Scalable Neural Decoder
for
Topological Surface Codes

Simon Trebst
University of Cologne

APS March Meeting
Chicago, March 2022

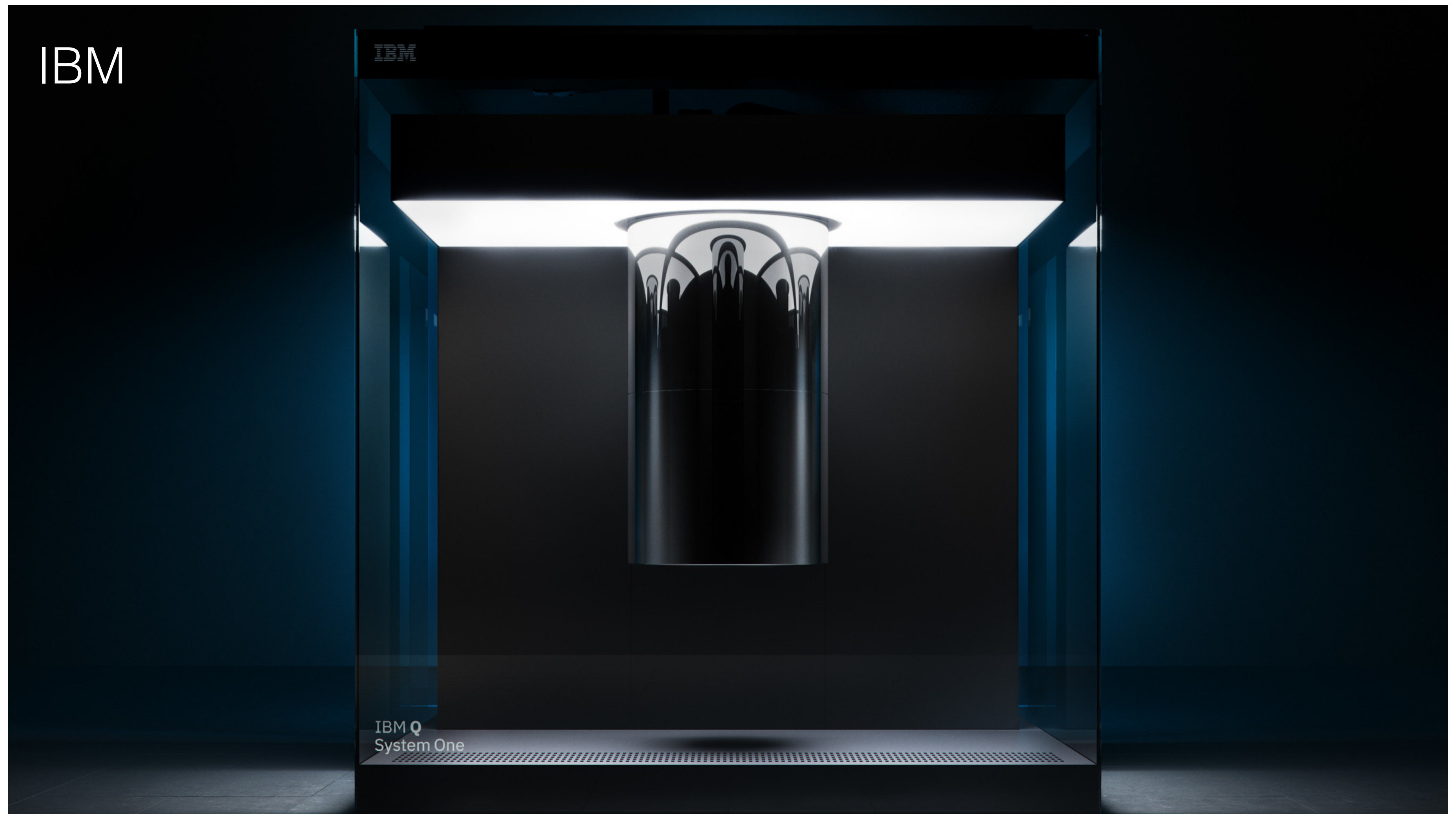
CRC183
ENTANGLED STATES OF MATTER

ML4Q
MATTER AND LIGHT FOR
QUANTUM COMPUTING

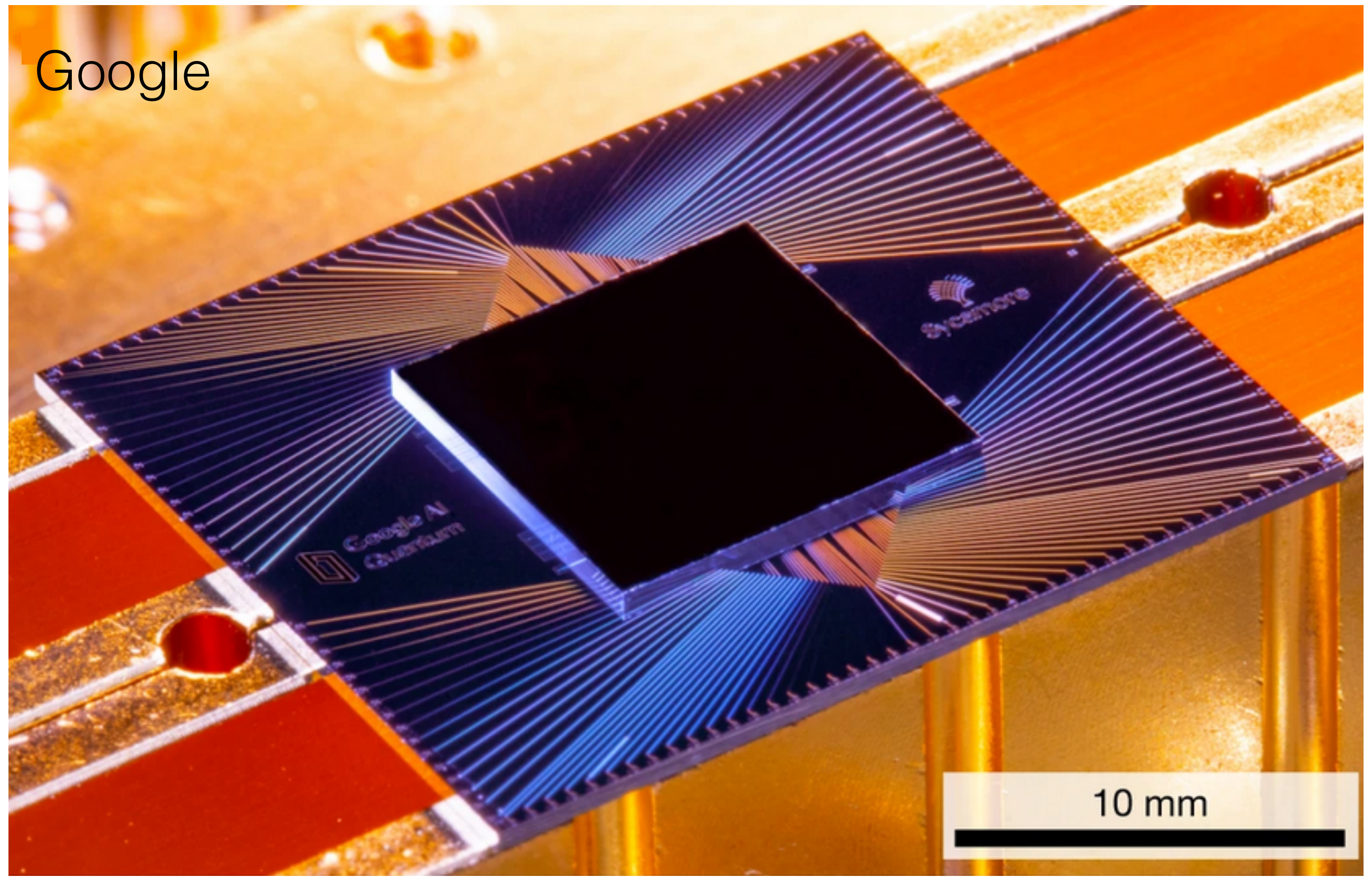
Quantum Computing in the NISQ era

An experimental pivot from of a **few pristine qubits** to the realization of circuit architectures of **50-100 qubits** but tolerating a significant level of **imperfections.**

Noisy intermediate scale quantum devices

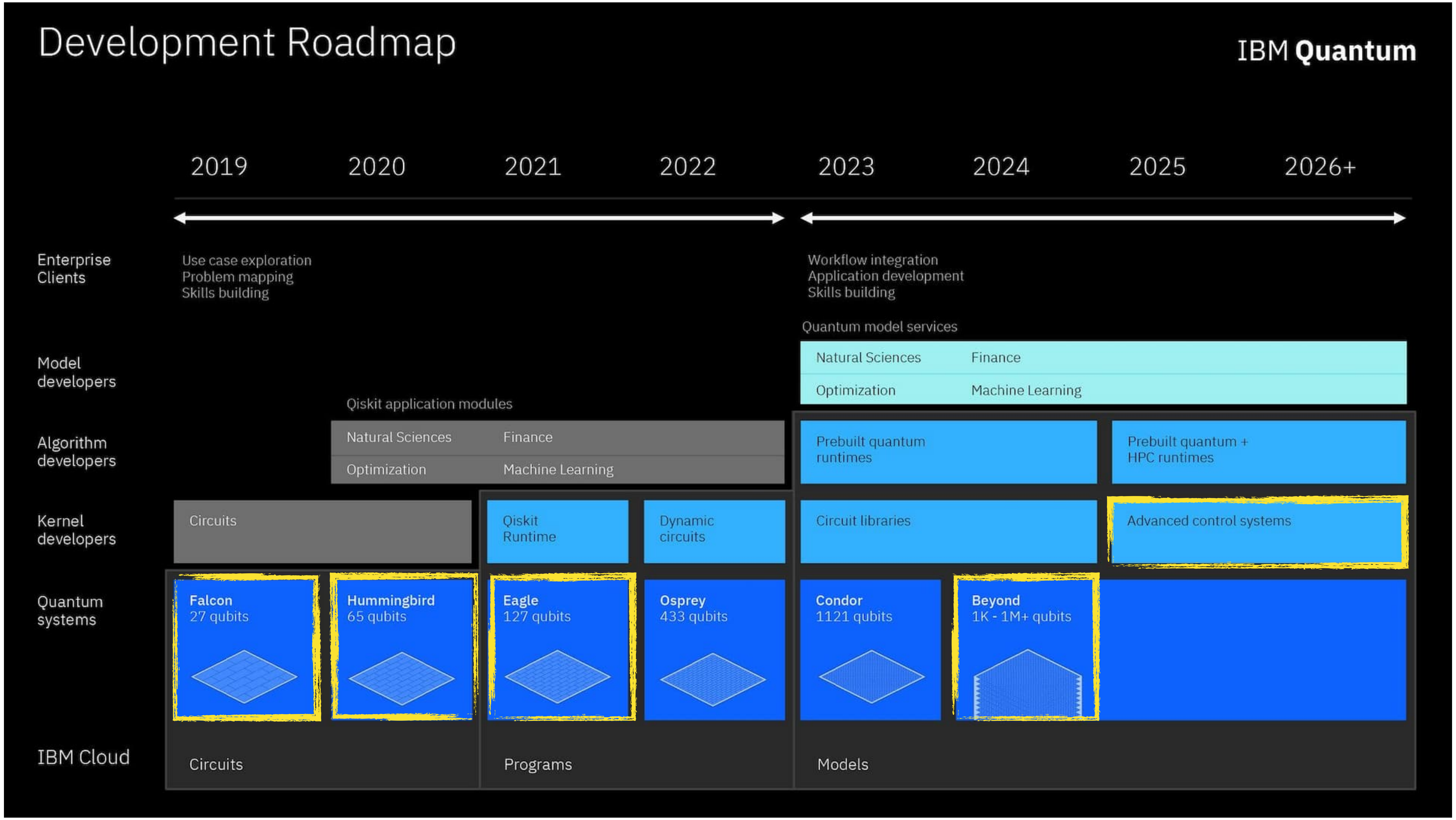


Eagle generation — 127 qubits



Sycamore chip — 53 qubits

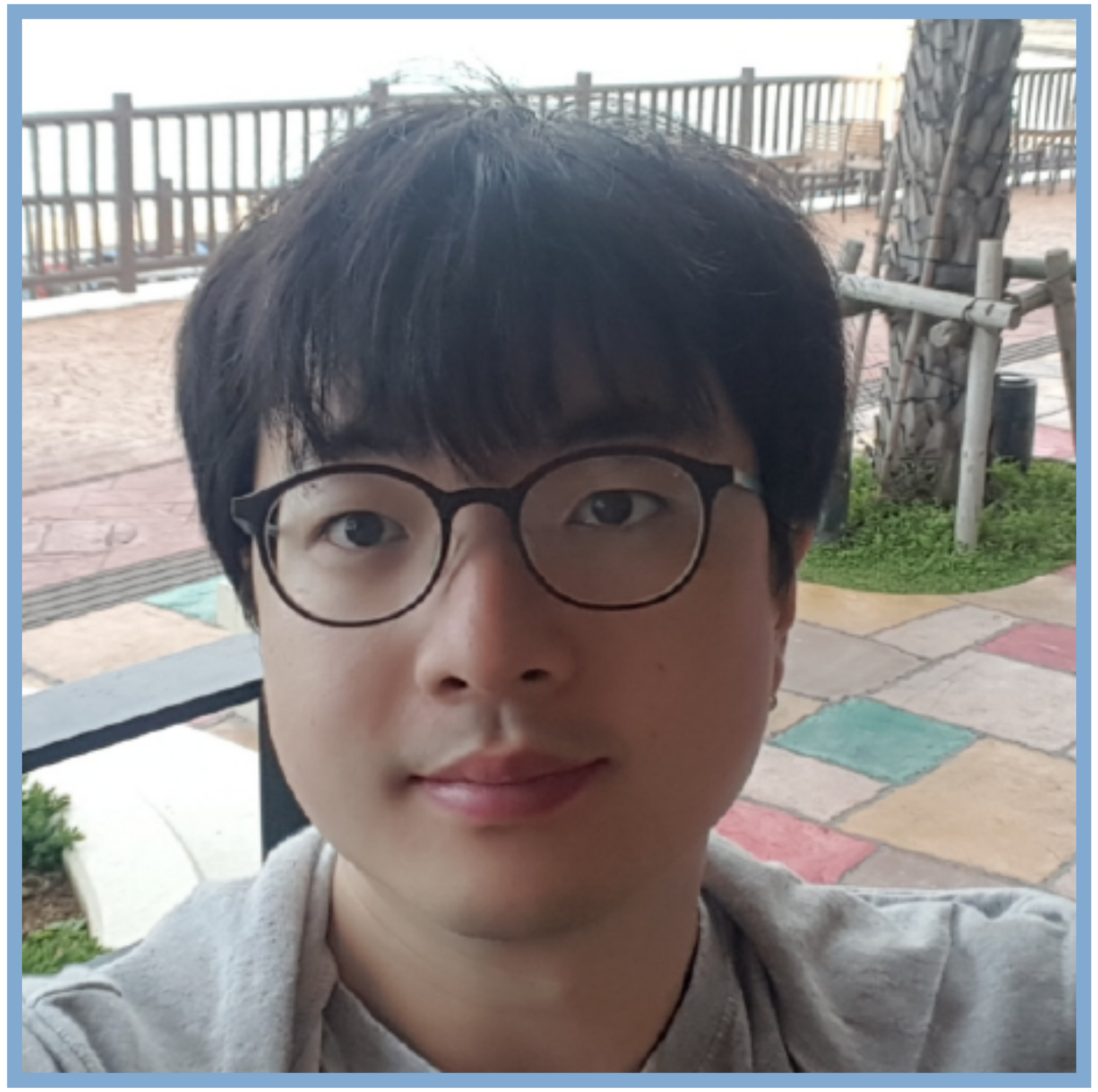
Quantum Computing in the NISQ era



https://research.ibm.com/blog/quantum-development-roadmap

meet the team

Kai Meinerz



Chae-Yeun Park

https://github.com/chaeyeunpark

Kai Meinerz, Chae-Yeun Park & ST,
PRL **128**, 080505 (2022).

quantum
error correction

error correction 101

Error correction in **classical computers** protects against bit flips

$1 \to 0$

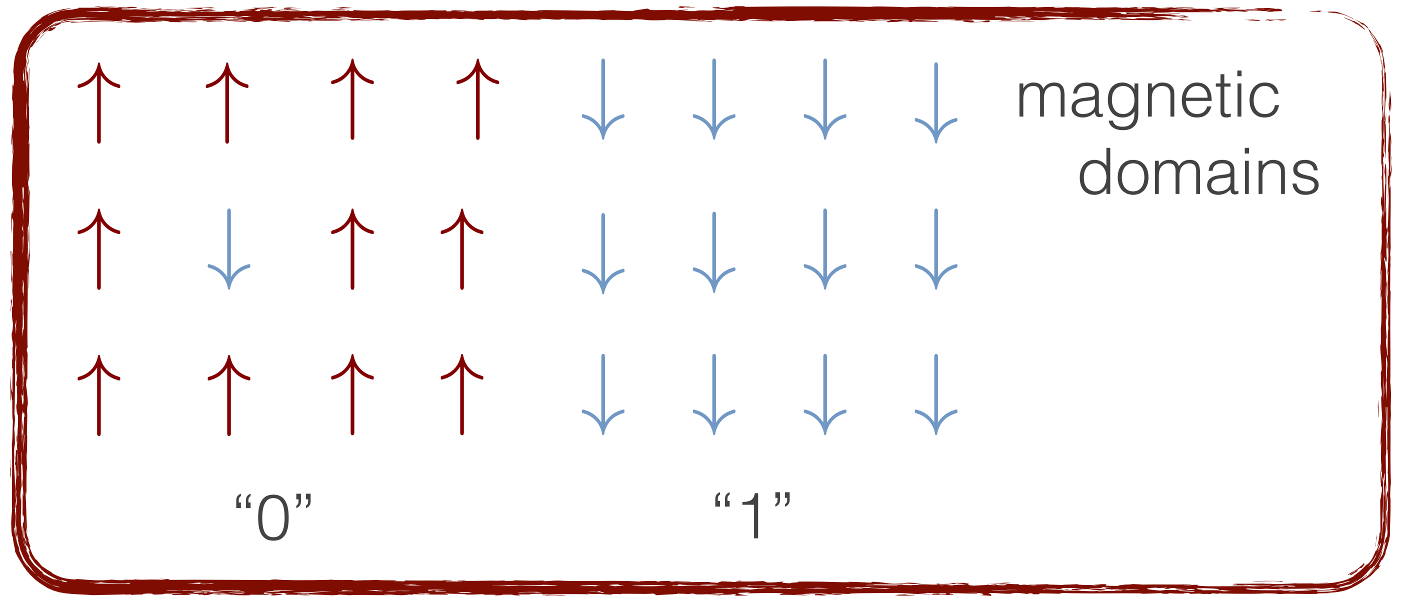
$0 \to 1$

bit flip

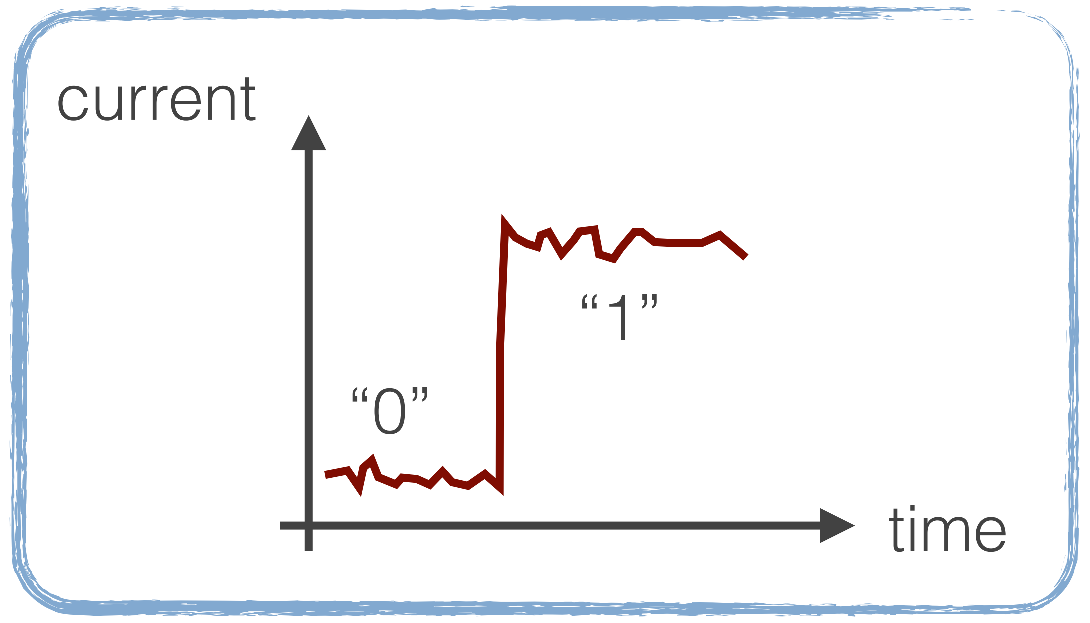
$111 \to 101$

$000 \to 101$

logical fault



principal concept
redundancy



Error correction in **quantum computers** is more complicated

$|1\rangle \to |0\rangle$

bit flip

$\alpha|1\rangle + \beta|0\rangle \quad \to \quad \alpha|1\rangle - \beta|0\rangle$

phase flip

$|\psi\rangle \quad \to \quad U|\psi\rangle$

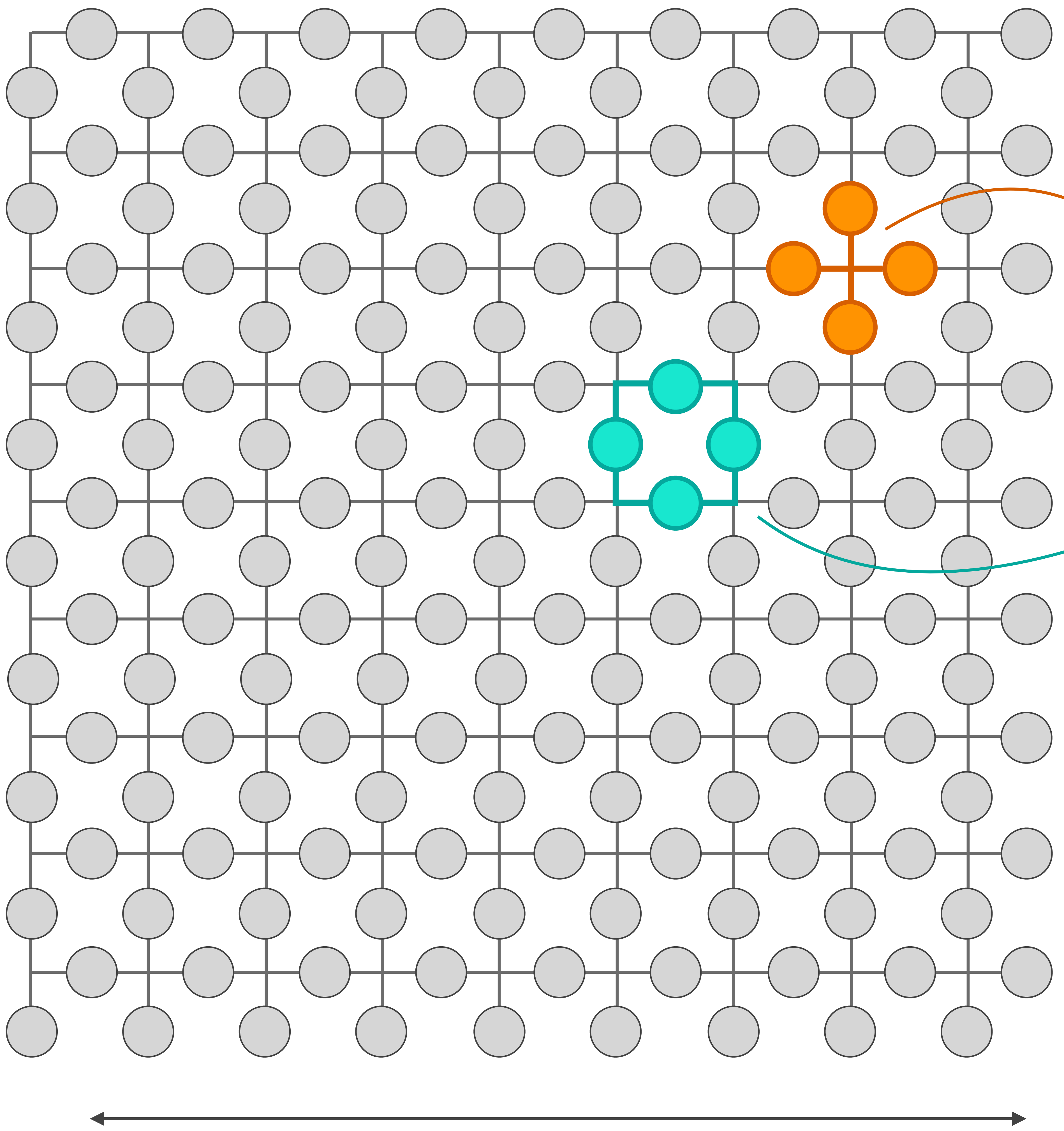
arbitrary unitary

no cloning theorem

$|\psi\rangle$ ~~$|\psi\rangle$~~ ~~$|\psi\rangle$~~

projective measurements disturb the quantum state
→ error detection is **destructive**

stabilizer codes



Kitaev's toric code

$$X_v = \prod_{i \in v} X_i$$

$$Z_p = \prod_{i \in p} Z_i$$

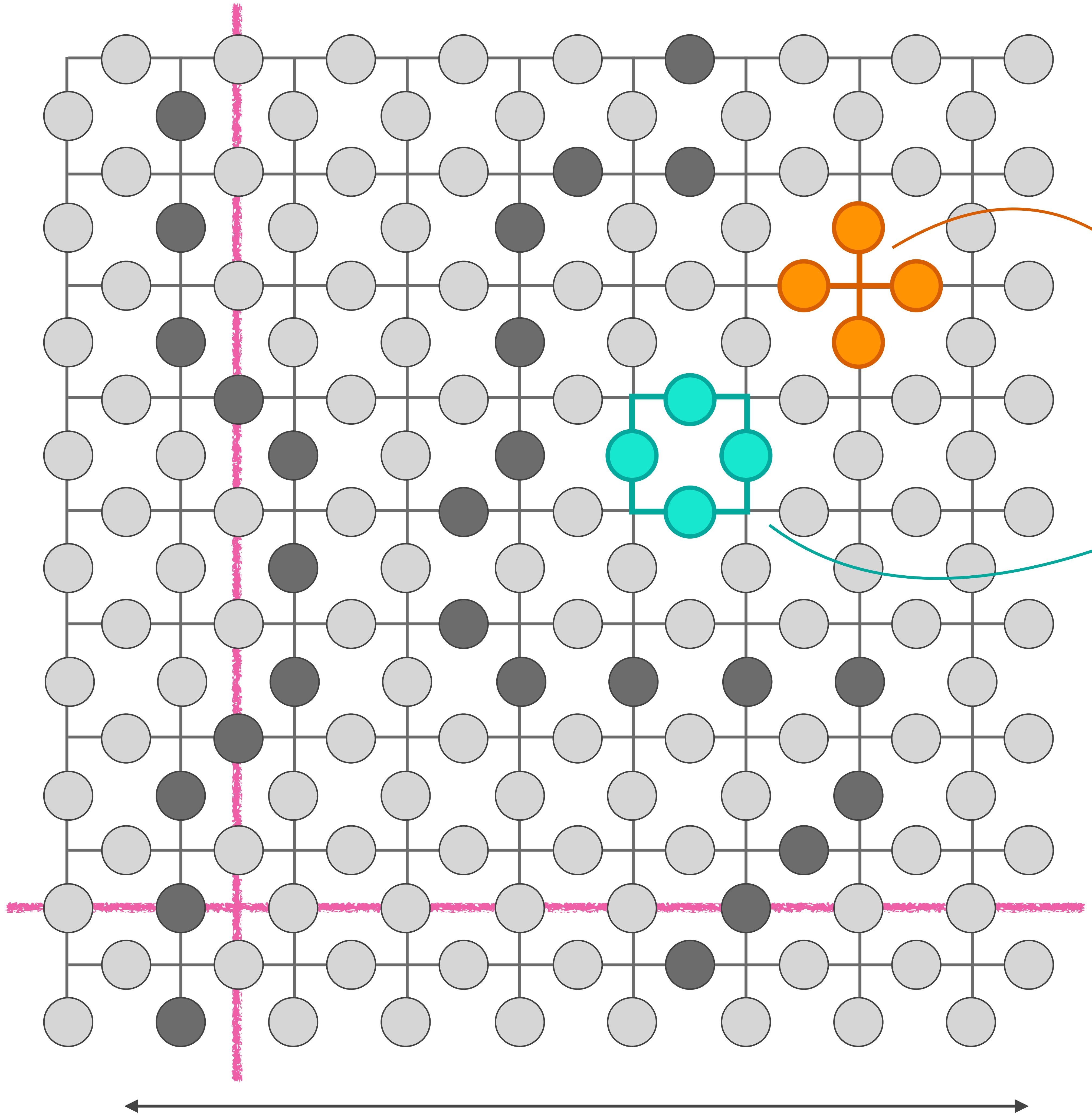
stabilizer operators

ground states

$$X_v|\psi\rangle = 1 \cdot |\psi\rangle \quad \forall v$$

$$Z_p|\psi\rangle = 1 \cdot |\psi\rangle \quad \forall p$$

stabilizer codes



Kitaev's toric code

$$X_v = \prod_{i \in v} X_i$$

$$Z_p = \prod_{i \in p} Z_i$$

stabilizer operators

ground states

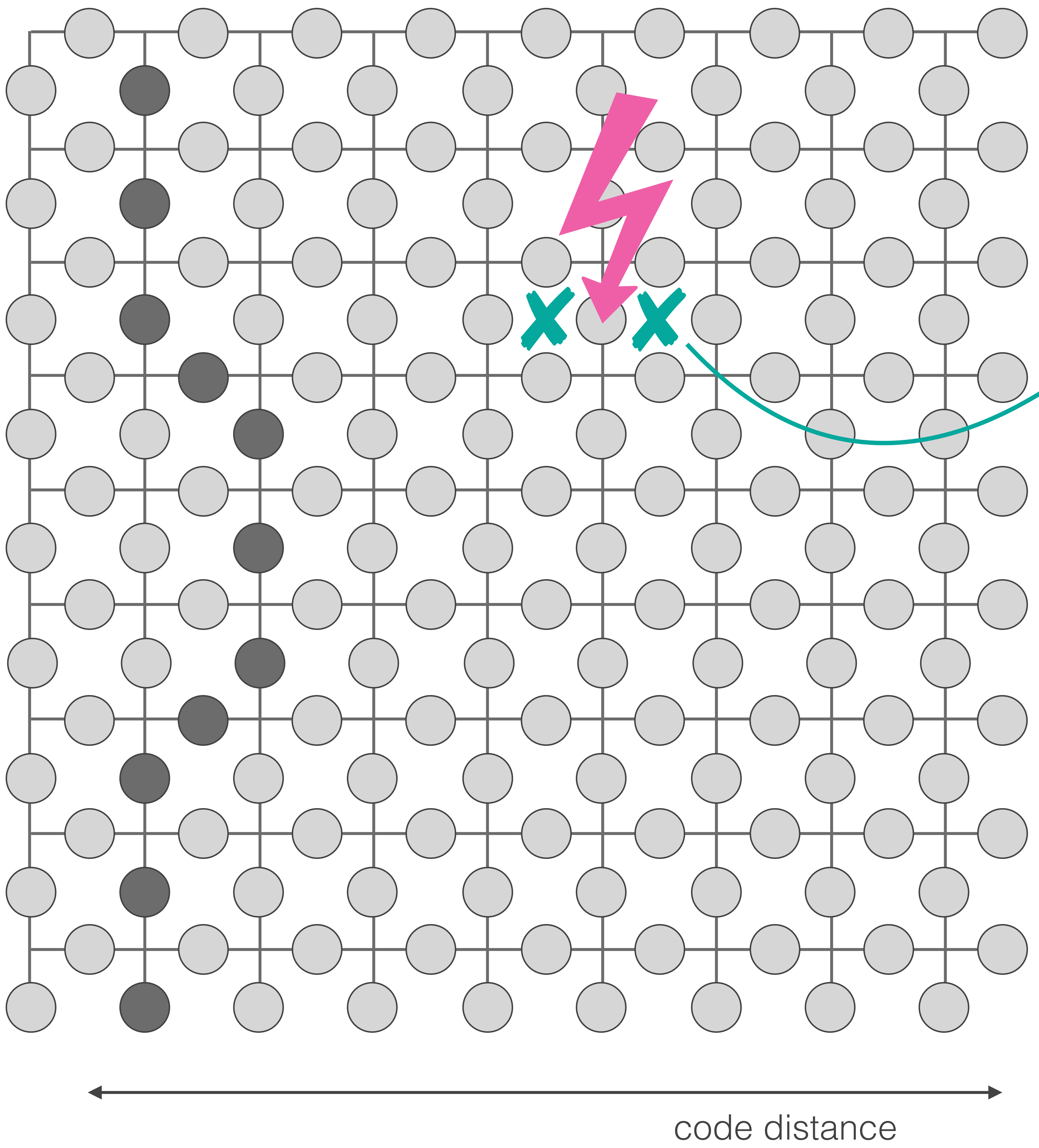
$$X_v|\psi\rangle = 1 \cdot |\psi\rangle \quad \forall v$$

$$Z_p|\psi\rangle = 1 \cdot |\psi\rangle \quad \forall p$$

code space is four-dimensional
$\Rightarrow$ encodes two logical qubits

topological order is stable to local perturbations

"bit flip" errors



The presence of a bit flip error is detected via a **syndrome**

$$Z_p = \prod_{i \in p} Z_i = -1$$

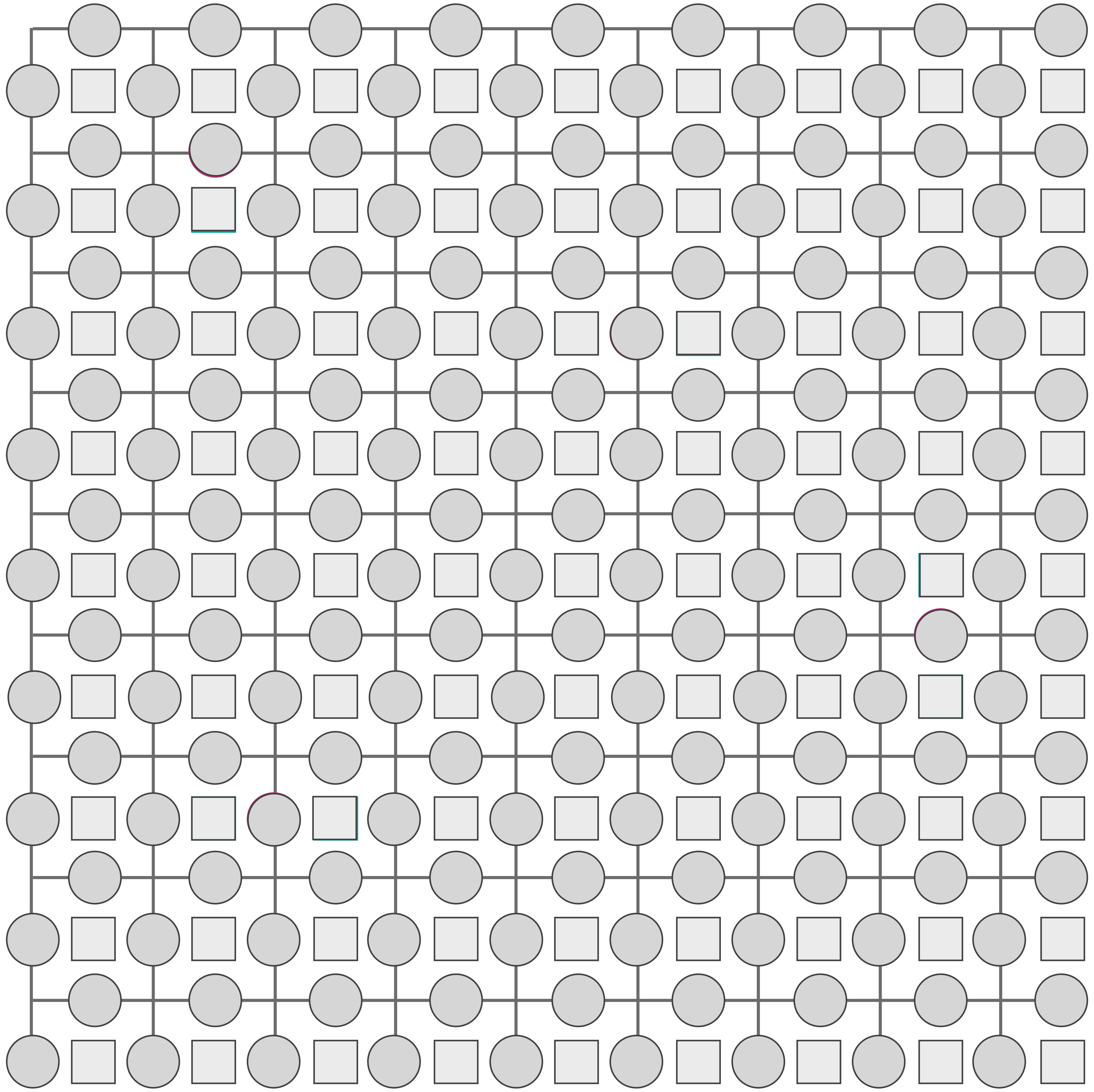
non-destructive measurement

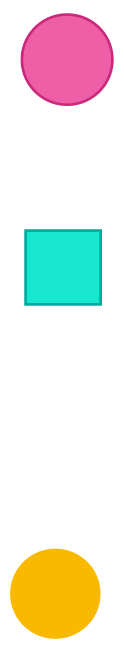
Stabilizer codes allow for non-destructive quantum error detection and correction.

Stabilizer Codes and Quantum Error Correction
D. Gottesmann, arXiv:quant-h/9705052

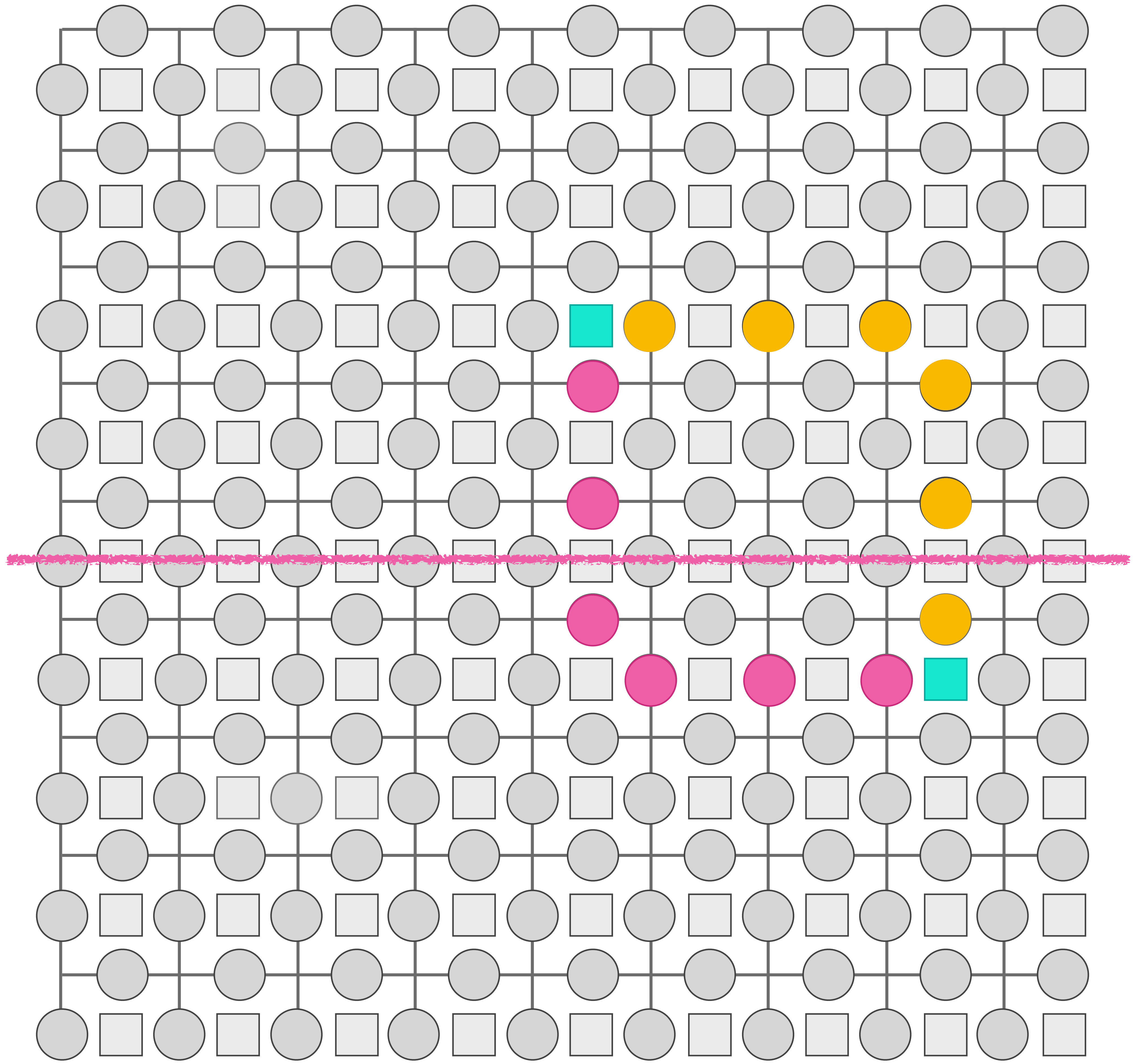
Stabilizer Formalism for Operator Quantum Error Correction
D. Poulin, Phys. Rev. Lett. 95, 230504 (2005)

“bit flip” errors & recovery



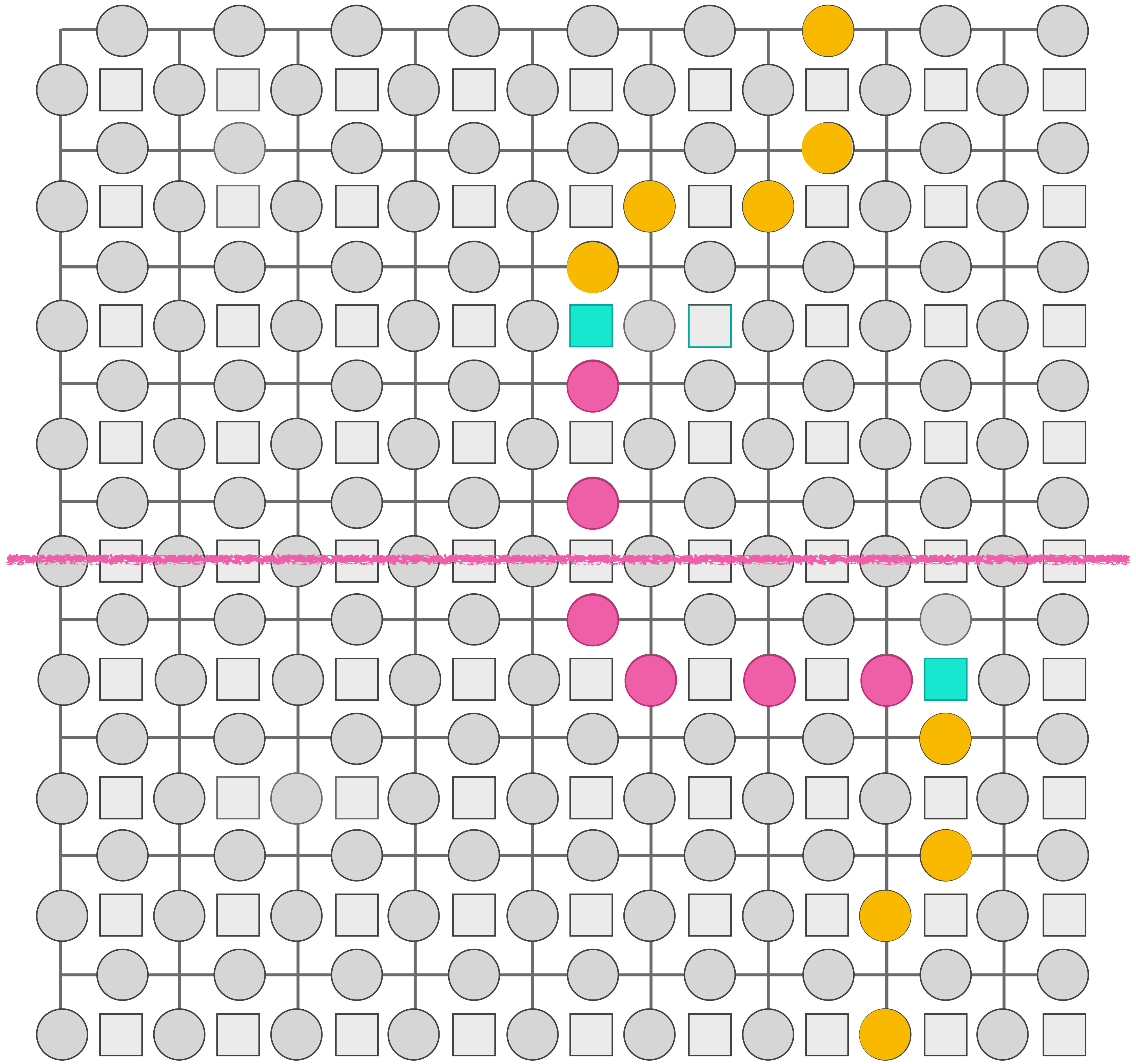


“bit flip” errors & recovery

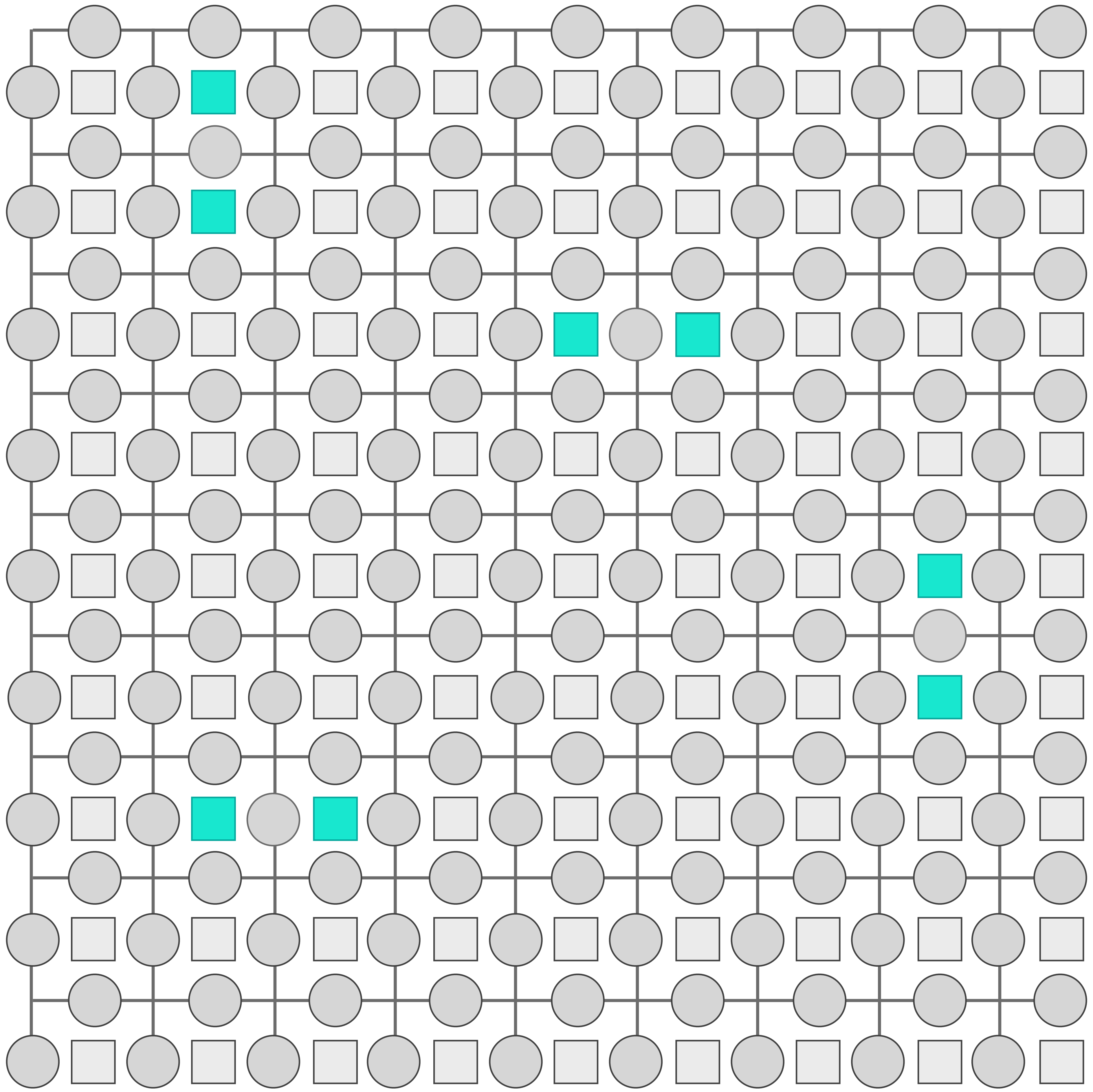


- bit flip error
- syndrome
- recovery operation X_v

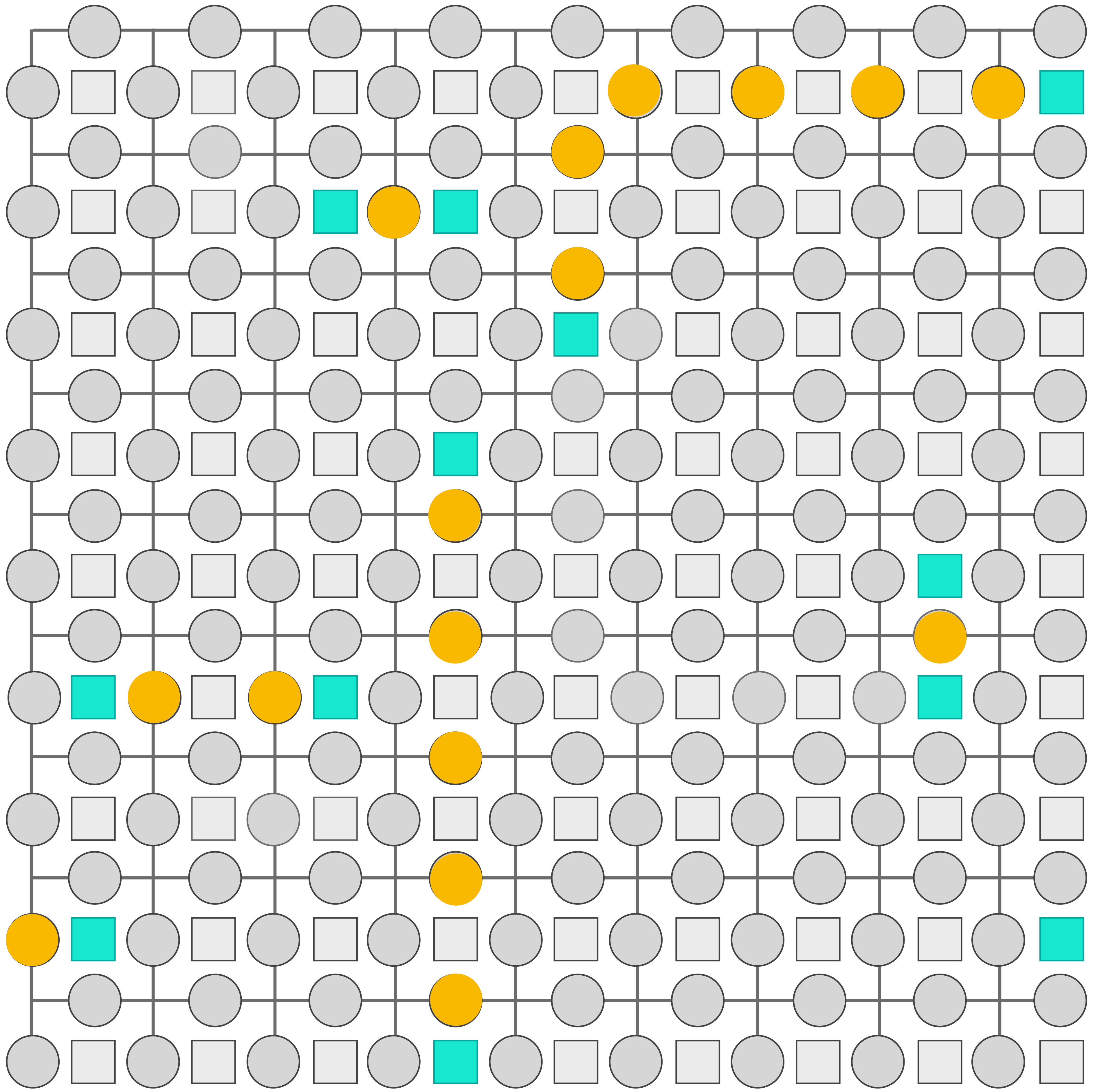
“bit flip” errors & recovery



multiple “bit flip” errors



decoding problem



bit flip error

syndrome

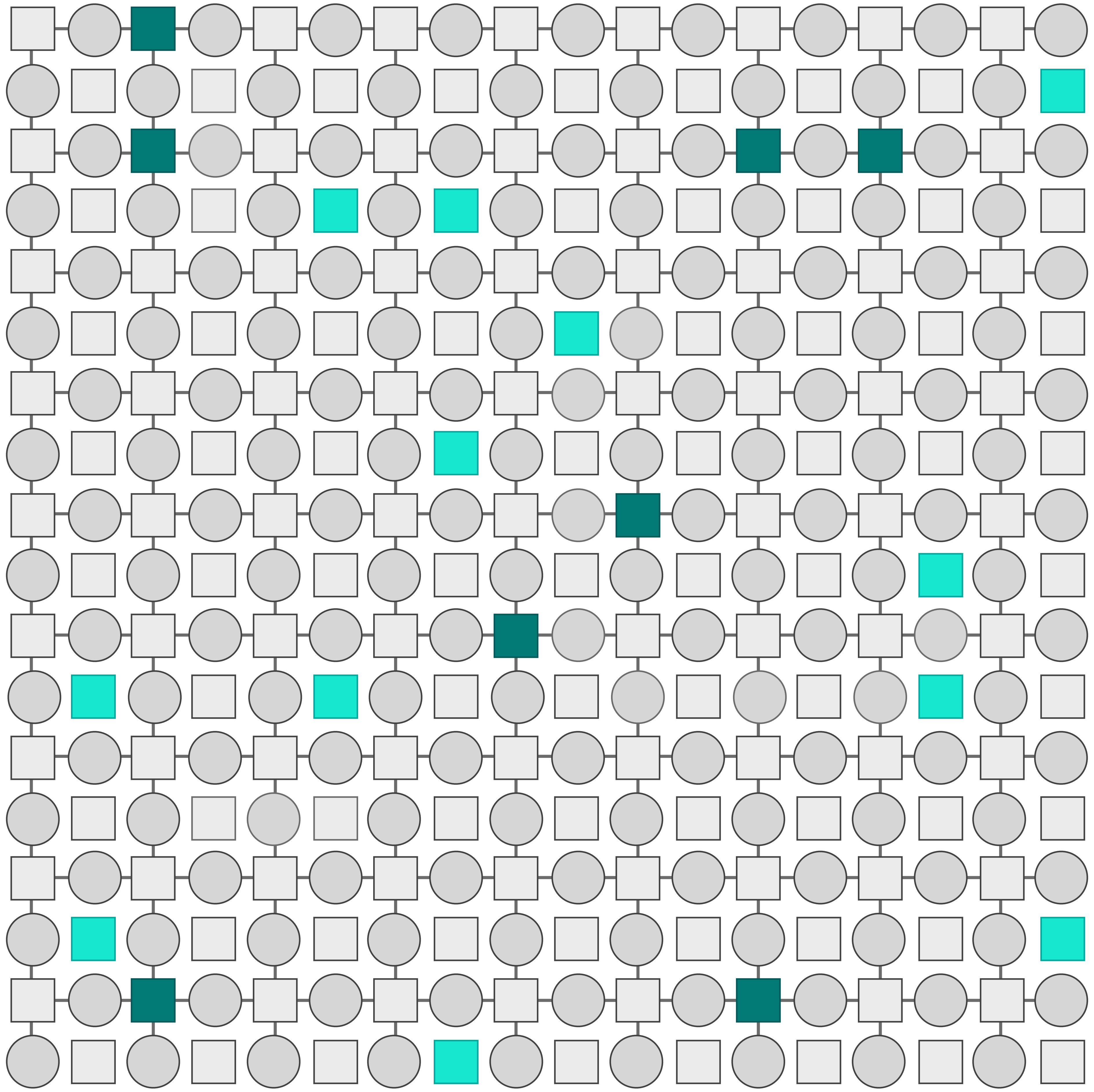
no obvious pairing of syndromes

goal of a **decoder** is to come up with a

recovery operation

without producing a logical fault

noise models



bit flip error

Z syndrome

X syndrome

no obvious pairing of syndromes

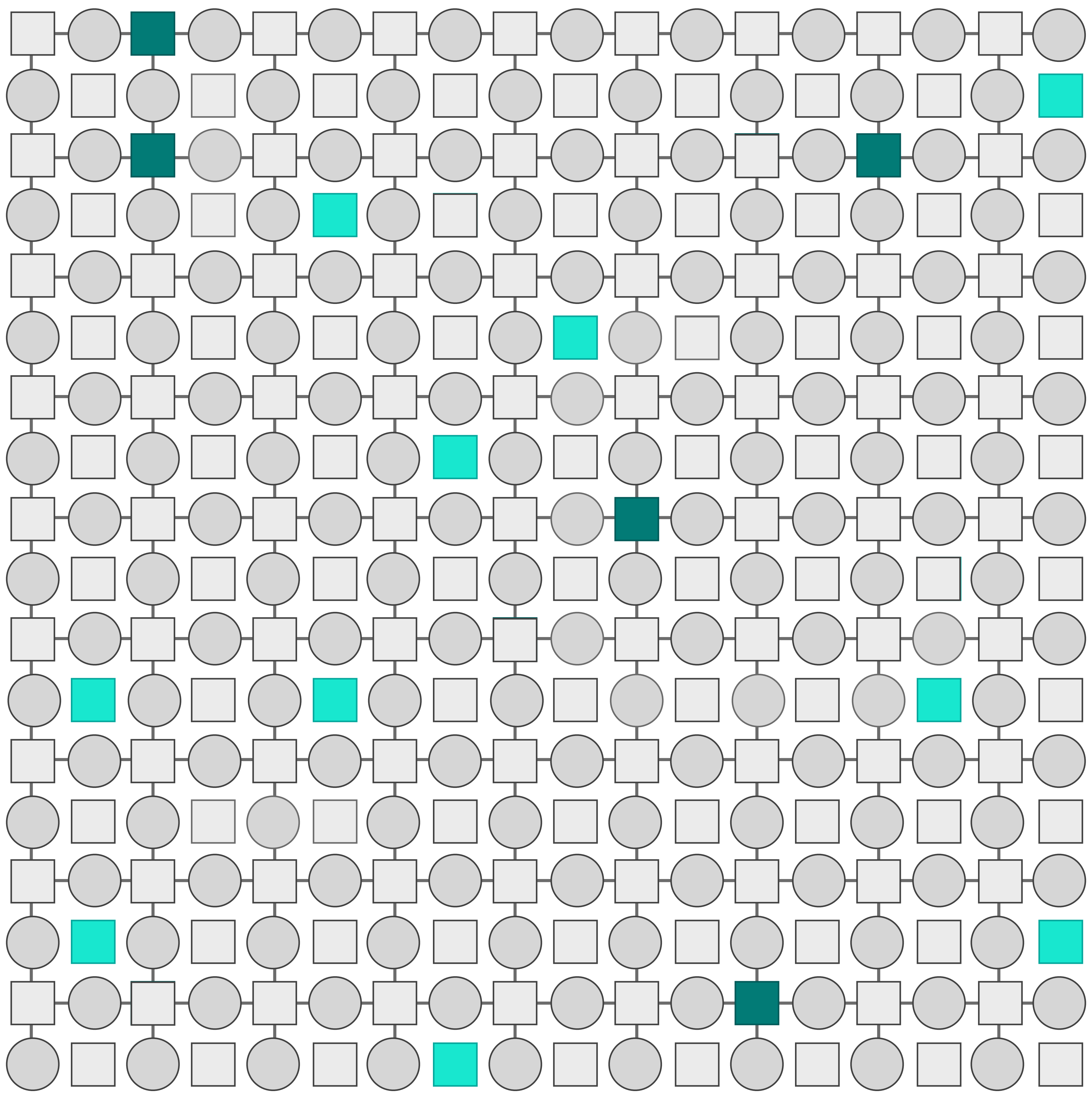
depolarizing noise

goal of a **decoder** is to come up with a

recovery operation

without producing a logical fault

noise models



bit flip error

Z syndrome

X syndrome

no obvious pairing of syndromes

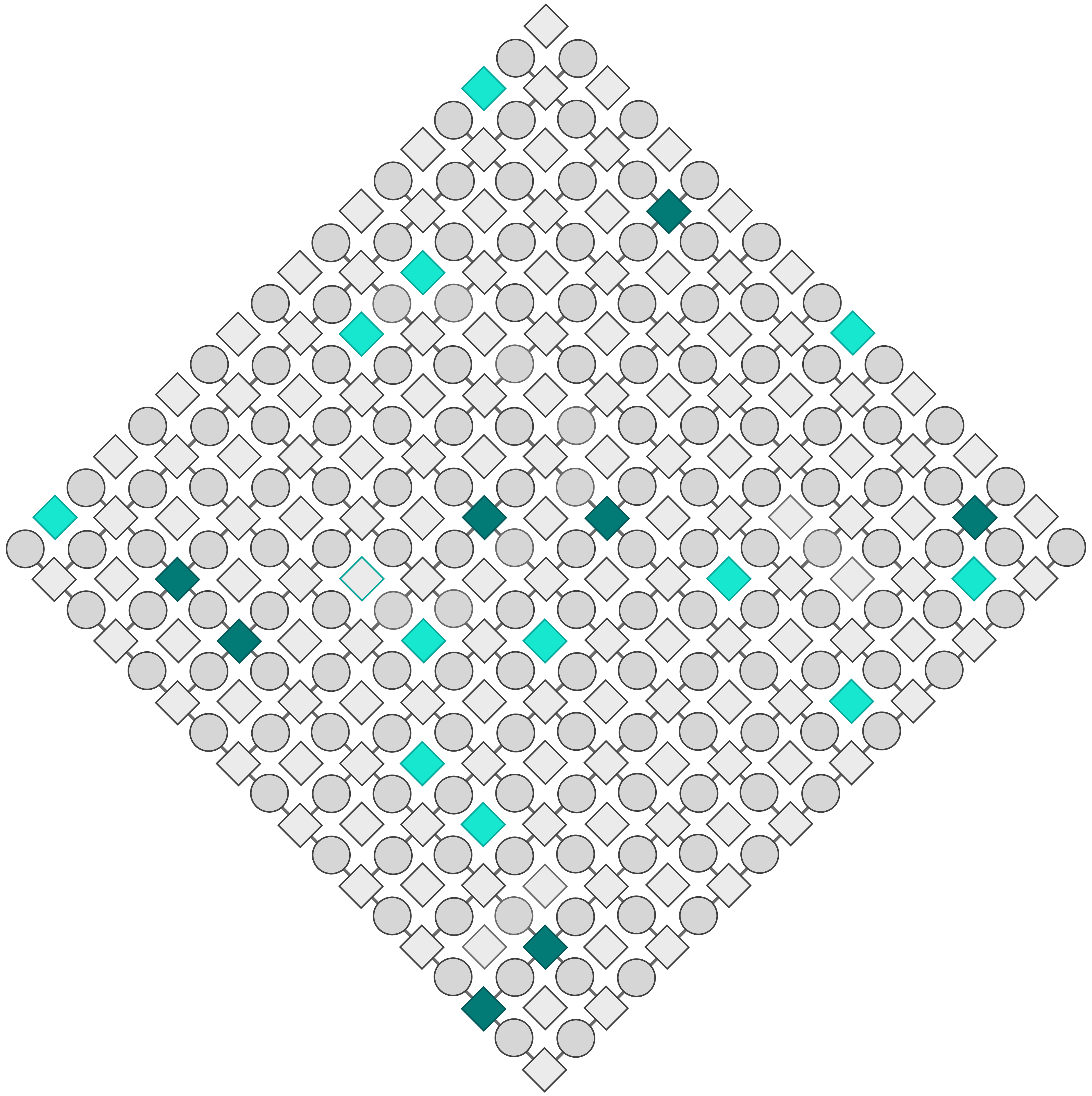
depolarizing noise + syndrome errors

goal of a **decoder** is to come up with a

recovery operation

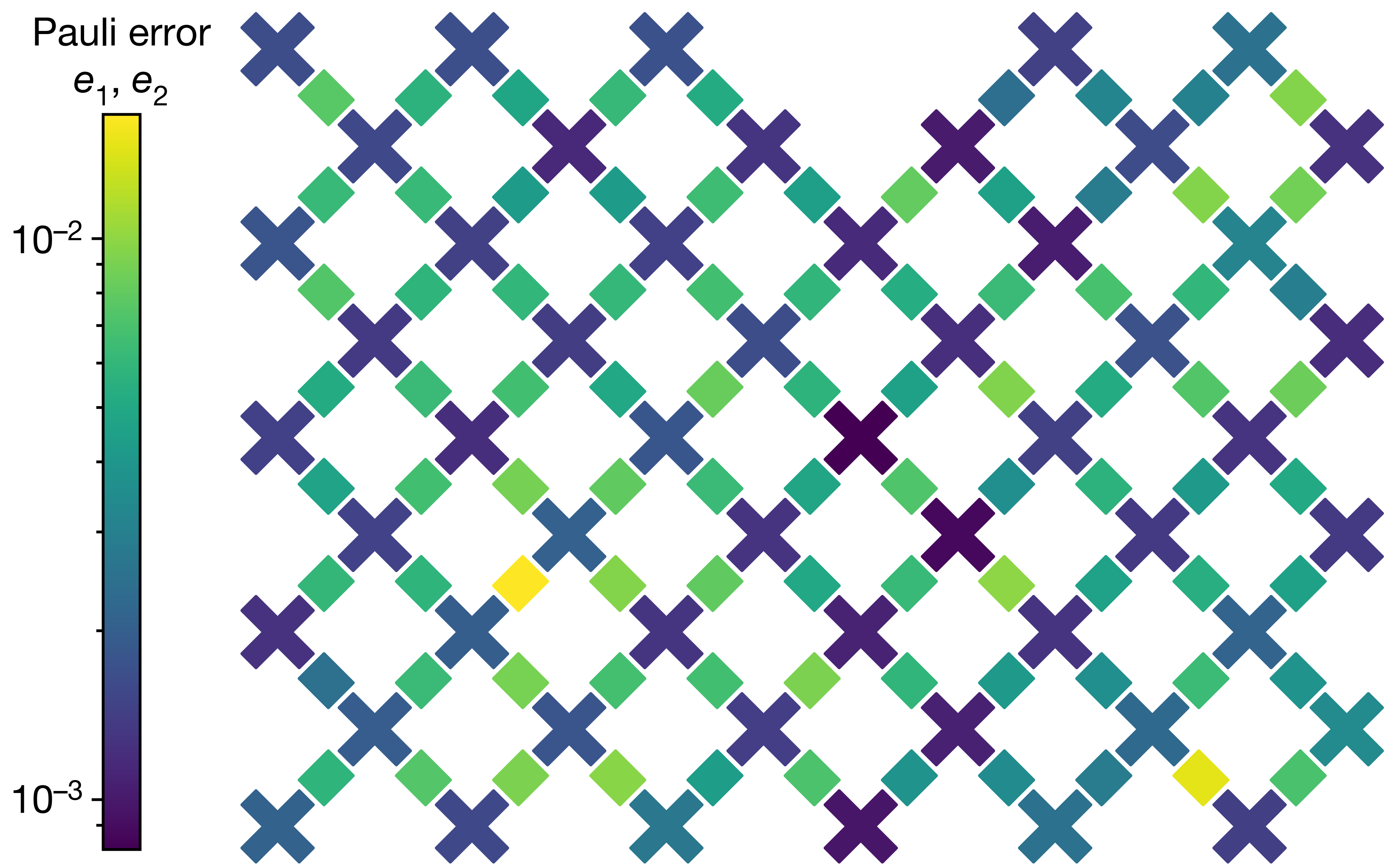
without producing a logical fault

stabilizer codes



Pauli and measurement errors

Average error	Isolated	Simultaneous
Single-qubit (e_1)	0.15%	0.16%
Two-qubit (e_2)	0.36%	0.62%
Two-qubit, cycle (e_{2c})	0.65%	0.93%
Readout (e_r)	3.1%	3.8%



Google's Sycamore chip

Realizing topologically ordered states on a quantum processor
K. J. Satzinger *et al.* (google team), arXiv:2104.01180

decoders

conventional decoders

minimum-weight perfect matching (**MWPM**)

Topological quantum memory
E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, J. Math. Phys. **43**, 4452 (2002)

renormalization group (**RG decoder**)

Fast Decoders for Topological Quantum Codes,
G. Duclos-Cianci and D. Poulin, Phys. Rev. Lett. **104**, 050504 (2010)

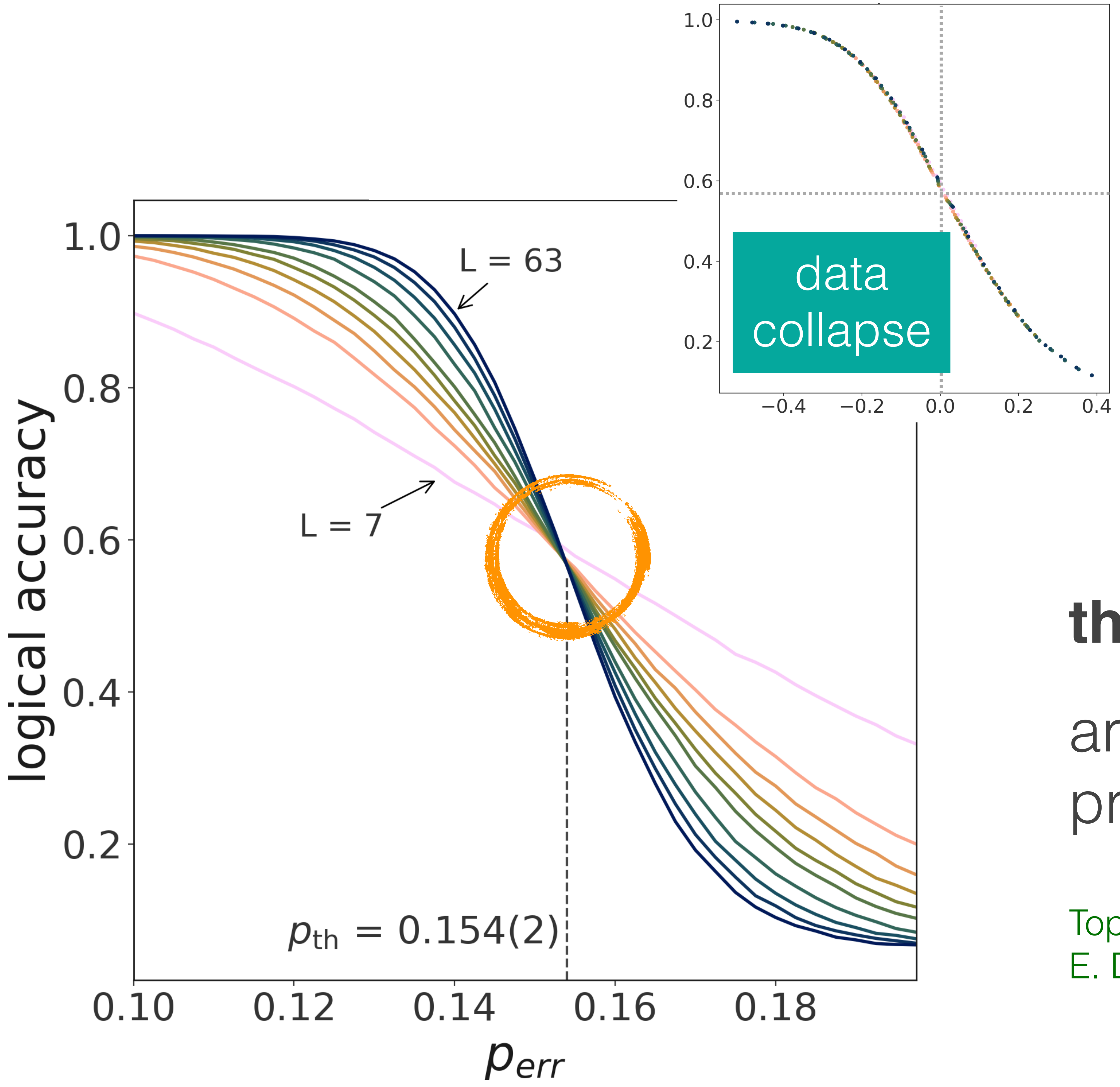
union find (**UF decoder**)

Almost-linear time decoding algorithm for topological codes
N. Delfosse and N. H. Nickerson, Quantum **5**, 595 (2021)

tensor networks (**TN decoder**)

General tensor network decoding of 2D Pauli codes
C. T. Chubb, arXiv:2101.04125 (2021).

minimum weight perfect matching



Open-source Python implementation **PyMatching**
Oscar Higgott, arXiv:2105.13082
https://github.com/oscarhiggott/PyMatching

threshold theorem

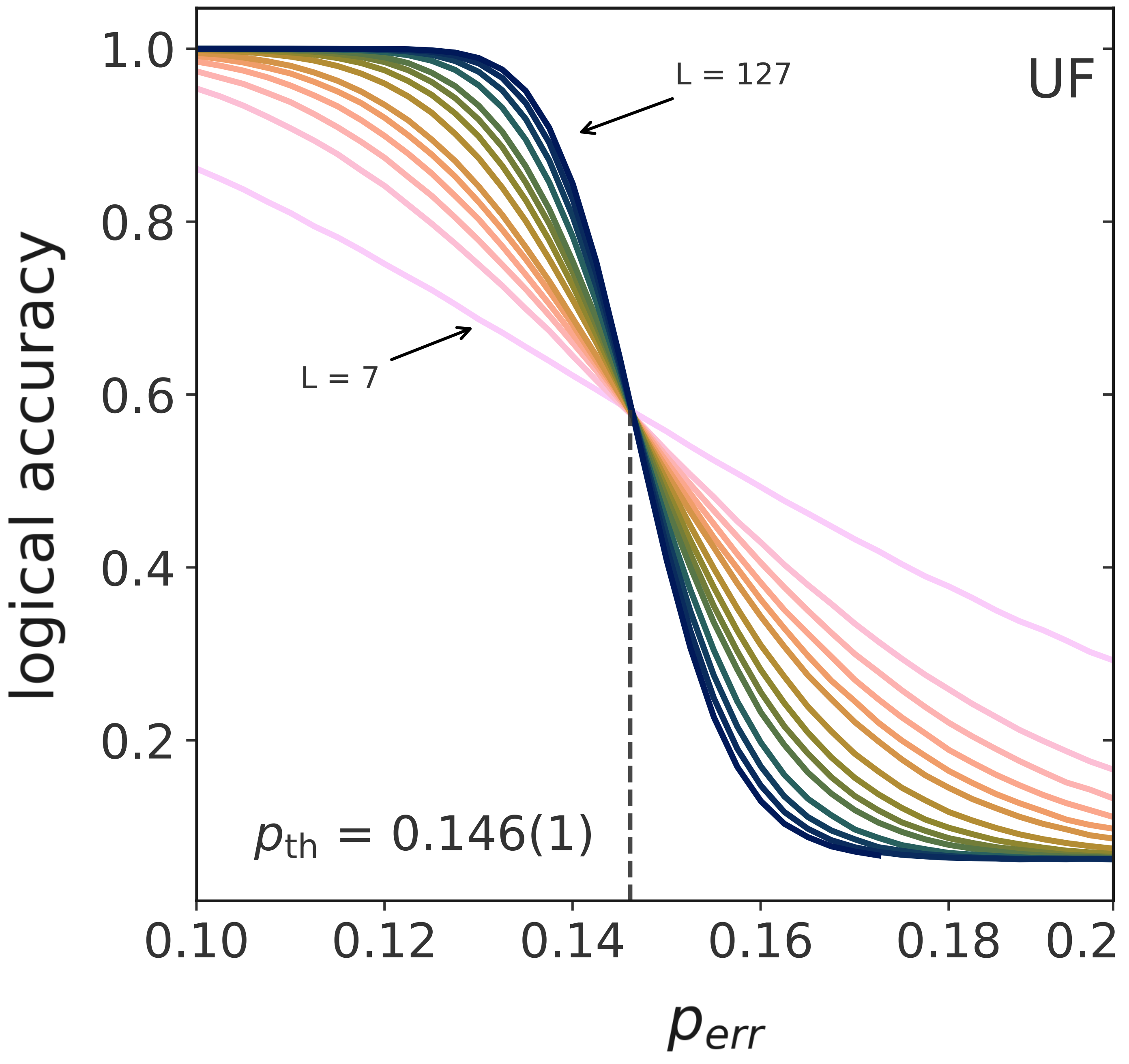
arbitrarily long, reliable computation is possible provided the error rate is below the threshold value

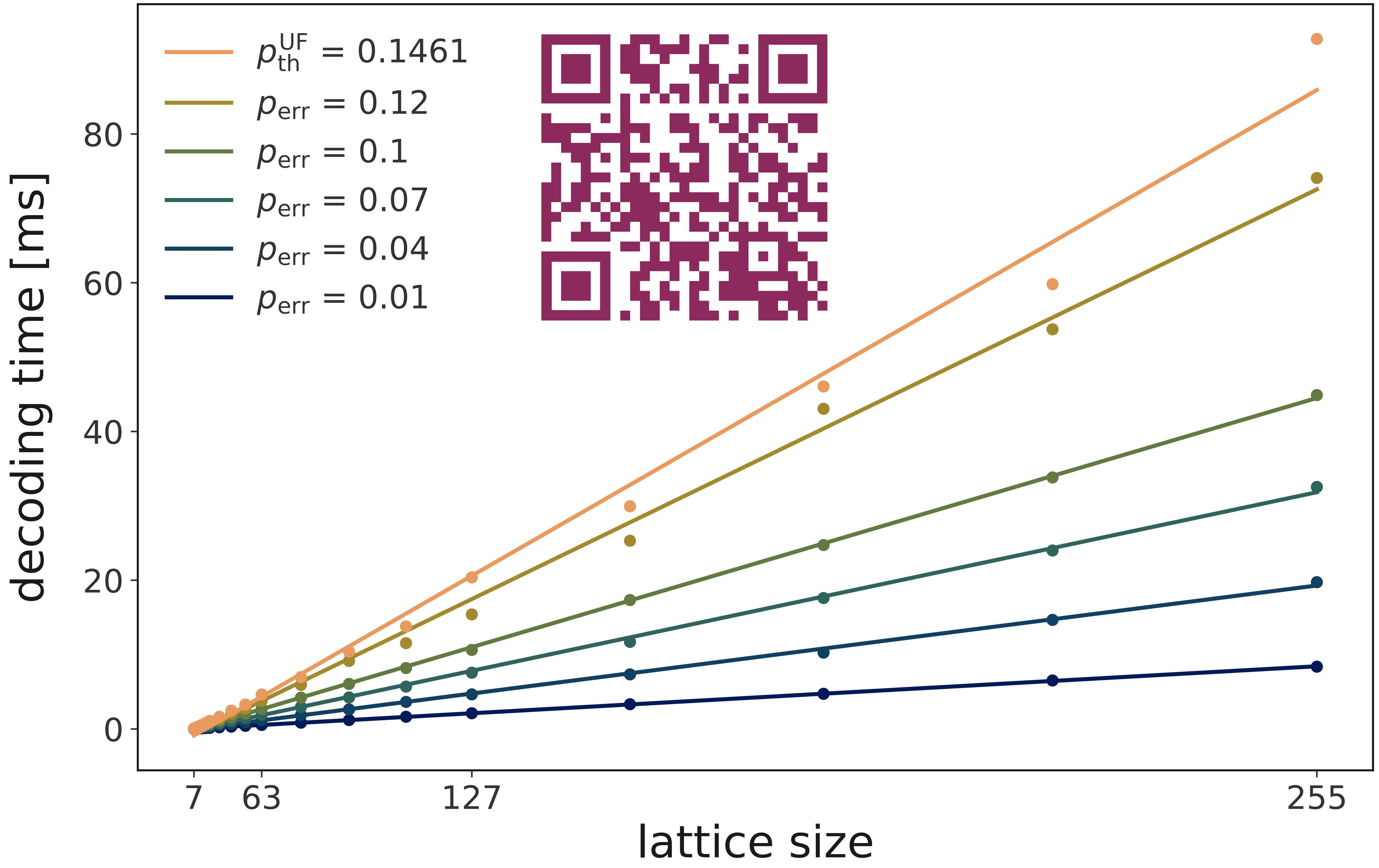
Topological quantum memory
E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, J. Math. Phys. **43**, 4452 (2002)

union find decoder

Almost-linear time decoding algorithm for topological codes
N. Delfosse and N. H. Nickerson, Quantum **5**, 595 (2021)

Open-source C++ implementation of the Union-Find decoder
Chae-Yeun Park & Kai Meinerz, https://github.com/chaeyeunpark



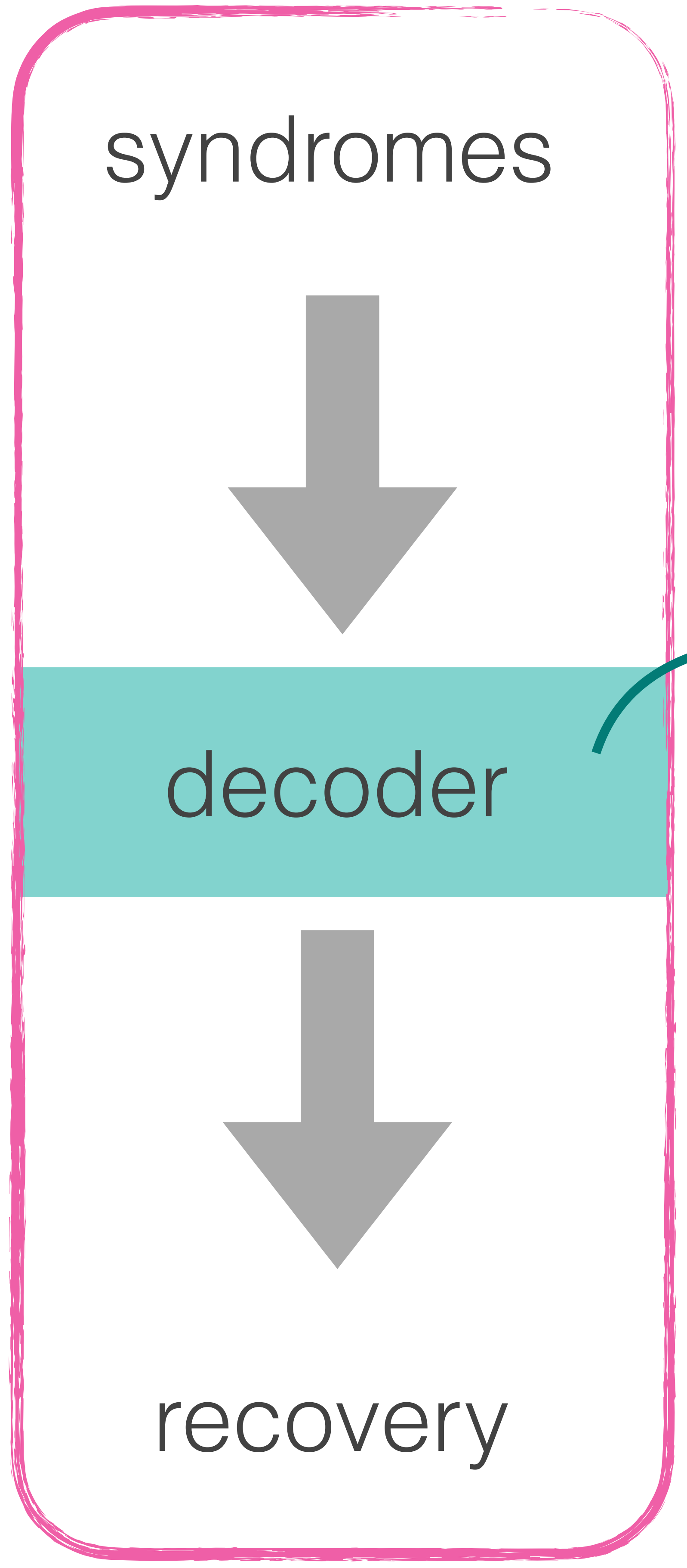


conventional decoders

	threshold	d_{max}	scaling
minimum-weight perfect matching (**MWPM**) Topological quantum memory E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, J. Math. Phys. **43**, 4452 (2002)	0.154	63*	$\mathcal{O}(n^3)$ $\mathcal{O}(d^{2.11})$
renormalization group (**RG decoder**) Fast Decoders for Topological Quantum Codes, G. Duclos-Cianci and D. Poulin, Phys. Rev. Lett. **104**, 050504 (2010)	0.164	128	$\mathcal{O}(d^2 \log d)$
union find (**UF decoder**) Almost-linear time decoding algorithm for topological codes N. Delfosse and N. H. Nickerson, Quantum **5**, 595 (2021)	0.146	255*	$\mathcal{O}(n \cdot \alpha(n))$
tensor networks (**TN decoder**) General tensor network decoding of 2D Pauli codes C. T. Chubb, arXiv:2101.04125 (2021).	0.188	64	$\mathcal{O}(n \log n + n\chi^3)$

*benchmark calculations by ourselves that go beyond the published literature

ML-assisted decoders



insert favoritve
machine learning
algorithm here

pattern recognition

dimensional reduction

high adaptability

easily parallelized

PRL **119,** 030501 (2017) PHYSICAL REVIEW LETTERS week ending 21 JULY 2017

Neural Decoder for Topological Codes

Giacomo Torlai and Roger G. Melko
Department of Physics and Astronomy, University of Waterloo, Ontario N2L 3G1, Canada
and Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
(Received 20 October 2016; published 18 July 2017)

PHYSICAL REVIEW LETTERS **122,** 200501 (2019)

Neural Belief-Propagation Decoders for Quantum Error-Correcting Codes

Ye-Hua Liu[1,*] and David Poulin[1,2,†]
[1]*Département de Physique & Institut Quantique, Université de Sherbrooke, J1K 2R1 Sherbrooke, Québec, Canada*
[2]*Canadian Institute for Advanced Research, M5G 1Z8 Toronto, Ontario, Canada*
(Received 26 November 2018; published 22 May 2019)

MACHINE LEARNING
Science and Technology

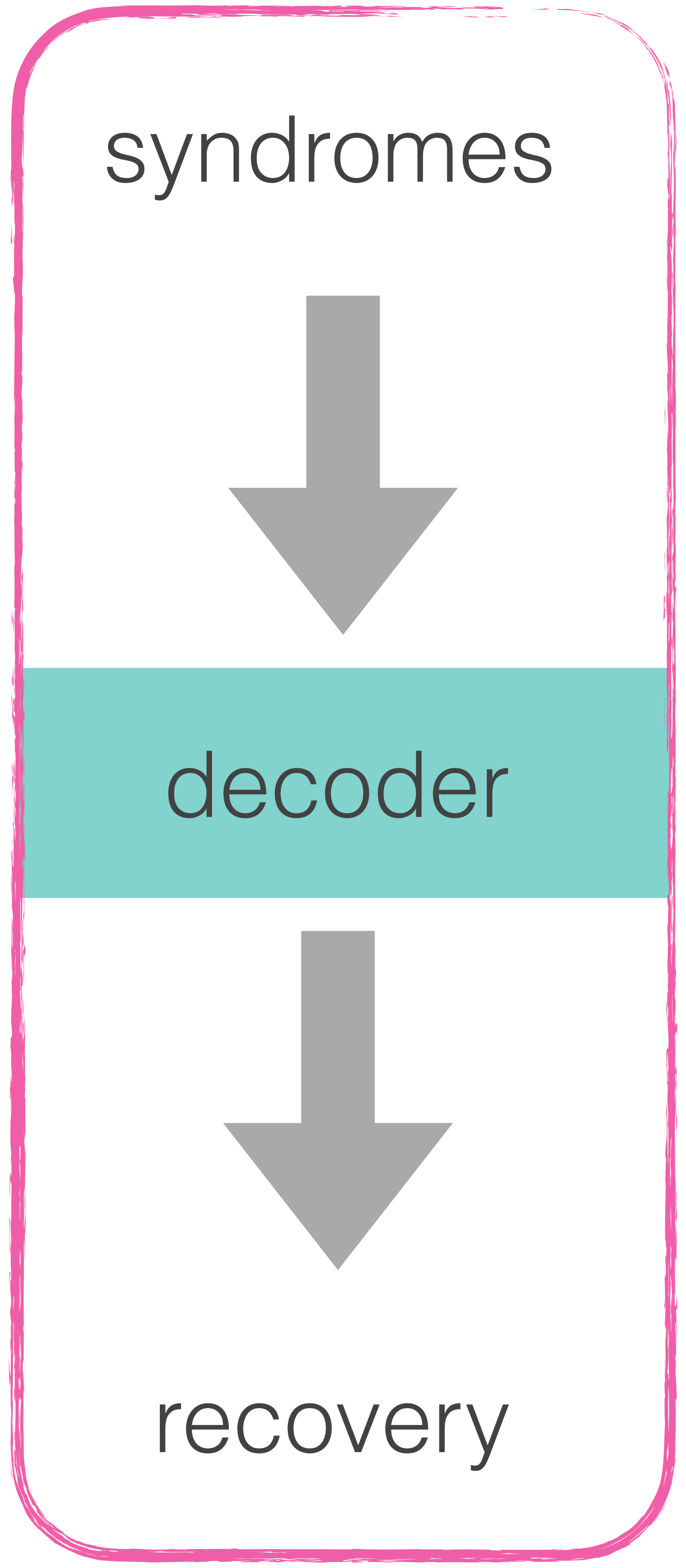
PAPER

Reinforcement learning decoders for fault-tolerant quantum computation

Ryan Sweke[1], Markus S Kesselring[1], Evert P L van Nieuwenburg[2] and Jens Eisert[1,3]
[1] Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany
[2] Niels Bohr International Academy, University of Copenhagen, 2100 Copenhagen, Denmark
[3] Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany
E-mail: rsweke@gmail.com

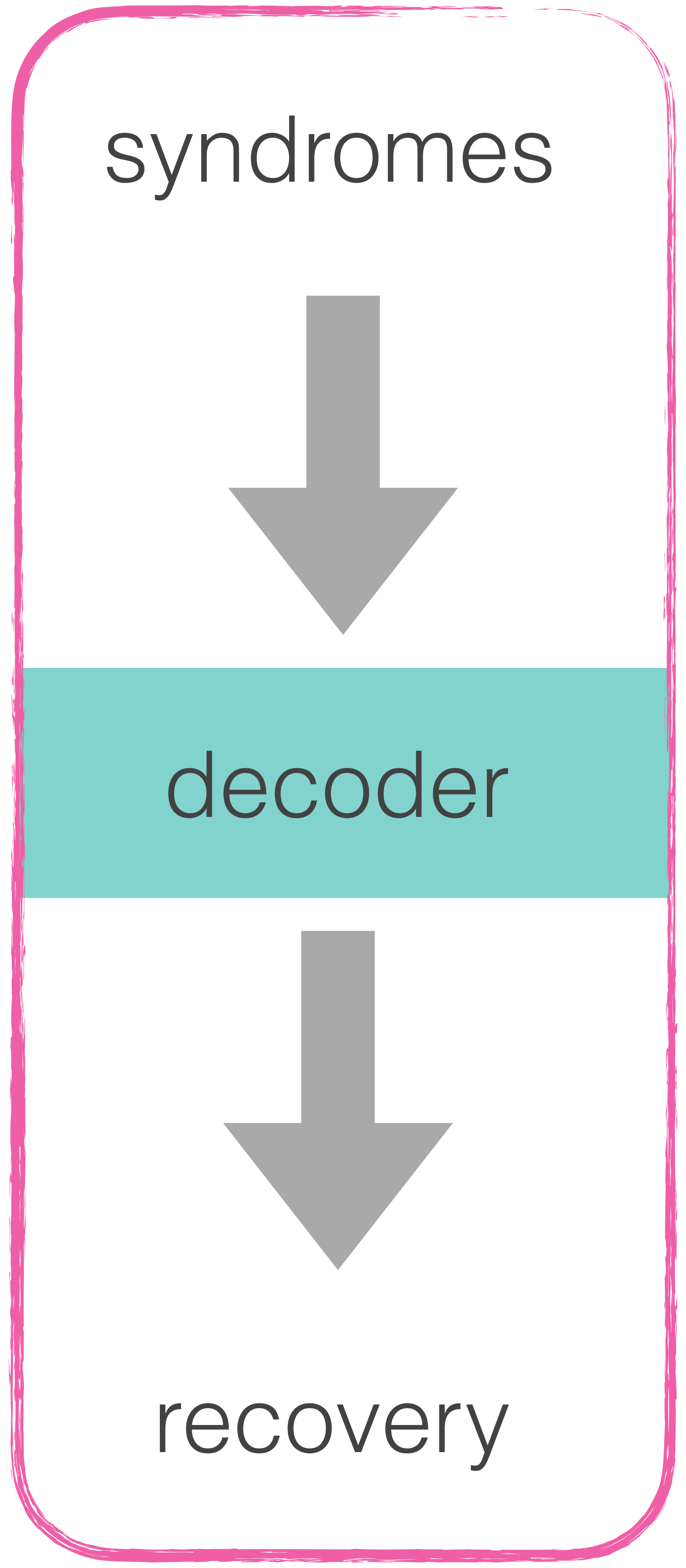
Keywords: quantum error correction, reinforcement learning, fault tolerant quantum computing

ML-assisted decoders



title	date	QECC	noise model	p_{th}	$d_{\max}$	algorithmic scaling
Decoding Small Surface Codes with Feedforward Neural Networks [16]	05/2017	SC	DP	~ 0.15	7	$\mathcal{O}$(MWPM)
Neural Decoder for Topological Codes [17]	07/2017	TC	i.i.d bit flip	~ 0.110	6	$> \mathcal{O}$(MWPM)
Deep Neural Network Probabilistic Decoder for Stabilizer Codes [18]	09/2017	TC	DP	0.164	11	$\gg \mathcal{O}$(MWPM)
Deep neural decoders for near term fault-tolerant experiments [19]	07/2018	SC	CLN	$\epsilon \sim 7.11 \times 10^{-4}$	5	n.a.
Neural network decoder for topological color codes with circuit level noise [20]	01/2019	CC	CLN	$\epsilon \sim 0.0023$	7	n.a.
Neural Belief-Propagation Decoders for Quantum Error-Correcting Codes [21]	05/2019	TC	i.i.d. X&Z	~ 0.07	10	n.a.
Quantum error correction for the toric code using deep reinforcement learning [22]	09/2019	TC	i.i.d. bit flip	~ 0.1	7	$\gg \mathcal{O}(n)$ (estimate)
Symmetries for a High Level Neural Decoder on the Toric Code [23]	10/2019	TC	DP	n.a.	7	$\mathcal{O}$(MWPM)
Deep Q-learning decoder for depolarizing noise on the toric code [24]	05/2020	TC	DP	~ 0.165	7 (9)	$\gg \mathcal{O}(n)$ (estimate)
Reinforcement learning for optimal error correction of toric codes [25]	06/2020	TC	i.i.d bit flip	0.103	9	$> \mathcal{O}$(MWPM)
Neural Network Decoders for Large-Distance 2D Toric Codes [26]	08/2020	TC	i.i.d. bit flip	~ 0.103	64	$> \mathcal{O}$(RG)
Determination of the semion code threshold using neural decoders [27]	09/2020	SM	DP	~ 0.105	13	n.a.
Reinforcement learning decoders for fault-tolerant quantum computation [28]	12/2020	SC	i.i.d. bit flip & DP	n.a.	5	$\gg \mathcal{O}(n)$ (estimate)
Scalable Neural Decoder for Topological Surface Codes **(this work)**	01/2021	TC	DP	0.167	255	$\mathcal{O}$(UF)

ML-assisted decoders



title	date	QECC	noise model	p_{th}	$d_{\max}$	algorithmic scaling
Decoding Small Surface Codes with Feedforward Neural Networks [16]	05/2017	SC	DP	~ 0.15	7	$\mathcal{O}$(MWPM)
Neural Dec						
Deep Neur Stabilizer C						
Deep neura experiments						
Neural net codes with						
Neural Beli Error-Corre						
Quantum er deep reinfor						(estimate)
Symmetries the Toric C						
Deep Q-lear the toric co						(estimate)
Reinforcem tion of toric						
Neural Net Toric Codes						
Determinati neural deco						
Reinforcement learning decoders for fault-tolerant quantum computation [28]	12/2020	SC	i.i.d. bit flip & DP	n.a.	5	$\gg \mathcal{O}(n)$ (estimate)
Scalable Neural Decoder for Topological Surface Codes **(this work)**	01/2021	TC	DP	0.167	255	$\mathcal{O}$(UF)

The **first generation** of ML-based decoders has typically **delivered only one of two benchmarks**

- **improve error threshold** at expense of scalability

OR

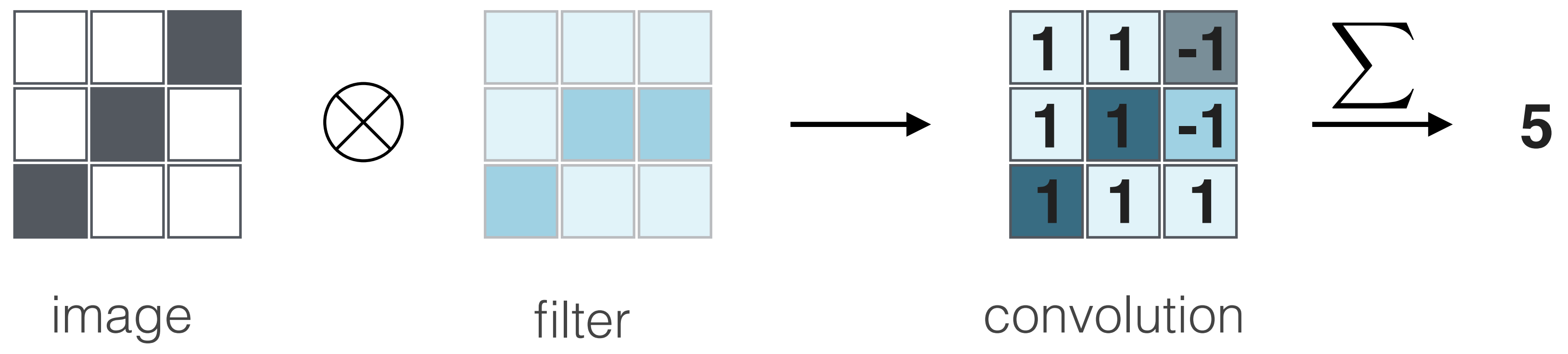
- **good scalability**, but reduced error thresholds

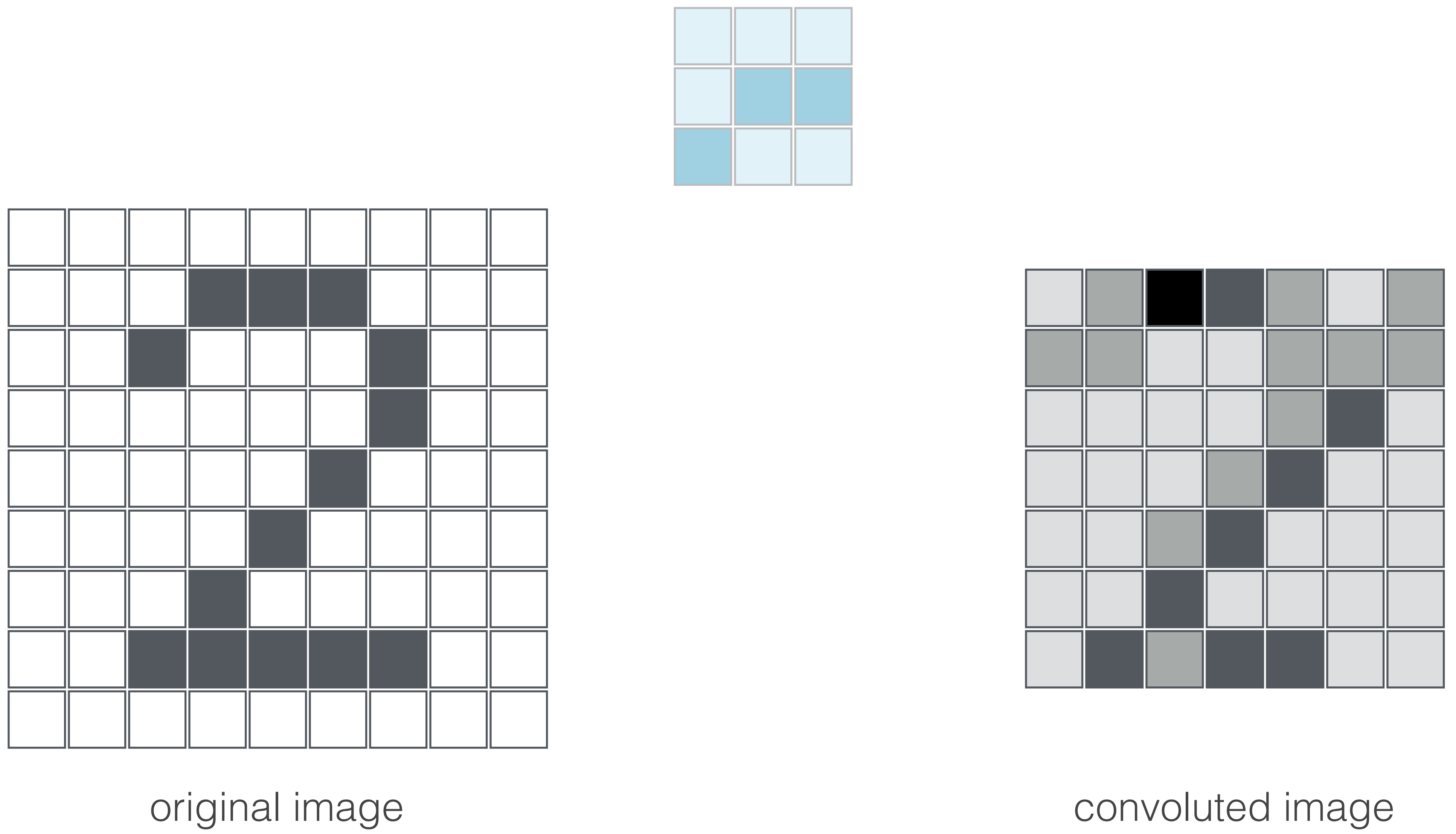
preprocessing

ML-assisted preprocessing

image classification

Convolutional neural networks (**CNN**) do preprocessing by looking for **recurring patterns**.



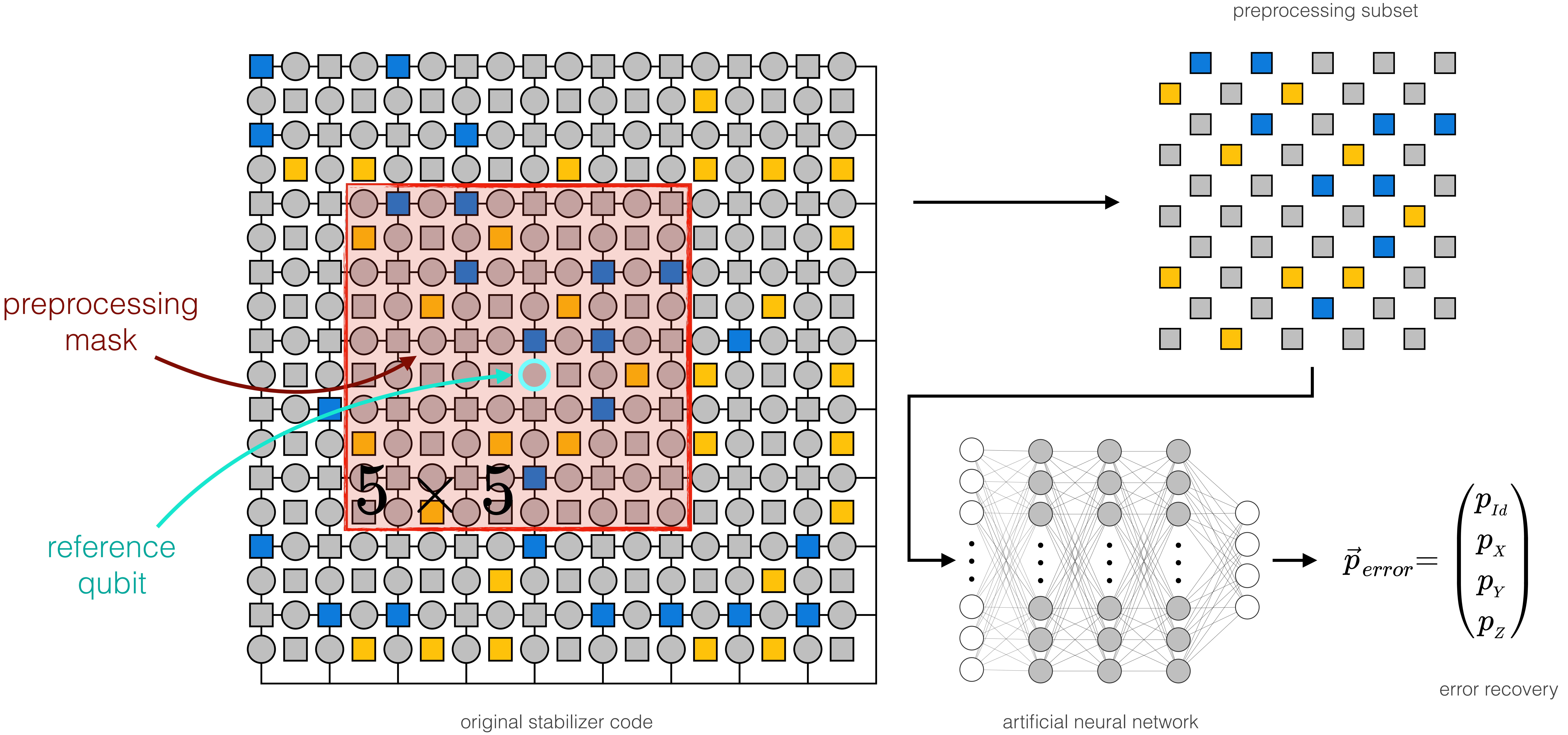


Adjustable filters help to identify and encode **local information** within the image at hand.

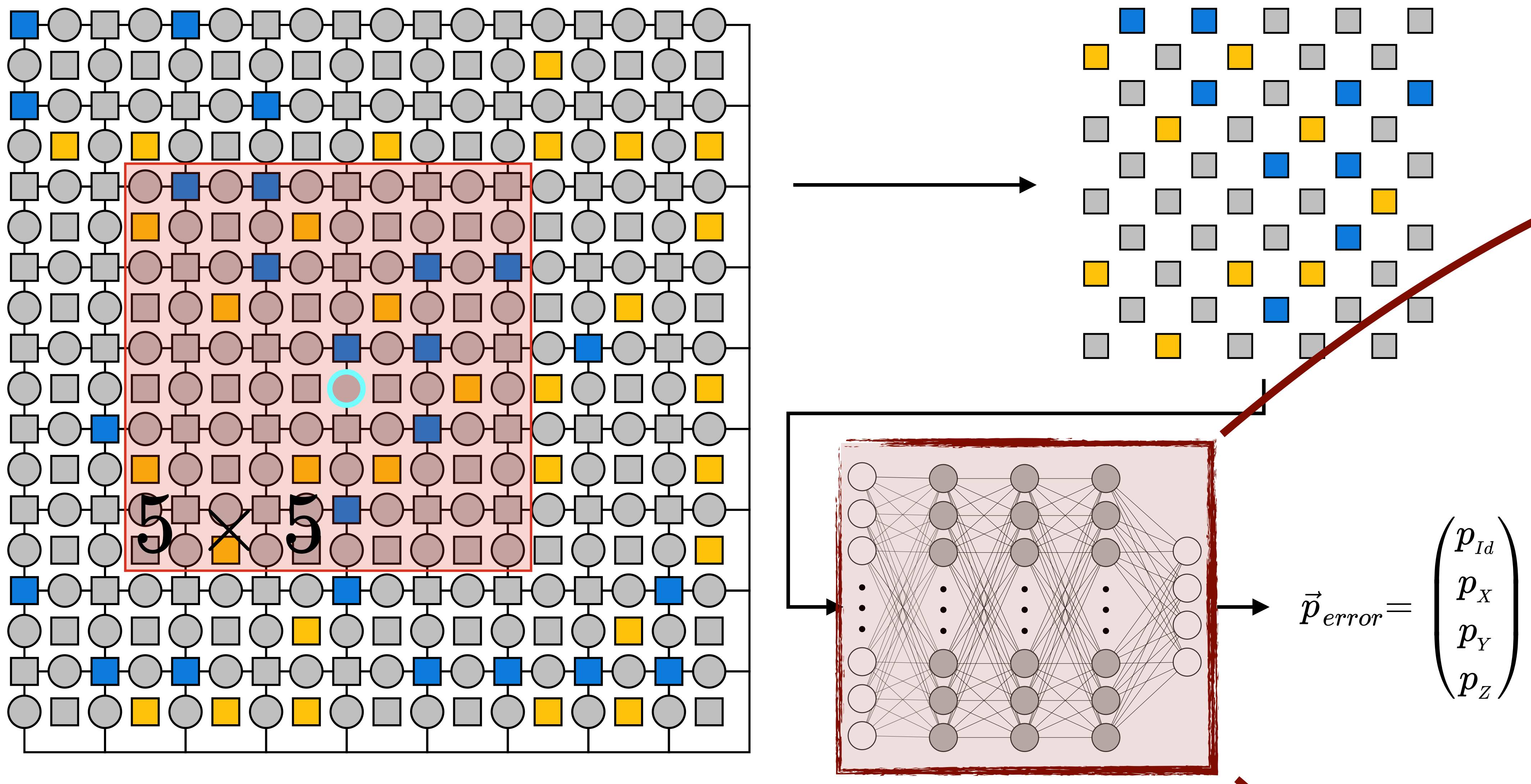
Slide filters across image and create new image encoding how well these filters fit.

ML-assisted preprocessing

quantum error correction



neural decoder setup



optimized network **architecture**

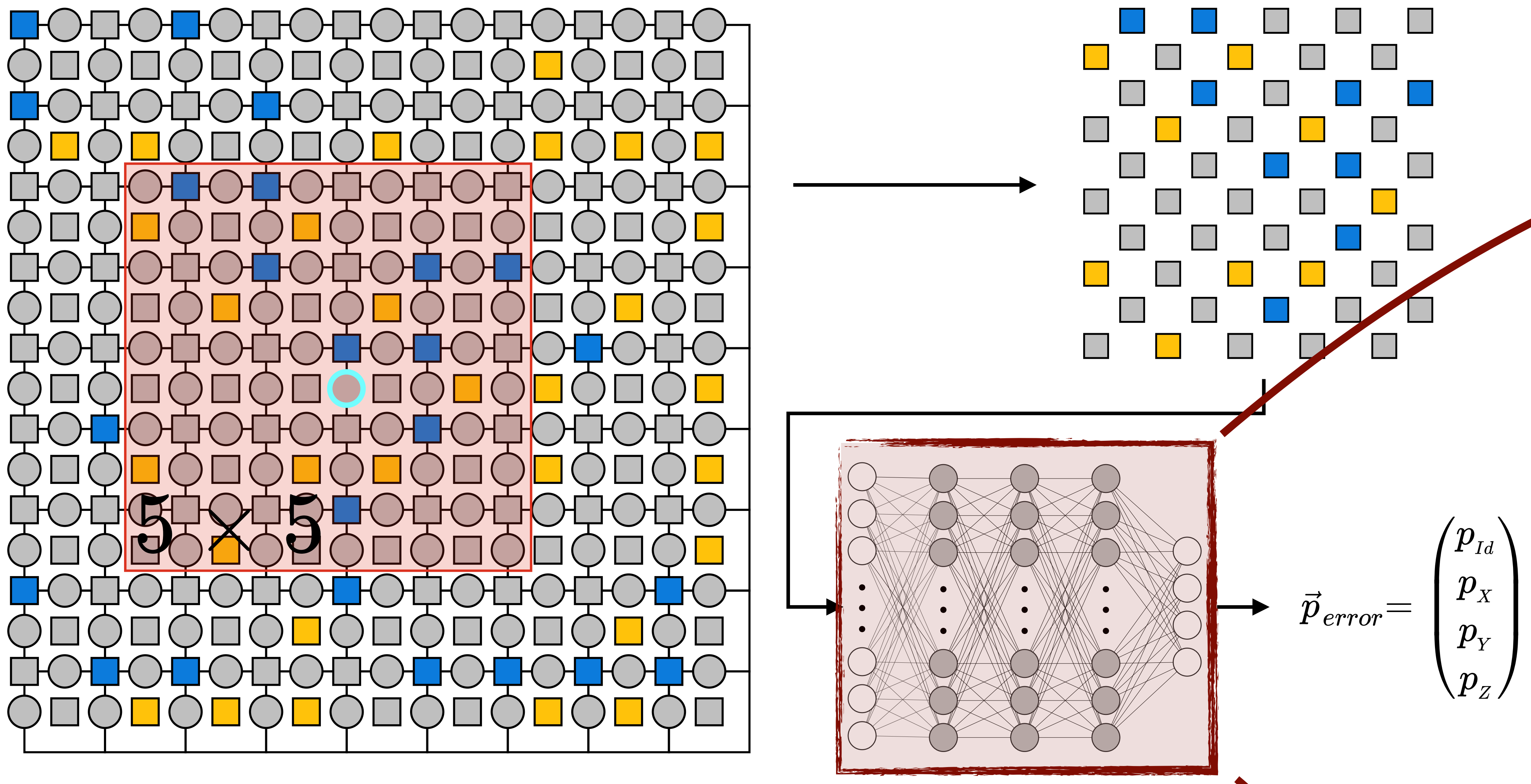
hidden layers	hidden nodes	total parameters		
		$\ell = 3$	$\ell = 5$	$\ell = 7$
1	32	740	1764	3300
1	64	1476	3526	6596
1	128	2948	7044	13188
1	256	5892	14084	26372
2	32	1796	2820	4356
2	64	5636	7684	10756
2	128	19460	23556	29700
2	256	71684	79876	92164
3	32	2852	3876	5412
3	64	9796	11844	14916
3	**128**	35972	**40068**	46212
3	256	137476	145668	157956

training parameters	
batch size	512
epochs	10^6
learning rate	0.001
optimizer	ADAM
loss function	categorical cross-entropy
training lattice size L_{train}	7
section distance ℓ	5

optimized network **parameters**

multiclass classification task

neural decoder setup



optimized network **architecture**

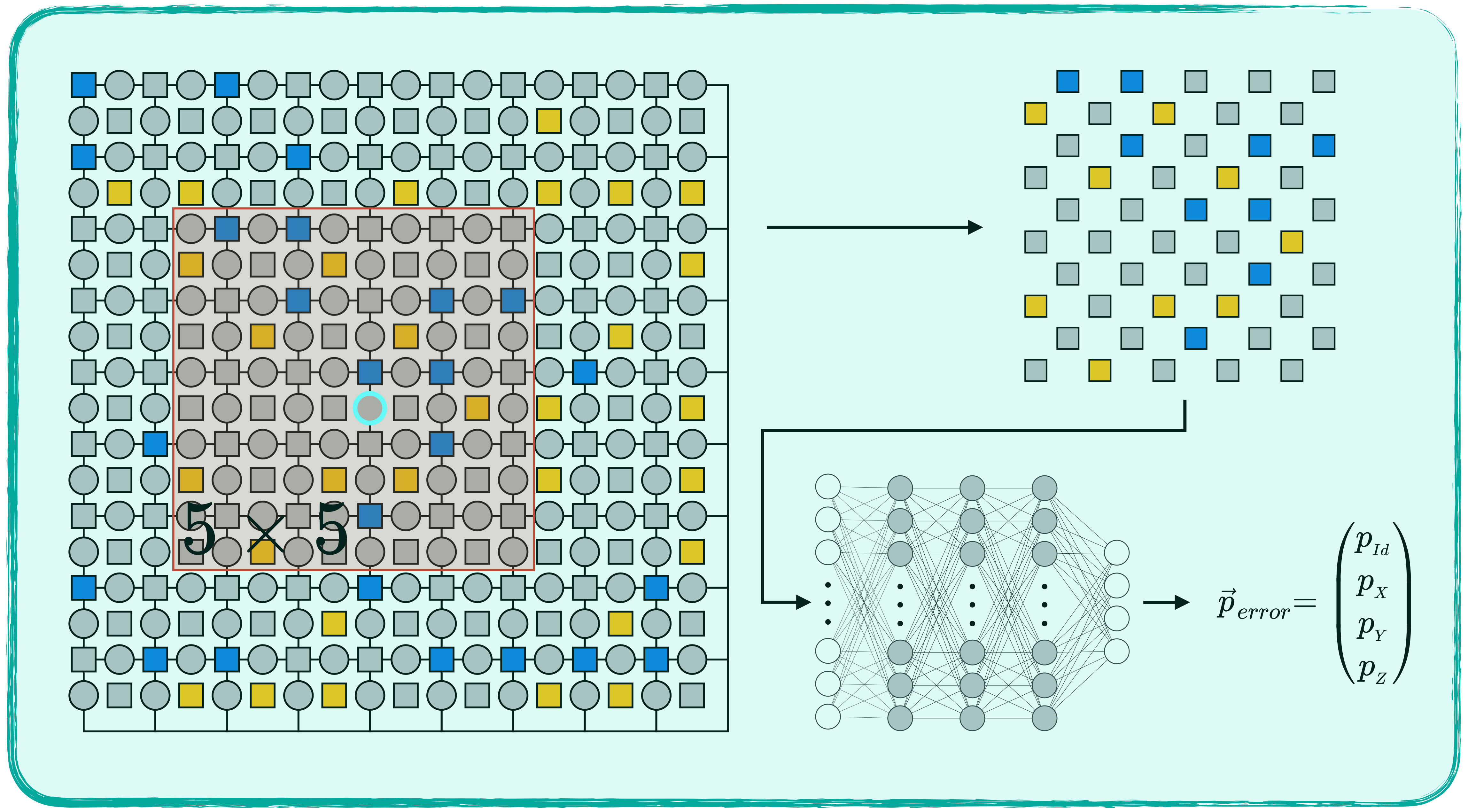
depolarizing noise	
$\ell = 5$ network parameters (optimal wall-clock time)	
hidden layers	3
hidden nodes per layer	128
total number free parameter	40 068
activation functions hidden layer	Relu
activation functions output layer	Softmax
$\ell = 7$ network parameters (optimal error threshold)	
hidden layers	5
hidden nodes per layer	512
total number free parameter	1 103 364
activation functions hidden layer	Relu
activation functions output layer	Softmax

training parameters	
batch size	512
epochs	10^6
learning rate	0.001
optimizer	ADAM
loss function	categorical cross-entropy
training lattice size L_{train}	7
section distance ℓ	5

optimized network **parameters**

multiclass classification task

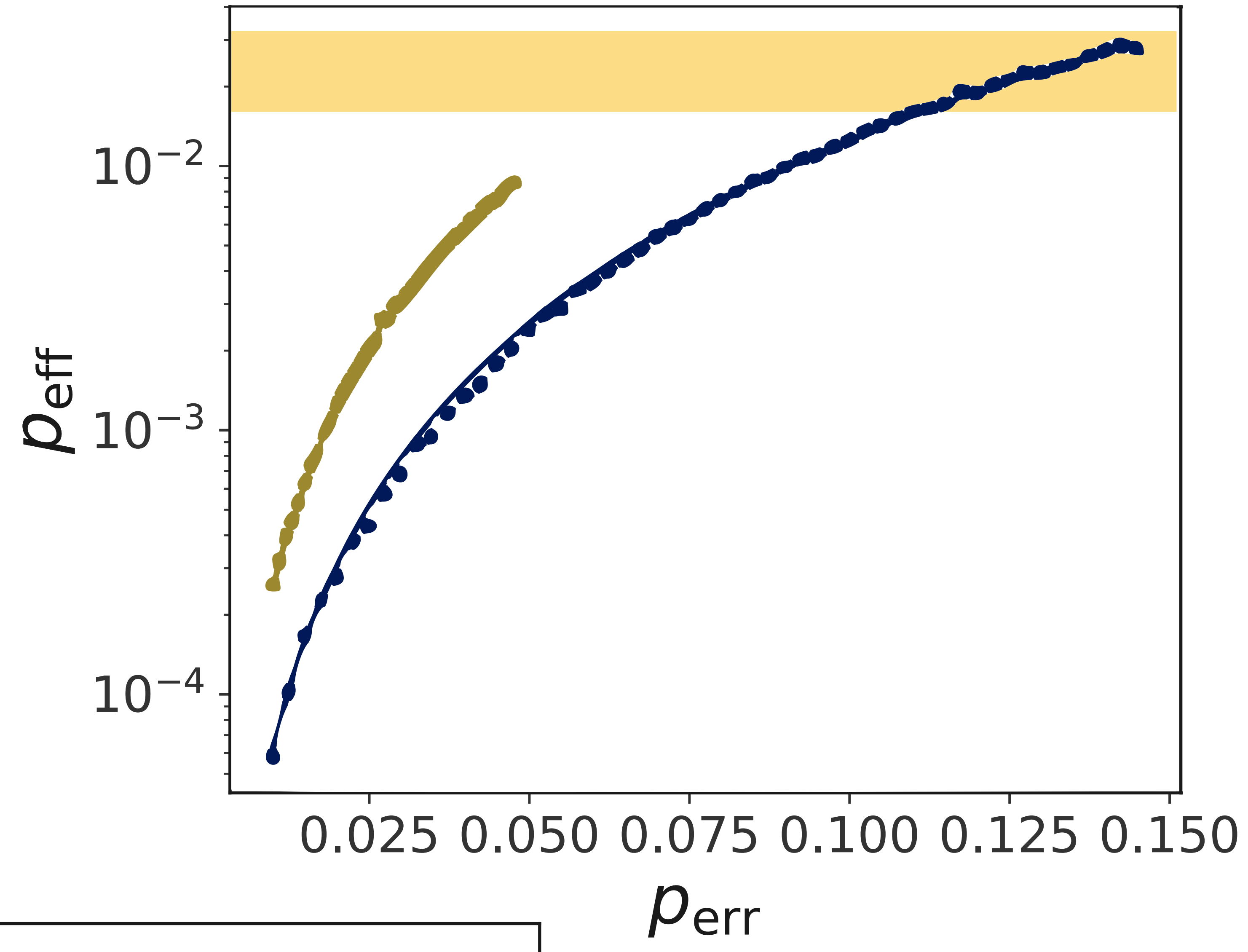
ML-assisted preprocessing

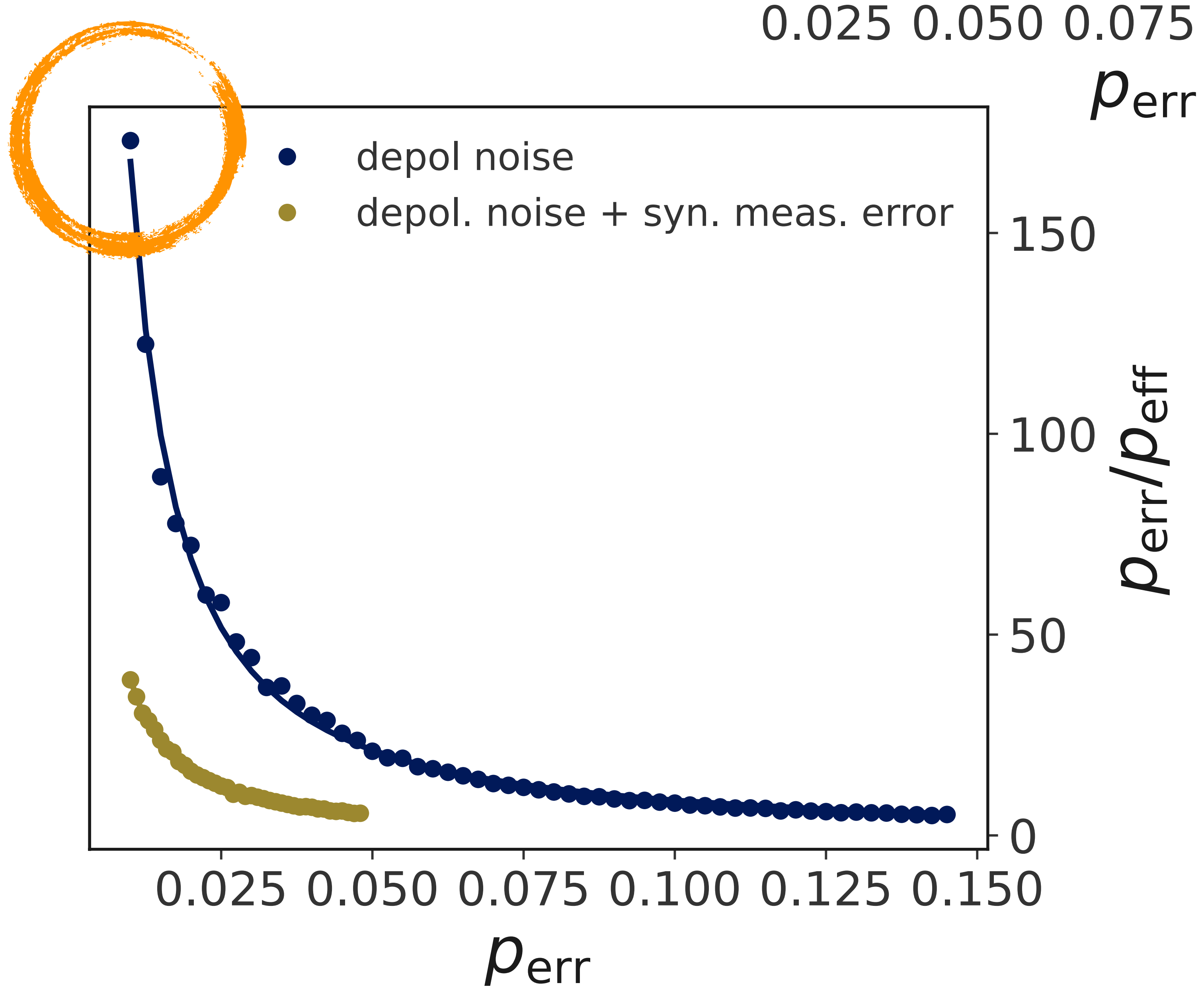


ML-assisted **preprocessing**

hierarchical decoding

conventional **decoder**



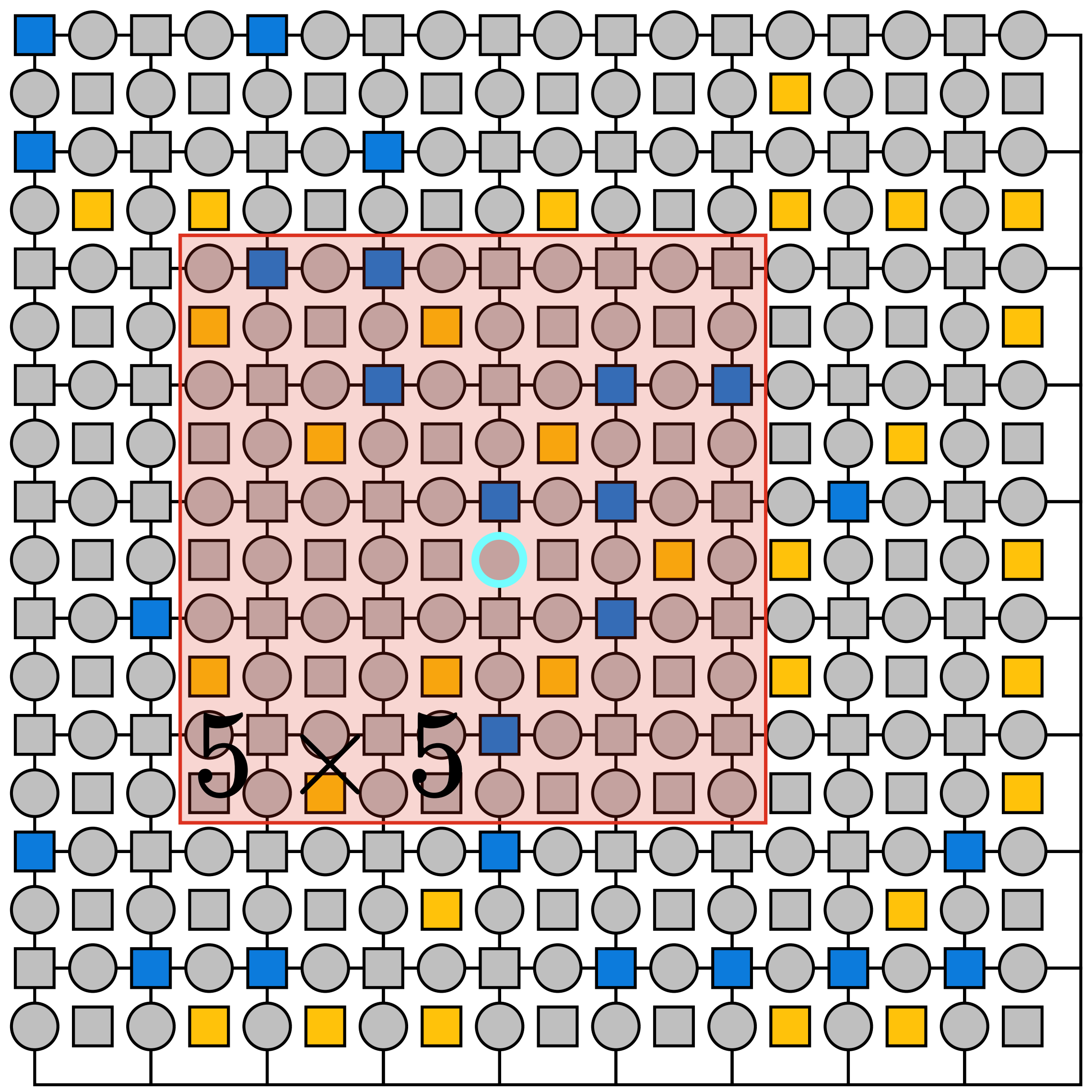


results

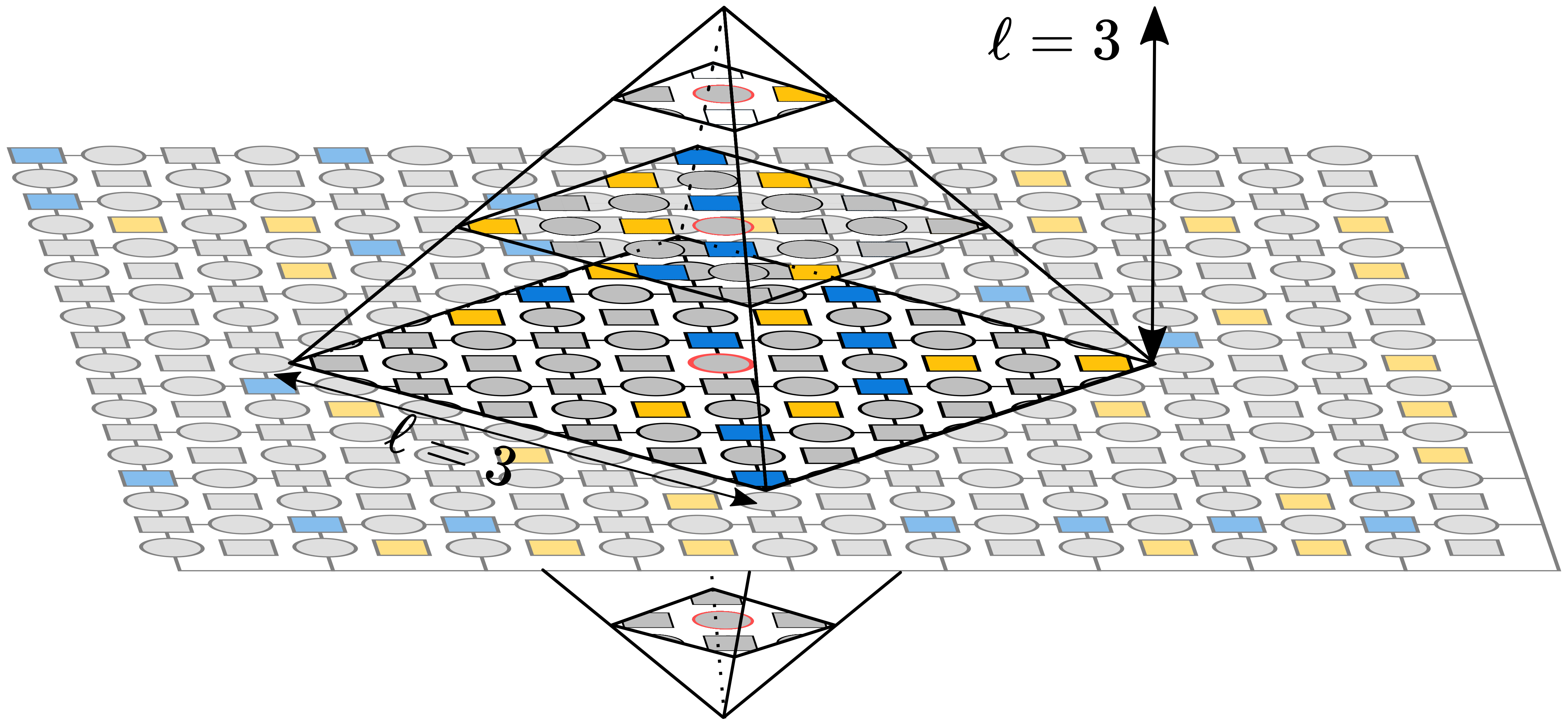
noise models

depolarizing noise

depolarizing noise + syndrome errors



two-dimensional setting

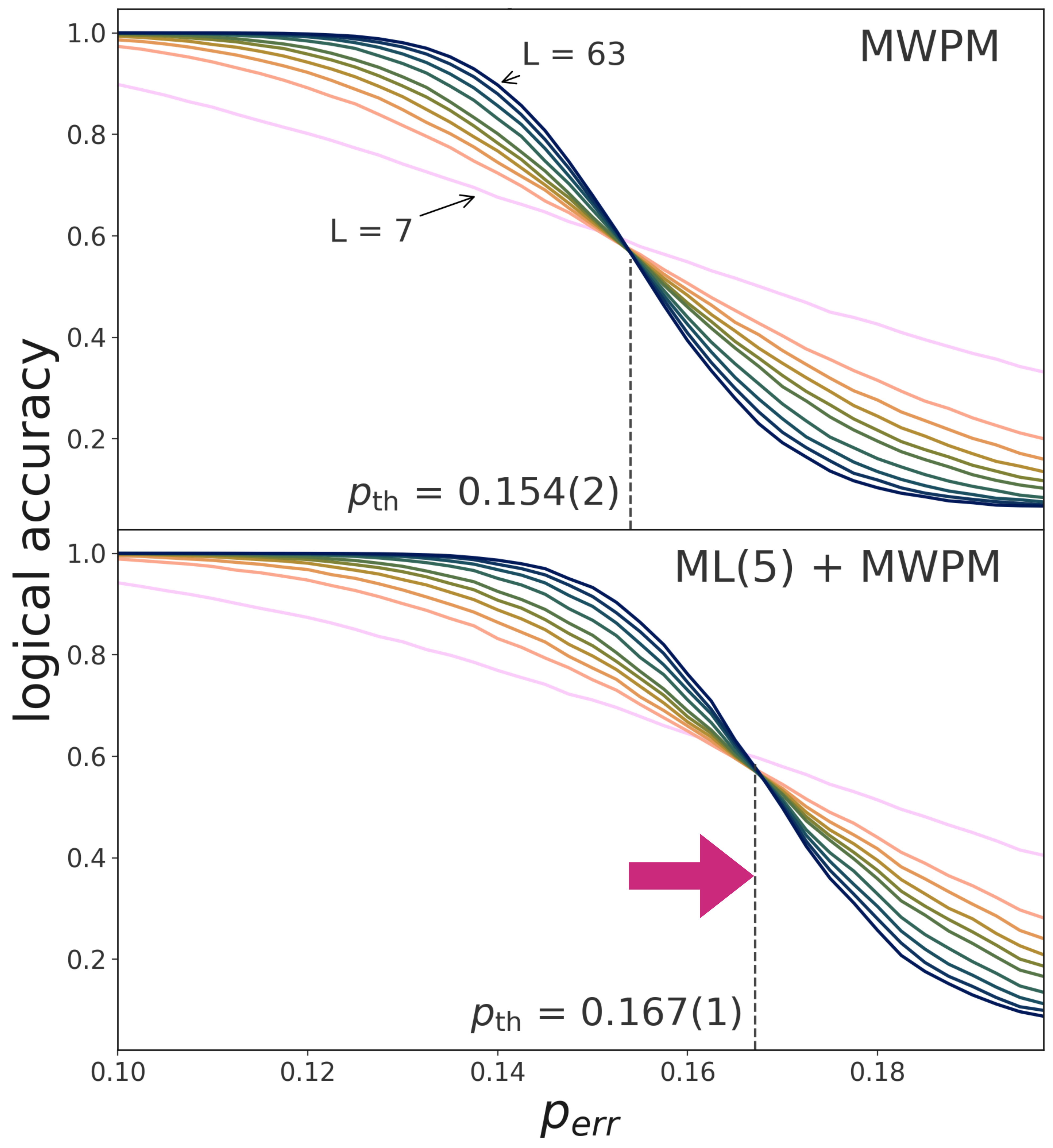


three-dimensional setting

smaller code distances

preprocessing + MWPM

depolarizing noise



threshold theorem

arbitrarily long, reliable computation is possible provided the error rate is below the threshold value

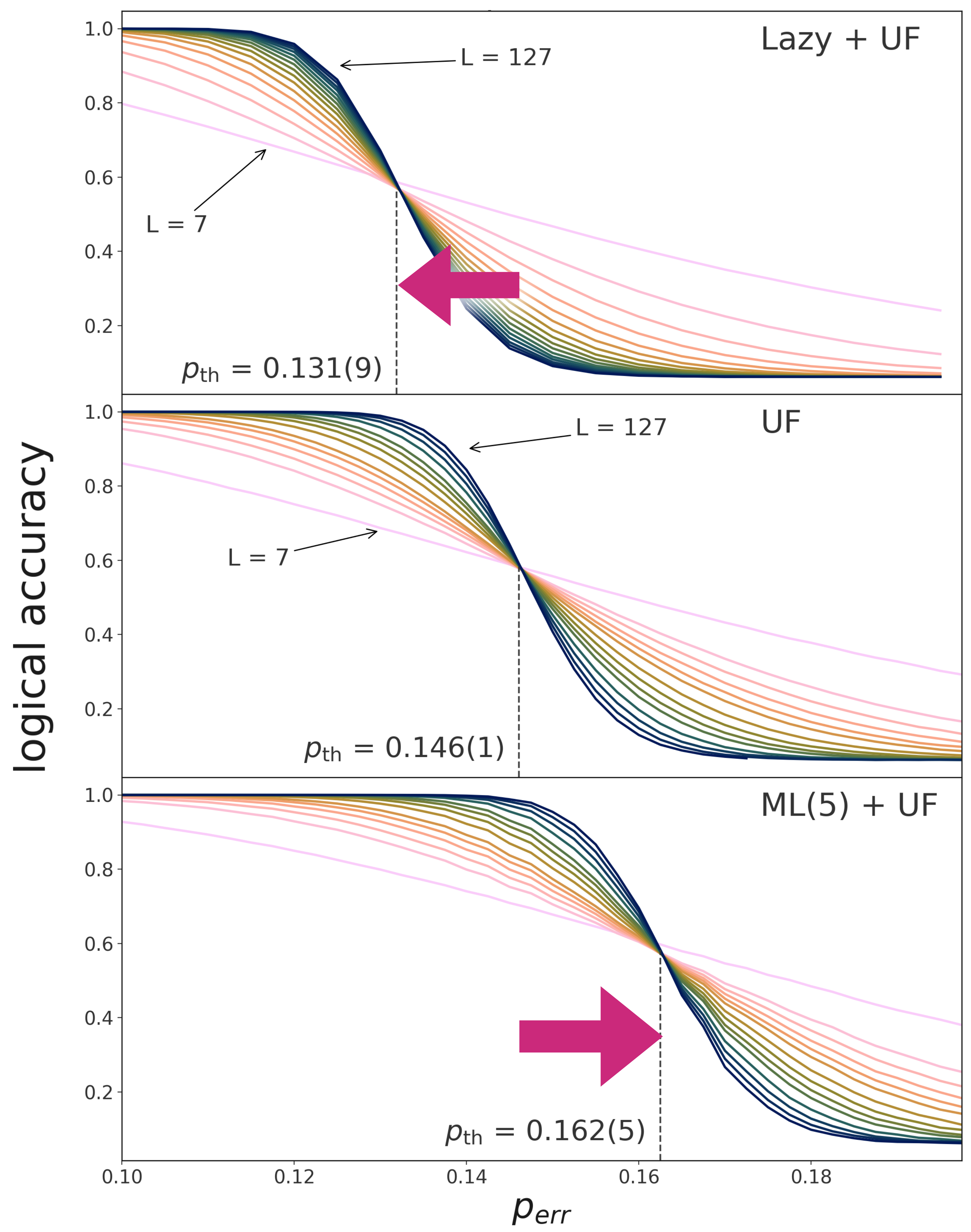
ML-assisted preprocessing allows us to **push up the error threshold** for MWPM

(about 10% increase)

preprocessing + UF

depolarizing noise

N. Delfosse, arXiv:2001.11427



"lazy preprocessing" (strictly local) is fast but **pushes down the error threshold** for UF

(about 10% decrease)

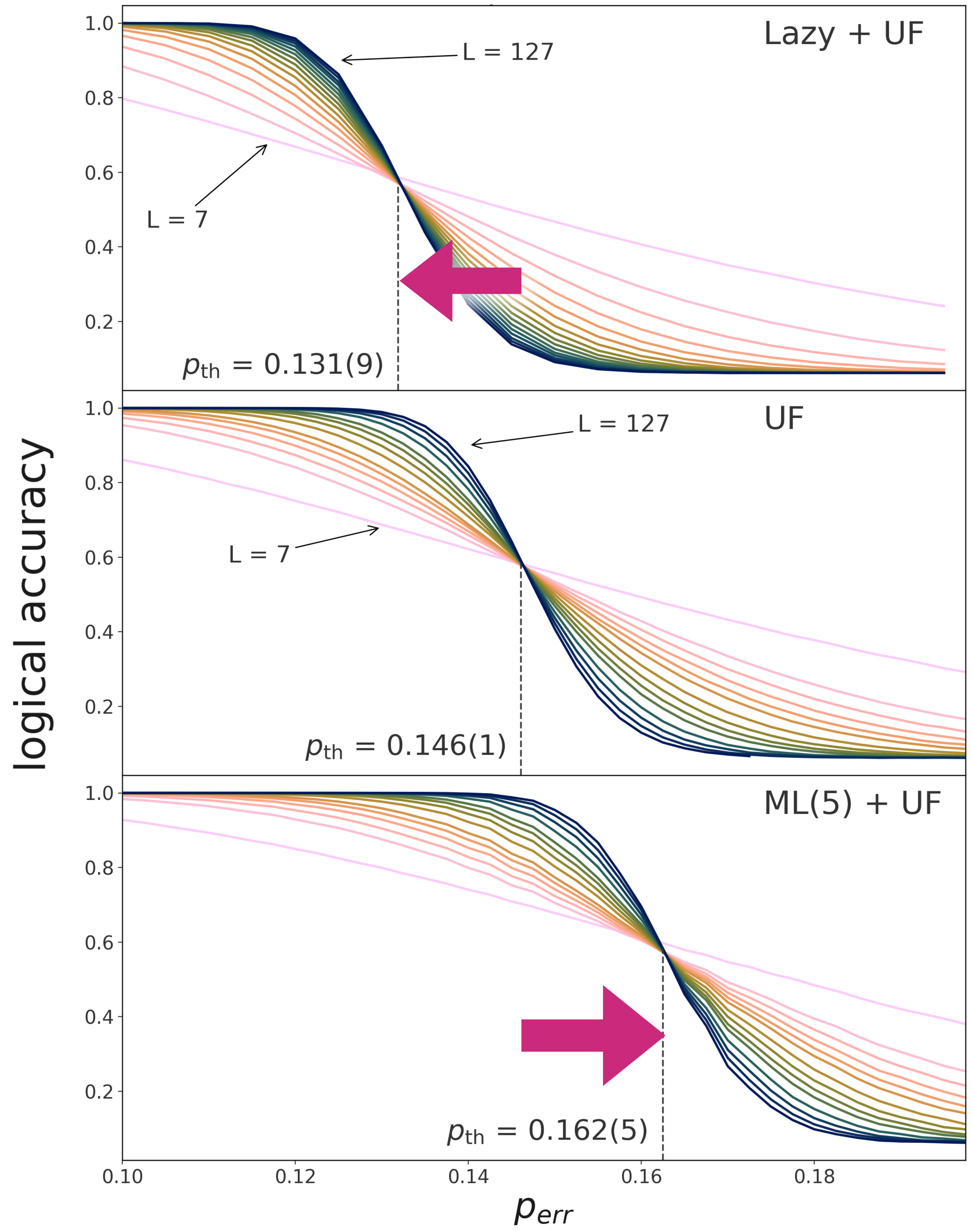
bare UF decoder has a **slightly lower threshold** than MWPM decoder

ML-assisted preprocessing allows us to **push up the error threshold** for UF

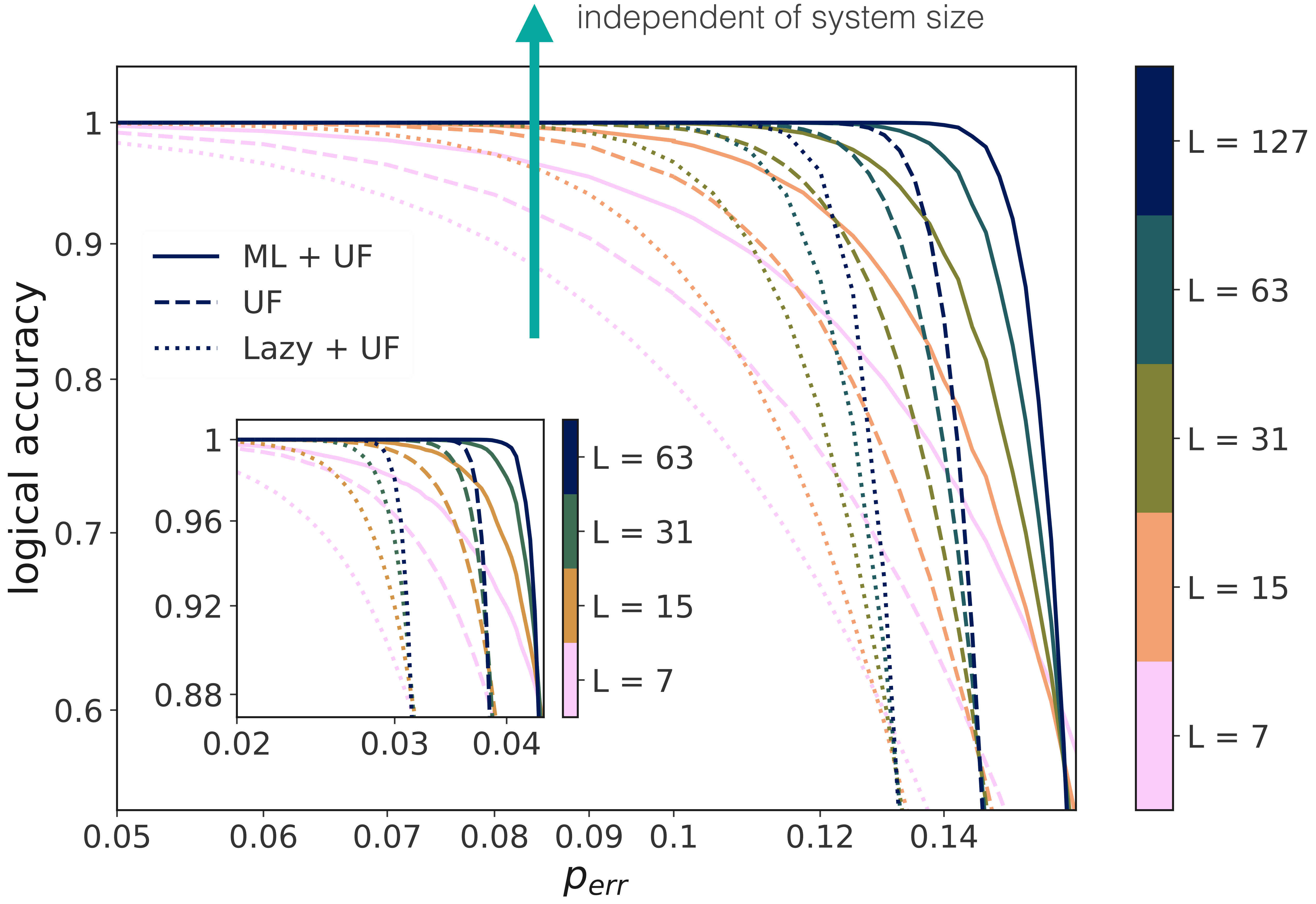
(about 10% increase)

preprocessing + UF

depolarizing noise

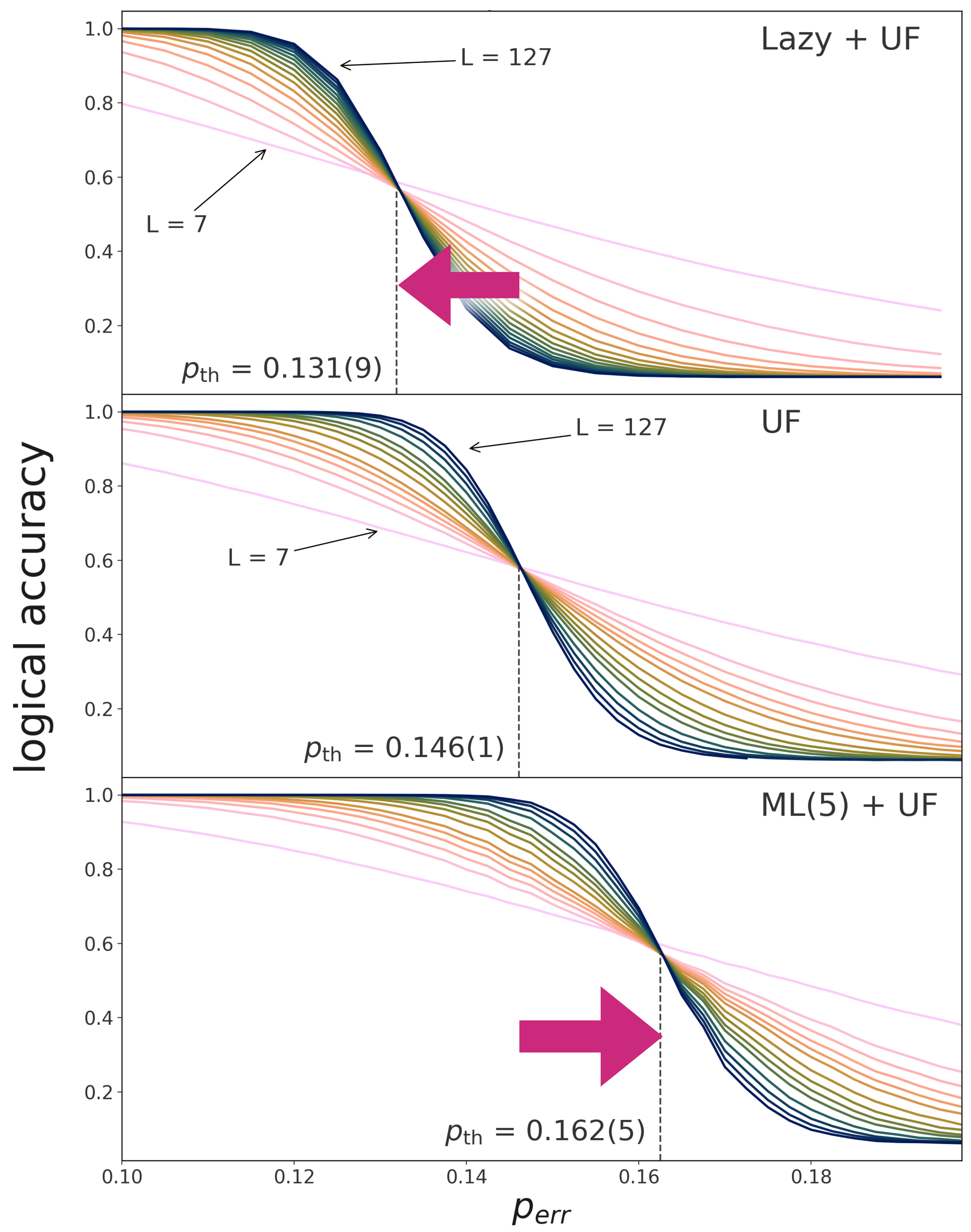


higher logical accuracy for ML+UF independent of system size

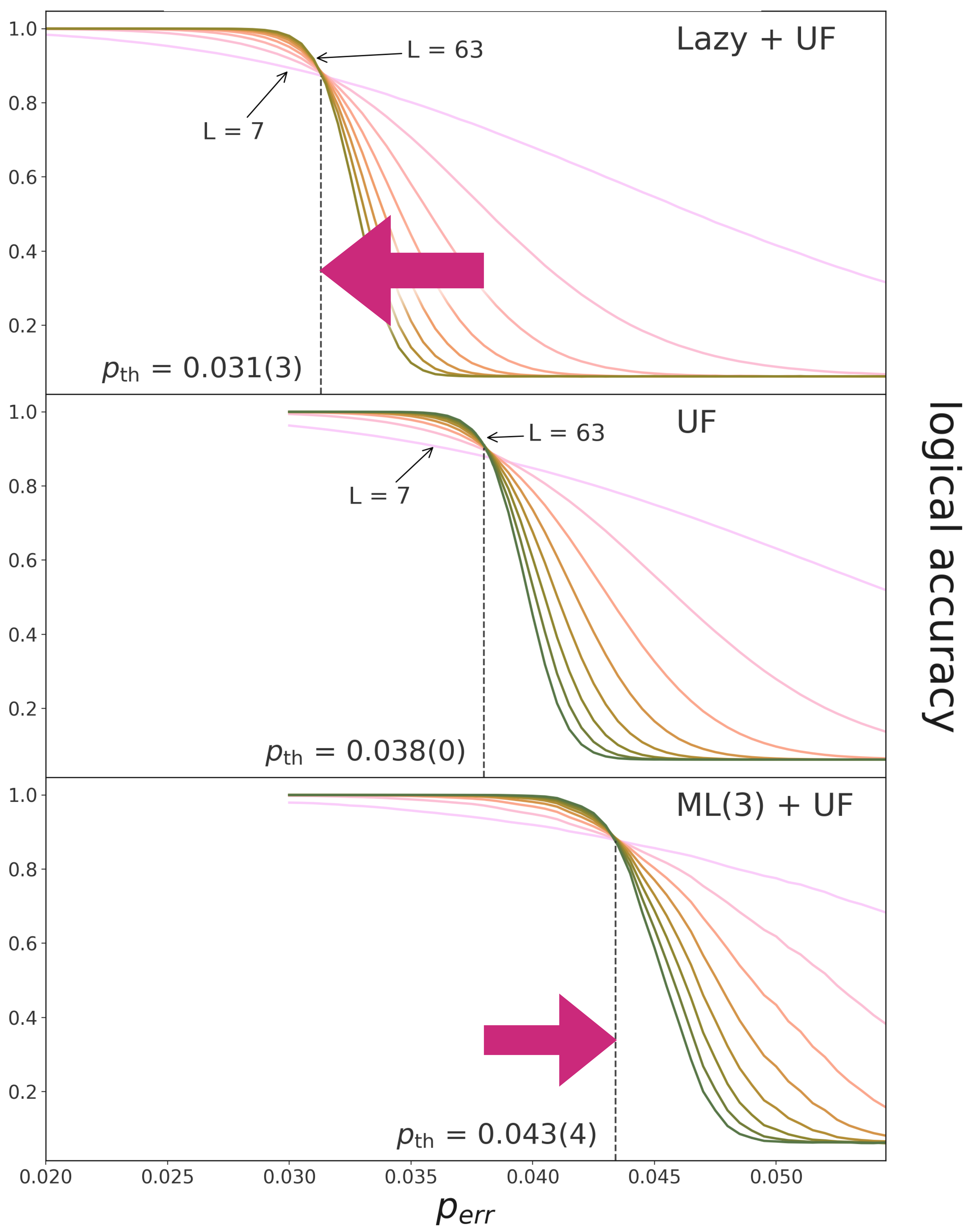


preprocessing + UF

depolarizing noise



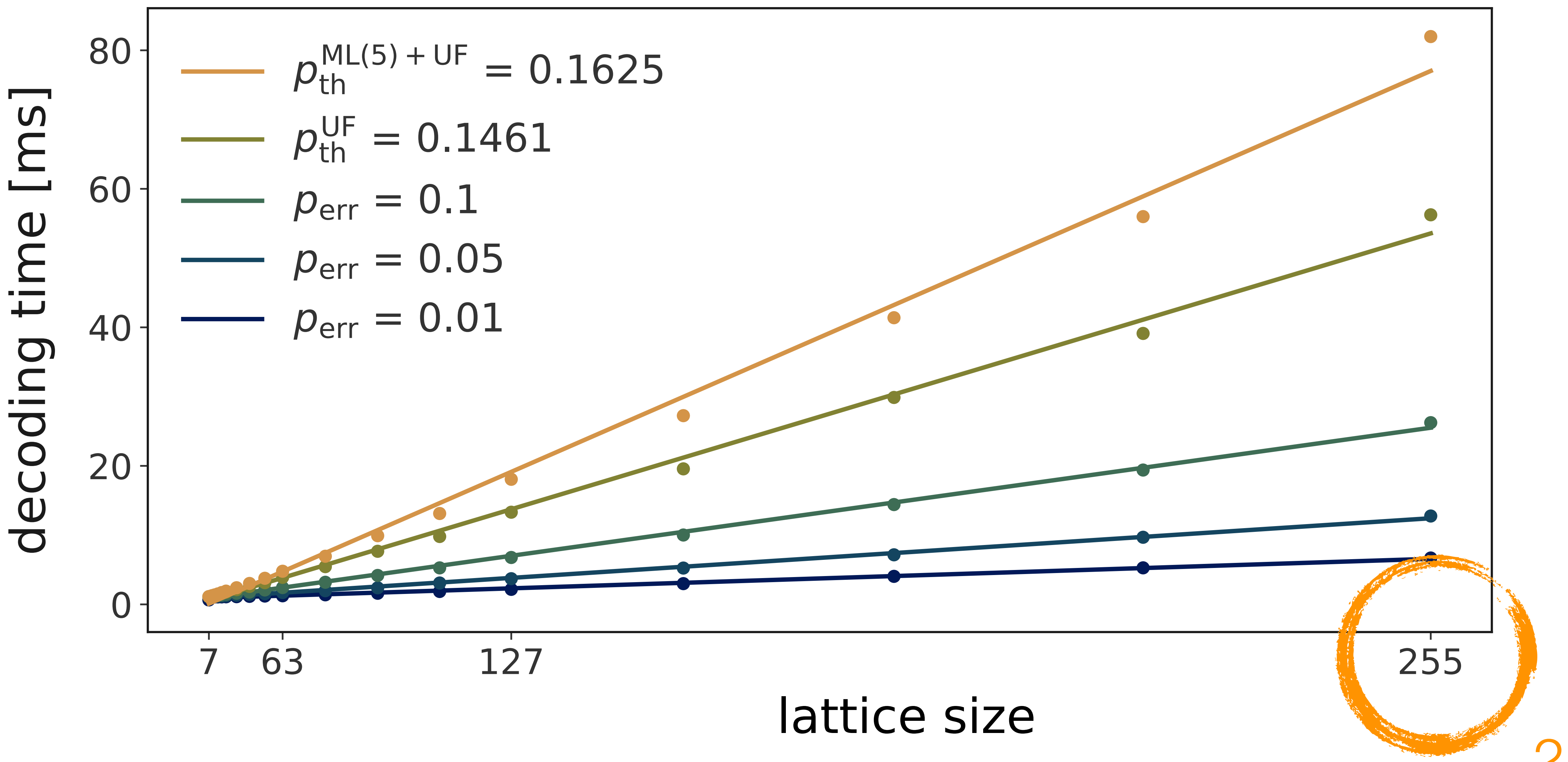
depolarizing noise + syndrome errors

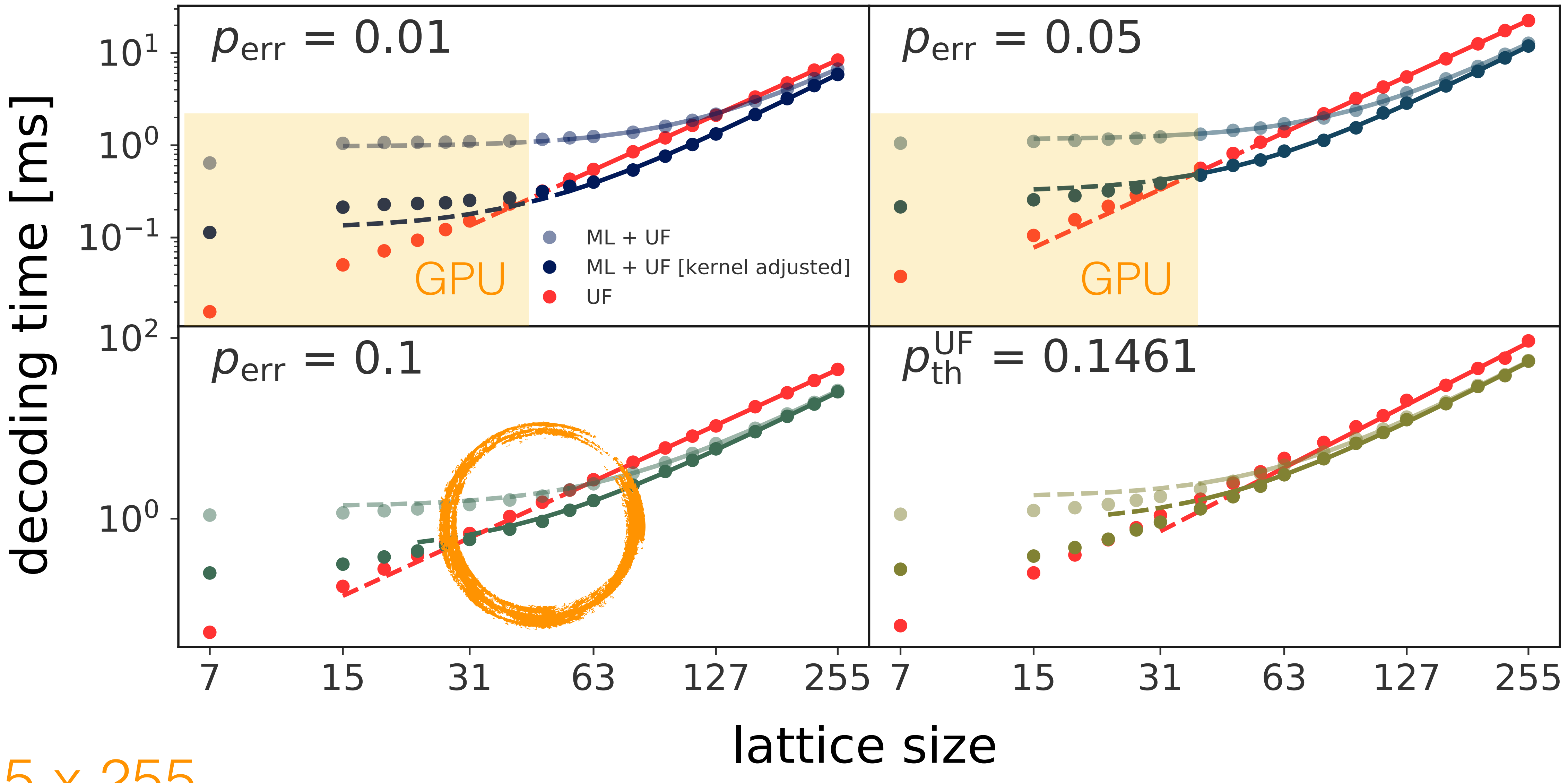


including
syndrome errors
do not qualitatively alter
these observations

compute times

algorithmic scaling





2 x 255 x 255
≈ **130,000 qubits**

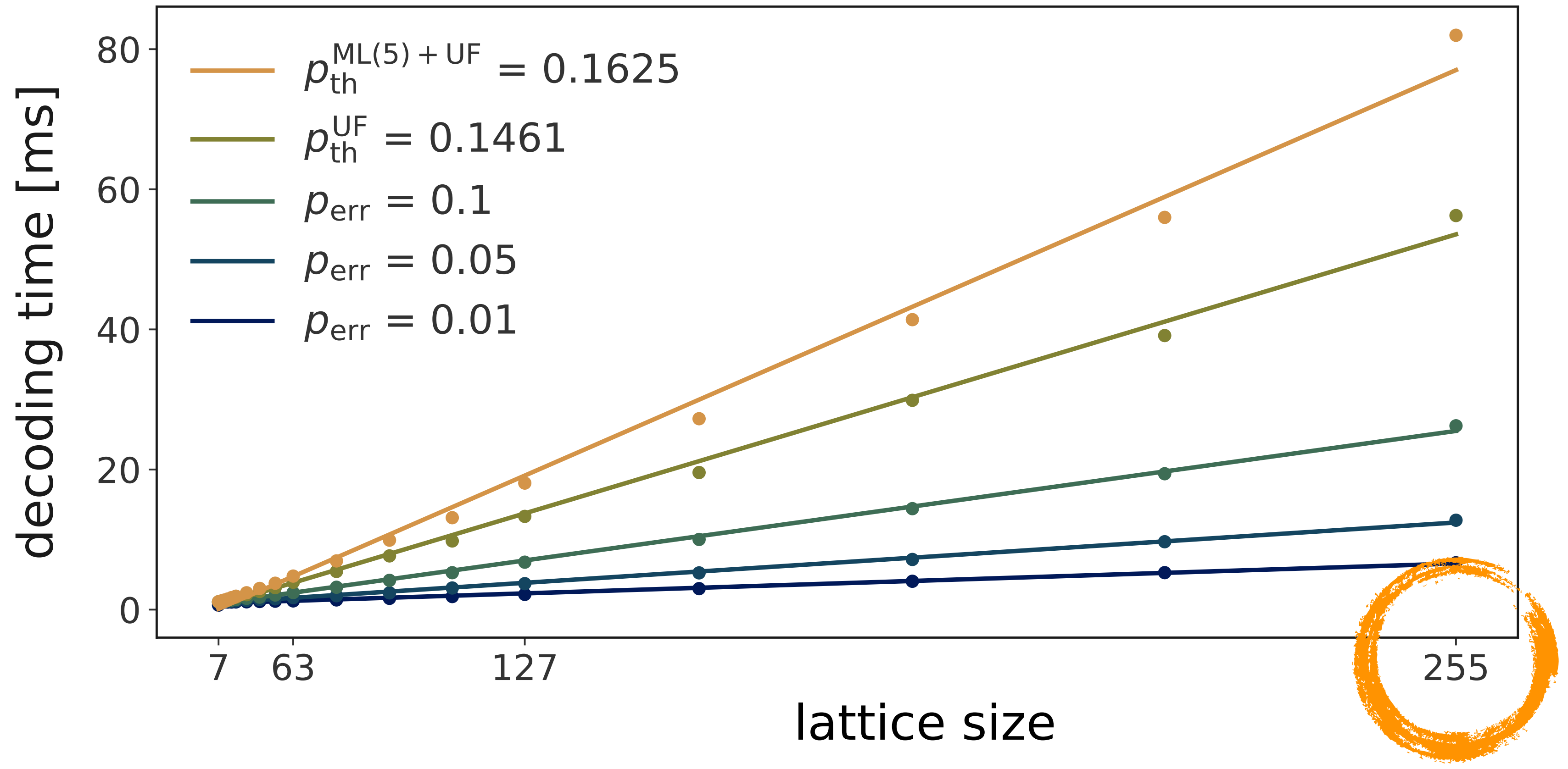
algorithmic scaling

- **asymptotic scaling** close to linear similar to bare UF decoder
- scalable to **tens of thousands** of qubits

comparison to bare UF decoder

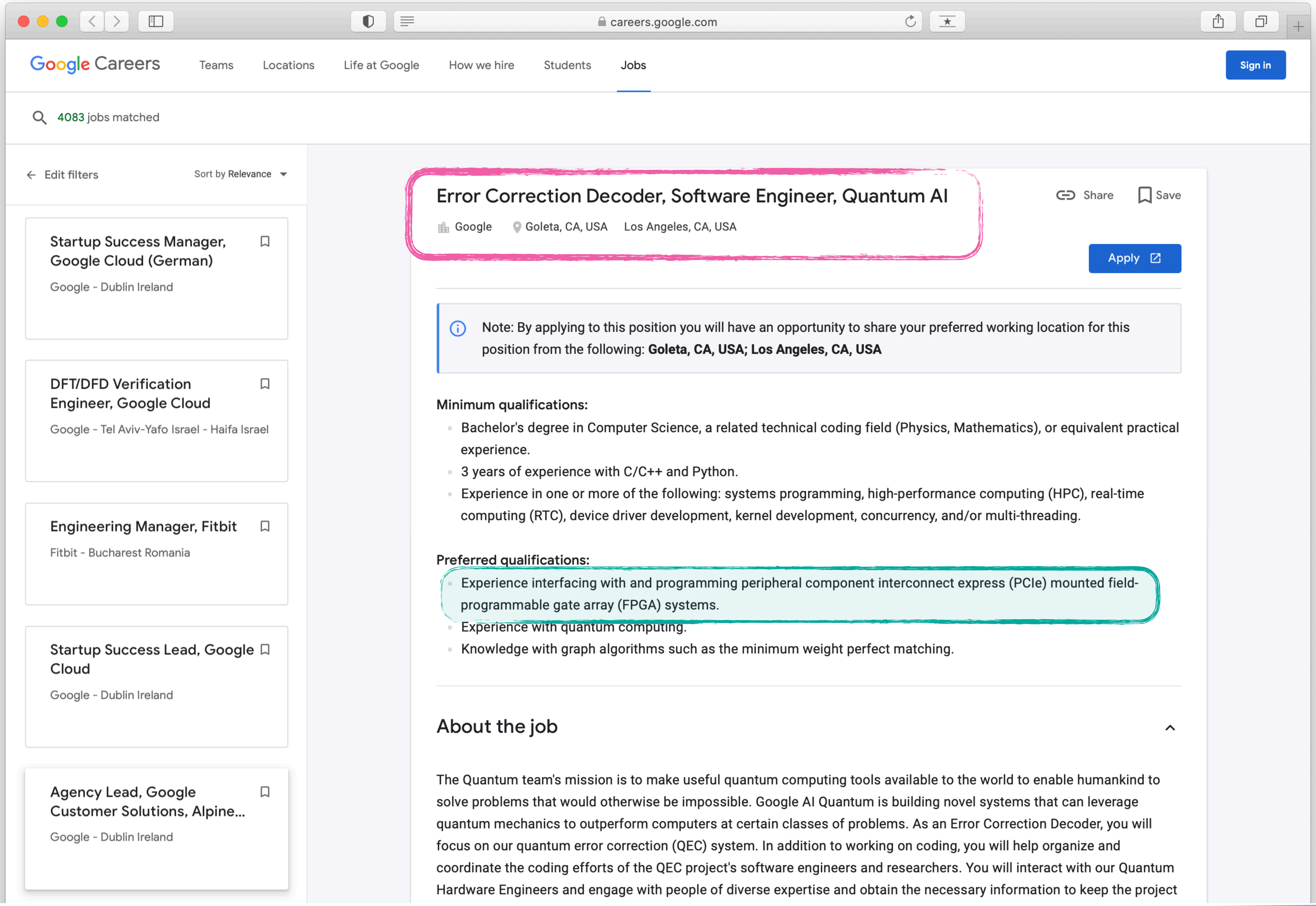
- **speedup** in asymptotic regime
- **breakdown** of (linear) scaling for small code distances

dedicated hardware



2 x 255 x 255
≈ **130,000 qubits**

- **tighter hardware integration**
  e.g. FPGAs or TPUs
- **in situ** decoding



compute times / comparison

depolarizing noise

threshold optimized

algorithm	p_{th}	$t_{p=0.01}$	$t_{p=0.05}$	$t_{p=0.1}$	$t_{p=0.1461}$
ML(7) + UF	**0.167(0)**	10.5	25.1	43.4	78.6
ML(5) + UF	0.162(5)	**6.7**	**12.8**	**26.2**	**56.2**
Lazy + UF	0.131(9)	6.9	20.7	51.1	—
UF	0.146(1)	8.4	22.5	44.9	92.8
ML(7) + MWPM	0.167(1)	~ 210	~ 530	~ 650	~ 980
ML(5) + MWPM	0.163(8)	~ 270	~ 510	~ 650	~ 970
MWPM	0.154(2)	~ 560	~ 840	~ 1100	~ 1300

speed optimized

preprocessing always wins independent of objective (speed vs. threshold)

depolarizing noise + syndrome errors

algorithm	p_{th}	$t_{p=0.01}$	$t_{p=0.02}$	$t_{p=0.03}$	$t_{p=0.0378}$
ML(3) + UF	0.043(4)	12.1	13.5	**15.4**	**17.8**
Lazy + UF	0.031(3)	**11.1**	**12.8**	16.6	—
UF	0.037(8)	11.5	13.4	15.7	18.9
ML(3) + MWPM*	**0.044(5)**	14.6	25.8	81.5	229
MWPM	0.043(7)	211	239	273	294

more nuanced observation:
lazy decoding for small error rates
ML-decoding for large error rates

conclusions

summary

Our hierarchical decoder with **ML-assisted preprocessing** simultaneously achieves

- **improved error threshold** for depolarizing noise (even w/ syndrome errors)
- **algorithmic scalability** up to tens of thousands of qubits
- **real-life wall-clock times** that best even those of the bare UF algorithm

This can be further improved by closer **hardware integration**, e.g. in situ decoding using FPGAs.

Paper

Scalable Neural Decoder for Topological Surface Codes
Kai Meinerz, Chae-Yeun Park & ST, PRL **128**, 080505 (2022).

Code

Open-source C++ implementation of the Union-Find decoder
Chae-Yeun Park & Kai Meinerz, https://github.com/chaeyeunpark

Institute for Theoretical Physics, Cologne / Germany