Hybrid Recurrent Architectures for Quantum-Classical NLP

Stephen Clark

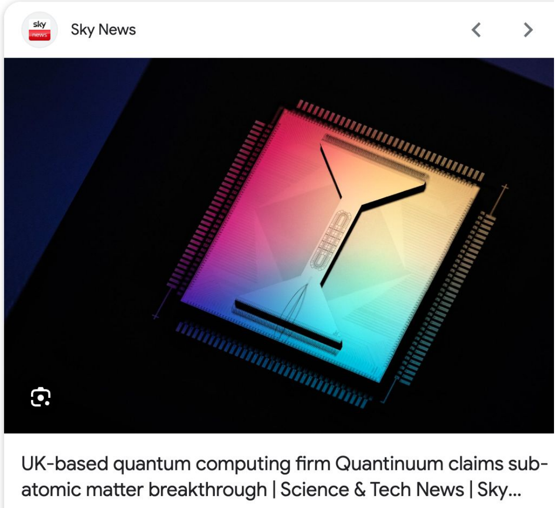
Global Software Technology Summit

University of Edinburgh

2 June 2023

QUANTINUUM

Quantum Computing



QUANTINUUM

Talk Outline

- Introduction to quantum computing / quantum circuits
- Our hybrid quantum RNN architectures
- Sentiment analysis experiments (in simulation)

The State of a Qubit

$$|\psi\rangle$$

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \qquad \alpha, \beta \in \mathbb{C} \qquad |\psi\rangle \in \mathbb{C}^2$$

superposition

QUANTINUUM

The State of a Qubit

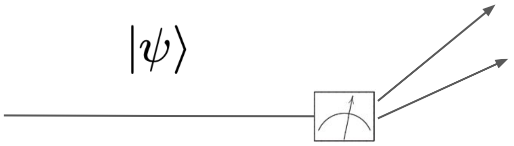
$$|\psi\rangle$$

amplitudes

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \qquad \alpha, \beta \in \mathbb{C} \qquad \boxed{|\alpha|^2 + |\beta|^2 = 1}$$

superposition

Measuring a Qubit



$|0\rangle$ with probability $|\alpha|^2$

$|1\rangle$ with probability $|\beta|^2$

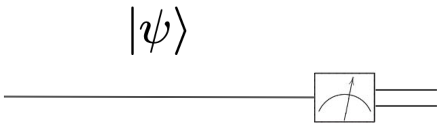
$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

superposition

$$|\alpha|^2 + |\beta|^2 = 1$$

QUANTINUUM

Measuring a Qubit - Scalar Output



$$\cancel{|\psi\rangle = \alpha|0\rangle + \beta|1\rangle}$$

~~*superposition*~~

1 with probability $|\alpha|^2$

-1 with probability $|\beta|^2$

QUANTINUUM

Measuring a Qubit - Scalar Output

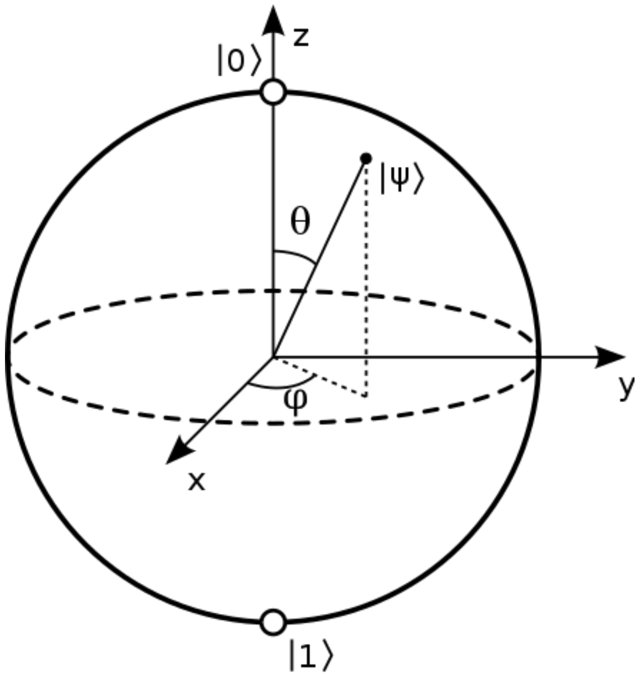
$|\psi\rangle$

$\langle M \rangle \in [-1, 1]$

$$\cancel{|\psi\rangle = \alpha|0\rangle + \beta|1\rangle}$$

~~*superposition*~~

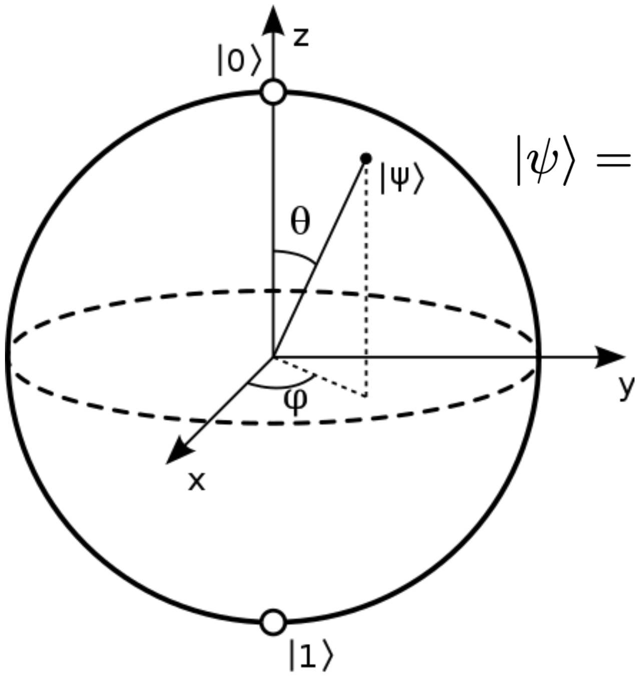
The Bloch Sphere Representation of a Qubit



$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

https://en.wikipedia.org/wiki/Bloch_sphere

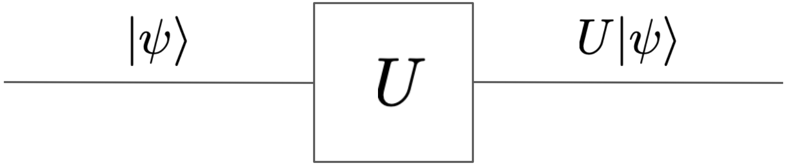
The Bloch Sphere Representation of a Qubit



$$|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle$$

https://en.wikipedia.org/wiki/Bloch_sphere

Unitary Transformations of a Qubit

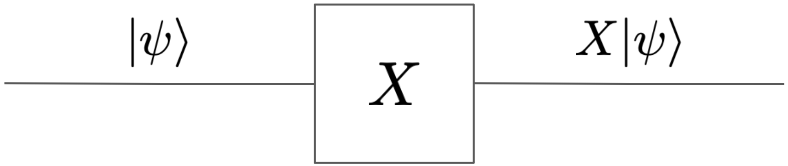


$$U : \alpha|0\rangle + \beta|1\rangle \mapsto \alpha'|0\rangle + \beta'|1\rangle$$

$$\boxed{|\alpha'|^2 + |\beta'|^2 = 1}$$

1-Qubit Quantum Gates

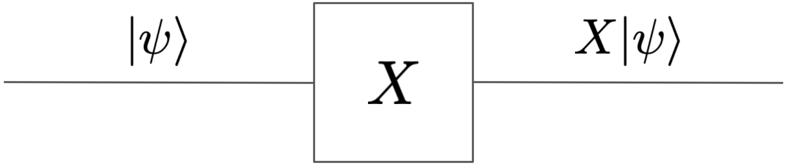
quantum Not gate



$$X : |0\rangle \mapsto |1\rangle$$
$$X : |1\rangle \mapsto |0\rangle$$

1-Qubit Quantum Gates

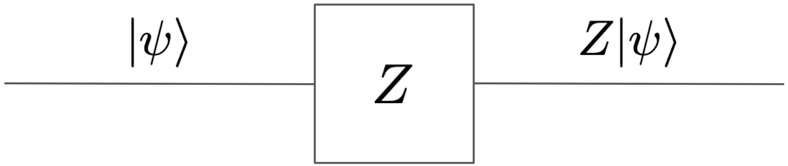
quantum Not gate *acts linearly*



$$X : \alpha|0\rangle + \beta|1\rangle \mapsto \alpha|1\rangle + \beta|0\rangle$$

1-Qubit Quantum Gates

Pauli Z Gate

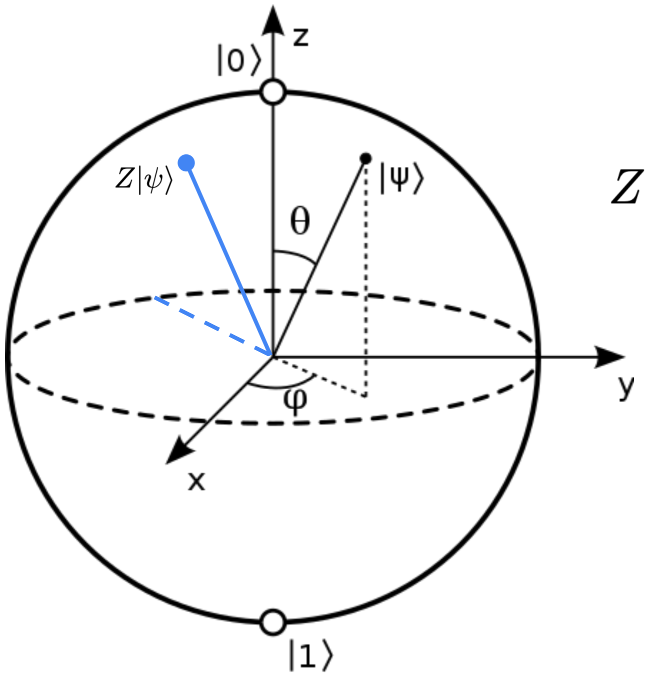


$$Z : \alpha|0\rangle + \beta|1\rangle \mapsto \alpha|0\rangle - \beta|1\rangle$$

QUANTINUUM

1-Qubit Quantum Gates

Pauli Z Gate *rotates about the Z axis*



$$Z : \alpha|0\rangle + \beta|1\rangle \mapsto \alpha|0\rangle - \beta|1\rangle$$

QUANTINUUM

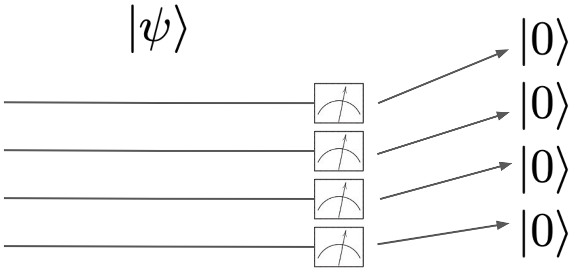
The State of Many Qubits

$$|\psi\rangle$$

$$|\psi\rangle \in \mathbb{C}^{2^4}$$

$$|\psi\rangle = \alpha_{0000}|0000\rangle + \alpha_{0001}|0001\rangle + \alpha_{0010}|0010\rangle + \ldots \alpha_{1111}|1111\rangle$$

Measuring Many Qubits



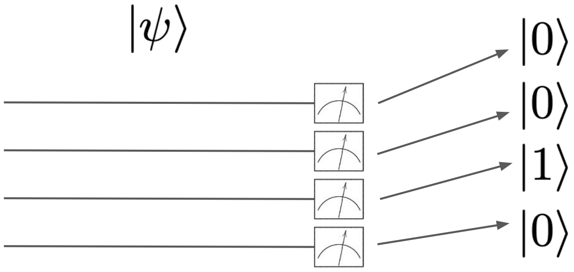
$$|\psi\rangle \in \mathbb{C}^{2^4}$$

$$|\psi\rangle = \alpha_{0000}\boxed{|0000\rangle} + \alpha_{0001}|0001\rangle + \alpha_{0010}|0010\rangle + \ldots \alpha_{1111}|1111\rangle$$

$$\boxed{|\alpha_{0000}|^2}$$

QUANTINUUM

Measuring Many Qubits



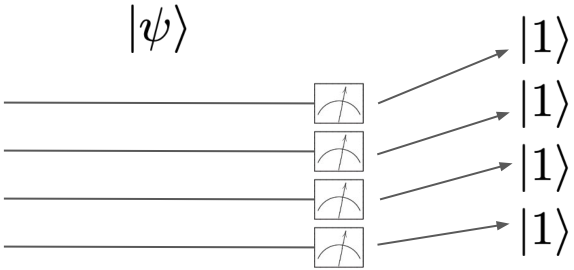
$$|\psi\rangle \in \mathbb{C}^{2^4}$$

$$|\psi\rangle = \alpha_{0000}|0000\rangle + \alpha_{0001}|0001\rangle + \alpha_{0010}\boxed{|0010\rangle} + \ldots \alpha_{1111}|1111\rangle$$

$$\boxed{|\alpha_{0010}|^2}$$

Measuring Many Qubits

$|\psi\rangle \in \mathbb{C}^{2^4}$



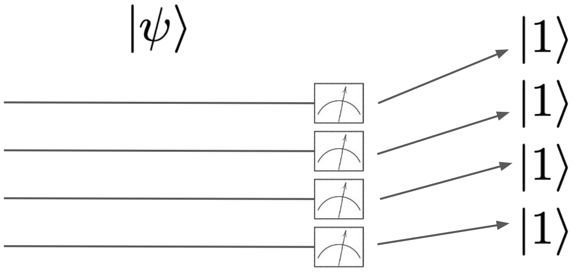
$$|\psi\rangle = \alpha_{0000}|0000\rangle + \alpha_{0001}|0001\rangle + \alpha_{0010}|0010\rangle + \ldots \alpha_{1111}\boxed{|1111\rangle}$$

$$\boxed{|\alpha_{1111}|^2}$$

QUANTINUUM

Measuring Many Qubits

$$|\psi\rangle \in \mathbb{C}^{2^4}$$



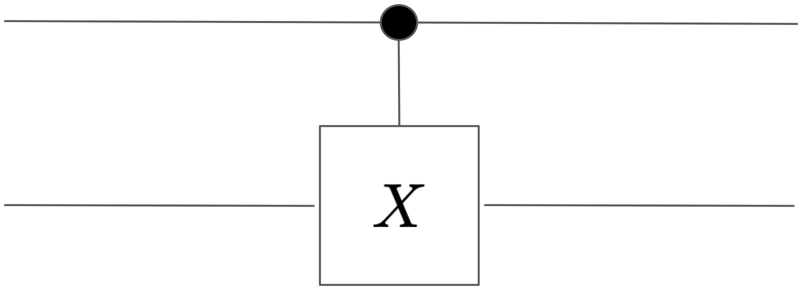
$$|\psi\rangle = \alpha_{0000}|0000\rangle + \alpha_{0001}|0001\rangle + \alpha_{0010}|0010\rangle + \ldots \alpha_{1111}|1111\rangle$$

$$\boxed{\sum_{b \in \{0,1\}^4} |\alpha_b|^2 = 1}$$

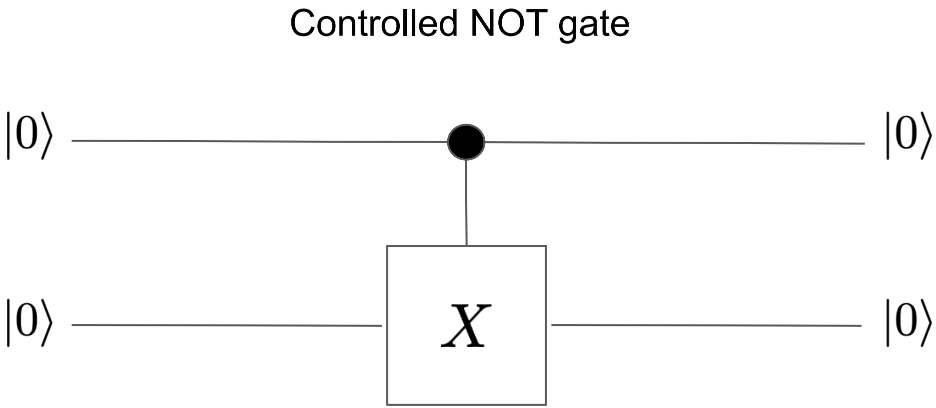
QUANTINUUM

Entangling Qubits

Controlled NOT gate



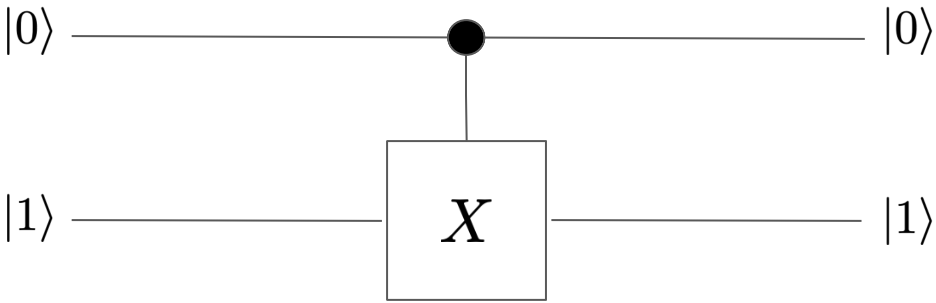
Entangling Qubits



QUANTINUUM

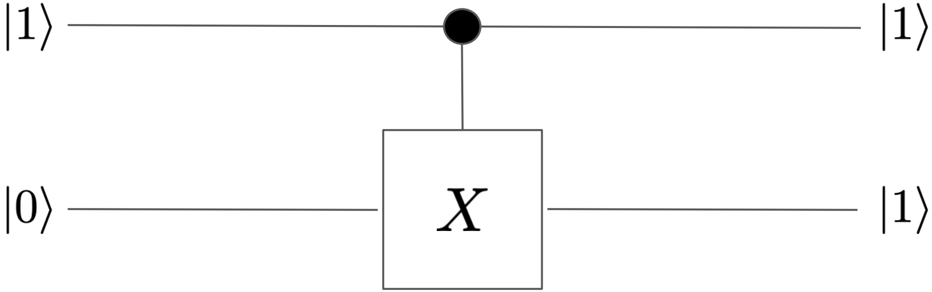
Entangling Qubits

Controlled NOT gate



Entangling Qubits

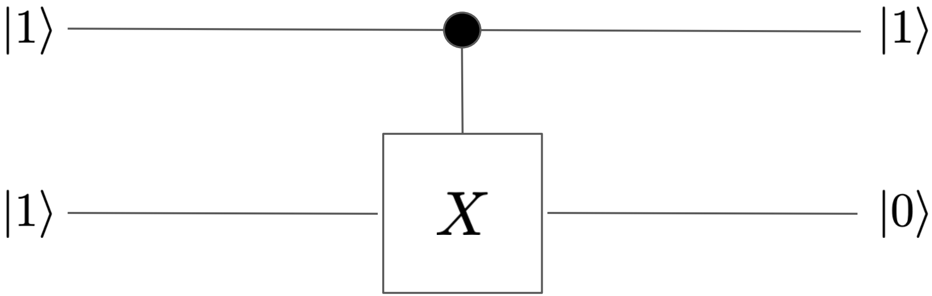
Controlled NOT gate



QUANTINUUM

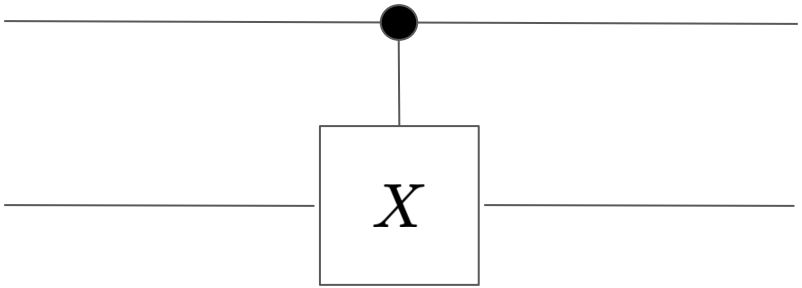
Entangling Qubits

Controlled NOT gate



Entangling Qubits

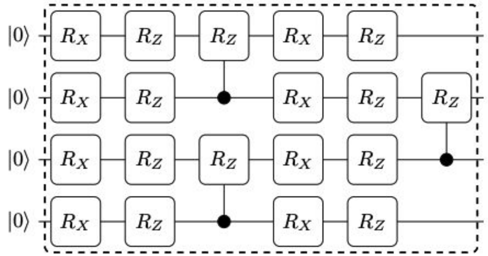
Controlled NOT gate *acts linearly*



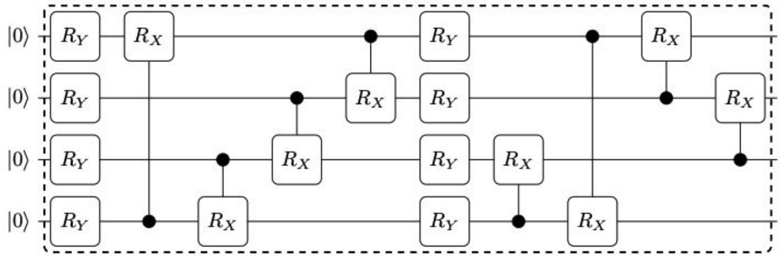
$$CX : \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle \mapsto \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|11\rangle + \alpha_{11}|10\rangle$$

QUANTINUUM

Quantum Circuits



Circuit 7

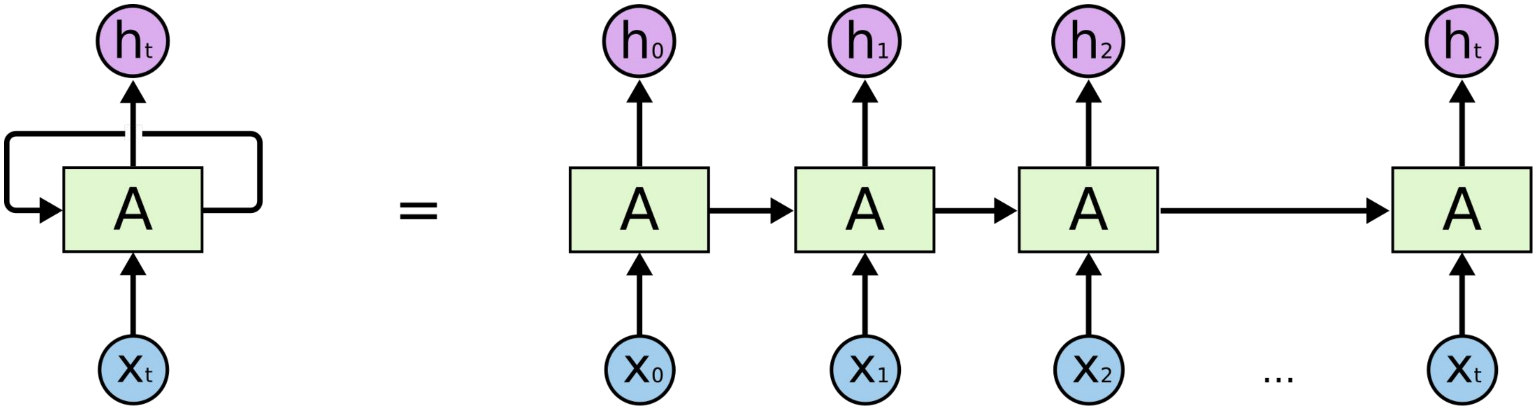


Circuit 14

Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms

Sukin Sim,[1,2,*] Peter D. Johnson,[2] and Alán Aspuru-Guzik[2,3,4,5,†]

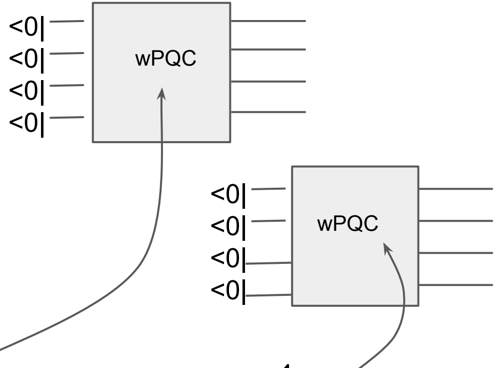
Recurrent Neural Networks (RNNs)

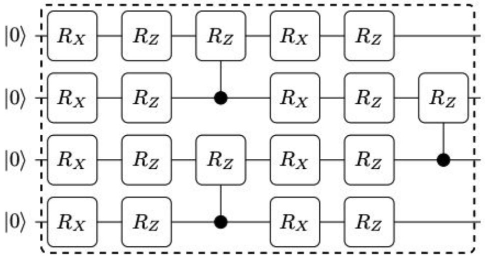


$$h_t = f(x_t\mathbf{U} + h_{t-1}\mathbf{W})$$

From Colah's blog: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Parameterised Quantum Circuits (PQCs)





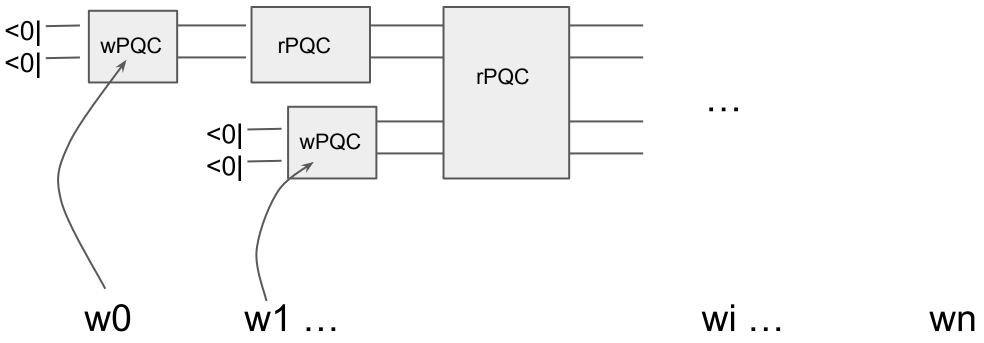
Circuit 7

w0 w1 = … wi … wn

(Rx1, Rx2, …, Rz1, …, CRz1, …)

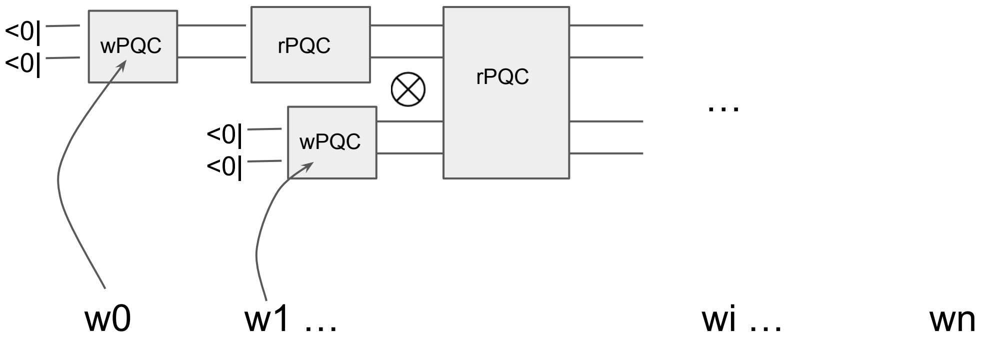
Angle encoding

qRNN Take One



QUANTINUUM

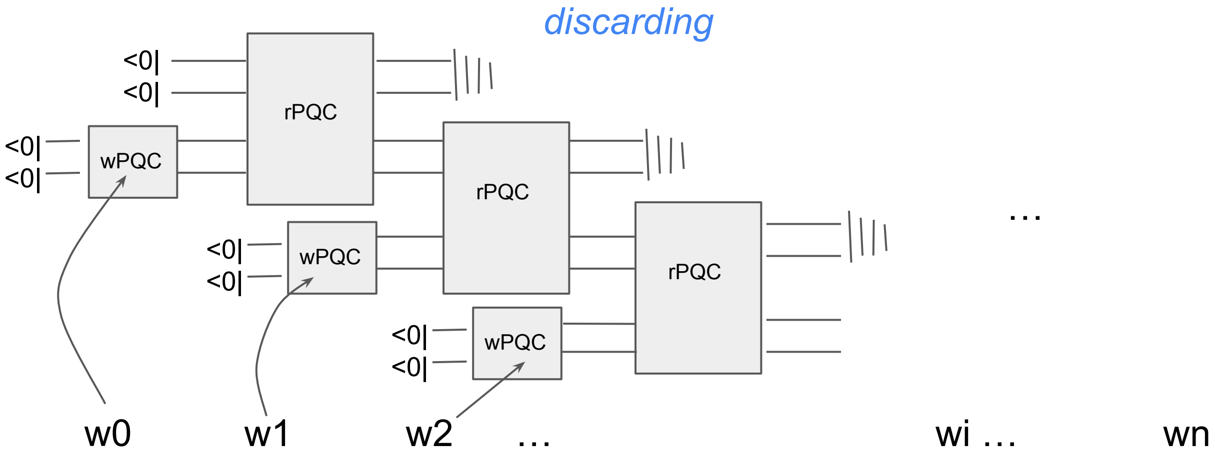
qRNN Take One



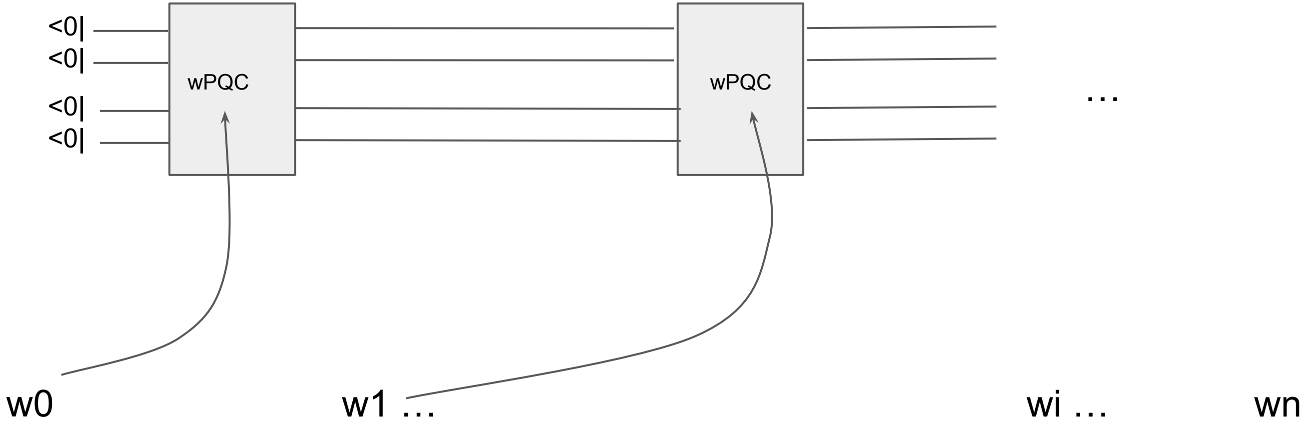
tensor product

QUANTINUUM

qRNN Take One

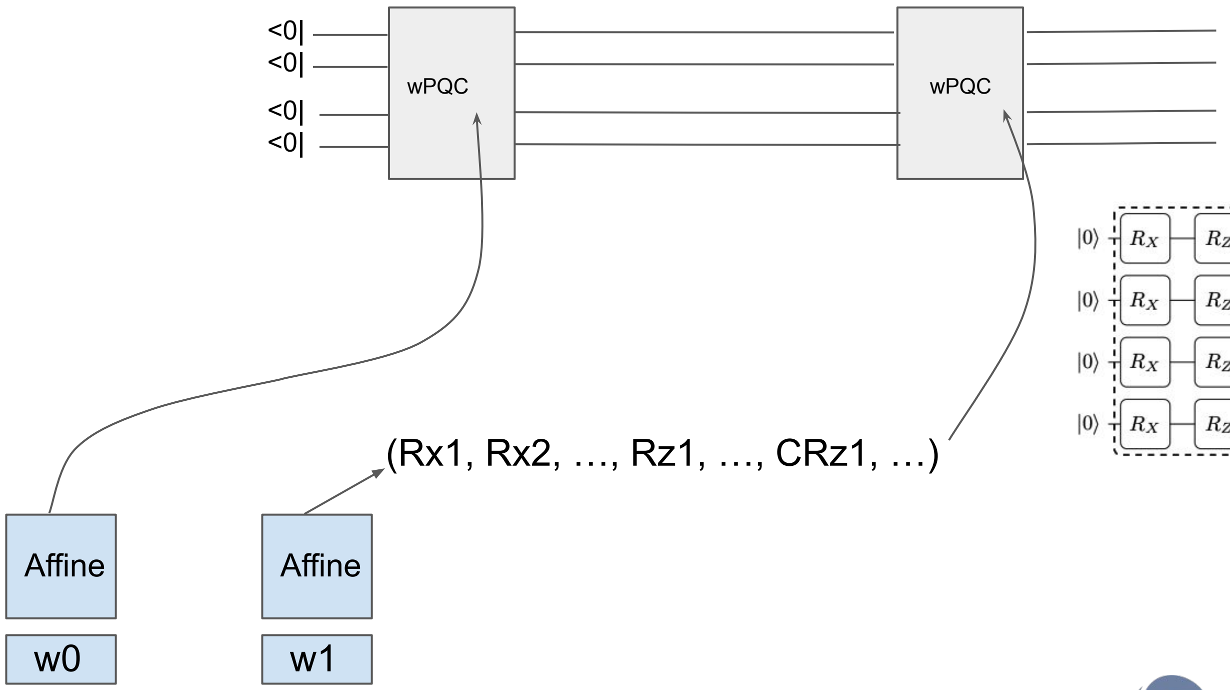


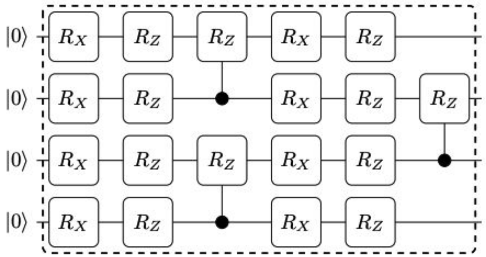
qRNN Take Two



QUANTINUUM

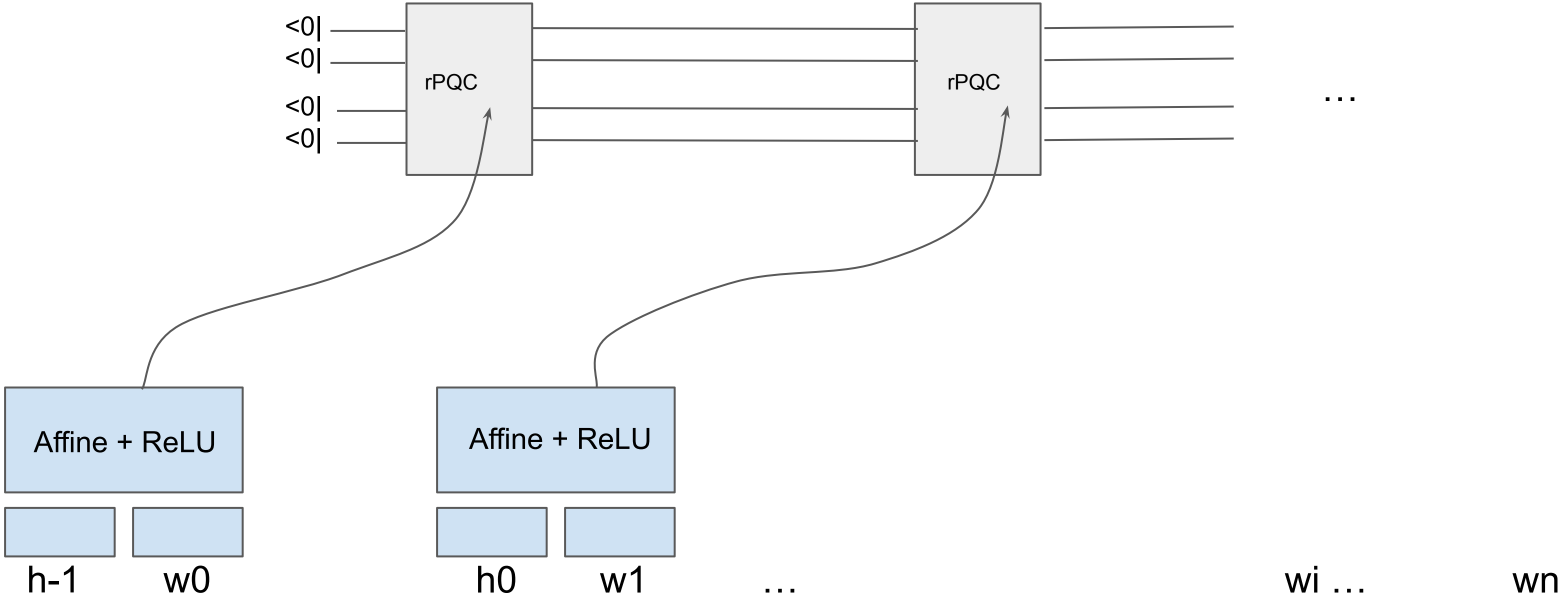
qRNN Take Two



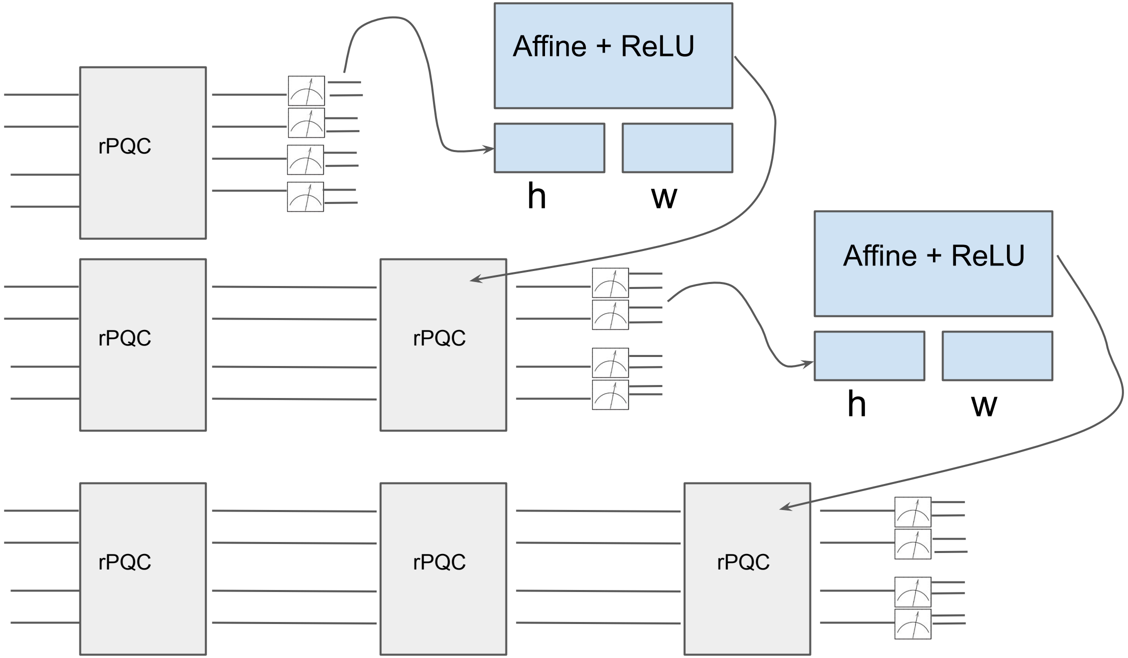


Circuit 7

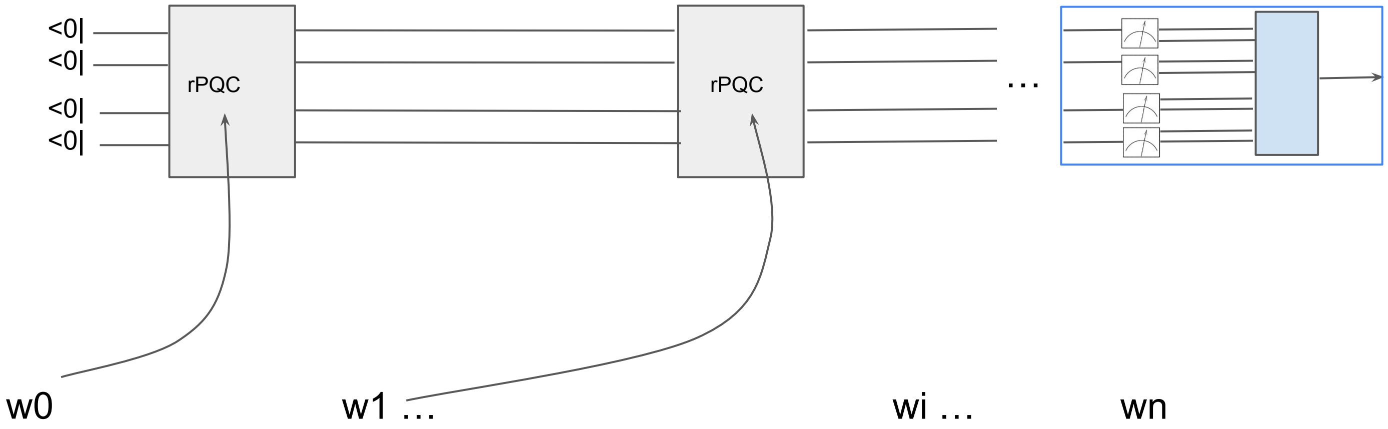
qRNN Take Two (Variant)



qRNN Take Two (Variant, “Unrolled”)

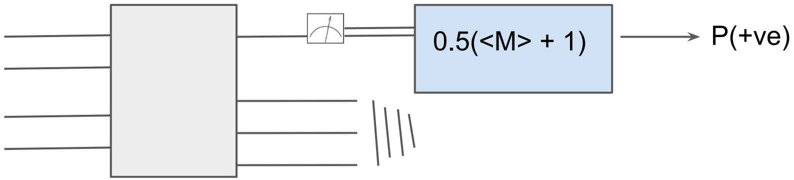


Output

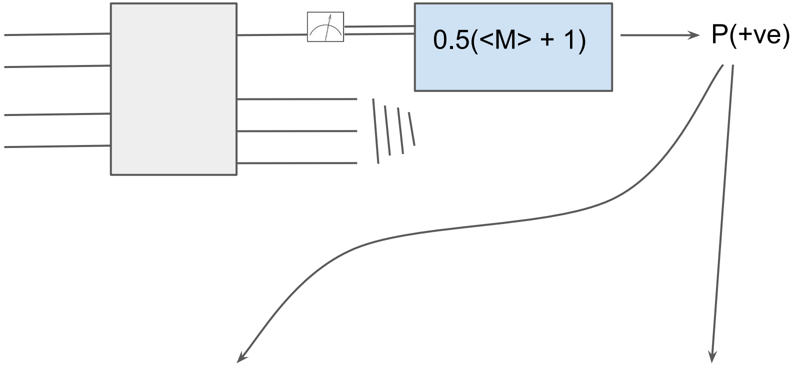


QUANTINUUM

Probabilistic Output

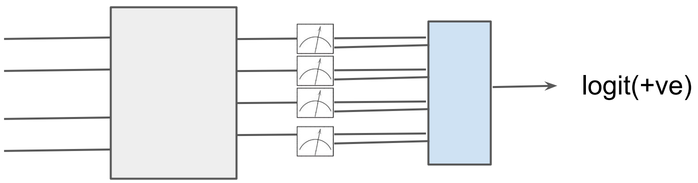


Probabilistic Output for Training



$$-(y_i \log(p_i) + (1 - y_i) \log(1 - p_i))$$

Neural Output



The Task

- Sentiment analysis (Rotten Tomatoes dataset)
- 8,530 training examples (well balanced); 1,066 dev examples
- Simple binary classification task

`if you sometimes like to go to the movies to have fun , wasabi is a good place to start .`	1
`emerges as something rare , an issue movie that's so honest and keenly observed that it doesn't feel like one .`	1
`simplistic , silly and tedious .`	0
`it's so laddish and juvenile , only teenage boys could possibly find it funny .`	0

Baseline / Goal

- Goal is *not* to beat the s-o-t-a
- Goal (at this stage) is to be competitive with a classical vanilla RNN

Hybrid Toolkit

- Requirements for classical simulation:
  - easily interfaces with PyTorch (or TensorFlow, JAX, …)
  - fast to train on real-world datasets
  - accommodates batching
  - **essentially PyTorch ML library with complex number linear algebra**

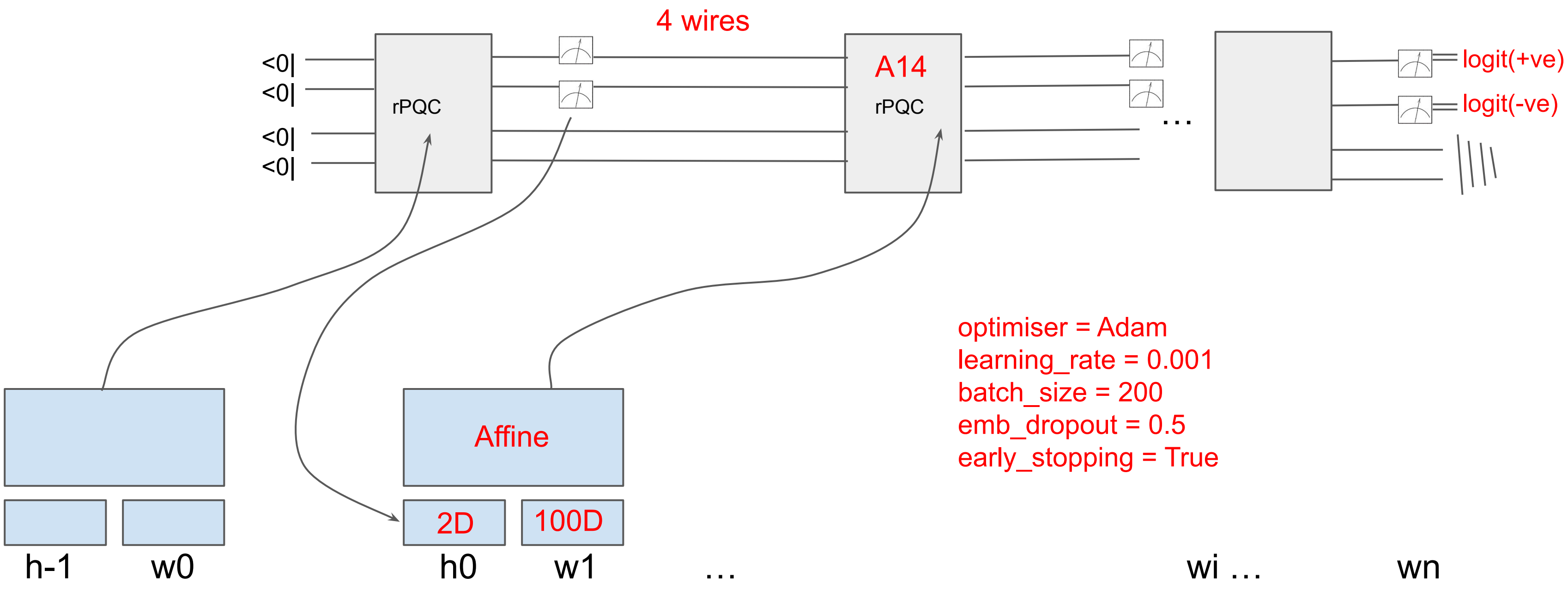
Torch Quantum

A PyTorch Library for Quantum Simulation and Quantum Machine Learning

Faster, Scalable, Easy Debugging, Easy Deployment on Real Machine

QUANTINUUM

Experimental Settings



optimiser = Adam
learning_rate = 0.001
batch_size = 200
emb_dropout = 0.5
early_stopping = True

QUANTINUUM

Results on Dev Set

4 wires	Dev acc
Classical RNN	75.6
“Stairs” (w/feedback, measure 2)	76.5
“Stairs” (wo/feedback, measure 2)	75.1
“Cups” (w/feedback, measure all + affine)	76.6
“Cups” (wo/feedback, measure all + affine)	76.2

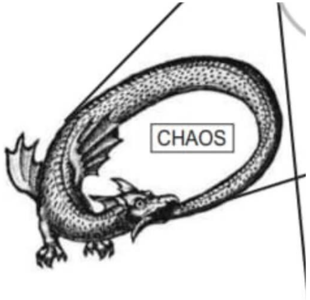
QUANTINUUM

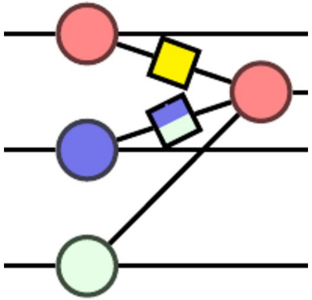
Results (# Wires)

# wires	Dev acc “cups” w/A14 + feedback
Classical RNN	75.6
1	72.1 (rxzx)
2	75.8
4	76.6
10	77.3 (A7, +ReLU on input)

QUANTINUUM

The Oxford Hybrid NLP Team





Wenduan Xu, Konstantinos Meichanetzidis, Douglas Brown, Gabriel Matos,
Charlie London, Richie Yeung, Carys Harvey, Stephen Clark

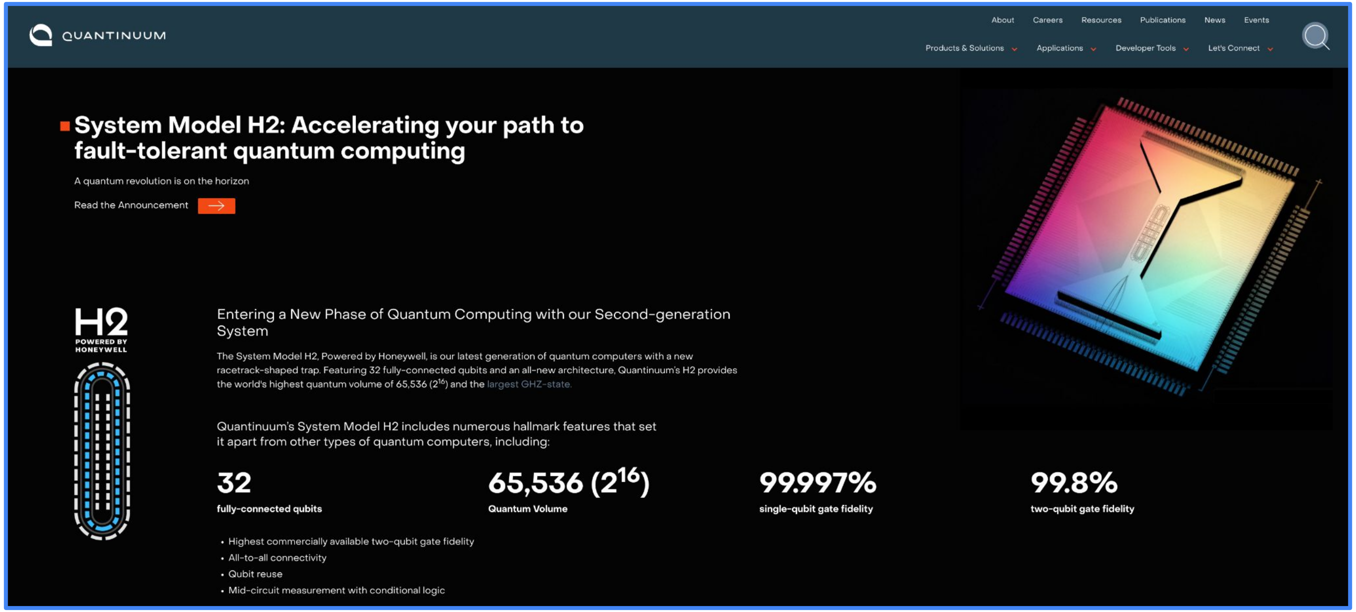
QUANTINUUM

Future Work

- Apply the models to more tasks
  - sequence labelling, language modelling, translation, …
- Apply pre-training / fine-tuning paradigm
- Develop more hybrid architectures
  - based on CNNs (e.g. MERA-like), transformers, …
- Run on quantum hardware

QUANTINUUM

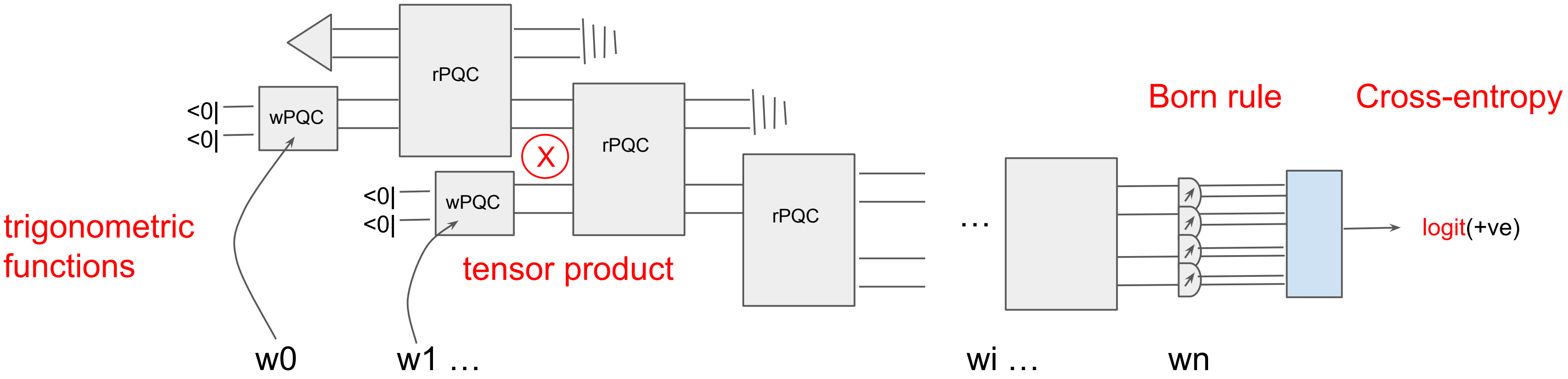
The Future is (Almost) Here



QUANTINUUM

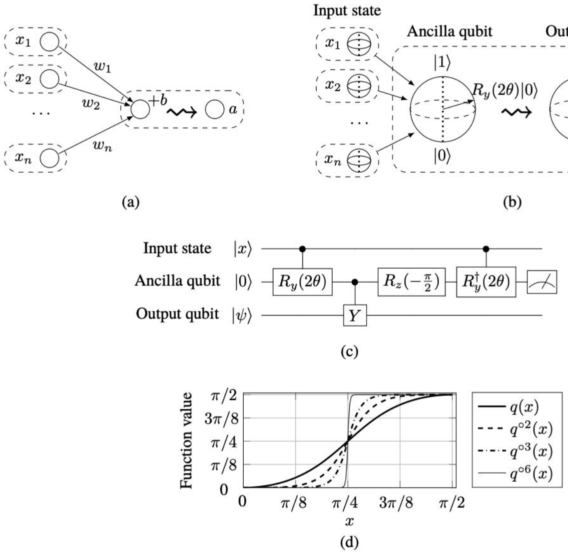
Extra Slides

Where are the Non-linearities?



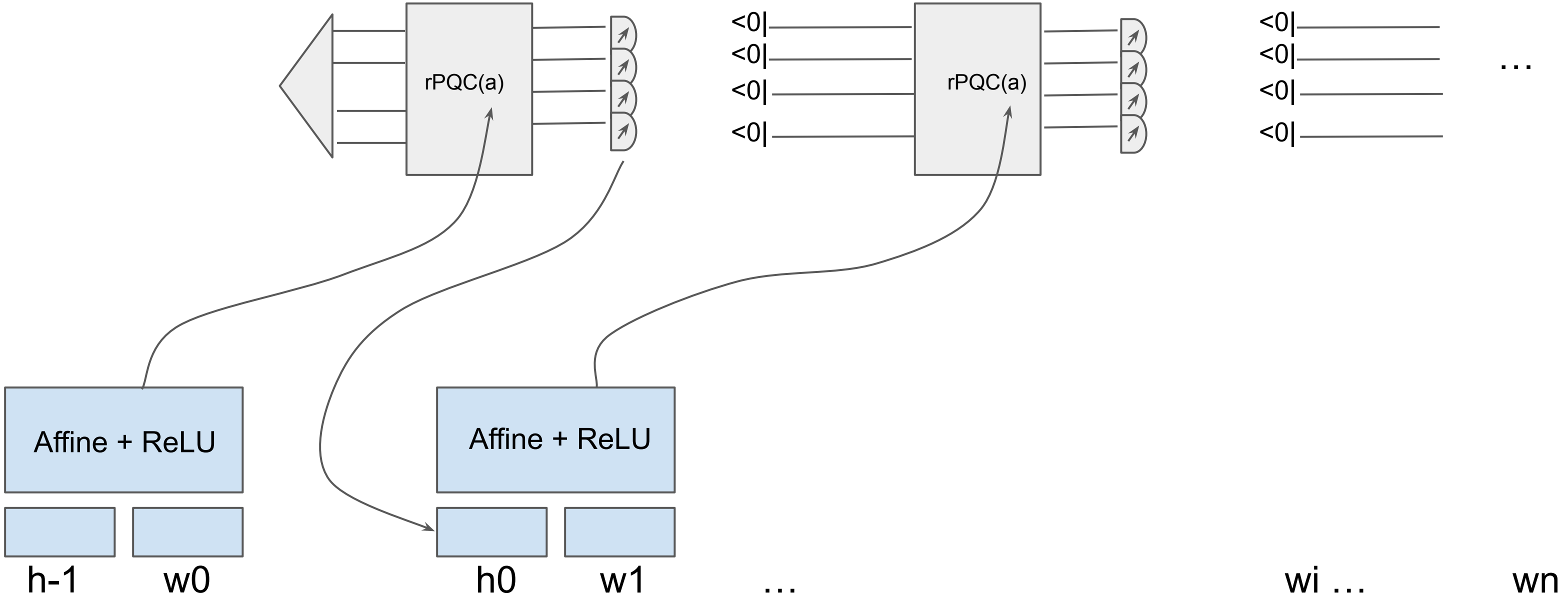
Where are the Non-linearities?

Quantum neuron?



Quantum Neuron: an elementary building block for machine learning on quantum computers, Cao et al., 2017

qRNN Take Two (Another Variant)



Hybrid Toolkits

PENNYLANE

lambeq

TensorFlow Quantum

Torch Quantum

Qiskit

QUANTINUUM

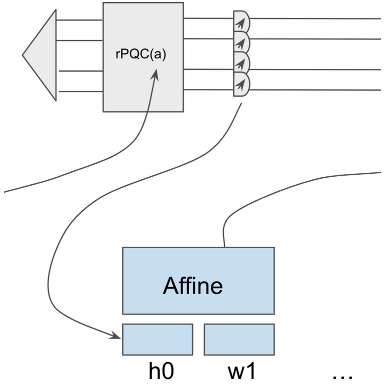
Summary of Architectures - URNN

Unitary Evolution Recurrent Neural Networks

Martin Arjovsky * MARJOVSKY@DC.UBA.AR
Amar Shah * AS793@CAM.AC.UK
Yoshua Bengio
Universidad de Buenos Aires, University of Cambridge,
Université de Montréal. Yoshua Bengio is a CIFAR Senior Fellow.

Assessing the Unitary RNN as an End-to-End Compositional Model of Syntax

Jean-Philippe Bernardy Shalom Lappin
Centre for Linguistic Theory and Studies in Probability
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg



Oct 2022

projUNN: efficient method for training deep networks with unitary matrices

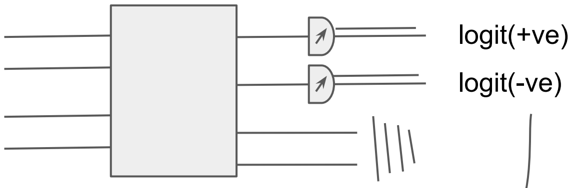
Bobak T. Kiani
MIT
bkiani@mit.edu

Randall Balestriero
Meta AI, FAIR
rbalestriero@fb.com

Yann LeCun
NYU & Meta AI, FAIR
yann@fb.com

Seth Lloyd
MIT & Turing Inc.
slloyd@mit.edu

Logits Output



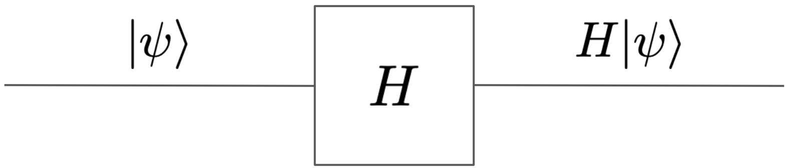
$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \quad CE = -\sum_i^C t_i log(f(s)_i)$$

Where does the Power Come From?

- Superposition
- Entanglement
- Interference

1-Qubit Quantum Gates

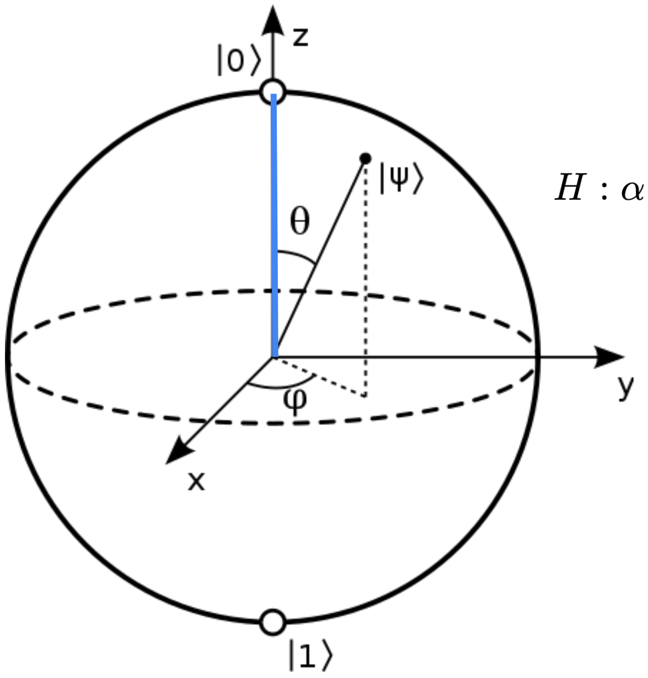
Hadamard Gate



$$H : \alpha|0\rangle + \beta|1\rangle \mapsto \alpha\frac{|0\rangle + |1\rangle}{\sqrt{2}} + \beta\frac{|0\rangle - |1\rangle}{\sqrt{2}}$$

1-Qubit Quantum Gates

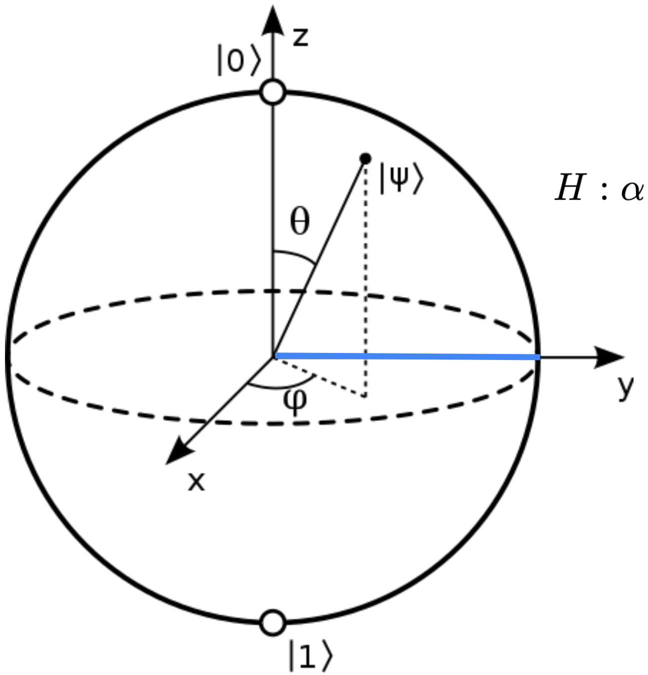
Hadamard Gate *introduces superposition*



$$H : \alpha|0\rangle + \beta|1\rangle \mapsto \alpha\frac{|0\rangle + |1\rangle}{\sqrt{2}} + \beta\frac{|0\rangle - |1\rangle}{\sqrt{2}}$$

1-Qubit Quantum Gates

Hadamard Gate *introduces superposition*



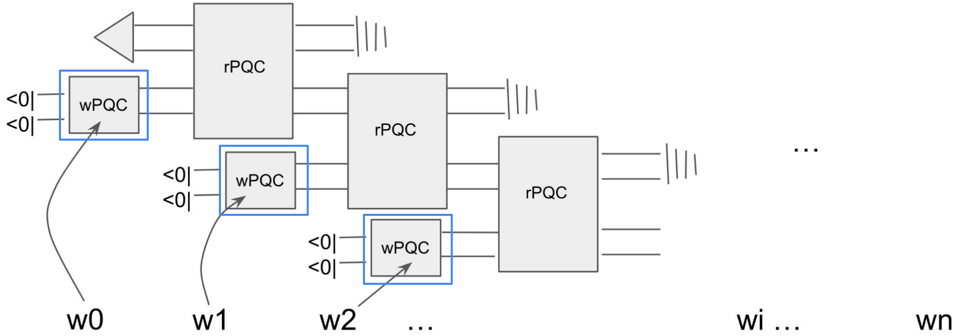
$$H : \alpha|0\rangle + \beta|1\rangle \mapsto \alpha\frac{|0\rangle + |1\rangle}{\sqrt{2}} + \beta\frac{|0\rangle - |1\rangle}{\sqrt{2}}$$

QUANTINUUM

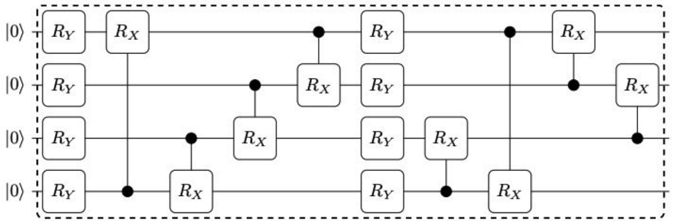
“Stairs” Architecture in Practice

- We added density matrices to TorchQuantum
- Choice of PQC:

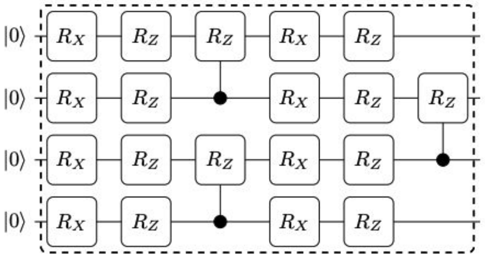
```
'2x4_ryzxy':
    [
        {'input_idx': [0], 'func': 'ry', 'wires': [0]},
        {'input_idx': [1], 'func': 'ry', 'wires': [1]},
        {'input_idx': [2], 'func': 'rz', 'wires': [0]},
        {'input_idx': [3], 'func': 'rz', 'wires': [1]},
        {'input_idx': [4], 'func': 'rx', 'wires': [0]},
        {'input_idx': [5], 'func': 'rx', 'wires': [1]},
        {'input_idx': [6], 'func': 'ry', 'wires': [0]},
        {'input_idx': [7], 'func': 'ry', 'wires': [1]},
    ],
```



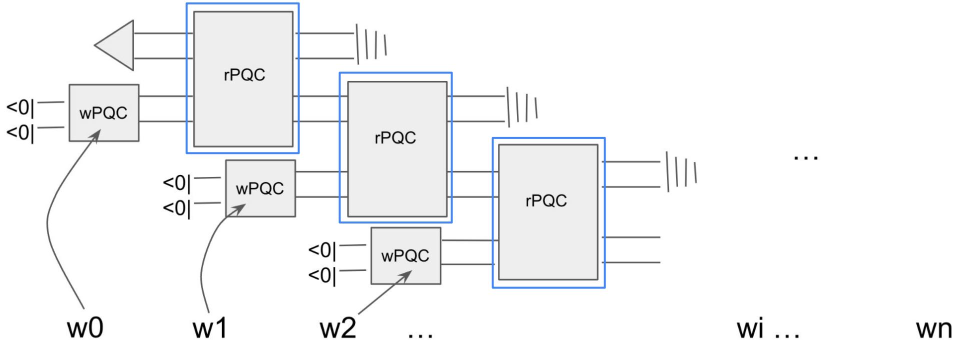
“Stairs” Architecture in Practice



Circuit 14



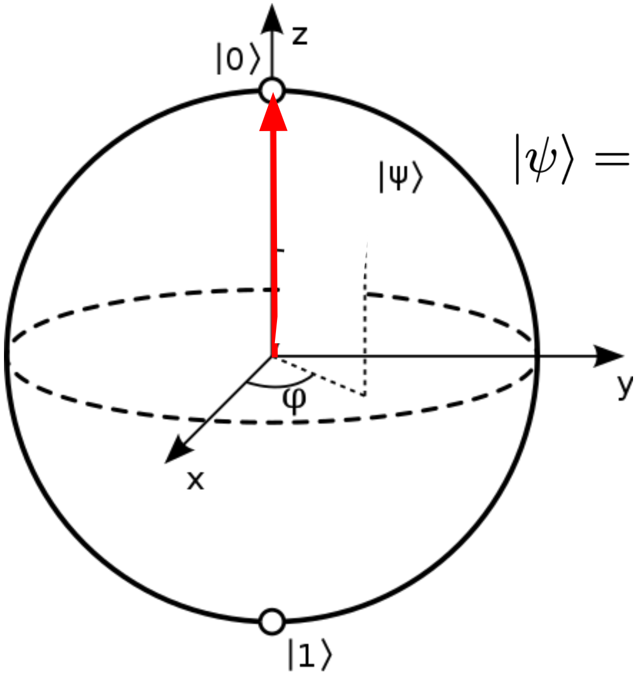
Circuit 7



Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms

Sukin Sim,[1,2,*] Peter D. Johnson,[2] and Alán Aspuru-Guzik[2,3,4,5,†]

“The Collapse of the Wave Function”

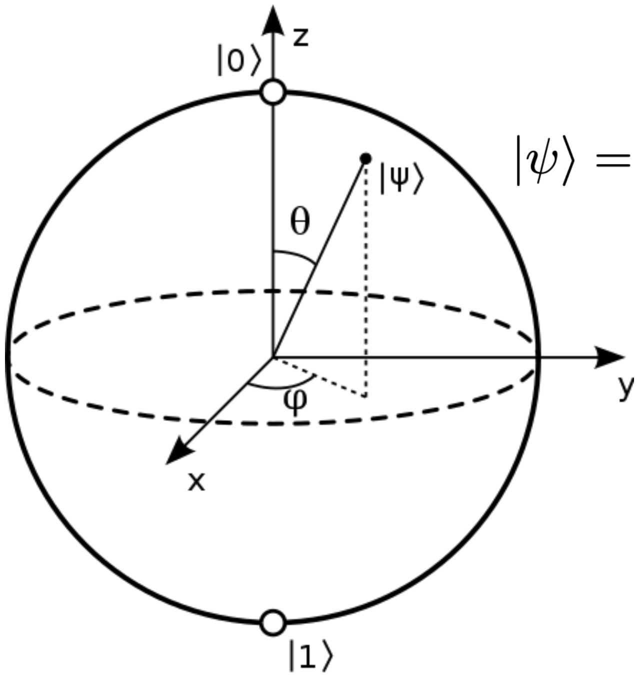


$$|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle$$

$$\boxed{\cos^2(\theta/2)}$$

https://en.wikipedia.org/wiki/Bloch_sphere

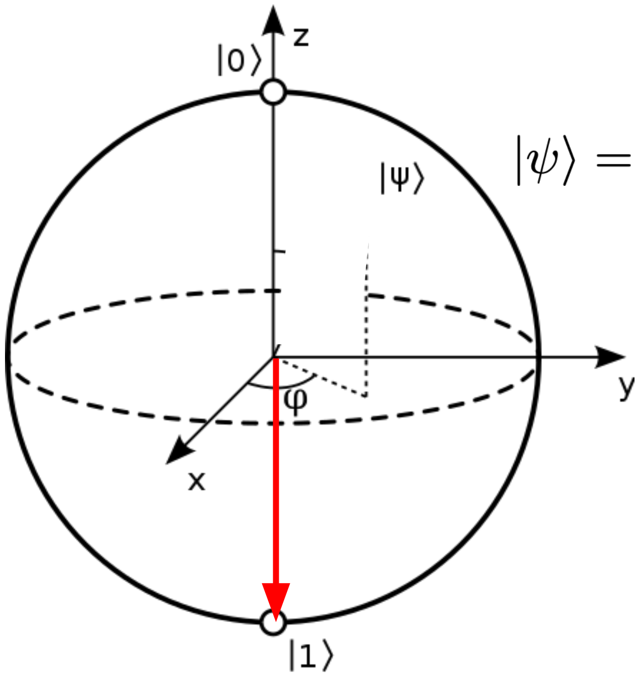
The Bloch Sphere Representation of a Qubit



$$|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle$$

https://en.wikipedia.org/wiki/Bloch_sphere

“The Collapse of the Wave Function”

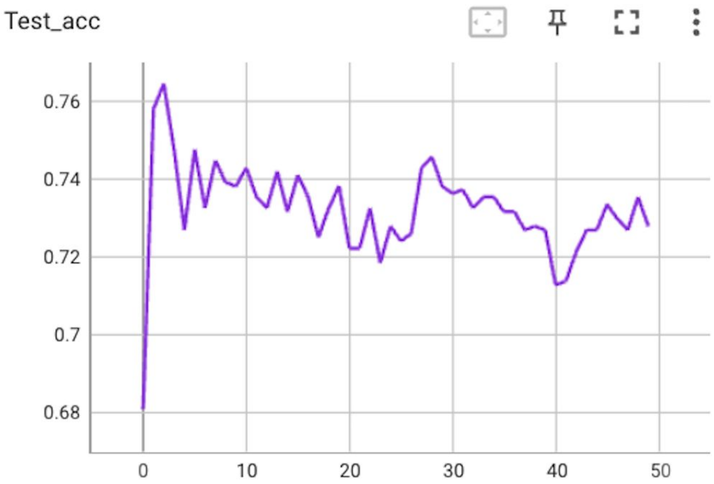


$$|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle$$

$$\boxed{\sin^2(\theta/2)}$$

https://en.wikipedia.org/wiki/Bloch_sphere

Learning Curve



NVidia A30 GPU, PyTorch 1.12:

~5 secs / epoch for 1 wire (pure state)
~11 secs / epoch for 2 wires (pure state)
~14 secs / epoch for 4 wires (pure state)
~26 secs / epoch for 8 wires (pure state)