Efficient estimation of trainability for Variational Quantum Algorithms

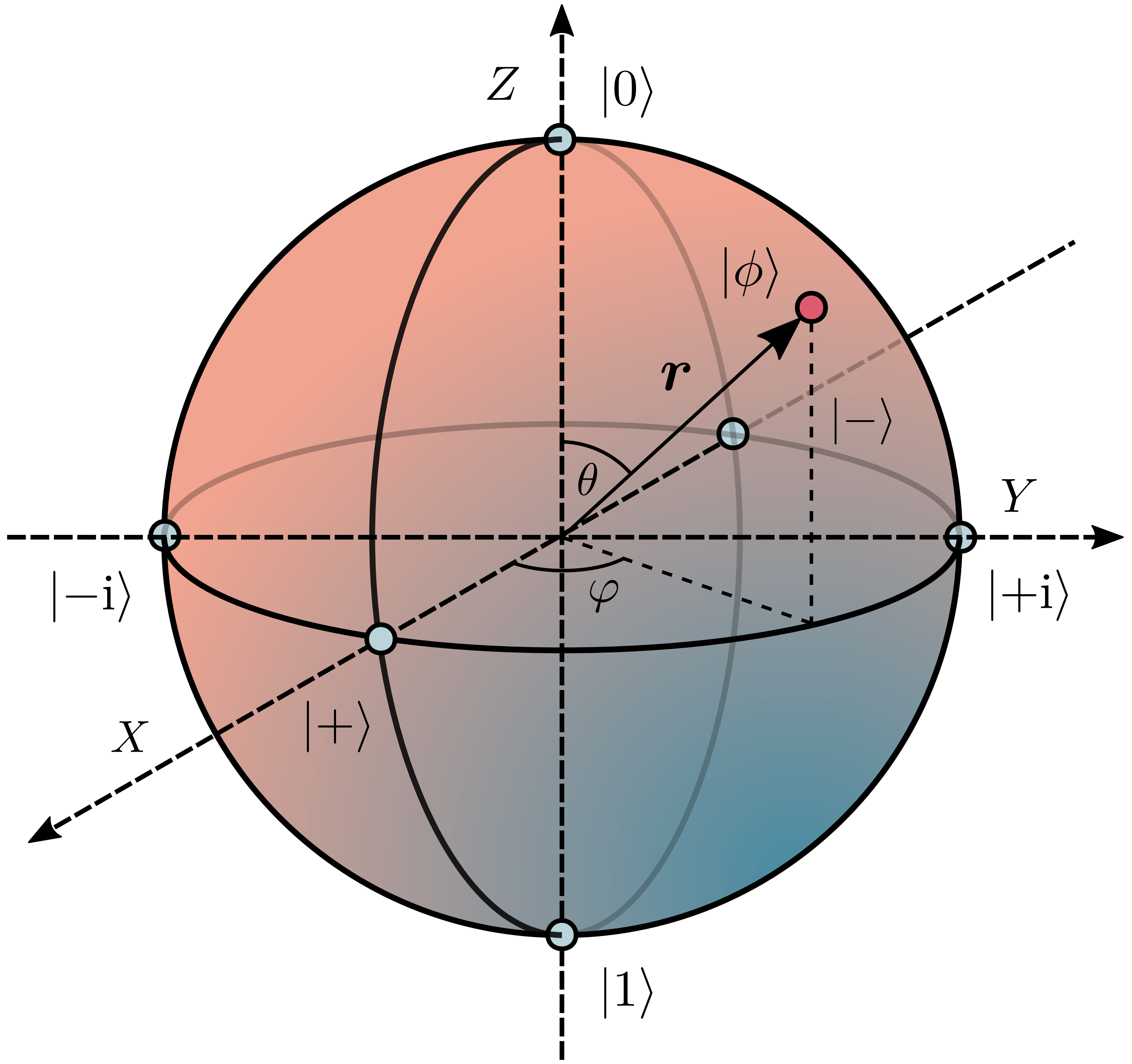
Congrès Général des 150 ans de la SFP
MC10 Physique et Intelligence Artificielle
06.07.2023

Valentin Heyraud

Lab. Matériaux et Phénomènes Quantiques - Université Paris Cité

"Efficient estimation of trainability for Variational Quantum Circuits",
V. Heyraud, Z. Li, K. Donatella, A. Le Boité, and C. Ciuti, arXiv:2302.04649

MPQ Université Paris Cité SIR TEQ cnrs



Summary

1. Variational Quantum Algorithms
2. Trainability and Barren Plateaus
3. Efficient estimation of the trainability
4. Perspectives

Variational Quantum Algorithms

Example: the Variational Quantum Eigensolver

McClean et al., “The theory of variational hybrid quantum-classical algorithms”, *New J. of Phys.* (2016)
Peruzzo et al., “A variational eigenvalue solver on a photonics quantum processor”, *Nat. Commun.* (2014)

Variational Quantum Algorithms

Example: the Variational Quantum Eigensolver

- Objective: Finding the ground state of a given N-spins Hamiltonian

$$\hat{H} = \sum_{\alpha} w_\alpha \hat{P}_\alpha \quad \text{with} \quad \hat{P}_\alpha \in \left\{ \hat{1}, \hat{X}, \hat{Y}, \hat{Z} \right\}^{\otimes N}$$

McClean et al., "The theory of variational hybrid quantum-classical algorithms", *New J. of Phys.* (2016)
Peruzzo et al., "A variational eigenvalue solver on a photonics quantum processor", *Nat. Commun.* (2014)

Variational Quantum Algorithms

Example: the Variational Quantum Eigensolver

- Objective: Finding the ground state of a given N-spins Hamiltonian

$$\hat{H} = \sum_{\alpha} w_\alpha \hat{P}_\alpha \quad \text{with} \quad \hat{P}_\alpha \in \left\{ \hat{1}, \hat{X}, \hat{Y}, \hat{Z} \right\}^{\otimes N}$$

- Variational state

$$|\phi_{\boldsymbol{\theta}}\rangle = \hat{U}\left(\boldsymbol{\theta}\right)|0\rangle^{\otimes N},$$

McClean et al., “The theory of variational hybrid quantum-classical algorithms”, *New J. of Phys.* (2016)
Peruzzo et al., “A variational eigenvalue solver on a photonics quantum processor”, *Nat. Commun.* (2014)

Variational Quantum Algorithms

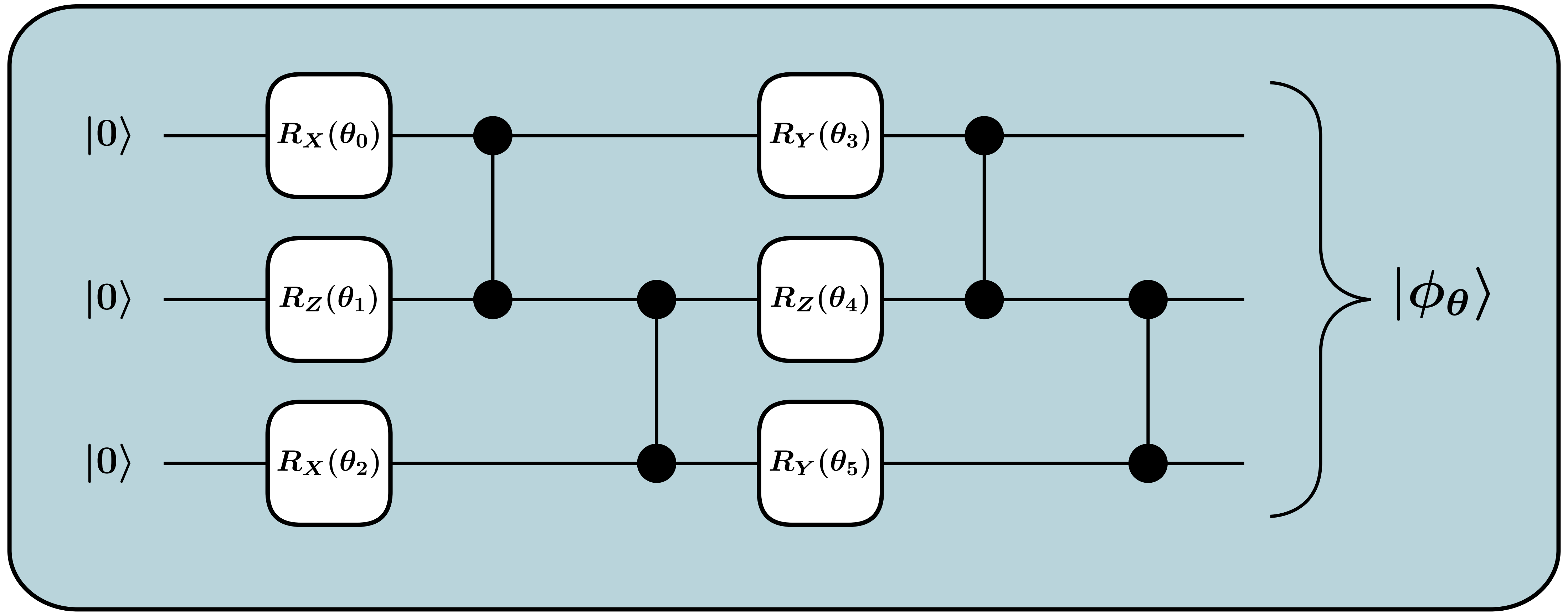
Example: the Variational Quantum Eigensolver

- Objective: Finding the ground state of a given N-spins Hamiltonian

$$\hat{H} = \sum_{\alpha} w_\alpha \hat{P}_\alpha \quad \text{with} \quad \hat{P}_\alpha \in \left\{ \hat{1}, \hat{X}, \hat{Y}, \hat{Z} \right\}^{\otimes N}$$

- Variational state

$$|\phi_{\boldsymbol{\theta}}\rangle = \hat{U}\left(\boldsymbol{\theta}\right) |0\rangle^{\otimes N},$$



Hardware Efficient Ansatz circuit

McClean et al., “The theory of variational hybrid quantum-classical algorithms”, *New J. of Phys.* (2016)
Peruzzo et al., “A variational eigenvalue solver on a photonics quantum processor”, *Nat. Commun.* (2014)

Variational Quantum Algorithms

Example: the Variational Quantum Eigensolver

- Objective: Finding the ground state of a given N-spins Hamiltonian

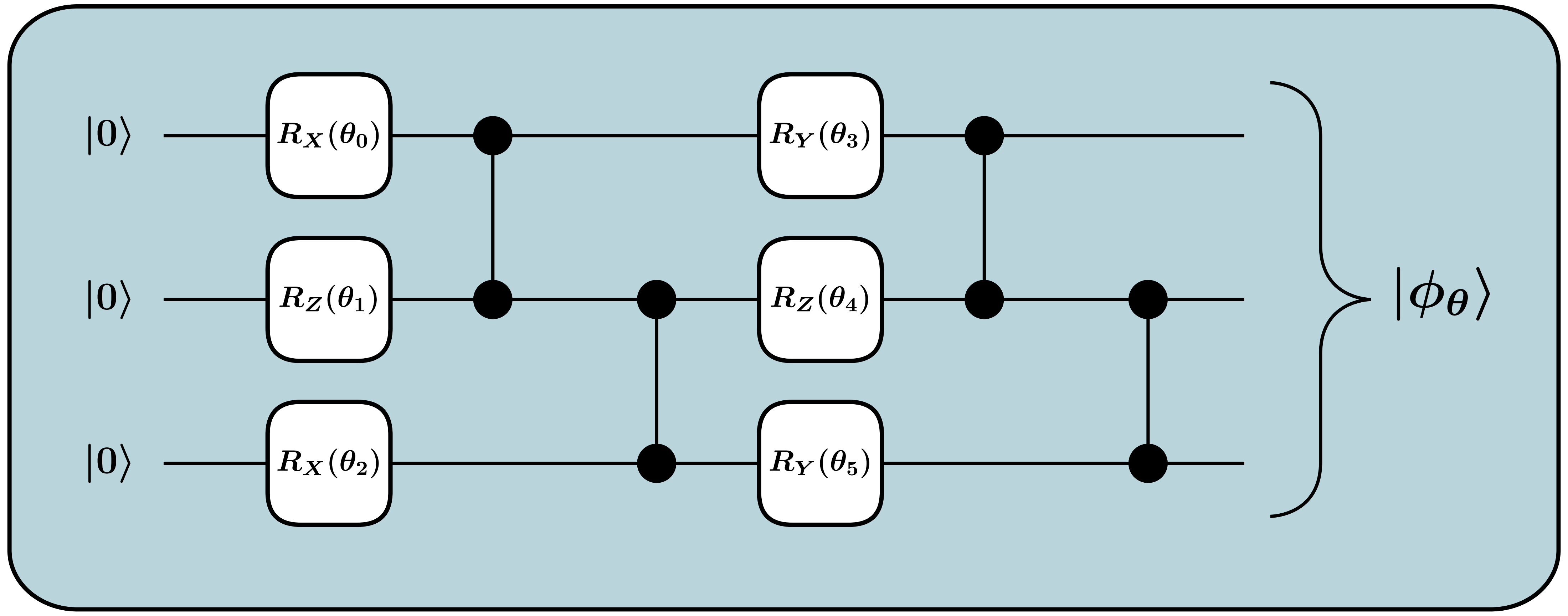
$$\hat{H} = \sum_{\alpha} w_\alpha \hat{P}_\alpha \quad \text{with} \quad \hat{P}_\alpha \in \left\{ \hat{1}, \hat{X}, \hat{Y}, \hat{Z} \right\}^{\otimes N}$$

- Variational state

$$|\phi_{\boldsymbol{\theta}}\rangle = \hat{U}\left(\boldsymbol{\theta}\right)|0\rangle^{\otimes N},$$

- Parameterized unitary

$$\hat{U}\left(\boldsymbol{\theta}\right) = \prod_{i=1}^{M} \left(e^{-\mathrm{i}\frac{\theta_\mathrm{i}}{2}\hat{\mathrm{P}}_\mathrm{i}} \right) \hat{W}_i \ , \qquad \hat{P}_i \in \left\{ \hat{X}, \hat{Y}, \hat{Z} \right\}$$



Hardware Efficient Ansatz circuit

McClean et al., “The theory of variational hybrid quantum-classical algorithms”, *New J. of Phys.* (2016)
Peruzzo et al., “A variational eigenvalue solver on a photonics quantum processor”, *Nat. Commun.* (2014)

Variational Quantum Algorithms

Training of the ansatz

Variational Quantum Algorithms

Training of the ansatz

- Minimization of the energy

$$E\left(\boldsymbol{\theta}\right) = \langle \phi_{\boldsymbol{\theta}} | \hat{H} | \phi_{\boldsymbol{\theta}} \rangle , \quad \boldsymbol{\theta}* = \text{argmin}\, E\left(\boldsymbol{\theta}\right)$$

Variational Quantum Algorithms

Training of the ansatz

- Minimization of the energy

  $$E\left(\boldsymbol{\theta}\right) = \langle \phi_{\boldsymbol{\theta}} | \hat{H} | \phi_{\boldsymbol{\theta}} \rangle, \quad \boldsymbol{\theta}* = \mathrm{argmin}\, E\left(\boldsymbol{\theta}\right)$$

- Initialisation: random choice $\boldsymbol{\theta}_0 \sim \mathbb{P}\left(\mathrm{d}\boldsymbol{\theta}_0\right)$

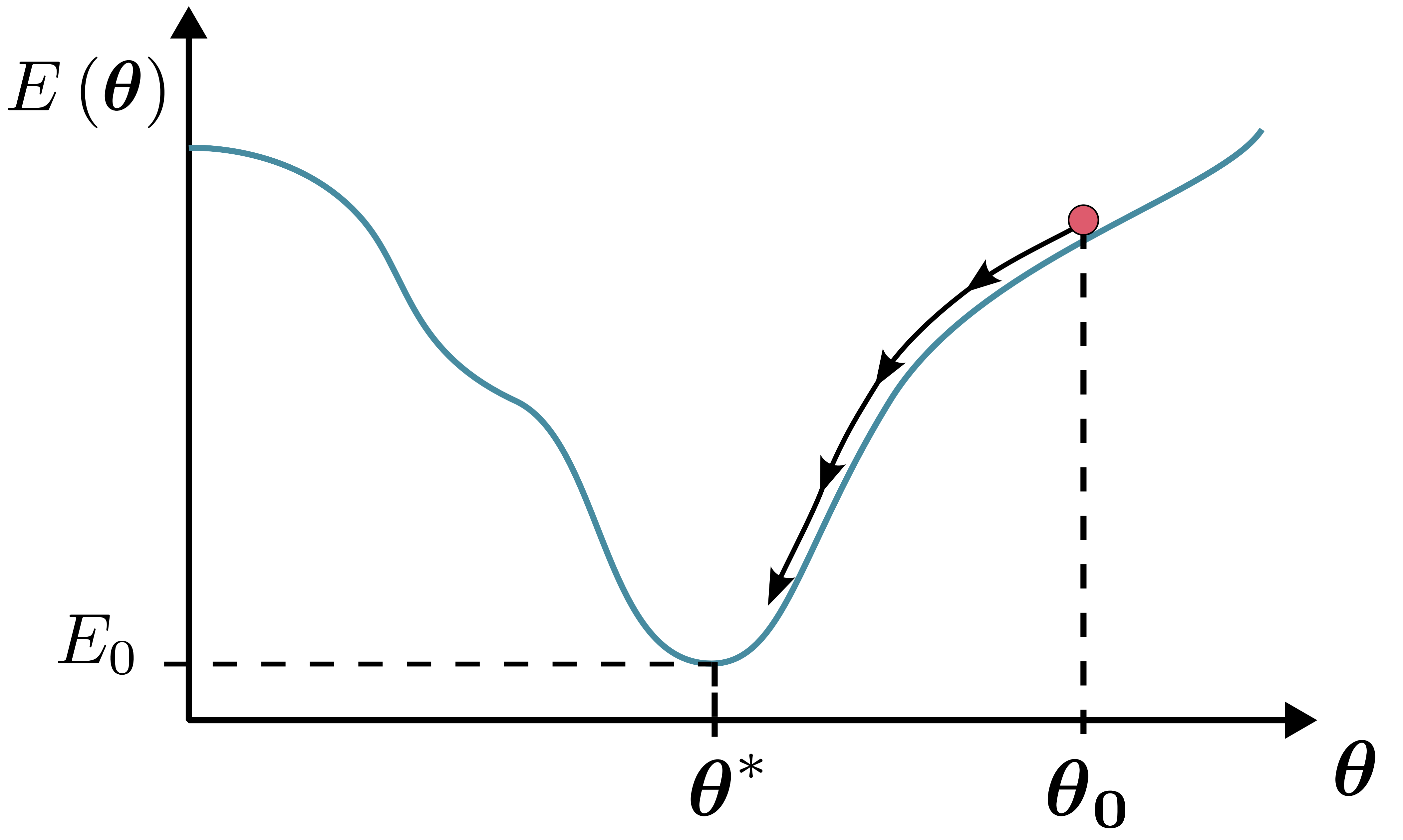
Variational Quantum Algorithms

Training of the ansatz

- Minimization of the energy

  $E\left(\boldsymbol{\theta}\right) = \langle \phi_{\boldsymbol{\theta}} | \hat{H} | \phi_{\boldsymbol{\theta}} \rangle , \quad \boldsymbol{\theta}^{*} = \operatorname{argmin} E\left(\boldsymbol{\theta}\right)$

- Initialisation: random choice $\boldsymbol{\theta}_0 \sim \mathbb{P}\left(\mathrm{d}\boldsymbol{\theta}_0\right)$



Gradient descent in the energy landscape

Variational Quantum Algorithms

Training of the ansatz

- Minimization of the energy

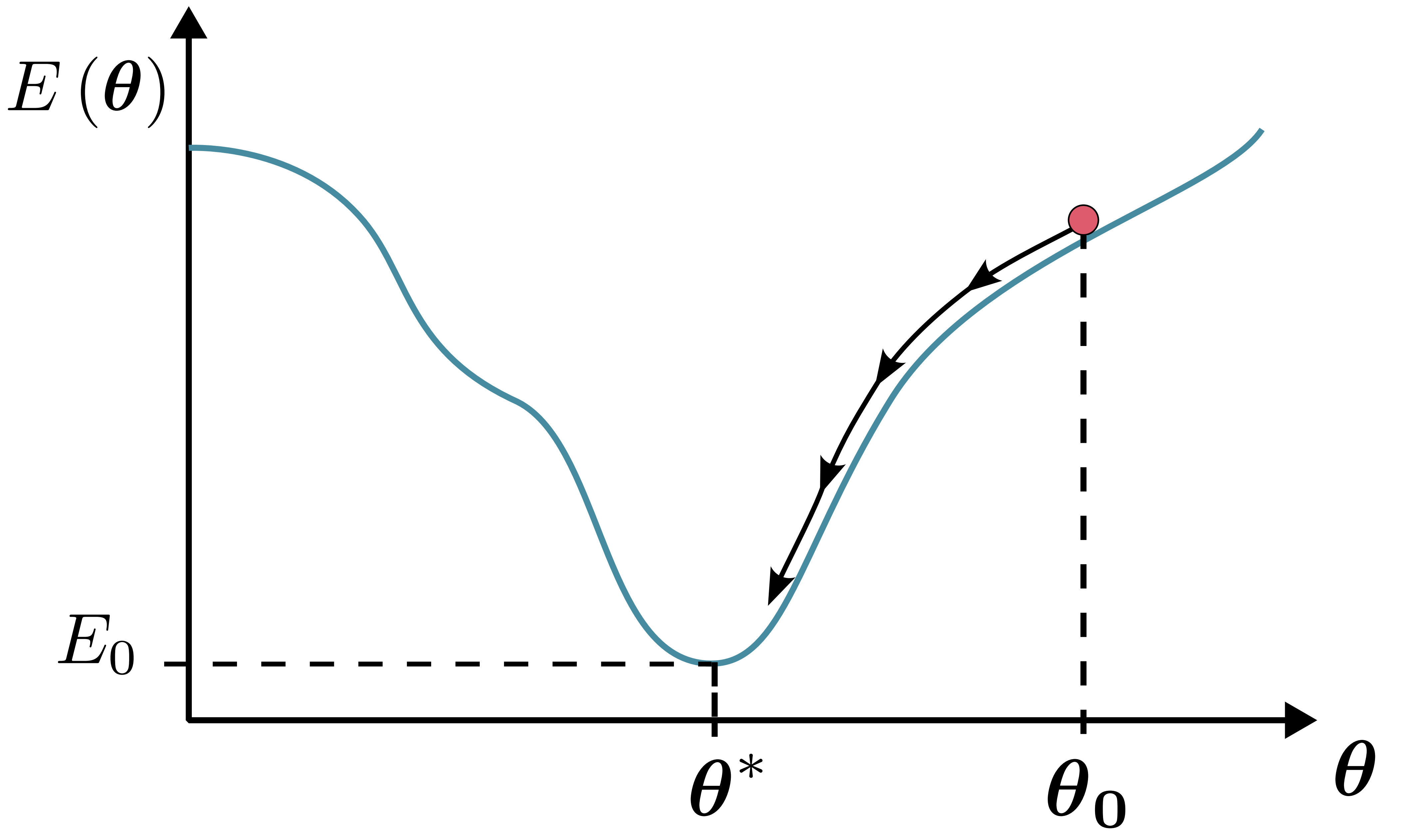
  $$E\left(\boldsymbol{\theta}\right) = \langle \phi_{\boldsymbol{\theta}} | \hat{H} | \phi_{\boldsymbol{\theta}} \rangle , \quad \boldsymbol{\theta}^* = \operatorname{argmin} E\left(\boldsymbol{\theta}\right)$$

- Initialisation: random choice $\boldsymbol{\theta}_0 \sim \mathbb{P}\left(\mathrm{d}\boldsymbol{\theta}_0\right)$

- Gradient descent

$$\boldsymbol{\theta}_{k+1} \leftarrow \boldsymbol{\theta}_k - \eta \nabla_{\boldsymbol{\theta}} E\left(\boldsymbol{\theta}_k\right),$$

$$\partial_i E\left(\boldsymbol{\theta}\right) = \frac{1}{2}\left(E\left(\boldsymbol{\theta} + \frac{\pi}{2}\boldsymbol{e}_i\right) - E\left(\boldsymbol{\theta} - \frac{\pi}{2}\boldsymbol{e}_i\right)\right)$$



Gradient descent in the energy landscape

Trainability and Barren Plateaus

The Barren Plateaus phenomenon

McClean et al., “Barren plateaus in quantum neural network training landscapes”, *Nat. Commun.* (2018)

Trainability and Barren Plateaus

The Barren Plateaus phenomenon

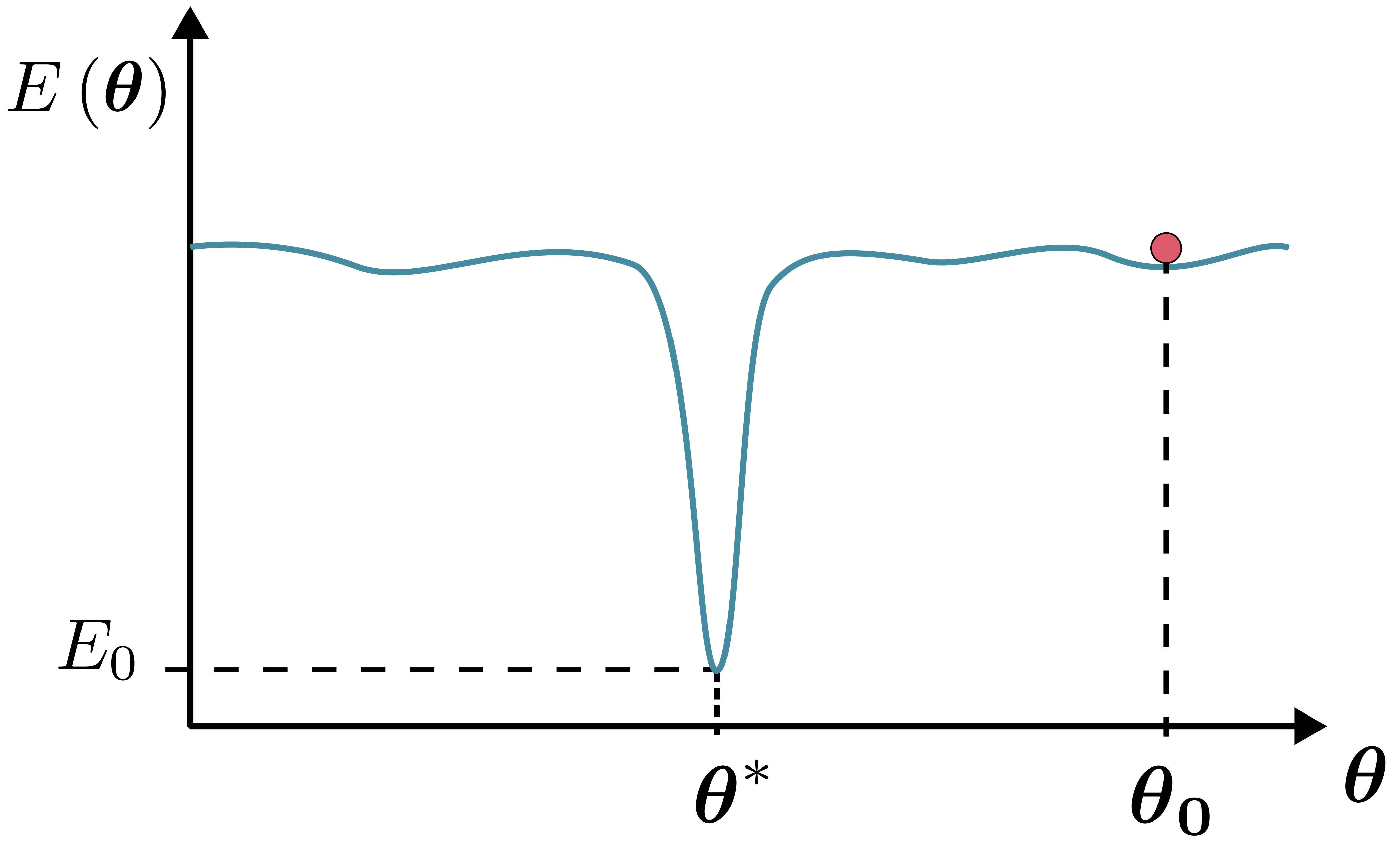
- Exponential vanishing of the energy gradient:

McClean et al., “Barren plateaus in quantum neural network training landscapes”, *Nat. Commun.* (2018)

Trainability and Barren Plateaus

The Barren Plateaus phenomenon

- Exponential vanishing of the energy gradient:



Flat energy landscape

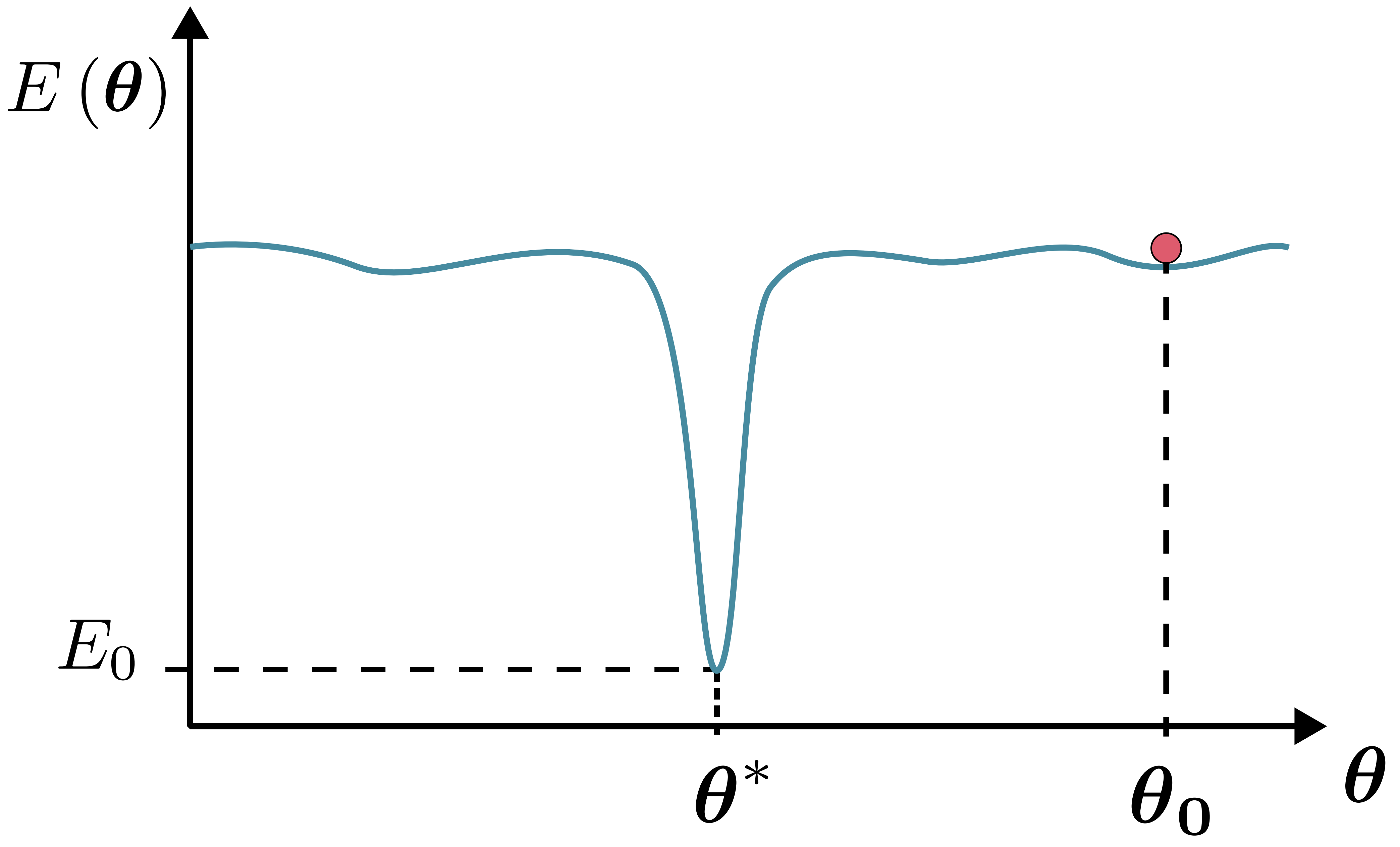
McClean et al., “Barren plateaus in quantum neural network training landscapes”, *Nat. Commun.* (2018)

Trainability and Barren Plateaus

The Barren Plateaus phenomenon

- Exponential vanishing of the energy gradient:

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \sim \mathcal{O}\left(\frac{1}{2^N}\right) \implies \mathbb{P}\left(\left|\partial_i E\left(\boldsymbol{\theta}\right)\right| \geq \epsilon\right) \leq \mathcal{O}\left(\frac{\epsilon^2}{2^N}\right)$$



Flat energy landscape

McClean et al., “Barren plateaus in quantum neural network training landscapes”, *Nat. Commun.* (2018)

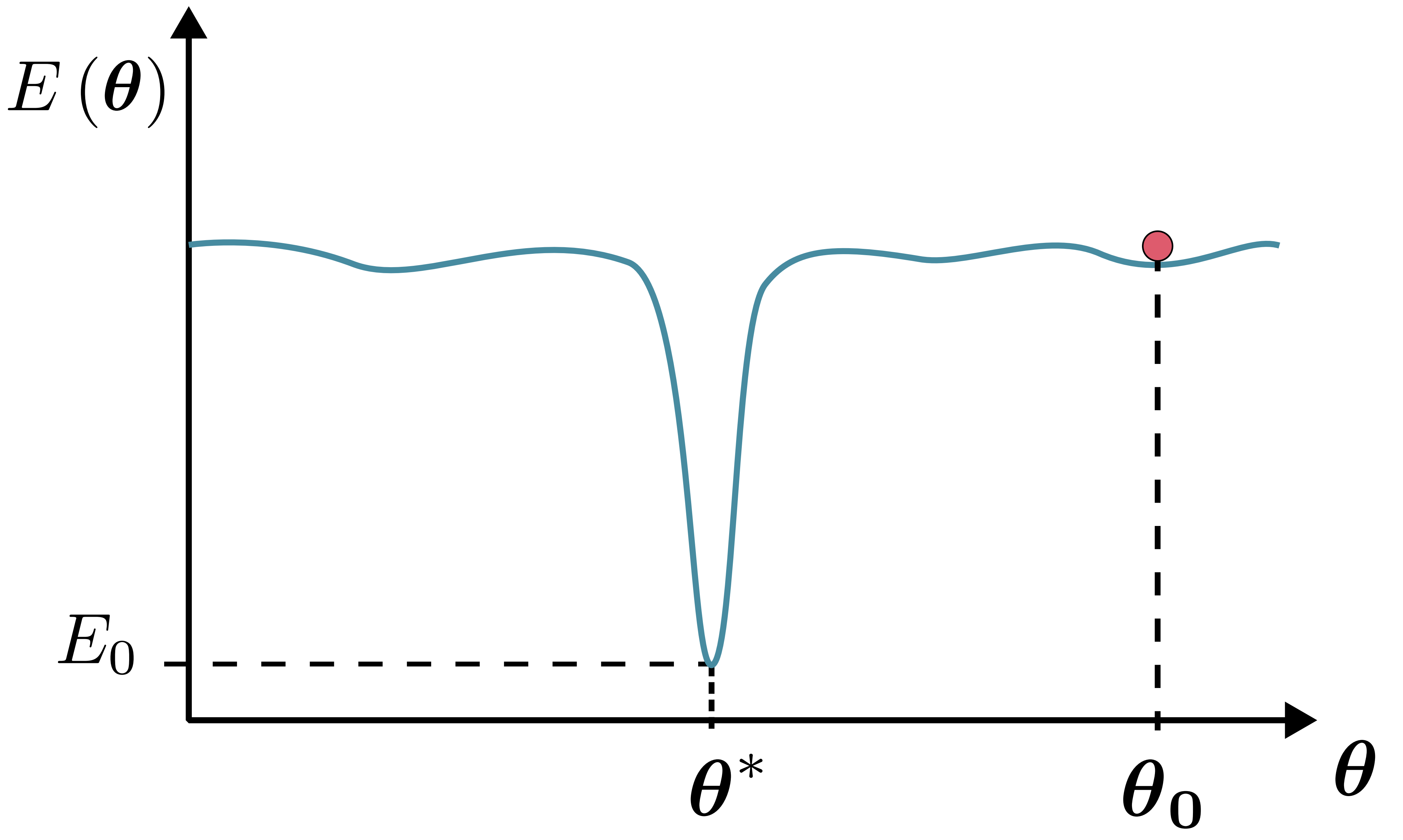
Trainability and Barren Plateaus

The Barren Plateaus phenomenon

- Exponential vanishing of the energy gradient:

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \sim \mathcal{O}\left(\frac{1}{2^N}\right) \implies \mathbb{P}\left(\left|\partial_i E\left(\boldsymbol{\theta}\right)\right| \geq \epsilon\right) \leq \mathcal{O}\left(\frac{\epsilon^2}{2^N}\right)$$

- Ansatz difficult to train!



Flat energy landscape

McClean et al., "Barren plateaus in quantum neural network training landscapes", *Nat. Commun.* (2018)

Trainability and Barren Plateaus

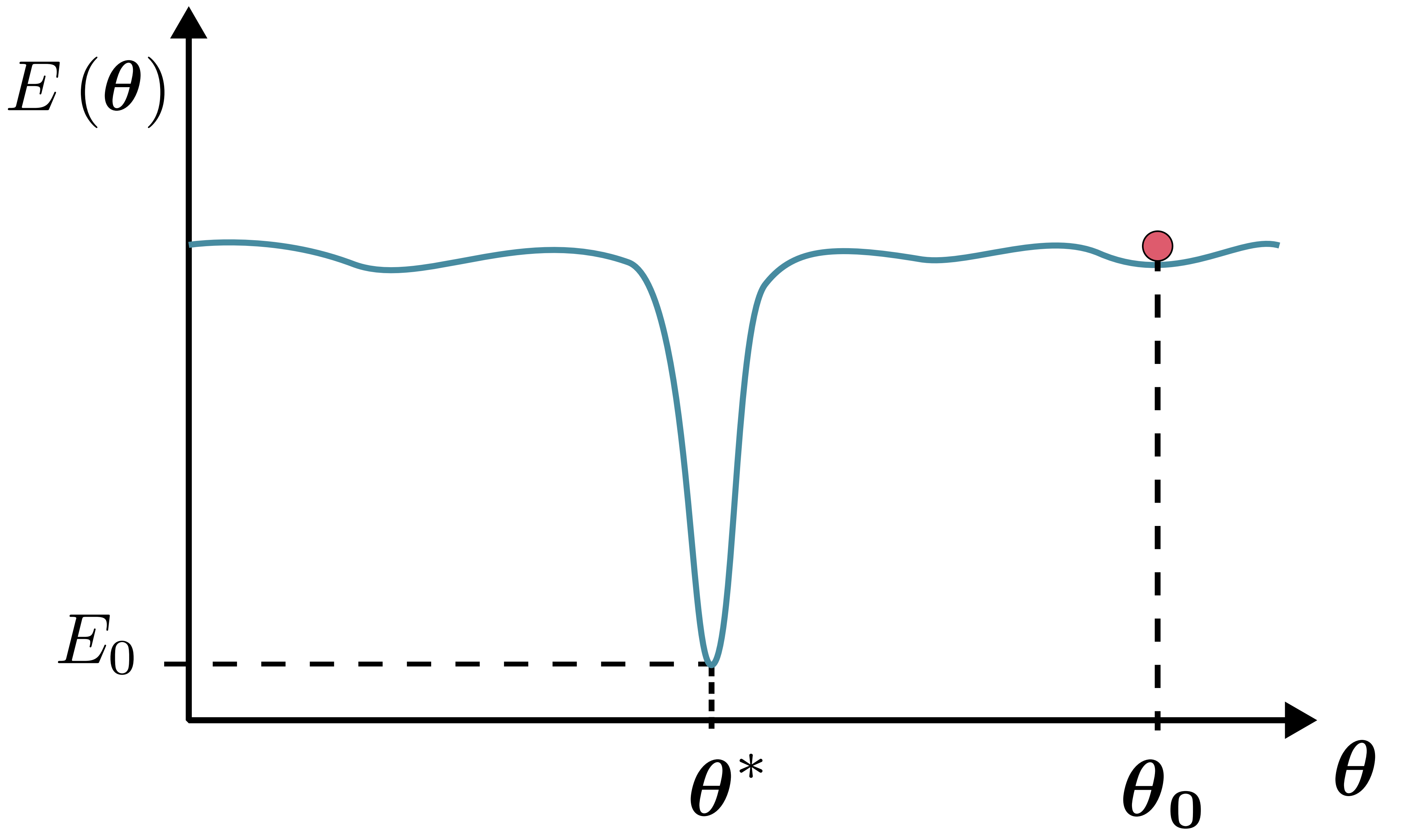
The Barren Plateaus phenomenon

- Exponential vanishing of the energy gradient:

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \sim \mathcal{O}\left(\frac{1}{2^N}\right) \implies \mathbb{P}\left(\left|\partial_i E\left(\boldsymbol{\theta}\right)\right| \geq \epsilon\right) \leq \mathcal{O}\left(\frac{\epsilon^2}{2^N}\right)$$

- Ansatz difficult to train!

$$\nabla E\left(\boldsymbol{\theta}\right) \simeq 0 \implies \boldsymbol{\theta}_{k+1} \simeq \boldsymbol{\theta}_k$$



Flat energy landscape

McClean et al., "Barren plateaus in quantum neural network training landscapes", *Nat. Commun.* (2018)

Trainability and Barren Plateaus

The Barren Plateaus phenomenon

- Exponential vanishing of the energy gradient:

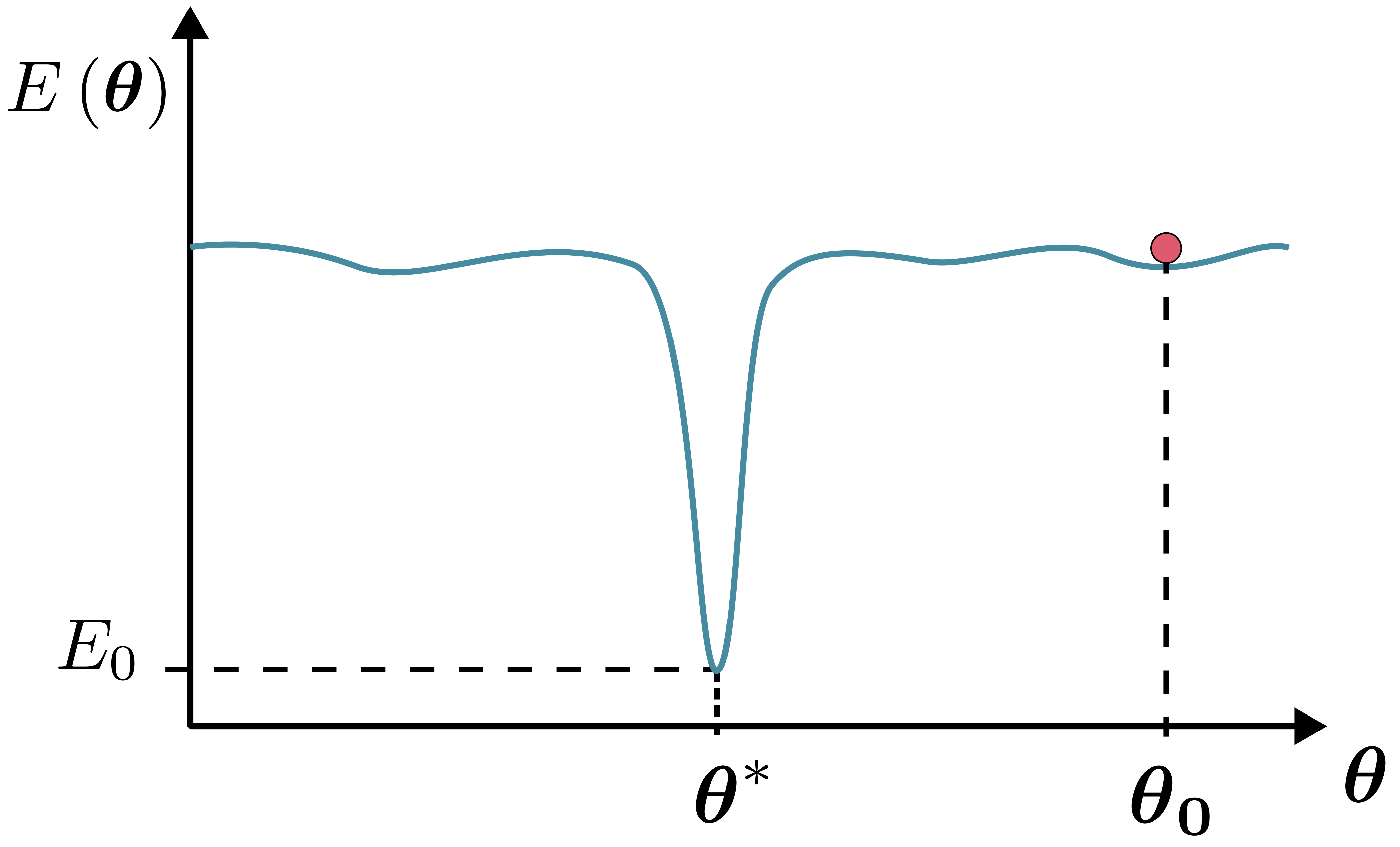
$$\mathbb{E}_{\boldsymbol{\theta}}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \sim \mathcal{O}\left(\frac{1}{2^N}\right) \implies \mathbb{P}\left(\left|\partial_i E\left(\boldsymbol{\theta}\right)\right| \geq \epsilon\right) \leq \mathcal{O}\left(\frac{\epsilon^2}{2^N}\right)$$

- Ansatz difficult to train!

$$\nabla E\left(\boldsymbol{\theta}\right) \simeq 0 \implies \boldsymbol{\theta}_{k+1} \simeq \boldsymbol{\theta}_k$$

- Multiple origins:

  Noise, expressivity, entanglement, global cost functions …



Flat energy landscape

McClean et al., “Barren plateaus in quantum neural network training landscapes”, *Nat. Commun.* (2018)

Trainability and Barren Plateaus

How to estimate the trainability ?

Trainability and Barren Plateaus

How to estimate the trainability ?

- Avoiding barren plateaus: Specific ansatz architectures and initialization methods

Trainability and Barren Plateaus

How to estimate the trainability ?

- Avoiding barren plateaus: Specific ansatz architectures and initialization methods
- Estimating the average gradient amplitude

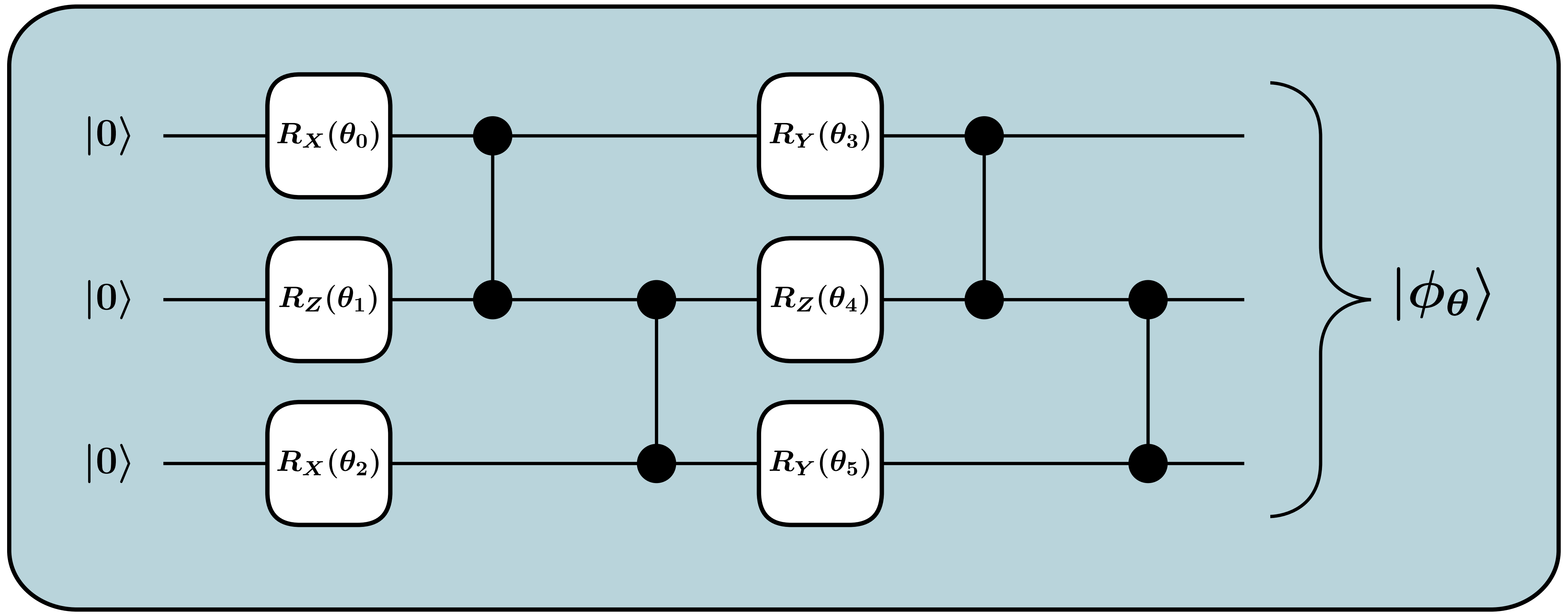
$$\mathbb{E}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \simeq \frac{1}{K}\sum_{j=1}^{K} \partial_i E\left(\boldsymbol{\theta}_j\right)^2$$

Trainability and Barren Plateaus

How to estimate the trainability ?

- Avoiding barren plateaus: Specific ansatz architectures and initialization methods
- Estimating the average gradient amplitude

$$\mathbb{E}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \simeq \frac{1}{K}\sum_{j=1}^{K} \partial_i E\left(\boldsymbol{\theta}_j\right)^2$$

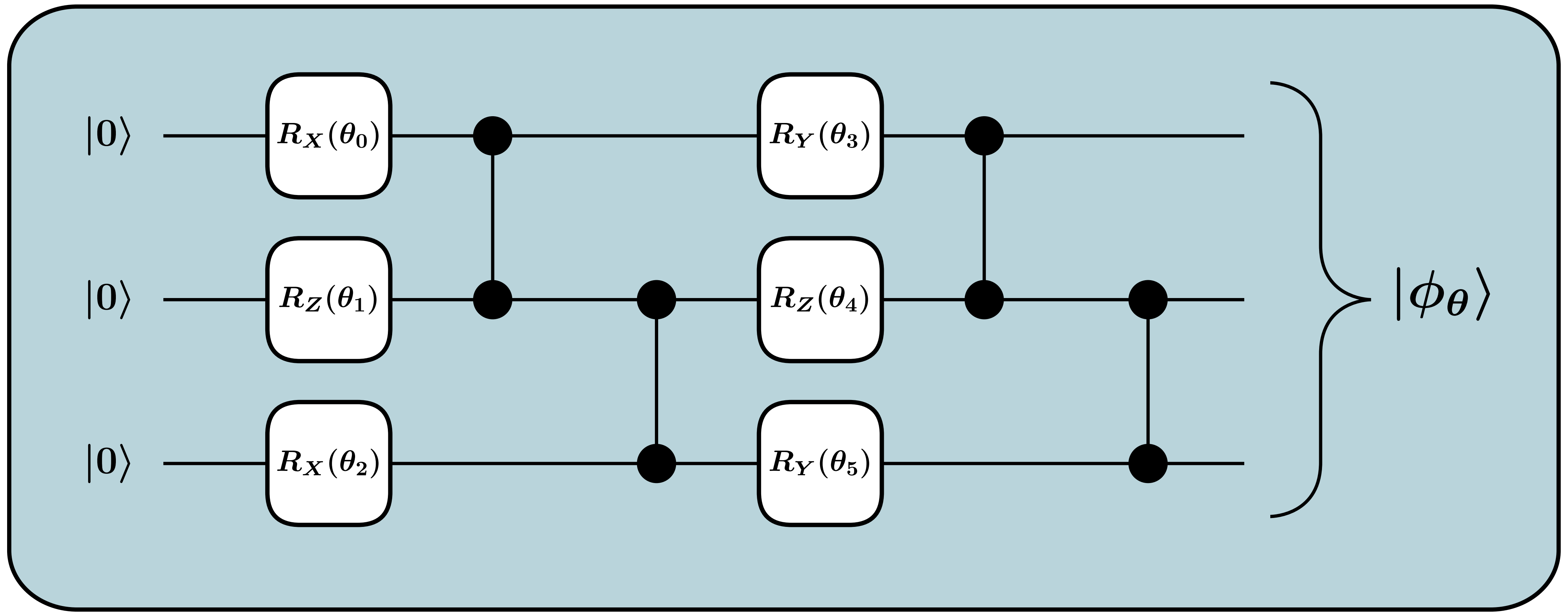


Trainability and Barren Plateaus

How to estimate the trainability ?

- Avoiding barren plateaus: Specific ansatz architectures and initialization methods
- Estimating the average gradient amplitude

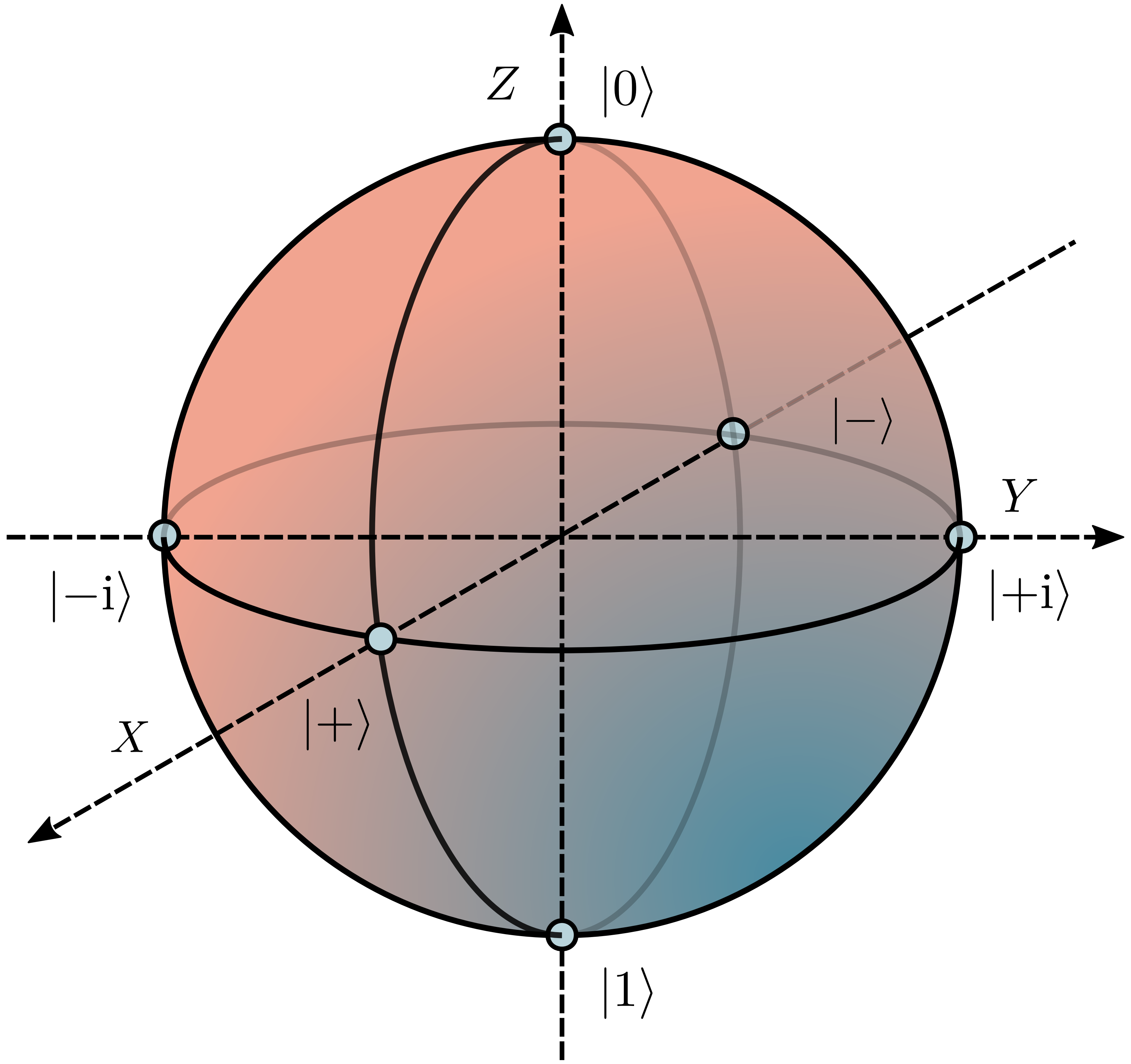
$$\mathbb{E}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \simeq \frac{1}{K}\sum_{j=1}^{K} \partial_i E\left(\boldsymbol{\theta}_j\right)^2$$



- Classical simulation complexity $\sim \mathcal{O}\left(2^N\right)$

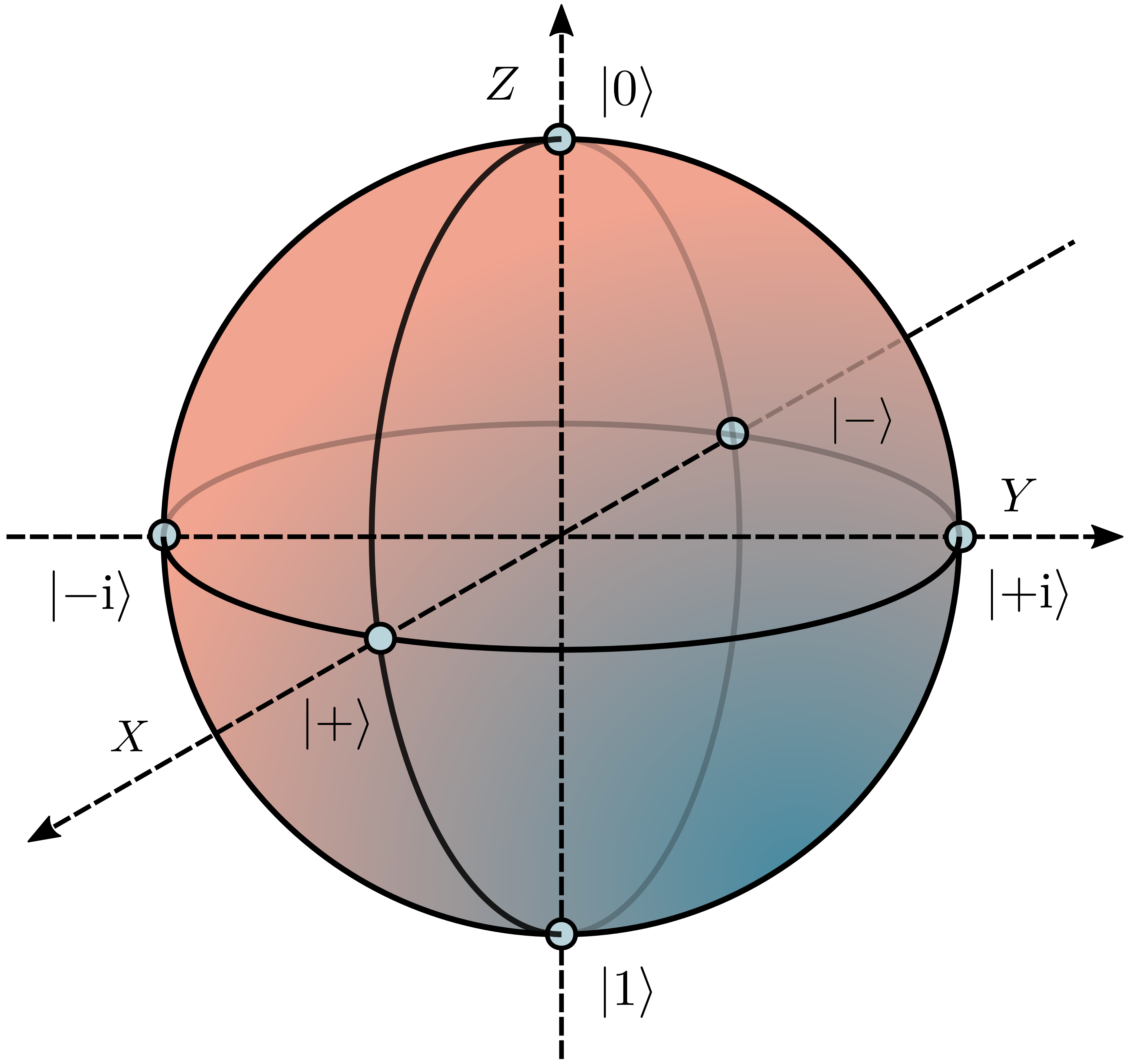
$\implies$ Too costly!

Interlude: Clifford circuits



Bloch sphere

Interlude: Clifford circuits

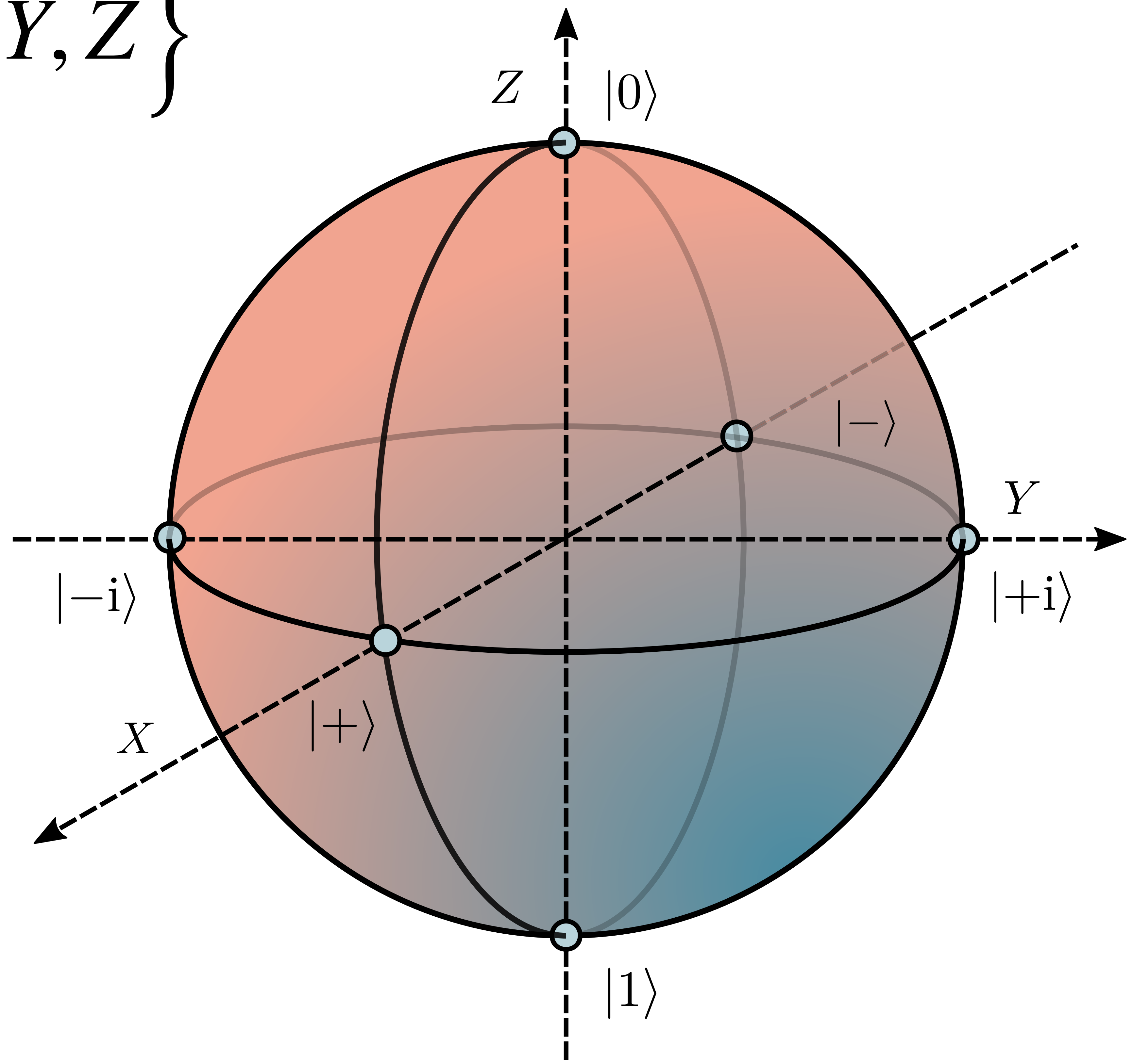


Bloch sphere

Interlude: Clifford circuits

- Clifford gates: gates mapping Pauli operators to Pauli operators

$$\hat{U} \in C_N \iff \hat{U}^\dagger \hat{P} \hat{U} = \hat{P}', \quad \text{with} \quad \hat{P}, \hat{P}' \in \left\{ \hat{1}, \hat{X}, \hat{Y}, \hat{Z} \right\}^{\otimes N}$$



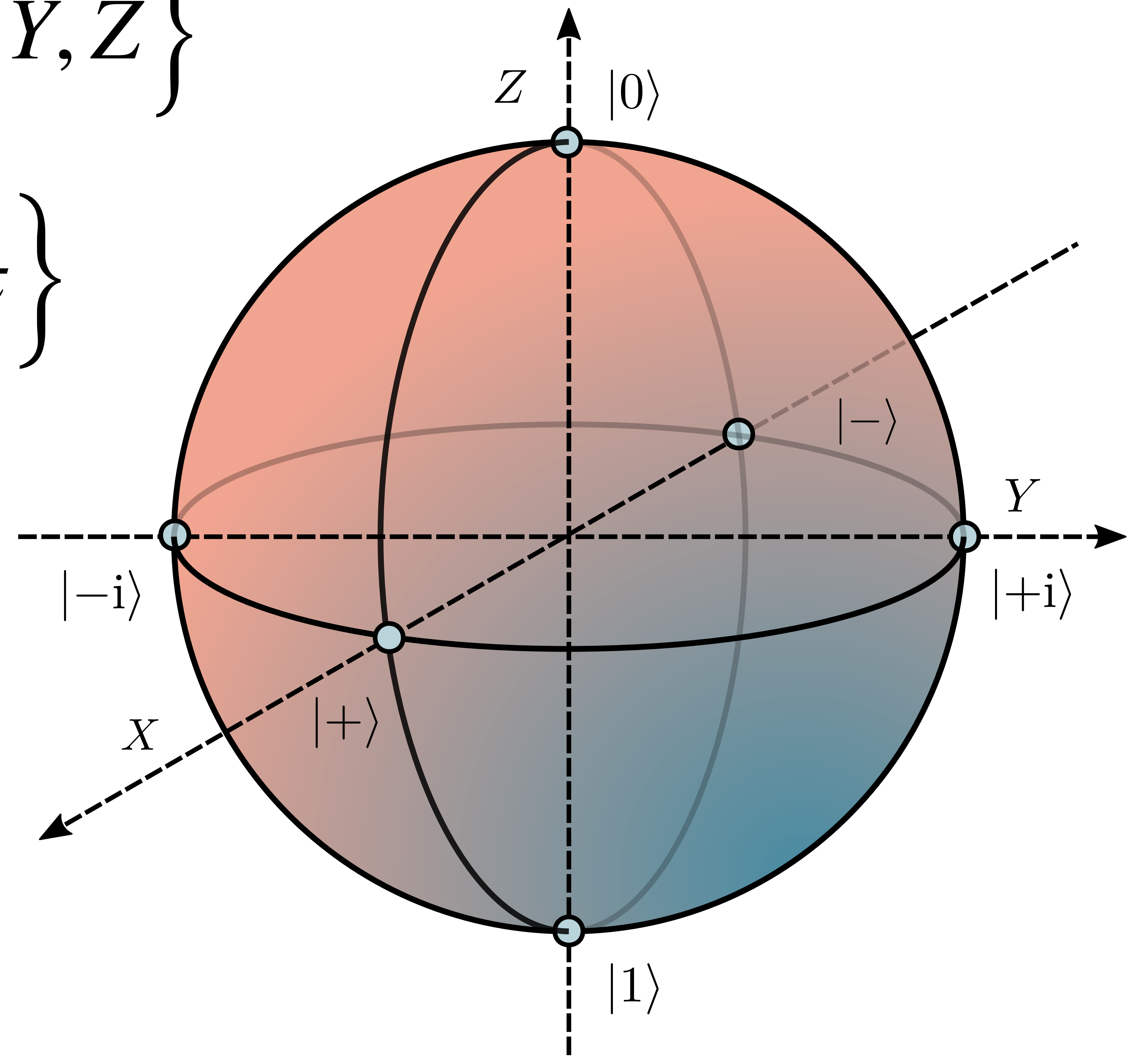
Bloch sphere

Interlude: Clifford circuits

- Clifford gates: gates mapping Pauli operators to Pauli operators

$$\hat{U} \in C_N \iff \hat{U}^\dagger \hat{P} \hat{U} = \hat{P}', \quad \text{with} \quad \hat{P}, \hat{P}' \in \left\{ \hat{1}, \hat{X}, \hat{Y}, \hat{Z} \right\}^{\otimes N}$$

- Example: X, Y and Z rotations with $\theta \in \left\{ -\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi \right\}$



Bloch sphere

Interlude: Clifford circuits

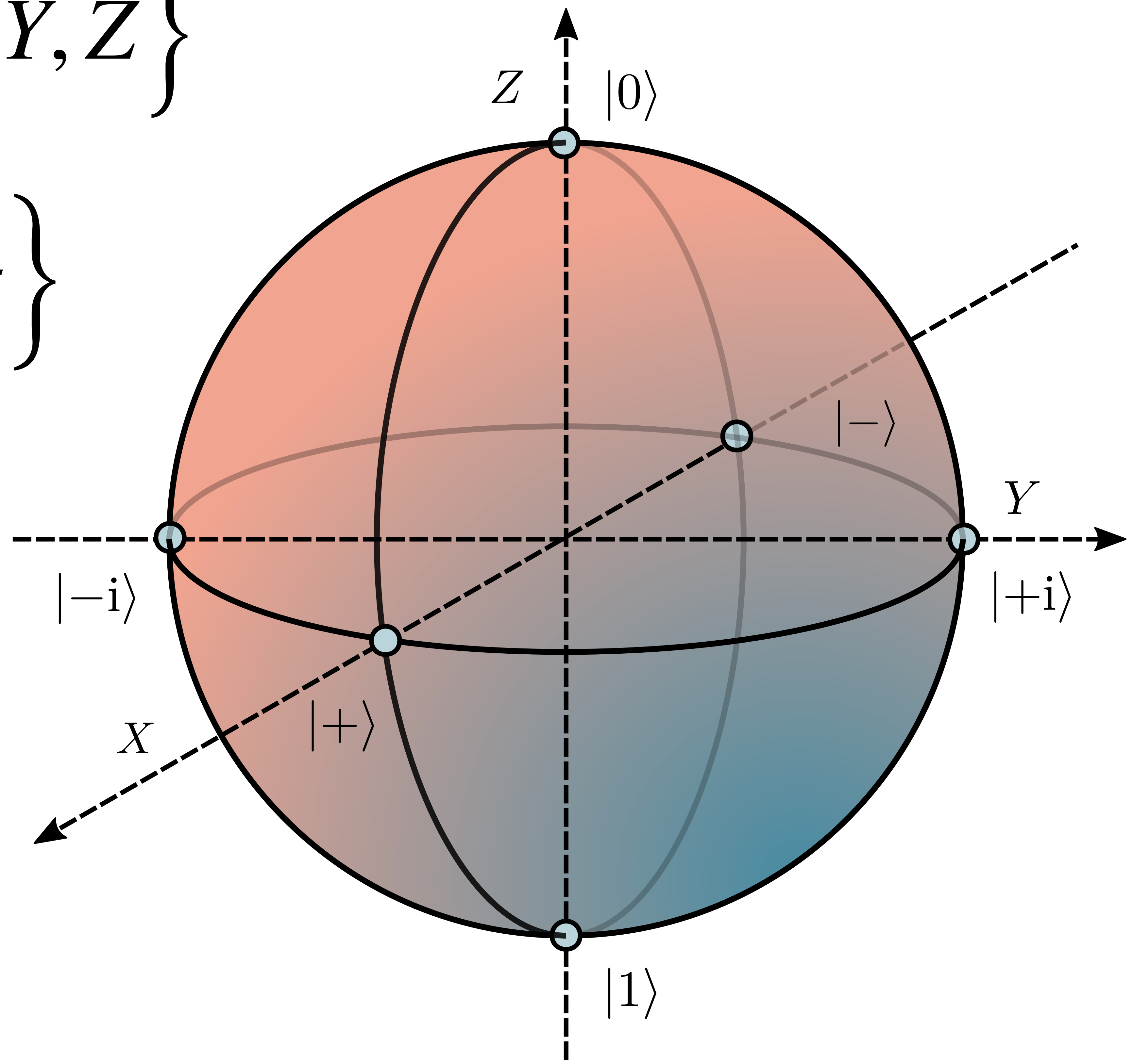
- Clifford gates: gates mapping Pauli operators to Pauli operators

$$\hat{U} \in C_N \iff \hat{U}^\dagger \hat{P} \hat{U} = \hat{P}', \quad \text{with} \quad \hat{P}, \hat{P}' \in \left\{ \hat{1}, \hat{X}, \hat{Y}, \hat{Z} \right\}^{\otimes N}$$

- Example: X, Y and Z rotations with $\theta \in \left\{ -\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi \right\}$

- Classical simulation of Clifford circuits:

  complexity $\sim \mathcal{O}\left(N^p\right)$



Bloch sphere

Efficient estimation of the trainability

Using Clifford circuits to estimate averages

Gottesman, “The Heisenberg Representation of Quantum Computers”, *arXiv:quant-ph/9807006* (1998)
Aaronson and Gottesman, “Improved simulation of stabiliser circuits”, *PRA* (2004)

Efficient estimation of the trainability

Using Clifford circuits to estimate averages

- Under some conditions, sampling Clifford angles is enough to estimate average quantities:

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \simeq \frac{1}{K}\sum_{j=1}^{K} \partial_i E\left(\boldsymbol{\theta}_j\right)^2$$

Gottesman, "The Heisenberg Representation of Quantum Computers", *arXiv:quant-ph/9807006* (1998)
Aaronson and Gottesman, "Improved simulation of stabiliser circuits", *PRA* (2004)

Efficient estimation of the trainability

Using Clifford circuits to estimate averages

- Under some conditions, sampling Clifford angles is enough to estimate average quantities:

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \simeq \frac{1}{K}\sum_{j=1}^{K}\partial_i E\left(\boldsymbol{\theta}_j\right)^2 \quad \leftarrow \quad \boldsymbol{\theta}_j \in \left\{-\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi\right\}^M$$

Gottesman, “The Heisenberg Representation of Quantum Computers”, *arXiv:quant-ph/9807006* (1998)
Aaronson and Gottesman, “Improved simulation of stabiliser circuits”, *PRA* (2004)

Efficient estimation of the trainability

Using Clifford circuits to estimate averages

- Under some conditions, sampling Clifford angles is enough to estimate average quantities:

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \simeq \frac{1}{K}\sum_{j=1}^{K} \partial_i E\left(\boldsymbol{\theta}_j\right)^2 \quad \leftarrow \quad \boldsymbol{\theta}_j \in \left\{-\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi\right\}^M$$

Clifford circuits

Gottesman, "The Heisenberg Representation of Quantum Computers", *arXiv:quant-ph/9807006* (1998)
Aaronson and Gottesman, "Improved simulation of stabiliser circuits", *PRA* (2004)

Efficient estimation of the trainability

Using Clifford circuits to estimate averages

- Under some conditions, sampling Clifford angles is enough to estimate average quantities:

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\partial_i E\left(\boldsymbol{\theta}\right)^2\right] \simeq \frac{1}{K}\sum_{j=1}^{K} \partial_i E\left(\boldsymbol{\theta}_j\right)^2 \quad \leftarrow \quad \boldsymbol{\theta}_j \in \left\{-\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi\right\}^M$$

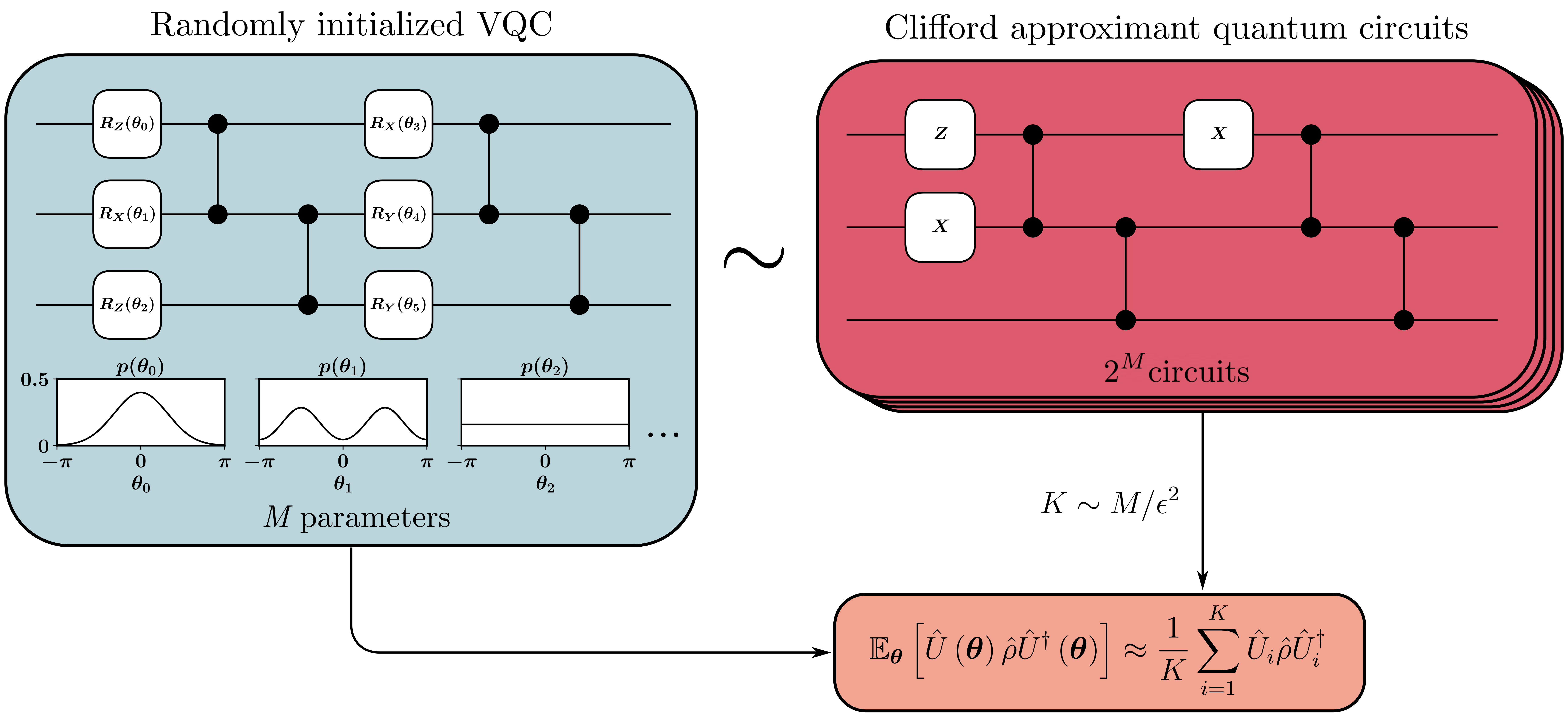
Clifford circuits

- One can estimate the trainability with a complexity $\sim \mathcal{O}\left(N^p M^q\right)$

Gottesman, "The Heisenberg Representation of Quantum Computers", *arXiv:quant-ph/9807006* (1998)
Aaronson and Gottesman, "Improved simulation of stabiliser circuits", *PRA* (2004)

Efficient estimation of the trainability

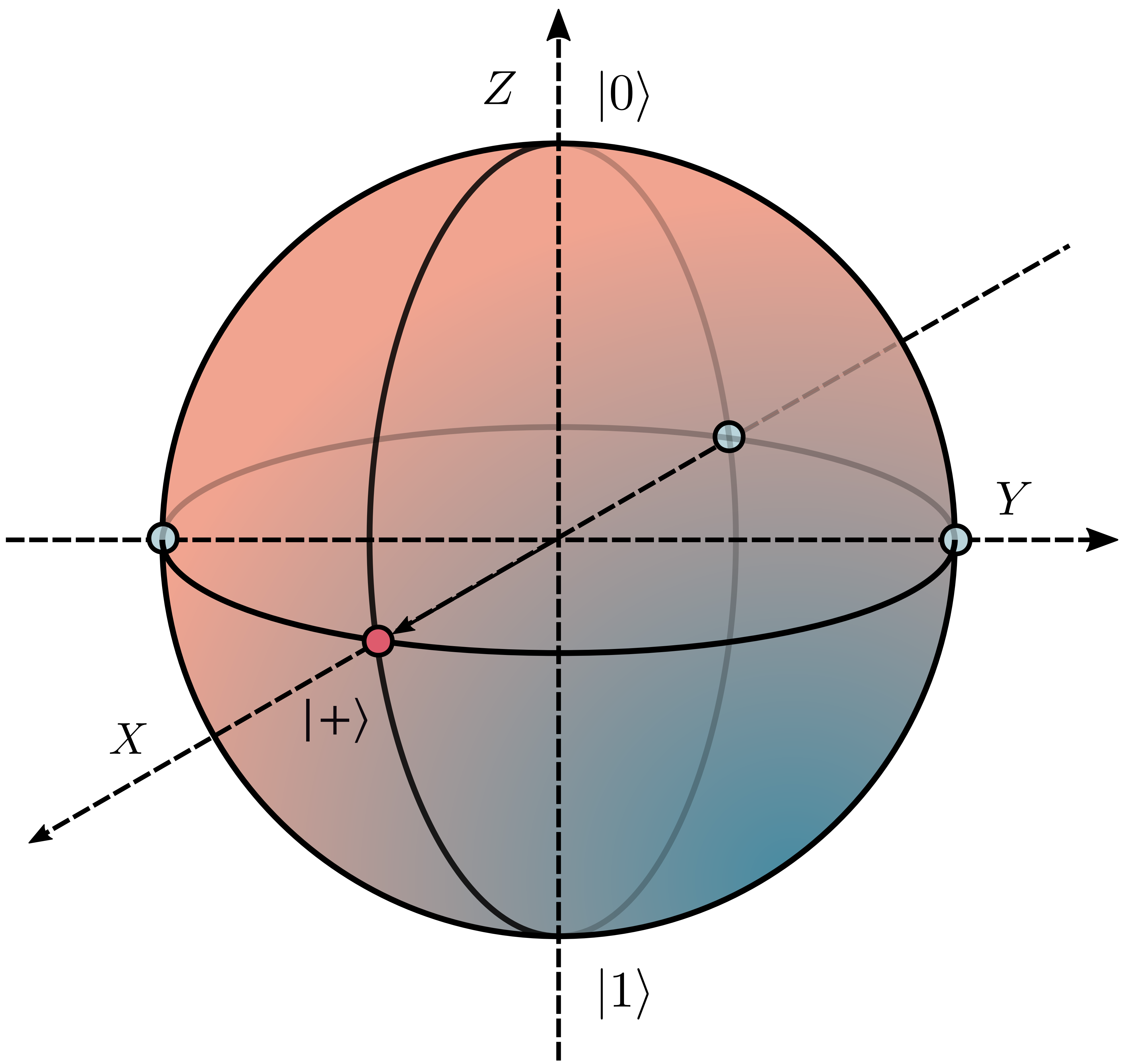
Using Clifford circuits to estimate averages



"Efficient estimation of trainability for Variational Quantum Circuits", V. Heyraud, Z. Li, K. Donatella, A. Le Boité, and C. Ciuti, arXiv:2302.04649

Efficient estimation of the trainability

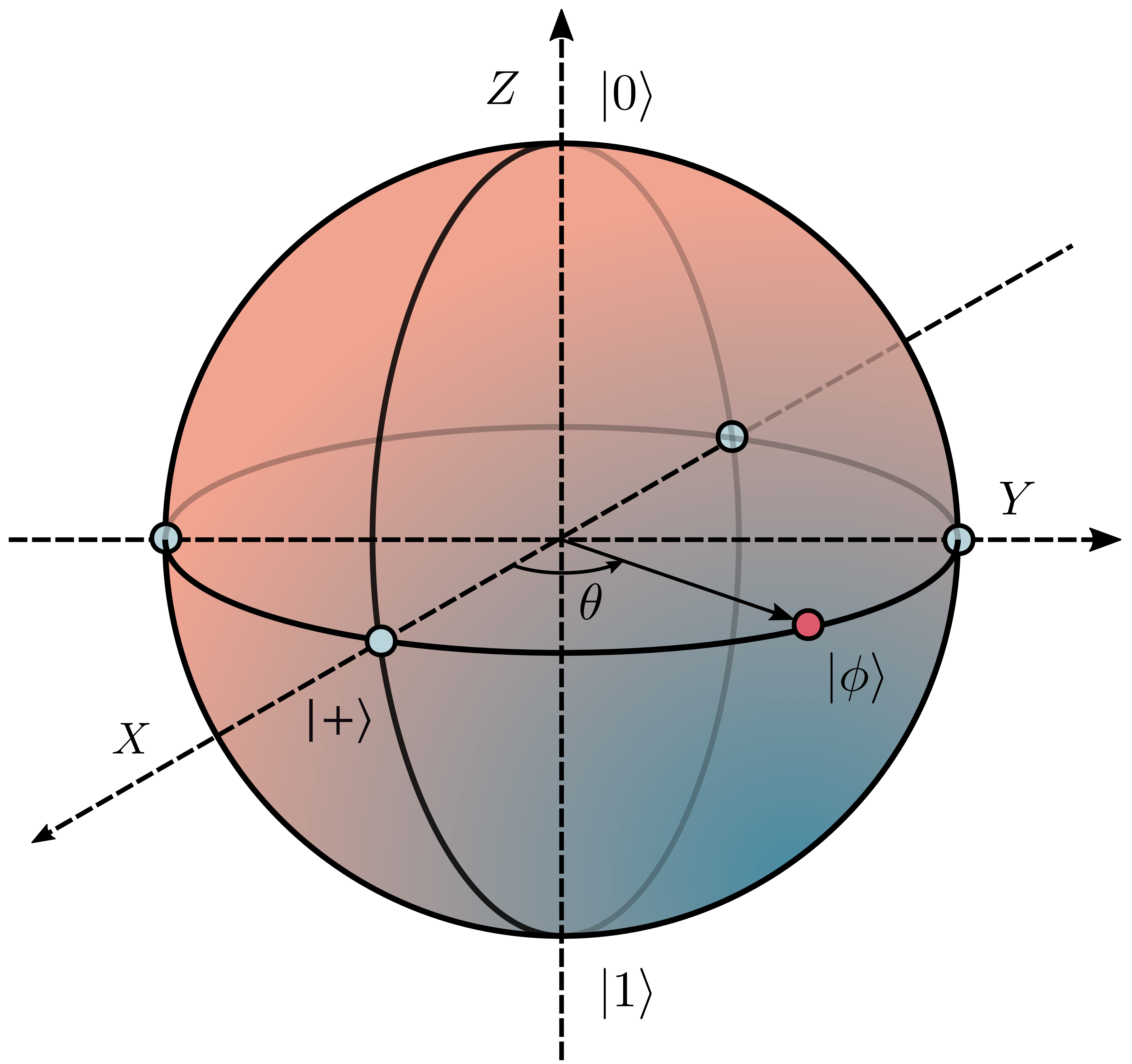
Physical intuition: decoherence



$$|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{\mathrm{i}\theta}|1\rangle\right)$$

Efficient estimation of the trainability

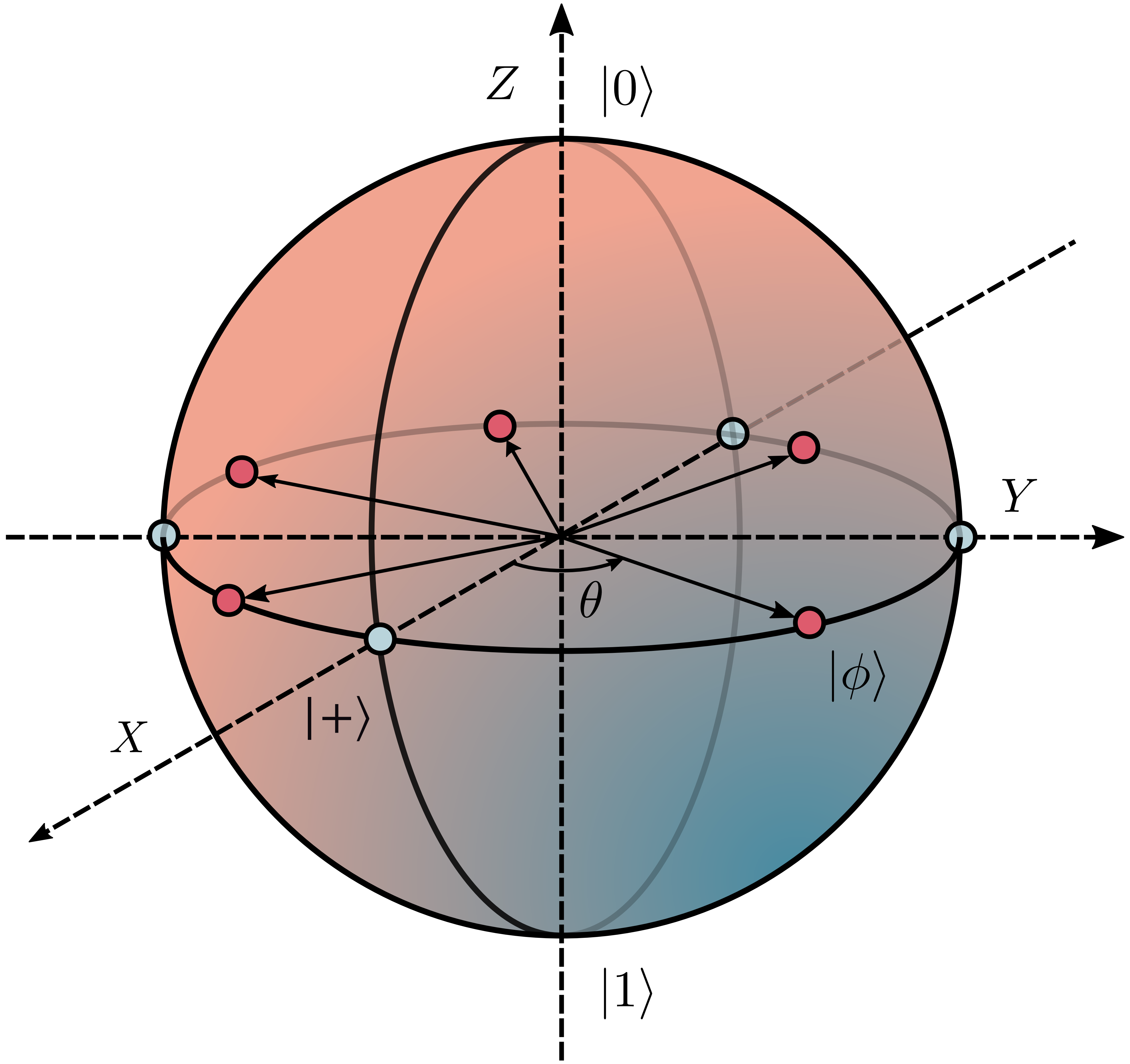
Physical intuition: decoherence



$$|\phi\rangle = \hat{R}_Z(\theta)|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{\mathrm{i}\theta}|1\rangle\right)$$

Efficient estimation of the trainability

Physical intuition: decoherence

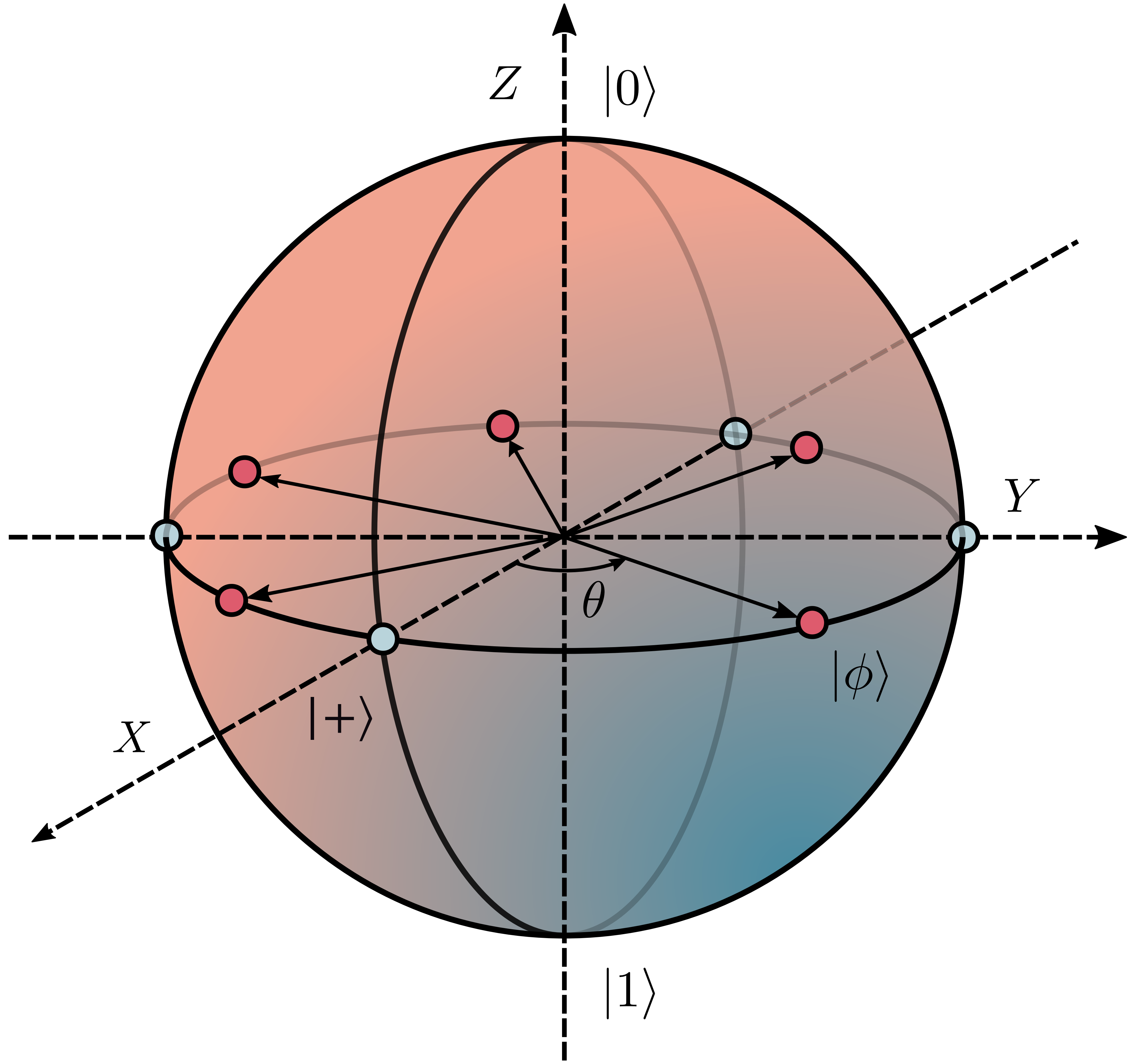


$$|\phi\rangle = \hat{R}_Z(\theta)\,|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{\mathrm{i}\theta}|1\rangle\right)$$

$$|\phi_j\rangle = \hat{R}_Z\left(\theta_j\right)|+\rangle\,, \quad j = 1, \ldots, K$$

Efficient estimation of the trainability

Physical intuition: decoherence



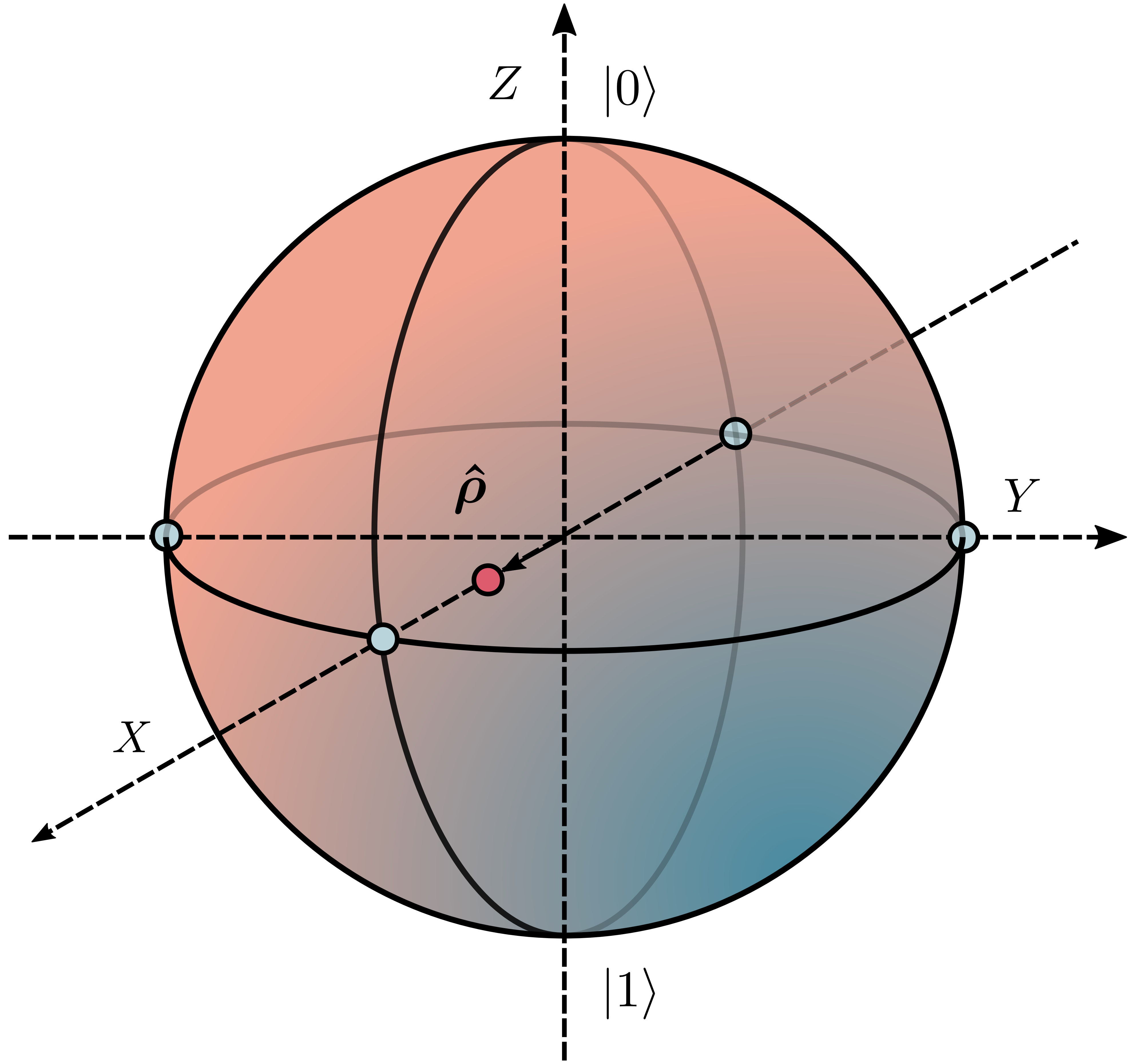
$$|\phi\rangle = \hat{R}_Z(\theta)|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{\mathrm{i}\theta}|1\rangle\right)$$

$$|\phi_j\rangle = \hat{R}_Z\left(\theta_j\right)|+\rangle, \quad j = 1, \ldots, K$$

$$\hat{\rho} = \frac{1}{K}\sum_{j=1}^{K}|\phi_j\rangle\langle\phi_j|$$

Efficient estimation of the trainability

Physical intuition: decoherence



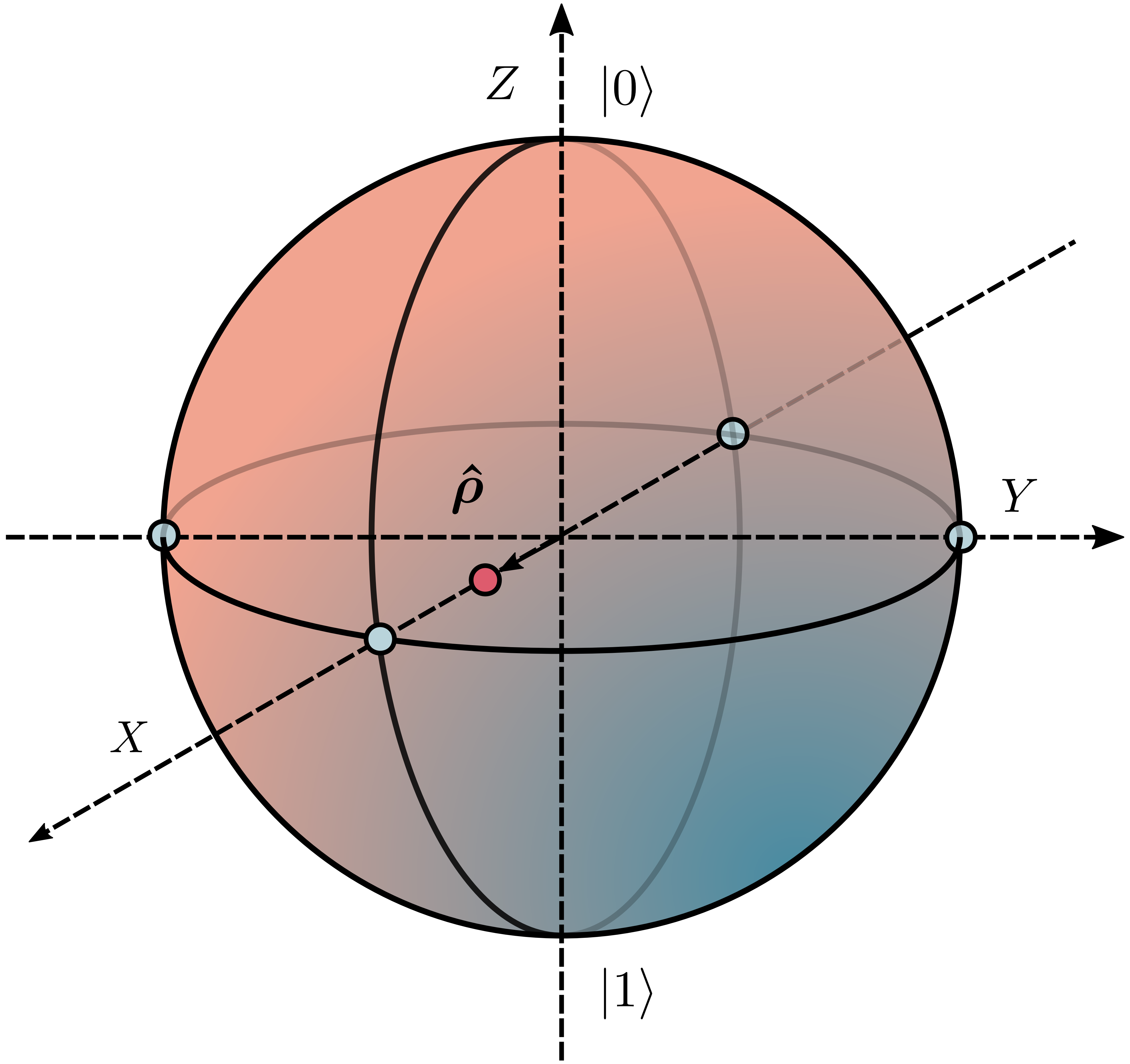
$$|\phi\rangle = \hat{R}_Z(\theta)\,|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{\mathrm{i}\theta}|1\rangle\right)$$

$$|\phi_j\rangle = \hat{R}_Z\left(\theta_j\right)|+\rangle\,, \quad j = 1, \ldots, K$$

$$\hat{\rho} = \frac{1}{K}\sum_{j=1}^{K} |\phi_j\rangle\langle\phi_j|$$

Efficient estimation of the trainability

Physical intuition: decoherence



$$|\phi\rangle = \hat{R}_Z(\theta)|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i\theta}|1\rangle\right)$$

$$|\phi_j\rangle = \hat{R}_Z\left(\theta_j\right)|+\rangle, \quad j = 1, \dots, K$$

$$\hat{\rho} = \frac{1}{K}\sum_{j=1}^{K} |\phi_j\rangle\langle\phi_j|$$

$$\frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \longrightarrow \hat{\rho} \simeq \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right)$$

Perspectives

Perspectives

- The method can be used to evaluate many average quantities

Perspectives

- The method can be used to evaluate many average quantities
- Limited to independent rotations and fixed Clifford gates

Perspectives

- The method can be used to evaluate many average quantities
- Limited to independent rotations and fixed Clifford gates
- Current investigations:
  Optimisation of the circuit architecture using reinforcement learning methods

Thank you !

Paper available on arXiv !

“Efficient estimation of trainability for Variational Quantum Circuits”,
V. Heyraud, Z. Li, K. Donatella, A. Le Boité, and C. Ciuti, arXiv:2302.04649