The Theory of Learning from Data as a Function of Noise

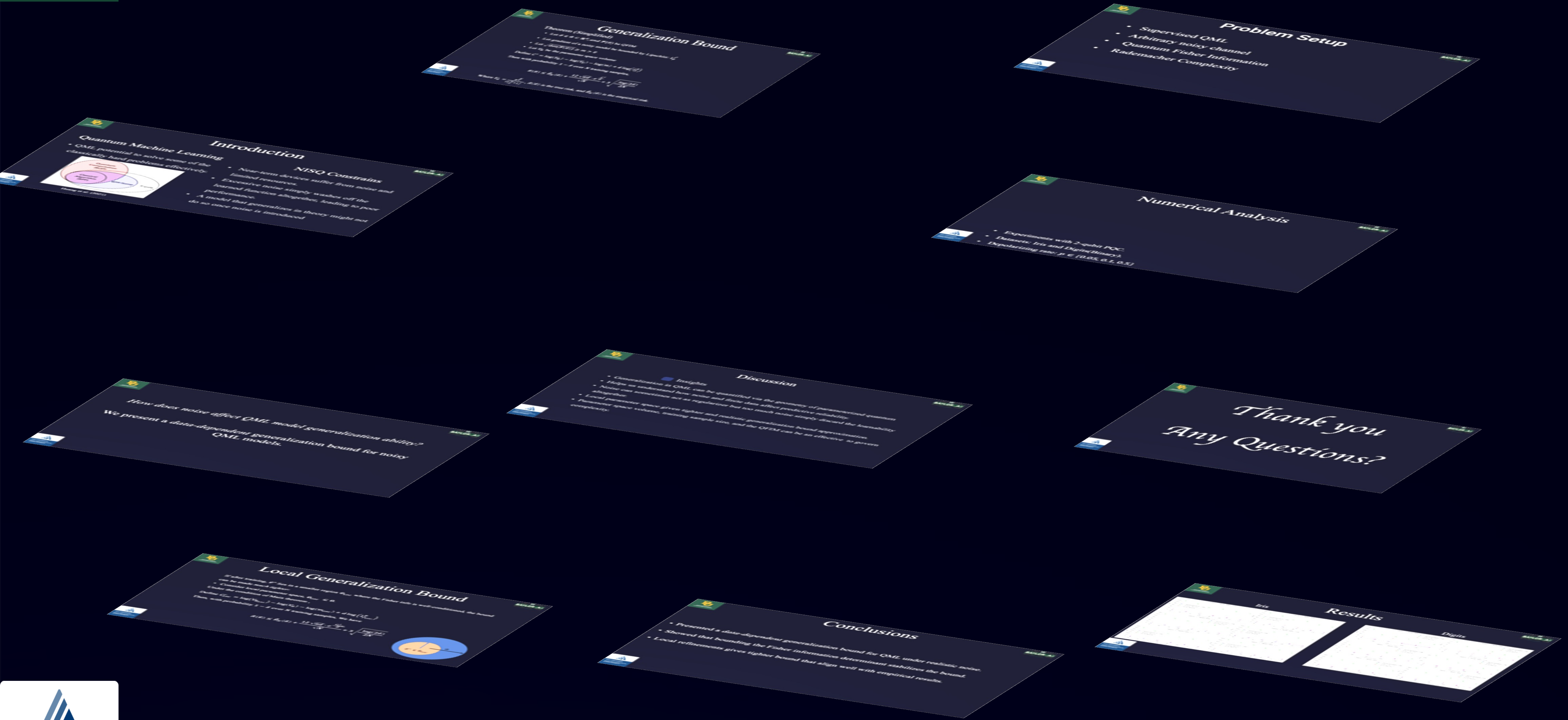
Bikram Khanal **and Pablo Rivas**

First International Workshop on Quantum Computing and Artificial Intelligence (QC+AI 2025 @ AAAI)

OJ

ENGINEERING & COMPUTER SCIENCE

BAYLOR.AI LAB
FOR AI OTRHOPRAXY

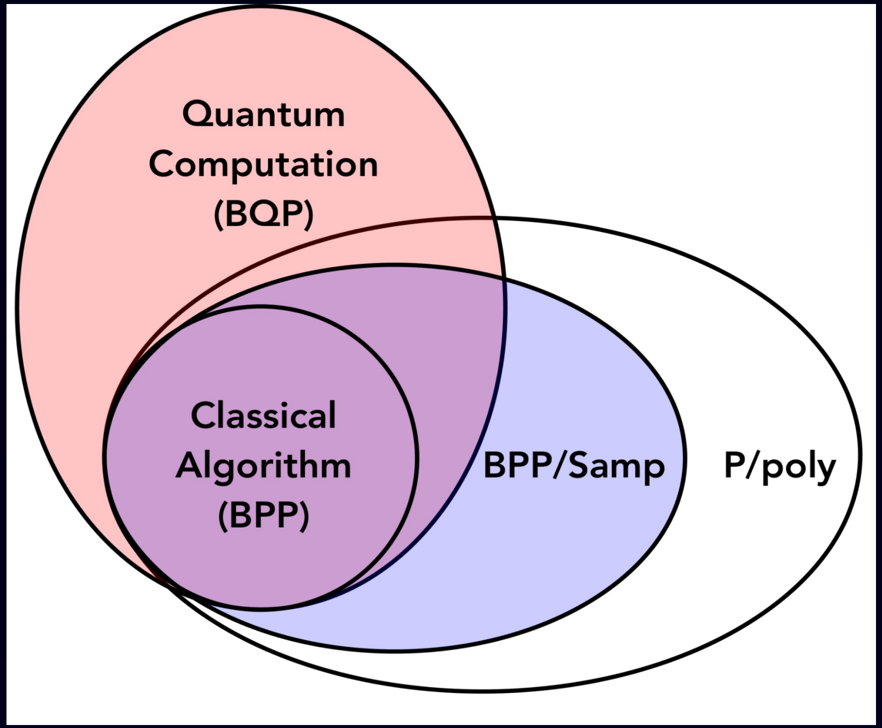


AAAI CONFERENCE
Annual AAAI Conference on Artificial Intelligence

Introduction

Quantum Machine Learning

- QML potential to solve some of the classically hard problems effectively.



Huang et al. (2021)

NISQ Constrains

- Near-term devices suffer from noise and limited resources.
- Excessive noise simply washes off the learned function altogether, leading to poor performance.
- A model that generalizes in theory might not do so once noise is introduced

ENGINEERING & COMPUTER SCIENCE

BAYLOR.AI LAB

FOR AI OTRHOPRAXY

How does noise affect QML model generalization ability?

We present a *data-dependent* generalization bound for noisy QML models.

AAAI CONFERENCE

Annual AAAI Conference on Artificial Intelligence

Problem Setup

- Supervised QML
- Arbitrary noisy channel
- Quantum Fisher Information
- Rademacher Complexity

Generalization Bound

Theorem (Simplified)

- Let $\theta \in \Theta \subset \mathbb{R}^d$, *and* $\boldsymbol{\mathcal{F}}(\theta)$ be QFIM
- Let gradient of a noisy model be bounded by Lipschitz: L_p^f
- Let $\sqrt{\det(\boldsymbol{\mathcal{F}}(\theta))} \geq m > 0$
- Let V_Θ be the parameter space volume

Define $C' = \log(V_\Theta) - \log(V_d) - \log(m) + d\log\left(L_f^p\right)$

Then with probability $1 - \delta$ over N training samples,

$$R(\theta) \leq \hat{R}_N(\theta) + \frac{12\sqrt{\pi d} \cdot e^{\frac{C'}{d}}}{\sqrt{N}} + 3\sqrt{\frac{\log(2\delta)}{2N}}$$

Where $V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}$, $R(\theta)$ is the true risk, and $\hat{R}_N(\theta)$ is the empirical risk.

Local Generalization Bound

If after training, θ^* lies in a smaller region Θ_{Loc} where the Fisher info is well-conditioned, the bound can be made *much tighter*:

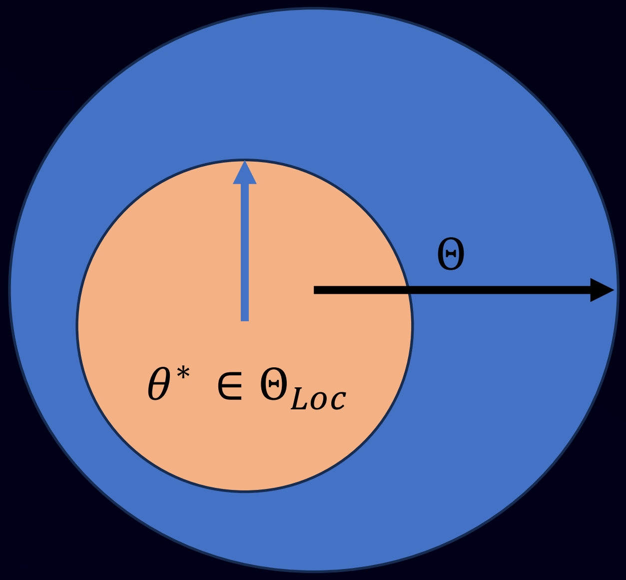
- Consider local parameter space, $\Theta_{Loc} \subseteq \Theta$.

Under the condition of Main theorem ,

Define $C'_{Loc} = \log(V_{\Theta_{Loc}}) - \log(V_d) - \log(m_{Loc}) + d\log\left(L^p_{f\,Loc}\right)$

Then, with probability $1-\delta$ over N training samples, We have

$$R(\theta) \leq \hat{R}_N(\theta) + \frac{12\sqrt{\pi d} \cdot e^{\frac{C'_{Loc}}{d}}}{\sqrt{N}} + 3\sqrt{\frac{\log(2\delta)}{2N}}$$



Numerical Analysis

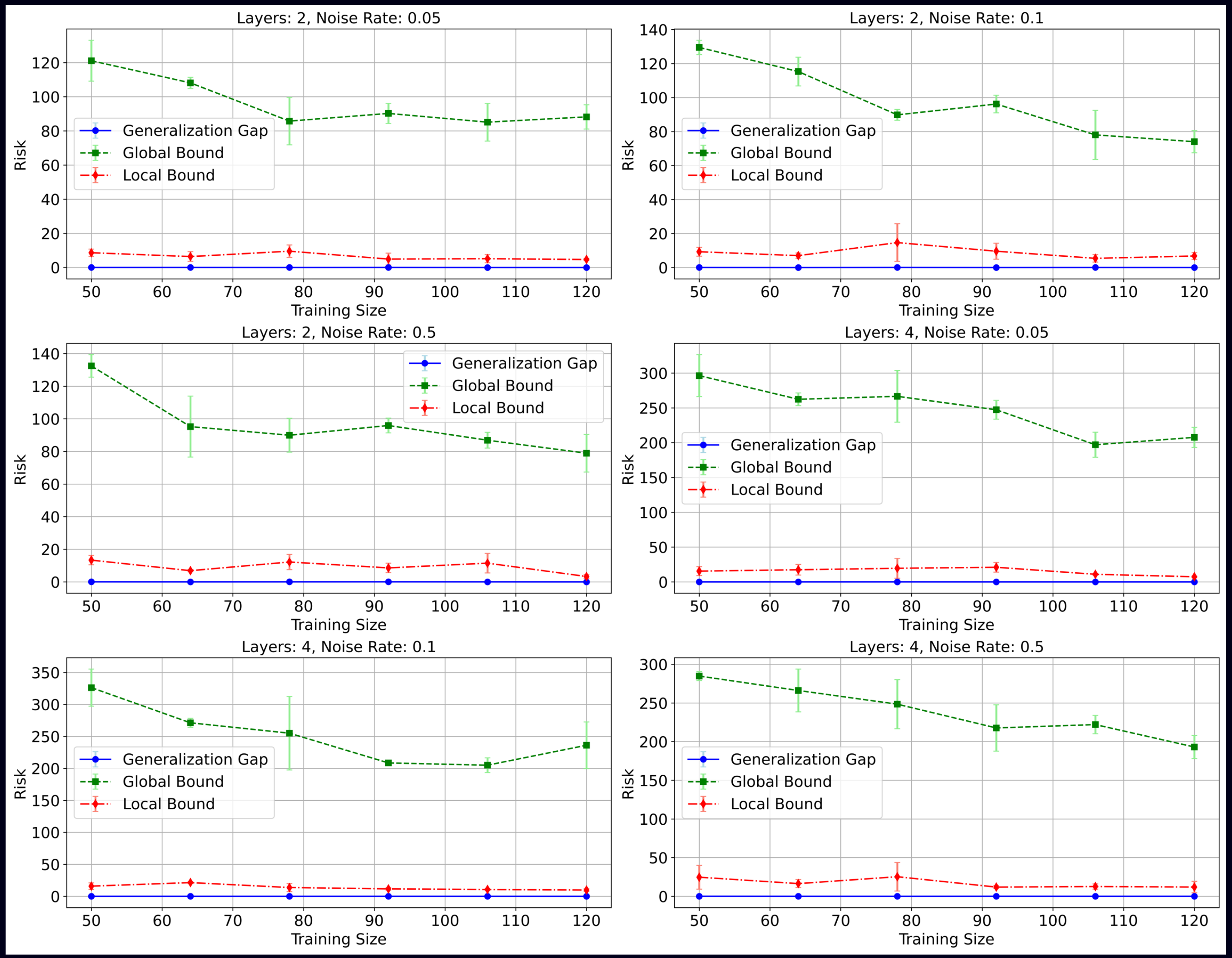
- Experiments with 2-qubit PQC.
- Datasets: Iris and Digits(Binary).
- Depolarizing rate: $p \in \{0.05, 0.1, 0.5\}$

ENGINEERING & COMPUTER SCIENCE

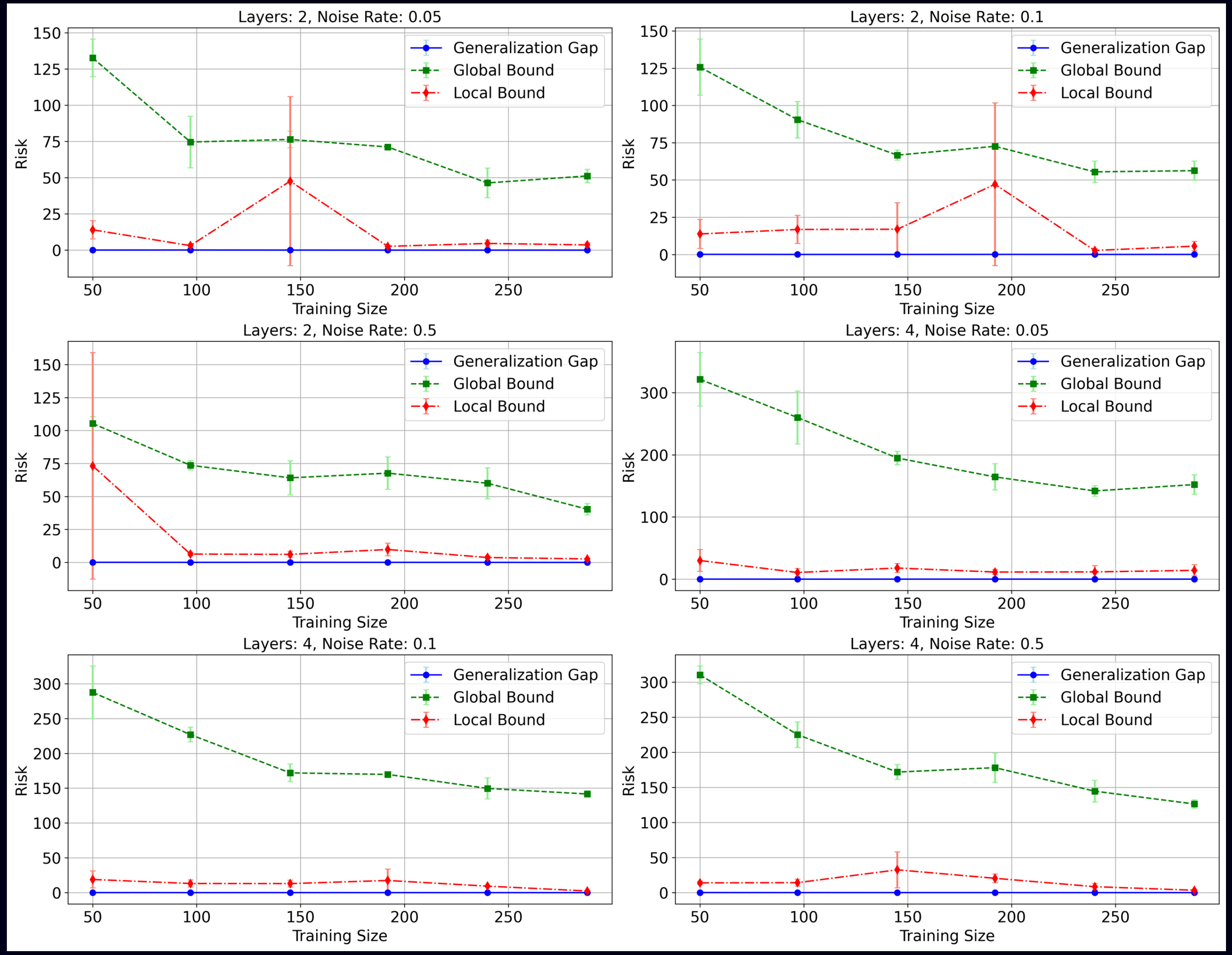
BAYLOR.AI LAB
FOR AI ORTHOPRAXY

Results

Iris



Digits



AAAI CONFERENCE
Annual AAAI Conference on Artificial Intelligence

ENGINEERING & COMPUTER SCIENCE

BAYLOR.AI LAB

Discussion

Insights

- Generalization in QML can be quantified via the geometry of parameterized quantum
- Helps us understand how noise and finite data affect predictive reliability.
- Noise can sometimes act as regularizer but too much noise simply discard the learnability altogether.
- Local parameter space given tighter and realistic generalization bound approximation.
- Parameter space volume, training sample size, and the QFIM can be an effective to govern complexity.

AAAI CONFERENCE
Annual AAAI Conference on Artificial Intelligence

Conclusions

- Presented a *data-dependent* generalization bound for QML under realistic noise.
- Showed that bounding the Fisher information determinant stabilizes the bound.
- Local refinements gives tighter bound that align well with empirical results.

Thank you

Any Questions?