Stochastic Gradient Descent for Hybrid Quantum-Classical Optimization

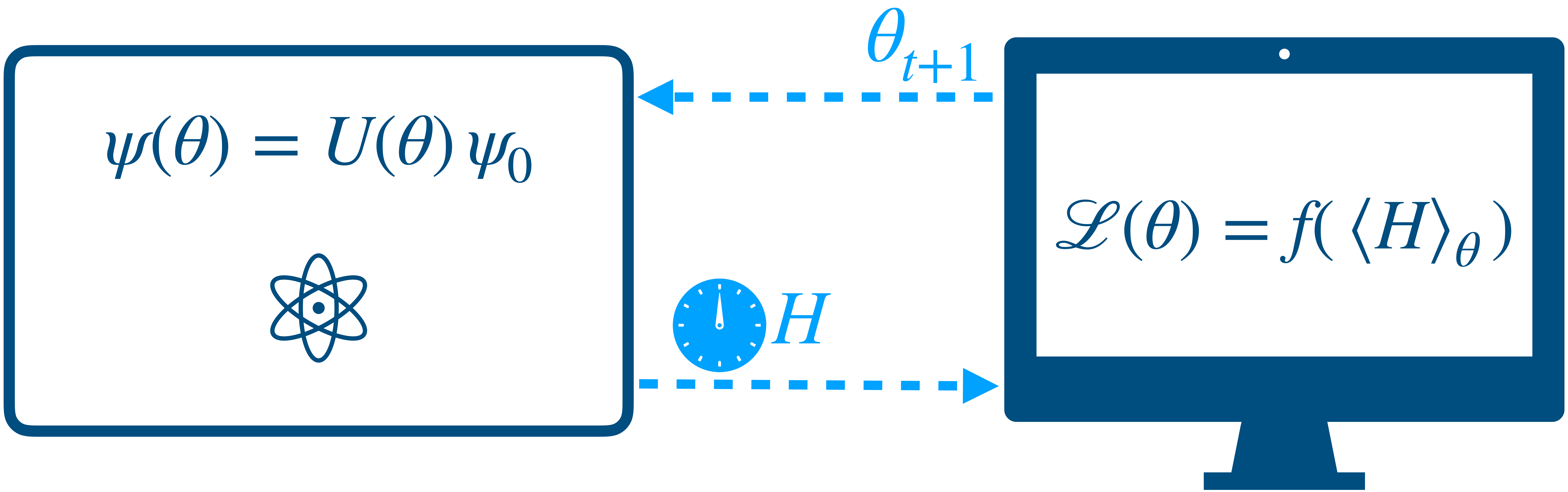
APS March Meeting 2020

Frederik Wilde
f.wilde@fu-berlin.de
frederikwil.de/slides/aps2020

Freie Universität Berlin

Hybrid quantum-classical algorithms

QAOA, VQE, quantum assisted ML, ...



Optimization

0-th order:

SPSA, swarm optimization, genetic algorithm, etc.

$\longrightarrow$ good for few parameters

1st order:

- gradient is expensive (no autograd)
- "exact" gradient is accessible

$\longrightarrow$ Parameter-shift rule $\partial_\theta \langle H \rangle_\theta \sim \langle H \rangle_{\theta+s} - \langle H \rangle_{\theta-s}$

[M. Schuld et. al, PRA 2019]

Sources of stochasticity

$$\mathscr{L}(\theta) = \sum_{i} \left[\left\langle \sum_{j} h_j \right\rangle_{(x_i,\, \theta)} - y_i \right]^2$$

1. Measurements
2. Observable components
3. Data
4. Parameter-shift terms

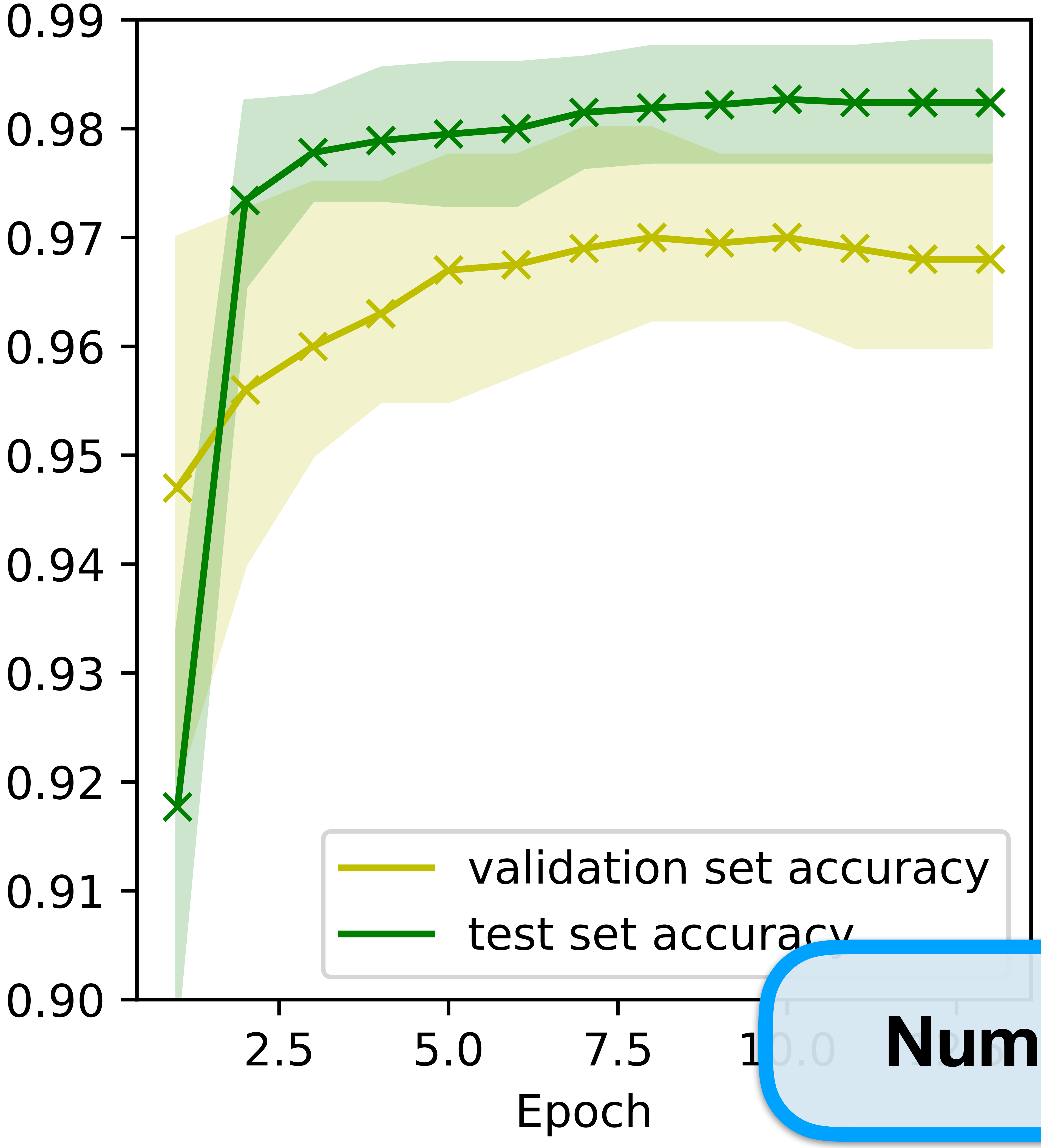
Convergence

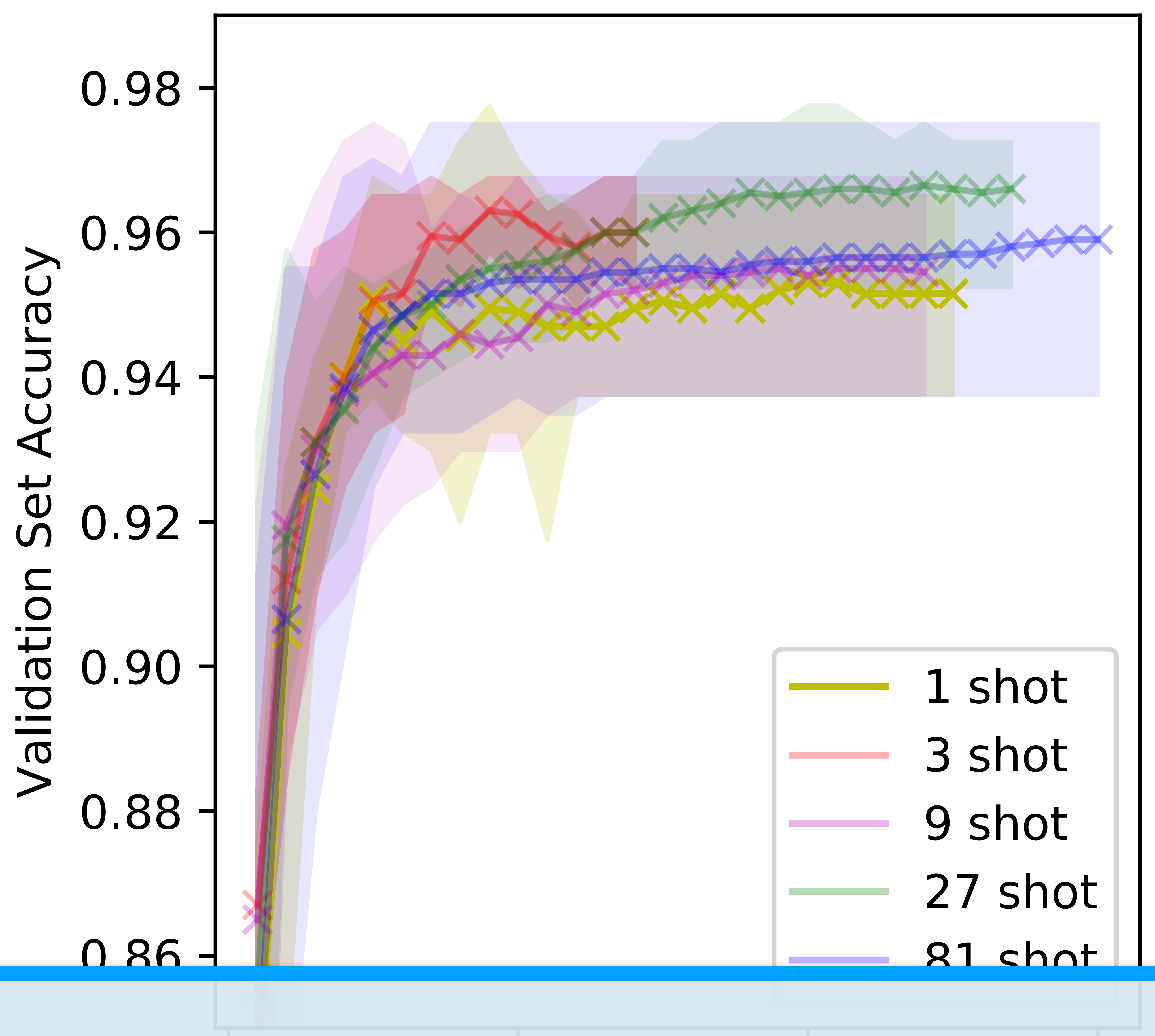
Problem: $\mathbb{E}(\mathscr{L}(X)) \neq \mathscr{L}(\mathbb{E}(X))$

$\longrightarrow$ Solved for polynomial loss functions

Provable convergence under certain assumptions about $\mathscr{L}$

MNIST Classifier





Number of shots is a hyper parameter.

8 qubits, 400 parameters, batch size = 1

What now?

- Non-polynomial loss functions
- Dynamical study of SGD
  e.g. iCANS [J. Kübler et. al, 1909.09083] and
  Langevin dynamics
- Beyond parameter shift rule
- #(measurement shots) and barren plateaus

Thanks for your attention

Co-authors:

Ryan Sweke
Freie Universität Berlin

Johannes Jakob Meyer
Freie Universität Berlin

Maria Schuld
Xanadu Inc. and University of KwaZulu-Natal South Africa

Paul K. Fährmann
Freie Universität Berlin

Barthélémy Meynard-Piganeau
Ecole Polytechnique Paris

Jens Eisert
Freie Universität Berlin

arxiv: `1910.01155`

Slides at:
`frederikwil.de/slides/aps2020`

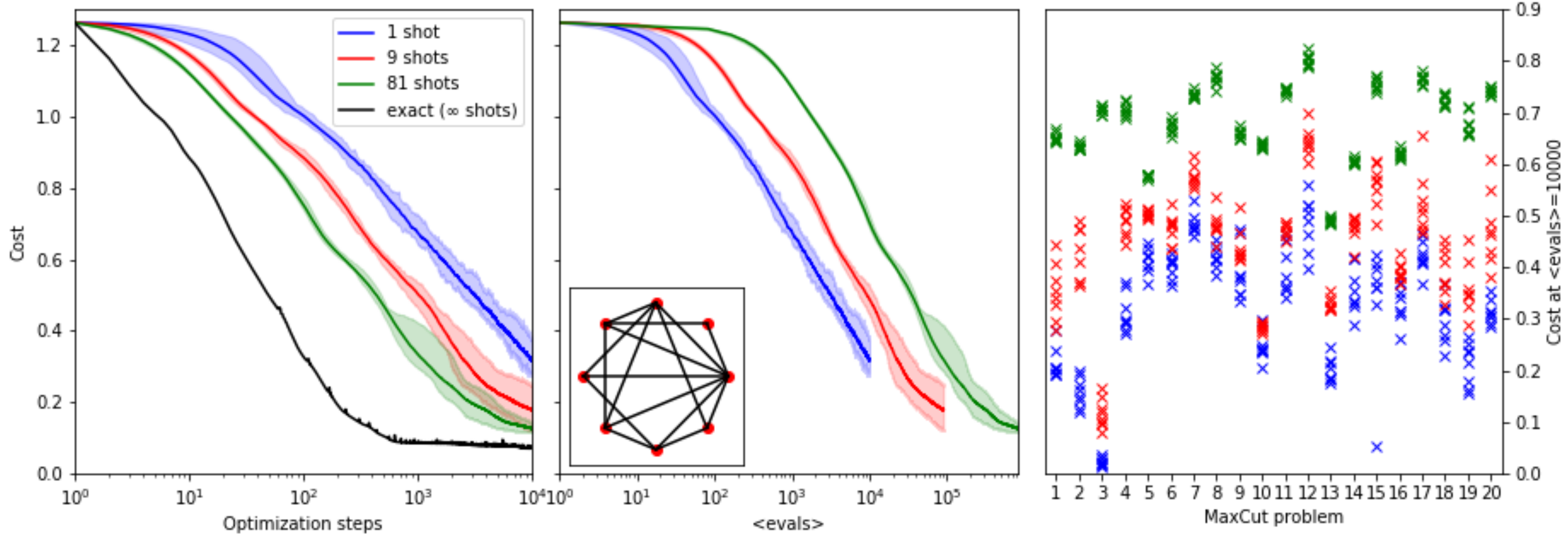
`f.wilde@fu-berlin.de`

Sources of stochasticity

$$\mathscr{L}(\theta) = \sum_i \left[\left\langle \sum_j h_j \right\rangle_{(x_i,\,\theta)} - y_i \right]^2$$

$$\partial_\theta \mathscr{L} = \sum_i 2 \left[\sum_j \langle h_j \rangle_{(x_i,\,\theta)} - y_i \right] \sum_j \frac{1}{2} \left(\langle h_j \rangle_{(x_i,\,\theta + \frac{\pi}{2})} - \langle h_j \rangle_{(x_i,\,\theta - \frac{\pi}{2})} \right)$$

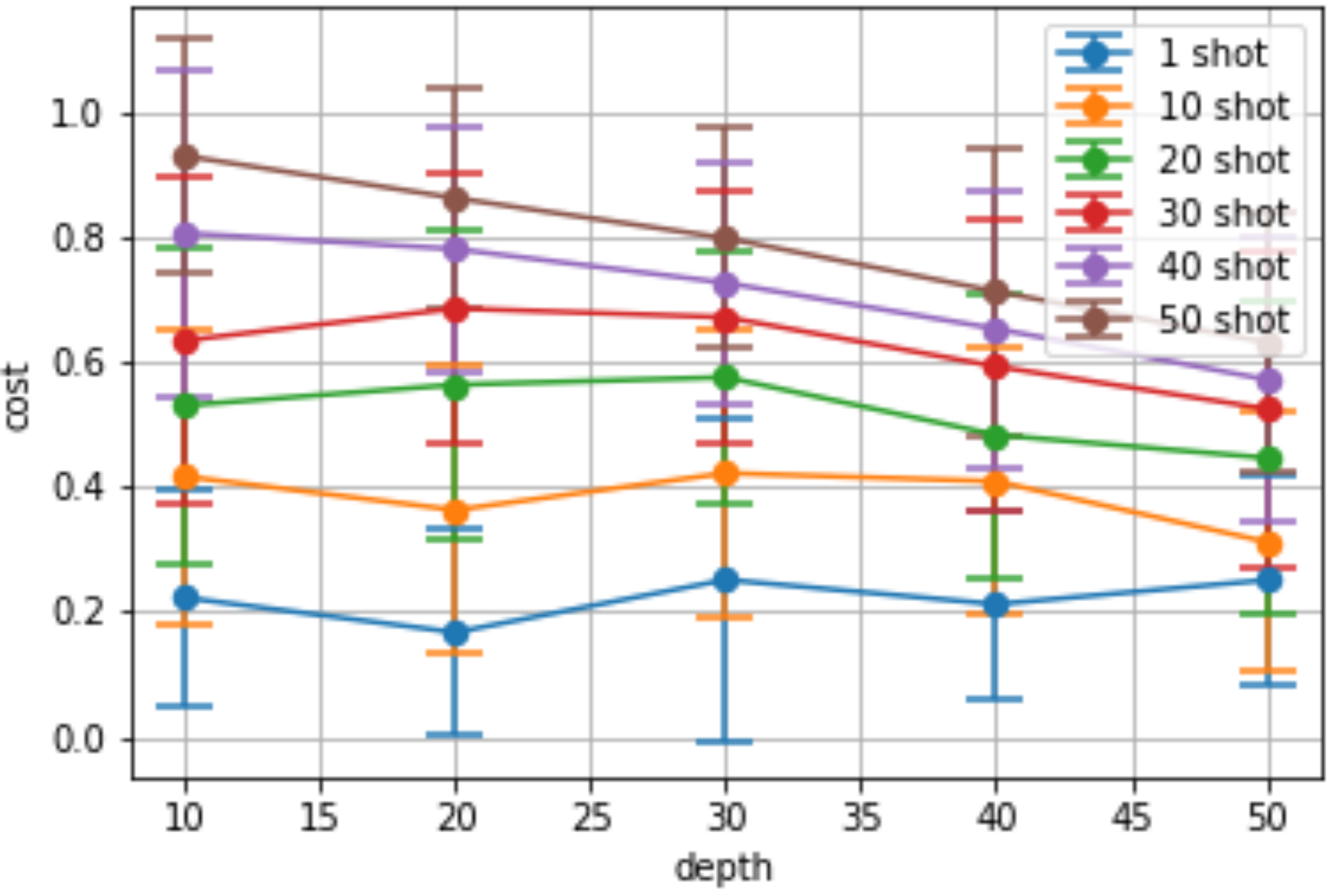
MAXCUT with QAOA

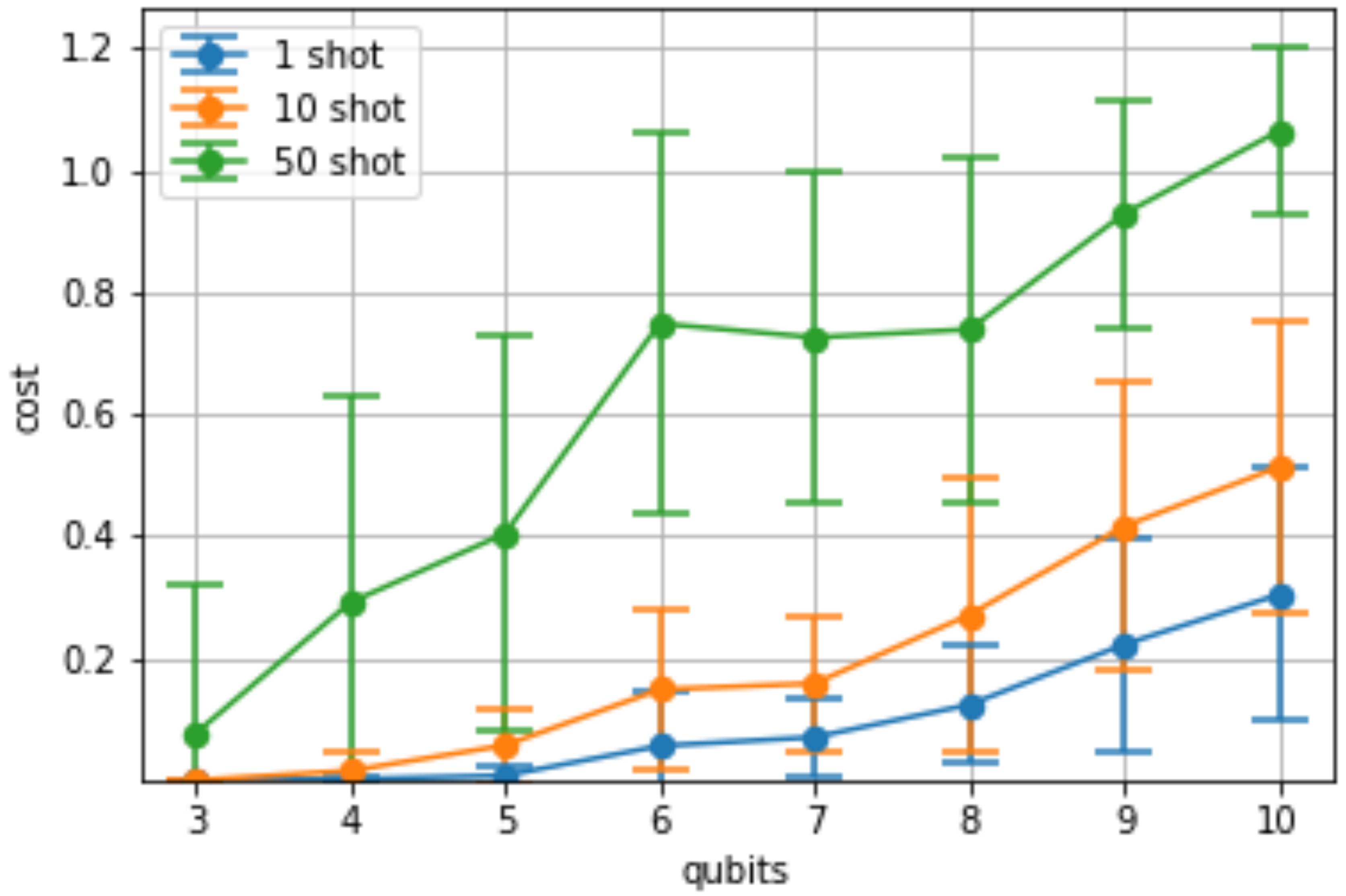


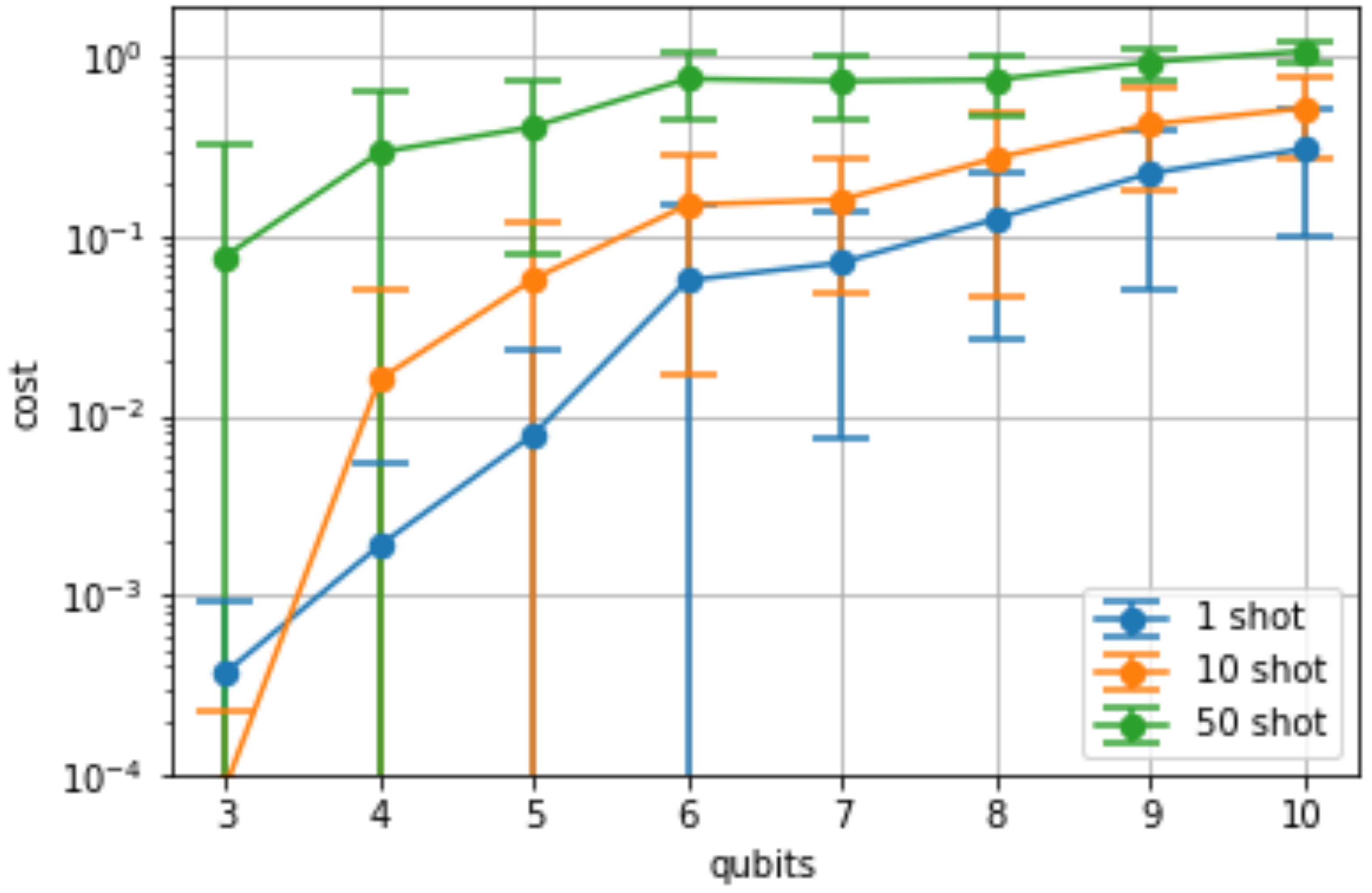
8 qubits, p=50, random graphs $|V| = 8, |E| = 16$

Scaling

QAOA on Erdős–Rényi graphs (edge probability 30%) ~20 samples/data-point (unpublished)







Corrections for polynomial loss functions

- Let X represent the measurement
- Expand $\mathscr{L}(X)$ around $\mathbb{E}(X) = x_0$
- $\mathscr{L}(X) = \mathscr{L}(x_0) + \mathscr{L}'(x_0)\,(X - x_0) + \sum_{n=2}^{s} \frac{1}{n!}\mathscr{L}^{(n)}(x_0)\,(X - x_0)^n$
- $\mathbb{E}\left[\mathscr{L}(X)\right] = \mathbb{E}\left[\mathscr{L}(x_0)\right] + \sum_{n=2}^{s} \frac{1}{n!}\mathscr{L}^{(n)}(x_0)\,\mathbb{E}\left[(X - x_0)^n\right]$