# Classically Approximating Variational Quantum Machine Learning with Random Fourier Features

Team Members: Jerry Xu, Nicholas Ng, Matthew Stefansson, Ying Qi Wen

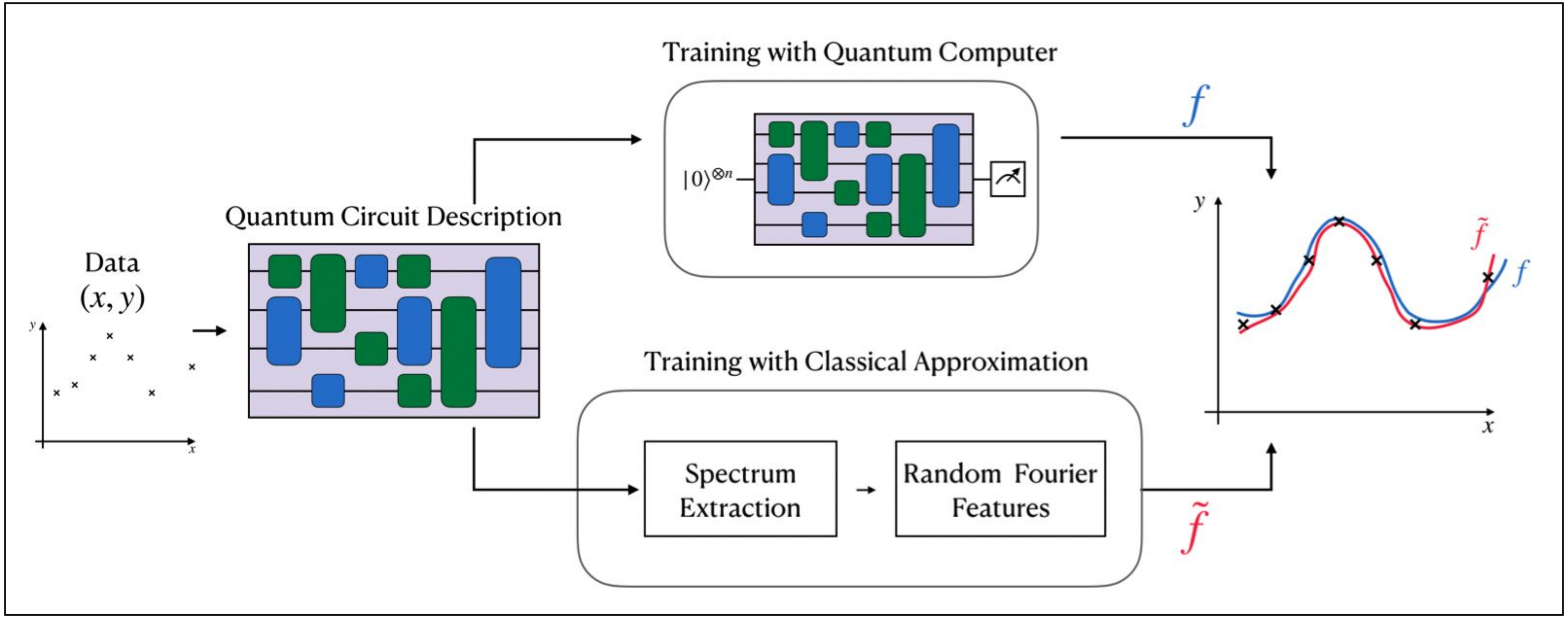
# Introduction

- Near-term quantum computing applications rely on VQCs
- VQCs trained with classical optimization of gates' parameters

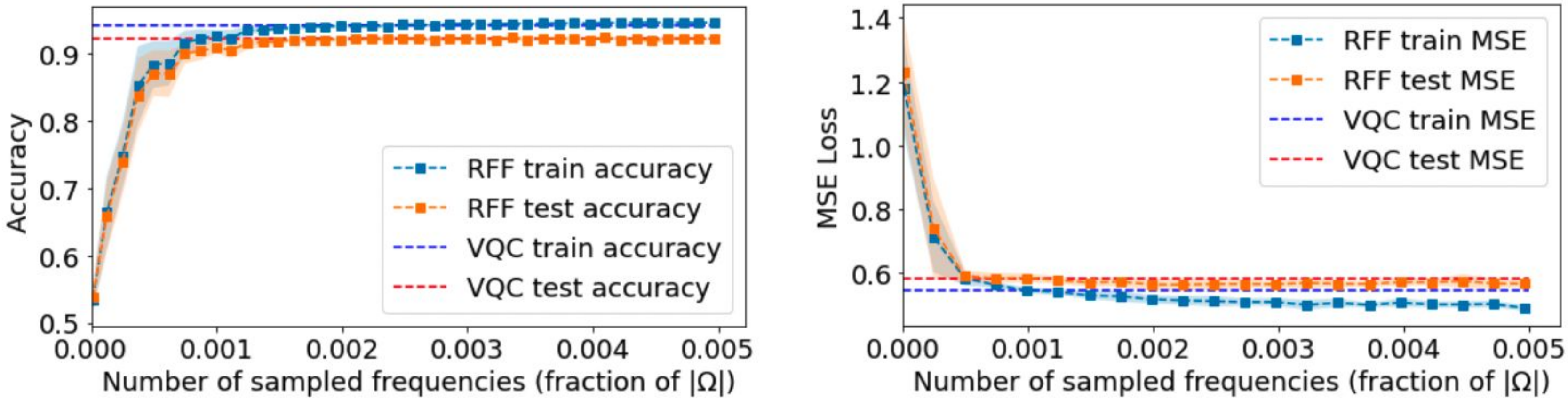
Major Challenge: Scalability of VQCs

- Idea to adapt 3 sampling strategies using Random Fourier Features (RFF)

# Why the Paper is Interesting



# The Novel Test Results




(a) Fashion-MNIST dataset (classification)

(b) California Housing dataset (regression)

# Underlying Theory

- "It is known that quantum models from VQCs are equivalent to kernel methods"
- High-dimensional kernels can be approximated using Random Fourier Features
- The kernel of the VQC is built from frequencies $\omega \in \Omega$
- 3 sampling methods: Distinct Sampling, Tree Sampling, Grid Sampling
- Use the sampled frequencies to solve ML problems

$$\tilde{\phi}(x) = \frac{1}{\sqrt{D}} \begin{bmatrix} cos(\omega_i^T x) \\ sin(\omega_i^T x) \end{bmatrix}_{i \in [\![1, D]\!]}$$

$$\tilde{k}(x, y) \simeq \tilde{\phi}(y)^T \tilde{\phi}(x)$$

$$\tilde{f} = \tilde{\mathbf{w}}^T \tilde{\phi}(x)$$

$$\mathbf{w}^* = (\mathbf{\Phi}^T \mathbf{\Phi} + M\lambda I_p)^{-1} \mathbf{\Phi}^T \boldsymbol{y}$$

# Software Implementation Approach

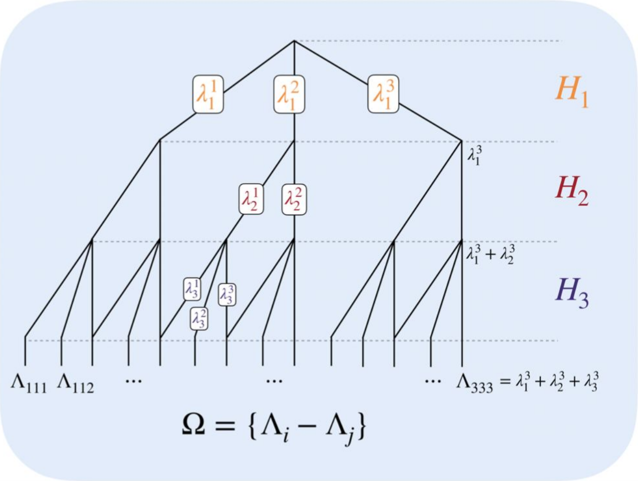
1. Create a dataset and target function
2. Diagonalize the Hamiltonians of the VQC's encoding gates
3. Obtain their eigenvalues to get all frequencies
4. Sample the frequencies
5. Construct $\boldsymbol{\varphi}$ and $\boldsymbol{\Phi}$ (i.e. matrix with each row i corresponds to $\boldsymbol{\varphi}(x_i)^T$)
6. Obtain the weight **w** and the approximate function

$$\tilde{f} = \tilde{\mathbf{w}}^T \tilde{\phi}(x) \qquad \mathbf{w}^* = (\boldsymbol{\Phi}^T \boldsymbol{\Phi} + M\lambda I_p)^{-1} \boldsymbol{\Phi}^T \boldsymbol{y}$$

# Software Implementation Highlights

- RFF with Distinct Sampling
  - Using qml.fourier.circuit_spectrum to obtain the frequencies
- RFF with Tree Sampling
  - Compute frequencies from eigenvalues of the encoding gates
  - Redundant frequencies are more likely to be sampled than unique frequencies
- RFF with Grid Sampling
  - Create a grid of frequencies between zero and the highest frequency
  - Sample from the grid instead of the actual frequencies in Ω
- Used all of the frequencies

$$\Lambda_i = \lambda_1^{i_1} + \cdots + \lambda_L^{i_L}$$



# Demo

# Issues Encountered

Notation inconsistencies

- Tried emailing author (no response)
- Have to work based on assumptions
- Reproducibility of the paper past simple VQC's.

$$D = \Omega\left(\frac{dC_1(1+\lambda)^2}{\lambda^4\epsilon^2}\left[log(dL^2|\mathcal{X}|)+log\frac{C_2(1+\lambda)}{\epsilon\lambda^2}-log\delta\right]\right) \tag{22}$$

# Thank you for Listening

Any Questions?