Clique: Better Than Worst-Case Decoding for Quantum Error Correction

Gokul Subramanian Ravi[1], Jonathan M. Baker[1], Arash Fayyazi[2], Sophia Fuhui Lin[1], Ali Javadi-Abhari[3], Massoud Pedram[2], Frederic T. Chong[1]

1: UChicago, 2: USC, 3: IBM

Best of both worlds approach, combining two schools of QEC decoding, for 100x–10,000x gains!!

arXiv:2208.08547

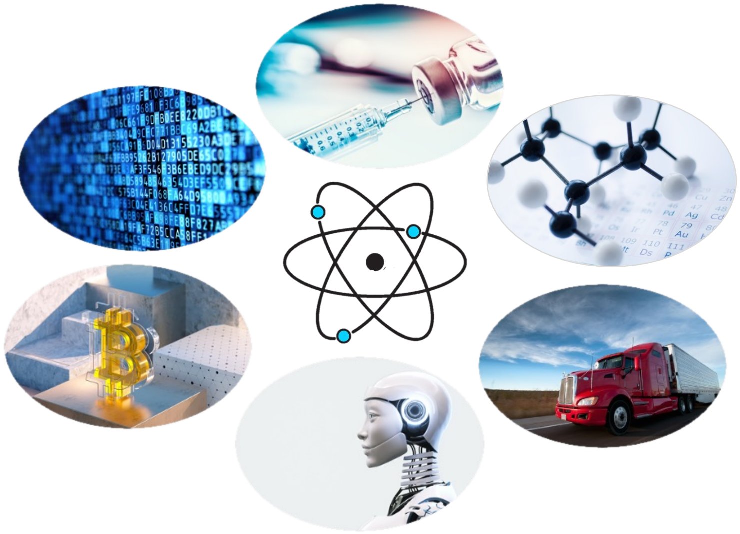
Error correction for fault tolerant quantum systems

Physical qubit

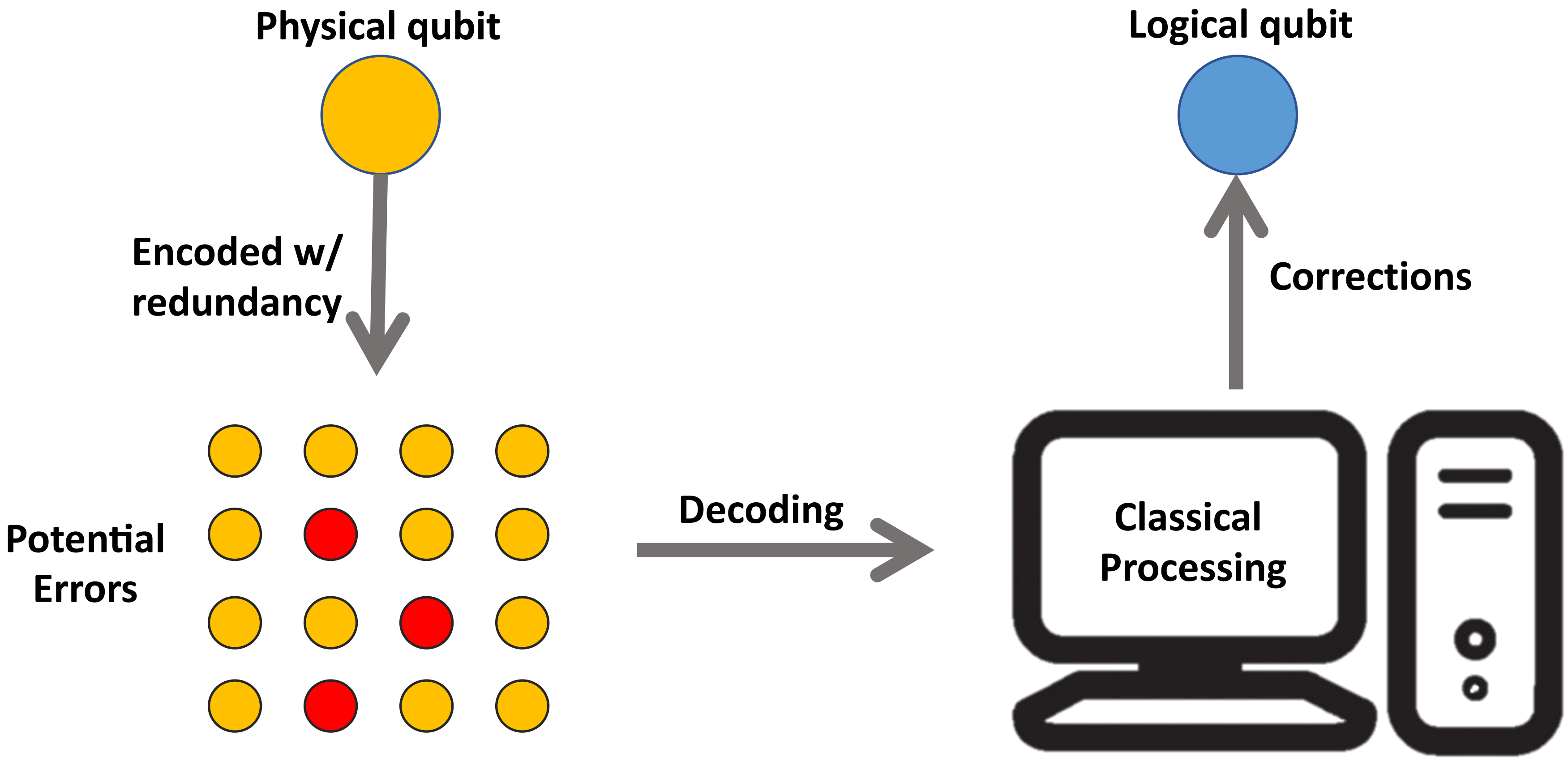
ER = $10^{-2} - 10^{-4}$

Logical qubit

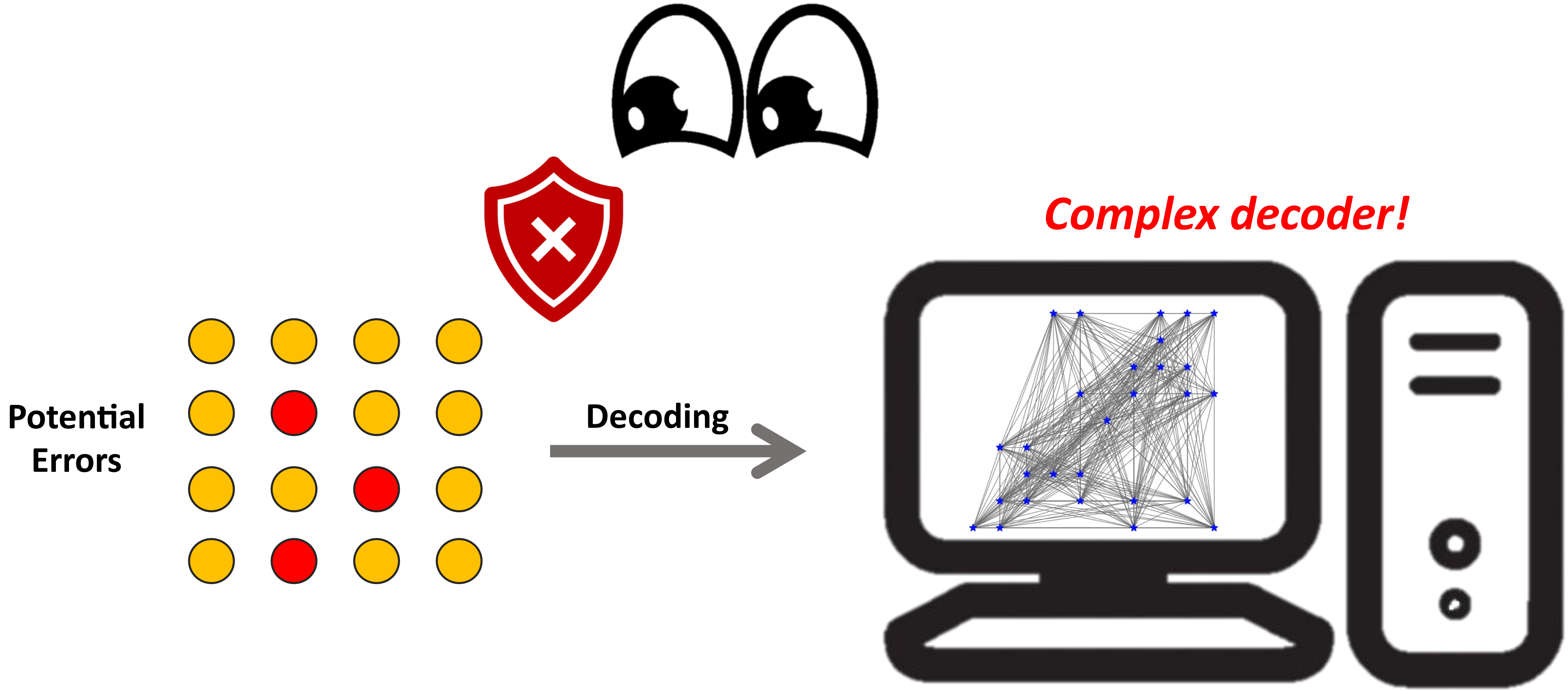
ER = $10^{-6} - 10^{-15}$



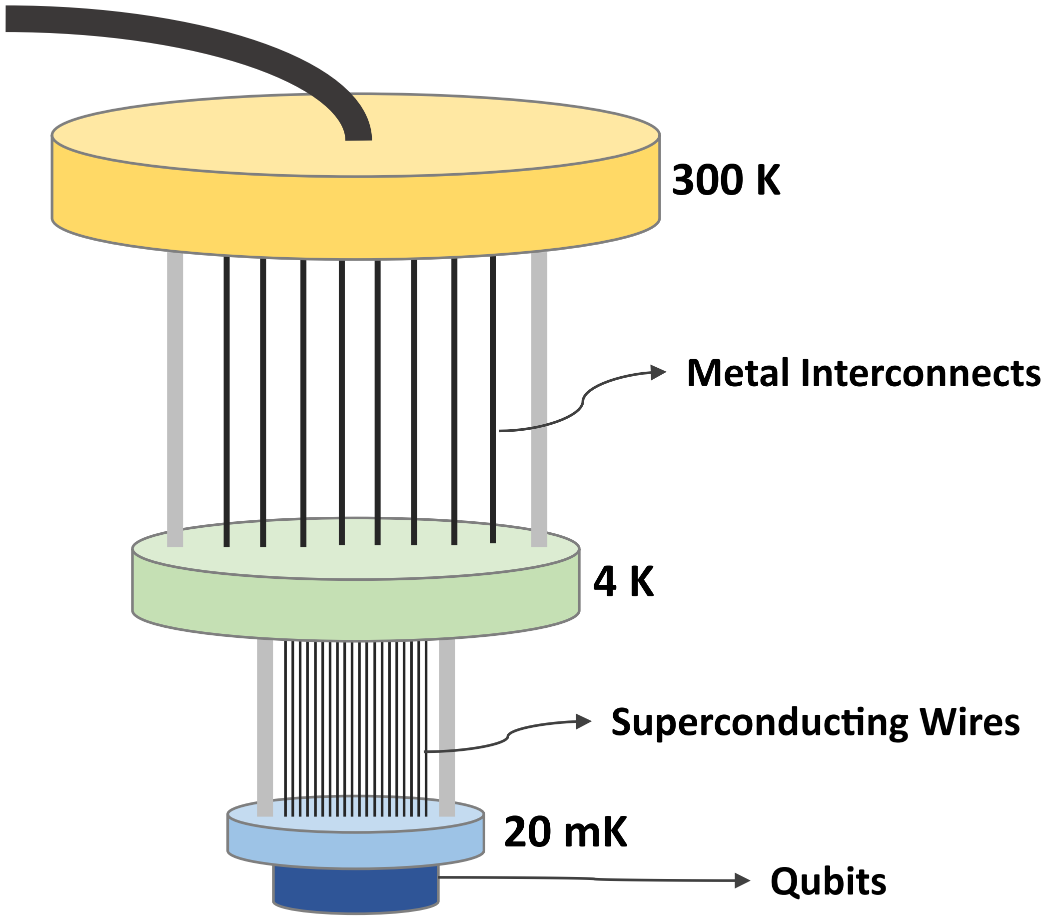
Error correction for fault tolerant quantum systems



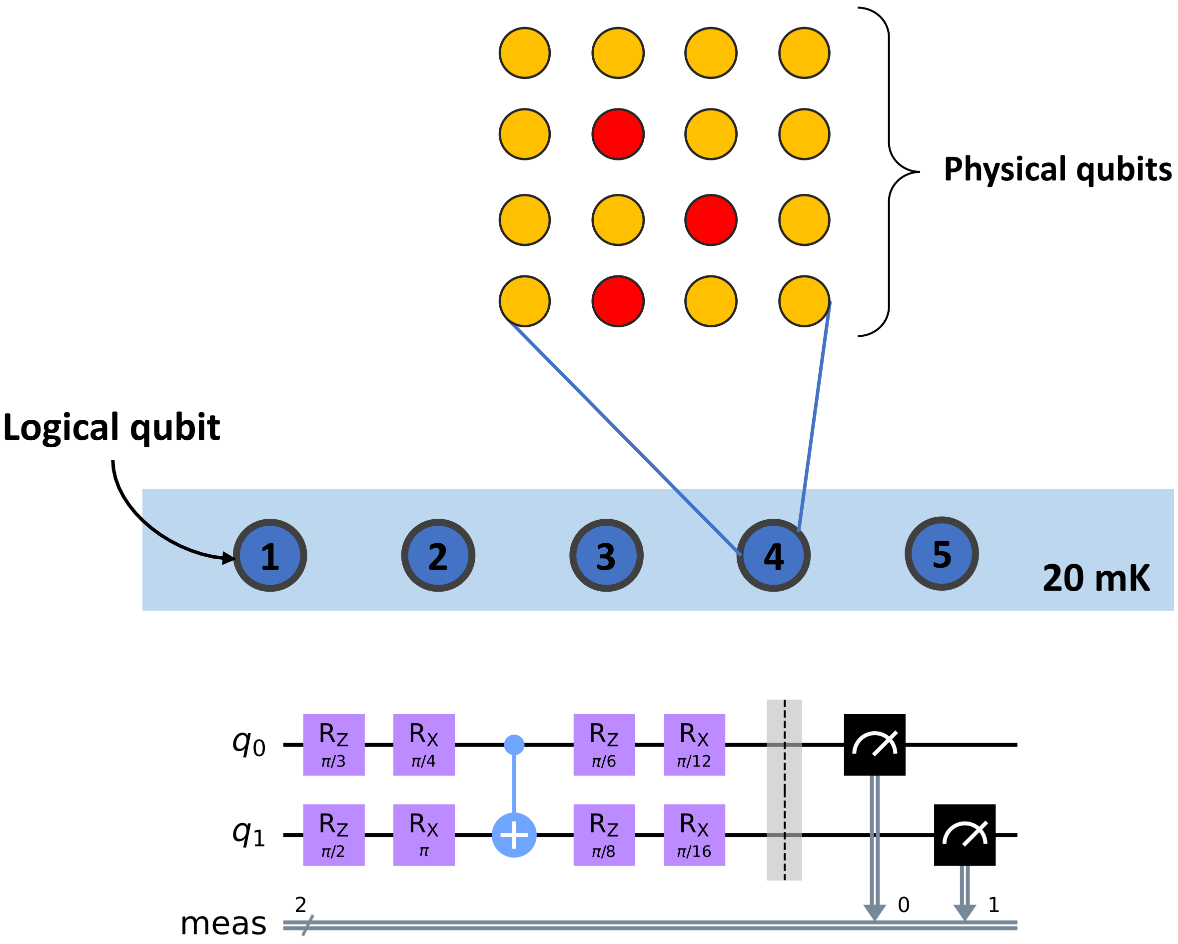
Error correction for fault tolerant quantum systems

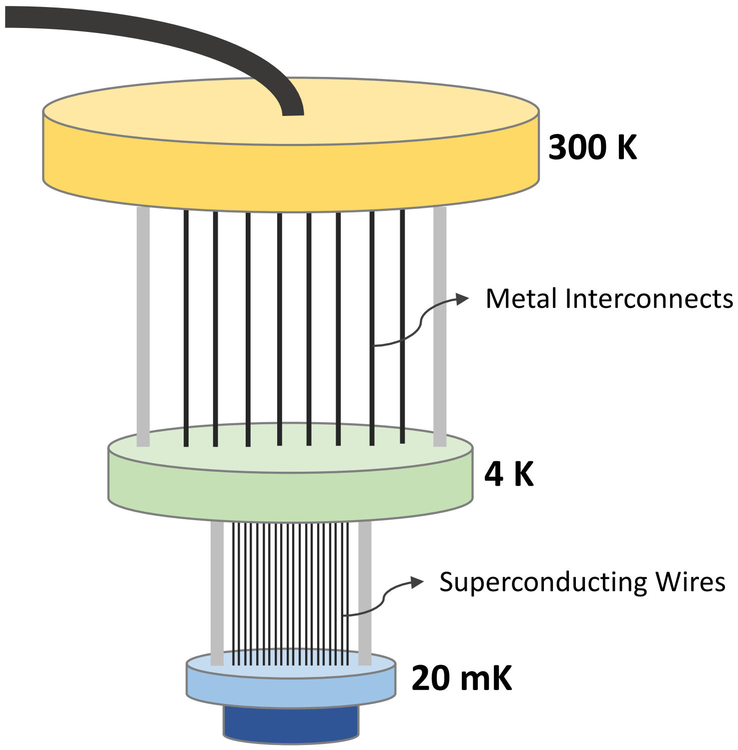


Scope: Cryogenic quantum systems



Scope: Cryogenic quantum systems

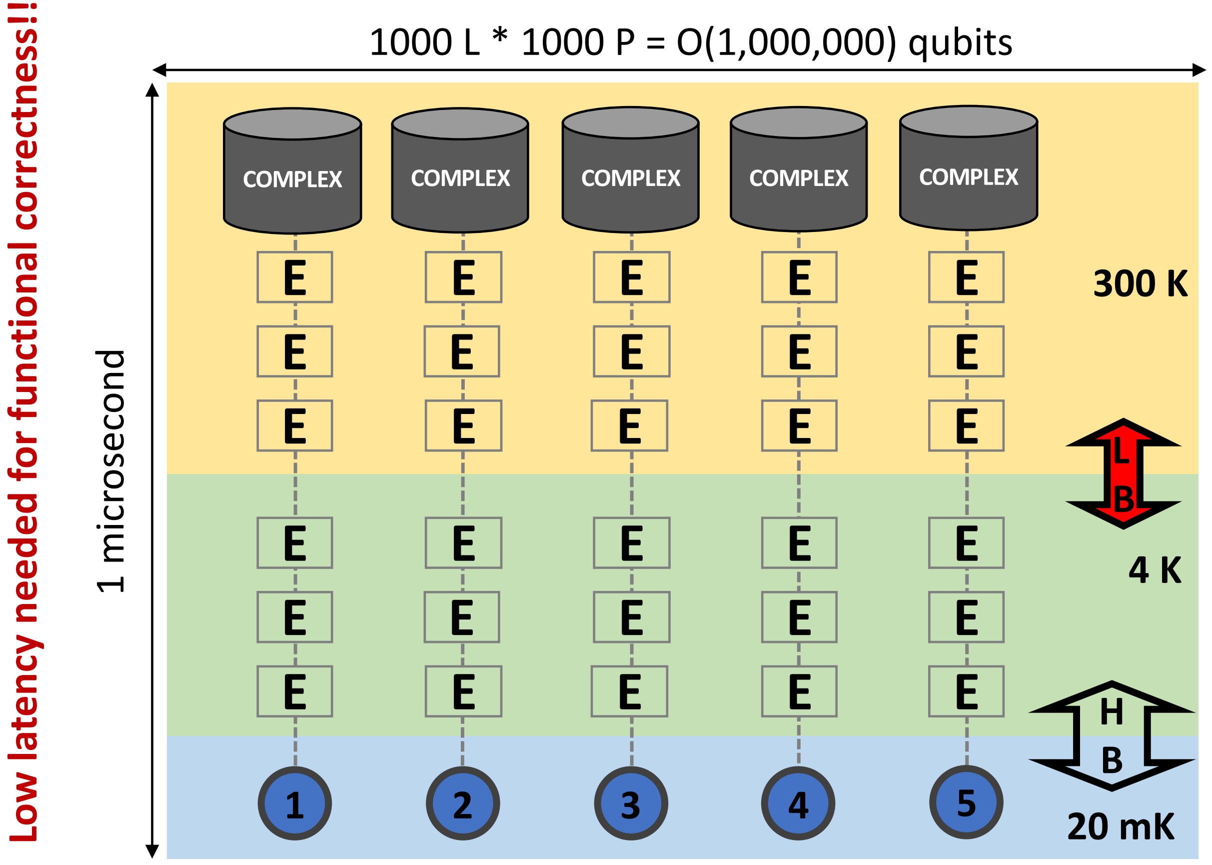




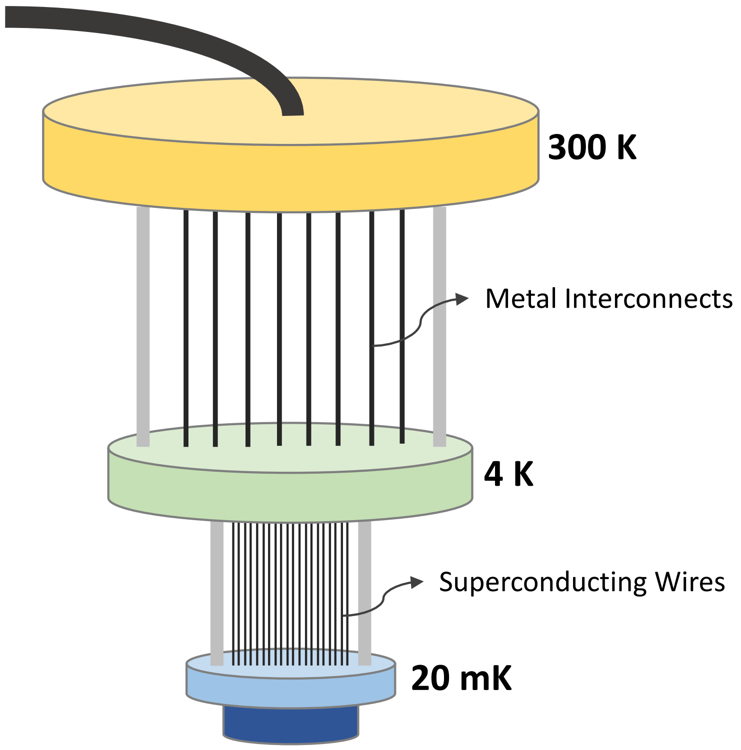
System-level view: Traditional outside-fridge QEC decoding

Tbps I/O bandwidth → <u>bandwidth bottleneck!</u>

[Fowler, PR-A ’12]
..
..
[Das, HPCA ’22]



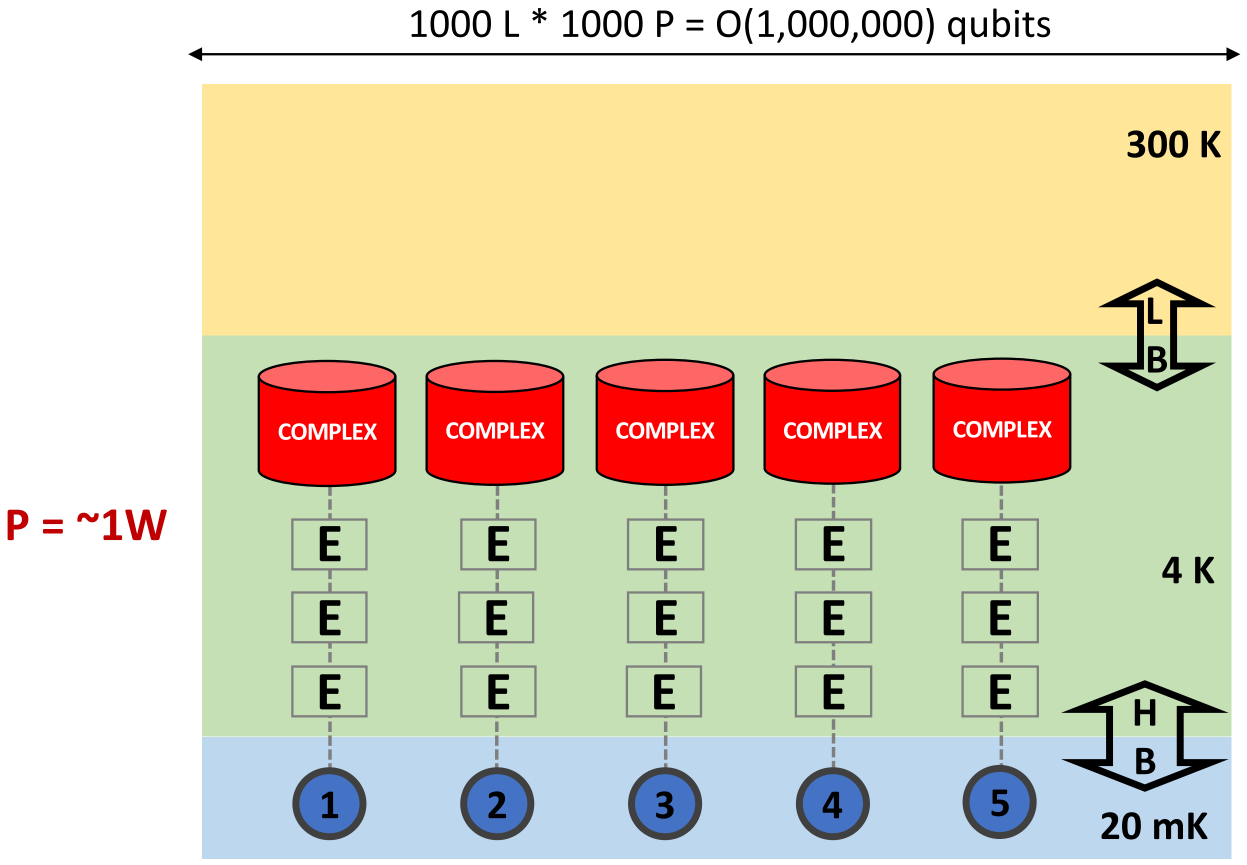
E: Error Signatures

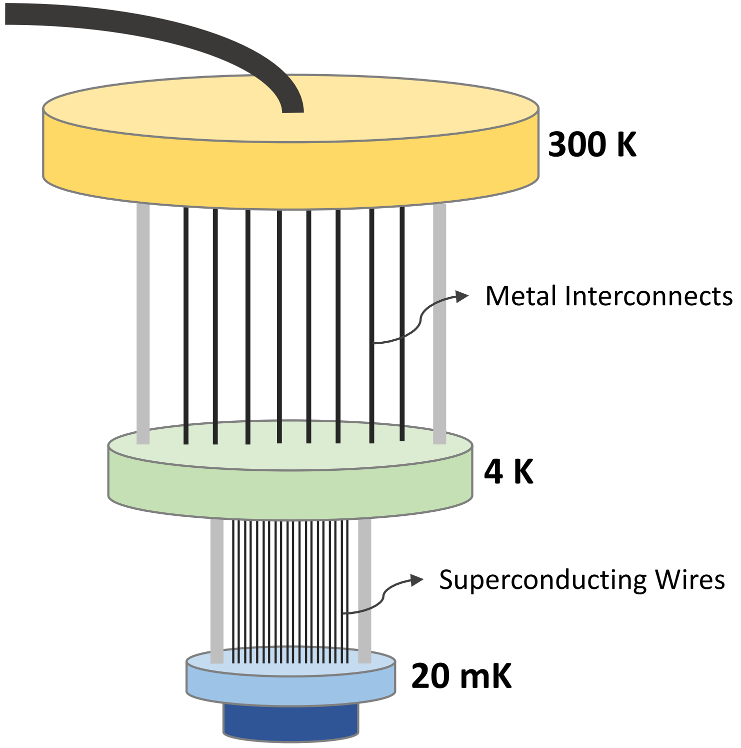


System-level view: Cryogenic inside-fridge QEC decoding

Limited cryogenic power budget (~1W) cryo-resource __bottleneck__!

[Holmes, ISCA '20]
[Byun, ISCA '22]
[Ueno, HPCA '22]

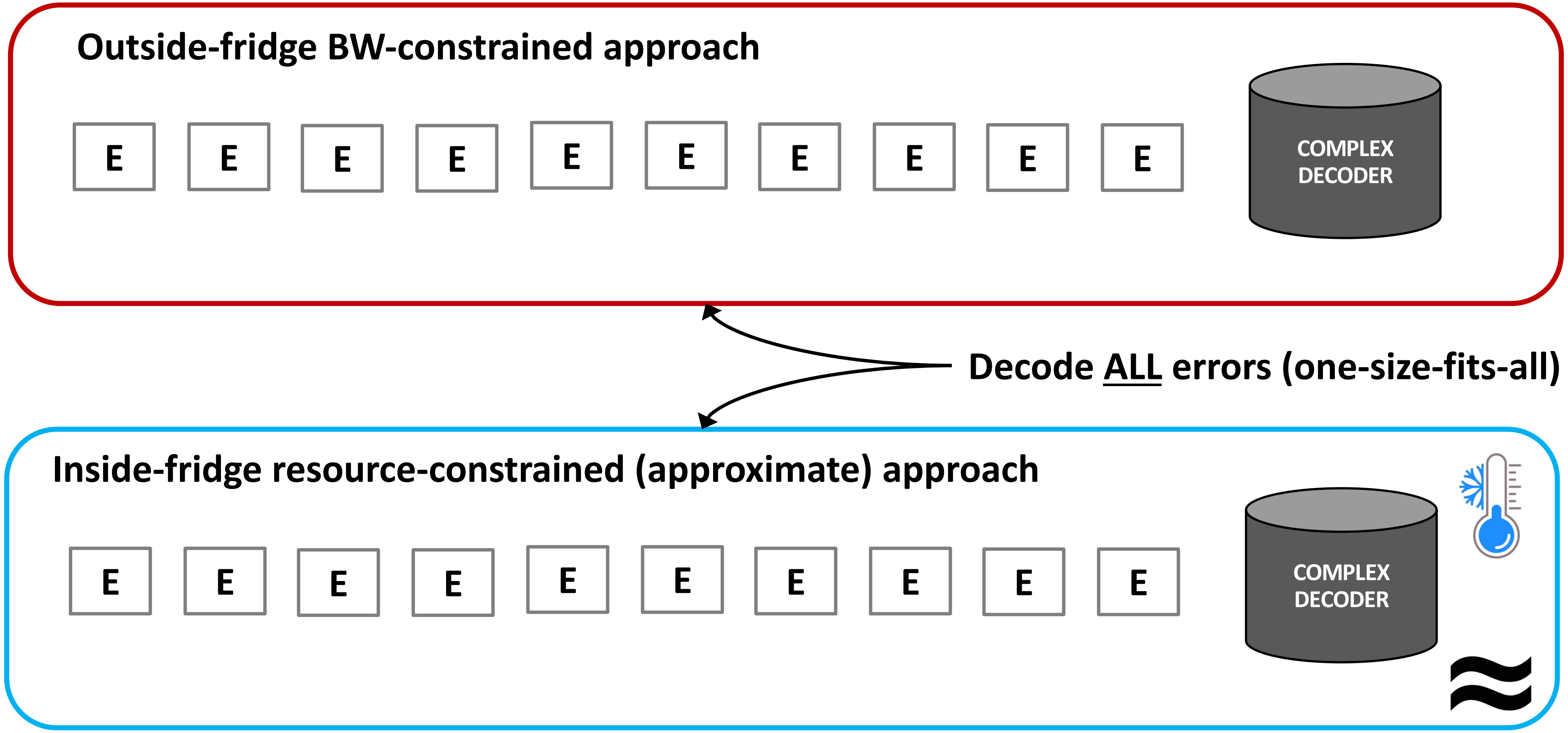


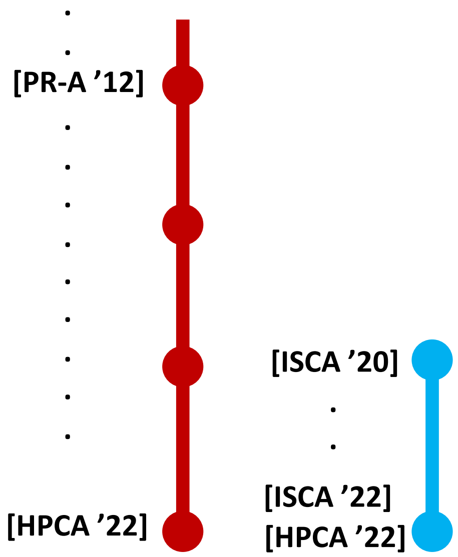


E: Error Signatures

Better than worst case approach to decoding

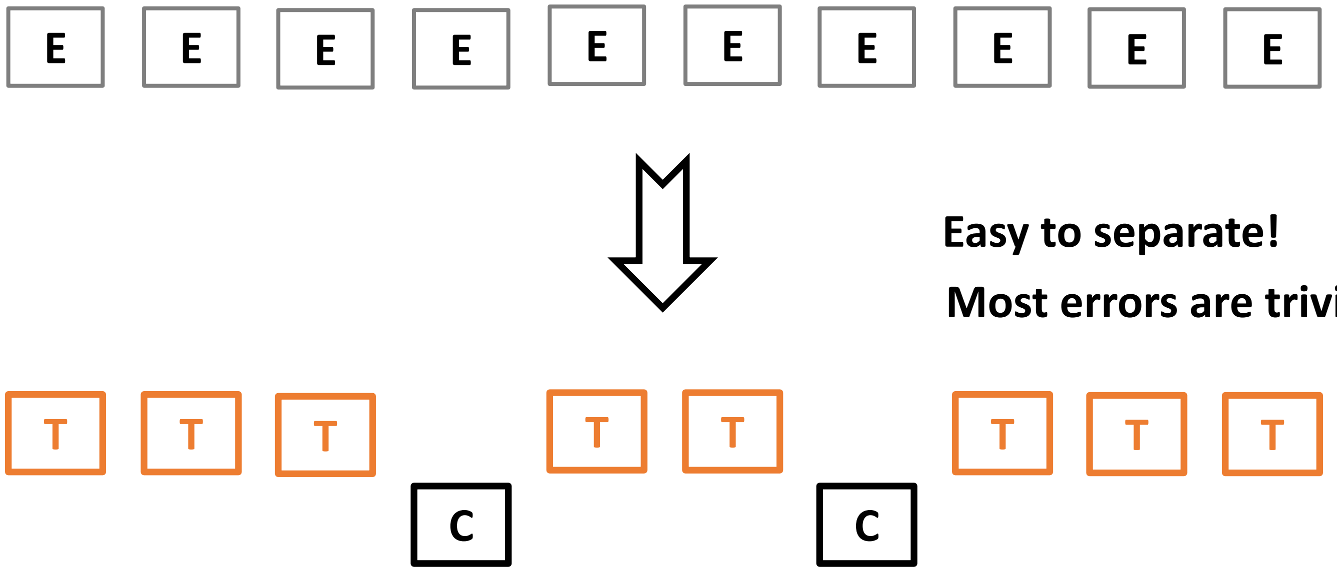
<u>Key Insight:</u> Not all errors hard to decode → Separate common trivial errors from rare complex errors.





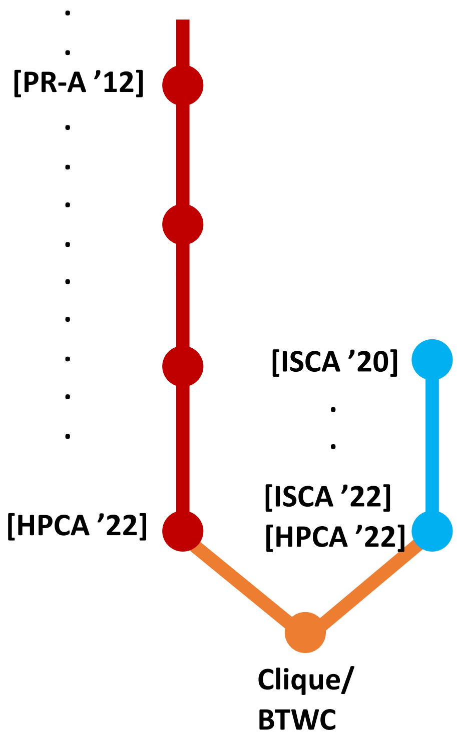
Better than worst case approach to decoding

<u>Key Insight:</u> Not all errors hard to decode → Separate common trivial errors from rare complex errors.



Easy to separate!

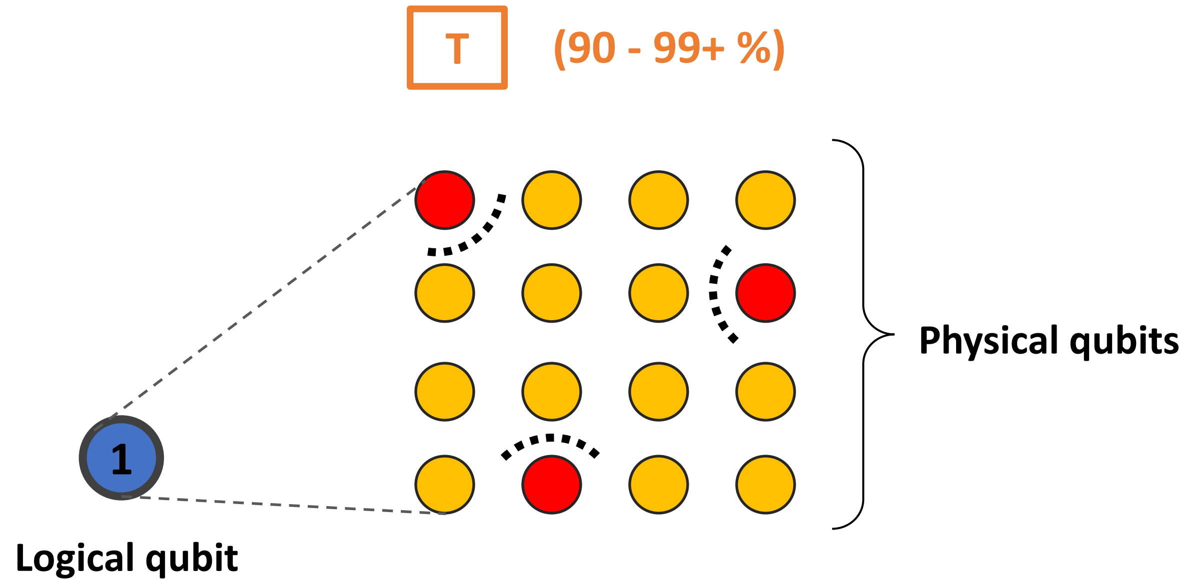
Most errors are trivial!

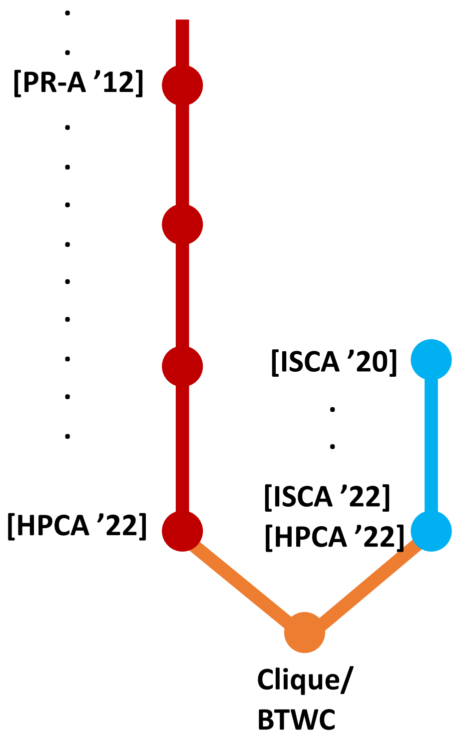


T: Trivial-to-decode
C: Complex-to-decode

Better than worst case approach to decoding

Isolated errors: Trivial to decode and very common!!



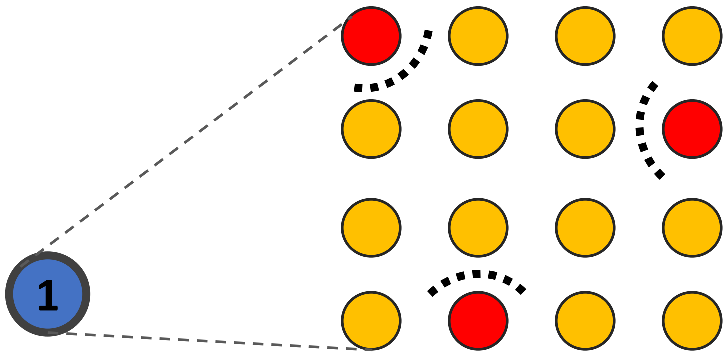


T: Trivial-to-decode
C: Complex-to-decode

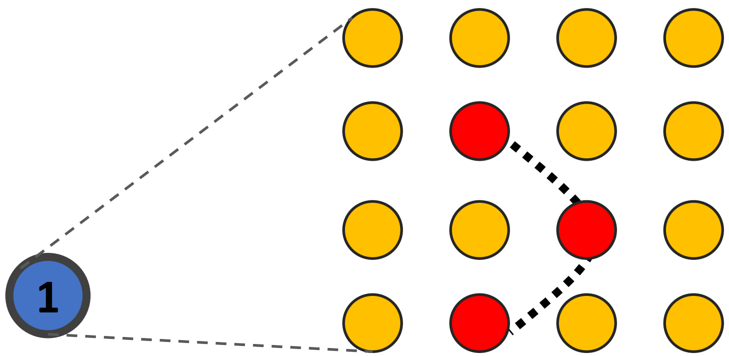
Better than worst case approach to decoding

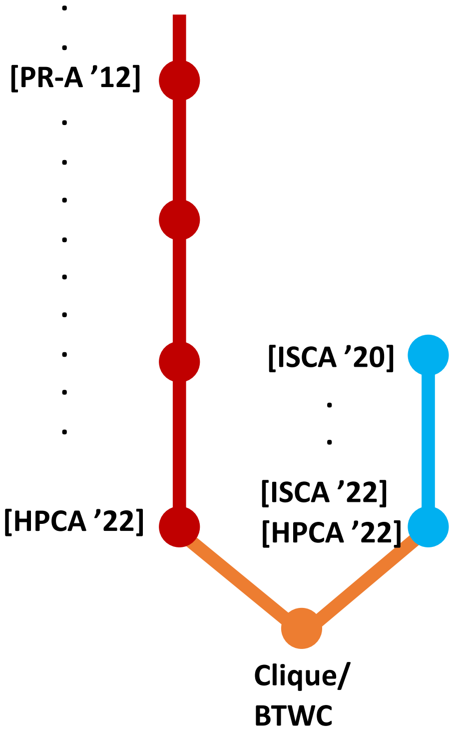
Error chains: Hard to decode and very rare!!

T (90 - 99+ %)



C (<1 - 10 %)

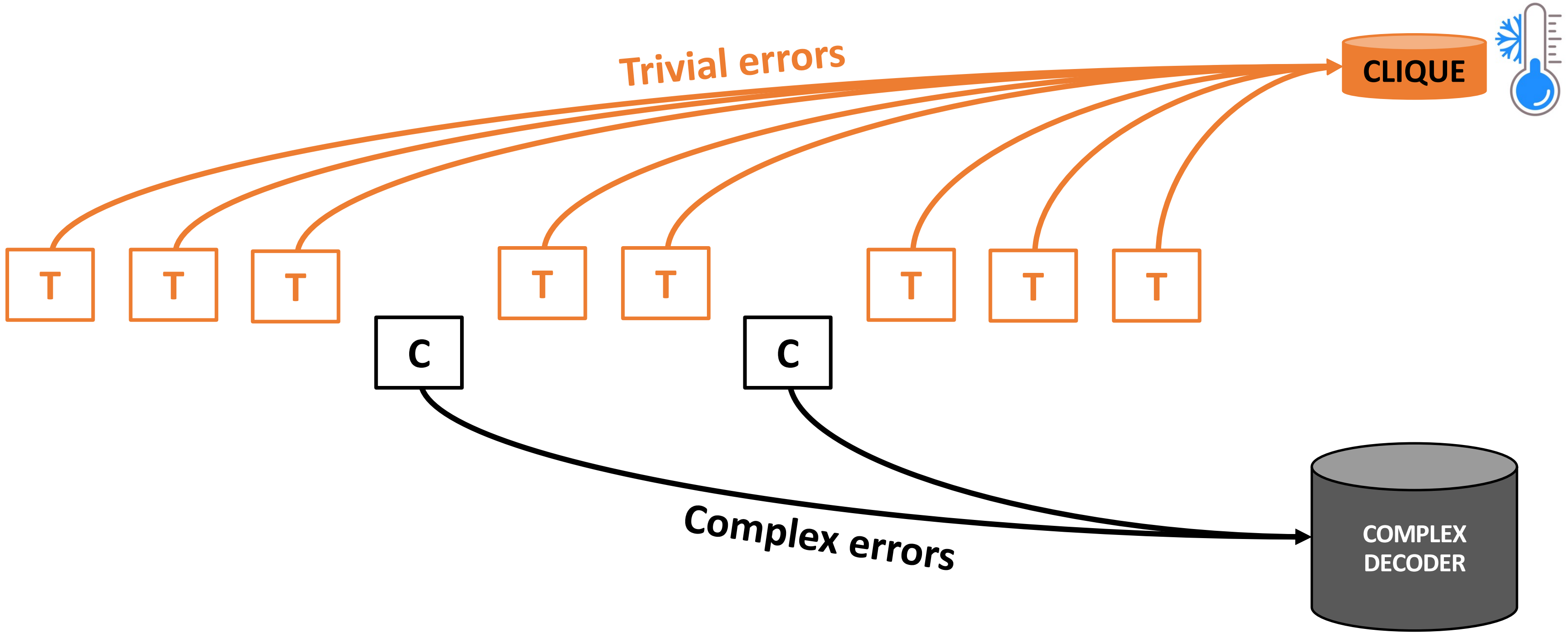




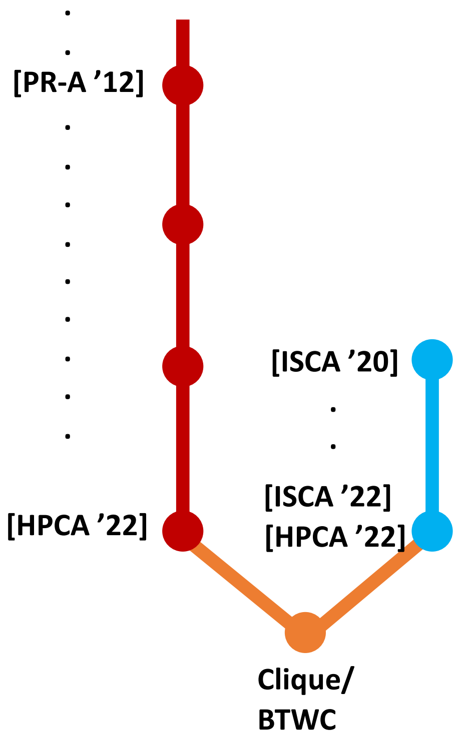
T: Trivial-to-decode
C: Complex-to-decode

Better than worst case approach to decoding

Common trivial errors → simple cryogenic 'Clique' decoder.
Rare complex errors → outside-fridge SOTA complex decoder.

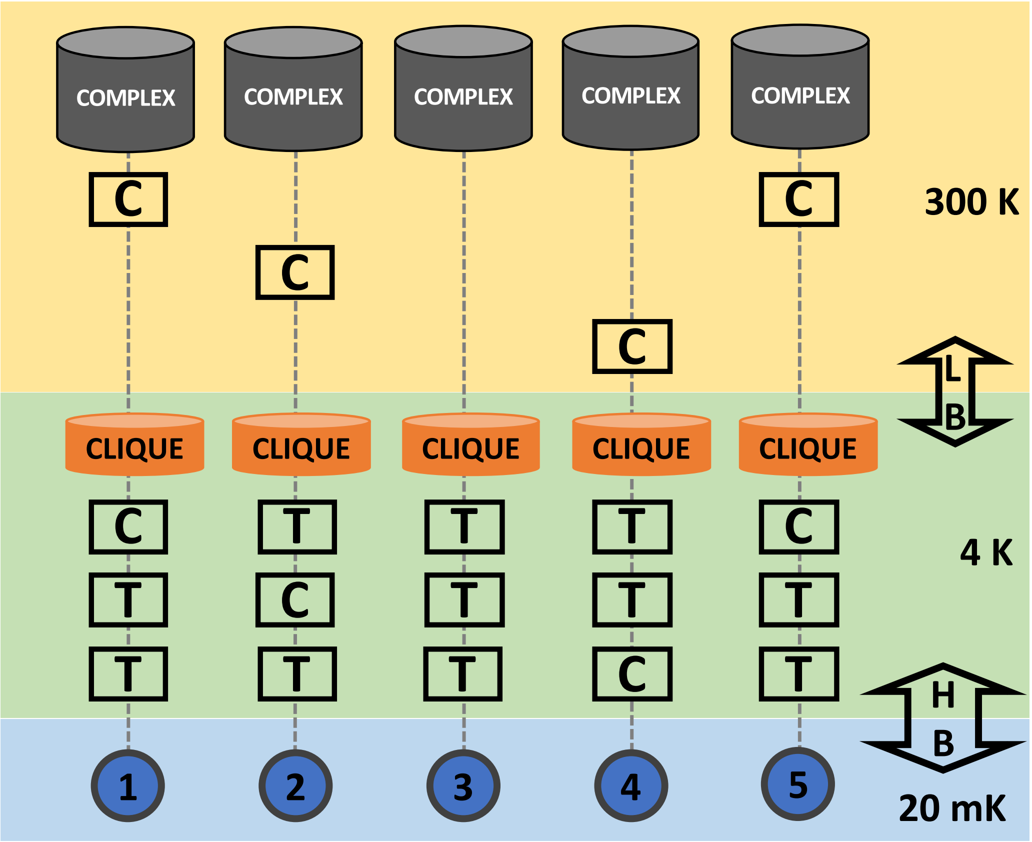


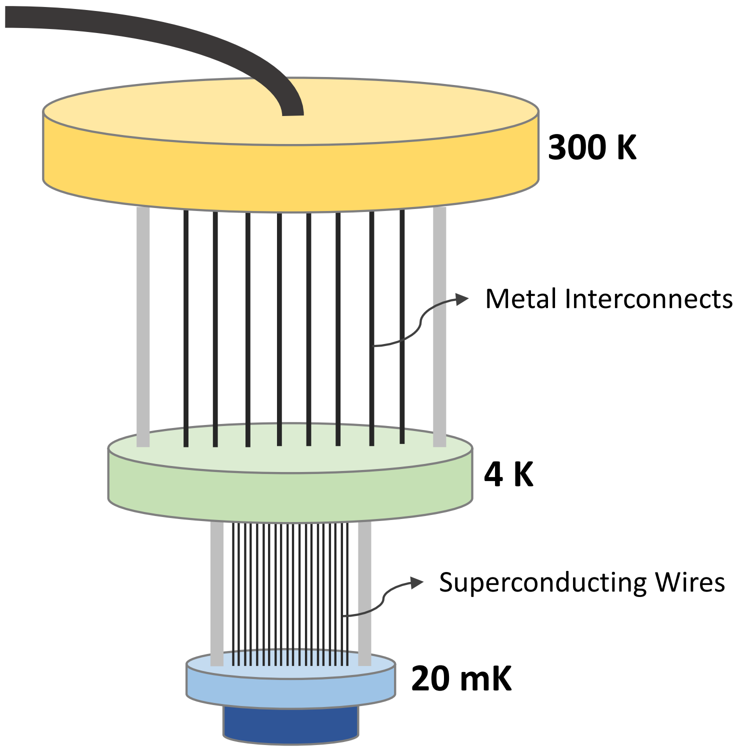
T: Trivial-to-decode
C: Complex-to-decode



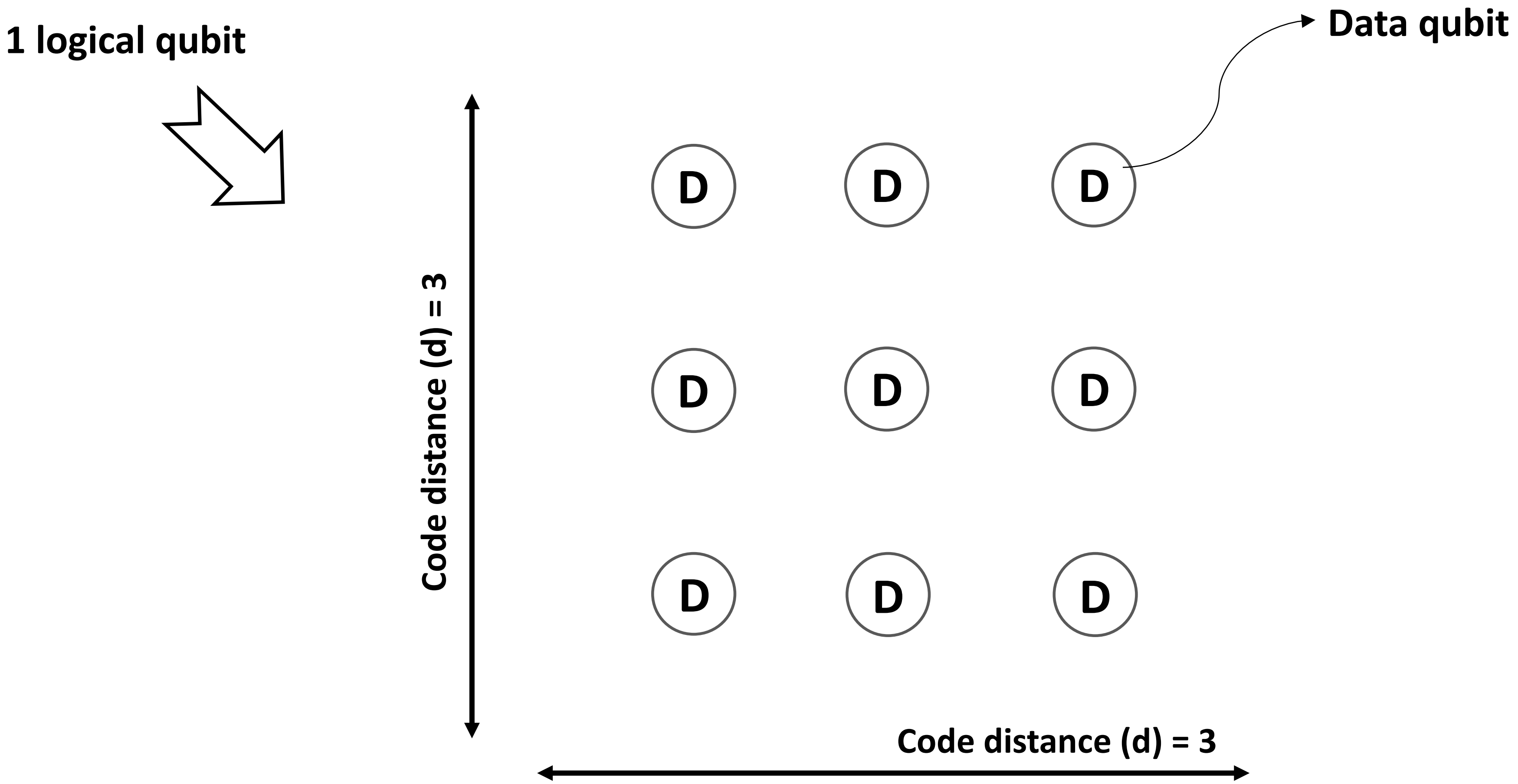
System-level view: Better than worse-case decoding

Reduced outside-fridge decoding → No bandwidth bottleneck!
Reduced inside-fridge decoding HW → No cryo-resource bottleneck!



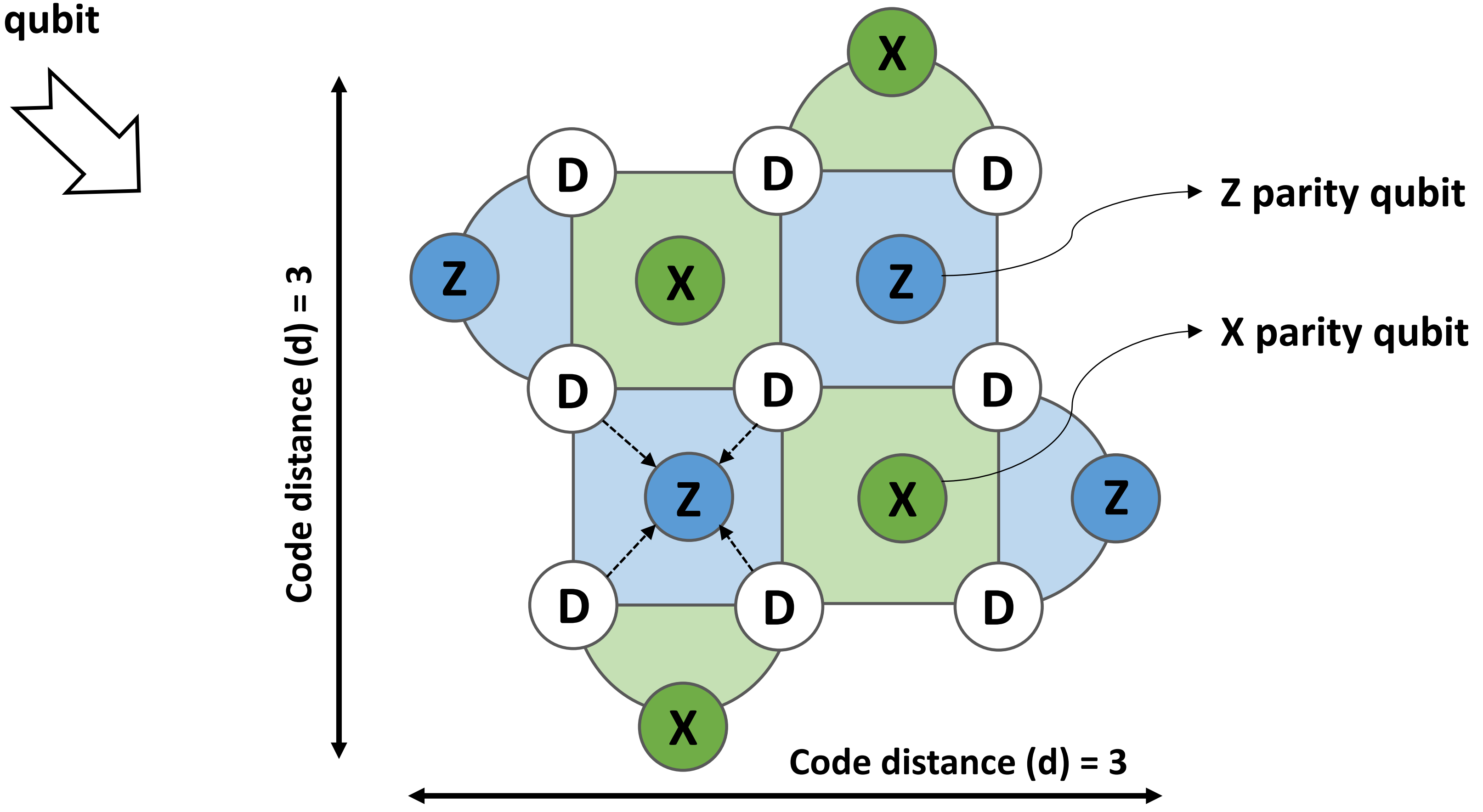


How does QEC work? (Surface code)

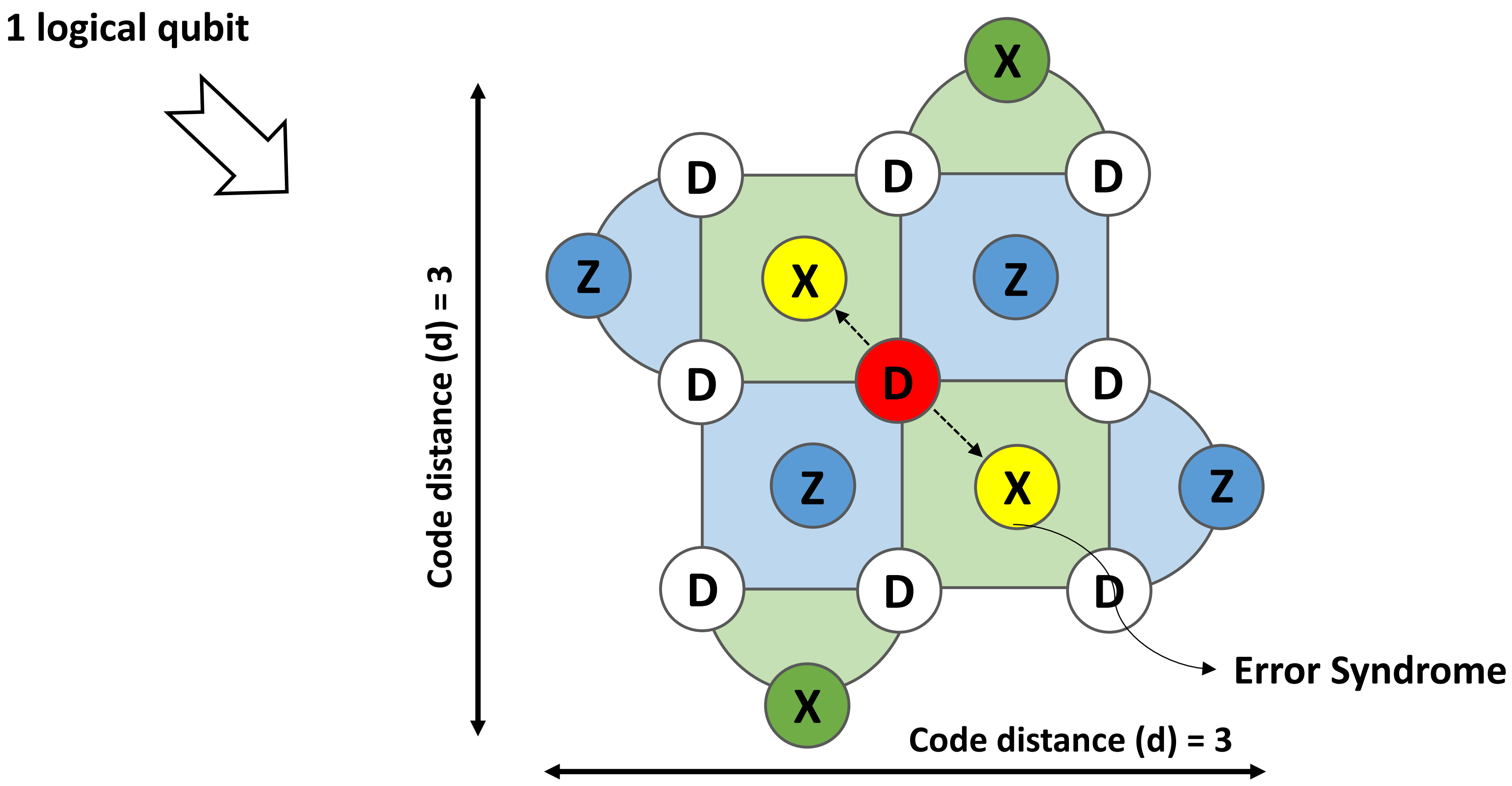


How does QEC work? (Surface code)

1 logical qubit

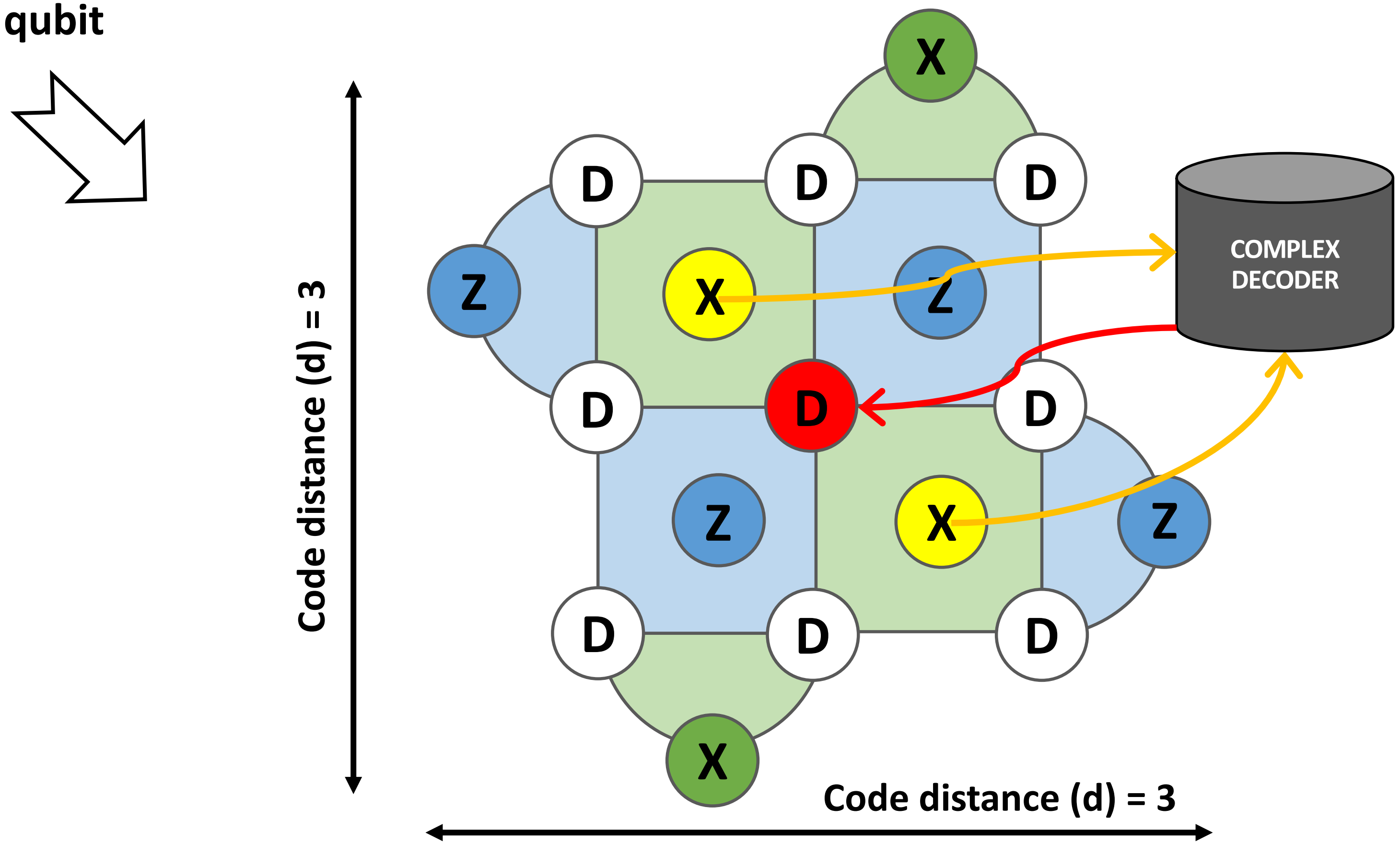


How does QEC work? (Surface code)

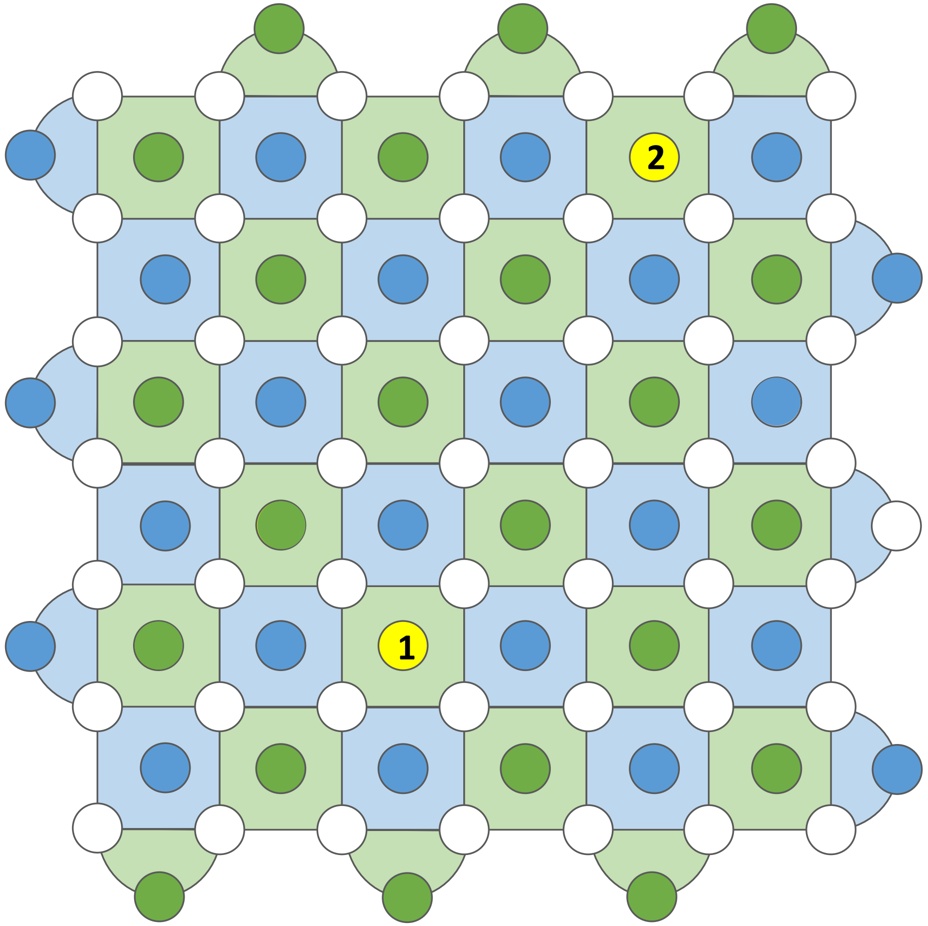


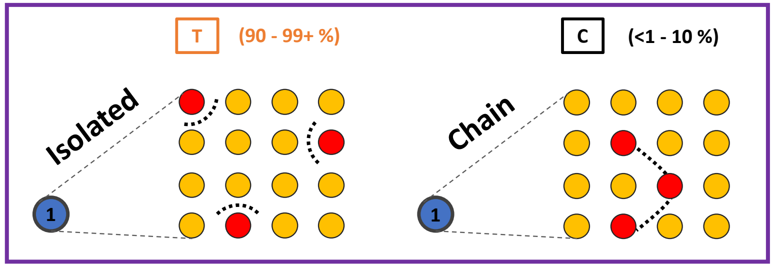
How does QEC work? (Surface code)

1 logical qubit

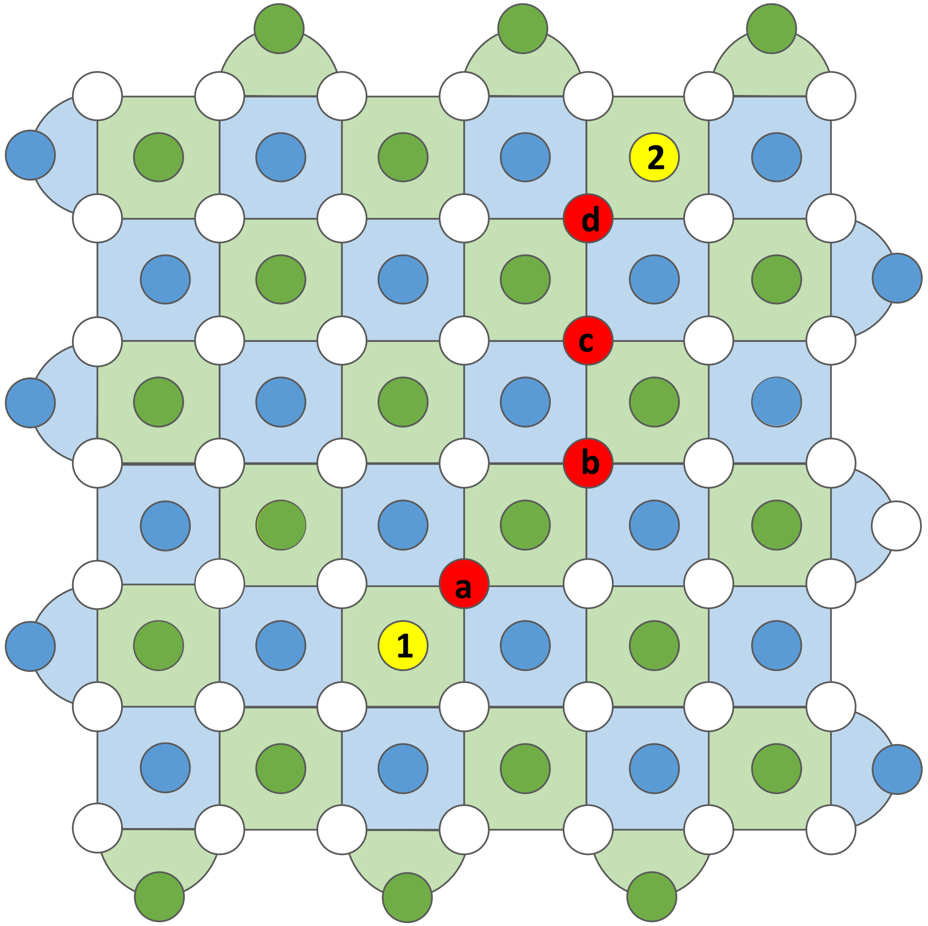


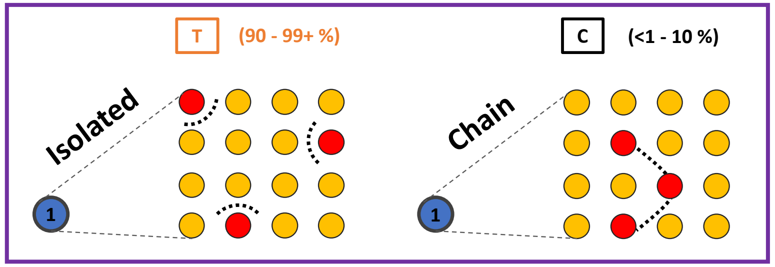
Why are error chains hard to decode?



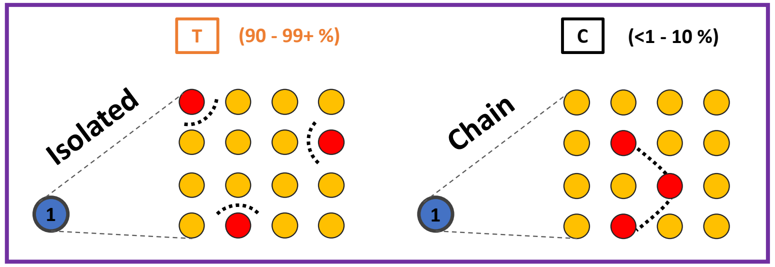


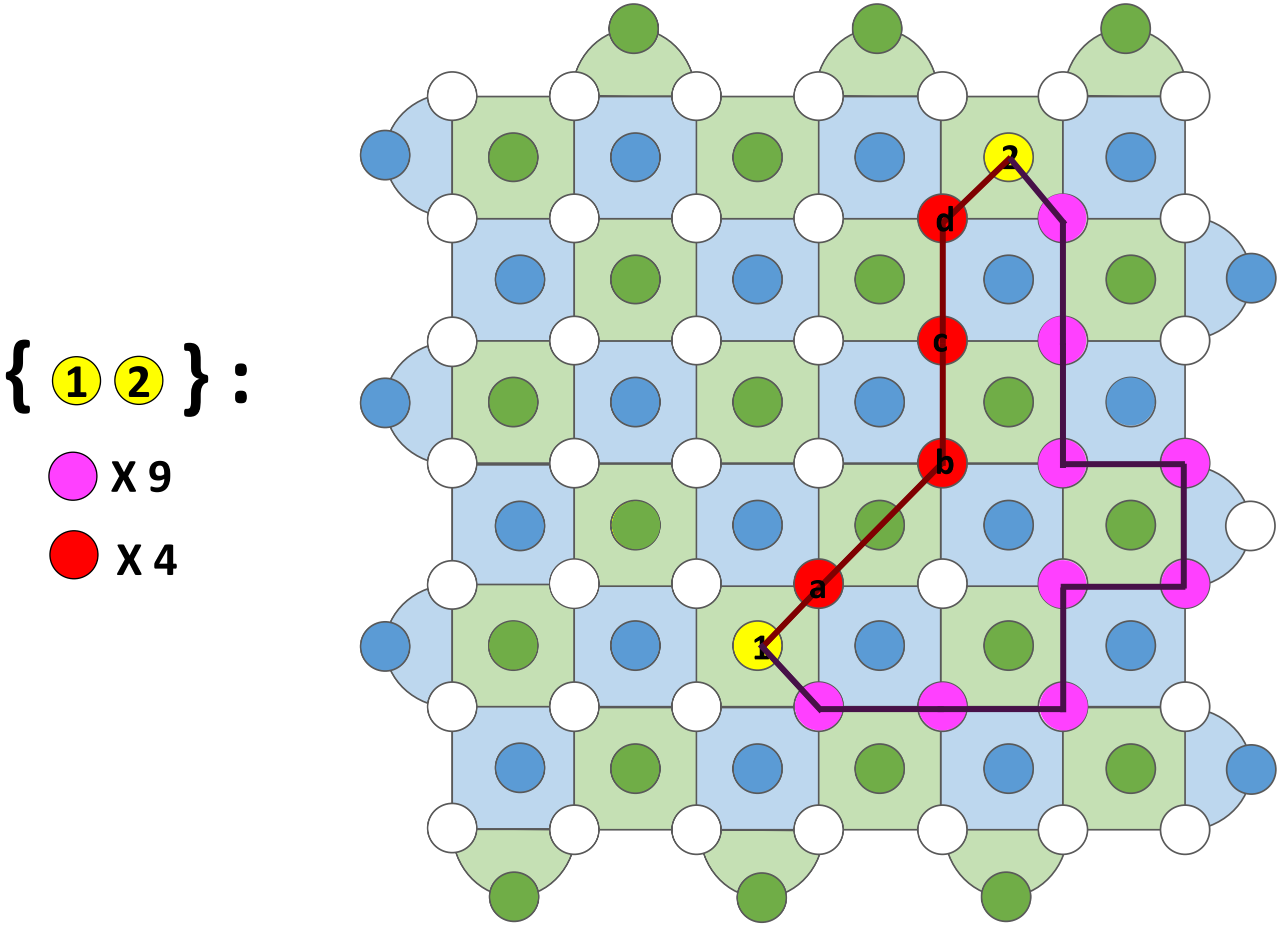
Why are error chains hard to decode?





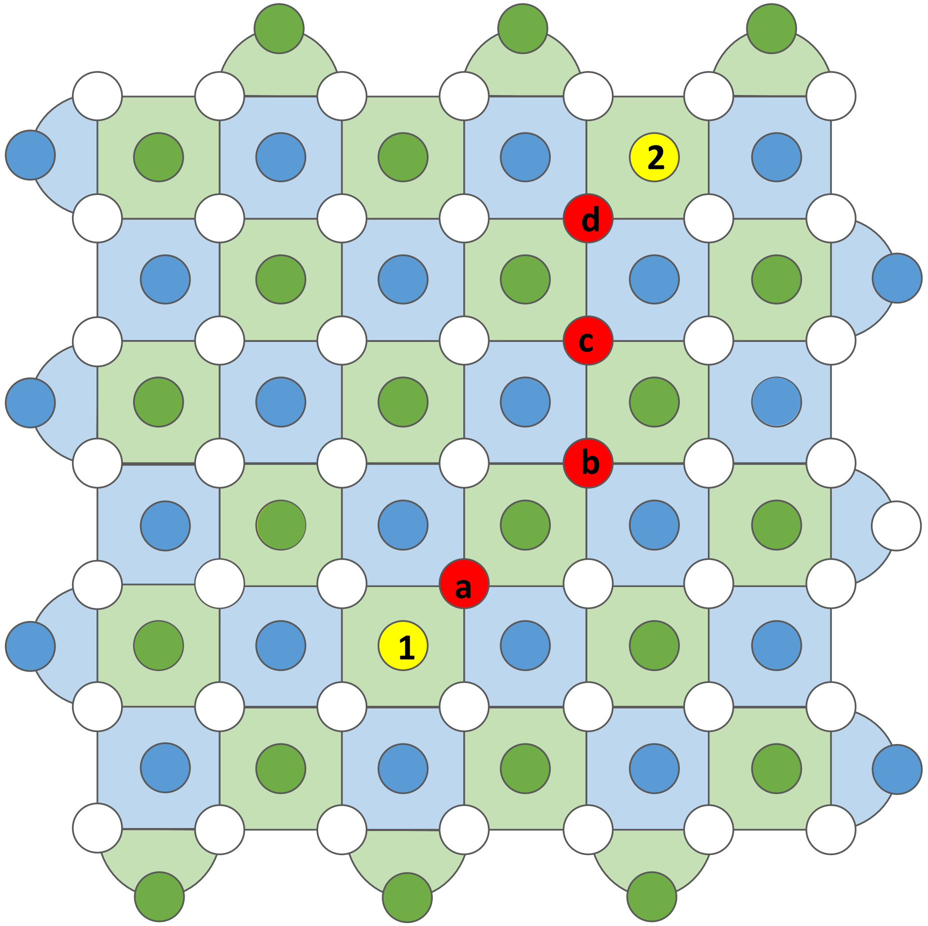
Why are error chains hard to decode?

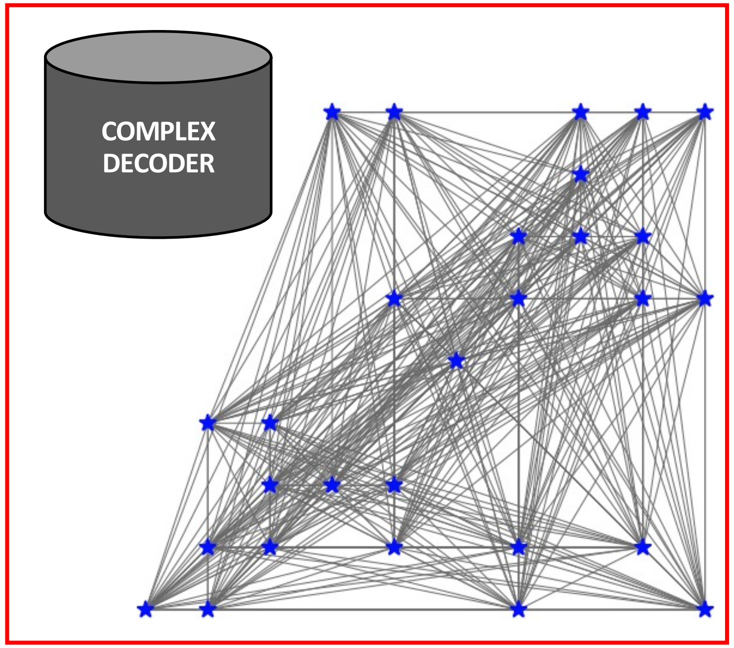




Chained data errors trigger non-local syndromes which are challenging to pair and decode.

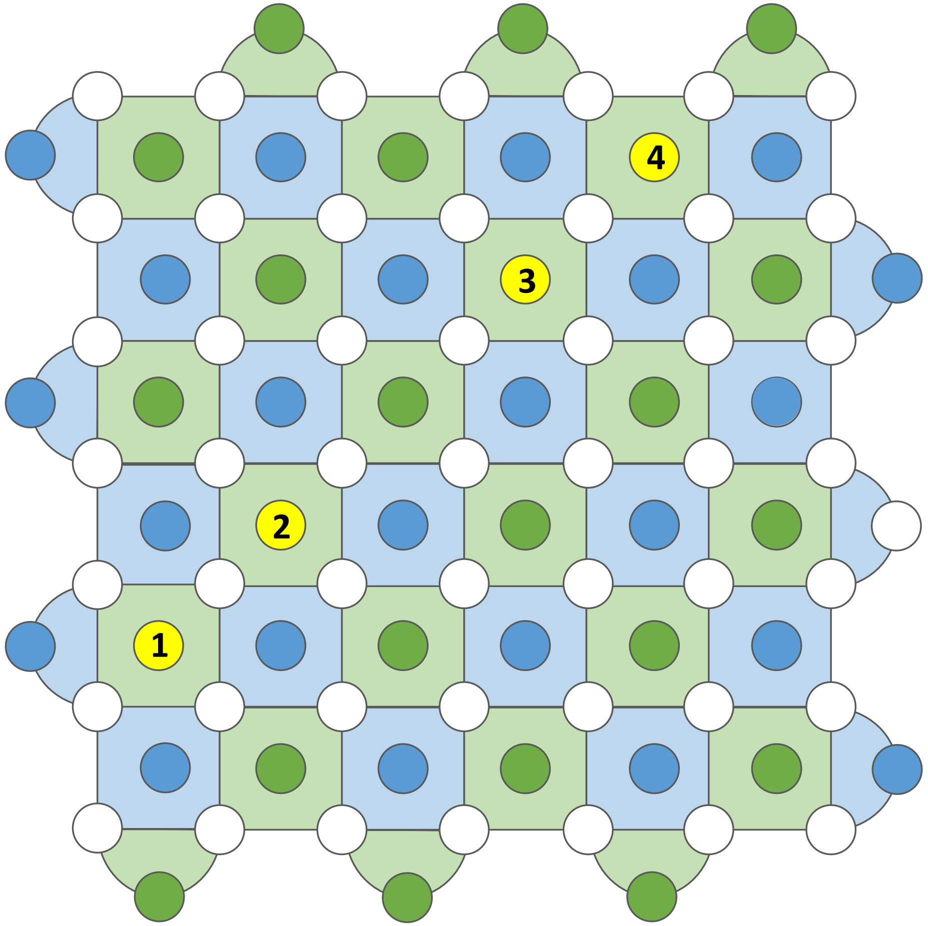
Why are error chains hard to decode?

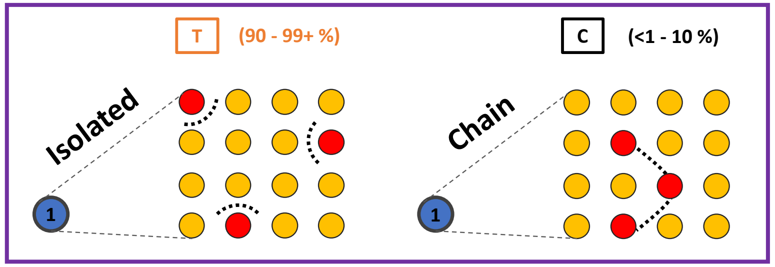




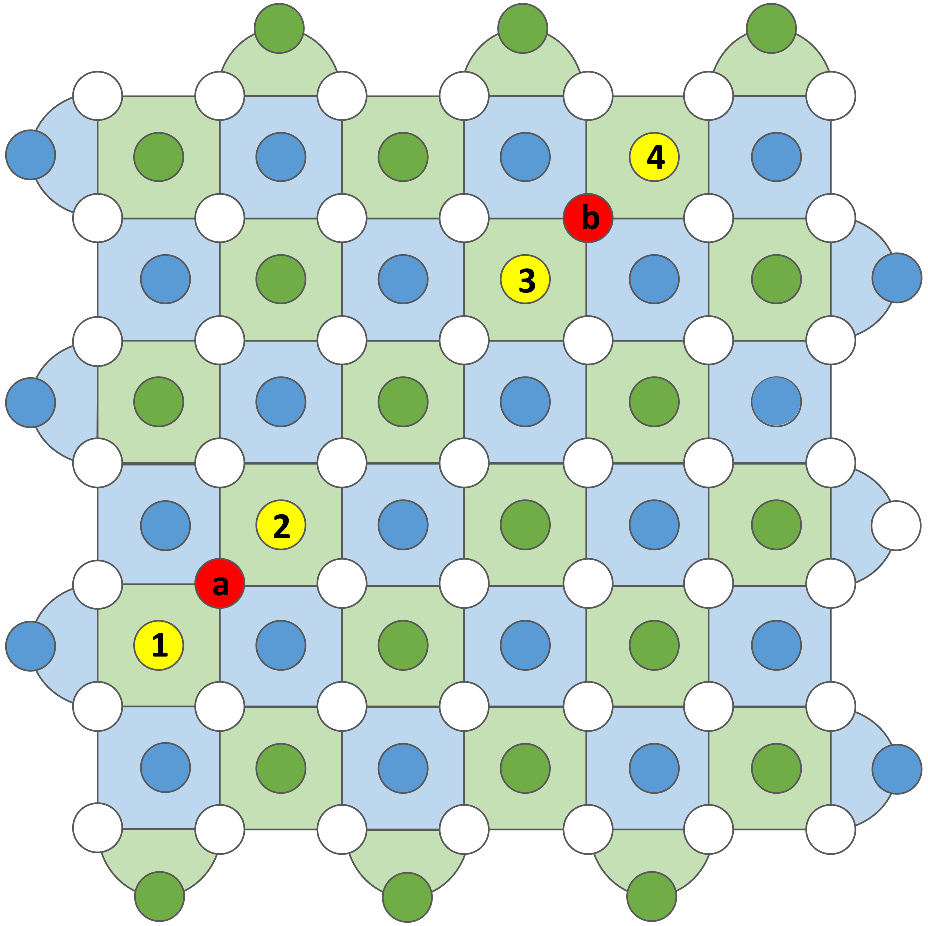
Minimum Weight Perfect Matching

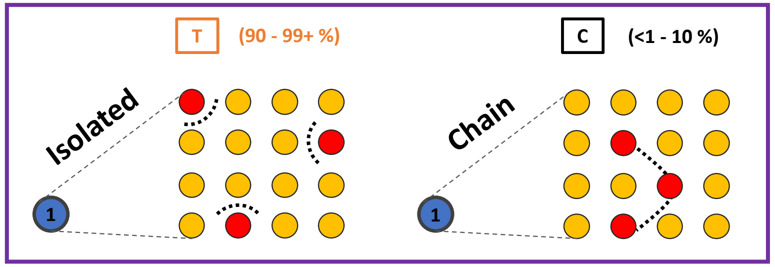
Why are isolated errors trivial to decode?



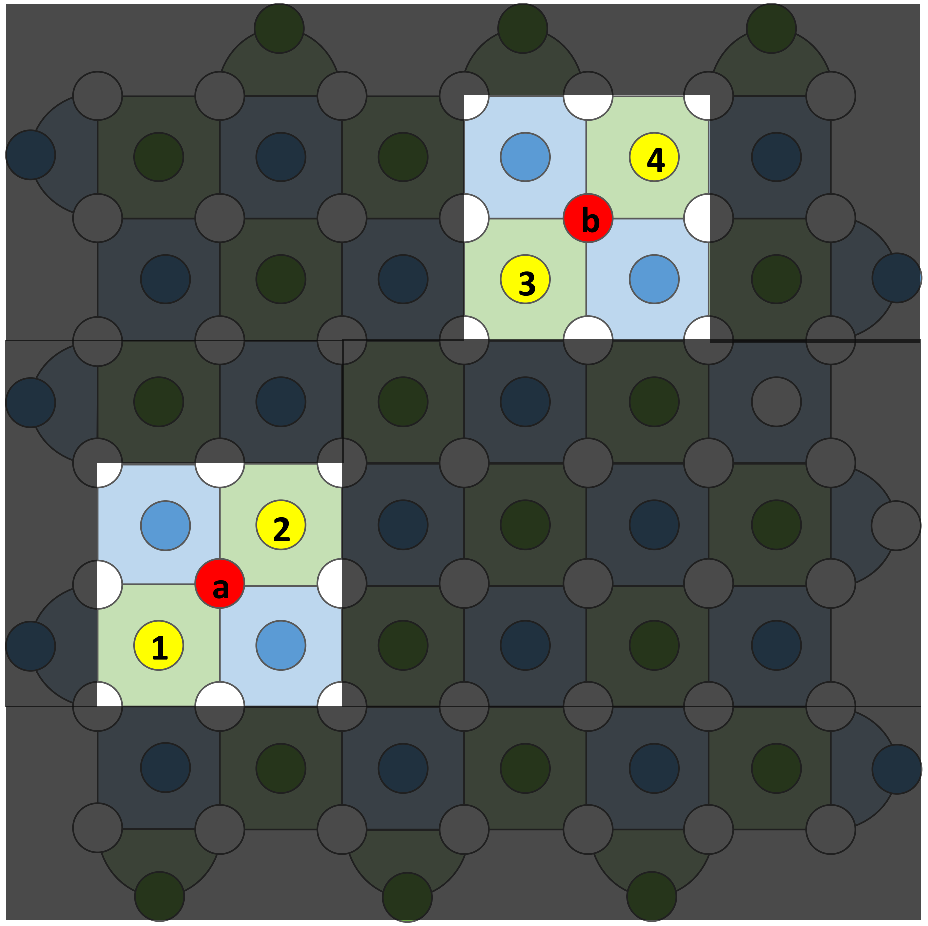


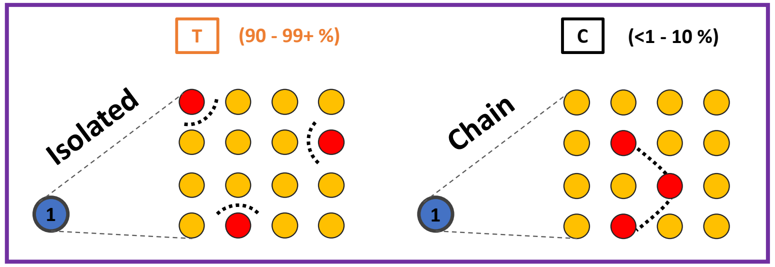
Why are isolated errors trivial to decode?



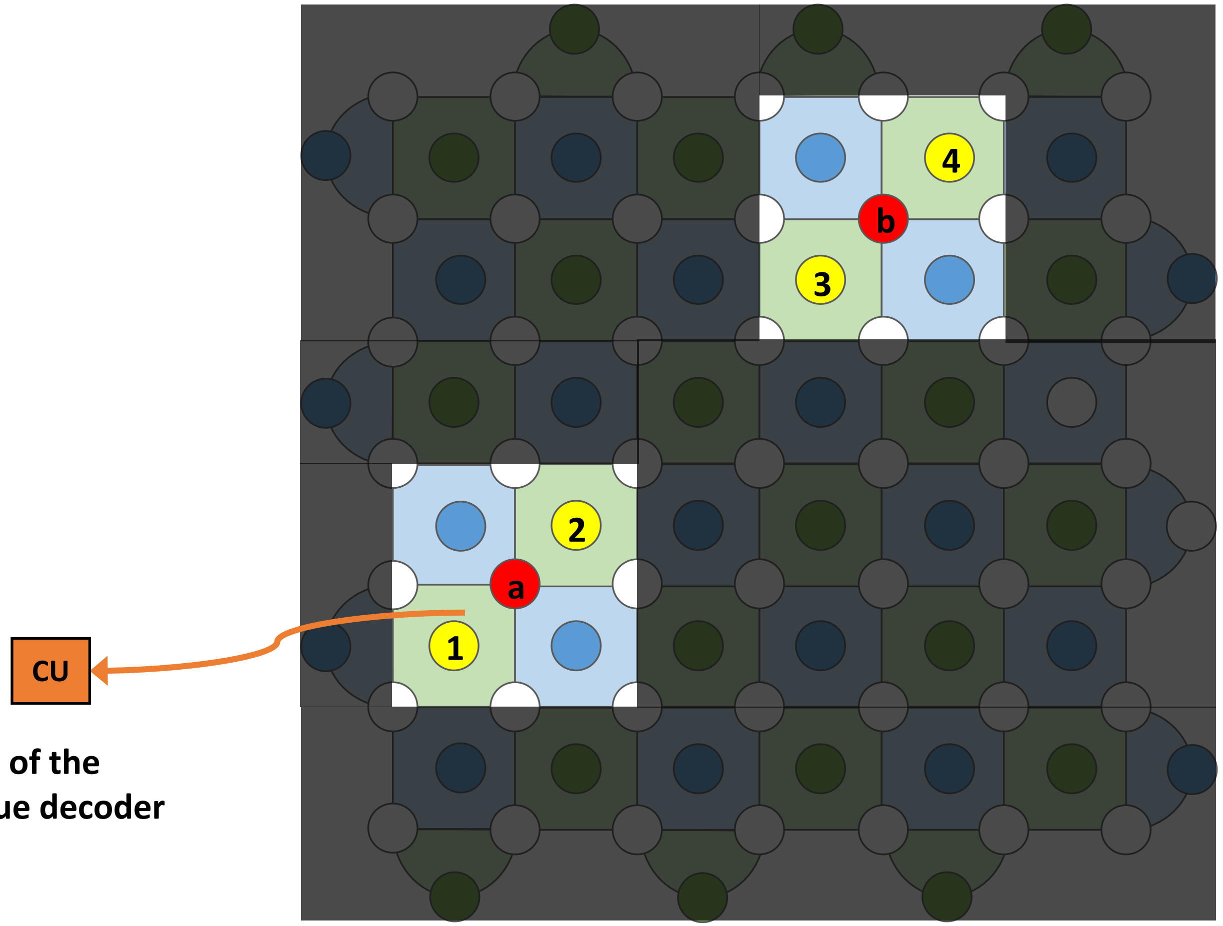


Why are isolated errors trivial to decode?

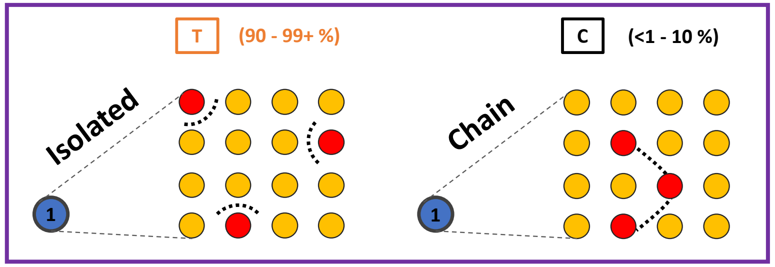




How clique trivially decodes isolated errors

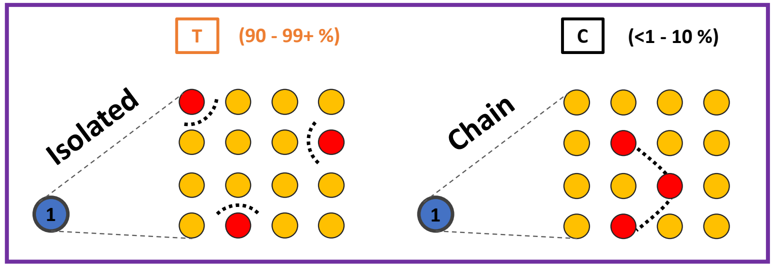


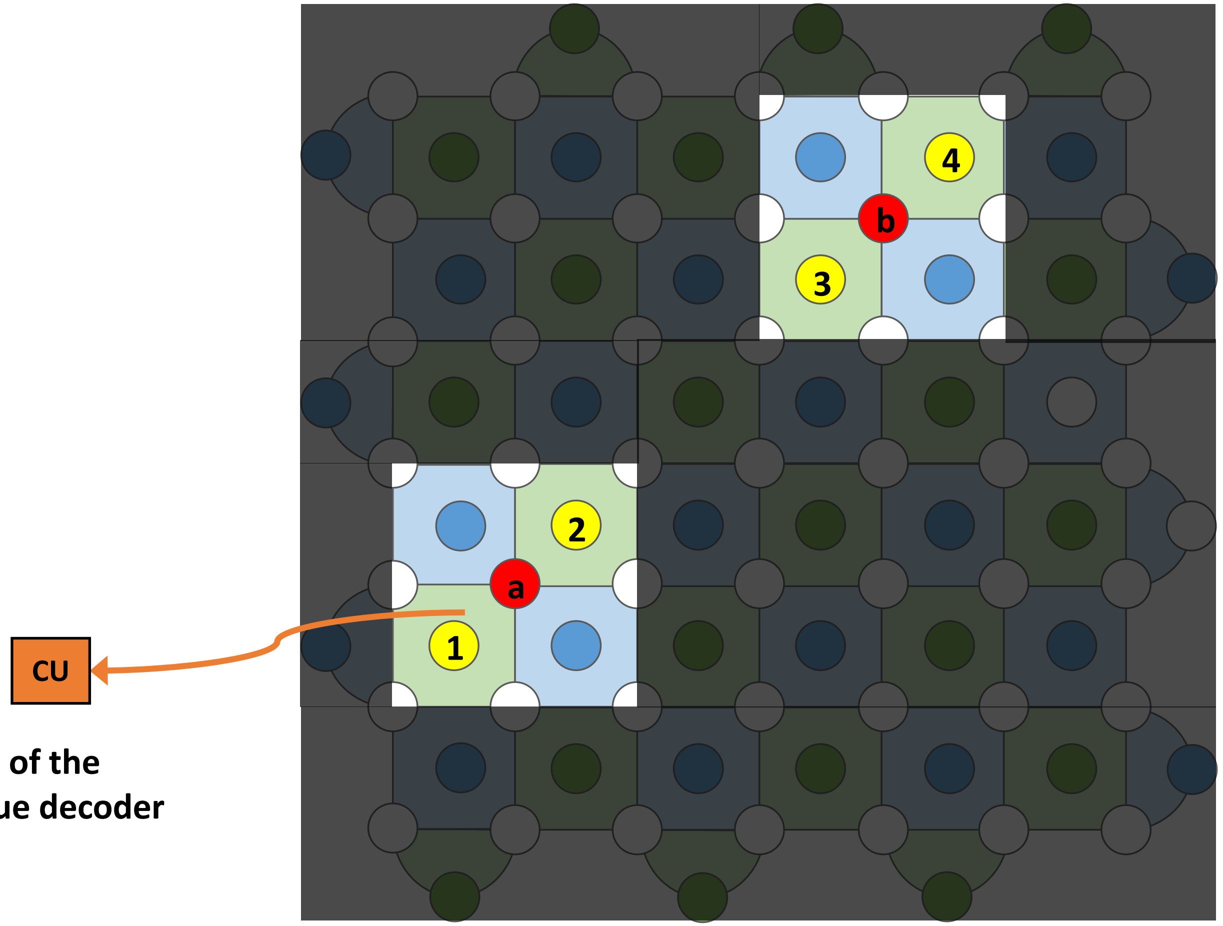
Unit of the
Clique decoder



Isolated data errors only trigger locally paired syndromes which are easy to decode.

How clique trivially decodes isolated errors





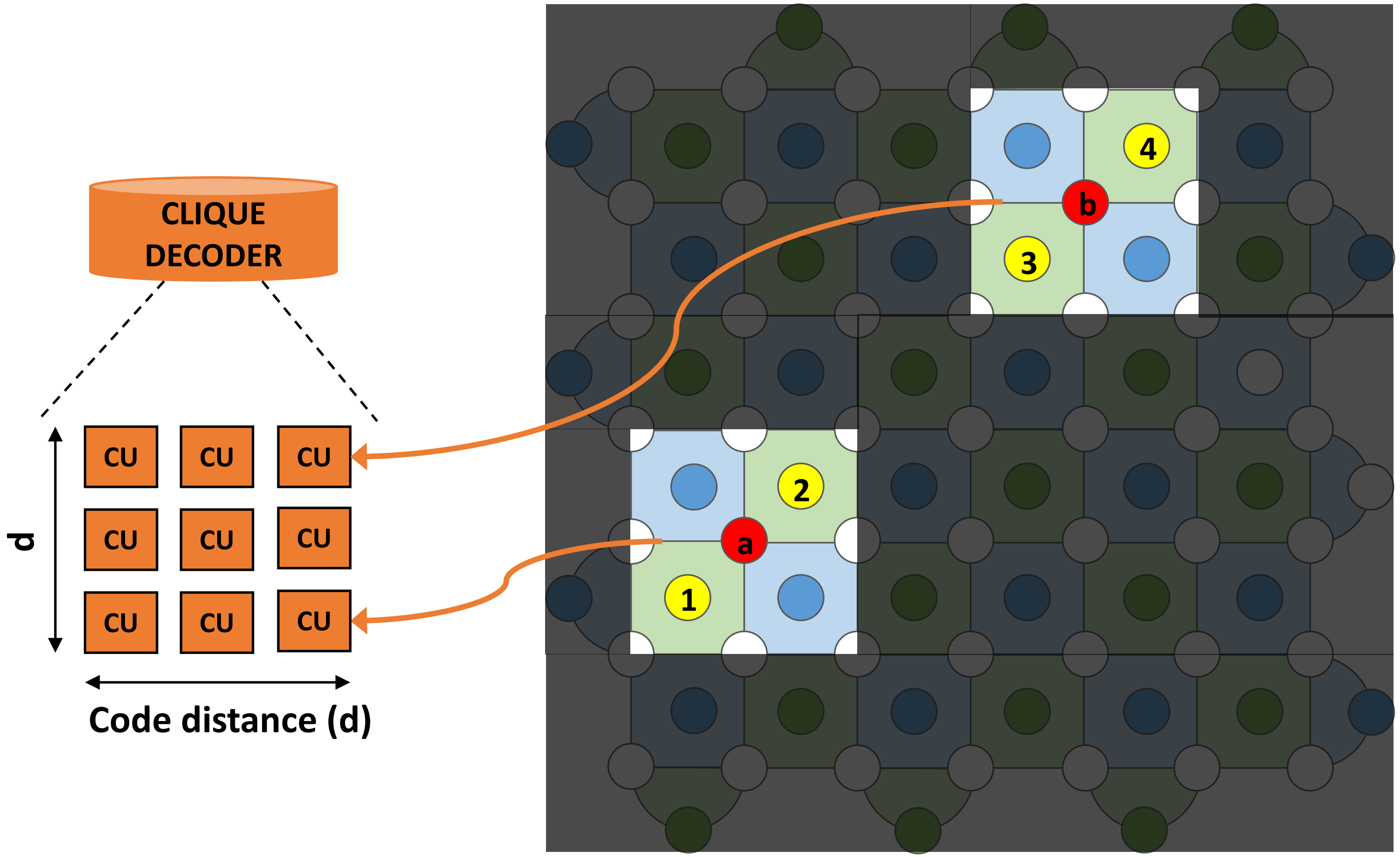
Unit of the
Clique decoder

Additional subtleties to detecting if all syndromes only correspond to isolated data errors!

Additional logic required to handle syndrome measurement errors!

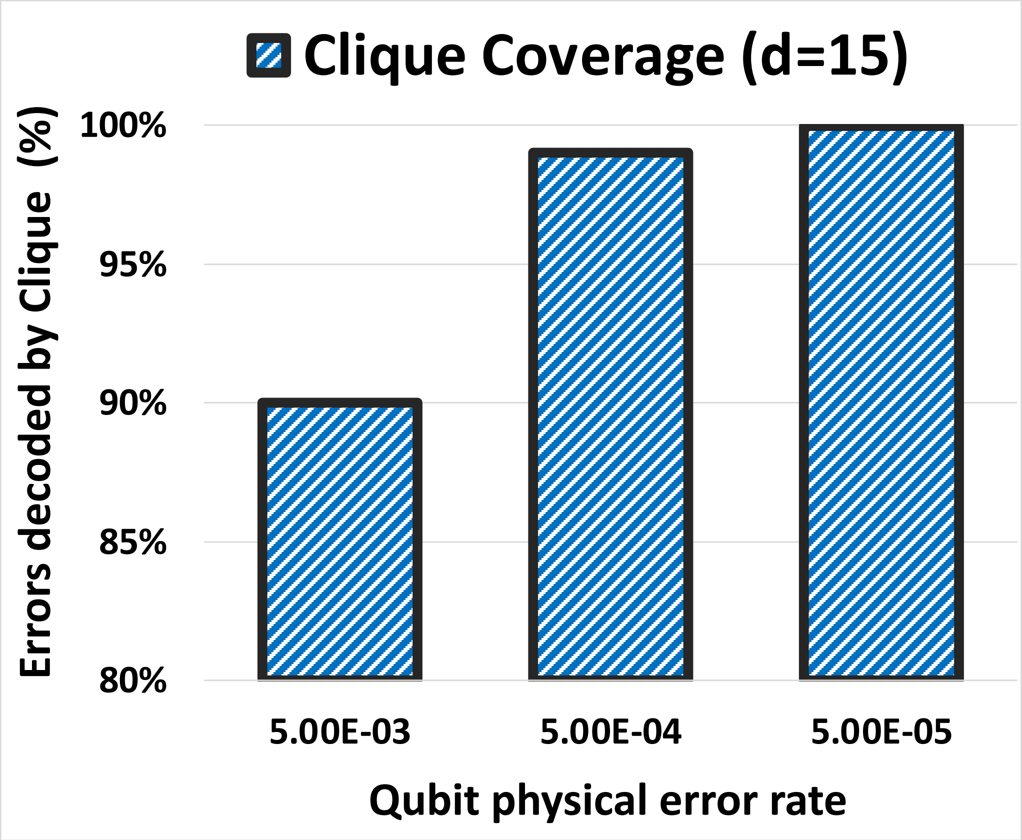
Clique decoder architecture

CU: ~10 combinational gates.
Clique decoder: d^2 CUs.
Linear Clique scaling wrt. physical qubits.



Quantitative benefits: Fridge I/O bandwidth reduction

90 - 100% of decodes handled trivially by Clique, largely eliminating outside-fridge decoding.

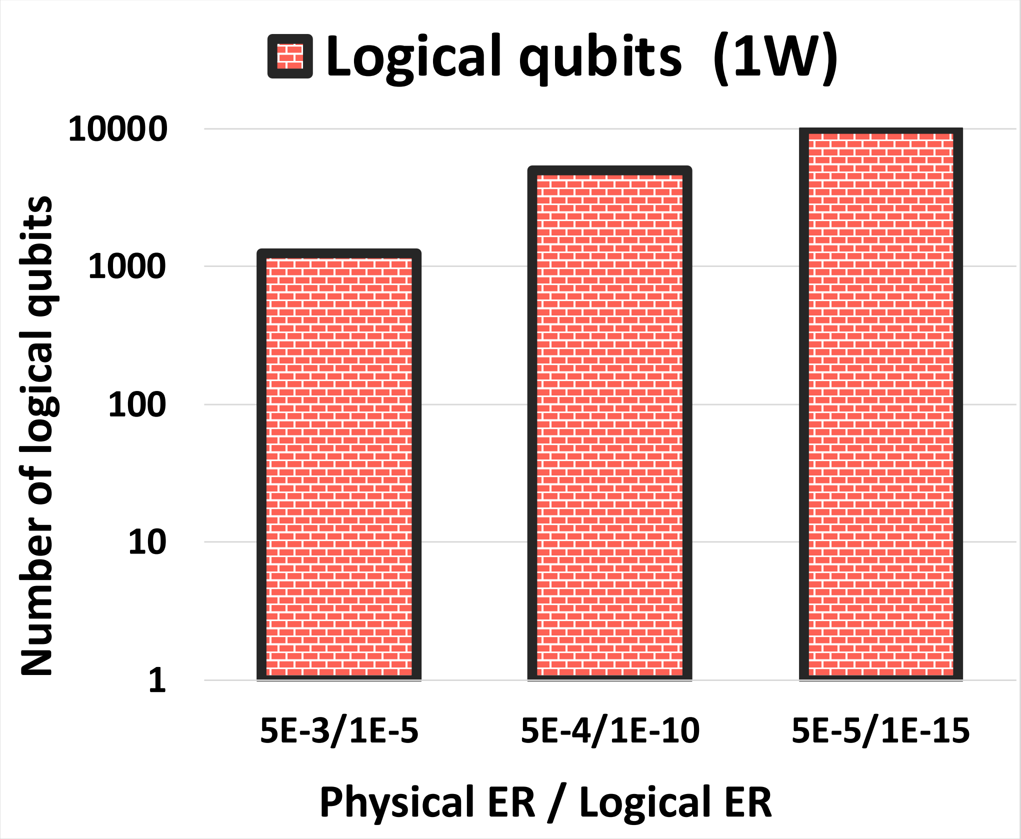


<u>Comparison to AFS compression [Das, HPCA ‘22]:</u> **Clique BW reduction is 10-10,000x greater than AFS which is an entirely off-chip decoding scheme but employs data compression on error I/O data.**

Quantitative benefits: Cryo-resource requirement

Clique supports 2.5M physical qubits at 1W power → 1000s of logical qubits.

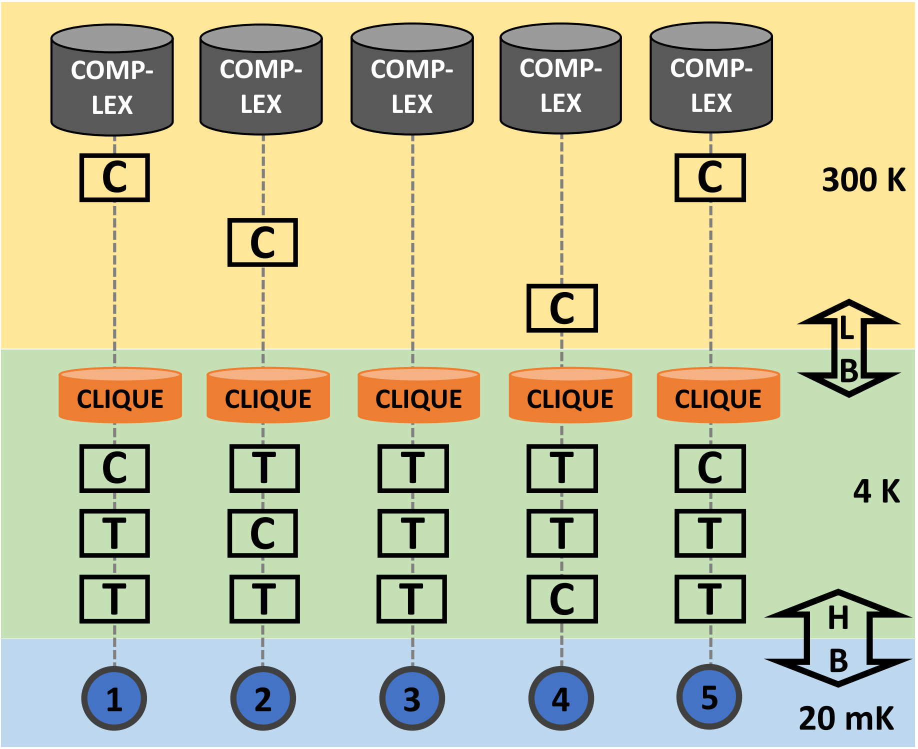
2.5M physical qubits!



Comparison to NISQ+ [Holmes, ISCA '20]:

At d=9, Clique requires 25-80x lower on-chip resources compared to NISQ+, an approximate fully cryogenic decoder. Greater benefits at higher code distances.

Key Takeaways



1. QEC decoding suffers severe bottlenecks: bandwidth, area, power, thermal.
2. BTWC approach: common trivial errors can be handled separately from rare complex errors
3. Clique: A lightweight cryogenic decoder for accurately decoding and correcting common-case trivial errors.
4. High fridge I/O bandwidth reduction and low cryo-resource requirement (2-4 orders of magnitude benefits over SOTA).

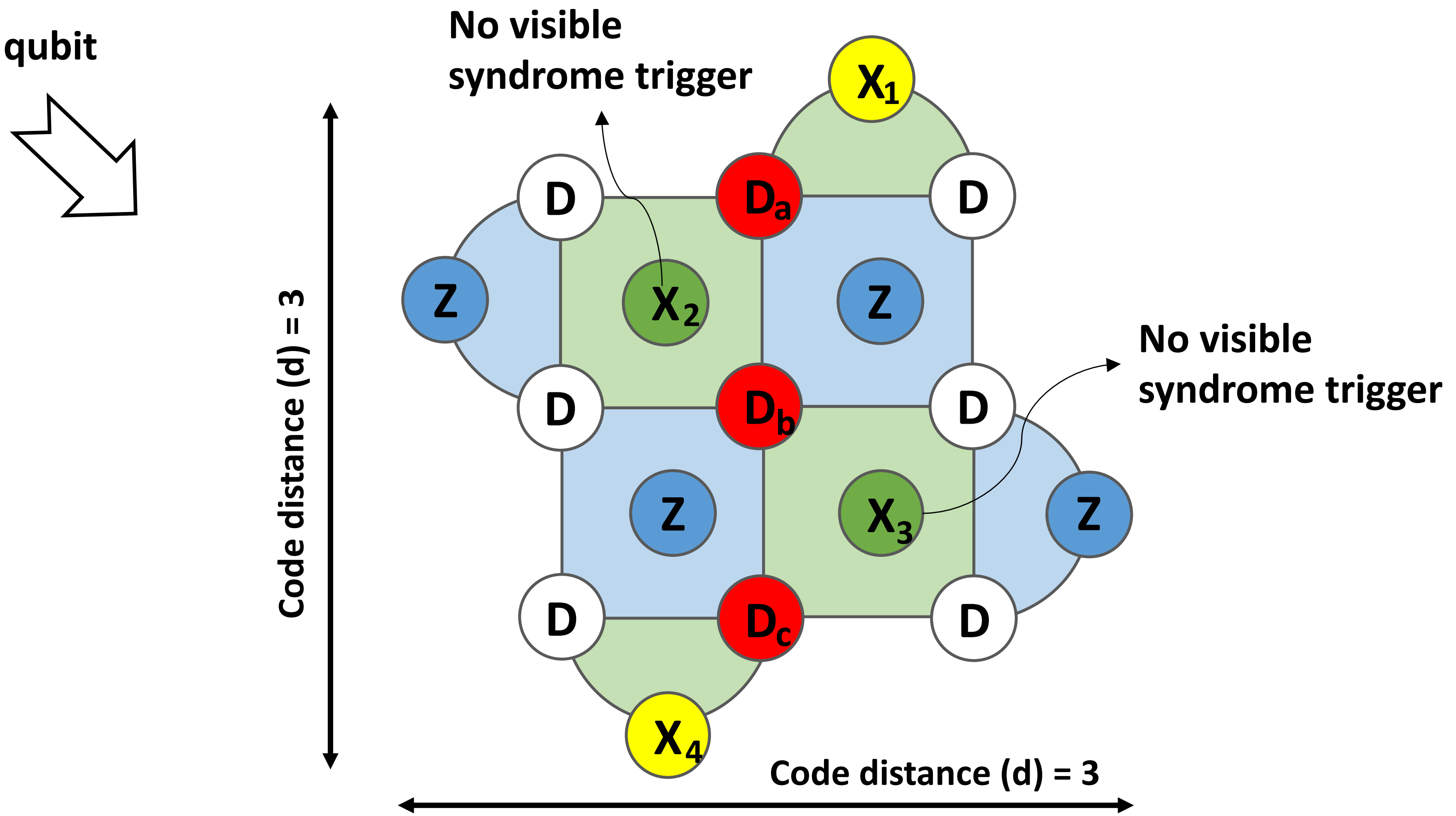
Thank you!

gravi@uchicago.edu

CCF-1730082/1730449, NSF Phy-1818914, NSF 2110860 , DOE (Accelerated Research for Quantum Computing Program), NSF OMA-2016136, Q-NEXT DOE NQI Center, CIF (NSF 2030859) , IBM/CQE Postdoc Fellowship, NSF (grants CCF-2119184, CNS-1956180, CNS-1952050, CCF-1823032, CNS-1764039), ARO (grant W911NF1920321), DOE Early Career Award (grant DESC0014195 0003), NSF Grant DGE-1842474.

How does QEC work? (Surface code)

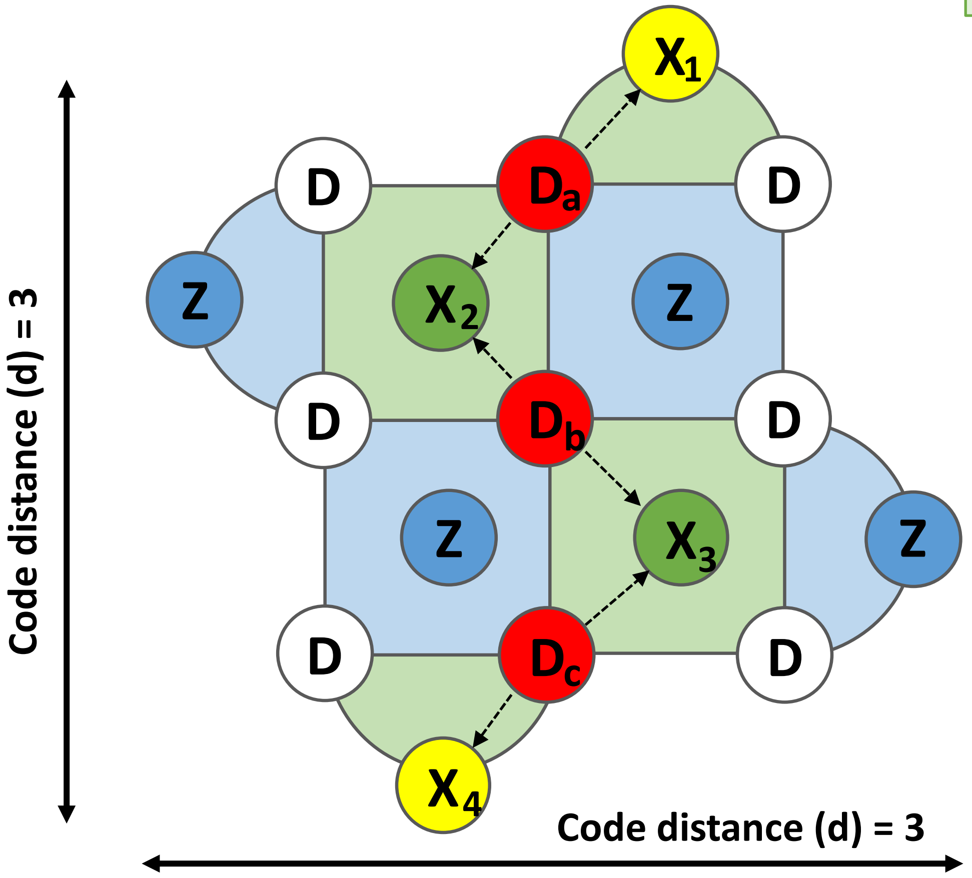
1 logical qubit



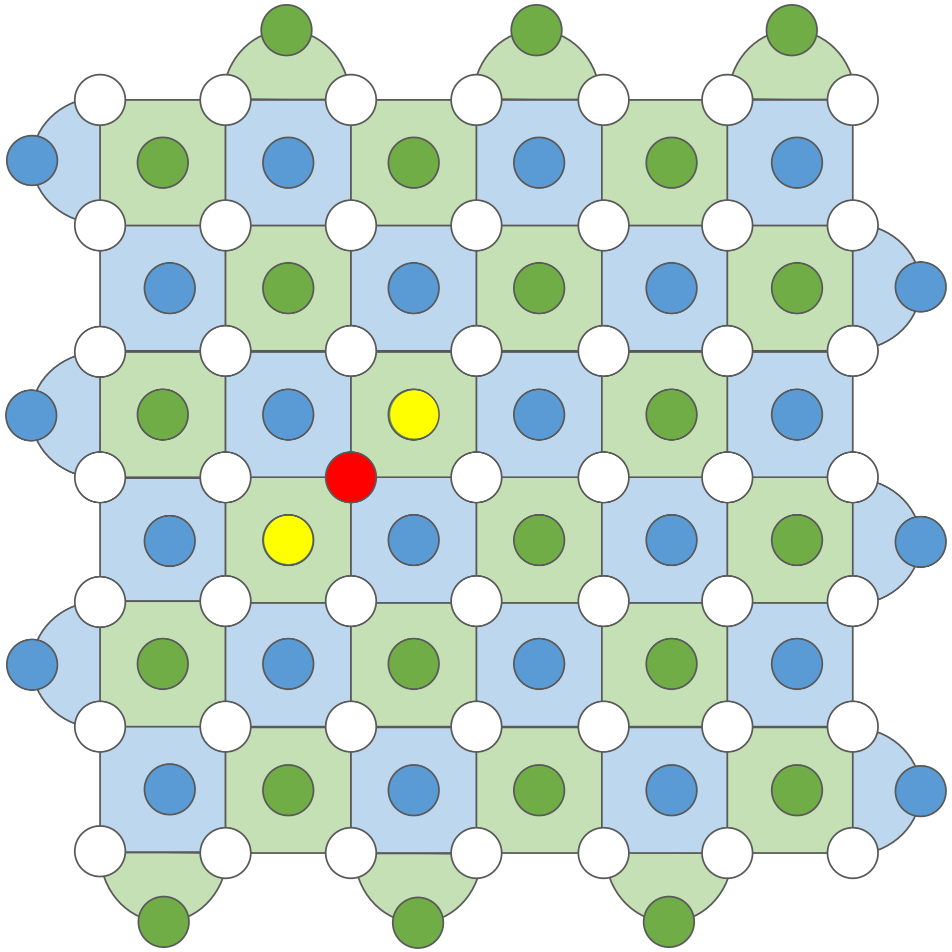
How does QEC work? (Surface code)

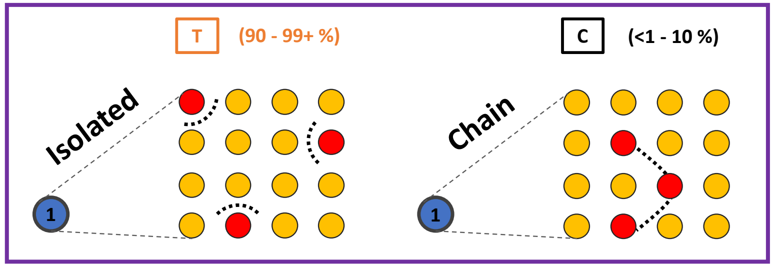
Error chains generate less syndrome information.

1 logical qubit



Why are isolated errors much more common than error chains?



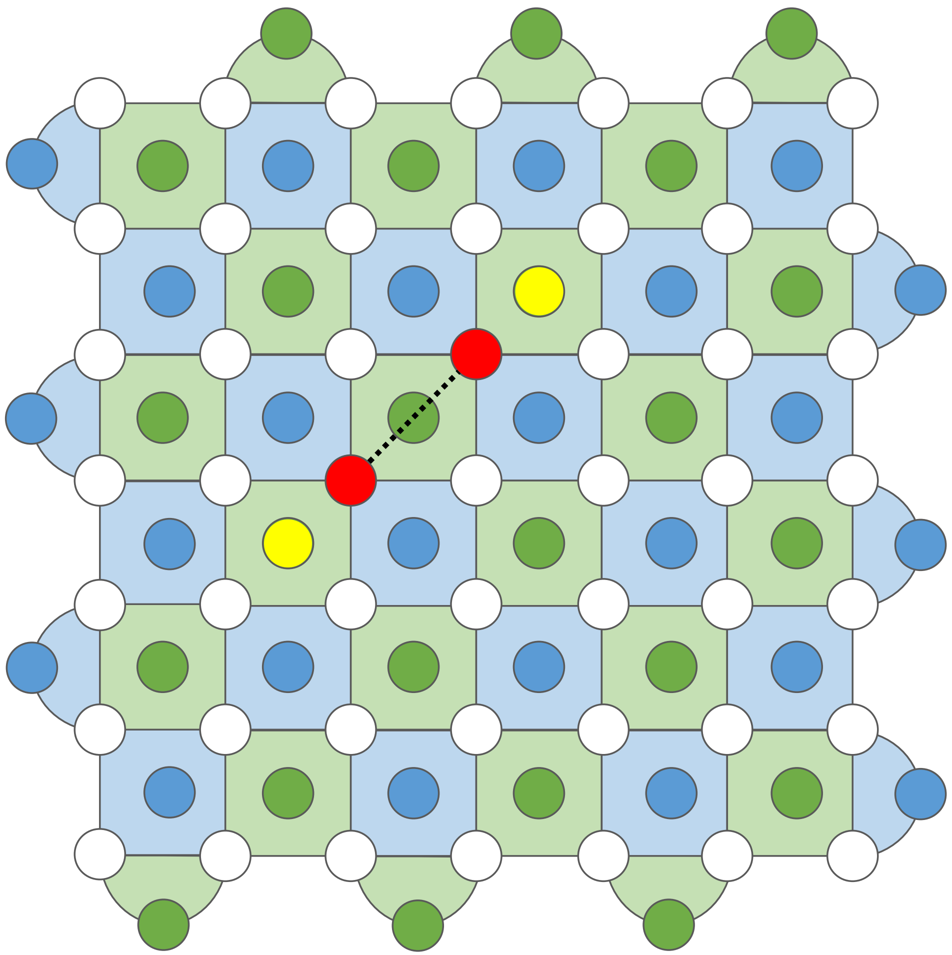


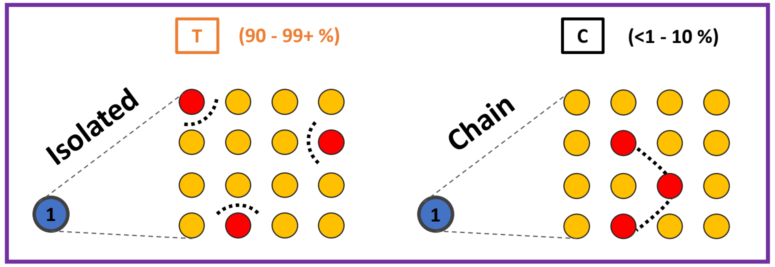
1 logical qubit encoded in
49 physical data qubits (d=7)

PER= 10^{-3} (1 in 1000), N = 49

P (1 error in block) = N * PER = 4.9%

Why are isolated errors much more common than error chains?





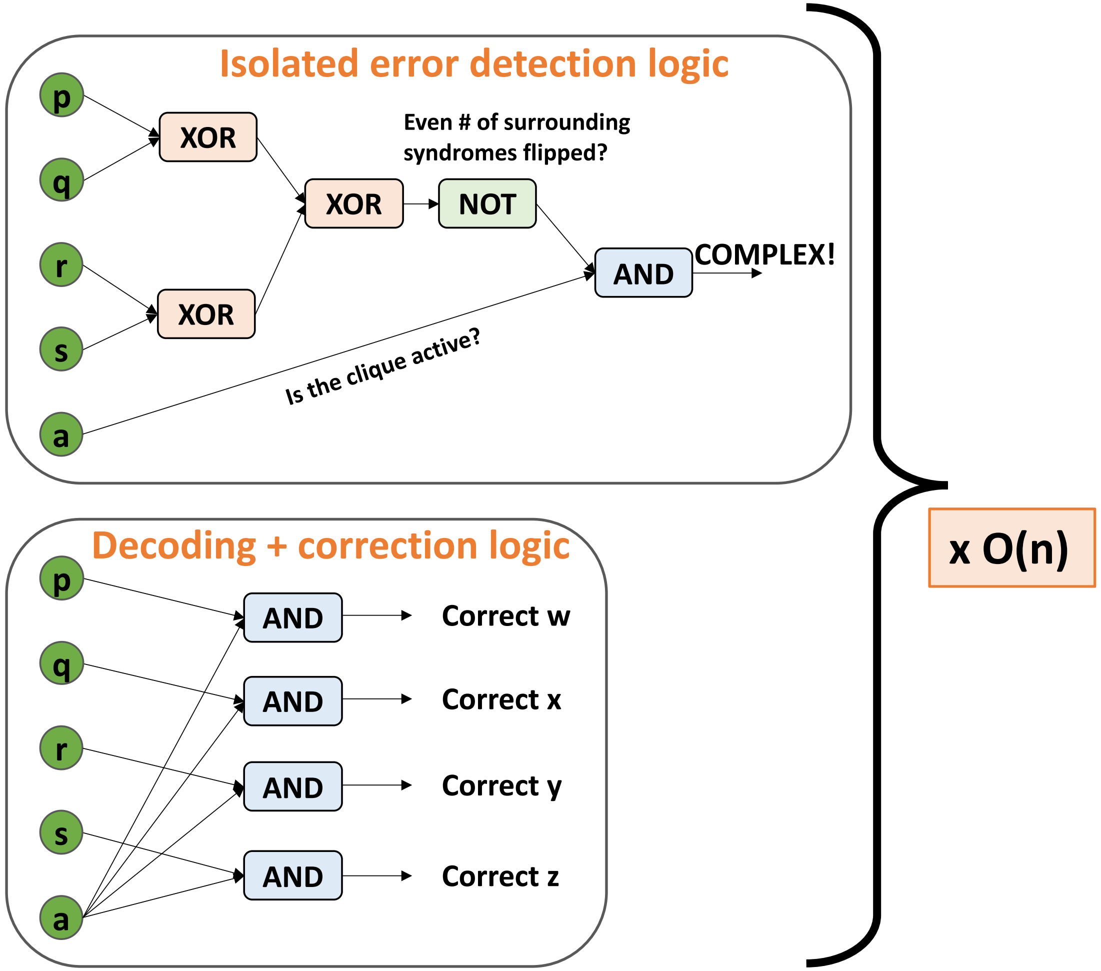
1 logical qubit encoded in
49 physical data qubits (d=7)

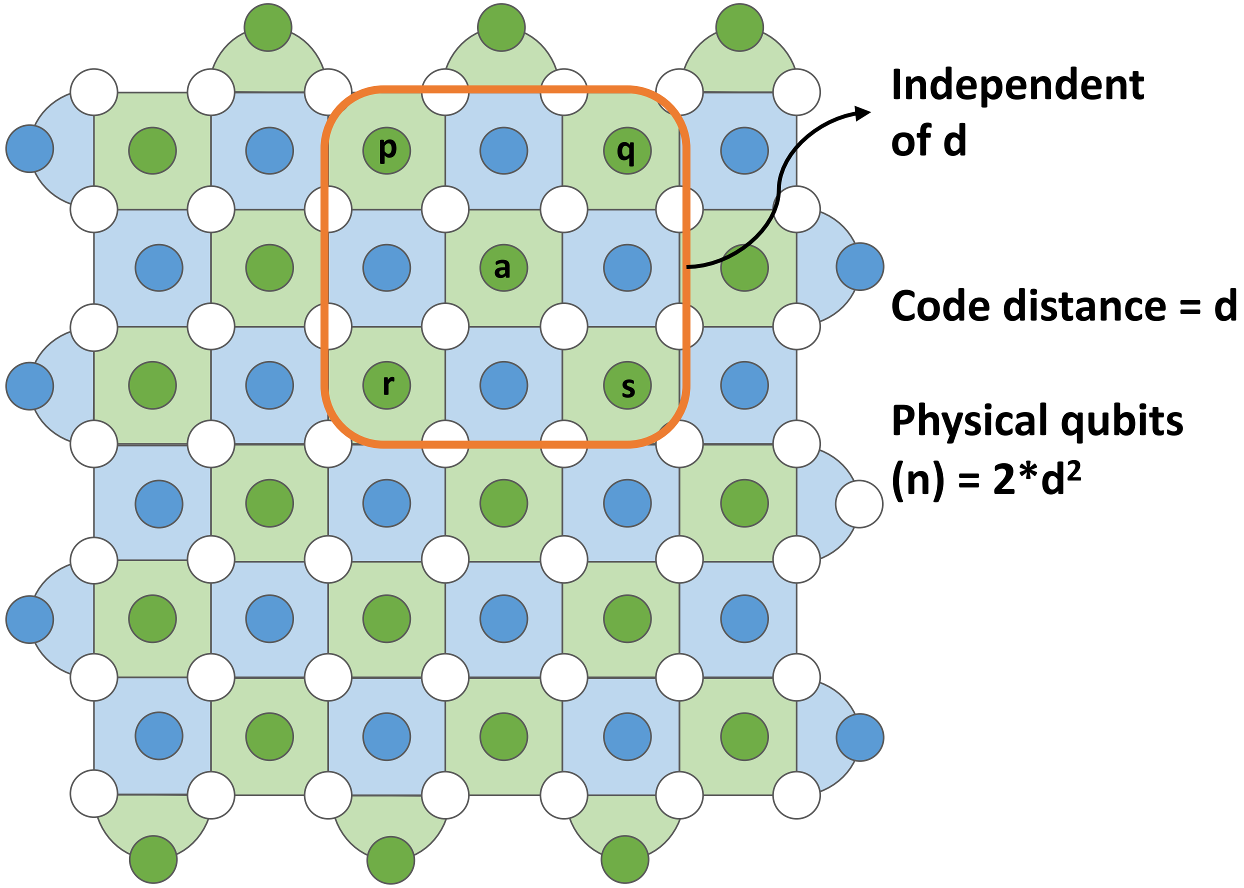
PER= 10^{-3} (1 in 1000), N = 49

P (2 adjacent errors anywhere in block)
= 6 * N * PER^2 = 0.03%
(160x less likely than the isolated case)

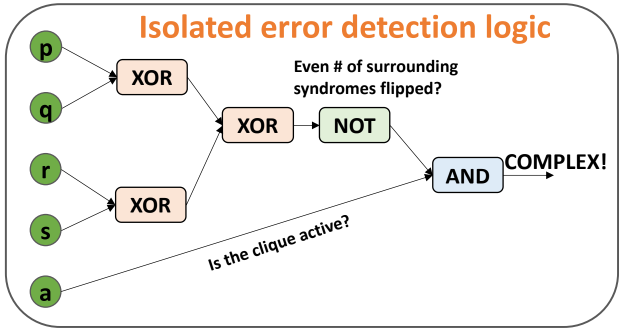
Clique decoder hardware design

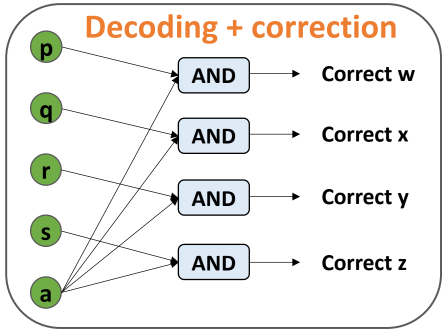
Lightweight hardware suited to cryo-domain: < 10 combinational logic gates per clique unit.
Total Clique decoder cost scales linearly in the number of qubits.

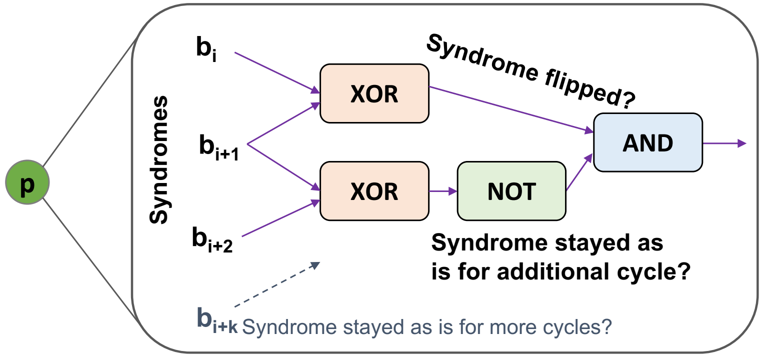




Clique decoder hardware design

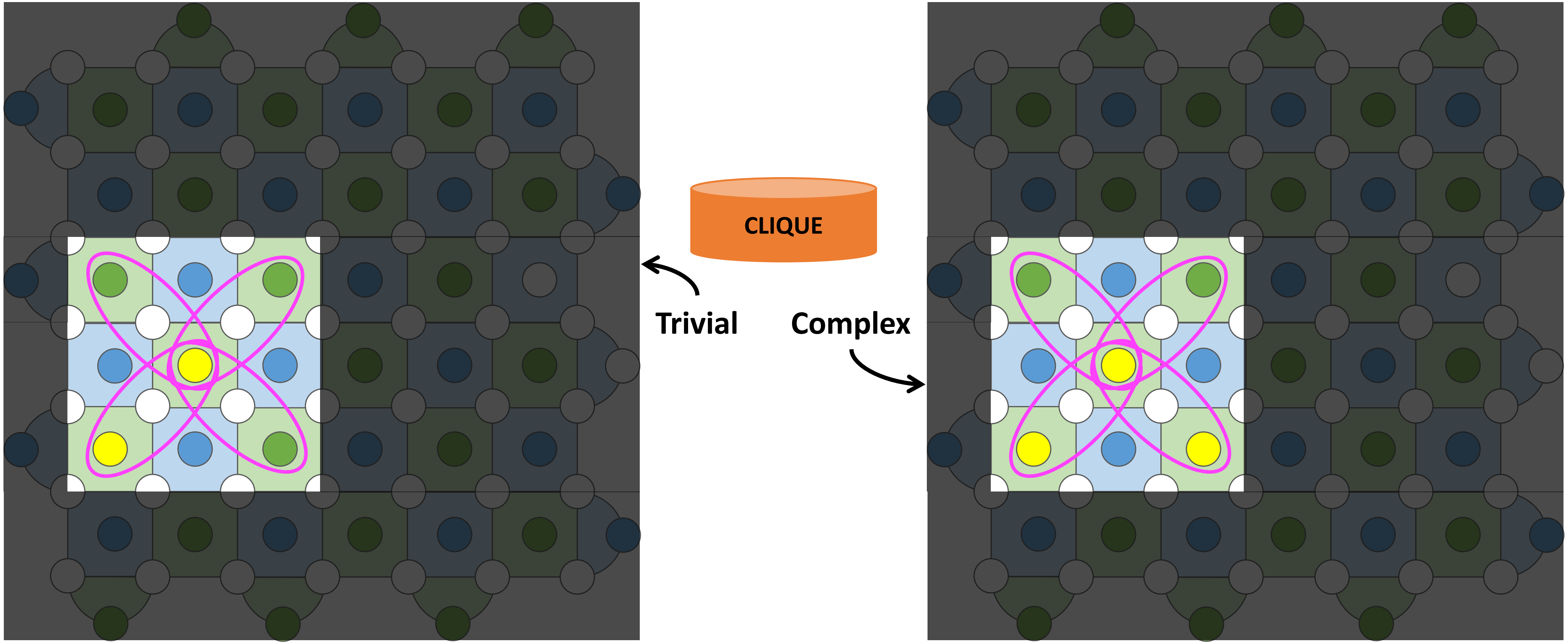






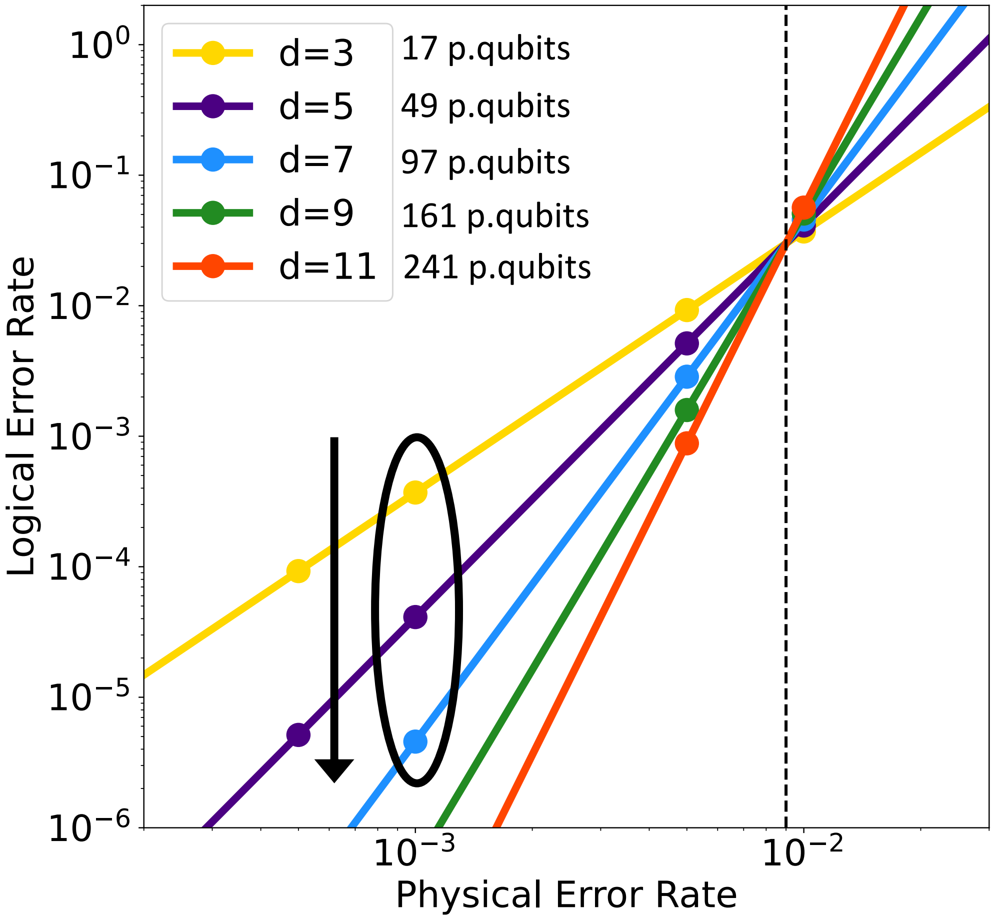
How to trivially detect isolated errors?

***Isolated error litmus test:* If the center of a clique is set, and if an odd number of neighbor syndromes are set, the clique can be trivially decoded.**

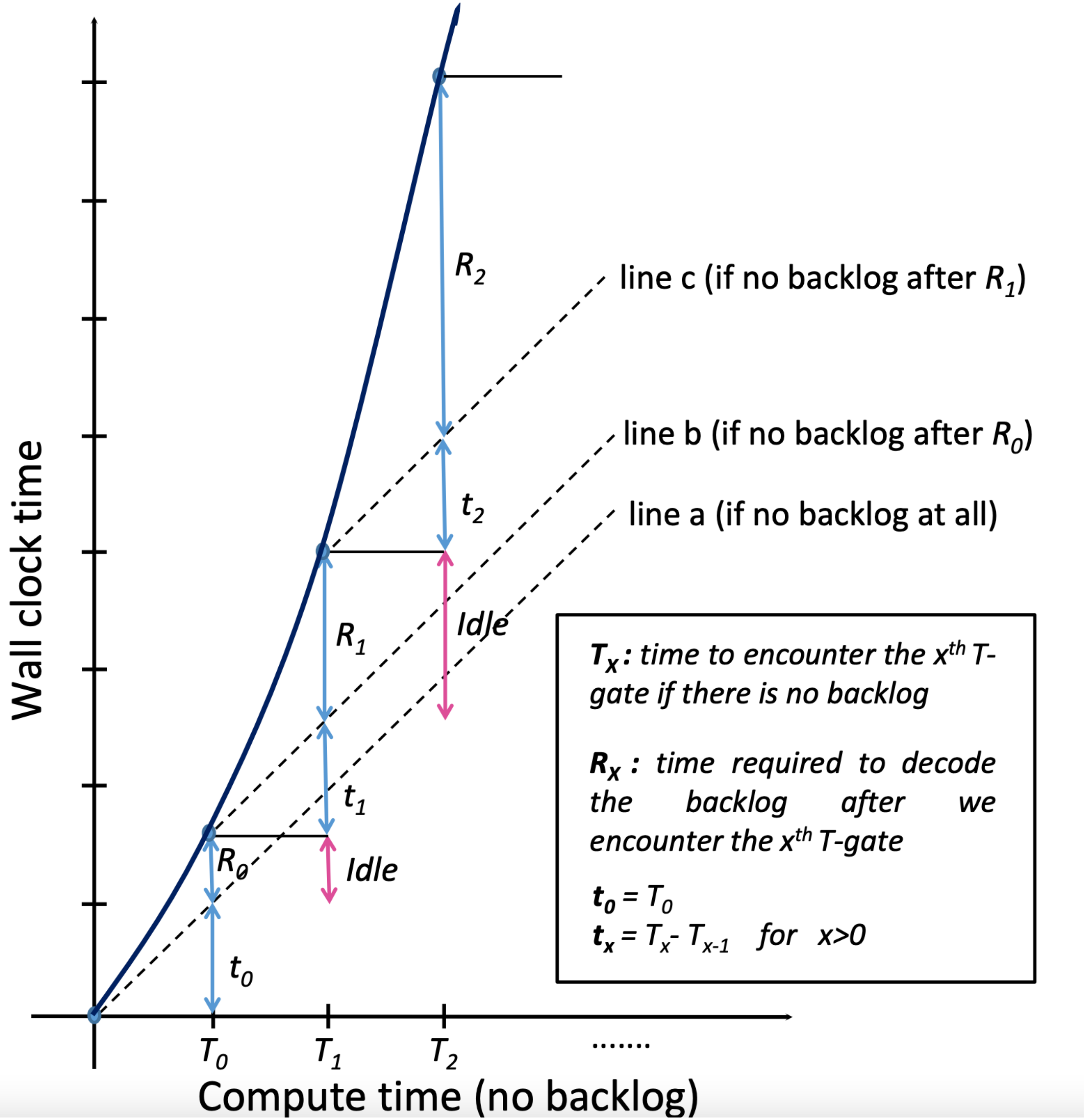


How QEC works: Surface code

Better logical qubits possible with higher code distance, but with increased overheads.

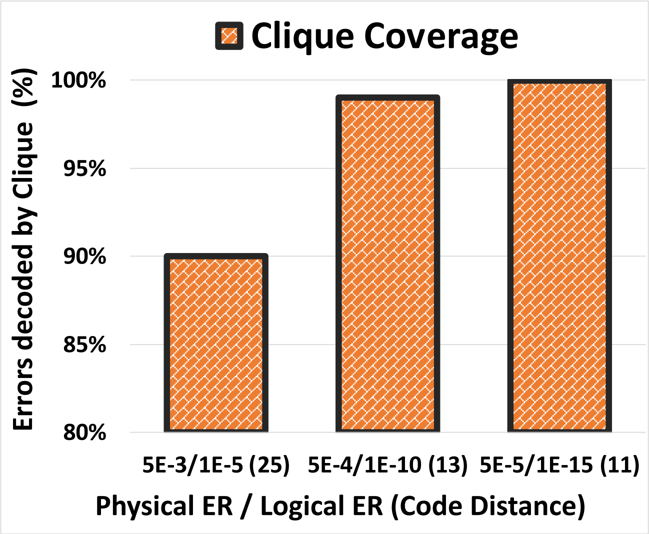


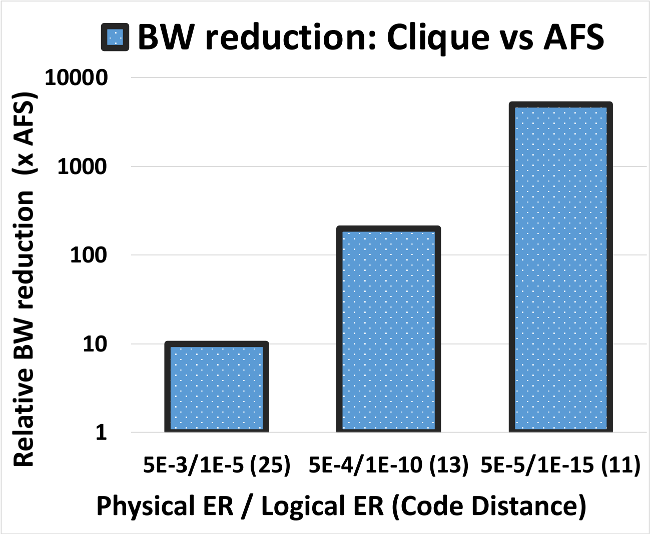
Logical error decreases with code distance

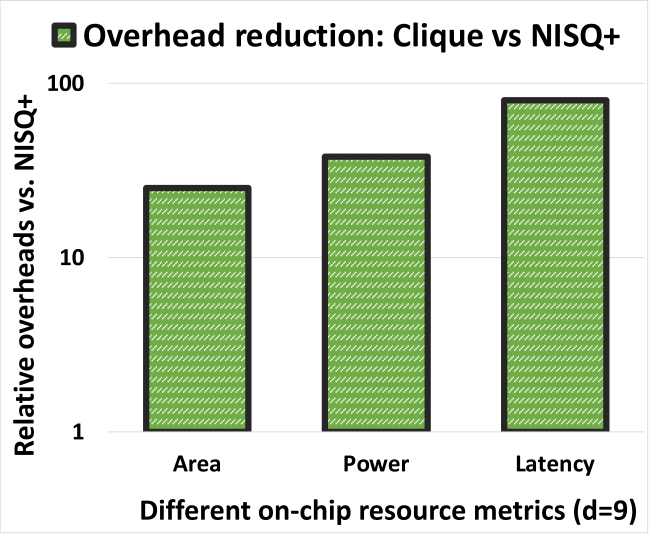


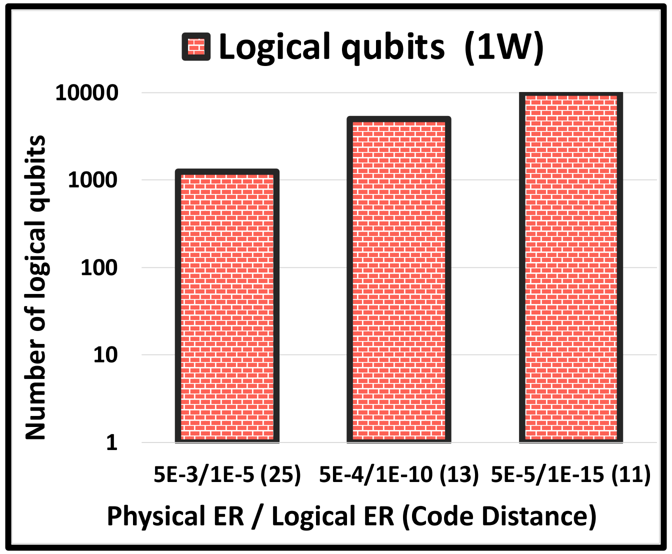
Wall clock time
Compute time (no backlog)
T_0
T_1
T_2
....
t_0
t_1
t_2
R_0
R_1
R_2
Idle
Idle
line c (if no backlog after R_1)
line b (if no backlog after R_0)
line a (if no backlog at all)
T_x : *time to encounter the x^{th} T-gate if there is no backlog*
R_x : *time required to decode the backlog after we encounter the x^{th} T-gate*
$t_0 = T_0$
$t_x = T_x - T_{x-1}$ *for* $x>0$

What are the quantitative benefits?

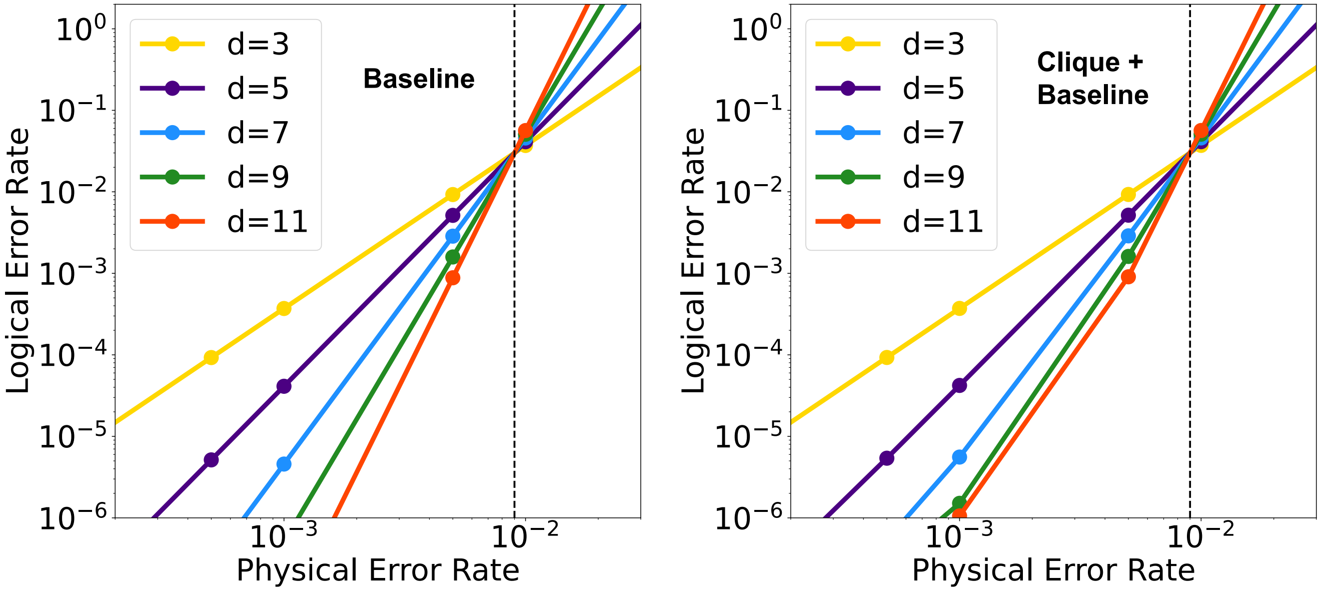






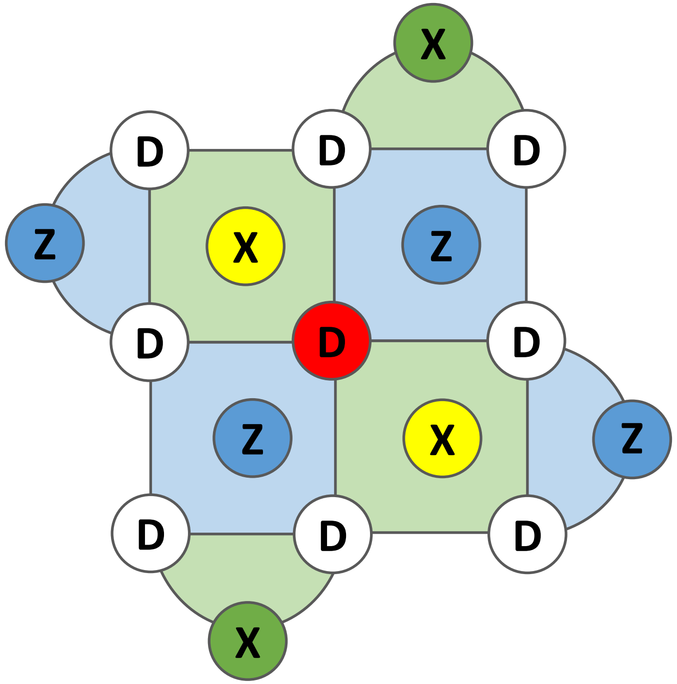


Logical error rate

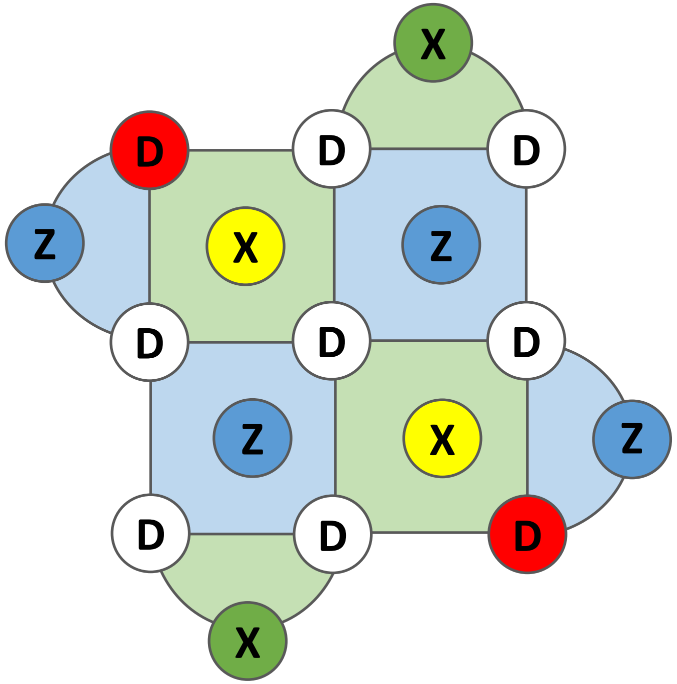


How does QEC decoding work?

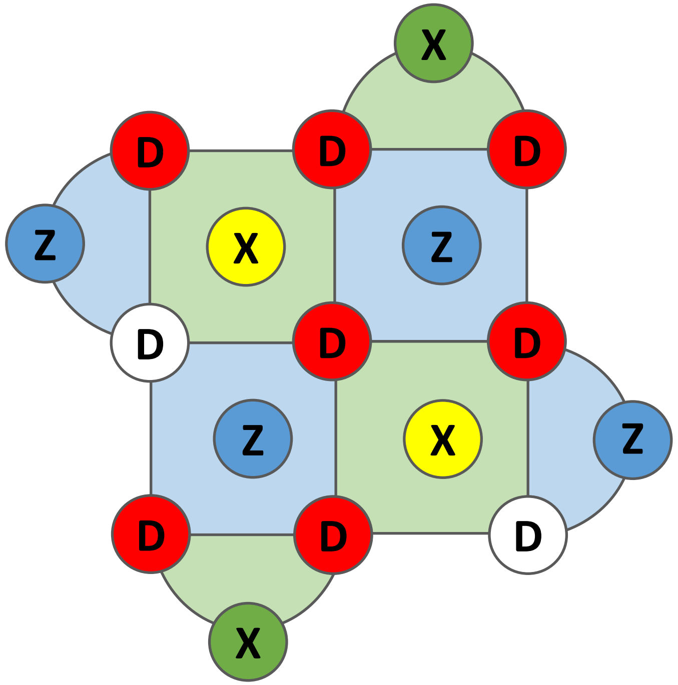
Likelihood of 1 data qubit error: P = 10^{-2}



Likelihood of 2 errors: $P^2 = 10^{-4}$



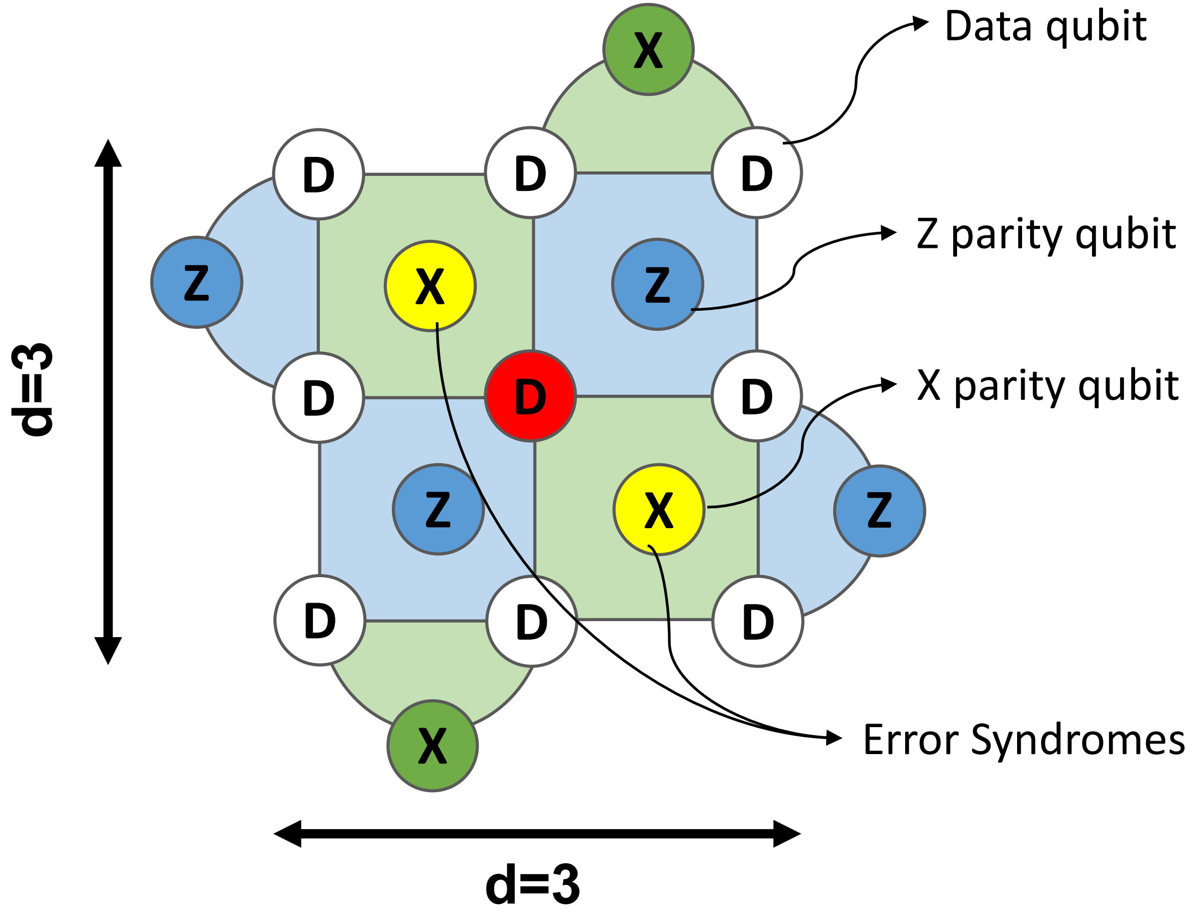
Likelihood of 7 errors: $P^7 = 10^{-14}$



The decoder returns a solution that is most likely: a decoding that produces the lowest number of data errors that satisfies the error syndrome pattern.

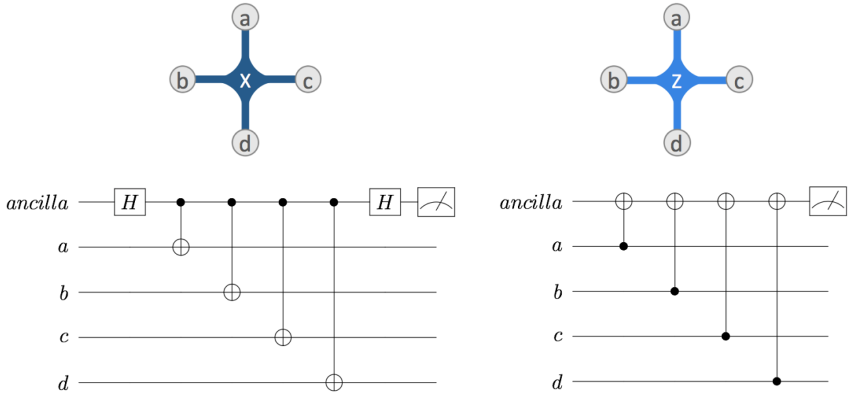
Background: Surface Codes

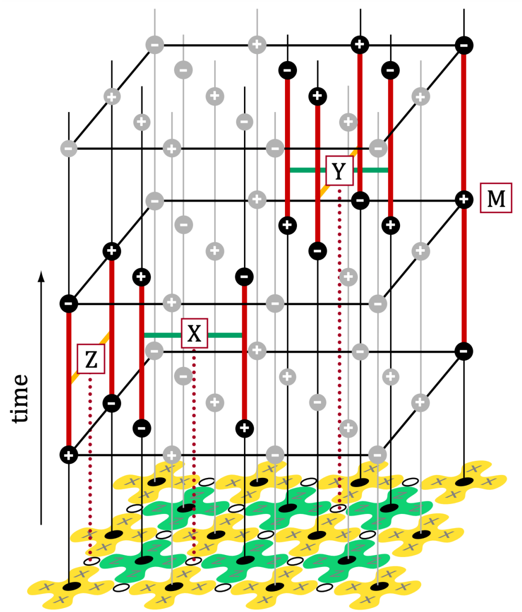
1 logical qubit w/ rotated surface code



Support for high physical error rates

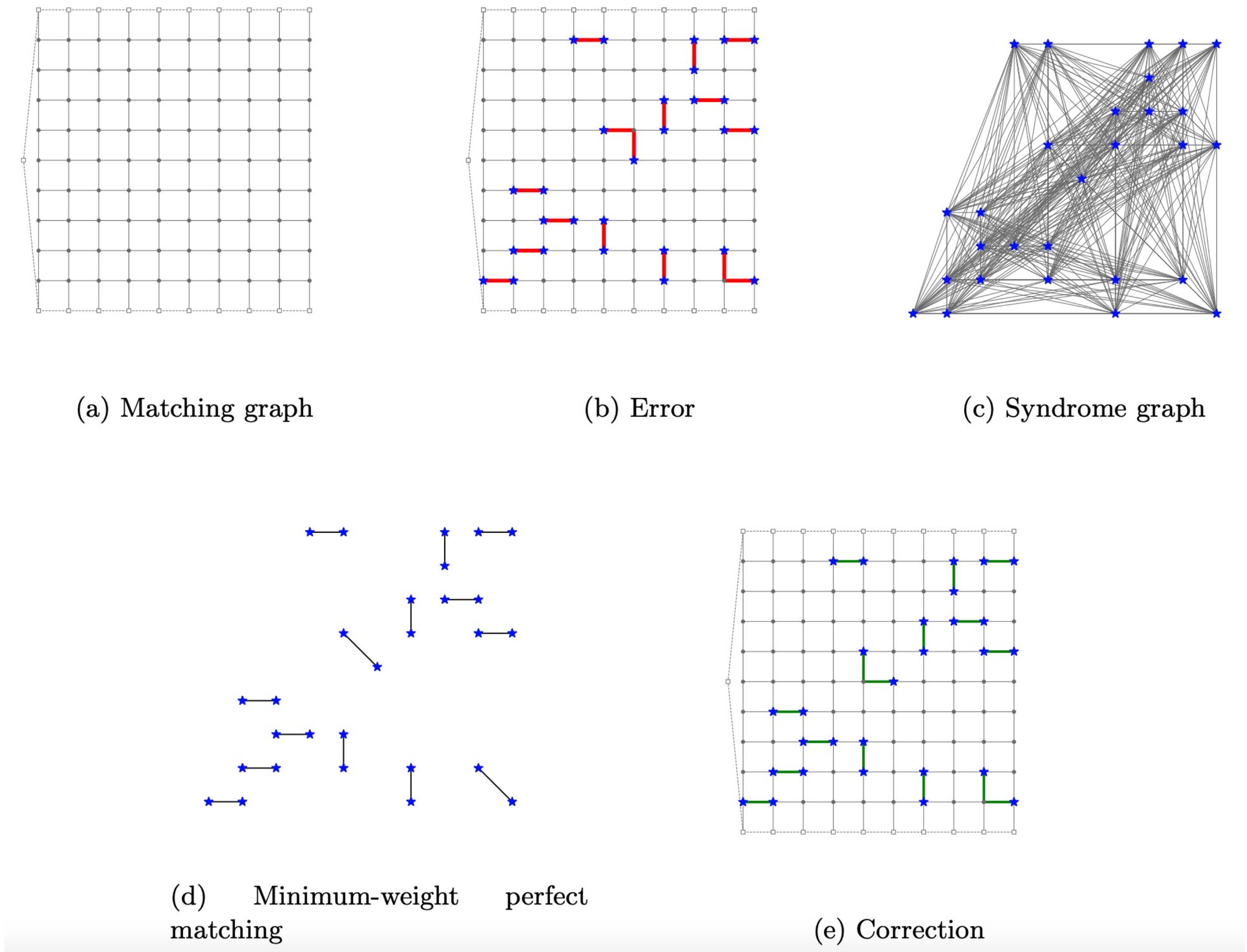
NISQ+ [Holmes2020]





Surface Codes… [Fowler2012]

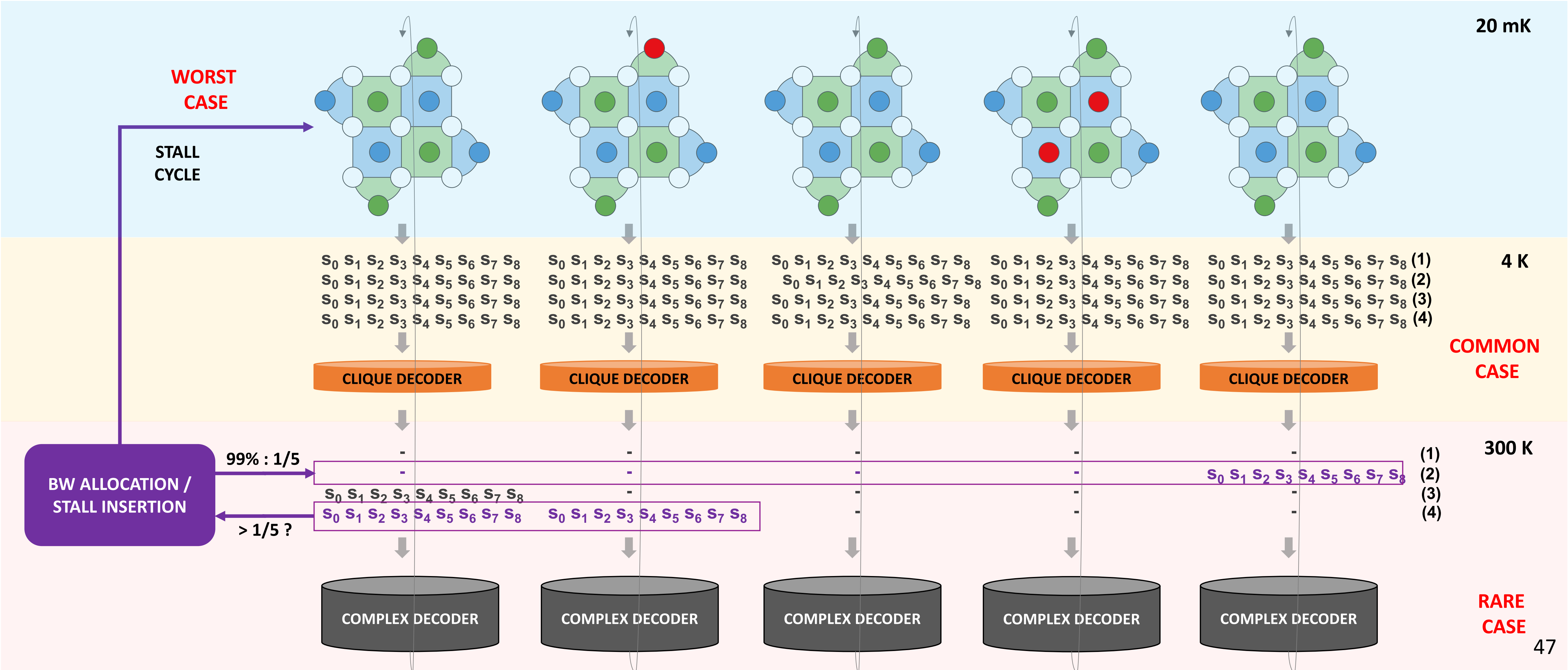
Decoding is complex at large code distances



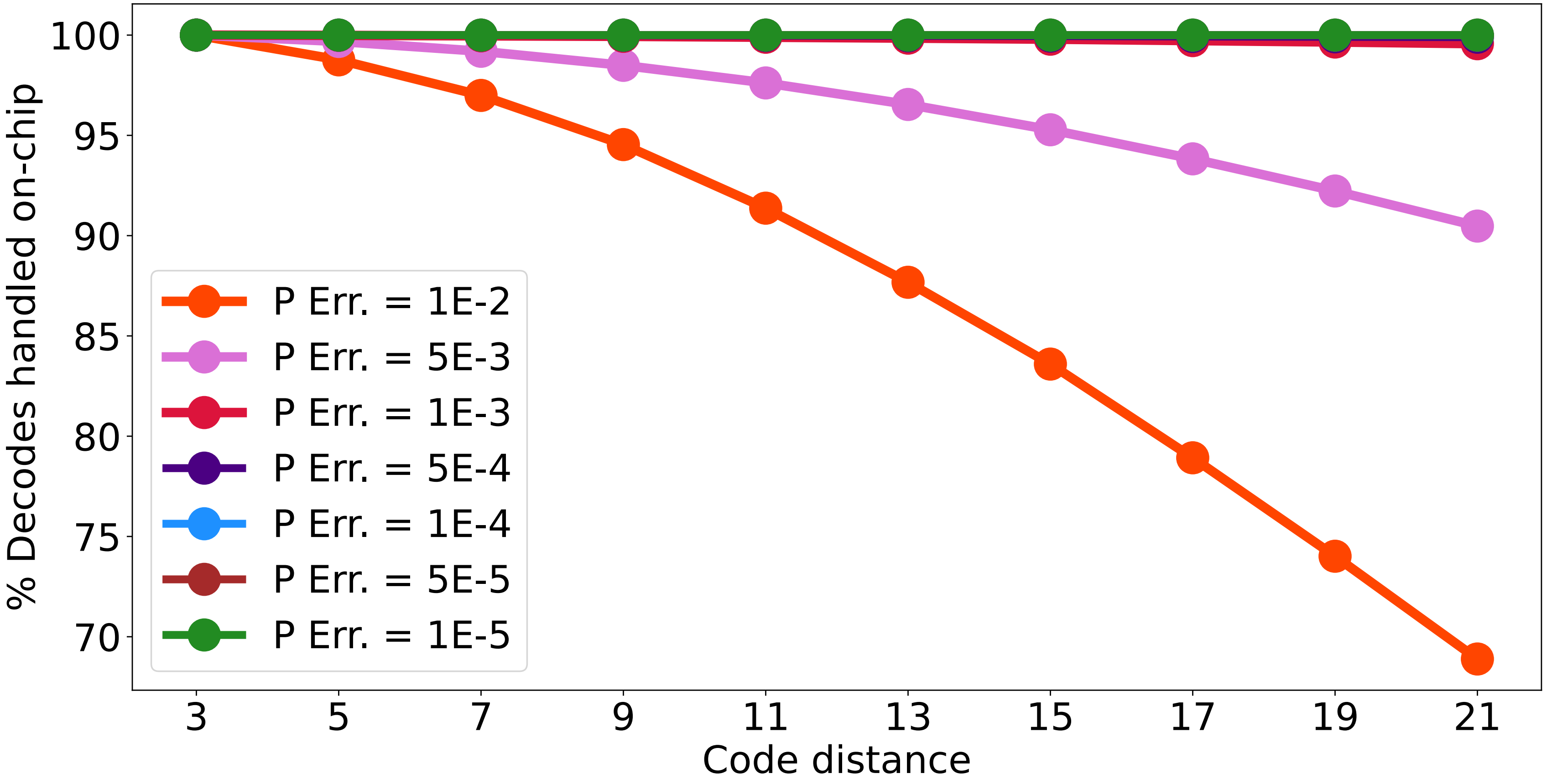
1) Increases with code distance.
2) Multiplied by number of logical qubits.

pyMatching [Higgott2021]

Proposal: Better than worst case decoding for QEC



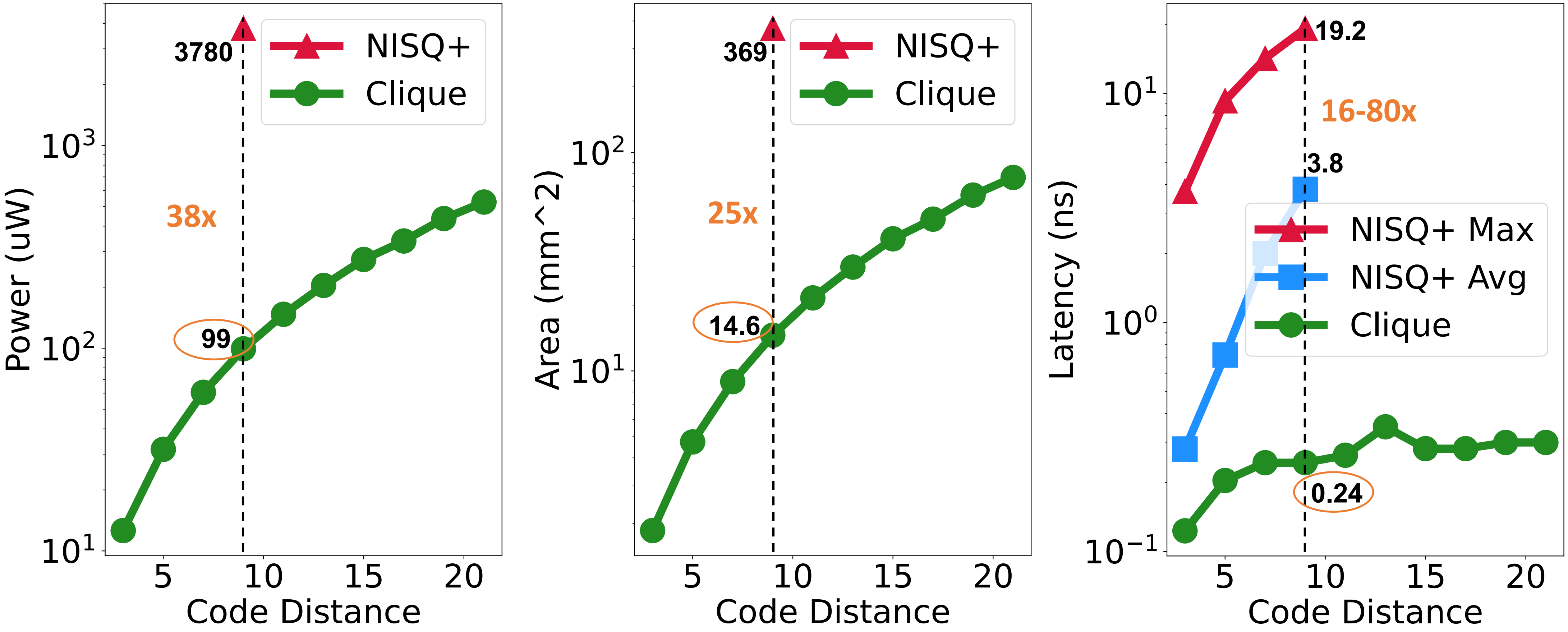
Results: Clique Decoder Coverage



10-1000x greater bandwidth reduction compared to AFS which is entirely off-chip decoding but employs data compression on the syndrome data that is be sent off chip

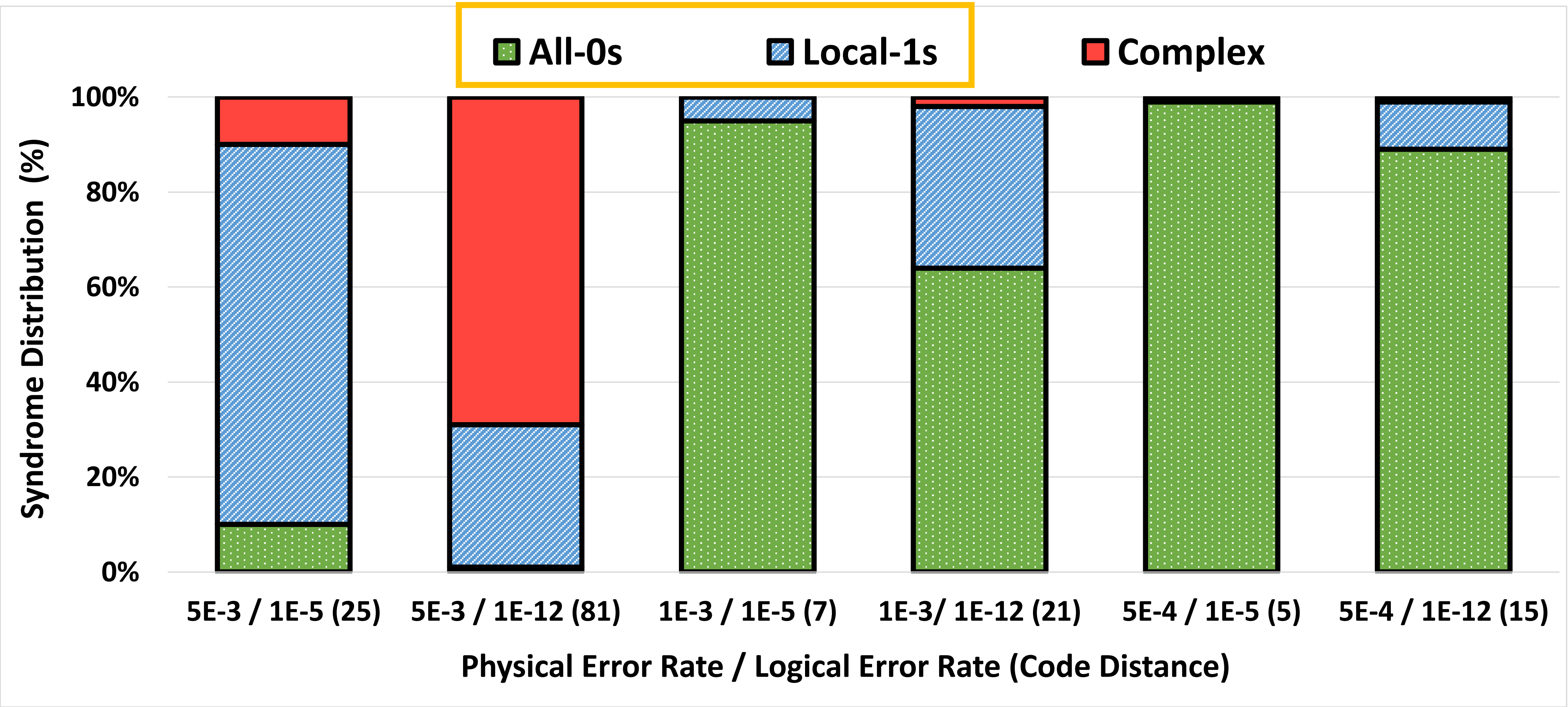
AFS: Accurate, Fast, and Scalable Error-Decoding for Fault Tolerant Quantum Computers

Results: Overheads compared to NISQ+

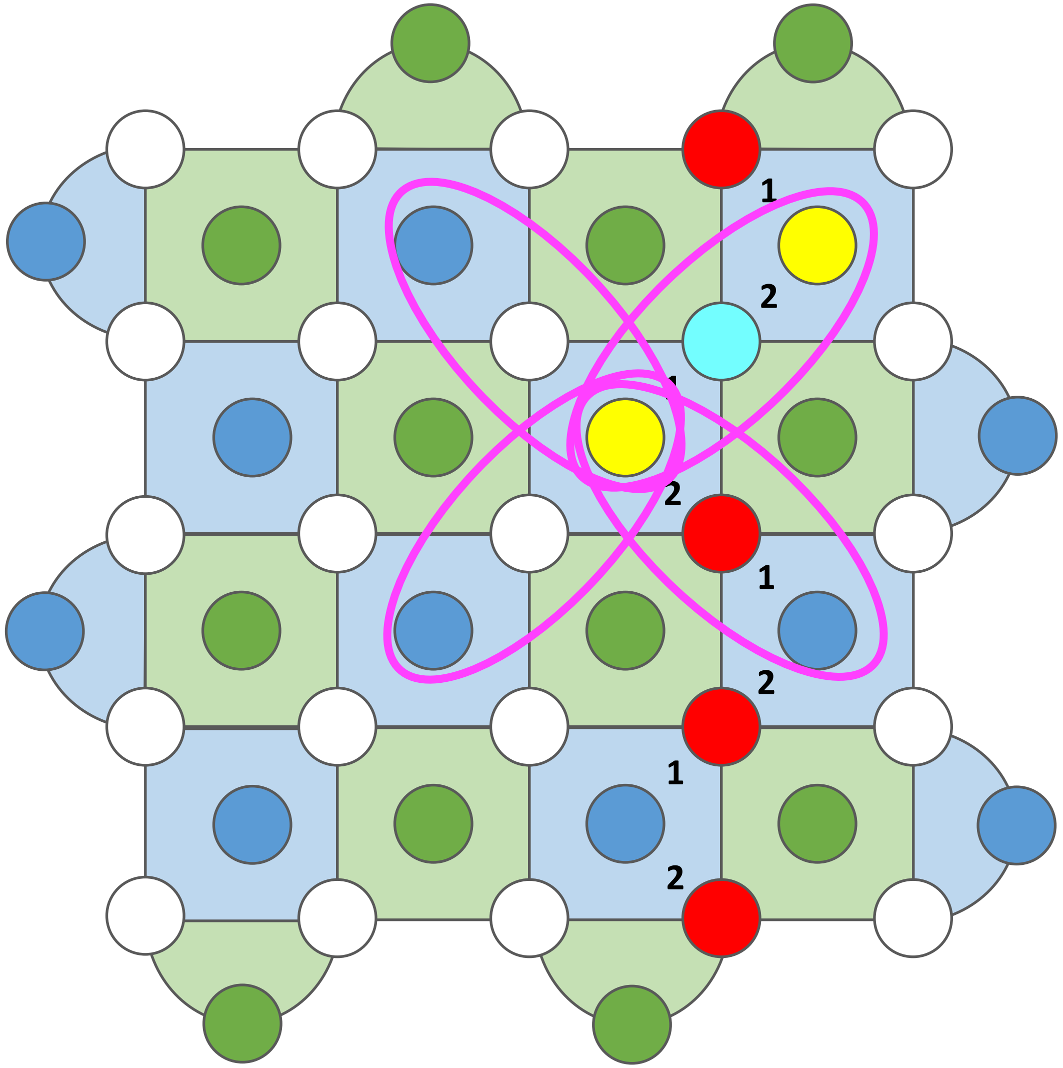


NISQ+: Boosting quantum computing power by approximating quantum error correction

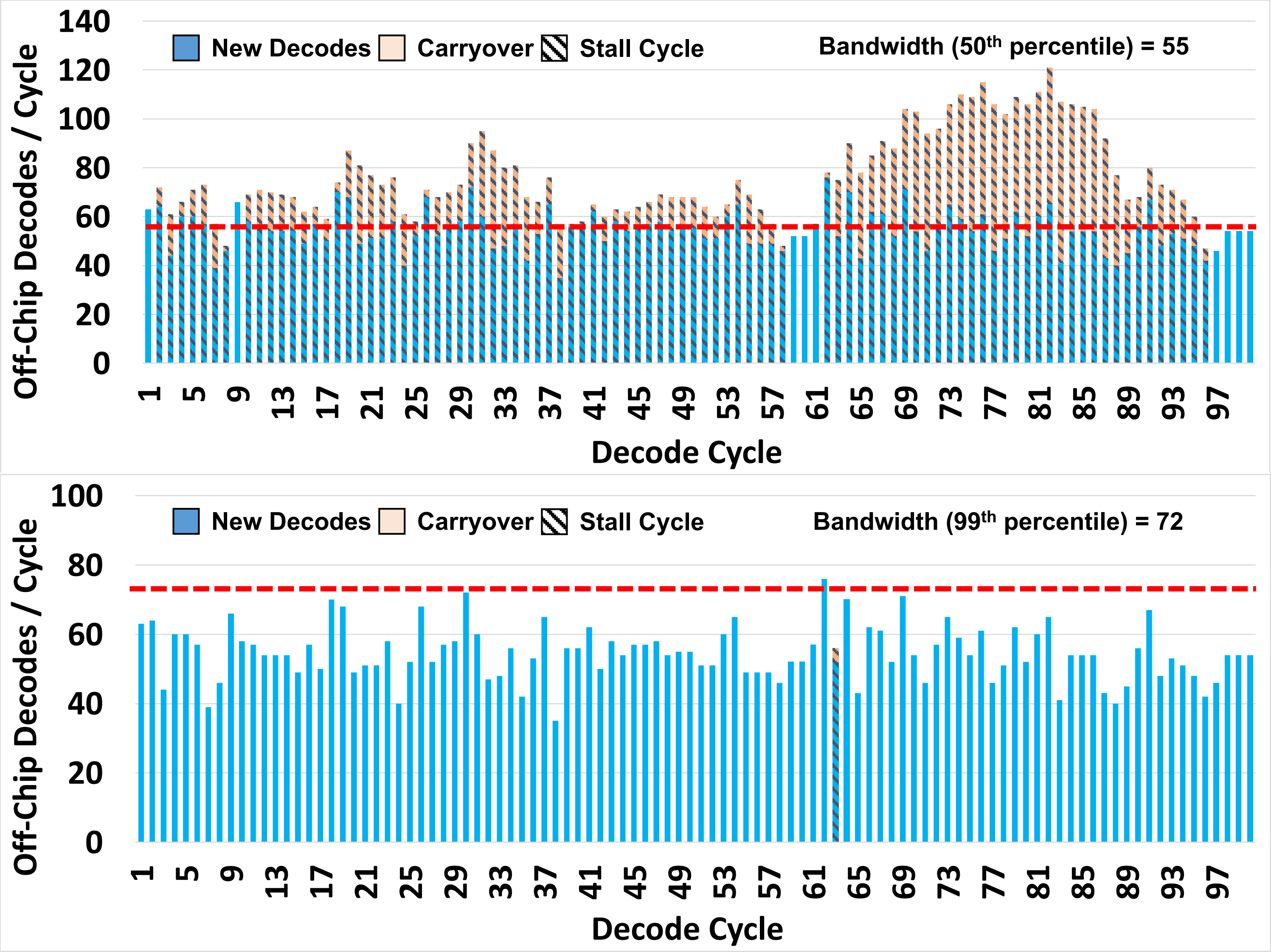
Observation: Error distribution vs Error rates



Logical errors (both Clique and complex decoder)



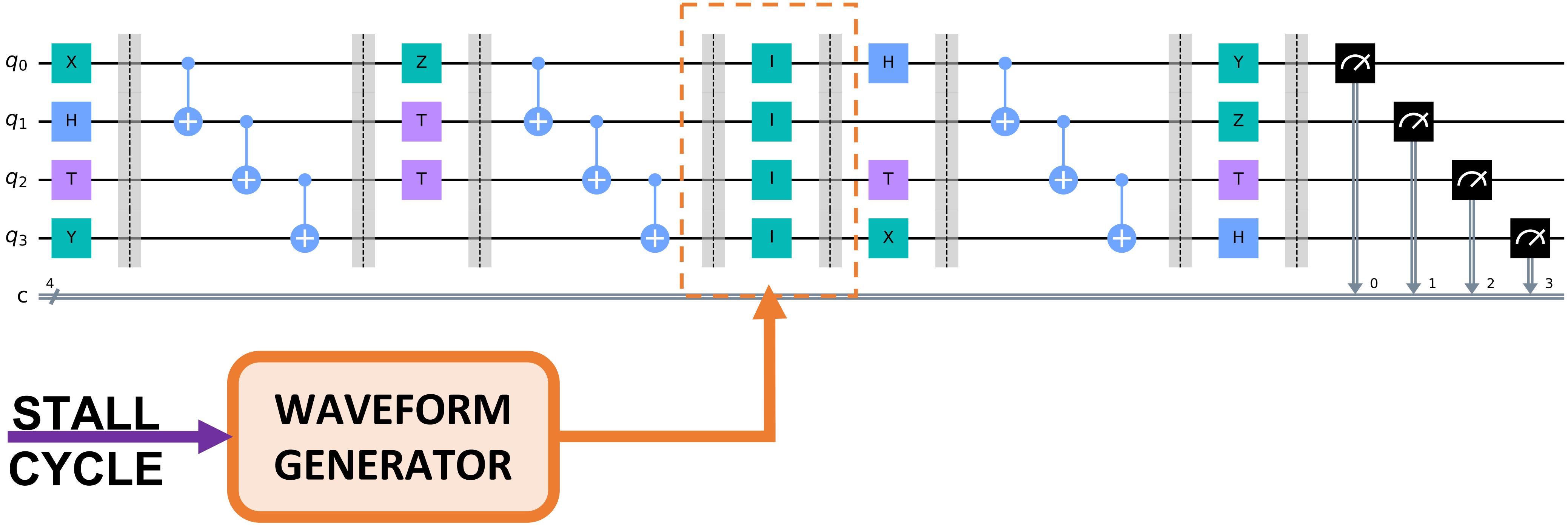
Statistical Off-chip Bandwidth Allocation



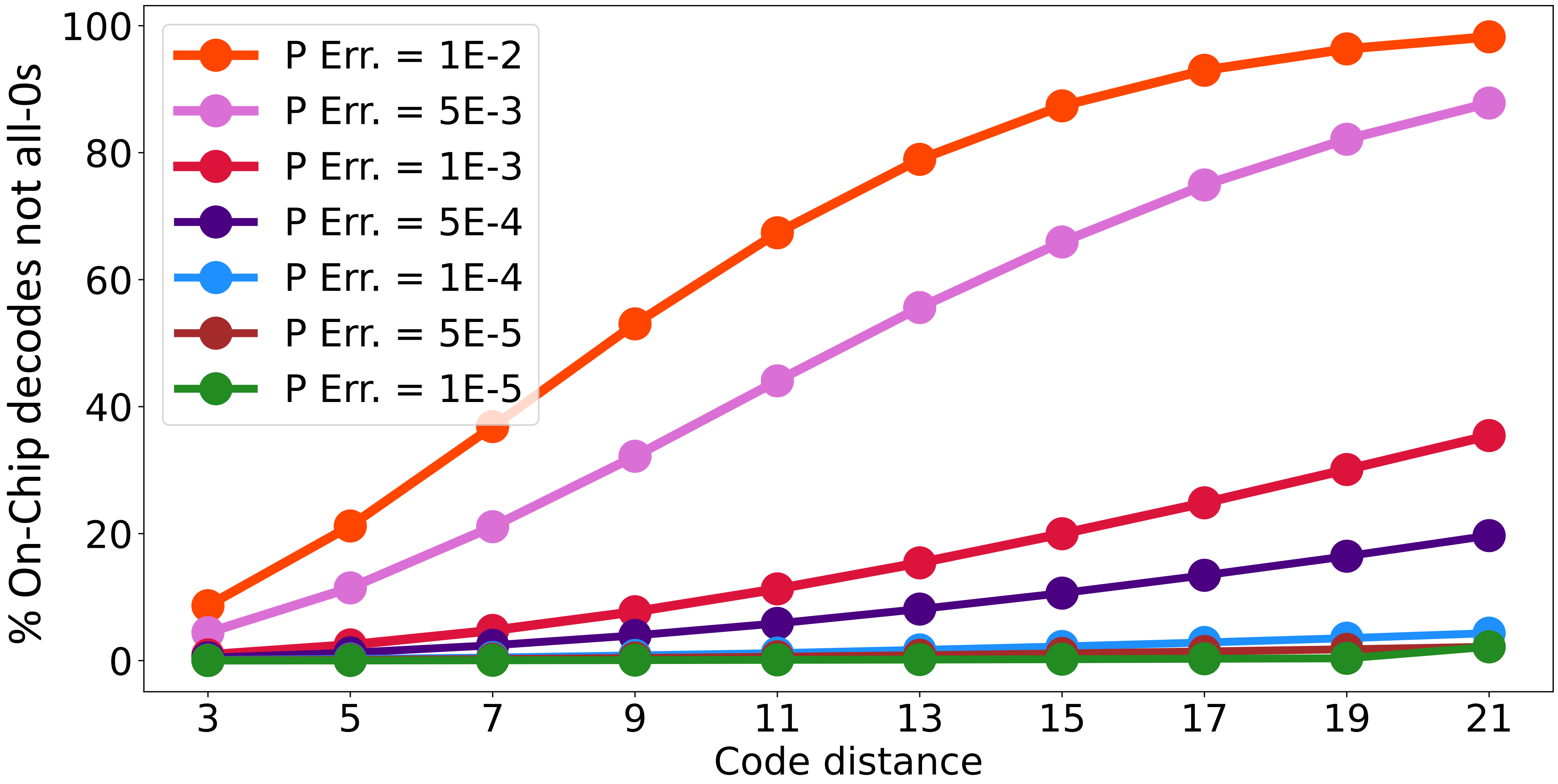
1000 logical qubits

Errors need to be resolved every cycle*

Idle cycle insertion on stall cycle



Results: Clique Decoder Coverage – not all 0s



Results: Bandwidth Allocation vs Stalling trade-offs