University of Pavia

Quantum computing models of artificial neurons

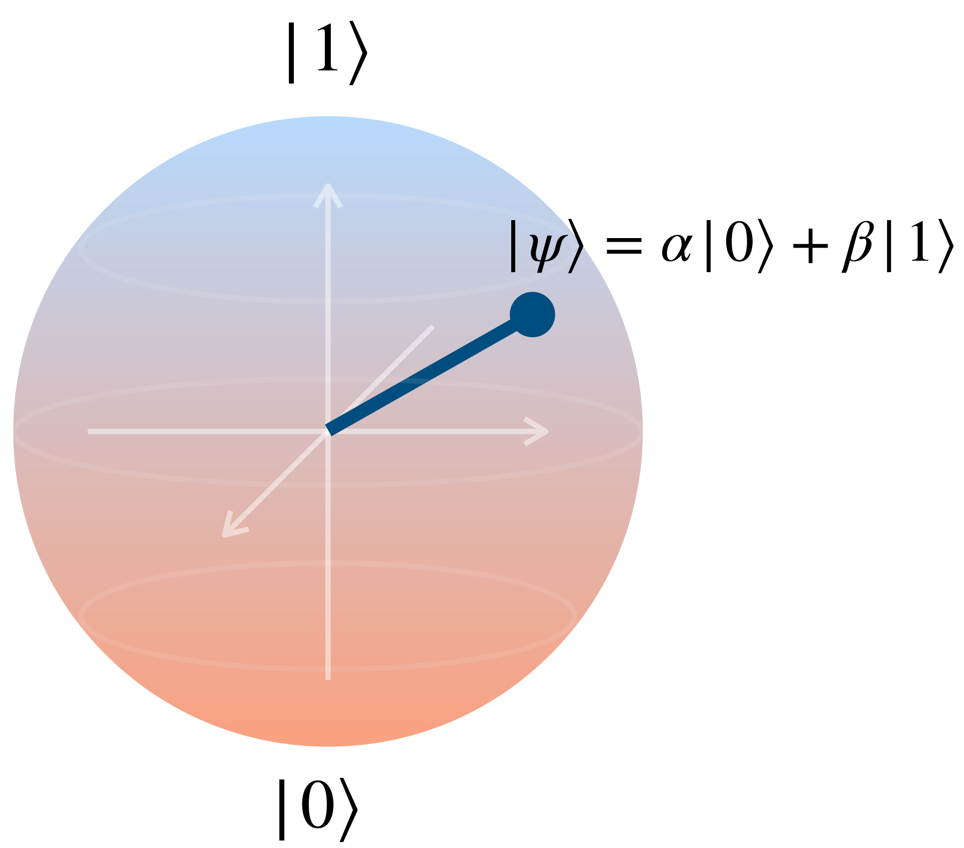
F. Tacchino *et al.*, npj Quantum Information 5, 26 (2019)
S. Mangini *et al.*, Machine Learning: Science and Technology (2020)

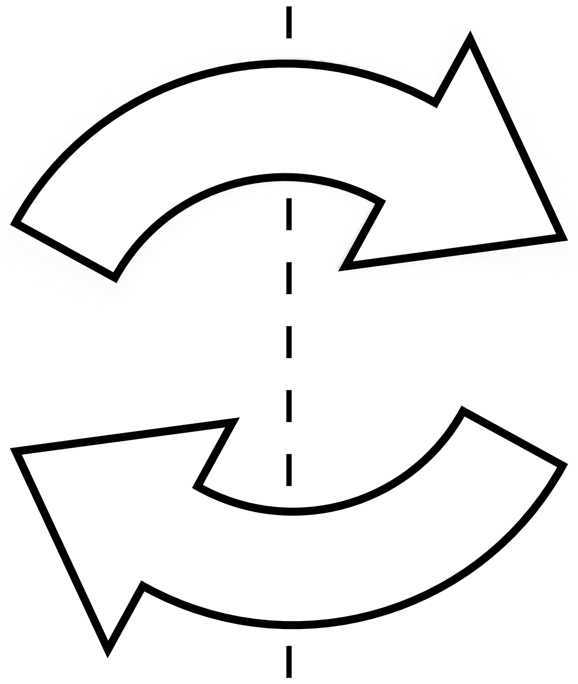
YIQIS 2020

30 September 2020, Stefano Mangini

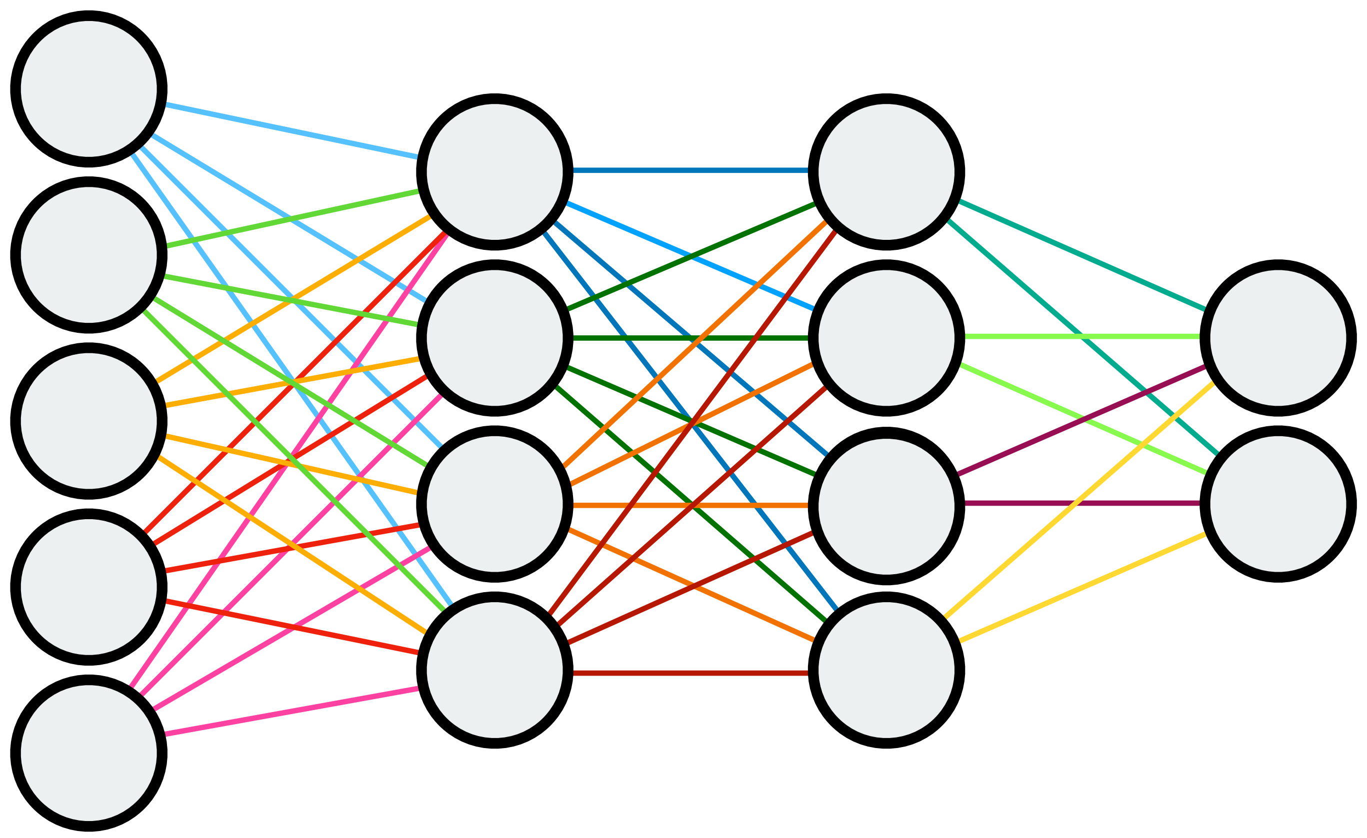
Quantum Machine Learning: what and why?

Quantum computing





Machine Learning



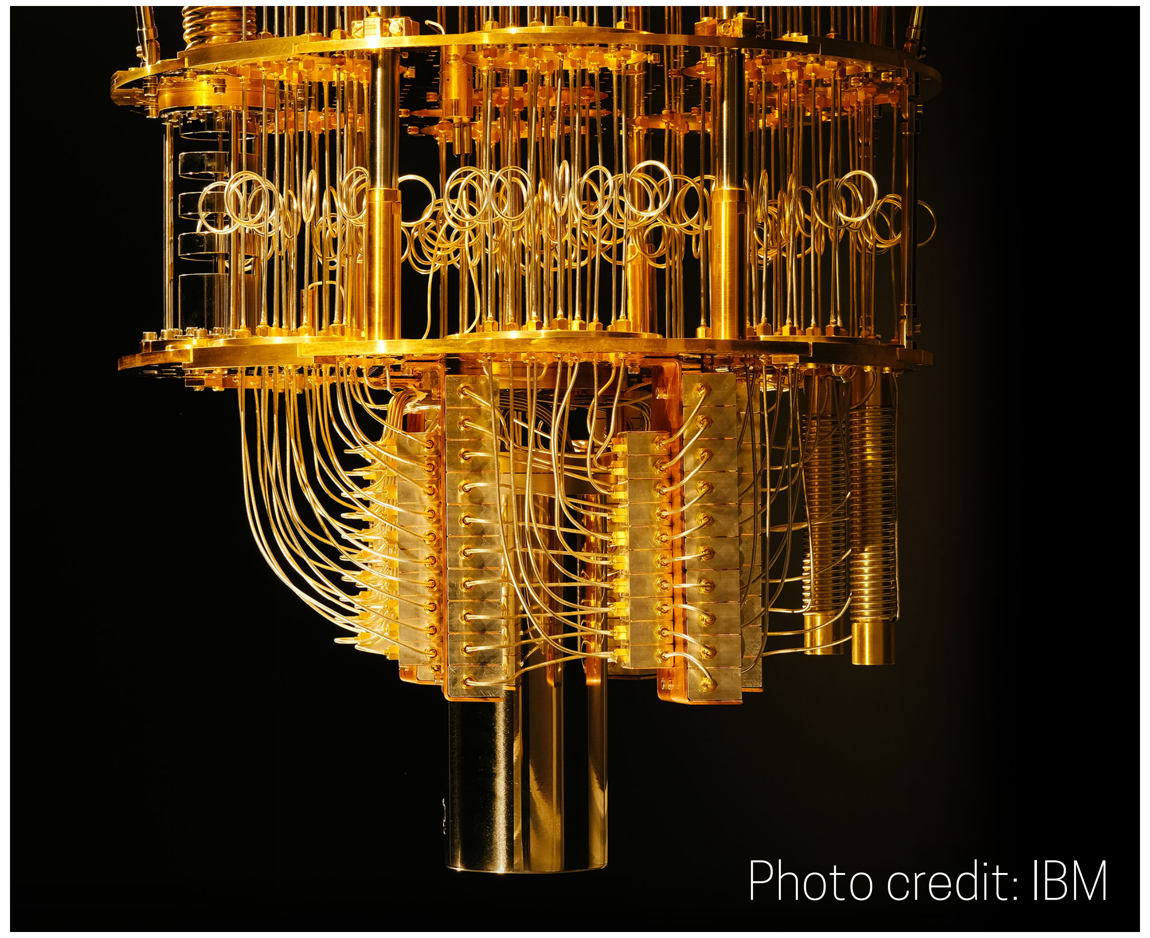


Photo credit: IBM

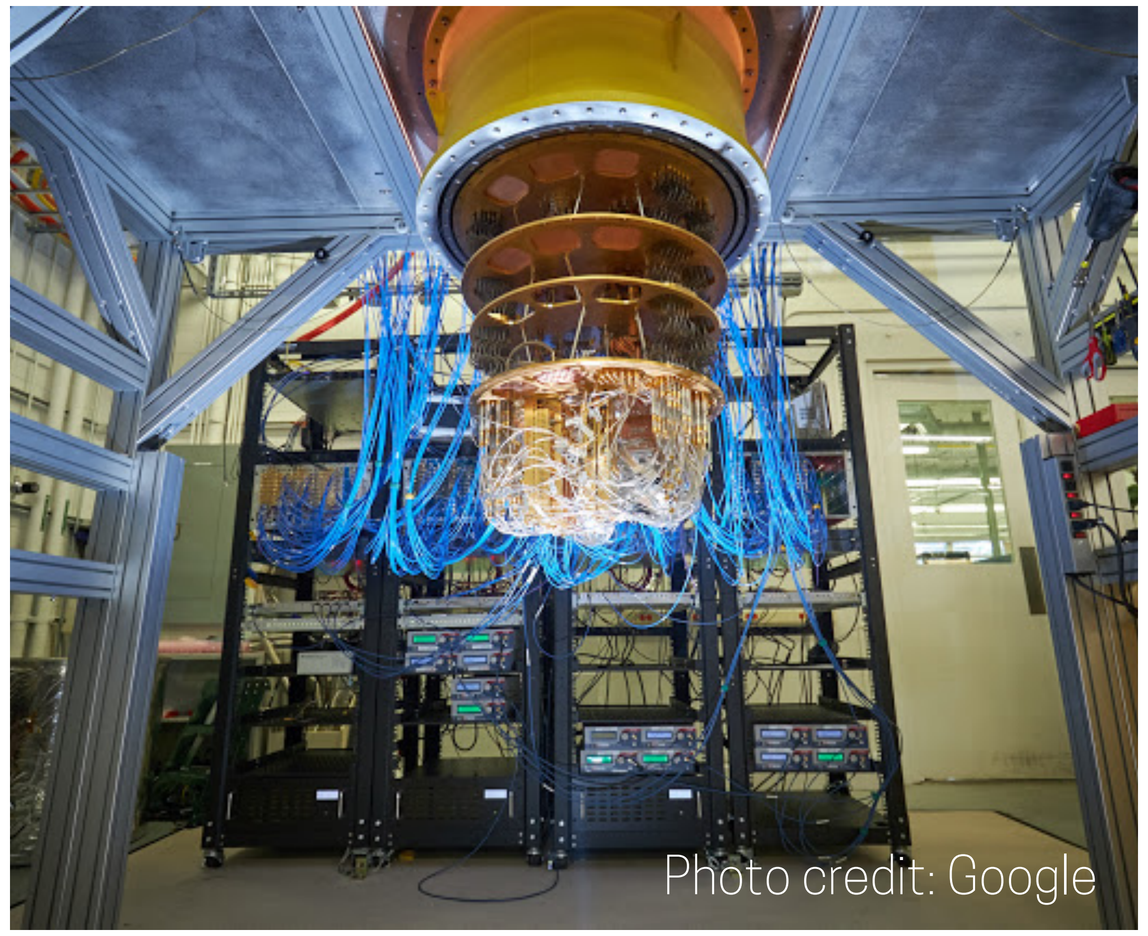
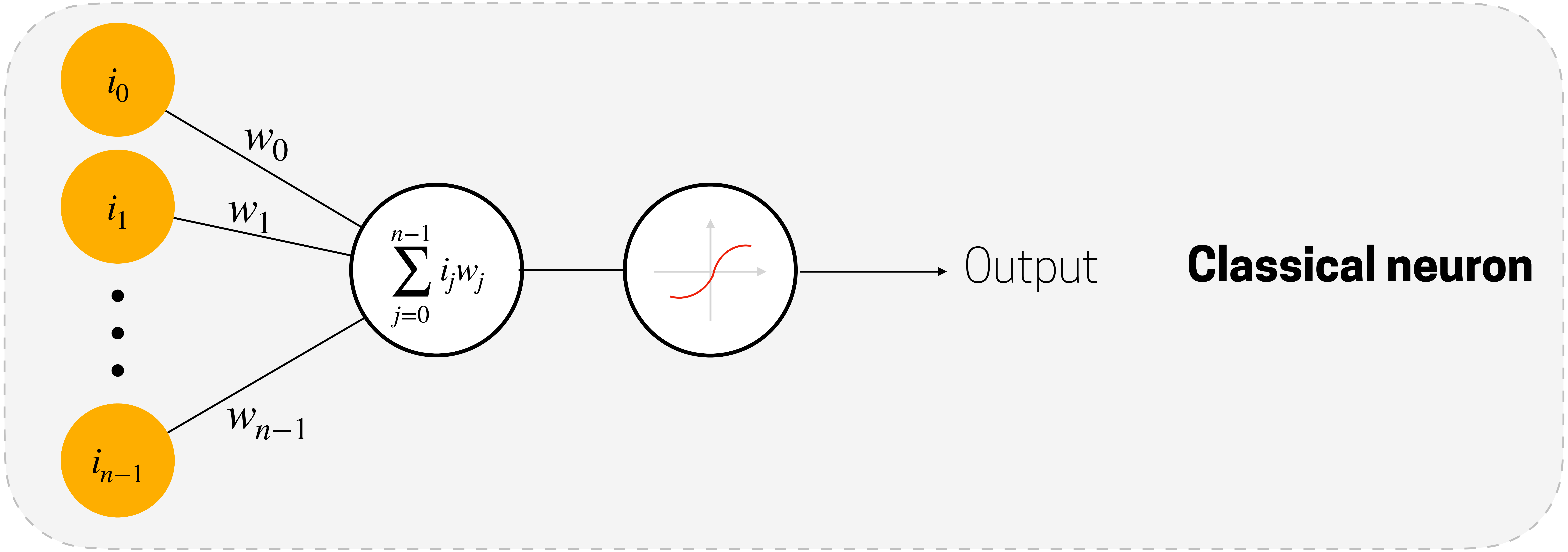


Photo credit: Google

The classical perceptron model

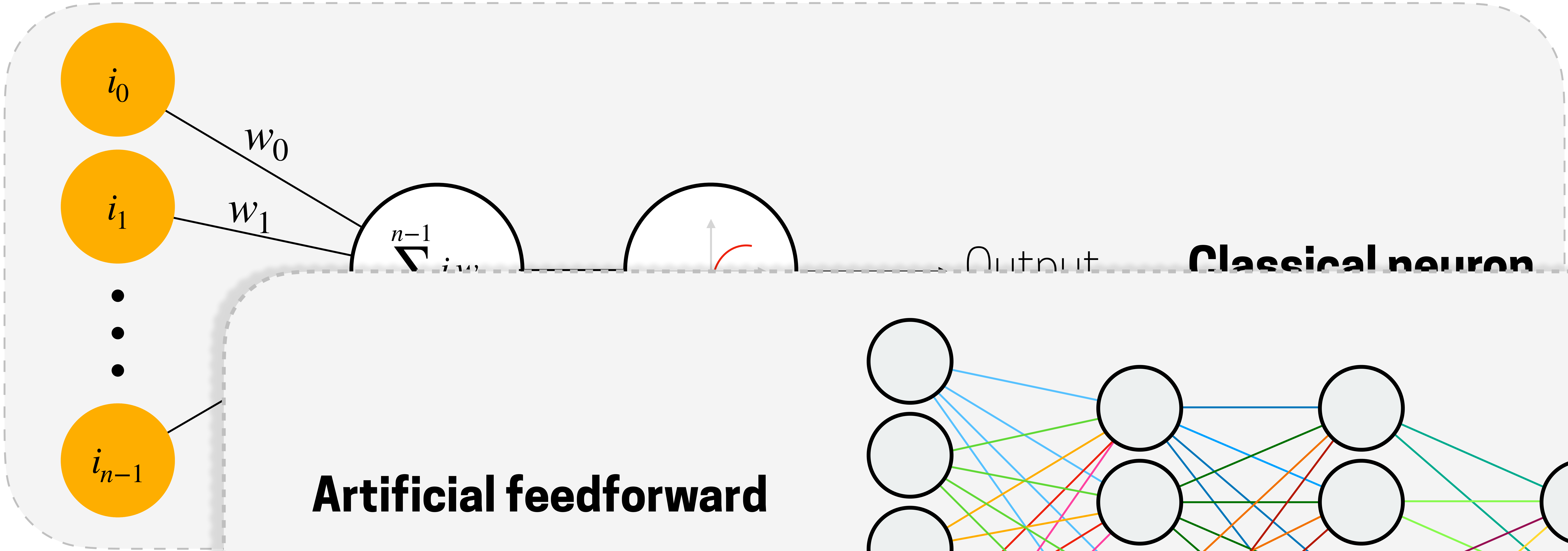


Input vector $\vec{i}$

Weights (and bias) vector $\vec{w}$

Output $y = f(\vec{i} \cdot \vec{w})$

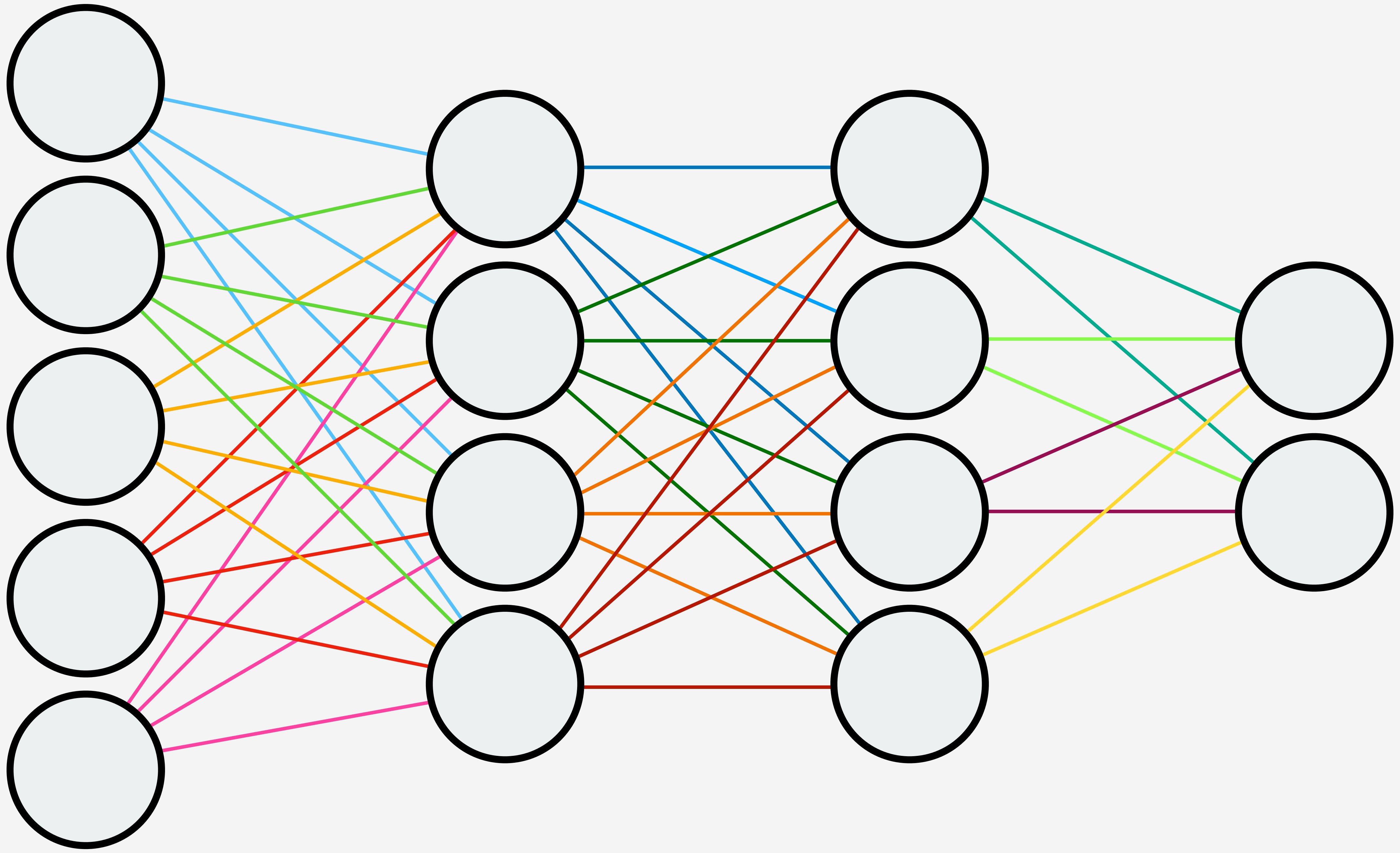
The classical perceptron model



Input ve

Weights

Artificial feedforward neural network



Into the quantum domain

$$\vec{i} = (i_0, i_1, \ldots, i_{n-1})$$

$$\vec{w} = (w_0, w_1, \ldots, w_{n-1})$$

$$i_j, w_j \in \{-1, 1\}$$

The $\boldsymbol{n}$-bit long input and weight vector can be encoded in the amplitudes ± 1 of a balanced superposition of the computational basis states of $N = \log_2 n$ qubits

Consider the quantum states

$$|\psi_i\rangle = \frac{1}{\sqrt{2^N}} \sum_{j=0}^{2^N-1} i_j |j\rangle$$

$$|\psi_w\rangle = \frac{1}{\sqrt{2^N}} \sum_{j=0}^{2^N-1} w_j |j\rangle$$

$$\langle \psi_i | \psi_w \rangle = \sum_{j,k=0}^{2^N-1} i_j w_k \langle j | k \rangle = \sum_{j=0}^{2^N-1} i_j w_j$$

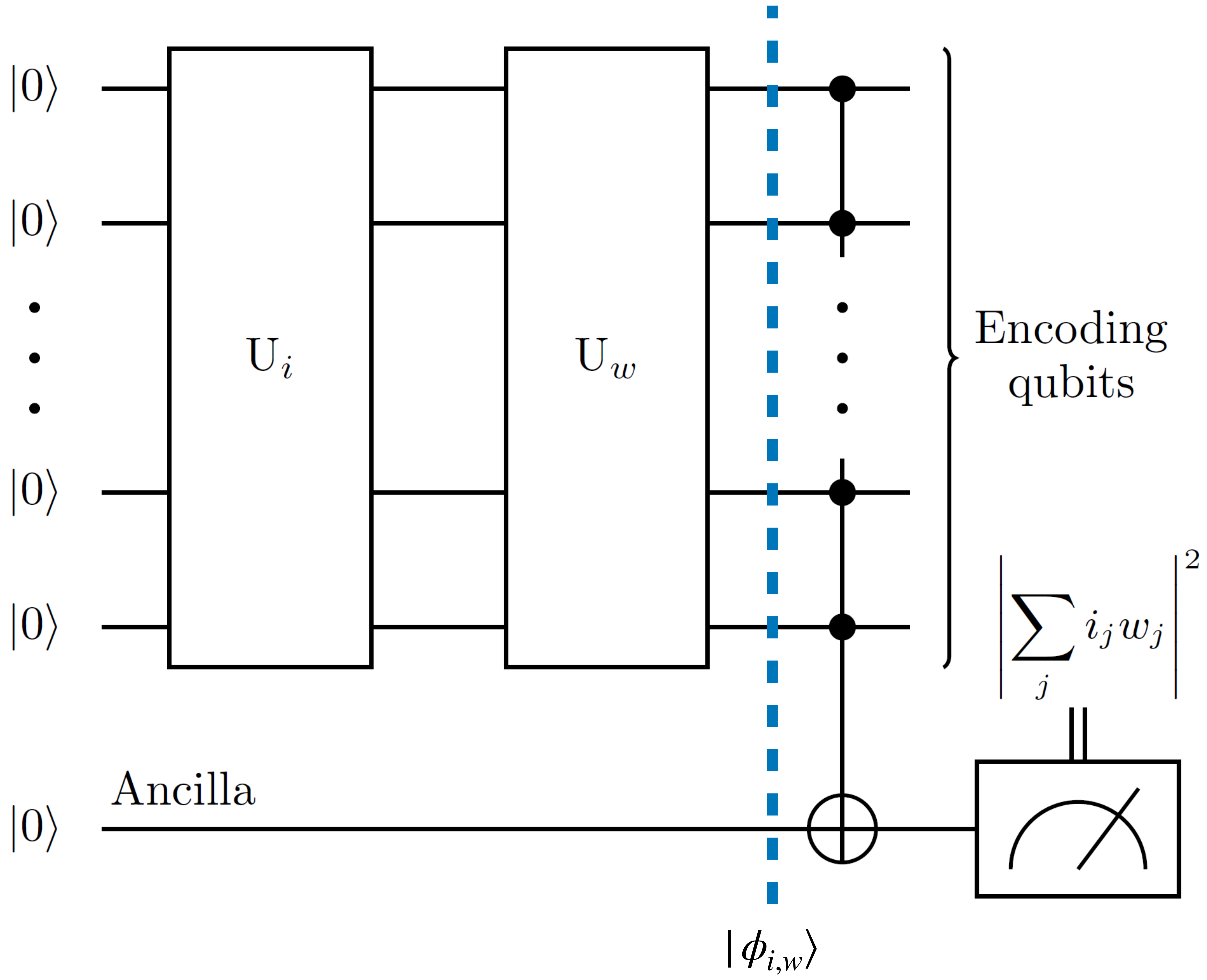
This class of states are known as Real Equally Weighted (REW) states

N qubits used to encode 2^N classical bits

F. Tacchino et al., npj Quantum Information 5, 26 (2019)

Quantum artificial neuron

Quantum circuit implementation



Quantum input state preparation

$$U_i|0\rangle^{\otimes N} = |\psi_i\rangle$$

Inner product (weighted sum)

$$U_w|\psi_w\rangle = |1\rangle^{\otimes N}$$

because:

$$\langle\psi_w|\psi_i\rangle = \langle\psi_w|U_w^\dagger U_w|\psi_i\rangle = \langle\mathbf{1}|\underbrace{U_w U_i|\mathbf{0}\rangle}_{|\phi_{i,w}\rangle}$$

Measurement of ancilla yields $|1\rangle$ with probability $\left|\sum_j i_j w_j\right|^2$

Activation function!

Classification of checkboard patterns

$$\vec{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_{2^N-1} \end{pmatrix} \quad \vec{i} = \begin{pmatrix} i_0 \\ i_1 \\ \vdots \\ i_{2^N-1} \end{pmatrix}$$

$N = 2$

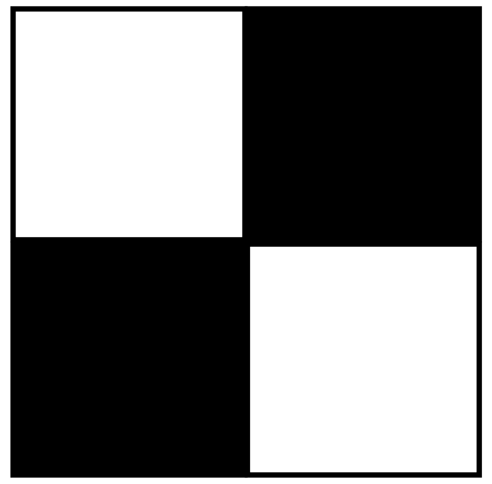
$\vec{i} = (i_0, i_1, i_2, i_3)$

i_0	i_1
i_2	i_3

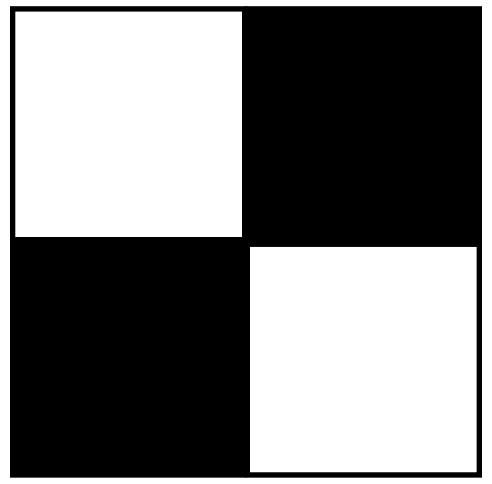
white if $i_j, w_j = +1$

black if $i_j, w_j = -1$

$\vec{i} =$ $(1, -1, 1, -1)$



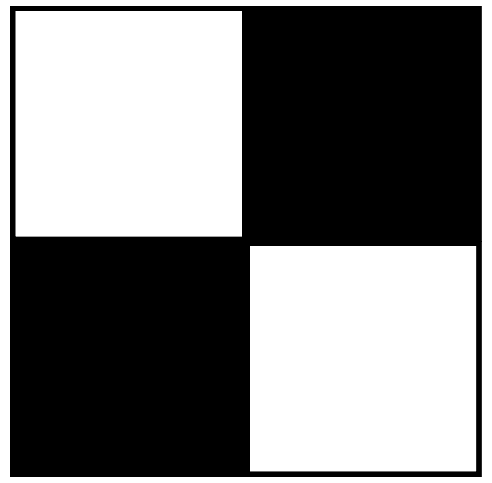
$\vec{w} =$ $(1, -1, 1, -1)$



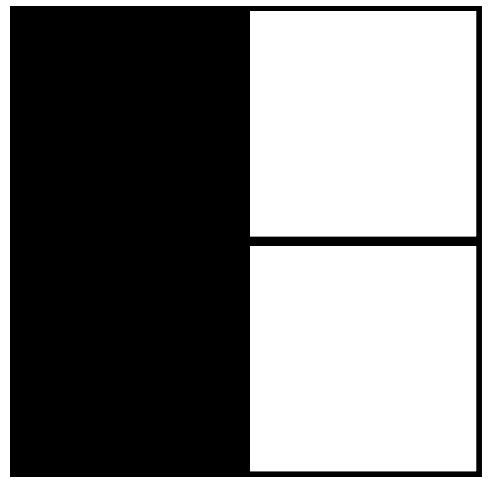
$\langle \psi_w | \psi_i \rangle = \vec{i} \cdot \vec{w} = 1$

Perfect activation

$\vec{i} =$ $(1, -1, 1, -1)$



$\vec{w} =$ $(-1, 1, 1, -1)$



$\langle \psi_w | \psi_i \rangle = \vec{i} \cdot \vec{w} = 0$

No activation

From Binary to Continuous values

Binary $|\psi_i\rangle = \frac{1}{\sqrt{2^N}} \sum_{j=0}^{2^N-1} i_j |j\rangle$ with $i_j = \pm 1$ but $e^{i\theta} = \begin{cases} 1 & \theta = 0 \\ -1 & \theta = \pi \end{cases}$

By means of **phase encoding** we load the data on the quantum states!

input $\vec{\theta} = (\theta_0, \theta_1, \ldots, \theta_{n-1})$

weights $\vec{\phi} = (\phi_0, \phi_1, \ldots, \phi_{n-1})$

$\theta_j, \phi_j \in [0,\pi]$ (not 2π due to periodicity)

Quantum states:

$$|\psi_\theta\rangle = \frac{1}{\sqrt{2^N}} \sum_{j=0}^{2^N-1} e^{i\theta_j} |j\rangle$$

$$|\psi_\phi\rangle = \frac{1}{\sqrt{2^N}} \sum_{j=0}^{2^N-1} e^{i\phi_j} |j\rangle$$

$\longrightarrow$

$$|\langle\psi_\phi|\psi_\theta\rangle|^2 = \left|\sum_{j}^{2^N-1} e^{i(\theta_j-\phi_j)}\right|^2 = \ldots =$$

$$= \frac{1}{2^N} + \frac{1}{2^{2N-1}} \sum_{i<j}^{2^N-1} \cos((\theta_j - \phi_j) - (\theta_i - \phi_i))$$

S. Mangini *et al.*, Machine Learning: Science and Technology (2020)

Some useful remarks

The activation function

$$f(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{1}{2^N} + \frac{1}{2^{2N-1}} \sum_{i<j}^{2^N-1} \cos((\theta_j - \phi_j) - (\theta_i - \phi_i))$$

- If $\boldsymbol{\theta} = \boldsymbol{\phi}$

$$f(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{1}{2^N} + \frac{1}{2^{2N-1}} \sum_{i<j}^{2^N-1} = \frac{1}{2^N} + \frac{1}{2^{2N-1}} \frac{2^N(2^N-1)}{2} = 1$$

which also holds if

$$\boldsymbol{\theta} = \boldsymbol{\phi} + \boldsymbol{\Delta}$$

Color invariance!

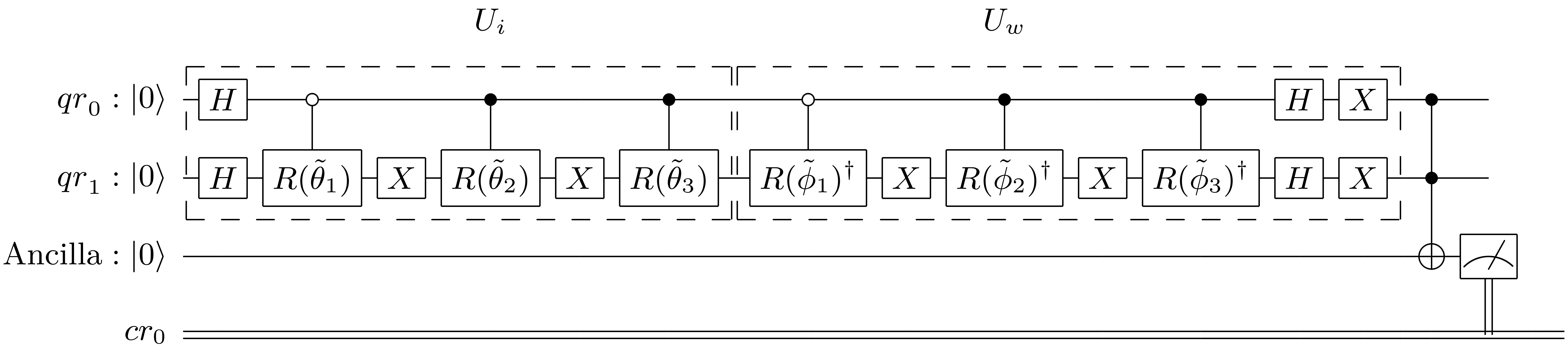
- If $\boldsymbol{\theta} = \boldsymbol{\phi} + \boldsymbol{\Delta}, \quad \Delta_j \sim \text{Unif}(-a/2, a/2)$

$$\langle f(\boldsymbol{\theta}, \boldsymbol{\phi}) \rangle \approx 1 - O(a^2)$$

Noise resilience!

Actual quantum circuit implementation

N = 2 qubits



NOT gate

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

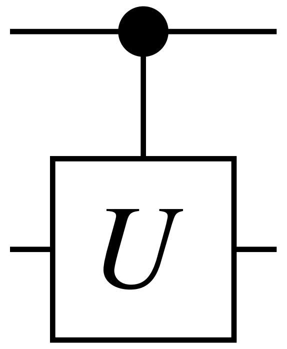
Hadamard gate

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Phase shift gate

$$R(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix}$$

Controlled gate



$$CU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & U_{11} & U_{12} \\ 0 & 0 & U_{21} & U_{22} \end{bmatrix}$$

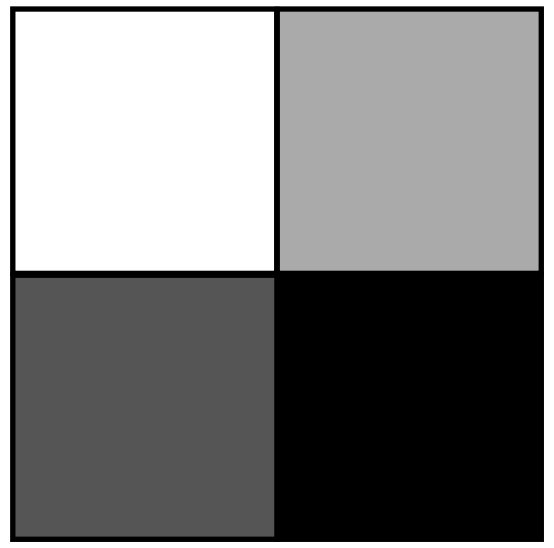
Classification of grayscale patterns

Grayscale images

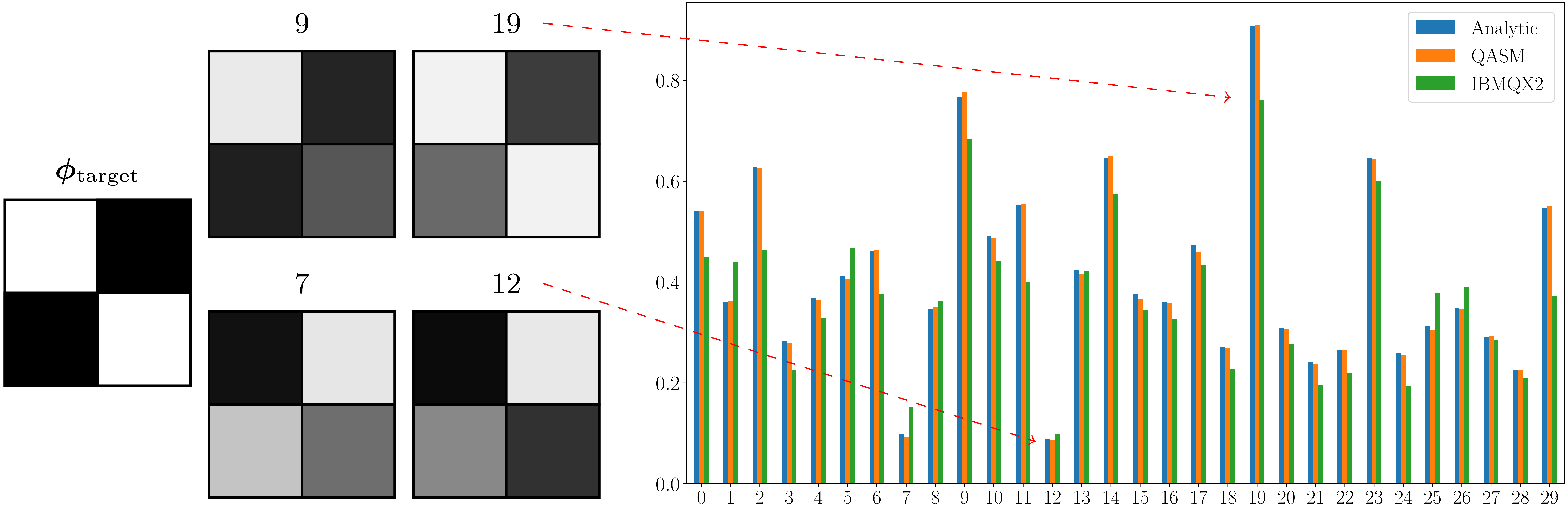
$\vec{i} = (255,170,85,0)$

$i_j \in [0,255]$

255	170
85	0



Normalization $\vec{i} \to \frac{\pi/2}{255}\vec{i}$



Learning

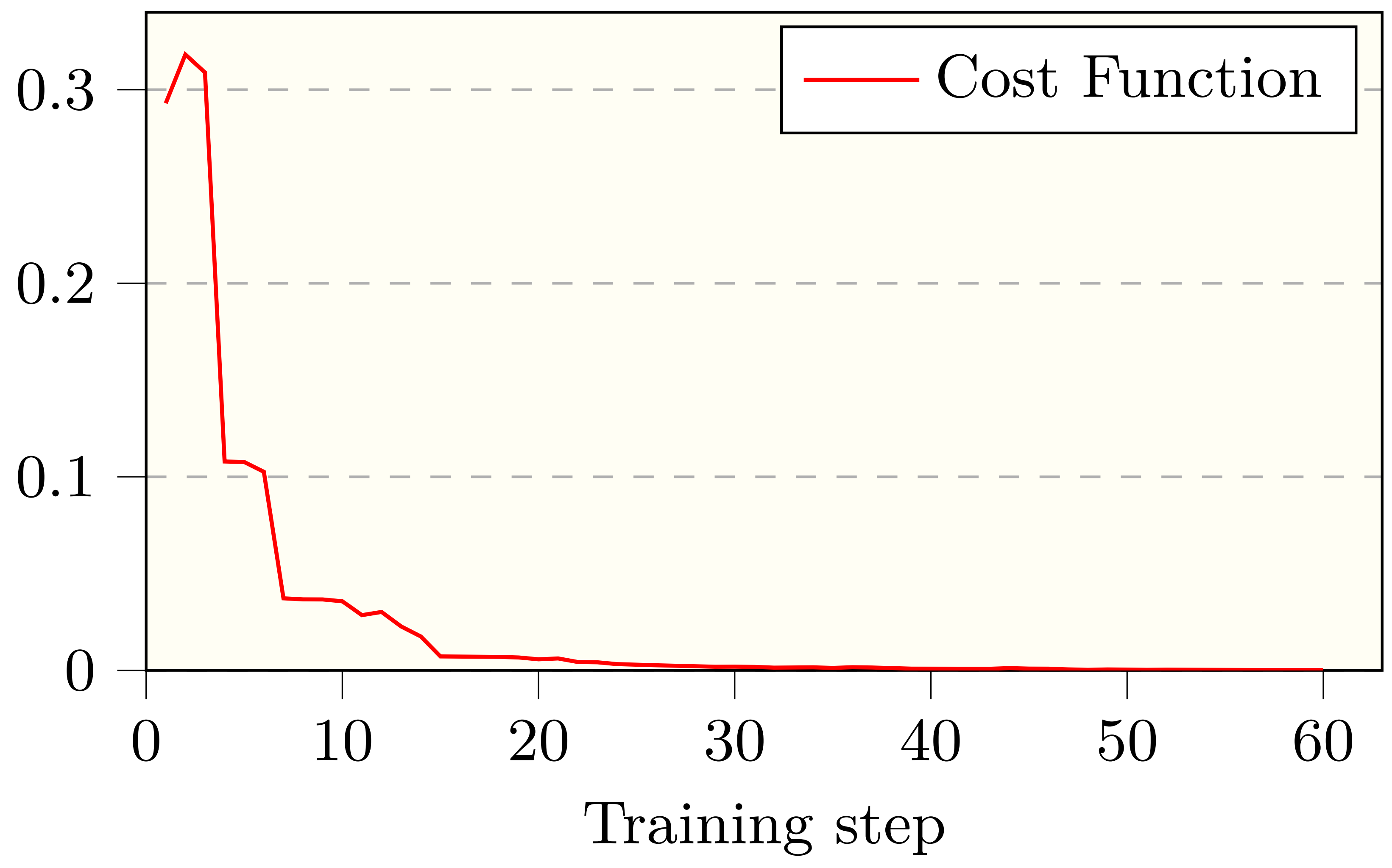
Key aspect of NN-based algorithm is training, which can now be implemented by means of a classical optimizator based on gradient descent.

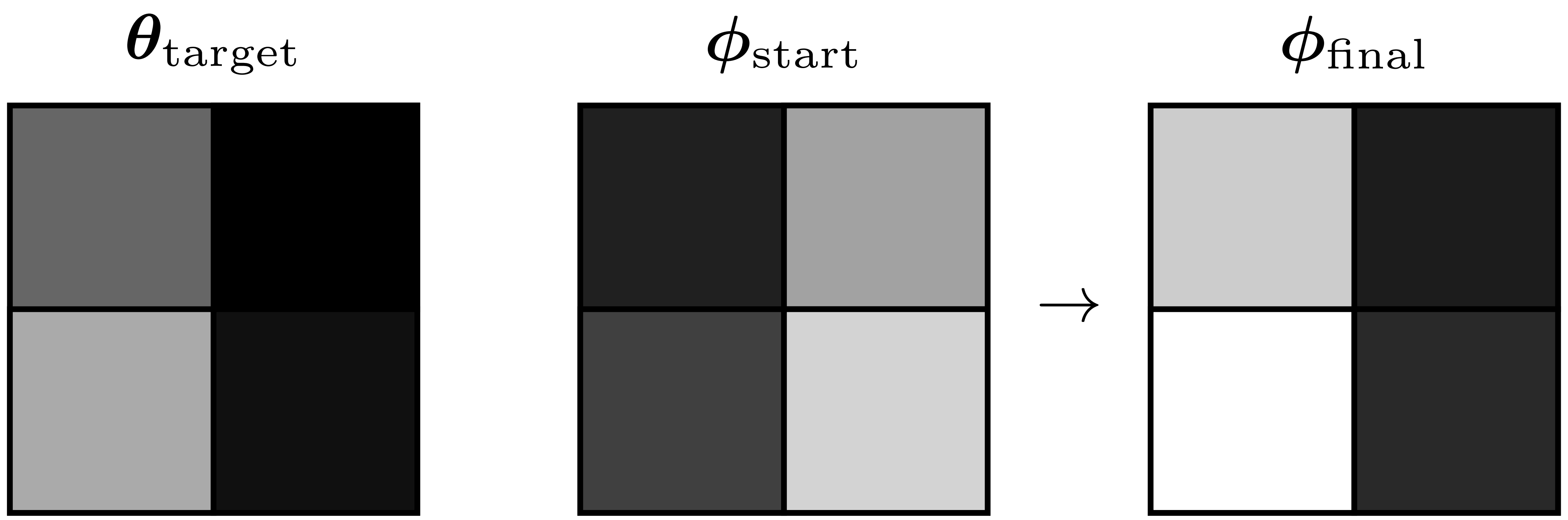
$$\mathscr{L}(\boldsymbol{\phi}) = \sum_{i=0}^{M} (y_i - \tilde{y}_i)^2$$

y_i = correct label

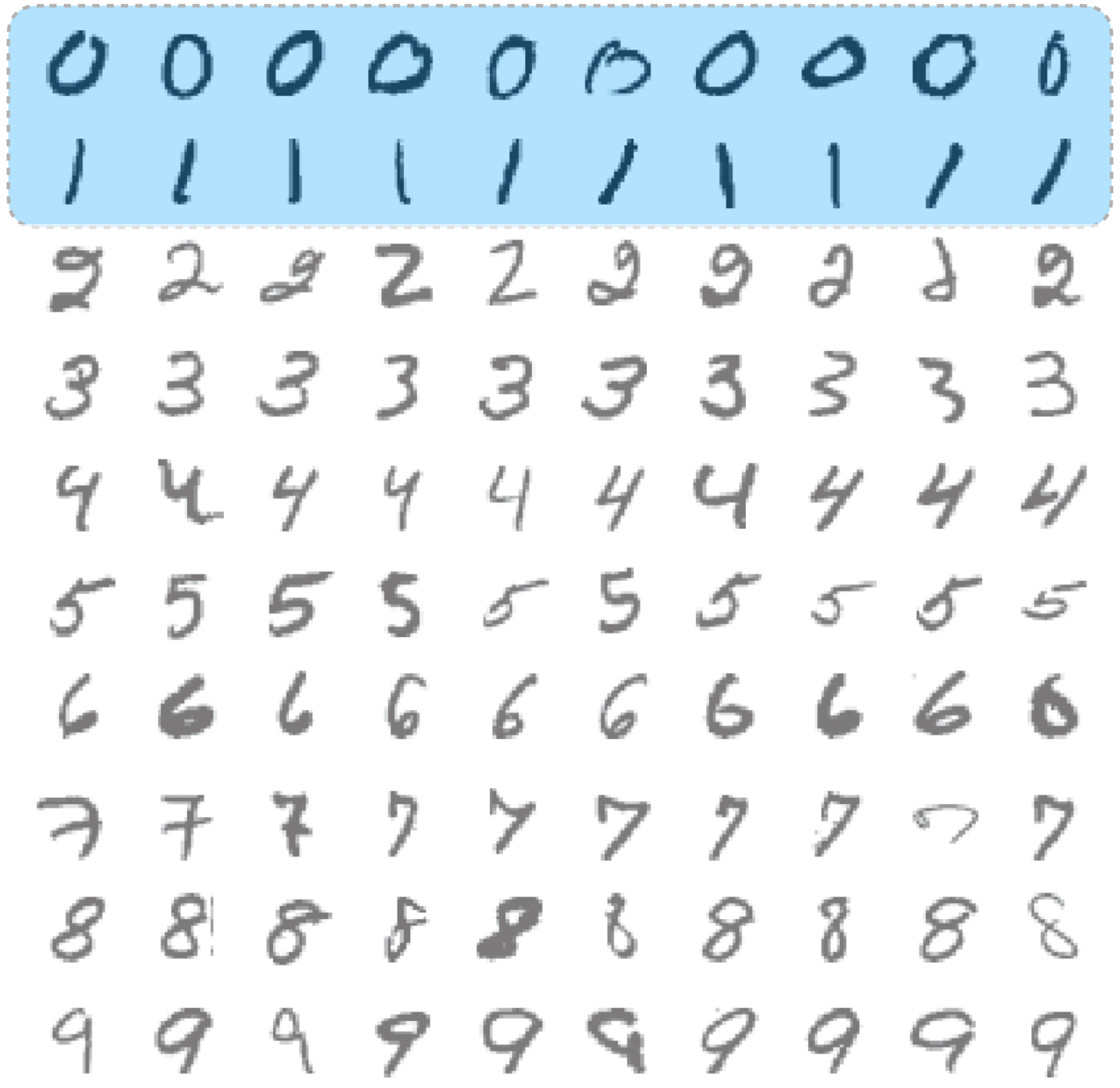
M = size of the dataset

$\tilde{y}_i$ = predicted label ⟶ $\tilde{y}_i = \begin{cases} 1 & \text{if } f(\boldsymbol{\theta}, \boldsymbol{\phi}) > \textit{threshold} \\ 0 & \text{otherwise} \end{cases}$

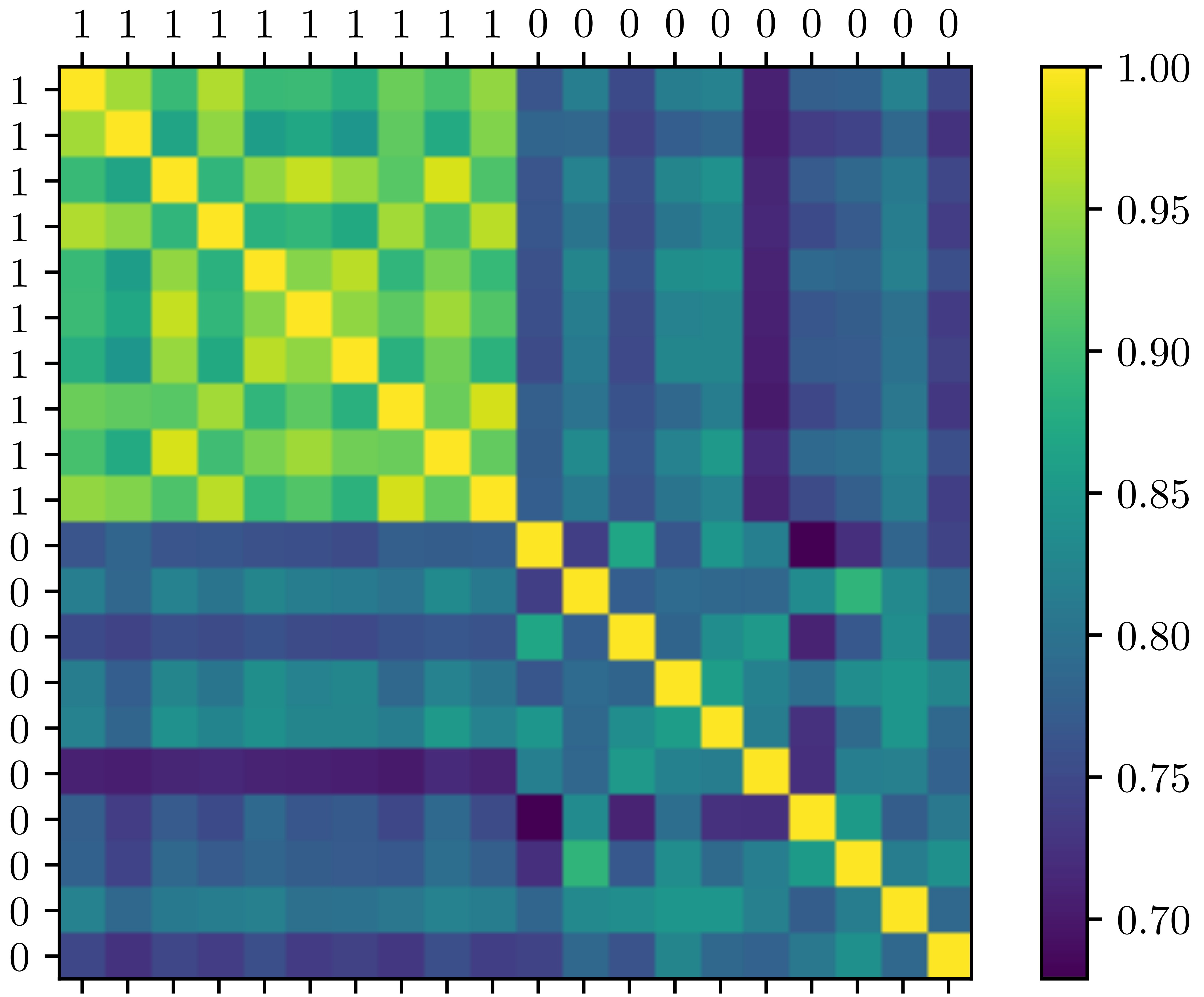




MNIST



$$\left|\langle\psi_w|\psi_i\rangle\right|^2$$



Accuracy $\sim 98\,\%$

Take home message

Key points

- Encode classical data through Phase encoding
- Color invariance and Noise resilience
- Suitable for optimization using gradient descent techniques
- Successfully tested on real quantum hardware (IBMQ)

References

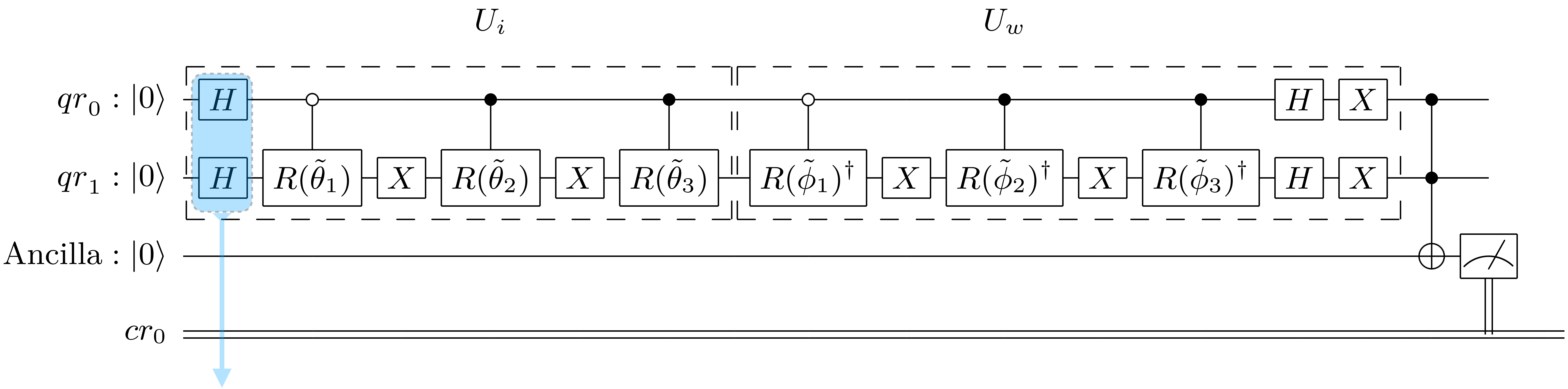
F. Tacchino *et al.*, npj Quantum Information 5, 26 (2019), DOI: https://doi.org/10.1038/s41534-019-0140-4

S. Mangini *et al.*, Machine Learning: Science and Technology (2020), DOI: https://doi.org/10.1088/2632-2153/abaf98

Group members: Chiara Macchiavello, Dario Gerace, Daniele Bajoni (UniPv), Francesco Tacchino (IBM Quantum)

Thank you for the attention!

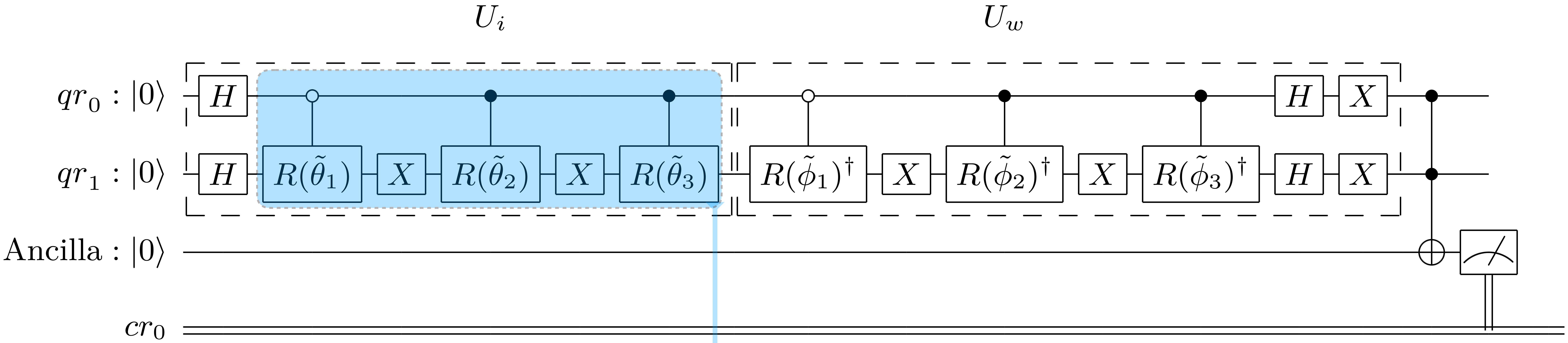
N = 2 qubits



Superposition

$$|00\rangle \rightarrow \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$$

N = 2 qubits



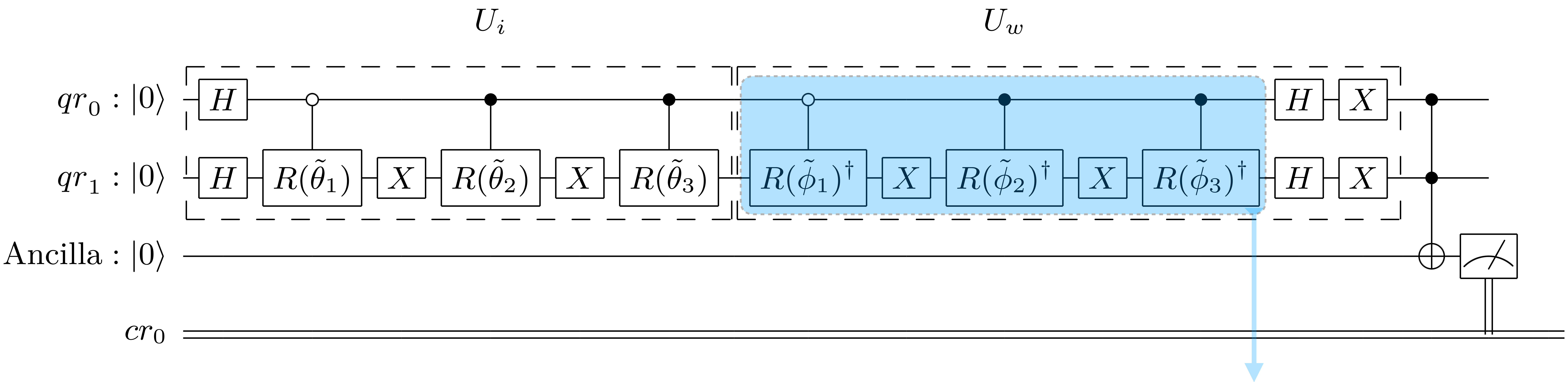
Superposition

$$|00\rangle \to \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$$

Phase encoding

$$\frac{1}{2}\left(00\rangle + e^{i\theta_1}|01\rangle + e^{i\theta_2}|10\rangle + |e^{i\theta_3}11\rangle\right)$$

N = 2 qubits



Superposition

$$|00\rangle \rightarrow \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$$

Weights

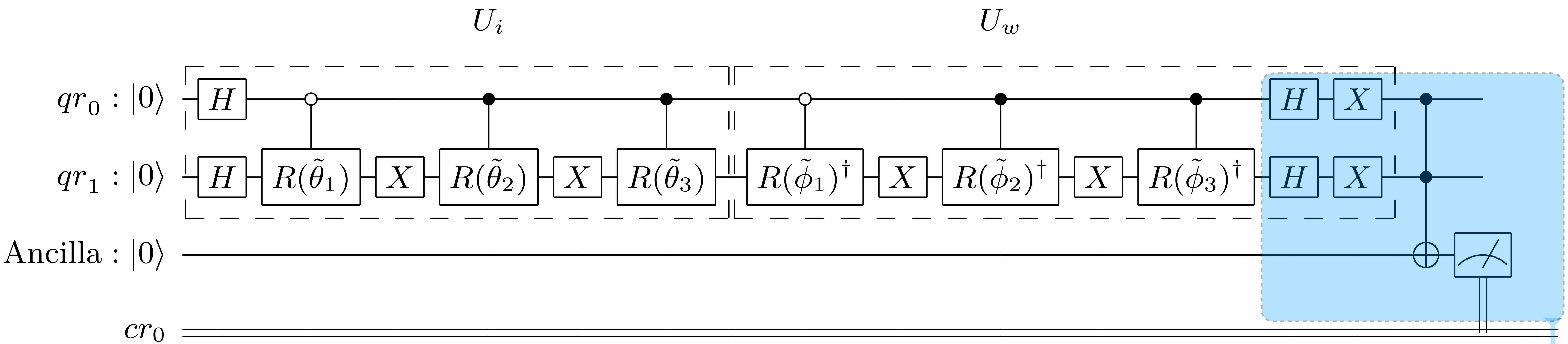
$$\frac{1}{2}\left(|00\rangle + e^{i(\theta_1-\phi_1)}|01\rangle + e^{i(\theta_2-\phi_2)}|10\rangle + |e^{i(\theta_3-\phi_3)}11\rangle\right)$$

Phase encoding

$$\frac{1}{2}\left(00\rangle + e^{i\theta_1}|01\rangle + e^{i\theta_2}|10\rangle + |e^{i\theta_3}11\rangle\right)$$

Actual quantum circuit implementation

N = 2 qubits



Superposition

$$|00\rangle \rightarrow \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$$

Weights

$$\frac{1}{2}\left(|00\rangle + e^{i(\theta_1-\phi_1)}|01\rangle + e^{i(\theta_2-\phi_2)}|10\rangle + |e^{i(\theta_3-\phi_3)}11\rangle\right)$$

Result

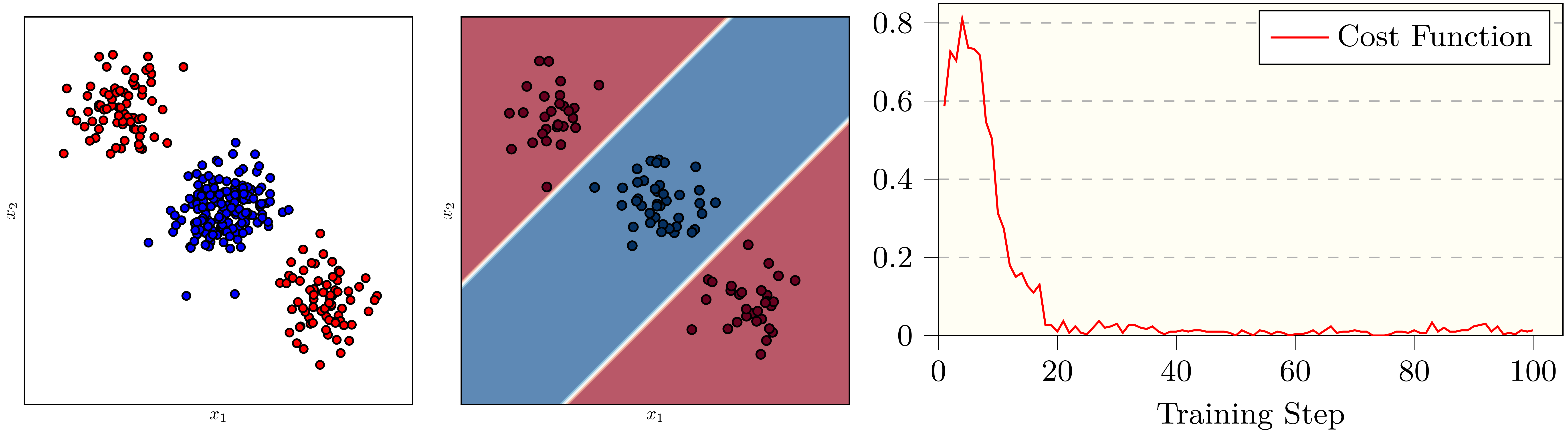
$$\left|\langle\psi_w|\psi_i\rangle\right|^2$$

Phase encoding

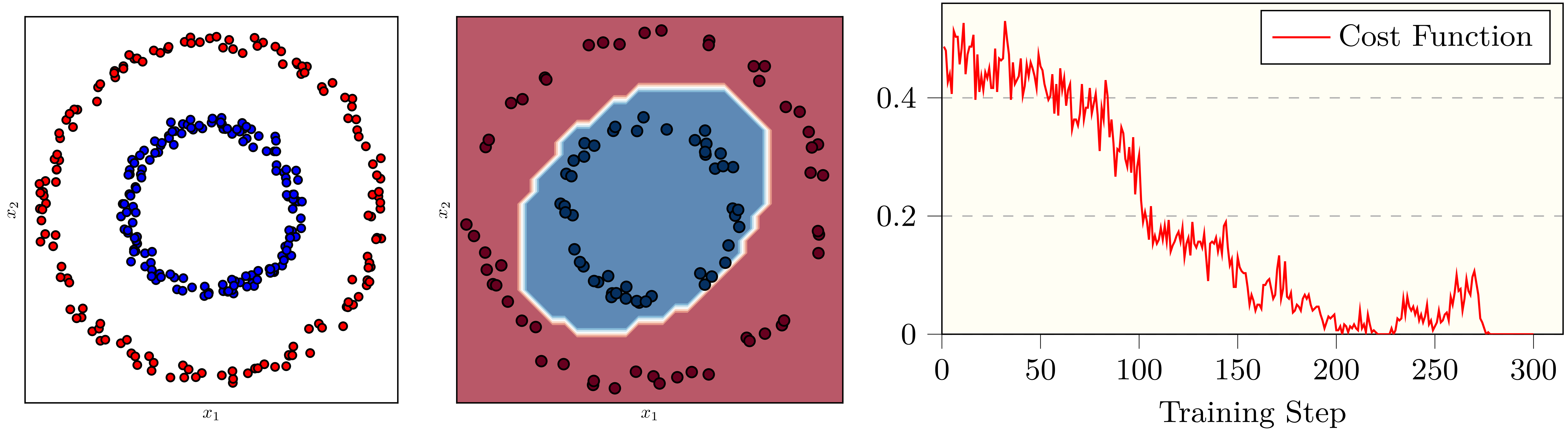
$$\frac{1}{2}\left(00\rangle + e^{i\theta_1}|01\rangle + e^{i\theta_2}|10\rangle + |e^{i\theta_3}11\rangle\right)$$

Classification of 2D data

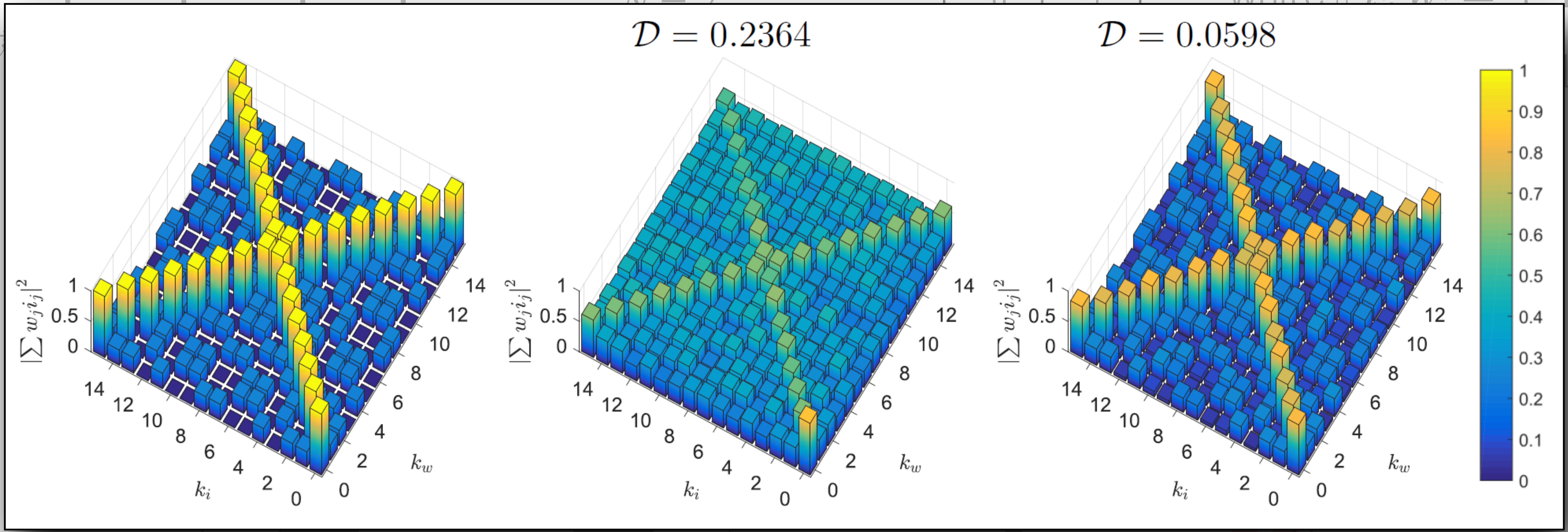
N=1 qubit



N=2 qubit



Classification of checkboard patterns



$\vec{i} = (1,-1,1,-1)$

$\vec{w} = (-1,1,1,-1)$

$\langle\psi_w|\psi_i\rangle = \vec{i}\cdot\vec{w} = 0$

No activation

State preparation

For binary encoding of data

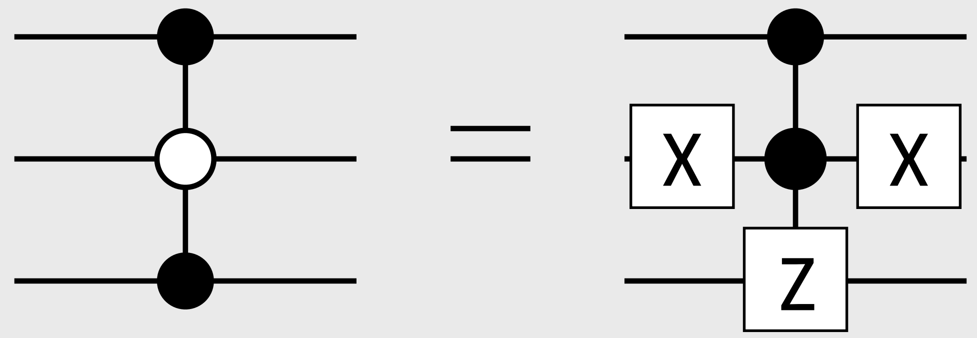
Brute-force approach

⋮

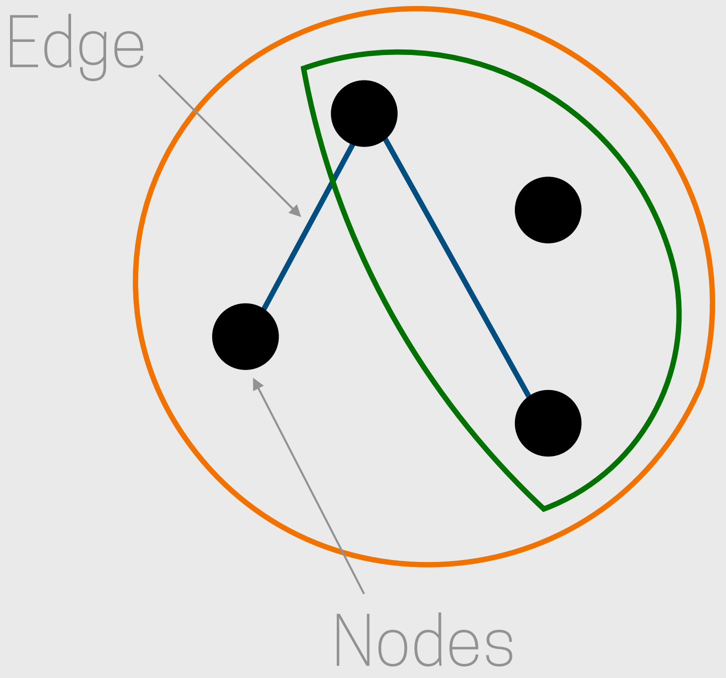
$$|010\rangle \rightarrow i_{010}|010\rangle$$
$$|110\rangle \rightarrow i_{110}|110\rangle$$

⋮

Requires $O(n)$ operations



Hypergraph states



$$\longleftrightarrow \quad |\psi_i\rangle = \frac{1}{\sqrt{2^N}} \sum_{j=0}^{2^N-1} i_j |j\rangle$$

REW state

Requires $O(n)$ operations
but lower multi-qubit operations
(at most one N-controlled gate)