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

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

1: UChicago, 2: USC, 3: IBM

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

Encoded w/ redundancy

Physical qubit

Potential Errors

Decoding

Logical qubit

Classical Processing

Corrections
Error correction for fault tolerant quantum systems

Potential Errors

Decoding

Complex decoder!
Scope: Cryogenic quantum systems
Scope: Cryogenic quantum systems
System-level view: Traditional outside-fridge QEC decoding

Tbps I/O bandwidth → bandwidth bottleneck!

Low latency needed for functional correctness!!

1 microsecond

1000 L * 1000 P = O(1,000,000) qubits

E: Error Signatures

[Fowler, PR-A ’12]
...
[Das, HPCA ’22]
System-level view: Cryogenic inside-fridge QEC decoding

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

1000 L * 1000 P = O(1,000,000) qubits

P = ~1W

E: Error Signatures

Metal Interconnects
Superconducting Wires

[Holmes, ISCA ’20]
[Byun, ISCA ’22]
[Ueno, HPCA ’22]
Better than worst case approach to decoding

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

Outside-fridge BW-constrained approach

 decode all errors (one-size-fits-all)

Inside-fridge resource-constrained (approximate) approach

Be>er than worst case approach to decoding

Key Insight: Not all errors hard to decode → Separate common trivial errors from rare complex errors.
Better than worst case approach to decoding

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

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 (90 - 99+ %)

Logical qubit

Physical qubits

T: Trivial-to-decode
C: Complex-to-decode
Better than worst case approach to decoding

Error chains: Hard to decode and very rare!!

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

T: (90 - 99+ %)
C: (<1 - 10 %)

[Clique/ BTWC]

[HPCA '22]
[ISCA '20]
[ISCA '22]
[HPCA '22]

[PR-A '12]
Better than worst case approach to decoding

Common trivial errors $\rightarrow$ simple cryogenic ‘Clique’ decoder.
Rare complex errors $\rightarrow$ 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!

Metal Interconnects

Superconducting Wires

System-level view: Better than worse-case decoding

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

Metal Interconnects

Superconducting Wires
How does QEC work? (Surface code)
How does QEC work? (Surface code)

1 logical qubit

Code distance (d) = 3

Z parity qubit

X parity qubit
How does QEC work? (Surface code)
How does QEC work? (Surface code)

1 logical qubit

Code distance \((d) = 3\)

COMPLEX DECODER
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?
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

Isolated data errors only trigger locally paired syndromes which are easy to decode.
How clique trivially decodes isolated errors

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.

Code distance ($d$)
Quantitative benefits: Fridge I/O bandwidth reduction

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

Comparison to AFS compression [Das, HPCA ‘22]:
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.

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.
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
How does QEC work? (Surface code)

1 logical qubit

No visible syndrome trigger

Code distance (d) = 3

No visible syndrome trigger

Code distance (d) = 3
How does QEC work? (Surface code)

1 logical qubit

Error chains generate less syndrome information.

Code distance \(d = 3\)
Why are isolated errors much more common than error chains?

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

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

\[ P \text{ (1 error in block)} = N \times \text{PER} = 4.9\% \]
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)  
$= 6 \times N \times \text{PER}^2 = 0.03\%$  
(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.**

**Isolated error detection logic**

- Even # of surrounding syndromes flipped?
- Is the clique active?

**Decoding + correction logic**

- Correct w
- Correct x
- Correct y
- Correct z

**Code distance = d**

**Physical qubits**

\( (n) = 2^d \)

\( x \, O(n) \)
**Clique decoder hardware design**

**Isolated error detection logic**
- Even # of surrounding syndromes flipped?
- Is the clique active?

**Decoding + correction**
- Correct w
- Correct x
- Correct y
- Correct z

** Syndromes**
- Syndrome flipped?
- Syndrome stayed as is for additional cycle?
- Syndrome stayed as is for more cycles?
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.
Better logical qubits possible with higher code distance, but with increased overheads.

Logical error decreases with code distance.
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

![Graph showing logical error rate against physical error rate with different curves for baseline and clique + baseline with various d values.]
How does QEC decoding work?

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.

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

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

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

1 logical qubit w/ rotated surface code

Data qubit

Z parity qubit

X parity qubit

Error Syndromes

Support for high physical error rates

Surface Codes... [Fowler2012]
Decoding is complex at large code distances

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

Bandwidth (50th percentile) = 55

Errors need to be resolved every cycle*

1000 logical qubits
Idle cycle insertion on stall cycle
Results: Clique Decoder Coverage – not all 0s
Results: Bandwidth Allocation vs Stalling trade-offs

- PE=1E-2 / D=9
- PE=5E-3 / D=5
- PE=5E-3 / D=21