Low overhead fault tolerance for universal quantum computation

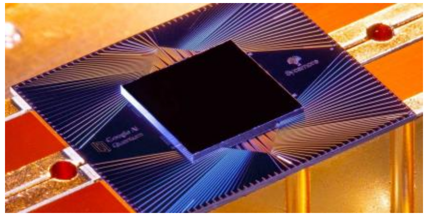
Madelyn Cain (mcain@g.harvard.edu)

Mathematical Picture Language Seminar

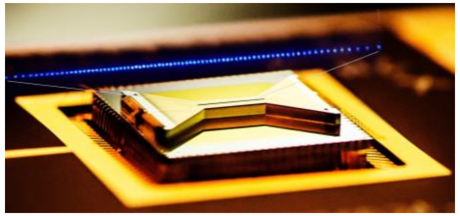
Frontier of quantum information science

Platforms for quantum computation

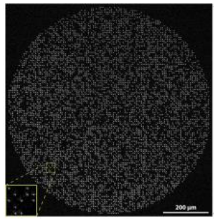
Superconducting qubits, trapped ions, neutral atoms…



Google



Monroe group



Endres group

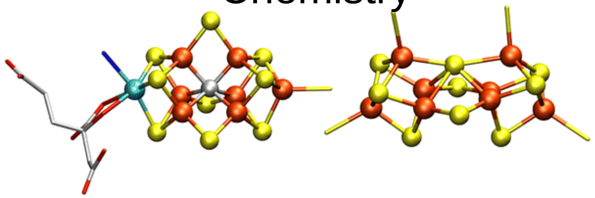
State-of-the-art gate error rates $10^{-2} - 10^{-3}$

Large-scale applications

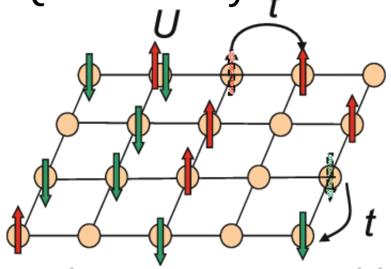
Cryptography

$$n = p \cdot q$$

Chemistry



Quantum dynamics



Yamada, Imamura, Machida

Required gate error rates $10^{-10} - 10^{-15}$

Gidney, Ekera, Fowler, Beverland, Hastad...

Suppressing errors is *the central challenge* in quantum computation.

Quantum error correction (QEC) is the only known realistic route to suppress errors for large algorithms.

Quantum error correction (QEC)

Initially, people thought QEC, and therefore quantum computing, would be *fundamentally impossible.*

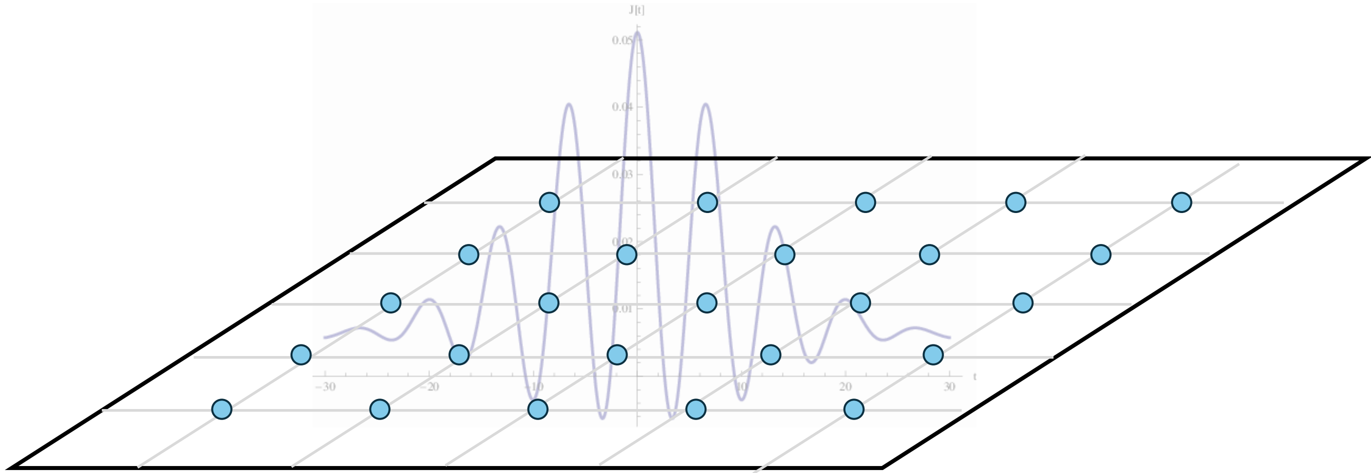
Classical error correction: duplicate information (repetition code) 0 → 0000000

Quantum error correction: conceptual challenge, cannot copy quantum information $|\psi\rangle \rightarrow |\psi\rangle|\psi\rangle|\psi\rangle$

QEC: theoretical breakthrough

Use *entanglement* to store information *nonlocally* to encode a *logical qubit.*

Use local *stabilizer measurements* to detect errors.



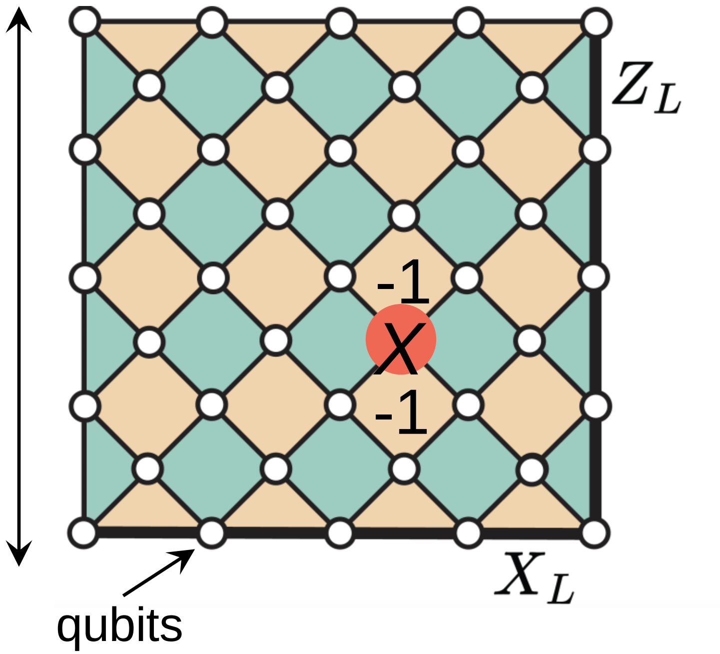
Logical qubit

Peter Shor, Andrew Steane, Alexei Kitaev, John Preskill, Laflamme, Calderbank, Gottesman…

QEC example: surface code

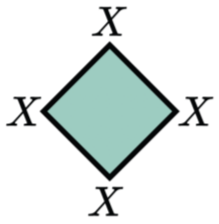
Product of two repetition codes

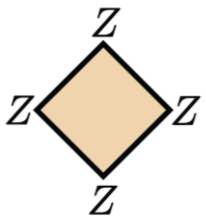
Code distance *d*



X stabilizer
(detects *Z* errors)

Z stabilizer
(detects *X* errors)





Error-free state: all stabilizers +1

Decoding: infer which error could have produced the measured syndrome so it can be undone

Threshold theorem

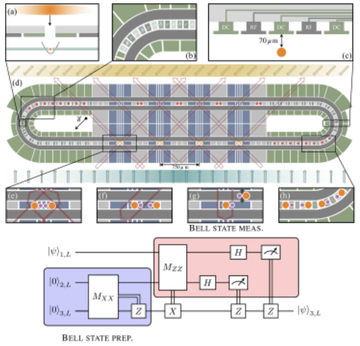
Logical error probability ~ $\left(\frac{p}{p_{\text{th}}}\right)^{O(d)}$

Threshold error rate $p_{\text{th}} \approx 1\%$ offers *realistic route* to extremely low errors

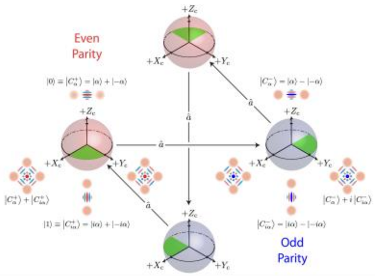
It was the advent of QEC that really allowed quantum computing to take off.

Breakthroughs in experimental QEC

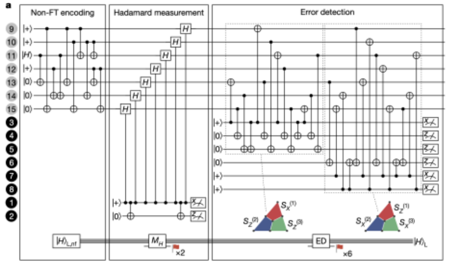
Many experimental platforms exploring QEC



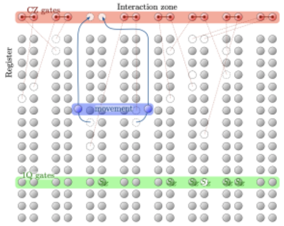
Quantinuum PRX 13, 4 (2023)
arXiv 2404.16728



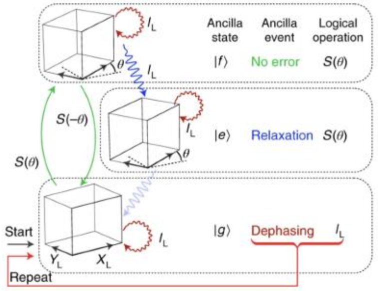
Ofek et al. Nature (2016)



Postler et al. Nature (2022)



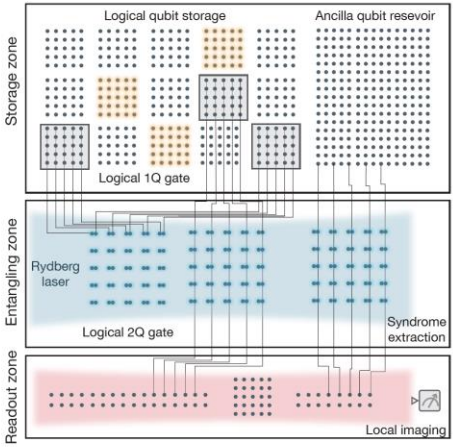
Reichardt et al. (2024)



Reinhold et al Nat. Phys (2020)

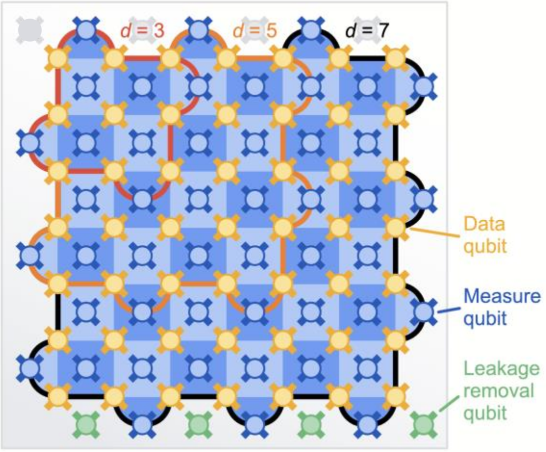
Recent breakthroughs:

Logical algorithms on 10’s of logical qubits



Bluvstein et al. Nature (2024)

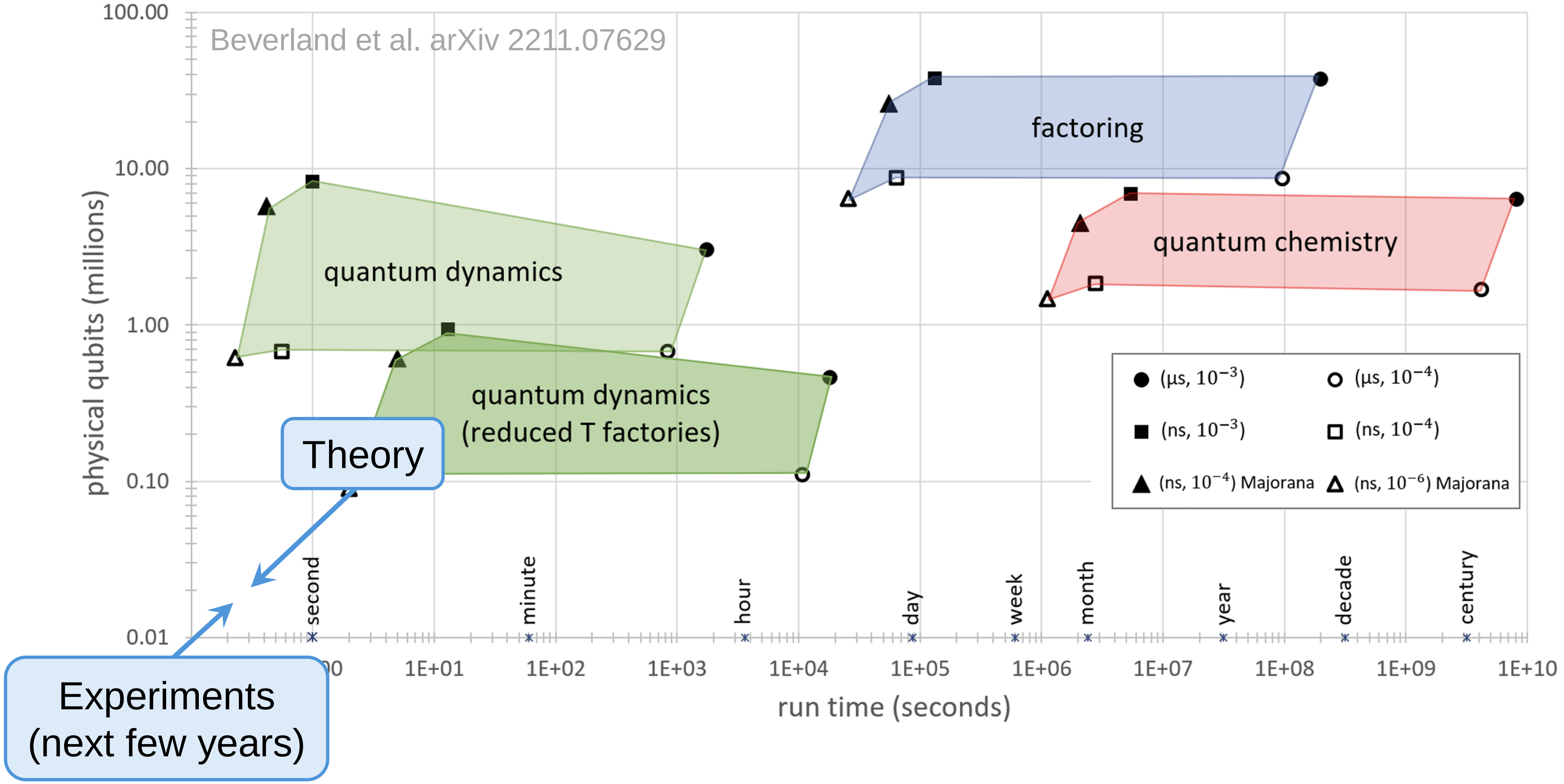
Sub-threshold scaling of a logical qubit



Google Quantum AI and Collaborators. Nature (2024)

Next few years, experimentally exploring deep quantum circuits with 1000’s-10000’s of qubits.
Can we implement useful fault-tolerant algorithms?

The race for useful QEC

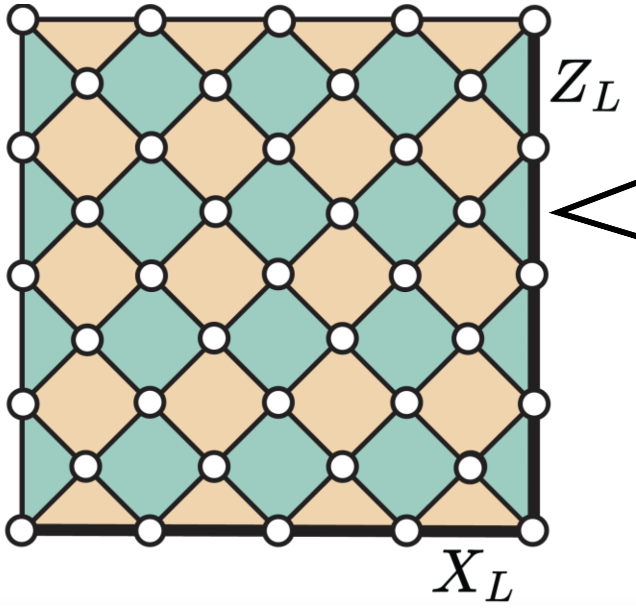


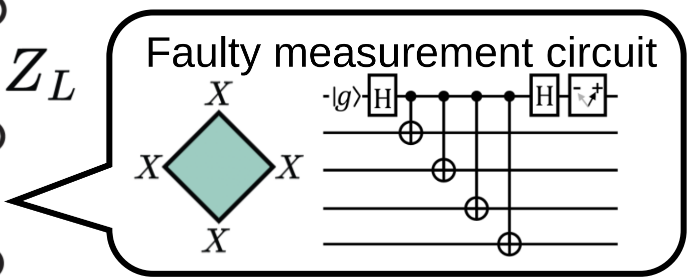
Key frontier:

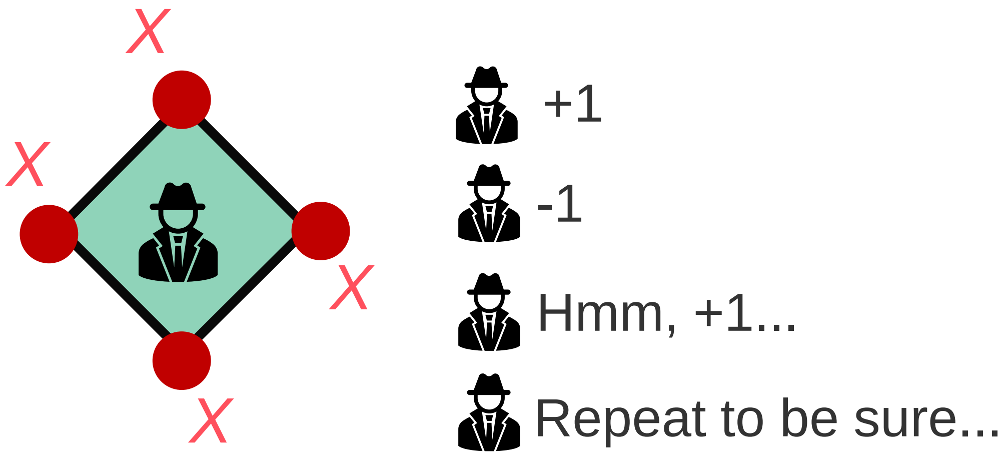
How to close the gap between hardware and useful error-corrected quantum computation?

Challenge of scaling QEC

Substantial resources due to ensuring *fault tolerance:* operations robust against noise







To know our stabilizers with exponential certainty in code distance, we need to measure O(d) times!

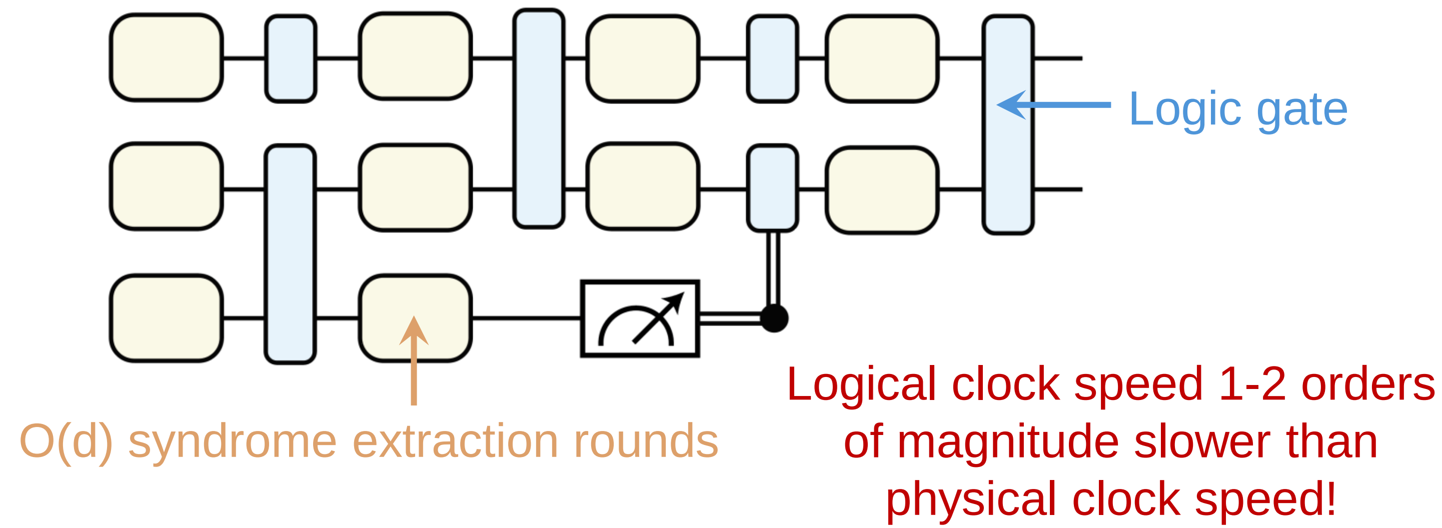
Large resource overhead: d is ~ 30 in large-scale problems!

D. Gottesman, arXiv:0904.2557; M. Beverland, S. Huang, V. Kliuchnikov, arXiv:2401.12017

Challenge of scaling QEC

Common approach to construct fault-tolerant algorithms: compose individual fault-tolerant operations

D. Gottesman, arXiv:0904.2557; M. Beverland, S. Huang, V. Kliuchnikov, arXiv:2401.12017vv



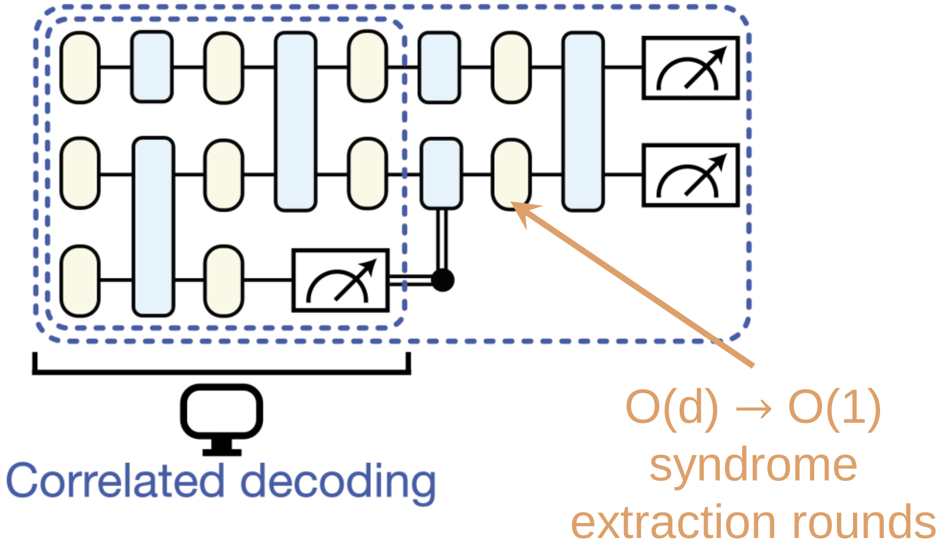
Intuitively, this ensures that we can confidently correct each operation independently.

Can we relax this requirement by considering the fault tolerance of the algorithm as a whole?

D. Gottesman, arXiv:0904.2557; M. Beverland, S. Huang, V. Kliuchnikov, arXiv:2401.12017

Our result

Using transversal gates and clean magic state inputs, and assuming fast decoding, we reduce the syndrome extraction overhead *for universal computation* by O(d).



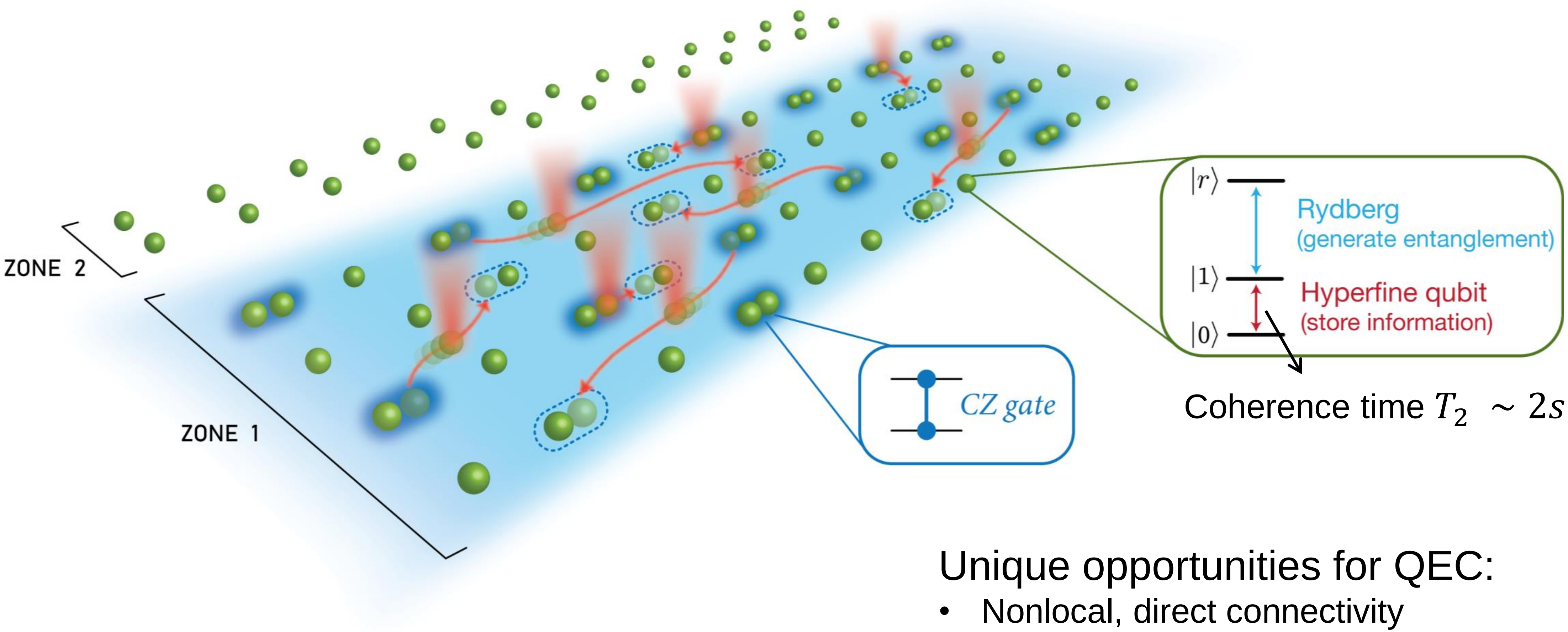
Theoretical implications: new understandings in fault tolerance
Practical implications: ≥ 30× circuit depth reduction for reconfigurable architectures

Outline

1. Exploring QEC with neutral atoms *“What we learn from experiments”*

2. Reducing QEC overhead with new theories of fault tolerance *“Enabling experiments of the future”*
    - Transversal Clifford circuits
    - Universal computation

Reconfigurable atom arrays



Coherence time $T_2 \sim 2s$

Bluvstein *et al. Nature* 604, 451-456 (2022)

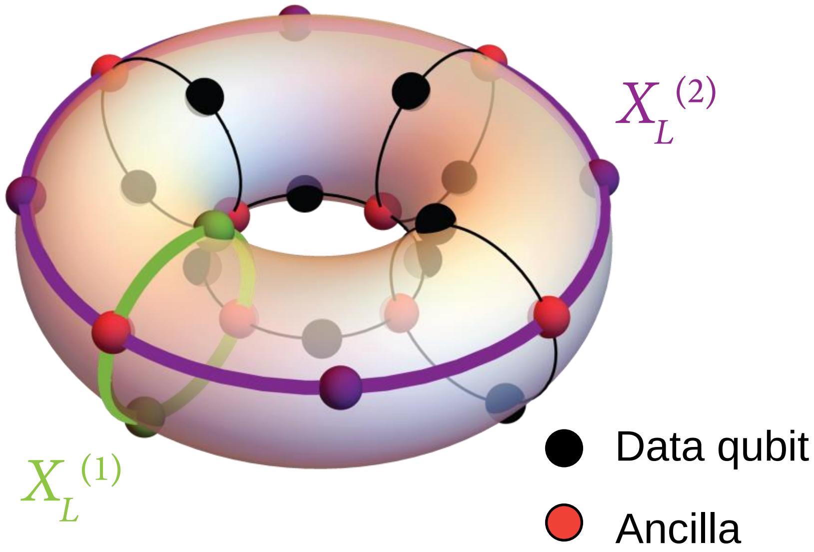
Pioneering work and recent exciting developments: Weiss, Saffman, Browaeys, Grangier, Regal, Endres, Kaufman, Bernien, Thompson, Ni, Bakr, Bloch, Covey ...

Unique opportunities for QEC:

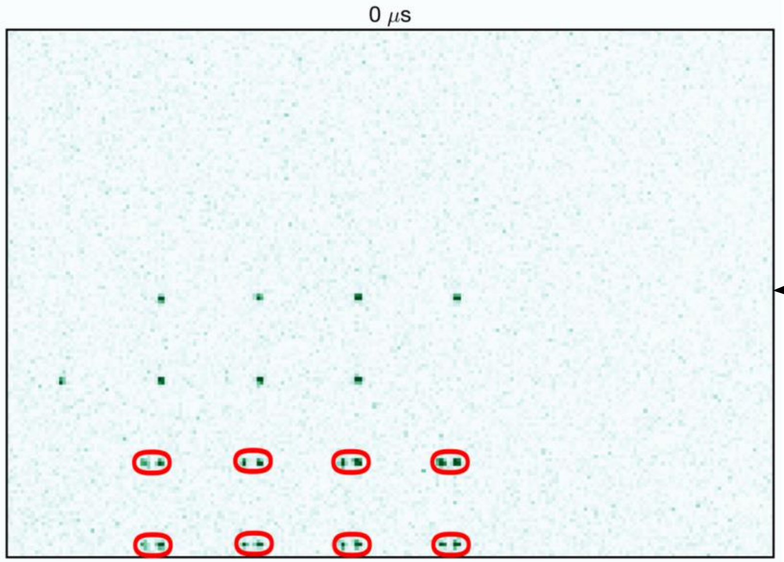
- Nonlocal, direct connectivity
- Efficiently rearrange atoms in parallel with O(1) classical controls

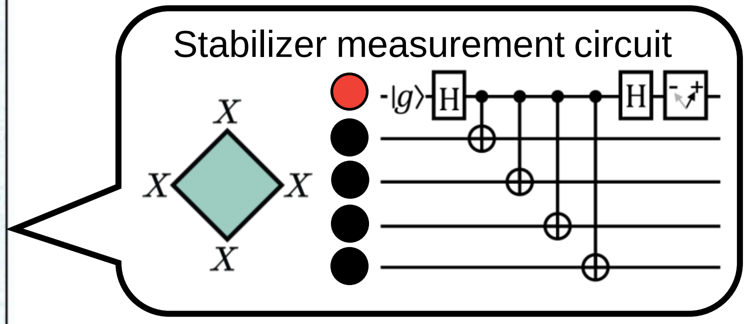
Toric code (on a torus)

Realizing error correcting codes with nonlocal connectivity



QEC primitive: syndrome extraction





Circuit simply programmed by AOD waveform (two voltages!)

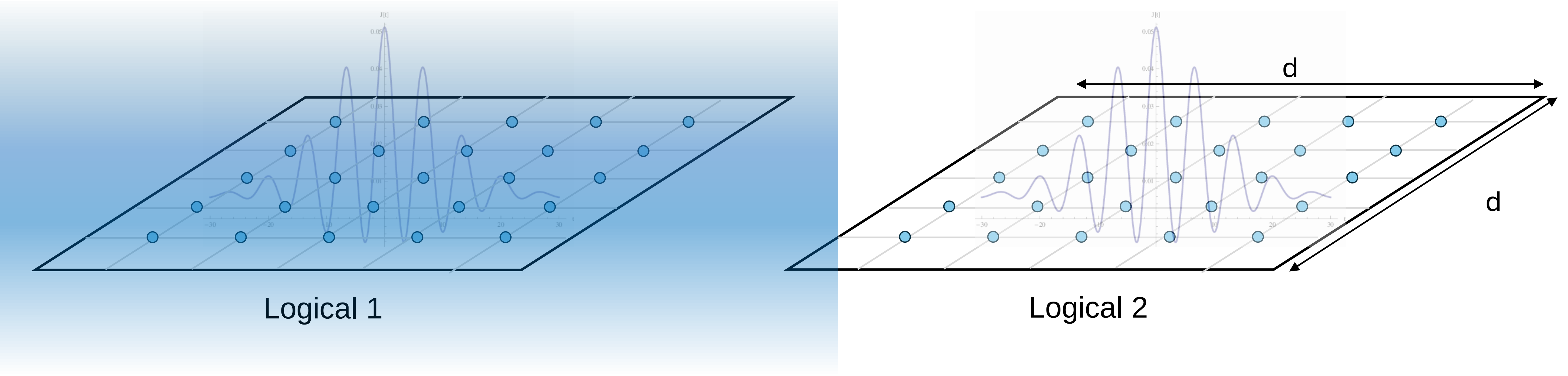
Parallel control over many qubits with O(1) controls

Bluvstein *et al. Nature* 604, 451-456 (2022)

Kitaev Toric code 1997

Nonlocal control: efficient logical operations

Transversal gate: generate logical operation by applying the same operation to the physical qubits

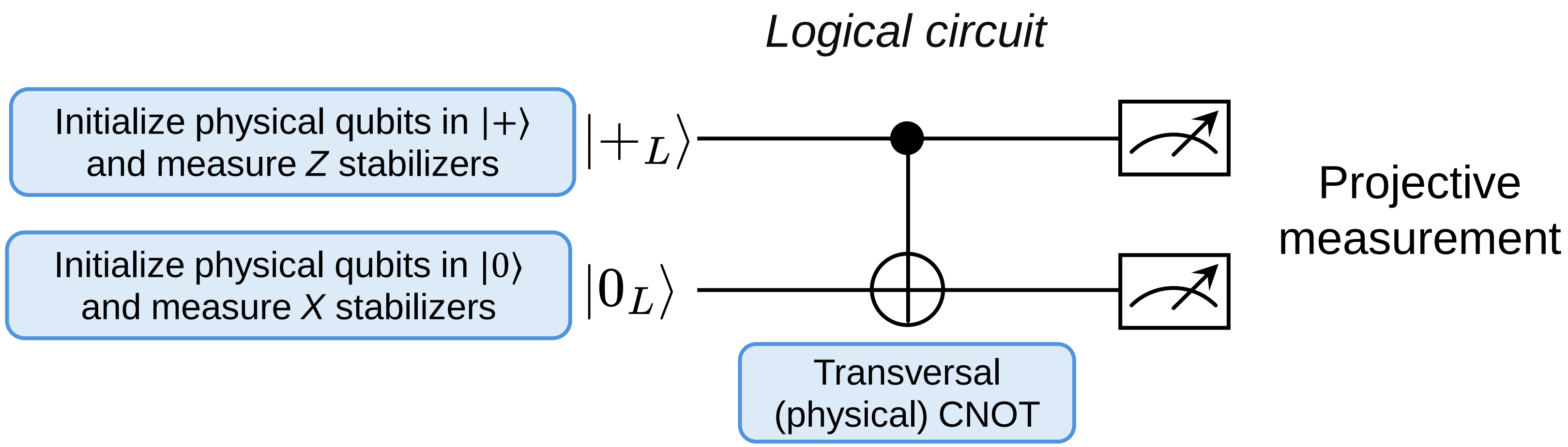


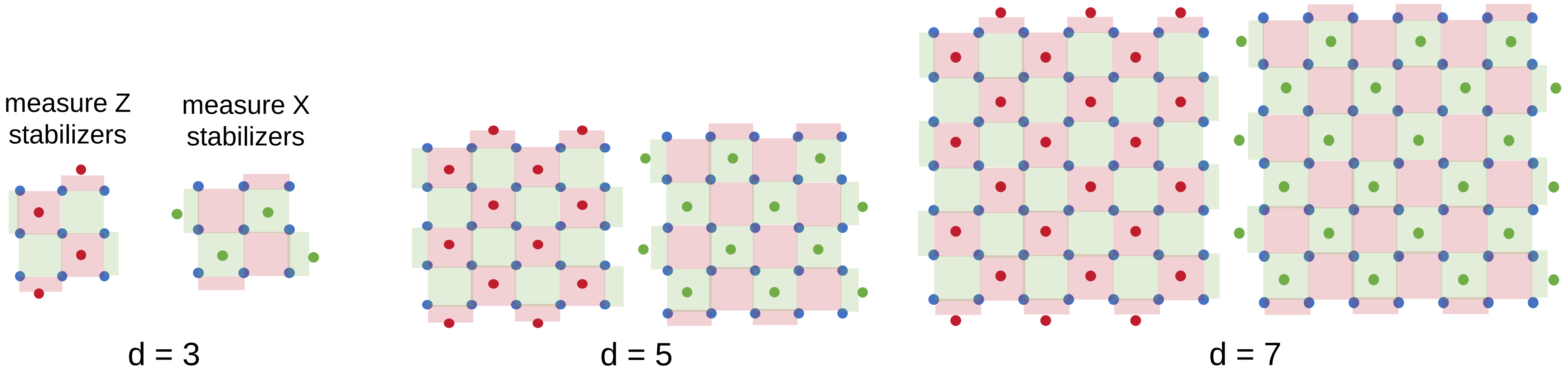
Transversal CNOT: directly interact the delocalized degrees of freedom via physical CNOTs

- *Inherently fault-tolerant:* errors cannot spread within code block
- Single *logical* qubit control, instead of single *physical* qubit control: O(1) classical optical controls needed

All physical qubits receive the same instruction and act like one big atom

Studying entangling gates with code distance





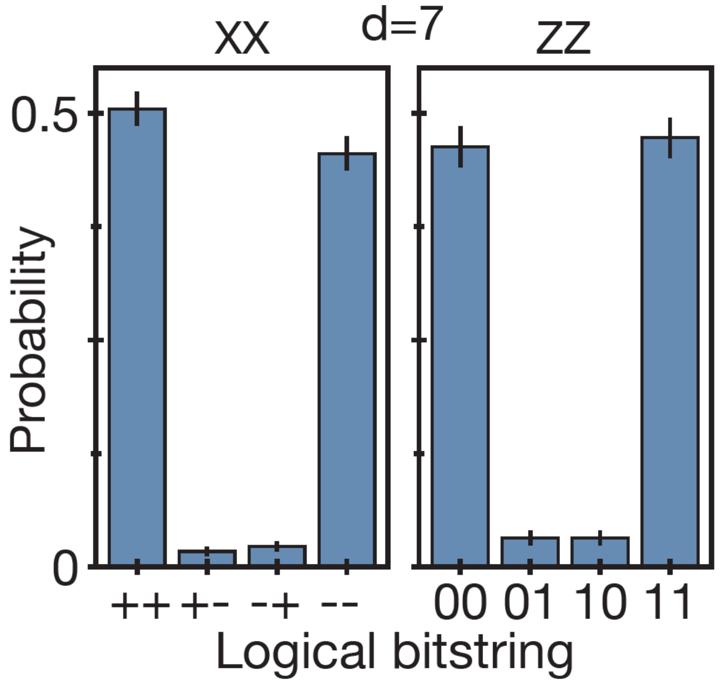
Threshold behavior: entangling operation improves with code distance

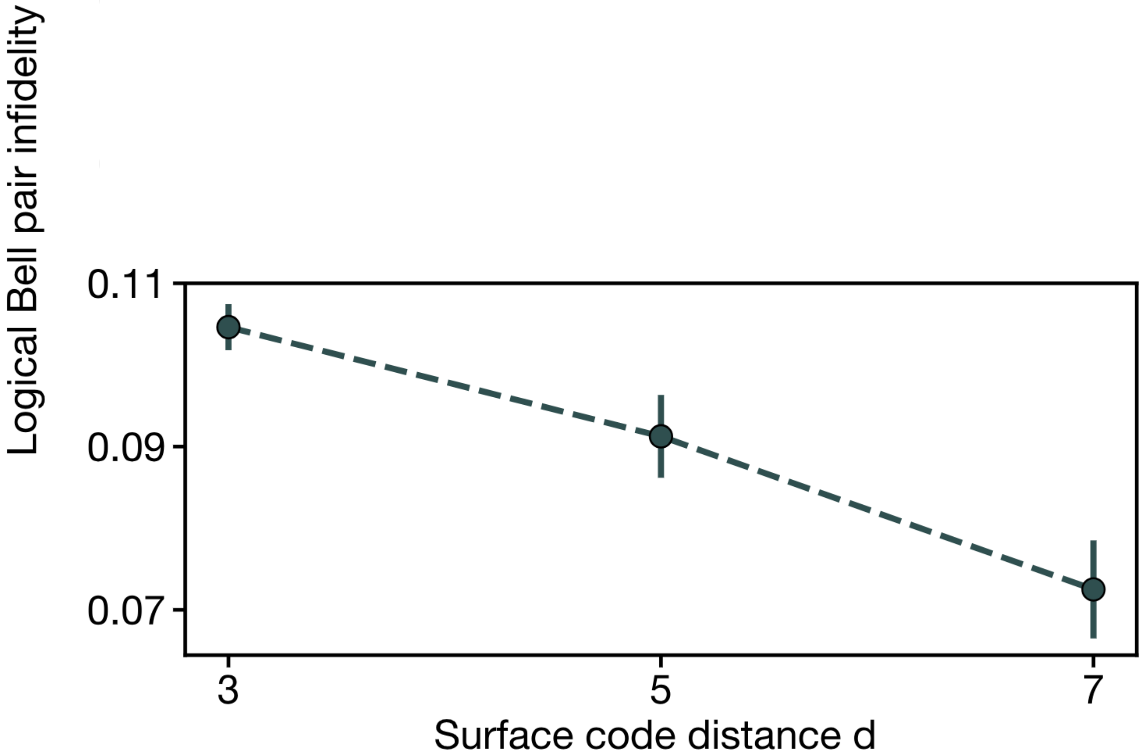
Prior work: Entanglement of two d=2 surface codes via lattice surgery (Innsbruck, Nature, 2021)
Google, Nature, 2023 and recently Google Quantum AI and Collaborators. Nature (2024)

Improving entanglement with code distance

Number of physical qubits per Bell pair

Compute fidelity bound from logical XX and ZZ expectations



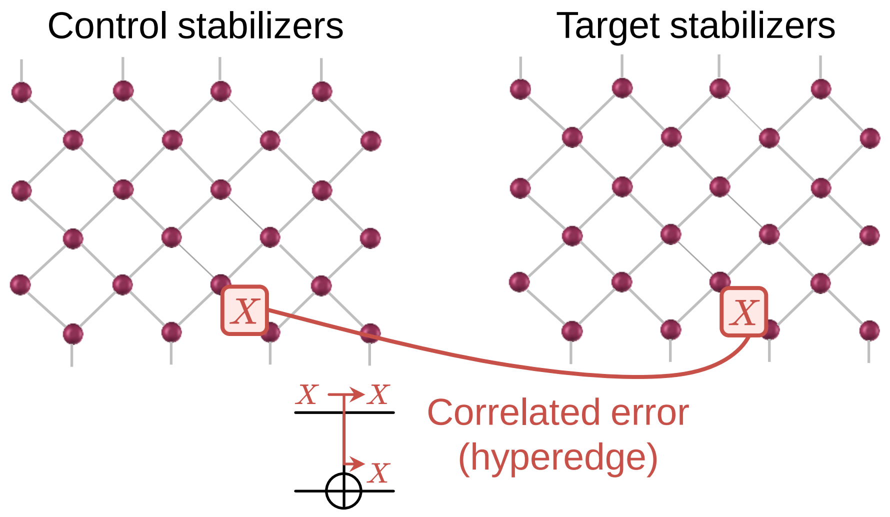


Bluvstein,…,MC et al., Nature 626, 58-65 (2024)

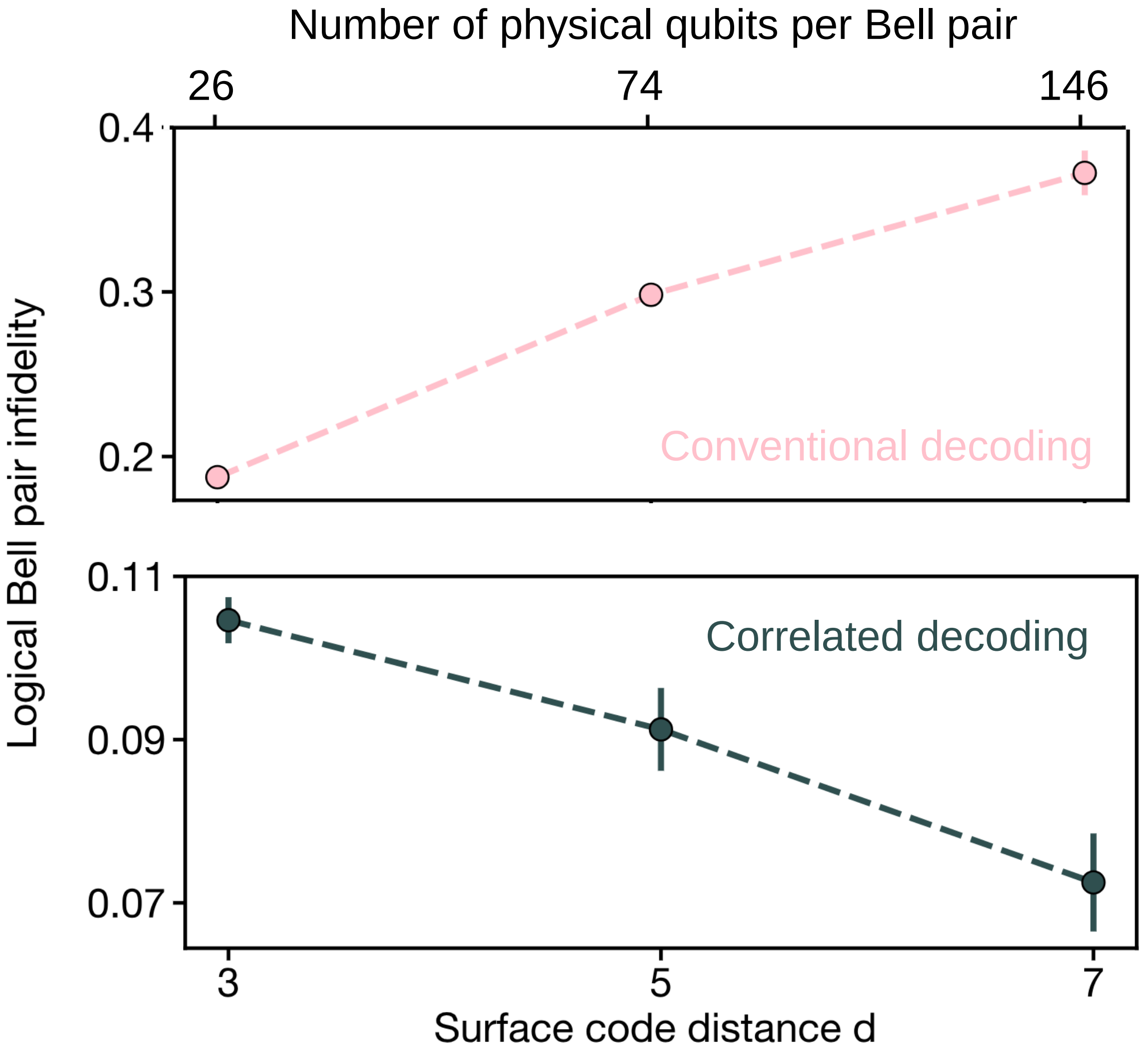
Improving entanglement with code distance

Key insight: decode the logical qubits in the algorithm jointly

MC et al. PRL (2024) arXiv: 2403.03272



O(d) syndrome extraction rounds
→ O(1) rounds

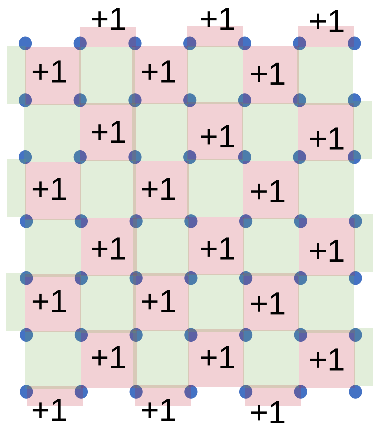


Bluvstein,…,MC et al., Nature 626, 58-65 (2024)

Criticism of logical CNOT experiment

$|+_L\rangle$ initialization was non fault-tolerant!

- Initialize all physical qubits in $|+\rangle$
- Measure Z stabilizers once



X stabilizers +1 (reliable)
Z stabilizers random (unreliable)

With O(1) rounds and applying corrections, the resulting state is a mixture of product states (not close to any logical codeword) due to *uncertain Z stabilizers*.

See Hastings, PRL 2011; Haah Simon's talk "What is your logical qubit" (YouTube)

To reconstruct the logical state to exponential precision in d, we need O(d) rounds!
However, do we even to reconstruct the state?

Correlated decoding

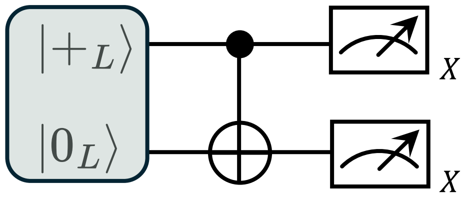
We don't need to reconstruct the many-body state, we need to reconstruct the *ideal joint logical measurement distribution* for $|00\rangle_L + |11\rangle_L$.

- $X_L Z_L$ and $Z_L X_L$ basis: 50/50 random
- $X_L X_L$ and $Z_L Z_L$ basis: +1 ideally

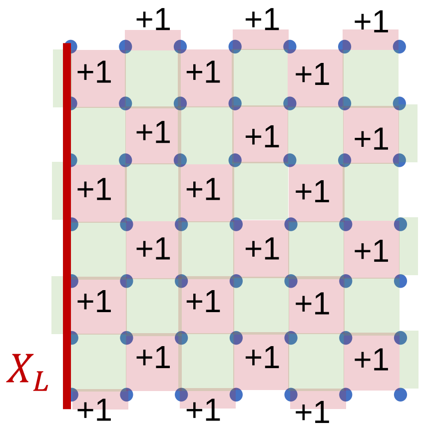
Don't even need to decode!

Must get right! Accurately predict by decoding logical qubits jointly.

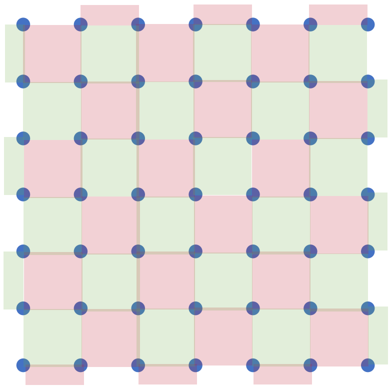
Consider predicting $X_L X_L$:



Control $|+_L\rangle$



Target $|0_L\rangle$



Correlated decoding

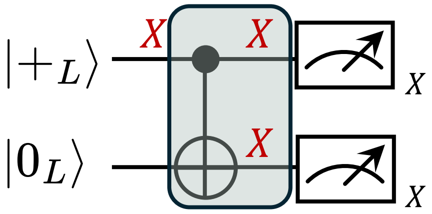
We don't need to reconstruct the many-body state, we need to reconstruct the *ideal joint logical measurement distribution* for $|00\rangle_L + |11\rangle_L$.

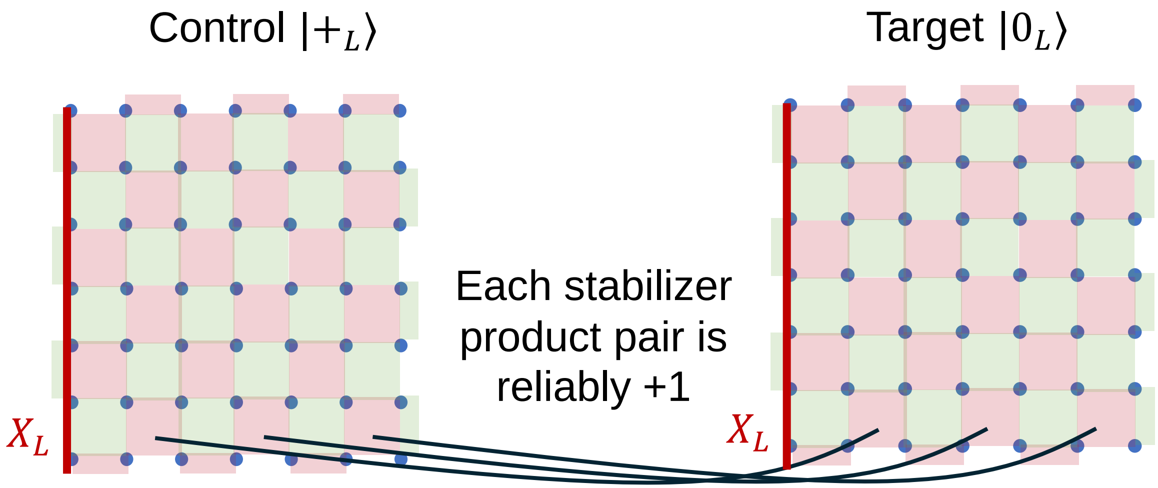
- $X_L Z_L$ and $Z_L X_L$ basis: 50/50 random
- $X_L X_L$ and $Z_L Z_L$ basis: +1 ideally

Don't even need to decode!

Must get right! Accurately predict by decoding logical qubits jointly.

Consider predicting $X_L X_L$:





Correlated decoding

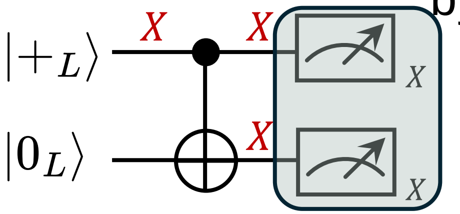
We don't need to reconstruct the many-body state, we need to reconstruct the *ideal joint logical measurement distribution* for $|00\rangle_L + |11\rangle_L$.

- $X_L Z_L$ and $Z_L X_L$ basis: 50/50 random
- $X_L X_L$ and $Z_L Z_L$ basis: +1 ideally

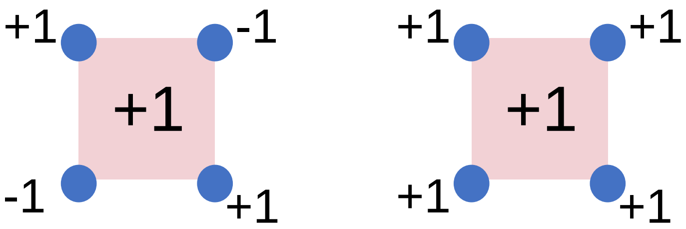
Don't even need to decode!

Must get right! Accurately predict by decoding logical qubits jointly.

Consider predicting $X_L X_L$:



Key idea: When we projectivity measure data qubits the $X_L X_L$ basis, we gain *reliable* stabilizer measurements!



Reliable +1 *X* stabilizer product

Correlated decoding

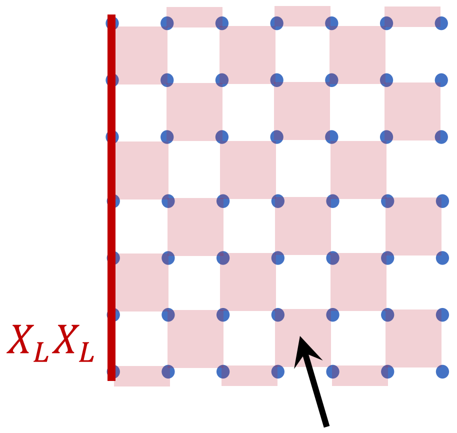
We don't need to reconstruct the many-body state, we need to reconstruct the *ideal joint logical measurement distribution* for $|00\rangle_L + |11\rangle_L$.

- $X_L Z_L$ and $Z_L X_L$ basis: 50/50 random
- $X_L X_L$ and $Z_L Z_L$ basis: +1 ideally

Don't even need to decode!

Must get right! Accurately predict by decoding logical qubits jointly.

Referee: In principle, you can predict $X_L X_L$ from the reliable stabilizer products!



X stabilizer product: anti-commute (detect) all Z errors on either code

Key idea: Reliable X stabilizers propagate the same way as the *X logical operator.*

When we measure the logical operator, we learn the right stabilizers (50/50 random stabilizers not needed).

Can we generalize?

Outline

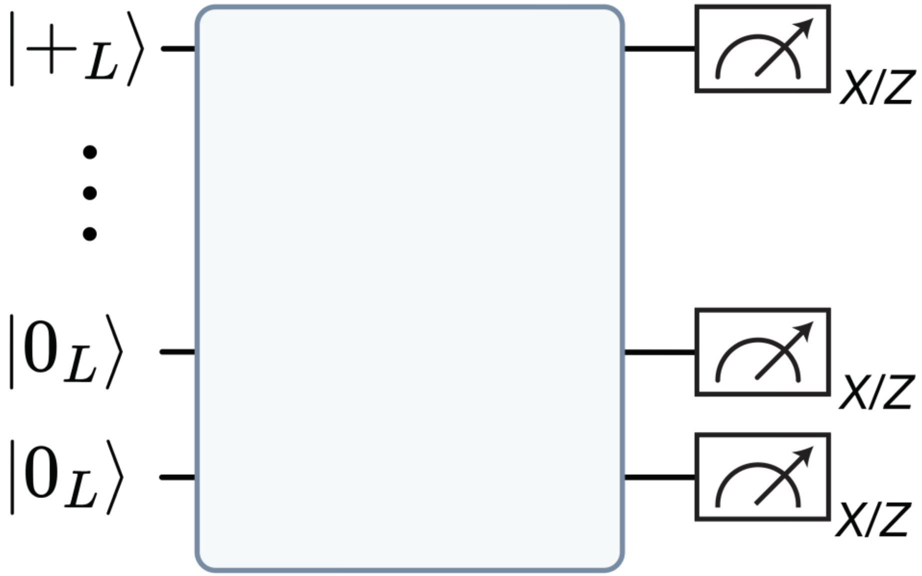
1. Exploring QEC with neutral atoms *“What we learn from experiments”*

2. Reducing QEC overhead with new theories of fault tolerance
   “Enabling experiments of the future”
   - Transversal Clifford circuits
   - Universal computation

Correlated decoding

Transversal Clifford circuit: $\{H_L, S_L, CNOT_L\}$
Paulis propagate to Paulis

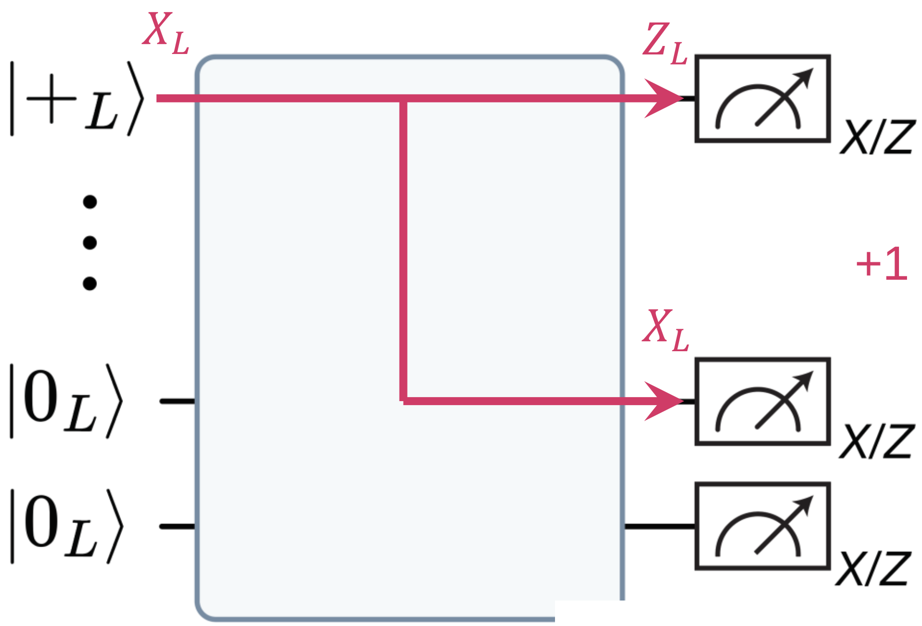
For example:
+1 X_L operator



Correlated decoding

Transversal Clifford circuit: $\{H_L, S_L, CNOT_L\}$
Paulis propagate to Paulis

For example:
+1 X_L operator



+1 $Z_L X_L$ operator

Pauli operators propagate deterministically

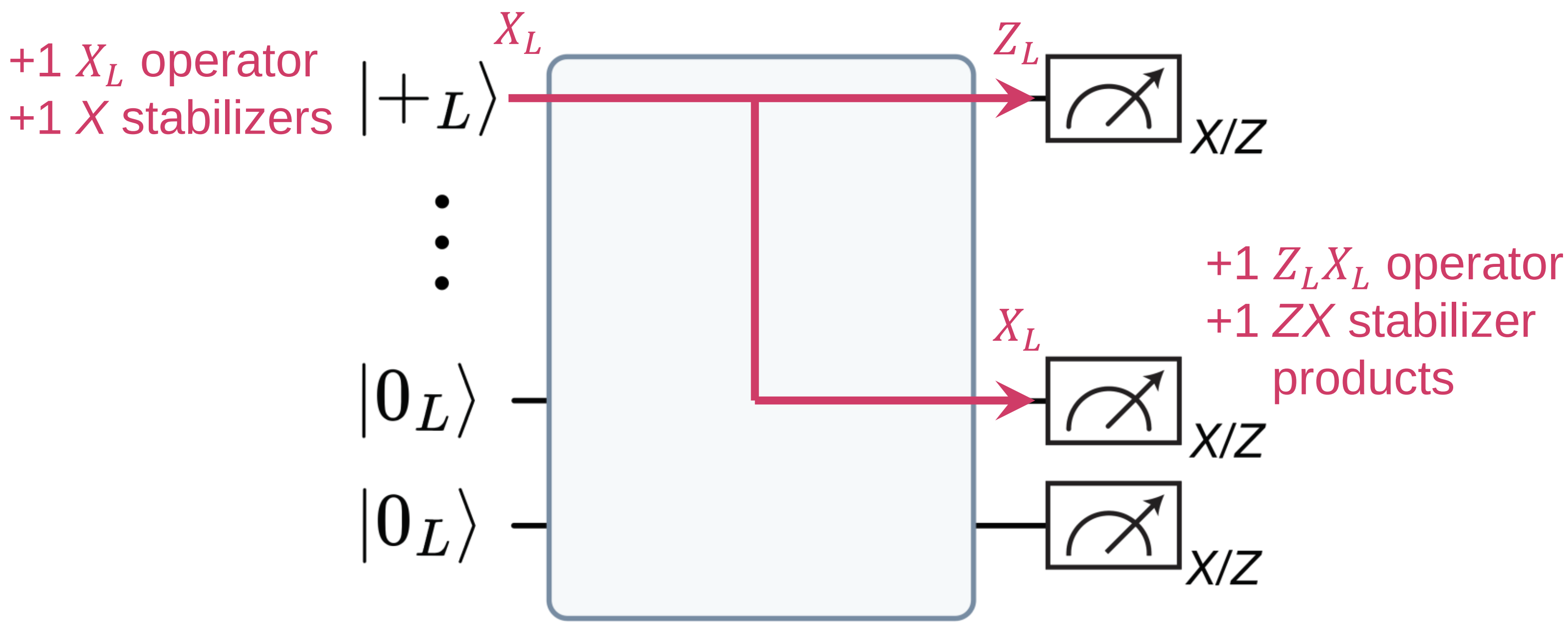
Fidelity:

- +1 logical operators *must be correct*
- Everything else is 50/50 random – *trivially correct!*

Can we predict the deterministic logical operators with correlated decoding? Yes!

Correlated decoding

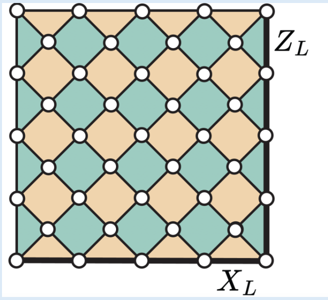
Transversal Clifford circuit: $\{H_L, S_L, CNOT_L\}$
Paulis propagate to Paulis



Logical operators and stabilizers propagate *the same way* through transversal gates!

Underlying physics: Clifford circuits propagate logical Pauli operators deterministically. These are inherently classical (i.e., not superpositions of operators).

→ Only need *one of our two classical codes* at a time to protect it.

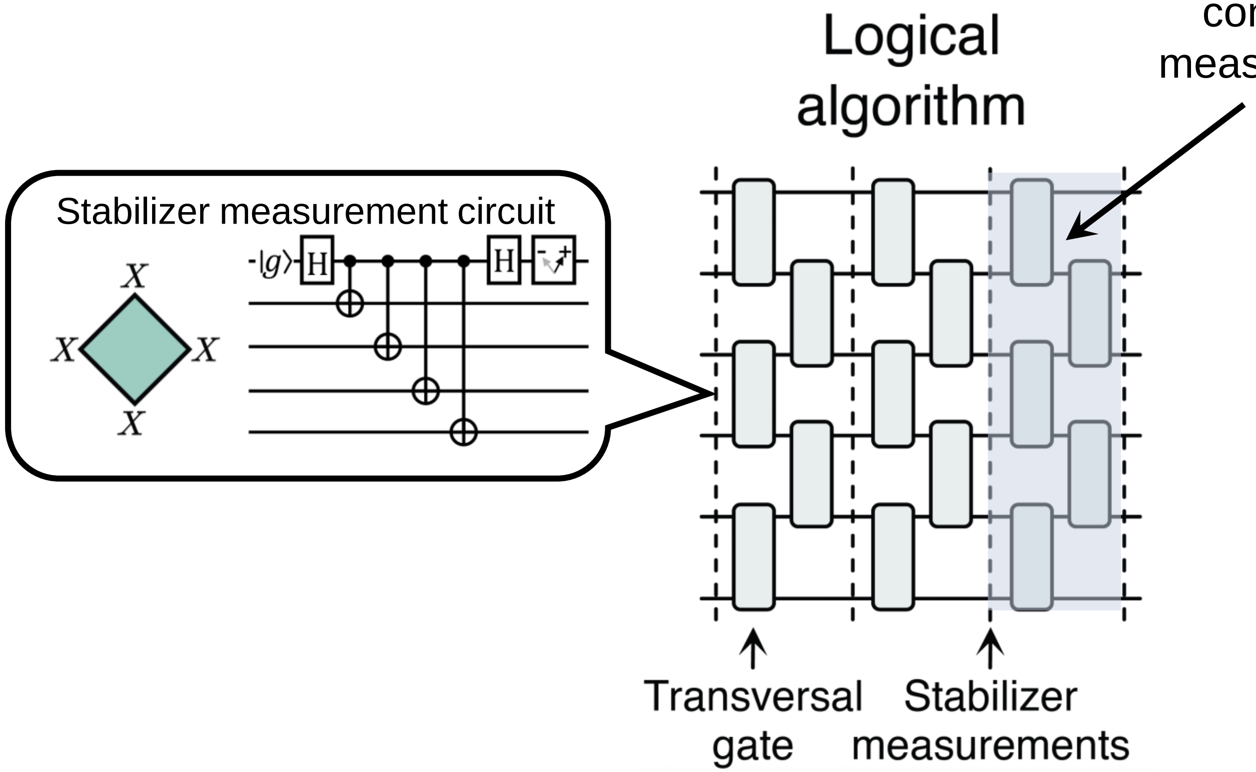


Stabilizers are reliable for the "useful" classical code.

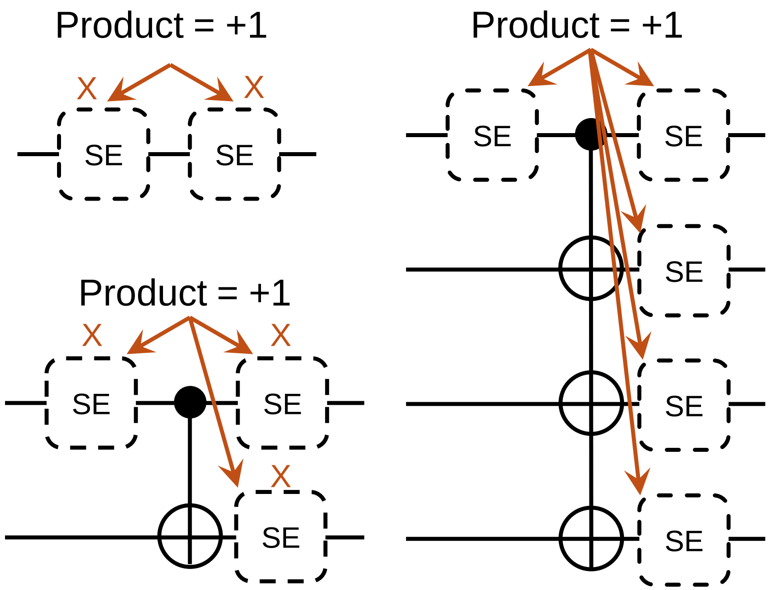
Decoding deep Clifford circuits

Although our final stabilizer measurements are reliable, errors can accumulate in deep circuits.
Need noisy ancilla stabilizer measurements to extract entropy and track errors through time.

Detect errors between consecutive stabilizer measurements via *checks*

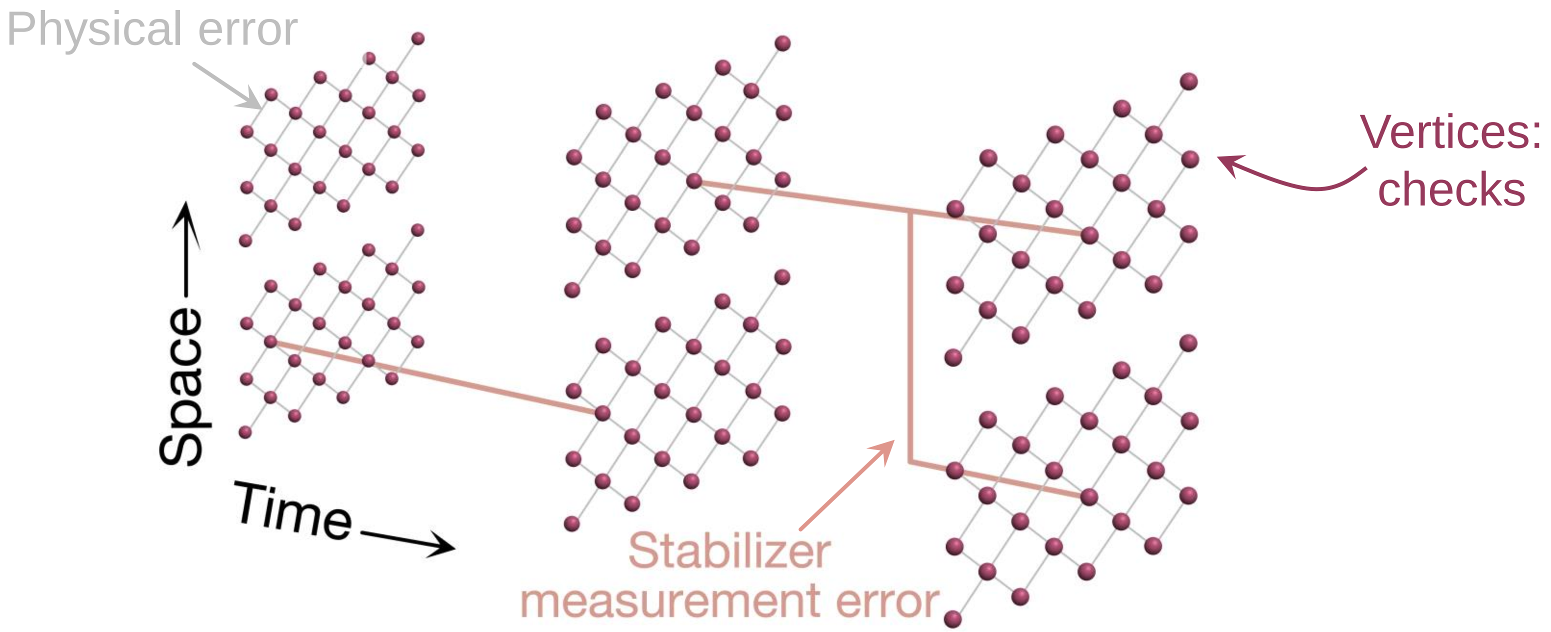


Checks: product of a stabilizer meas. and its previous measurement(s)



Errors which anti-commute with a stabilizer in a check flip it to -1

Decoding hypergraph



Decoding hypergraph:
input to decoder (use Stim)

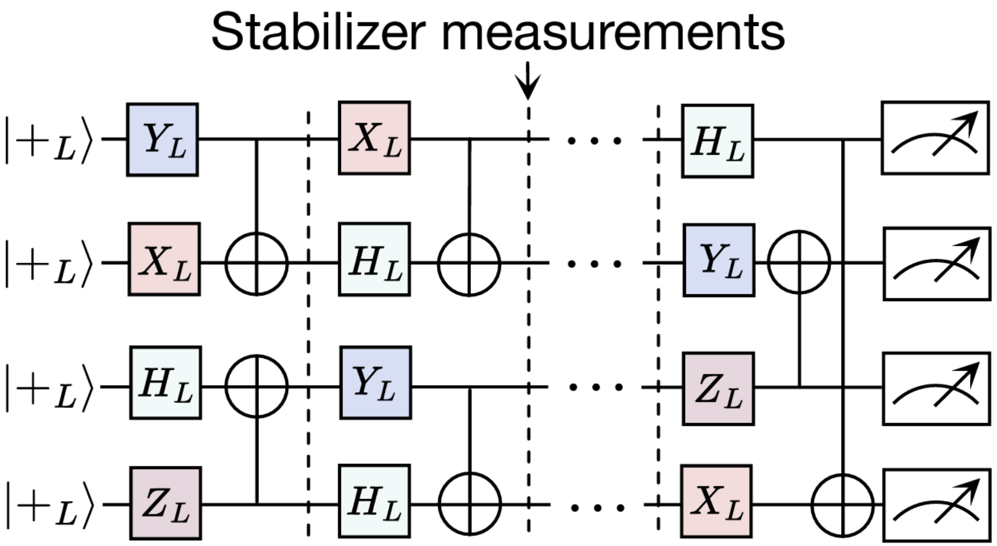
Decoder finds *likely error* consistent with syndrome.

With correlated decoding, track physical + stabilizer measurement errors *through* transversal gates – enables O(1) rounds per CNOT.

MC et al. PRL (2024) arXiv:2403.03272

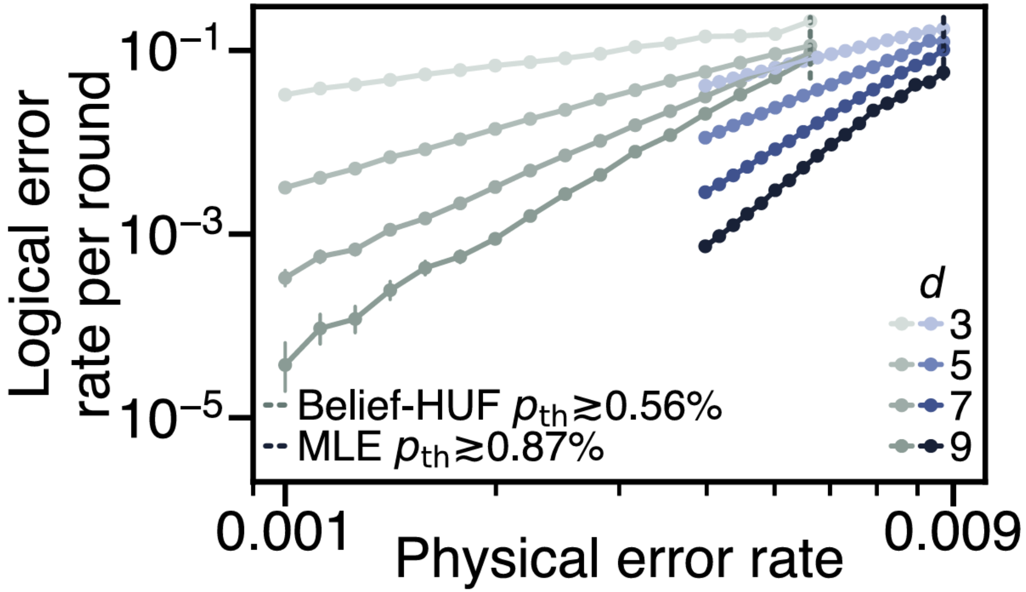
Correlated decoding → O(1) round per gate

Deep random logical Clifford circuits



1 SE round per CNOT

Two decoders: exact & approximate



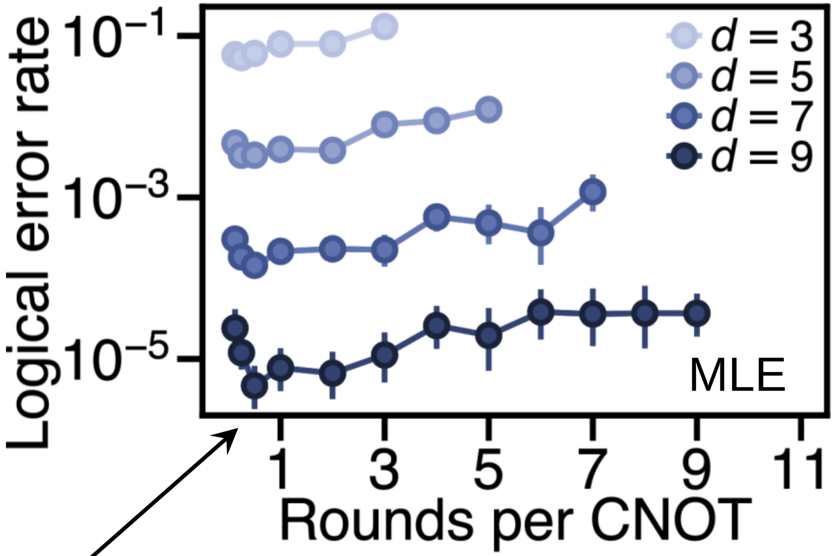
Threshold with 1 round per CNOT comparable to standard surface code threshold $\sim 1\%$ (compare to $d \approx 30$ rounds!)

MLE: Landahl et al., arXiv:1108.5738, HUF: Delfosse, Londe, Beverland IEEE 2022, BP+OSD Roffe, White, Burton, Campbell (2020)

MC et al. PRL (2024) arXiv:2403.03272

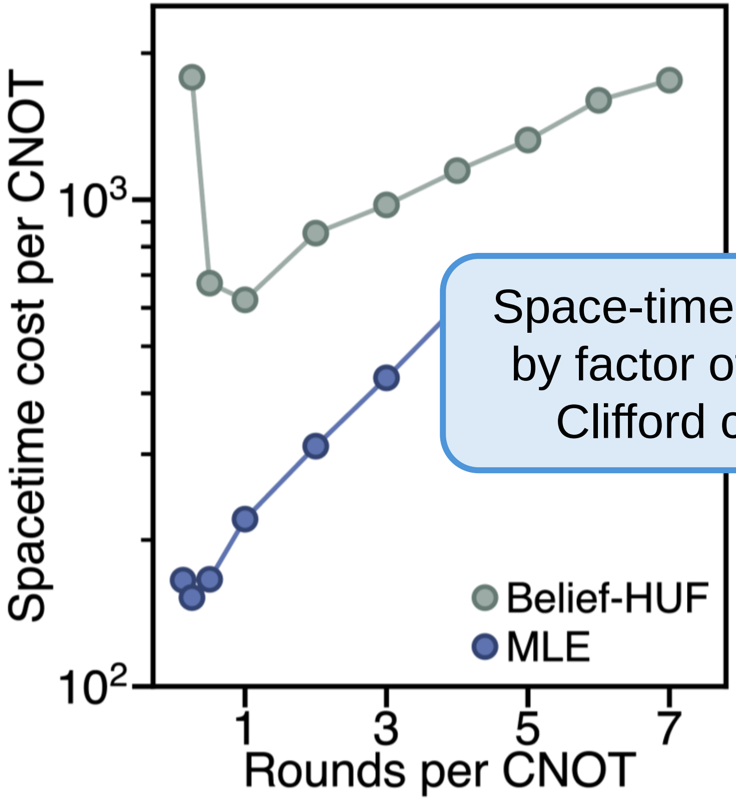
Correlated decoding → O(1) round per gate

Fix physical error rate and
vary rounds per CNOT



Logical error rate minimized
for < 1 round per CNOT:
multiple gates before SE

Compute spacetime cost =
$(\text{rounds per CNOT} + 1)d^2$
to reach target logical error rate 10^{-6}



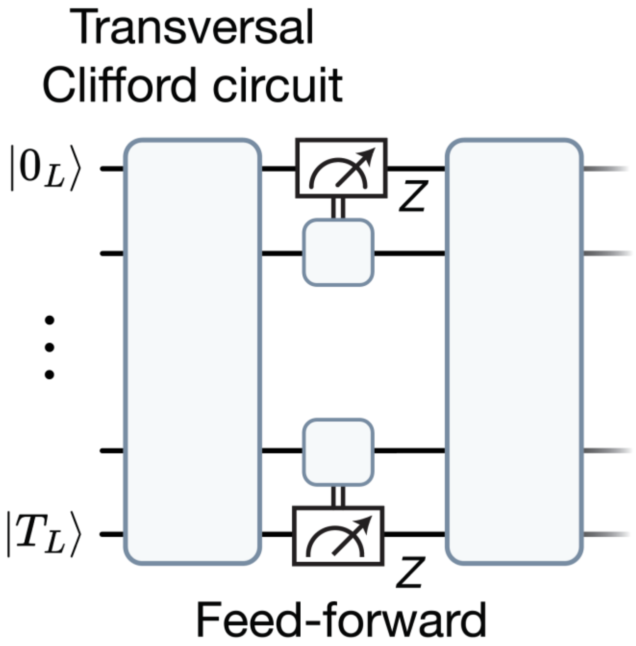
Space-time reduction
by factor of O(d) for
Clifford circuits!

MC et al. PRL (2024) arXiv:2403.03272

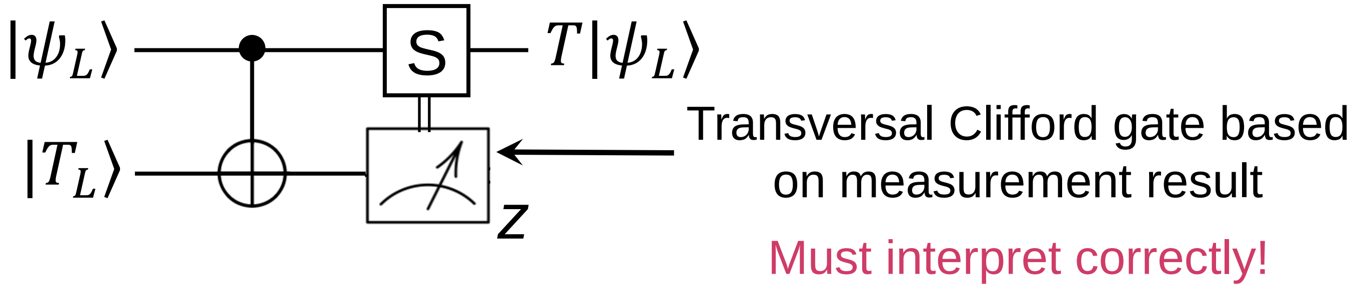
Generalization to universal computation

We need one more ingredient for universal computation: non-Clifford (magic) gates
Cannot do them transversally! (Eastin-Knill Theorem)

It suffices to have non-Clifford magic states $|T_L\rangle = |0_L\rangle + e^{i\pi/4}|1_L\rangle$ (assume reliable stabilizers) and feed-forward.



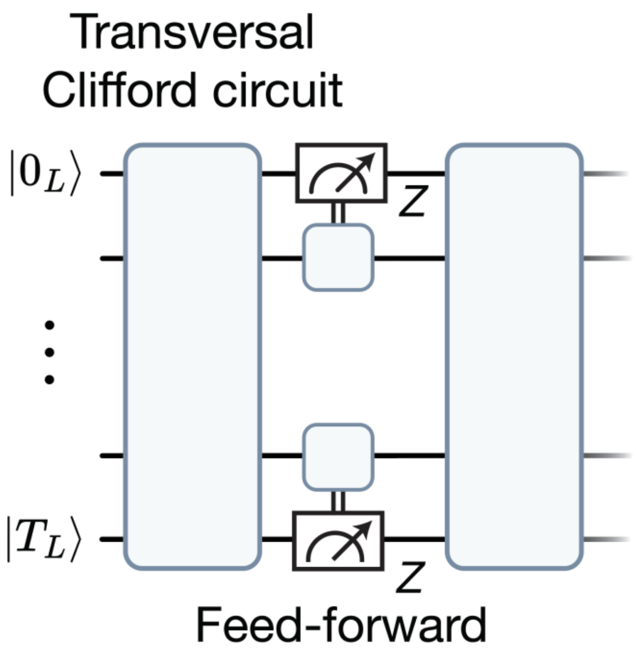
T gate teleportation

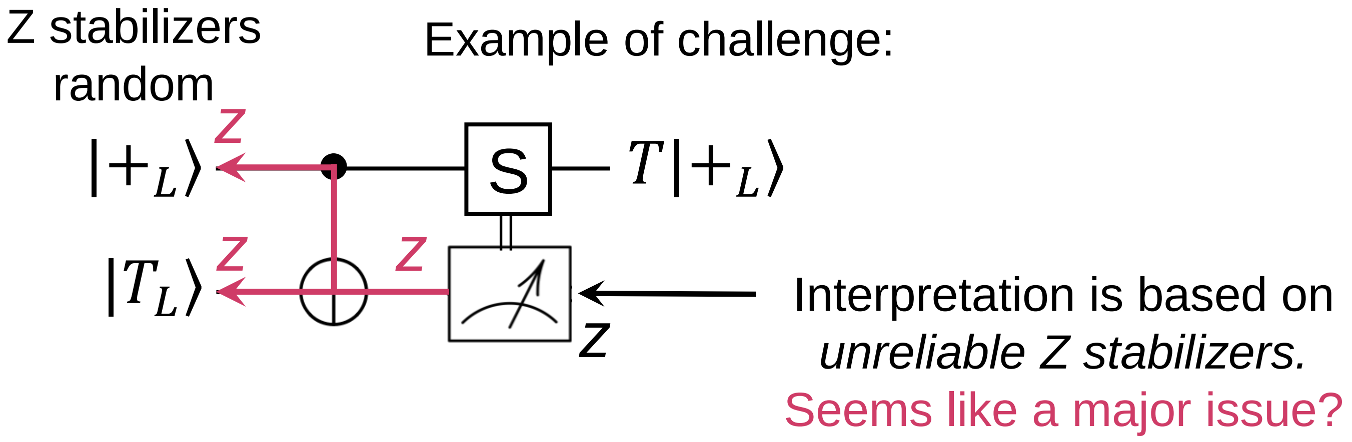


Generalization to universal computation

We need one more ingredient for universal computation: non-Clifford (magic) gates
Cannot do them transversally! (Eastin-Knill Theorem)

It suffices to have non-Clifford magic states $|T_L\rangle = |0_L\rangle + e^{i\pi/4}|1_L\rangle$ (assume reliable stabilizers) and feed-forward.



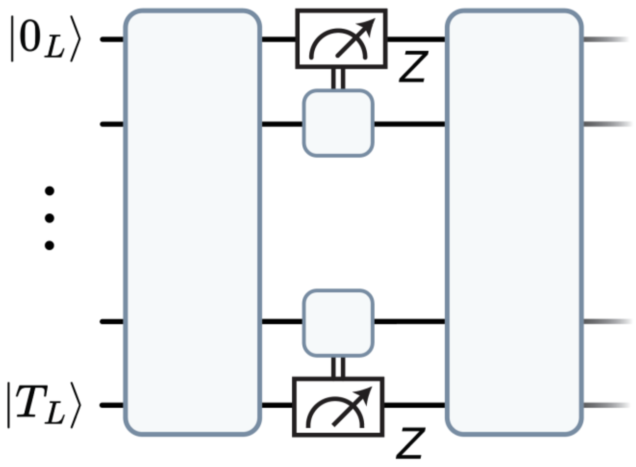


Generalization to universal computation

We need one more ingredient for universal computation: non-Clifford (magic) gates
Cannot do them transversally! (Eastin-Knill Theorem)

It suffices to have non-Clifford magic states $|T_L\rangle = |0_L\rangle + e^{i\pi/4}|1_L\rangle$ (assume reliable stabilizers) and feed-forward.

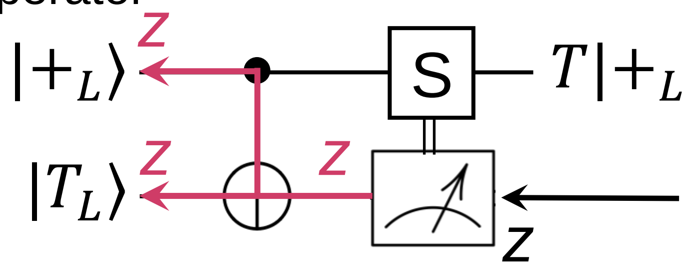
Transversal Clifford circuit



Feed-forward

Example of challenge:

+1 X_L operator

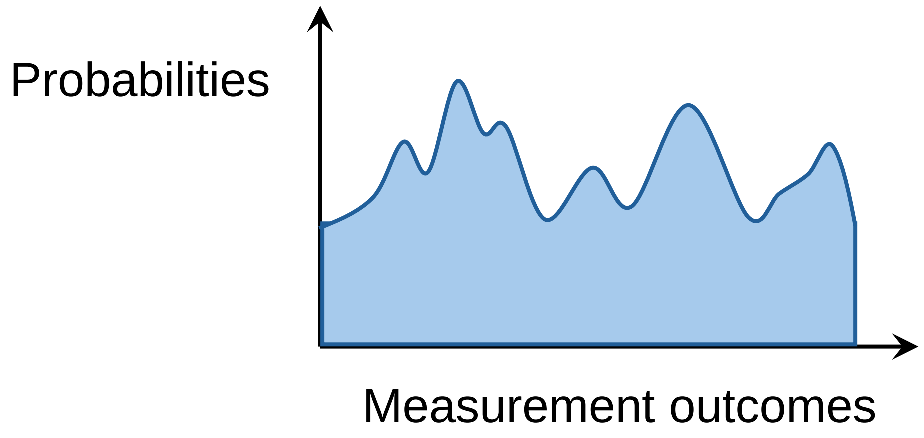


Interpretation is based on *unreliable Z stabilizers.*
Seems like a major issue?

However, this measurement anti-commutes with X_L - 50/50 random!
Can we play the same tricks as before, now with feedforward?

Reconstructing ideal measurement distribution

Minimal requirement for fault tolerance: reproduce the *ideal joint logical measurement distribution*.



$$P(z_j, z_{j-1}, \ldots) = P(z_j | z_{j-1}, z_{j-2}, \ldots) \cdot P(z_{j-1}, z_{j-2}, \ldots)$$

Joint measurement distribution

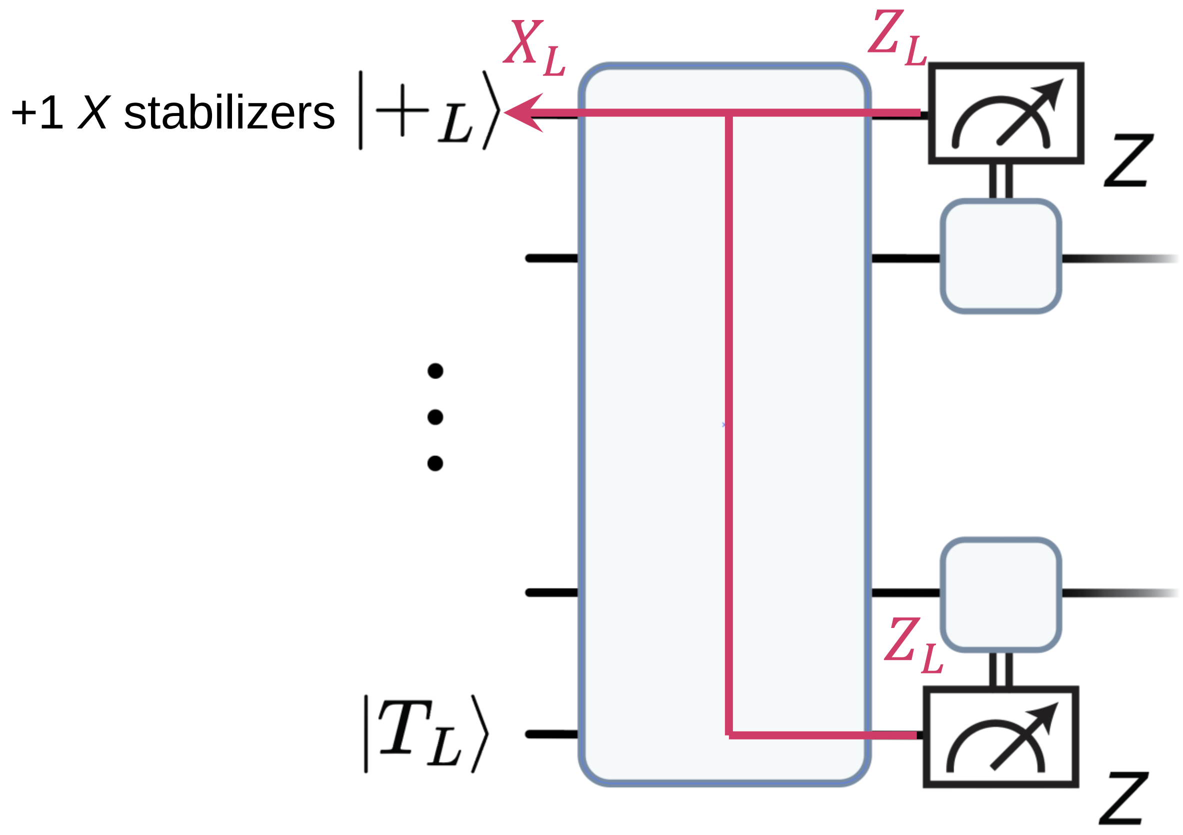
New measurement distribution, conditioned on previous results

Past measurement distribution

We don't need to know our stabilizers / state, we simply need to *sample new measurements from the ideal conditional distribution.*

Interpreting new measurements

Circuit executed so far (after feedforwards)
is a *transversal Clifford circuit*



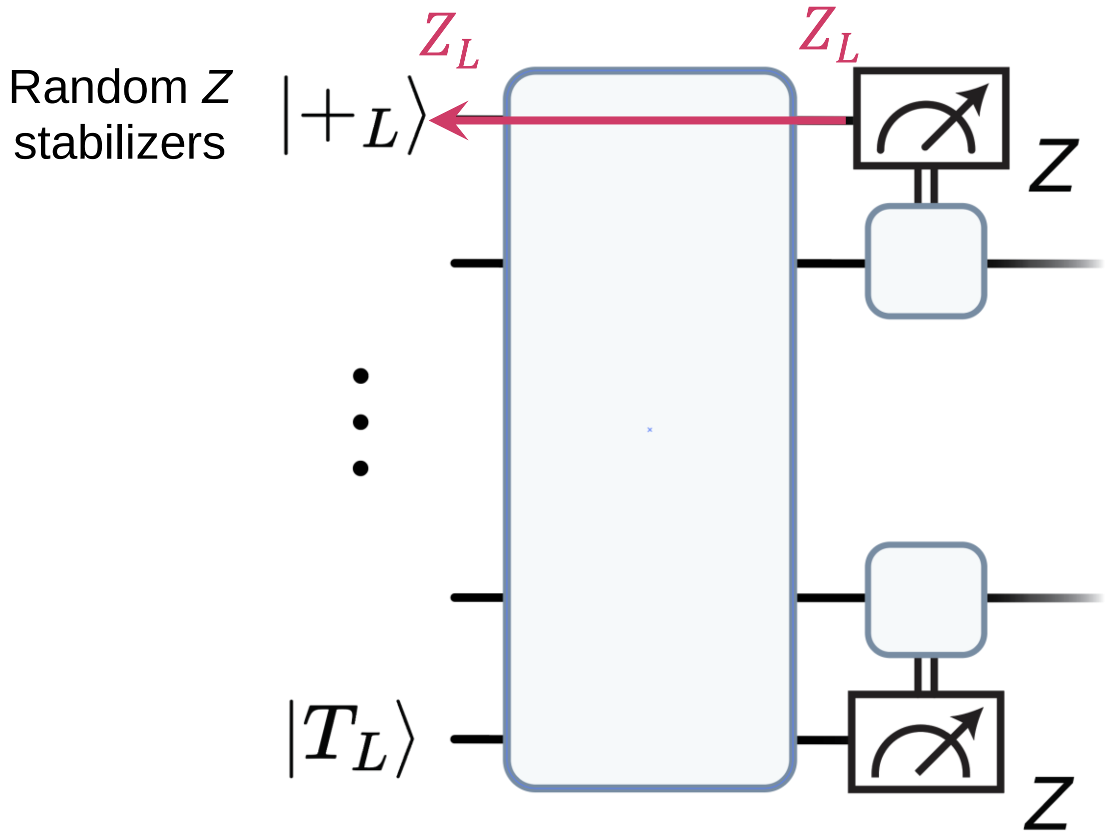
Analogous procedure to before:

1. Measurement product commutes with all logical Pauli operators at initialization

Then we *also have reliable stabilizers* → predict with correlated decoding

Interpreting new measurements

Circuit executed so far (after feedforwards) is a *transversal Clifford circuit*



Analogous procedure to before:

1. Measurement product commutes with all logical Pauli operators at initialization

   Then we *also have reliable stabilizers* → predict with correlated decoding

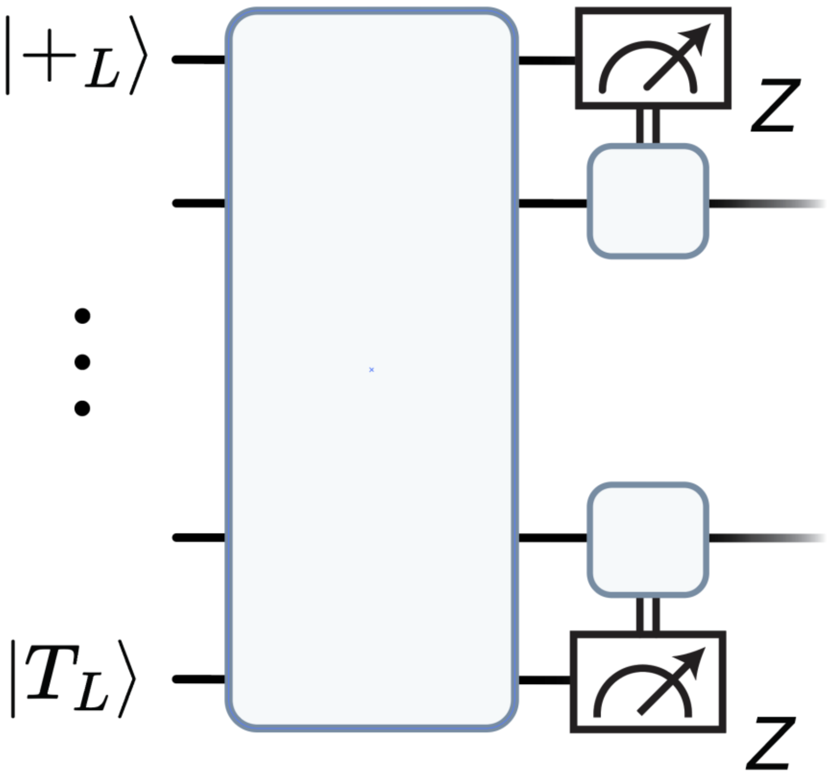
2. Measurement product anti-commutes with a logical Pauli operator at initialization

   50/50 random!

H. Zhou*, C. Zhao*, MC et al., arXiv:2406.17653

Interpreting new measurements

Circuit executed so far (after feedforwards)
is a *transversal Clifford circuit*



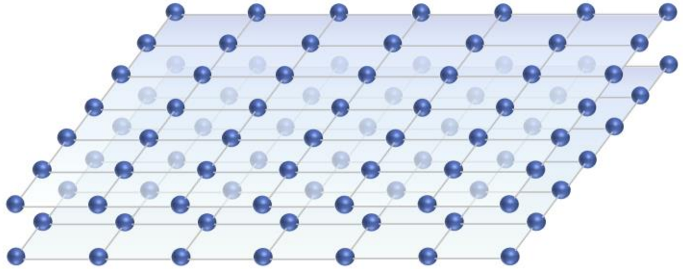
To interpret new measurements:

- Generate a basis of measurements which *commute with all logical Pauli operators at initialization* – decode with correlated decoding!
- All other measurements are 50/50 random – can be assigned randomly

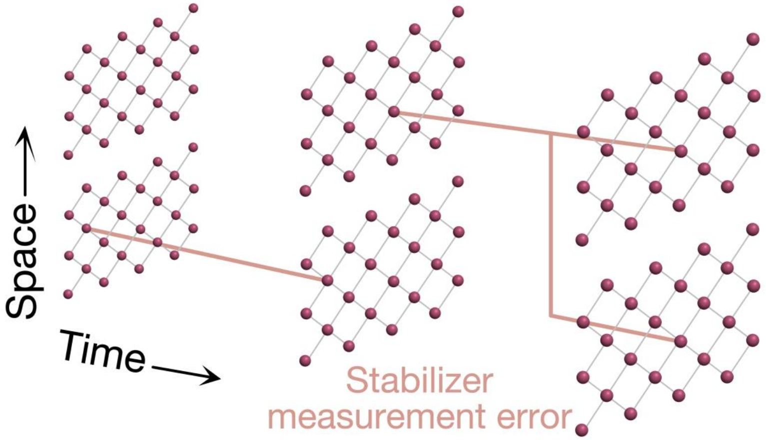
Threshold theorem: using this procedure, the logical error rate is exponentially suppressed with d for any qLDPC code.

H. Zhou*, C. Zhao*, MC et al., arXiv:2406.17653

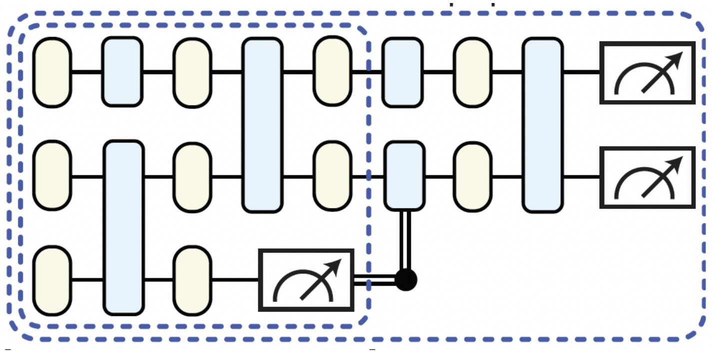
Summary



Transversal gates



Correlated decoding

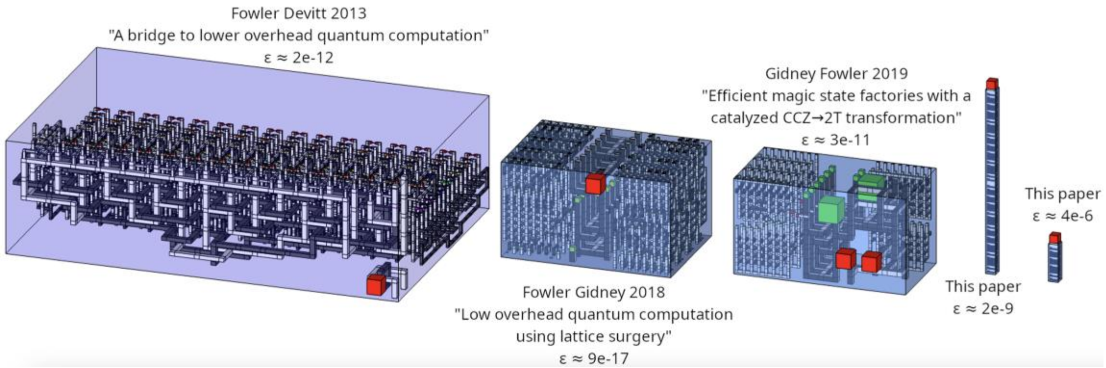


Handling feed-forward and non-Cliffords

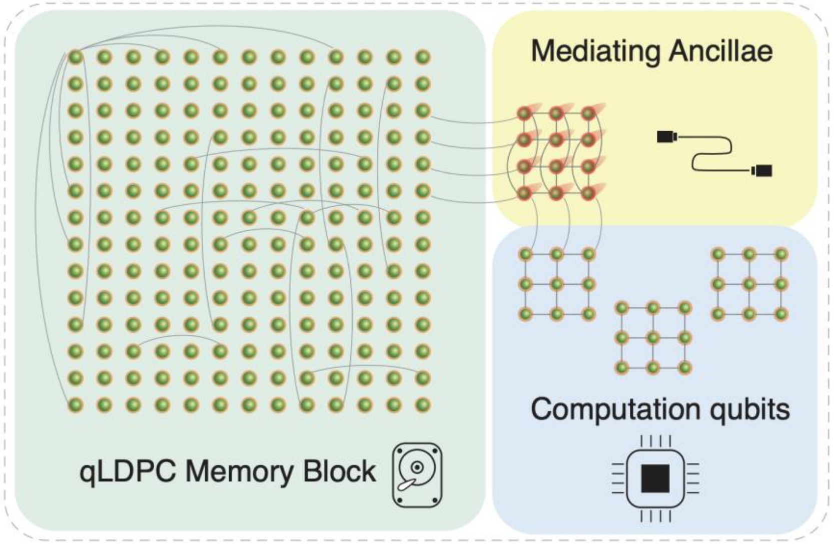
Factor of *d* circuit depth reduction for universal quantum computation

H. Zhou*, C. Zhao*, MC et al., arXiv:2406.17653
MC et al., arXiv:2403.03272
D Bluvstein,…, MC et al., Nature 2024

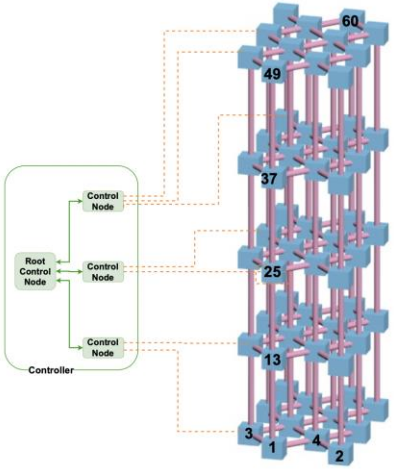
Outlook: low-overhead algorithms with QEC



Efficient ways to generate magic
E.g. cultivation, distillation
Gidney, Shutty, Jones arXiv: 2409.17595 (2024)
Recently explored experimentally by QuEra!
(used correlated decoding ☺) arXiv: 2412.15165



Low-space-overhead schemes,
e.g. high-rate qLDPC codes
Q. Xu*, P. Bonilla*, et al. Nature Physics 2024



Crucial to implement in practice:
develop fast decoders

Liyanage et al., QCE 2023

How much do the presented results + these developments reduce the resource estimates of algorithms?

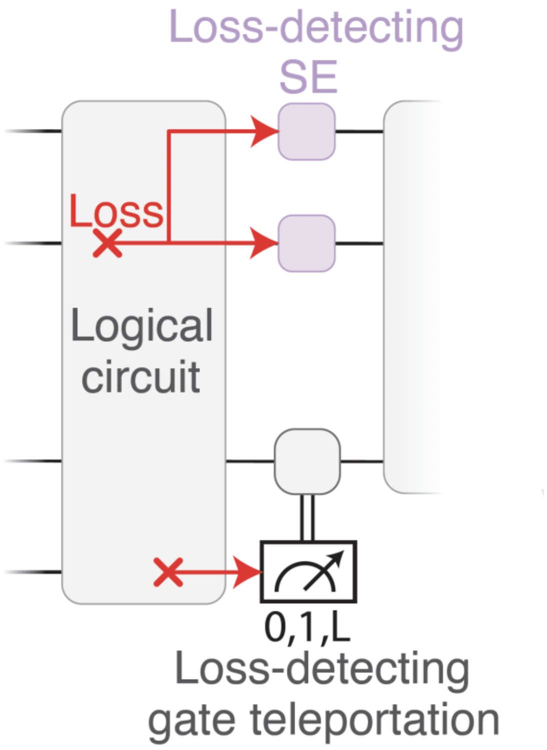
Outlook: tailor to realistic experimental errors

Atom loss: major source of experimental error, accounting for over half of entangling gate errors.

Knowledge of error/loss location → more accurate QEC!

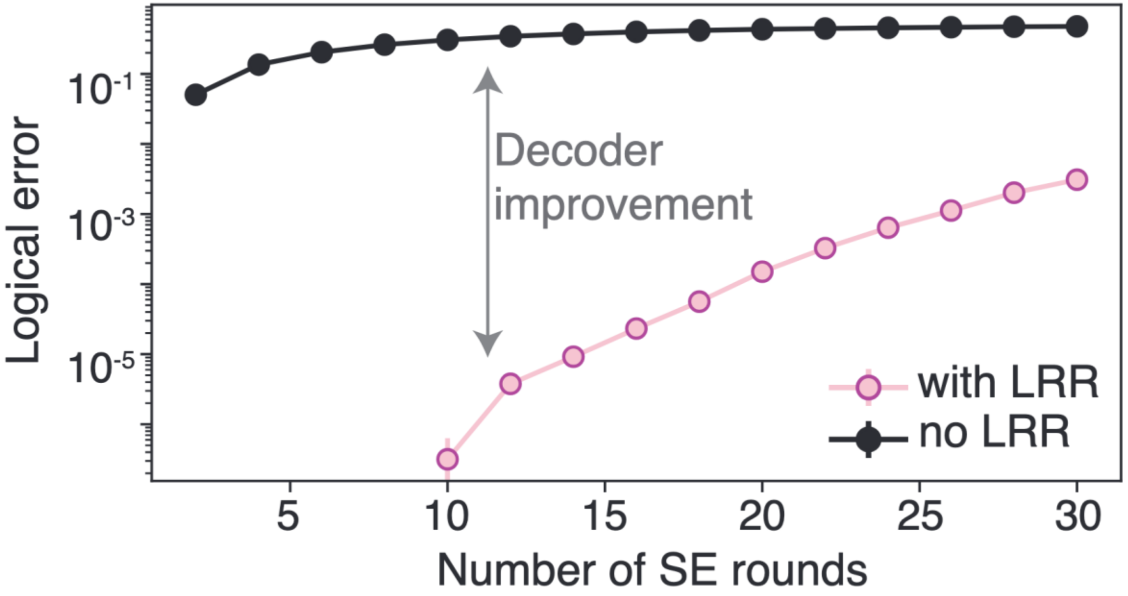
Loss-resolving readout (LRR): measure $|0\rangle, |1\rangle, |\text{loss}\rangle$

Integrate LRR with minimal overhead in logical algorithms



Gefen Baranes Pablo Bonilla

Gefen Baranes*, MC*, Pablo Bonilla* et al. (2025)

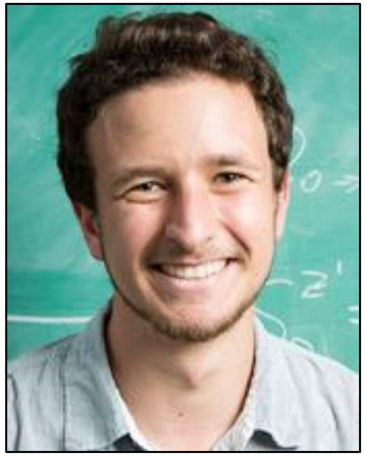


Foundational work: Wu et al. Nature Communications 13, 4657 (2022), Scholl et al. Nature 622, 273 (2023), Sahay et al. Phys. Rev. X 13, 041013 (2023), Ma et al. Nature 622, 279 (2023), Cong et al. Phys. Rev. X 12, 021049 (2022).

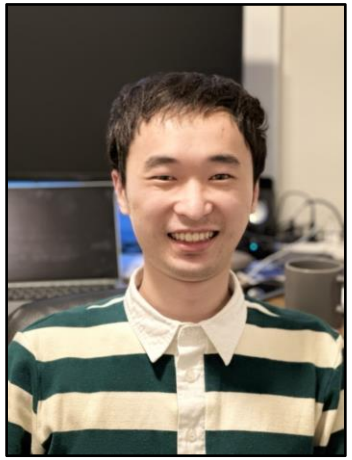
Recently: Reichardt et al. arXiv:2411.11822, C.-C. Wu et al. arXiv:2411.04664, Perrin et al. arXiv:2412.07841, Chow et al. arXiv:2405.10434

Acknowledgements

Harry Zhou (QuEra)



Dolev Bluvstein (Harvard)



Chen Zhao (QuEra)

Alex Kubica (AWS → Yale)

Arthur Jaffe (Harvard)

Mikhail Lukin (Harvard)

Funding: IARPA ELQ, CUA, AFOSR MURI, DARPA IMPAQT, DOE, NDSEG, Hertz

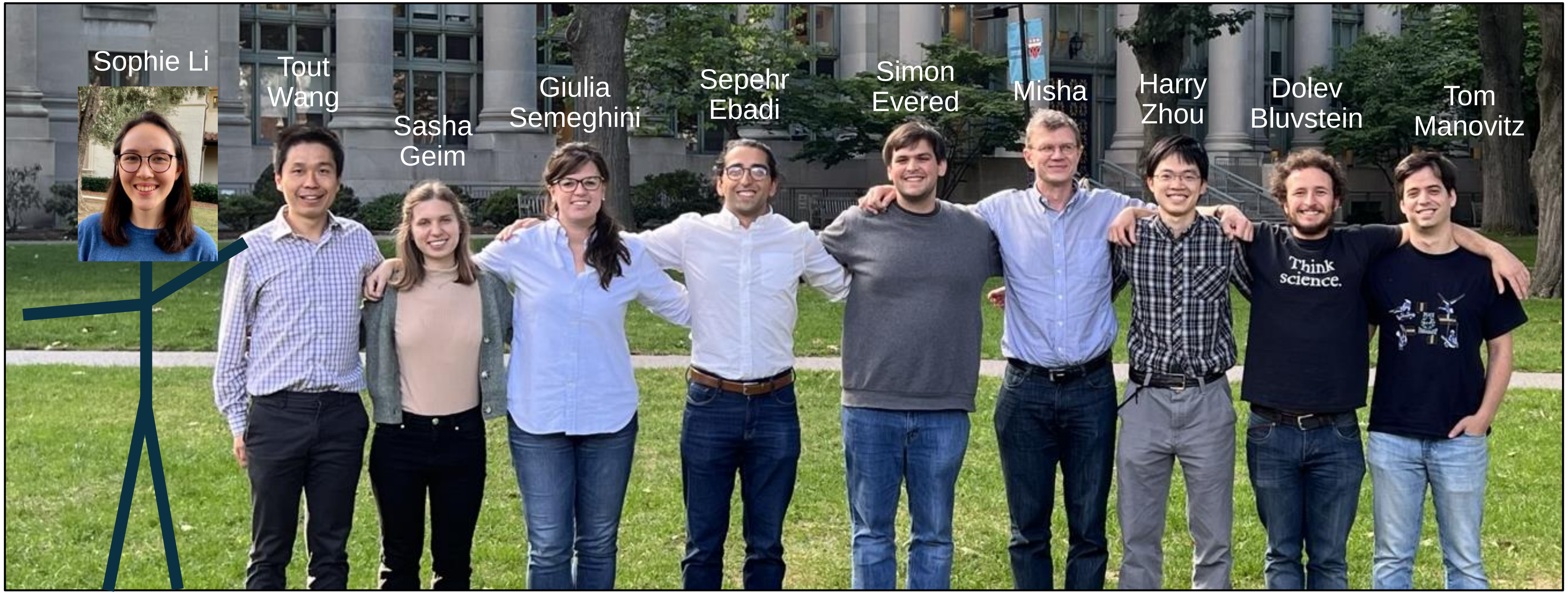
Casey Duckering (QuEra)

Hong-Ye Hu (Harvard)

Shengtao Wang (QuEra)

Nadine Meister (Harvard)

Pablo Bonilla (Harvard)



★ ***Atom Array II:***

Tout Wang, Tim Guo, Allen Chiu, Pavel Stroganov, Mohamed Abobeih, Simon Hollerith, Sebastian Geier

★ ***QuEra:***

P Sales Rodriguez

T Karolyshyn

PIs

M. Lukin

V. Vuletic

M. Greiner

Theory team

N. Maskara

I. Cong

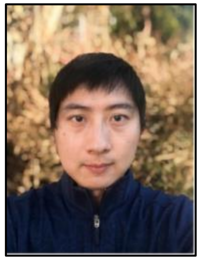
JP Bonilla Ataides

M. Kalinowski

M. Gullans

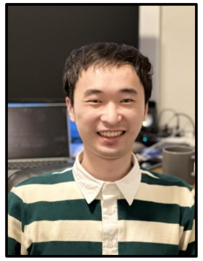
D. Hangleiter

X. Gao



N. Meister

C. Zhao



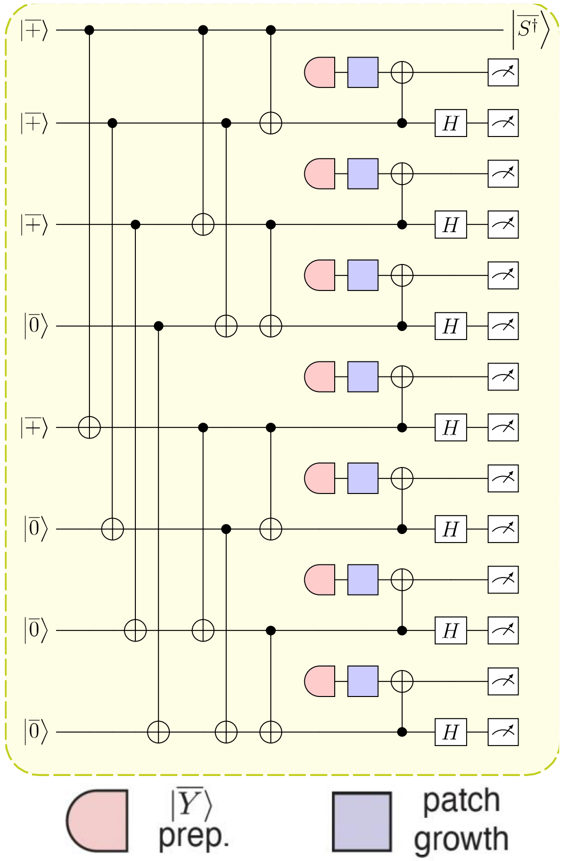
Funding: QuEra, NSF, CUA, Vannevar Bush, AFOSR MURI, DARPA, DOE, NDSEG, Hertz

Thank you!

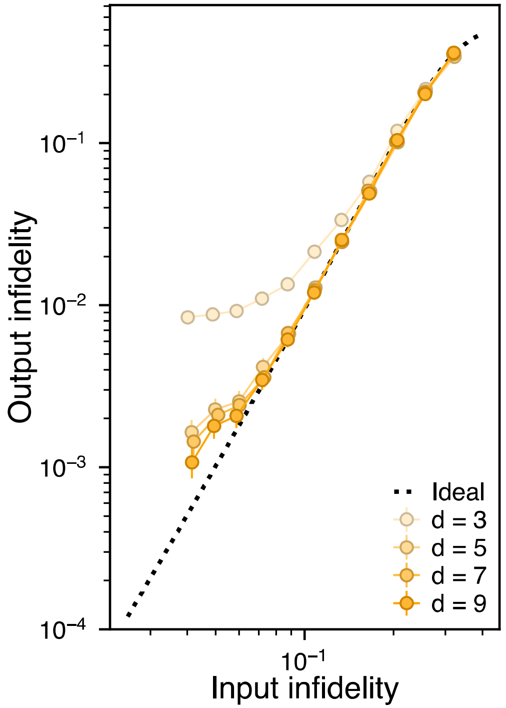
Practical performance: state distillation

Key subroutine in large-scale algorithms: generate good resource states (e.g. T states) by distilling from noisy states

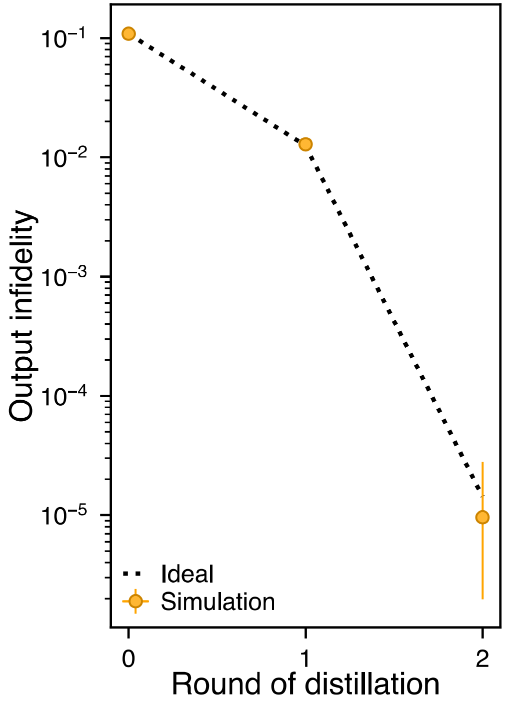
Here: Distill $|Y\rangle = S|+\rangle$



Output error vs. input error



Output error vs. # factory stages



Future research: in-depth analysis of multi-level magic state distillation.
Additional opportunities: explore magic state cultivation
Gidney, Shutty, Jones arXiv: 2409.17595 (2024)

H. Zhou*, C. Zhao*, MC et al., arXiv:2406.17653