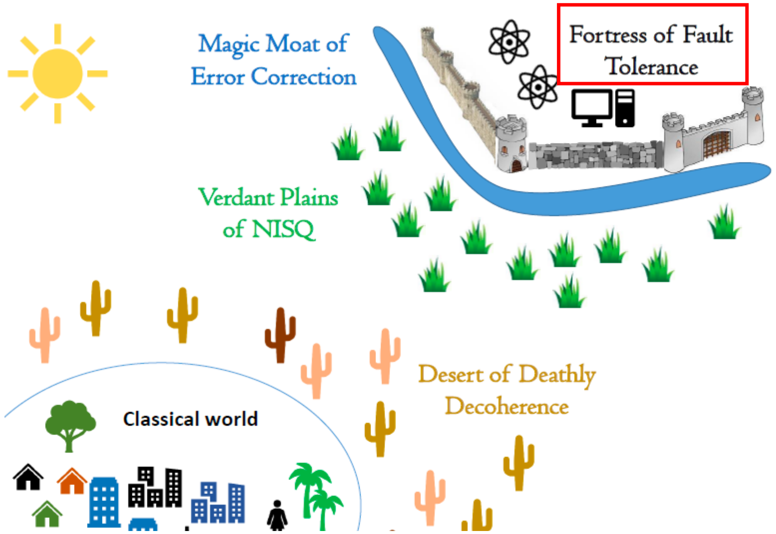


Lecture 9:
Fault tolerance

COMP3366
Quantum algorithms & computing architecture

Instructor: Yuxiang Yang

Department of Computer Science, HKU

Objectives:

- **[O1] Concepts:** Error propagation, logical gates, transversality, magic states, fault-tolerance, threshold theorem.
- **[O2] Problem solving:** Analysis of transversality and error threshold.

Overview of Fault-tolerance

- In the last lecture, we assumed that the error correction circuit (encoding, syndrome detection, …) does not have error, and we didn't discuss gates on the logical qubit.

- The goal of **fault-tolerance** is to enable reliable quantum computations, even when the computer's elementary components are imperfect.
- Compared to the last lecture, we have new issues:

1. Every operation in the circuit, including those involved in the error-correction, may introduce additional errors!
2. Errors may propagate and "replicate" themselves in the circuit!

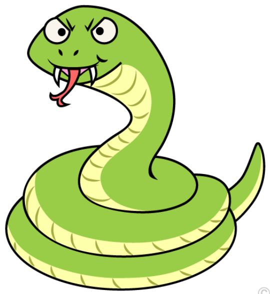
Error correction vs fault-tolerant computations

- What we did last lecture: we focused on a communication scenario, where quantum error correction codes are used to protect information during its transmission.

$b = 1$

Alice

$|1\rangle \longrightarrow X|1\rangle \longrightarrow |0\rangle$



$b = 0?$

Bob

- What we need: To perform computations in the presence of errors!

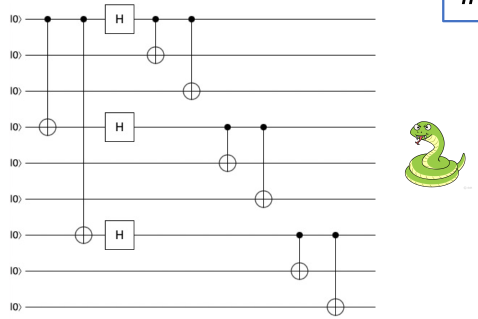
Error models in comparison

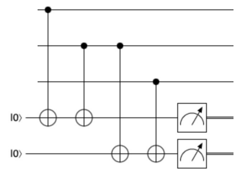
- The error model in the last lecture:

Error-free encoding

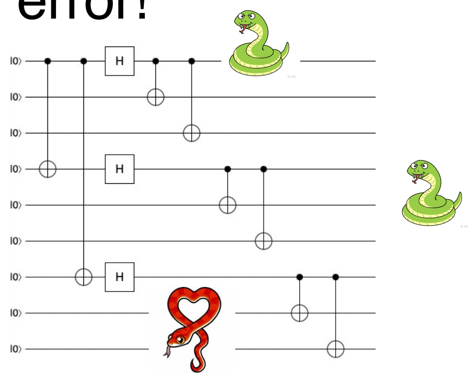
Possibly some errors in the physical qubits

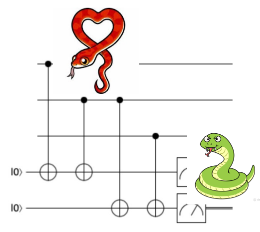
Error-free syndrome detection





- The much more realistic error model in this lecture:
  Everything you do, gates (including identity gates) and measurements, have a (small) chance of creating an error!

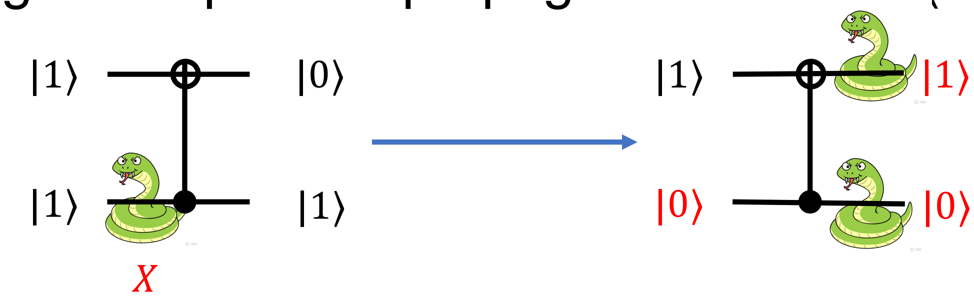




Part I:
Error propagation and transversal implementation of logical gates

Error propagation

- Under this realistic error model, a big issue is that any existing error in some of the qubits may propagate to other qubits!
- Example: a single bit flip X can propagate via CNOT ($CNOT_{1,2}X_1 = X_1X_2CNOT_{1,2}$).



- If we further apply CNOTs, the errors will keep reproducing …
- The (fatal) error propagation problem:
  One single error may eventually ruin the whole computation!
- The effect of error propagation depends heavily on the implementation of logical gates. If the logical gates are implemented smartly, error propagation can be mitigated.

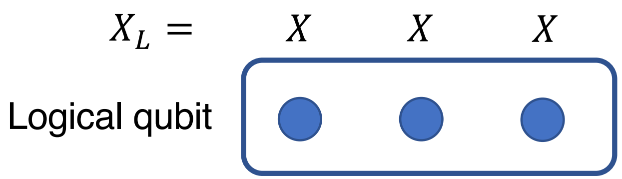
Implementing logical gates

- In a quantum computation protected by a QEC code, we define everything with respect to the logical qubit $|0_L\rangle, |1_L\rangle$.
- The logical gates can be defined as gates in this basis (and different codes have different logical gates). For example, $X_L = |0_L\rangle\langle 1_L| + |1_L\rangle\langle 0_L|$.
- The way to implement logical gates is not unique.
- For example, for the repetition code $|0/1_L\rangle = |000/111\rangle$, we require X_L to act as $|0_L\rangle\langle 1_L| + |1_L\rangle\langle 0_L| = |000\rangle\langle 111| + |111\rangle\langle 000|$ on the logical subspace.
- Alternatively, the gate $X_1X_2X_3$ has the same action as $|0_L\rangle\langle 1_L| + |1_L\rangle\langle 0_L|$ on any codeword $\alpha|0_L\rangle + \beta|1_L\rangle$.
  (although they act differently on other states).
- In this sense, we can denote this implementation as

$$X_L = X_1X_2X_3.$$

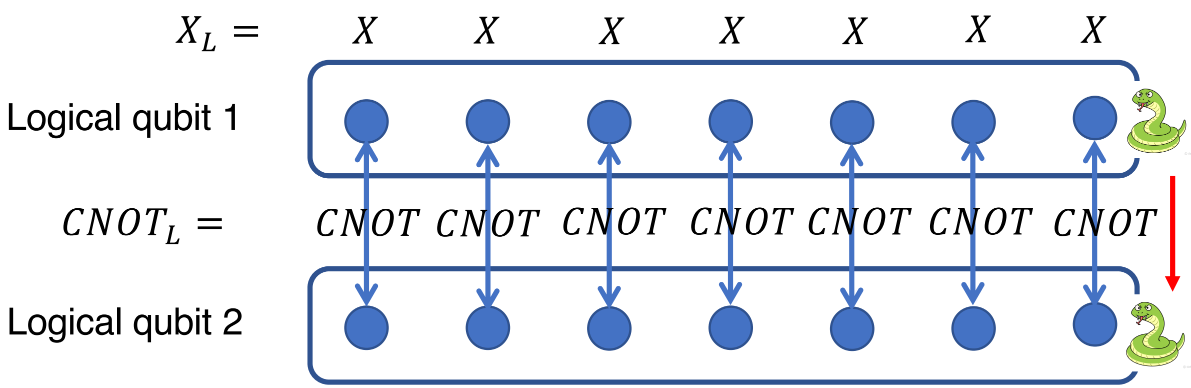
Transversal operations

- Question: How to deal with error propagation?
- Proposed solution: To do the logical operations transversally:
  1. The single-qubit logical operation U_L is done by physically doing $V_U \otimes V_U \otimes \cdots \otimes V_U$ (for some single-qubit gate V_U) on the physical qubits.
  2. The logical CNOT is done by doing one CNOT between 2 paired physical qubits, each from each logical qubit.
- For example, for the repetition code $|0/1_L\rangle = |000/111\rangle$:
- Not transversal $X_L = |000\rangle\langle 111| + |111\rangle\langle 000| + (I - |000\rangle\langle 000| - |111\rangle\langle 111|)$.
- Transversal $X_L = X_1 X_2 X_3$.



Benefits of transversality

- What we want: all logical gates (= single-qubit gates + CNOTs) to be done transversally.
- When a single-qubit error happens:
  1. The single-qubit logical operation U_L will not cause error propagation.
  2. The logical CNOT will propagate the error, but only to one more qubit in the other logical qubit. Then we have only one single-qubit error in each logical qubit, and we can correct that (since the code can correct any single-qubit error!)



- What we don't want: If one logical gate involves 2+ physical qubits in the same block, we may not be able to correct! (avoided by transversality)

The Steane code

- Let us consider the Steane code's transversality.
- Recall that the Steane code is defined by the encoding:

$$|0_L\rangle = \frac{|0000000\rangle + |1010101\rangle + |0110011\rangle + |1100110\rangle + |0001111\rangle + |1011010\rangle + |0111100\rangle + |1101001\rangle}{2\sqrt{2}}$$

$$|1_L\rangle = \frac{|1111111\rangle + |0101010\rangle + |1001100\rangle + |0011001\rangle + |1110000\rangle + |0100101\rangle + |1000011\rangle + |0010110\rangle}{2\sqrt{2}}$$

- It has 7 physical qubits, 1 logical qubit, and correct 1-qubit errors.
  Its syndrome detection is defined by the following observables:

	1	2	3	4	5	6	7
S_1				Z	Z	Z	Z
S_2		Z	Z			Z	Z
S_3	Z		Z		Z		Z
S_4				X	X	X	X
S_5		X	X			X	X
S_6	X		X		X		X

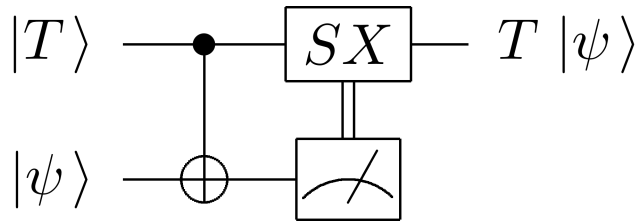
Logical operation of the 7-qubit code

- Let us check out if the 7-qubit code satisfies transversality.
- Since $\{CNOT, H, T\}$ are universal, it is enough to verify transversality for them!
- $|0_L\rangle = \frac{|0000000\rangle+|1010101\rangle+|0110011\rangle+|1100110\rangle+|0001111\rangle+|1011010\rangle+|0111100\rangle+|1101001\rangle}{2\sqrt{2}}$
- $|1_L\rangle = \frac{|1111111\rangle+|0101010\rangle+|1001100\rangle+|0011001\rangle+|1110000\rangle+|0100101\rangle+|1000011\rangle+|0010110\rangle}{2\sqrt{2}}$
- In the 7-qubit code:
  1. The logical Hadamard $H_L = H \otimes H \otimes H \otimes H \otimes H \otimes H \otimes H$.
  2. The logical CNOT = CNOT on each pair of qubits.
  3. But $(T \otimes T \otimes T \otimes T \otimes T \otimes T \otimes T)|0_L\rangle \neq |0_L\rangle$!

Exercise: Verify these 3 points.

Logical operation of the 7-qubit code

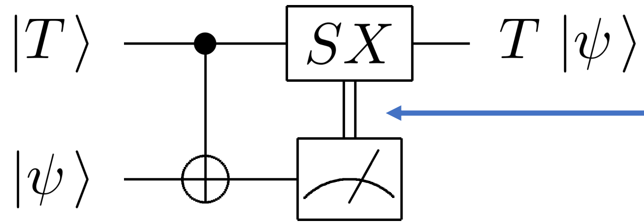
- In fact, this is not a coincidence:
- (Eastin-Knill no go theorem)
  There is no quantum error correction code (capable of correcting any single-qubit error) that implements every gate transversally.
- It seems that, between error correctability and transversality, we can only pick one.
- Luckily, there is a magic remedy to this issue …



Fault-tolerant T gate with magic

Exercise: Verify that this circuit indeed outputs $T|\psi\rangle$

- For any $|\psi\rangle$, the circuit below generates $T|\psi\rangle$:



Apply SX (I) when measurement outcome = 1 (= 0)

- Here $|T\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\frac{\pi}{4}}|1\rangle)$ is called the magic state and $S = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{2}} \end{pmatrix} = T^2$.
- What doe this tell us?
- We can implement T fault-tolerantly, if the above circuit can be done with FT. $CNOT$ and X are transversal, so it remains to check:
    1. The S gate can be done transversally.
    2. The measurement can be made robust against single-qubit errors.
    3. The magic state $|T\rangle$ can be generated in a way robust against single-qubit errors.

Fault-tolerant S gate

- For the S gate, it can be implemented with FT if we find a transversal version.
- Question: Is $S_L = S_1 S_2 \cdots S_7$?
- $|0_L\rangle = \frac{|0000000\rangle + |1010101\rangle + |0110011\rangle + |1100110\rangle + |0001111\rangle + |1011010\rangle + |0111100\rangle + |1101001\rangle}{2\sqrt{2}}$
- $|1_L\rangle = \frac{|1111111\rangle + |0101010\rangle + |1001100\rangle + |0011001\rangle + |1110000\rangle + |0100101\rangle + |1000011\rangle + |0010110\rangle}{2\sqrt{2}}$
- No:

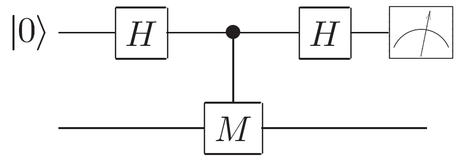
$$S_1 S_2 \cdots S_7 |0_L\rangle = |0_L\rangle$$
$$S_1 S_2 \cdots S_7 |1_L\rangle = -i|1_L\rangle$$

$$\boxed{S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} = T^2}$$

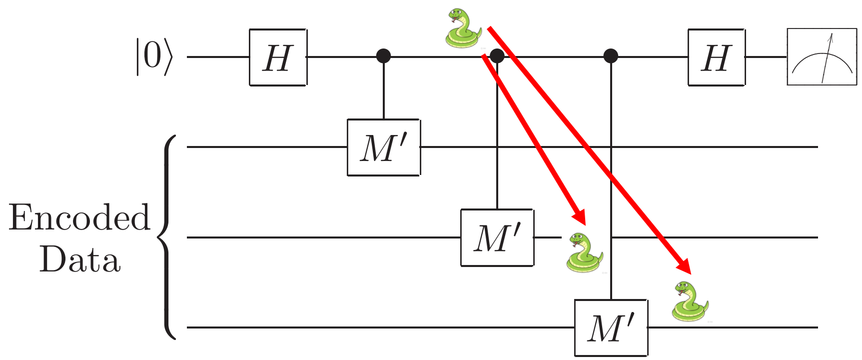
  How to get rid of the minus sign?
- Answer: Try $S_L = (Z_1 S_1)\cdots(Z_7 S_7)$!
- Conclusion: S_L can be done transversally (and thus fault-tolerantly)!

Fault-tolerant measurements

- The following is the Hadamard test for measuring a general unitary observable M:



- Question: Is it FT? (i.e., can error propagation in the circuit be stopped?)
- No! When the data is encoded in a QEC code:

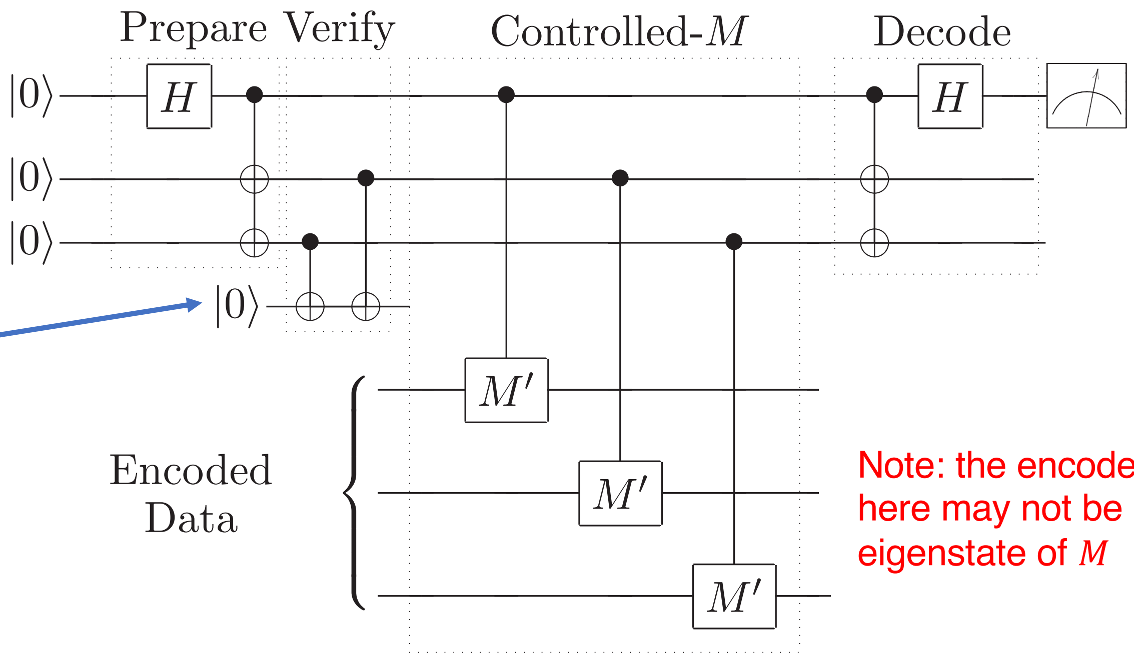


- An error in the ancilla will affect all encode qubits (and thus creating uncorrectable error)!

Build fault-tolerant measurements* (bonus material)

- To make the measurement of M FT, we prepare the ancilla using the repetition code:

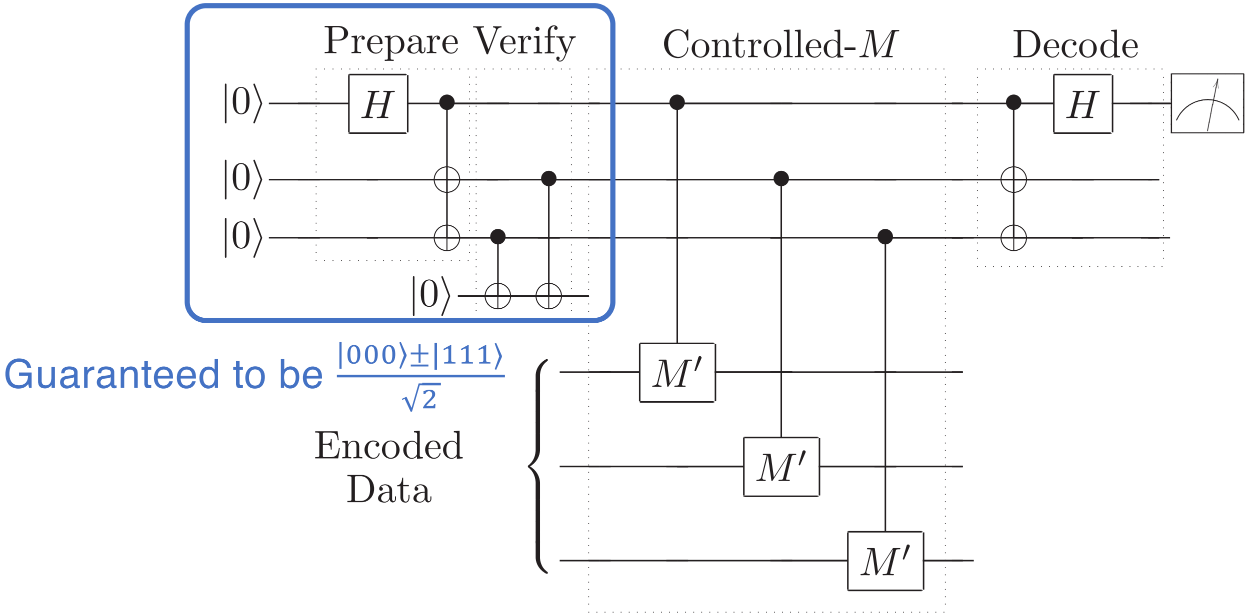
parity checks Z_1Z_2, Z_1Z_3, Z_2Z_3 to verify that the ancilla is in $\frac{|000\rangle+|111\rangle}{\sqrt{2}}$ or $\frac{|000\rangle-|111\rangle}{\sqrt{2}}$ (when there is a single Z error)



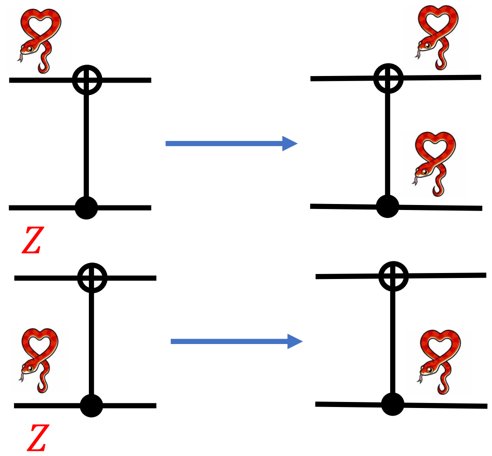
Note: the encoded data here may not be an eigenstate of M

- If a parity check fails (i.e., yielding the outcome 1), we start over from the beginning.

- Our goal: to show that the below circuit is robust against any single-qubit error in the measurement gadget or in any block of the encoded data.



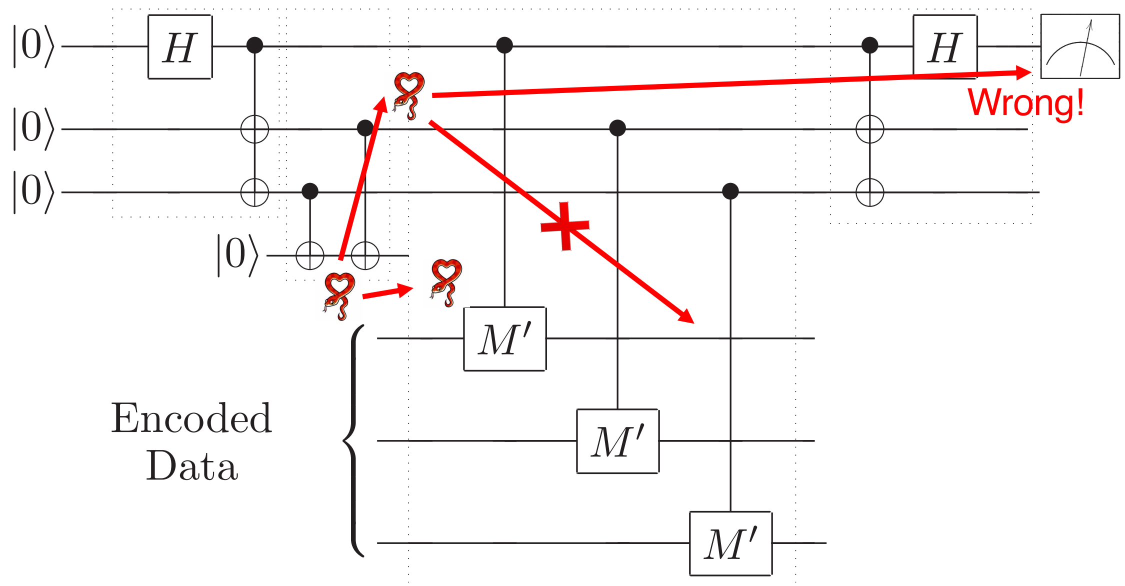
Exercise: Verify the error propagation relation below ($(Z \otimes Z)CNOT = CNOT(I \otimes Z)$):



- For this, we need to verify its robustness against single-qubit X error and Z error.
- Initial state preparation, with the help of parity verification, is already shown to be robust. We need to check the remaining part of the circuit.

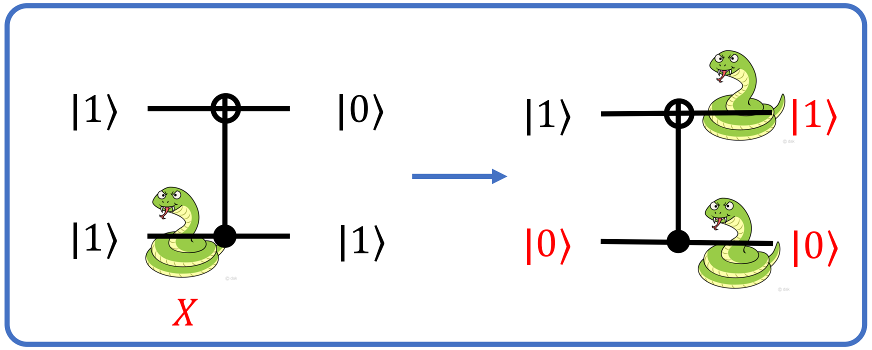
- If a Z error occurs in the extra qubit, it might propagate to the ancillary qubits but not to the encoded data:

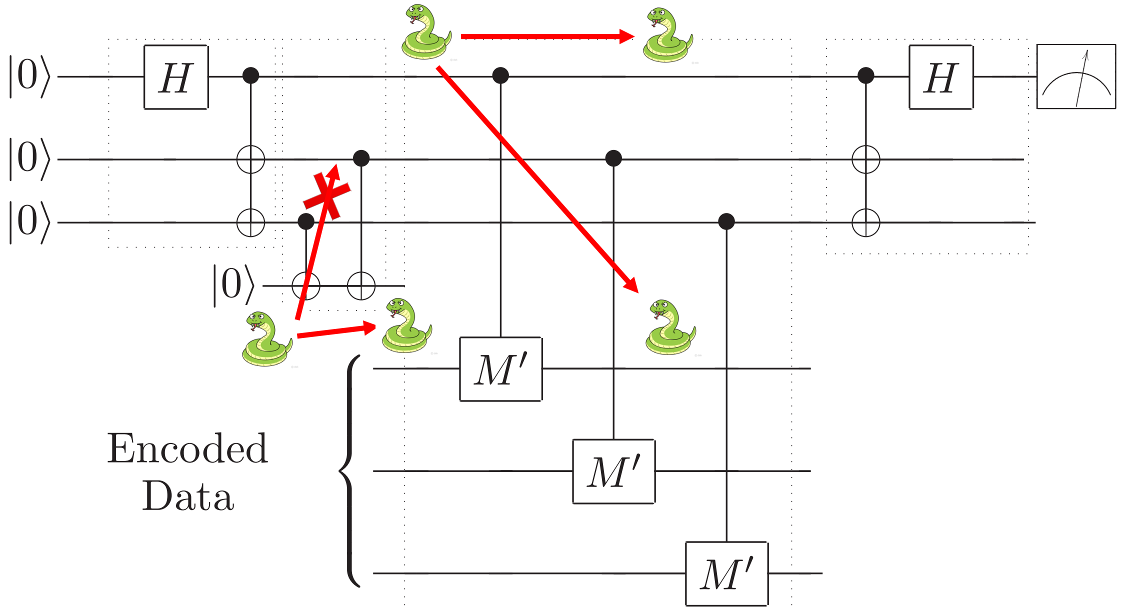
$$\frac{|000\rangle + |111\rangle}{\sqrt{2}} \rightarrow_Z \frac{|000\rangle - |111\rangle}{\sqrt{2}}$$



- Any Z error in the ancilla won't propagate to the encoded data.
- The only effect is that the Hadamard test might yield wrong outcome.
- Therefore, we can resolve it by repeating the Hadamard test for 3 times and make majority vote (e.g., $010 \rightarrow 0$).

- If a X error occurs to any of the first 3 qubits, it might propagate to the encoded data:





- But it is still fine! Because there are only 1 (propagated) error in the encoded data. If the encoded data is protected by a QEC code, the single-qubit error can be corrected.
- An X error in the additional qubit will not propagate, so this is also fine.

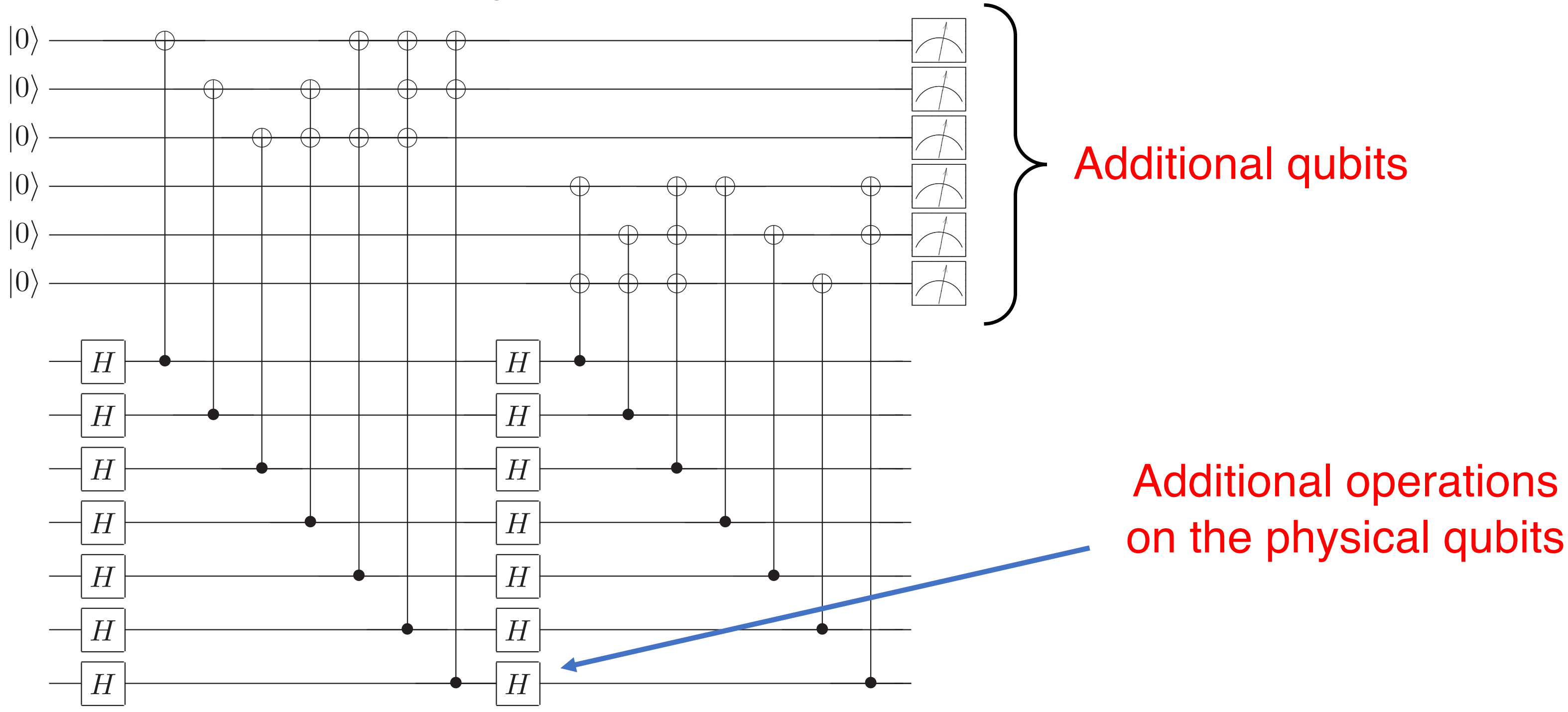
Fault-tolerant preparation of $|T\rangle$

- How do we get $|T\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i\frac{\pi}{4}}|1\rangle\right)$ with FT?
- Fact: $|T\rangle$ is the eigenstate of $e^{-\frac{i\pi}{4}}SX = \begin{pmatrix} 0 & e^{-i\frac{\pi}{4}} \\ e^{i\frac{\pi}{4}} & 0 \end{pmatrix}$!
- We have $|0\rangle = \frac{1}{\sqrt{2}}(|T\rangle + |T^{\perp}\rangle)$ for $|T^{\perp}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - e^{i\frac{\pi}{4}}|1\rangle\right)$.
- FT preparation of $|T\rangle$:
  1. FT measurement: Measure the observable $e^{-\frac{i\pi}{4}}SX$ (using the previous FT measurement circuit) on the input $|0\rangle$.
  2. If the outcome is $+1$, we get $|T\rangle$.
  3. If the outcome is -1, we either start over or apply (with FT) X to flip $|T^{\perp}\rangle$ into $|T\rangle$
- Conclusion:
  Although T is not transversal, it can also be implemented with FT!

Part II:
Full error analysis for QEC

Two issues with FT

1. Error propagation. → Solution: transversal gates + some magic ✅
2. Error-correction itself may introduce additional errors!



Syndrome measurement for the Steane code

Criterion for error reduction

- Question: when does QEC make the error smaller?

- Claim:

When the original circuit elements have error rate p, the effective error-rate of a FT implementation is

$$C \cdot p^2$$

for some $C > 0$ that depends only on the QEC code (and doesn't reply on p).

- Conclusion: QEC reduces the error, if $C \cdot p^2 < p$!

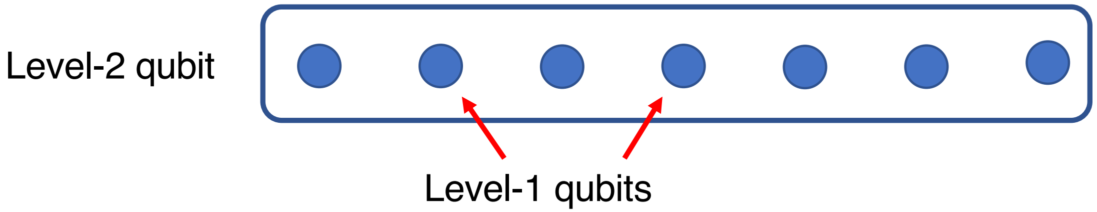
Constant C

- What is C here? How big it is?
- Without QEC, the computation may fail with a single error (w. prob. p).
- With QEC, the computation fails only if at least a pair of errors occur in the same logical qubit (w. prob. p). The number of such pairs is C.
  Note that the circuit involves more physical gates (for encoding, syndrome, recovery etc.), which could lead to a pretty large constant C.
- How large C is?
  It depends on the QEC code …
  For the Steane code, $C \sim 10^4$ (see p479 of Nielson & Chuang's textbook for details).
- A QEC code's performance is not only determined by the ratio $\frac{\text{physical qubits}}{\text{logical qubits}}$.
  It should also have as small C as possible!

Part II:
Threshold theorem

Error rate of the logical qubit

- In the 7-qubit code, we arranged 7 physical qubits, each with an error rate p_1 (i.e., each physical qubit has probability p_1 to go wrong), into a “single” logical qubit:



- As shown previously, the new qubit has error rate
$$p_2 < C \cdot p_1^2$$
for some constant C, which is independent of p_1.
- Observation:
$p_2 < p_1$ if we can make $p_1 < p^* := 1/C$!

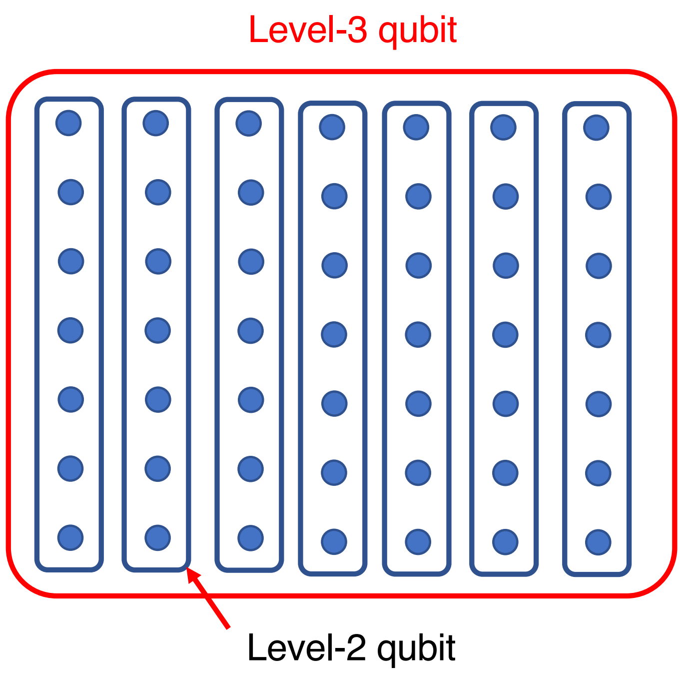
Encoding a few more times!

- Idea:
  Making the error rate even smaller, by applying the 7-qubit encoding one more time and arranging 7 level-2 qubits (and thus 49 level-1 qubits) into a larger qubit (level-3 qubit)!

- Question: The new qubit's error rate?
  Answer:

$$p_3 < p^* * \left(\frac{p_2}{p^*}\right)^2 < p^* \cdot \left(\frac{p_1}{p^*}\right)^4 !$$

- Code concatenation:
  Repeating this procedure for a few more times!



Code concatenation

- For a level-k qubit, its error rate satisfies $p_k < p^* \cdot \left(\frac{p_{k-1}}{p^*}\right)^2$ and thus

$$p_k < p^* \cdot \left(\frac{p_1}{p^*}\right)^{2^{k-1}}.$$

- Error-size tradeoff of concatenated codes: As long as

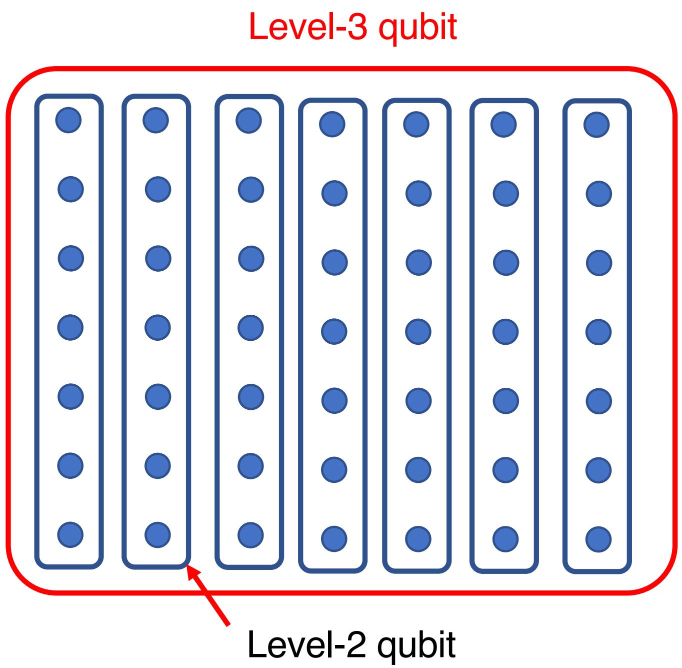
$$p_1 < p^*$$

depends only on the code!

depends only on how good the qubit engineering is!

we can make p_k arbitrarily small!

- As a price to pay, we need -- if we use the 7-qubit code -- 7^{k-1} physical qubits for a single logical qubits!



Threshold theorem

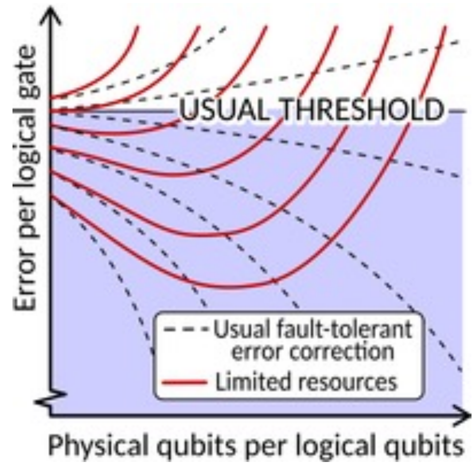
- (**Threshold theorem**)
  There exists an error rate threshold $p^* > 0$ such that any (large) ideal quantum circuit with f gates can be simulated up to an arbitrarily small error ϵ, by a realistic quantum circuit whose element has an error rate $p < p^*$, whose gate number is
  $$f \cdot \mathrm{poly}(\log f).$$
- If the ideal circuit is poly-sized, the implementation is also poly-sized!
- Proof:
  1. By doing a k-concatenation of a n-qubit code, the circuit size grows as $n^k \cdot f \ldots$
  2. … while the error rate goes down double exponentially as $\left(\frac{p}{p^*}\right)^{2^k} \cdot f$
     (= gate error rate * # gates)
  3. To reach an approximation error $\epsilon \sim f \cdot \left(\frac{p}{p^*}\right)^{2^k}$, we need $k \sim \log\log f$ layers of concatenations, and the circuit size grows as $f \cdot n^{\log\log f} = f \cdot \mathrm{poly}(\log f)$.

Assumptions in threshold theorem

- Constant error rate:
  In the threshold theorem, it is assumed that the single-qubit error rate is a constant.

- In practice, this might not be true.
  Depending on the implementation, the single-qubit error rate may increase with the scale of computation or the duration of the computation.

- The threshold theorem can be modified accordingly.

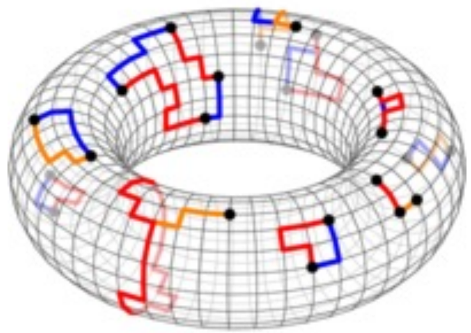


PRX Quantum **2**, 040335

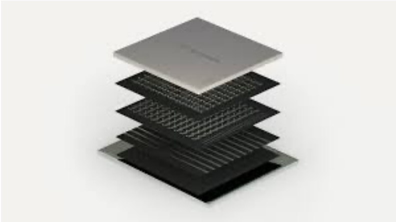
Part III:
How far are we from full FT:
Recent progresses and challenges

Threshold values

- The 7-qubit Steane code:
  $\sim 10^{-5}$ (theoretical lower bound)/ $\sim 10^{-3}$ (tested value)
- The 25 qubit Bacon-Shor code:
  $\sim 2\times10^{-4}$
- Surface codes:
  $\sim 10^{-2}$ (best among known constructions)
- Even in the best case, we need thousands of physical qubits per logical qubit!
- The best quantum computer so far (IBM Eagle) has only 127 physical qubits.
- Within the near-term future, if we do full error correction, there will be only 1-2 logical qubits left for computation!



A surface code



IBM Eagle (2021)

LDPC codes: the most promising for FT

- Quantum low-density parity check (LDPC) codes:
  Codes whose syndrome detection involves only short Pauli strings, say, those with $O(1)$-length, and each qubit is involved only in $O(1)$ stabilizers.
- Benefit: low computational cost for identifying the error.
  $\Rightarrow$ Also potentially smaller C!
- It is believed that all practically useful codes should be LDPC codes.

… I have championed the idea of using high-rate low-density parity check (LDPC) codes for fault tolerance.

Daniel Gottesman,
Pioneer of QEC

The "big picture" of QC, QEC & FT

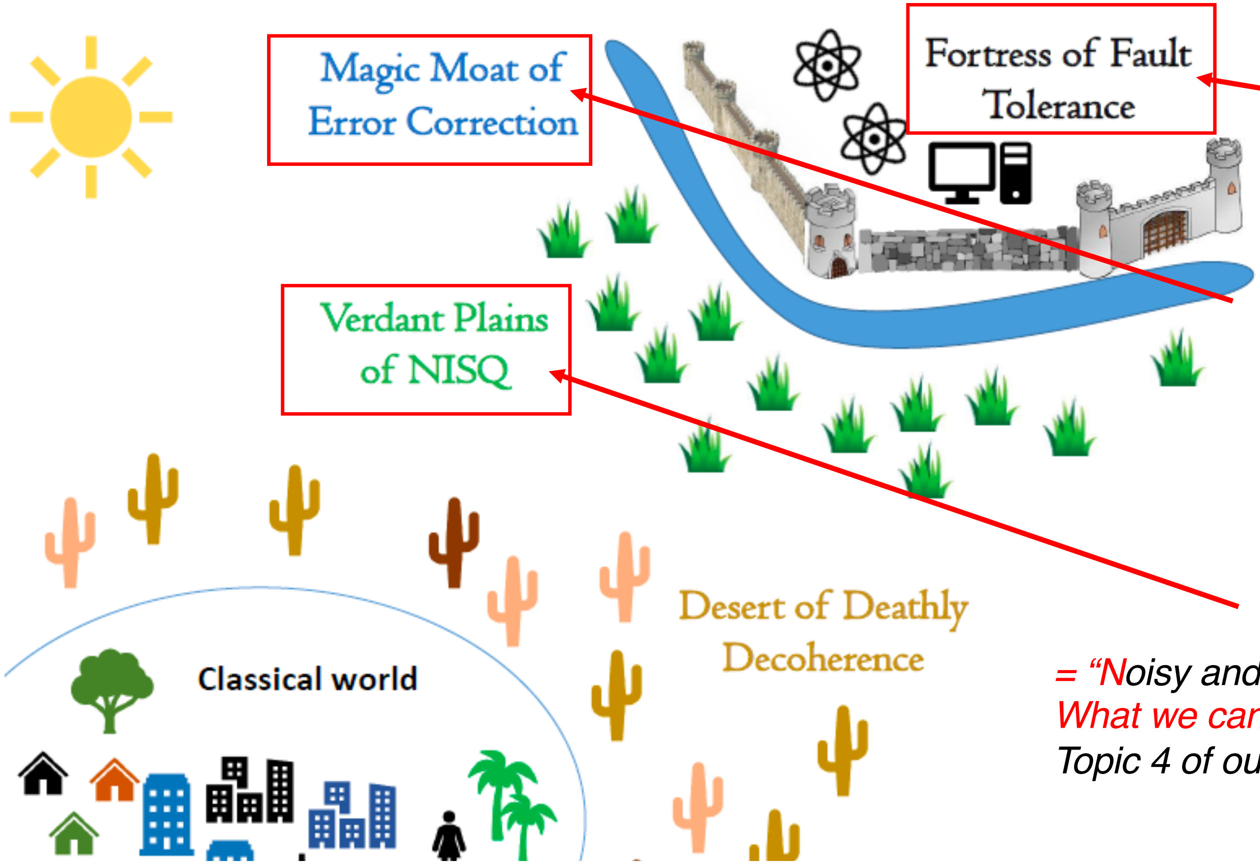


Figure by Gottesman/Munro

The ultimate goal (10+ years);
Pre-requisite of Shor, Grover, HHL in application

Key technique for FT;
The most important research direction of QC at present.

= "Noisy and Intermediate-Scale Quantum"
What we can do now (0-10 years).
Topic 4 of our course

Summary

- Error model for quantum computation, Error propagation, Fault tolerance.

- Realization of fault-tolerance:
  Transversal gates, FT measurements, magic states.

- Threshold theorem

Homework

- Review the lecture slides; you may find the review questions in the next slides helpful.
  Try the exercises in the slides and discuss with your classmates.
- Attempt Q5 in Assignment 3.
- Optional: Read p475-493 of *Quantum Computation and Quantum Information* by Nielsen and Chuang.

Review questions

- What is the difference between physical and logical qubits?
- Try to recover the proof of no-copying theorem by yourself.
- In the syndrome detection for the 3-qubit repetition code, why can't we introduce 3 (instead of 2) ancillary qubits, and do 1 CNOT on each pair of 1 code qubit + 1 ancillary qubit (in total 3 CNOTs)?
  (That is, can't we measure Z_1, Z_2, Z_3?)
- Why must the stabilizers commute with each other? What if they do not?
- To detect errors, we made (syndrome) detection, which extracts information about a quantum state. Why (and under what condition) can we do this without collapsing the state?