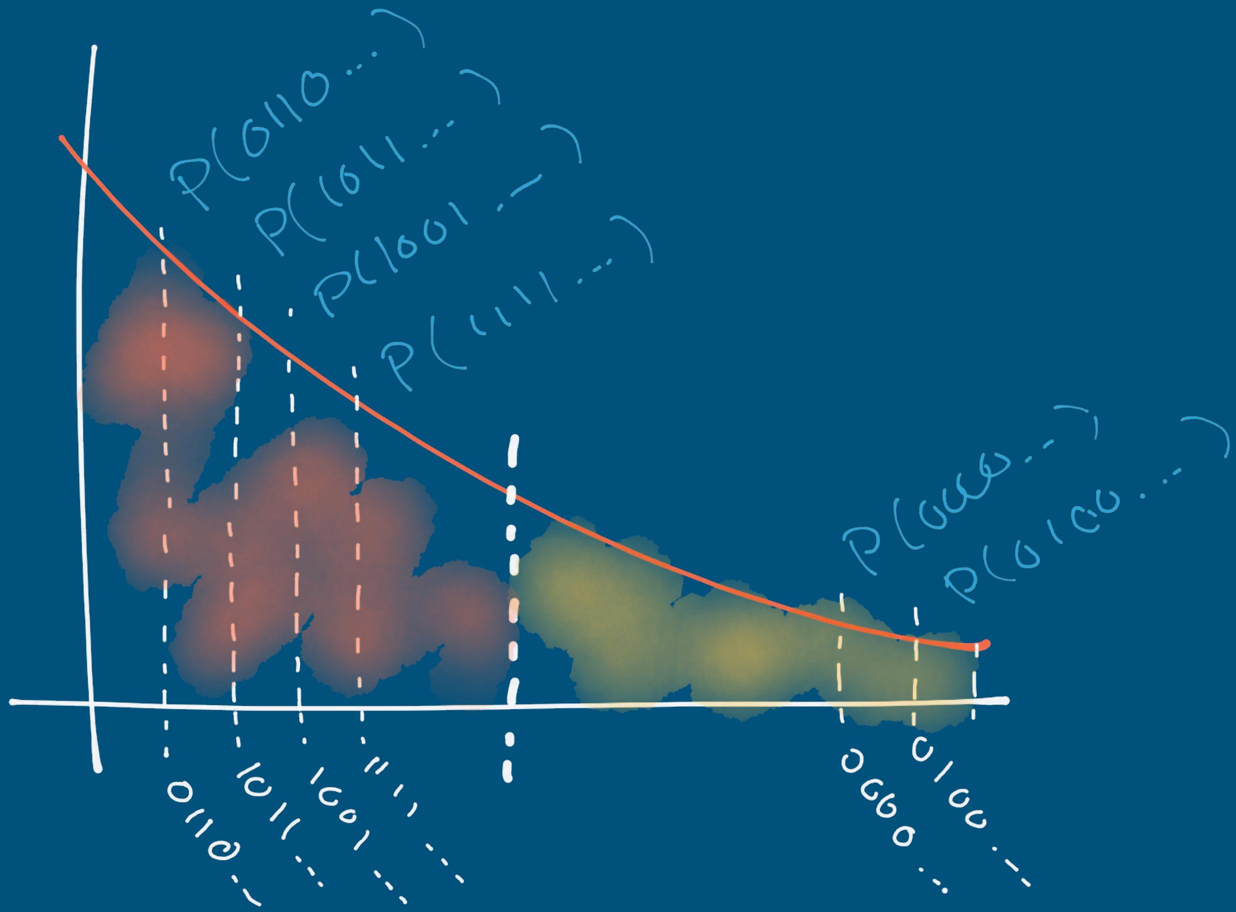
noise-aware qiskit

noise-unaware pytket

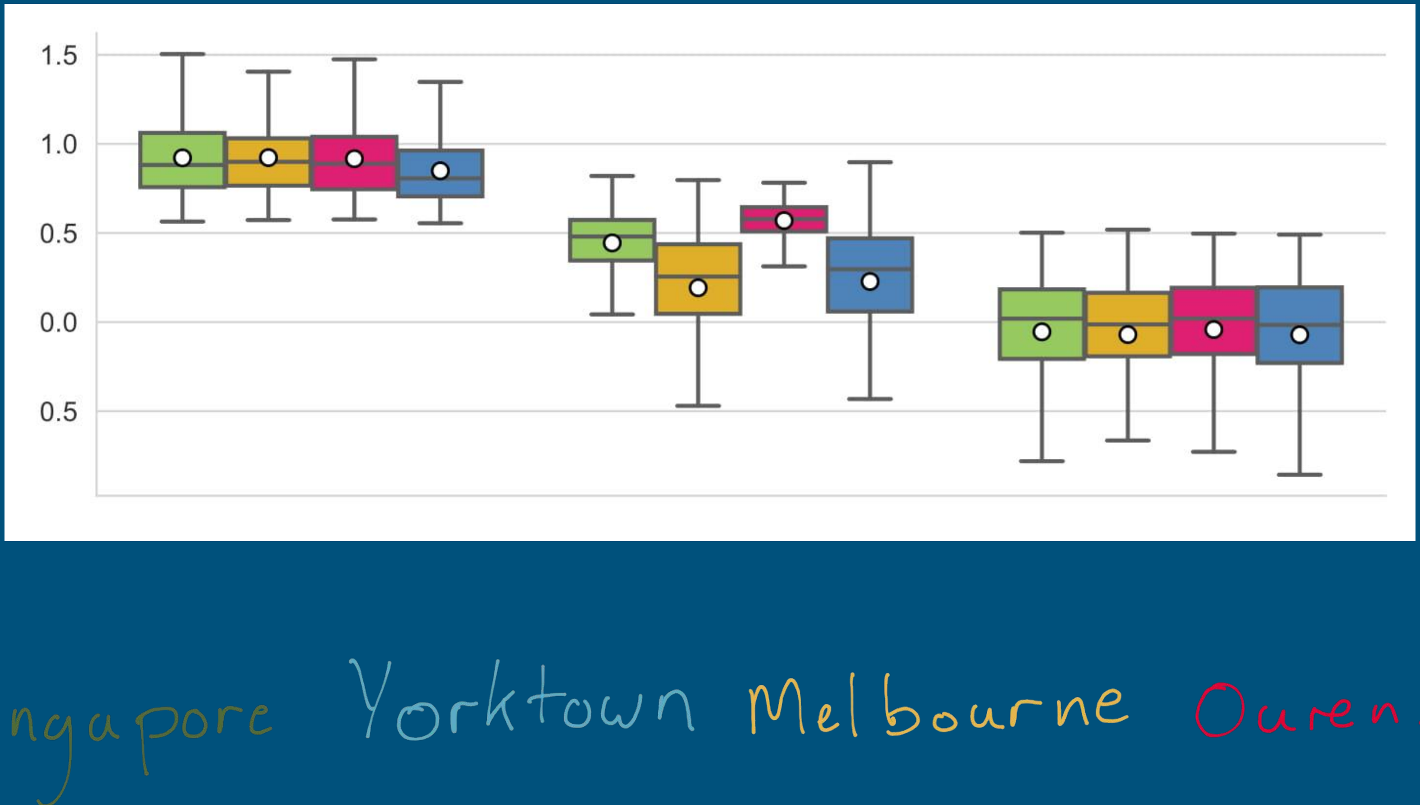
only pytket routing

noise-unaware qiskit




P(0110...)
P(1011...)
P(1001...)
P(1111...)
P(0000...)
P(0100...)
0110...
1011...
1001...
1111...
0000...
0100...

Pauli Gadgets Device Comparison



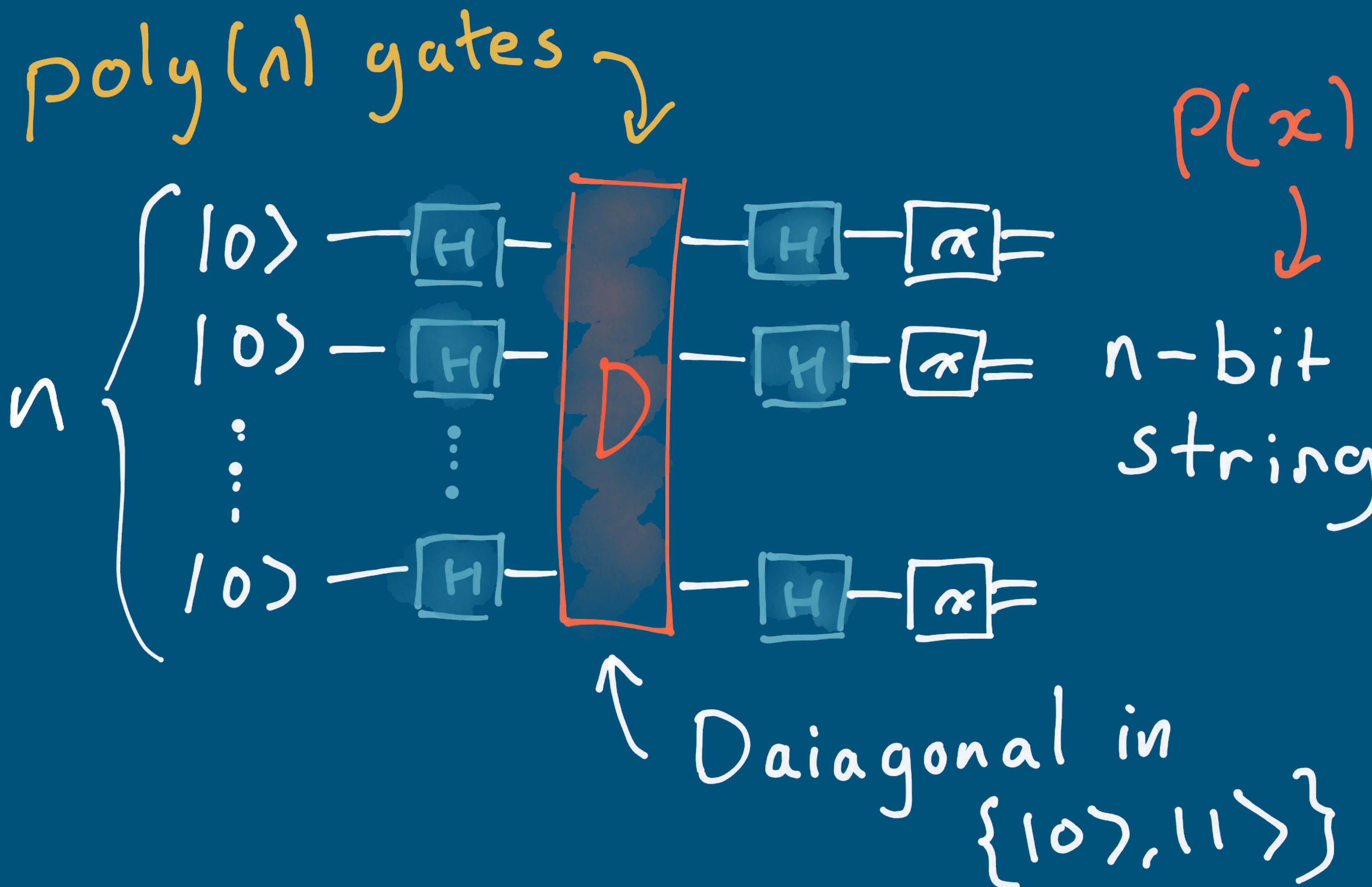
Pros and Cons

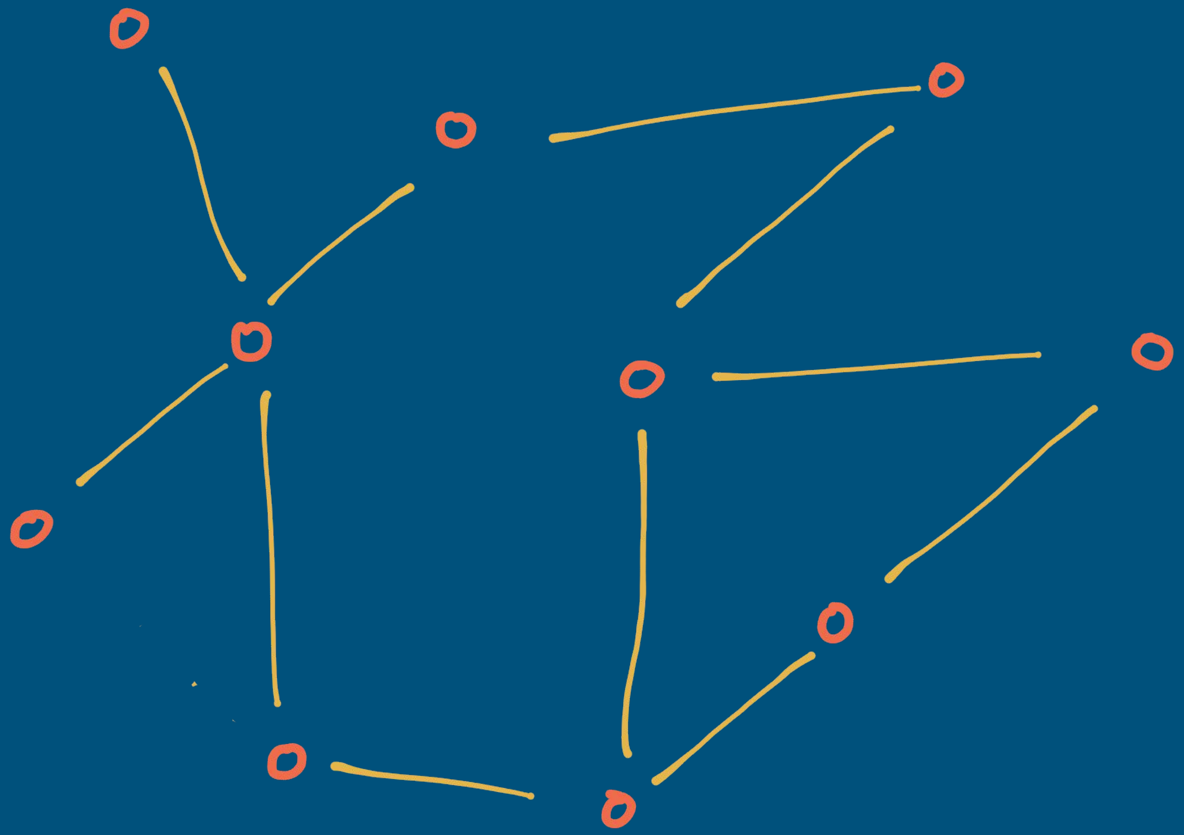
Advantages:

- Application motivated, not application specific.
- Motivated Unitary Coupled Cluster family of trial states used in VQE.
- Sample efficient.

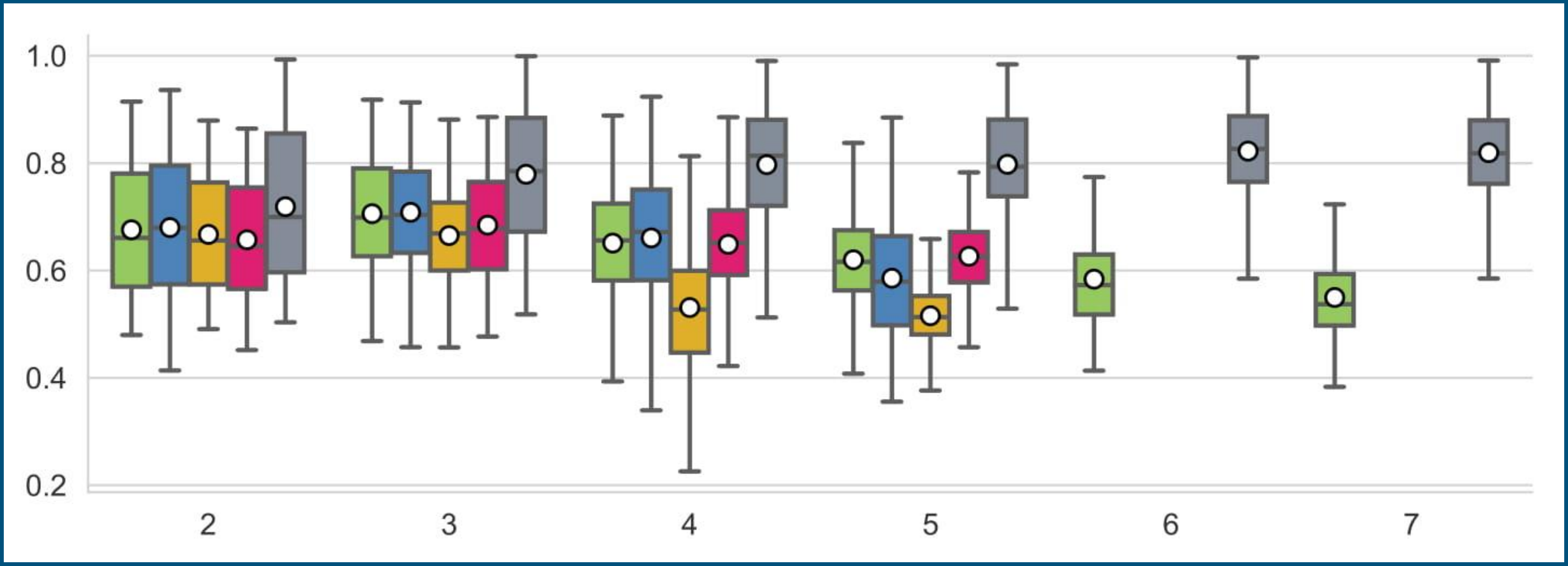
Disadvantages

- Very deep, very quickly


poly(n) gates
P(x)
n
|0⟩
|0⟩
|0⟩
H
D
x
n-bit string
Daiagonal in {|0⟩,|1⟩}



IQP Device Comparison



Singapore Yorktown Melbourne Ourense

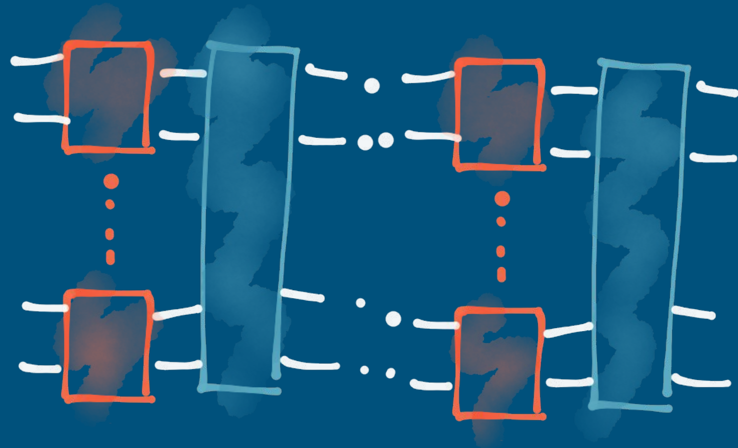
Pros and Cons

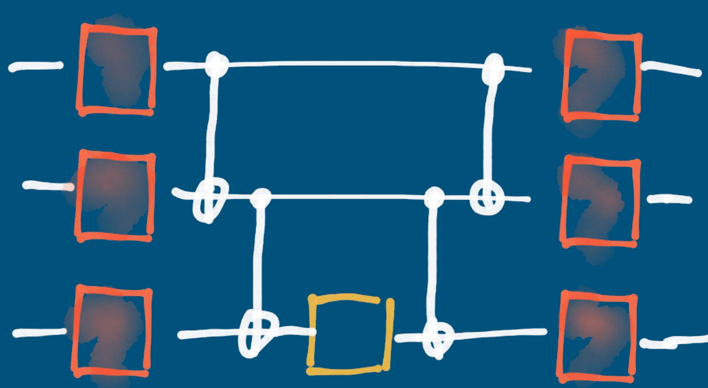
Advantages:

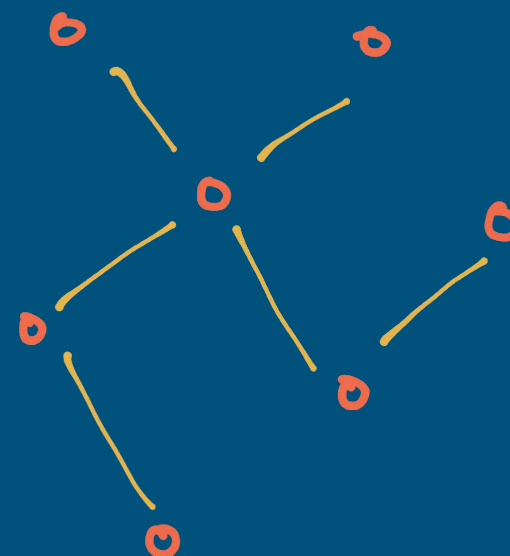
- Quantum Computational Supremacy results hold in the presence of noise, and on sparse architectures.
- Measure the impact of increasing circuit width independently of increasing circuit depth.

Disadvantages:

- Made up for by other circuits!







+ more figure of merits
+ Classical simulations

Summary

Measures practical performance:

- Covered a variety of depths and applications.

Well motivated figures of merit:

- Circuits and figures of merit are developed jointly.

Extensive results:

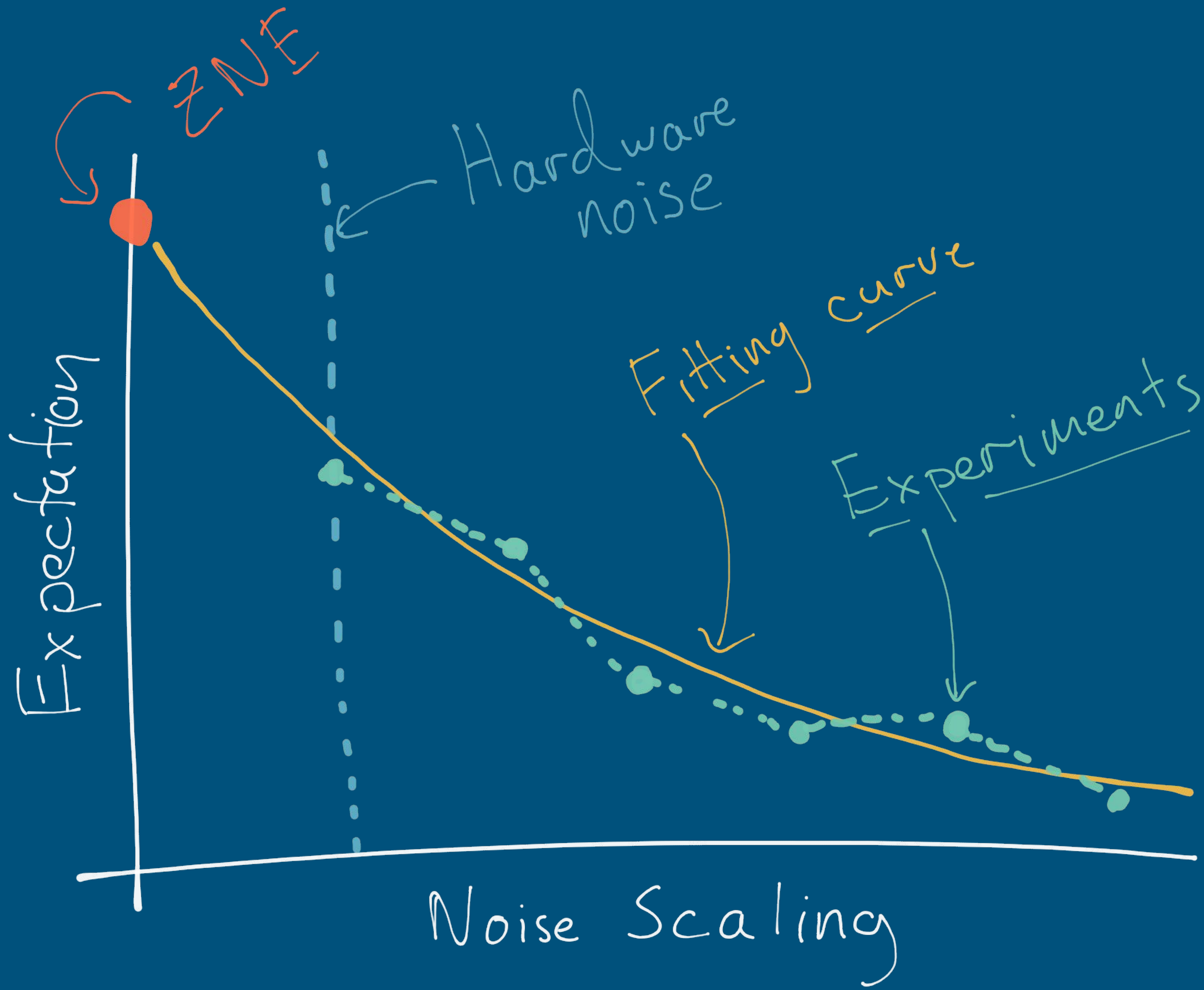
- Isolated best stack for applications.

Please see the paper: https://arxiv.org/abs/2006.01273

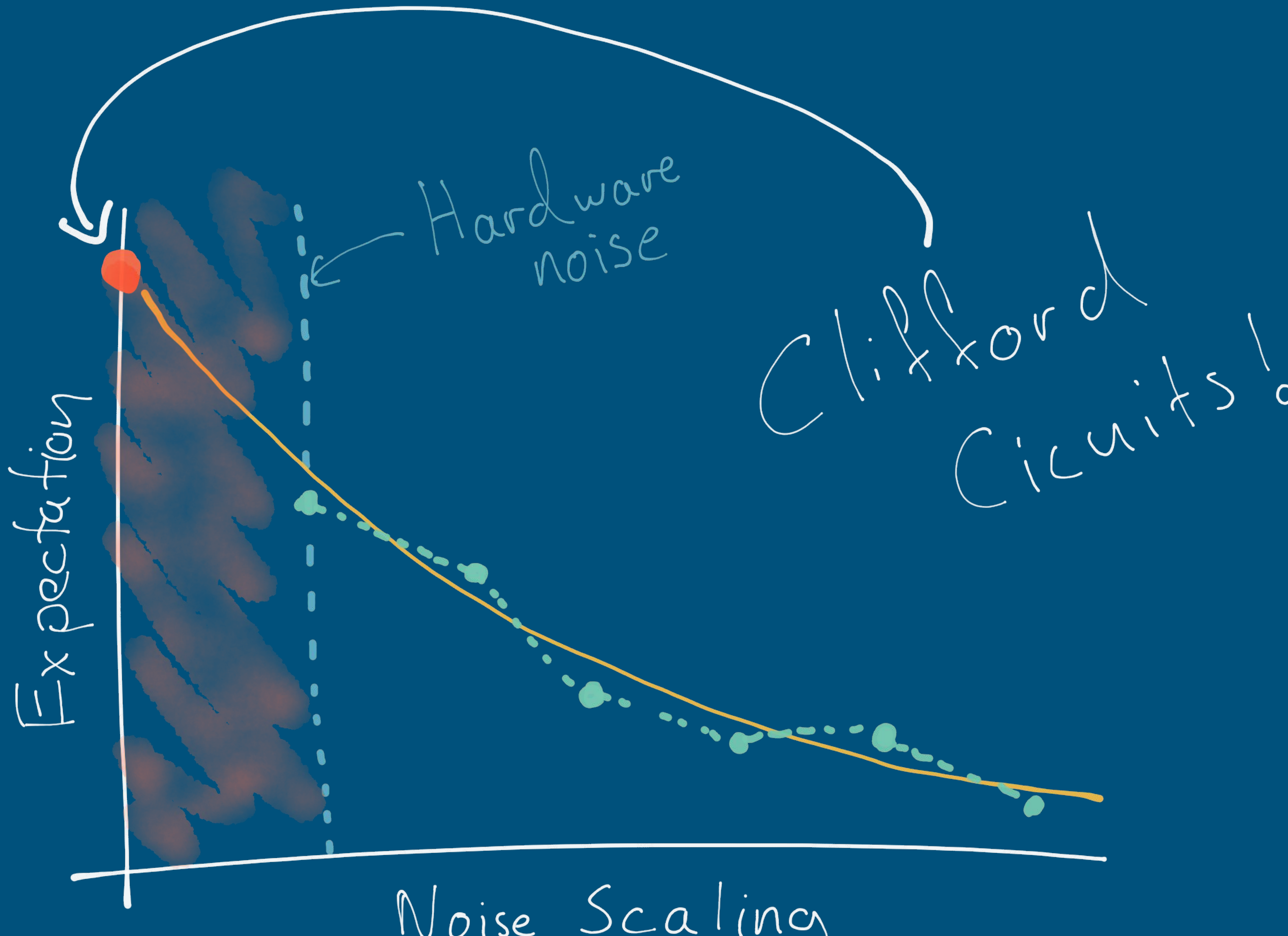
Extension of the Quantum Computing Stack

Error-Mitigation

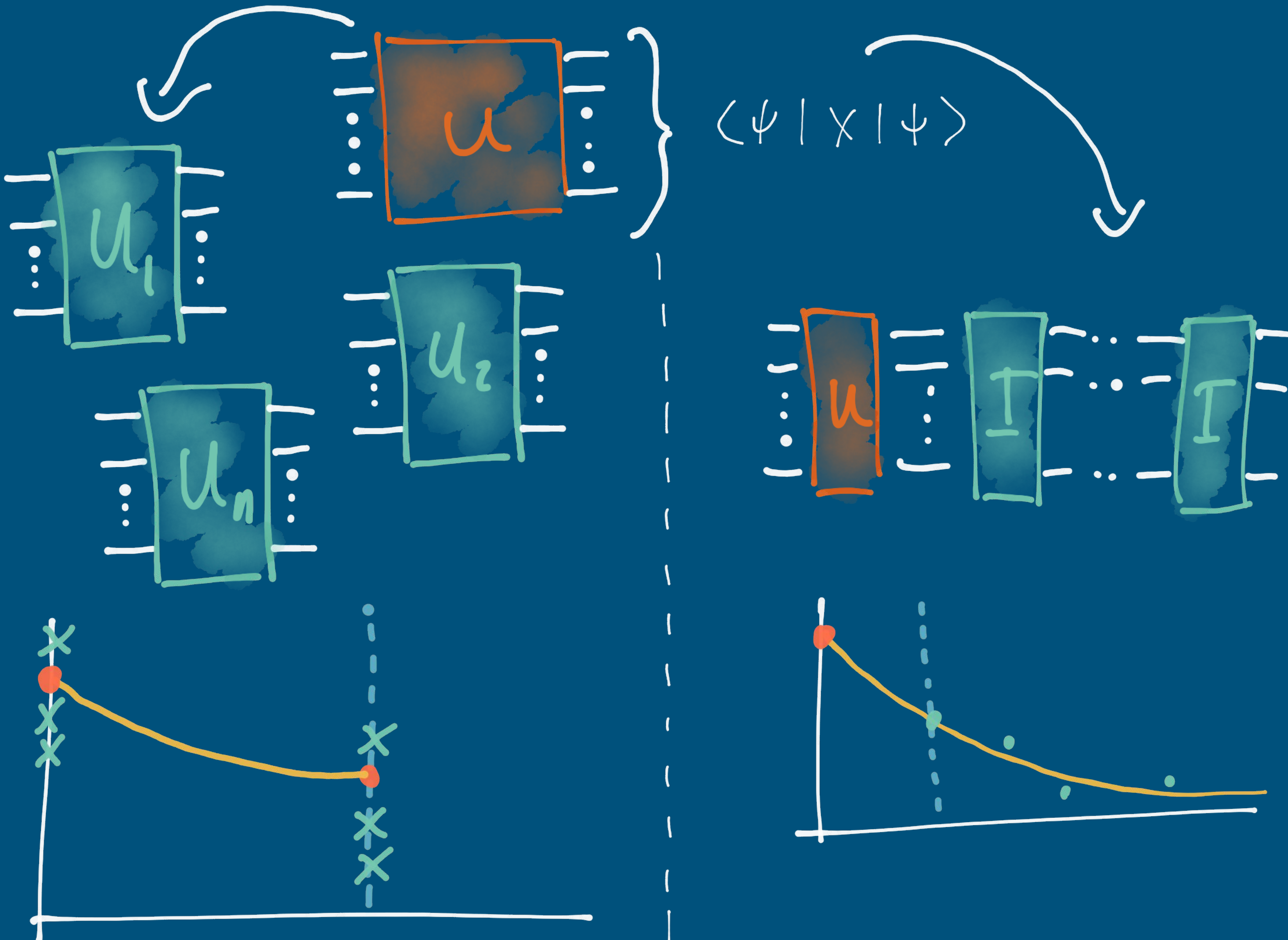
- Some examples of error mitigation
- A benchmarking methodology



Error mitigation for short-depth quantum circuits - https://arxiv.org/abs/1612.02058
Efficient variational quantum simulator incorporating active error minimisation - https://arxiv.org/abs/1611.09301

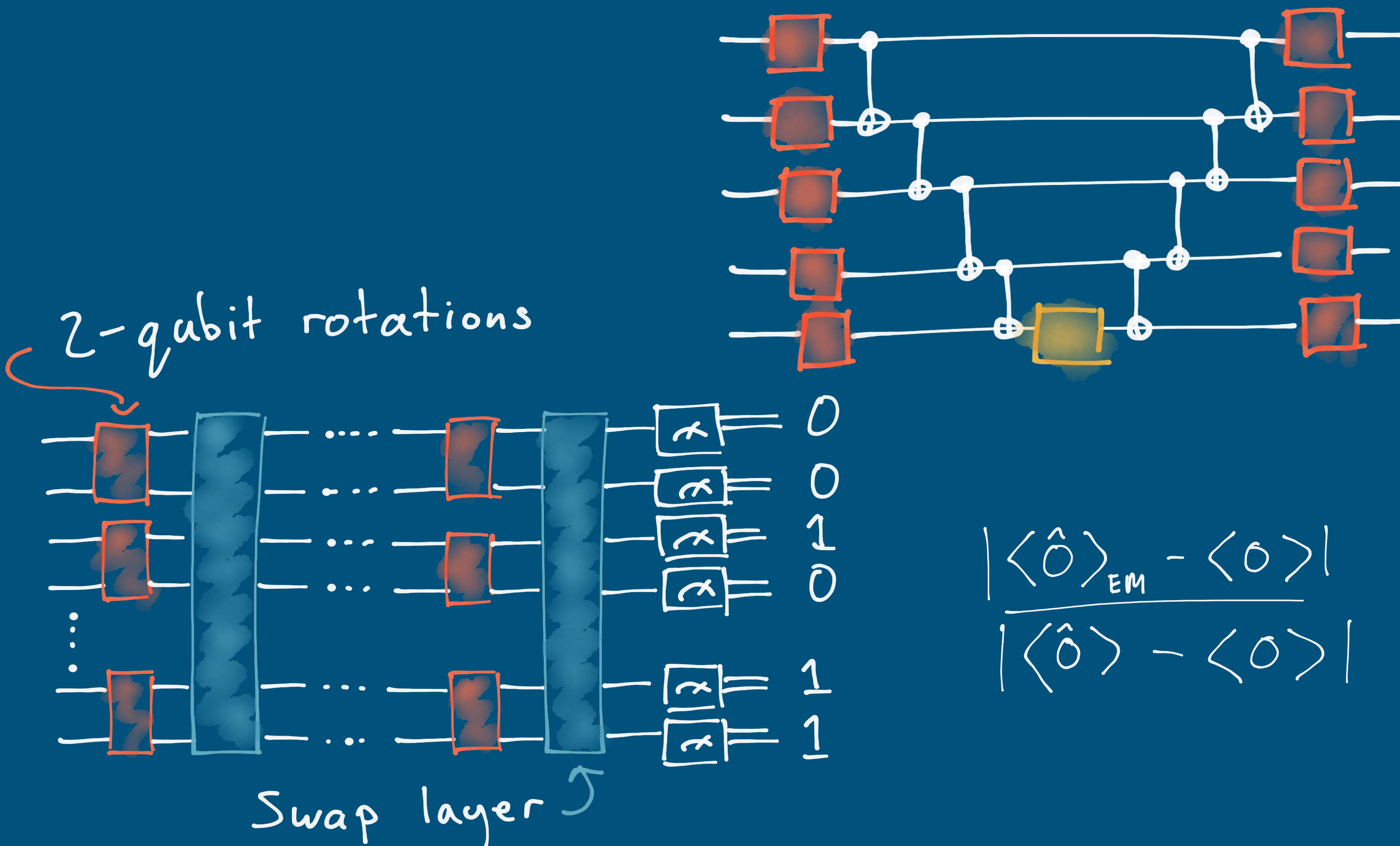


Error mitigation with Clifford quantum-circuit data -
https://arxiv.org/abs/2005.10189



U
⟨ψ | X | ψ⟩
U1
U2
Un
U
I
I

Benchmarking Circuits

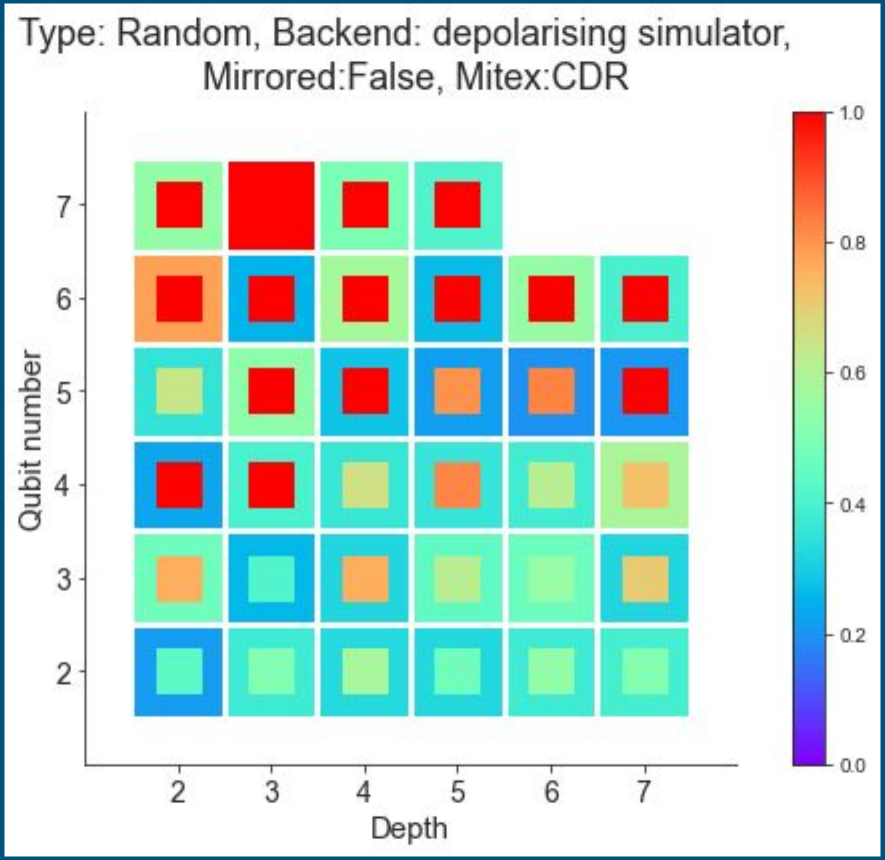


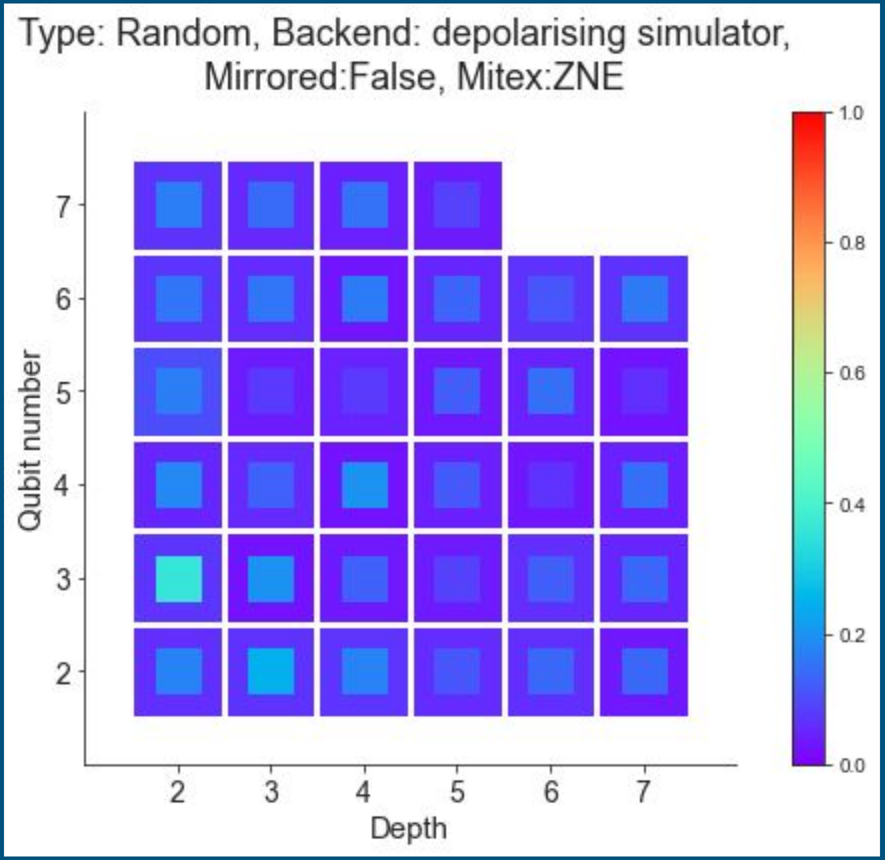
Qermit

pip install qermit

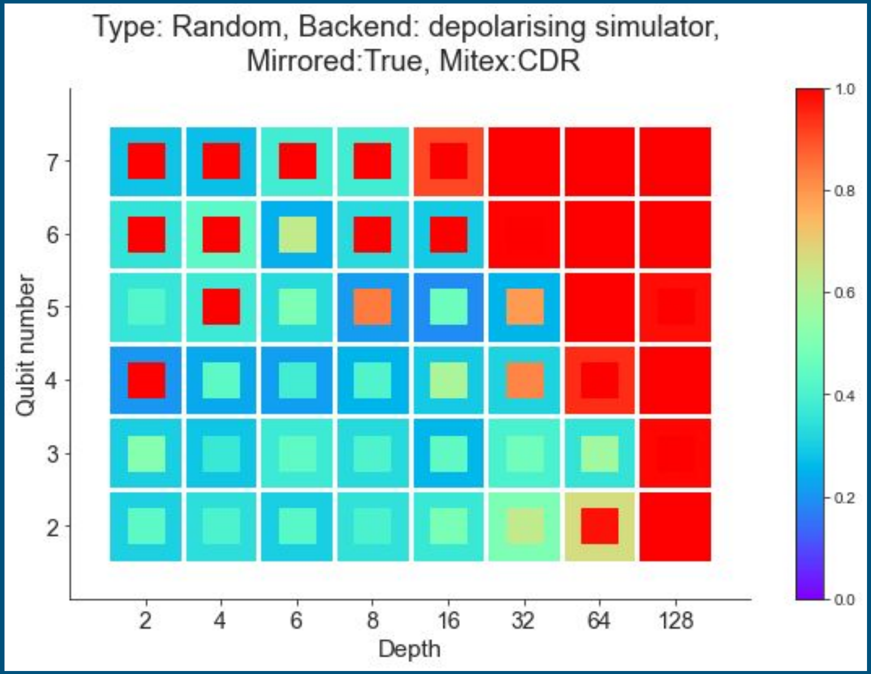
github.com/CQCL/qermit

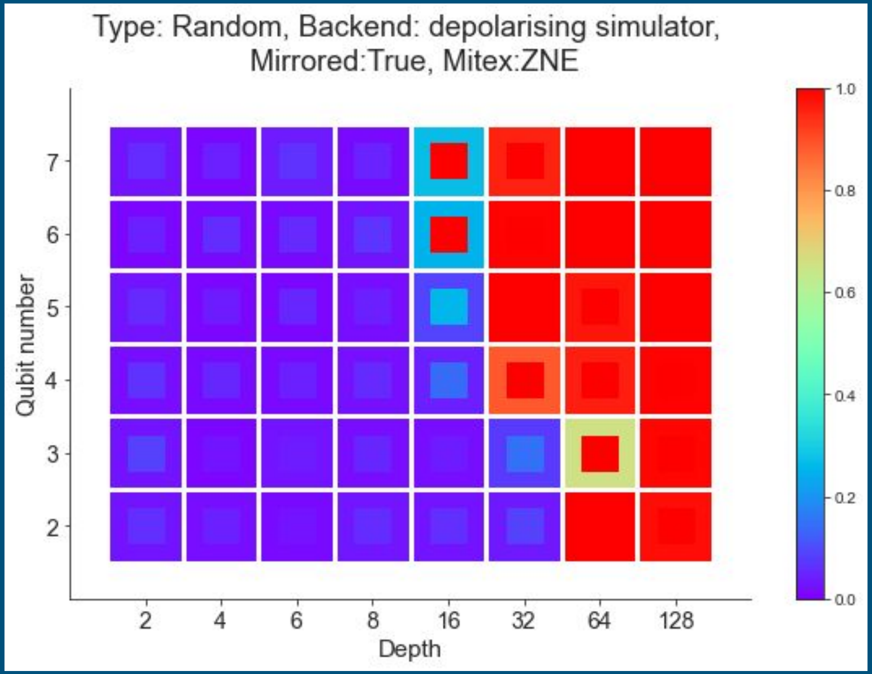
Classical Simulations



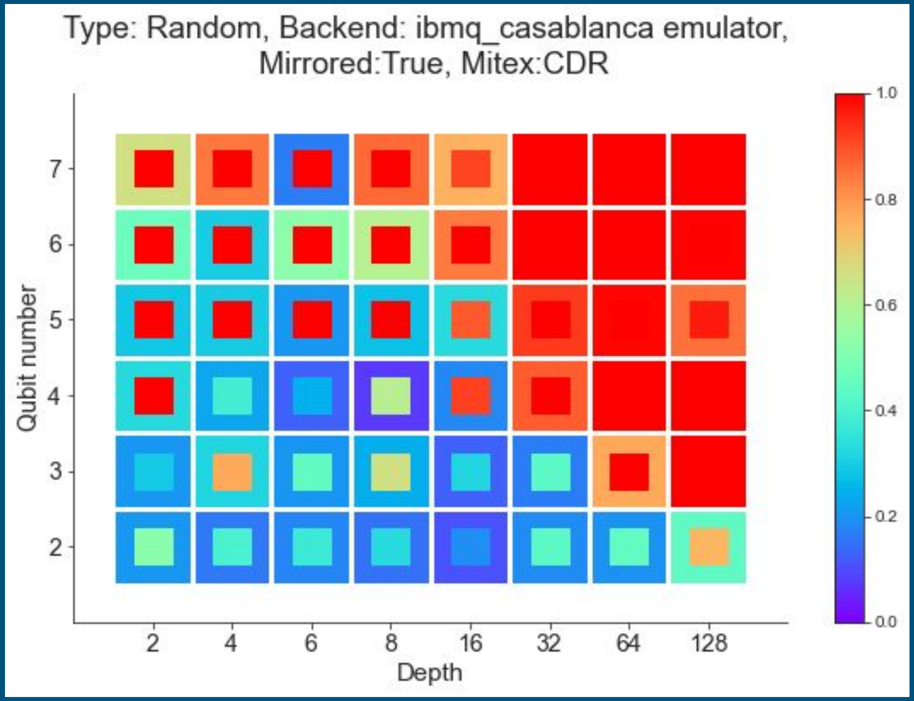


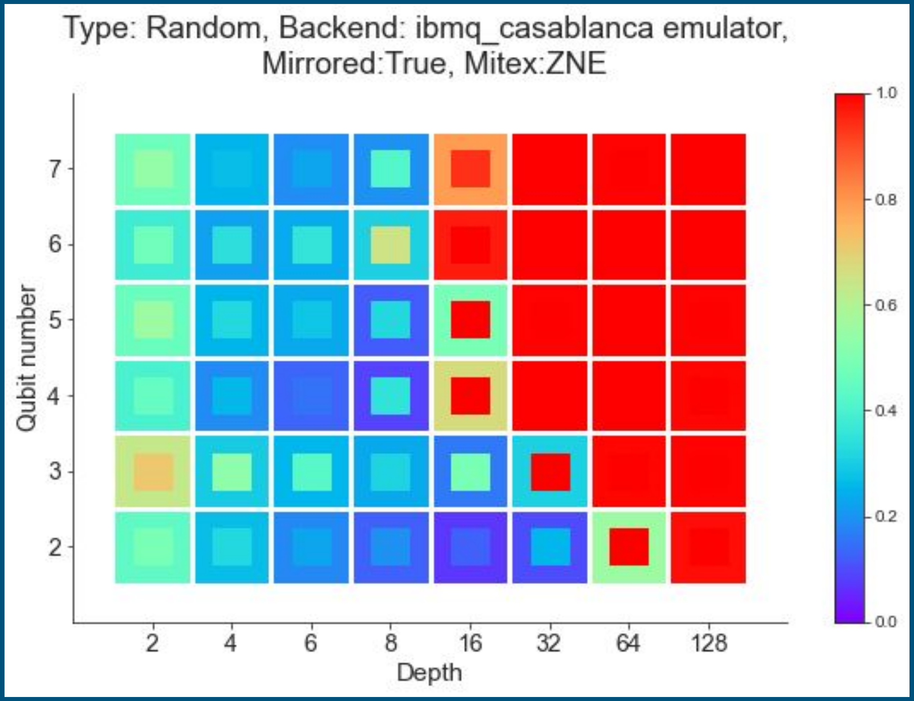
Mirrored Classical Simulations



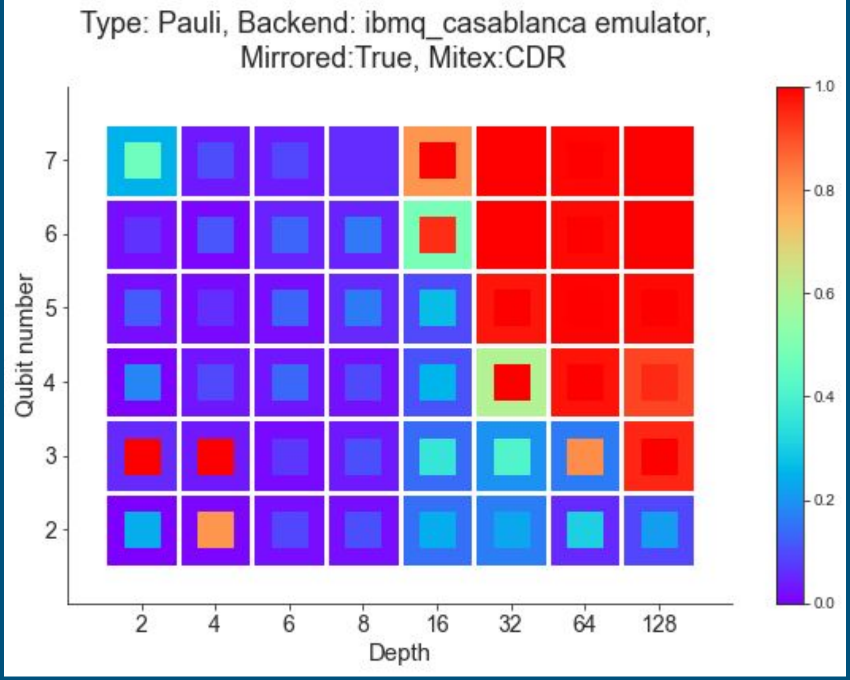


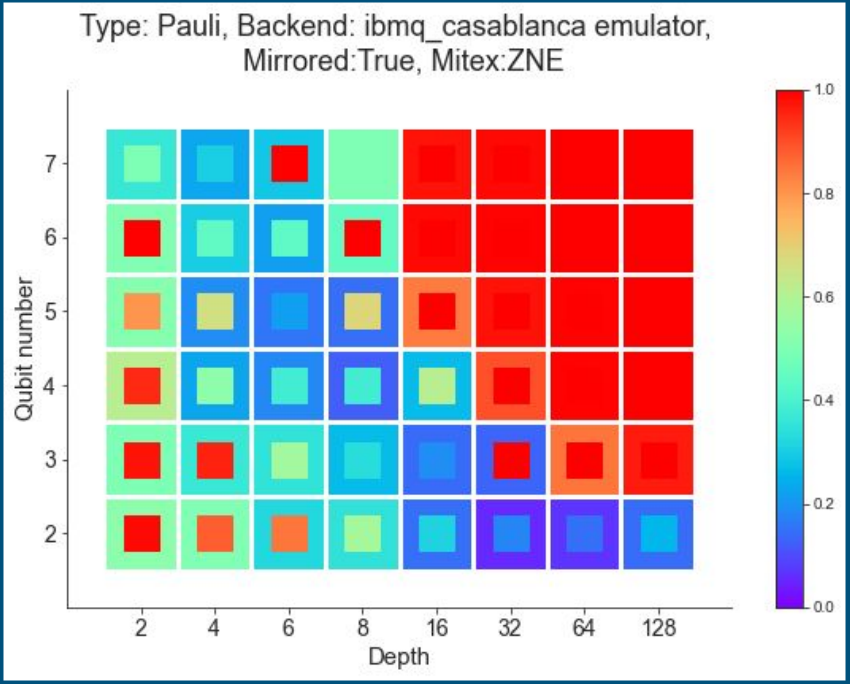
Mirrored Classical Emulation



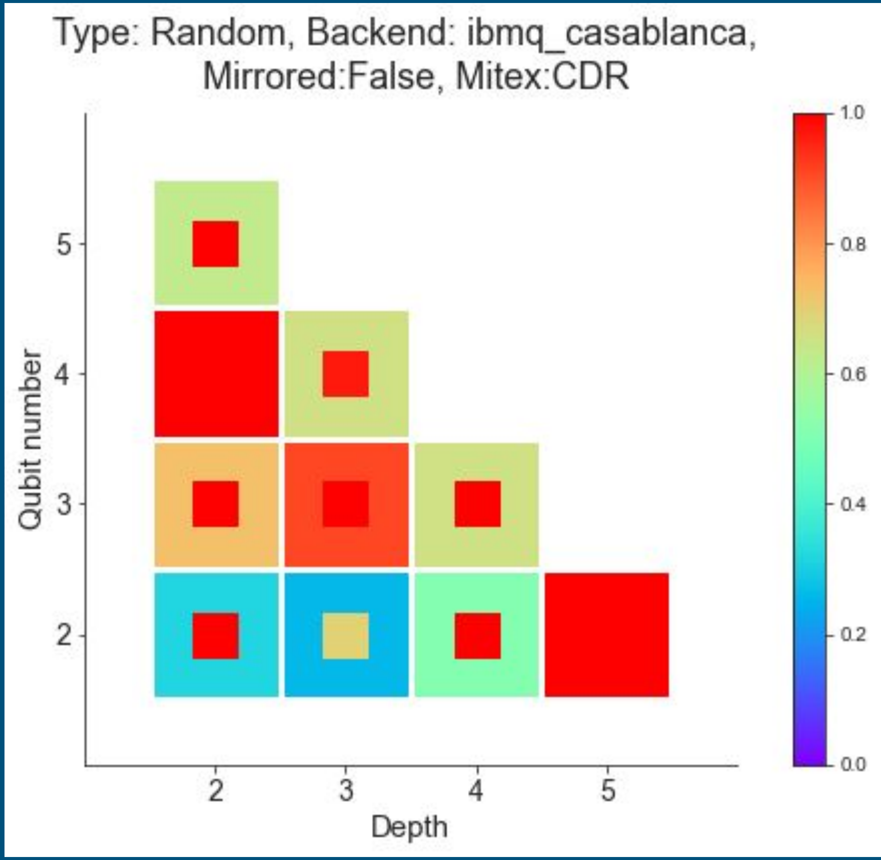


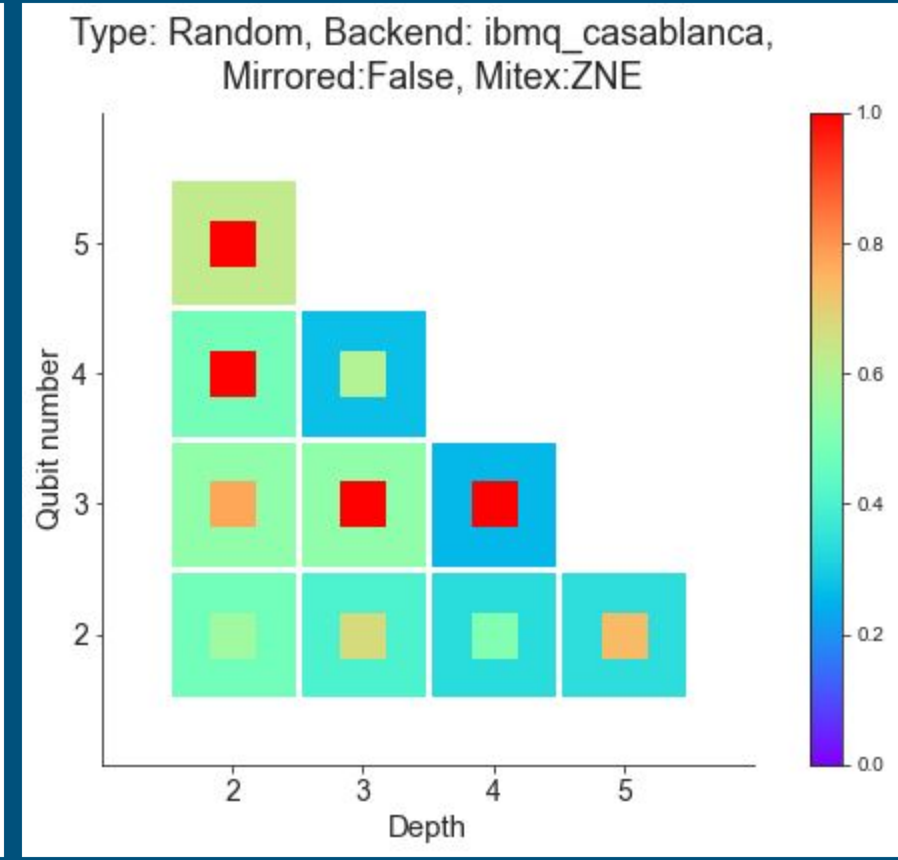
Pauli Gadget Classical Emulation





Real Device Performance





Conclusion

Application-Motivated, Holistic Benchmarking of a Full Quantum Computing Stack

- Covers many applications in a small suite
- Measures performance **in practice**
- Now includes error-mitigation

daniel.mills@cambridgequantum.com

Cheers

To you, and my collaborators