DPG SAMOP 2021

Understanding Variational Quantum Learning Models

JOHANNES JAKOB MEYER, FU BERLIN

@jj_xyz

10.1103/PhysRevA.103.032430

10.1103/PhysRevA.103.032430

The effect of data encoding on the expressive power of variational quantum machine learning models

Maria Schuld,[1] Ryan Sweke,[2] and Johannes Jakob Meyer[2]

[1] *Xanadu, Toronto, ON, M5G 2C8, Canada*

[2] *Dahlem Center for Complex Quantum Systems, Freie Universitt Berlin, 14195 Berlin, Germany*

(Dated: August 21, 2020)

10.1103/PhysRevA.103.032430

The effect of data encoding on the expressive power of variational quantum machine learning models

Maria Schuld,[1] Ryan Sweke,[2] and Johannes Jakob Meyer[2]

[1] *Xanadu, Toronto, ON, M5G 2C8, Canada*
[2] *Dahlem Center for Complex Quantum Systems, Freie Universitt Berlin, 14195 Berlin, Germany*

(Dated: August 21, 2020)

Maria Schuld
Xanadu

Ryan Sweke
FU Berlin

10.1103/PhysRevA.103.032430

Disclaimer: After uploading to the arxiv we were notified of work by Vidal and Theis[1] that has significant overlap with ours.

[1]Francisco Javier Gil Vidal and Dirk Oliver Theis. "Input Redundancy for Parameterized Quantum Circuits." *Frontiers in Physics* 8 (2020): 297.

Maria Schuld
Xanadu

Ryan Sweke
FU Berlin

Key Motivation

Key Motivation

Using quantum models for learning tasks is one of the key fields where NISQ devices are hoped to bring forth a quantum advantage

Key Motivation

Using quantum models for learning tasks is one of the key fields where NISQ devices are hoped to bring forth a quantum advantage

A lot of work has been done to understand the practical side, but little is known on the theory side

Key Motivation

Using quantum models for learning tasks is one of the key fields where NISQ devices are hoped to bring forth a quantum advantage

A lot of work has been done to understand the practical side, but little is known on the theory side

So we asked ourselves: **What functions can such models learn?**

Variational Quantum Learning Models

Variational Quantum Learning Models

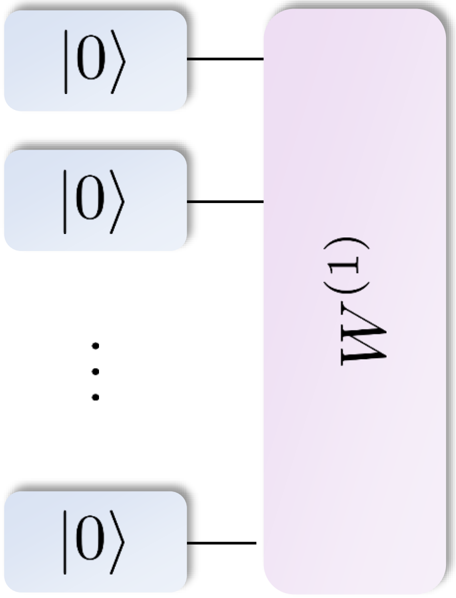
$|0\rangle$

$|0\rangle$

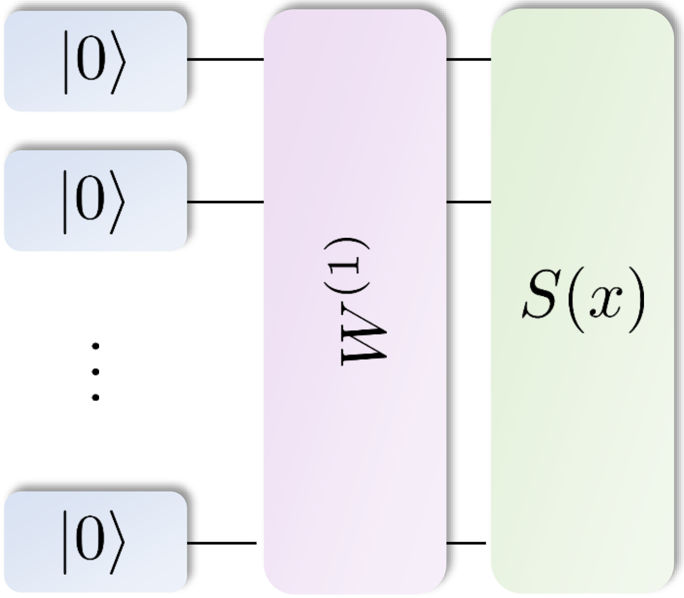
$\vdots$

$|0\rangle$

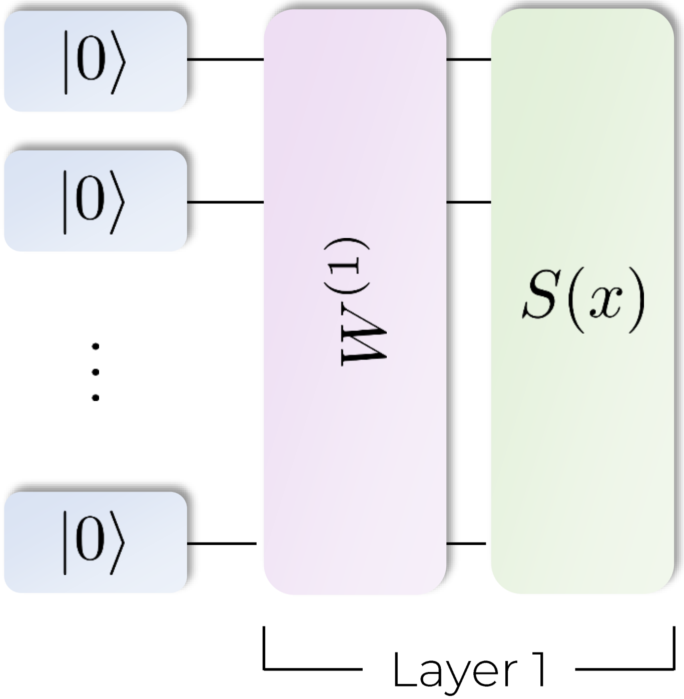
Variational Quantum Learning Models



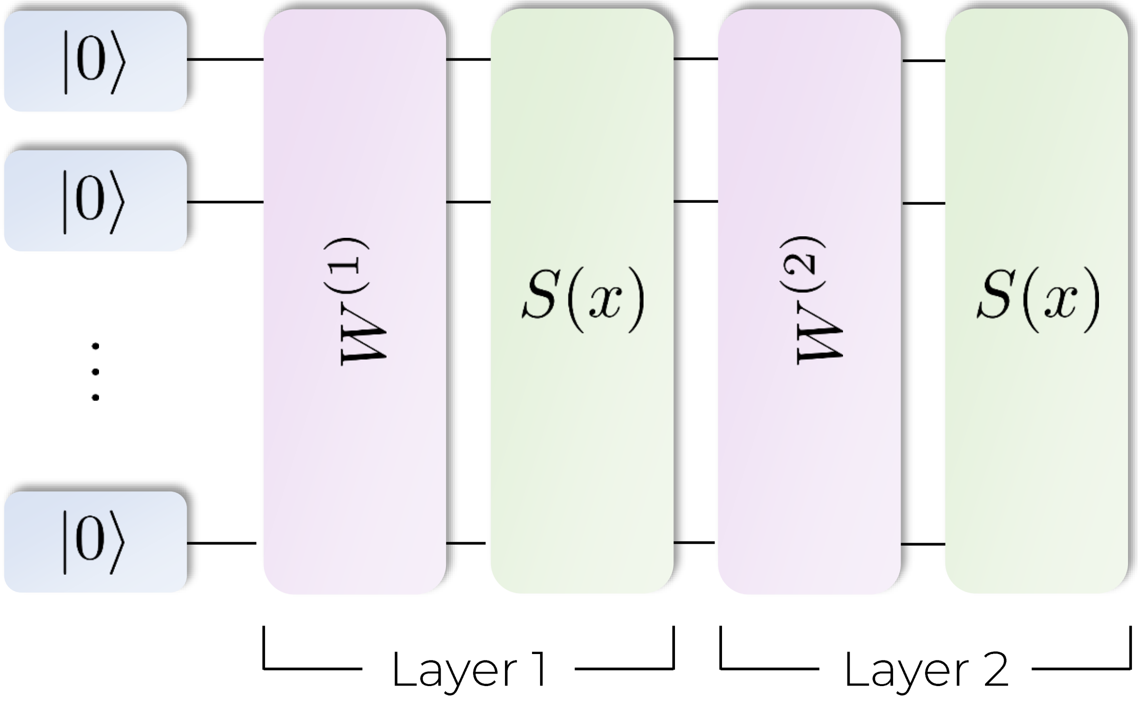
Variational Quantum Learning Models



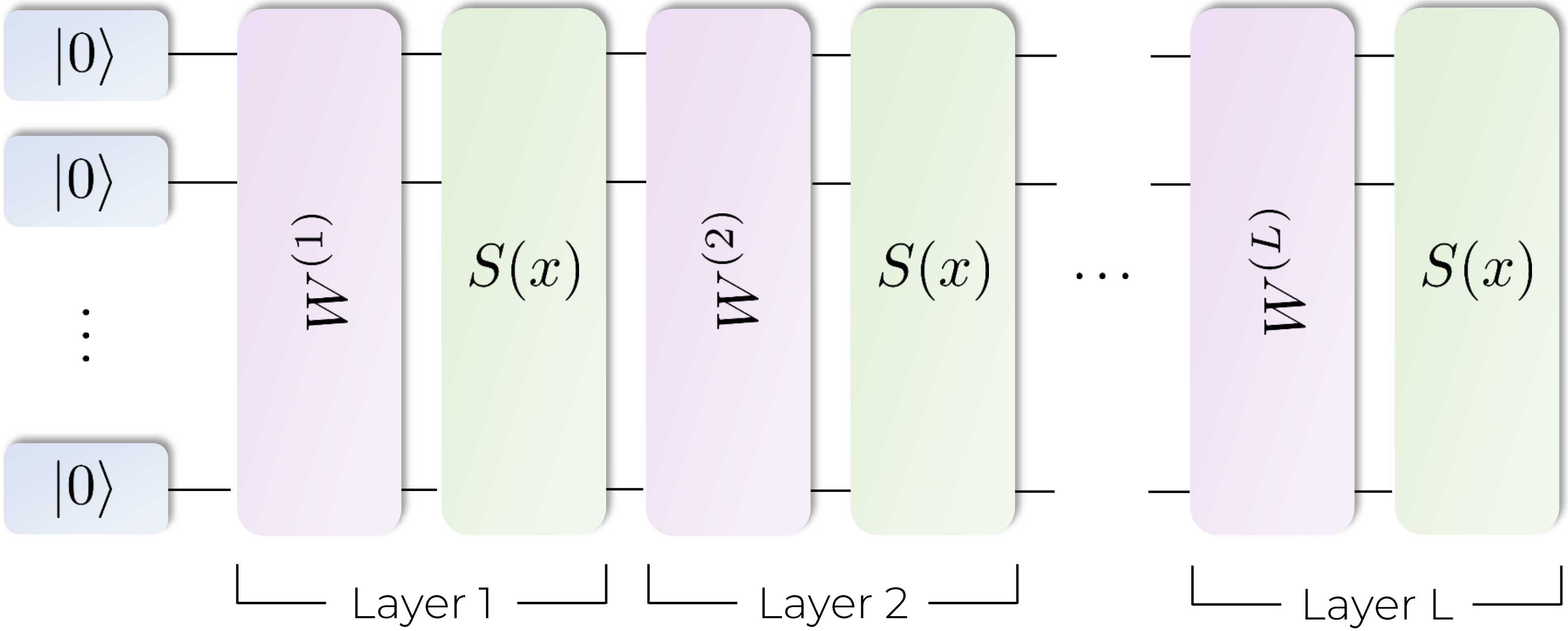
Variational Quantum Learning Models



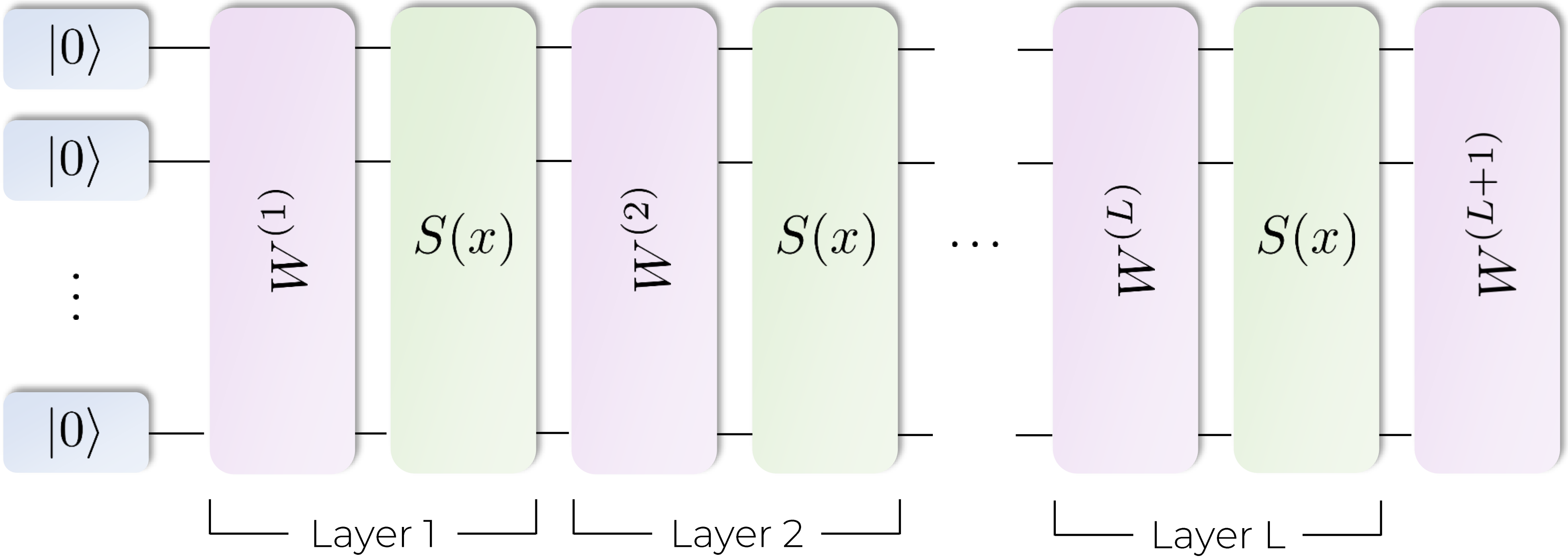
Variational Quantum Learning Models



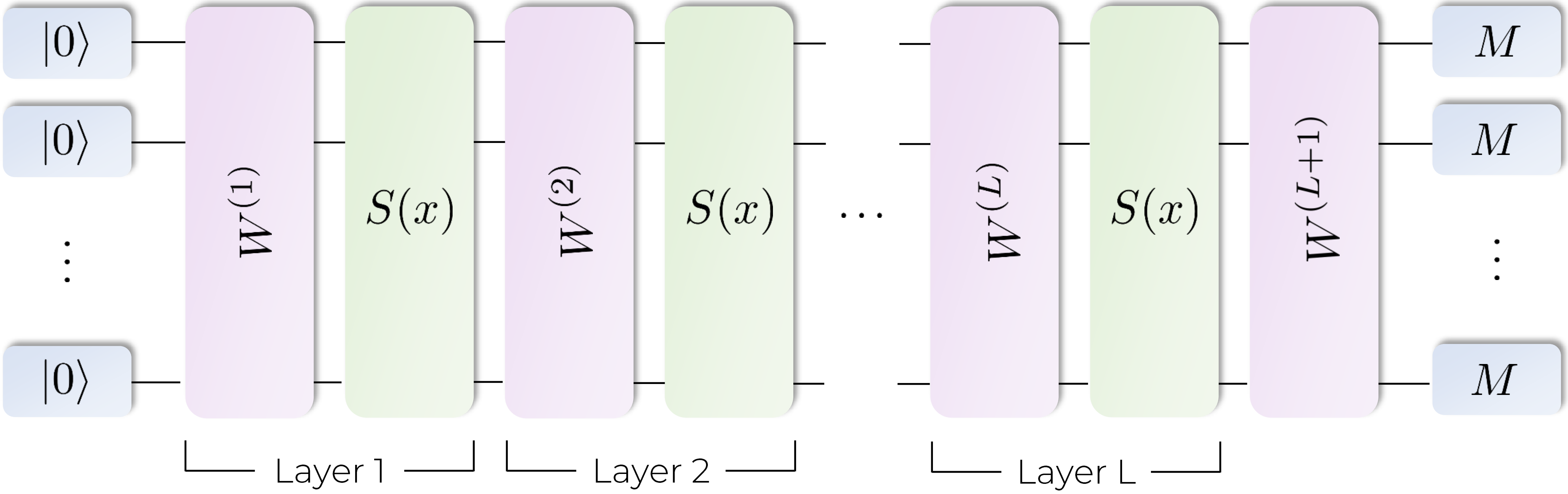
Variational Quantum Learning Models



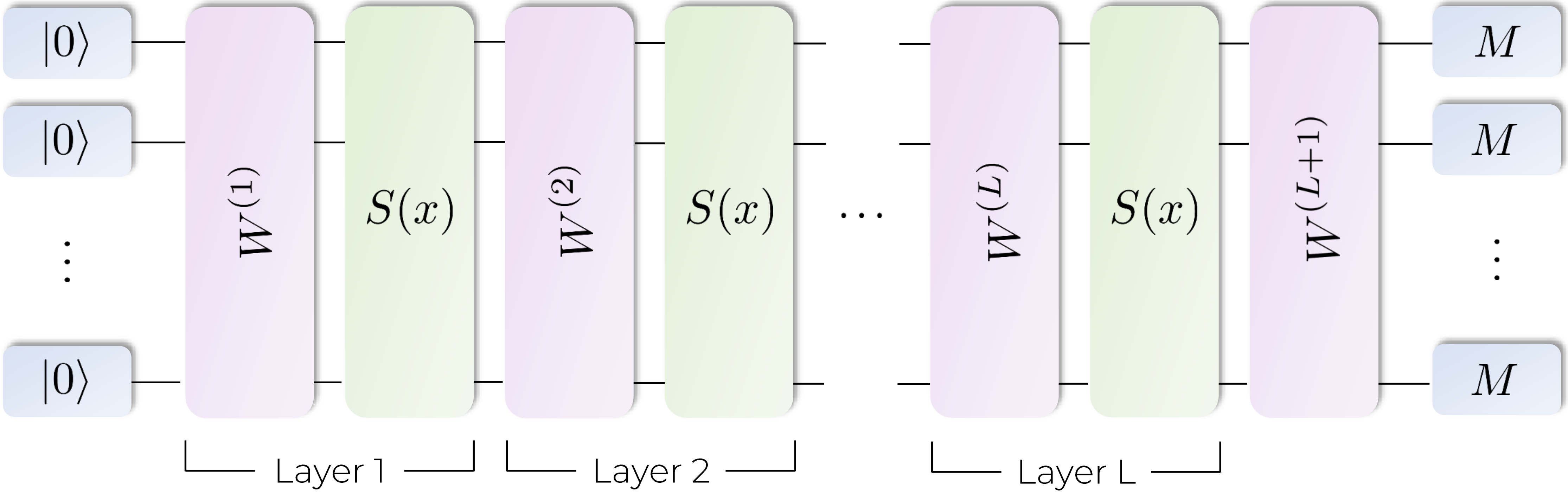
Variational Quantum Learning Models



Variational Quantum Learning Models



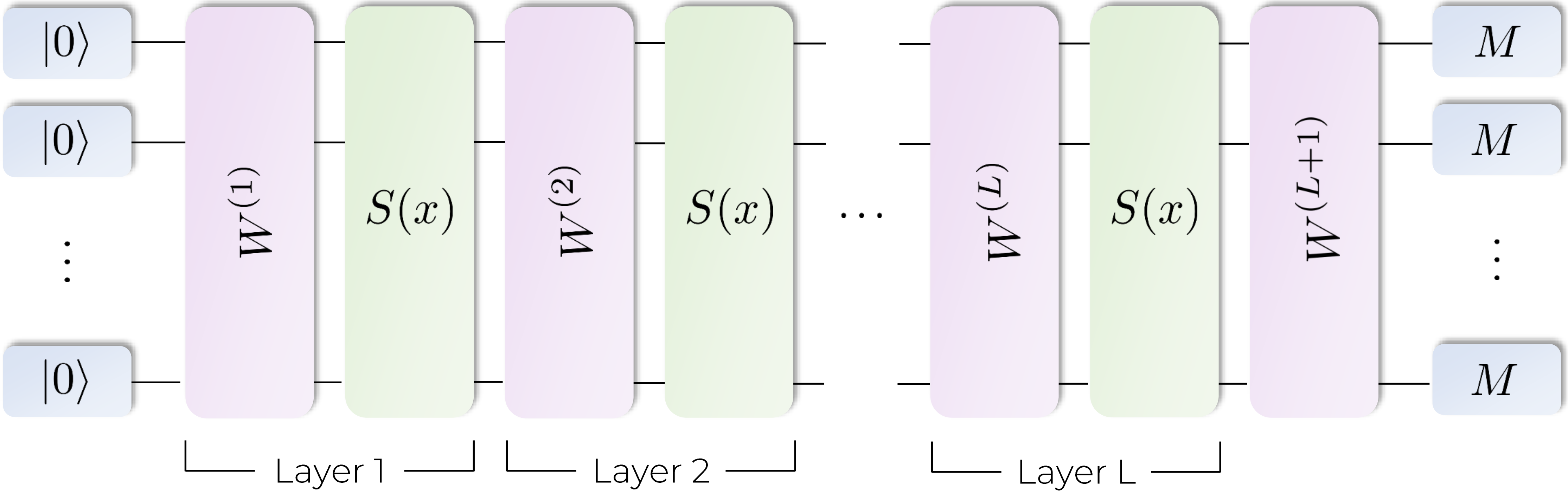
Variational Quantum Learning Models



$$f(x) = \langle M \rangle$$

Model Output

Variational Quantum Learning Models



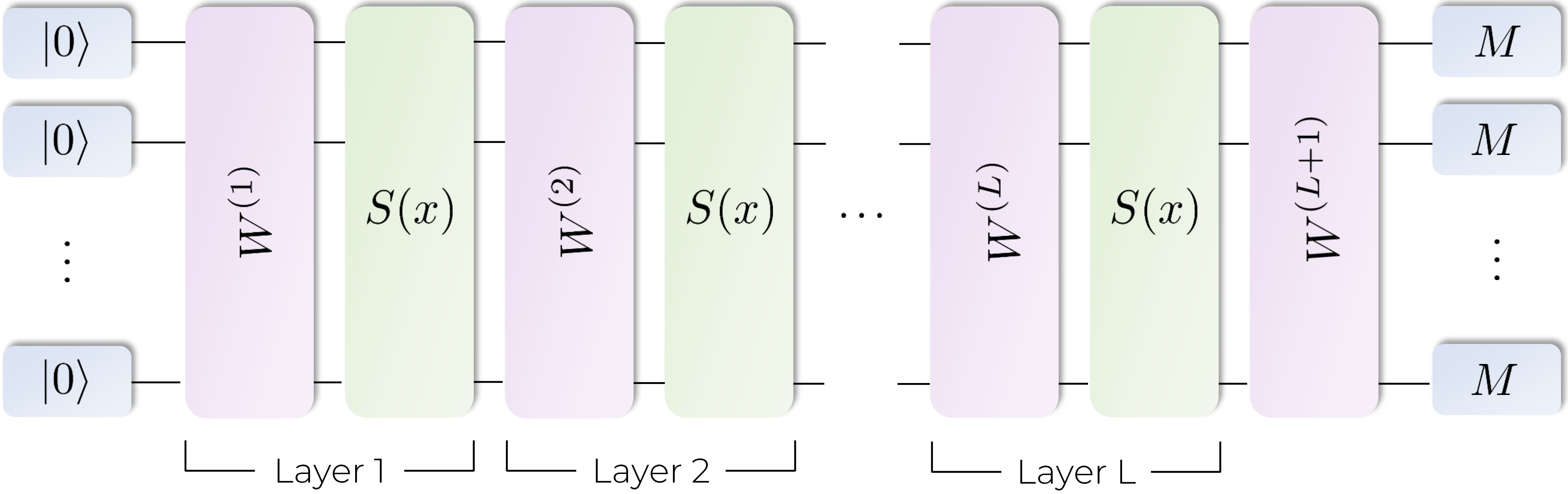
$$f(x) = \langle M \rangle$$

Model Output

$$S(x) = e^{-ixH}$$

Hamiltonian Evolution

Variational Quantum Learning Models



$$f(x) = \langle M \rangle$$

Model Output

$$S(x) = e^{-ixH}$$

Hamiltonian Evolution

$$W^{(l)}(\boldsymbol{\theta})$$

Trainable Blocks

Output State of the Quantum Model

Output State of the Quantum Model

The eigenvalues of the generator determine the frequencies

$$H|\lambda\rangle = \lambda|\lambda\rangle$$

$$S(x)|\lambda\rangle = e^{-ix\lambda}|\lambda\rangle$$

Output State of the Quantum Model

The eigenvalues of the generator determine the frequencies

$$H|\lambda\rangle = \lambda|\lambda\rangle$$

$$S(x)|\lambda\rangle = e^{-ix\lambda}|\lambda\rangle$$

and the frequencies accumulate between layers:

Output State of the Quantum Model

The eigenvalues of the generator determine the frequencies

$$H|\lambda\rangle = \lambda|\lambda\rangle$$

$$S(x)|\lambda\rangle = e^{-ix\lambda}|\lambda\rangle$$

and the frequencies accumulate between layers:

$$|\psi\rangle = \sum_{\lambda} \psi_{\lambda}|\lambda\rangle$$

Output State of the Quantum Model

The eigenvalues of the generator determine the frequencies

$$H|\lambda\rangle = \lambda|\lambda\rangle$$
$$S(x)|\lambda\rangle = e^{-ix\lambda}|\lambda\rangle$$

and the frequencies accumulate between layers:

$$|\psi\rangle = \sum_{\lambda} \psi_{\lambda}|\lambda\rangle$$
$$S(x)|\psi\rangle = \sum_{\lambda} \psi_{\lambda} e^{-ix\lambda}|\lambda\rangle$$

Output State of the Quantum Model

The eigenvalues of the generator determine the frequencies

$$H|\lambda\rangle = \lambda|\lambda\rangle$$

$$S(x)|\lambda\rangle = e^{-ix\lambda}|\lambda\rangle$$

and the frequencies accumulate between layers:

$$|\psi\rangle = \sum_{\lambda} \psi_{\lambda}|\lambda\rangle$$

$$S(x)|\psi\rangle = \sum_{\lambda} \psi_{\lambda} e^{-ix\lambda}|\lambda\rangle$$

$$WS(x)|\psi\rangle = \sum_{\lambda'\lambda} W_{\lambda'\lambda}\psi_{\lambda} e^{-ix\lambda}|\lambda'\rangle$$

Output State of the Quantum Model

The eigenvalues of the generator determine the frequencies

$$H|\lambda\rangle = \lambda|\lambda\rangle$$
$$S(x)|\lambda\rangle = e^{-ix\lambda}|\lambda\rangle$$

and the frequencies accumulate between layers:

$$|\psi\rangle = \sum_{\lambda} \psi_\lambda |\lambda\rangle$$
$$S(x)|\psi\rangle = \sum_{\lambda} \psi_\lambda e^{-ix\lambda}|\lambda\rangle$$
$$WS(x)|\psi\rangle = \sum_{\lambda'\lambda} W_{\lambda'\lambda}\psi_\lambda e^{-ix\lambda}|\lambda'\rangle$$
$$S(x)WS(x)|\psi\rangle = \sum_{\lambda'\lambda} W_{\lambda'\lambda}\psi_\lambda e^{-ix(\lambda+\lambda')}|\lambda'\rangle$$

Output State of the Quantum Model

The eigenvalues of the generator determine the frequencies

$$H|\lambda\rangle = \lambda|\lambda\rangle$$
$$S(x)|\lambda\rangle = e^{-ix\lambda}|\lambda\rangle$$

and the frequencies accumulate between layers:

$$|\psi\rangle = \sum_{\lambda} \psi_\lambda |\lambda\rangle$$
$$S(x)|\psi\rangle = \sum_{\lambda} \psi_\lambda e^{-ix\lambda}|\lambda\rangle$$
$$WS(x)|\psi\rangle = \sum_{\lambda'\lambda} W_{\lambda'\lambda}\psi_\lambda e^{-ix\lambda}|\lambda'\rangle$$
$$S(x)WS(x)|\psi\rangle = \sum_{\lambda'\lambda} W_{\lambda'\lambda}\psi_\lambda e^{-ix(\lambda+\lambda')}|\lambda'\rangle$$

The output state contains all possible sums of frequencies

Output of the Quantum Model

Output of the Quantum Model

Output is expectation value and therefore contains a complex conjugation

$$f_{M,\boldsymbol{\theta}}(x) = \langle \psi_{\boldsymbol{\theta}}(x) | M | \psi_{\boldsymbol{\theta}}(x) \rangle = \sum_{\omega \in \Omega} c_{\omega}(M, \boldsymbol{\theta}) e^{i\omega x}$$

Output of the Quantum Model

Output is expectation value and therefore contains a complex conjugation

$$f_{M,\boldsymbol{\theta}}(x) = \langle \psi_{\boldsymbol{\theta}}(x) | M | \psi_{\boldsymbol{\theta}}(x) \rangle = \sum_{\omega \in \Omega} c_\omega(M, \boldsymbol{\theta}) e^{i\omega x}$$

For one layer of encoding

$$\Omega = \{\lambda_j - \lambda_k \,|\, \lambda_j, \lambda_k \in \mathrm{spec}(H)\}$$

Output of the Quantum Model

Output is expectation value and therefore contains a complex conjugation

$$f_{M,\boldsymbol{\theta}}(x) = \langle \psi_{\boldsymbol{\theta}}(x) | M | \psi_{\boldsymbol{\theta}}(x) \rangle = \sum_{\omega \in \Omega} c_\omega(M, \boldsymbol{\theta}) e^{i\omega x}$$

For L layers of encoding

$$\Omega = \{\lambda_{j_1} + \cdots + \lambda_{j_L} - \lambda_{k_1} - \cdots - \lambda_{k_L} \,|\, \lambda_{j_l}, \lambda_{k_l} \in \mathrm{spec}(H)\}$$

Output of the Quantum Model

Output is expectation value and therefore contains a complex conjugation

$$f_{M,\boldsymbol{\theta}}(x) = \langle \psi_{\boldsymbol{\theta}}(x) | M | \psi_{\boldsymbol{\theta}}(x) \rangle = \sum_{\omega \in \Omega} c_\omega(M, \boldsymbol{\theta}) e^{i\omega x}$$

For L layers of encoding

$$\Omega = \{\lambda_{j_1} + \cdots + \lambda_{j_L} - \lambda_{k_1} - \cdots - \lambda_{k_L} \,|\, \lambda_{j_l}, \lambda_{k_l} \in \mathrm{spec}(H)\}$$

The accessible spectrum consists of all **sums of differences** of eigenvalues of the generator of the data encoding

Take-Home Message

Take-Home Message

Quantum learning models output Fourier series, repeated data encoding gives access to higher frequencies

Pauli-Encodings

Pauli-Encodings

Pauli rotations are the most popular encoding strategy, e.g.

$$S(x) = R_Z(x) = e^{-ixZ/2}$$

Pauli-Encodings

Pauli rotations are the most popular encoding strategy, e.g.

$$S(x) = R_Z(x) = e^{-ixZ/2}$$

Pauli rotations give an integer spectrum

Pauli-Encodings

Pauli rotations are the most popular encoding strategy, e.g.

$$S(x) = R_Z(x) = e^{-ixZ/2}$$

Pauli rotations give an integer spectrum

The number of available frequencies grows linearly in depth and width

Pauli-Encodings

Pauli rotations are the most popular encoding strategy, e.g.

$$S(x) = R_Z(x) = e^{-ixZ/2}$$

Pauli rotations give an integer spectrum

The number of available frequencies grows linearly in depth and width

But for general encodings the dependence can be exponential:

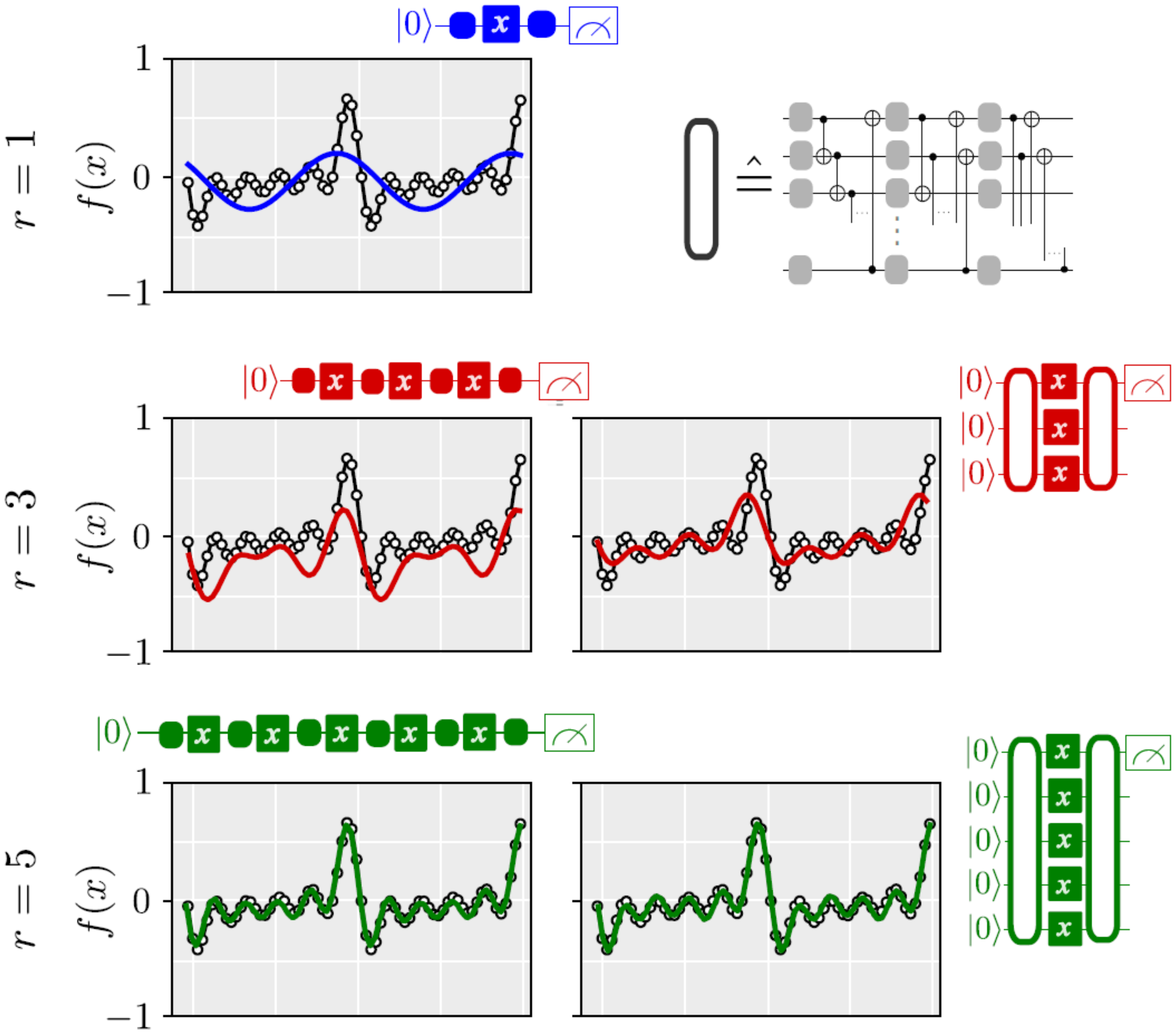
$$\#(\text{frequencies}) \leq \frac{d^{2L}}{2} - 1$$

d System dimension

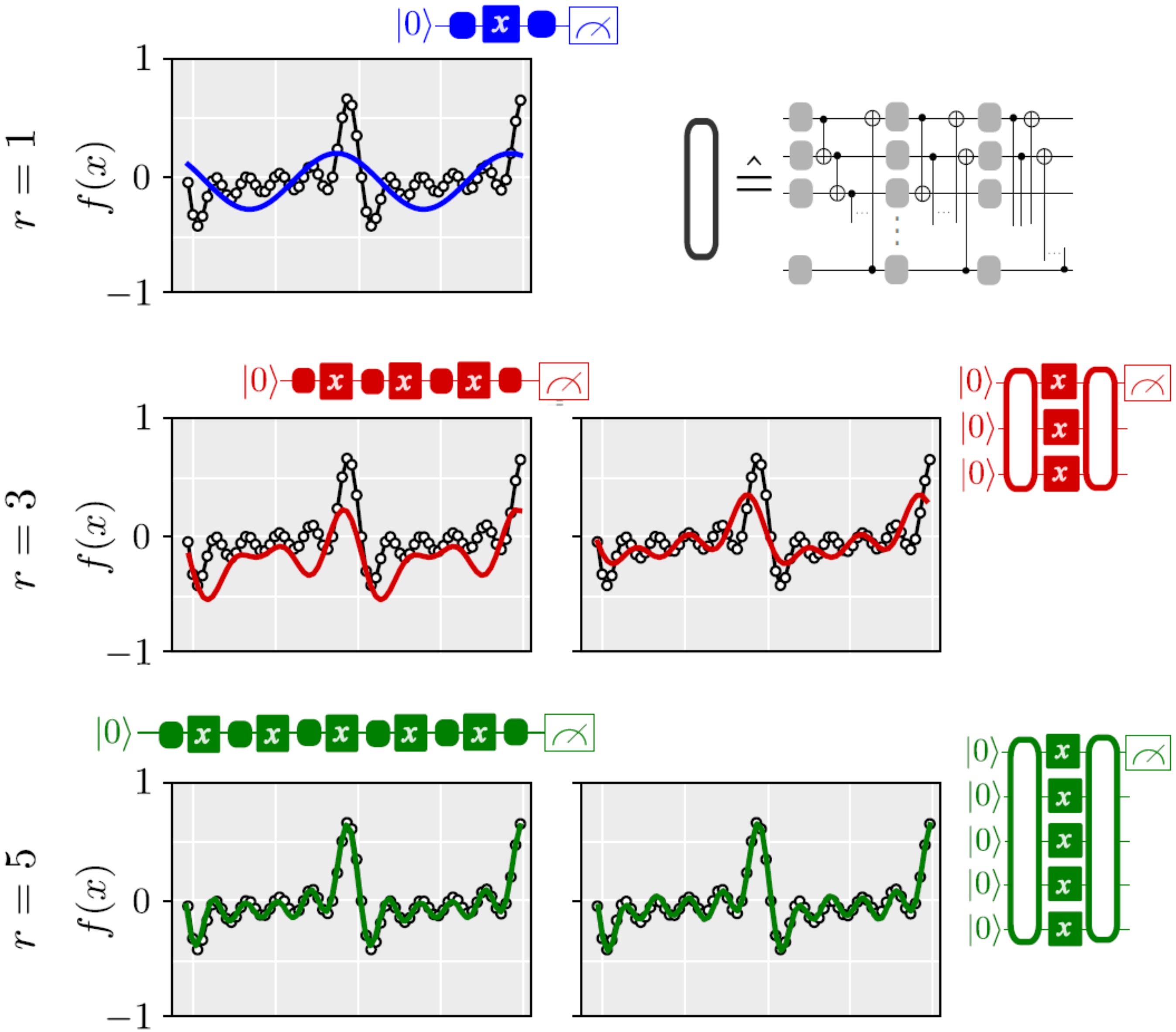
L Number of layers

Consequences for Learning Tasks

Consequences for Learning Tasks



Consequences for Learning Tasks



You can reproduce all figures from the paper at home!

arXiv:2106.03880

arXiv:2106.03880

Encoding-dependent generalization bounds for parametrized quantum circuits

Matthias C. Caro,[1,2] Elies Gil-Fuster,[3] Johannes Jakob Meyer,[3,4] Jens Eisert,[3,5] and Ryan Sweke[3]

[1]*Department of Mathematics, Technical University of Munich, 85748 Garching, Germany*
[2]*Munich Center for Quantum Science and Technology (MCQST), 80799 Munich, Germany*
[3]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[4]*QMATH, Department of Mathematical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark*
[5]*Helmholtz-Zentrum Berlin für Materialien und Energie, 14109 Berlin, Germany*

(Dated: June 9, 2021)

arXiv:2106.03880

Encoding-dependent generalization bounds for parametrized quantum circuits

Matthias C. Caro,[1,2] Elies Gil-Fuster,[3] Johannes Jakob Meyer,[3,4] Jens Eisert,[3,5] and Ryan Sweke[3]

[1]*Department of Mathematics, Technical University of Munich, 85748 Garching, Germany*
[2]*Munich Center for Quantum Science and Technology (MCQST), 80799 Munich, Germany*
[3]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[4]*QMATH, Department of Mathematical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark*
[5]*Helmholtz-Zentrum Berlin für Materialien und Energie, 14109 Berlin, Germany*

(Dated: June 9, 2021)

Matthias Caro
TU Munich

Elies Gil-Fuster
FU Berlin

Jens Eisert
FU Berlin

Ryan Sweke
FU Berlin

Generalization Bounds

Generalization Bounds

We want to quantify how good a trained learning model performs

Generalization Bounds

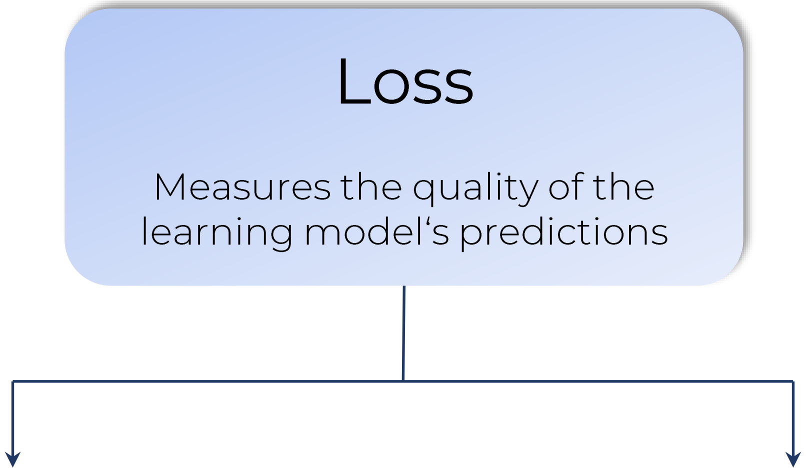
We want to quantify how good a trained learning model performs

Loss

Measures the quality of the learning model‘s predictions

Generalization Bounds

We want to quantify how good a trained learning model performs



Generalization Bounds

We want to quantify how good a trained learning model performs

Loss

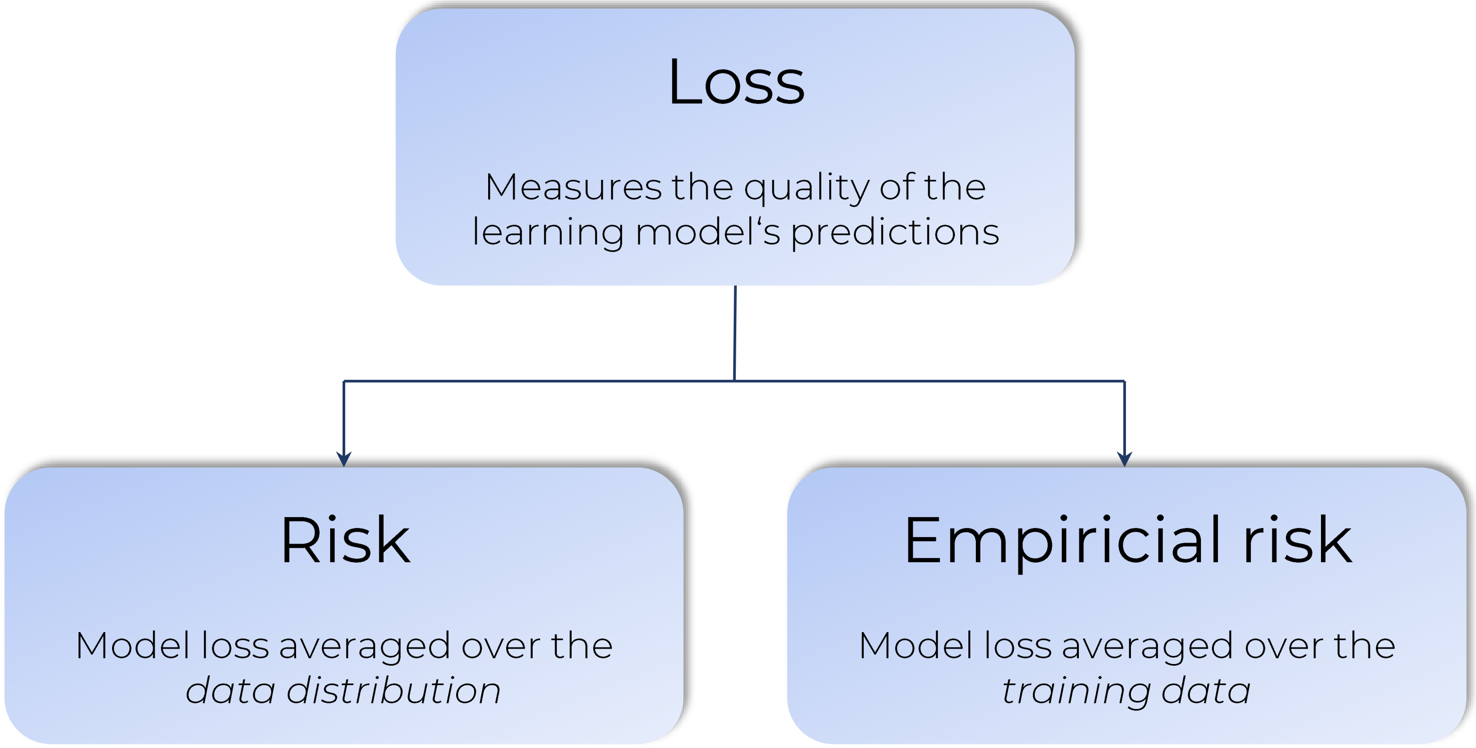
Measures the quality of the learning model‘s predictions

Risk

Model loss averaged over the *data distribution*

Generalization Bounds

We want to quantify how good a trained learning model performs



Generalization Bounds

We want to quantify how good a trained learning model performs

Loss

Measures the quality of the learning model's predictions

Risk

Model loss averaged over the *data distribution*

–

Empiricial risk

Model loss averaged over the *training data*

Generalization Bounds

We want to quantify how good a trained learning model performs

Loss

Measures the quality of the learning model's predictions

Risk		**Empiricial risk**			**Generalization bound**
Model loss averaged over the *data distribution*	$-$	Model loss averaged over the *training data*	$\vert$	$\leq$	Holds with high probability

Deriving a Generalization Bound

Deriving a Generalization Bound

Generalization bounds reflect the „complexity" of the model

Deriving a Generalization Bound

Generalization bounds reflect the „complexity“ of the model

The more different functions the model can learn, the more training data we need

Deriving a Generalization Bound

Generalization bounds reflect the „complexity“ of the model

The more different functions the model can learn, the more training data we need

Can be made explicit via *covering numbers*:
How many points do I need such that any possible output of the learning model is ϵ-close to at least one point?

Deriving a Generalization Bound

Generalization bounds reflect the „complexity“ of the model

The more different functions the model can learn, the more training data we need

Can be made explicit via *covering numbers*:
How many points do I need such that any possible output of the learning model is ϵ-close to at least one point?

We exploit the Fourier representation to derive covering numbers that depend explicitly on the data-encoding strategy

Results

Results

Models that use fixed Hamiltonians have a generalization bound that scales polynomially in the number of gates, guaranteeing efficient learning

Results

Models that use fixed Hamiltonians have a generalization bound that scales polynomially in the number of gates, guaranteeing efficient learning

Models using arbitrary Hamiltonians can have a generalization bound that scales exponentially

Take-Home Message

Take-Home Message

Quantum learning models with fixed Hamiltonians can learn efficiently

Thank you for your attention!

Paper Data-encoding and Fourier series	Paper Generalization bounds	Slides	Demo