A General Approach to Dropout in Quantum Neural Networks

<u>**Francesco Scala**</u>, Andrea Ceschini,
Massimo Panella, Dario Gerace

UNIVERSITÀ DI PAVIA
Department of Physics

SAPIENZA
UNIVERSITÀ DI ROMA

21/11/2023

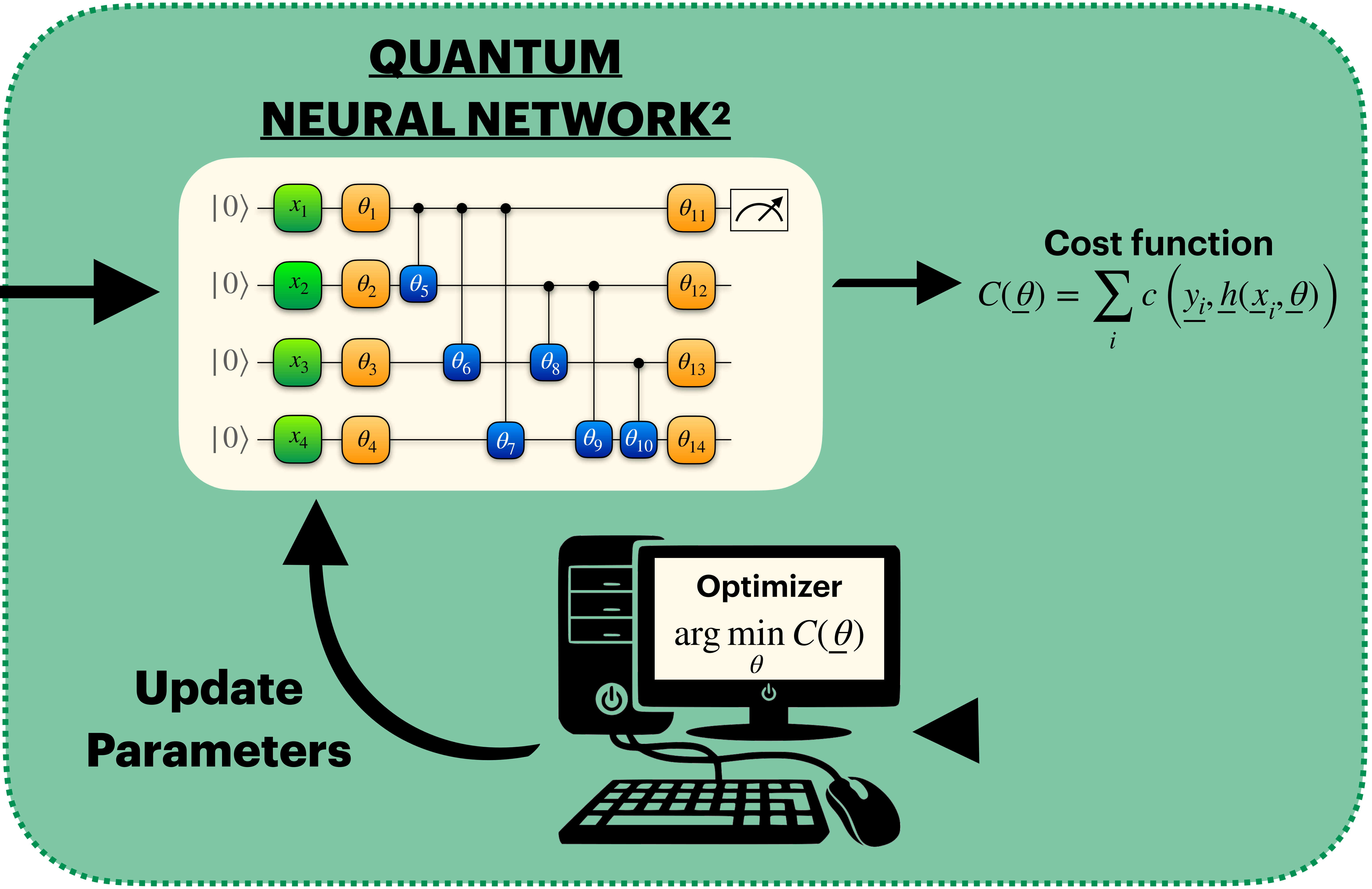
QTML
Quantum Techniques
in Machine Learning 2023

arXiv: 2310.04120 (2023)

Quantum Machine Learning[1]

Input data

$$\mathscr{A} = \{\underline{x}_i\}_{i=1}^{N}$$



Task

[1] J. Biamonte, et al., **Nature 549, 195 (2017)** [2] S. Mangini *et al.*, ***EPL* 134 10002 (2021)**

Overparametrized QNNs

$\mathscr{F}_x$ is the Quantum Fisher Information Matrix
→ **Hessian** of fidelity w.r.t. parameters

$$[\mathcal{F}_x(\boldsymbol{\theta})]_{ij} = 4\mathrm{Re}[\langle \partial_i \psi(x,\boldsymbol{\theta}) | \partial_j \psi(x,\boldsymbol{\theta})\rangle - \langle \partial_i \psi(x,\boldsymbol{\theta}) | \psi(x,\boldsymbol{\theta})\rangle \langle \psi(x,\boldsymbol{\theta}) | \partial_j \psi(x,\boldsymbol{\theta})\rangle]$$

A QNN is **overparametrized** if

$$\mathbb{E}_x\left[rank\left(\mathscr{F}_x(\theta_{rand})\right)\right]$$

does not increase when adding parametrized operations

T. Haug, et al., **PRX Quantum 2, 040309 (2021)**
M. Larocca, et al., **Nat. Comput. Sci. 3, 542 (2023)**

Overparametrized QNNs

$\mathscr{F}_x$ is the Quantum Fisher Information Matrix
→ **Hessian** of fidelity w.r.t. parameters

$$[\mathcal{F}_x(\boldsymbol{\theta})]_{ij} = 4\mathrm{Re}[\langle\partial_i\psi(x,\boldsymbol{\theta})|\partial_j\psi(x,\boldsymbol{\theta})\rangle - \langle\partial_i\psi(x,\boldsymbol{\theta})|\psi(x,\boldsymbol{\theta})\rangle\langle\psi(x,\boldsymbol{\theta})|\partial_j\psi(x,\boldsymbol{\theta})\rangle]$$

A QNN is **overparametrized** if

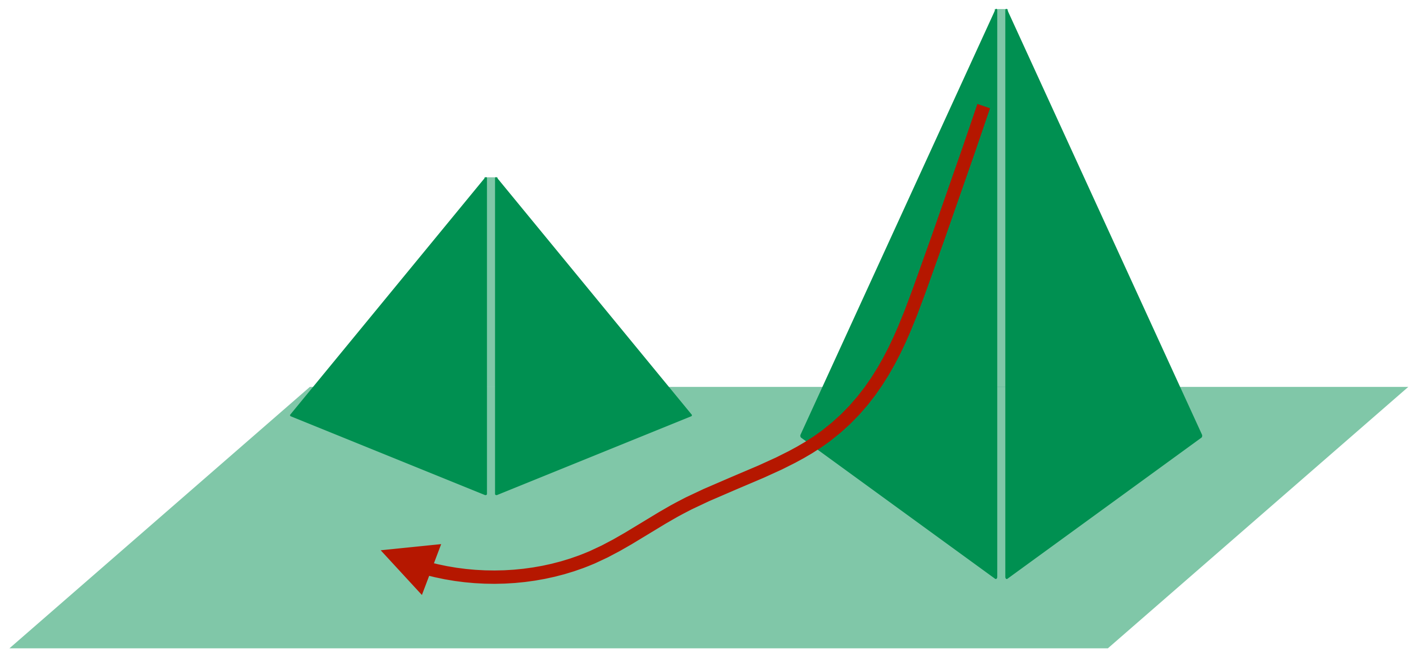
$$\mathbb{E}_x\left[rank\left(\mathscr{F}_x(\theta_{rand})\right)\right]$$

does not increase when adding parametrized operations

T. Haug, et al., **PRX Quantum 2, 040309 (2021)**
M. Larocca, et al., **Nat. Comput. Sci. 3, 542 (2023)**

IMPROVED TRAINABILITY

Computational phase transition
In the cost landscape



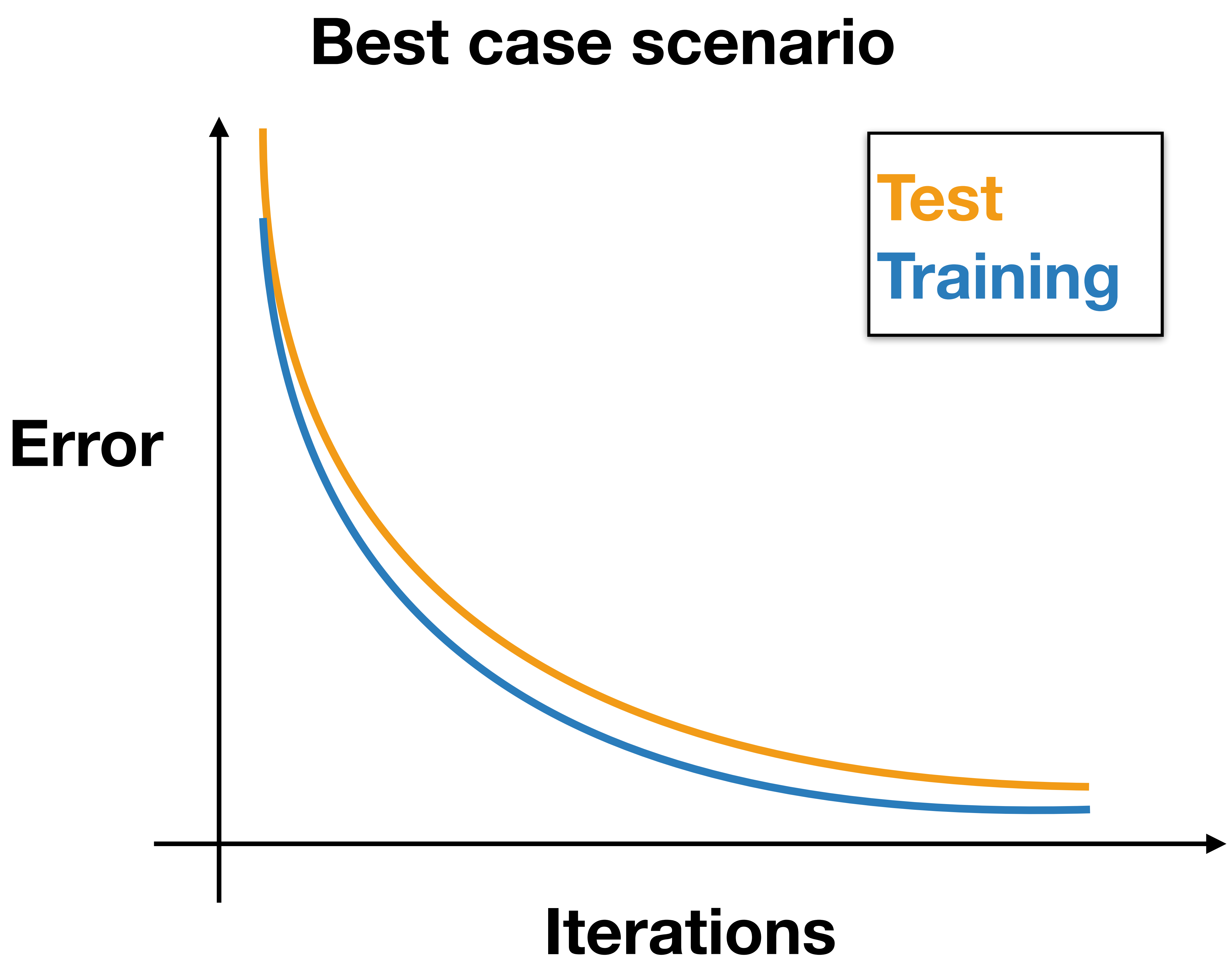
B. T. Kiani et al., **arXiv:2001.11897 (2020)**
J. Liu et al., **PRL 130, 150601 (2023)**

Generalization

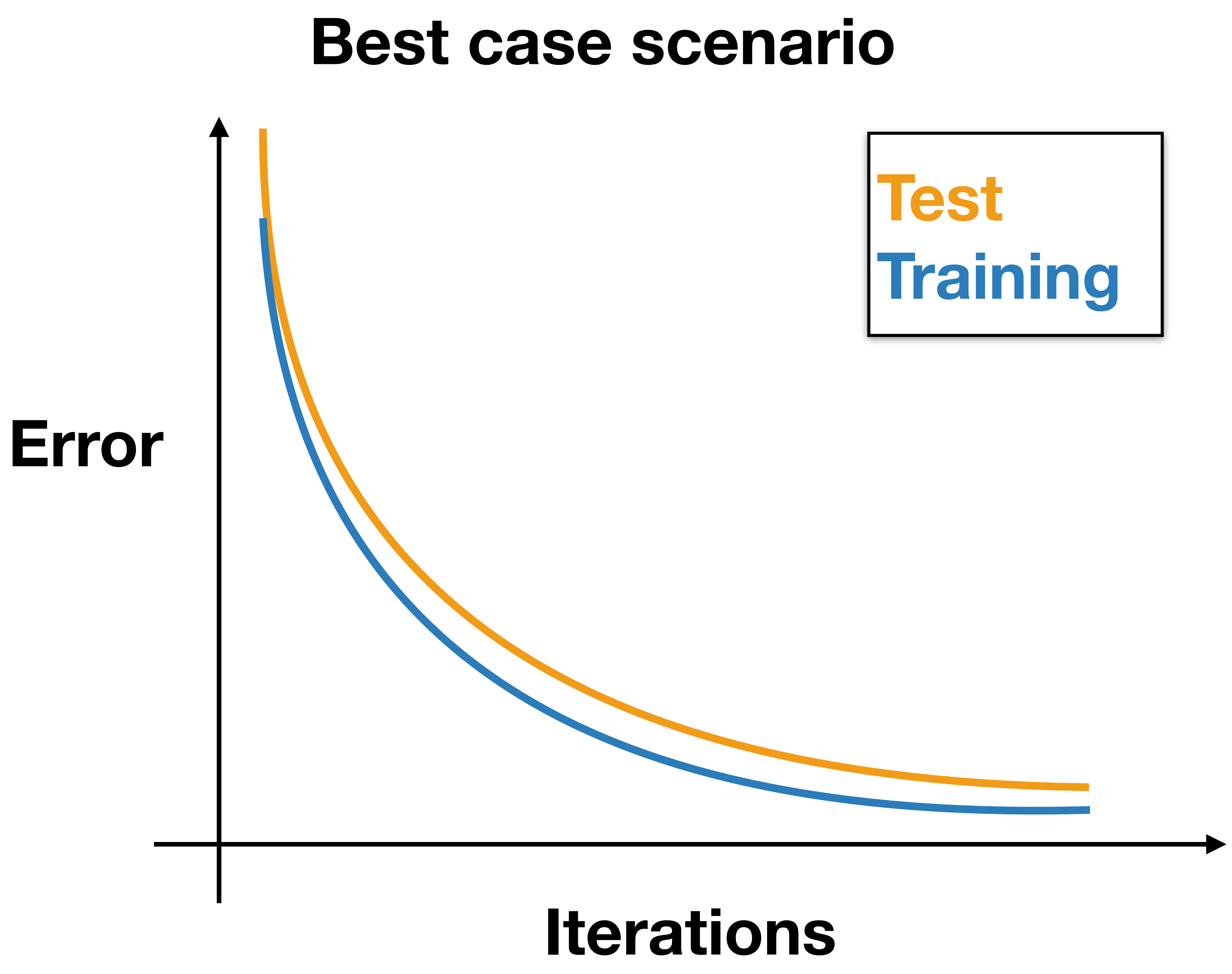
Training

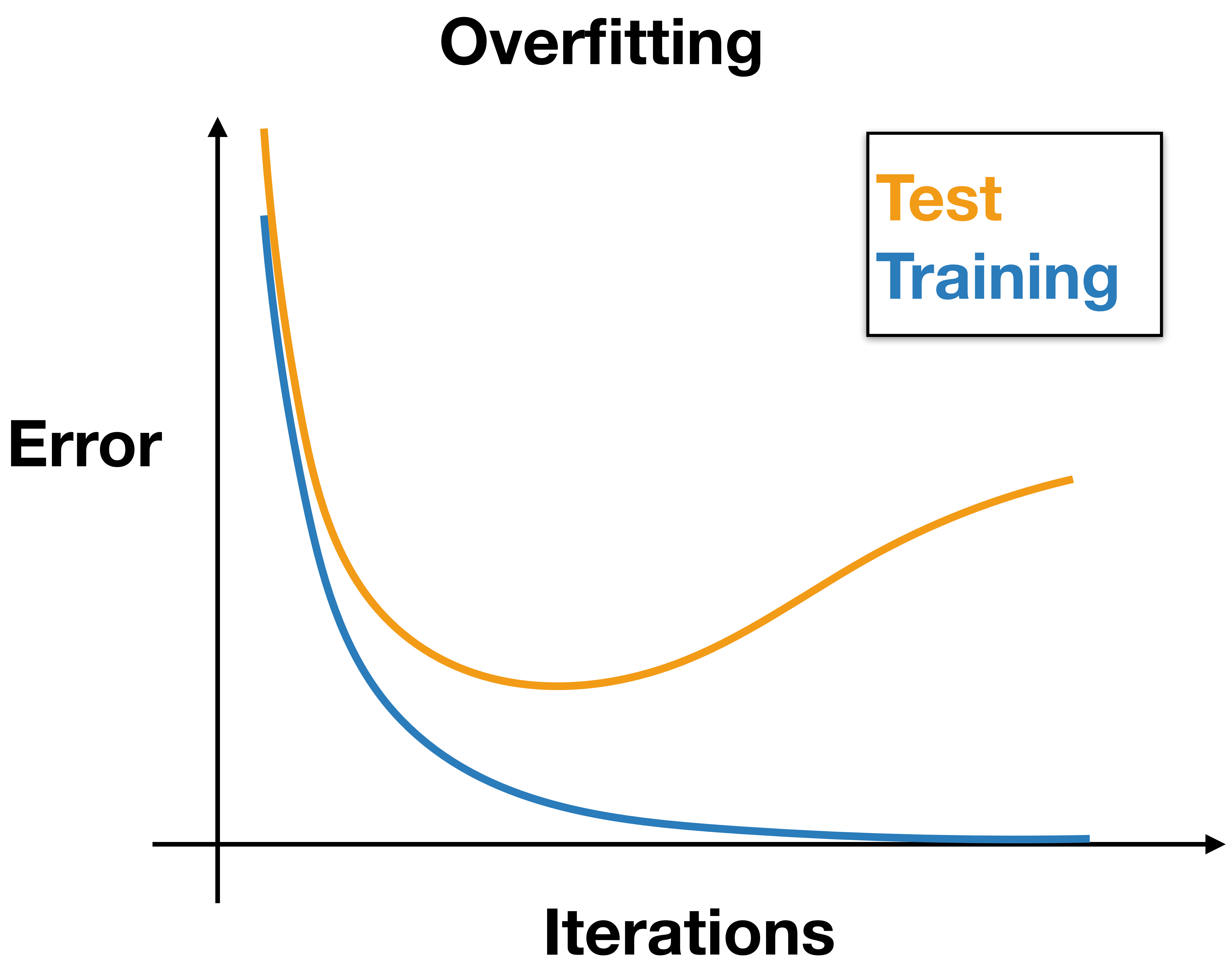
Test

Generalization and Overfitting



Generalization and Overfitting





Generalization and Overfitting

Best case scenario

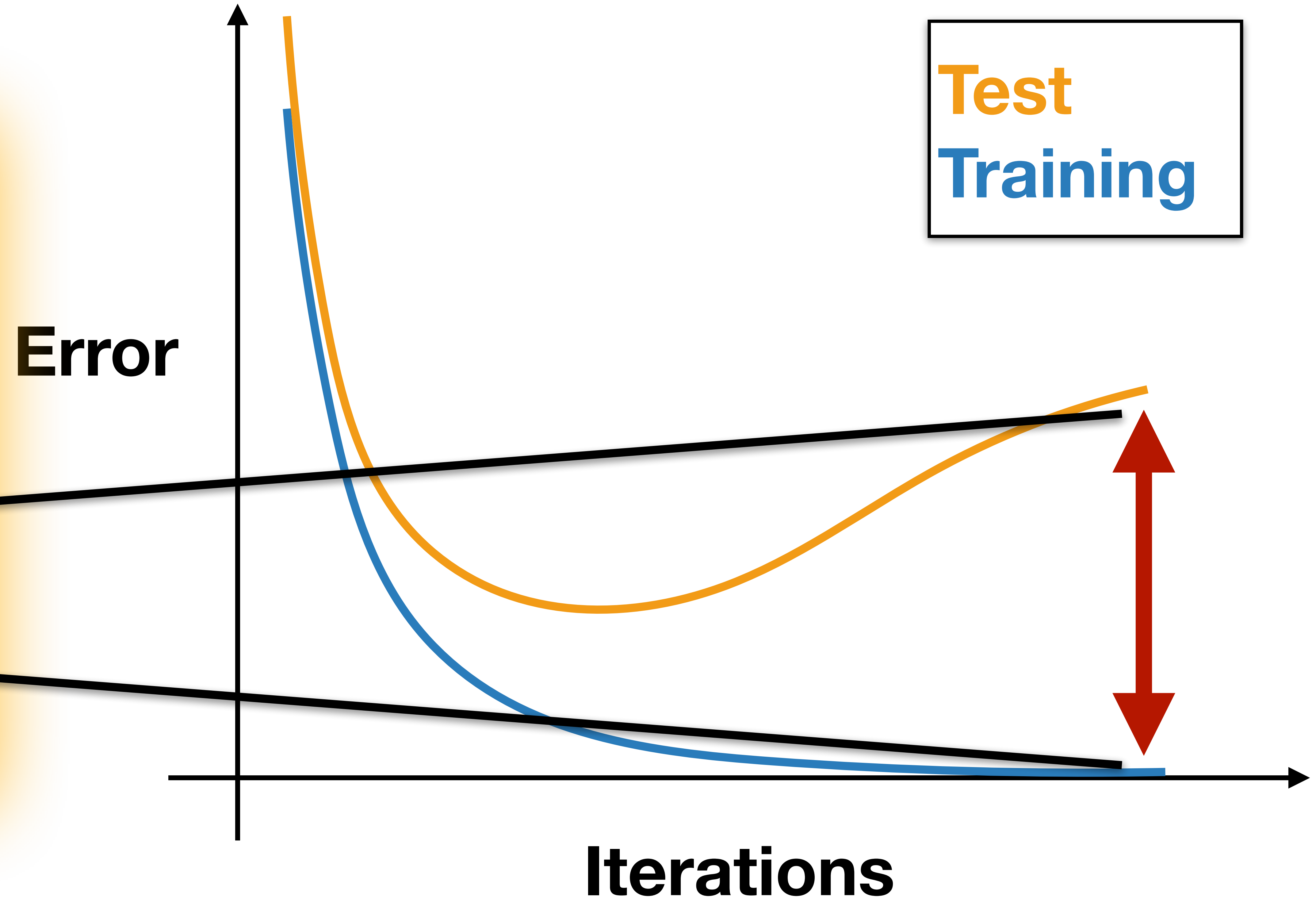
Iterations

Learn the **NOISE** in the data

High generalization gap witnesses

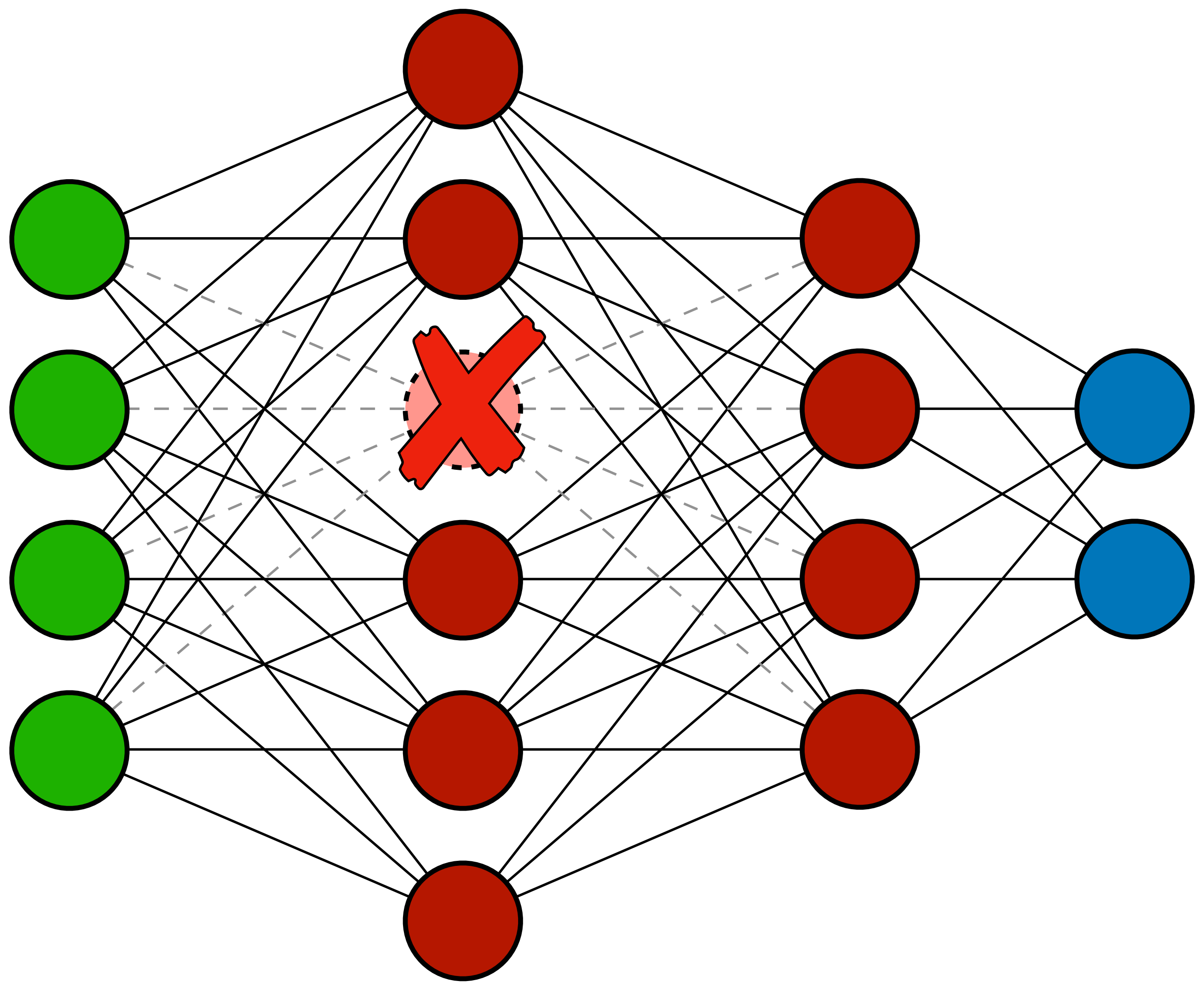
OVERFITTING

Overfitting



Classical Dropout

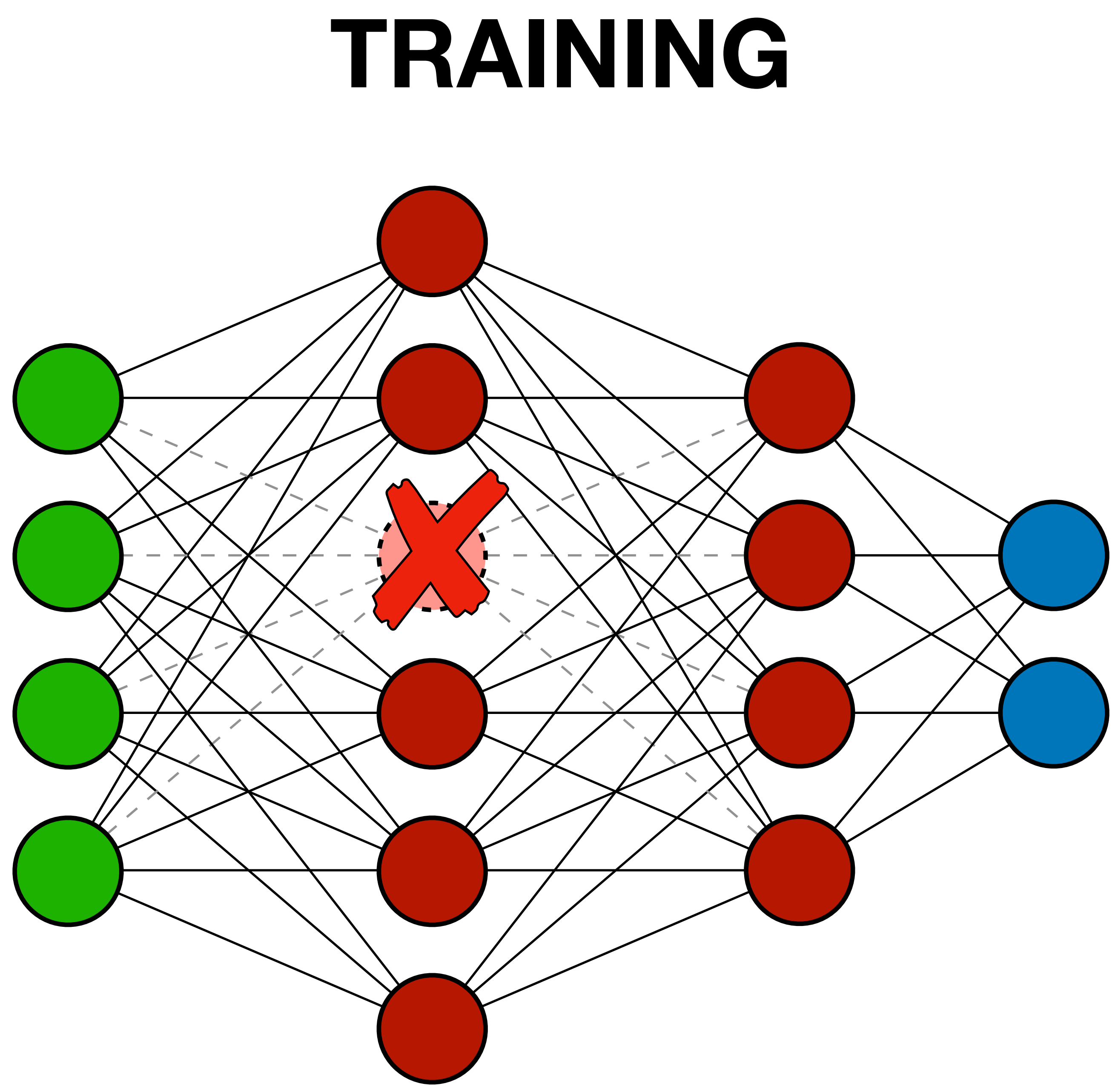
TRAINING



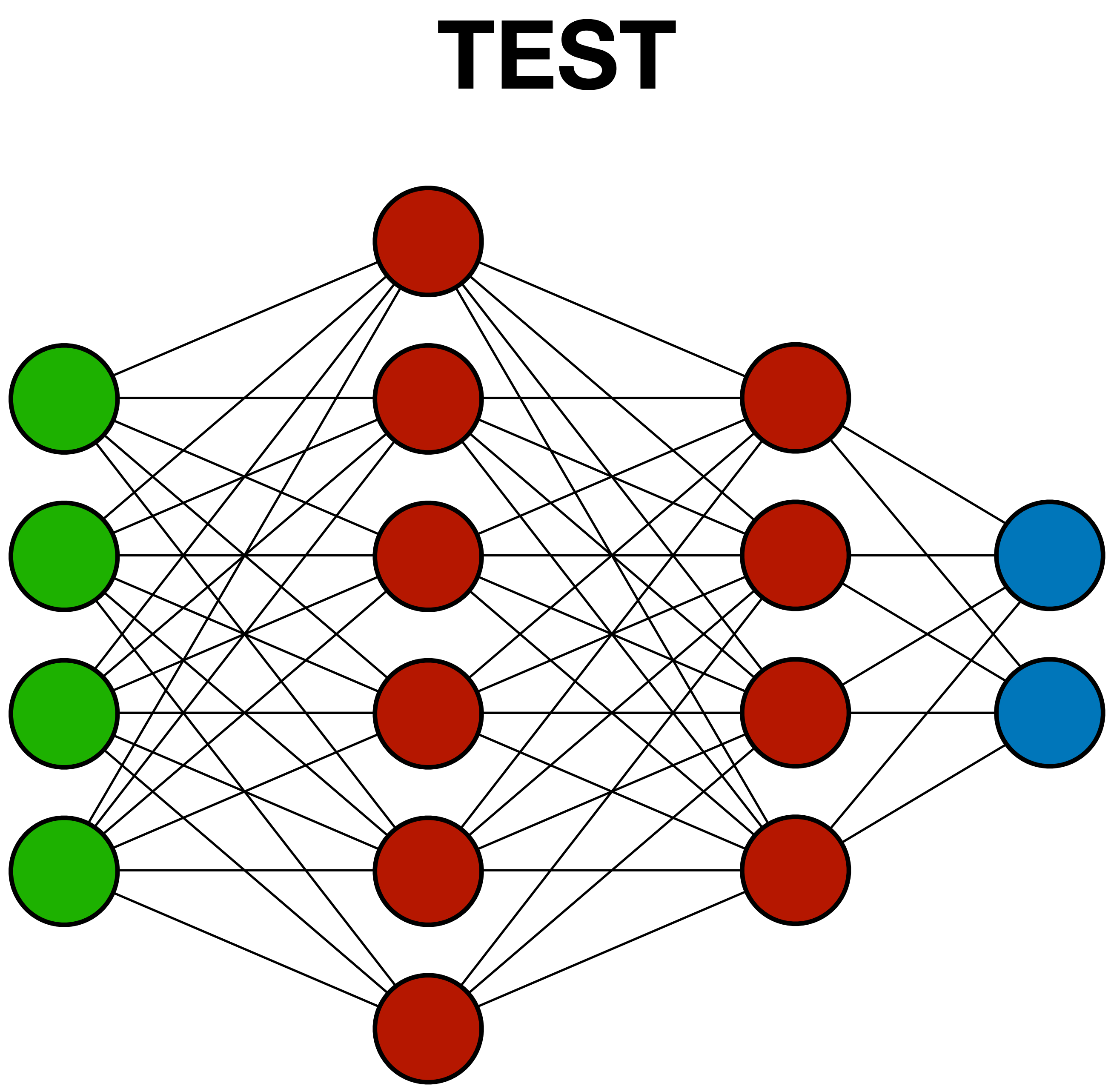
Randomly remove neurons in
OVERPARAMETRIZED NNs

N. Srivastava, et al., **Journal of Machine Learning Research (2014)**

Classical Dropout



Randomly remove neurons in **OVERPARAMETRIZED** NNs

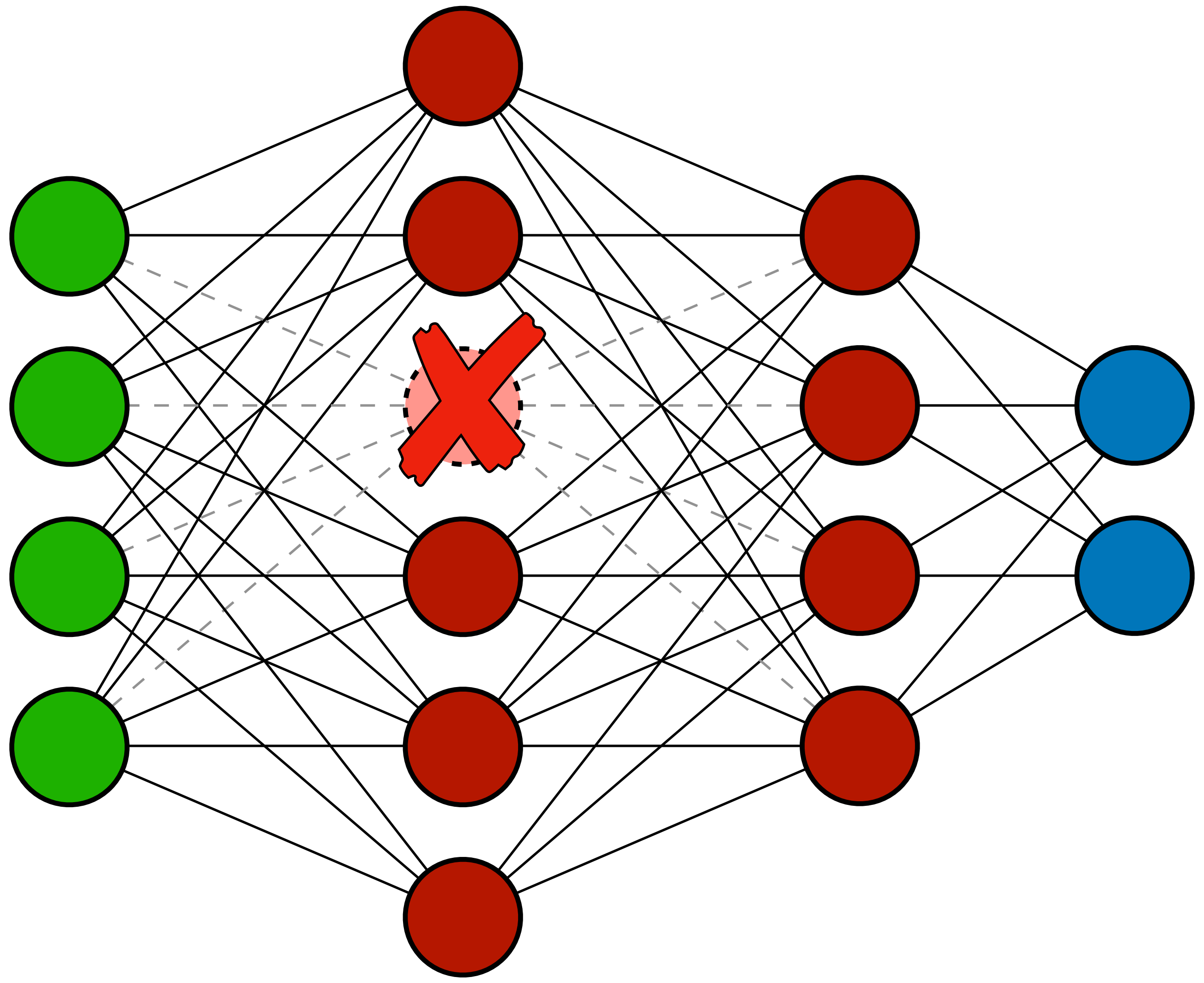


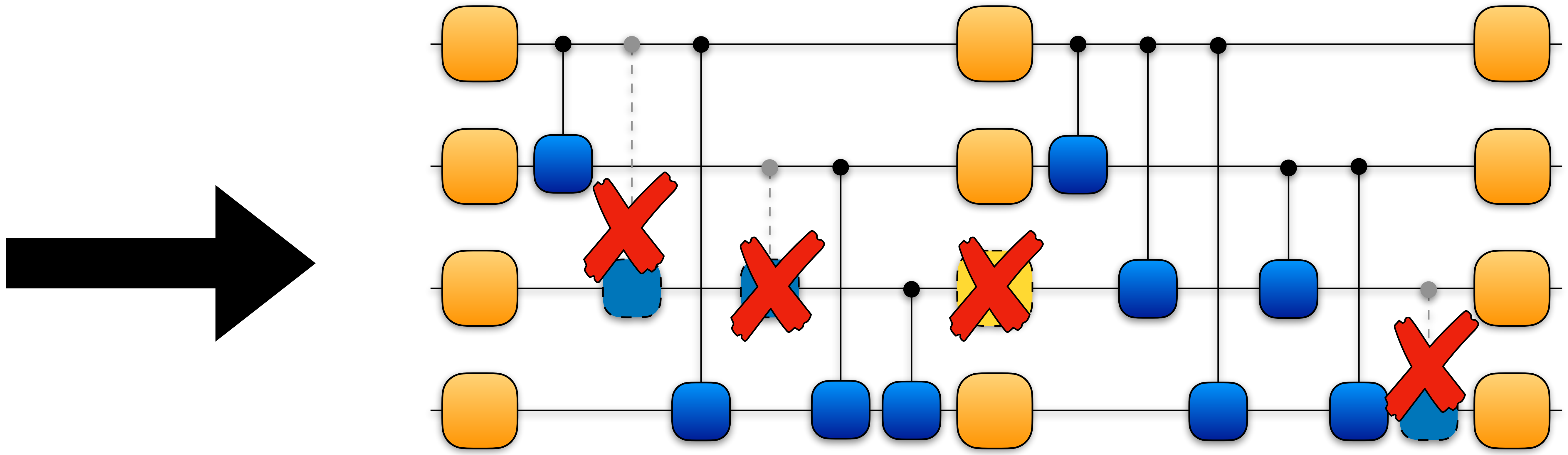
Employ the whole NN with **rescaled weights**

$$\underline{\theta}_{test} = (1 - p_{drop})\underline{\theta}_{train}$$

N. Srivastava, et al., **Journal of Machine Learning Research (2014)**

Quantum Dropout



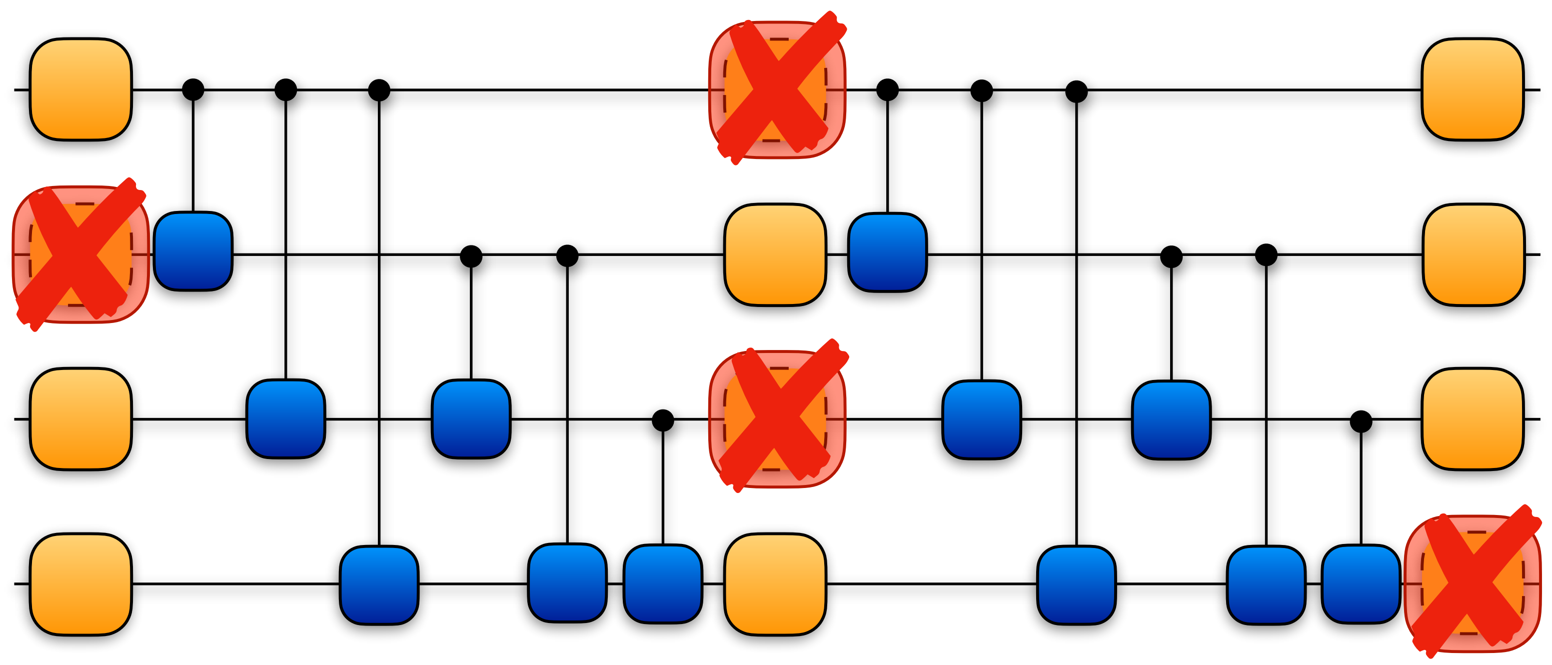


Randomly remove unitaries
during training in
OVERPARAMETRIZED QNNs

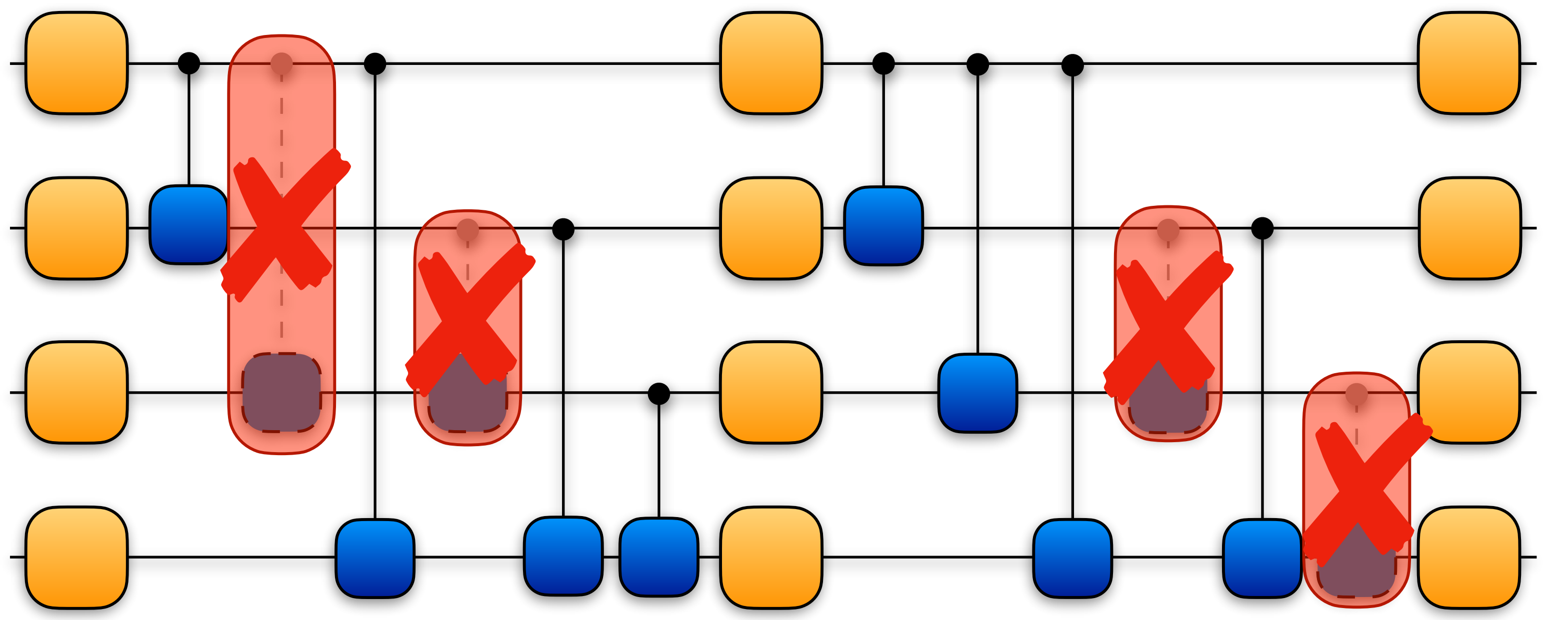
F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Different strategies

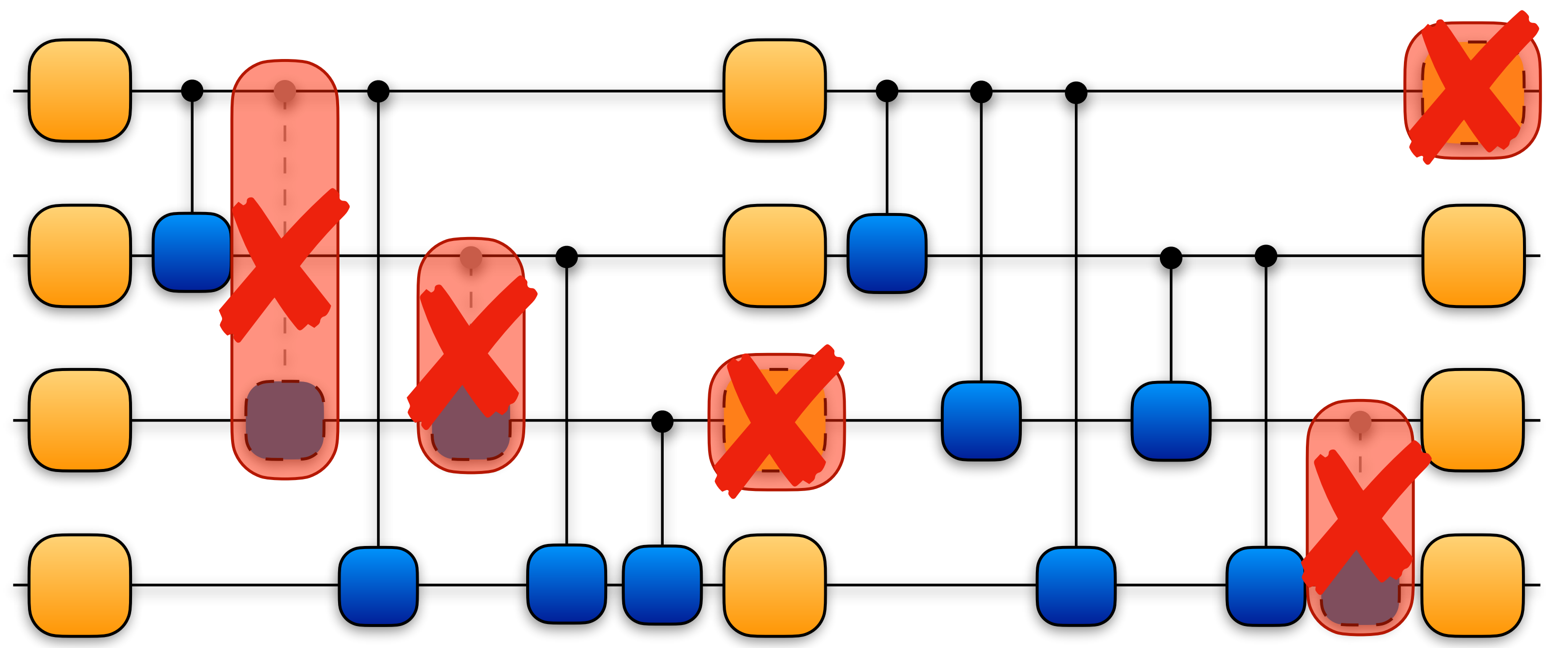
ROTATION dropout



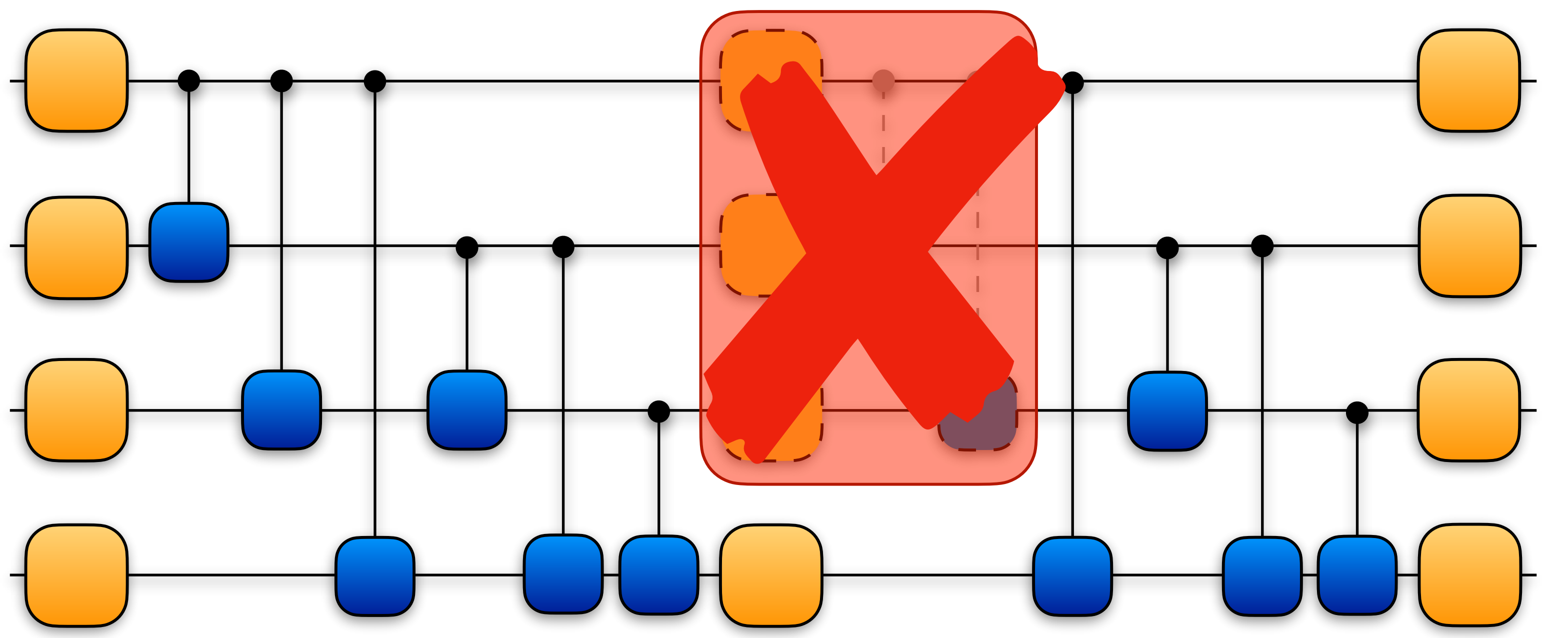
ENTANGLING dropout[1]



INDEPENDENT dropout



GROUP dropout

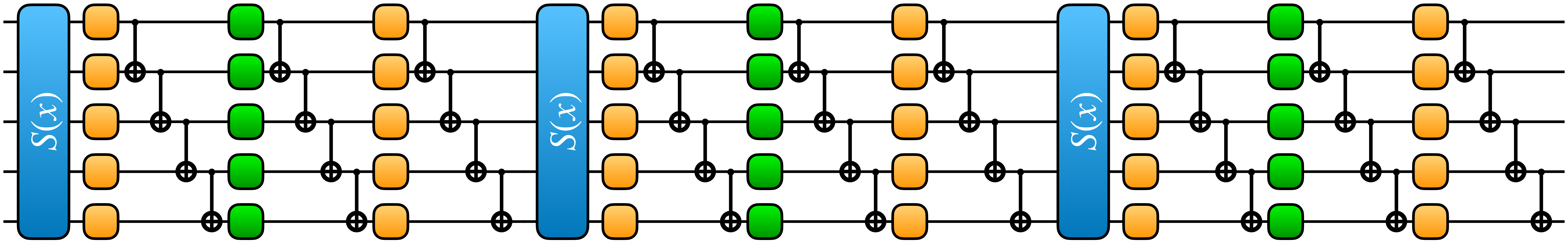


[1] M. Kobayashi, et al., **Quantum Mach. Intell. 4, (2022)**

F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Dropout scheme

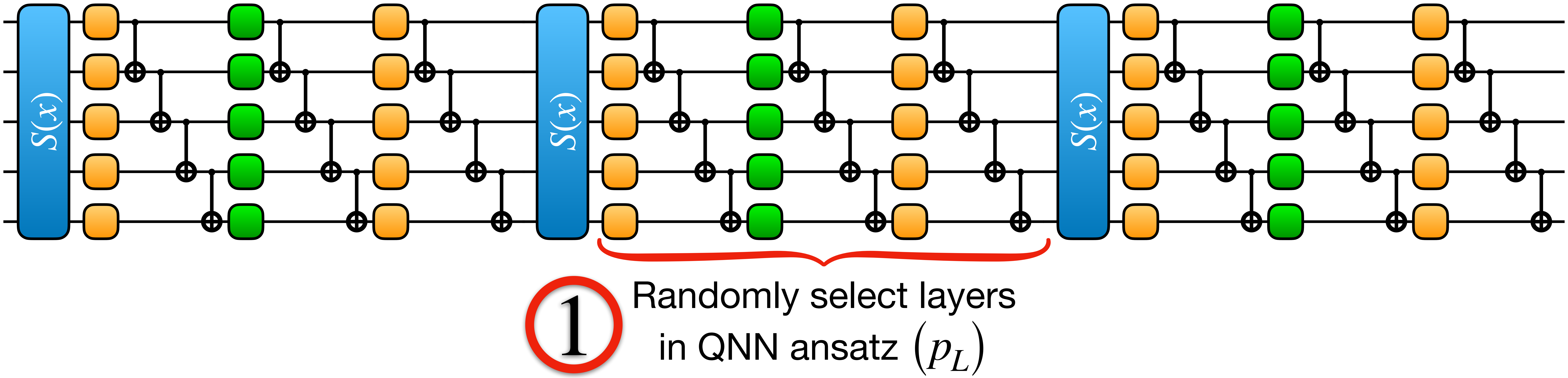
At each training step



F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Dropout scheme

At each training step

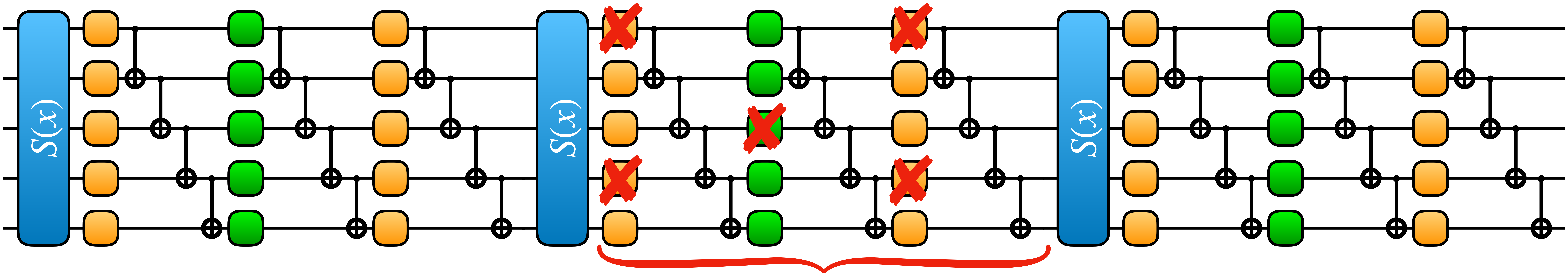


F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Dropout scheme

At each training step

2 Remove gates based on p_G following a strategy

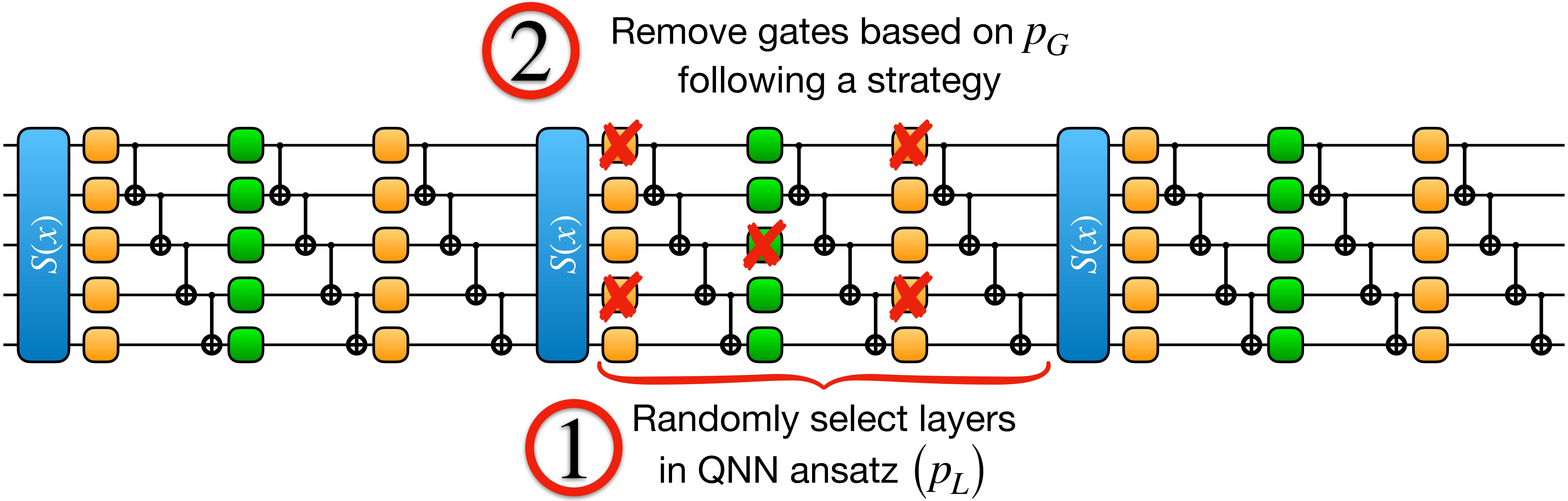


1 Randomly select layers in QNN ansatz (p_L)

F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Dropout scheme

At each training step



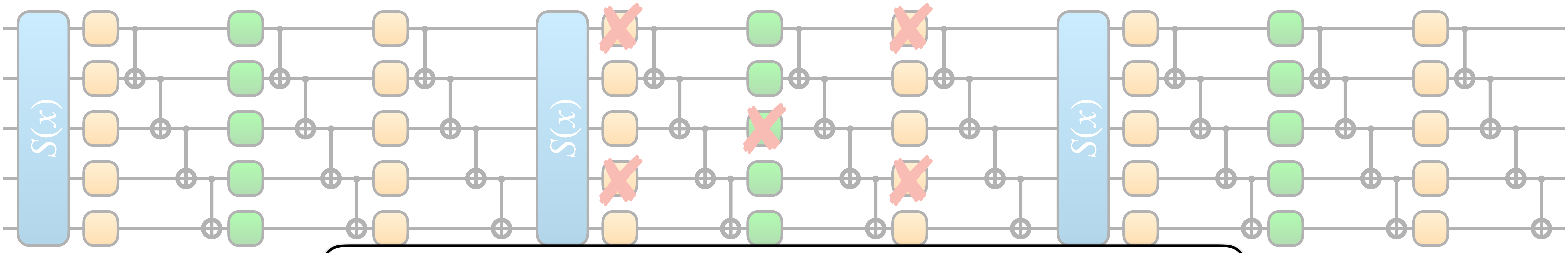
3 Optimization step: update the parameters

F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Dropout scheme

At each training step

2 Remove gates based on p_G following a strategy



3 Optimizat... update the ...

The circuit changes at each training step

F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Main results

1

Generalization

2

Expressibility and **entanglement** of QNN

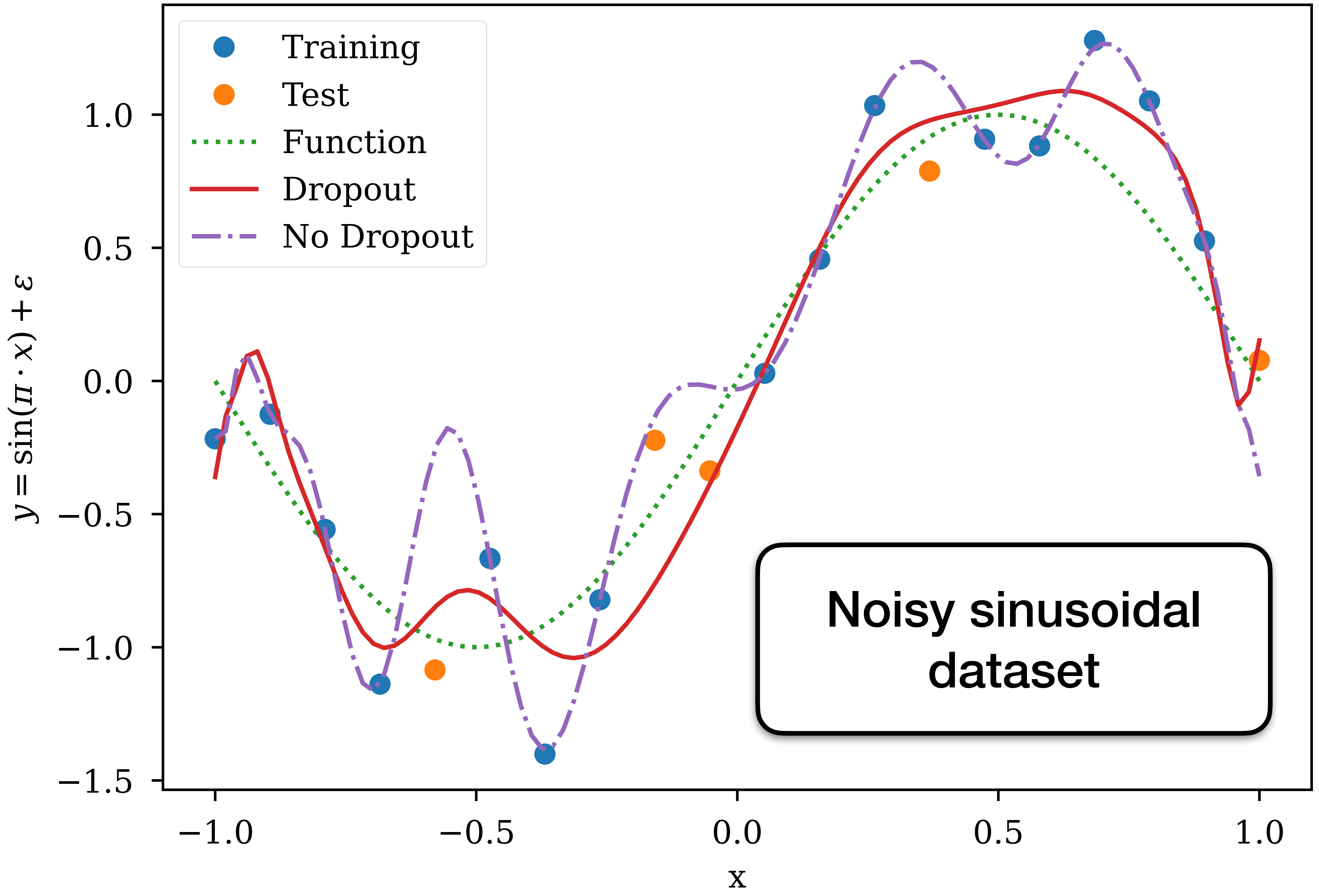
3

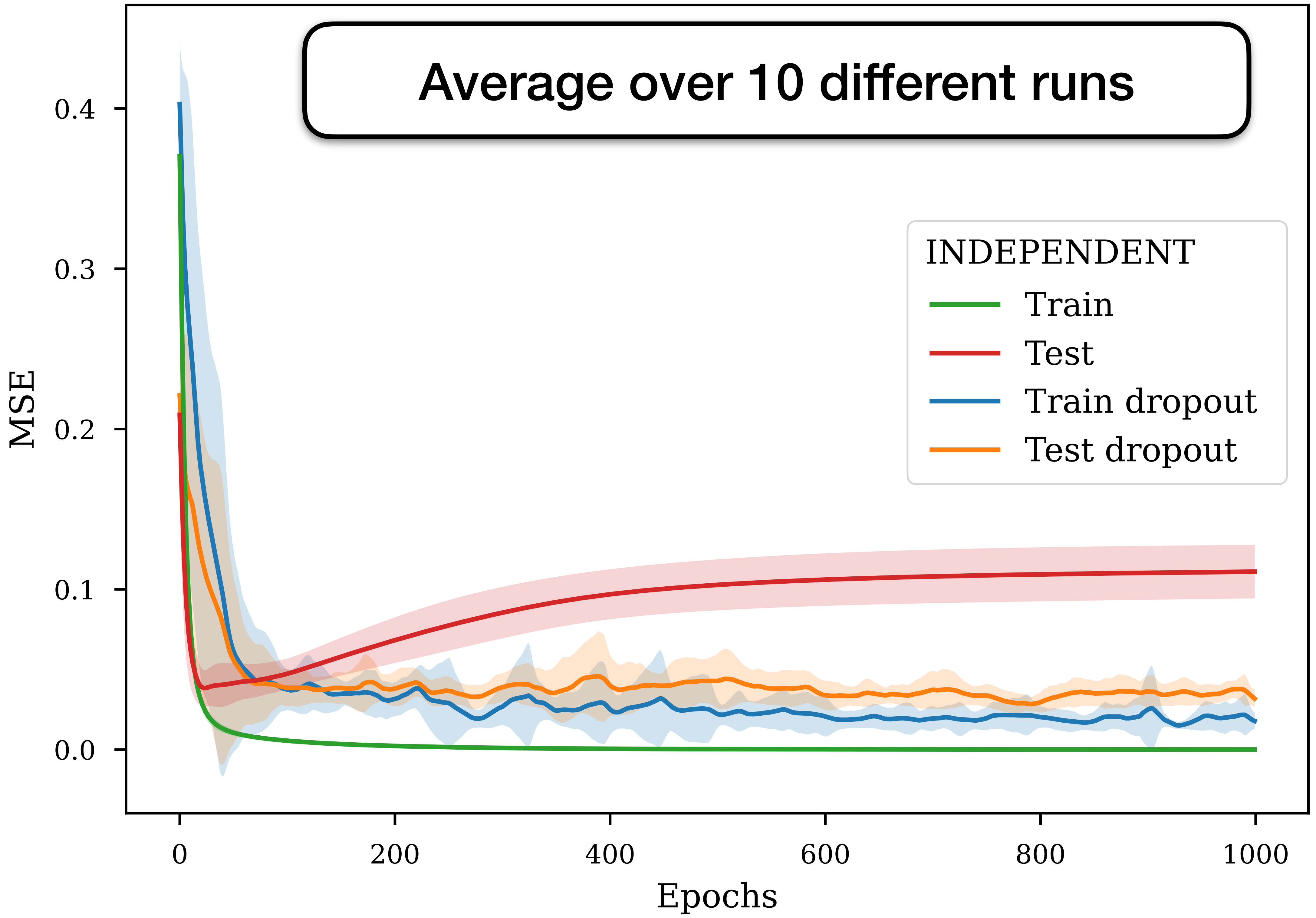
Parameters rescaling at test time

F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Generalization

Noiseless quantum simulations



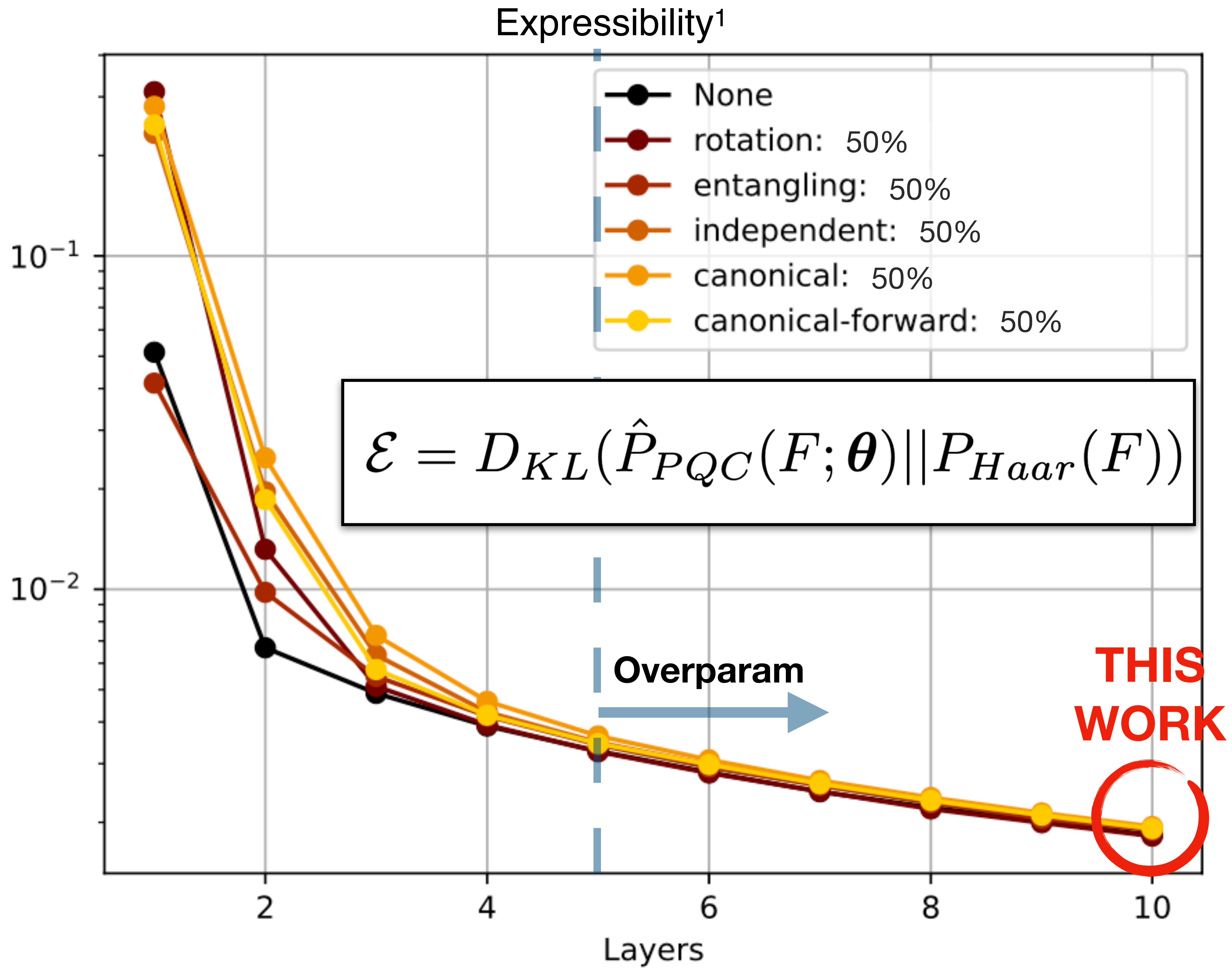


Optimal dropout percentages: 5-8% depending on the strategy

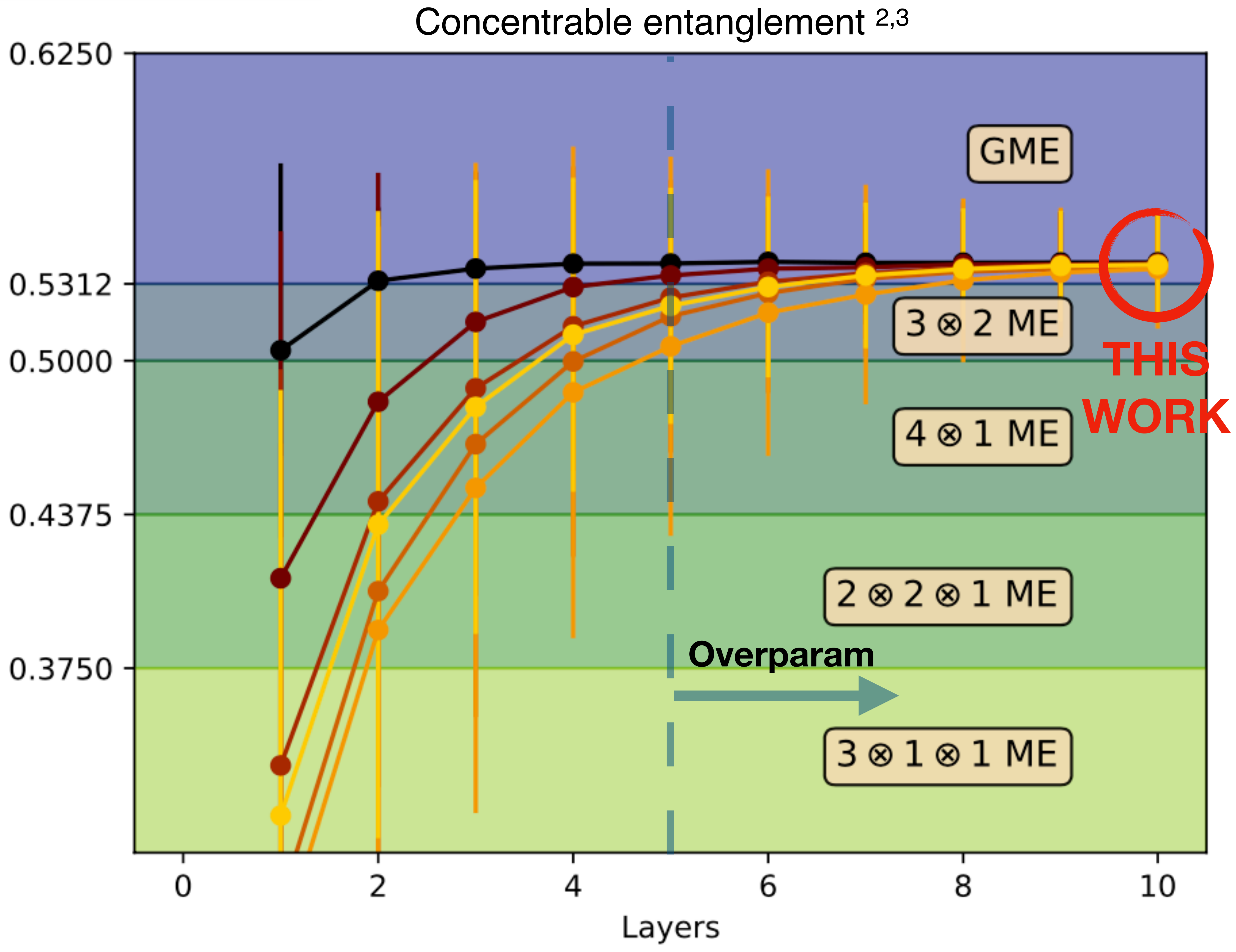
F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Expressibility and entanglement

Average over 15000 different states



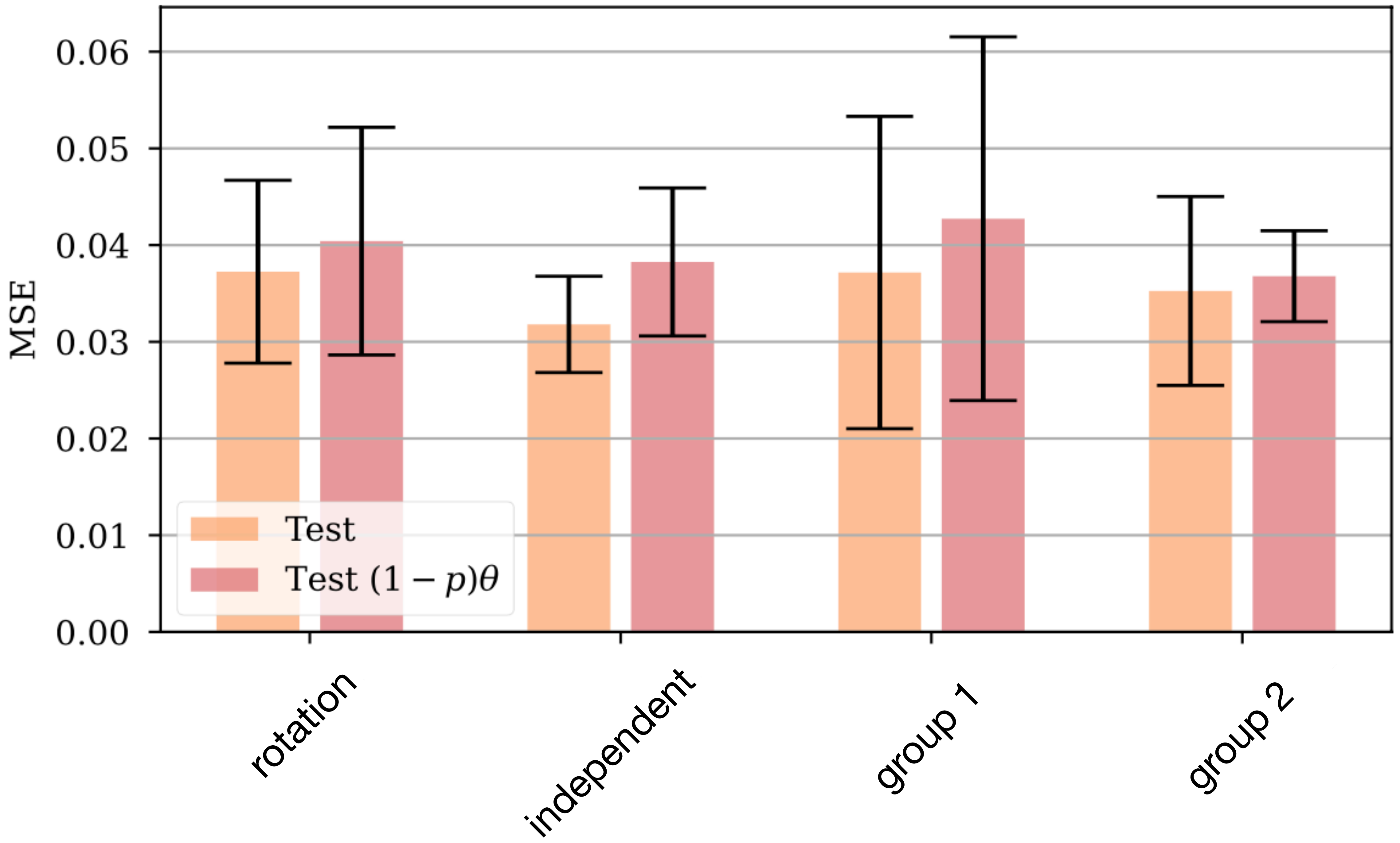
$$\mathcal{E} = D_{KL}(\hat{P}_{PQC}(F;\boldsymbol{\theta})||P_{Haar}(F))$$



[1] S. Sim, et al., **Adv. Quant. Tech. (2019)** [2] J. L. Beckey, et al., **PRL 127, 140501 (2021)** [3] L. Schatzki, et al., **arXiv 2209.07607 (2022)**

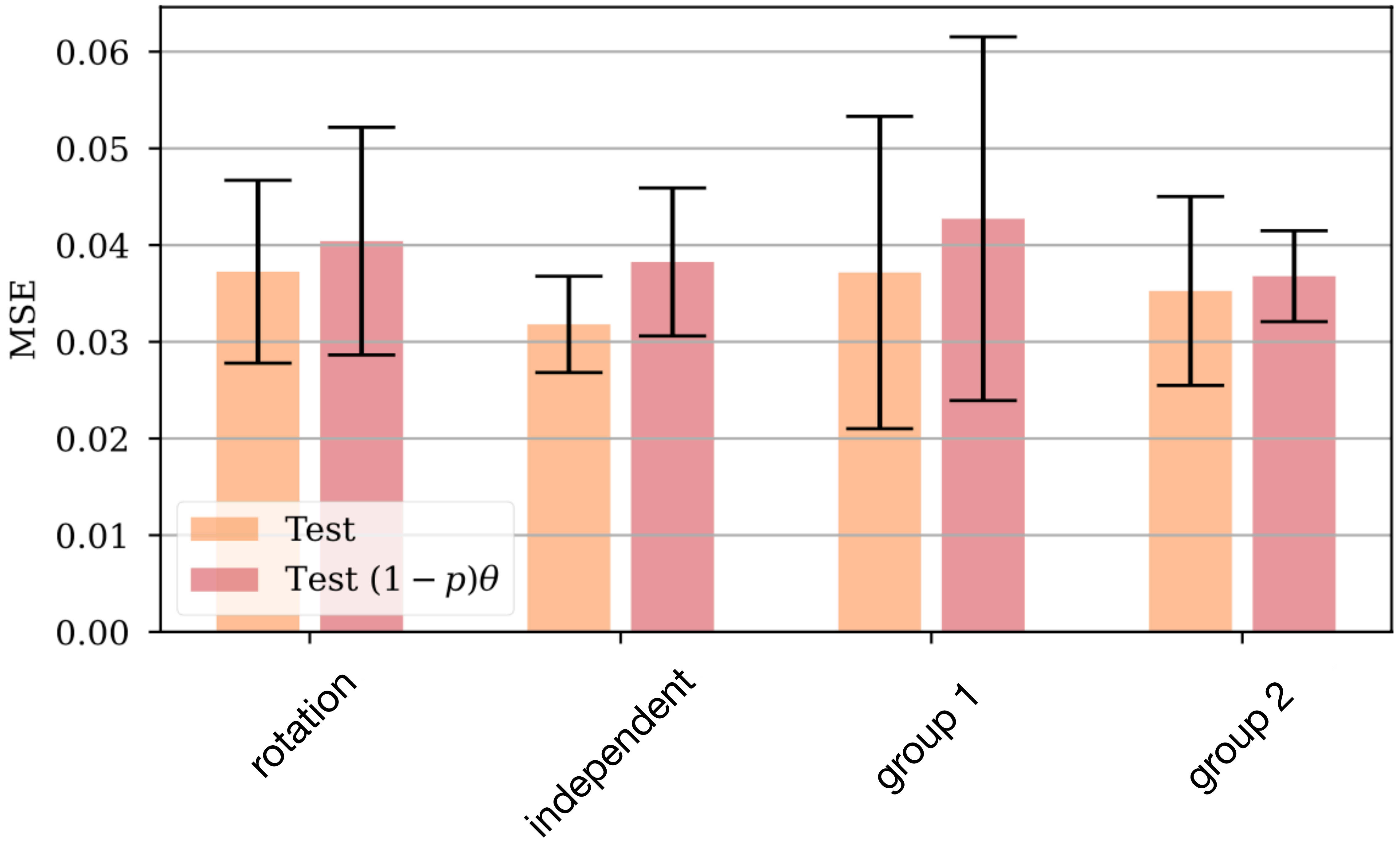
F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Parameters rescaling

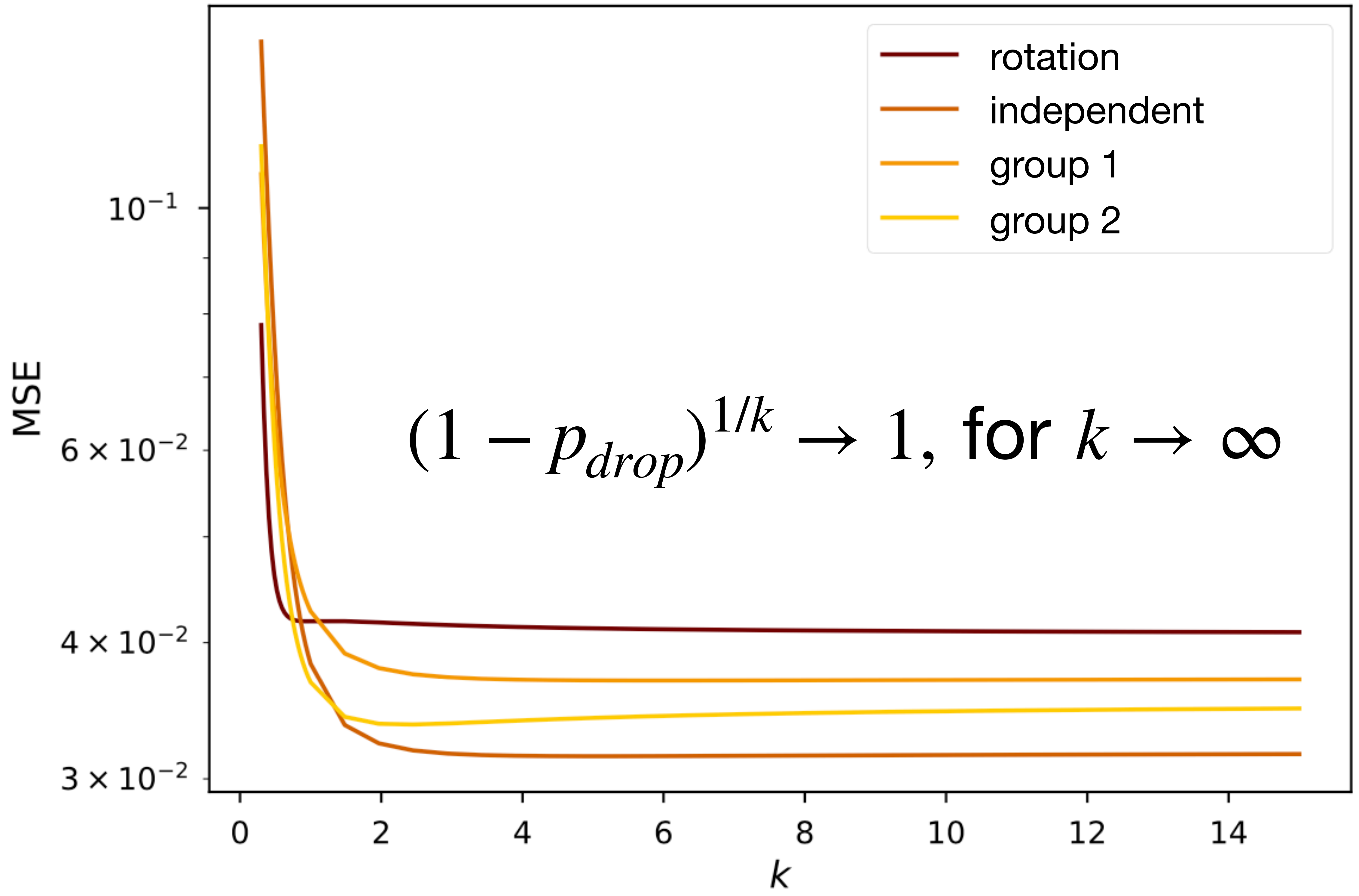


F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Parameters rescaling



$$\underline{\theta}_{test} = (1 - p_{drop})^{1/k} \underline{\theta}_{train}$$



$(1 - p_{drop})^{1/k} \rightarrow 1$, for $k \rightarrow \infty$

F. Scala, et al., A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

Aknowledgements

UNIVERSITÀ DI PAVIA
Department of Physics

SAPIENZA
UNIVERSITÀ DI ROMA

Francesco Scala

Dario Gerace

Andrea Ceschini

Massimo Panella

NQSTI
National Quantum Science
and Technology Institute

Summary

A General Approach to Dropout in Quantum Neural Networks, **arXiv: 2310.04120 (2023)**

To soon appear in Adv. Quant. Tech.
STAY TUNED

UNIVERSITÀ DI PAVIA

Quantum dropout does:

- Reduce **overfitting** allowing **generalization**
- Not reduce **expressibility nor entanglement** for highly overparametrized QNNs
- Not seem to need parameters **rescaling** at test time