Learning theory for quantum machines

Hsin-Yuan Huang (Robert)

Google Quantum AI

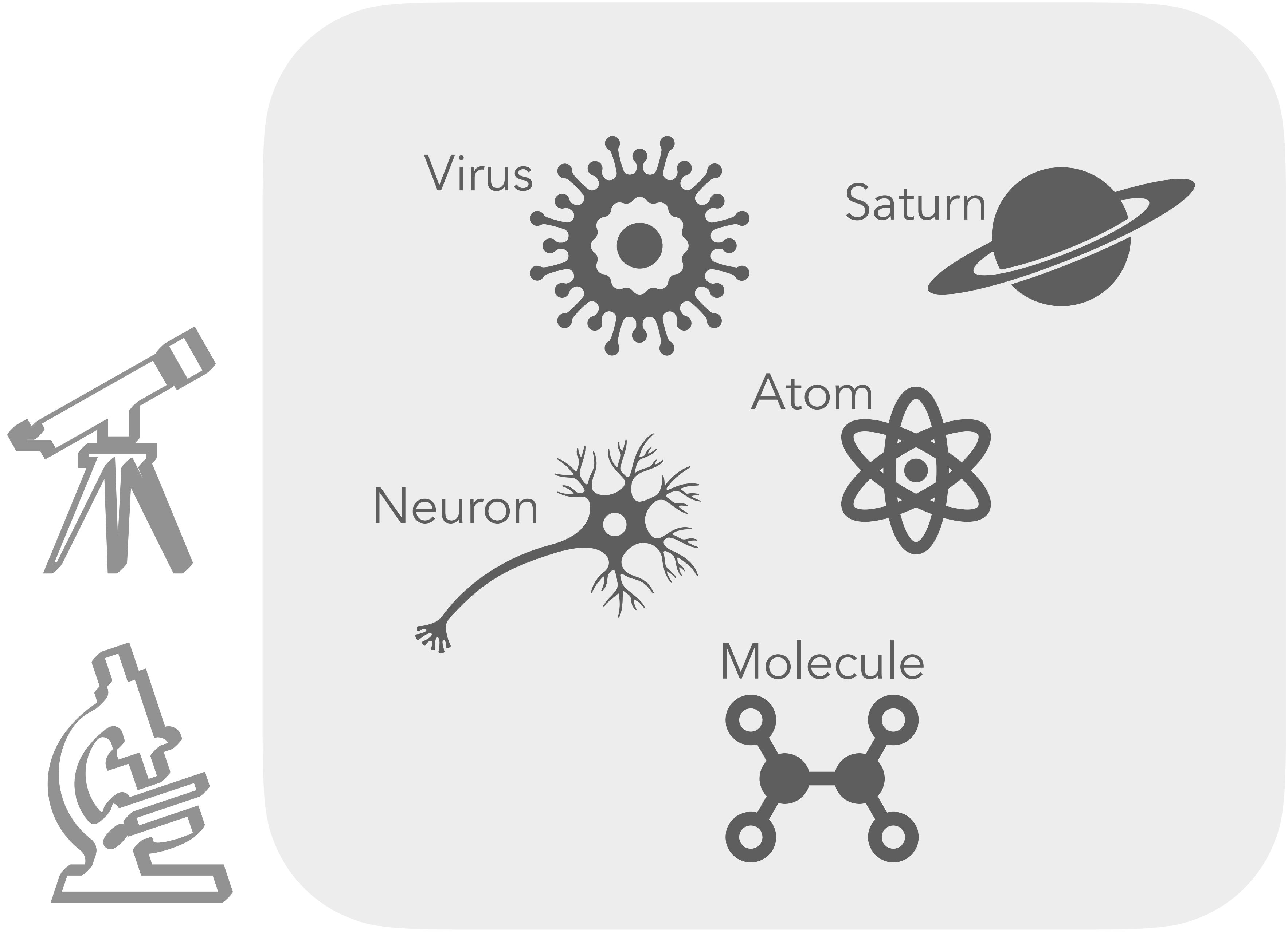
IQIM INSTITUTE FOR QUANTUM INFORMATION AND MATTER

Caltech

MIT

Motivation

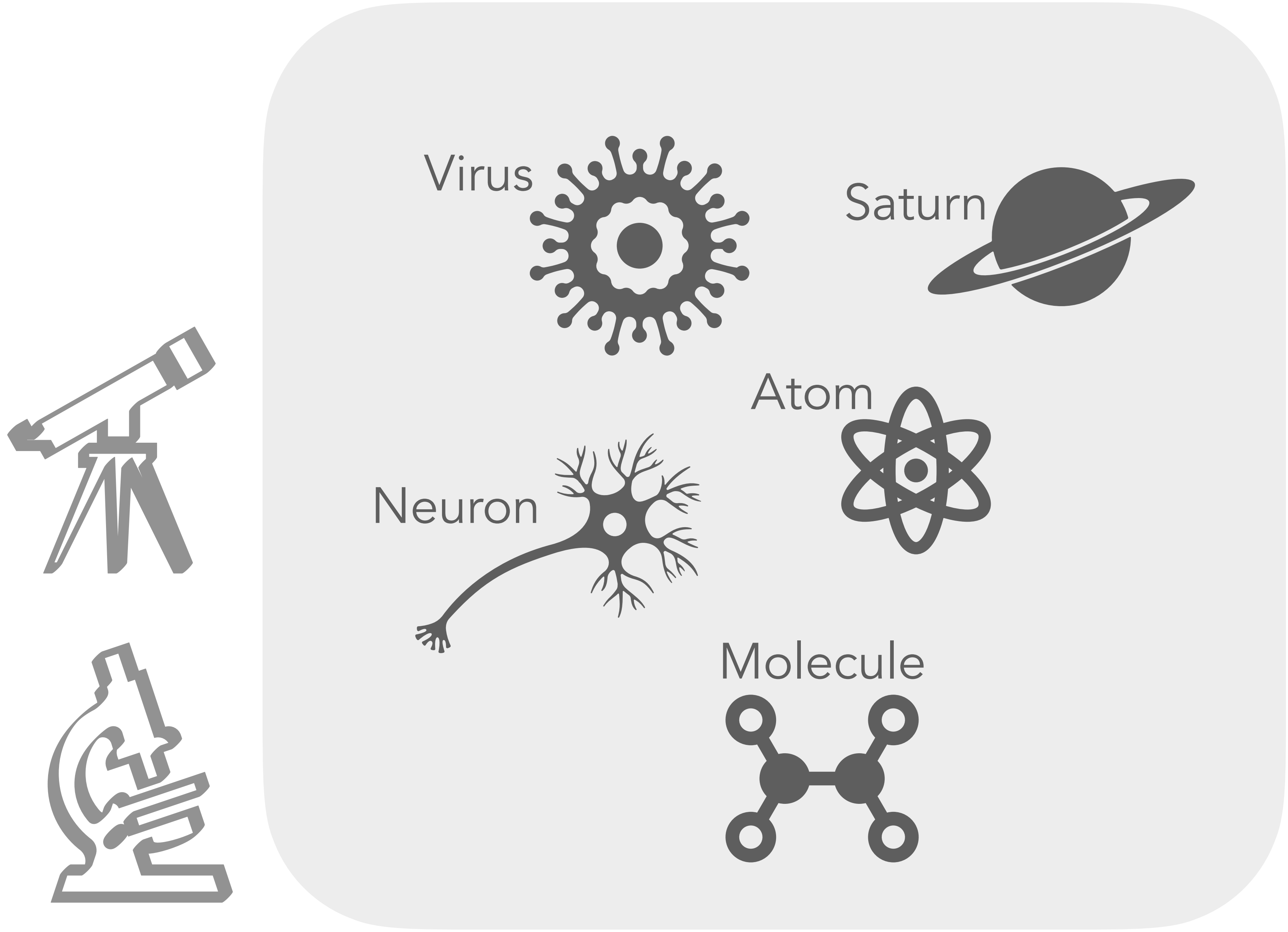
- A central goal of science is to learn how our universe operates.



Examples of scientific disciplines

Motivation

- A central goal of science is to learn how our universe operates.
- Because our universe is **inherently quantum**, a quantum machine that can learn has the potential to lead to many scientific advances.

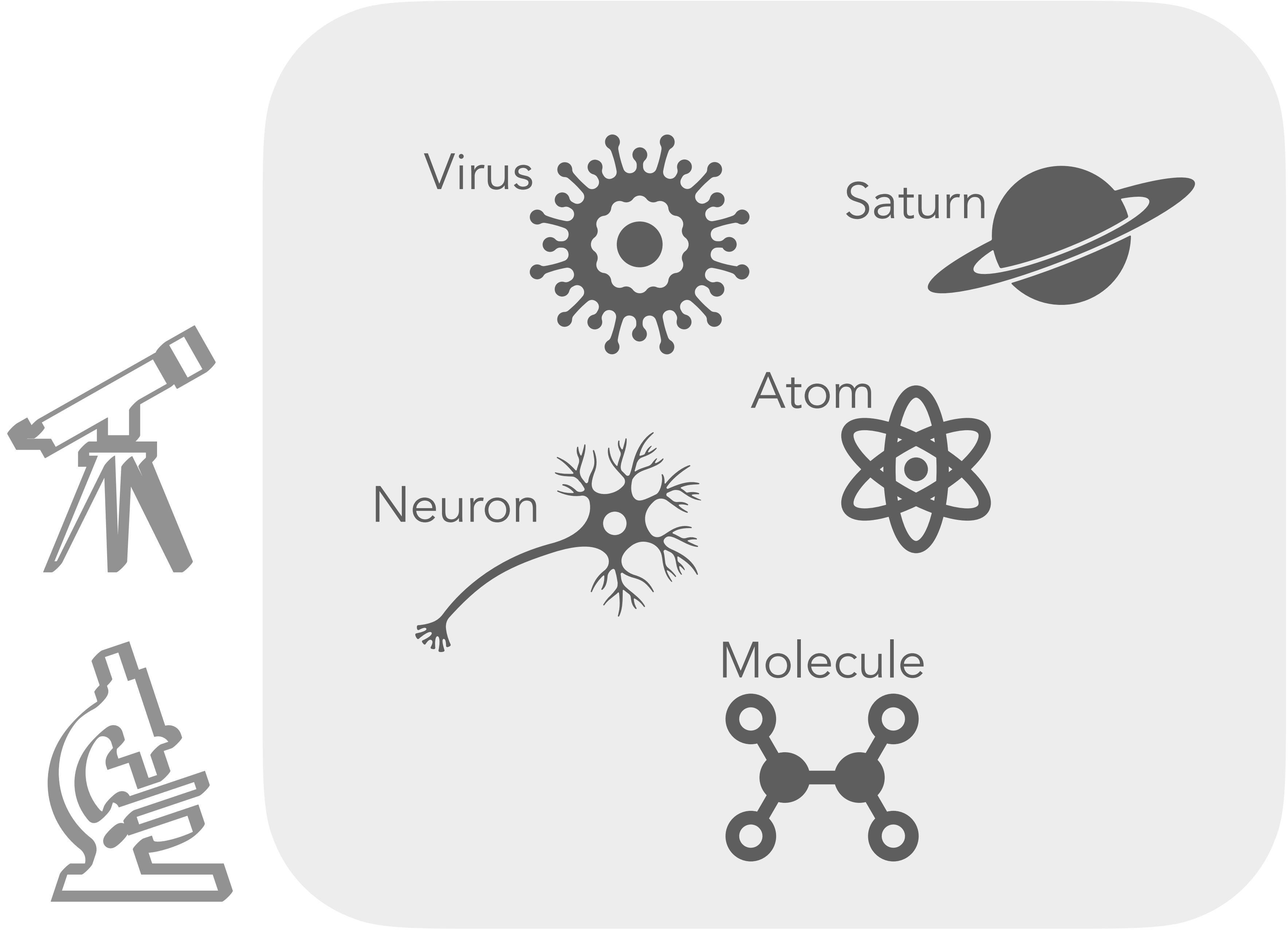


Examples of scientific disciplines

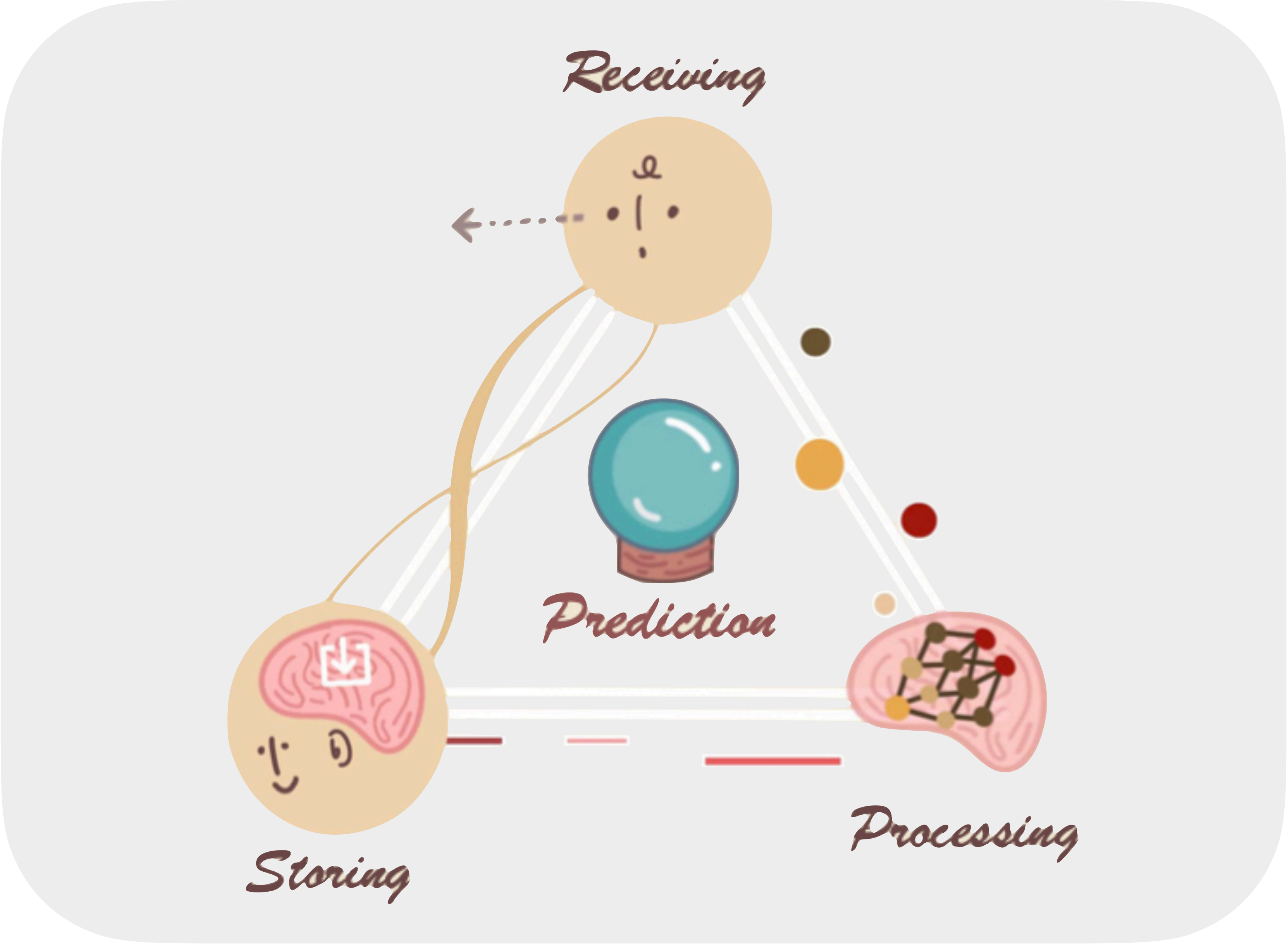
Motivation

- Learning is the combination of:

  1. receiving information about the universe, 2. processing that information to form models,
  3. storing the models and, subsequently, 4. using the models to predict in new scenarios.



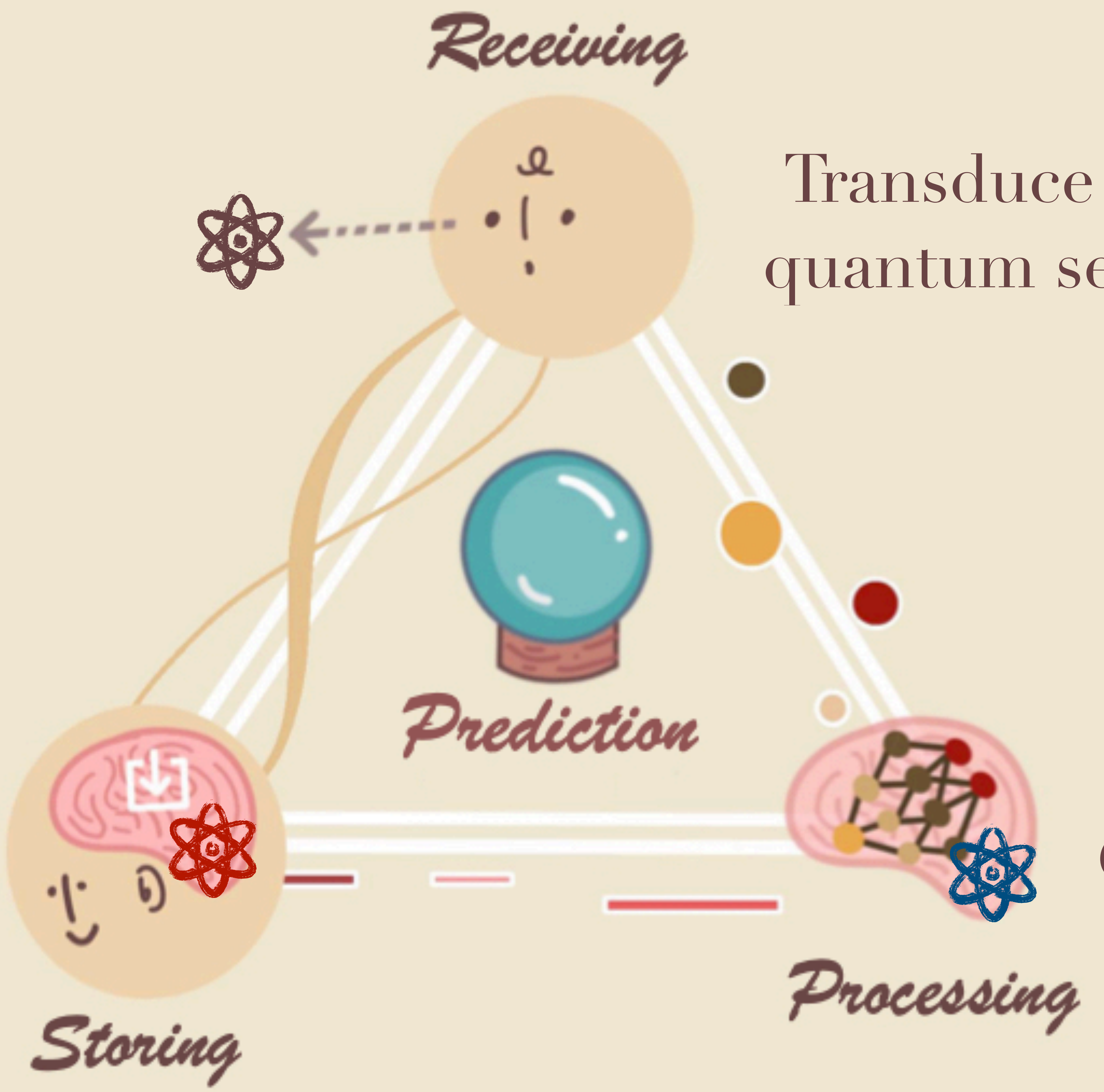
Examples of scientific disciplines



A cartoon depiction of learning

Quantum agent

Receive, process, and store quantum information



Transduce from quantum sensors

Quantum Memory

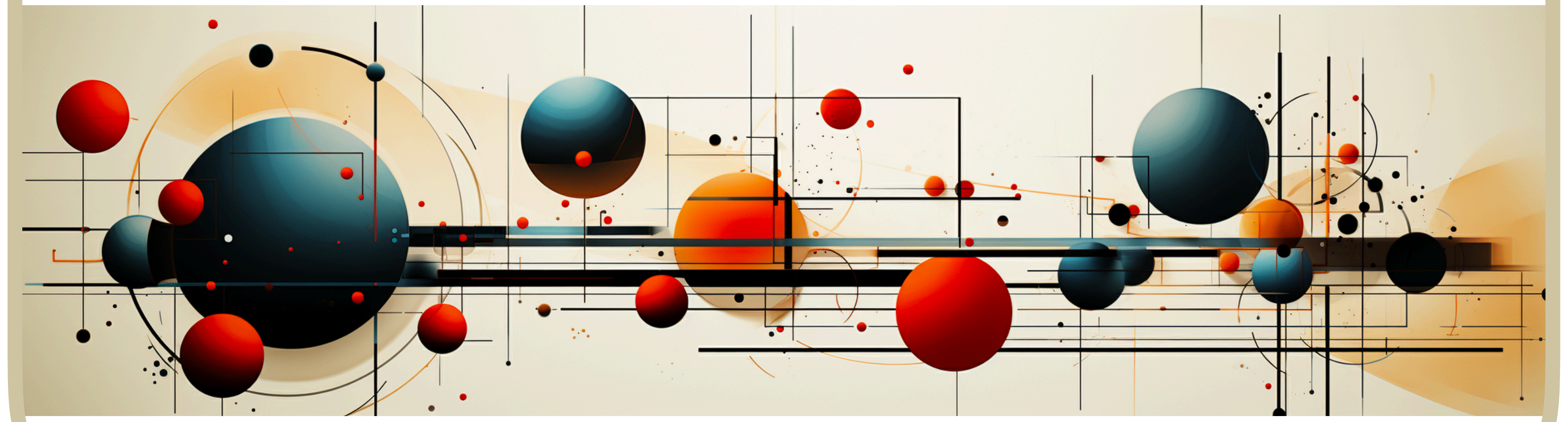
Quantum Computation

Overview

Learning theory for quantum machines

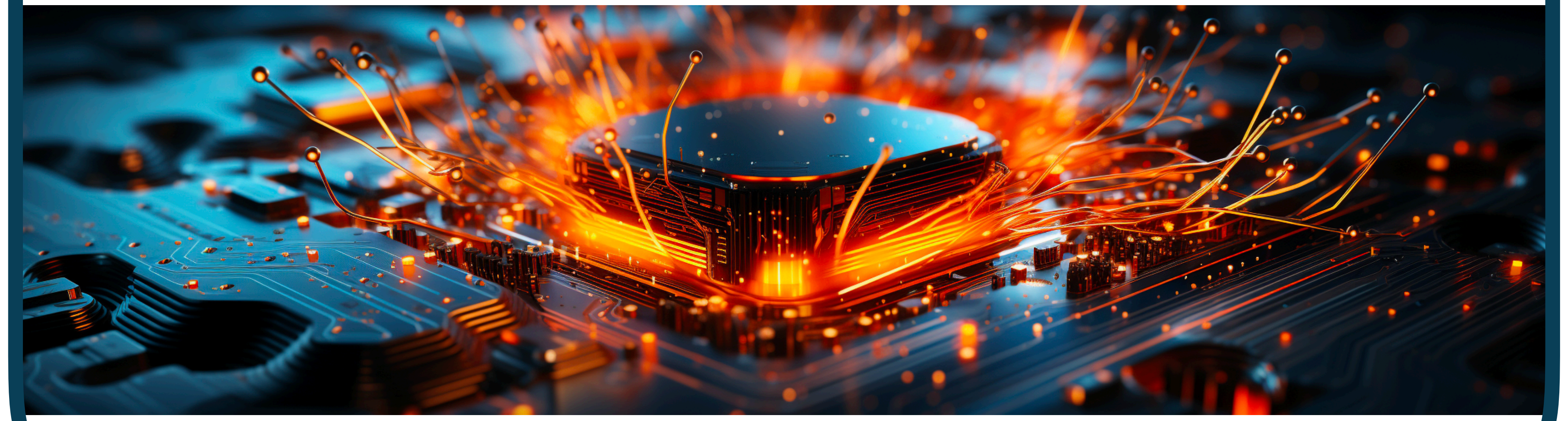
Foundation

How well can quantum machines predict?
How good is the generalization ability of quantum machines?



Quantum advantage

What can quantum machines learn that classical machines cannot?
How big can the advantage be?

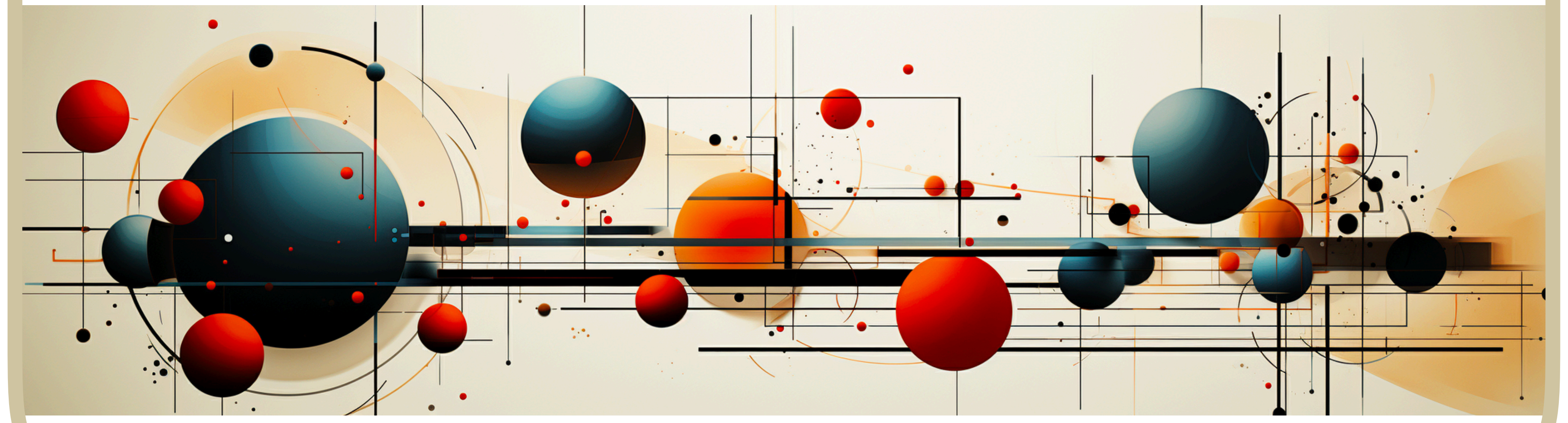


Overview

Learning theory for quantum machines

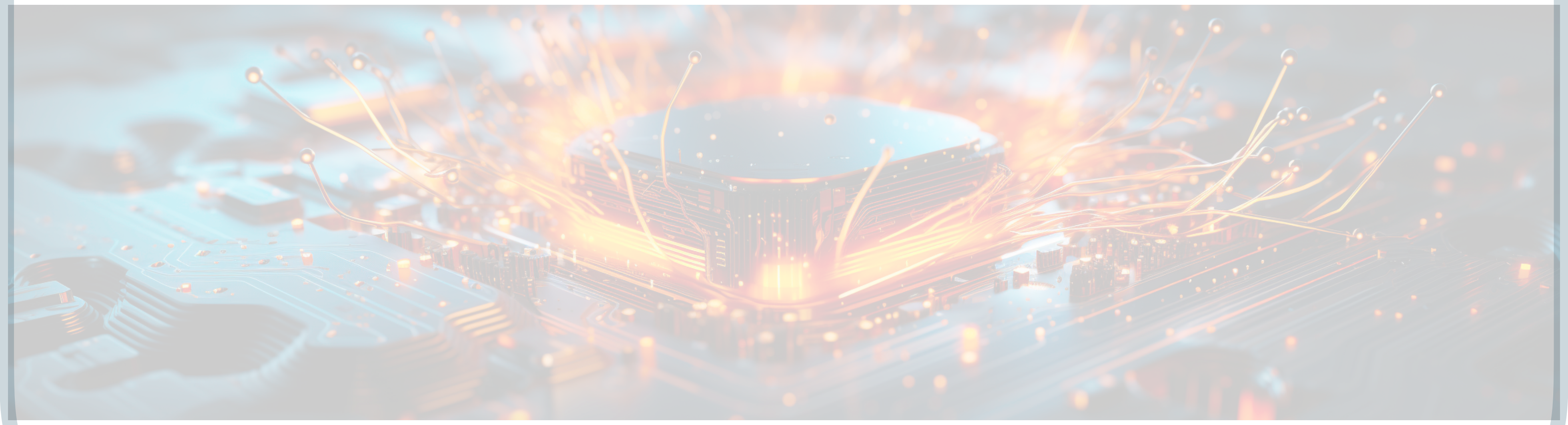
Foundation

How well can quantum machines predict?
How good is the generalization ability
of quantum machines?



Quantum advantage

What can quantum machines learn
that classical machines cannot?
How big can the advantage be?



Foundation

- How to understand the prediction performance of a quantum machine?

Foundation

- How to understand the prediction performance of a quantum machine?

Prediction error = Training error + Generalization error

Foundation

- How to understand the prediction performance of a quantum machine?

Prediction error = Training error + Generalization error

Error on unseen inputs

Foundation

- How to understand the prediction performance of a quantum machine?

Prediction error = Training error + Generalization error

Error on unseen inputs   Error on training data

Foundation

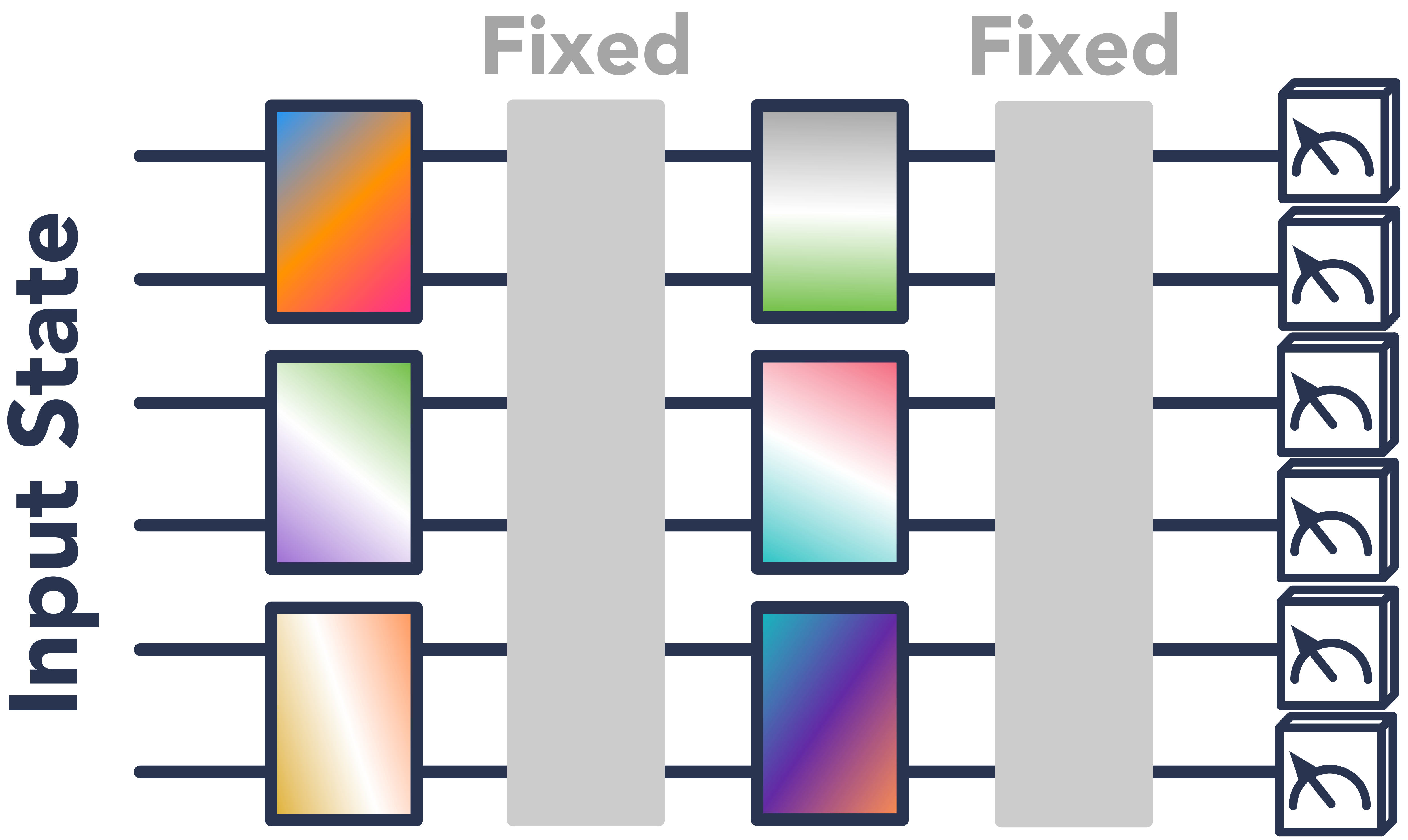
- How to understand the prediction performance of a quantum machine?

Prediction error = Training error + Generalization error

Error on unseen inputs   Error on training data

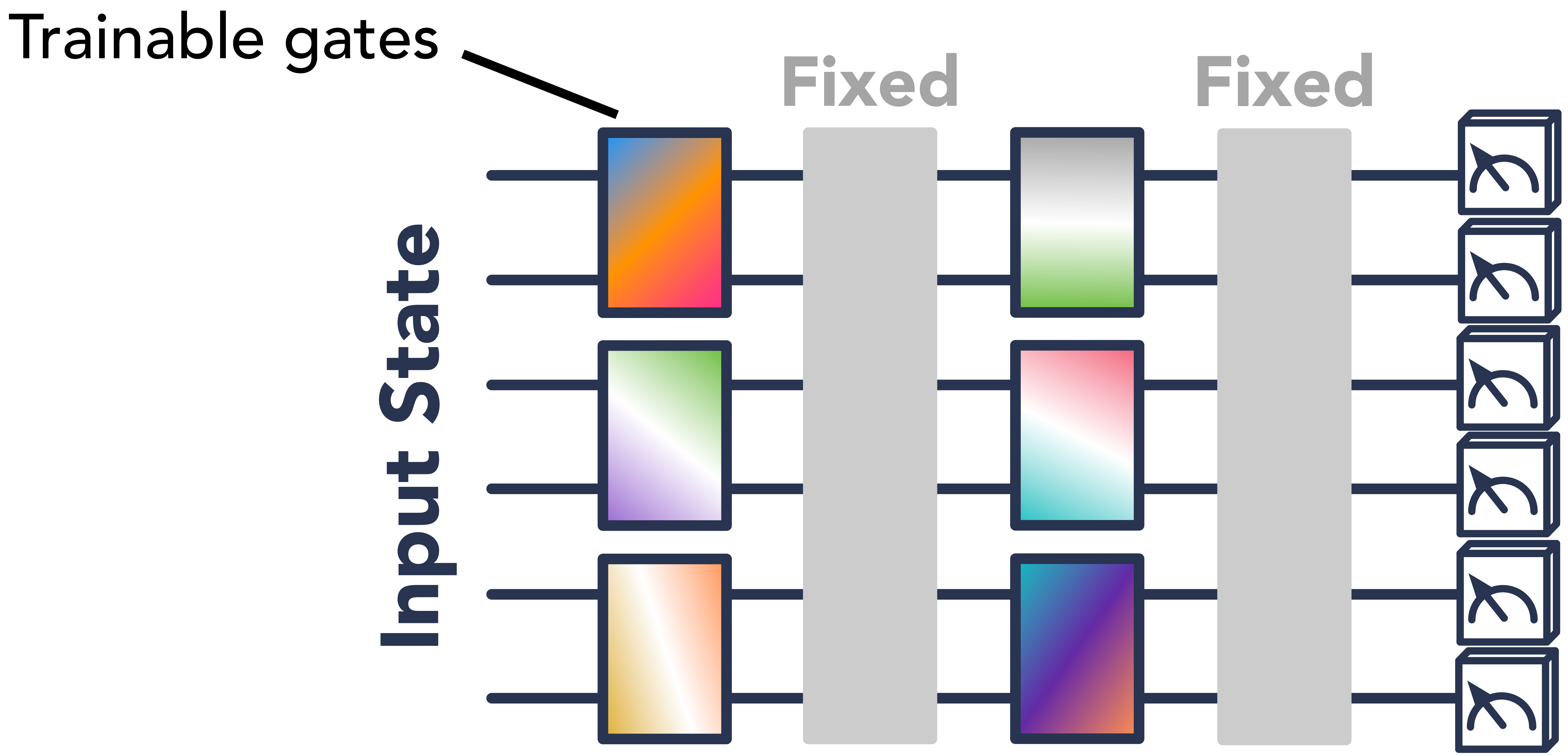
- The key is to understand the generalization error.

Trainable Quantum Machine



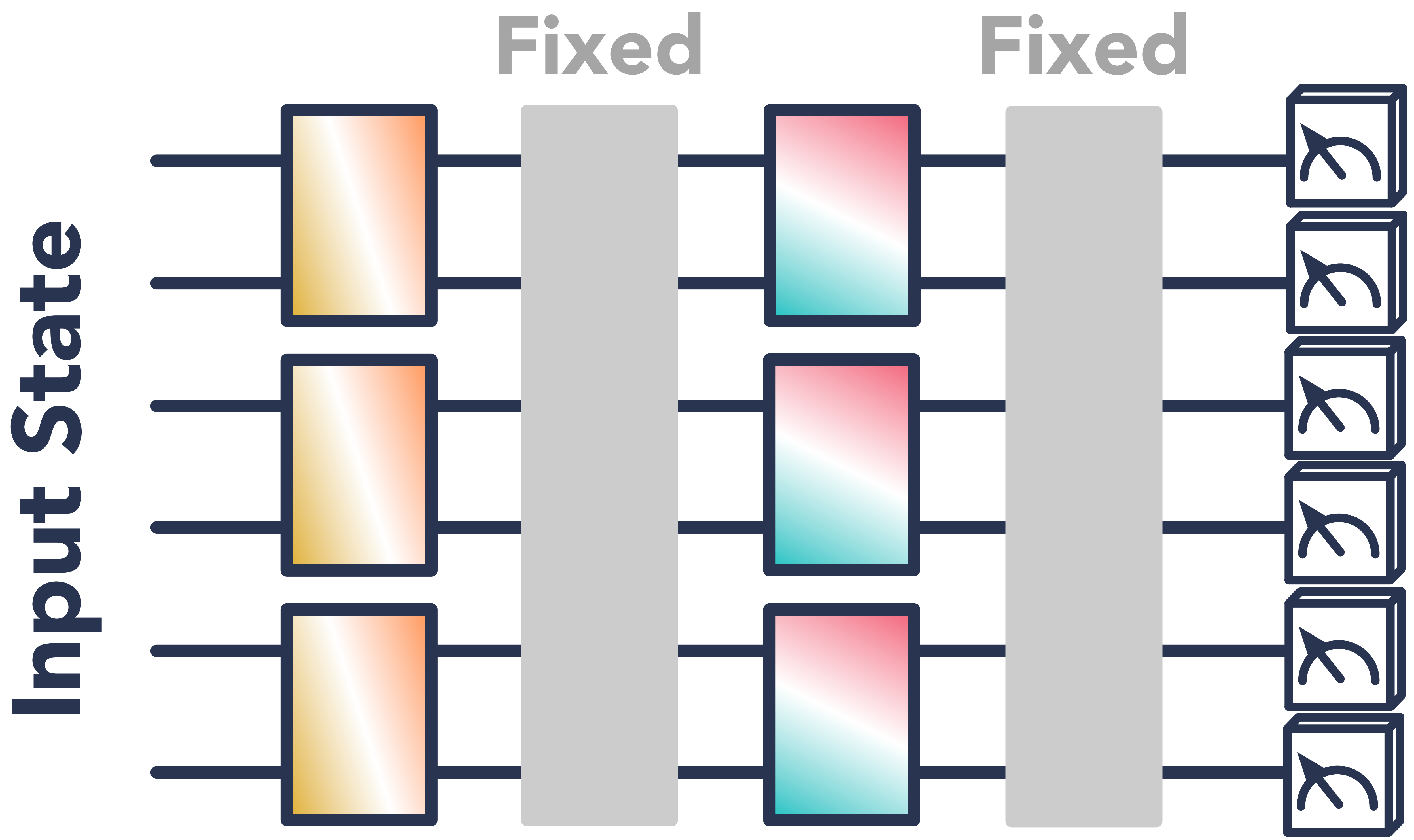
Basic Form

Trainable Quantum Machine



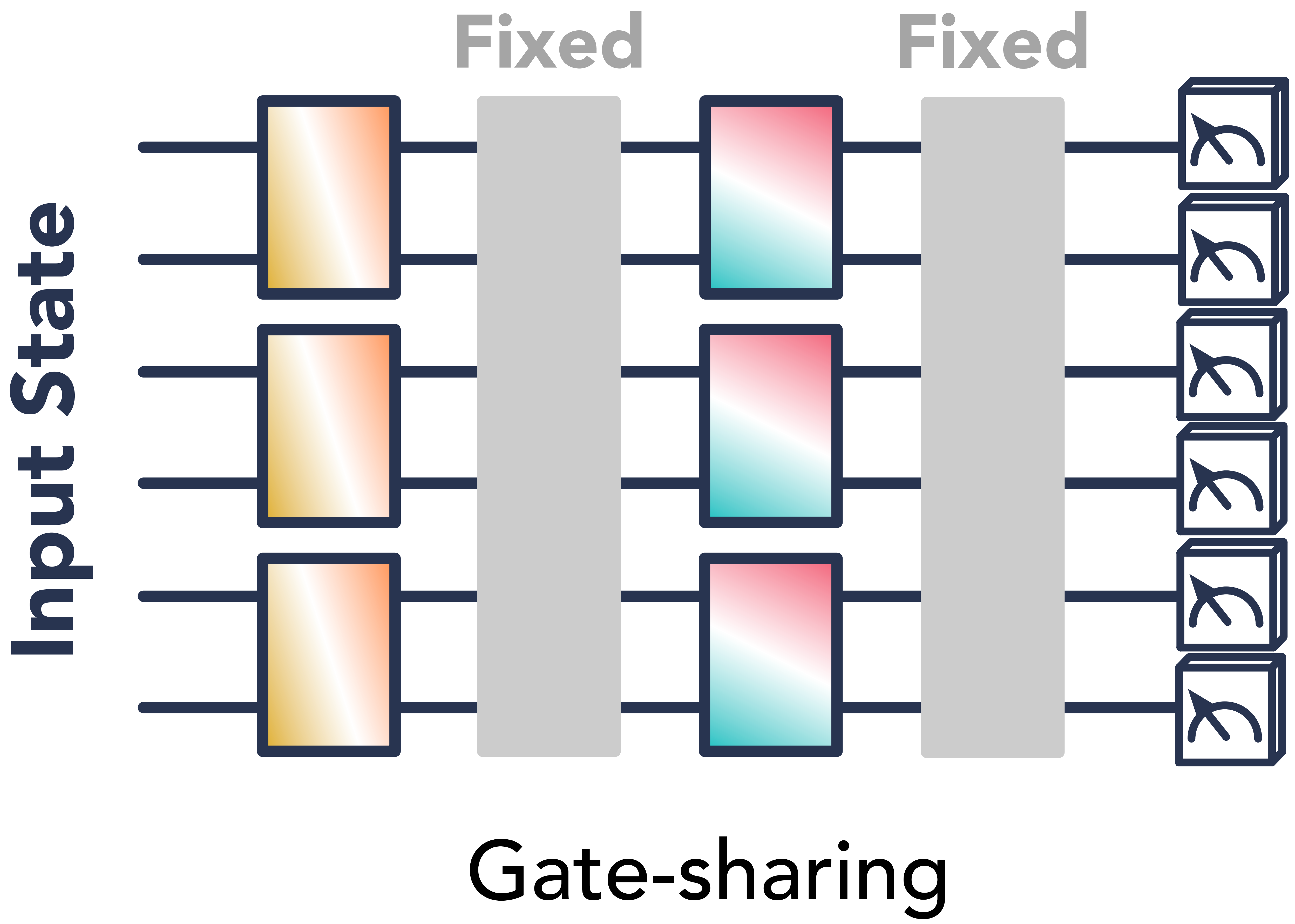
Basic Form

Trainable Quantum Machine



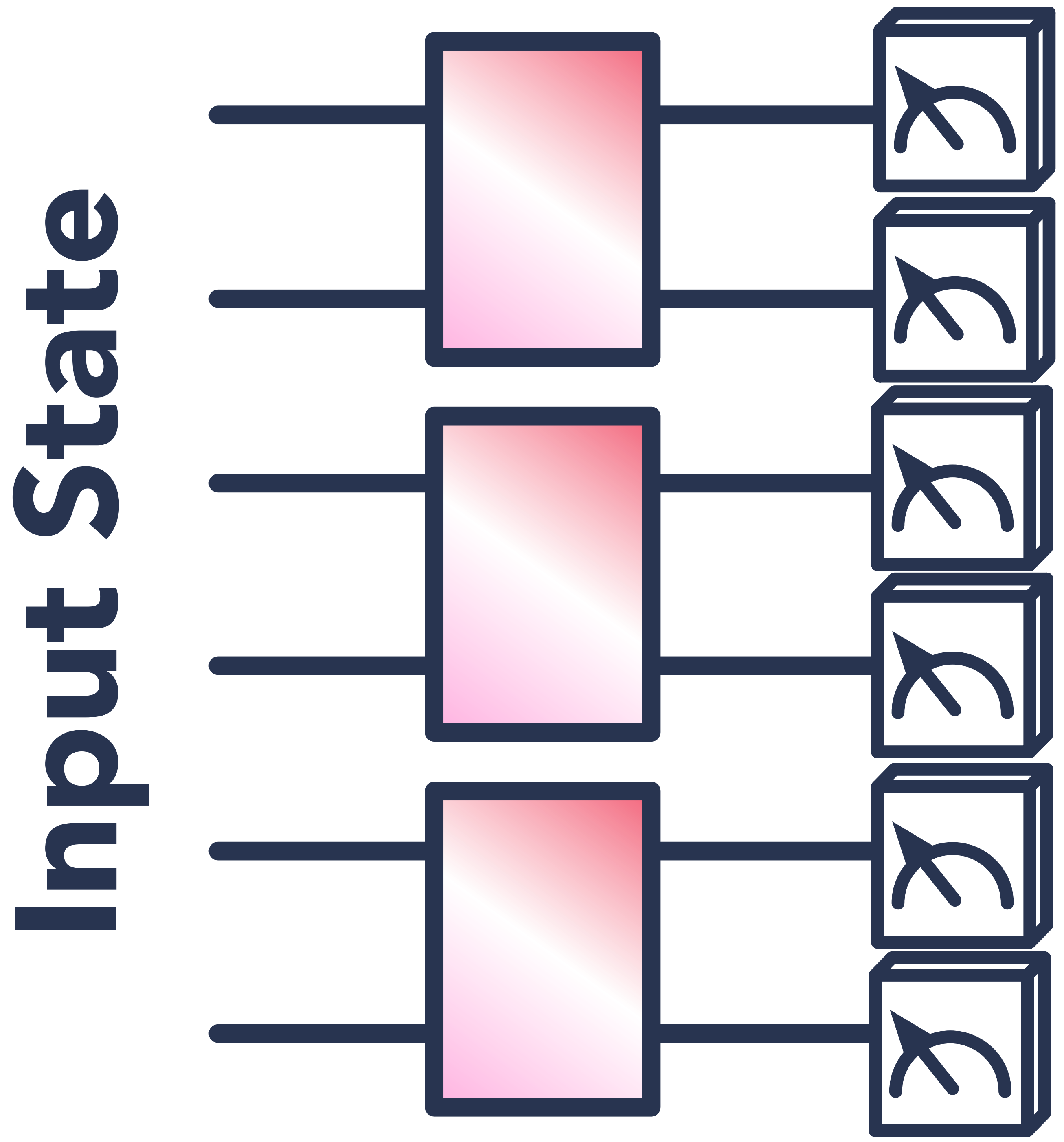
Gate-sharing

Trainable Quantum Machine



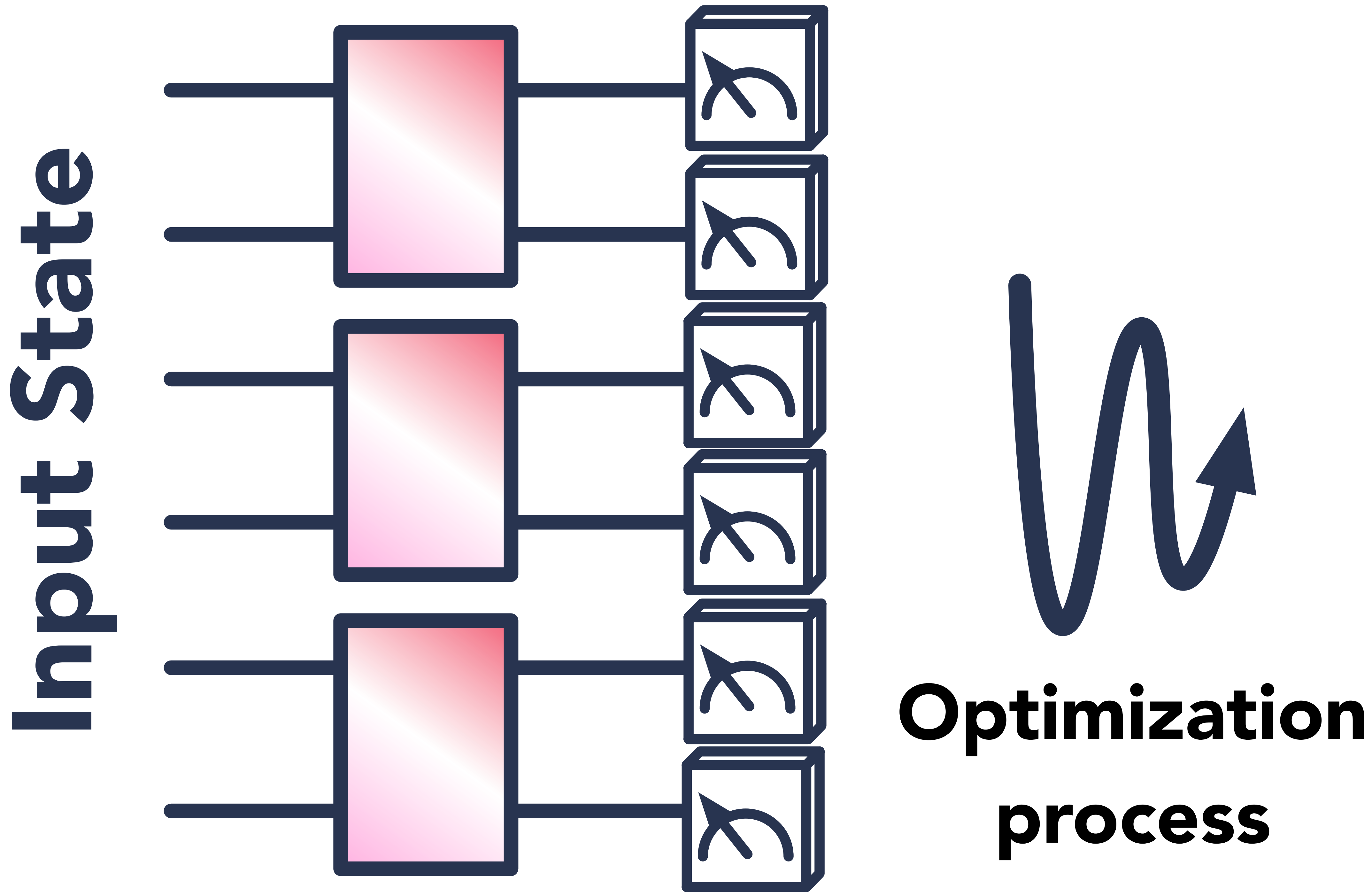
For example:
QCNN

Trainable Quantum Machine



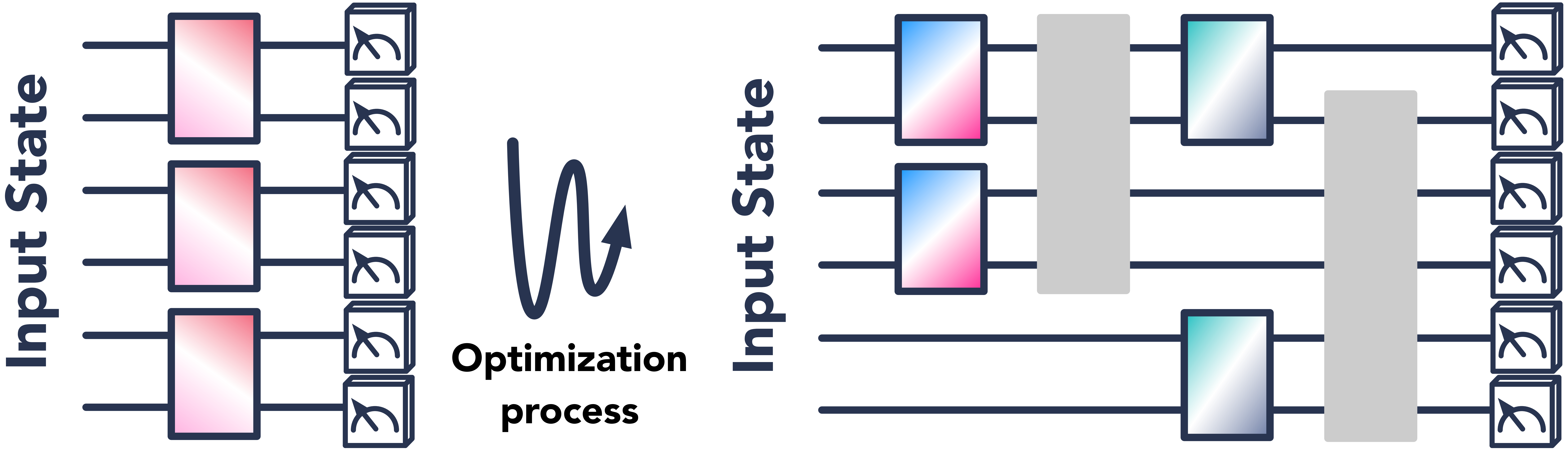
Gate-sharing Variable-structure

Trainable Quantum Machine



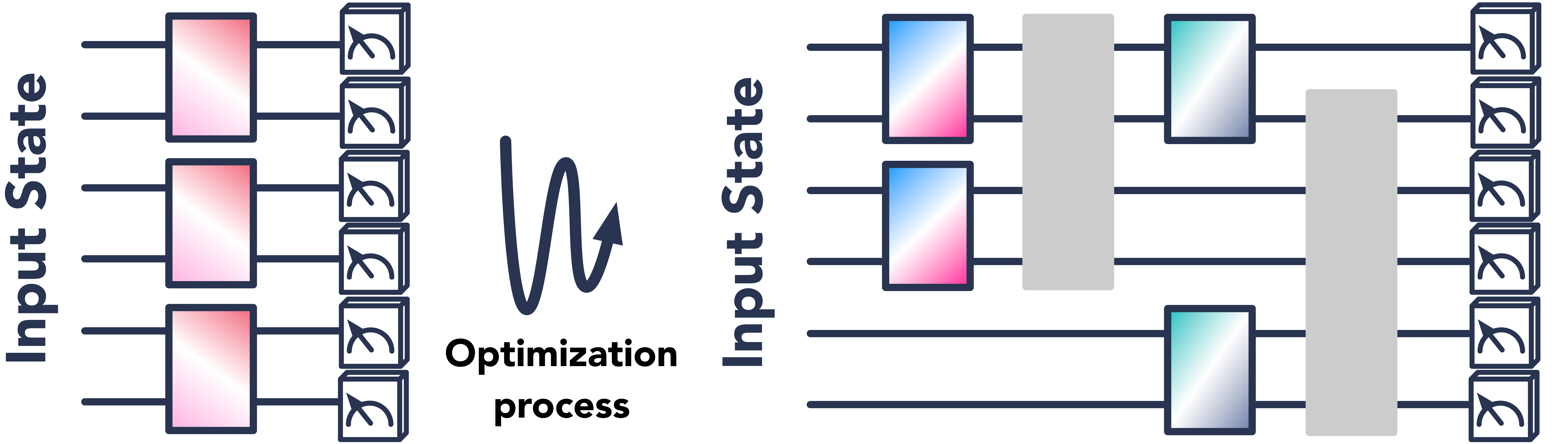
Gate-sharing Variable-structure

Trainable Quantum Machine



Gate-sharing Variable-structure

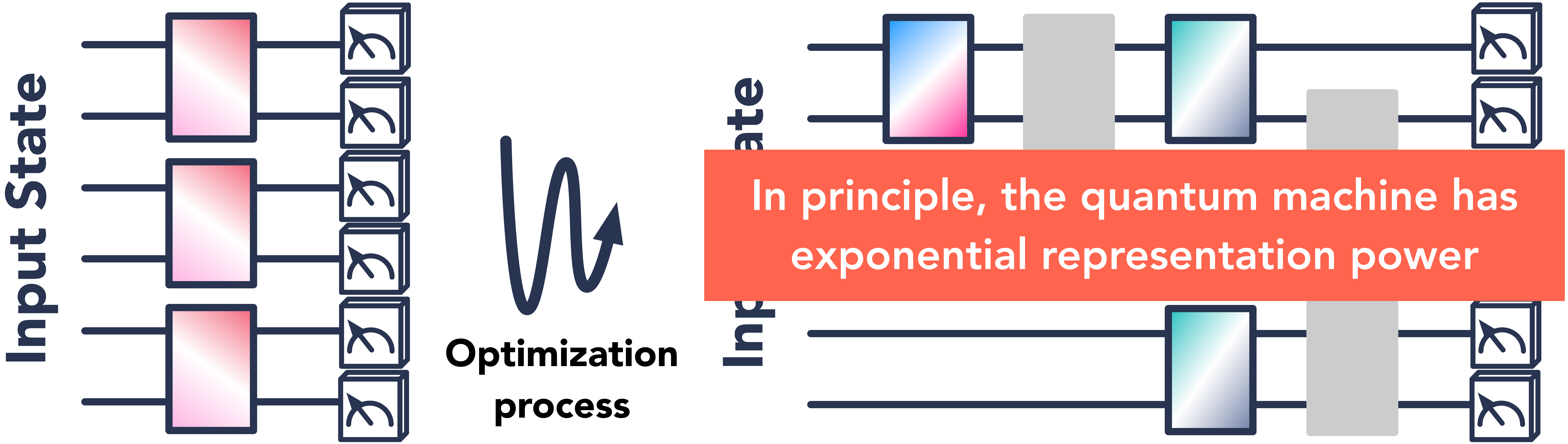
Trainable Quantum Machine



Gate-sharing Variable-structure

E.g., Adaptive VQE

Trainable Quantum Machine



In principle, the quantum machine has exponential representation power

Gate-sharing Variable-structure

E.g., Adaptive VQE

Generalization error

Prediction error − Training error = Generalization error

Error on unseen inputs — Error on training data

Generalization error

- What does generalization error depend on?

Prediction error − Training error = Generalization error

Error on unseen inputs   Error on training data

Generalization error

- What does generalization error depend on?
- Model, data, optimization process, … are all important factors.

Prediction error − Training error = Generalization error

Error on unseen inputs   Error on training data

Generalization error

Some empirical facts:

Generalization error

Some empirical facts:

1. Model: If the trainable machine has many trainable gates described by the same parameters, then generalization error is small.

Generalization error

Some empirical facts:

1. Model: If the trainable machine has many trainable gates described by the same parameters, then generalization error is small.

2. Data: If the data is purely random, the machine can grow to a large size, fit the training data perfectly, but does not generalize.

Generalization error

Some empirical facts:

1. Model: If the trainable machine has many trainable gates described by the same parameters, then generalization error is small.

2. Data: If the data is purely random, the machine can grow to a large size, fit the training data perfectly, but does not generalize.

3. Optimization: If the data is simple, the adaptive optimization finds a good model early. The machine remains small and generalize well.

Generalization error

- What does generalization error depend on?
- Model, data, optimization process, … are all important factors.

Prediction error − Training error = Generalization error

Error on unseen inputs — Error on training data

- We will see a type of generalization error bound for quantum machines.

Generalization error

- A crude but informative characterization of generalization error:

With N training samples, if the trained machine has T trainable gates, $\leq G_T$ possible structures, and each trainable gate is used $\leq M_T$ times,

then generalization error $= \mathcal{O}\left(\sqrt{\frac{T\log(M_T T)}{N}} + \sqrt{\frac{\log(G_T)}{N}}\right)$ w.h.p.

Generalization error

- A crude but informative characterization of generalization error:

With N training samples, if the trained machine has T trainable gates, $\leq G_T$ possible structures, and each trainable gate is used $\leq M_T$ times,

then generalization error $= \mathcal{O}\left(\sqrt{\frac{T\log(M_T T)}{N}} + \sqrt{\frac{\log(G_T)}{N}}\right)$ w.h.p.

1. Model: If the trainable machine has M trainable gates described by the same parameters, then generalization error is small.

Generalization error

$$\mathcal{O}\left(\sqrt{\frac{T\log(M_T T)}{N}} + \sqrt{\frac{\log(G_T)}{N}}\right)$$

With N training samples, if the trained machine has 1 trainable gate, 1 possible structure, and each trainable gate is used $\leq M$ times,

then generalization error $= \mathcal{O}\left(\sqrt{\frac{\log(M)}{N}}\right)$ w.h.p.

1. Model: If the trainable machine has M trainable gates described by the same parameters, then generalization error is small.

Generalization error

$$\mathcal{O}\left(\sqrt{\frac{T\log(M_T T)}{N}} + \sqrt{\frac{\log(G_T)}{N}}\right)$$

With N training samples, if the trained machine has M trainable gates, 1 possible structure, and each trainable gate is used 1 times,

then generalization error $= \mathcal{O}\left(\sqrt{\frac{M}{N}}\right)$ w.h.p.

1. Model: If the trainable machine has M trainable gates described by different parameters, then generalization error is small if $M \ll N$.

Generalization error

$$\mathcal{O}\left(\sqrt{\frac{T\log(M_T T)}{N}}+\sqrt{\frac{\log(G_T)}{N}}\right)$$

With N training samples, if the trained machine has N trainable gates, 1 possible structure, and each trainable gate is used 1 times,

then generalization error $= \mathcal{O}\left(\sqrt{\frac{N}{N}}\right) = \mathcal{O}(1)$ w.h.p.

2. Data: If the data is purely random, the machine can grow to a large size, fit the training data perfectly, but does not generalize.

Generalization error

$$\mathcal{O}\left(\sqrt{\frac{T\log(M_T T)}{N}} + \sqrt{\frac{\log(G_T)}{N}}\right)$$

With N samples, if the trained machine has $\mathcal{O}(1)$ trainable gates, $\mathcal{O}(1)$ possible structures, and each trainable gate is used $\mathcal{O}(1)$ times,

then generalization error $= \mathcal{O}\left(\sqrt{\frac{1}{N}}\right)$ w.h.p.

3. Optimization: If the data is simple, the adaptive optimization finds a good model early. The machine remains small and generalize well.

Generalization error

- A crude but informative characterization of generalization error:

With N training samples, if the trained machine has T trainable gates, $\leq G_T$ possible structures, and each trainable gate is used $\leq M_T$ times,

then generalization error $= \mathcal{O}\left(\sqrt{\frac{T\log(M_T T)}{N}} + \sqrt{\frac{\log(G_T)}{N}}\right)$ w.h.p.

Generalization error = Prediction error − Training error

Prediction error: Error on unseen inputs

Training error: Error on training data

Generalization error

Board Time

- A crude but informative characterization of generalization error:

With N training samples, if the trained machine has T trainable gates, $\leq G_T$ possible structures, and each trainable gate is used $\leq M_T$ times,

then generalization error $= \mathcal{O}\left(\sqrt{\frac{T\log(M_T T)}{N}} + \sqrt{\frac{\log(G_T)}{N}}\right)$ w.h.p.

Generalization error = Prediction error − Training error

Prediction error: Error on unseen inputs

Training error: Error on training data

Concentration

Board Time

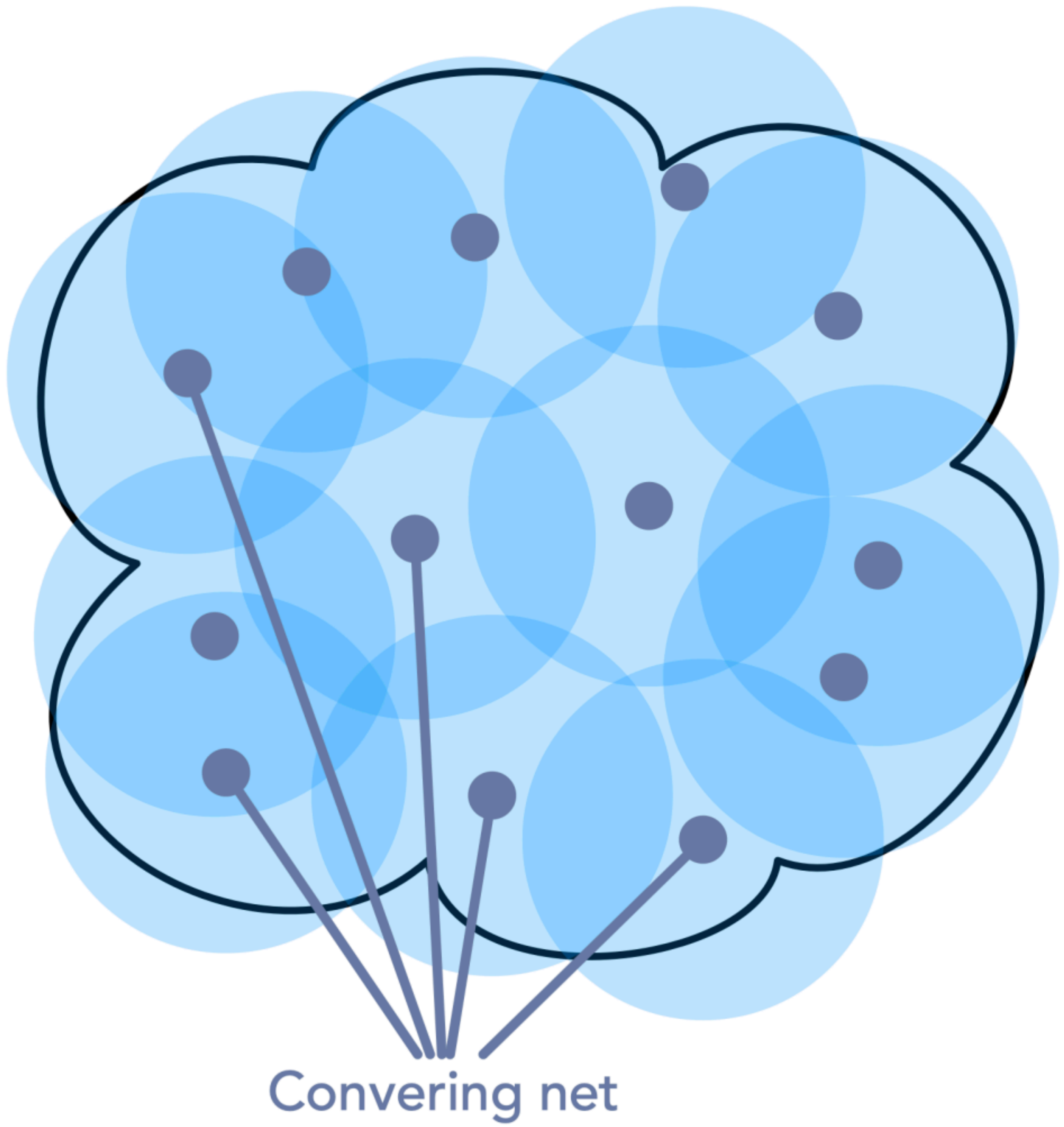
Let $X_1, \ldots, X_N$ be independent and identically distributed (i.i.d.) random variable in $[0,1]$. We have

$$\Pr\left[\mathbb{E}[X_i] \leq \frac{1}{N}\sum_{i=1}^{N} X_i + \epsilon\right] \geq 1 - \exp(-2N\epsilon^2)$$

This is known as *Hoeffding's concentration inequality*.

Covering Net

Board Time



To cover a trainable 2-qubit gate, we only need $(1/\epsilon)^{\mathcal{O}(1)}$ ϵ-radius $\|\cdot\|_\infty$-norm ball.

How many balls are needed to cover all quantum machines with T trainable 2-qubit gates?

Beyond training distribution

- We now have a good understanding for generalization error when the **training data** come from the same distribution as the **unseen inputs**.

Prediction error = Training error + Generalization error

Error on unseen inputs — Error on training data

- This kinds of generalization based on I.I.D. samples is useful.

Beyond training distribution

- We now have a good understanding for generalization error when the **training data** come from the same distribution as the **unseen inputs**.

Prediction error = Training error + Generalization error

Error on unseen inputs — Error on training data

- However, ideally, we want to generalize beyond the training distribution.

Beyond training distribution

- Suppose the training data only consists of **product state inputs**.
  Could the quantum machine predict well for **entangled state inputs**?

Prediction error = Training error + Generalization error

Error on unseen inputs   Error on training data

Beyond training distribution

- Suppose the training data only consists of **product state inputs**.
  Could the quantum machine predict well for **entangled state inputs**?

Prediction error = Training error + Generalization error

Error on unseen inputs   Error on training data

- While this seems impossible, one can actually do this!
- This ability is known as *"out-of-distribution generalization"*.

Beyond training distribution

- This theorem holds when training samples are random product states; But the prediction is on random entangled states.

With N training samples, if the trained machine has T trainable gates, $\leq G_T$ possible structures, and each trainable gate is used $\leq M_T$ times,

then generalization error $= \mathcal{O}\left(\sqrt{\frac{T\log(M_T T)}{N}} + \sqrt{\frac{\log(G_T)}{N}}\right)$ w.h.p.

Beyond training distribution

- This theorem holds when training samples are **random product states**; But the prediction is on **random entangled states**.

With N training samples, if the trained machine has T trainable gates, $\leq G_T$ possible structures, and each trainable gate is used $\leq M_T$ times,

then generalization error $= \mathcal{O}\left(\sqrt{\frac{T\log(M_T T)}{N}} + \sqrt{\frac{\log(G_T)}{N}}\right)$ w.h.p.

Board Time

Equivalence of predictions

Let $\mathscr{D}_1, \mathscr{D}_2$ be two distributions over n-qubit states, such that the distributions are *locally-scrambled*.

$$0.5\ (\textbf{prediction error under } \mathscr{D}_2)$$
$$\leq \textbf{prediction error under } \mathscr{D}_1 \leq$$
$$2\ (\textbf{prediction error under } \mathscr{D}_2)$$

Constraints are from the structure of unitaries.

Board Time

Relevant works

Generalization in QML
from few training data

[This tutorial + numerics]

Relevant works

Generalization in QML
from few training data

[This tutorial + numerics]

Out-of-distribution generalization
in learning quantum dynamics

[This tutorial + numerics]

Relevant works

Generalization in QML
from few training data

[This tutorial + numerics]

Learning quantum states and unitaries
of bounded gate complexity

[Covering-net learning is optimal]

Out-of-distribution generalization
in learning quantum dynamics

[This tutorial + numerics]

Relevant works

Generalization in QML from few training data

[This tutorial + numerics]

Learning quantum states and unitaries of bounded gate complexity

[Covering-net learning is optimal]

Out-of-distribution generalization in learning quantum dynamics

[This tutorial + numerics]

Understanding QML also requires rethinking generalization

[Looking at model class alone is not enough]

Take home message

- How to understand prediction error of trainable quantum machines?

Prediction error = Training error + Generalization error

Error on unseen inputs — Error on training data

- Structure of quantum mechanics imply bounded generalization error:

(A) Train well $\Longrightarrow$ predict well for trainable quantum machines

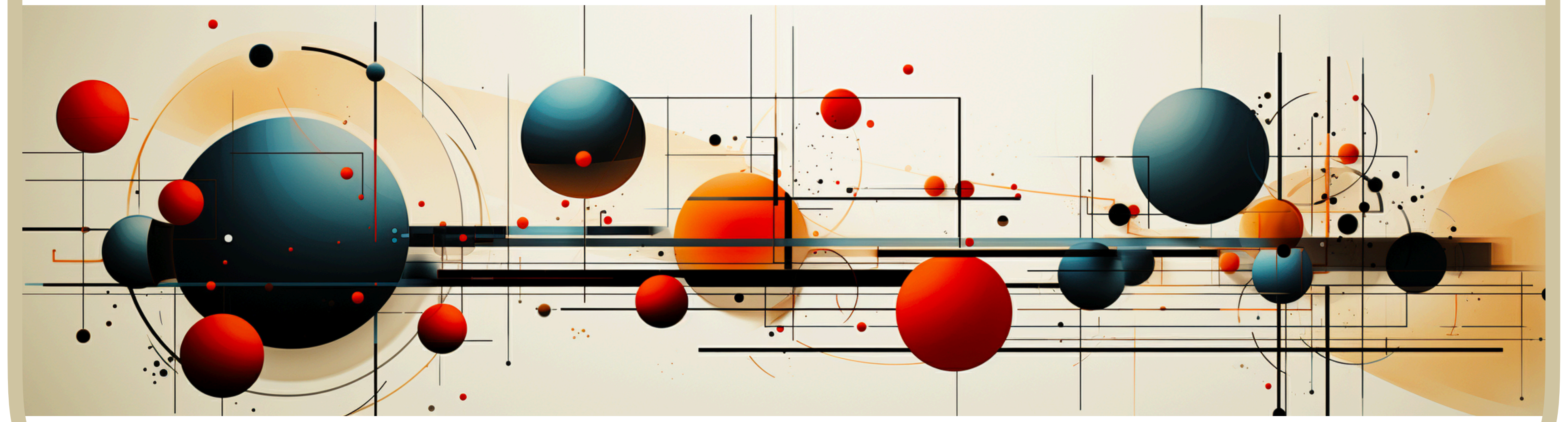
(B) Train well on product states $\Longrightarrow$ predict well on entangled states

Overview

Learning theory for quantum machines

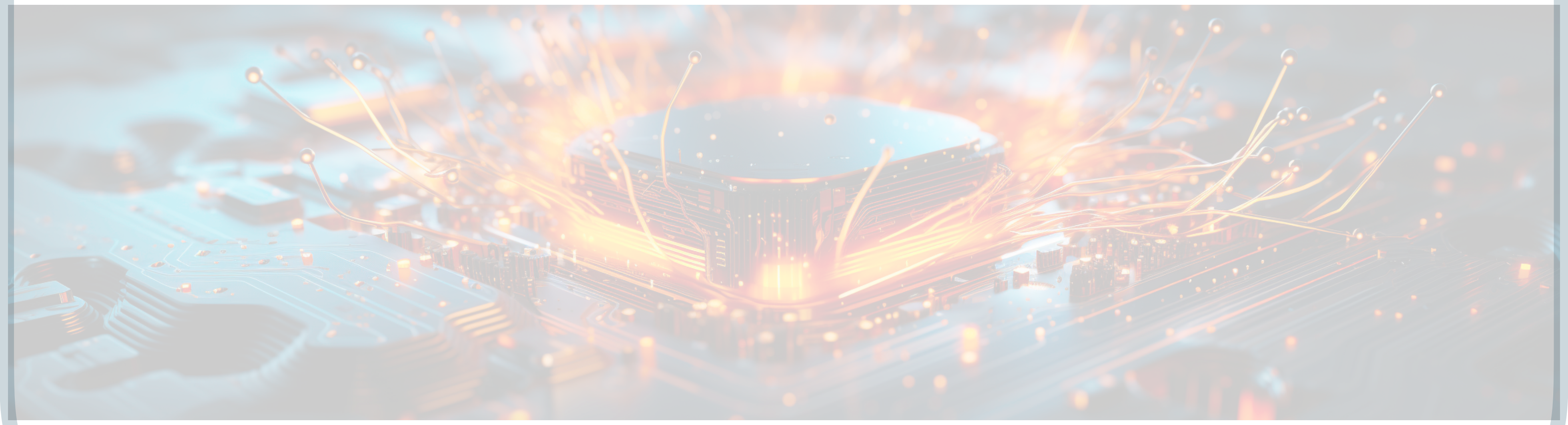
Foundation

How well can quantum machines predict?
How good is the generalization ability of quantum machines?



Quantum advantage

What can quantum machines learn that classical machines cannot?
How big can the advantage be?



Overview

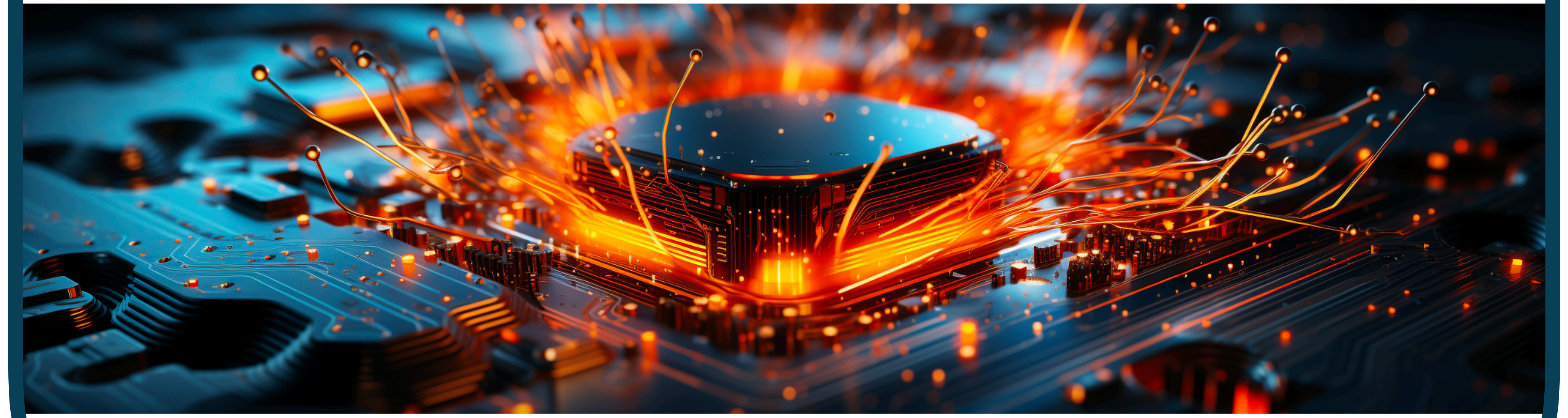
Learning theory for quantum machines

Foundation

How well can quantum machines predict?
How good is the generalization ability of quantum machines?

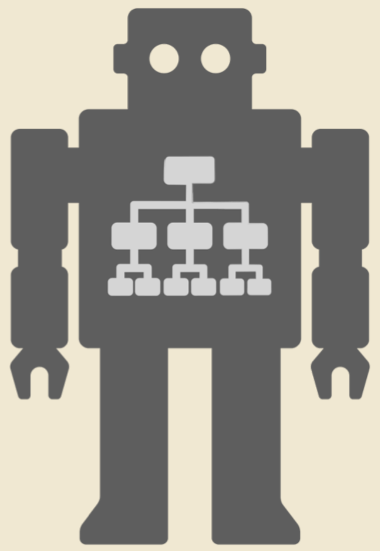
Quantum advantage

What can quantum machines learn that classical machines cannot?
How big can the advantage be?



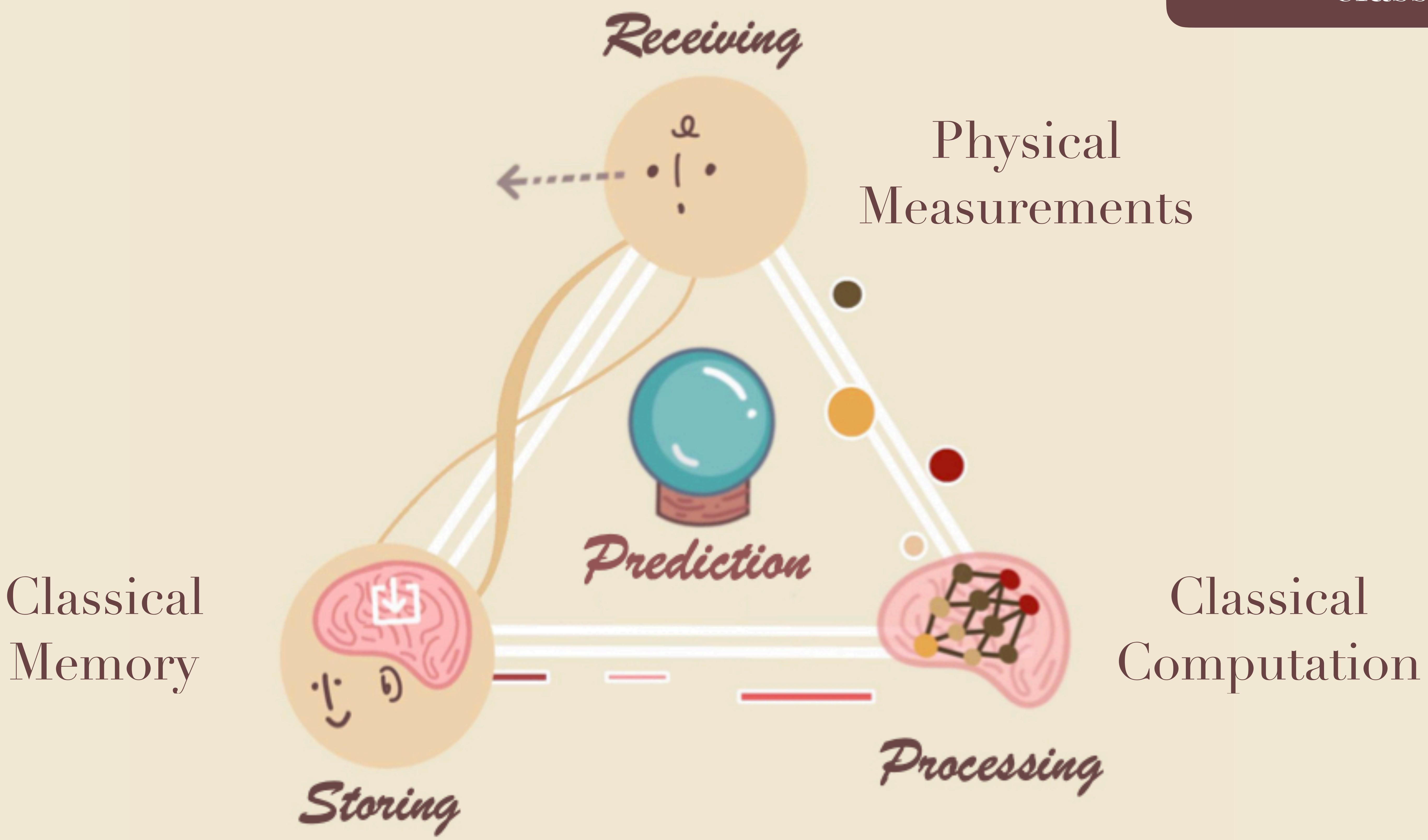
Quantum advantage

- When can quantum machines predict better than classical machines?

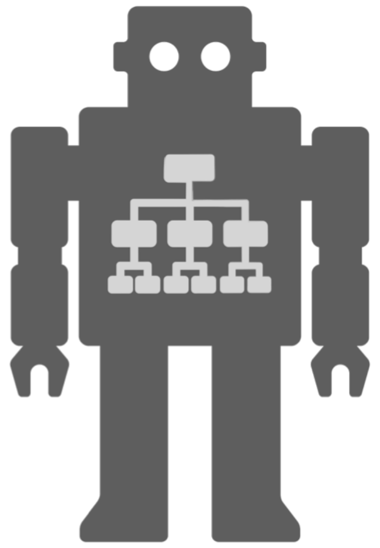


Classical agent

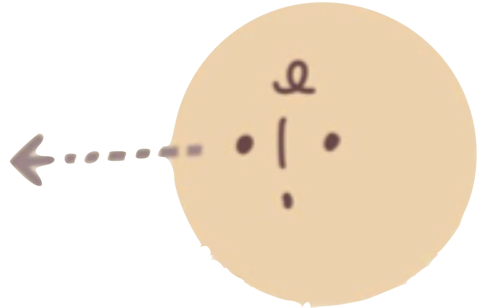
Receive, process, and store classical information



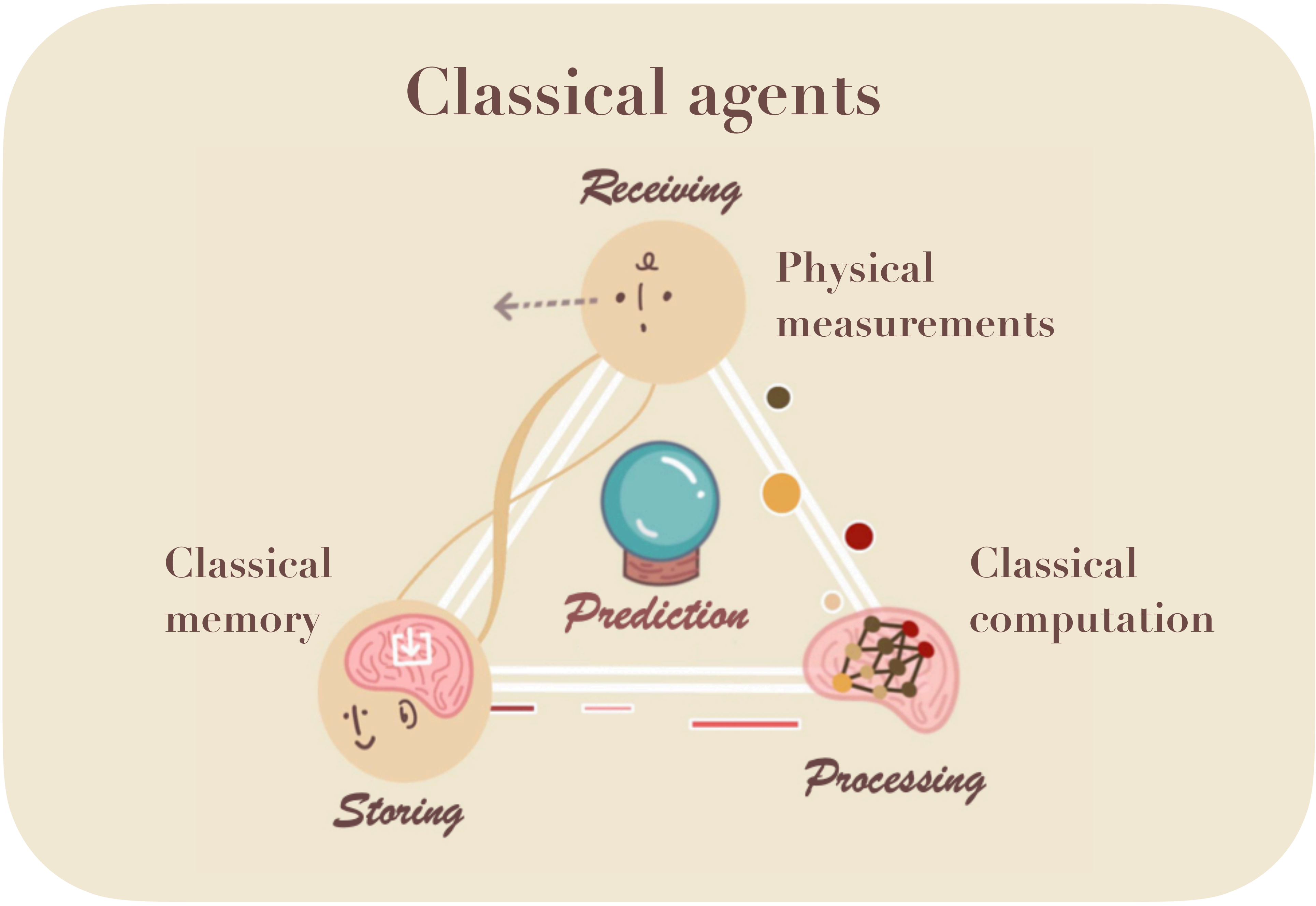
[HKP21] Huang, Kueng, Preskill. Information-theoretic bounds on quantum advantage in machine learning, *Physical Review Letters*, 2021.
[CCHL21] Chen, Cotler, Huang, Li. Exponential separations in learning with and without quantum memory, *FOCS*, 2021.
[HBC+] Huang, et all. Quantum advantage in learning from experiments, *Science*, 2022.



Learning physical dynamics

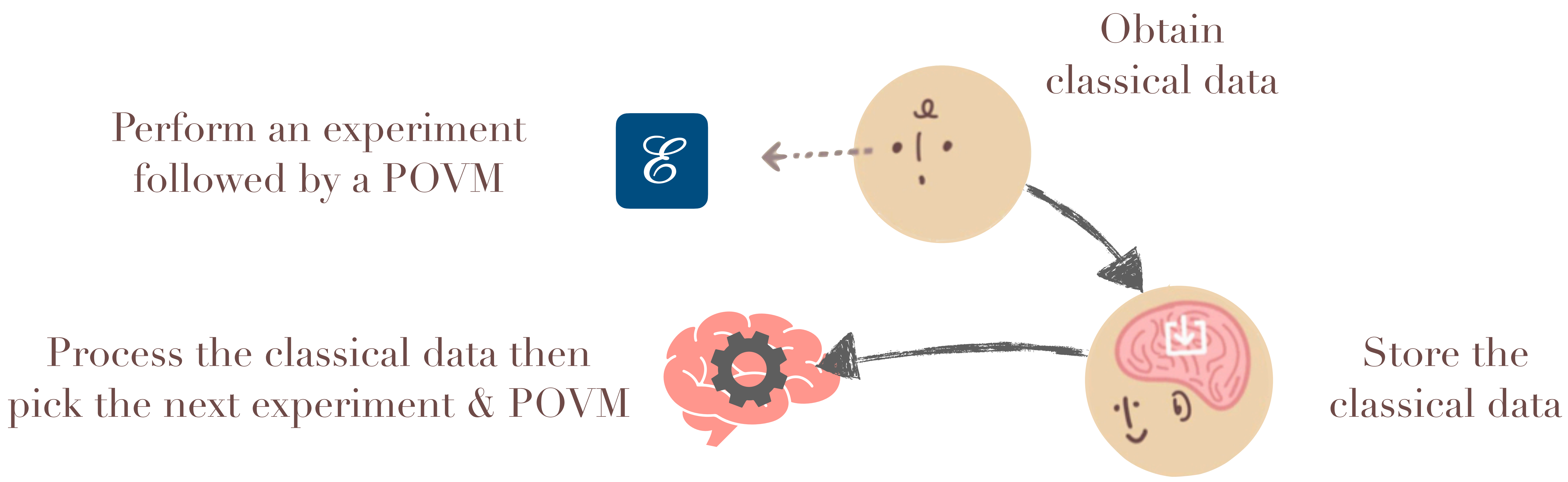


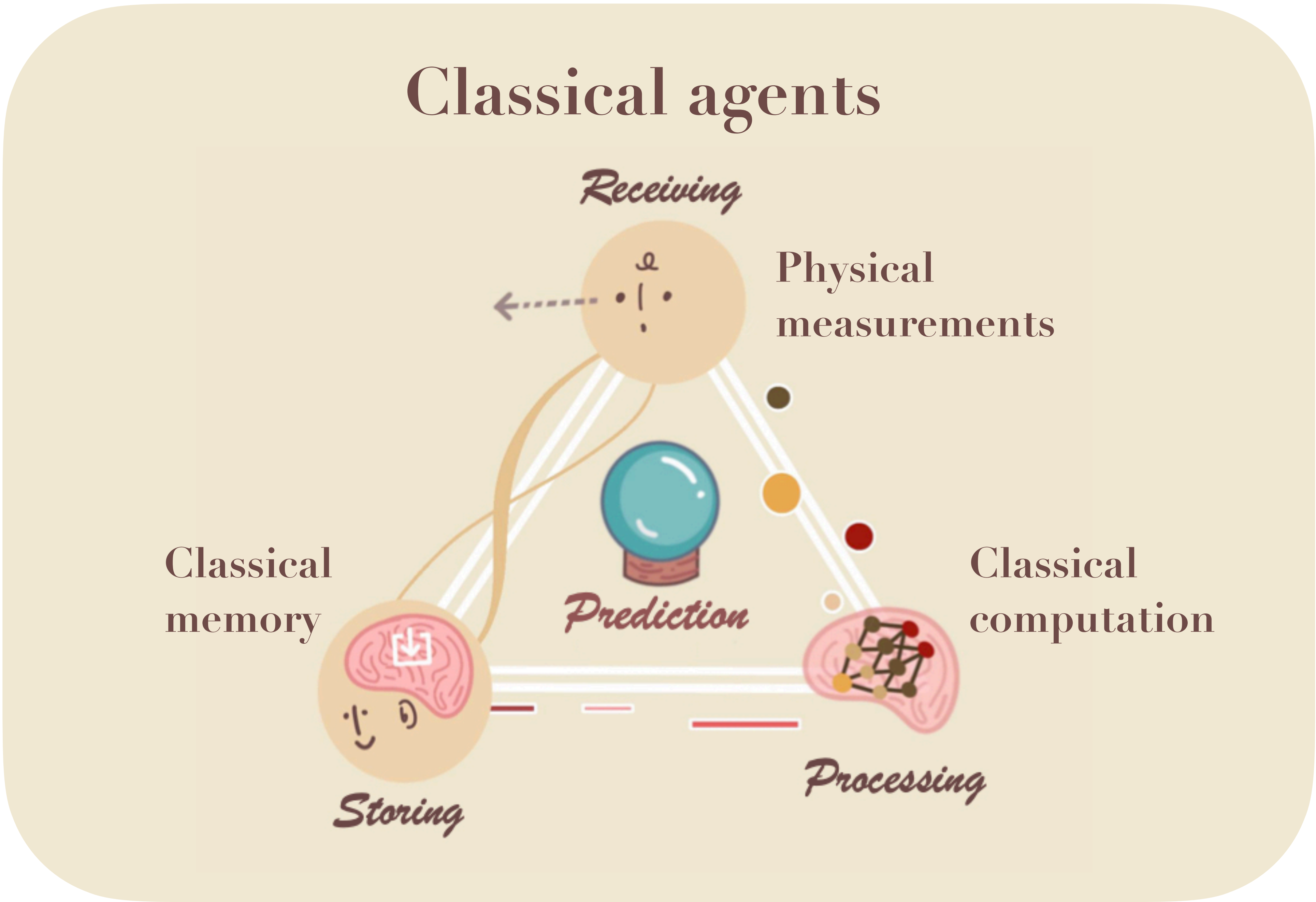
- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.



Learning physical dynamics

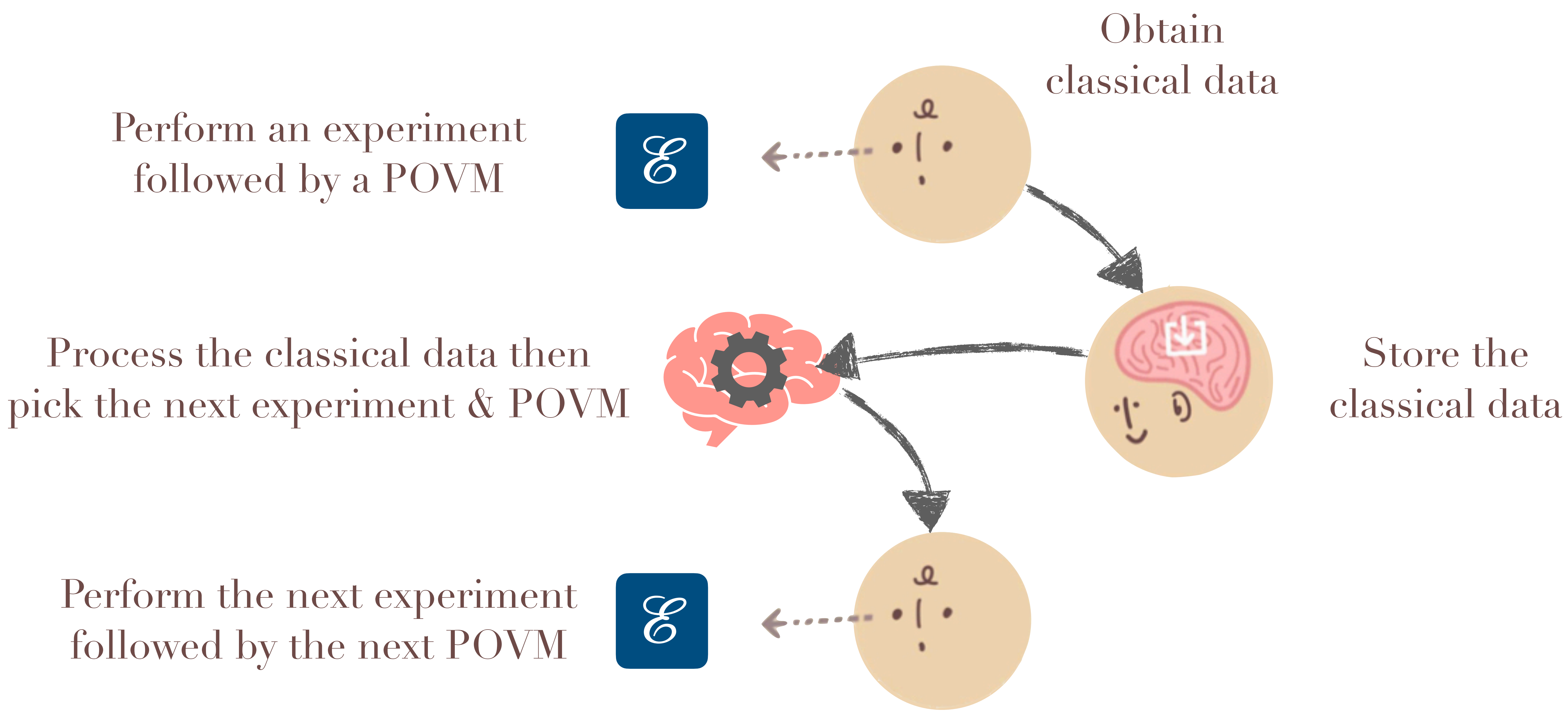
- We can also consider learning about an unknown physical dynamics $\mathcal{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathcal{E}$.

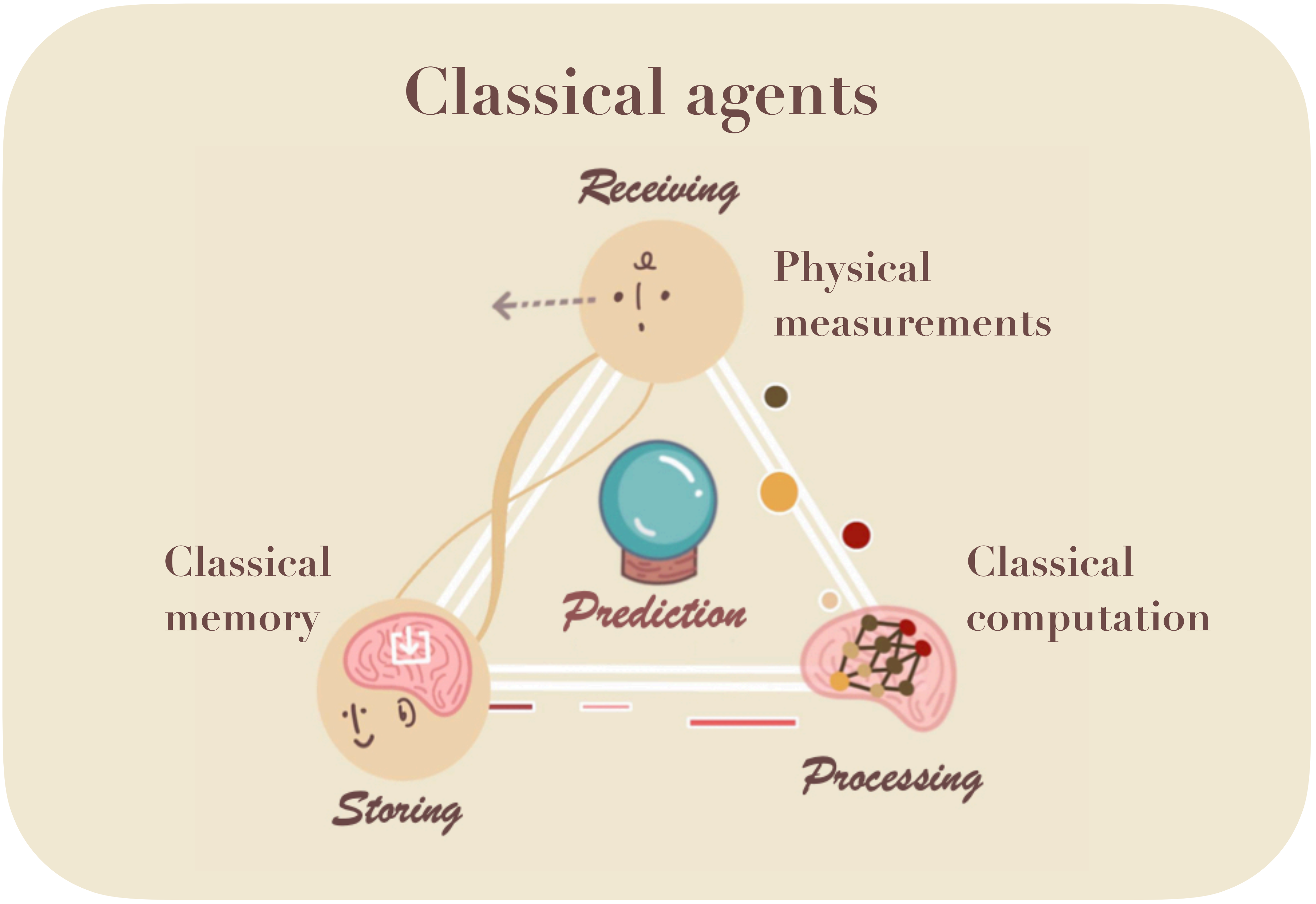




Learning physical dynamics

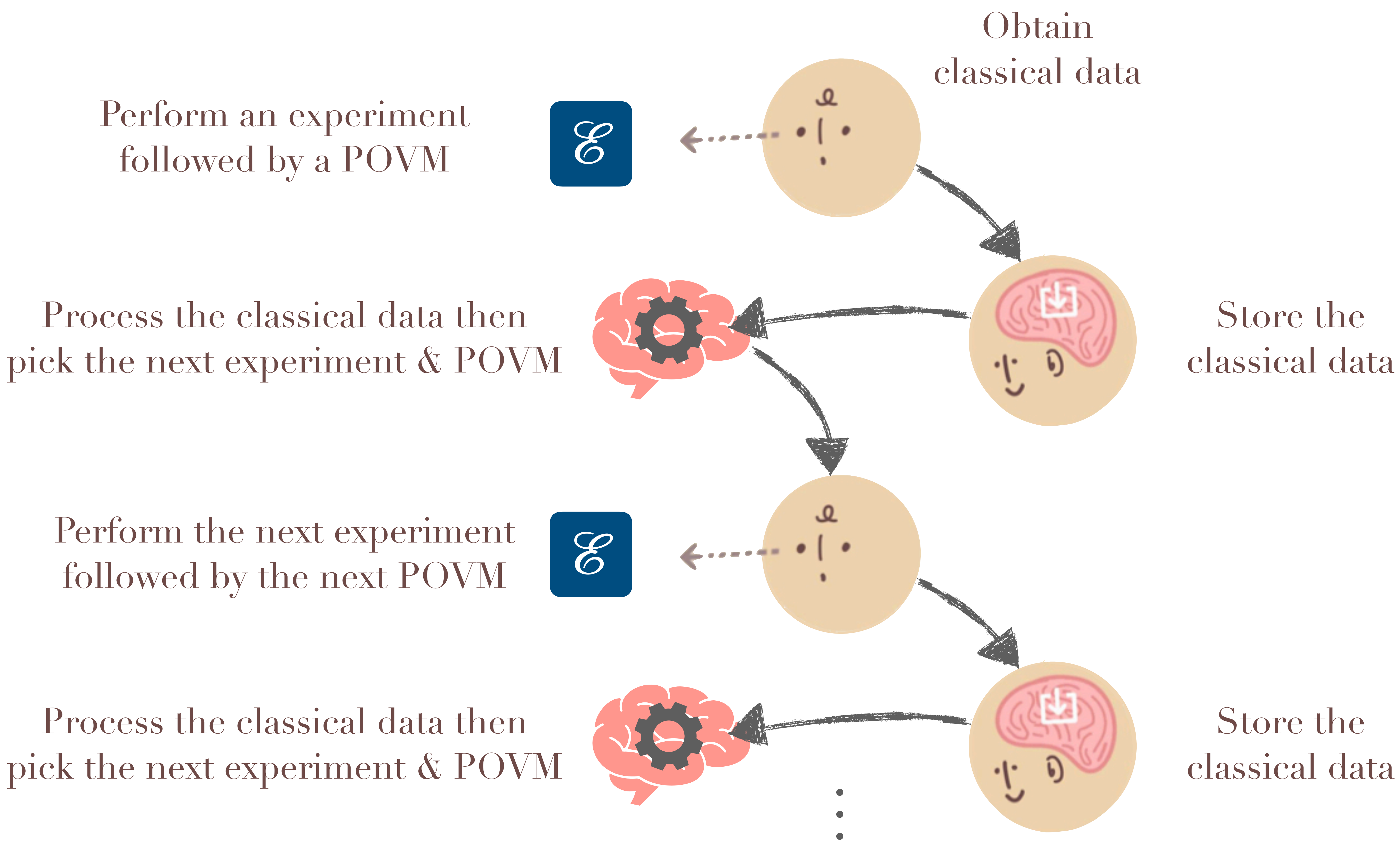
- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.

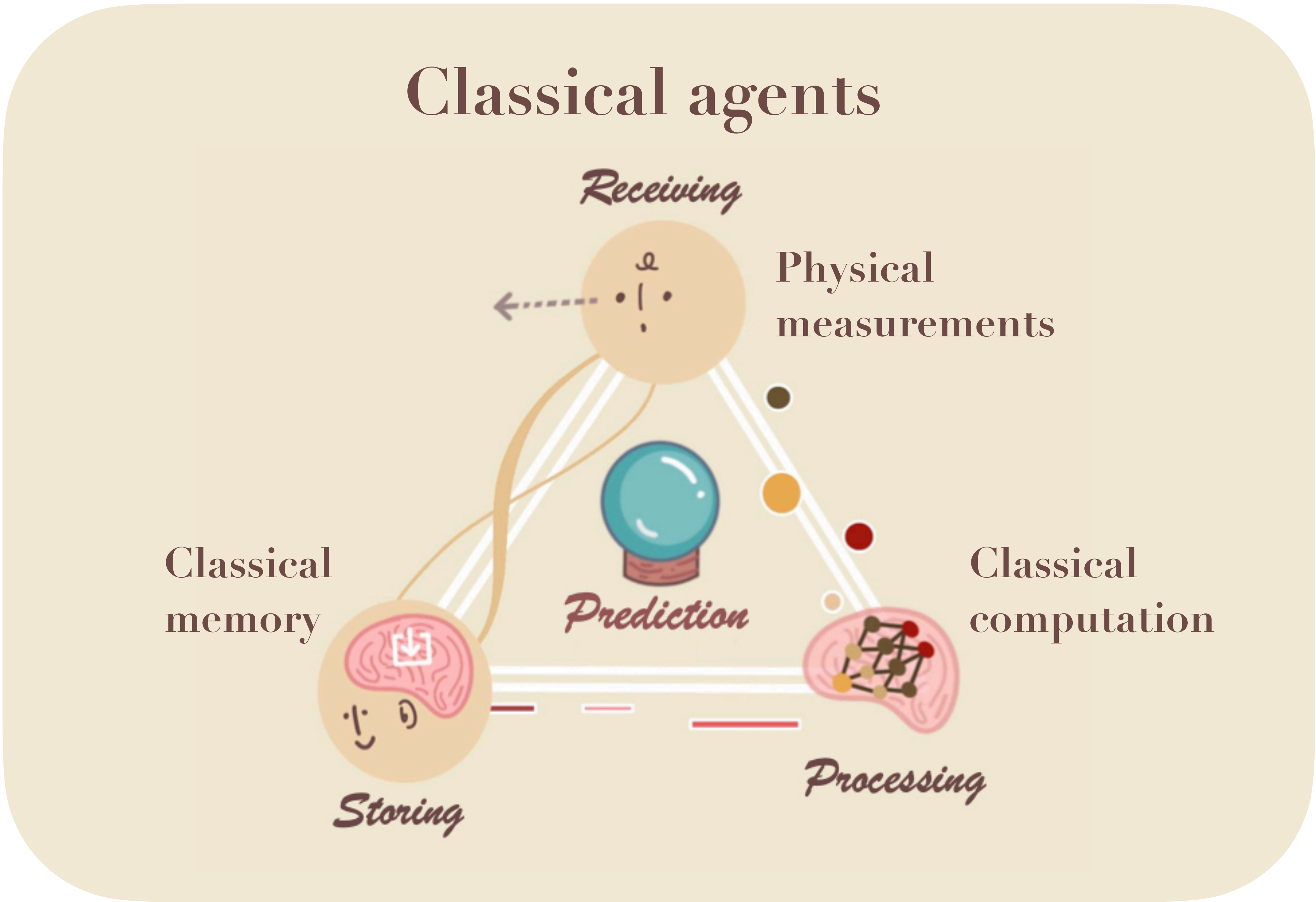




Learning physical dynamics

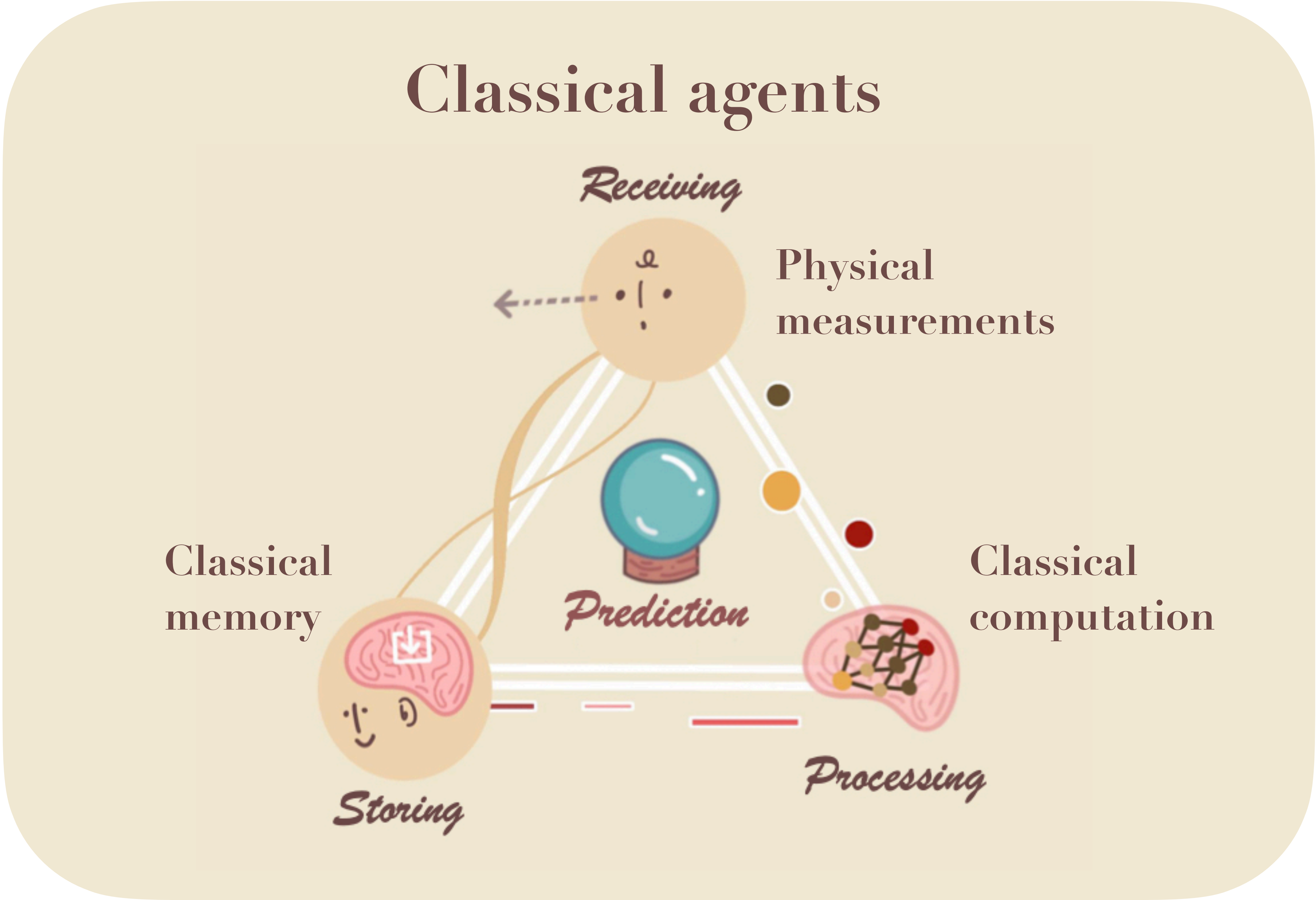
- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.





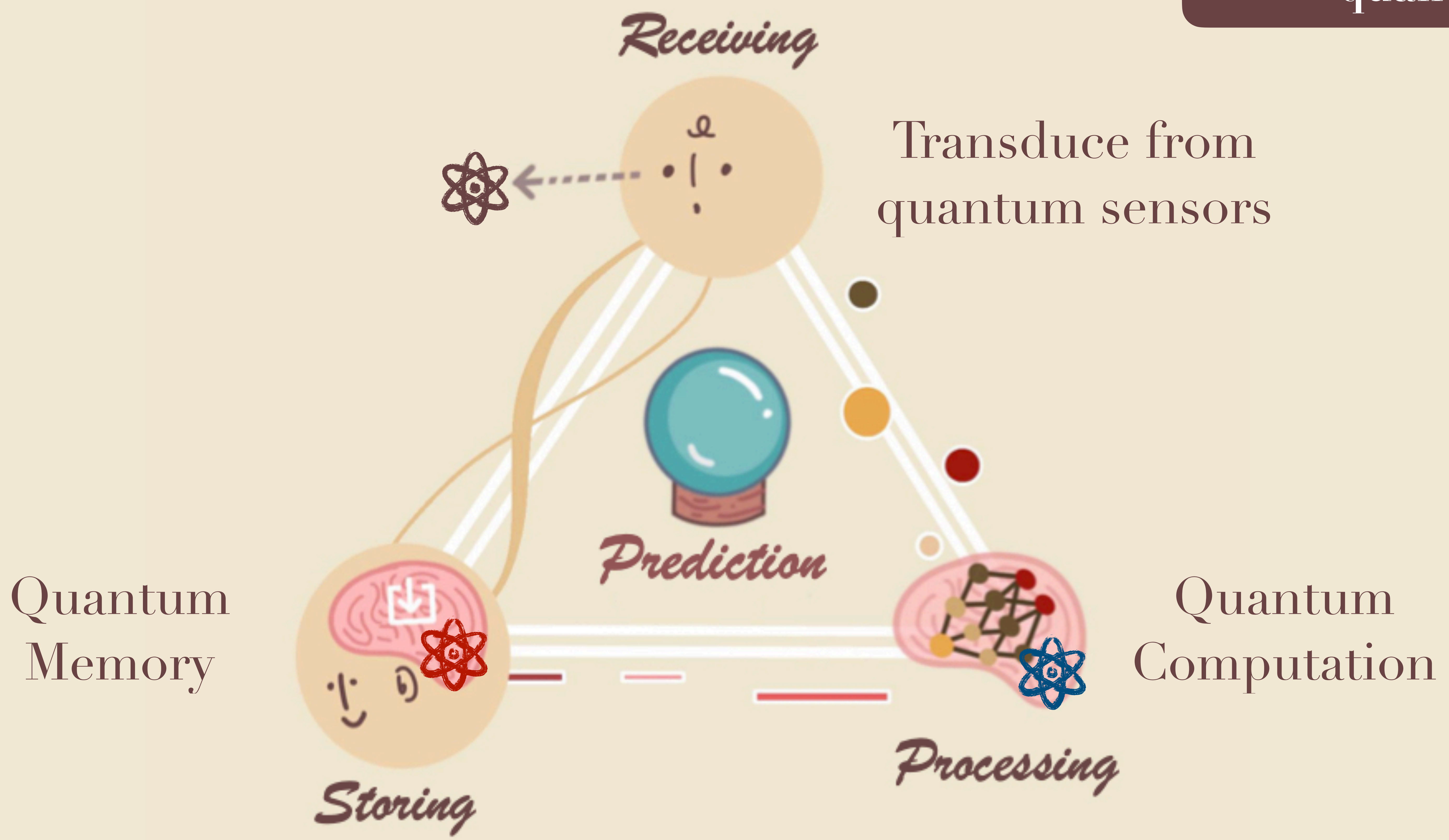
Learning physical dynamics

- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.

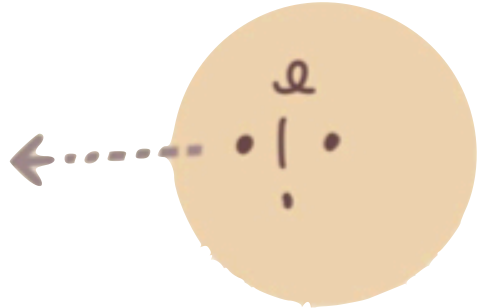


Quantum agent

Receive, process, and store quantum information

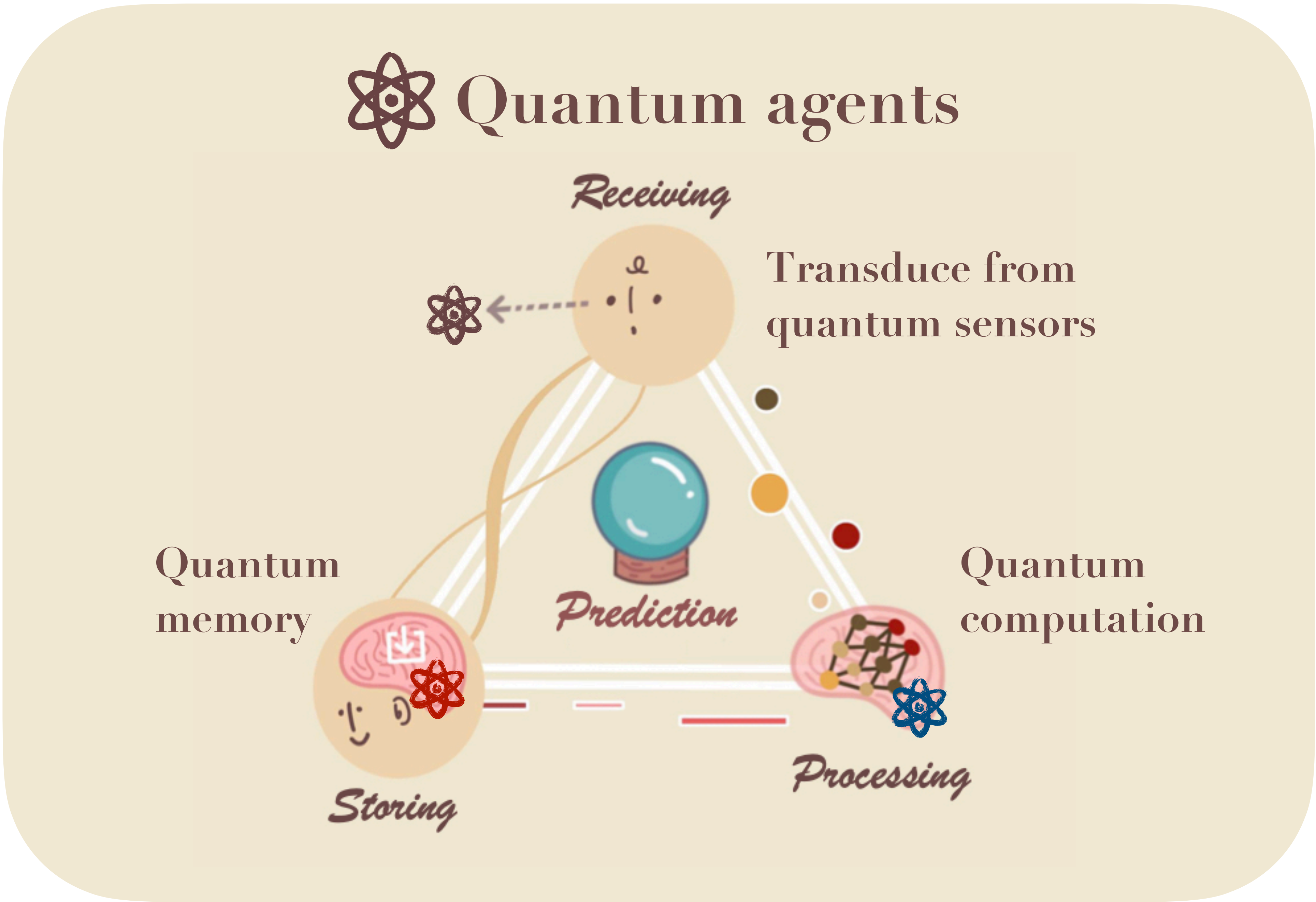


[HKP21] Huang, Kueng, Preskill. Information-theoretic bounds on quantum advantage in machine learning, *Physical Review Letters*, 2021.
[CCHL21] Chen, Cotler, Huang, Li. Exponential separations in learning with and without quantum memory, *FOCS*, 2021.
[HBC+] Huang, et all. Quantum advantage in learning from experiments, *Science*, 2022.



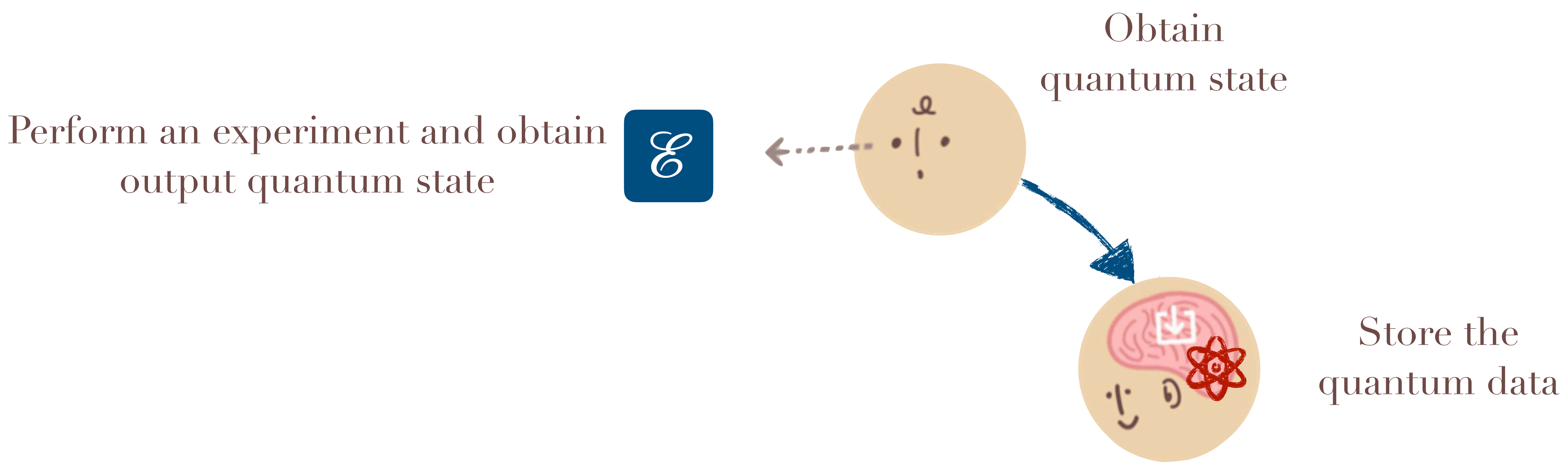
Learning physical dynamics

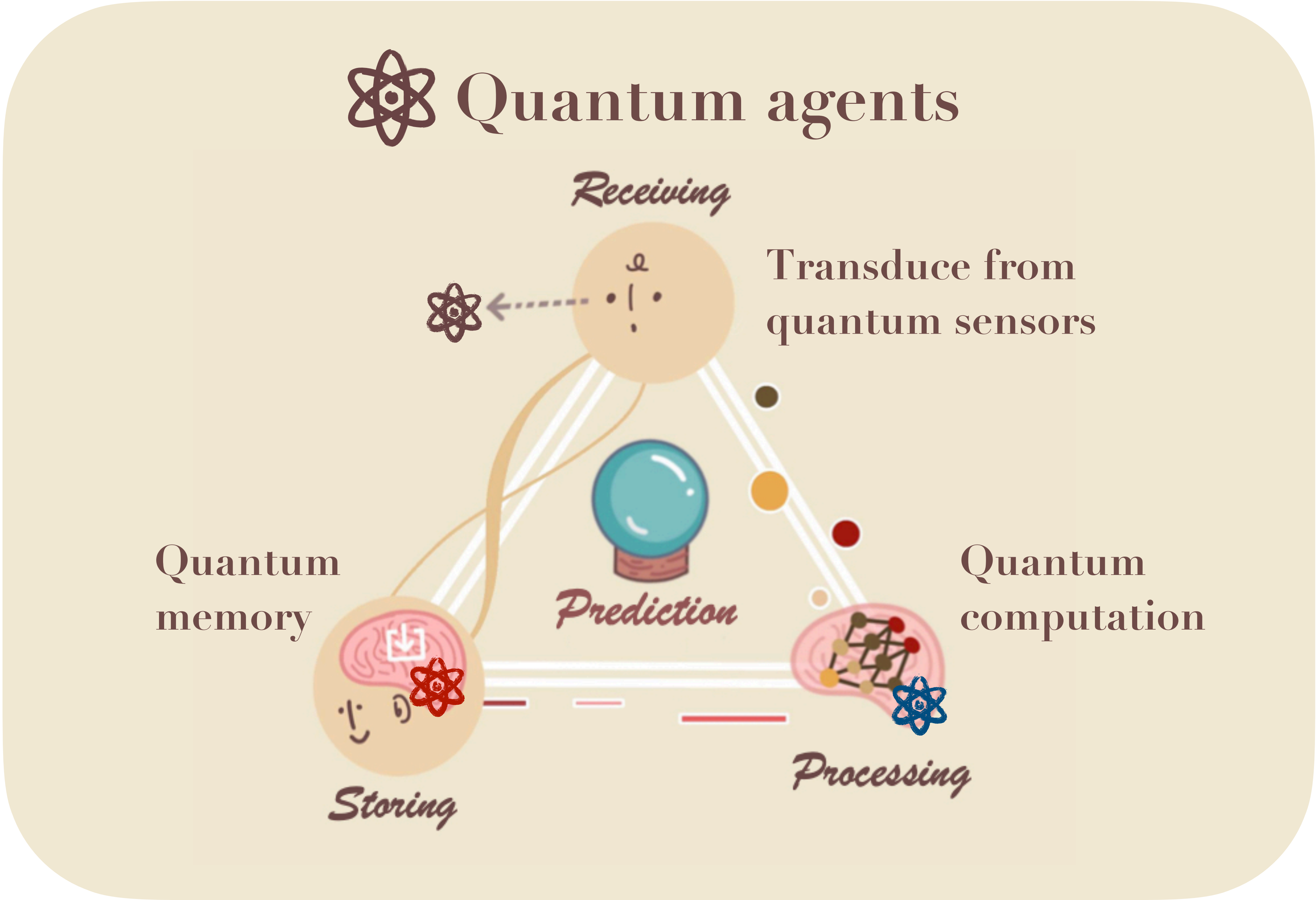
- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.



Learning physical dynamics

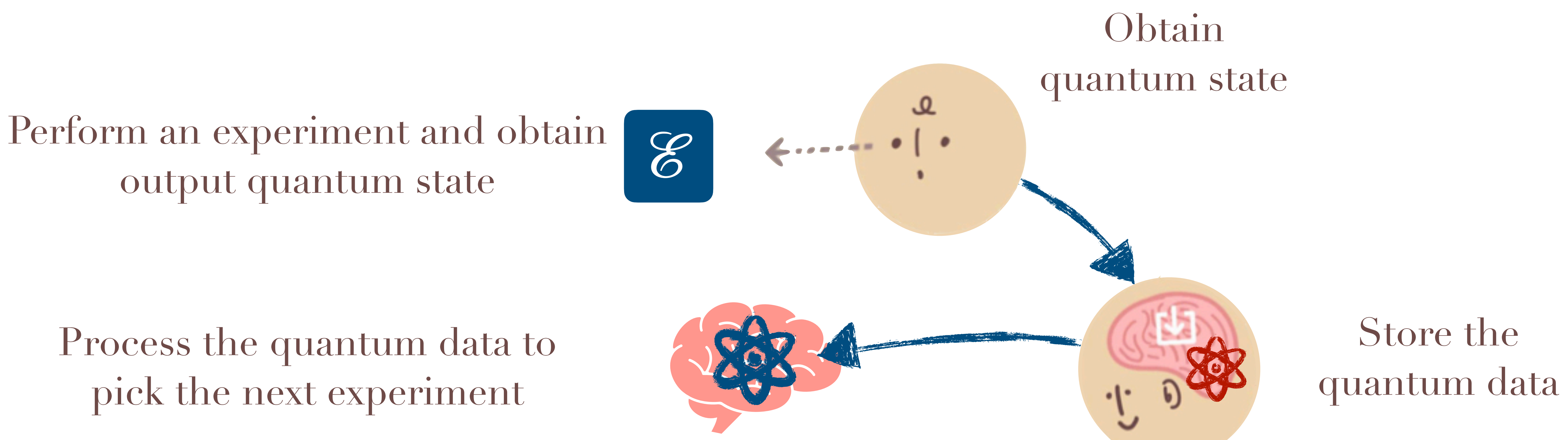
- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.

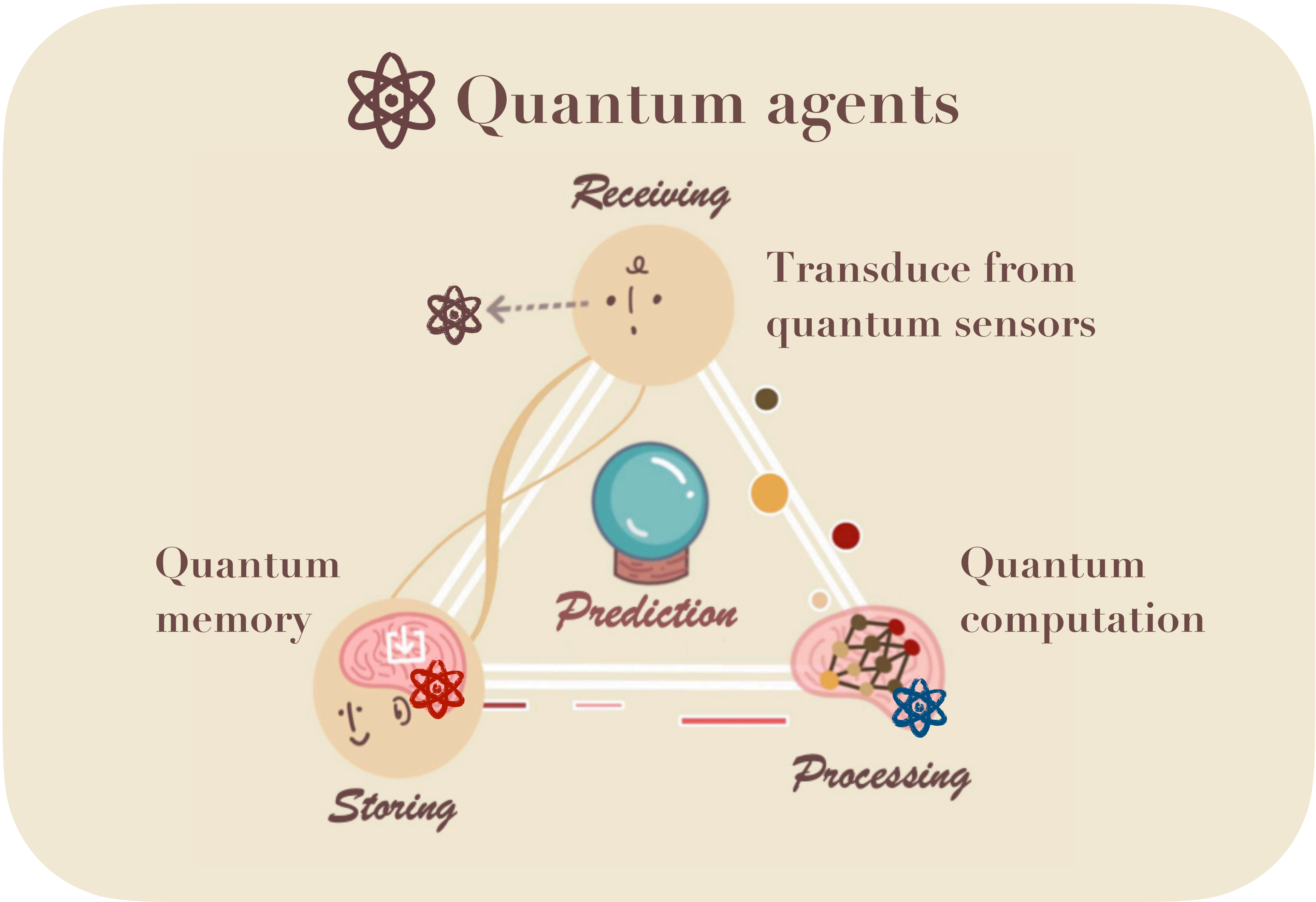




Learning physical dynamics

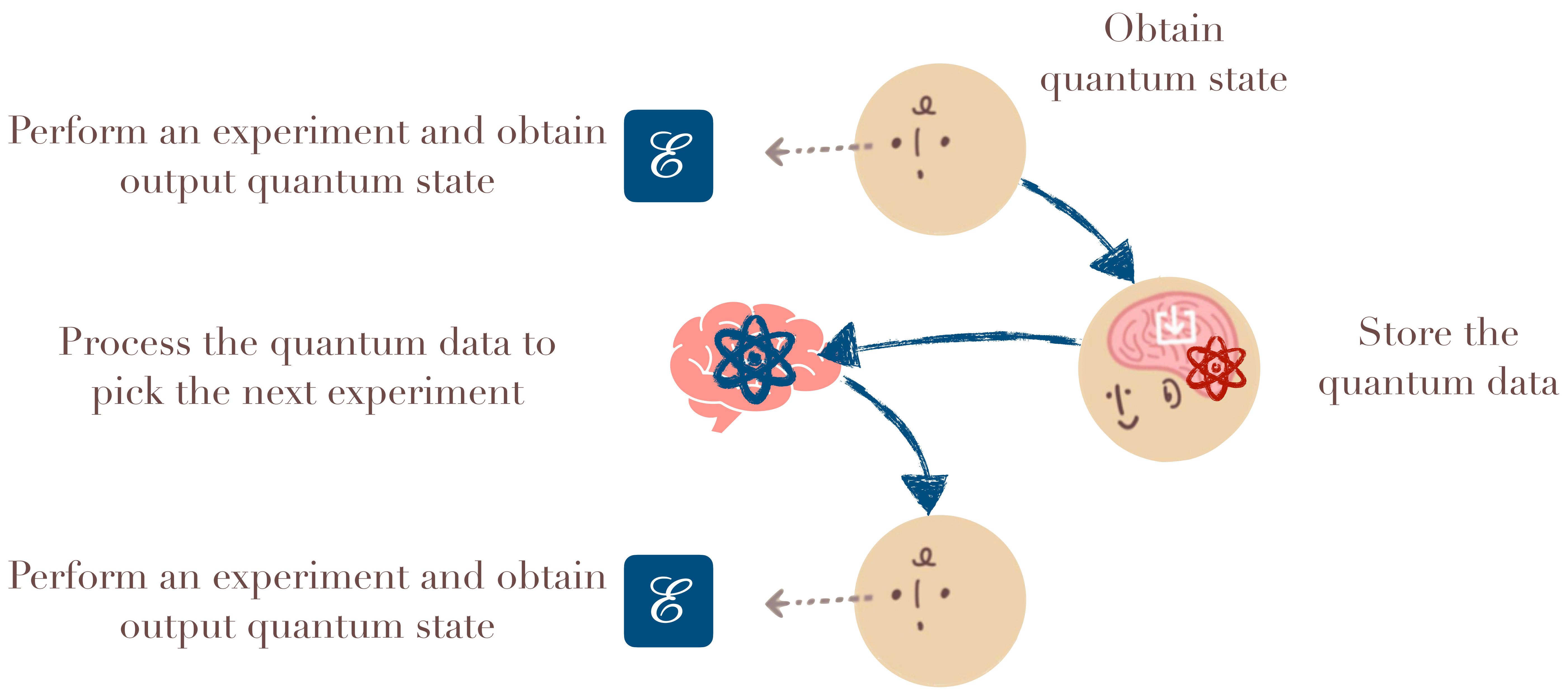
- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.

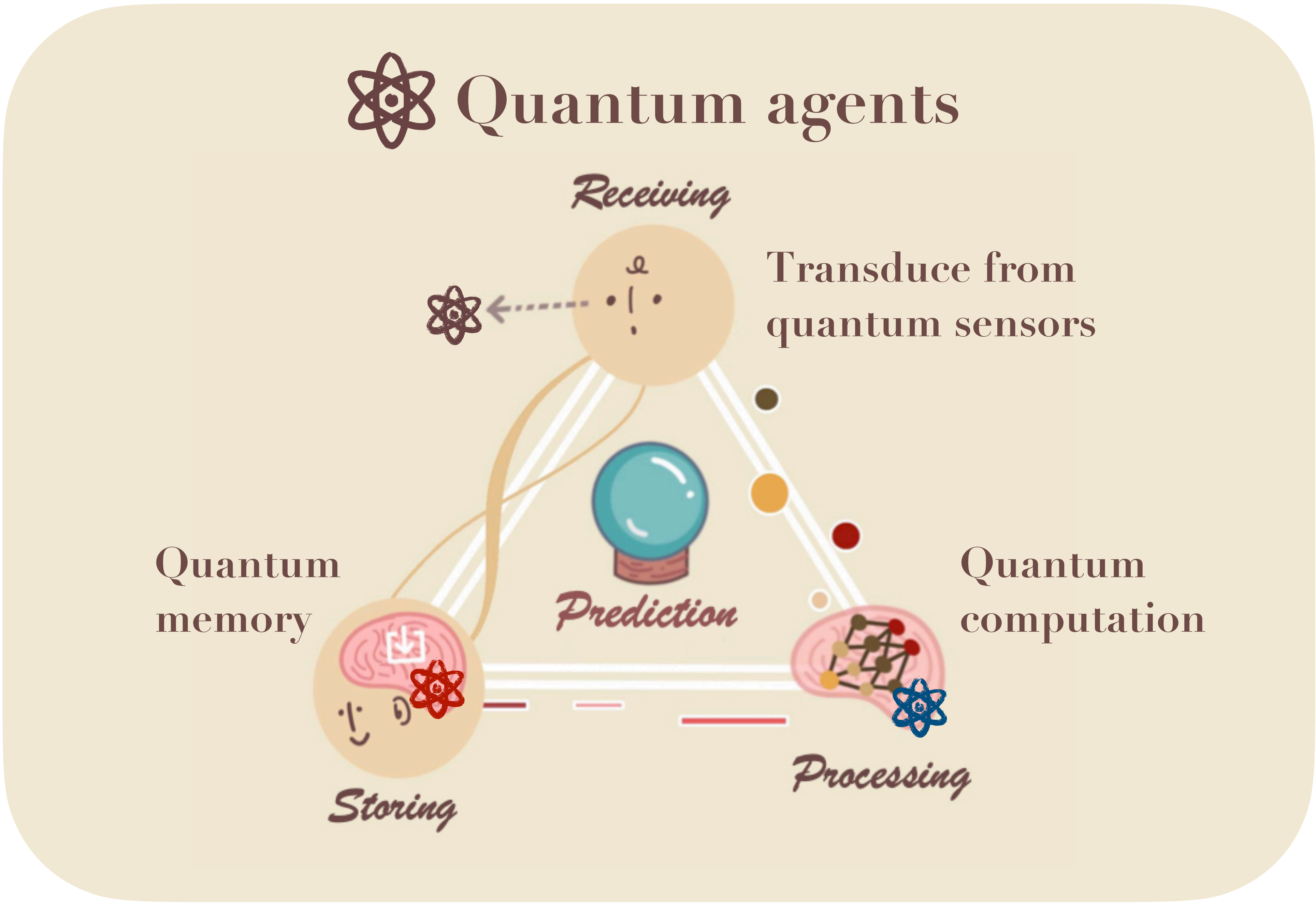




Learning physical dynamics

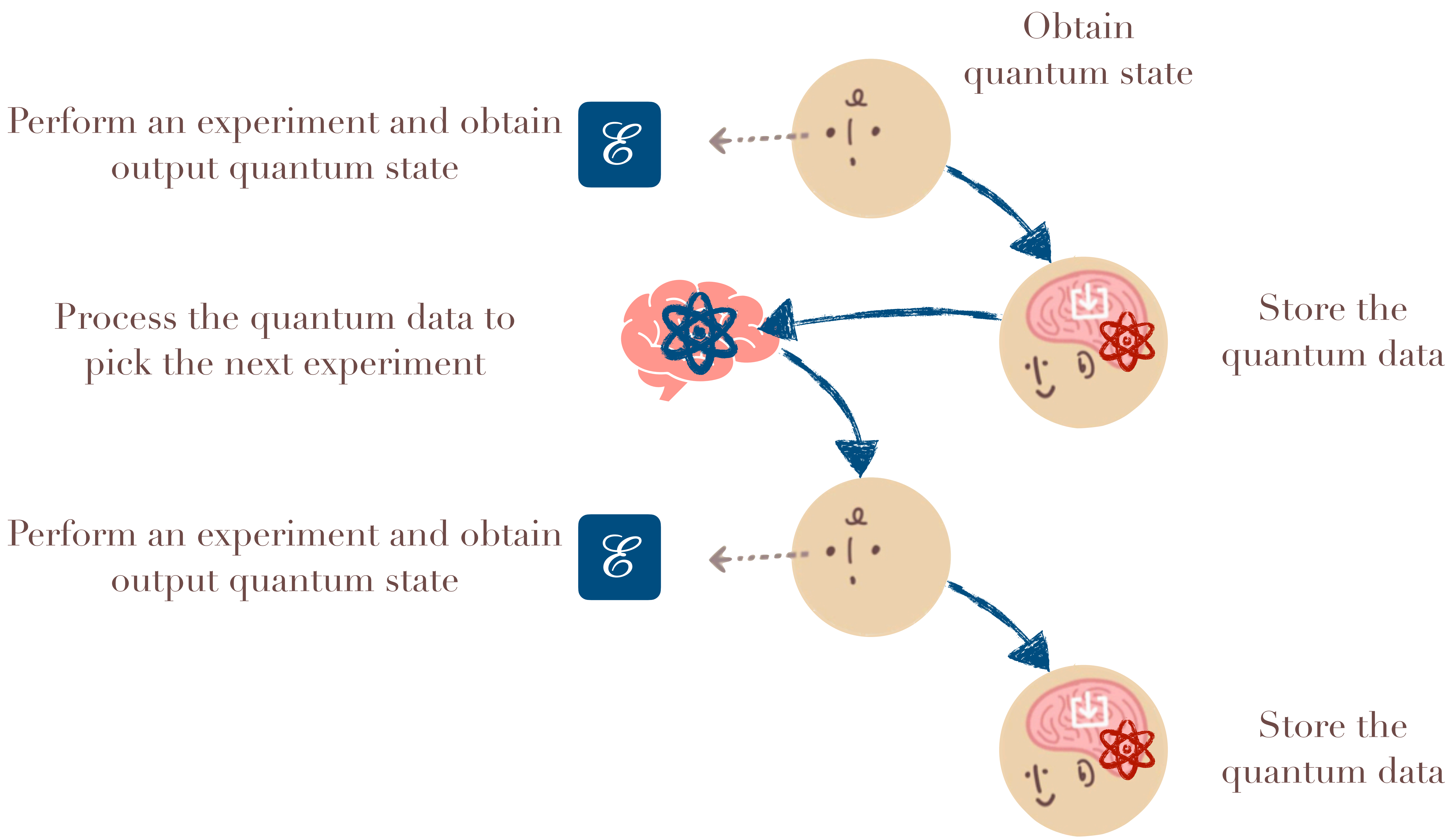
- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.

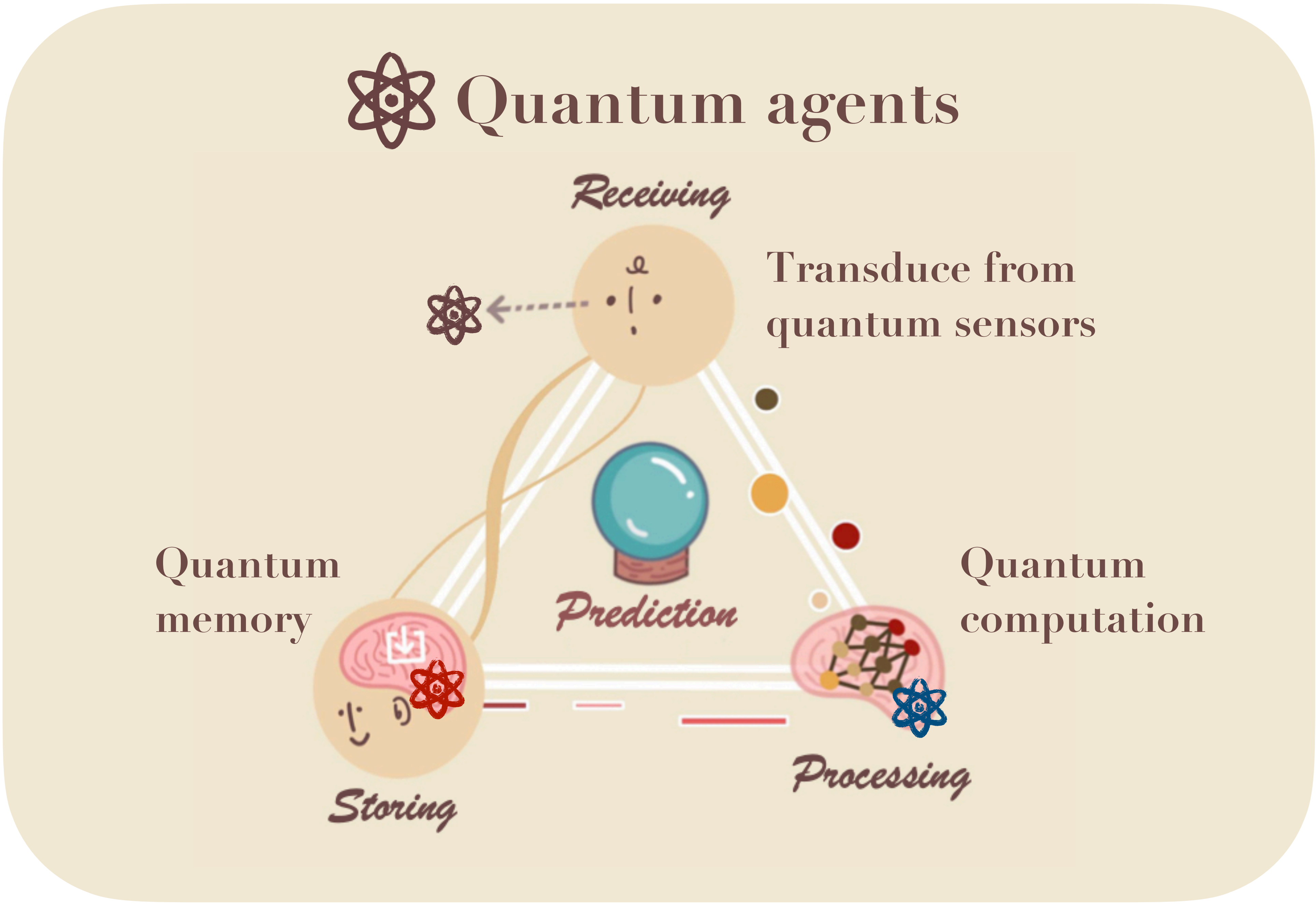




Learning physical dynamics

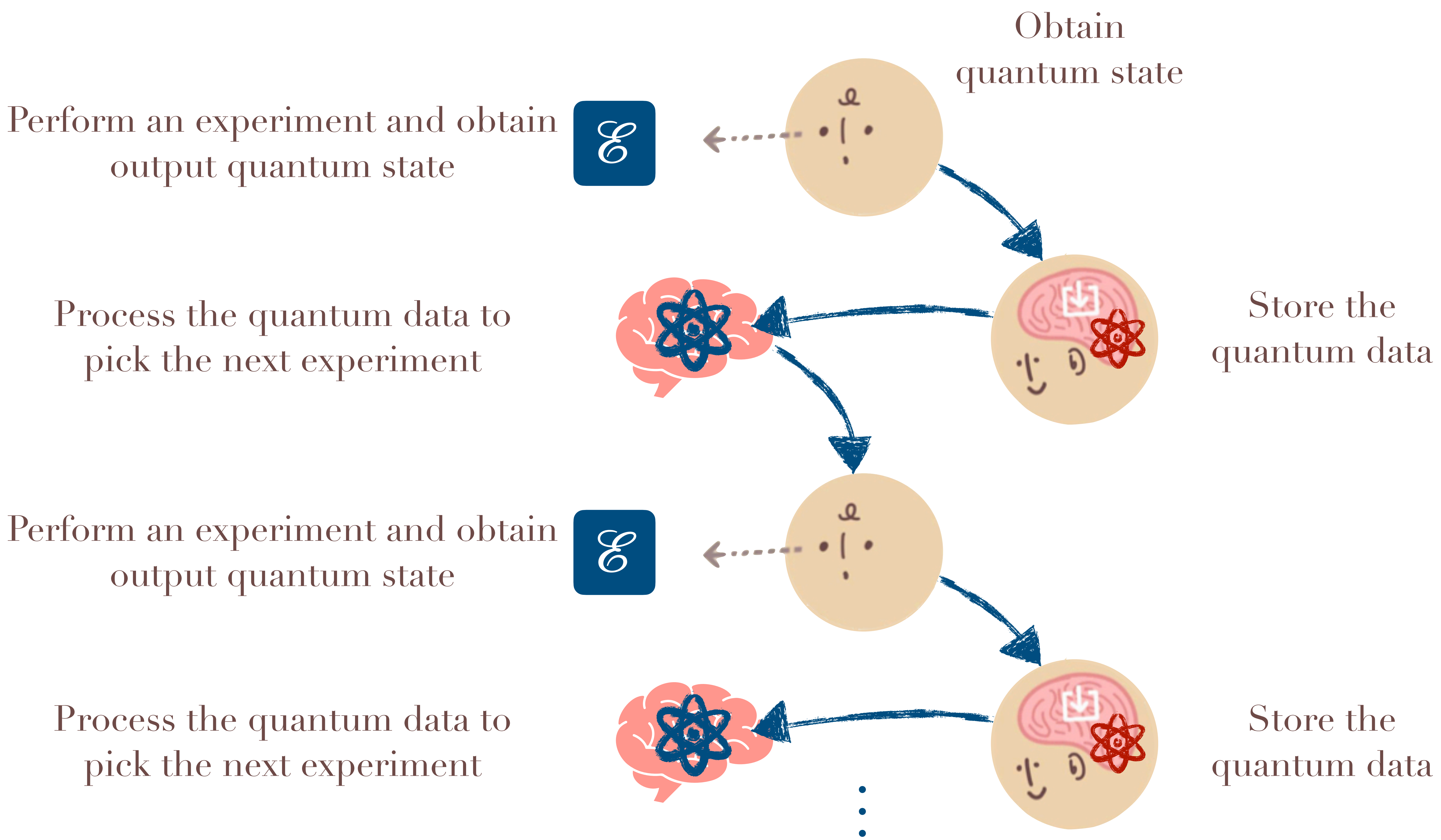
- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.

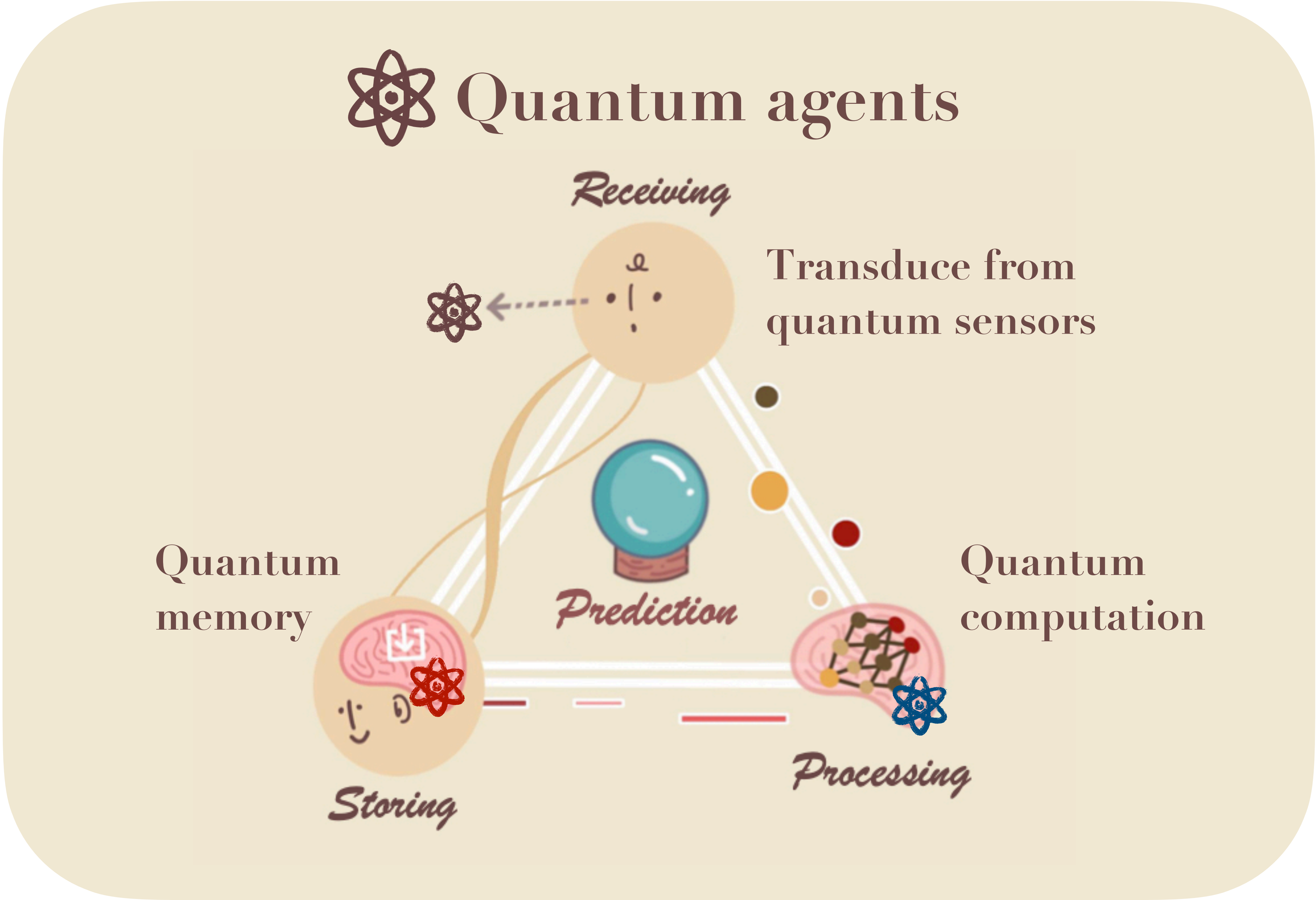




Learning physical dynamics

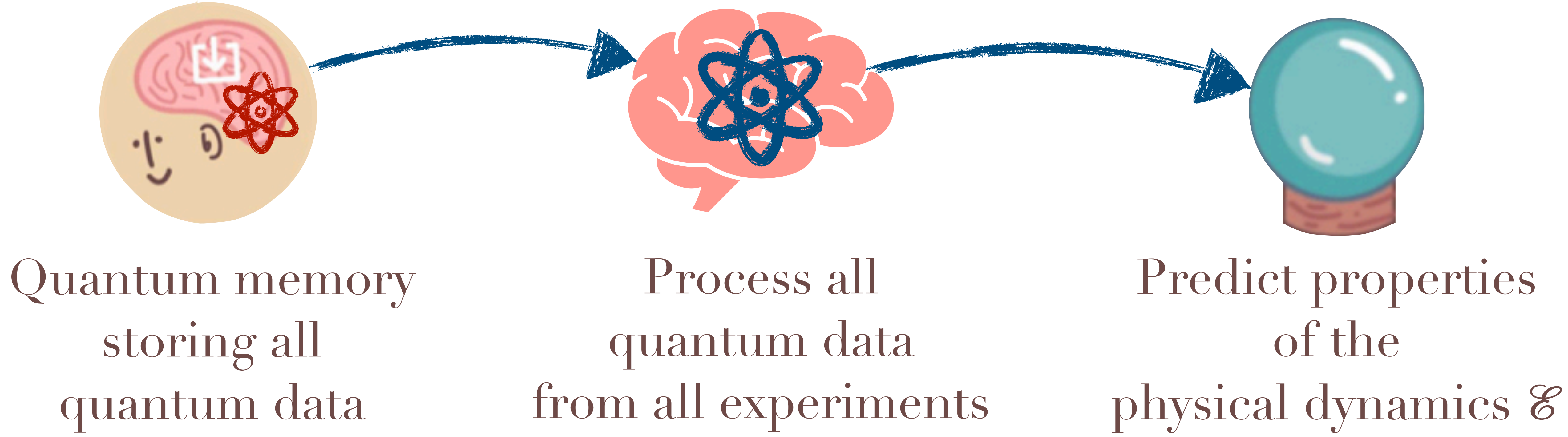
- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.

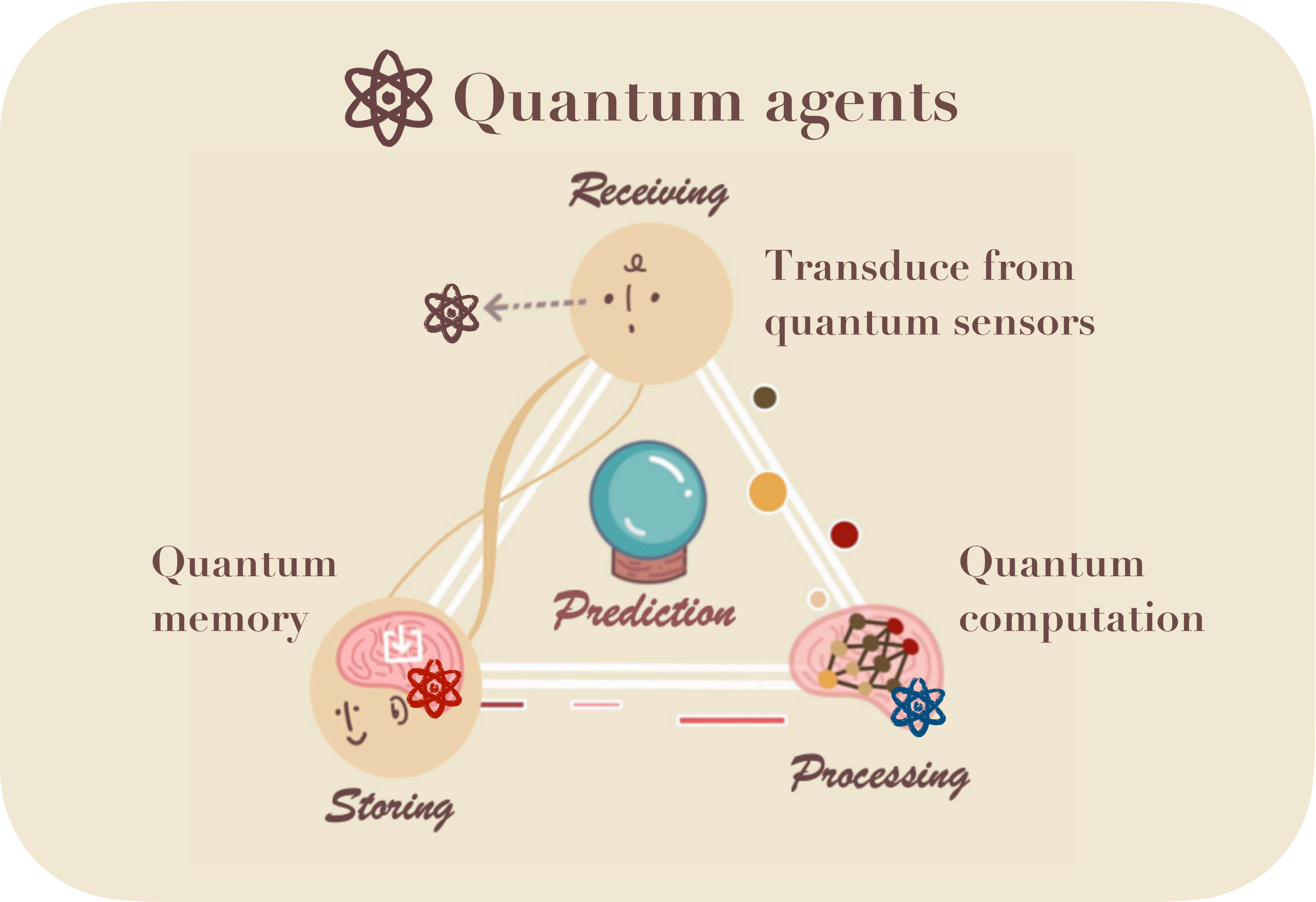




Learning physical dynamics

- We can also consider learning about an unknown physical dynamics $\mathscr{E}$ (quantum process).
- An experiment consists of a state preparation and an evolution under $\mathscr{E}$.





Quantum advantage in learning physical dynamics

- There is an unknown n-qubit process $\mathscr{E}$ that can be generated in $\mathrm{poly}(n)$ time.
- And there is a known distribution $\mathscr{D}$ over n-qubit states.
- Goal: Predict $\mathscr{E}(\rho)$ to a small trace distance for most of $\rho \sim \mathscr{D}$.

Theorem

Classical agent needs $\Omega(2^n)$ experiments to predict $\mathscr{E}(\rho)$ well for $\rho \sim \mathscr{D}$.

Quantum agent only needs $\mathrm{poly}(n)$ experiments to predict $\mathscr{E}(\rho)$ well for $\rho \sim \mathscr{D}$.

Quantum advantage in learning physical dynamics

- There is an unknown n-qubit process $\mathscr{E}$ that can be generated in $\mathrm{poly}(n)$ time.
- And there is a known distribution $\mathscr{D}$ over n-qubit states.
- Goal: Predict $\mathscr{E}(\rho)$ to a small trace distance for most of $\rho \sim \mathscr{D}$.

Theorem

Classical agent needs $\Omega(2^n)$ experiments to predict $\mathscr{E}(\rho)$ well for $\rho \sim \mathscr{D}$.

Quantum agent only needs $\mathrm{poly}(n)$ experiments to predict $\mathscr{E}(\rho)$ well for $\rho \sim \mathscr{D}$.

Quantum advantage in learning physical dynamics

Board Time

- There is an unknown n-qubit process $\mathscr{E}$ that can be generated in $\mathrm{poly}(n)$ time.
- And there is a known distribution $\mathscr{D}$ over n-qubit states.
- Goal: Predict $\mathscr{E}(\rho)$ to a small trace distance for most of $\rho \sim \mathscr{D}$.

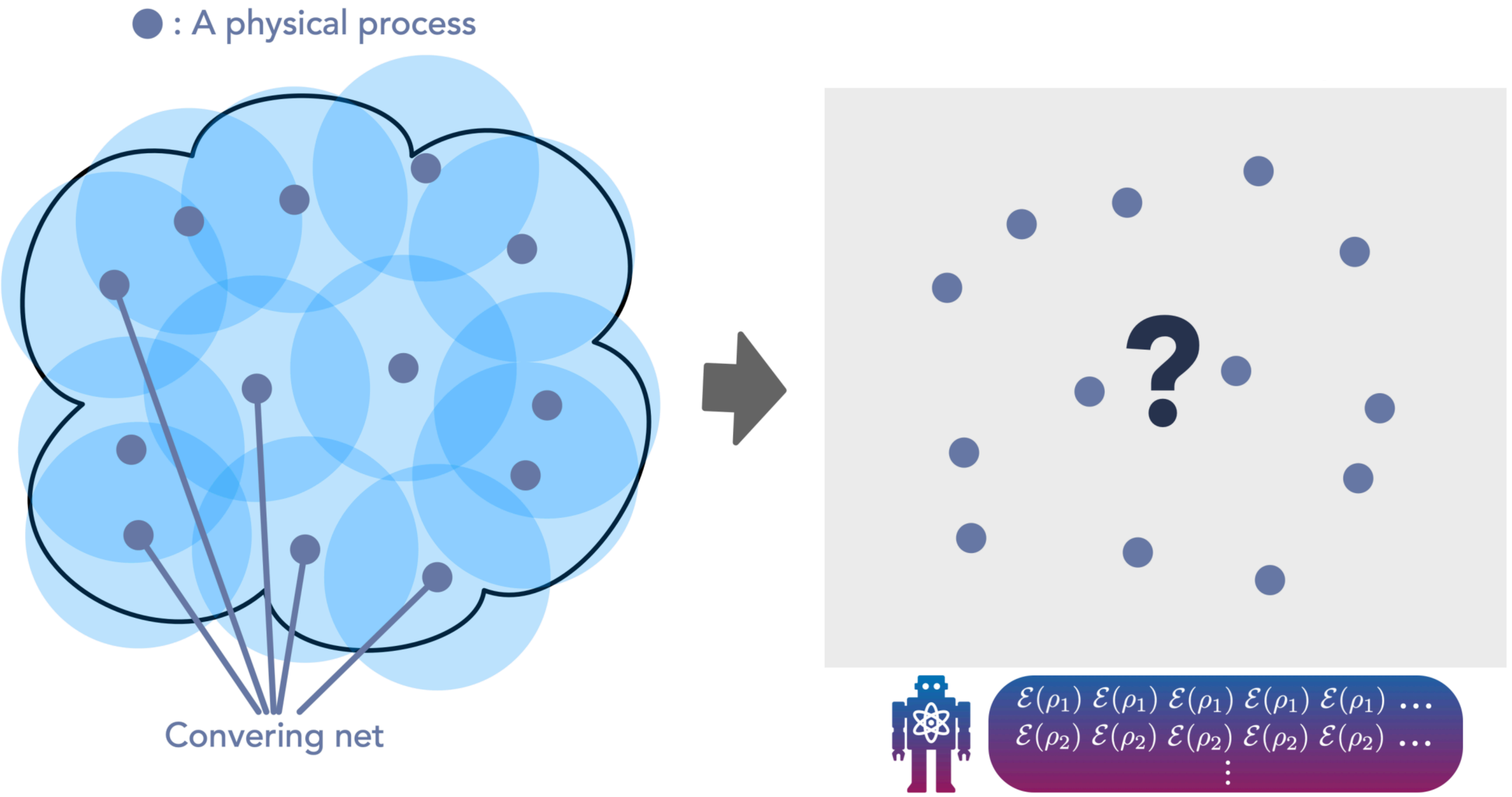
Theorem

Classical agent needs $\Omega(2^n)$ experiments to predict $\mathscr{E}(\rho)$ well for $\rho \sim \mathscr{D}$.

Quantum agent only needs $\mathrm{poly}(n)$ experiments to predict $\mathscr{E}(\rho)$ well for $\rho \sim \mathscr{D}$.

Covering Net

Board Time



Quantum advantage in learning physical dynamics

- There is an unknown n-qubit process $\mathscr{E}$ that can be generated in $\mathrm{poly}(n)$ time.
- And there is a known distribution $\mathscr{D}$ over n-qubit states.
- Goal: Predict $\mathscr{E}(\rho)$ to a small trace distance for most of $\rho \sim \mathscr{D}$.

Theorem

Classical agent needs $\Omega(2^n)$ experiments to predict $\mathscr{E}(\rho)$ well for $\rho \sim \mathscr{D}$.

Quantum agent only needs $\mathrm{poly}(n)$ experiments to predict $\mathscr{E}(\rho)$ well for $\rho \sim \mathscr{D}$.

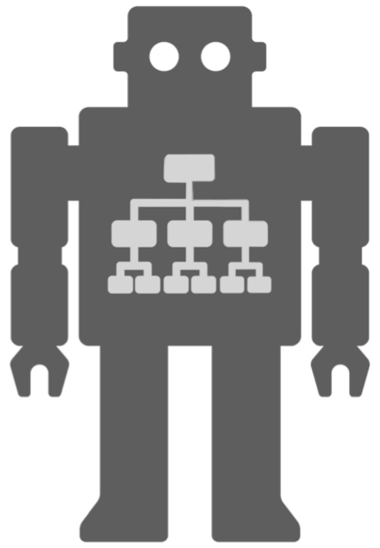
Quantum advantage in learning physical dynamics

- There is an unknown n-qubit process $\mathscr{E}$ that can be generated in $\mathrm{poly}(n)$ time.
- And there is a known distribution $\mathscr{D}$ over n-qubit states.
- Goal: Predict $\mathscr{E}(\rho)$ to a small trace distance for most of $\rho \sim \mathscr{D}$.

Theorem

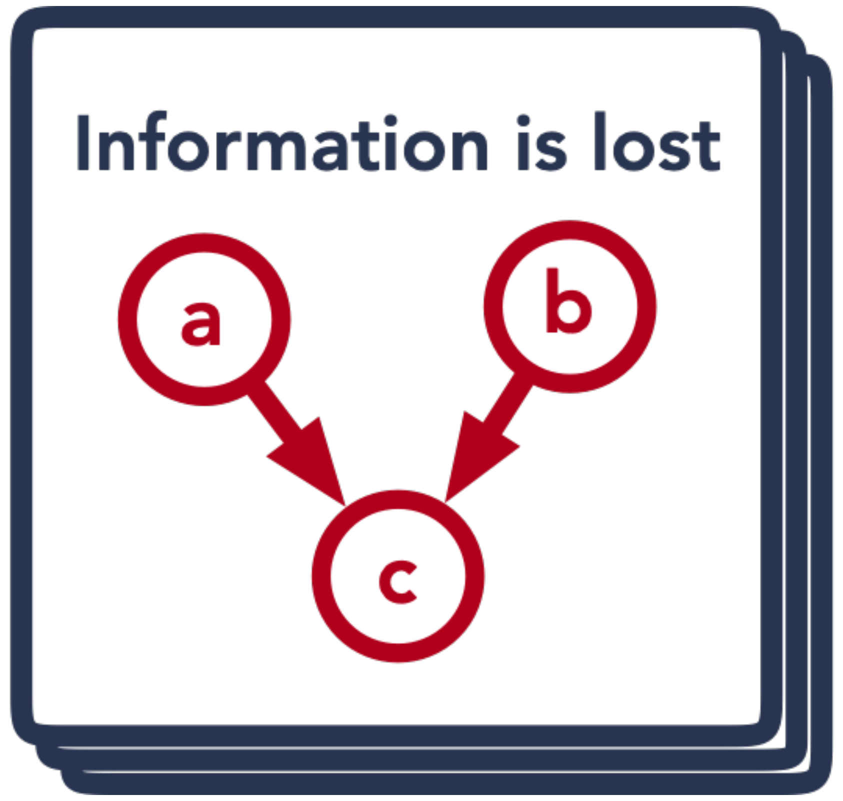
Classical agent needs $\Omega(2^n)$ experiments to predict $\mathscr{E}(\rho)$ well for $\rho \sim \mathscr{D}$.

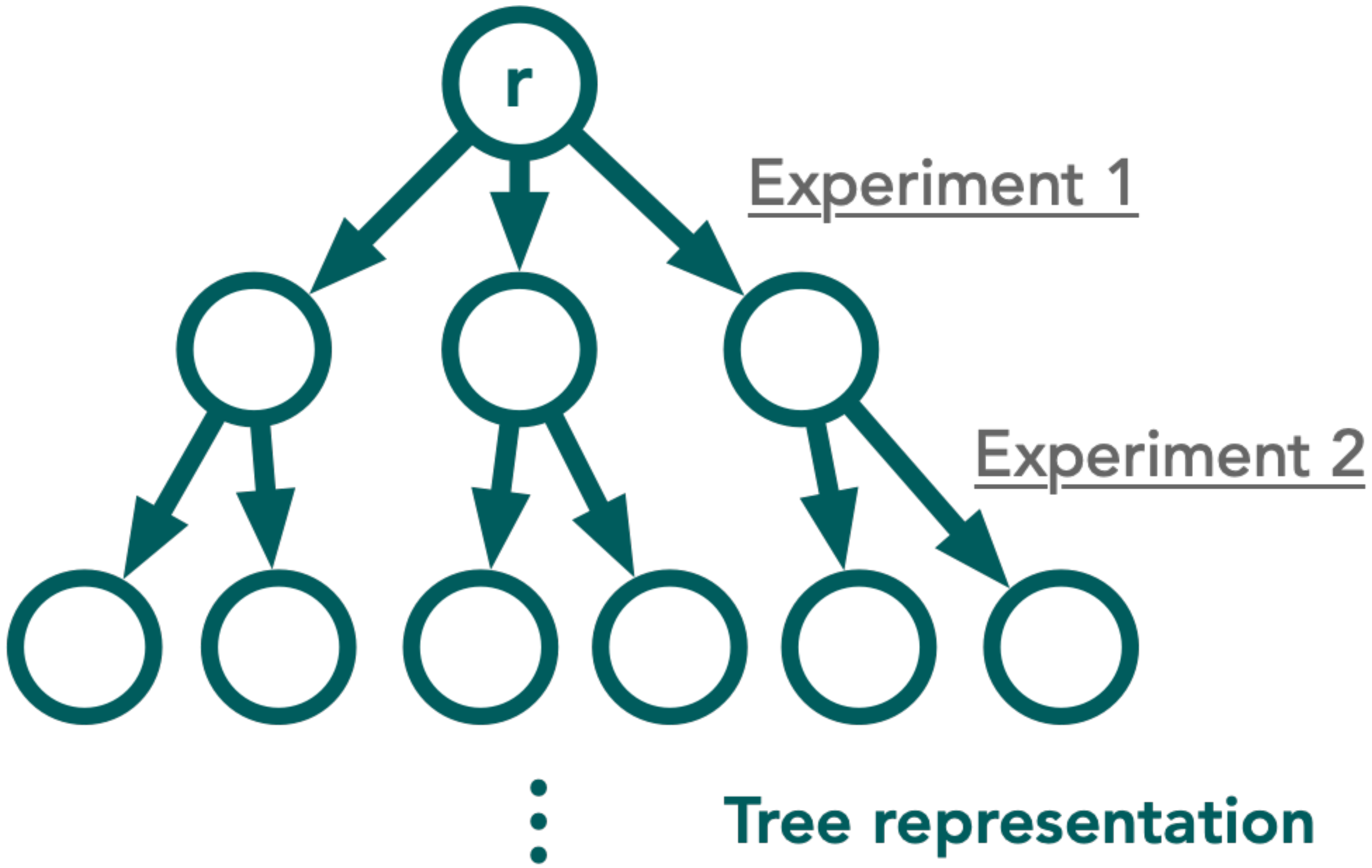
Quantum agent only needs $\mathrm{poly}(n)$ experiments to predict $\mathscr{E}(\rho)$ well for $\rho \sim \mathscr{D}$.

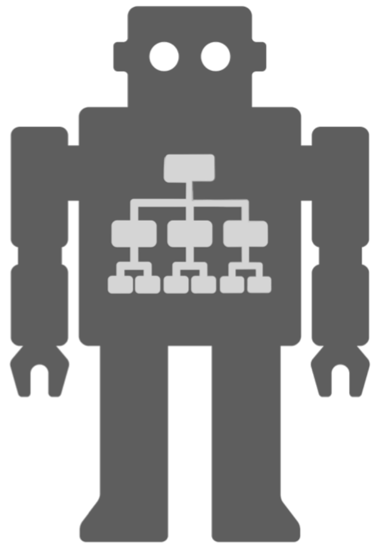


Tree Representation

- We consider a graphical representation of the memory state of the classical agent when learning a quantum process $\mathscr{E}$.

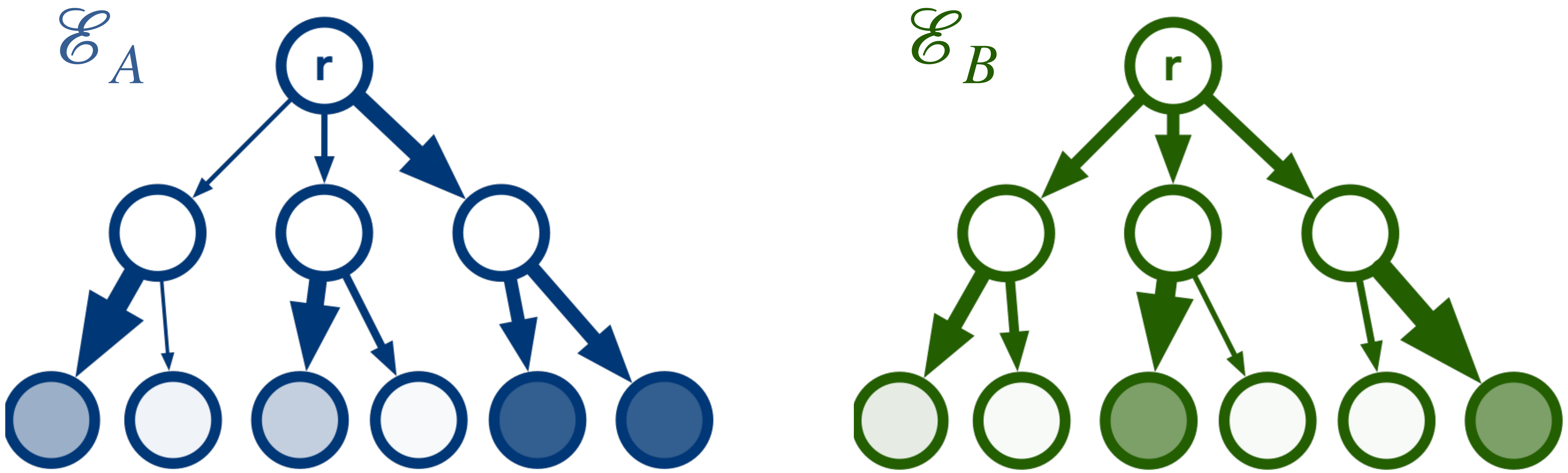


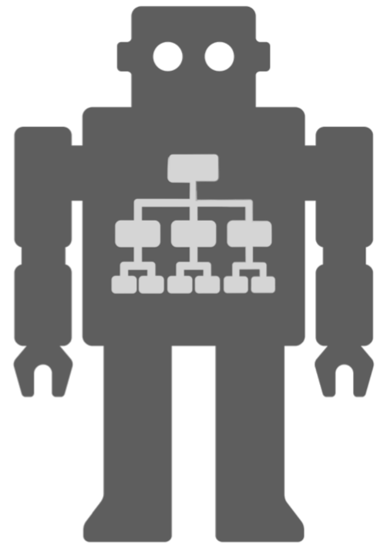




Tree Representation

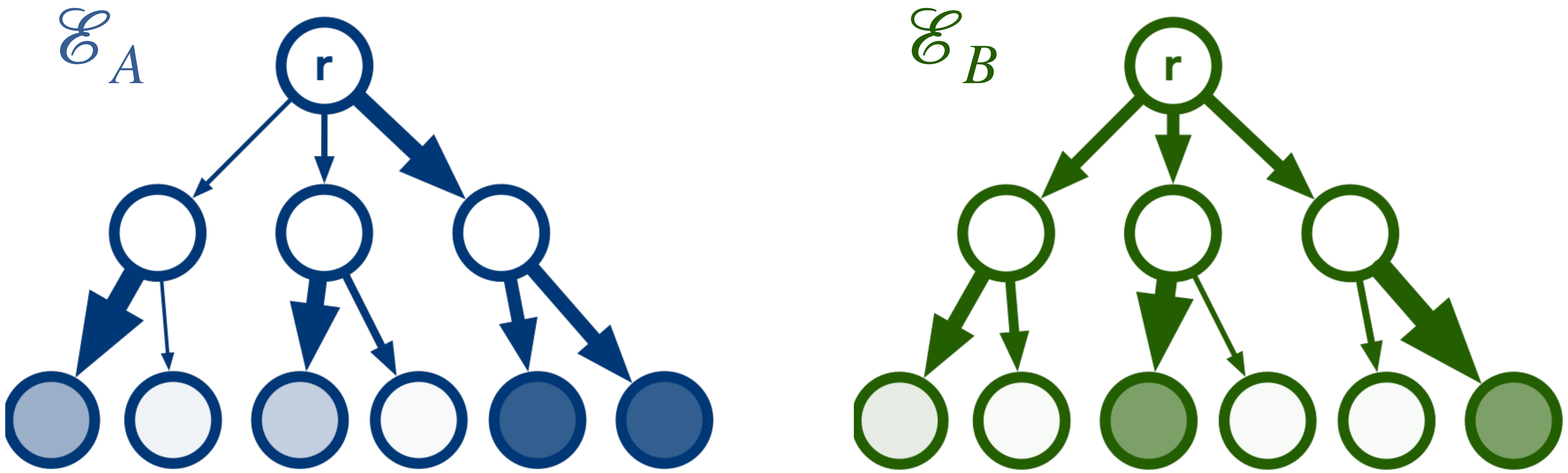
- We consider a graphical representation of the memory state of the classical agent when learning a quantum process $\mathscr{E}$.



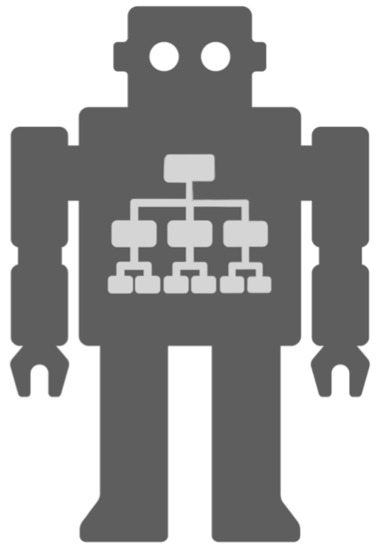


Tree Representation

- We consider a graphical representation of the memory state of the classical agent when learning a quantum process $\mathscr{E}$.

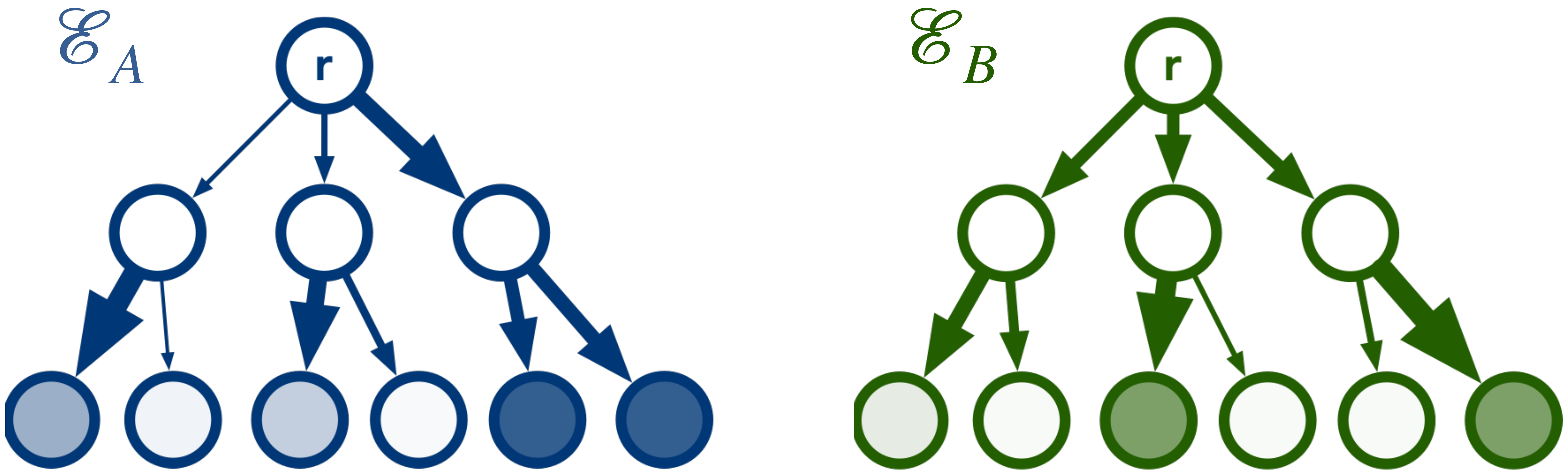


Probability distribution (bottom layer) sufficiently different $\equiv$ Classical agent can distinguish $\mathscr{E}_A$ and $\mathscr{E}_B$



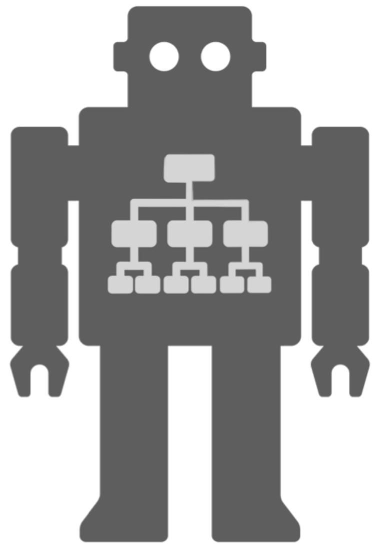
Tree Representation

- We consider a graphical representation of the memory state of the classical agent when learning a quantum process $\mathscr{E}$.



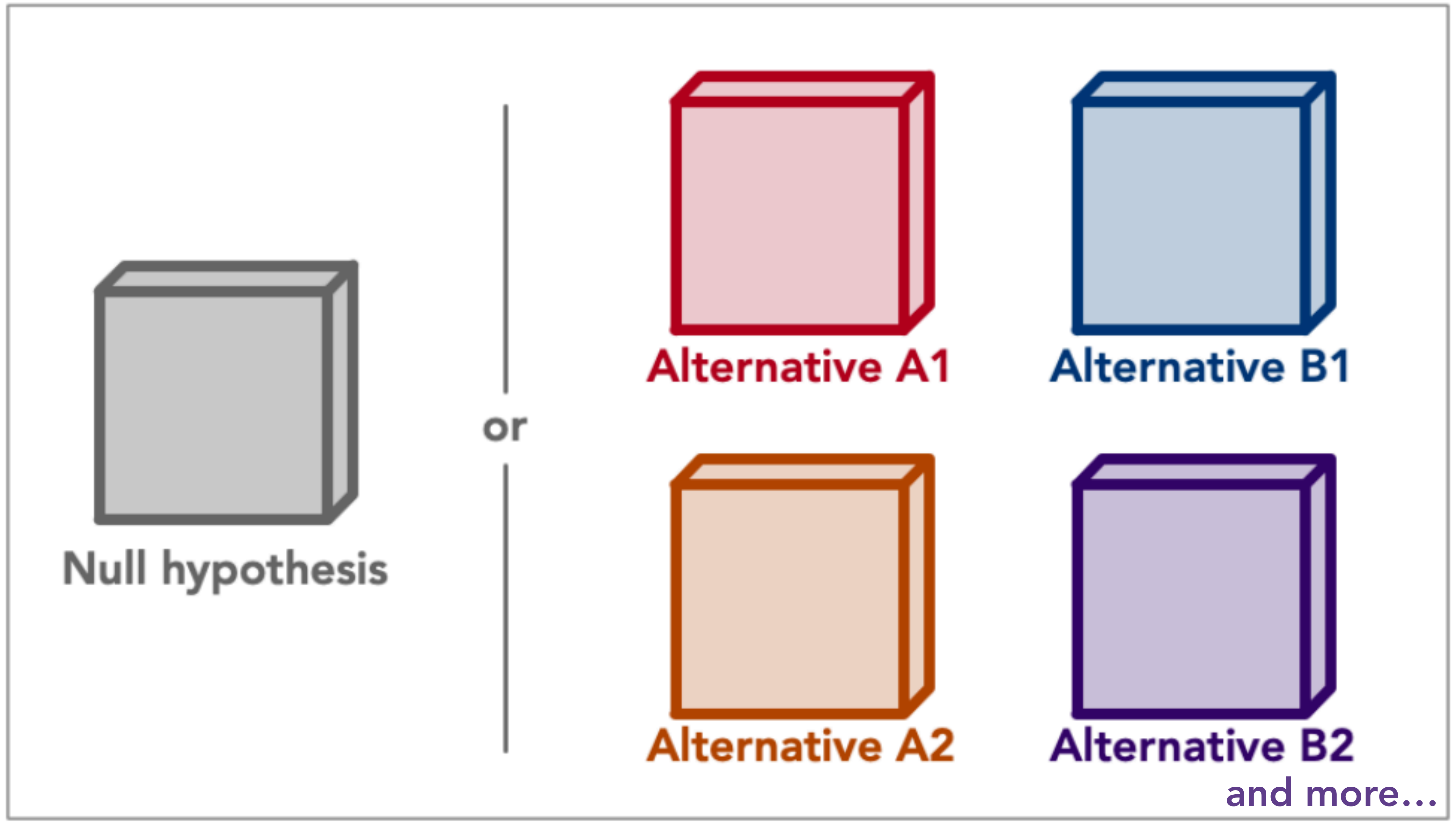
Probability distribution (bottom layer) sufficiently different $\equiv$ Classical agent can distinguish $\mathscr{E}_A$ and $\mathscr{E}_B$

More experiments done $\equiv$ Deeper the tree $\equiv$ More distinct the distribution

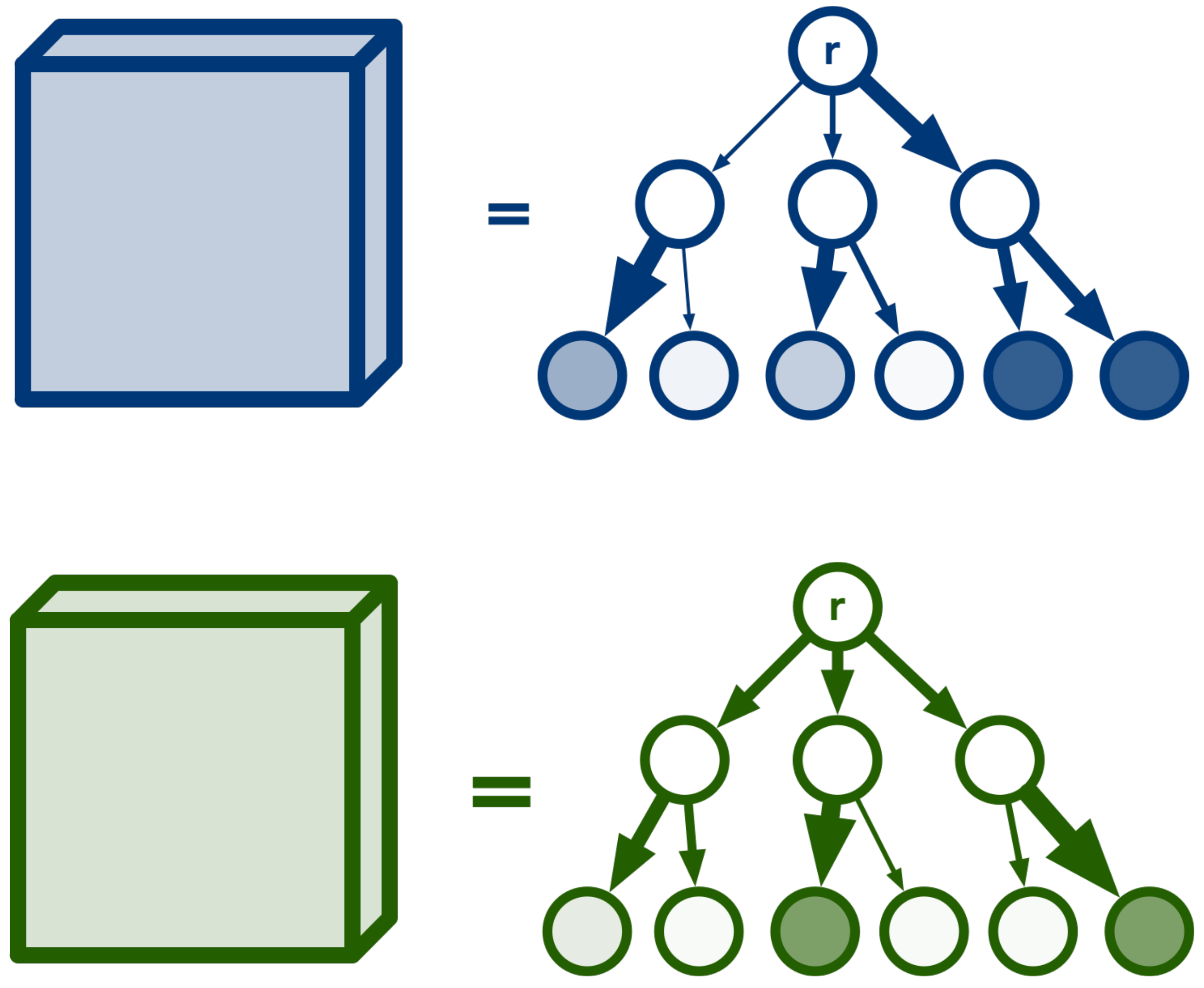


Reduction

- Consider $\mathcal{E}(\rho) = I/2^n$ (Null) vs. $\mathcal{E}(\rho) = (I \pm P)/2^n$ (Alternative).
- $P \in \{I, X, Y, Z\}^{\otimes n} \backslash \{I^{\otimes n}\}$.

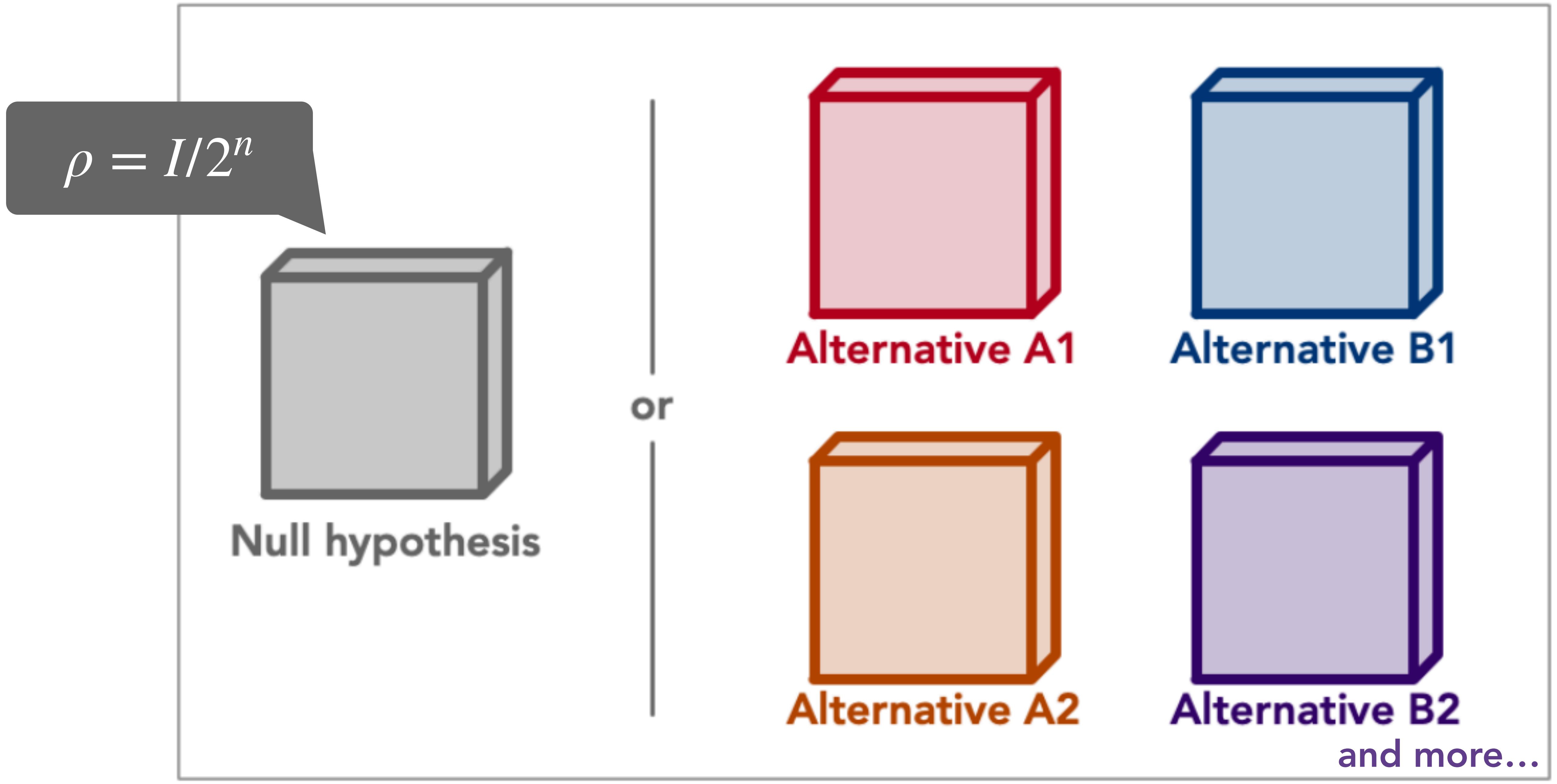


Many-versus-one distinguishing task

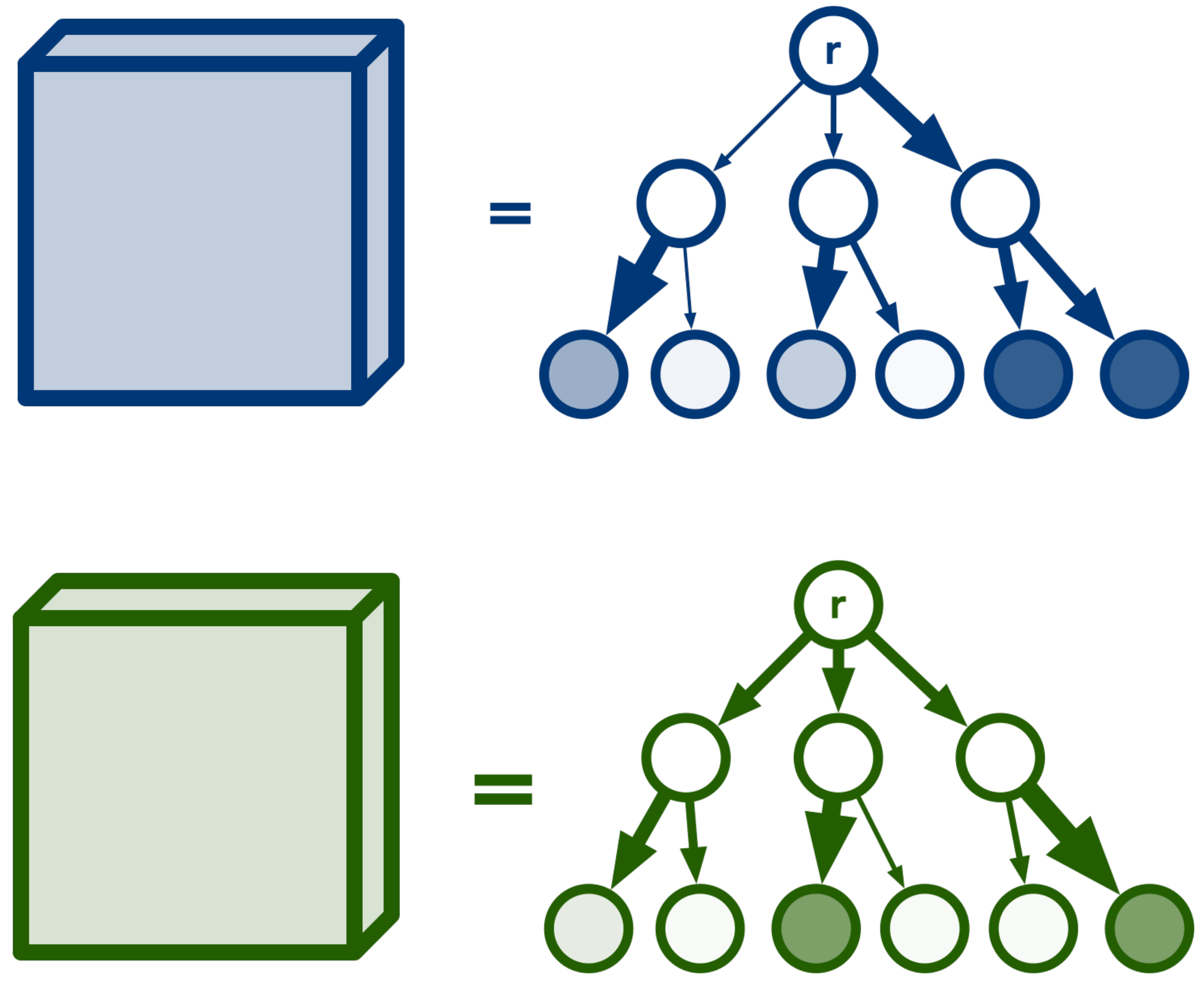


Reduction

- Consider $\mathscr{E}(\rho) = I/2^n$ (Null) vs. $\mathscr{E}(\rho) = (I \pm P)/2^n$ (Alternative).
- $P \in \{I, X, Y, Z\}^{\otimes n} \backslash \{I^{\otimes n}\}$.

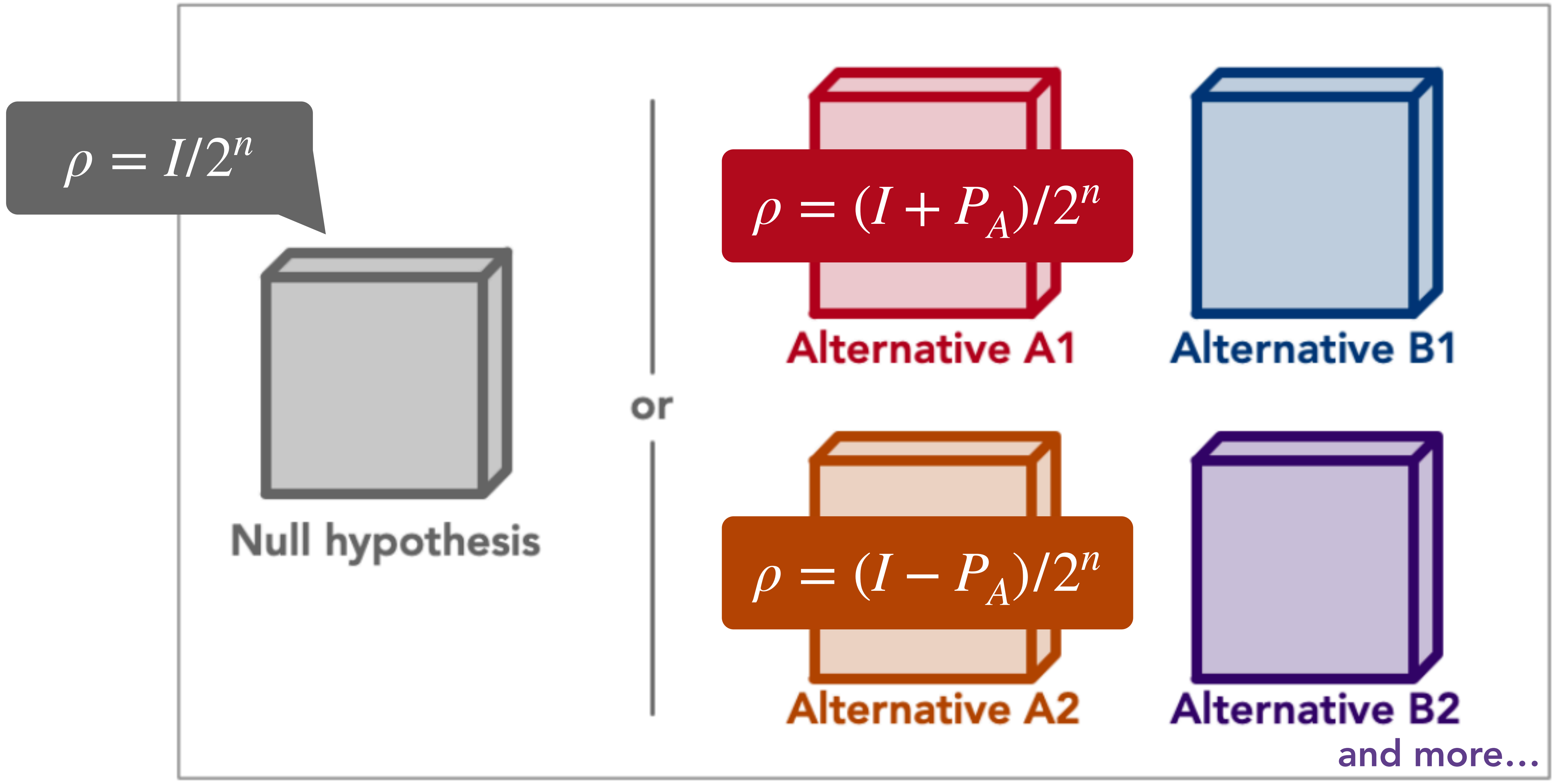


Many-versus-one distinguishing task

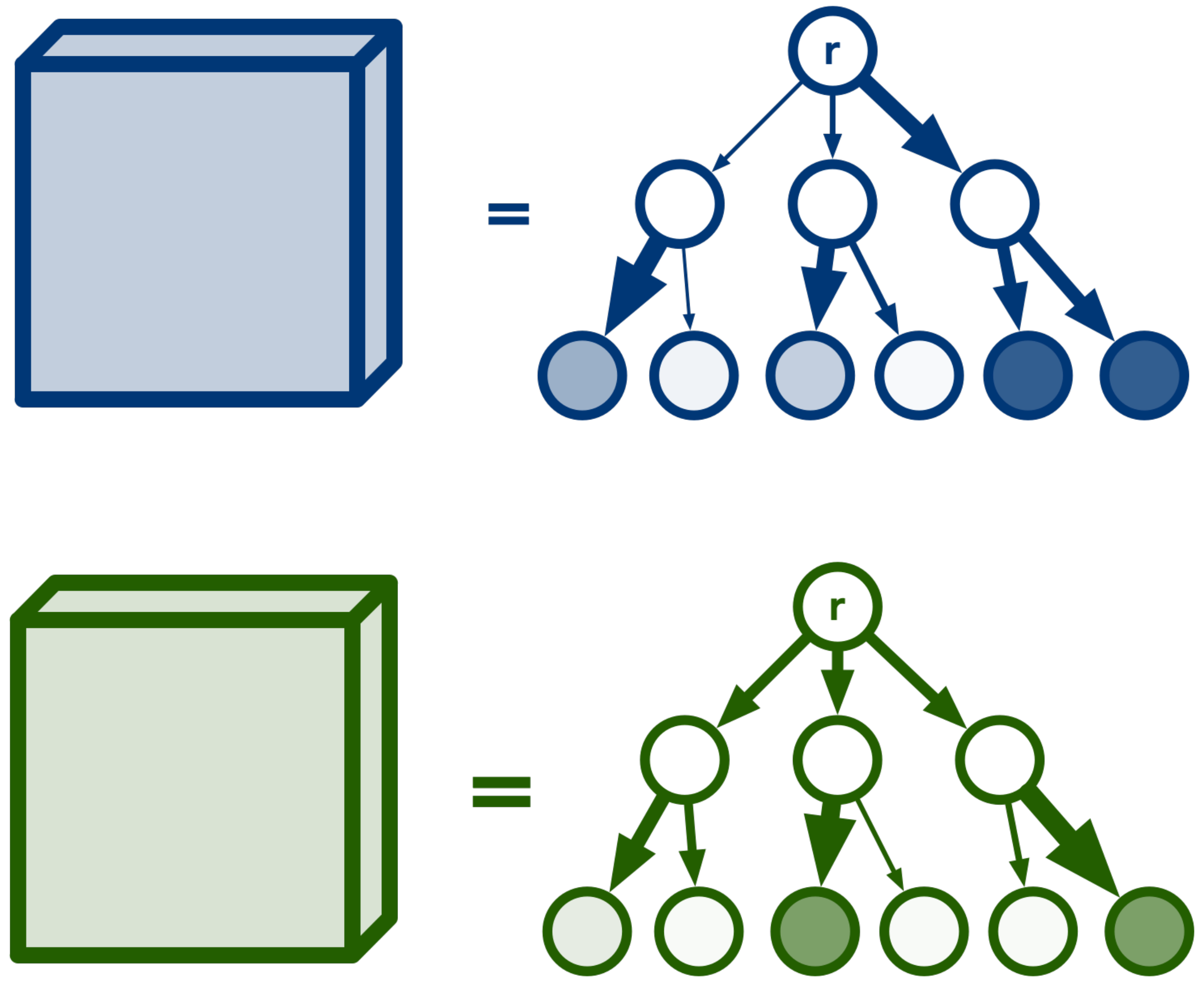


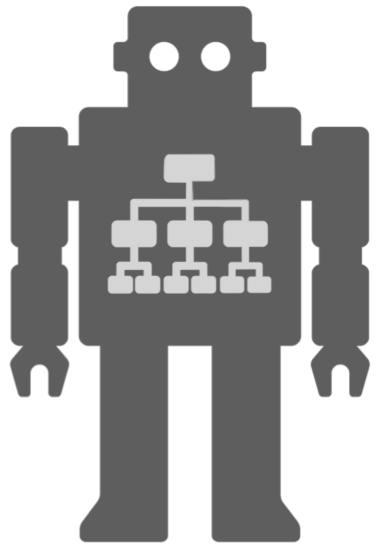
Reduction

- Consider $\mathscr{E}(\rho) = I/2^n$ (Null) vs. $\mathscr{E}(\rho) = (I \pm P)/2^n$ (Alternative).
- $P \in \{I, X, Y, Z\}^{\otimes n} \backslash \{I^{\otimes n}\}$.



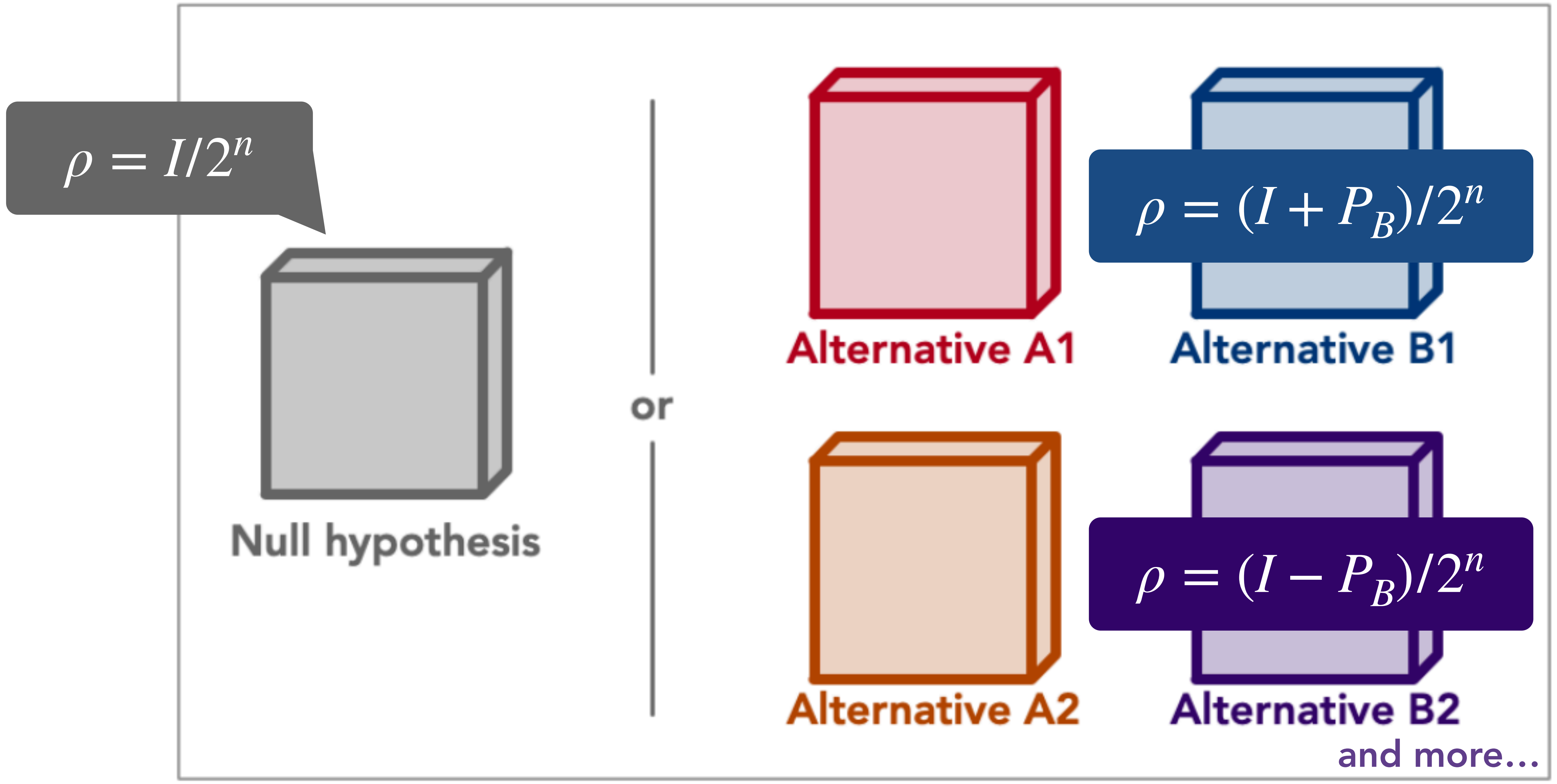
Many-versus-one distinguishing task



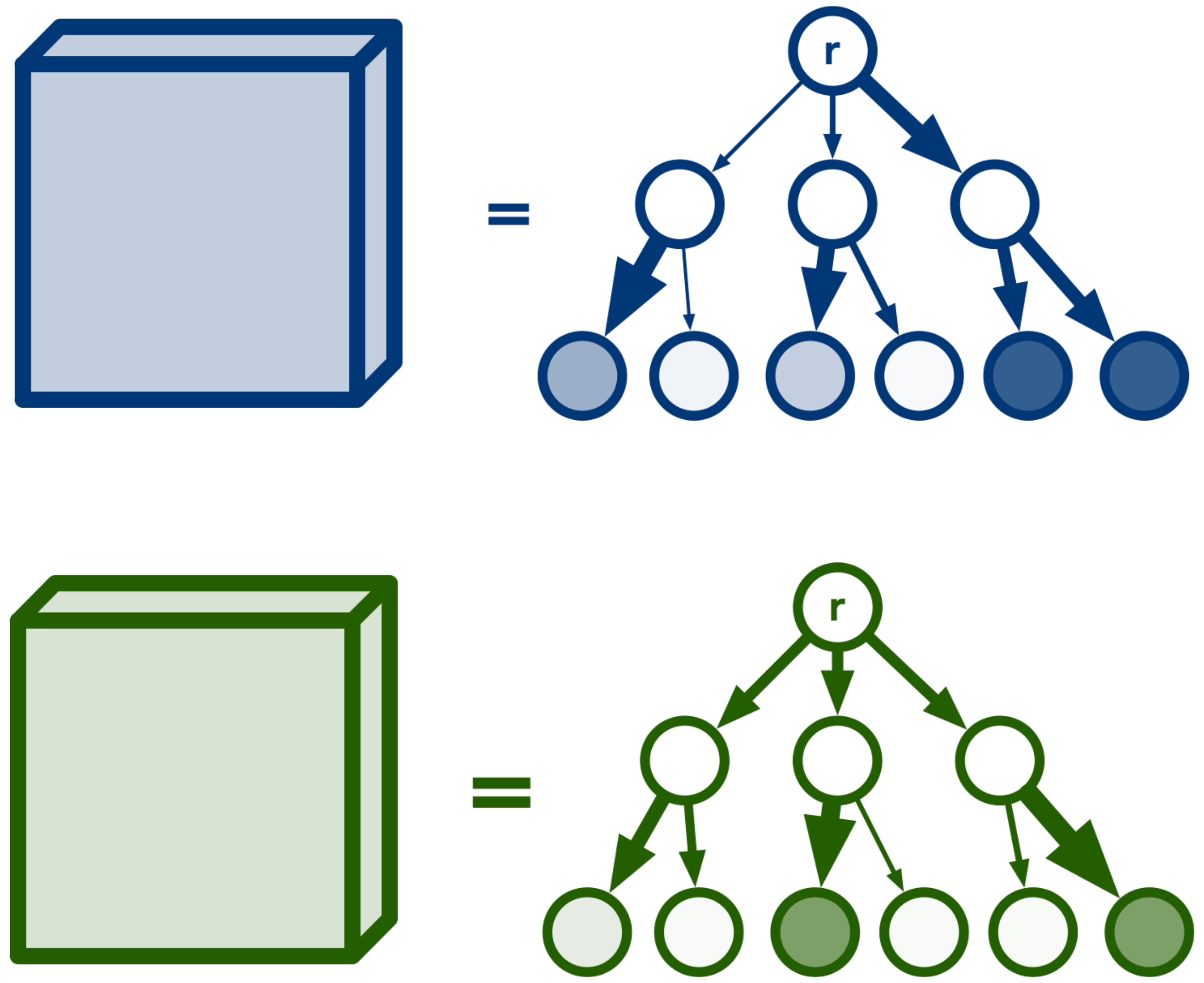


Reduction

- Consider $\mathscr{E}(\rho) = I/2^n$ (Null) vs. $\mathscr{E}(\rho) = (I \pm P)/2^n$ (Alternative).
- $P \in \{I, X, Y, Z\}^{\otimes n} \backslash \{I^{\otimes n}\}$.

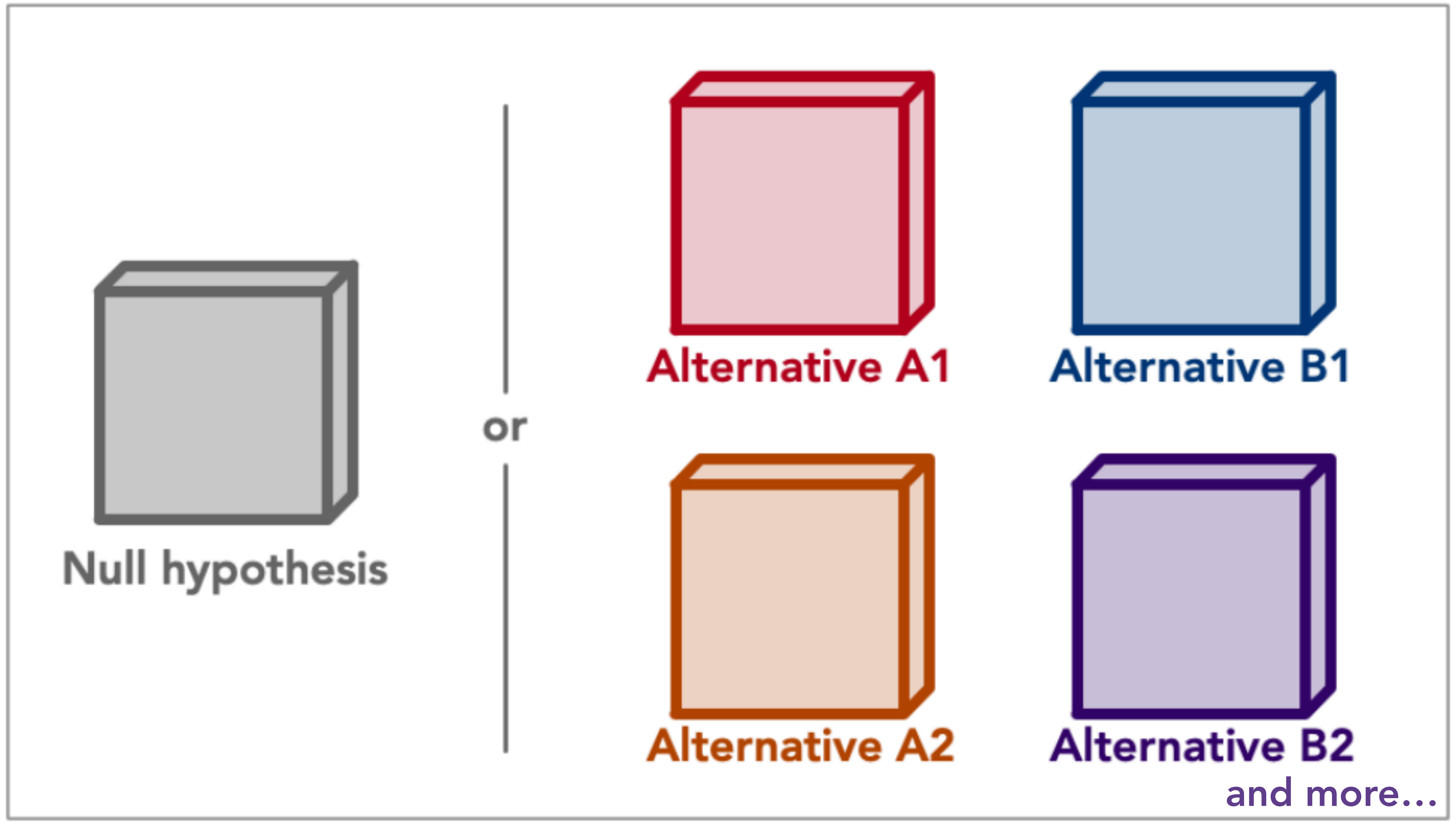


Many-versus-one distinguishing task

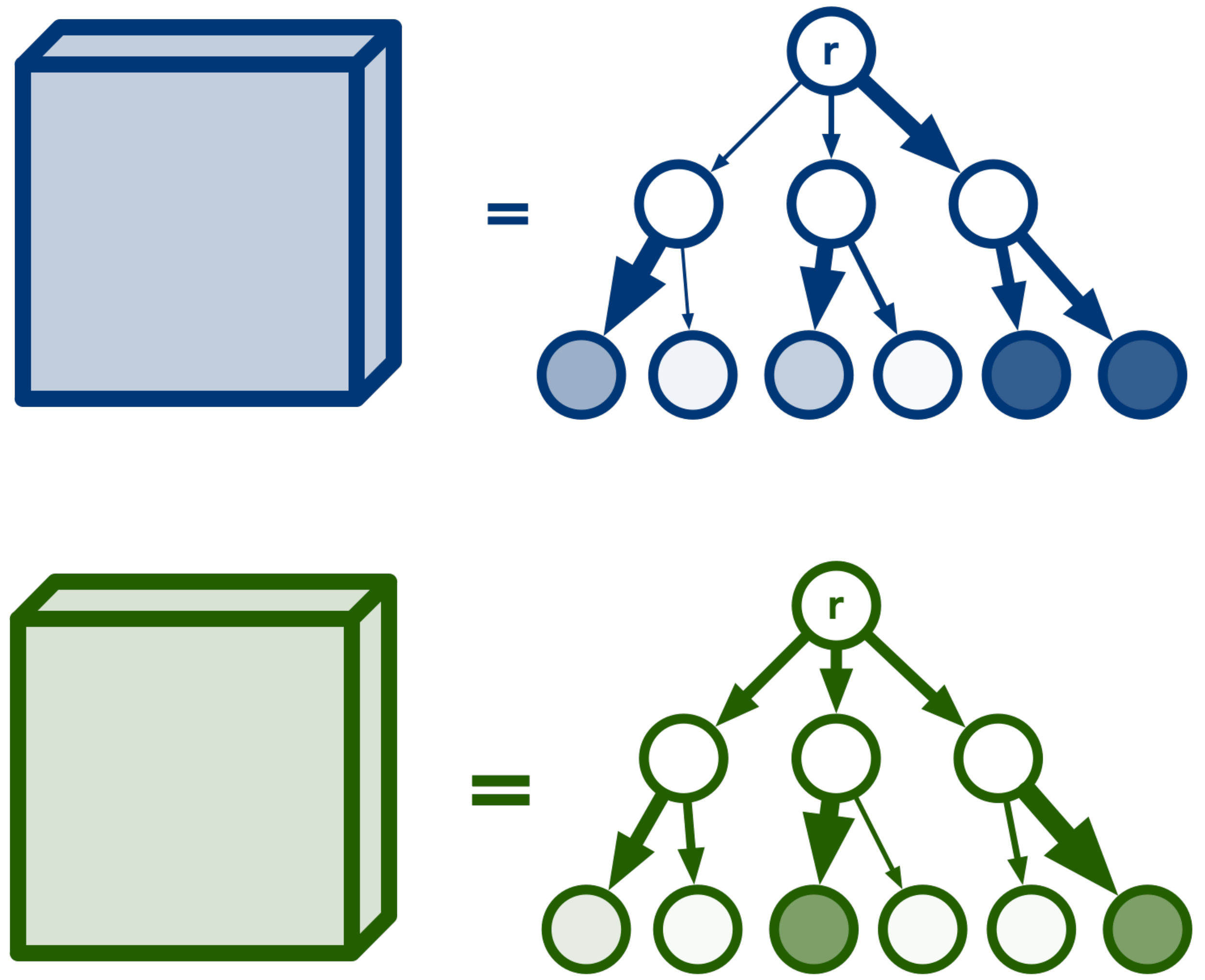


Information-theoretic

- Controlling the total variation (TV) distance between the leaf distribution in the null hypothesis and the alternative hypothesis.

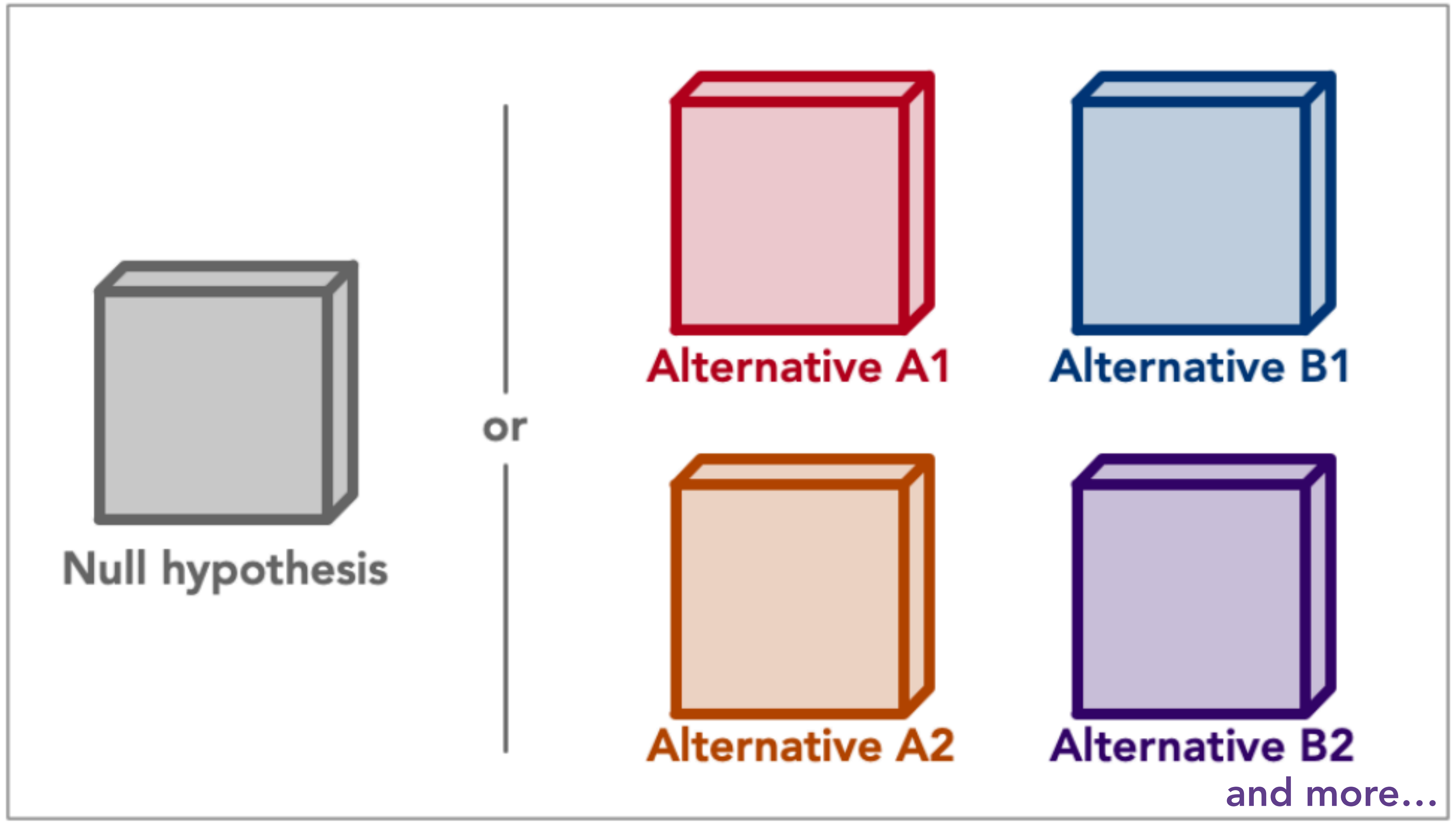


Many-versus-one distinguishing task

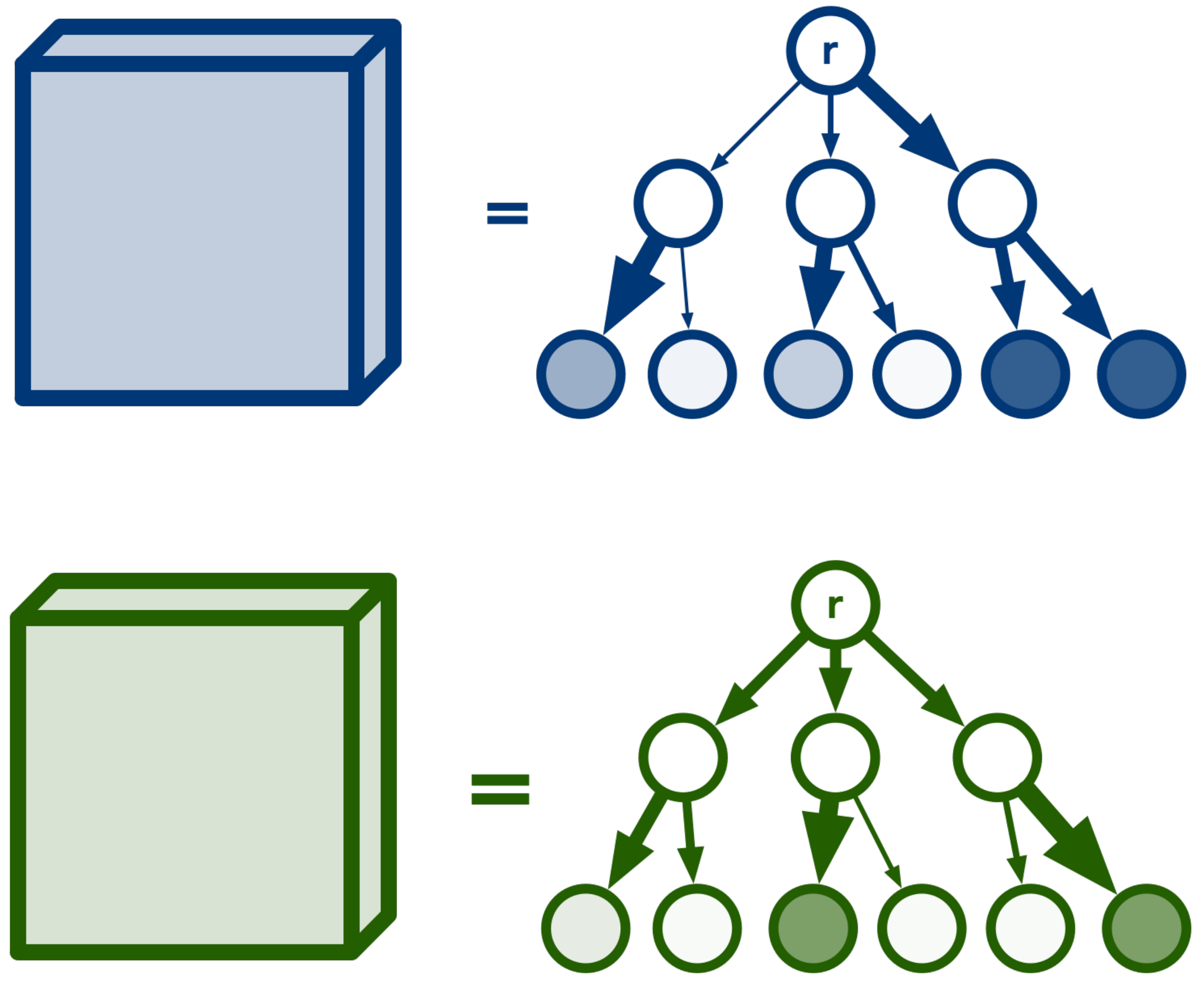


Information-theoretic

- Controlling the total variation (TV) distance between the leaf distribution in the null hypothesis and the alternative hypothesis.

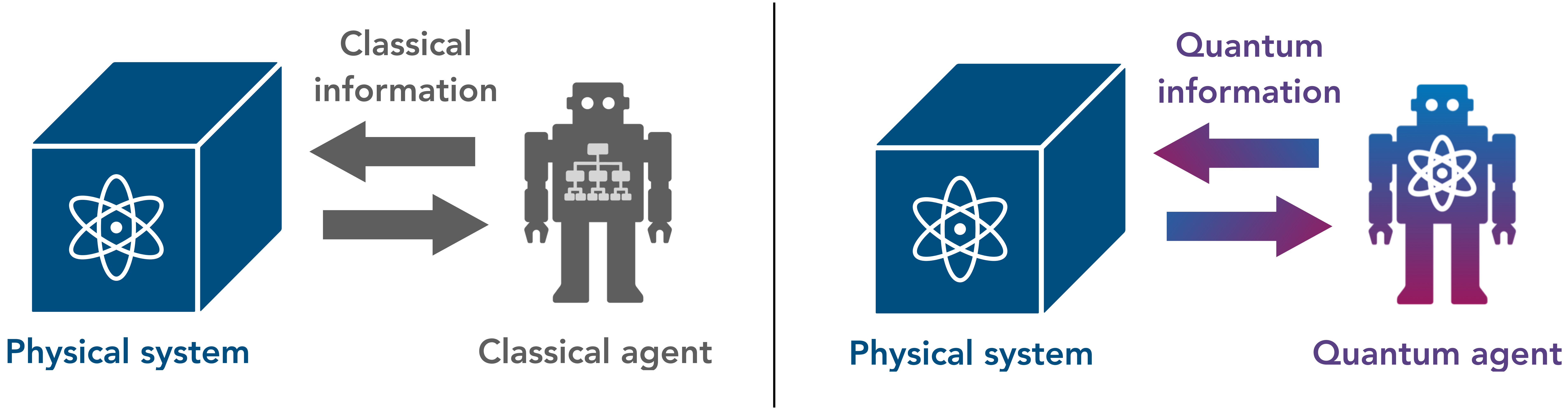


Many-versus-one distinguishing task



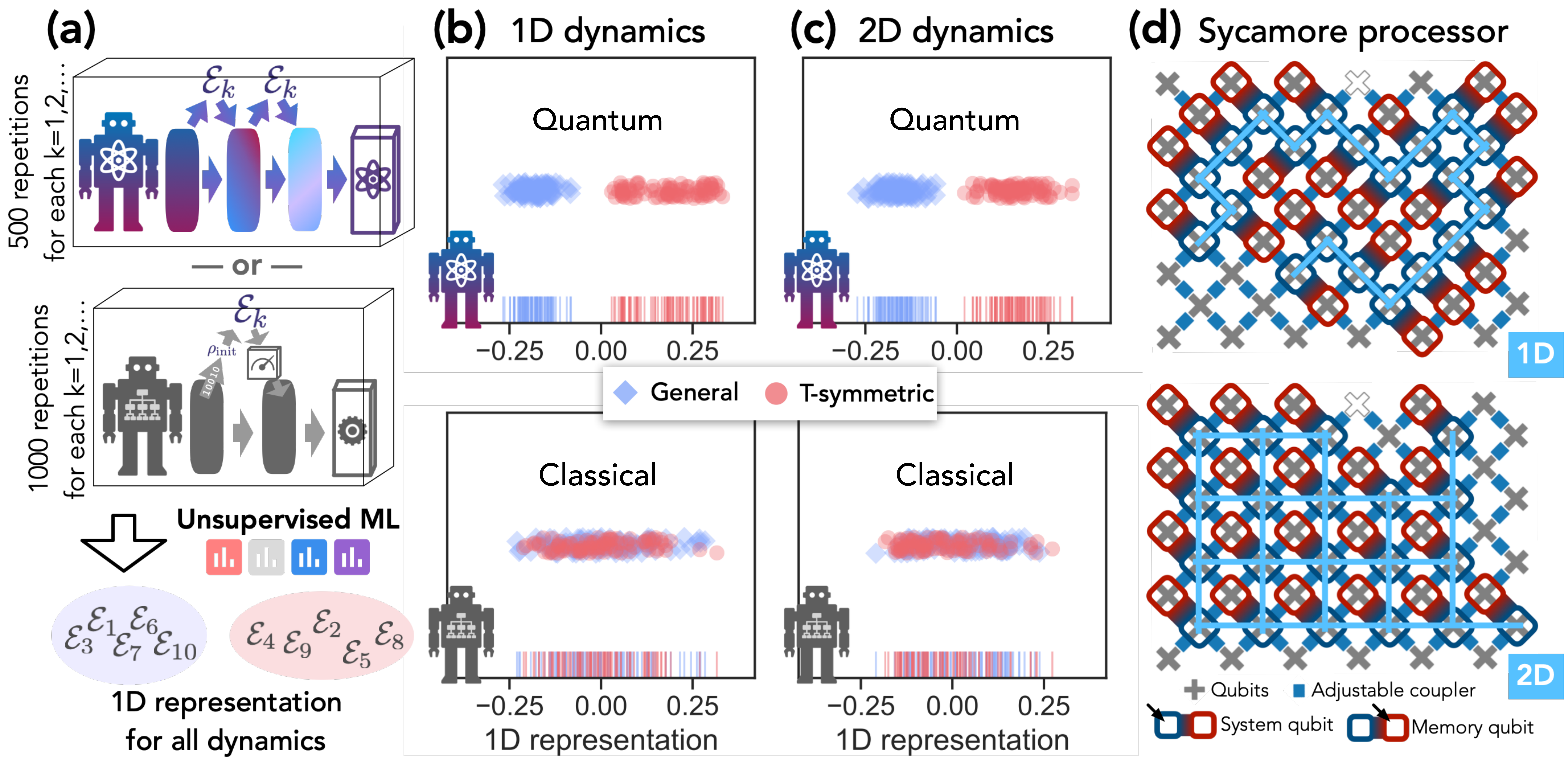
Quantum advantage in NISQ

- Do these quantum advantages persist in noisy quantum computers? Yes! Rigorous analysis in [HFP22], Experiments in [HBC+22].



[HFP22] Huang, Flammia, Preskill. Foundations for learning from noisy quantum experiments, *QIP*, 2022.
[HBC+22] Huang, Broughton, Cotler, Chen, Li, Mohseni, Neven, Babbush, Kueng, Preskill, McClean. Quantum advantage in learning from experiments, *Science*, 2022.

Demonstration on Sycamore: Quantum advantage in learning dynamics



[HFP22] Huang, Flammia, Preskill. Foundations for learning from noisy quantum experiments, *QIP*, 2022.
[HBC+22] Huang, Broughton, Cotler, Chen, Li, Mohseni, Neven, Babbush, Kueng, Preskill, McClean. Quantum advantage in learning from experiments, *Science*, 2022.

Relevant works

Exponential separation btw. learning w/ and w/o quantum memory

[This tutorial + more techniques]

Relevant works

Exponential separation btw. learning w/ and w/o quantum memory

[This tutorial + more techniques]

Quantum advantage in learning from experiments

[This tutorial + more experiments]

Relevant works

Exponential separation btw. learning w/ and w/o quantum memory

[This tutorial + more techniques]

Advantage of quantum control in many-body Hamiltonian learning

[Quadratic advantage in learning Hamiltonians]

Quantum advantage in learning from experiments

[This tutorial + more experiments]

Relevant works

Exponential separation btw. learning w/ and w/o quantum memory

[This tutorial + more techniques]

Advantage of quantum control in many-body Hamiltonian learning

[Quadratic advantage in learning Hamiltonians]

Quantum advantage in learning from experiments

[This tutorial + more experiments]

Learning quantum processes and Hamiltonians via Pauli transfer matrix

[Exponential advantage in learning entries of PTM]

Take home message

- Quantum advantage is the ultimate goal of quantum technology.
  (otherwise, we should just use the existing classical technology)

Take home message

- Quantum advantage is the ultimate goal of quantum technology.
  (otherwise, we should just use the existing classical technology)

- The advantage can be diverse: computation, information, memory, energy, ….

Take home message

- Quantum advantage is the ultimate goal of quantum technology.
  (otherwise, we should just use the existing classical technology)

- The advantage can be diverse: computation, information, memory, energy, ….
  [Previous focus]

Take home message

- Quantum advantage is the ultimate goal of quantum technology.
  (otherwise, we should just use the existing classical technology)

- The advantage can be diverse: computation, information, memory, energy, ...

  [——— This tutorial ———]

Take home message

- Quantum advantage is the ultimate goal of quantum technology.
  (otherwise, we should just use the existing classical technology)

- The advantage can be diverse: computation, information, memory, energy, ….
  [——— This tutorial ———]

- However, we should not fixate solely on quantum advantage.
- As we build the foundation of QML, quantum advantages naturally emerge.
  (e.g., the exponential advantage in learning poly-time physical processes)

Overview

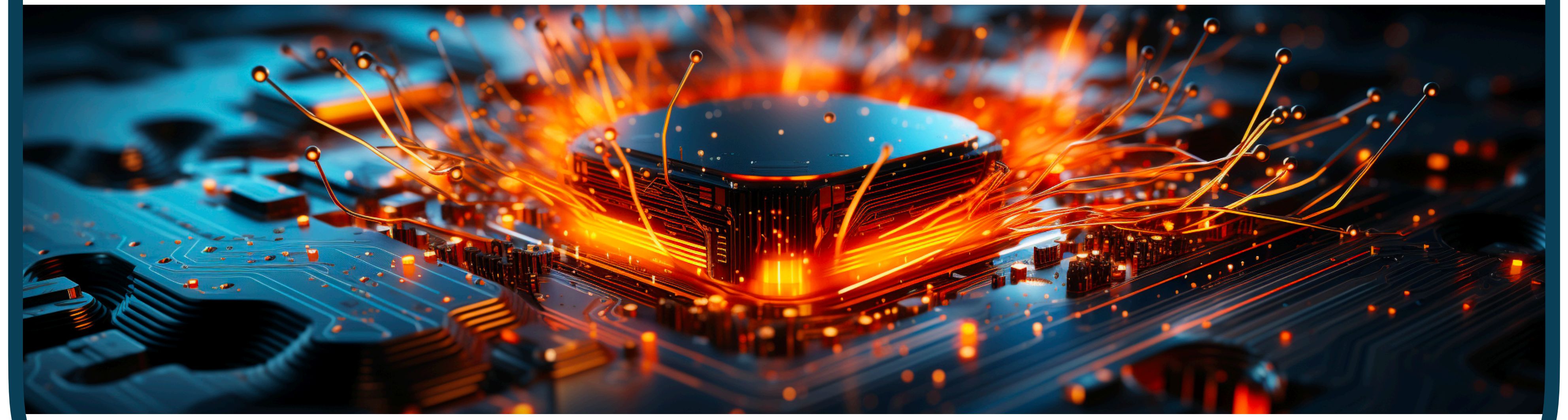
Learning theory for quantum machines

Foundation

How well can quantum machines predict?
How good is the generalization ability of quantum machines?

Quantum advantage

What can quantum machines learn that classical machines cannot?
How big can the advantage be?

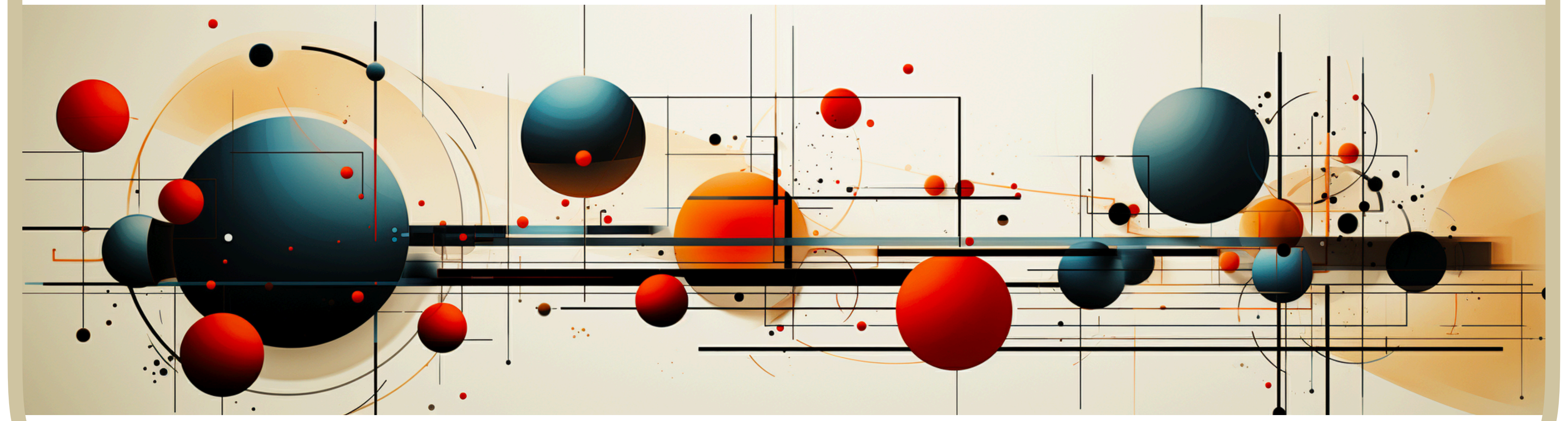


Overview

Learning theory for quantum machines

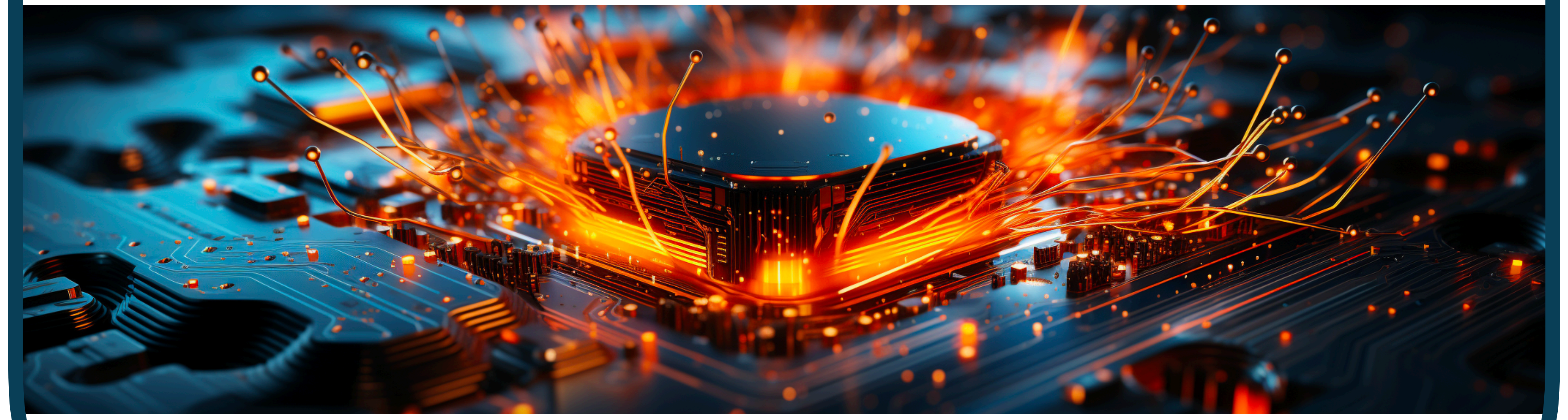
Foundation

How well can quantum machines predict?
How good is the generalization ability
of quantum machines?



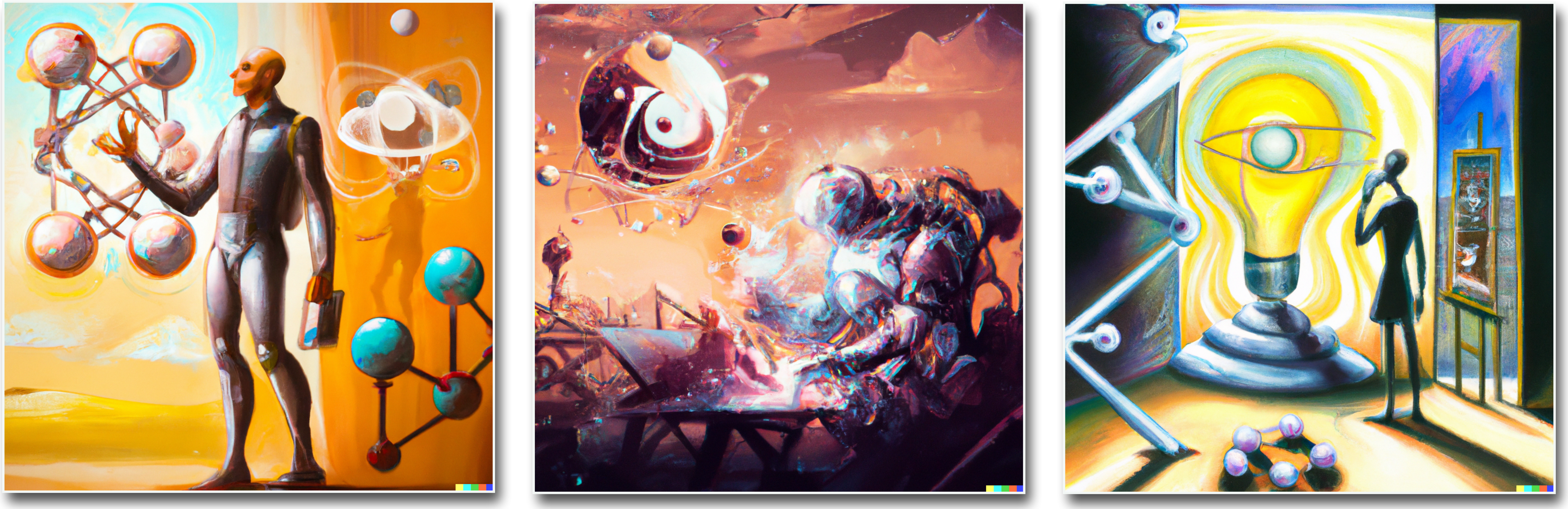
Quantum advantage

What can quantum machines learn
that classical machines cannot?
How big can the advantage be?



Long-term ambition

- How to build a quantum machine capable of learning and discovering new facets of our universe beyond humans and classical machines?



AI (2022) imaging itself learning and discovering new facets of our quantum universe

Q&A

- Questions about learning, about quantum agents, about shadows, about NISQ, about future directions are all welcomed. Ask anything!