Efficient Fault-Tolerant Quantum Computing

Martin Suchara

AT&T Labs

September 8, 2015

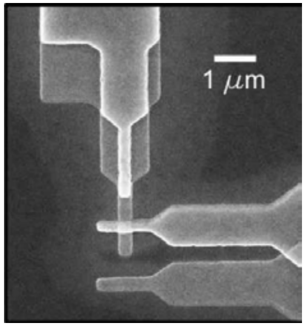
Applications of Quantum Computing

- Shor's factoring algorithm
  - Find prime factors of integer N
  - Quantum algorithm runs in polynomial time
  - Can be used to break public-key cryptography (RSA)

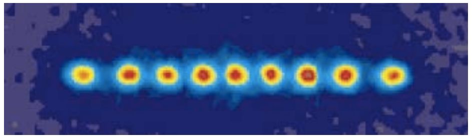
- How can we build a scalable reliable quantum computer?

Some Candidate Hardware Technologies

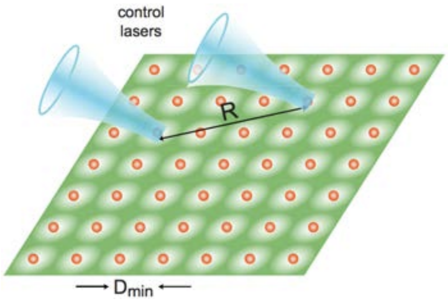
- Superconducting qubits
  - Josephson Junctions between superconducting electrodes



- Ion traps



  - Ions trapped in electromagnetic field, gates performed by applying lasers
- Neutral atoms
  - Ultracold atoms trapped by light waves in an optical lattice



3

Quantum Error Correction (QEC)

- Need error correction for reliable information storage and computation with unreliable technologies
- Much more challenging than classically
  - Analog nature of quantum operations
  - New kinds of errors: partial bit flips, phase flips, small shifts

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \rightarrow \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

- State can even leak out of the code space

Overview

I. How much resources needed by quantum computing?
   - *ICCD 2013 Conference, Quantum Information and Computation 2014*

II. Improved error decoding – MLE algorithms

III. New types of errors – correcting qubit leakage

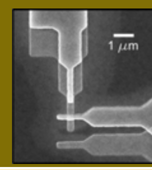
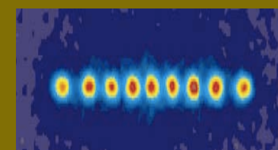
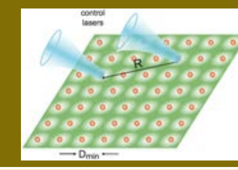
IV. Future directions

Shor's Factoring Algorithm – How Long to Break a Key?

- Factor a 1024-bit number
- Algorithm needs approximately 1.68×10^8 Toffoli gates and 6,144 logical qubits
  (Jones et al., 2012)

Gate	Occurrences	Parallelization Factor
CNOT	1.18×10^9	1
Hadamard	3.36×10^8	1
T or T†	1.18×10^9	2.33
Other gates	negligible	

Properties of Future Technologies (Estimates in IARPA QCS Project)

	Supercond. Qubits 	Ion Traps 	Neutral Atoms 
Average Gate Time (ns)	25	32,000	19,000
Worst Gate Error	1.00×10^{-5}	3.19×10^{-9}	1.47×10^{-3}
Memory Error	1.00×10^{-5}	2.52×10^{-12}	high

- The speed and reliability of gates varies
- Much higher errors than classical computers

QuRE: The Quantum Resource Estimator Toolbox

Algorithm Specs
- # of logical qubits
- # of logical gates
- Circuit parallelism

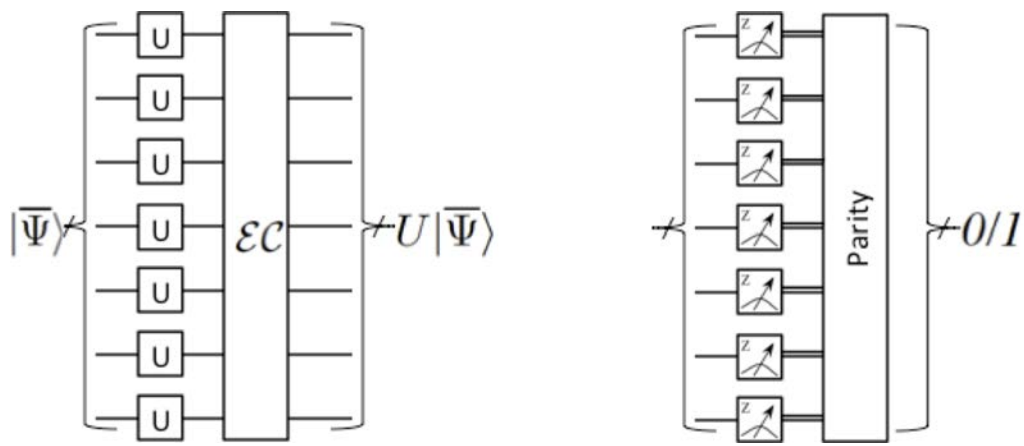
Analysis of Error Correction
- Estimate cost of each logical operation as a function of error correction "strength"

Technology Specs
- Gate times and fidelities
- Memory error rates

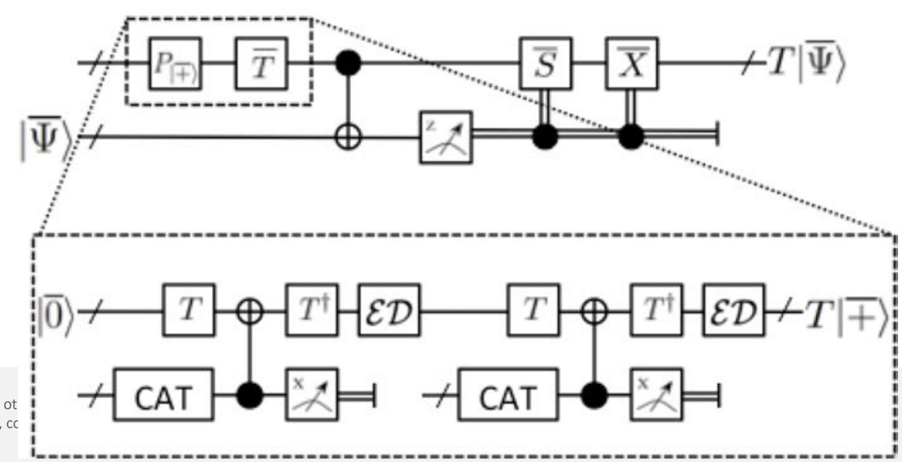
Automated Resource Estimate
- Find code distance / concatenations needed for successful error correction
- Estimate number of physical qubits, running time, physical gate and instruction count, etc.

Concatenated Error Correction Codes – Steane [[7,1,3]] Code

- Multiple qubits encode a single logical qubit
- Most operations transversal:

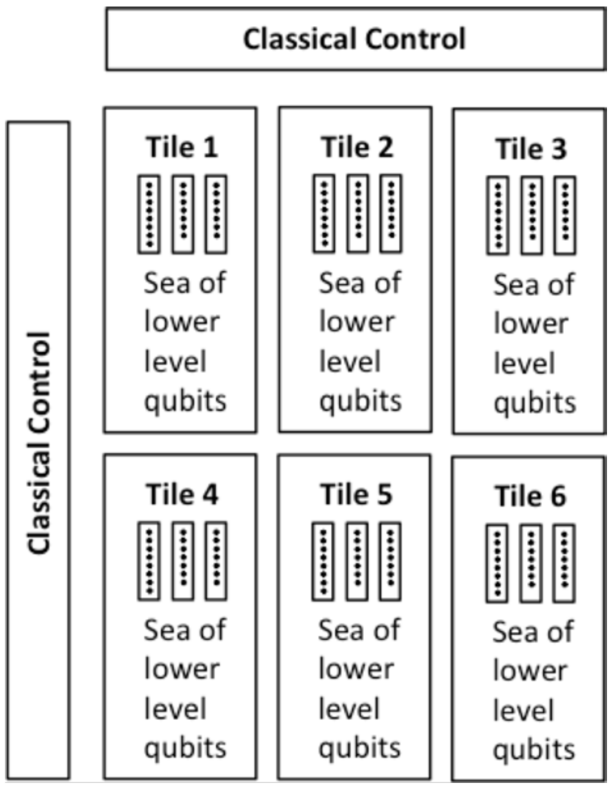


- Non-transversal T gate:



Concatenated Error Correction Codes – Tiled Qubit Layout

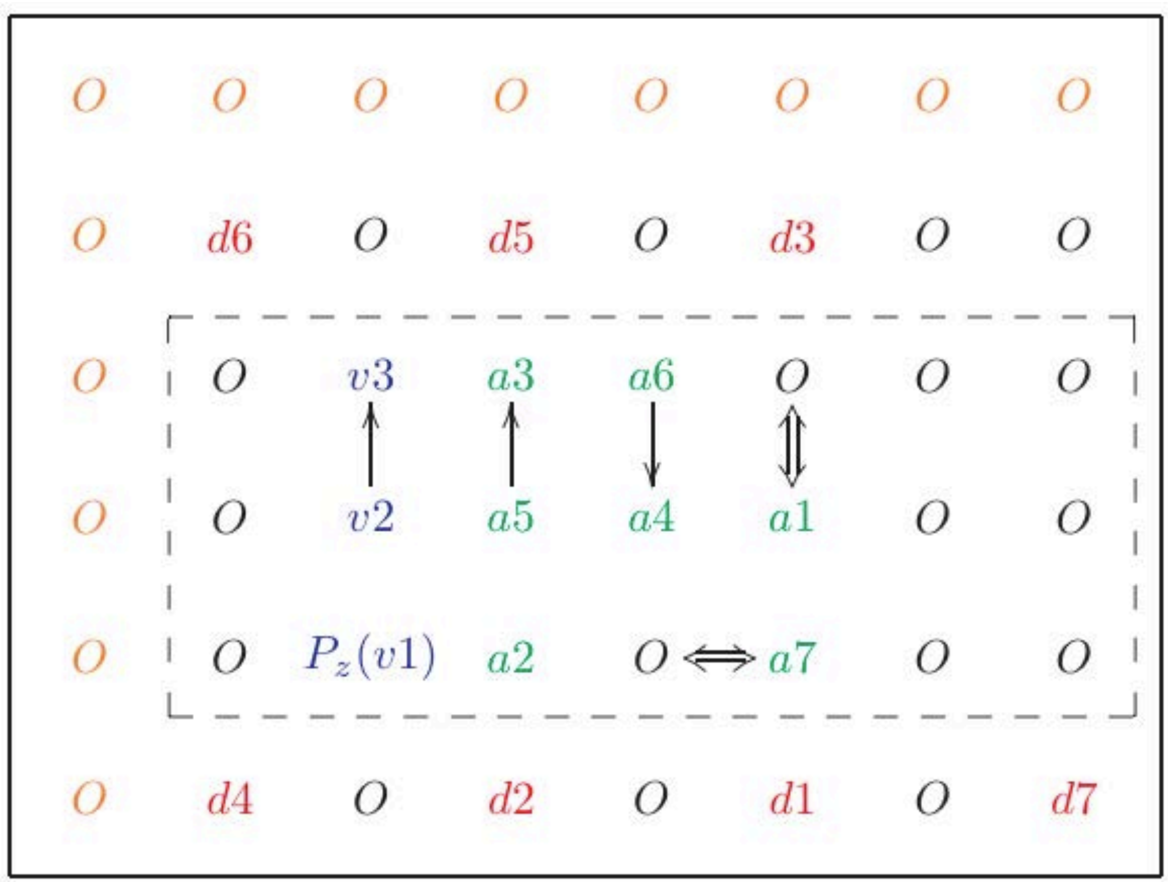
- Each logical qubit is stored in a separate tile



- Tiles are hierarchical
- Tiles must contain enough data and ancilla qubits
- Supported operations:
  - Error correct a tile
  - Apply fault-tolerant logical operation

Optimized Layout in Each Tile

(*Svore et al., 2006*)



"empty" qubit

data qubit

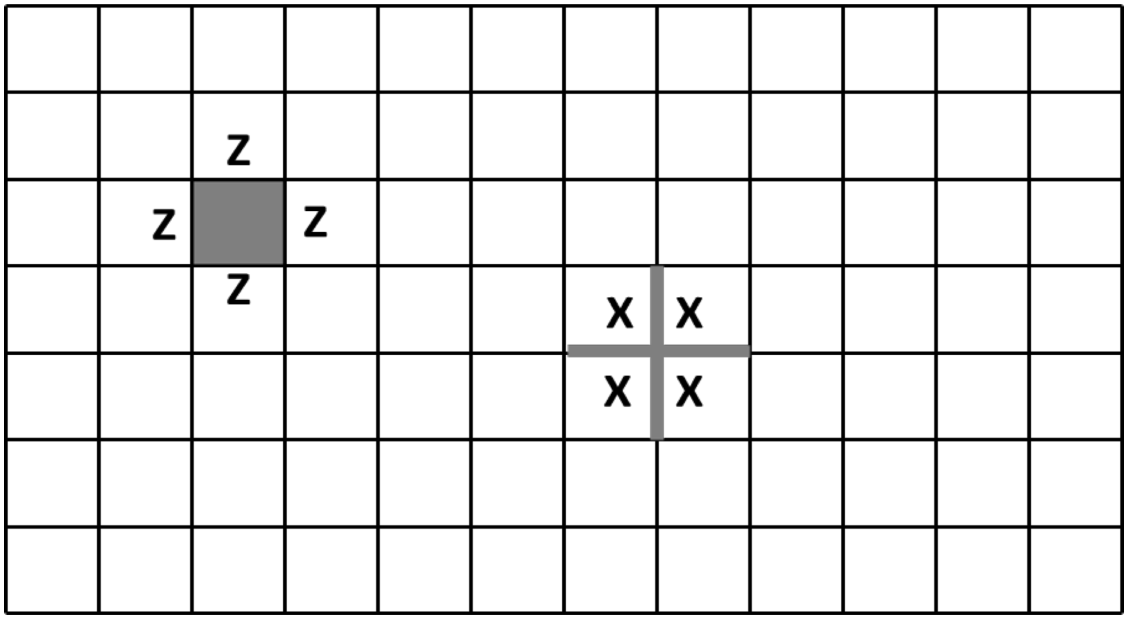
verification qubit

ancilla qubit

⟺ SWAP

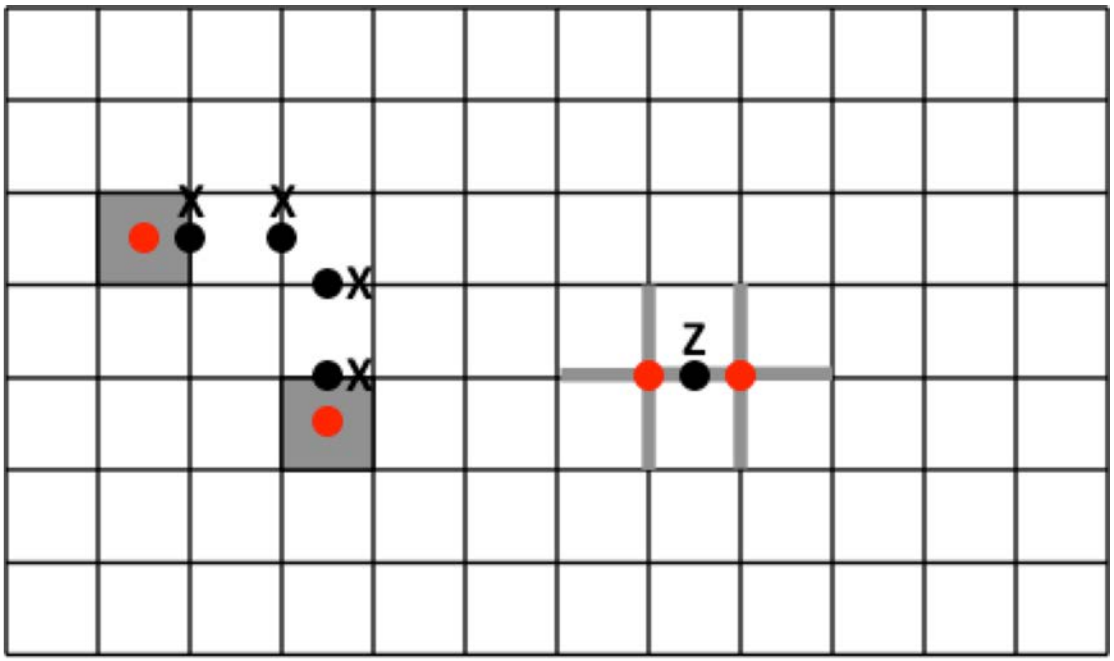
↑ CNOT

Topological Quantum Error Correction – The Surface Code



- Physical qubits on links in the lattice
- Measuring the shown “check” operators yields error syndromes

Topological Quantum Error Correction – Example of Errors and Syndromes

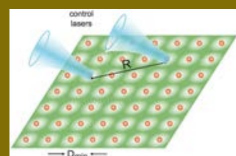
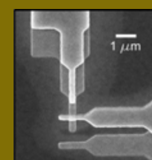
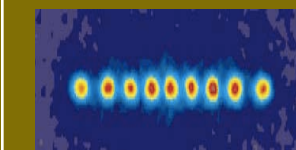


- Error correction performed continuously
- Guess a likely error consistent with observed syndromes

Topological Quantum Error Correction – Tiles Represent Logical Qubits

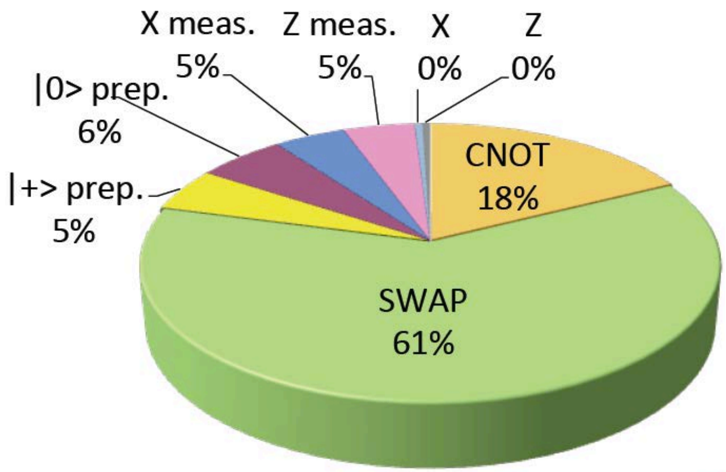
- Each logical qubit represented by a pair of holes
- CNOT gates performed by moving holes around each other

Numerical Results – Shor's Factoring

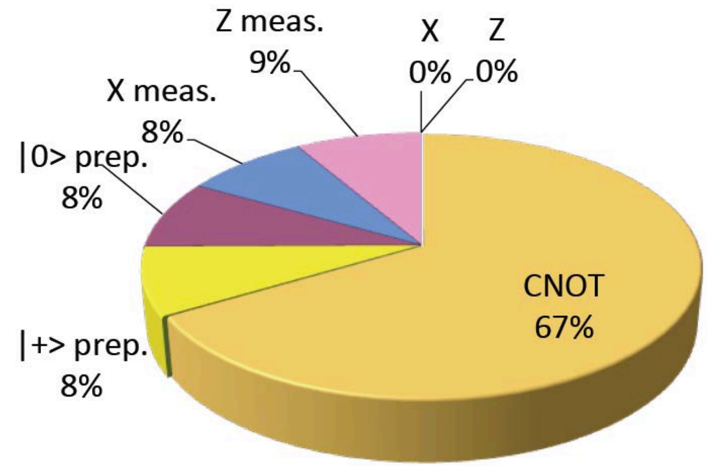
	Neutral Atoms	Supercond. Qubits	Ion Traps	
	$e = 1 \times 10^{-3}$ $t = 19,000$ ns	$e = 1 \times 10^{-5}$ $t = 25$ ns	$e = 1 \times 10^{-9}$ $t = 32,000$ ns	
				
Surface Code	2.6 years	**10.8 hours**	2.2 years	Time
	5.3×10^{8}	**4.6×10^{7}**	1.4×10^{8}	Qubits
	1.0×10^{21}	**2.6×10^{19}**	5.1×10^{19}	Gates
Steane Code	-	5.1 years	**58 days**	Time
	-	2.7×10^{12}	**4.6×10^{5}**	Qubits
	-	1.2×10^{32}	**4.1×10^{18}**	Gates

Qualitative Difference in Gate Composition

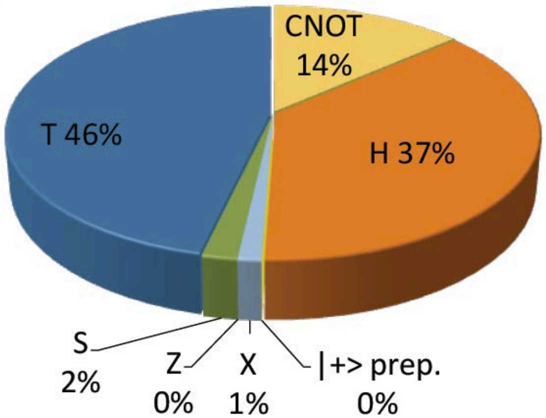
Steane code:



Surface code:



Logical circuit:



Fault tolerance increases number of qubits by 2-4 orders of magnitude and number of gates by 9-10 orders of magnitude!

T gates are the most expensive

Overview

I. How much resources needed by quantum computing?

II. Improved error decoding – MLE algorithms

- *Physical Review A, 2014*

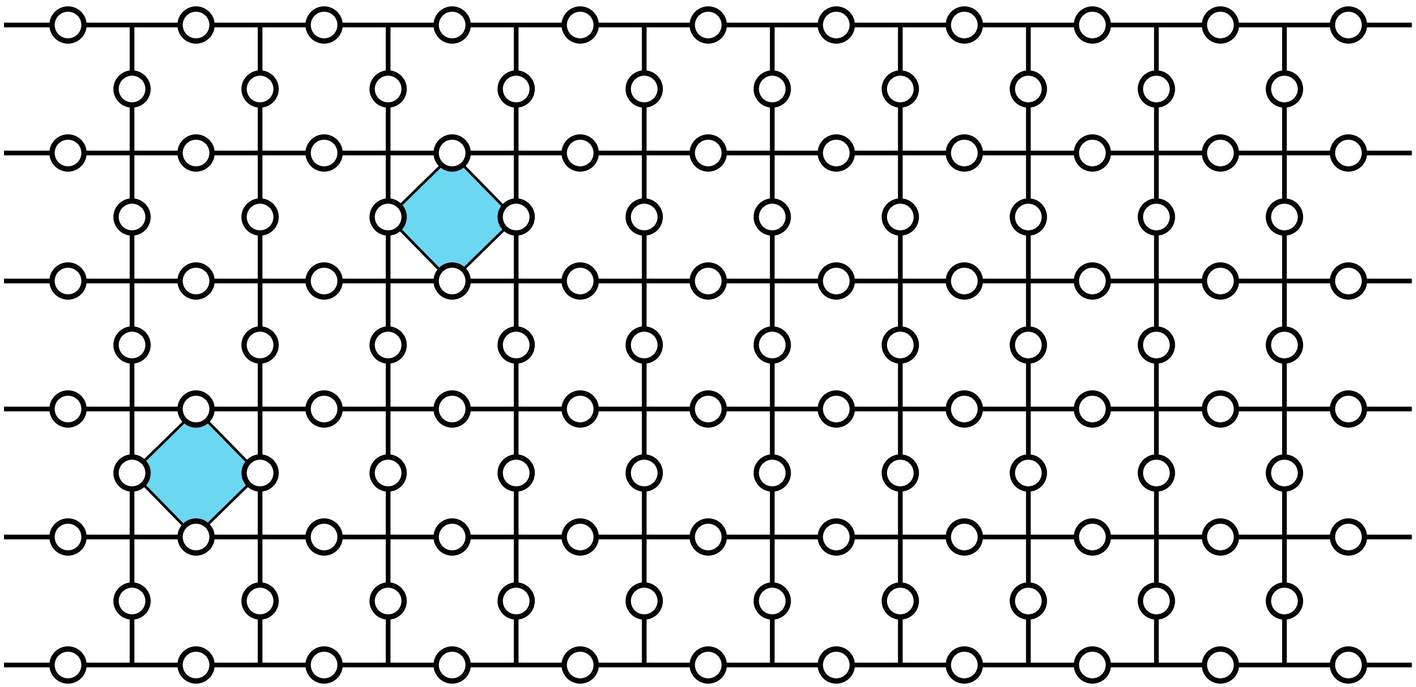
III. New types of errors – correcting qubit leakage

IV. Future directions

Goal – Construct a Maximum Likelihood Decoder for the Surface Code

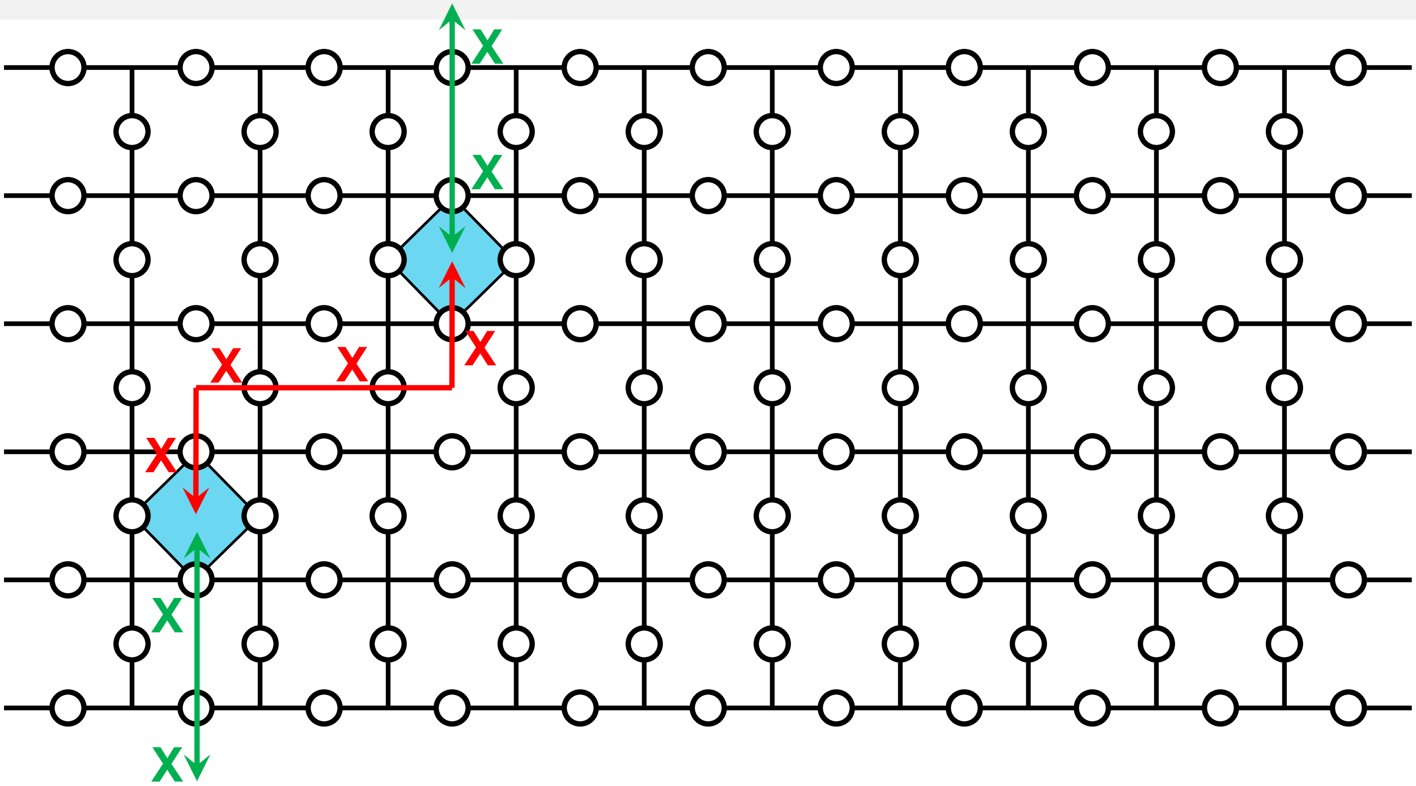
- Error decoders use a heuristic - minimum weight matching - to guess a likely error
  - Problem 1: ignores degeneracy, existence of multiple error chains between a specific syndrome pair
  - Problem 2: ignores correlations of X and Z errors in some error models
- Goal: construct a decoder that finds the most likely error given the observed syndrome

Inefficiency of Minimum Weight Matching (MWM) – Degeneracy



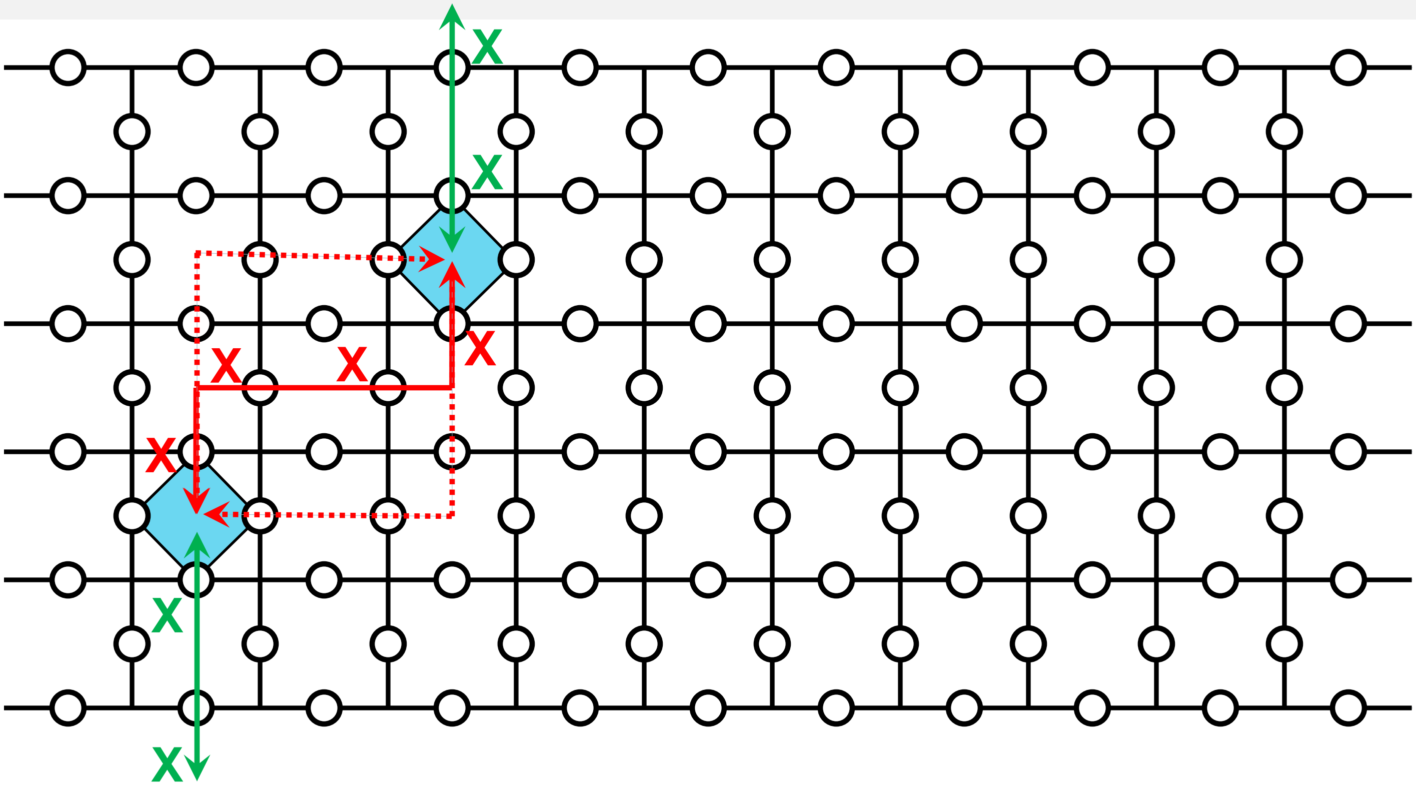
- A possible syndrome measurement

Inefficiency of Minimum Weight Matching (MWM) – Degeneracy



- Two ways of matching with equal cost, matching algorithm picks one at random

Inefficiency of Minimum Weight Matching (MWM) – Degeneracy



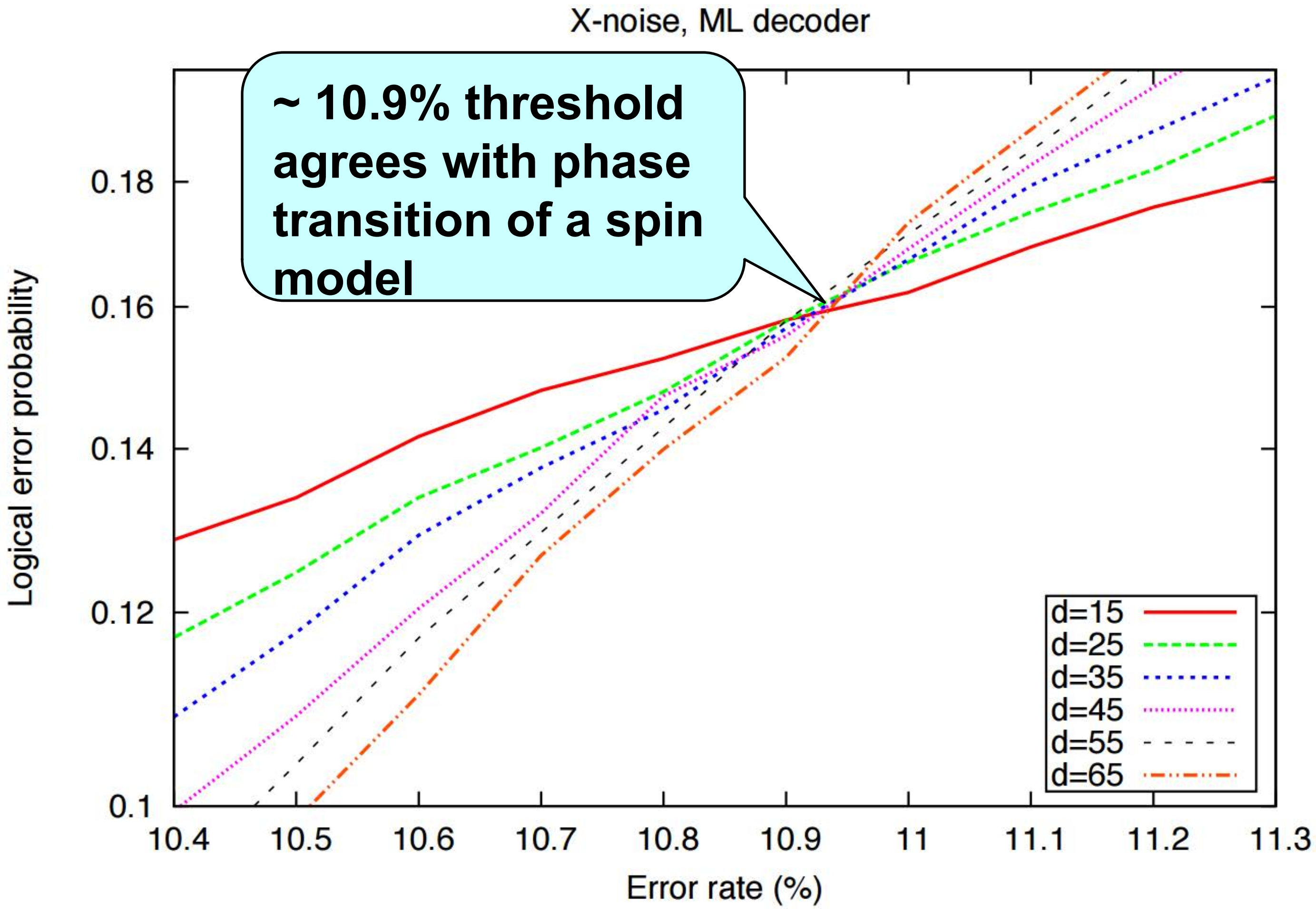
- The green error less likely than one of the red ones, which are equivalent

Our Maximum Likelihood Decoder (MLD)

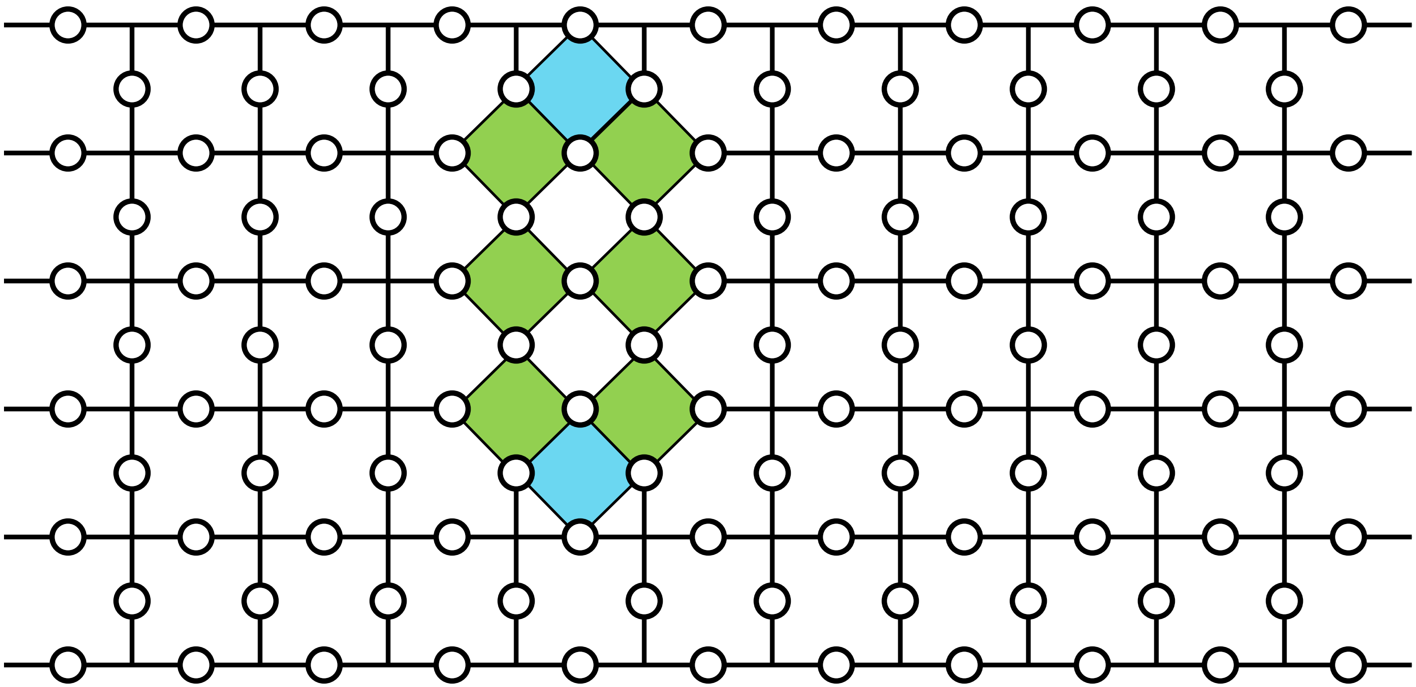
- Maximum likelihood decoder finds the most likely error given the observed syndrome
  - Must consider all error chains
  - Works by formulating problem as a matchgate quantum circuit that can be simulated efficiently
  - Exact solution in time $O(n^2)$ where n is the number of code qubits
  - Only works for X error noise:

$$\epsilon_X = \epsilon, \quad \epsilon_Y = \epsilon_Z = 0$$

Threshold of the ML Decoder, X-Noise Model

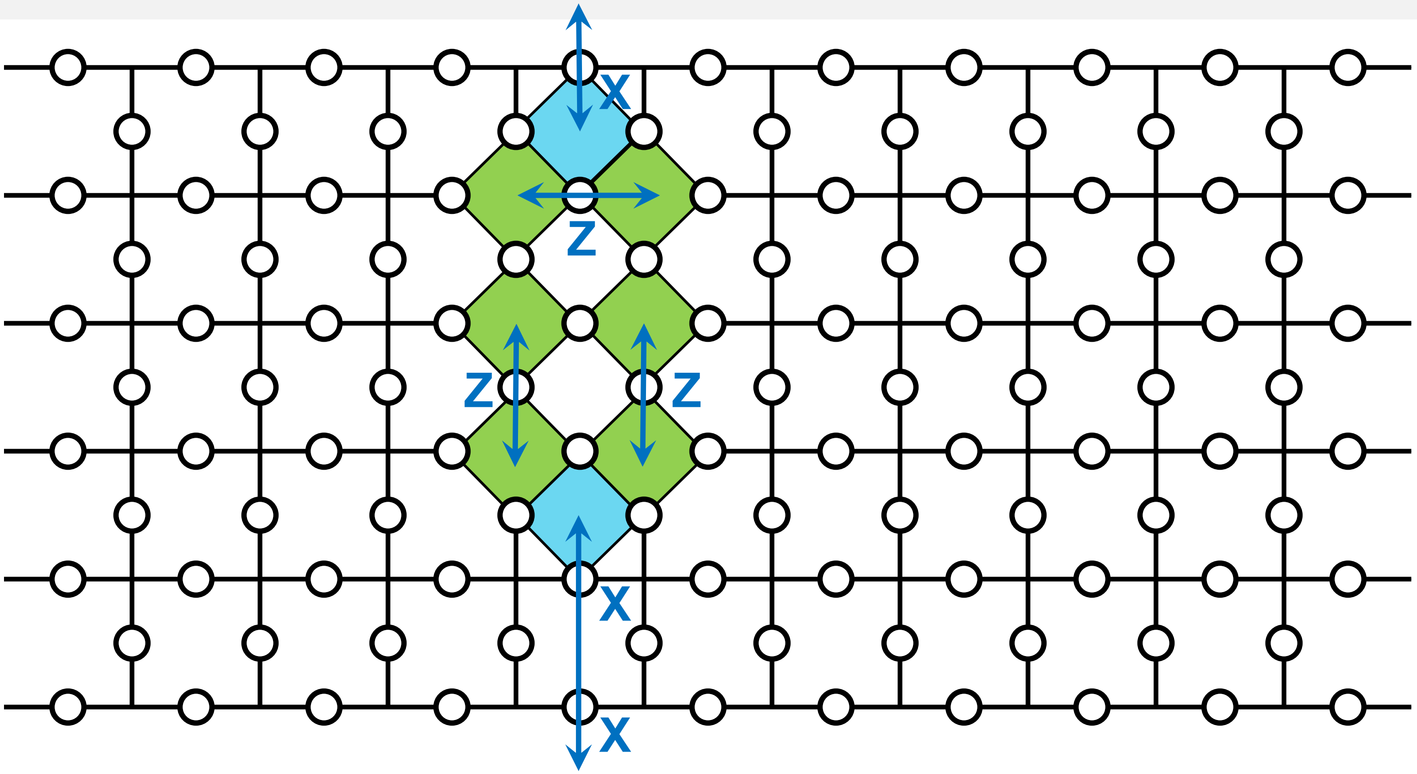


Inefficiency of Minimum Weight Matching (MWM) – Correlations



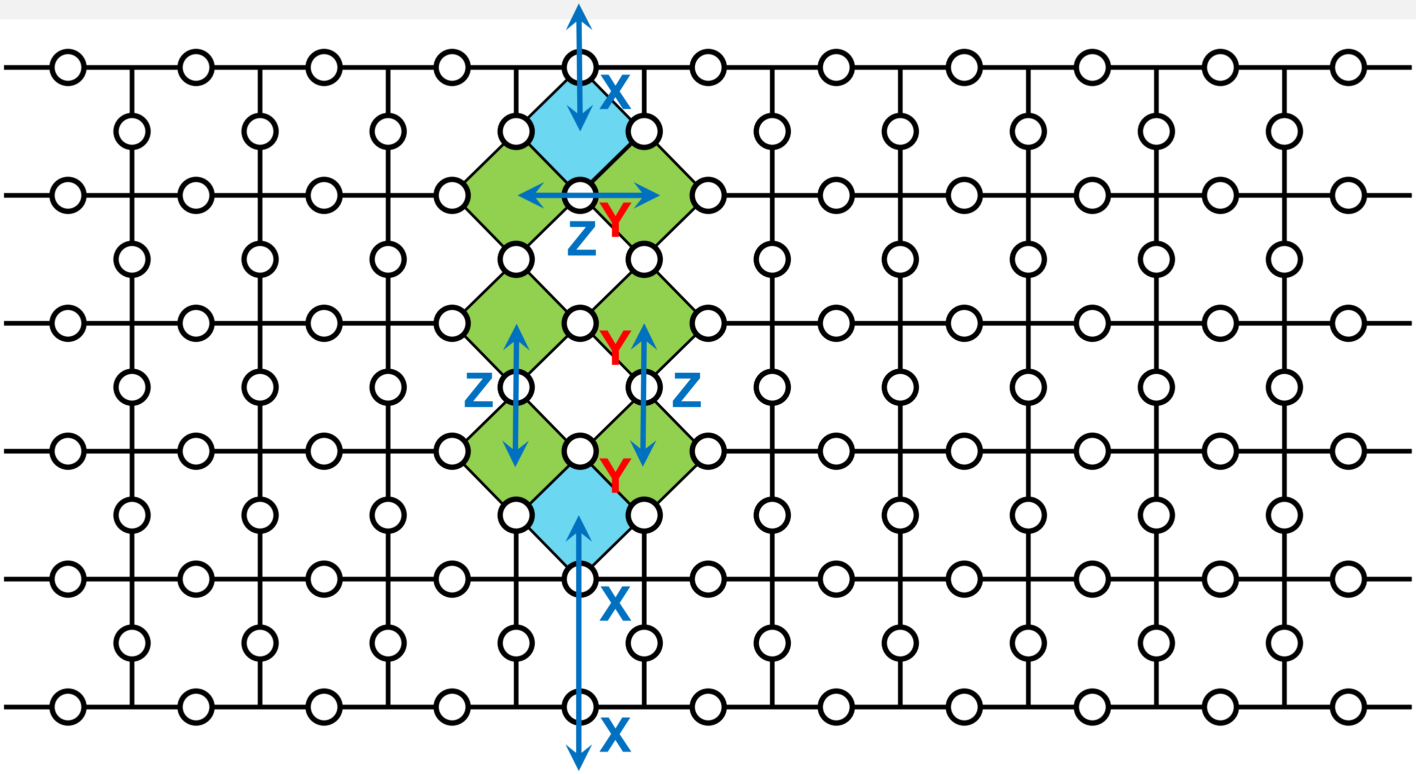
- Assume $\epsilon_X = \epsilon_Y = \epsilon_Z = \epsilon/3$
- Example of a syndrome measurement

Inefficiency of Minimum Weight Matching (MWM) – Correlations



- Pair up syndromes and correct errors

Inefficiency of Minimum Weight Matching (MWM) – Correlations



- Actual error (in red) much more likely cause of syndrome! Logical error occurs.

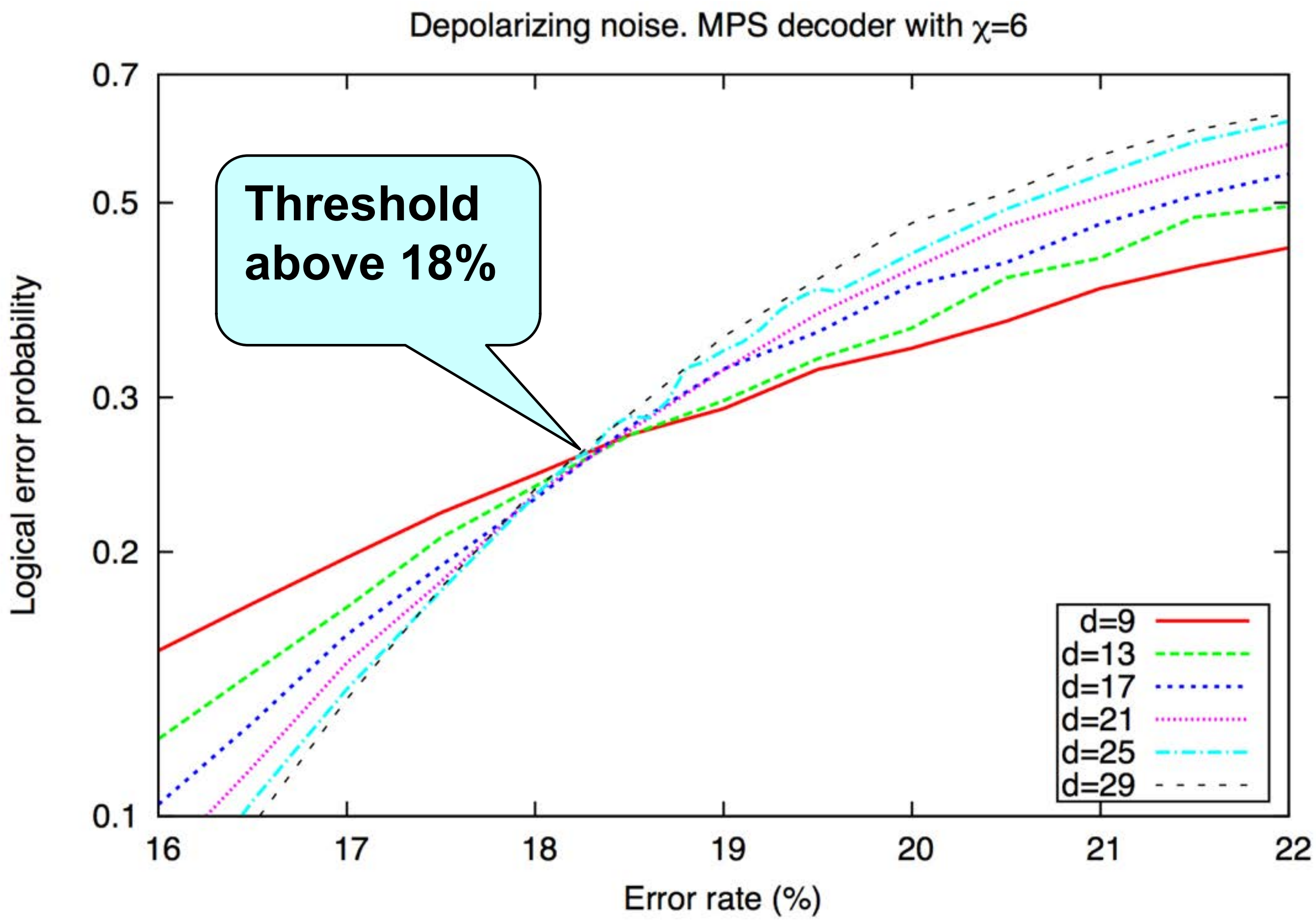
Approximate Solution for More General Noise Models

- Approximate algorithm that uses matrix product states (MPS)
  - Approximate solution in time O(n χ^3) where χ controls the approximation precision
  - Close to optimum for small χ
  - Works for the depolarizing noise model:

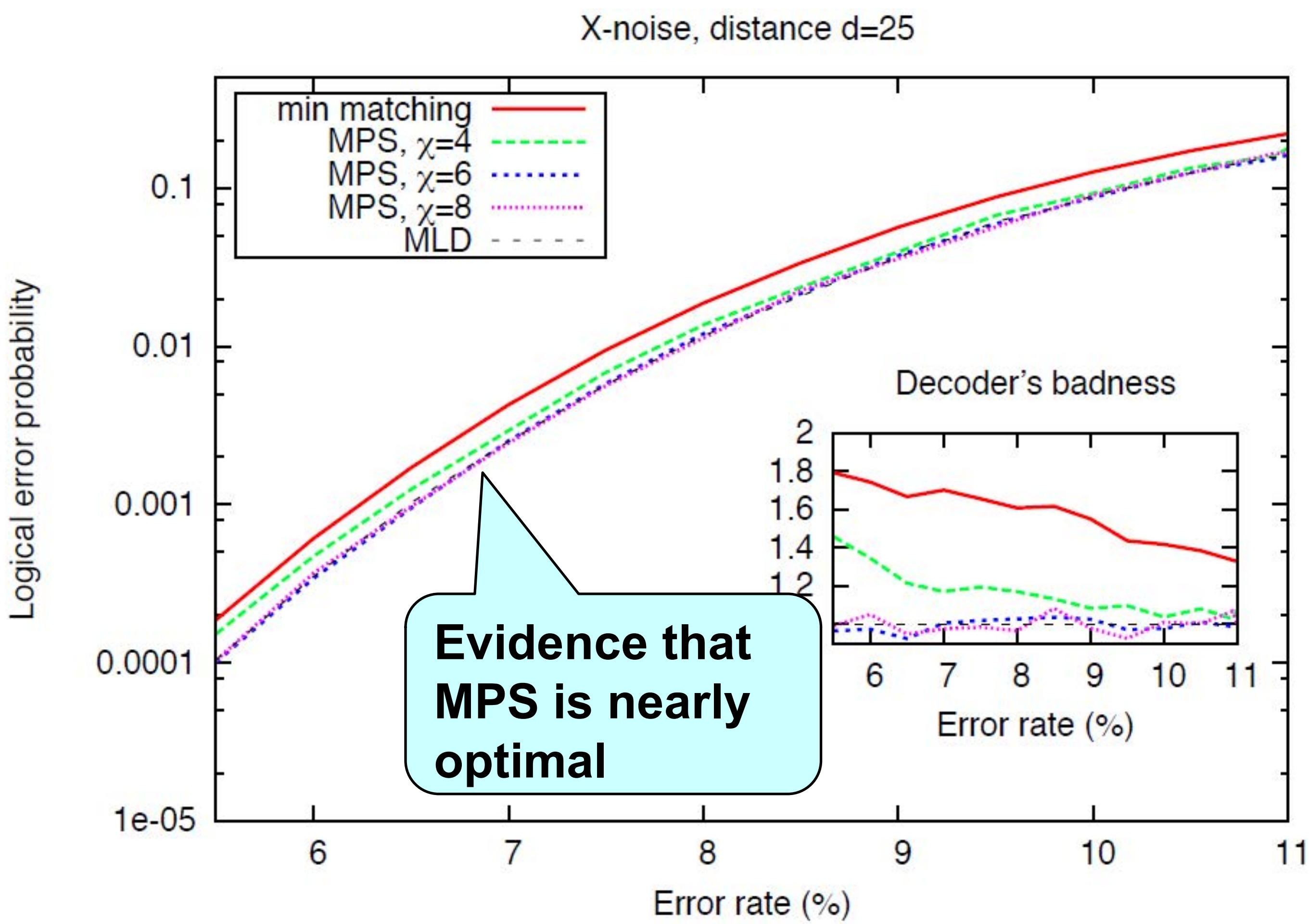
    $$\epsilon_X = \epsilon_Y = \epsilon_Z = \epsilon/3$$

  - Key step contracts a tensor network on the two-dimensional grid of the code

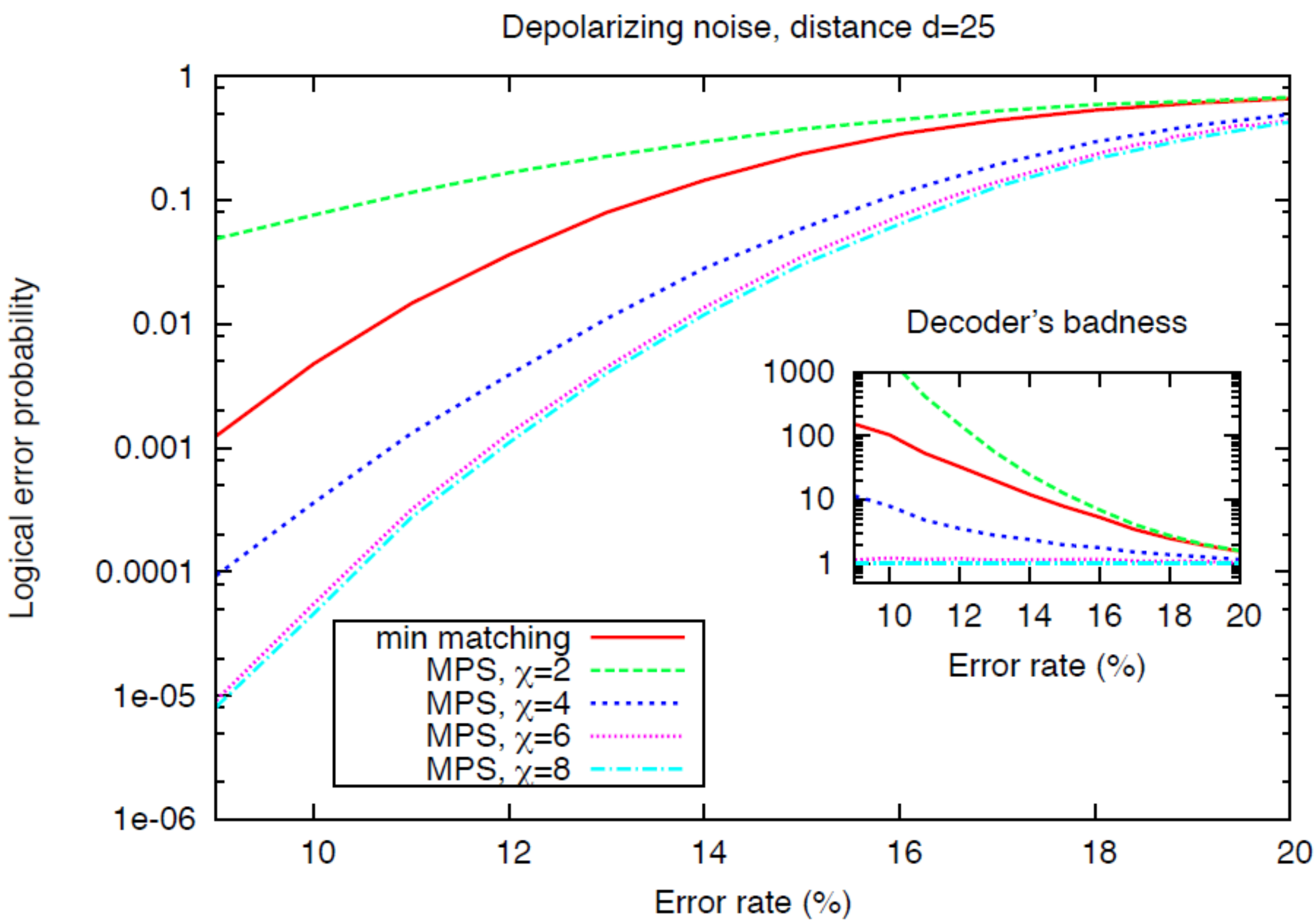
Threshold of the ML Decoder, X-Noise Model



Exact and Approximate ML Decoding, X-Noise Model



Approximate ML Decoding with the MPS Algorithm, Depolarizing Noise



100 fold or better improvement of logical error rates in depolarizing noise models!

Considering correlations much more important than degeneracy

Open question: maximum likelihood decoding with noisy syndrome extraction

Overview

I. How much resources needed by quantum computing?

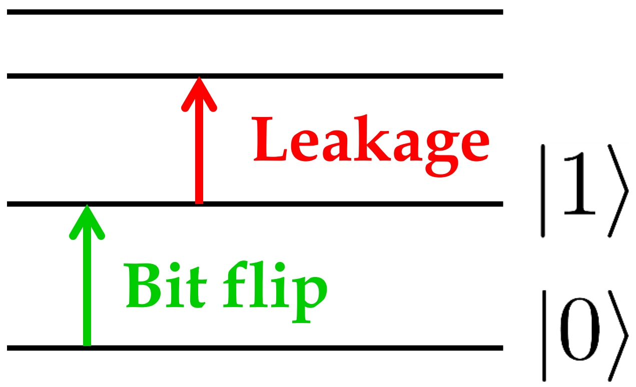
II. Improved error decoding – MLE algorithms

III. New types of errors – correcting qubit leakage
  - *Quantum Information and Computation 2015, ISIT 2015 Conference*

IV. Future directions

What is Qubit Leakage?

- Physical qubits are not ideal two-level systems and may leak out of the computational space



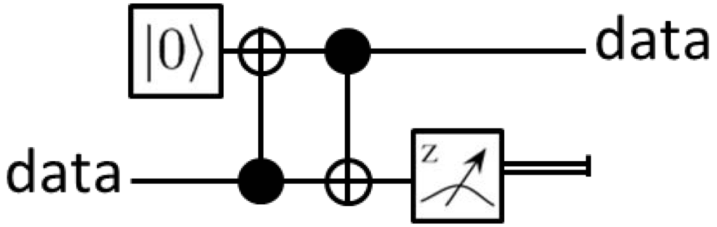
- With standard error correction techniques leaked qubits accumulate and spread errors

- Our work: simple model of leakage and comparisons of leakage reduction strategies

Leakage in the Literature

- Analysis of leakage reduction units based on quantum teleportation, threshold theorem for concatenated codes

  (Aliferis, Terhal 2005)



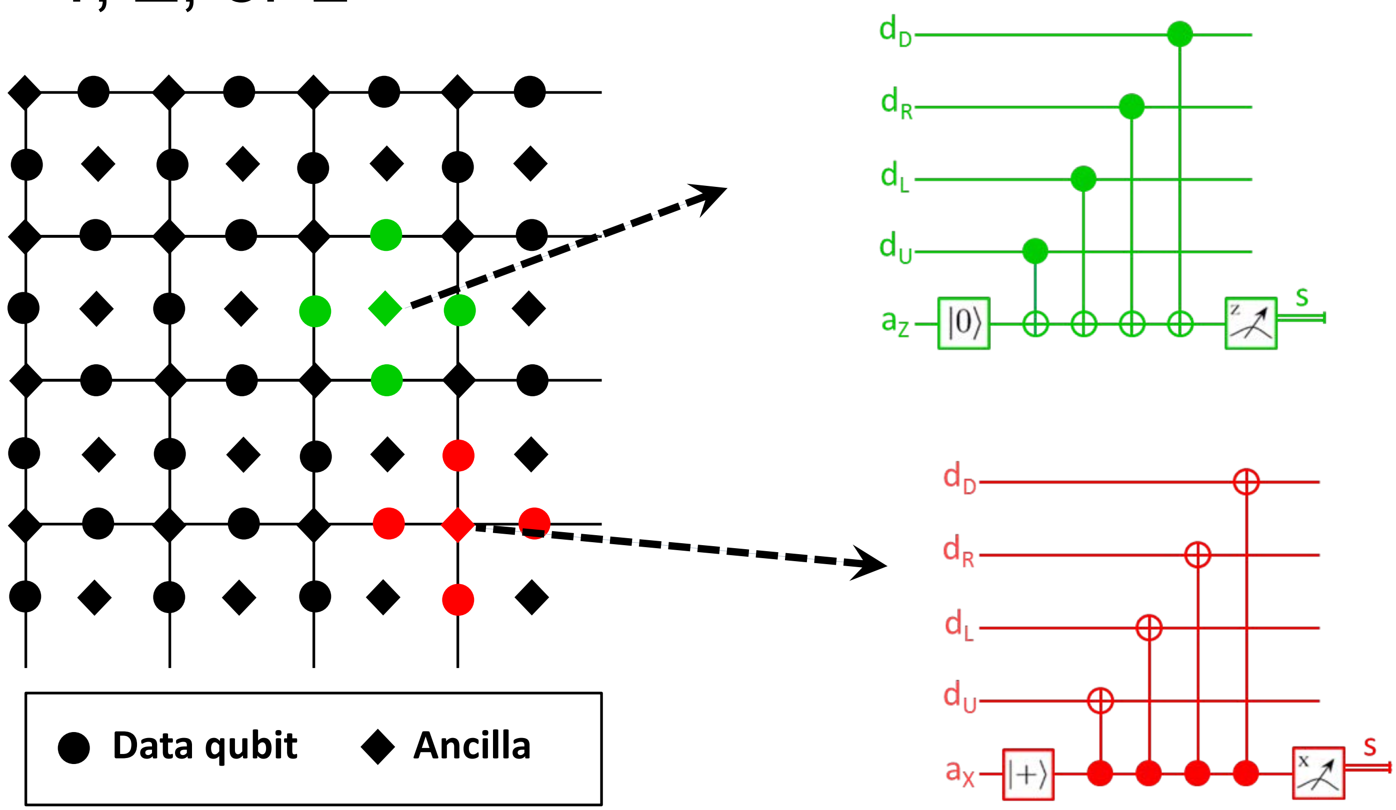
- Model of leakage for repetition code that labels leaked qubits

  (Fowler 2013)

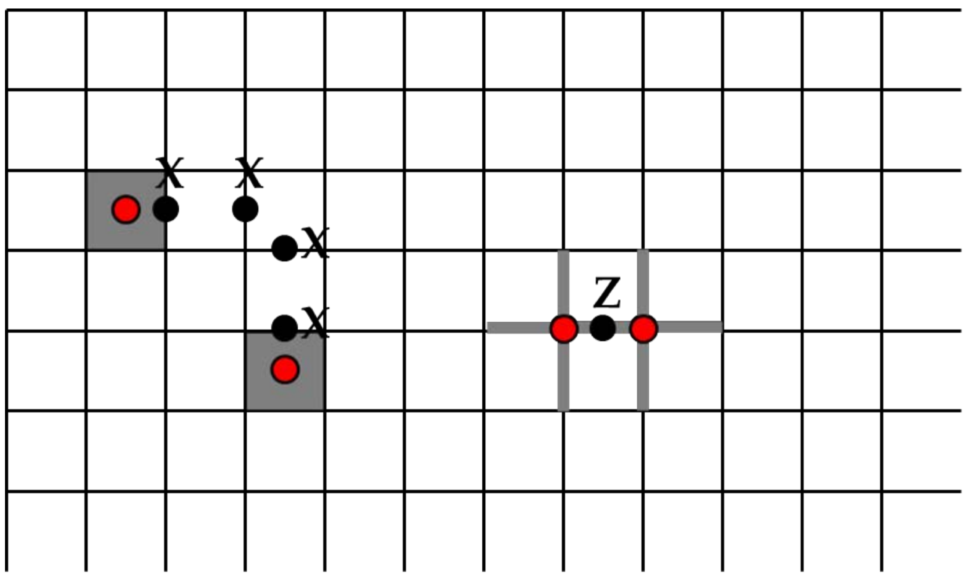
Simple Model of Error Correction with the Toric Code

- Label-based model: each qubit is in state I, X, Y, Z, or L

Syndrome extractions:

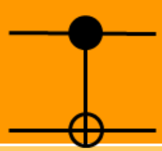
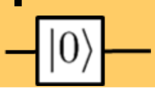


Monte Carlo Simulations: Measure Syndromes and Decode Errors



- Measure syndromes, build a 3-D syndrome history (a 2-D slice shown)
- Use minimum weight matching decoder and correct errors between matched syndrome pairs

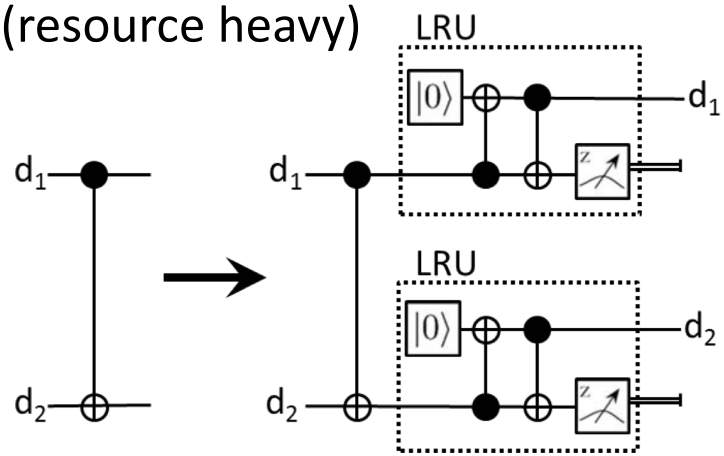
Leakage Model of Gates

Gate	Possible Errors	Leakage Errors
Identity	X, Y, Z	if leaked relaxes w/ prob. pd, doesn't increase leakage
CNOT 	IX, XX, XZ, etc.	if leaked, applies random Pauli to the other qubit; leaks w/ prob. pu and relaxes w/ prob. pd
Preparation 	orthogonal state	leaks w/ prob. pu
Measurement 	incorrect	if leaked, always measures 1 (also consider leakage detection)

Leakage Reduction Circuits

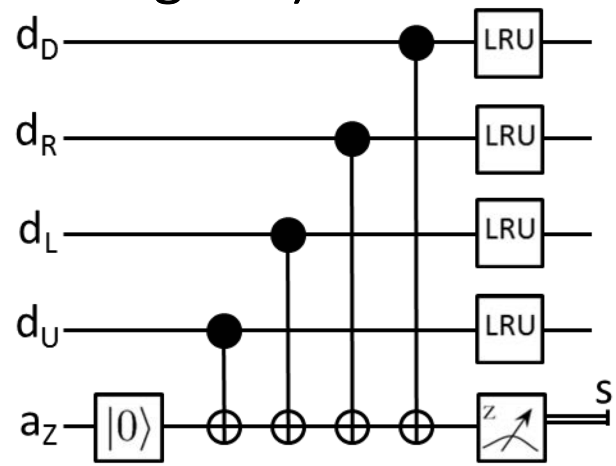
1. Full-LRU:

(resource heavy)



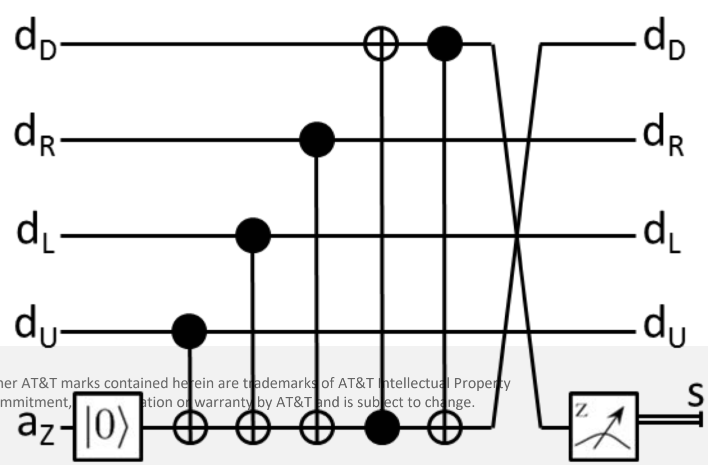
2. Partial-LRU:

(fewer gates)



3. Quick circuit:

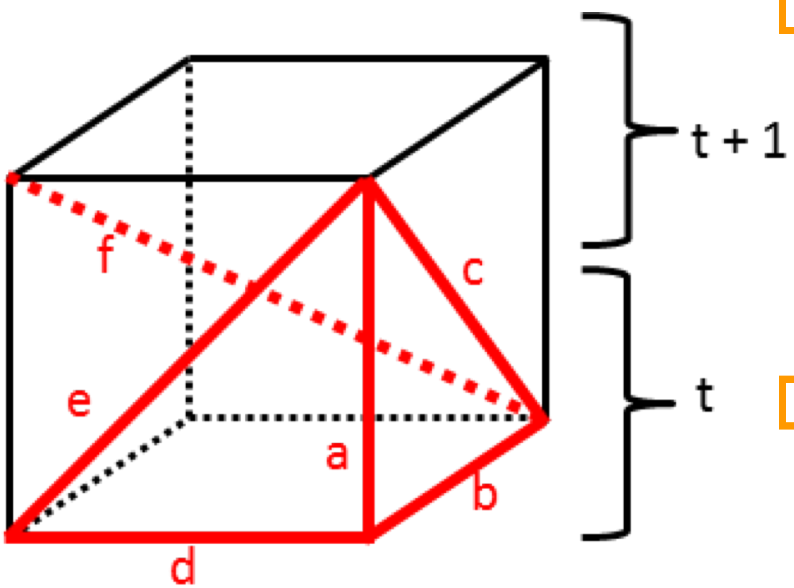
(swap data and ancilla)



The Standard and Heralded Leakage (HL) Decoders

- **Standard Decoder** only relies on syndrome history to decode errors
- **HL Decoder** uses leakage detection when qubits are measured
  - Partial information about leakage locations
  - Error decoder must be modified

Standard Decoder for the Toric Code



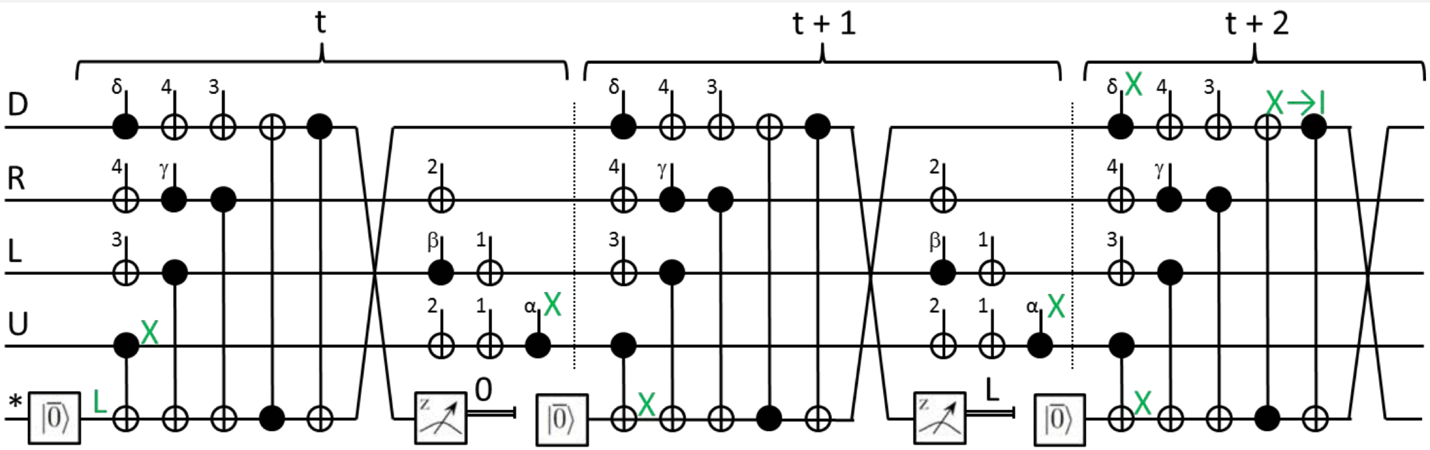
- Decoding graphs for X and Z errors built up using this unit cell

  (Fowler 2011)

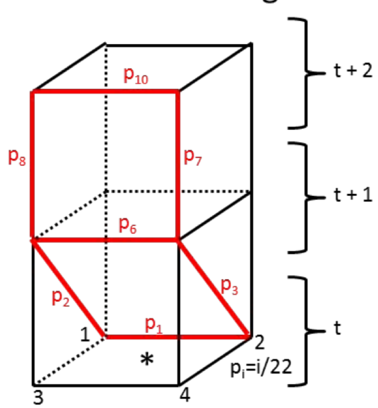
- Corrects error chains between pairs of matched syndromes

- Need to adjust edge weights for each leakage suppressing circuit (Full-LRU, Partial-LRU, Quick circuit)

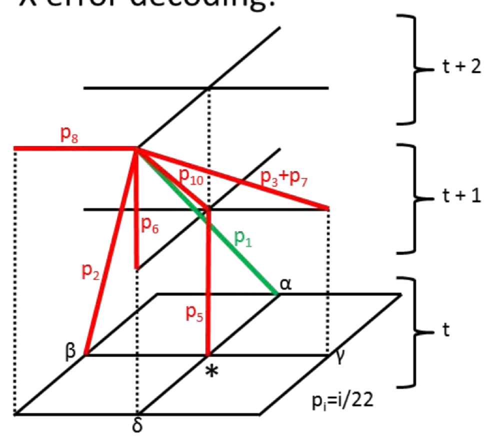
HL Decoder – Quick Circuit

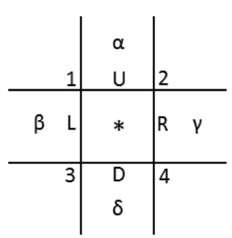


Z error decoding:

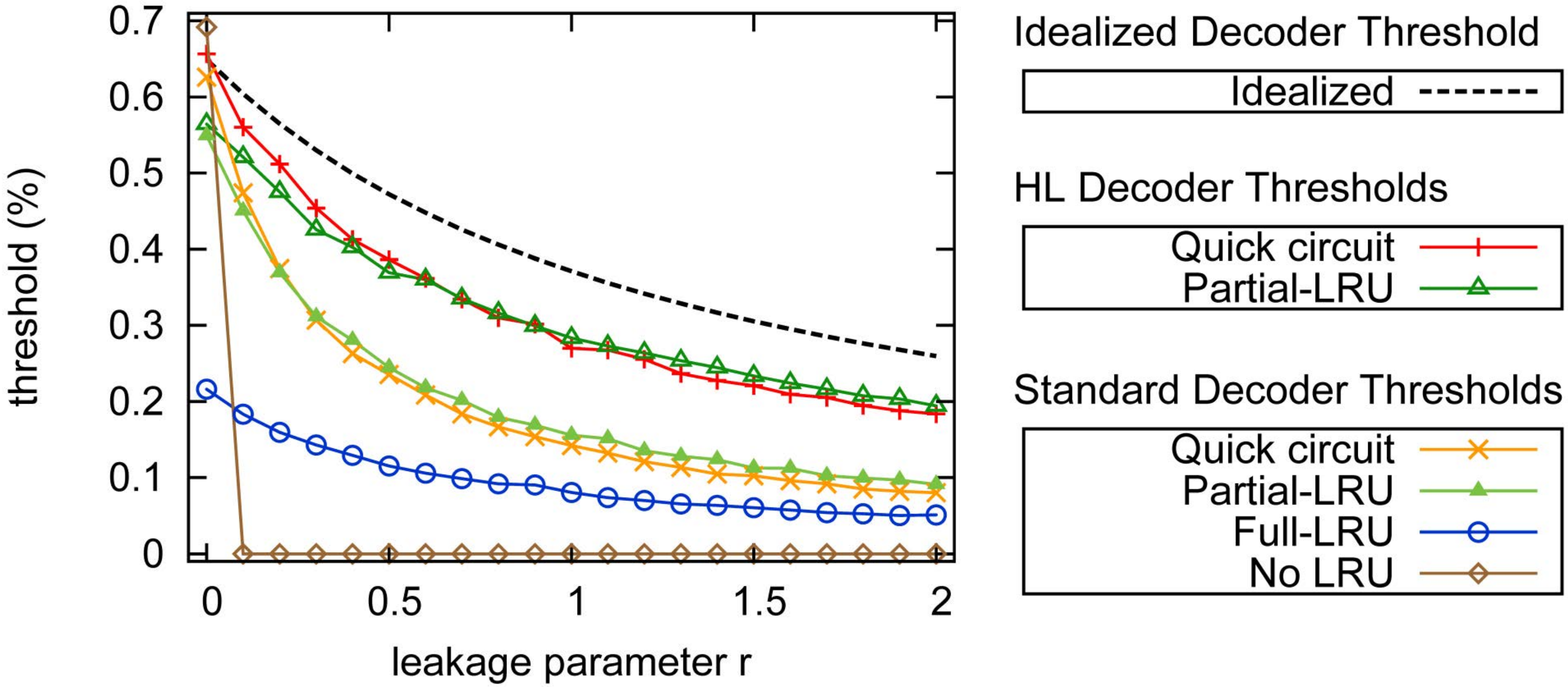


X error decoding:



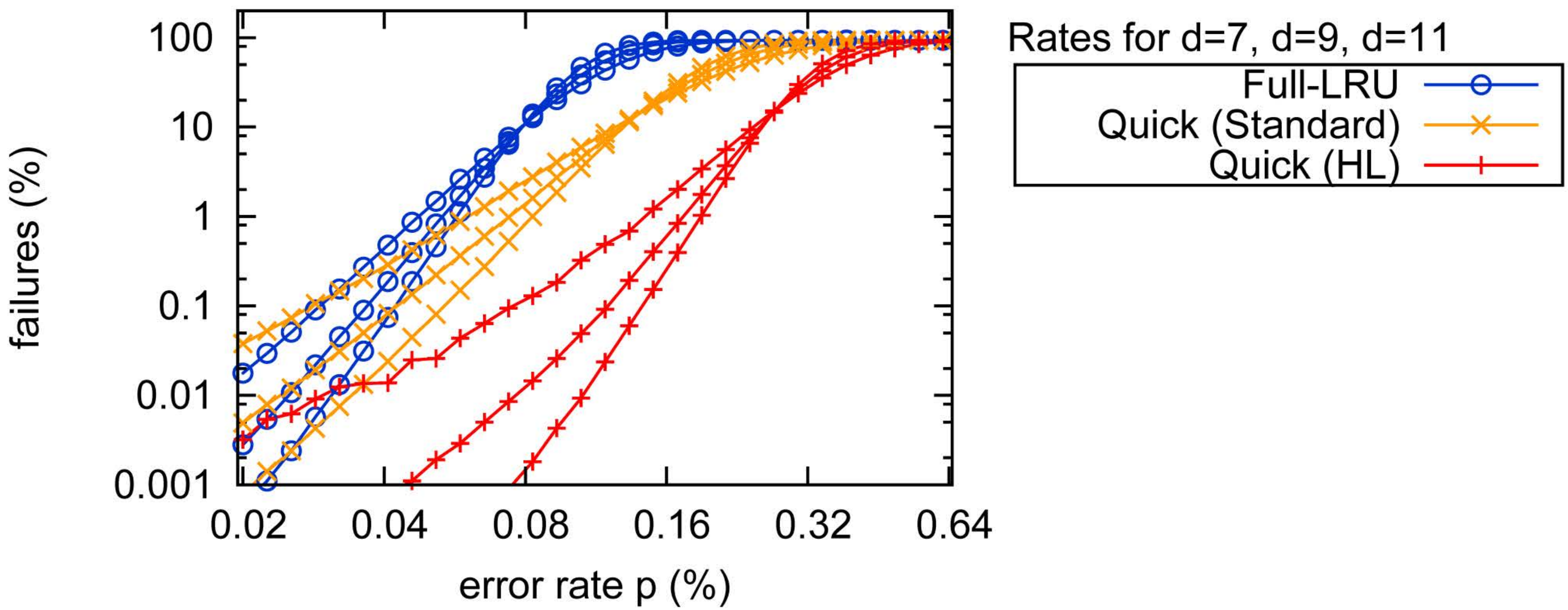


Threshold Comparison



- More complicated circuits have lower threshold
- HL decoder helps boost the threshold

Decoding Failure Rates



- Full-LRU performs well at low error rates

Leakage reduction is necessary

A model of leakage and systematic exploration of parameter space

A simple leakage reduction circuit that only adds a single CNOT gate and new decoders are effective

Overview

I. How much resources needed by quantum computing?

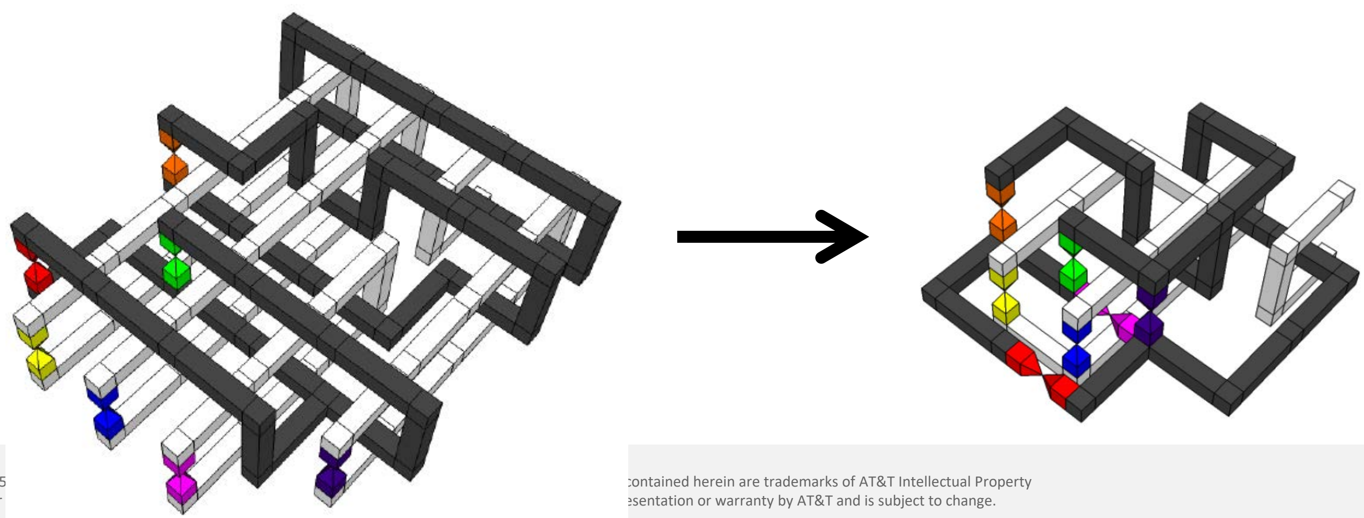
II. Improved error decoding – MLE algorithms

III. New types of errors – correcting qubit leakage

IV. Future directions

Improving the Efficiency of Computation with the Surface Code

- Implementation of CNOT gates by braiding of smooth and rough holes in the surface
  - A few known transformation rules can make the volume of the code smaller
  - Complete set of rules and their optimal use?



Decreasing the Overhead of T Gates

- T gates are very expensive
- New codes with transversal T gates
  - No go theorem of Bravyi and Konig doesn’t apply to subsystem codes
  - Are there 2-D subsystem codes with transversal T gates and good enough error suppression?

Thank You!