Quantum AI

Magic State Cultivation

Craig Gidney
Simons Institute Quantum Colloquium
April 2025

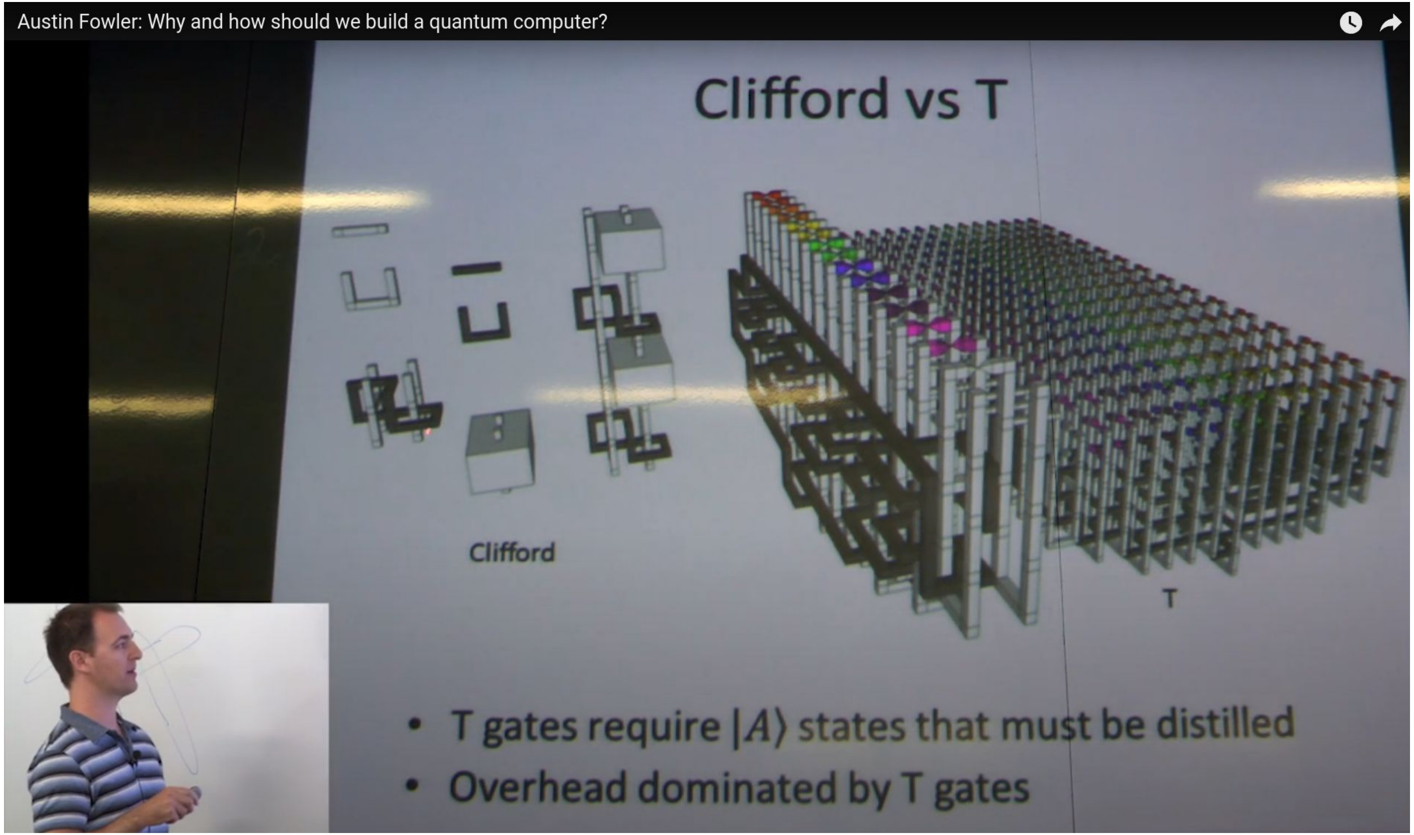
Google

10 years ago: **94% of quantum computation is T factories**

states translates to then needing about 1200 AA factories working in parallel, and thus $1200 \times 8 \times 10^5 \approx 10^9$ physical qubits. The remainder of Shor's algorithm requires about $2N = 4000$ logical qubits, which in a $d_2 = 34$ distance surface code takes about $4000 \times 14500 \approx 5.6 \times 10^7$ additional physical qubits, adding a fairly negligible 6% to the AA factory footprint, for a total of about a billion physical qubits.

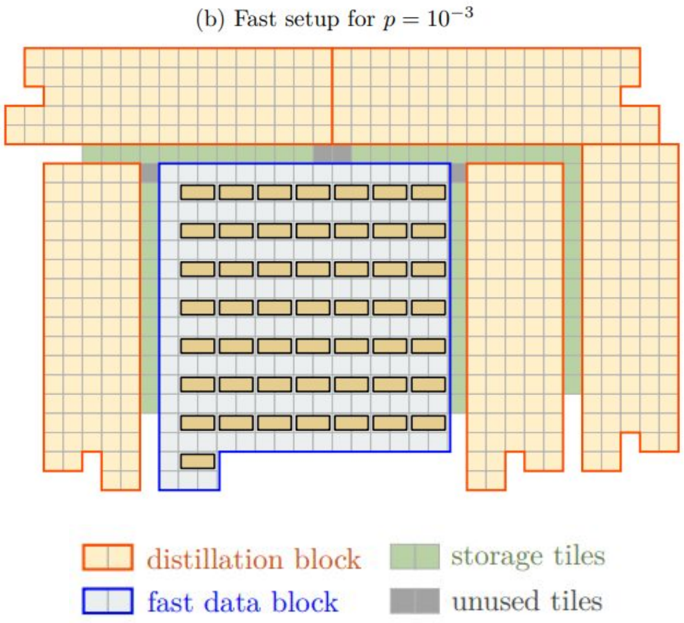
Fowler et al 2012
"Surface codes: Towards practical large-scale quantum computation"
arXiv:1208.0928

10 years ago: **94% of quantum computation is T factories**

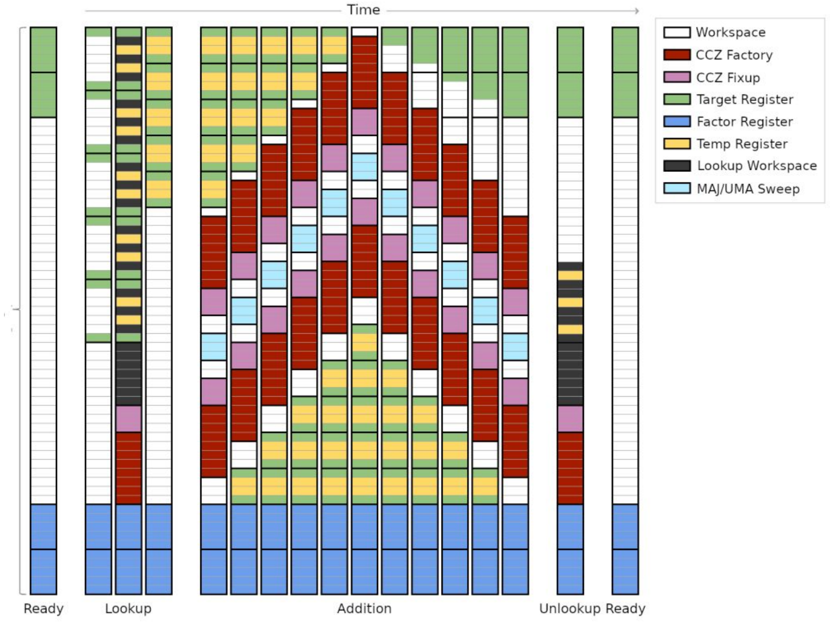


youtube.com/watch?v=XbFoXT73xVQ

5 years ago: **30-70% of quantum computation is T factories**

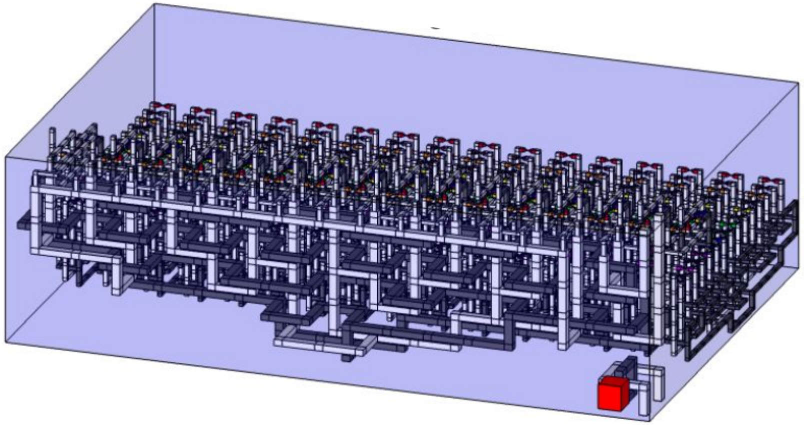


Litinski 2018
"A Game of Surface Codes: Large-Scale Quantum Computing with Lattice Surgery"
[arXiv:1808.02892]



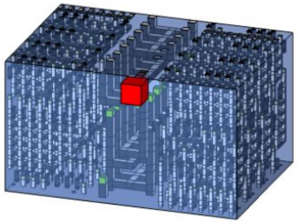
Gidney et al 2019
"How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits"
[arXiv:1808.02892]

2014



1e-12 error

2019



1e-12

This Talk

1e-9 1e-6

(assuming 1e-3 physical noise)

Major improvements over the past 10 years

- Postselect the injection

  Li 2014 arXiv:1410.7808

- Use lattice surgery instead of braiding

  Horsman+ arXiv:1111.4022

  Fowler+ arXiv:1808.06709

  Litinski arXiv:1808.02892

- Reduce distance where the distillation code protects the computation code

  Litinski arXiv:1905.06903

- Use complementary gap to get better attempts-vs-errors tradeoffs

  Bombin+ arXiv:2212.00813

  Gidney+ arXiv:2312.04522

Major improvements over the past 10 years

- Postselect the injection

  Li 2014 arXiv:1410.7808

- Use lattic

ects the computation code

- U ry gap to get better attempts-vs-errors tradeoffs

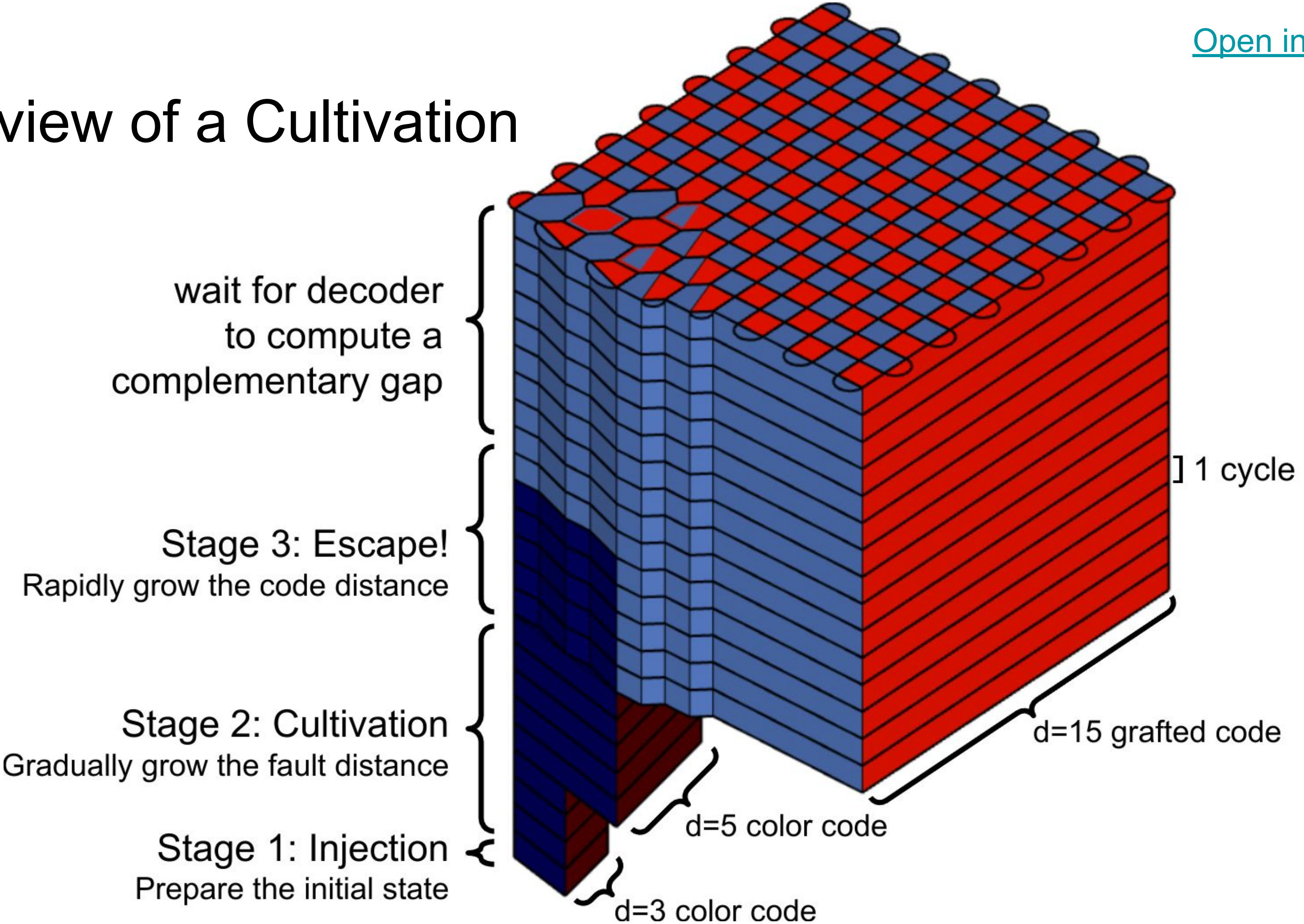
  + arXiv:2212.00813

  Gidney+ arXiv:2312.04522

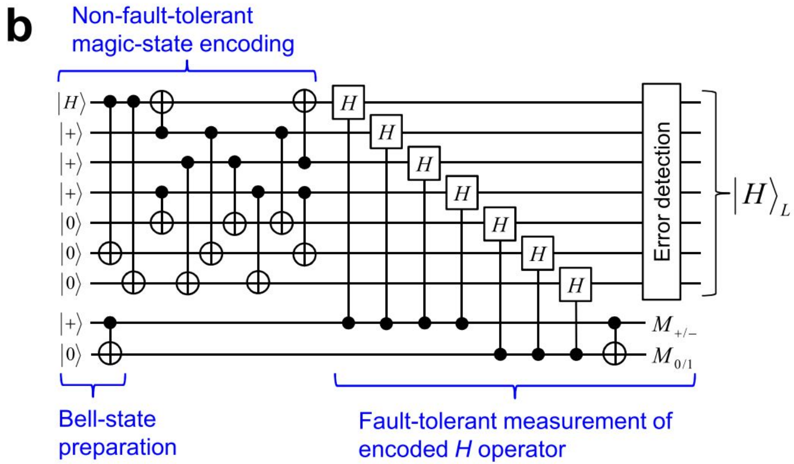
Very different from a list of theory advances!

[Open in Crumble]

Overview of a Cultivation



History of magic state cultivation - 2015



Goto 2015

"Minimizing resource overheads for fault-tolerant preparation of encoded states of the Steane code"

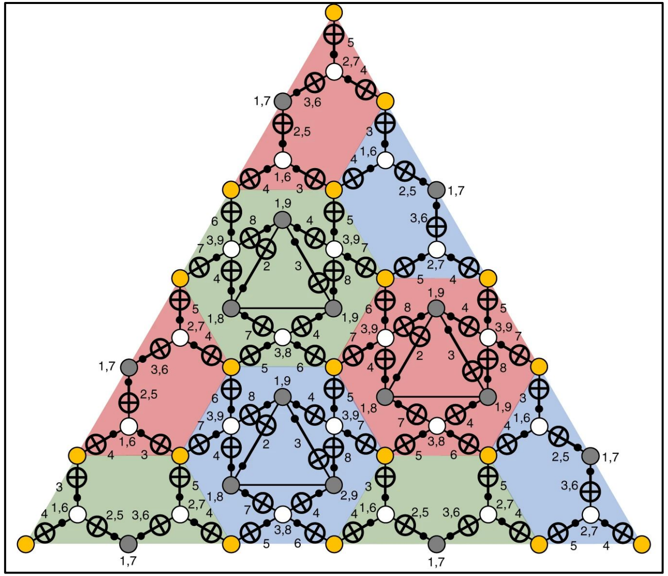
(no arXiv preprint?) [doi: 10.1038/srep19578](https://doi.org/10.1038/srep19578)

- Encode |T⟩ into a distance 3 color code

- Measure transversal Clifford to verify |T⟩
  - Prepare a bell pair |00⟩+|11⟩
  - Use it to control a transversal H_{XY}
  - Phase kickback produces |00⟩-|11⟩ if |T⟩ is wrong
  - Measure XX to distinguish |00⟩±|11⟩
  - Measure ZZ to flag hook errors

Cons: d=2 fault distance, ends in small code

History of magic state cultivation - 2020

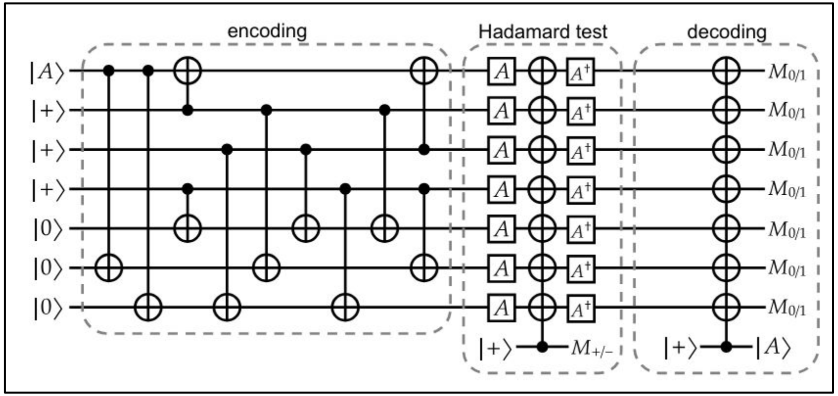


Chamberland et al 2020
"Very low overhead fault-tolerant magic state preparation using redundant ancilla encoding and flag qubits"
arXiv:2003.03049

- Can target any suppression d
- Encode |T⟩ into a distance 2d+1 color code
- Control transversal Clifford with a cat state
- Alternate measuring stabilizers and Clifford
    - Requires just-in-time decoding

Cons: 1e-4 noise, ends in small code

History of magic state cultivation - 2024



- Cultivation works at 10^{-3} noise
- Only need simple grid connectivity

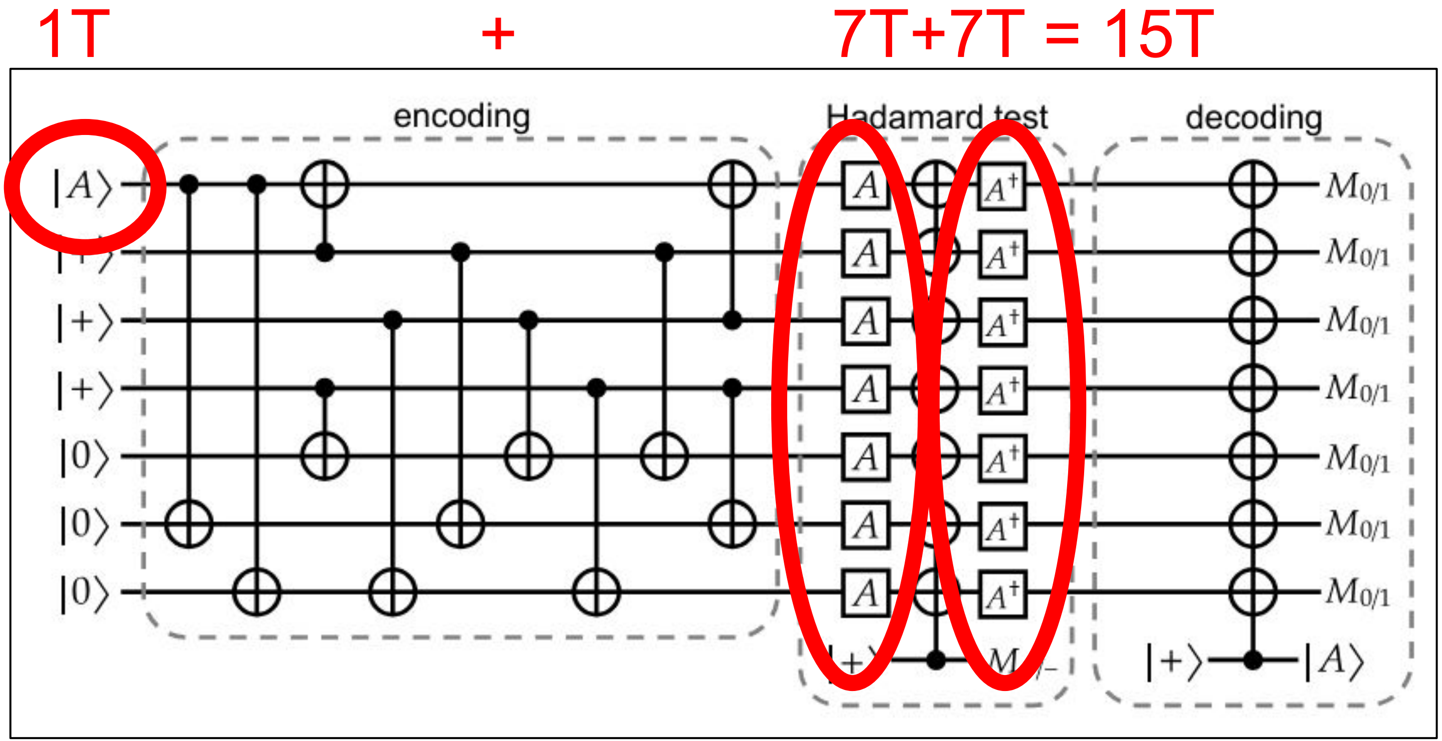
Itogawa et al 2024

"Even more efficient magic state distillation by zero-level distillation"

[arXiv:2403.03991]

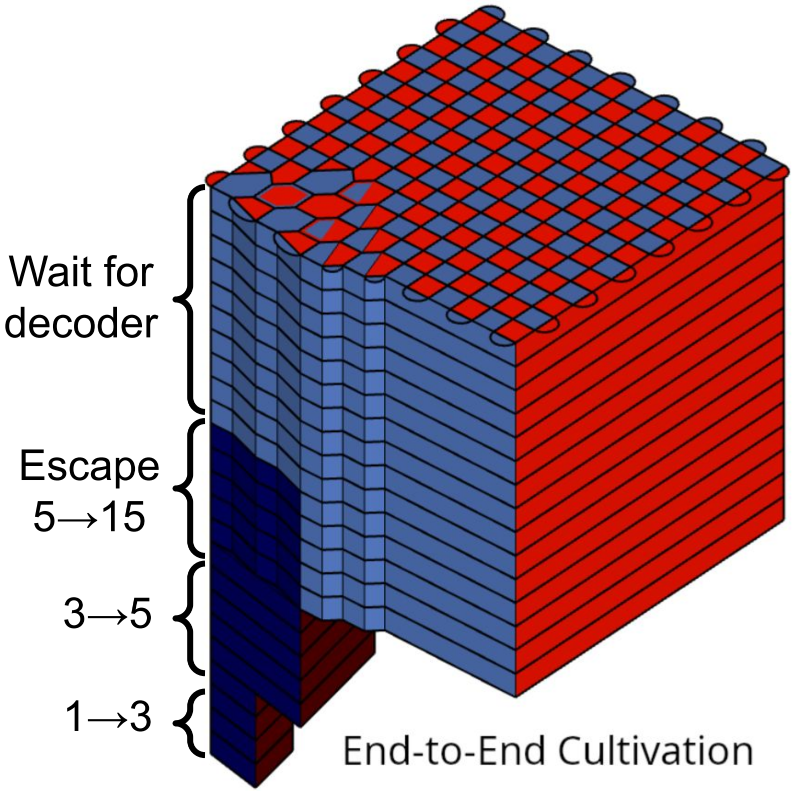
Cons: d=2 fault distance, ends in small code

This is a 15-to-1 distillation jammed into a color code



Another tell: **35 triplets** of T -> ZT errors that make it fail

Our Improvements



Decrease costs, enabling postselection

- Grow color code incrementally
- Pair up T checks to reduce their cost
- Use superdense color code cycle

End with the state in a safe place

- Escape to a big surface code by "grafting"
- Good soft information via complementary gap

Was 1e-4 logical error → Now 2e-9

Reasons for the new name

Cultivation

- Gradually improve one state

- Key operations are physical
  (within one color code)

- Can't reach arbitrarily low error
  (cost wall from postselection)

- Ends with an exciting escape scene

Distillation

- Turn many states into fewer better states

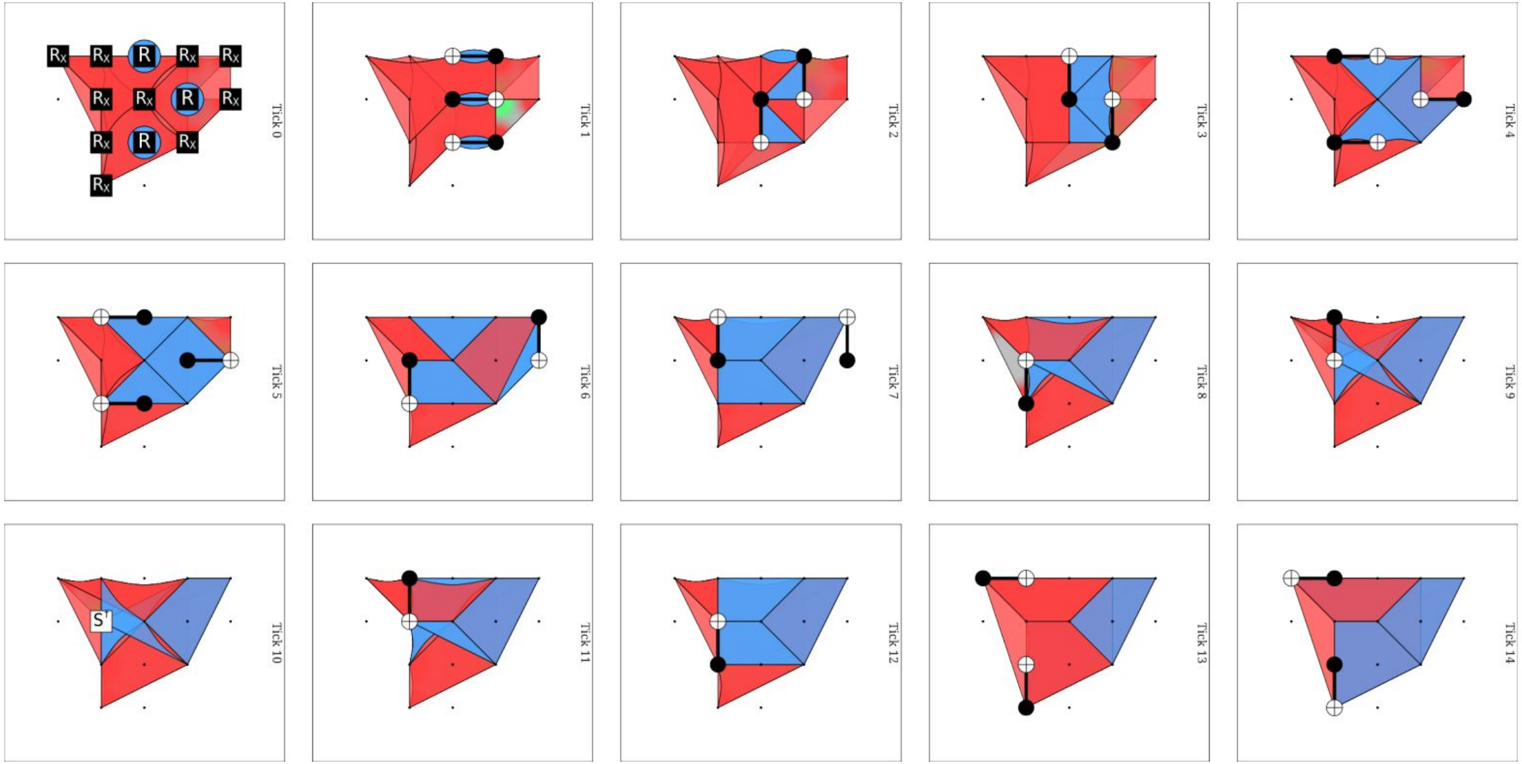
- Key operations are logical
  (across many surface codes)

- Reaches arbitrarily low error
  (just keep iterating it)

- States always in sufficiently large codes

Stage 1: Injection

[Open in Crumble]



Stage 2: Cultivation

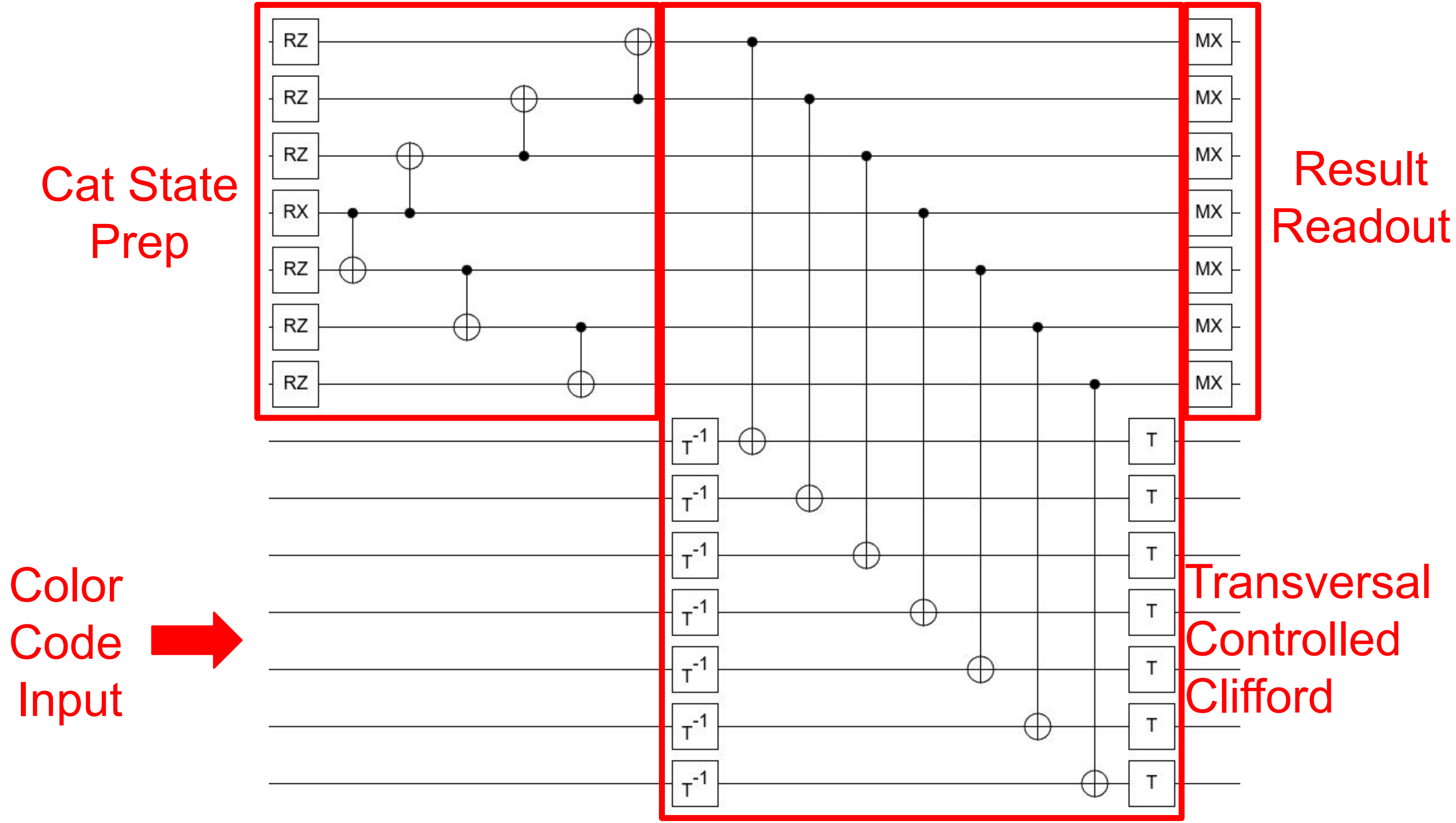
Grow the color code

Stabilize the color code

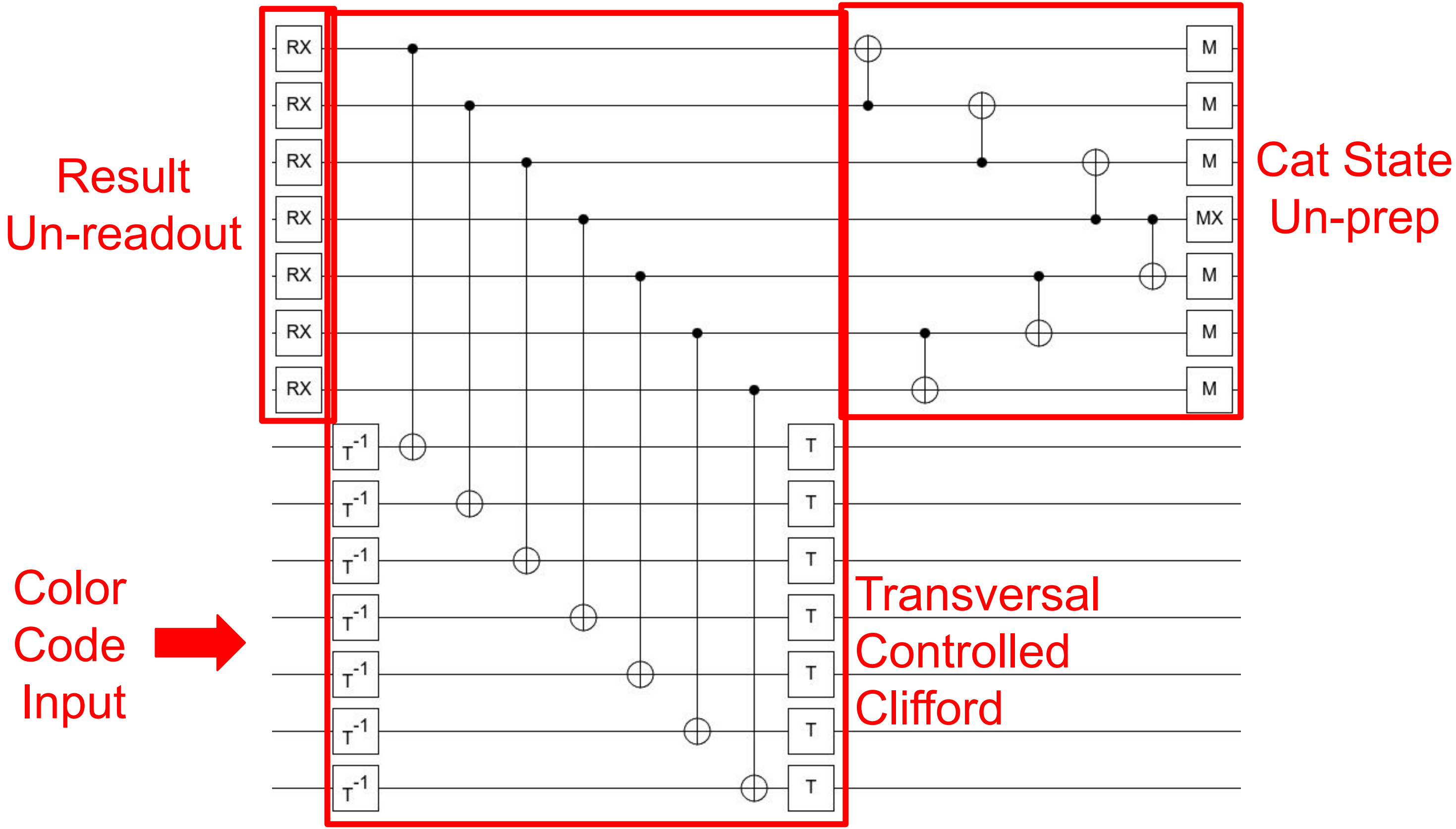
Double-check the T state

Repeat (until approaching the postselection cost wall)

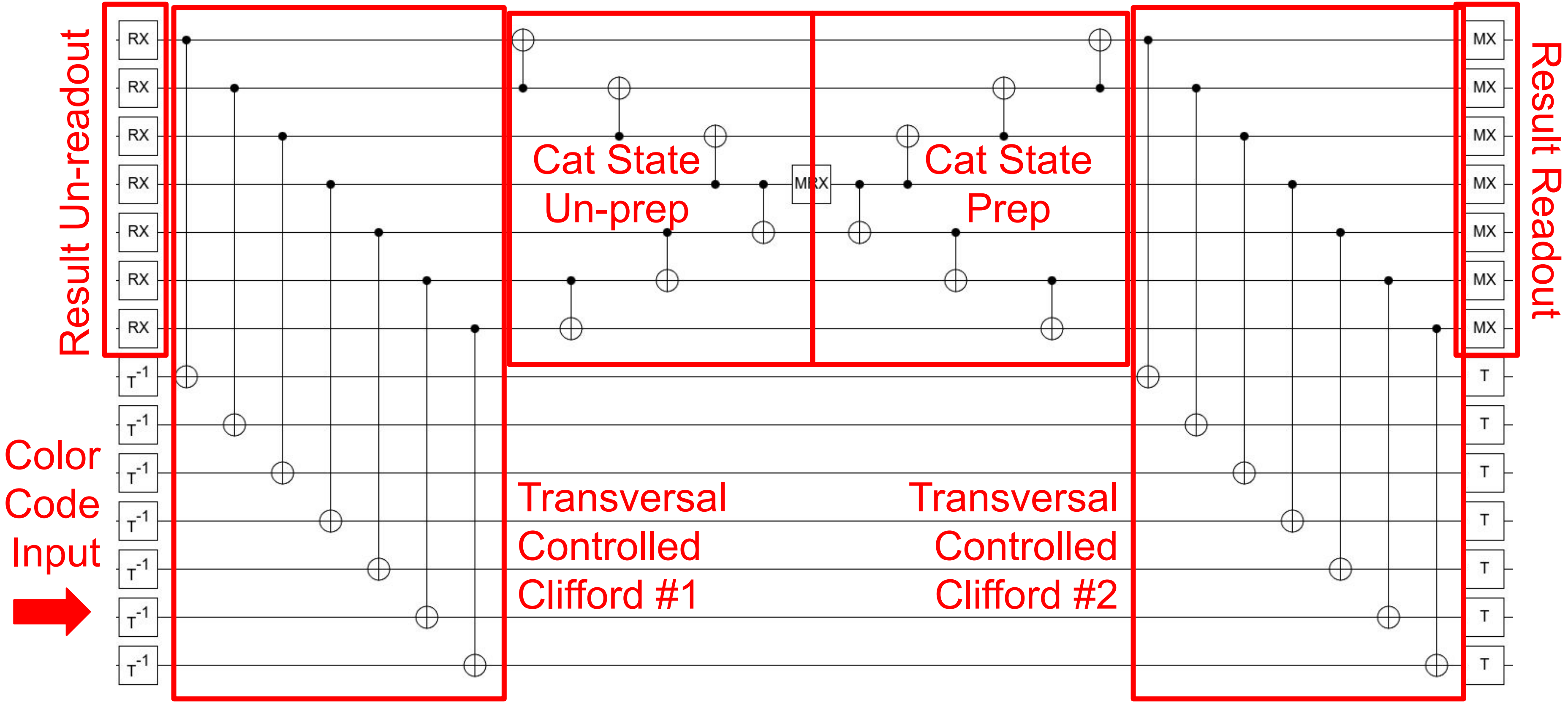
Example Kickback Measurement by Cat State



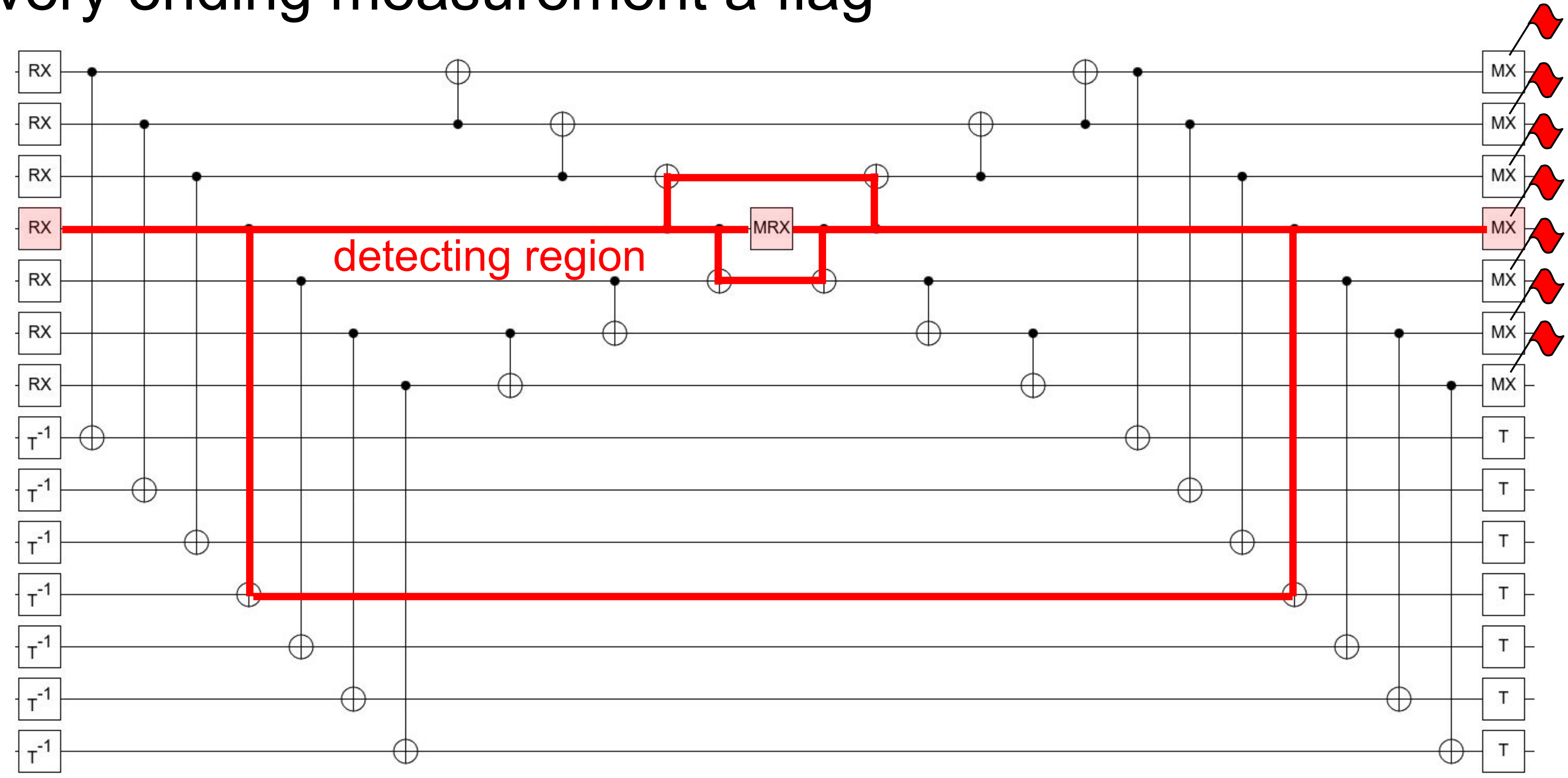
Reverse Example Kickback Measurement by Cat State



Double-Check Kickback Measurement



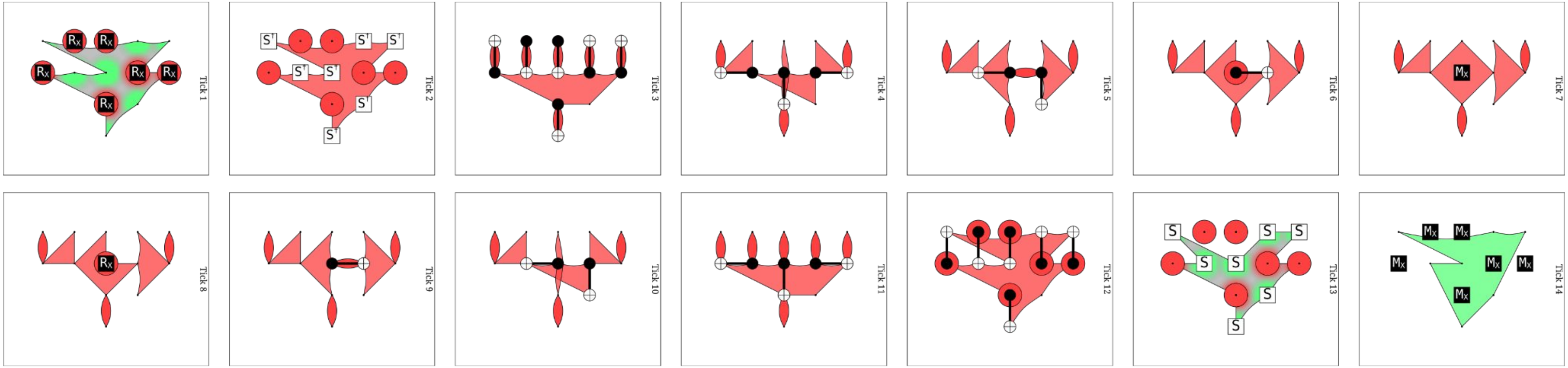
Every ending measurement a flag



Exact Kickback Measurement (d=3)

Time-Reversed Check (Top Row)

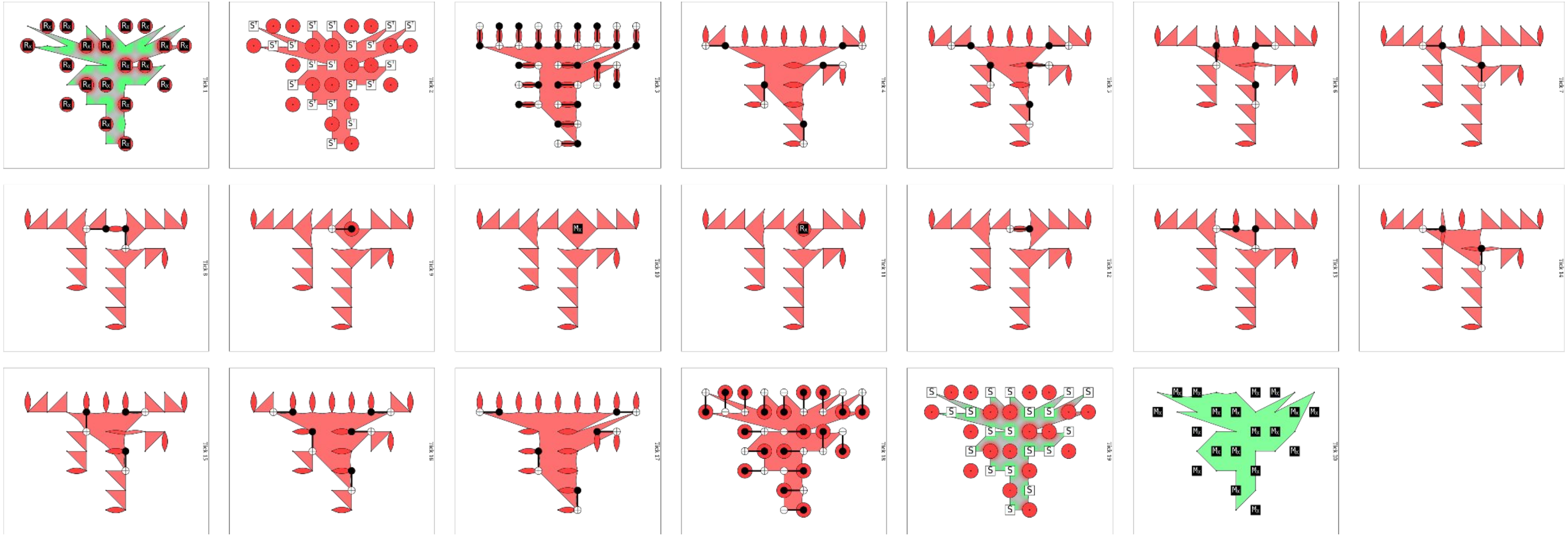
Open in Crumble



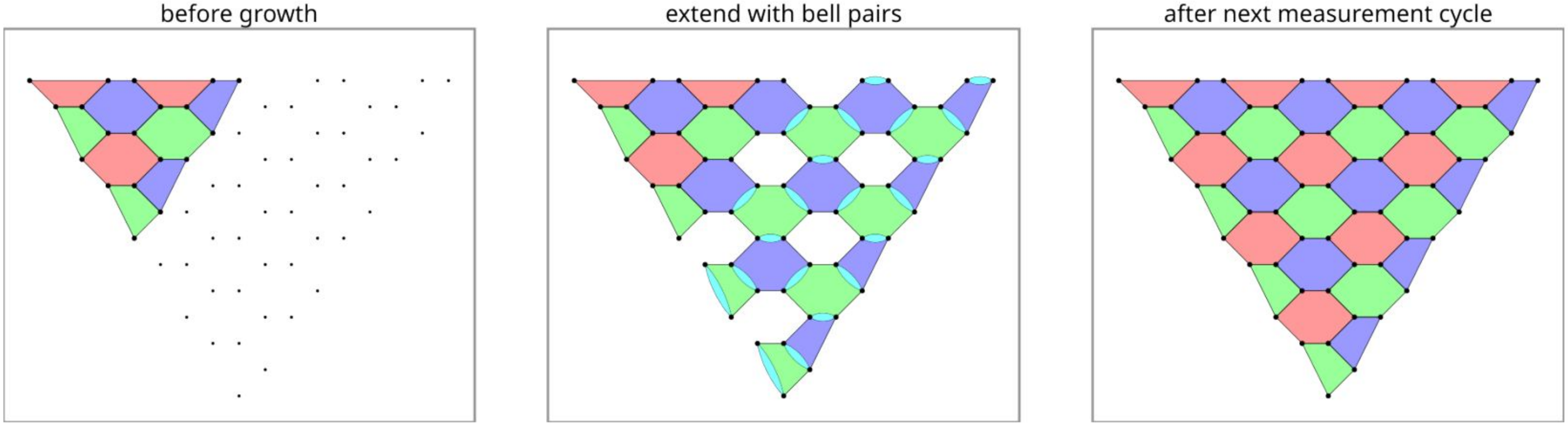
Time-Forward (Bottom Row)

Exact Kickback Measurement (d=5)

Open in Crumble



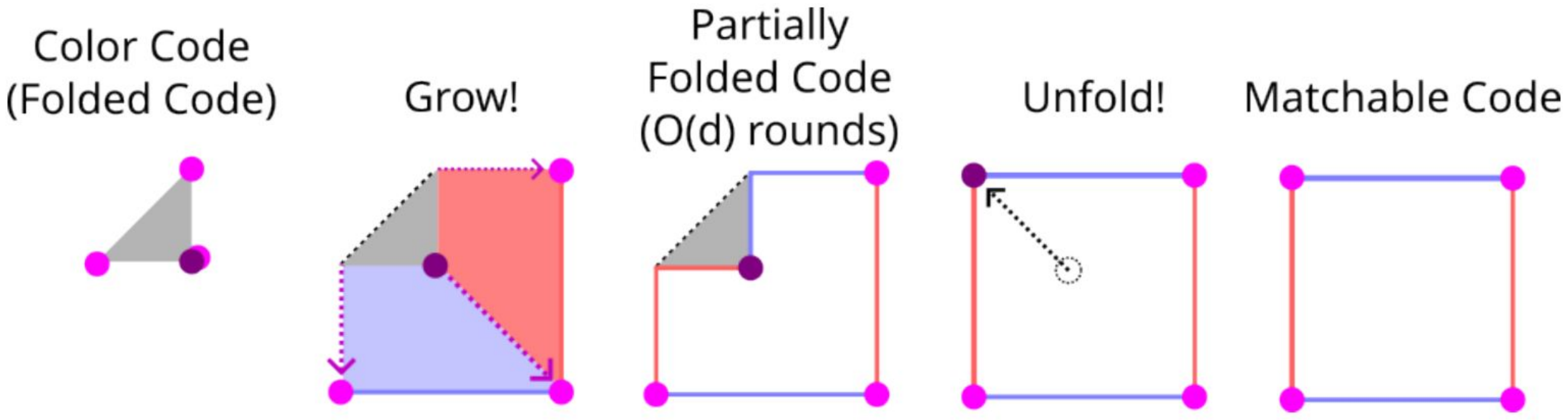
Escaping small code distance by growing the color code



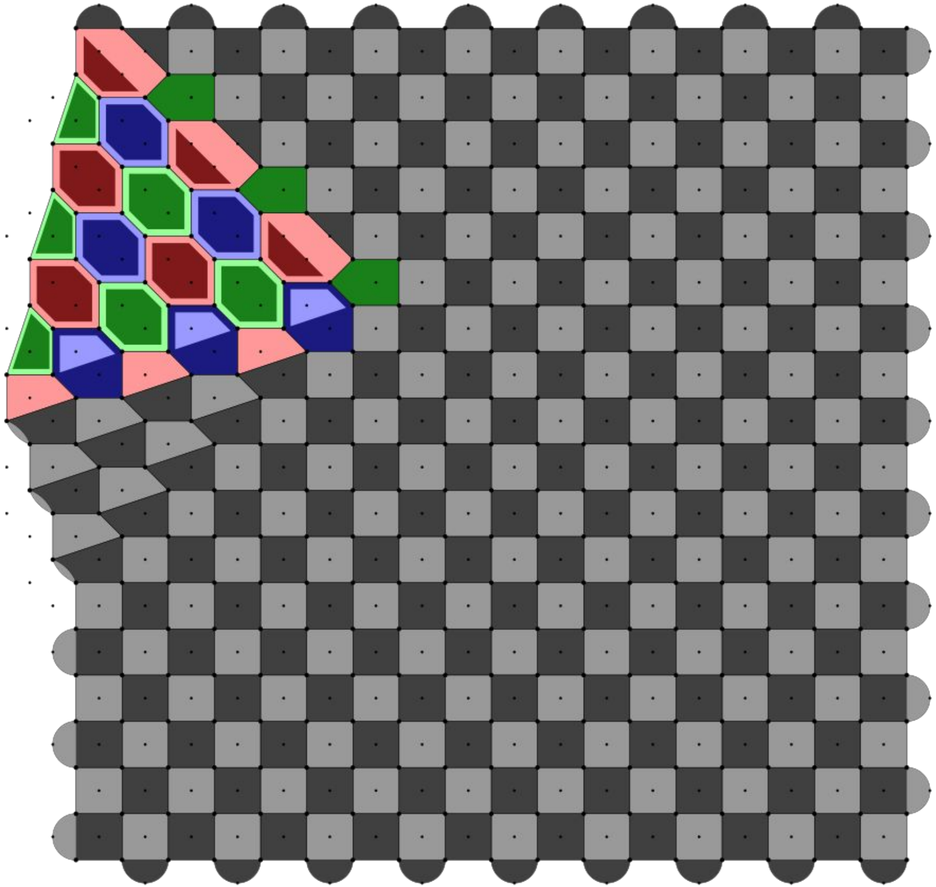
Chromobius was so bad at this I almost dropped the whole idea

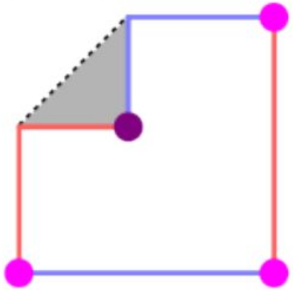
Definitely not *intrinsically* bad. **We need a better decoder**.

Stage 3: Escape!



Grafted Color/Surface Code



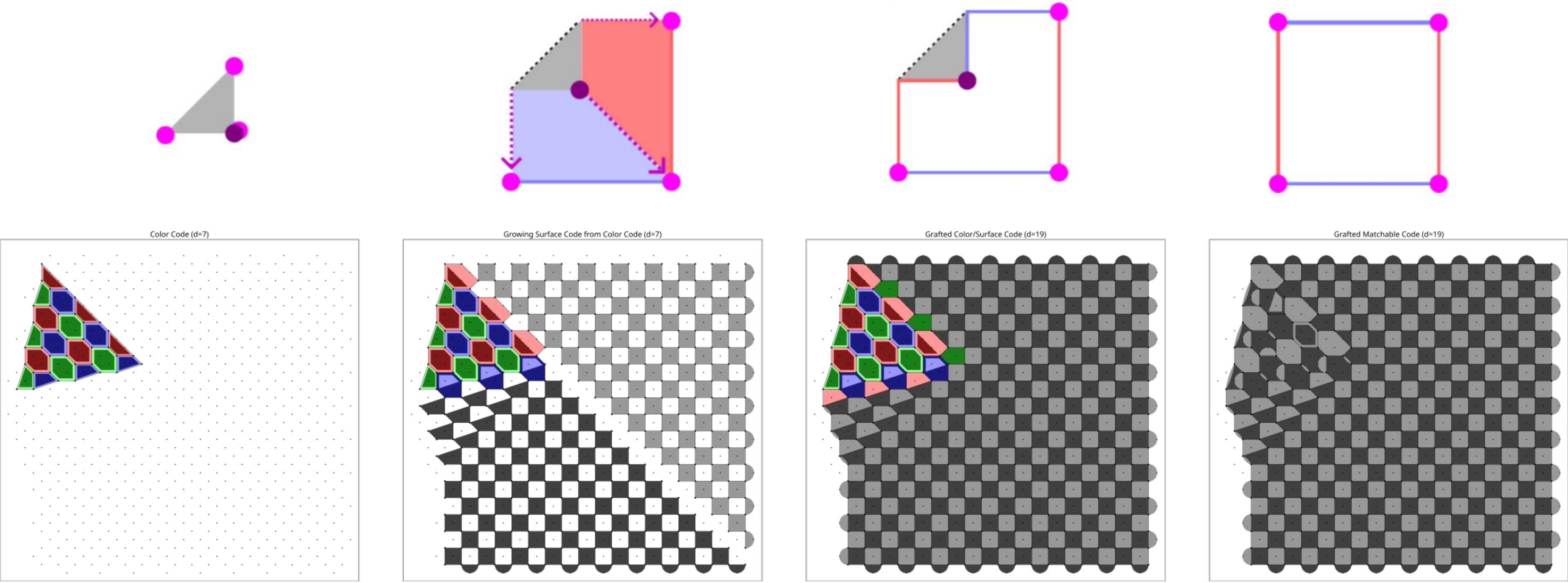


Light = X
Dark = Z

Red/Green/Blue:
Color Code Region

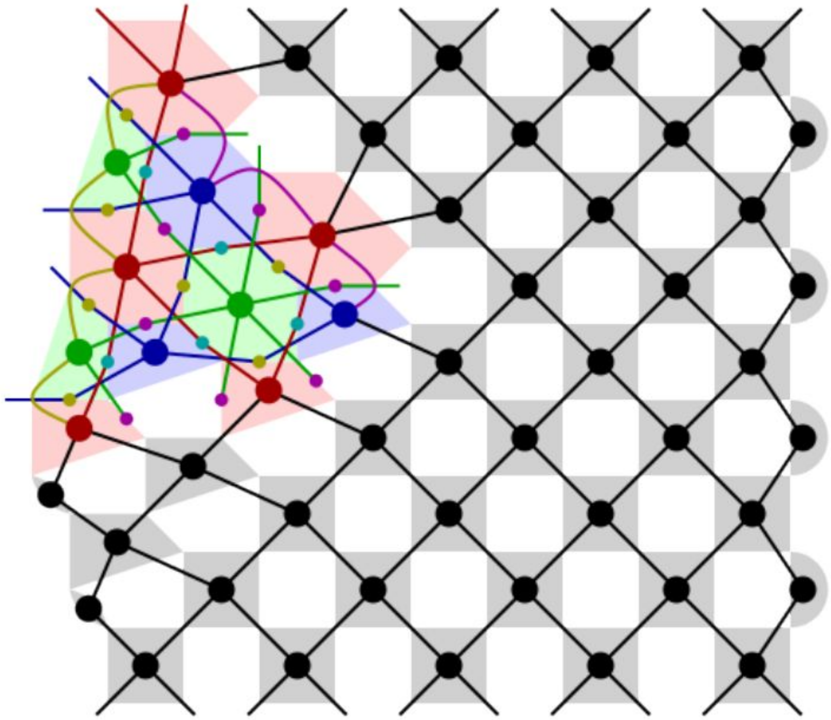
Overlapping stabilizers are shown inset to avoid ambiguity

Stage 3: Escape!

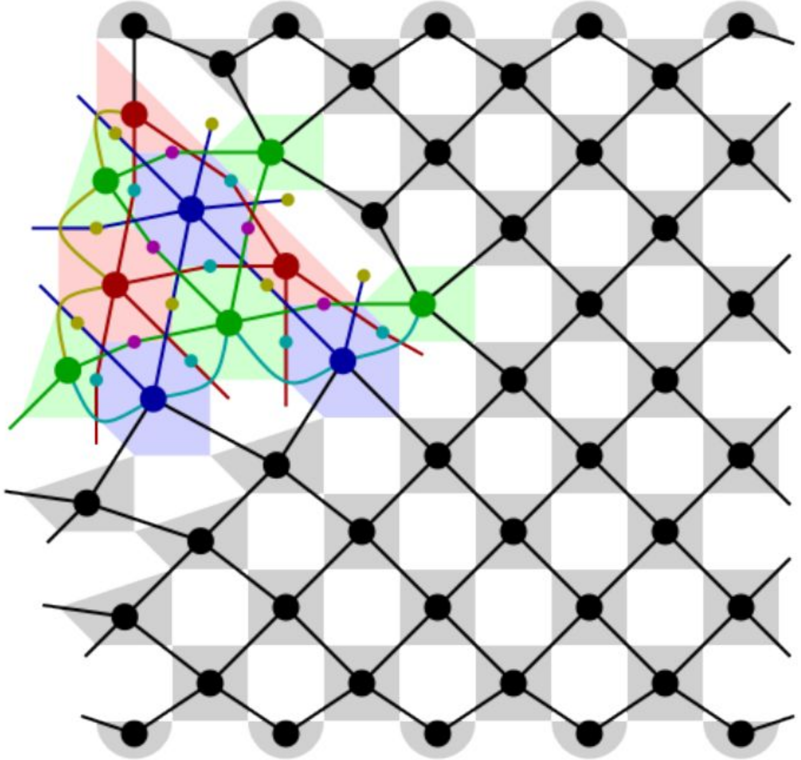


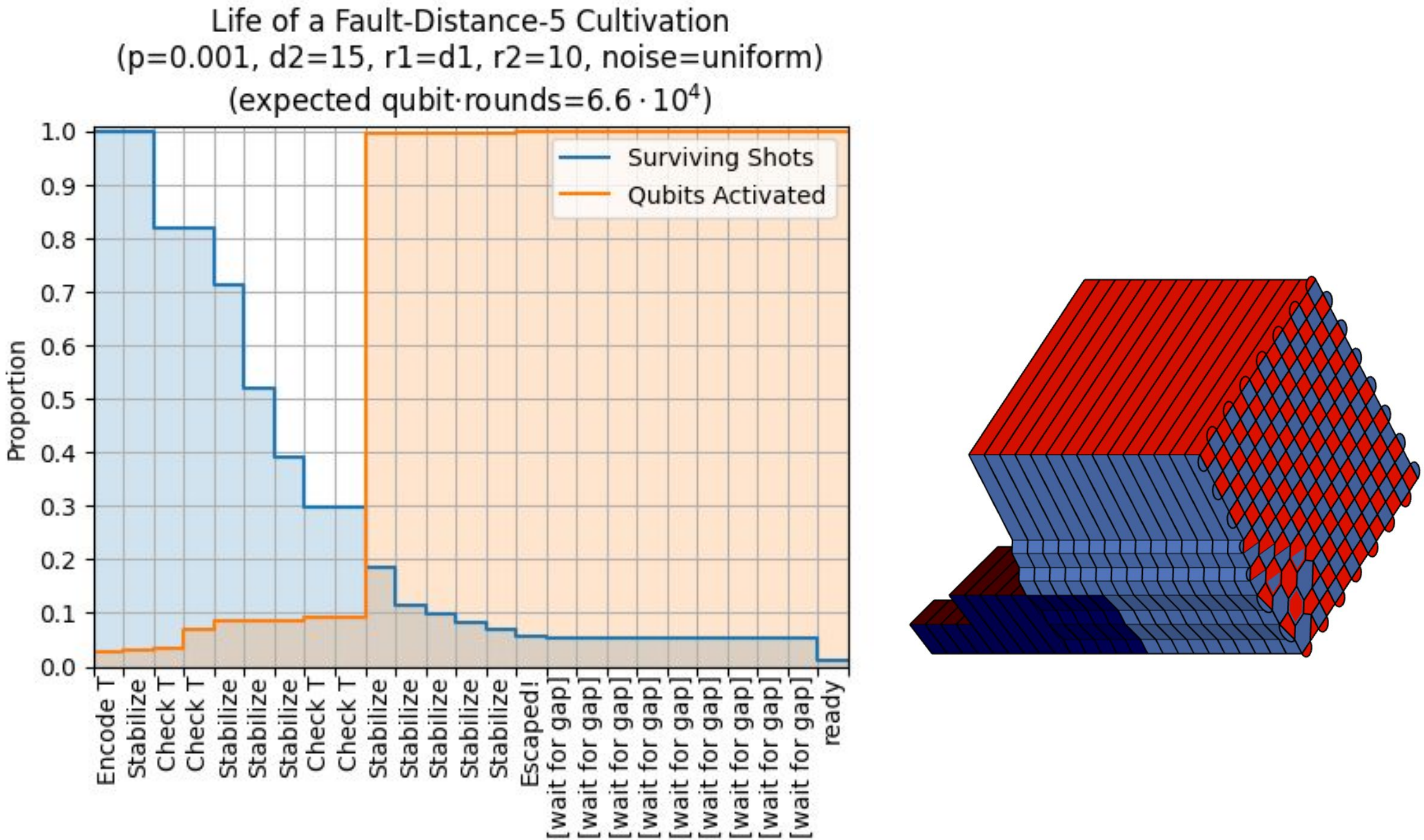
Decoding the Color Code Region with a Matcher

X Subgraph



Z Subgraph





Life of a Fault-Distance-5 Cultivation
(p=0.001, d2=15, r1=d1, r2=10, noise=uniform)
(expected qubit·rounds=$6.6 \cdot 10^4$)
Surviving Shots
Qubits Activated
Proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Encode T
Stabilize
Check T
Check T
Stabilize
Stabilize
Stabilize
Check T
Check T
Stabilize
Stabilize
Stabilize
Stabilize
Stabilize
Escaped!
[wait for gap]
[wait for gap]
[wait for gap]
[wait for gap]
[wait for gap]
[wait for gap]
[wait for gap]
[wait for gap]
[wait for gap]
ready

Benchmarking

Obstacle: How do we sample this?

State vector simulation

- ✗ Too large. Escape stage has >100 qubits.

Obstacle: How do we sample this?

State vector simulation

- ✗ Too large. Escape stage has >100 qubits.

Stabilizer simulation

- ✗ Too picky. Can't handle the T gates.

Obstacle: How do we sample this?

State vector simulation

- ✗ Too large. Escape stage has >100 qubits.

Stabilizer simulation

- ✗ Too picky. Can't handle the T gates.

Tensor network simulation

- ✗ Too interconnected. Dozens of color code cycles.

Obstacle: How do we sample this?

State vector simulation

- ✗ Too large. Escape stage has >100 qubits.

Stabilizer simulation

- ✗ Too picky. Can't handle the T gates.

Tensor network simulation

- ✗ Too interconnected. Dozens of color code cycles.

Stabilizer decomposition

- ✓ arxiv:2202.09202 does some comparable sized simulations
- ✗ Too slow. Minutes per shot. Want billions of shots.

Obstacle: How do we sample this?

State vector simulation

- ✗ Too large. Escape stage has >100 qubits.

Stabilizer simulation

- ✗ Too picky. Can't handle the T gates.

Tensor network simulation

- ✗ Too interconnected. Dozens of color code cycles.

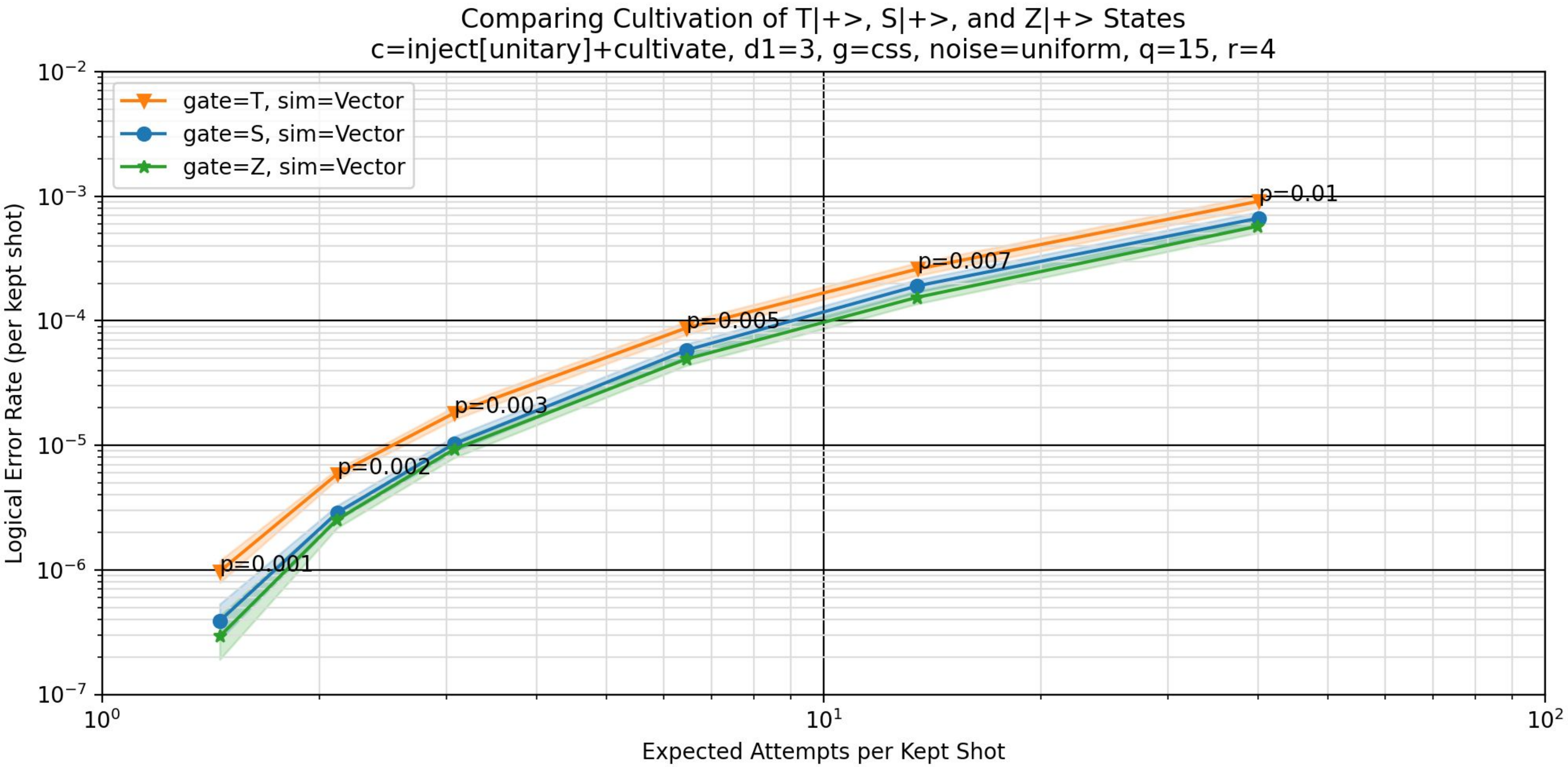
Stabilizer decomposition

- ✓ arxiv:2202.09202 does some comparable sized simulations
- ✗ Too slow. Minutes per shot. Want billions of shots.

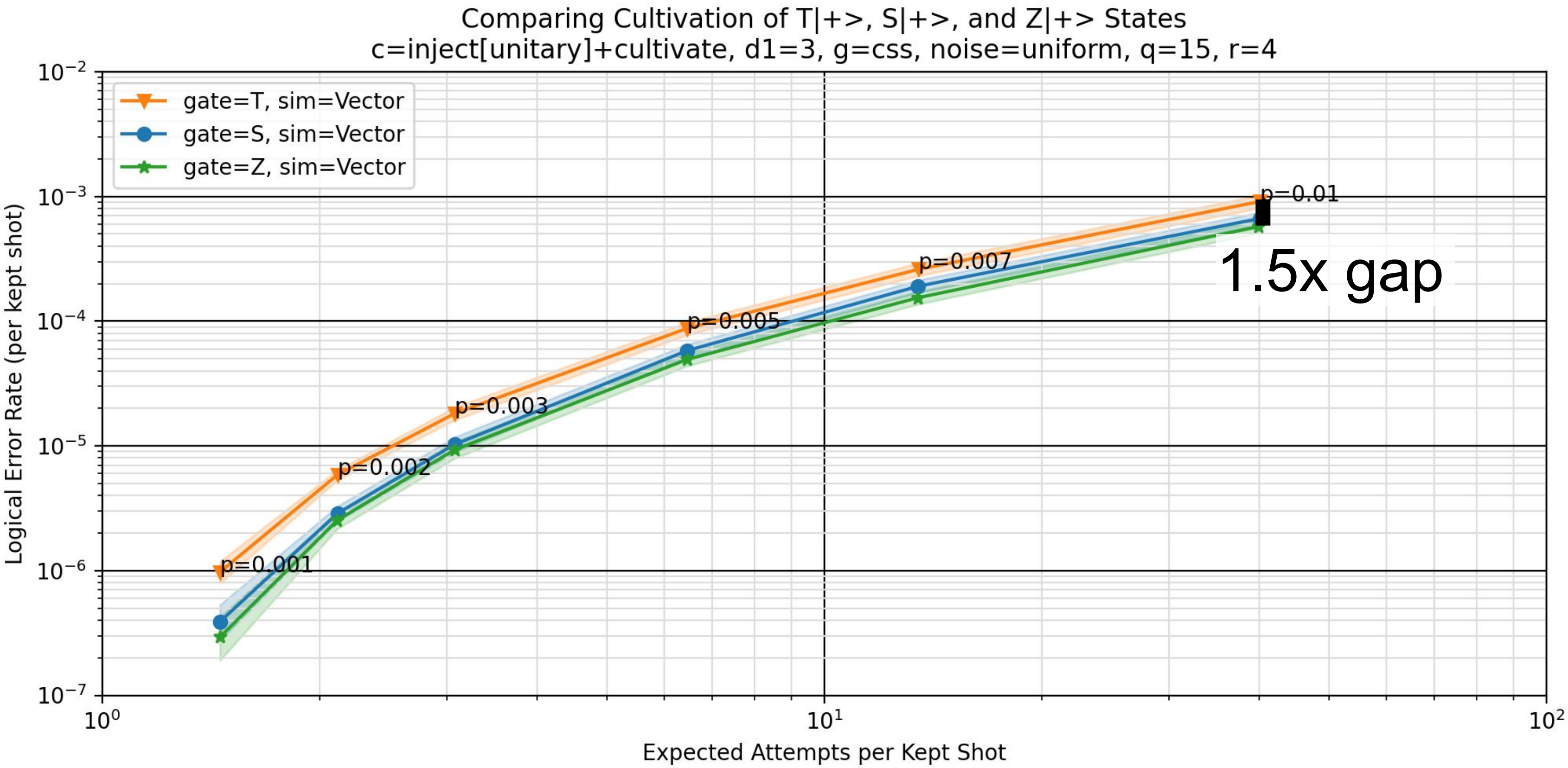
Replace T gates with S gates

- ✓ Easy and fast to simulate. Circuit still functions (distills $S|+\rangle$ instead of $T|+\rangle$).
- **? How accurate is this?**

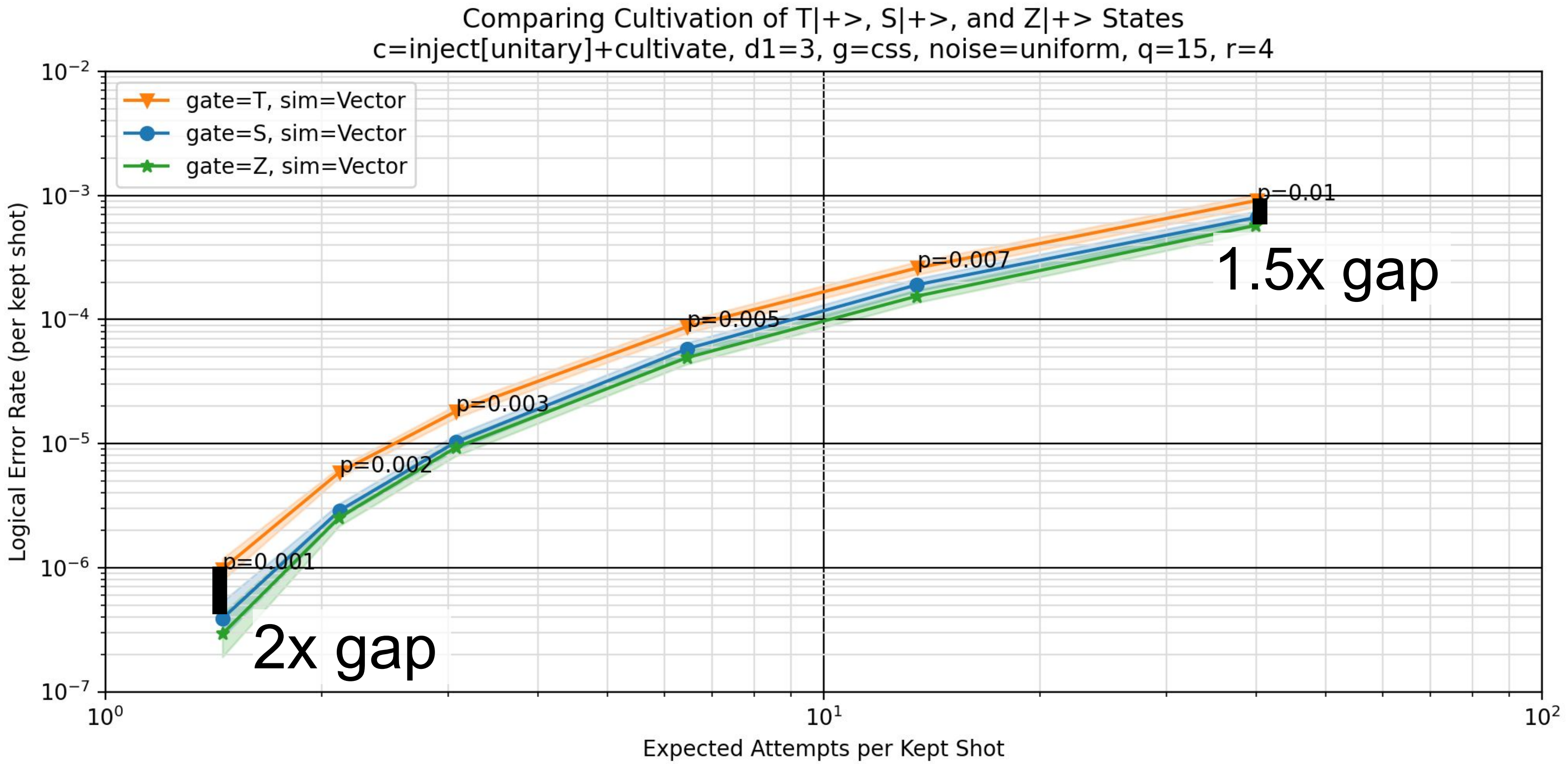
S gate circuit as a proxy for the T gate circuit



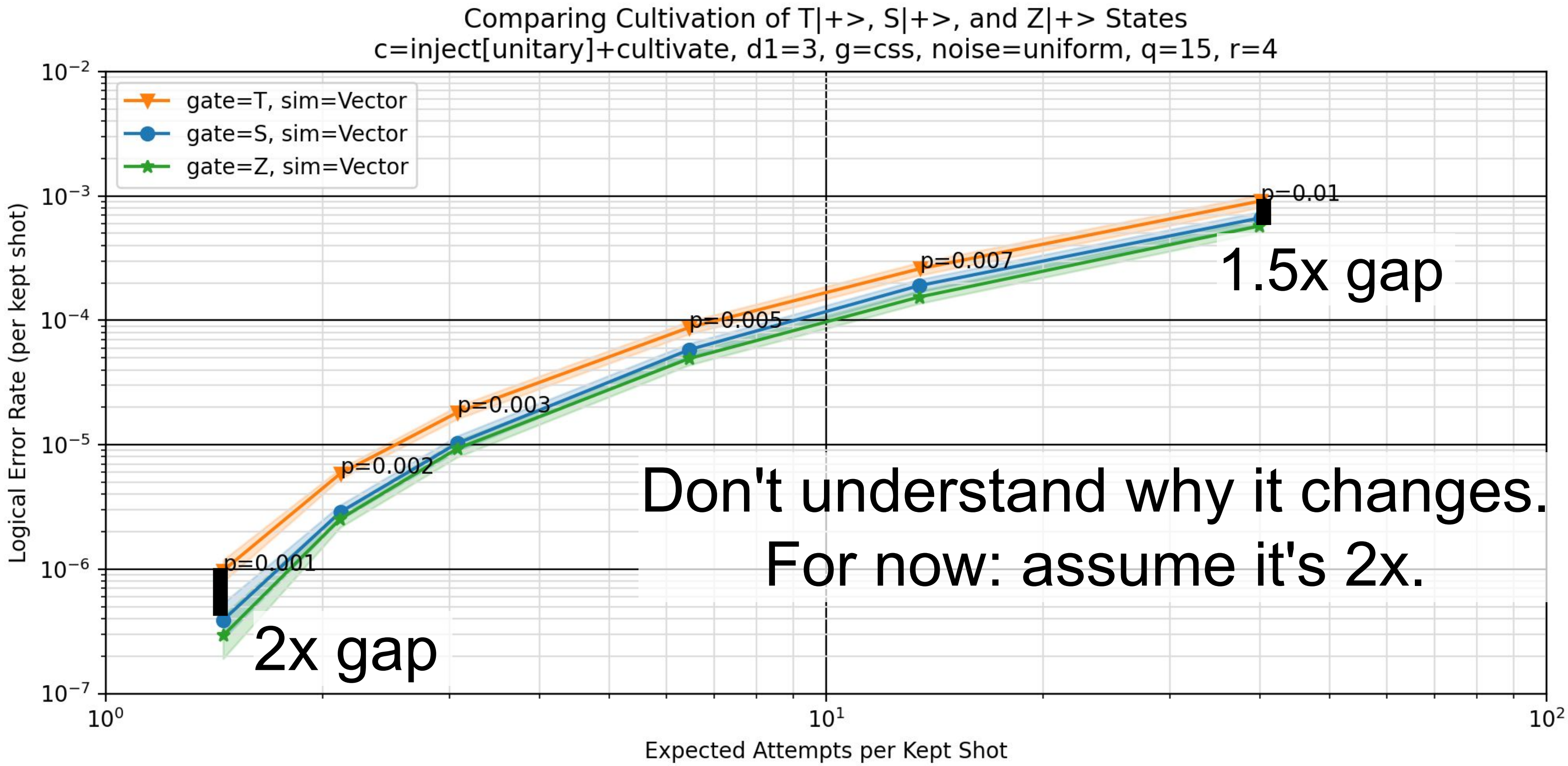
S gate circuit as a proxy for the T gate circuit



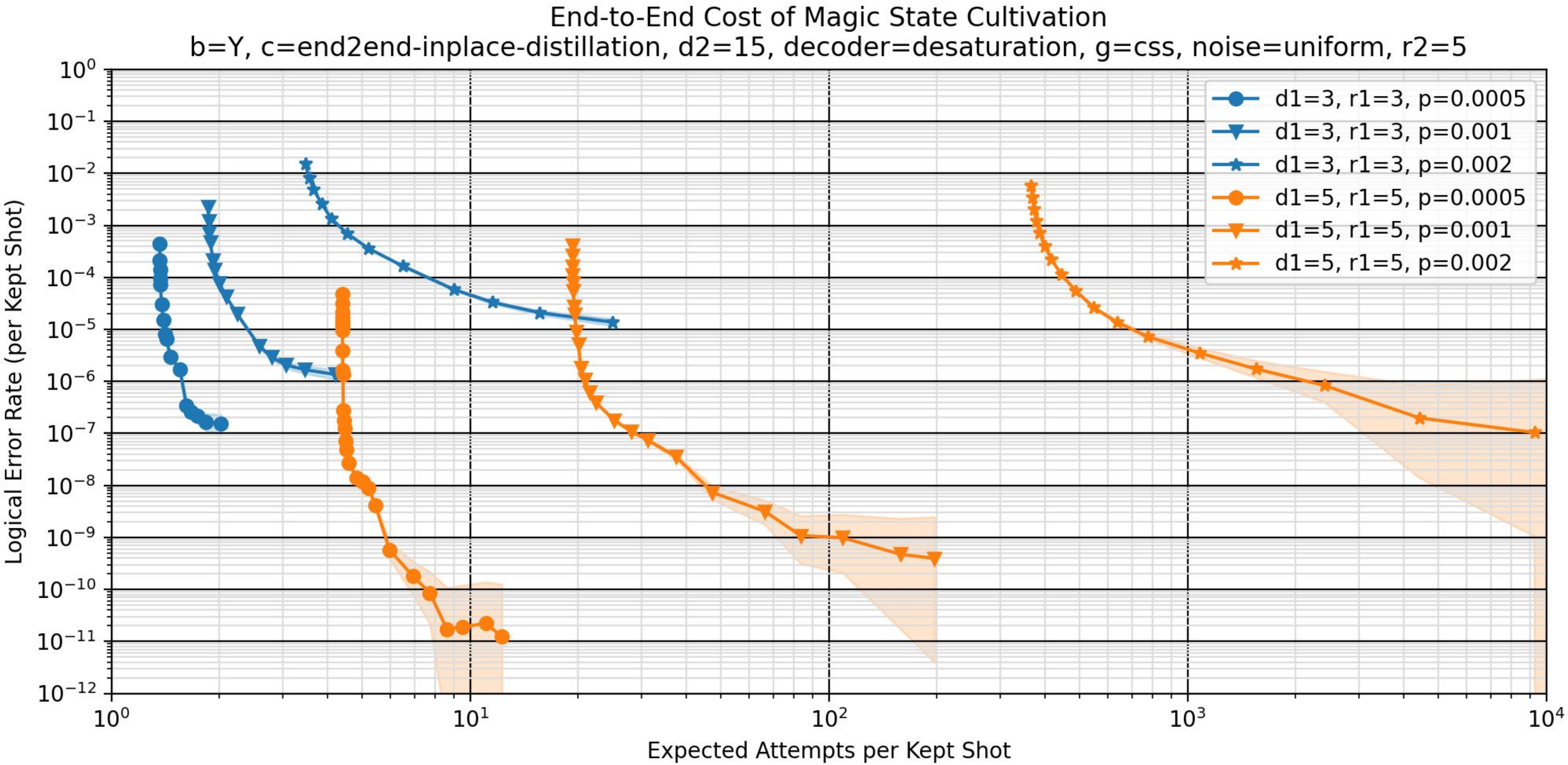
S gate circuit as a proxy for the T gate circuit



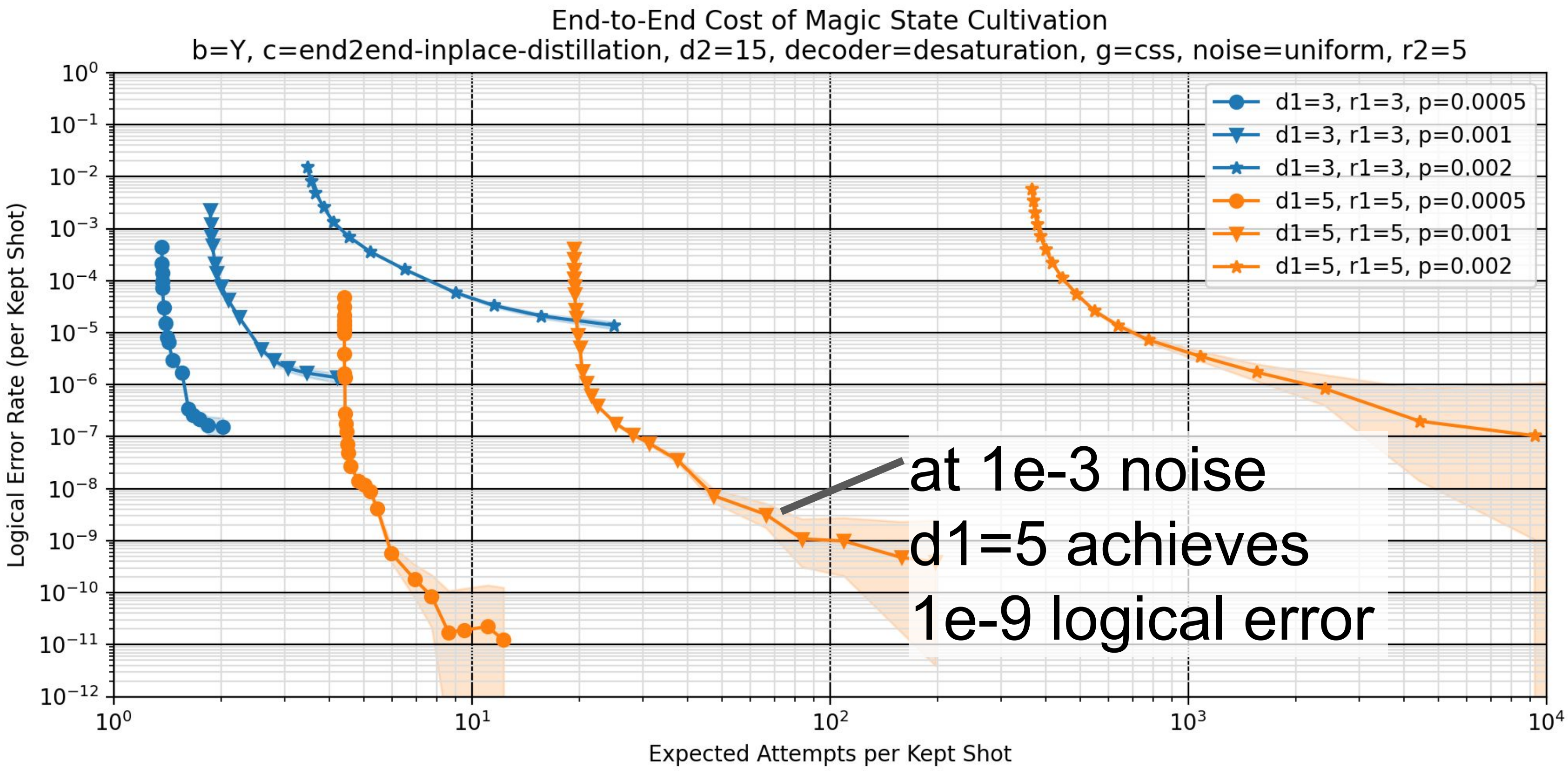
S gate circuit as a proxy for the T gate circuit



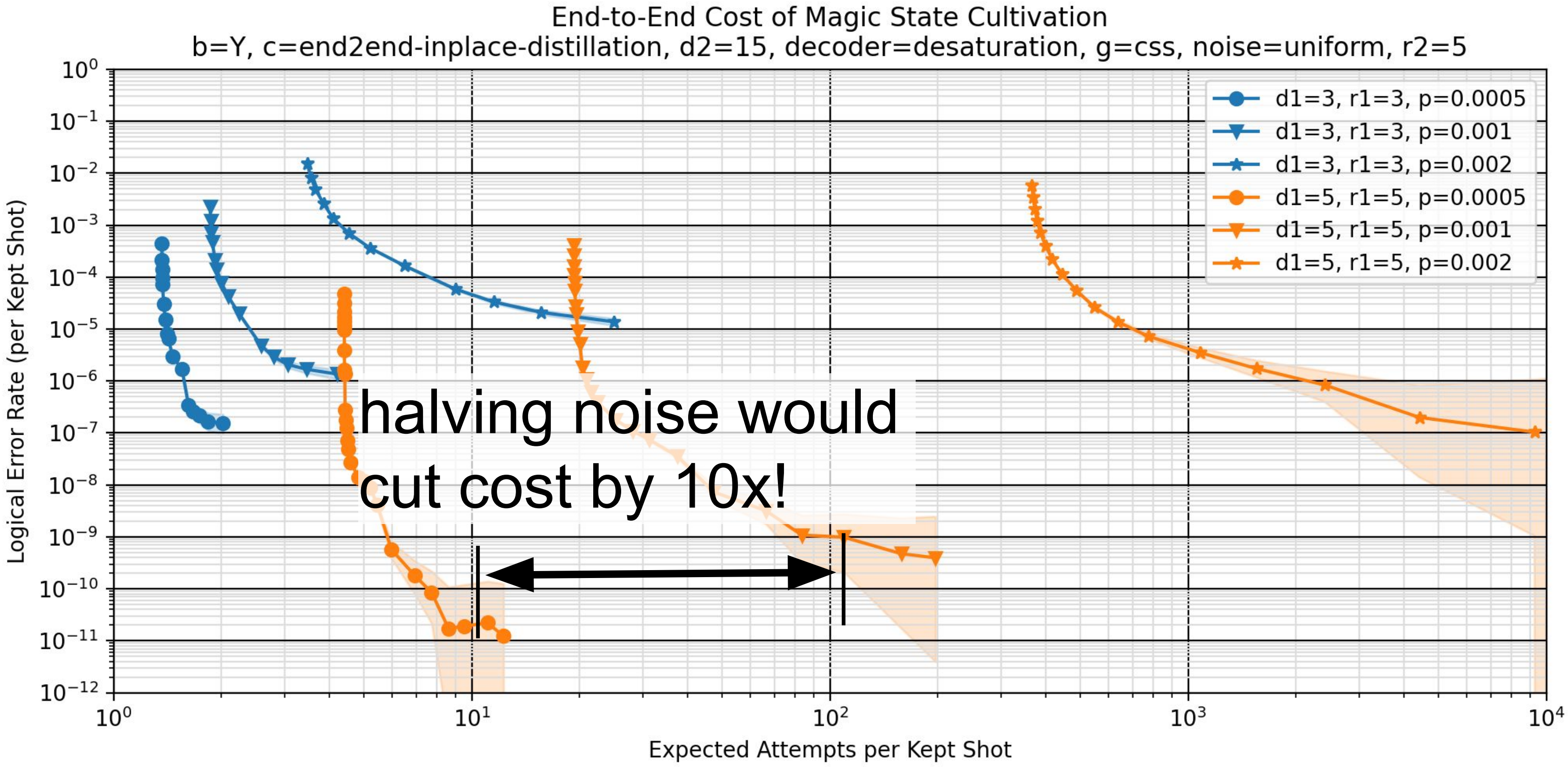
End-to-end simulations



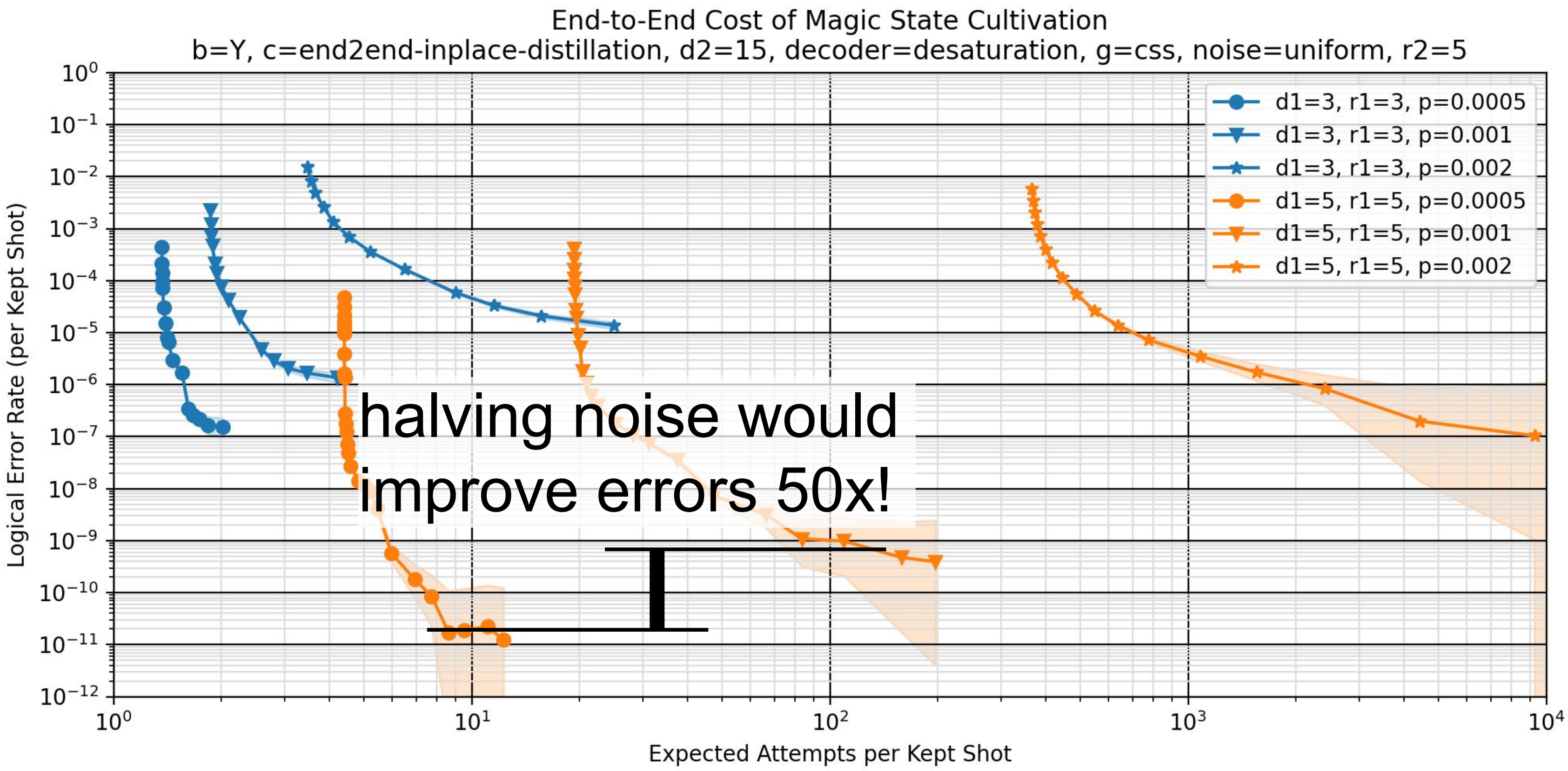
End-to-end simulations



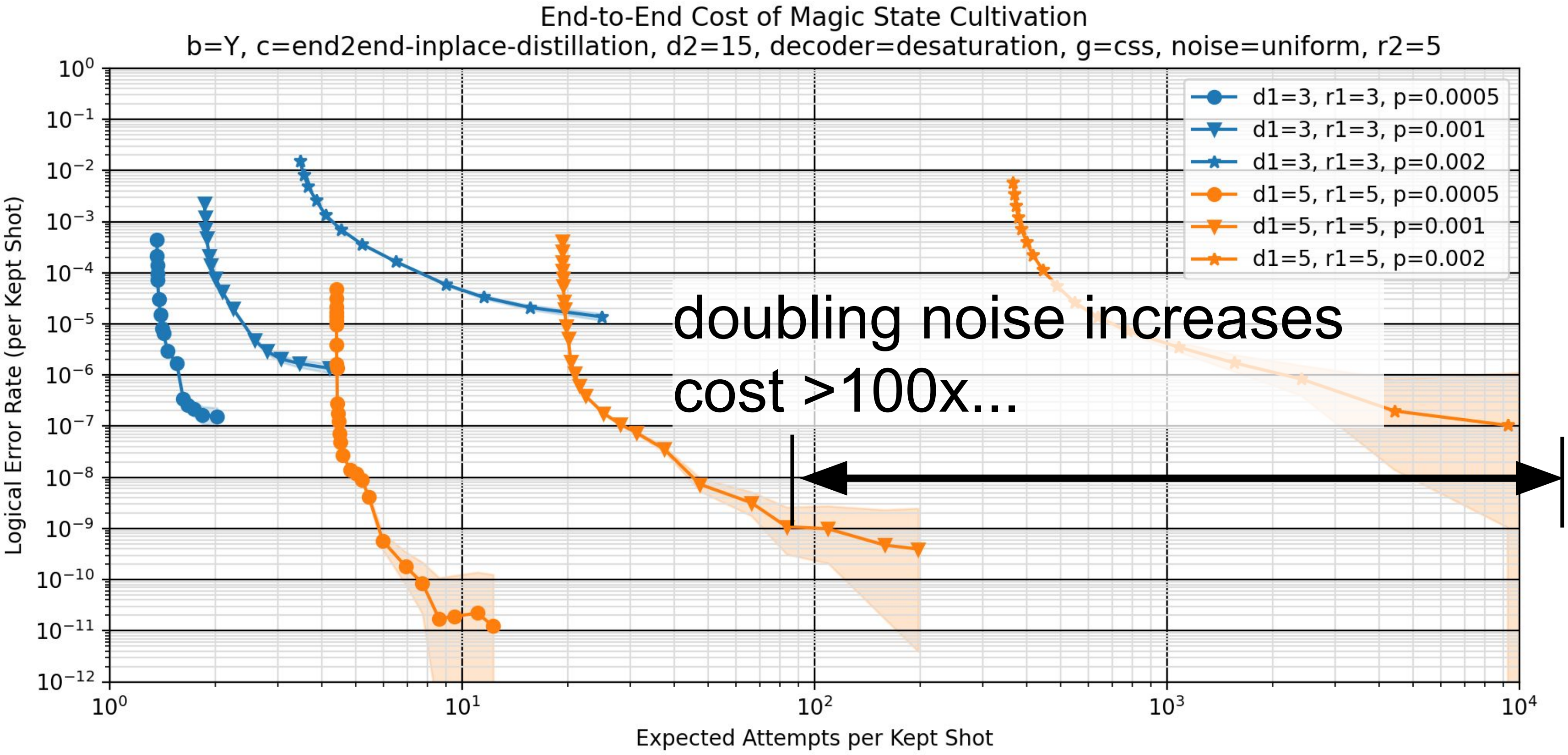
End-to-end simulations



End-to-end simulations

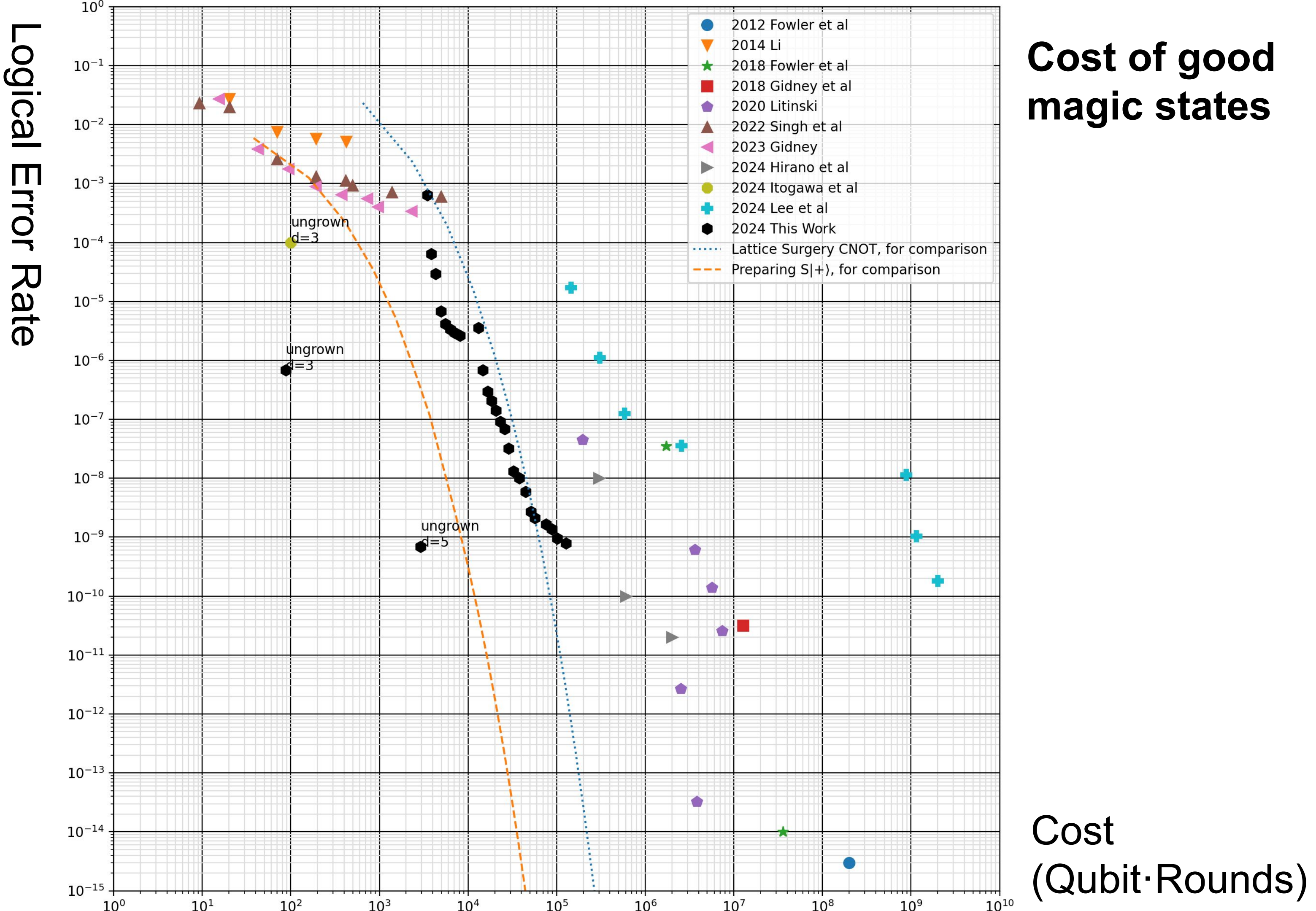


End-to-end simulations

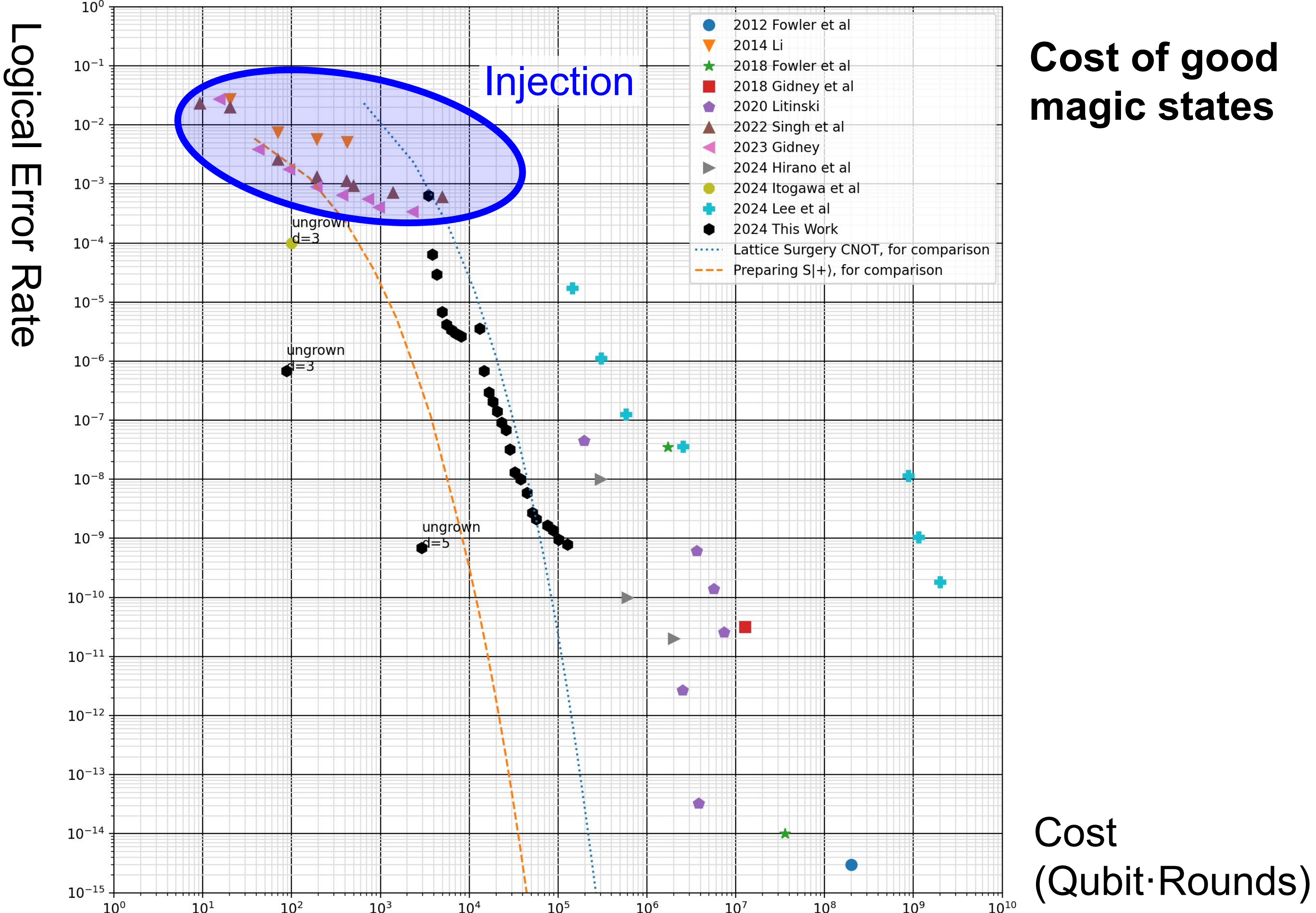


Contrasting With Other Techniques

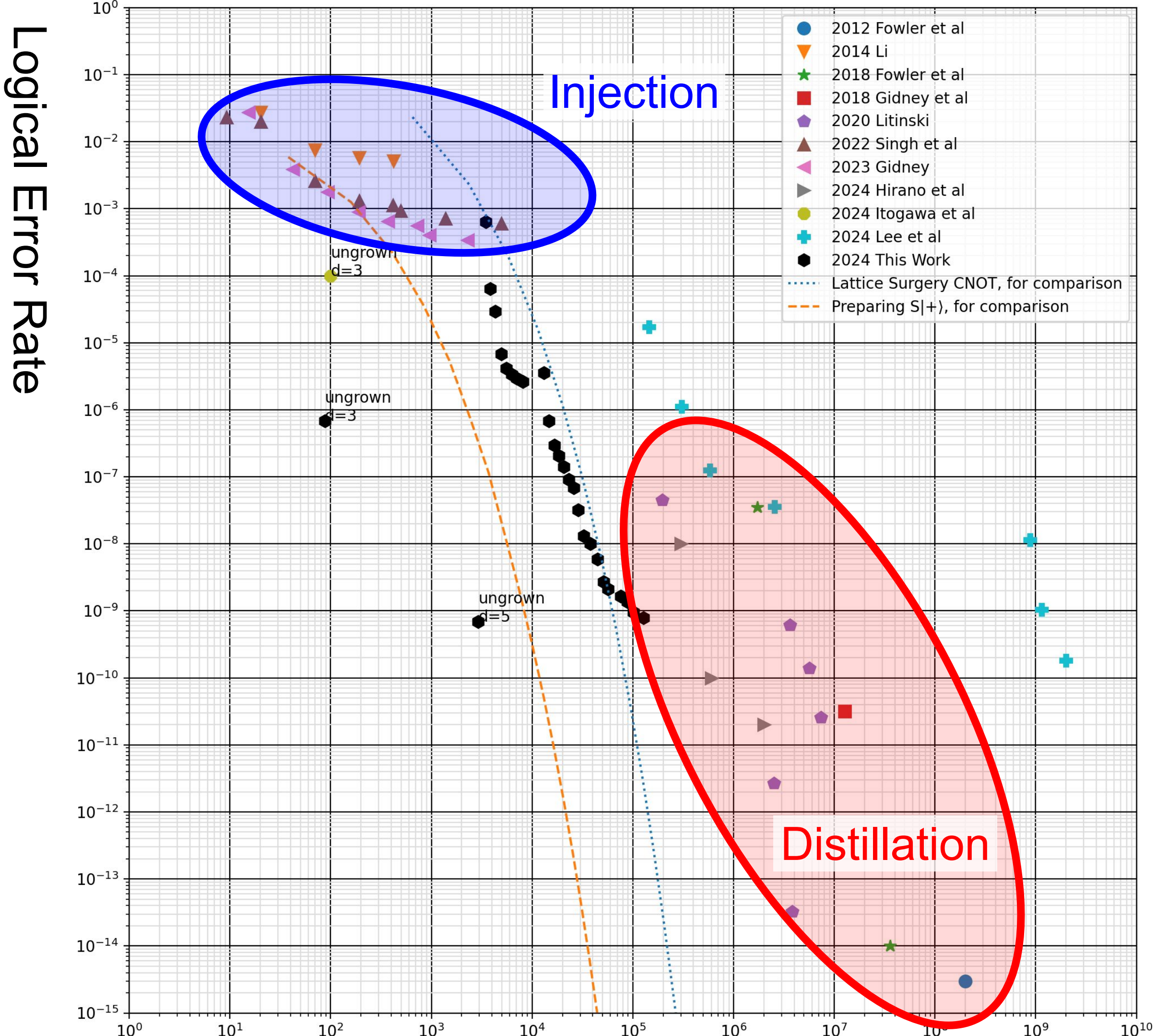
Cost of good magic states



Cost of good magic states

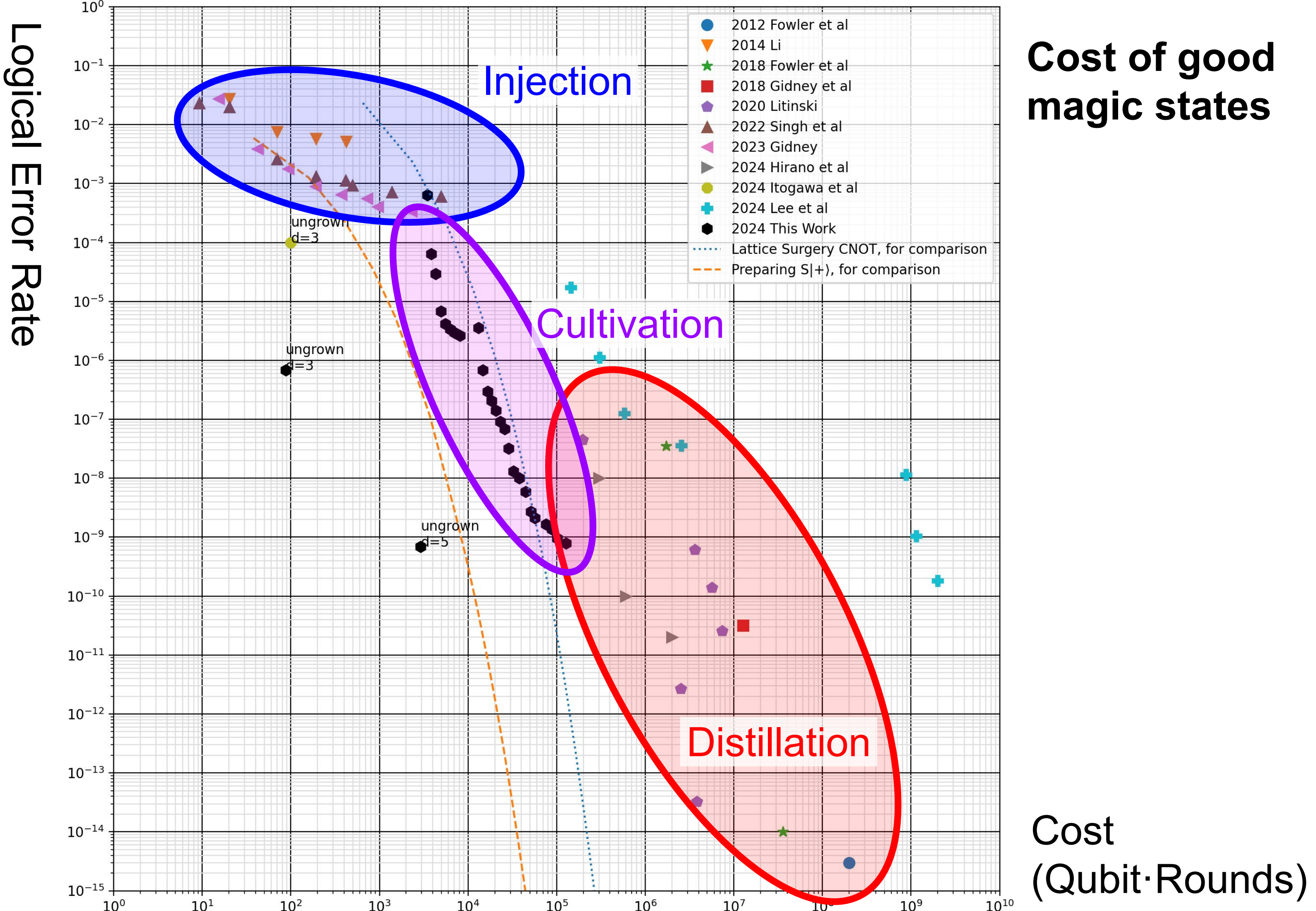


Cost of good magic states

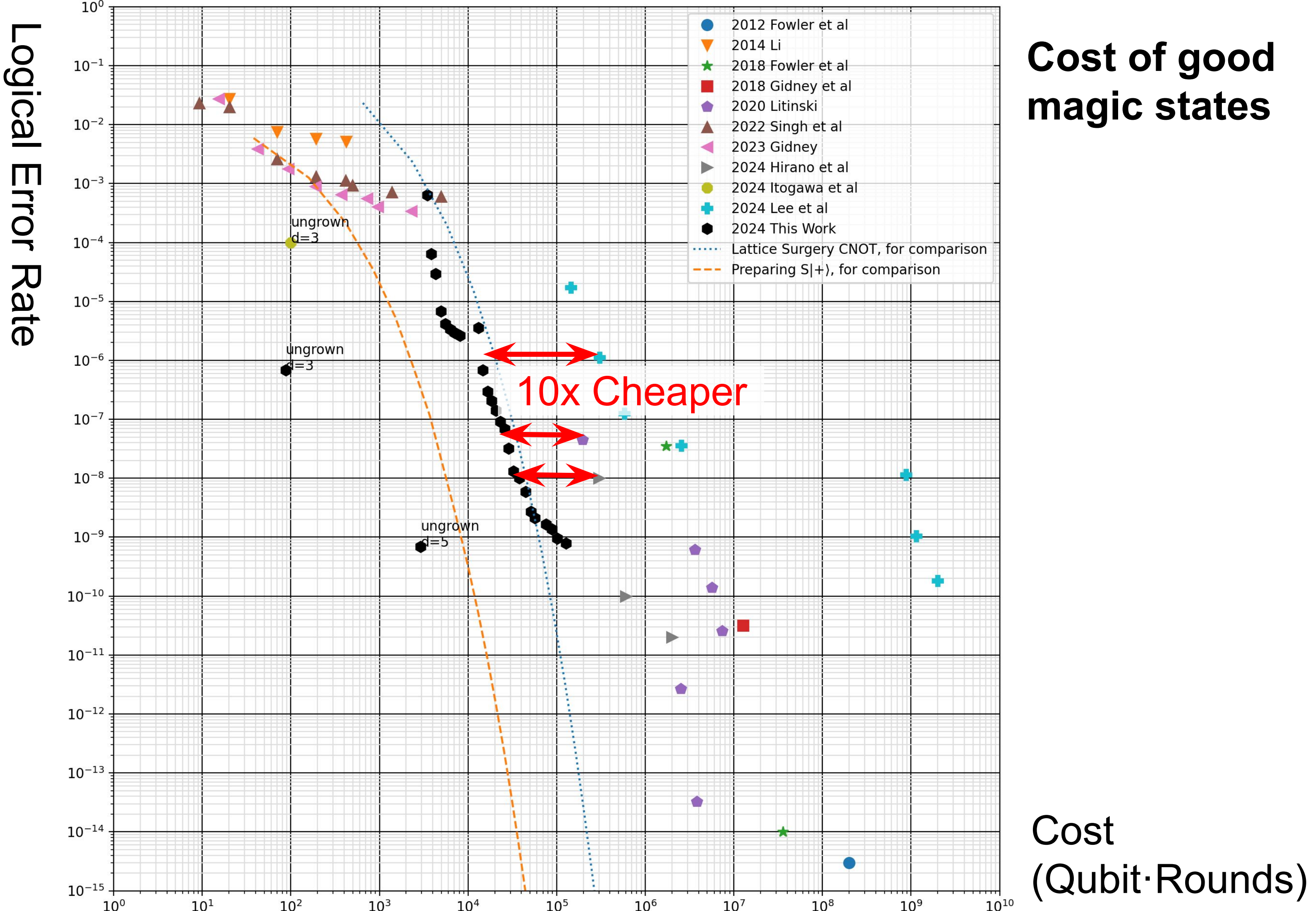


Cost
(Qubit·Rounds)

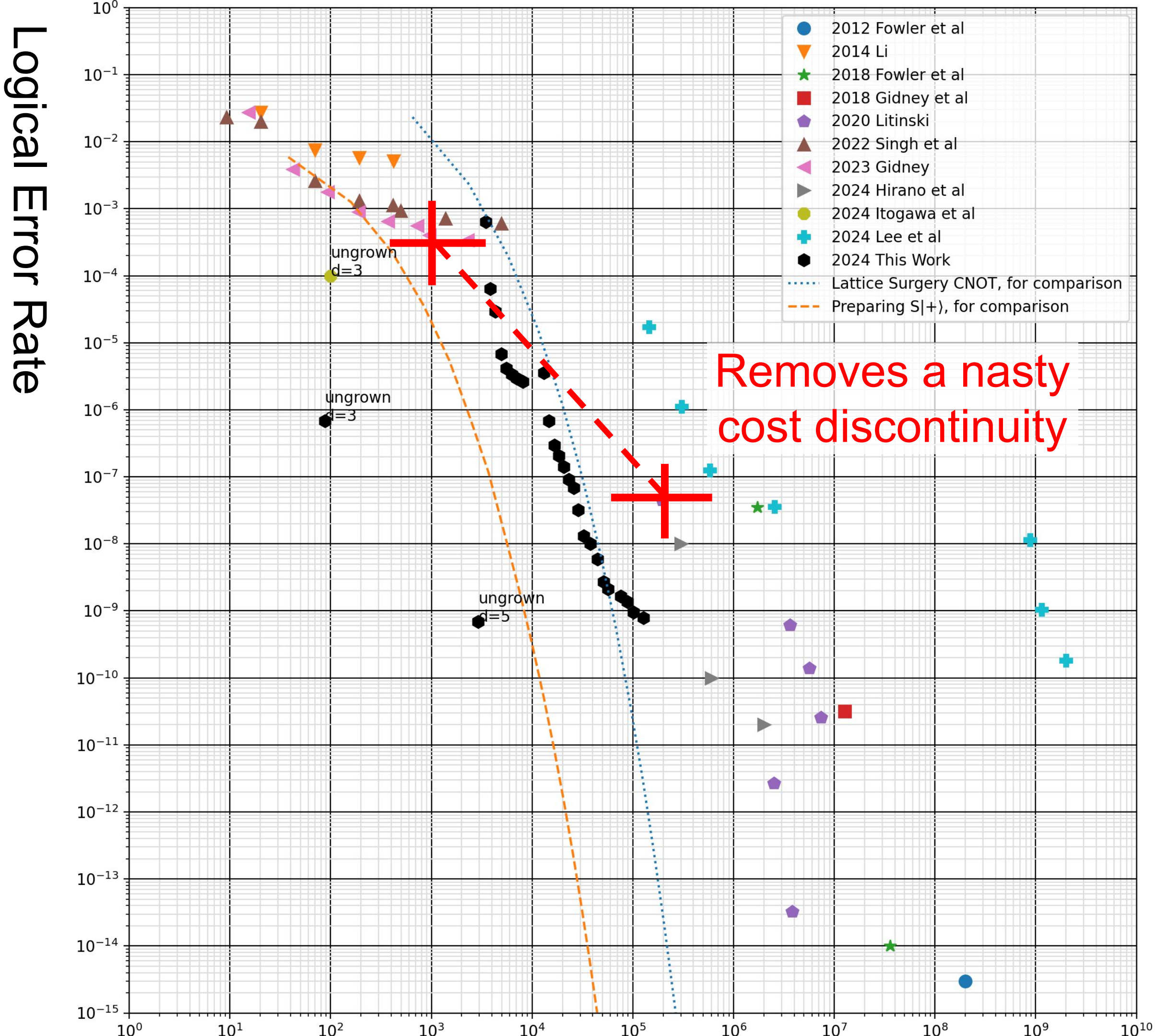
Cost of good magic states



Cost of good magic states

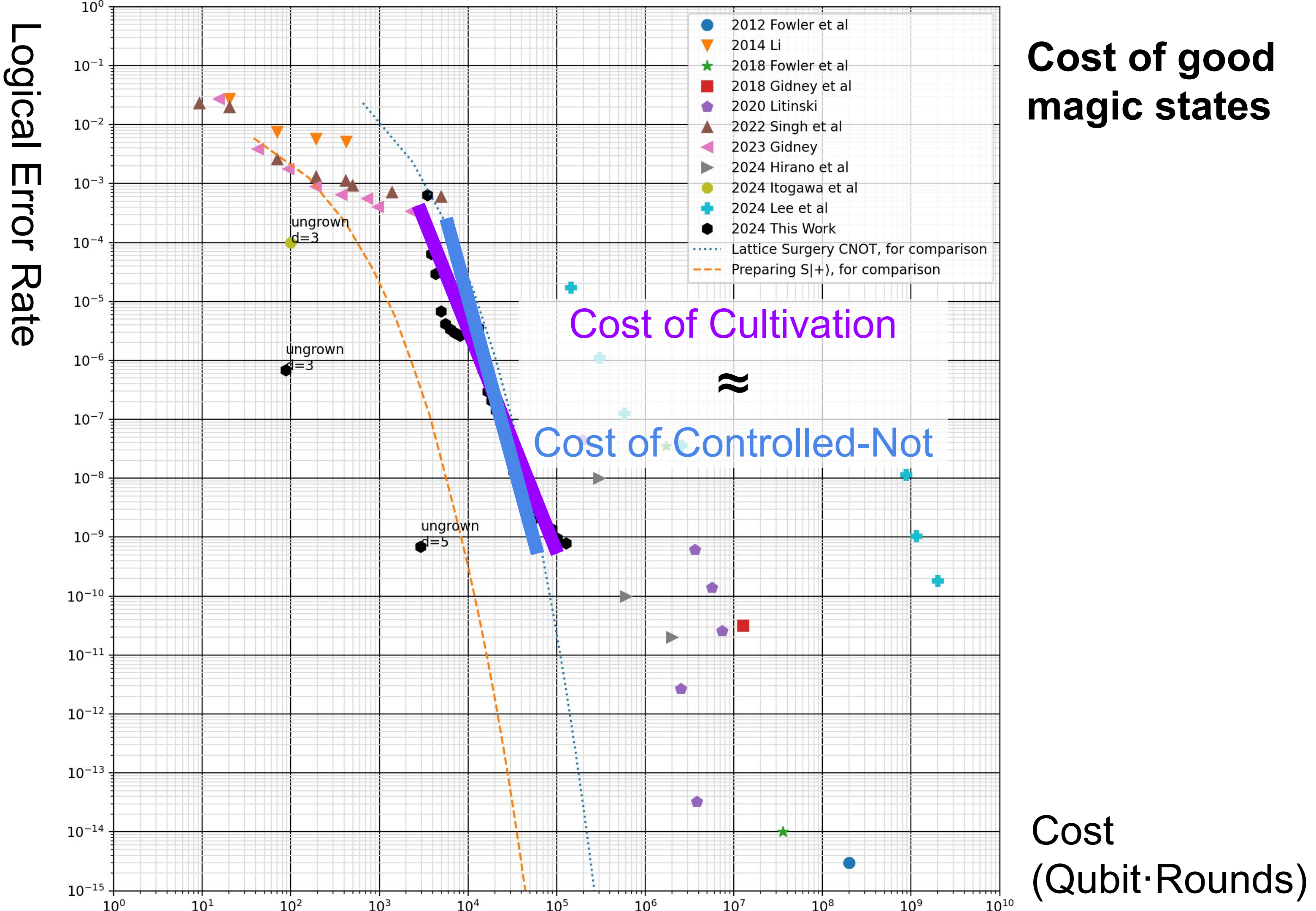


Cost of good magic states



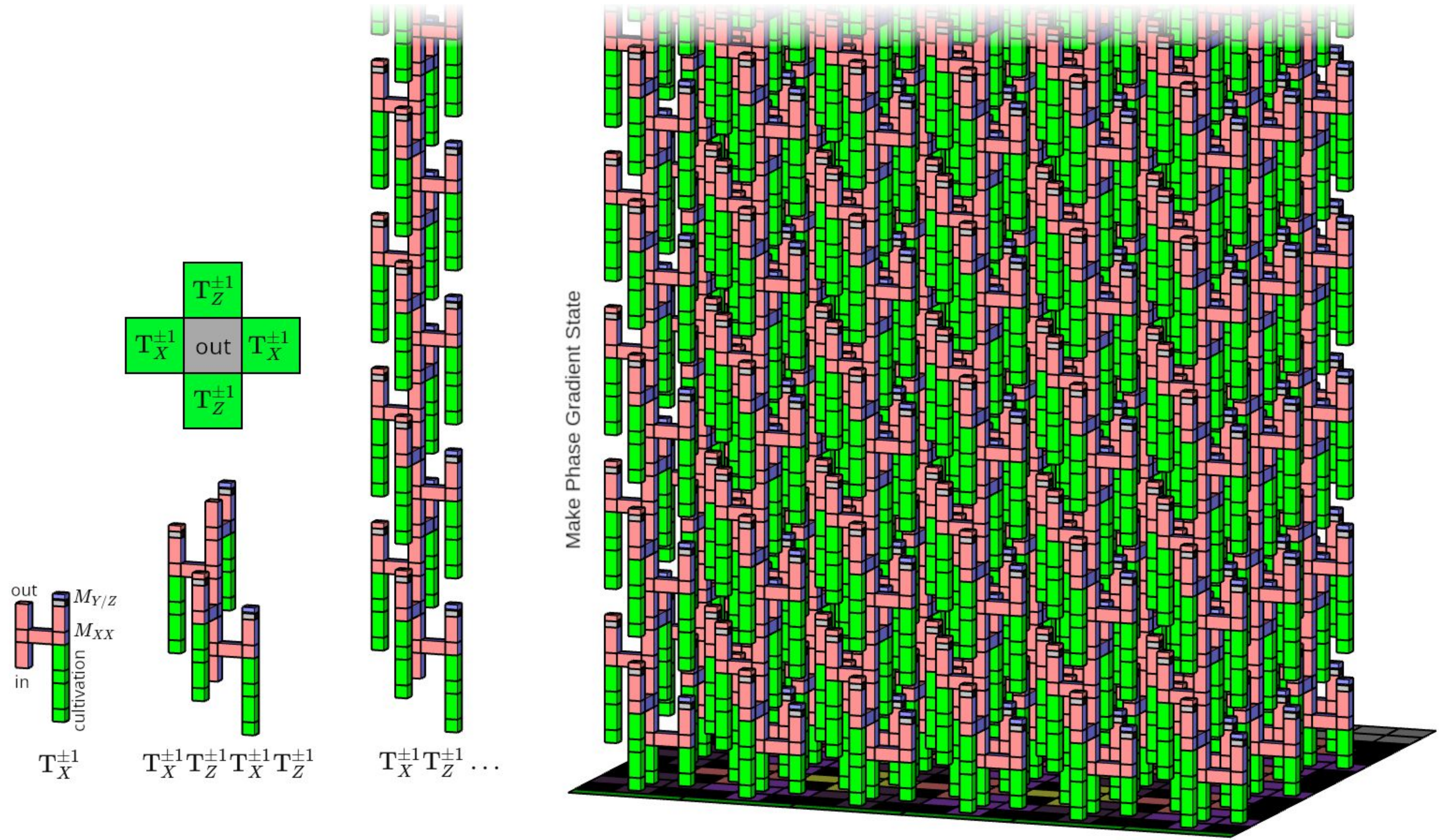
Cost (Qubit·Rounds)

Cost of good magic states



An example of possible consequences: rotating compactly by using more T gates

Qubit Index (k in $Z^{2^{-k}}\|+\rangle$)	Init Basis	T Gate Directions (the signs in $T_X^{\pm 1}, T_Z^{\pm 1}, T_X^{\pm 1}, T_Z^{\pm 1}, \ldots$)	Finish Clifford	T Count	Infidelity
0	R_X		Z	0	0
1	R_Y			0	0
2	R_Z	+	H	1	0
3	R_Z	+-+----++------++-+----+-----++-+++--+++----+++-++	H,Y	50	0
4	R_Z	----+++-+-+---+++-++--+-+-+-++++--+++-+-++--+-	H,Z	46	4.1e-16
5	R_Z	+-+--+----+---++-+++++--+++-++---++-++----+-++-+	S,H	48	6.8e-16
6	R_Z	---+++-------++--+++-+--++-++--+-+-+++++++----+-+	X	49	4.1e-17
7	R_Z	++---+++++--+---++++-+---+-++-+-+++-+++--++------+	H	50	6.9e-16
8	R_Z	++-++-+++--++-++-++++-++--++++++----------+--	Y	45	3.0e-16
9	R_Z	+---+--++++++++--+--++--++-+-++---+-++----+-+-++	H	48	2.3e-16
10	R_Z	+++--+--+-+++++++-++++-+-++-+-+++-++---+++++-+---		49	5.0e-17
11	R_Z	++-+--++-++--++-----+-+-+-+---+++-------+-+--++	Z	47	9.7e-16
12	R_Z	++------++++++--+--+-++-+----+-+++++--+++---+	Y	45	6.8e-16
13	R_Z	+++---+--++++-++--+++--++--+-++-+++++--+-++--+	H,Z	46	6.9e-16
14	R_Z	+++--+-----+++---++++-+-+++++-+++----+++---+-+--+		49	3.9e-16
15	R_Z	+---++-+--+-++---+--+++-++-++++--++++-+++-+--+	H,Y	46	6.2e-16
16	R_Z	+--+-+-++--+++++-+-++---+++++-+-++++--+---+-++-+	H,Y	48	4.9e-16
17	R_Z	+-++----+---++++-+-+++-+-++++-+-+----+-++--		43	5.3e-16
18	R_Z	++---+--++--+---+--+-+---+---------++++++++--++--	X	49	1.8e-16
19	R_Z	+--+--+-+-+-------+---+--+---++-+--+-+++-+-+++-++	Z	49	1.3e-16
20	R_Z	----+-++++--+-++--+++--+++--++-------+-+-+---++--		49	9.7e-16
21	R_Z	++---++-++-+--++----+-+-+--+-++++--++-++---+++++	H	48	6.1e-16
22	R_Z	+-++-+-+----+++--+-++--+-+-+-++-++---++-+--+++-+--++	H,Z	52	3.4e-16
23	R_Z	+++---+---+-+--++--+------+--++--+-+---+---+++-	X	47	5.9e-16
24	R_Z	++++-----+-++---+-+++---+-+-++++-+++-++++-++++--+	X	49	2.5e-16
25	R_Z	+-+-++---+++-++++-+-+-++++-+-+-++++-+++---++-+-++		49	2.9e-16
26	R_X			0	5.5e-16
27	R_X			0	1.4e-16



$T_Z^{\pm 1}$
$T_X^{\pm 1}$
out
$T_X^{\pm 1}$
$T_Z^{\pm 1}$
out
$M_{Y/Z}$
M_{XX}
in
cultivation
$T_X^{\pm 1}$
$T_X^{\pm 1} T_Z^{\pm 1} T_X^{\pm 1} T_Z^{\pm 1}$
$T_X^{\pm 1} T_Z^{\pm 1} \ldots$
Make Phase Gradient State

Summary

Impact on large scale architecture?

Amdahl's law:

- If 50% of the computation is distillation, max easy improvement is 2x

Jevon's paradox:

- When X gets cheaper, you use more of X
- Sometimes so much more that total spend on X goes up instead of down!
- Example: carry lookahead adder runs faster but uses more T states

Unclear how this all plays out.

Glass half empty

Non-Clifford gates still force a time ordering (prevent arbitrary parallelization)

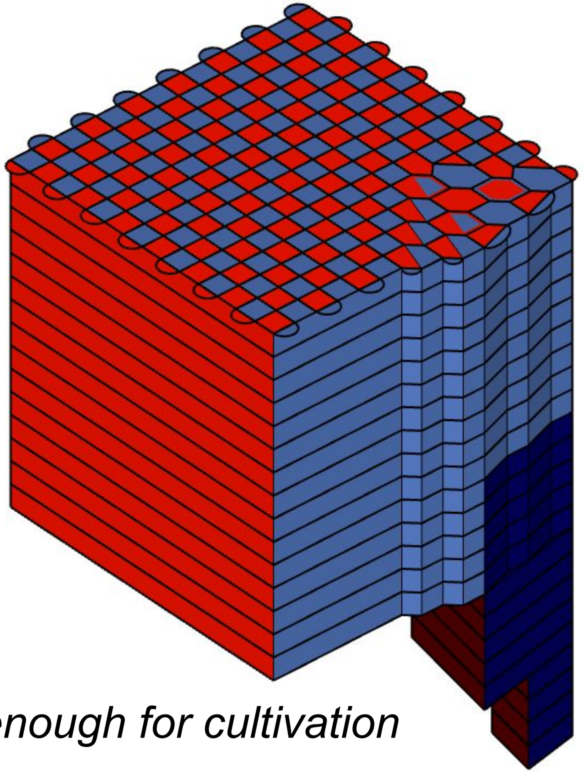
Non-Clifford gates still require feedback to complete

Good estimates can't just count Ts anymore

If T states are as cheap as CNOTs... why are CNOTs so expensive??

Architectures enabling transversal CNOTs might still be T state dominated

Summary



Cultivation is 10x cheaper than prior work

One T gate ≈ two CNOT gates

Cultivation reaches **2e-9** error (from noise strength 1e-3)

within 10x of relevance to factoring 2048 bit numbers

Cultivation fits in small spaces

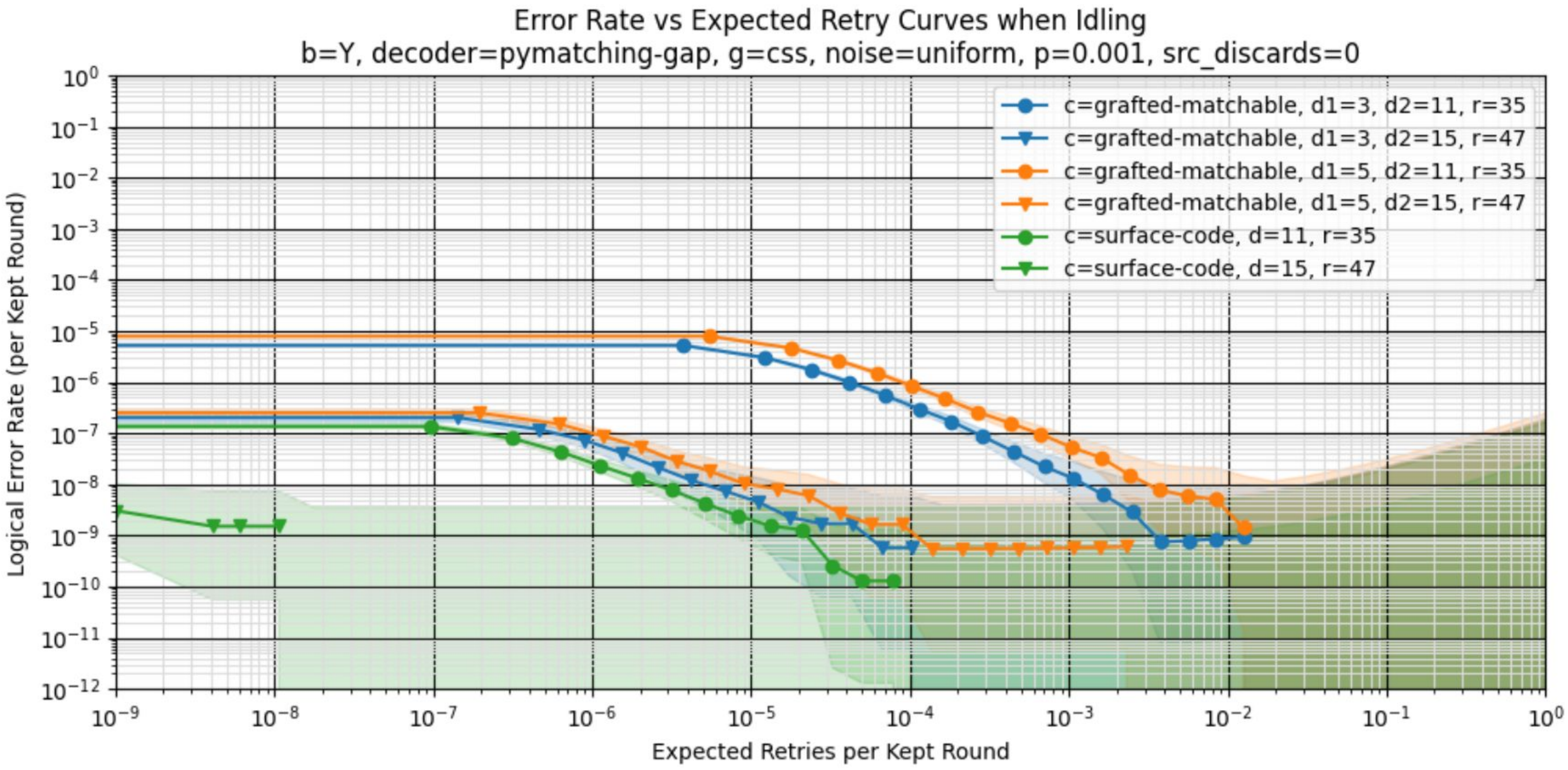
A hallway wide enough to route logical qubits is large enough for cultivation

Reducing noise improves cultivation dramatically

If physical qubits improve 10% per year, we might never need distillation

Bonus Figures

Cost of waiting longer



What if we still need distillation?