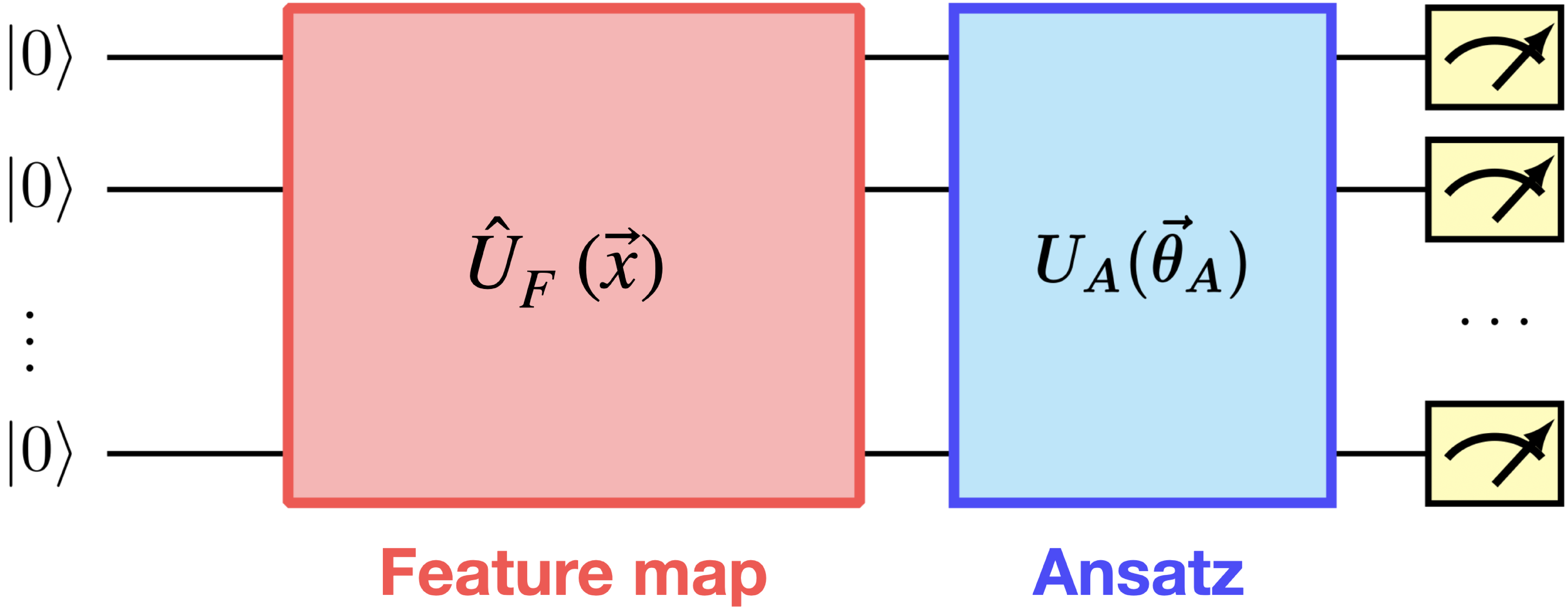
Let Quantum Neural Networks Choose Their Own Frequencies

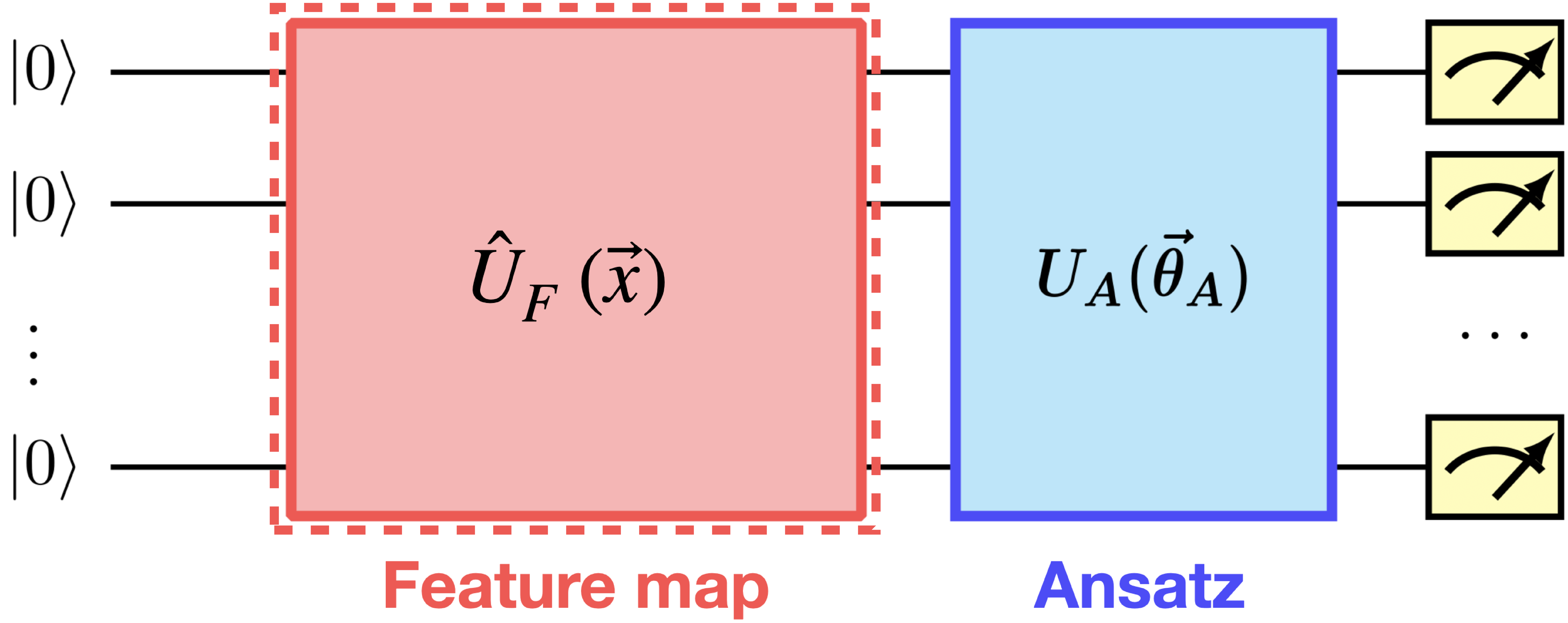
Ben Jaderberg, Antonio Gentile, Youssef Berrada, Elvira Shishenina, Vincent E. Elfving

https://arxiv.org/abs/2309.03279

(Variational) quantum machine learning models



(Variational) quantum machine learning models



Feature maps

$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\hat{G}_m(\gamma_m)\,\phi(\vec{x})}$$

Feature maps

$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\hat{Y}_m x}$$

$$\hat{G}_m(\gamma_m) = \hat{Y}_m$$

$$\phi(\vec{x}) = x$$

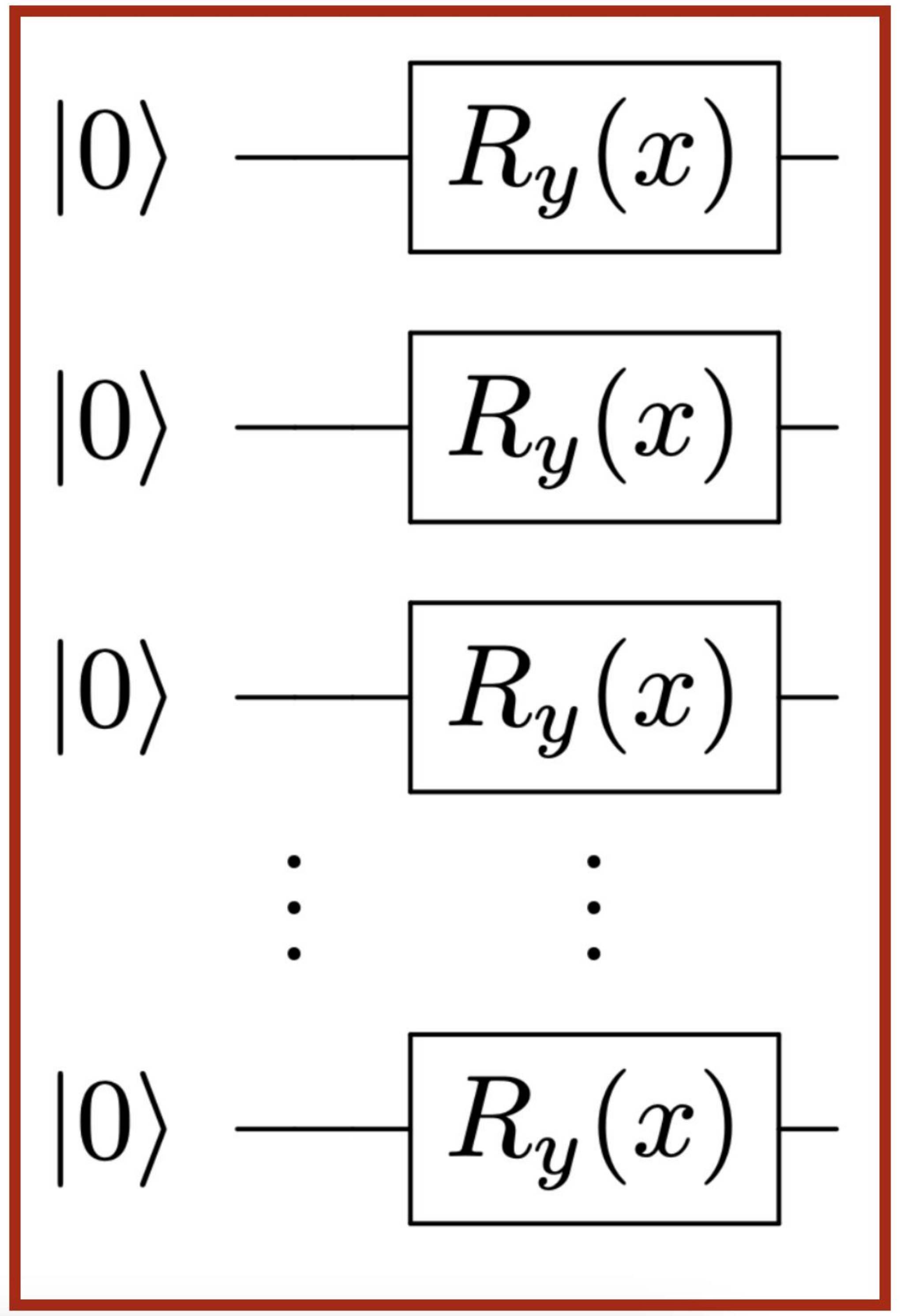
$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\hat{G}_m(\gamma_m)\,\phi(\vec{x})}$$

Feature maps

$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\hat{Y}_m x}$$

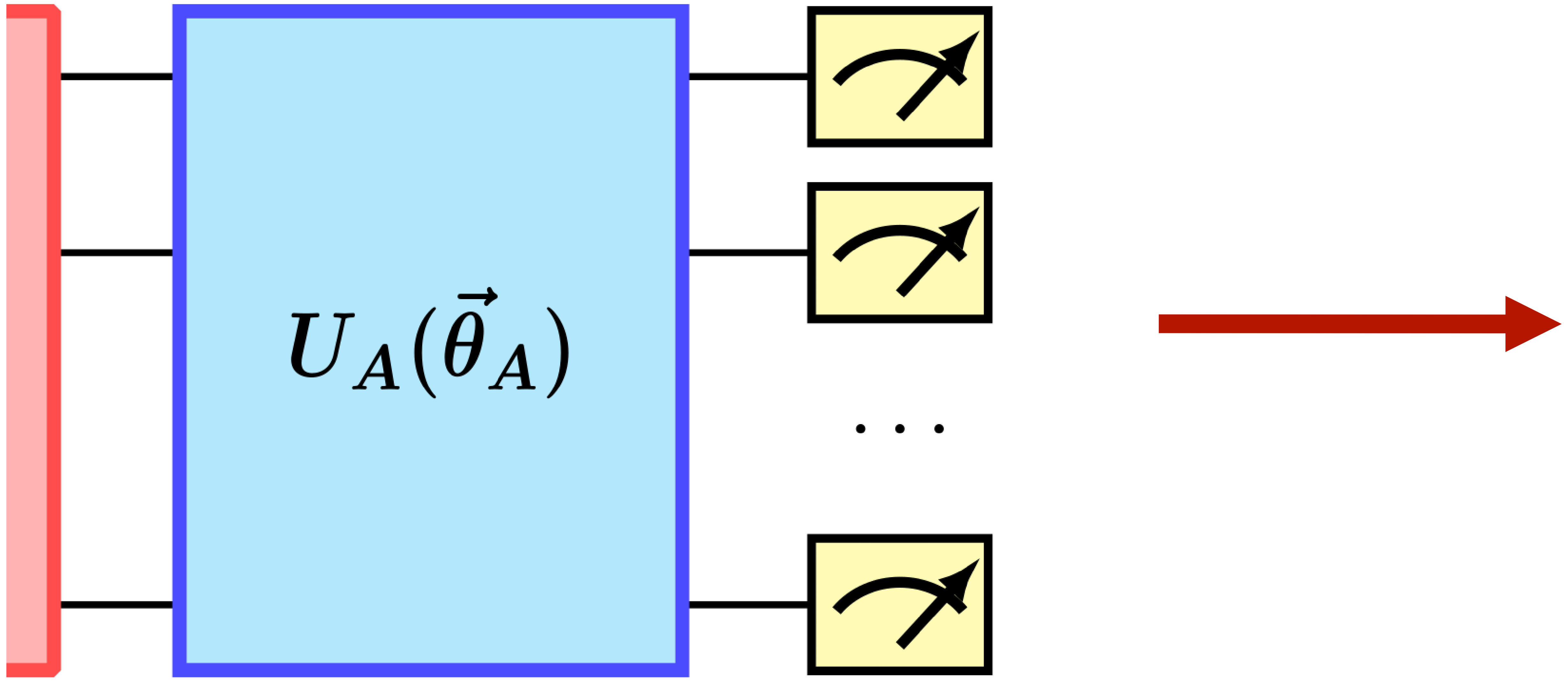
$$\hat{G}_m(\gamma_m) = \hat{Y}_m$$

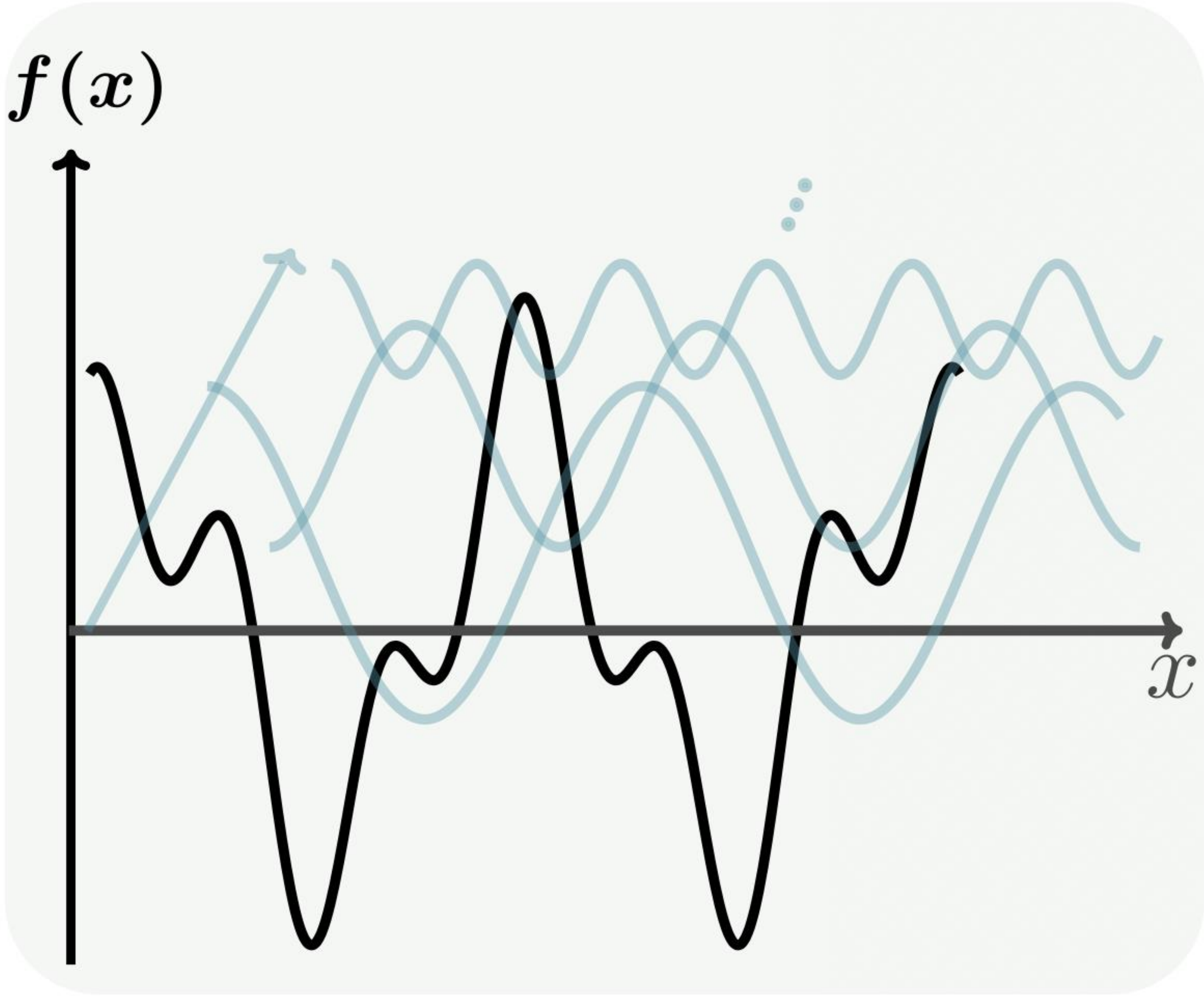
$$\phi(\vec{x}) = x$$



$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\hat{G}_m(\gamma_m)\,\phi(\vec{x})}$$

Quantum models





$$f(\vec{x}, \vec{\theta}_A) = \sum_{\vec{\omega}_j \in \Omega} \vec{c}_j(\vec{\theta}_A) e^{i\vec{\omega}_j \cdot \phi(\vec{x})}$$

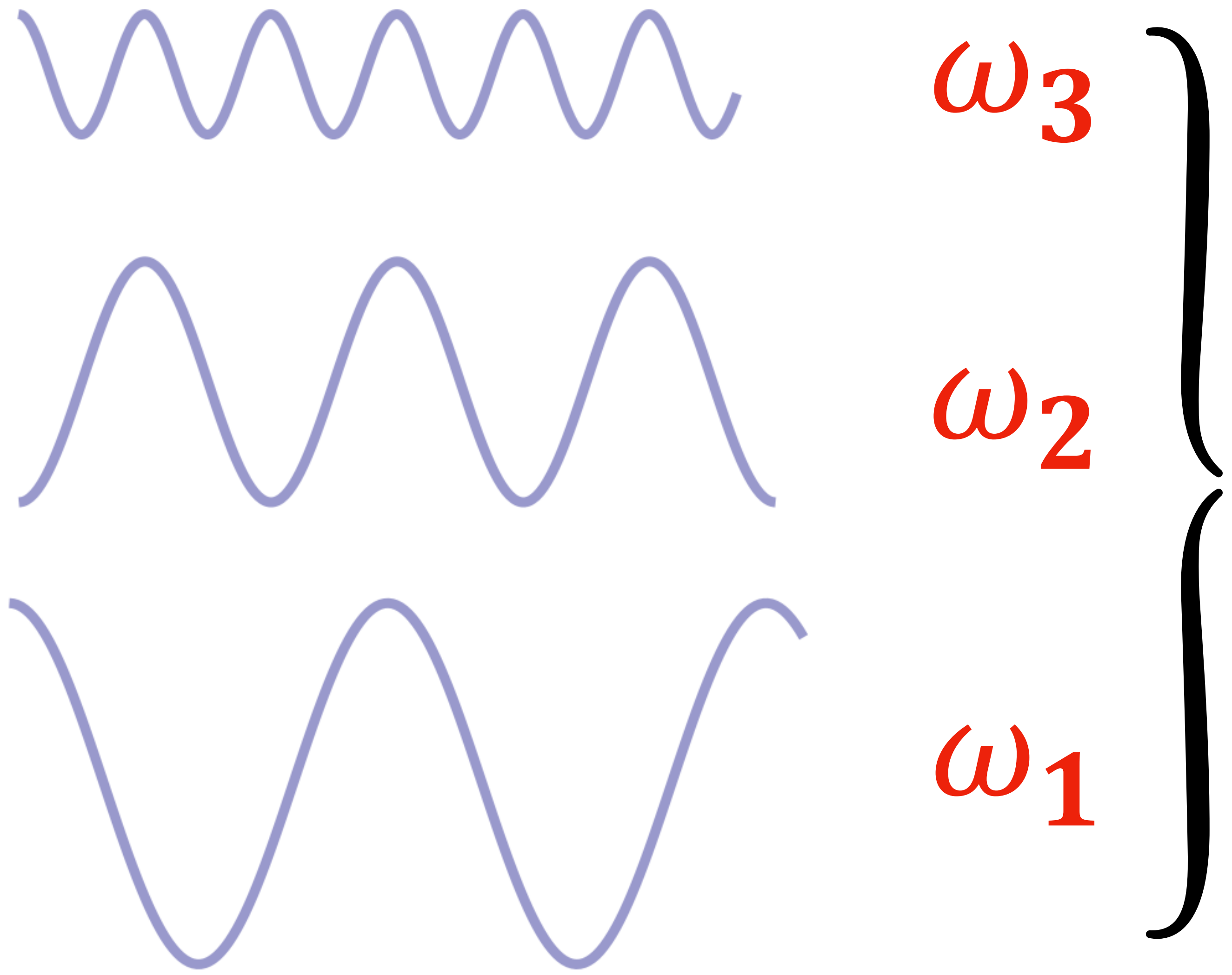
The measured output of a quantum model is a partial Fourier series

Schuld, Maria, Ryan Sweke, and Johannes Jakob Meyer. "Effect of data encoding on the expressive power of variational quantum-machine-learning models." Physical Review A 103.3 (2021): 032430.

Quantum models

$$\hat{G} = \sum_m \gamma_m \hat{Y}_m \in \mathbb{R}^{2^n \times 2^n}$$

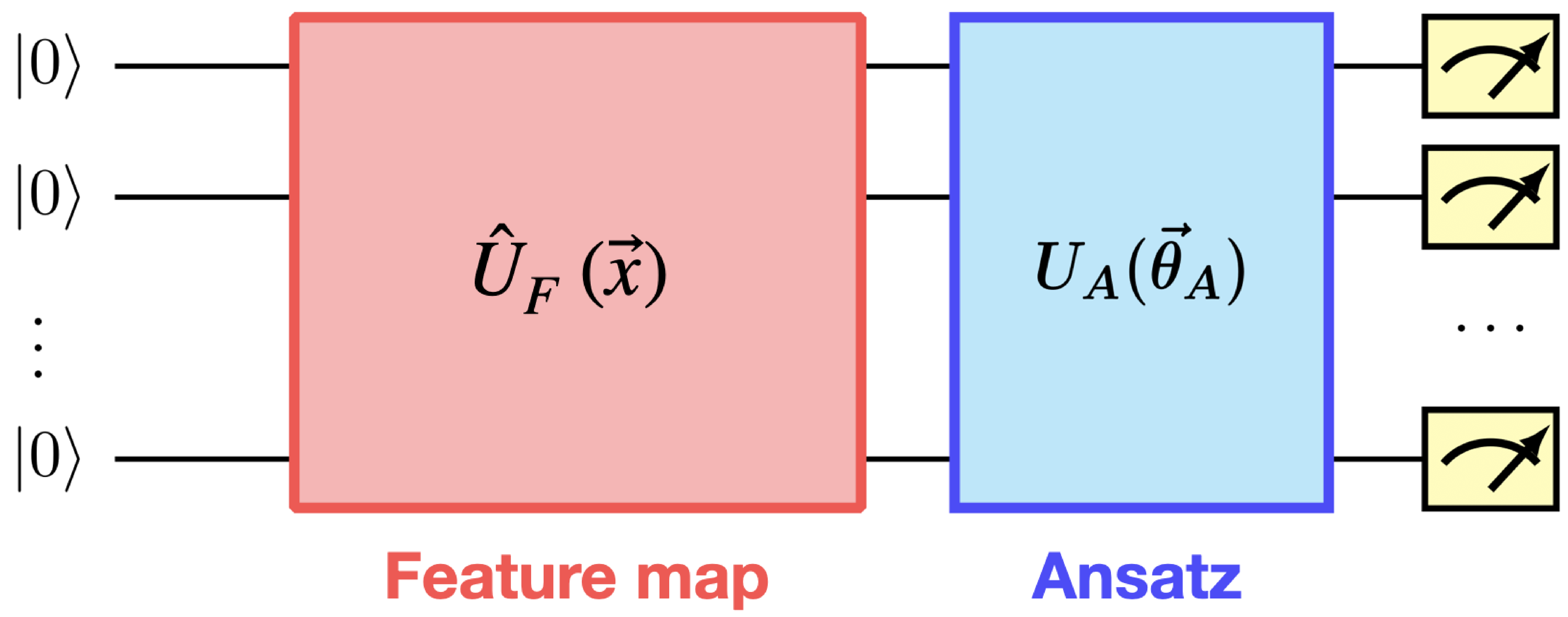
$$\hat{G}\,|\psi\rangle = \lambda_j\,|\psi\rangle$$

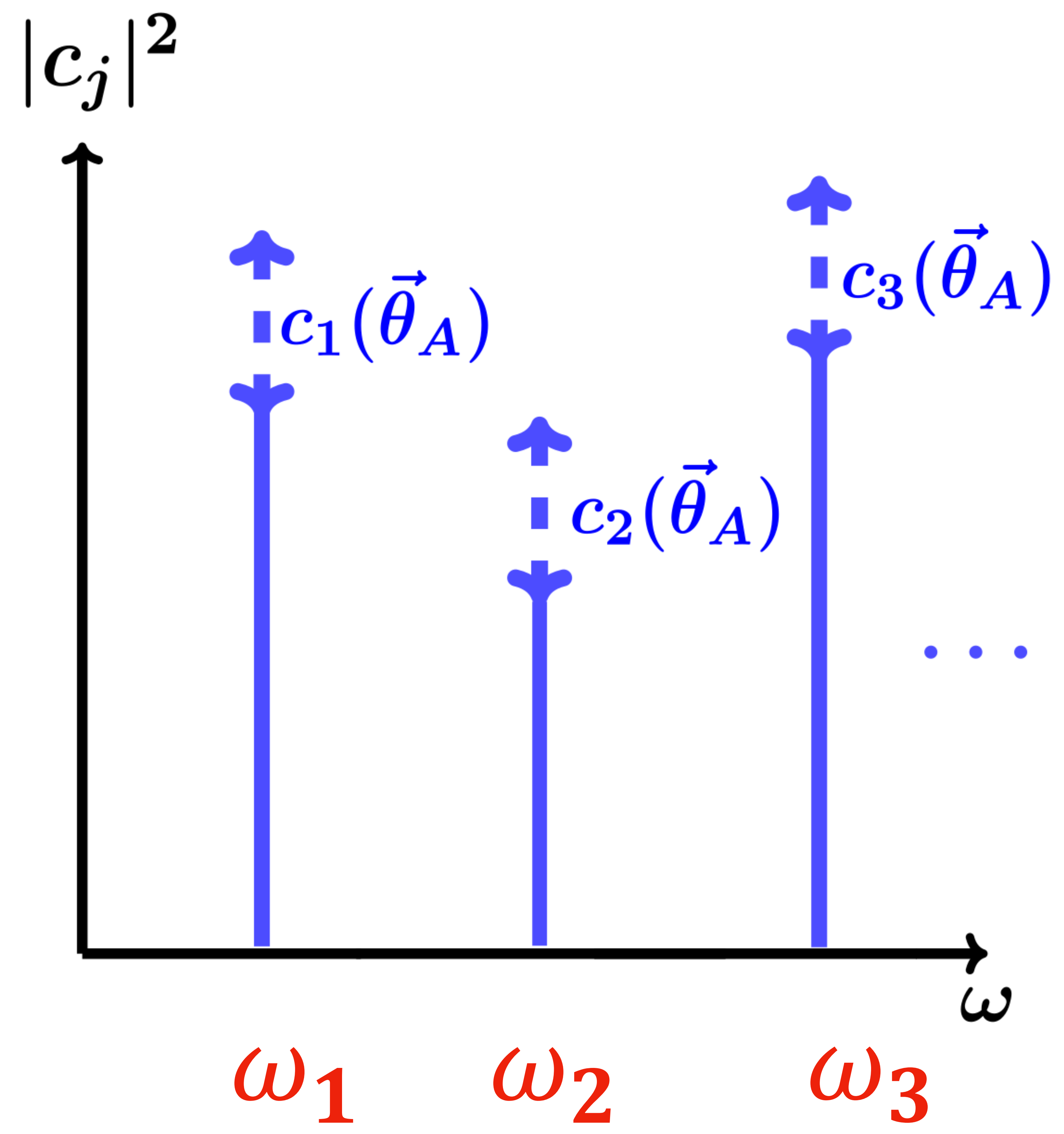


$$\Omega = \{\lambda_j\}$$

The frequencies of the Fourier series depend on the eigenvalues of the generator

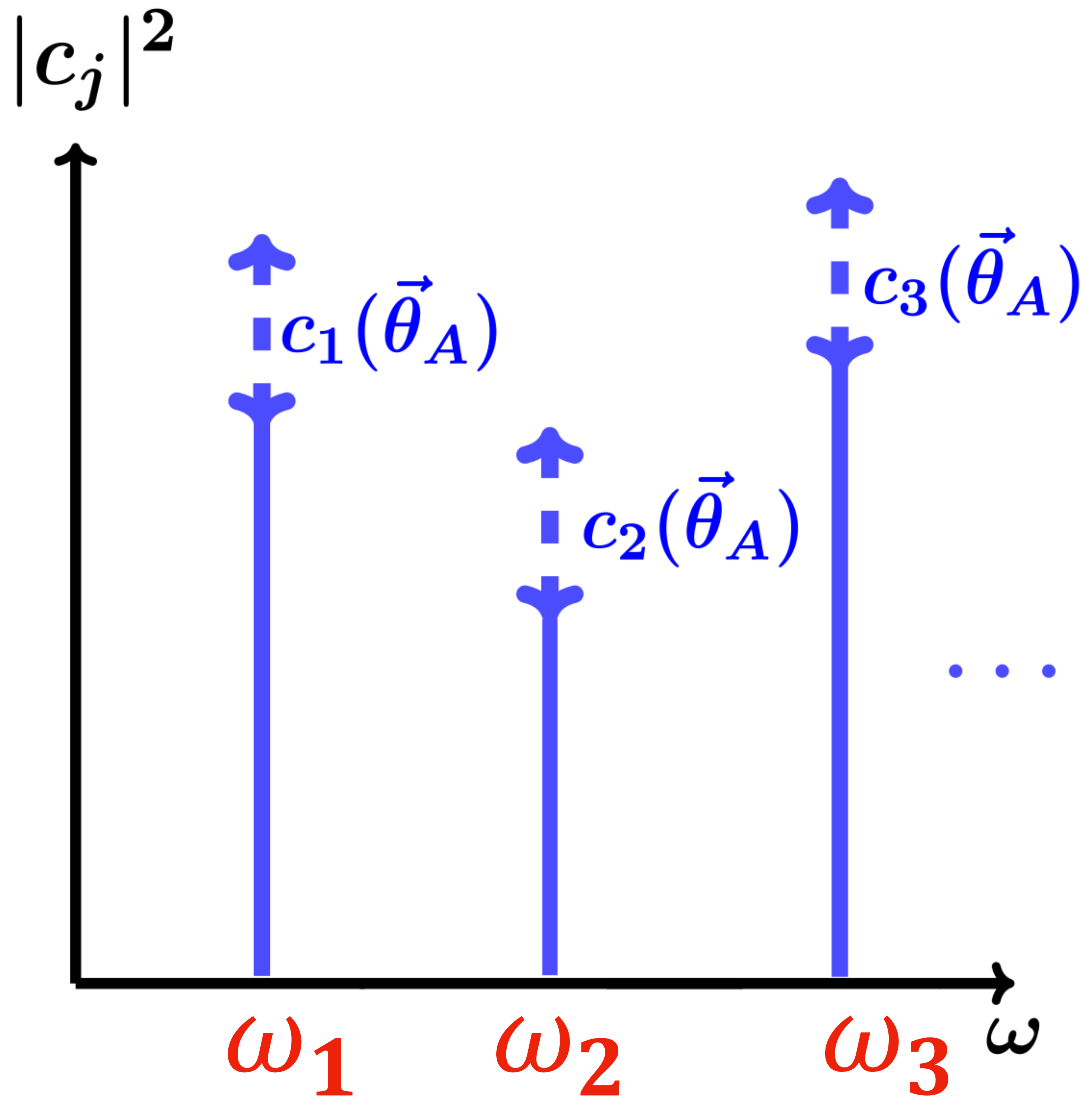
Quantum models

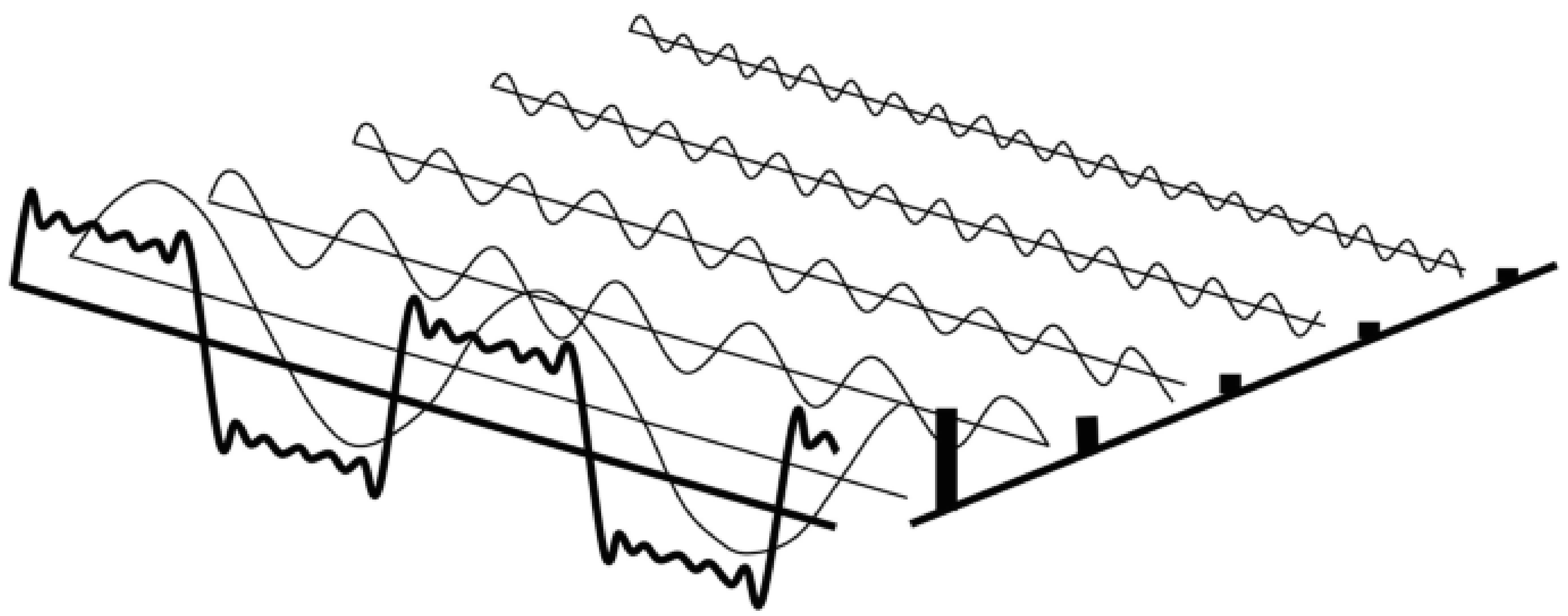




Conventional quantum models have trainable coefficients and fixed frequencies

Quantum models





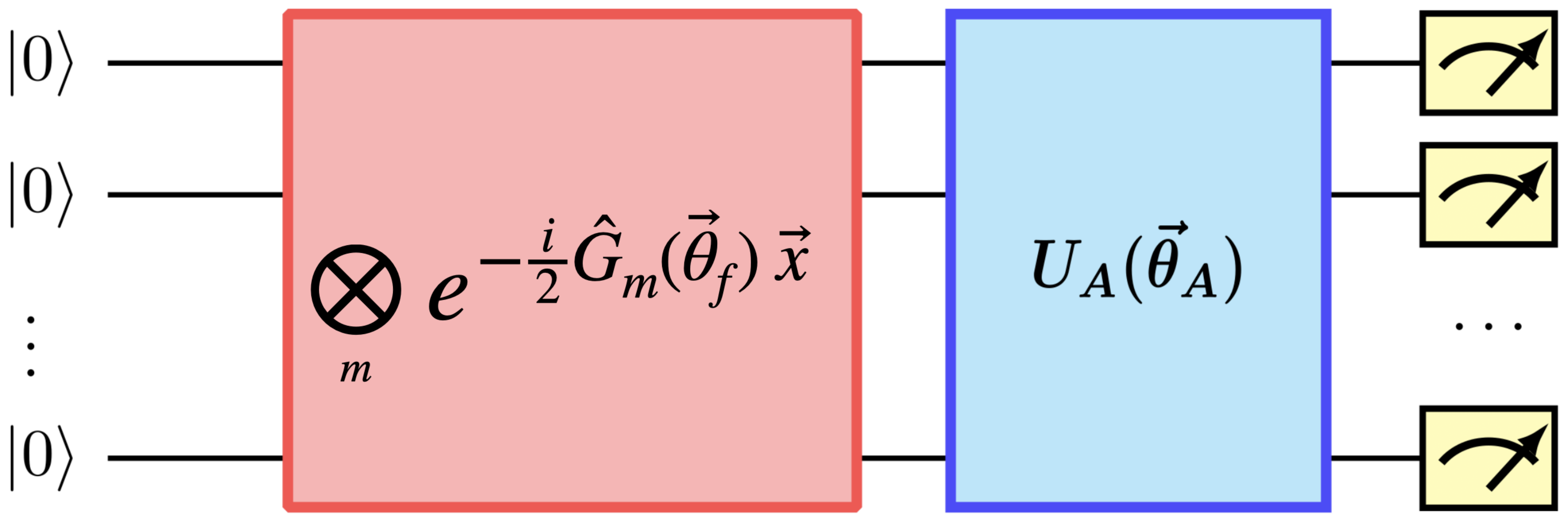
$$e^{ix} + e^{i2x} + e^{i3x} + \dots$$

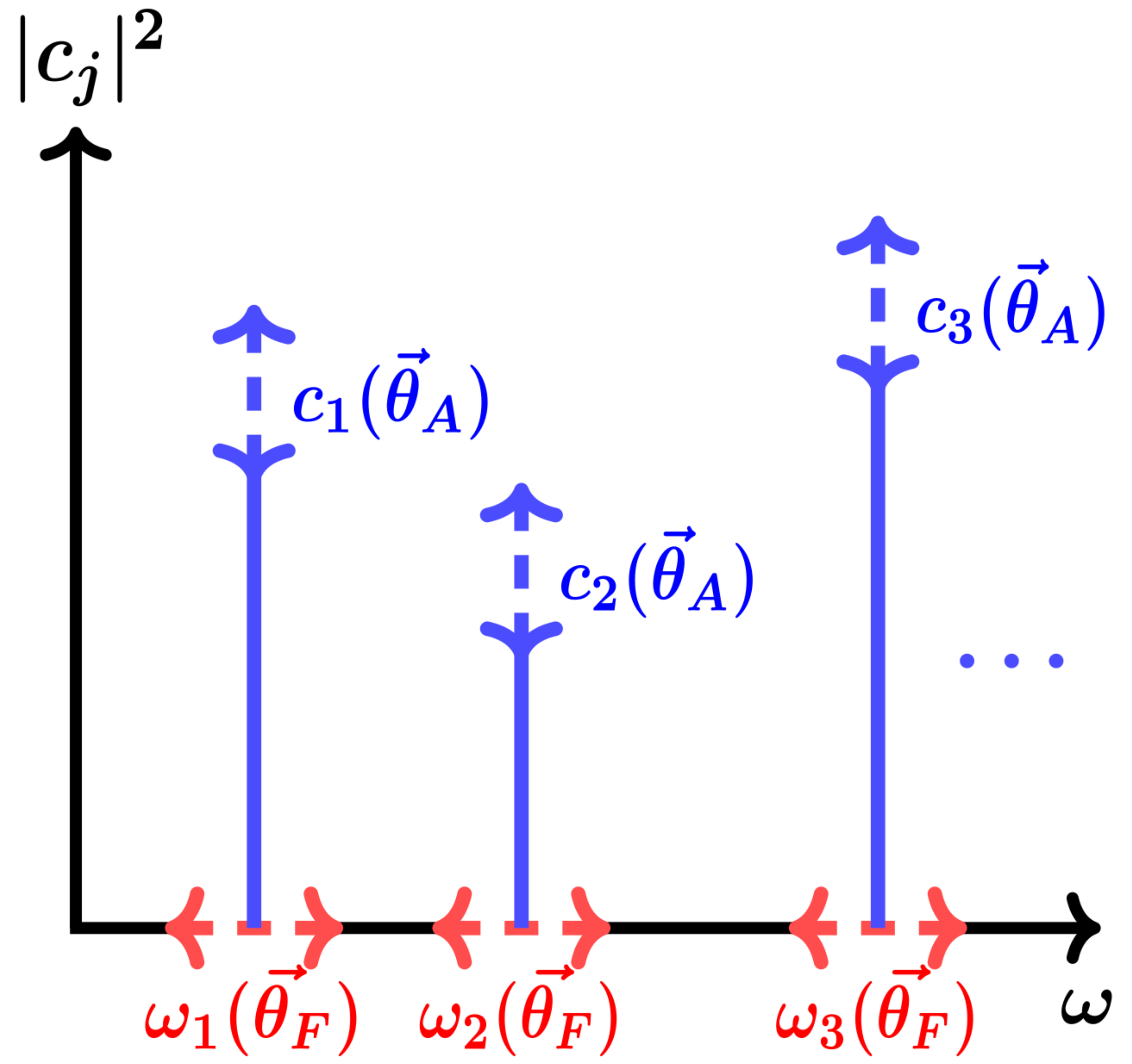
Orthogonal basis functions

Conventional quantum models have regularly spaced frequencies

They are universal in the asymptotic limit, but is that useful in practise?

Trainable frequency quantum models

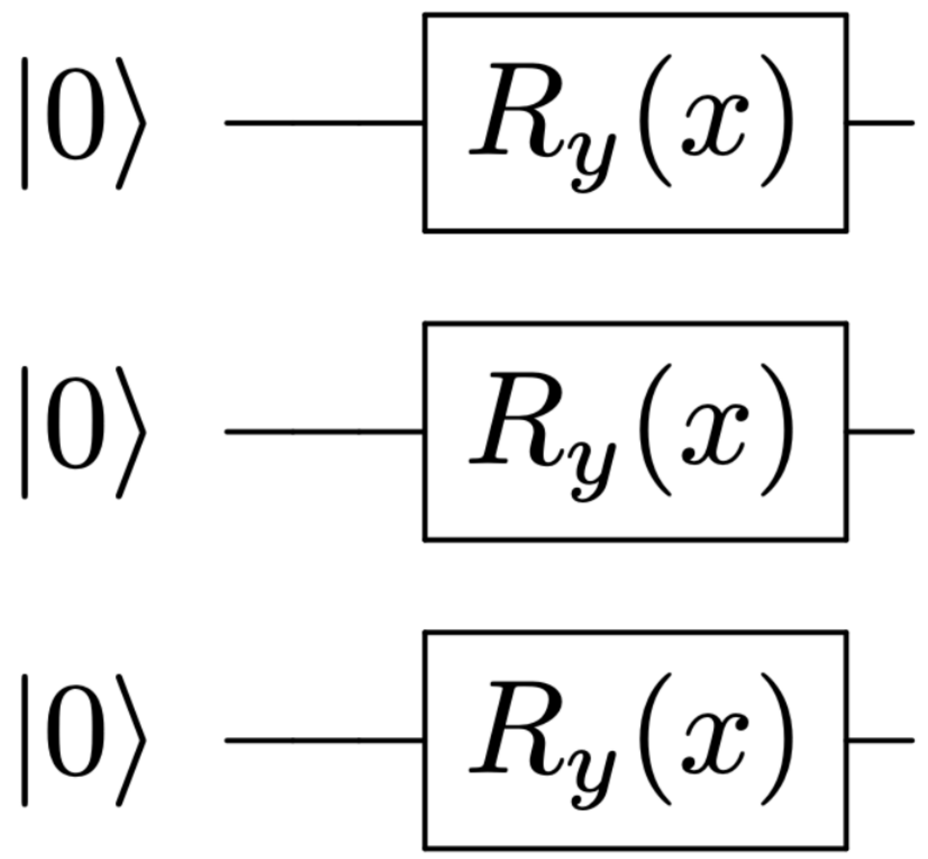




Including $\boldsymbol{\theta}_f$ in the feature map generator leads to a trainable frequency model

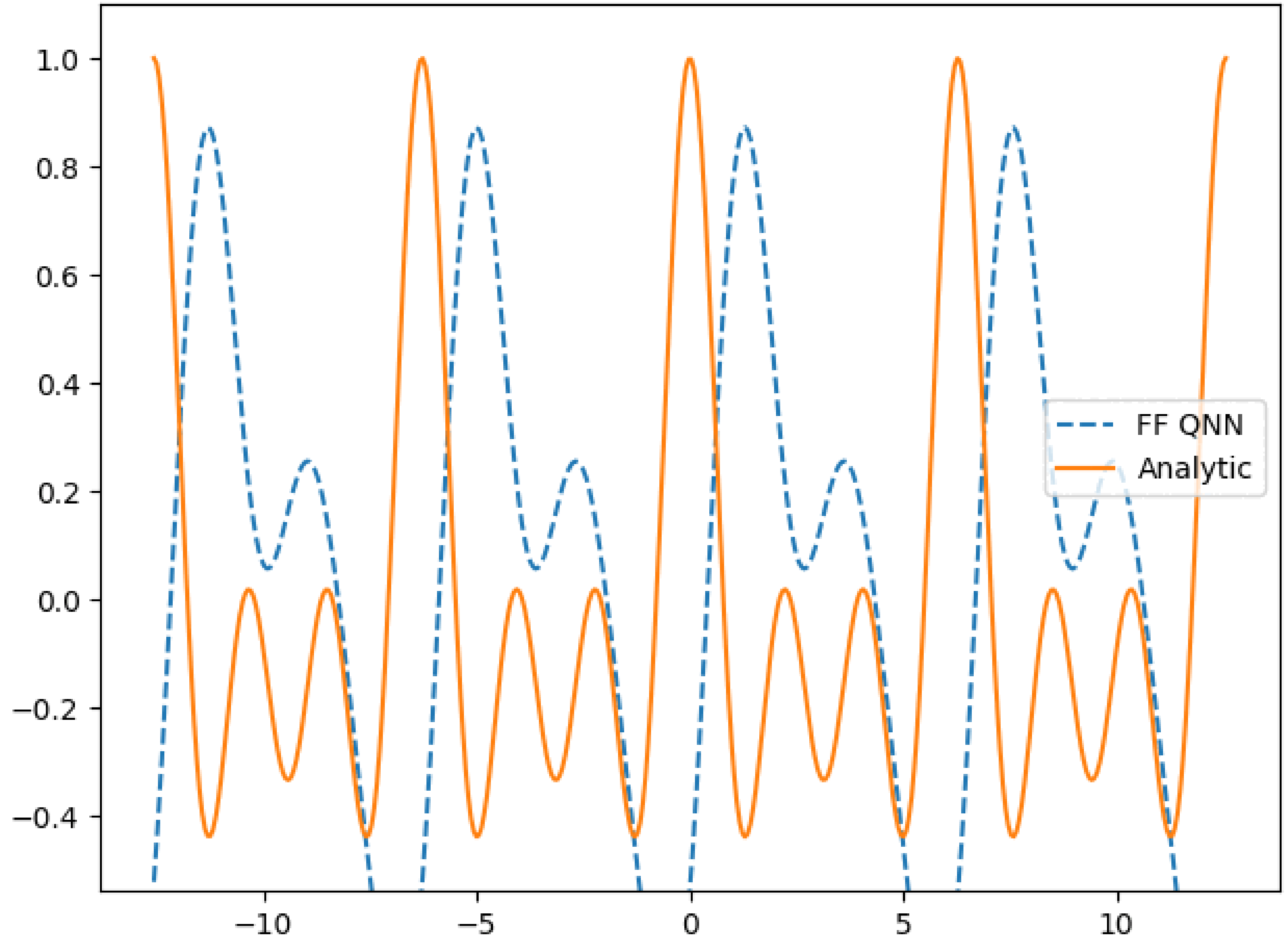
Proof-of-principle results

N=3 qubits
L=4 HEA layers



$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\hat{Y}_m x}$$

$$\Omega_d = \{1, 2, 3\}$$

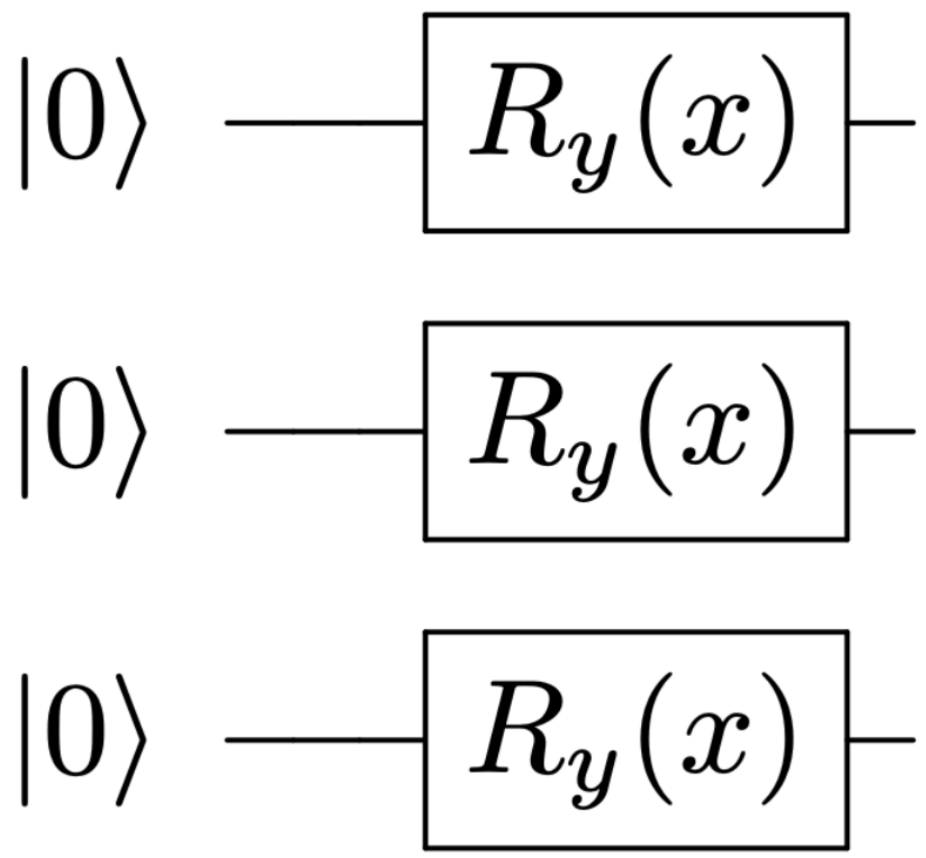


$$\lambda = \{-\tfrac{3}{2}, -\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{3}{2}\}$$

$$\Delta = \{1, 2, 3\}$$

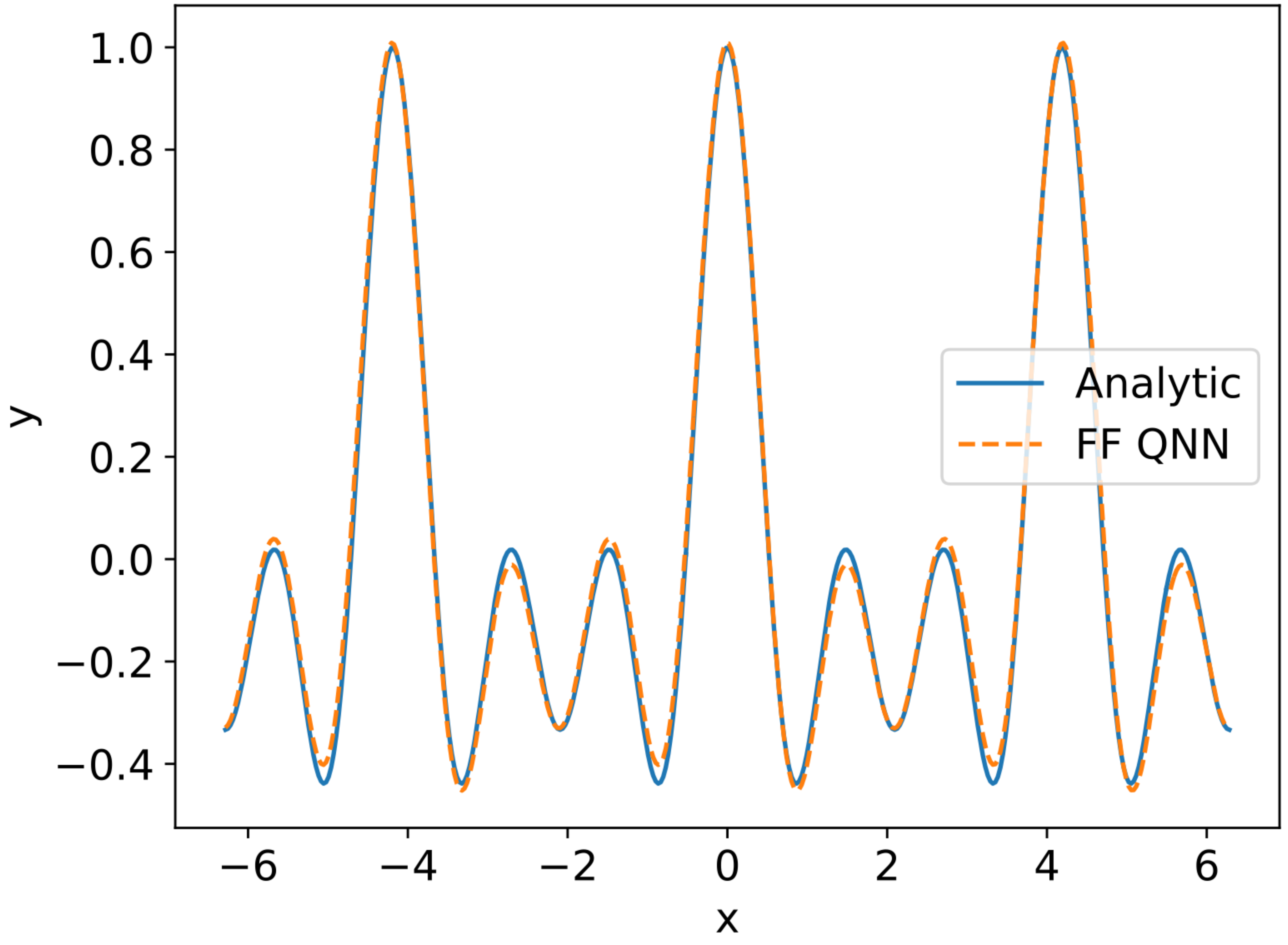
Proof-of-principle results

N=3 qubits
L=4 HEA layers



$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\hat{Y}_m x}$$

$$\Omega_d = \{1.5, 3, 4.5\}$$



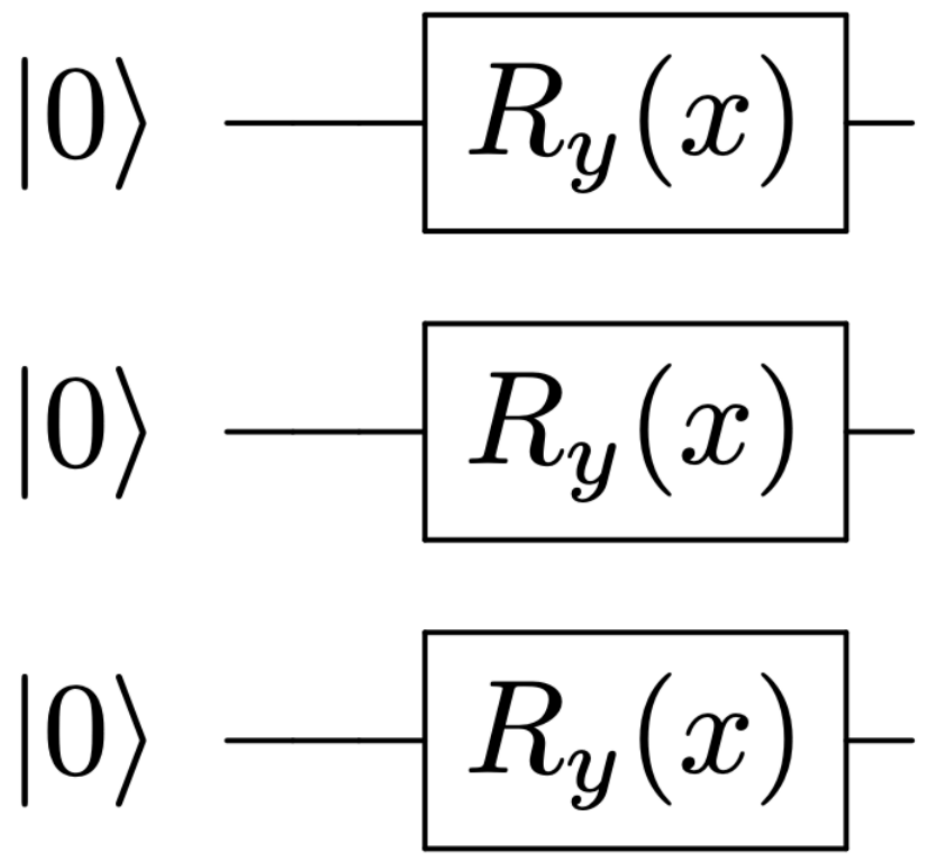
$$\lambda = \{-\tfrac{3}{2}, -\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{3}{2}\}$$

$$\Delta = \{1, 2, 3\}$$

Given a trivial classical resource of trainable scaling of the input

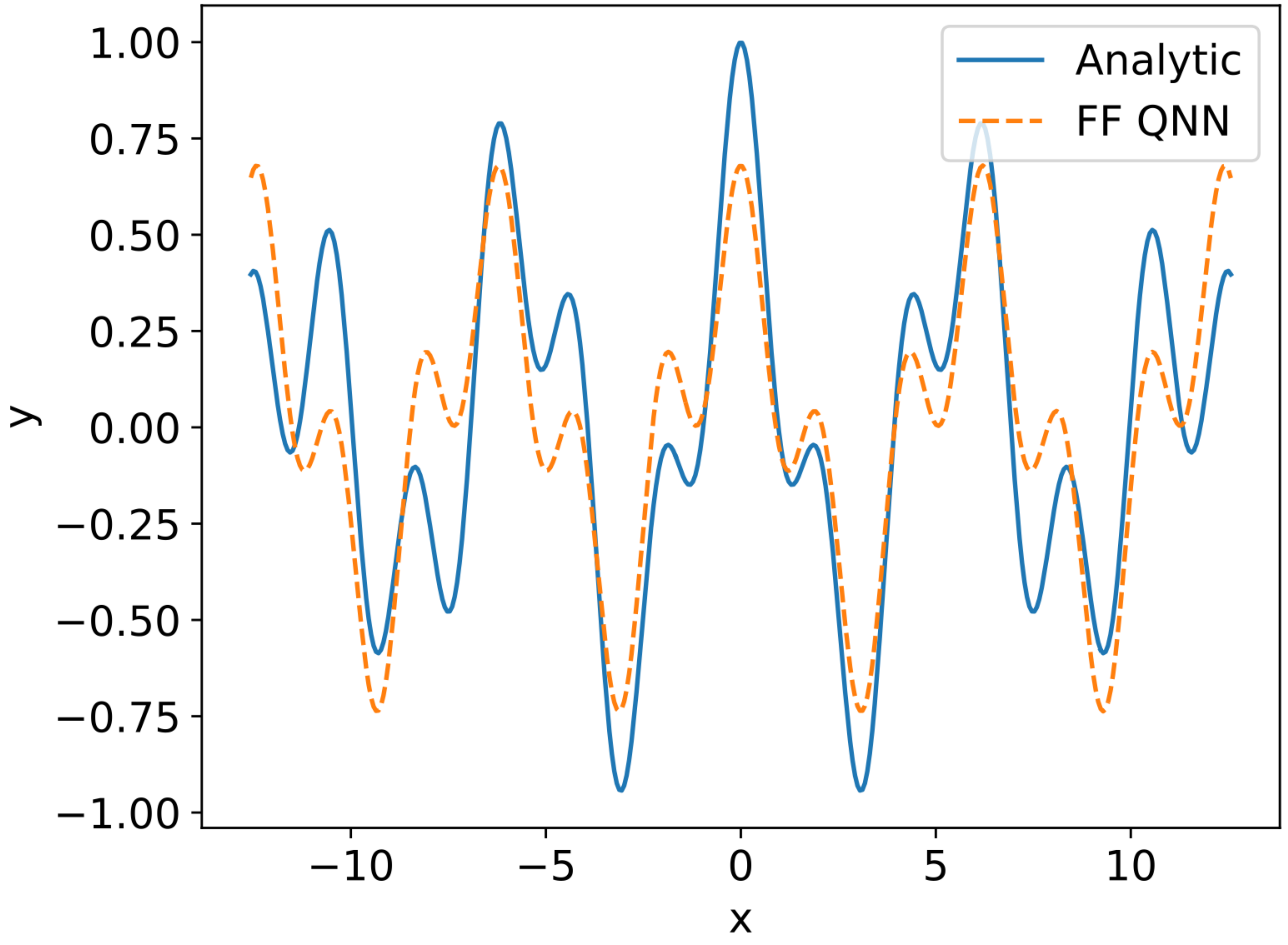
Proof-of-principle results

N=3 qubits
L=4 HEA layers



$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\hat{Y}_m x}$$

$$\Omega_d = \{1, 1.2, 3\}$$



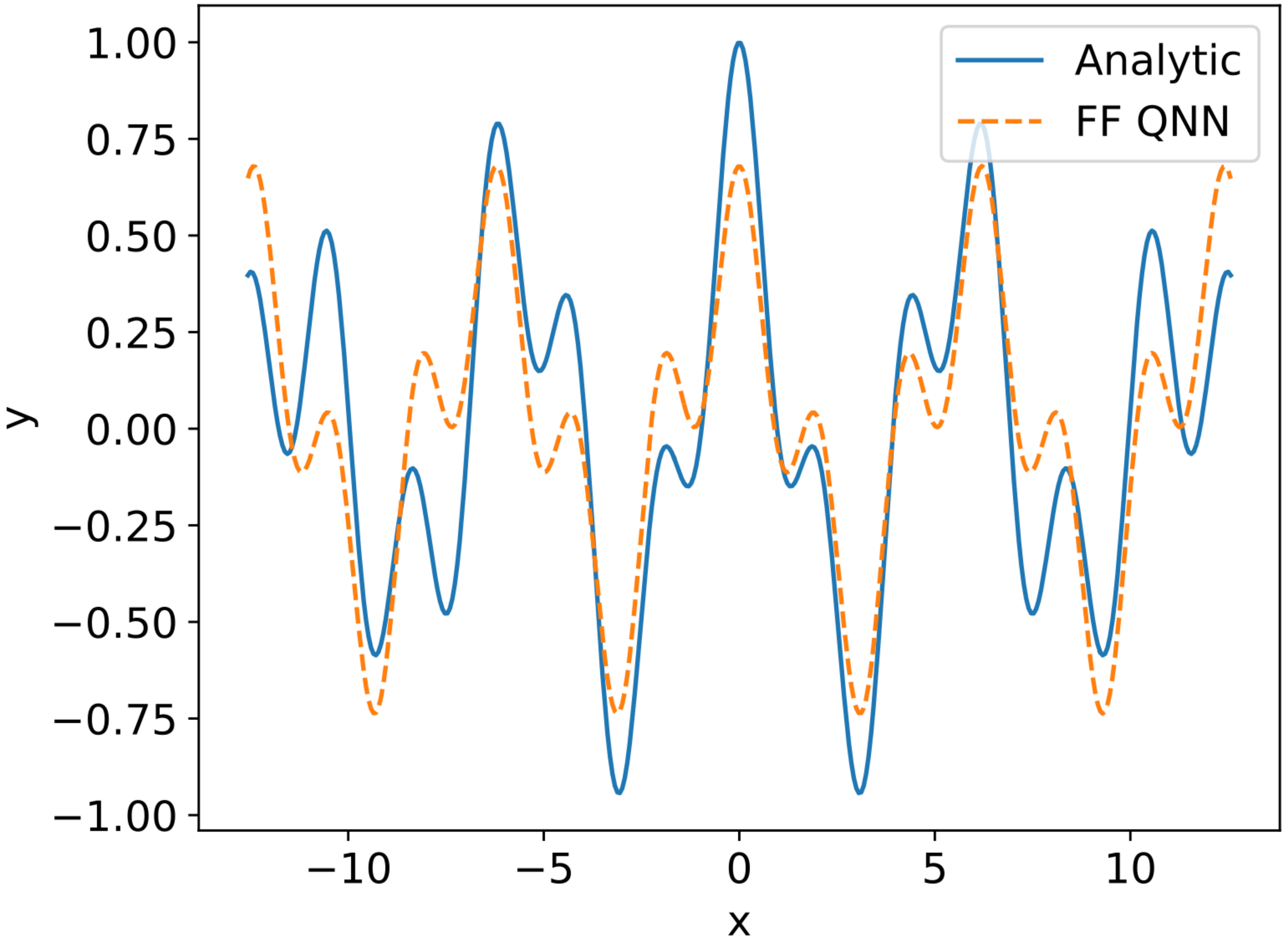
$$\lambda = \{-\tfrac{3}{2}, -\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{3}{2}\}$$

$$\Delta = \{1, 2, 3\}$$

Proof-of-principle results

No global scaling of the data can enable the fixed generator eigenspectrum to contain gaps with unequal spacing.

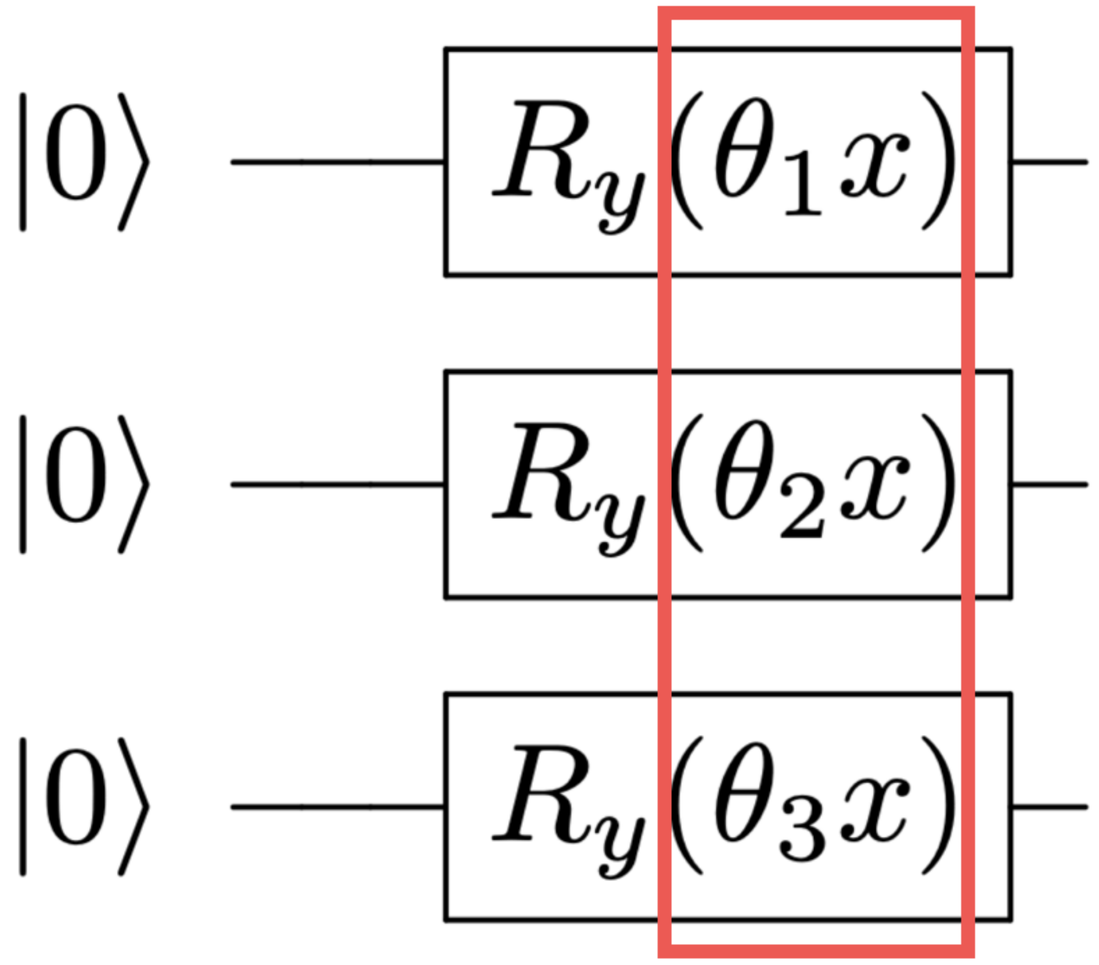
$$\Omega_d = \{1, 1.2, 3\}$$



$$\Delta = \{1, 2, 3\}$$

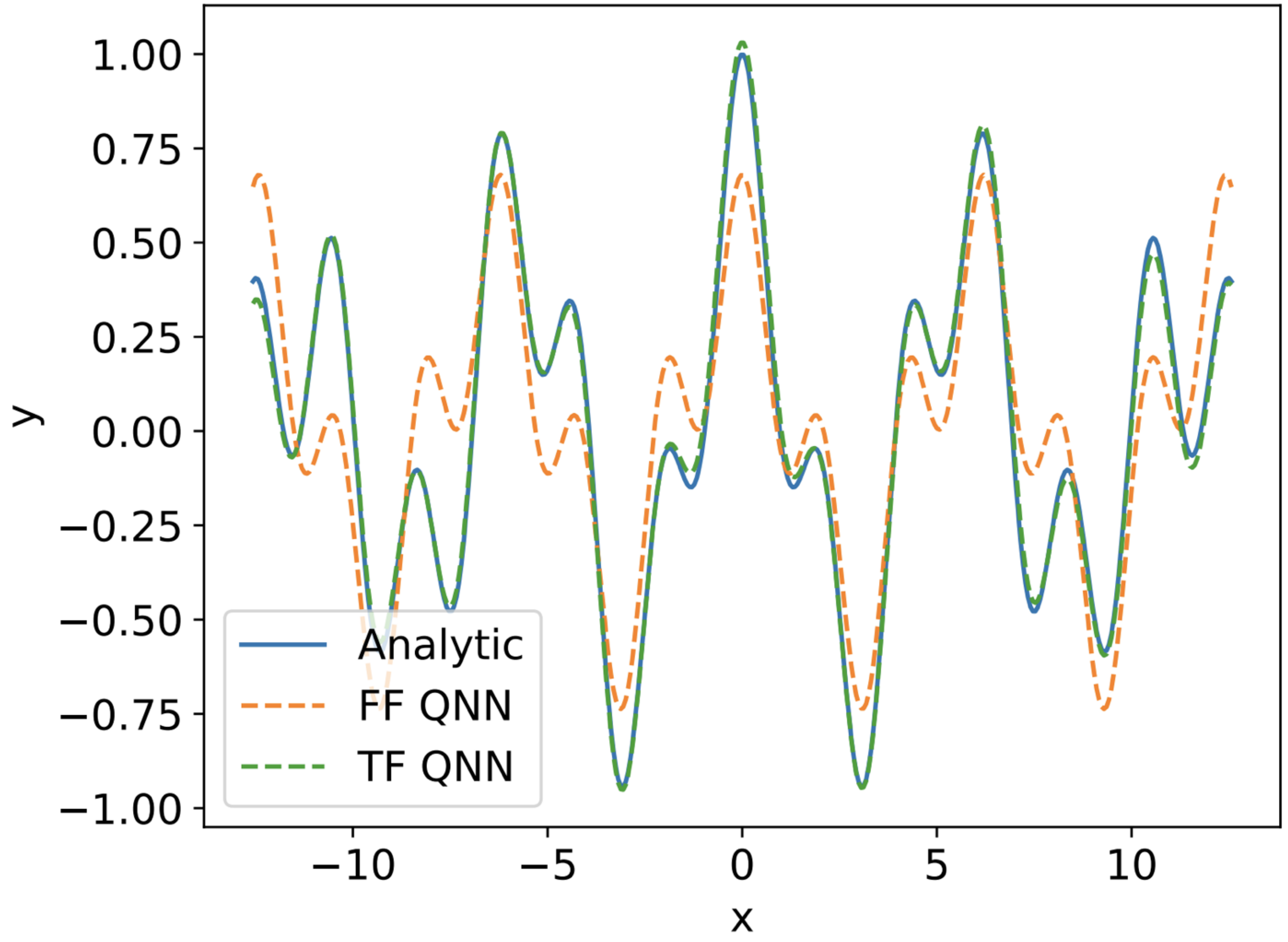
Proof-of-principle results

N=3 qubits
L=4 HEA layers



$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\theta_m \hat{Y}_m x}$$

$$\Omega_d = \{1, 1.2, 3\}$$



$$\lambda = \{-\frac{1}{2}(\theta_1 + \theta_2 + \theta_3), \ldots, \frac{1}{2}(\theta_1 + \theta_2 + \theta_3)\}$$

Proof-of-principle results

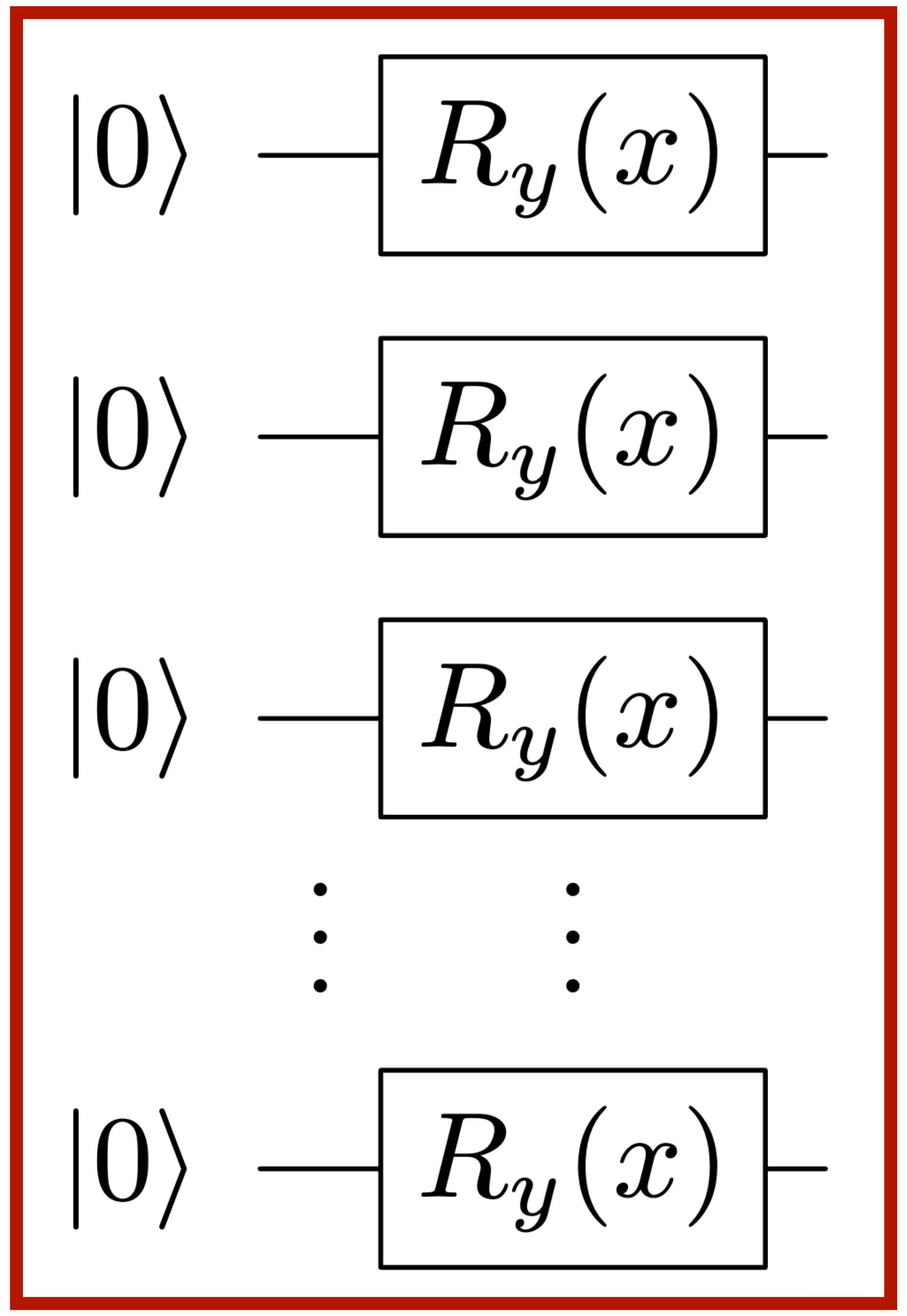
$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\gamma_m \hat{Y}_m \phi(\vec{x})}$$

Spectrum

Simple	$\gamma_m = 1$
Tower	$\gamma_m = m$
Exponential	$\gamma_m = 2^{(m-1)}$

Proof-of-principle results

$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\gamma_m \hat{Y}_m \phi(\vec{x})}$$

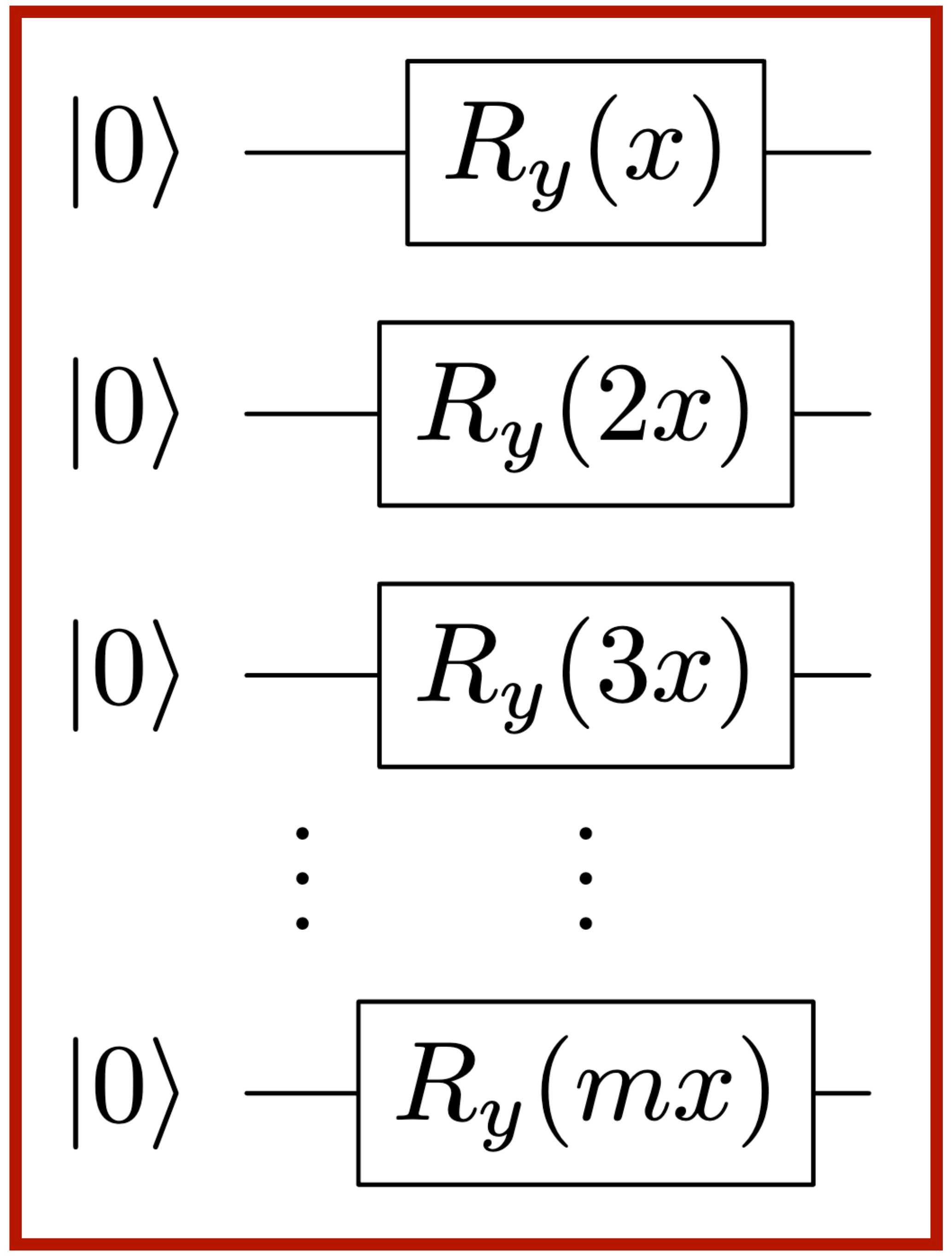


Spectrum

Simple	$\gamma_m = 1$
Tower	$\gamma_m = m$
Exponential	$\gamma_m = 2^{(m-1)}$

Proof-of-principle results

$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\gamma_m \hat{Y}_m \phi(\vec{x})}$$



Spectrum

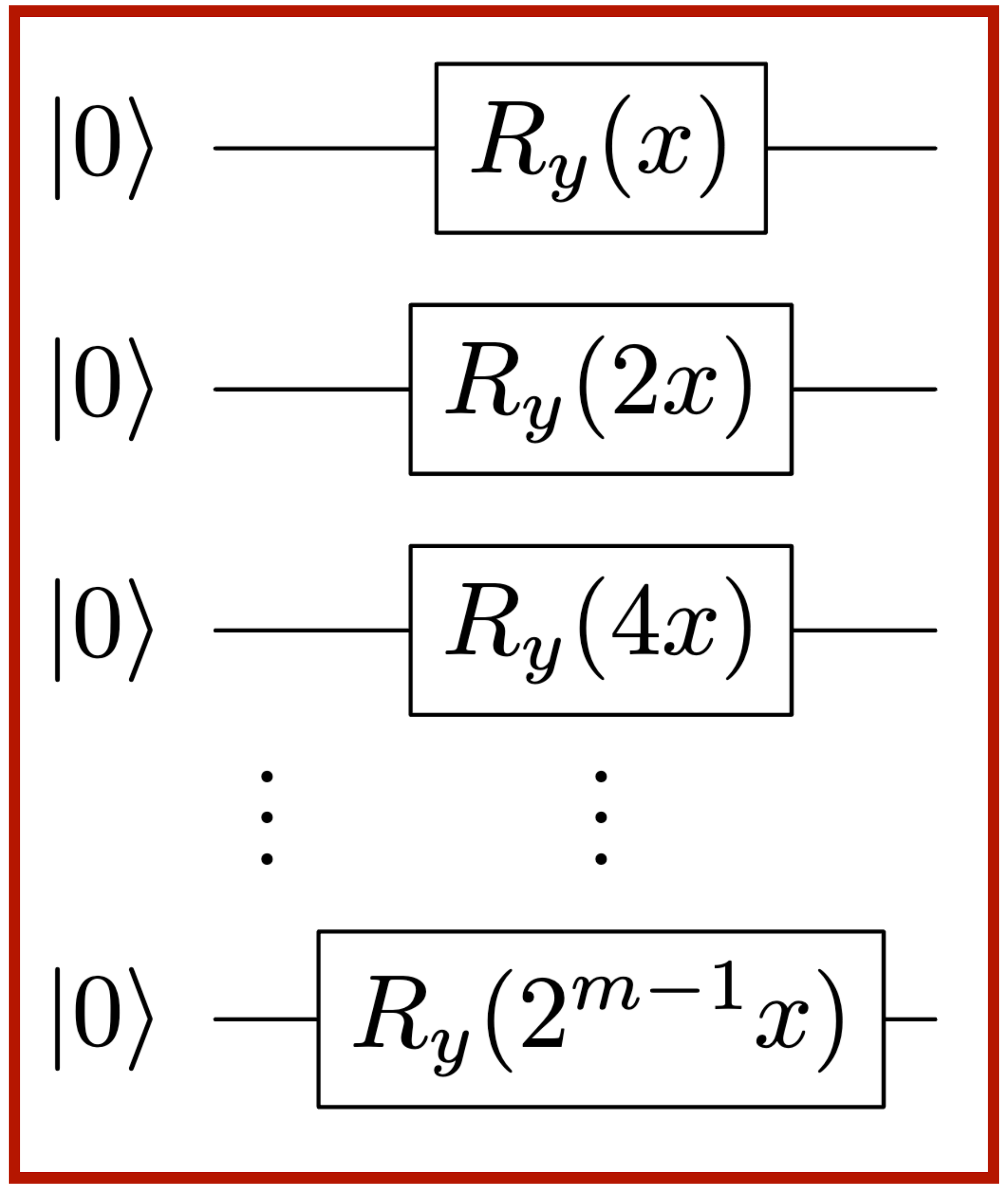
Simple $\gamma_m = 1$

Tower $\gamma_m = m$

Exponential $\gamma_m = 2^{(m-1)}$

Proof-of-principle results

$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\gamma_m \hat{Y}_m \phi(\vec{x})}$$



Spectrum

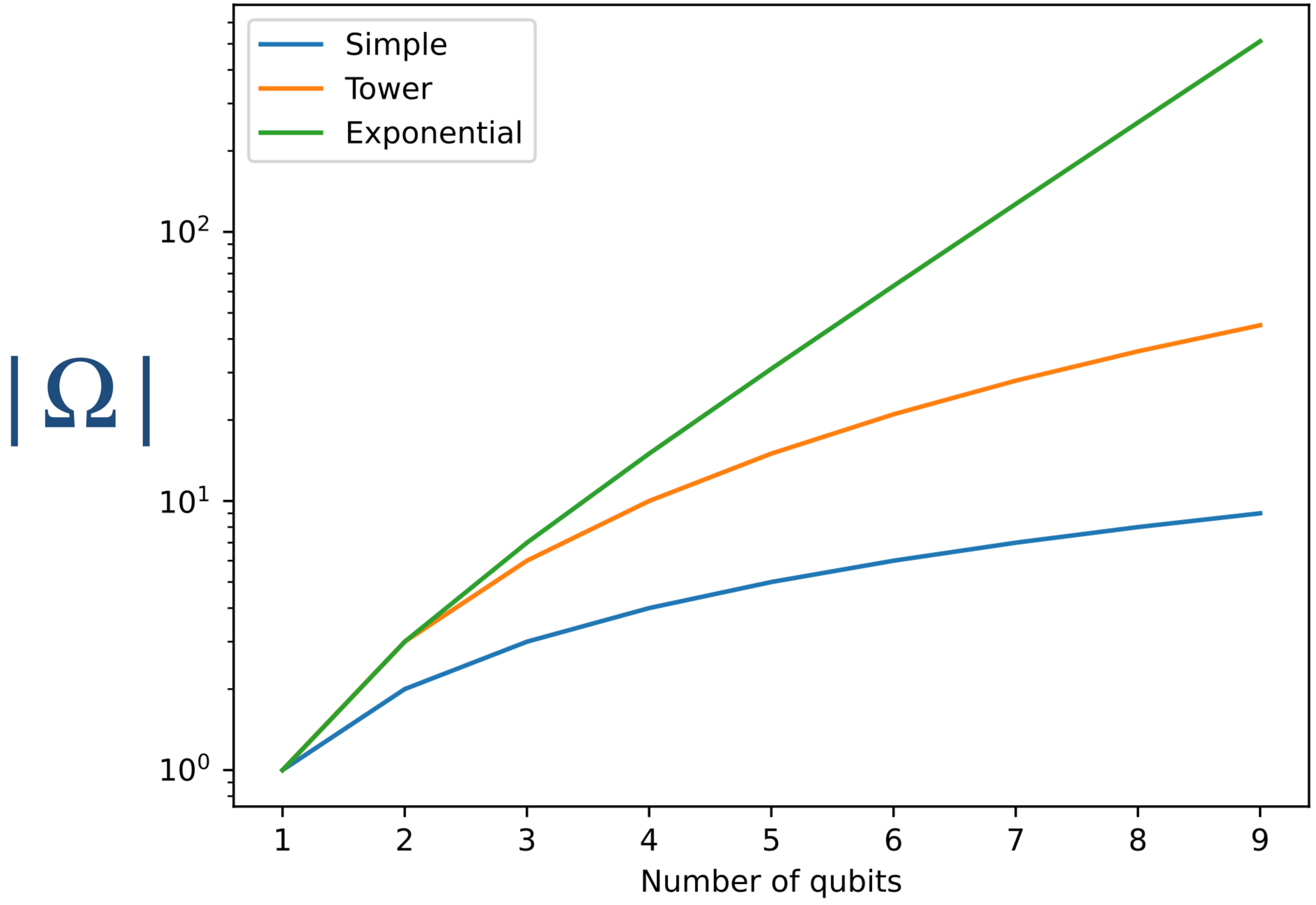
Simple	$\gamma_m = 1$
Tower	$\gamma_m = m$
Exponential	$\gamma_m = 2^{(m-1)}$

Proof-of-principle results

$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\gamma_m \hat{Y}_m \phi(\vec{x})}$$

Spectrum

Simple	$\gamma_m = 1$
Tower	$\gamma_m = m$
Exponential	$\gamma_m = 2^{(m-1)}$



Proof-of-principle results

FFFM

$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\gamma_m \hat{Y}_m x}$$

Spectrum

Simple $\gamma_m = 1$

Tower $\gamma_m = m$

Exponential $\gamma_m = 2^{(m-1)}$

TFFM

$$\hat{U}_F(\vec{x}) = \bigotimes_m e^{-\frac{i}{2}\theta_m \hat{Y}_m x}$$

Spectrum

Simple $\theta_m = 1$

Tower $\theta_m = m$

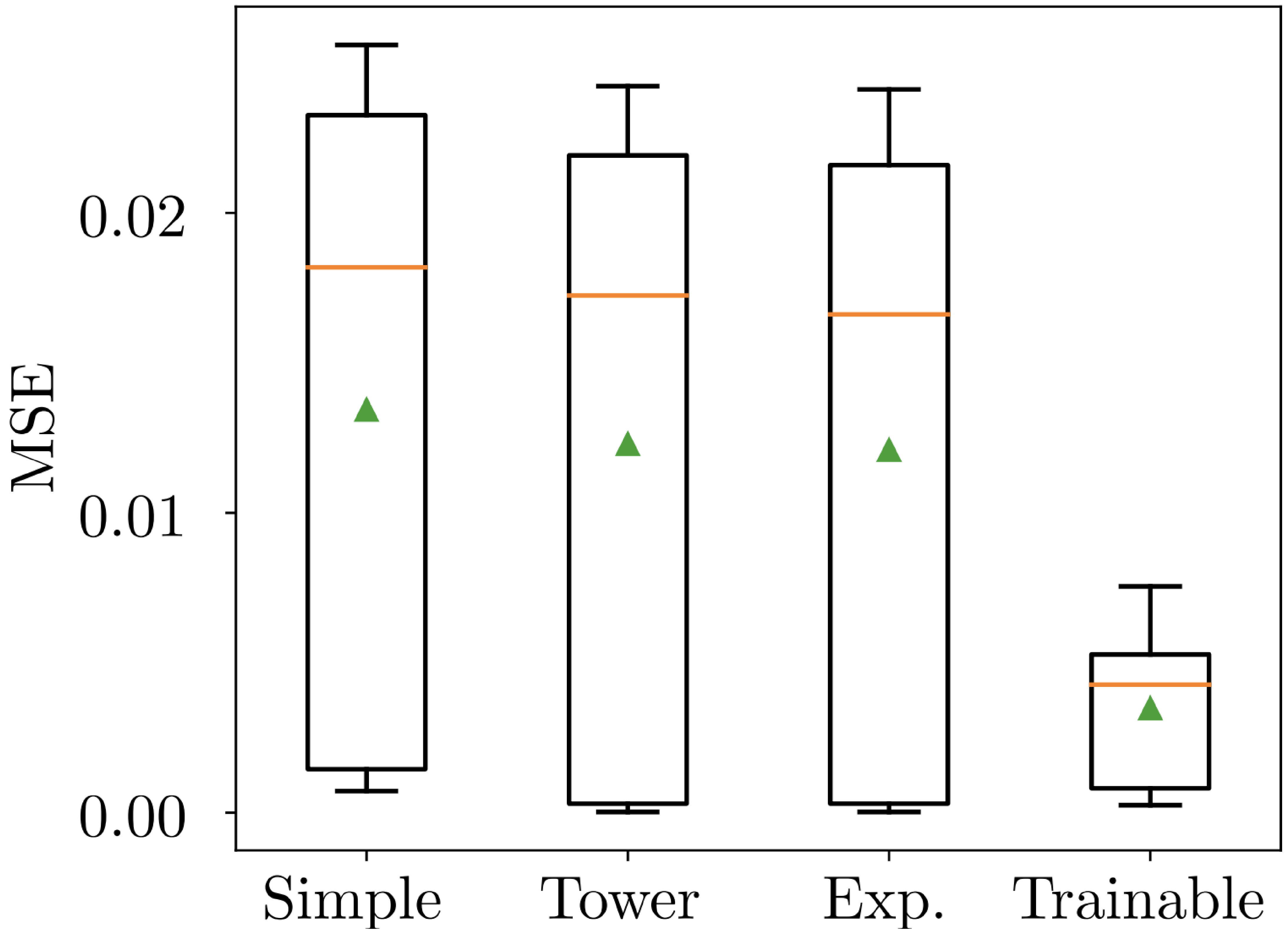
Exponential $\theta_m = 2^{(m-1)}$

Proof-of-principle results

$$y(x) = \frac{1}{|\Omega_d|} \sum_{\omega_d \in \Omega_D} \cos(\omega_d x)$$

$$\Omega_{d_1} = \{1\}$$
$$\Omega_{d_2} = \{1,3\}$$
$$\Omega_{d_3} = \{1,2,3\}$$
$$\cdots$$
$$\Omega_{d_7} = \{1, \frac{4}{3}, \frac{5}{3}, 2, \ldots, 3\}$$

N=3 qubits
L=4 HEA layers



Practical benefit of TF models

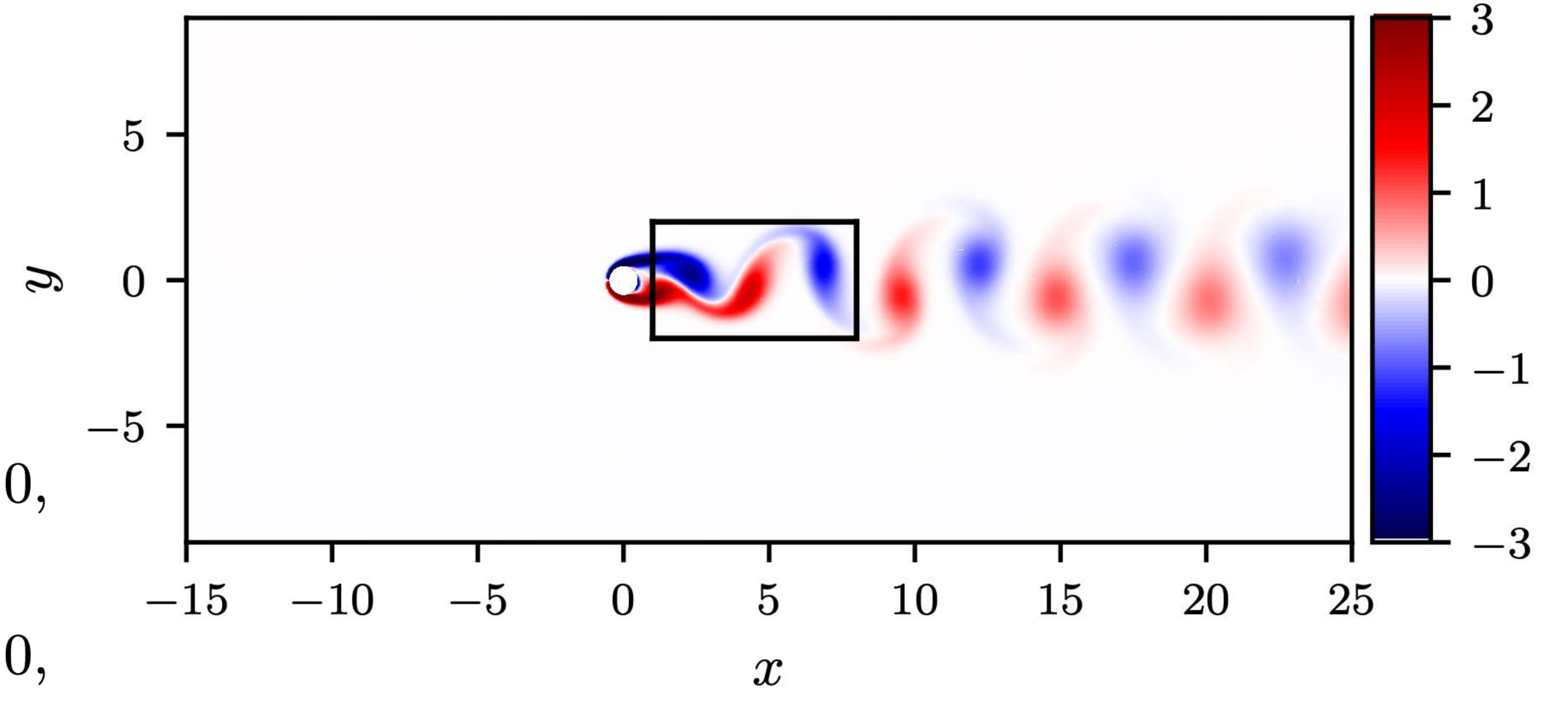
Goal: Learn a solution to the Navier-Stokes equations for fluid passing over a circular cylinder.

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - \frac{1}{Re}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \frac{\partial p}{\partial x} = 0,$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} - \frac{1}{Re}\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) + \frac{\partial p}{\partial y} = 0,$$

$$Re = 100$$

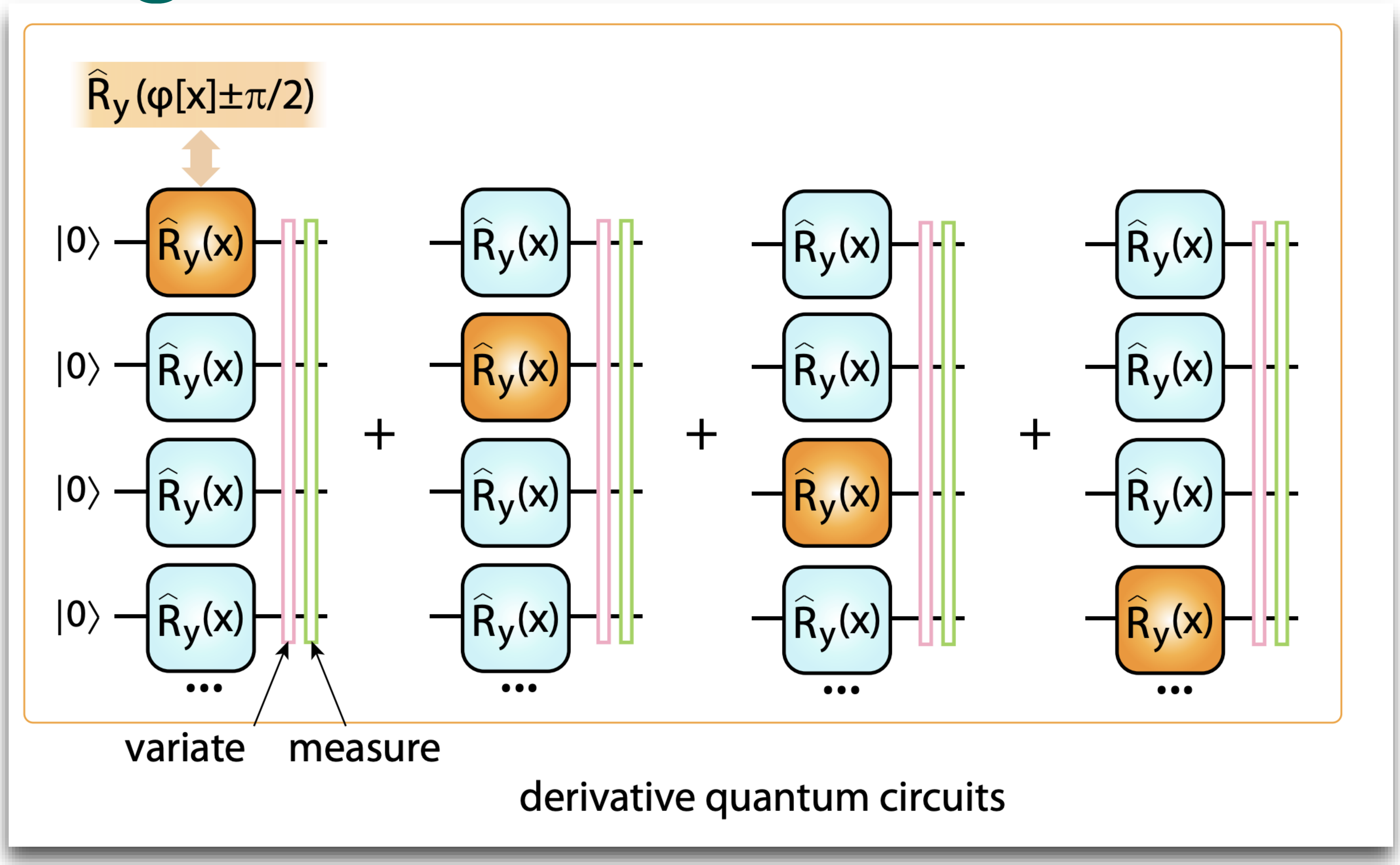
N=6, L=8, 192 parameters



600 batch size
5,000 training iterations

Raissi, Maziar, Paris Perdikaris, and George Em Karniadakis. "Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations." arXiv preprint arXiv:1711.10566 (2017).

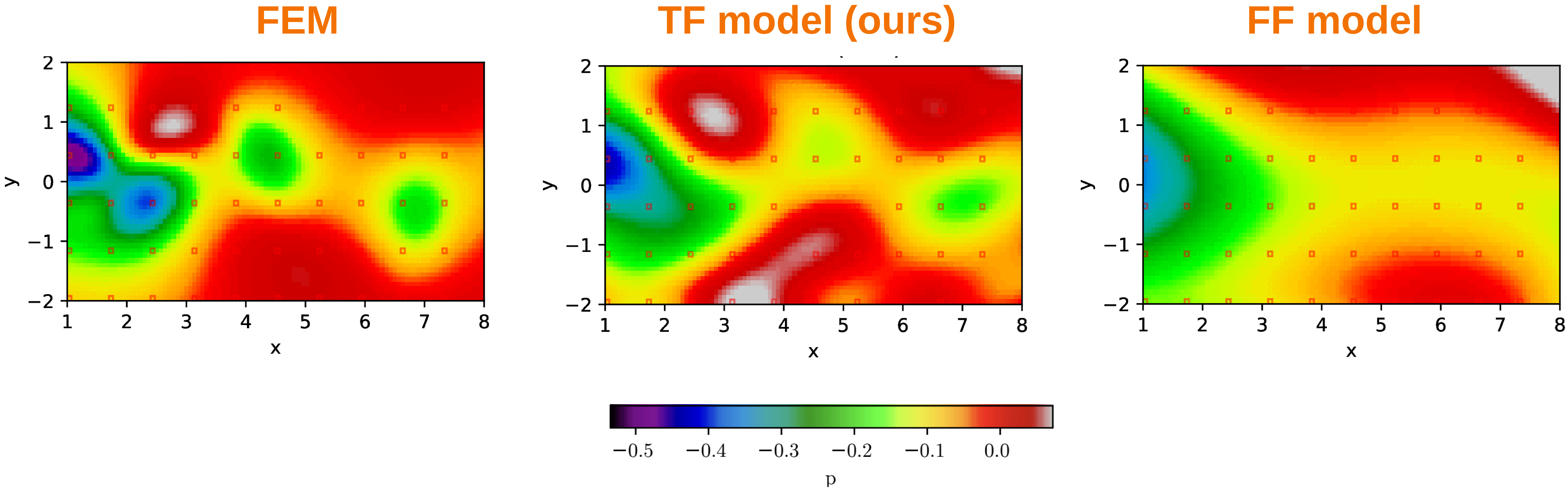
Differentiable quantum circuits (DQC) algorithm



Given a QNN representing f(x), we can obtain df/dx using derivative quantum circuits

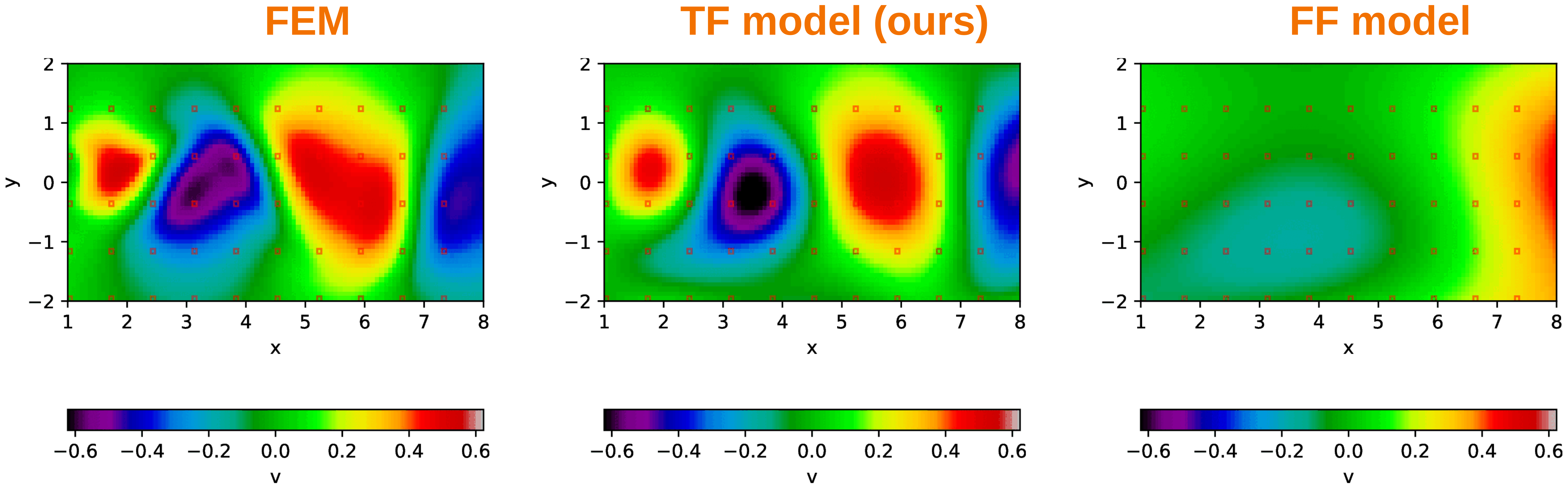
Kyriienko, Oleksandr, Annie E. Paine, and Vincent E. Elfving. "Solving nonlinear differential equations with differentiable quantum circuits." Physical Review A 103.5 (2021): 052416.

Practical benefit of TF models



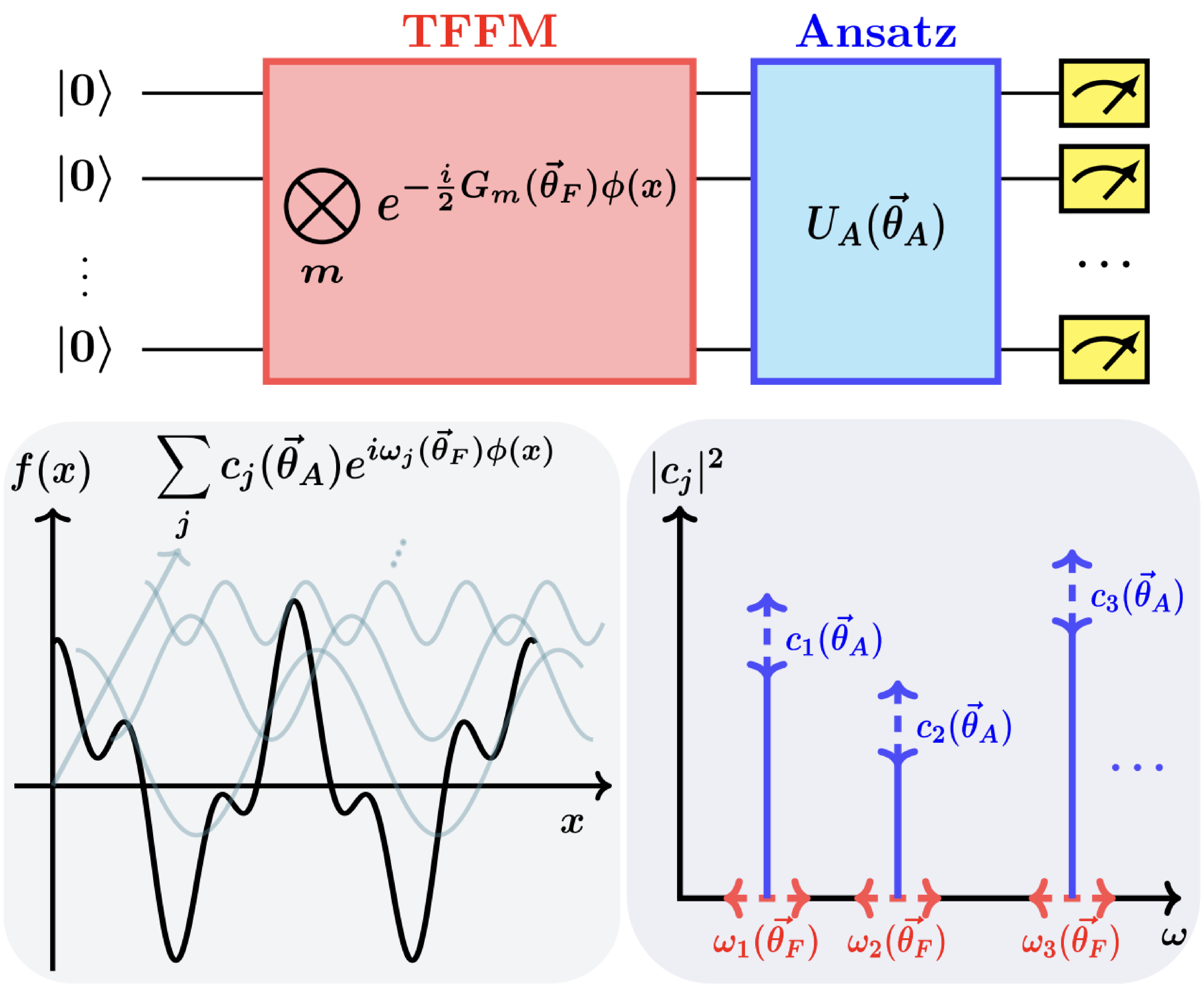
Practical benefit of TF models

Practical benefit of TF models



Conclusion

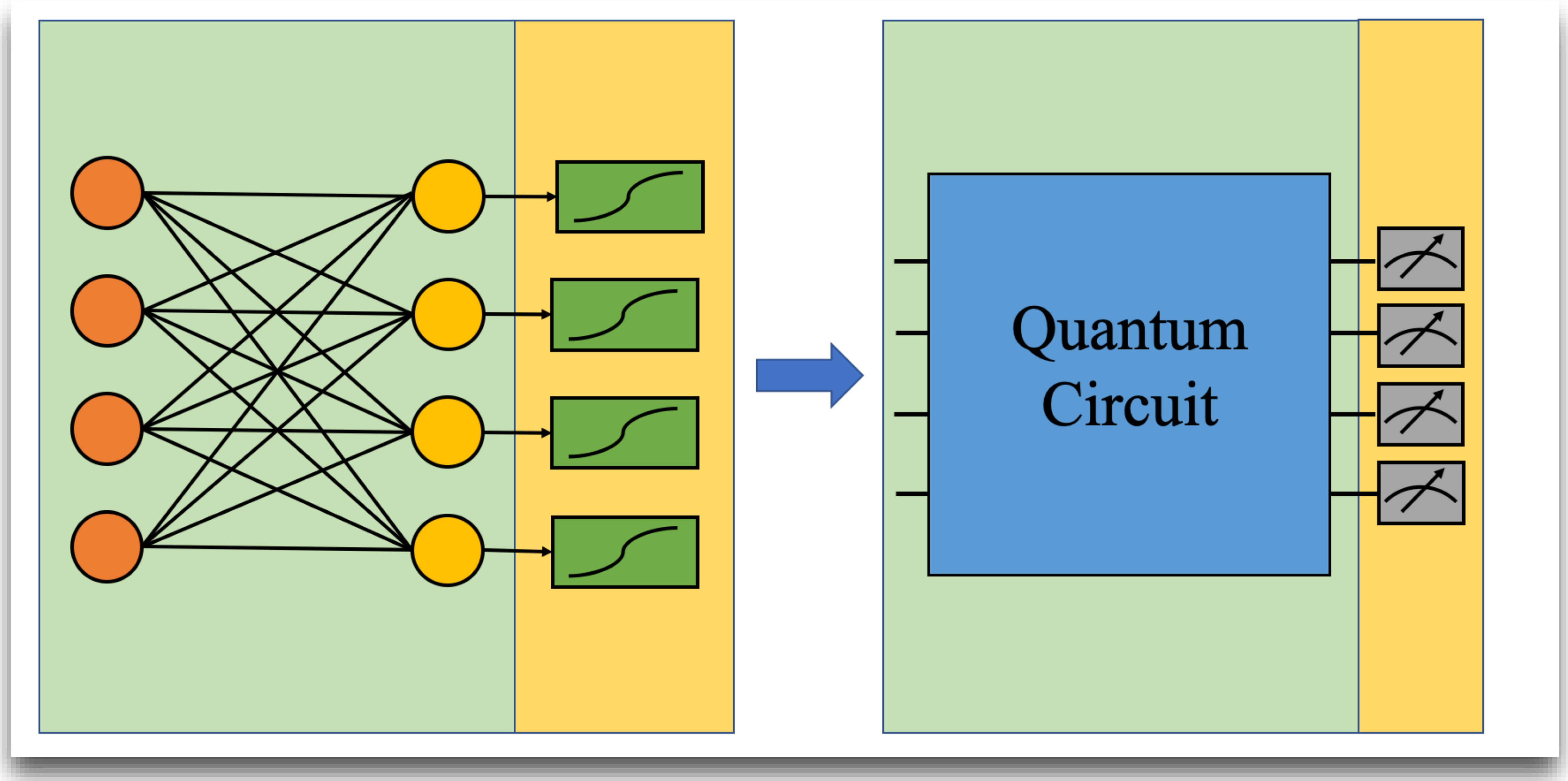
- The output of conventional quantum models are Fourier series with regularly spaced fixed frequencies.
- Introducing trainable parameters into the feature map generator leads to models with trainable frequencies.
- We can expect practical improvements on problems where the optimal spectral decomposition is (a) non-regularly spaced or (b) has unknown spectral richness.
- This is often the case in reality!



Discussion

New insight

Hybrid quantum-classical networks are using a classical neural network to set the frequencies of a quantum neural network.



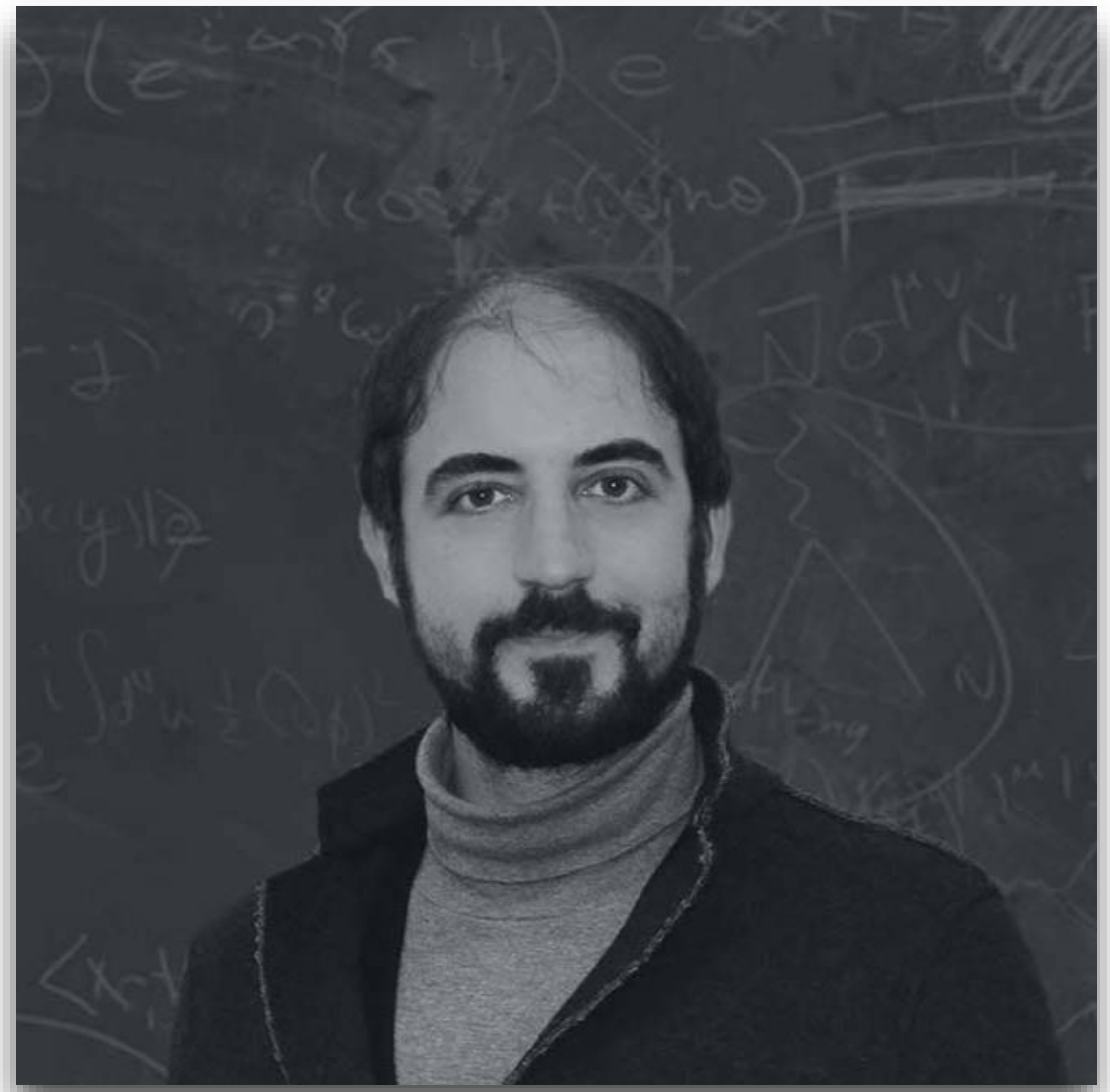
Mari, Andrea, et al. "Transfer learning in hybrid classical-quantum neural networks." *Quantum* 4 (2020): 340.
Jaderberg, Ben, et al. "Quantum self-supervised learning." Quantum Science and Technology 7.3 (2022): 035005.

Acknowledgements

PASQAL

IBM

Vincent Elfving



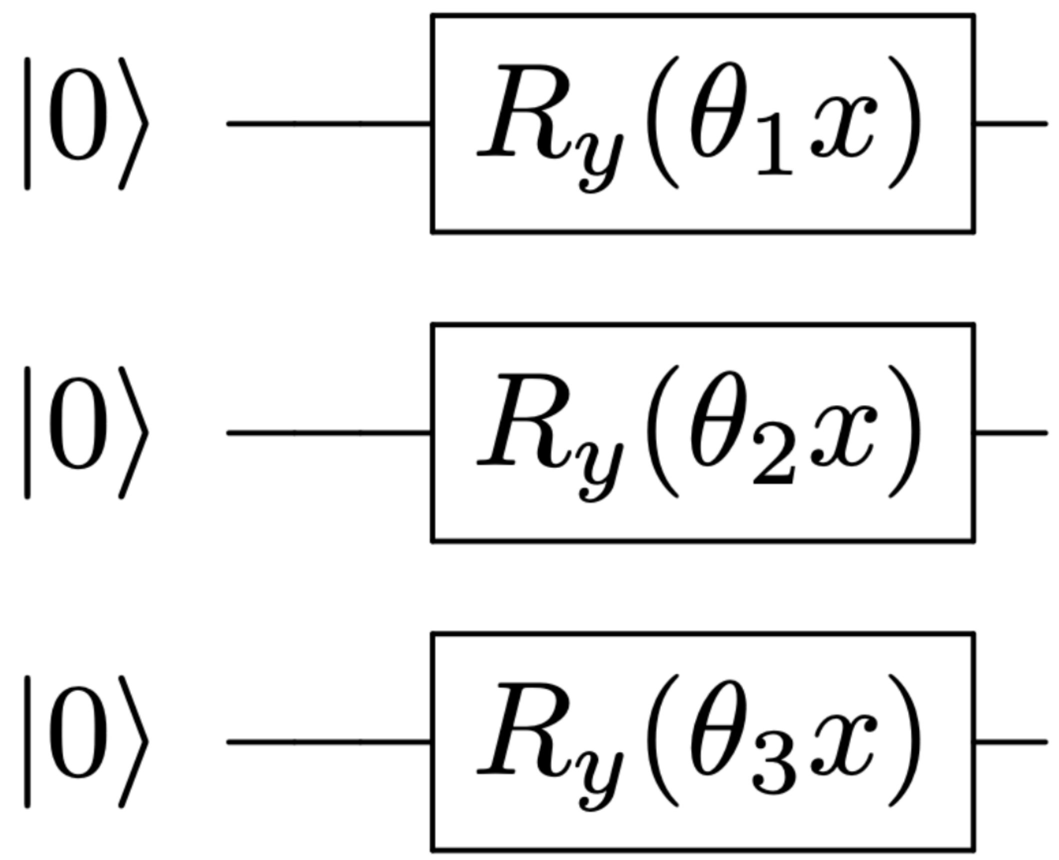
Andrea Gentile

@benjaderberg

benjamin.jaderberg@ibm.com

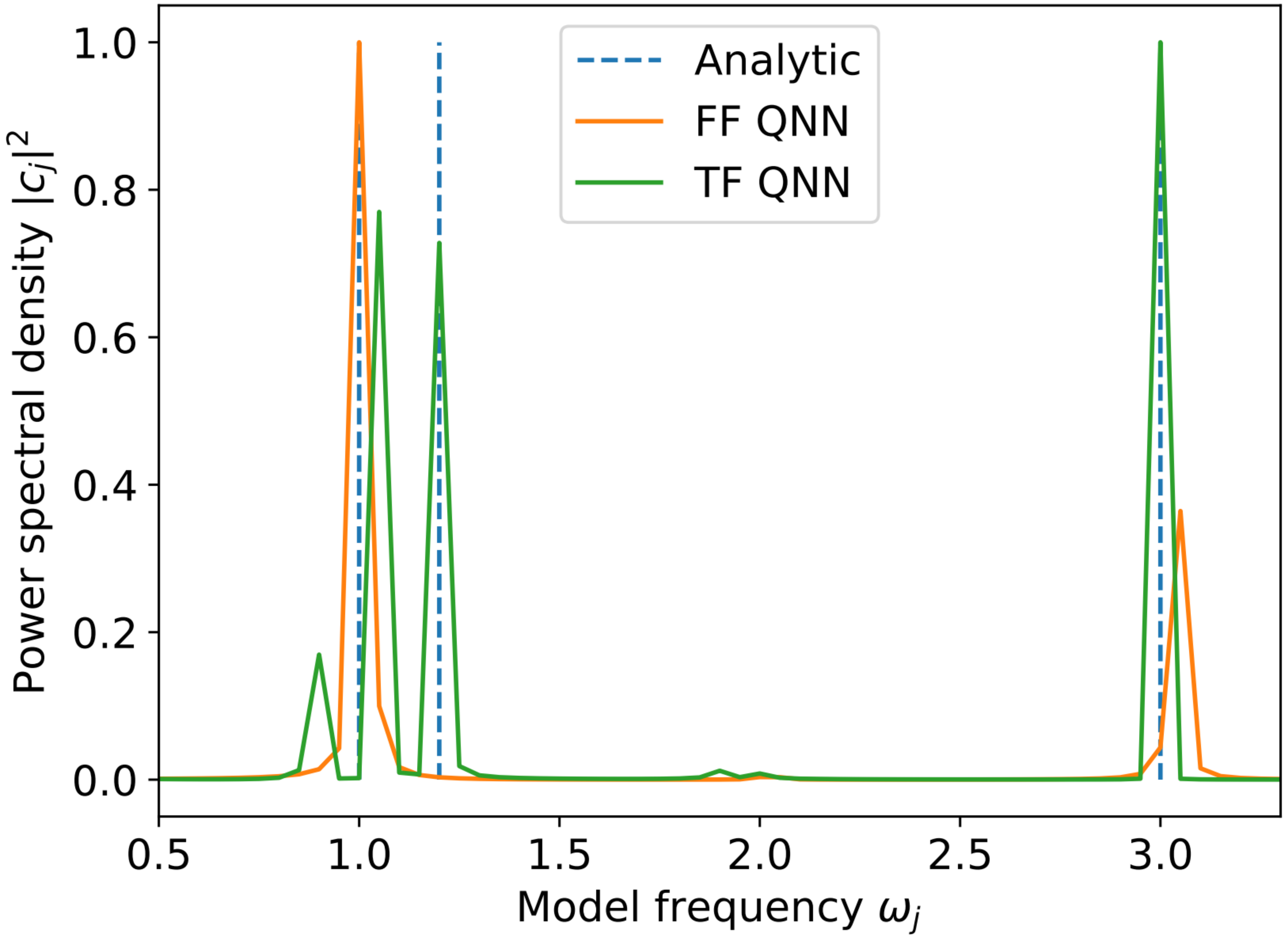
Proof-of-principle results

N=3 qubits
L=4 HEA layers



$\vec{\theta}_F = \{0.89, 1.05, 1.04\}$

$\Omega_d = \{1, 1.2, 3\}$



Practical benefit of TF models

Goal:

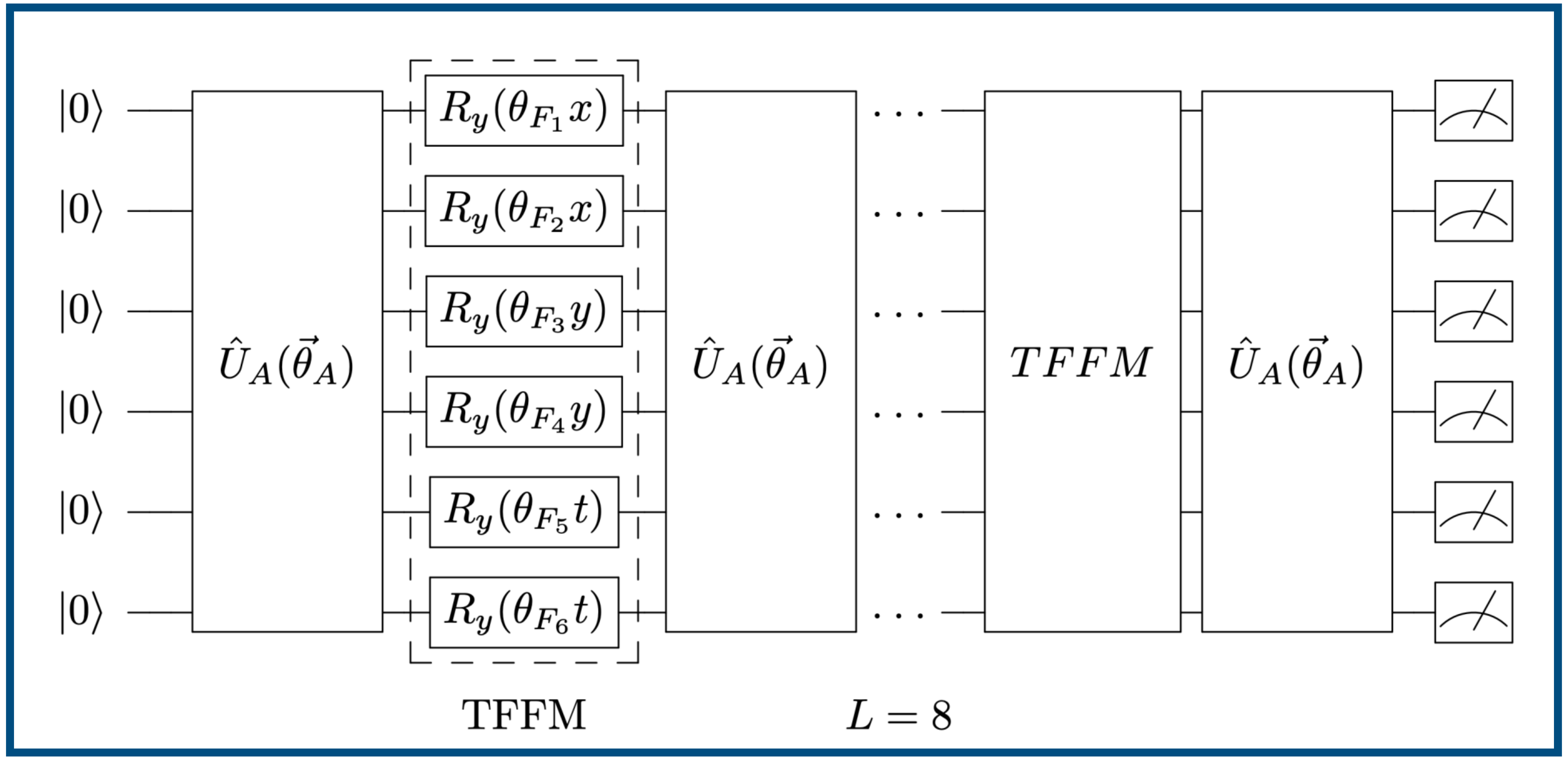
Predict u, v and p

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - \frac{1}{Re}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \frac{\partial p}{\partial x} = 0,$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} - \frac{1}{Re}\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) + \frac{\partial p}{\partial y} = 0,$$

+ 1% of data as boundary condition

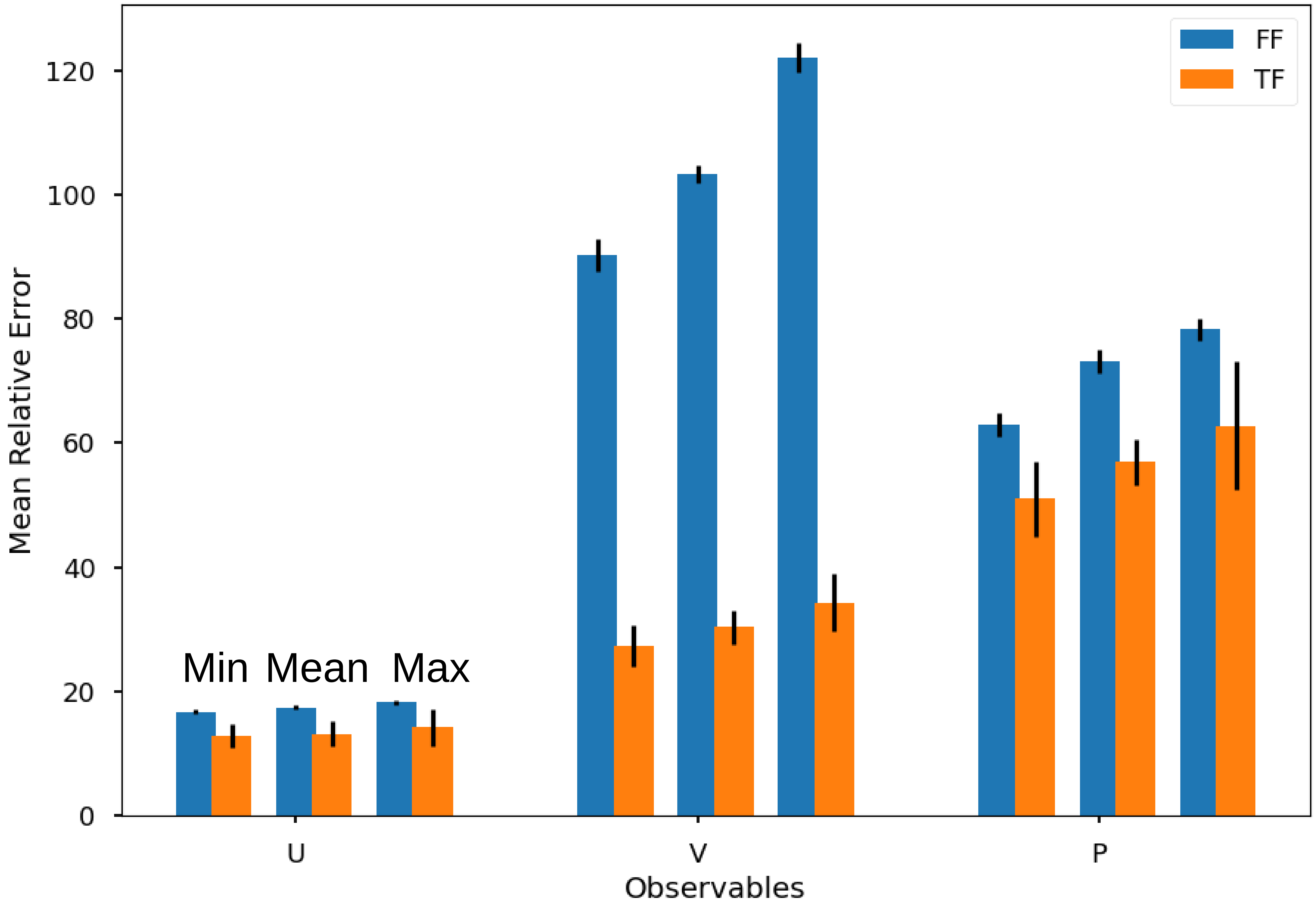
N=6, L=8, 192 parameters



600 batch size

5,000 training iterations

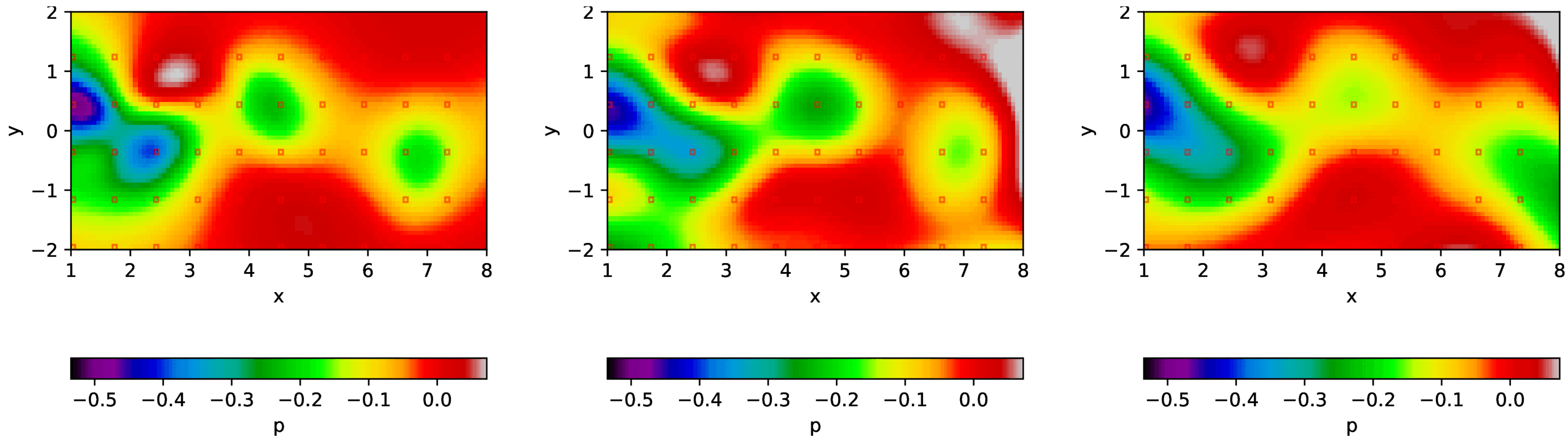
Practical benefit of TF models



Practical benefit of TF models

Overparameterised regime: L=64, 804 parameters

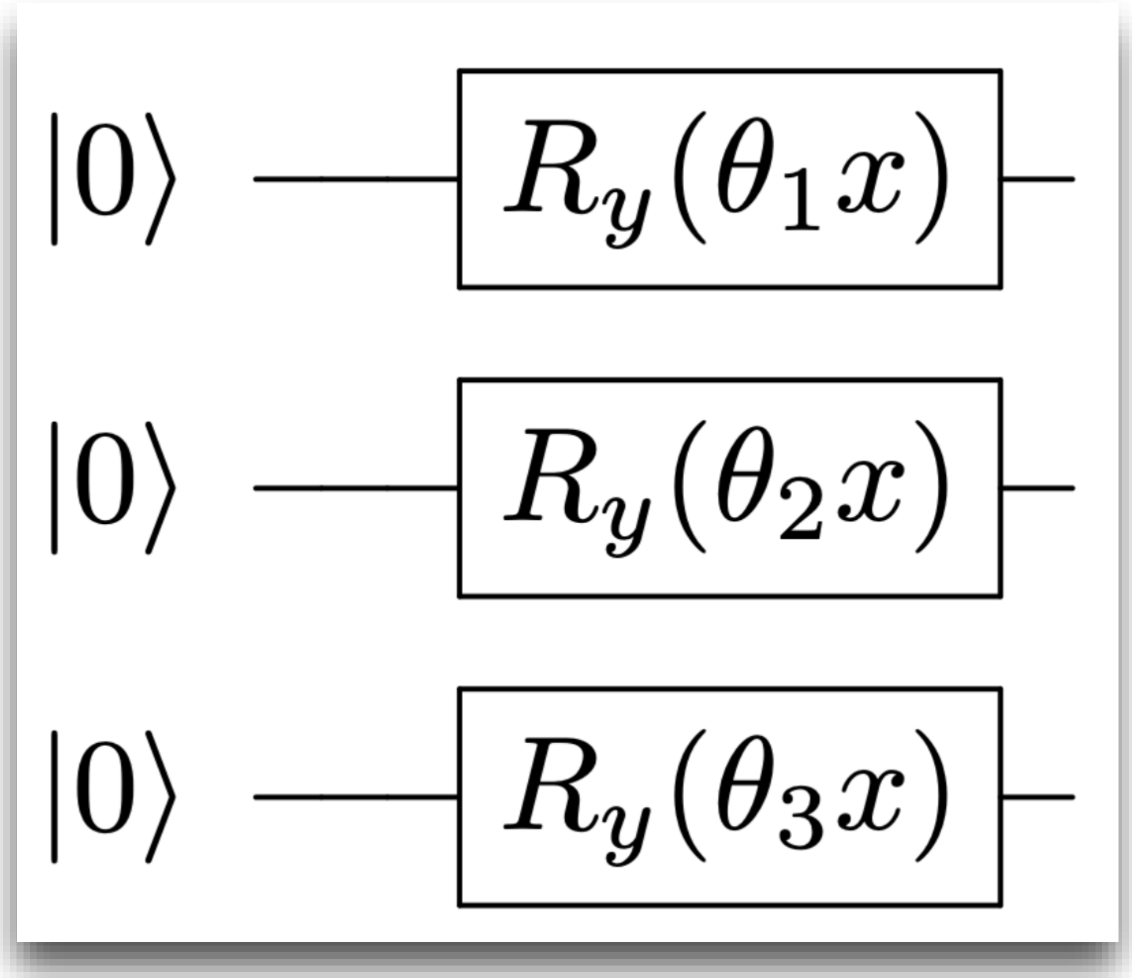
FEM **TF model** **FF model**



Discussion

TF models can fall-back to conventional FF models

There is a trivial initialisation of θ which allows us to prepare regularly spaced orthogonal frequencies as the starting point. If this is the optimal spectrum, we will remain close to it (not rigorous).



Simple $\theta_m = 1$

Tower $\theta_m = m$

Exponential $\theta_m = 2^{(m-1)}$