Quantum Data Management in the NISQ Era

Rihan Hai

Available next week

TU Delft
Delft University of Technology

Agenda

- Quantum computing for data management (ICDE'24 tutorial)
- **Data management for quantum computing**

A little more about me

- https://infinidata-team.github.io/

InfiniData

What we focus on

Empowering the Future of Data, Today

AI in data lakes
Multimodal data & GPU acceleration

Federated Learning
Data privacy and security

Quantum Data Management
Data Management for Quantum Computing and Quantum Internet

Wenbo Sun

Danning Zhan

Aditya Shankar (with L. Chen)

Tim Littau

A little more about me

InfiniData

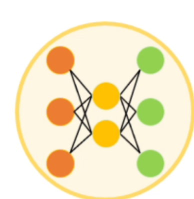
What we focus on

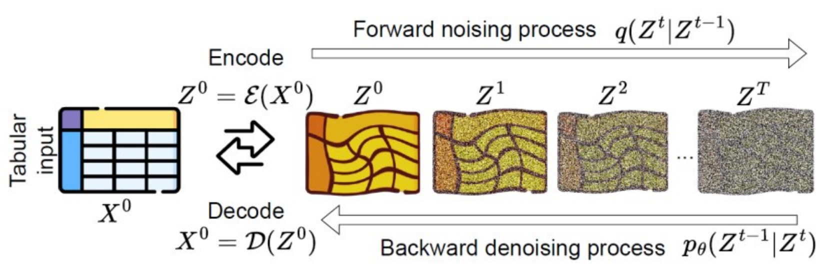
Empowering the Future of Data, Today

AI in data lakes
Multimodal data & GPU acceleration

Federated Learning
Data privacy and security

Quantum Data Management
Data Management for Quantum Computing and Quantum Internet

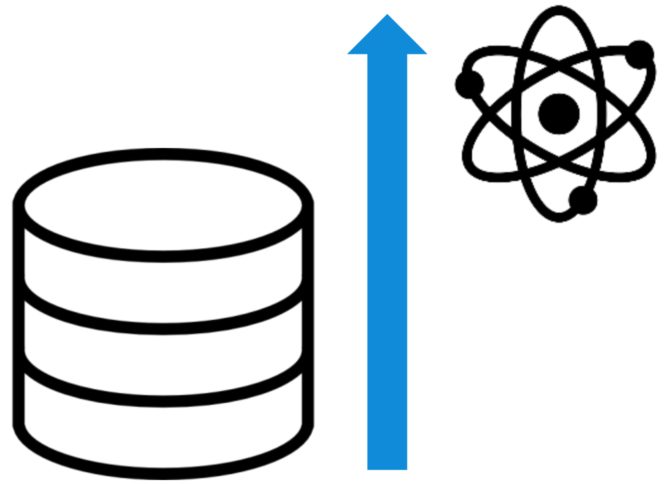




Amalur (CIDR’22, ICDE’23, TKDE’23 ICDE’24, TKDE’24, CIKM’24demo)

SiloFuse (ICDE’24)

ICDE’24 tutorial



Quantum computing for data management

DBMS

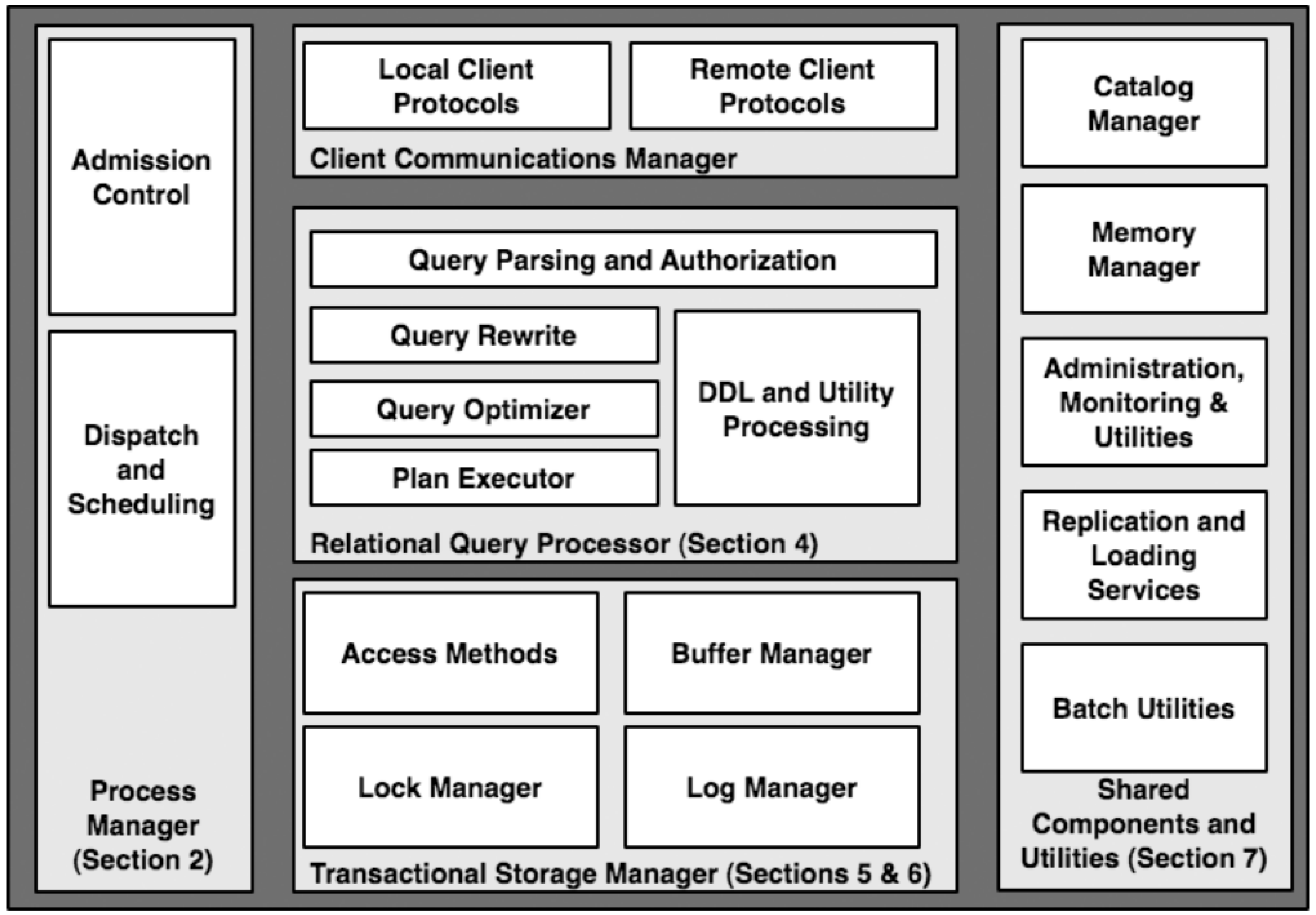
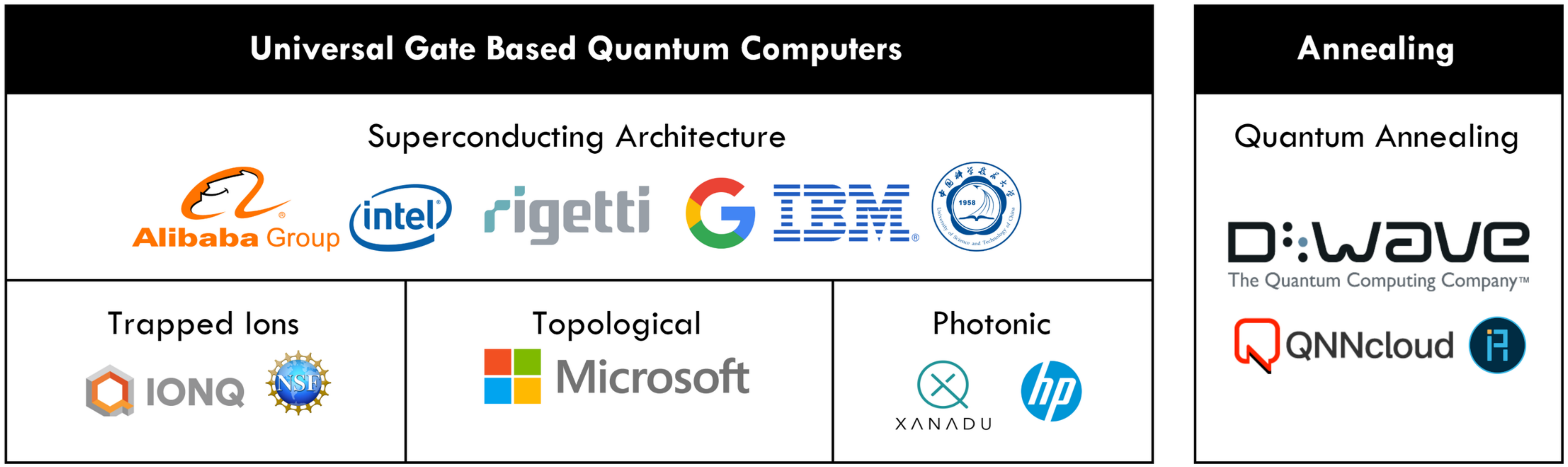


Fig. 1.1 Main components of a DBMS.

Hellerstein, Joseph M., Michael Stonebraker, and James Hamilton. "Architecture of a database system." *Foundations and Trends® in Databases* 1.2 (2007): 141-259.

Quantum computing for data management



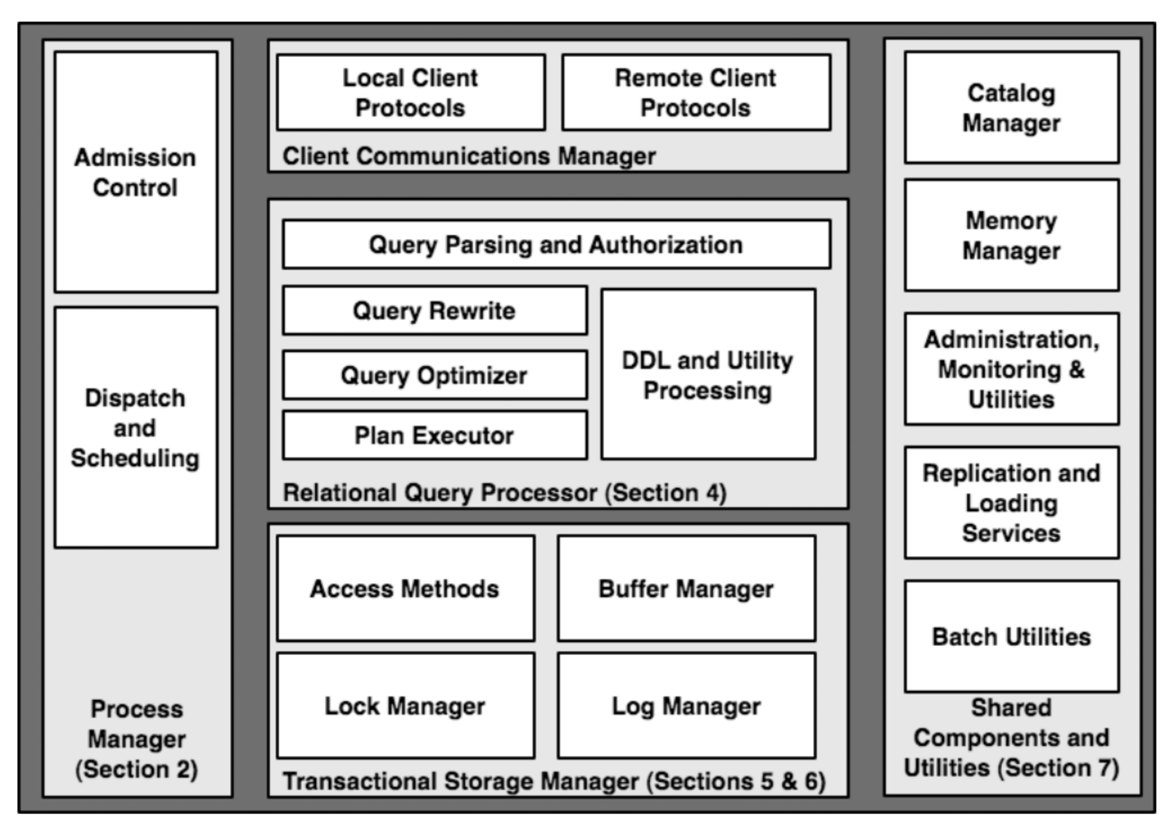


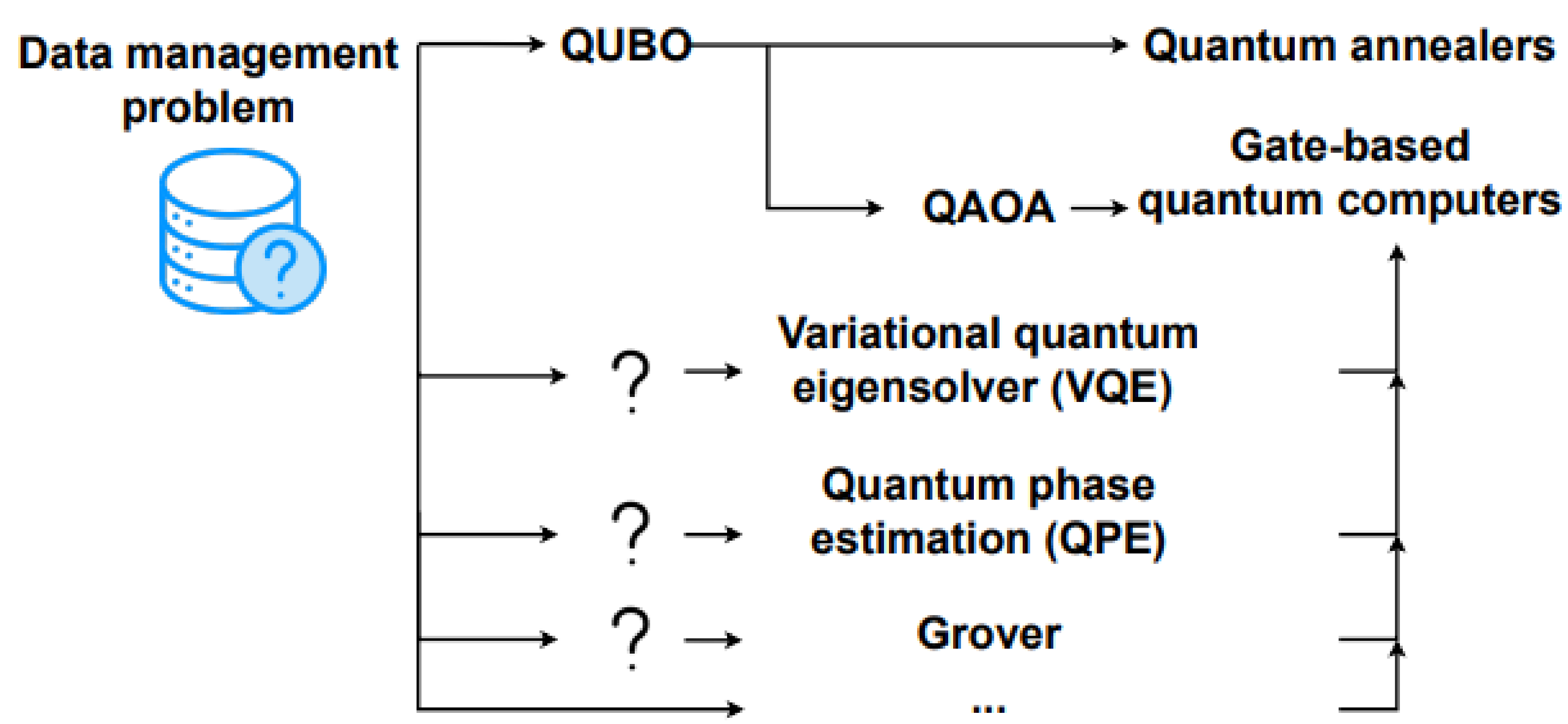
Fig. 1.1 Main components of a DBMS.

DB Problems Solved Using QPUs

Reference	DB problem	Subproblem	Formulation	Intermediate quantum algorithm	Quantum computer
I. Trummer et al., VLDB’16	Query optimization	Multiple query optimization	QUBO	–	Annealing-based
T. Fankhauser et al., IEEE Access, 2023				QAOA	Gate-based
M. Schonberger et al., SIGMOD23		Join ordering		QAOA	Gate-based & annealing-based
N. Nayak et al., BiDEDE ’23				QAOA, VQE	Gate-based & annealing-based
T. Winker et al., BiDEDE ’23			–	VQC	Gate-based
K. Fritsch et al., VLDB’23 Demo	Data integration	Schema matching	QUBO	QAOA	Gate-based & annealing-based
T. Bittner et al., IDEAS’20, OJCC S. Groppe et al., IDEAS’21	Transaction management	Two-phase locking	QUBO	–	Annealing-based

Roadmap

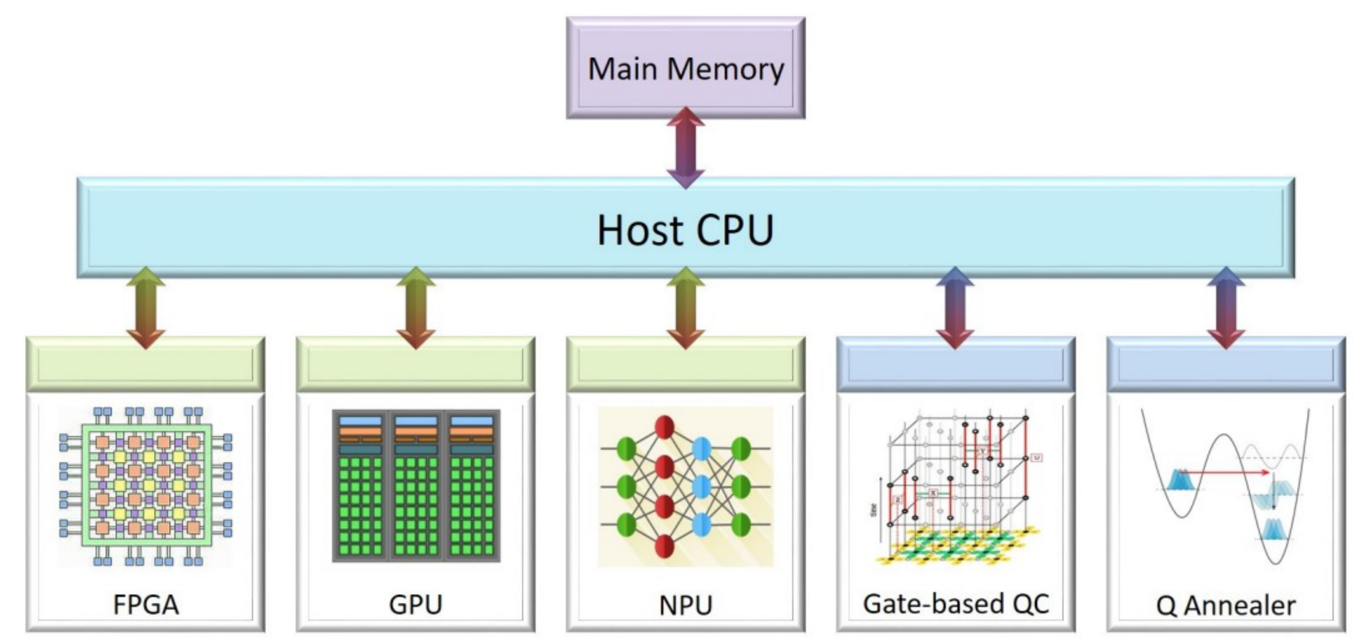
- Solving data management problems on quantum computers
  - Problem benefit from **quantum advantage**, and practically useful
    - Optimization problem
    - Classical approaches have scaling limits
    - Yet it does not require to load a large classical dataset
  - Convert a data management solution to quantum algorithms
  - **Constraints** of current quantum hardware

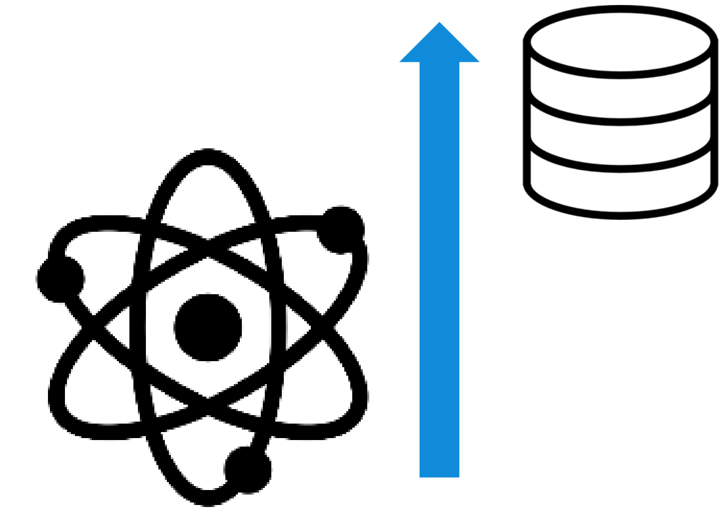


Research Opportunities

- DB problem **reformulation**
- **Hybrid** approach on classical and quantum computers
- Optimization given quantum computer **constraints**

Quantum computer will enhance, not replace, current HPC systems

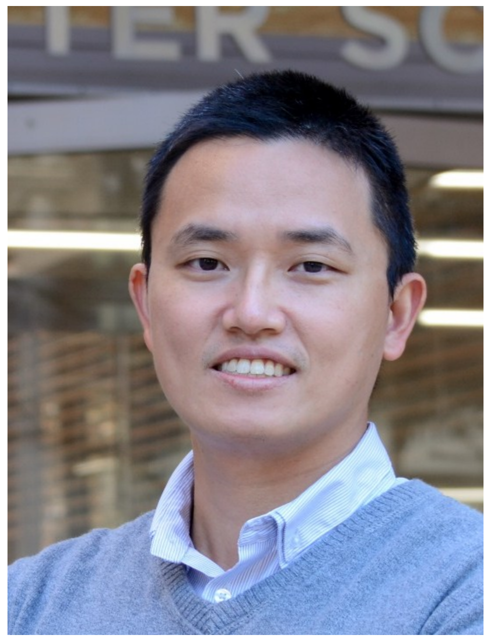




Data management for quantum computing

Many thanks to my collaborators

Floris Geerts
University of Antwerp



Shih-Han Hung
National Taiwan University

Tim Coopmans
Leiden University

Classical data vs. quantum data

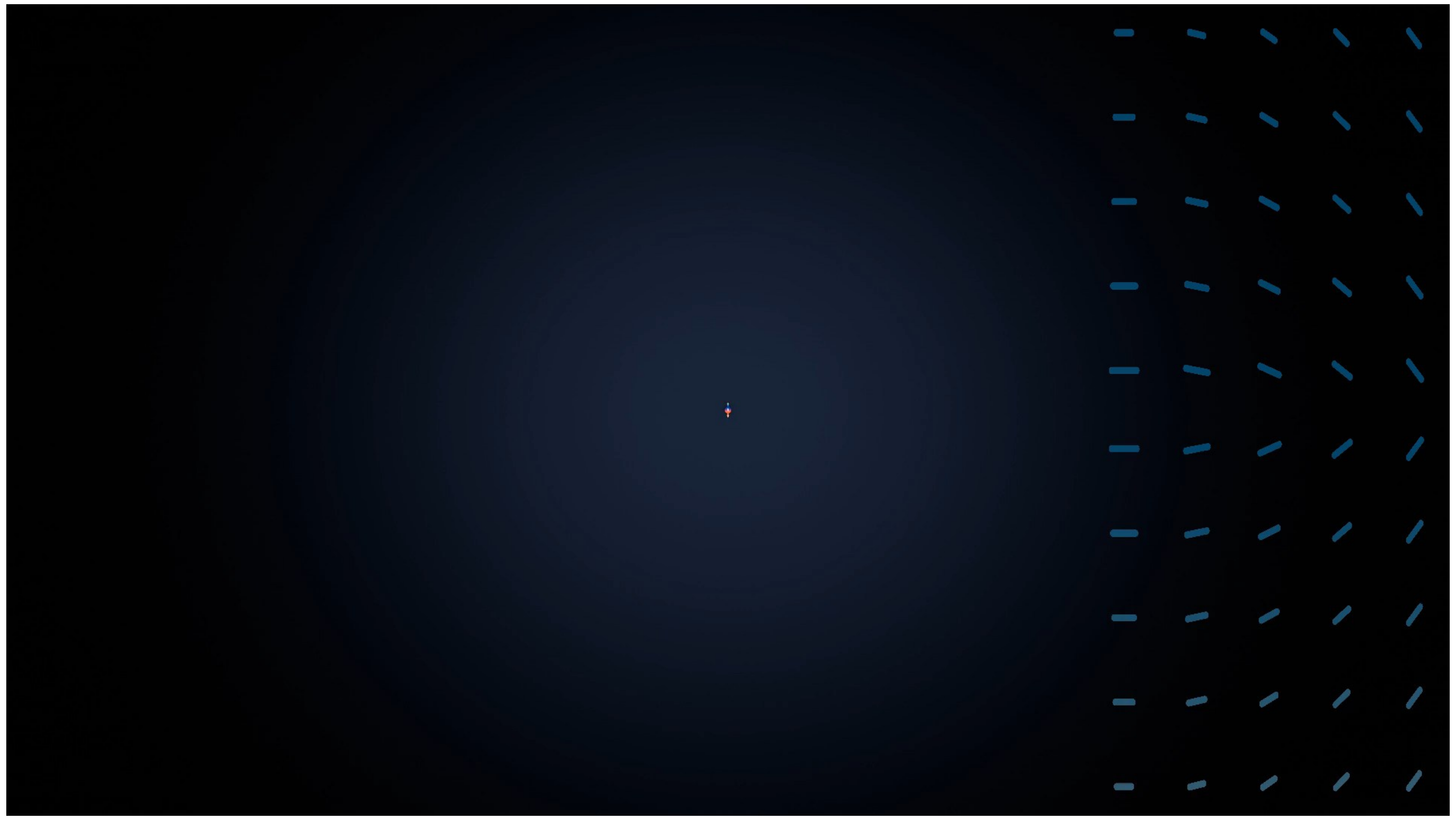
- Classical data
  - Information that is collected, processed, and stored with traditional computing methods
  - Stored and queried using DBMS such as relational databases, document stores, graph databases, and vector databases

- Quantum data
  - Information collected and processed using quantum computing devices that follow the rules of quantum mechanics to their advantage
  - Represented by qubits

Unique features of quantum data

1. Quantum data is probabilistic

Superposition

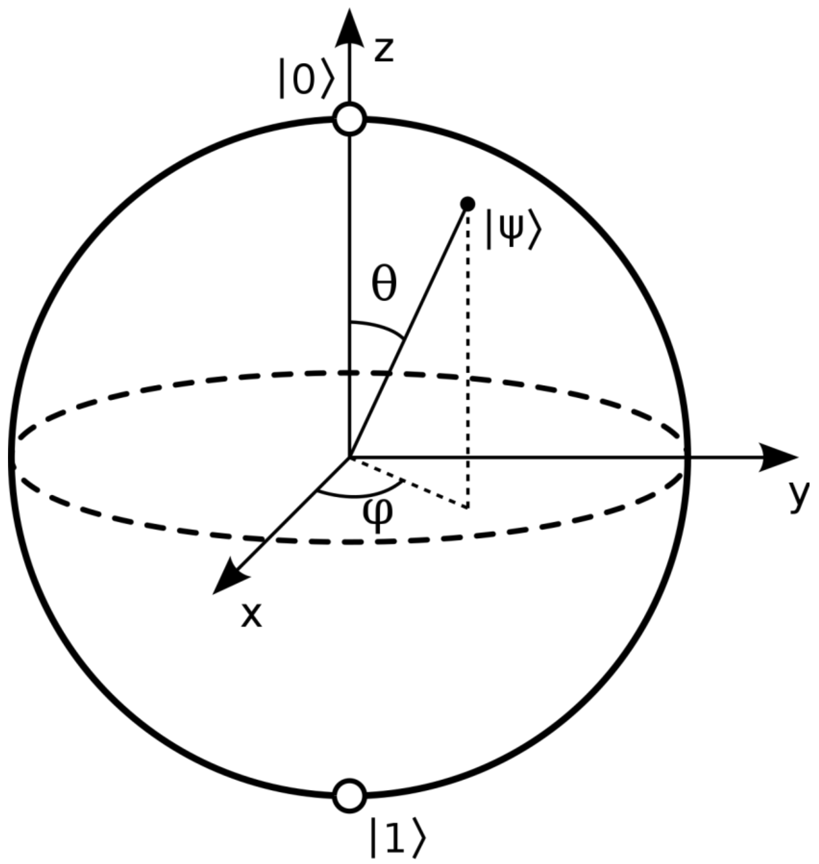


What is a qubit (for us)?

- A qubit is a linear combination of basis states

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

$$\alpha, \beta \in \mathbb{C} \text{ with } |\alpha|^2 + |\beta|^2 = 1$$



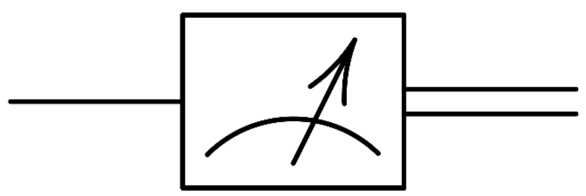
Unique features of quantum data

1. Quantum data is probabilistic

- α, β are called probability amplitudes
- When measuring, $|\alpha|^2$ is probability of finding qubit in state $|0\rangle$

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

$$\alpha, \beta \in \mathbb{C} \text{ with } |\alpha|^2 + |\beta|^2 = 1$$

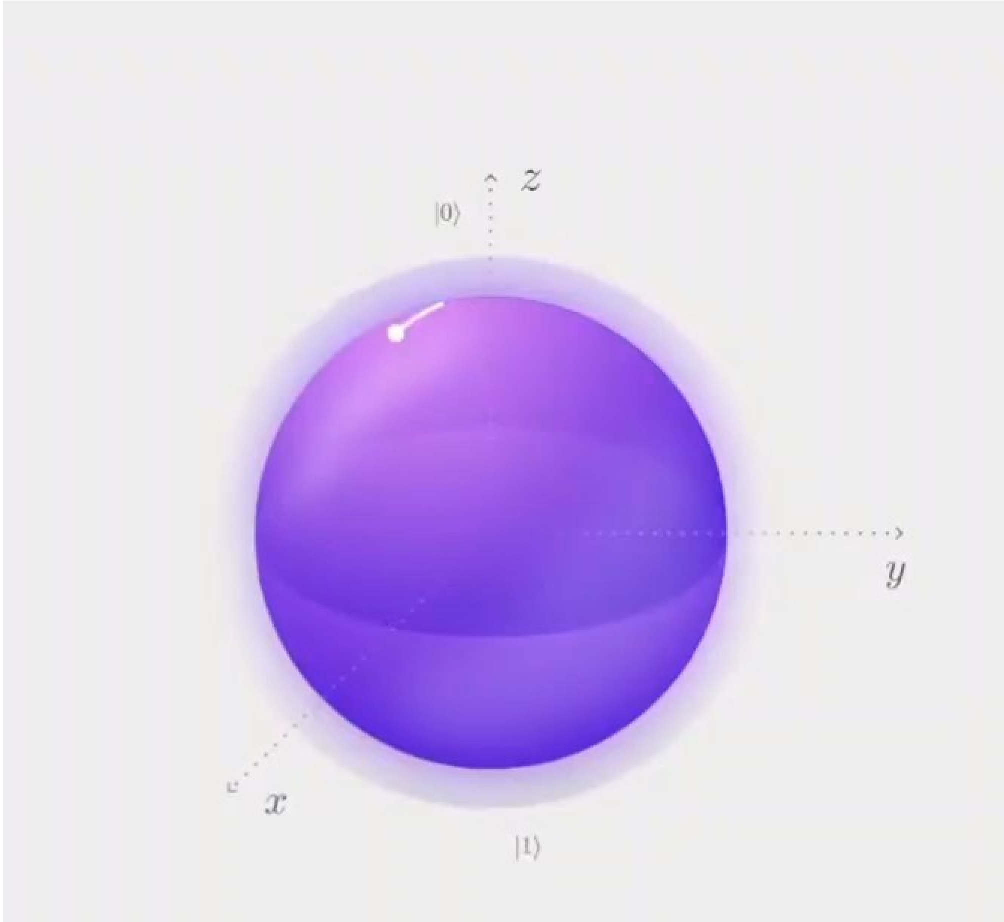
$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$  $|0\rangle$ or $|1\rangle$

Unique features of quantum data

2. Quantum data is fragile

Quantum noise results from unwanted coupling environment

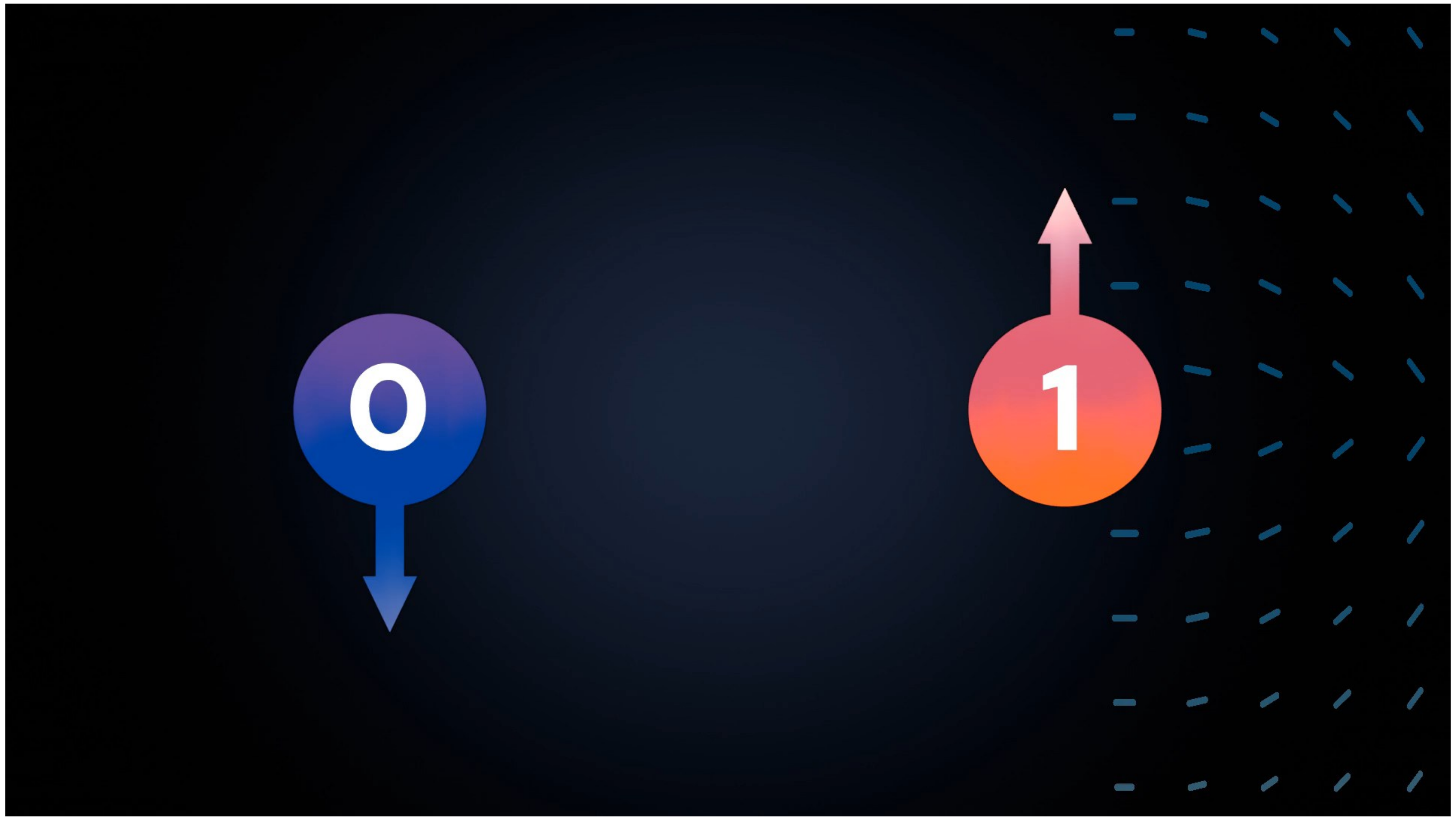
- Depolarizing
- Bit & phase flipping
- Amplitude & phase damping



Unique features of quantum data

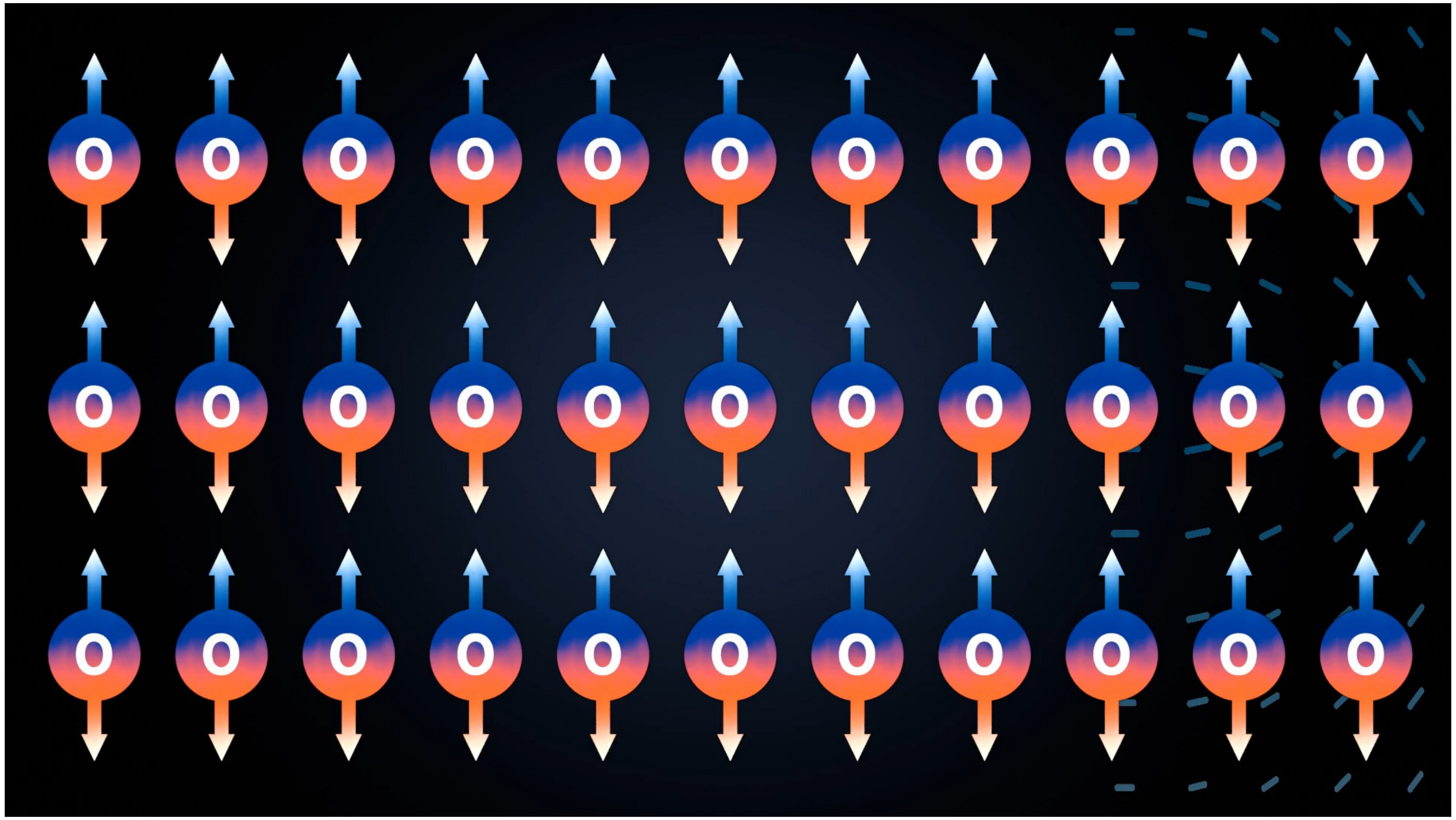
3. Quantum data can be entangled

Entanglement

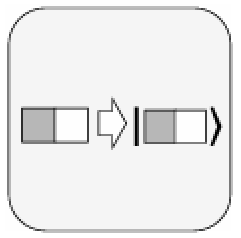
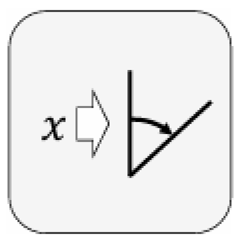
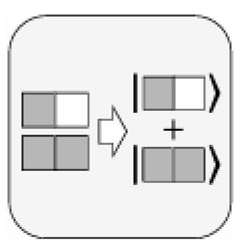
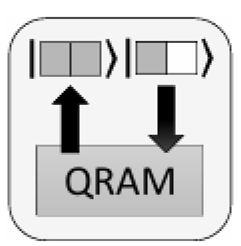
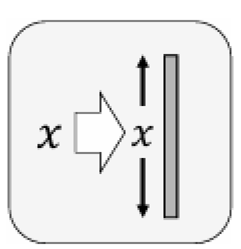


From quantum data to classical data

- Qubit fate (0 or 1) determined upon measurement



From classical data to quantum data

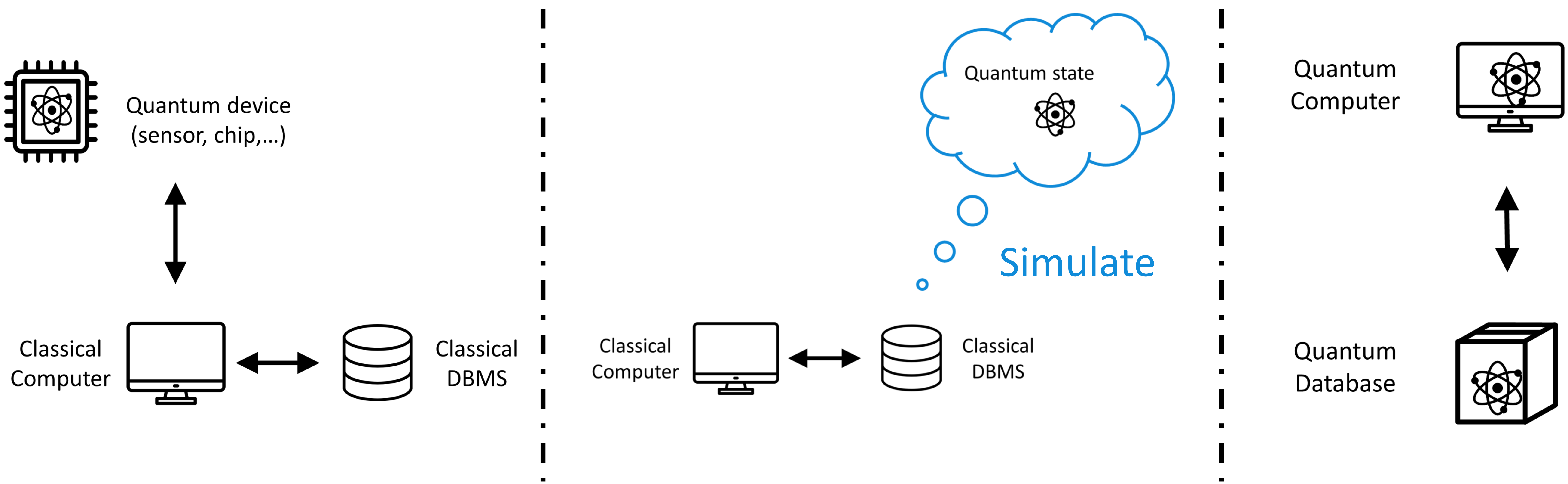
Encoding Pattern		Encoding	Req. Qubits
	BASIS ENCODING [1]	$x_i \approx \sum_{i=-k}^{m} b_i 2^i \mapsto \|b_m \ldots b_{-k}\rangle$	$l = k+m$ per data-point
	**ANGLE ENCODING**	$x_i \mapsto cos(x_i)\|0\rangle + sin(x_i)\|1\rangle$	1 per data-point
	QUAM ENCODING [1]	$X \mapsto \sum_{i=0}^{n-1} \frac{1}{\sqrt{n}} \|x_i\rangle$	l
	**QRAM ENCODING**	$X \mapsto \sum_{n=0}^{n-1} \frac{1}{\sqrt{n}} \|i\rangle \|x_i\rangle$	$\lceil \log n \rceil + l$
	AMPLITUDE ENCODING [1]	$X \mapsto \sum_{i=0}^{n-1} x_i \|i\rangle$	$\lceil \log n \rceil$

M. Weigold, J. Barzen, F. Leymann and M. Salm, "Expanding Data Encoding Patterns For Quantum Algorithms," *2021 IEEE 18th International Conference on Software Architecture Companion (ICSA-C)*, Stuttgart, Germany, 2021, pp. 95-101, doi: 10.1109/ICSA-C52384.2021.00025.

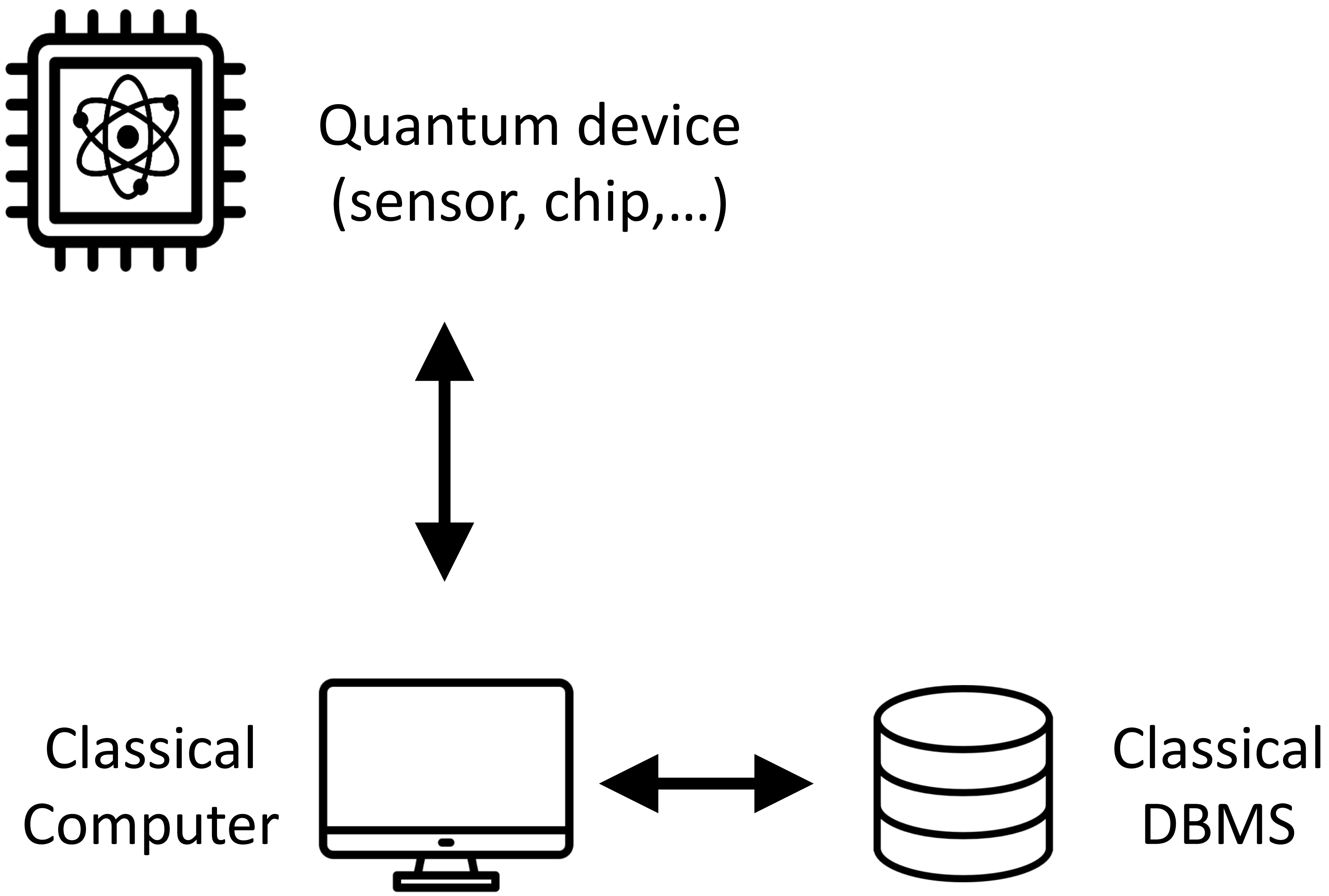
DB for QC: where to start?

Landscape: data management for quantum computing

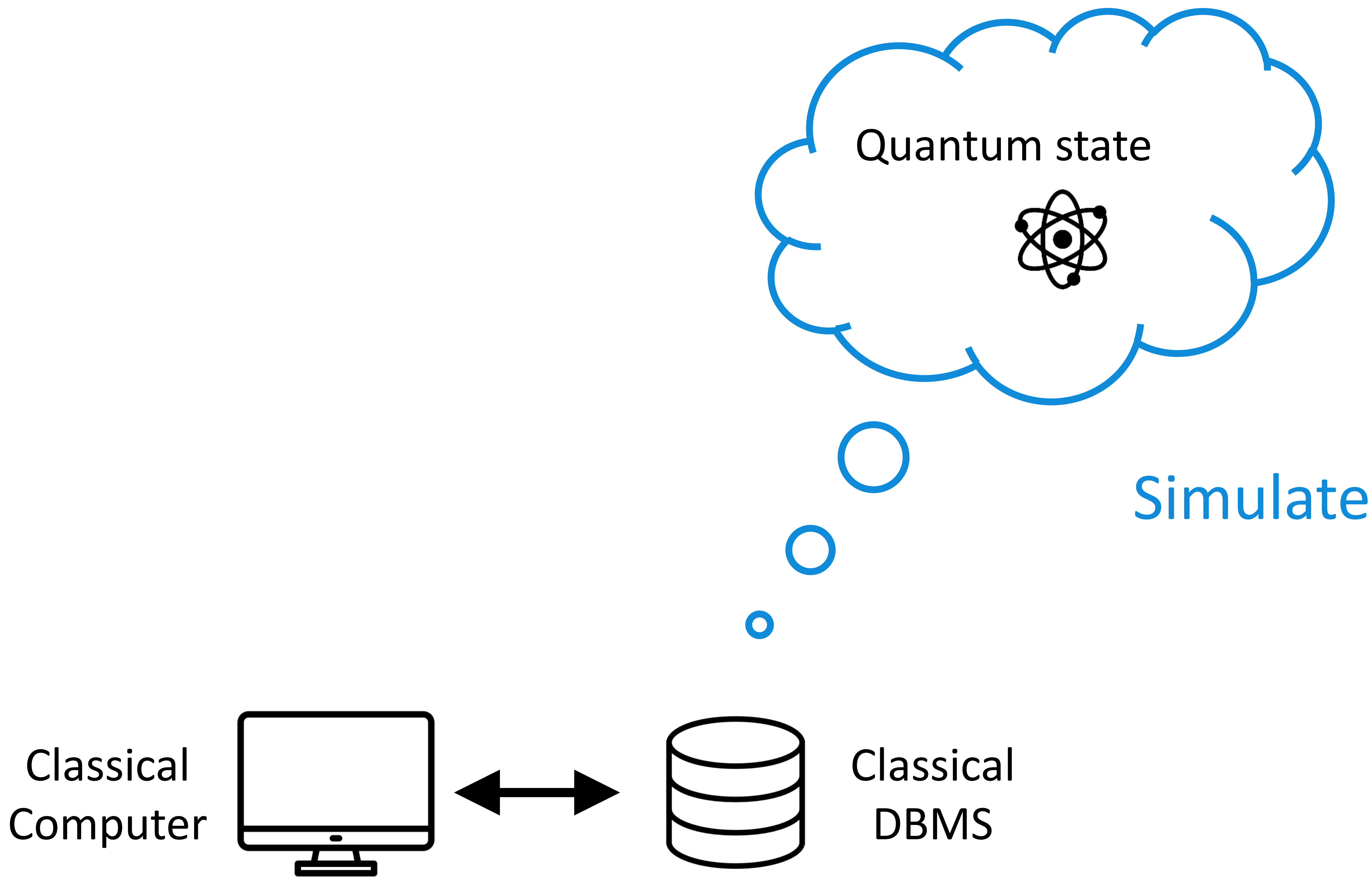
I Classical | II Classical-Quantum | III Quantum



I Classical data

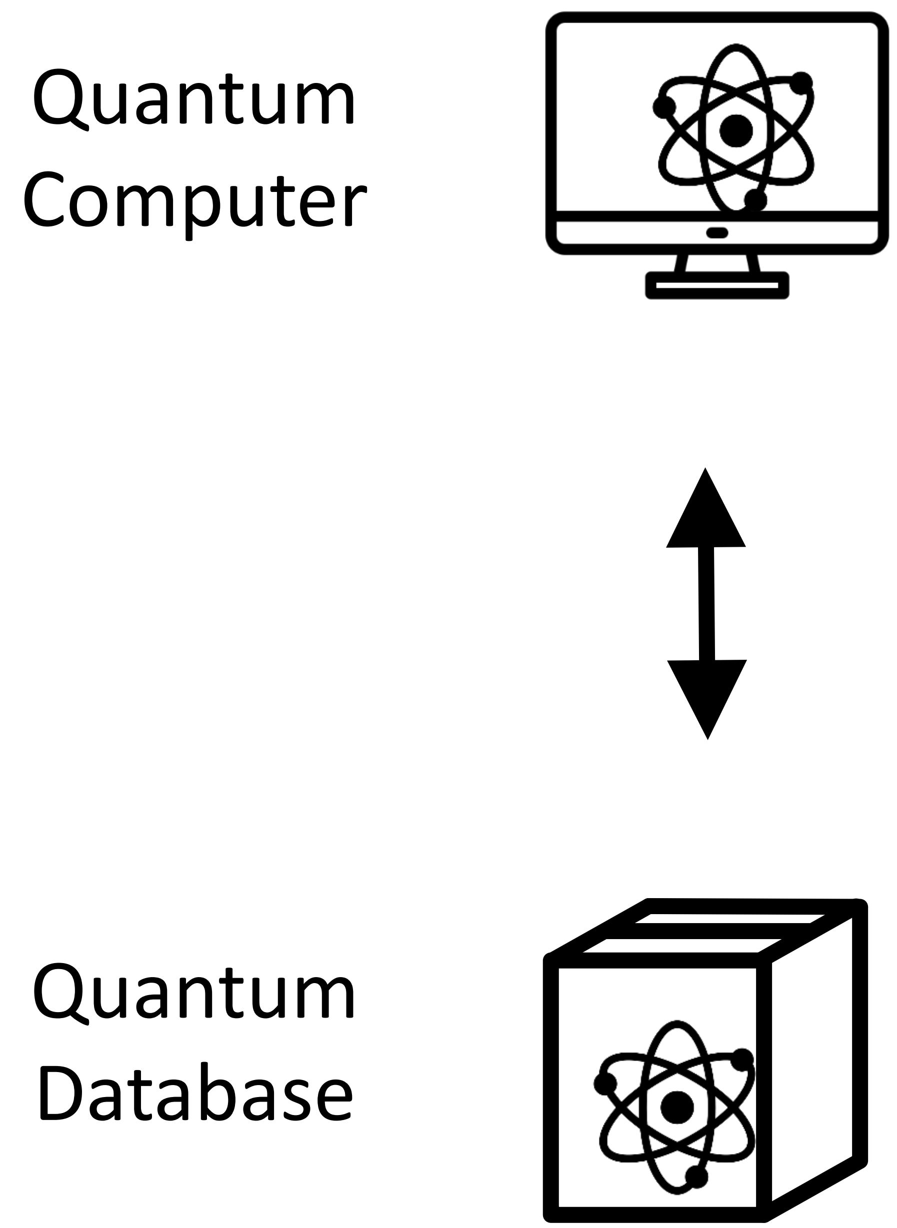


II Classical-Quantum



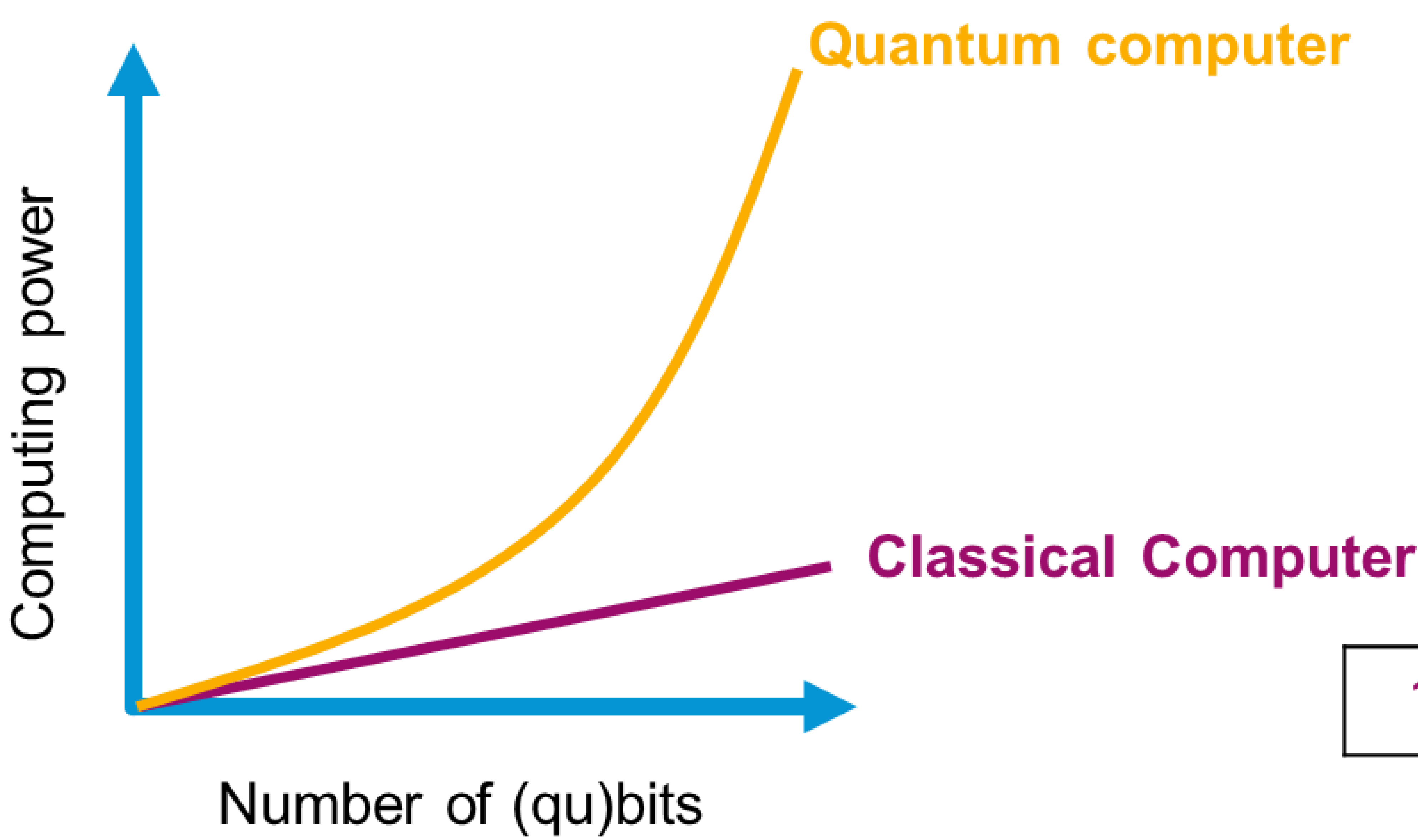
III Quantum

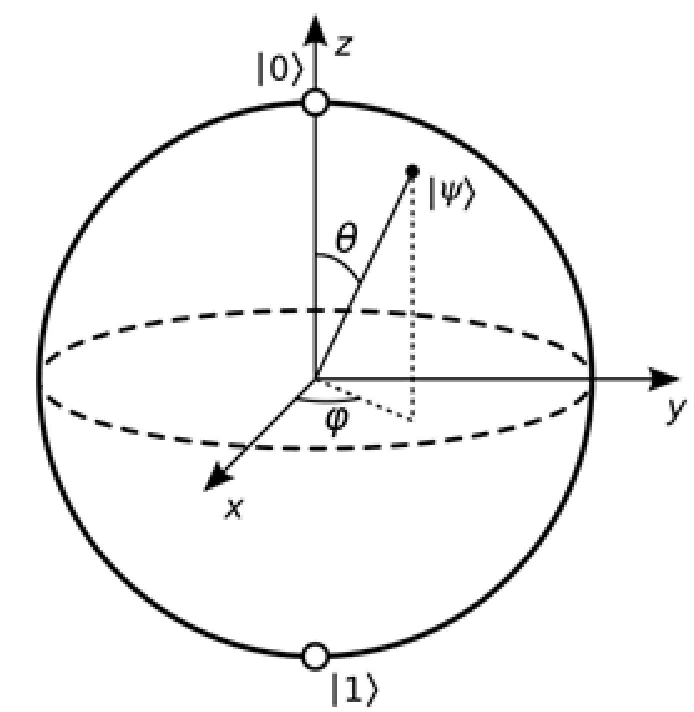
Fault-Tolerant Quantum Computing (FTQC) Era



III Quantum

- Quantum computing = Enormous computing power

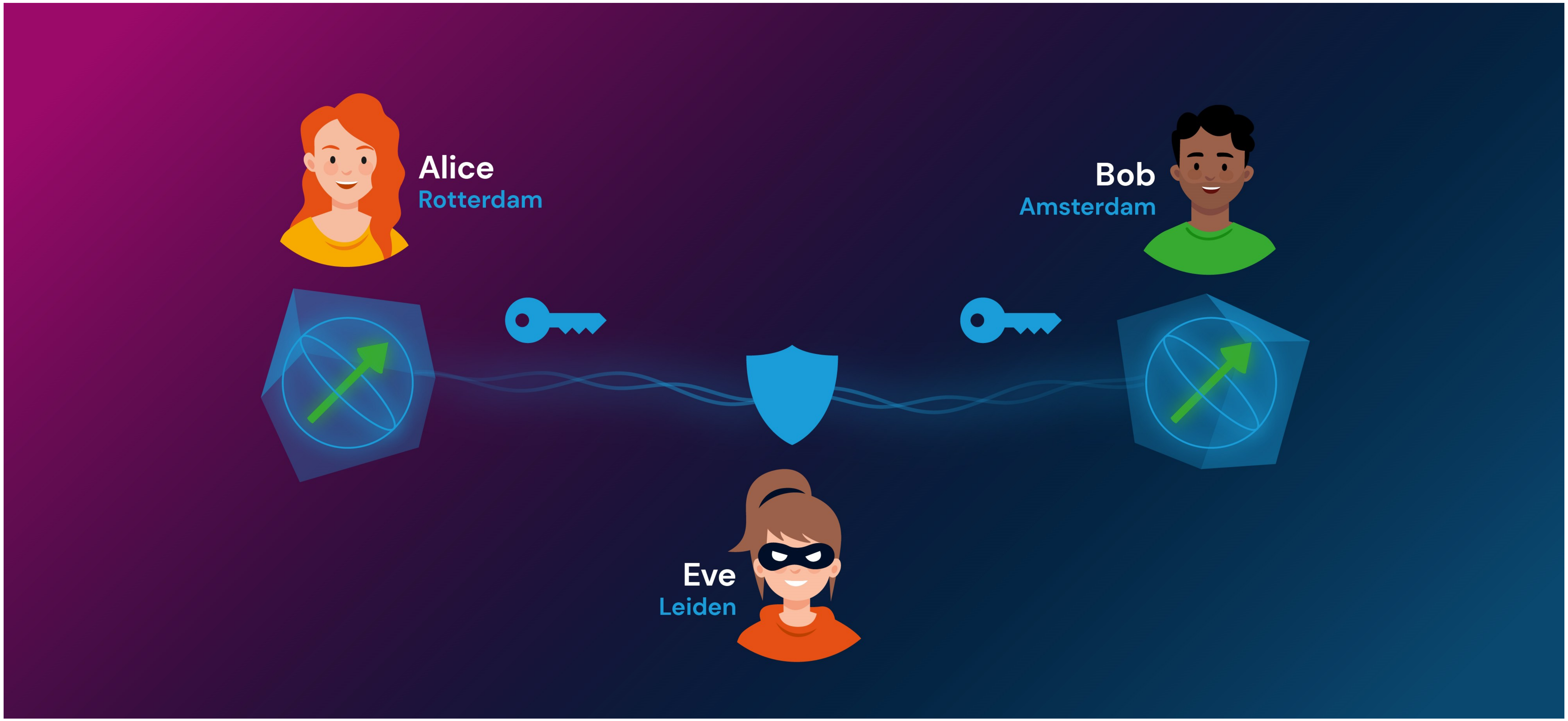




1	1	0	1

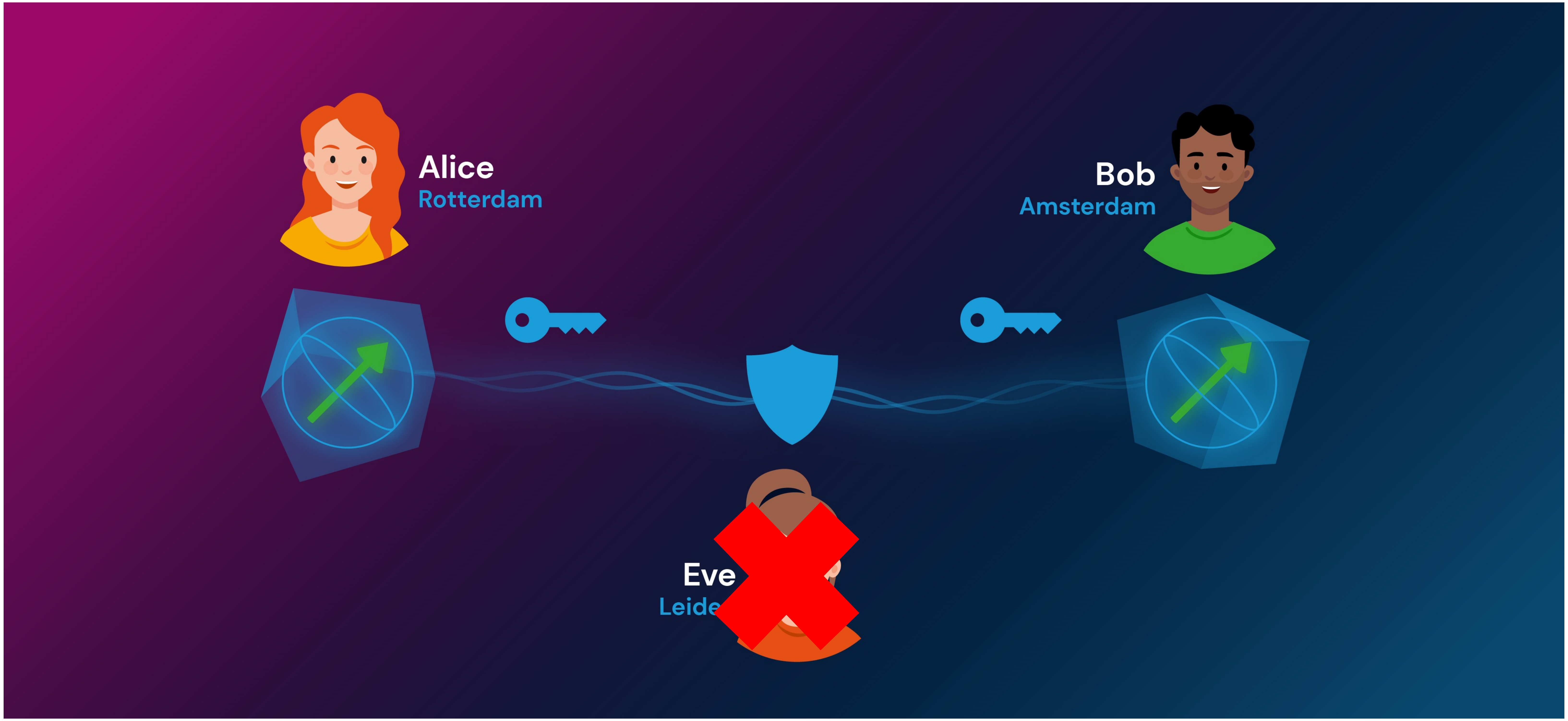
III Quantum

- Quantum communication = Inherently safe communication



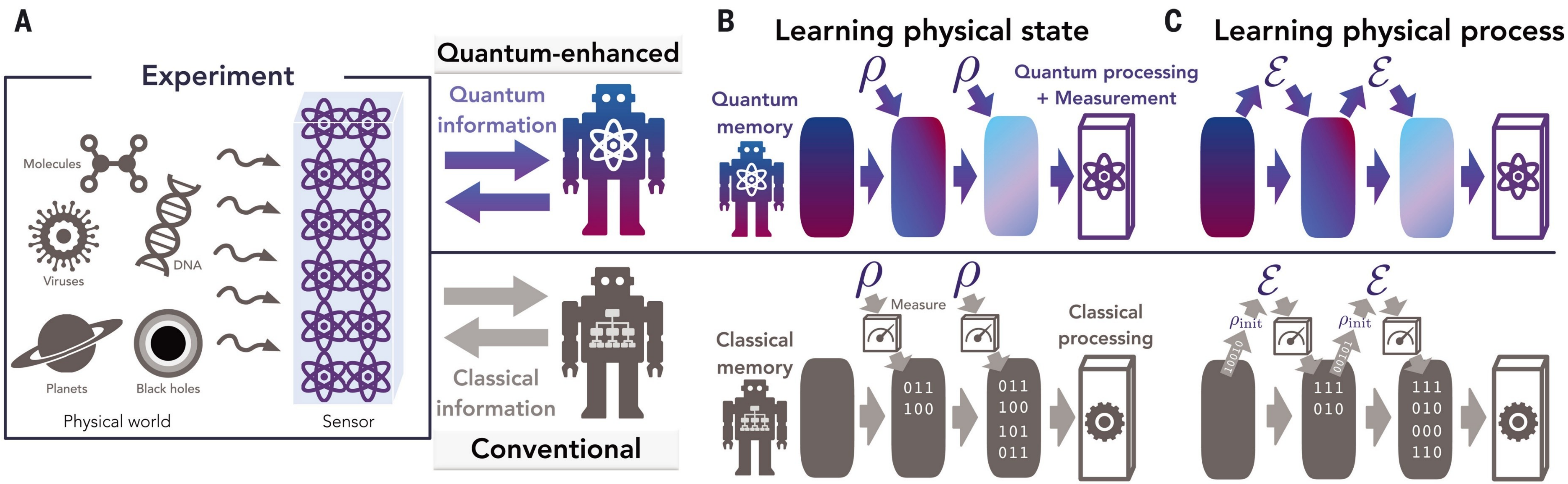
III Quantum

- Quantum communication = Inherently safe communication



III Quantum

- Quantum data is collected from quantum sensing systems; then stored and processed via the quantum memory of a quantum computer

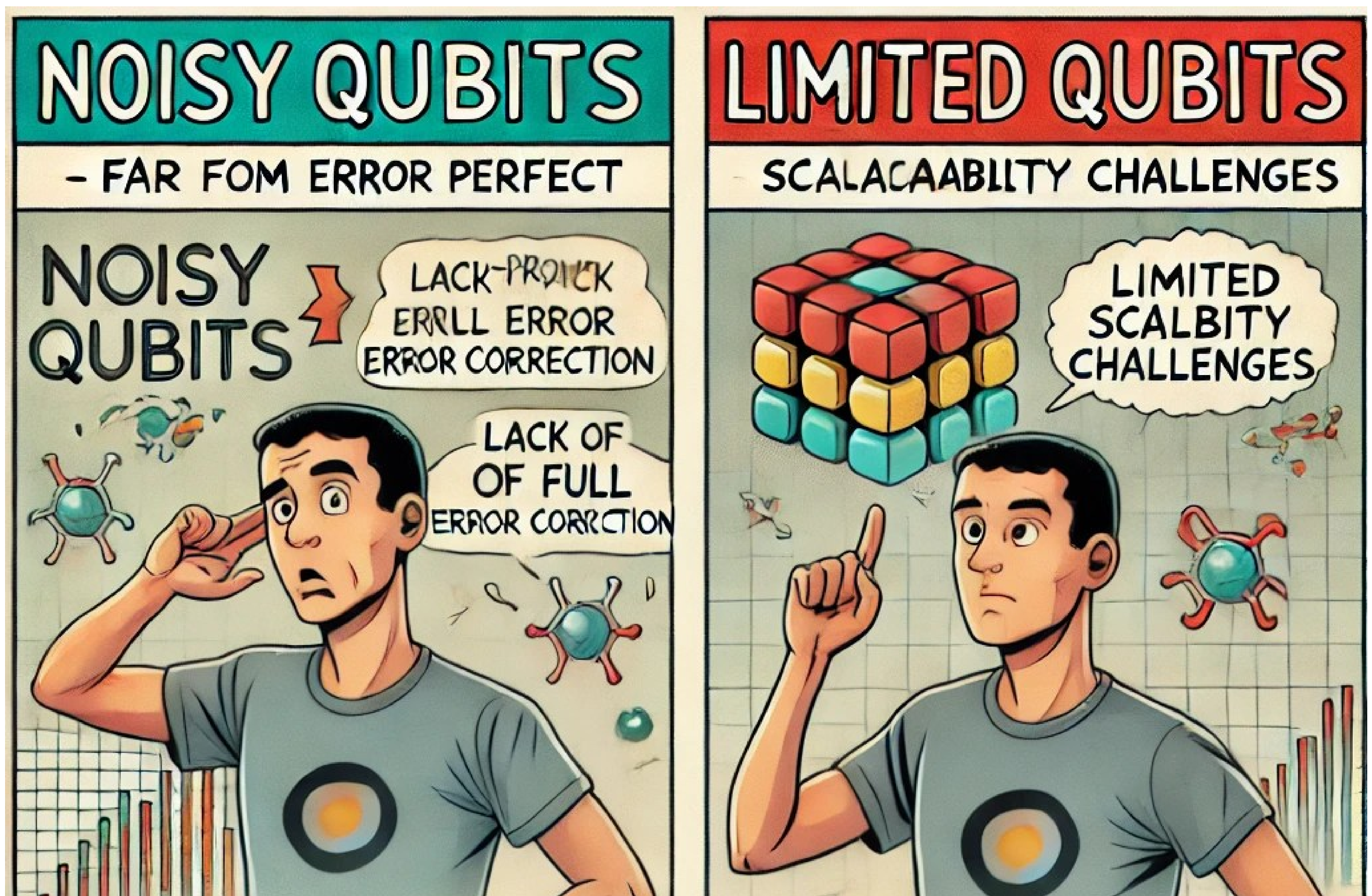


Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li,Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill,et al. 2022. Quantum advantage in learning from experiments. Science 376, 6598(2022), 1182–1186

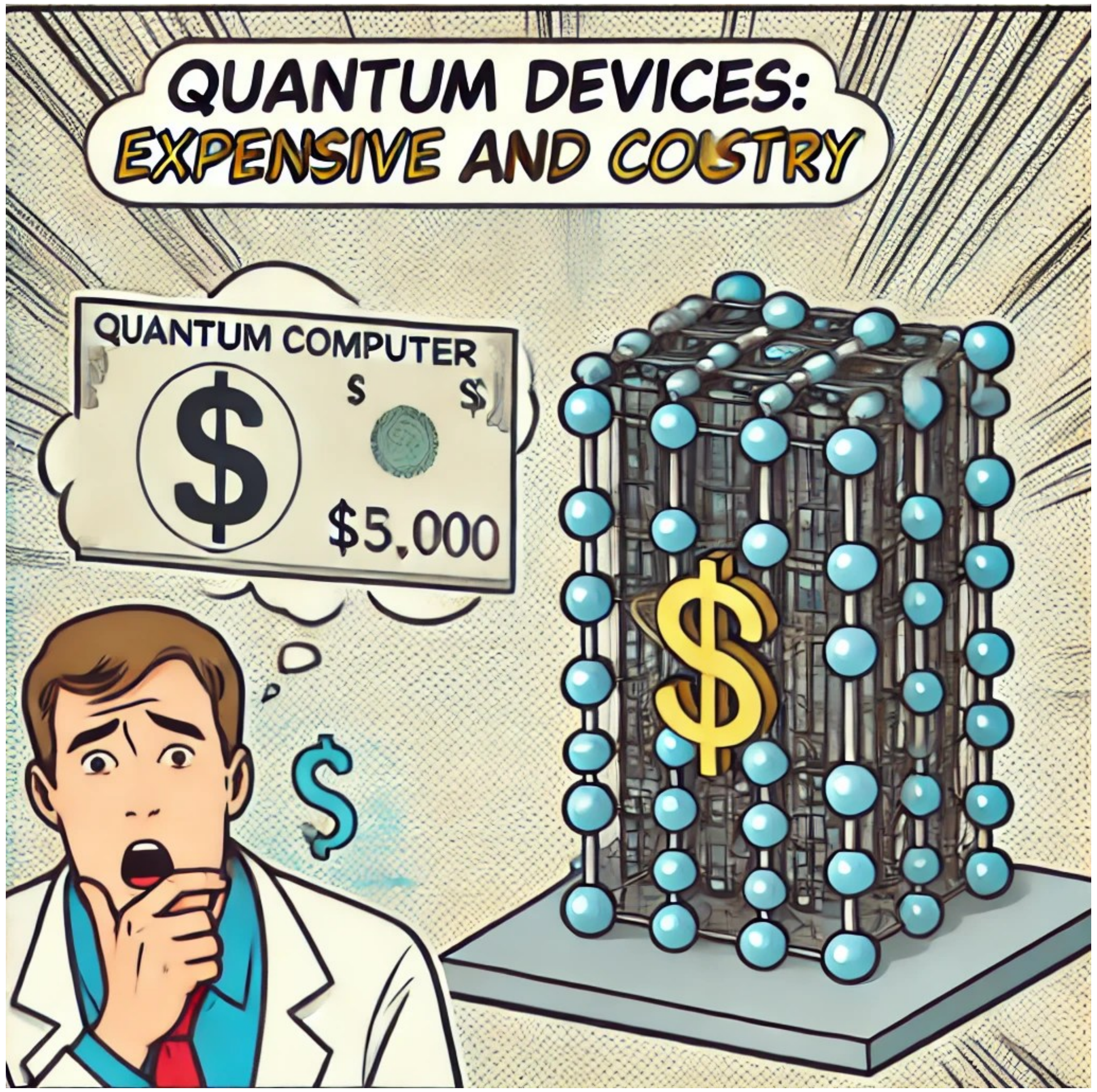
Noisy Intermediate-Scale Quantum (NISQ)

Quantum Computing in the NISQ era and beyond

John Preskill



Noisy Intermediate-Scale Quantum (NISQ)



Databases to the rescue

Can DB technologies boost the development of quantum computing?

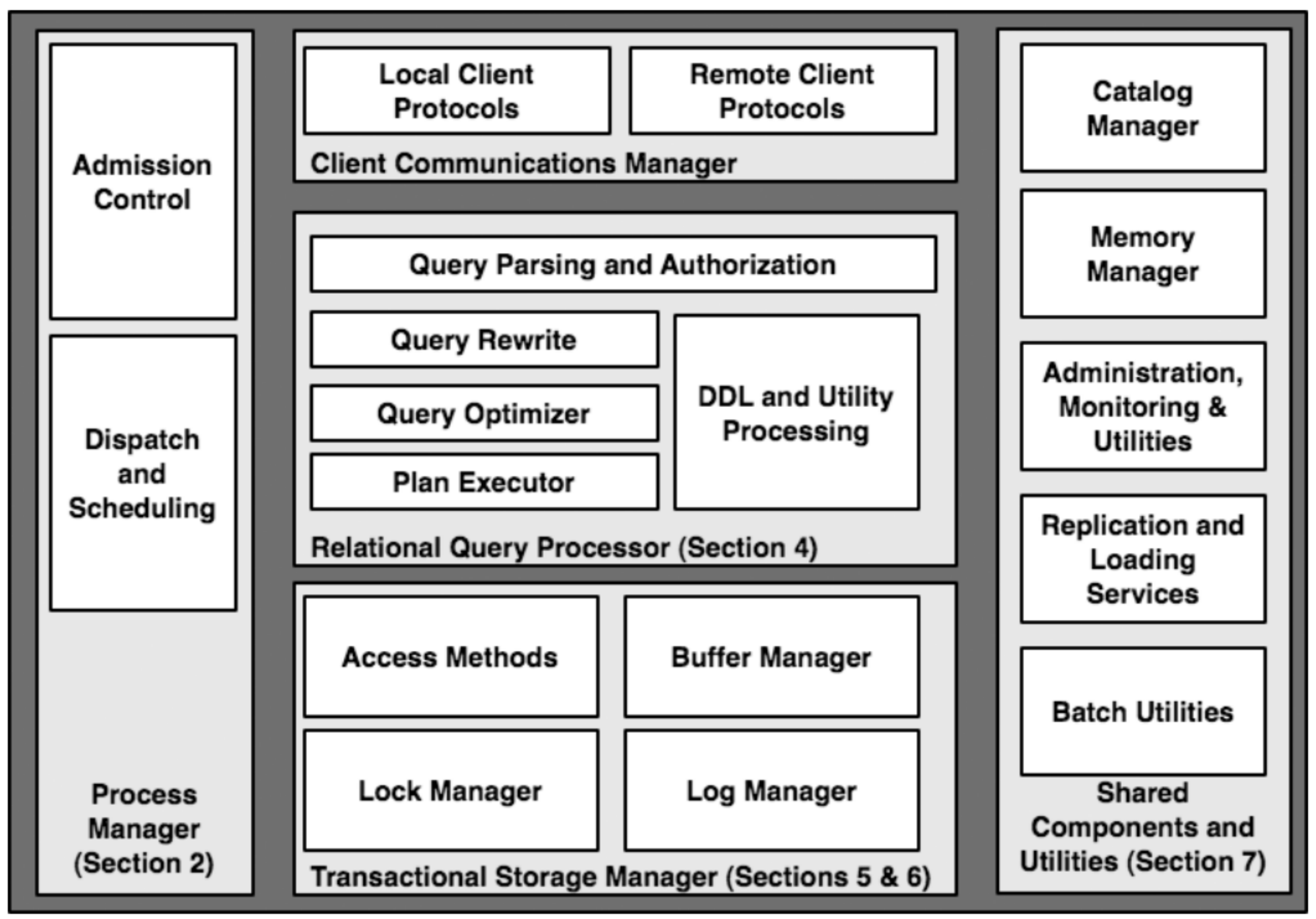
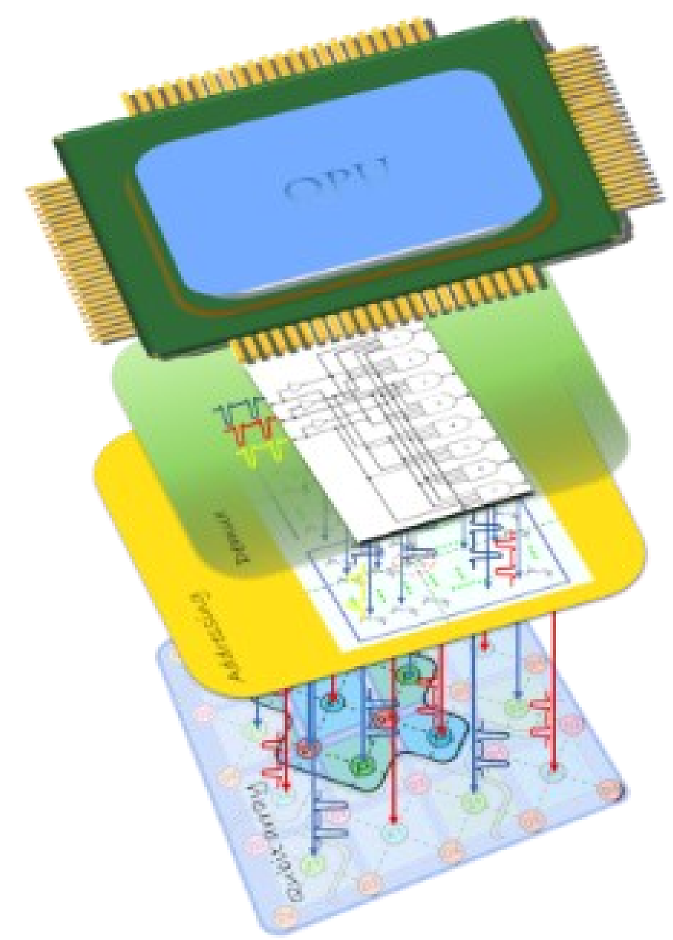


Fig. 1.1 Main components of a DBMS.



Quantum computing

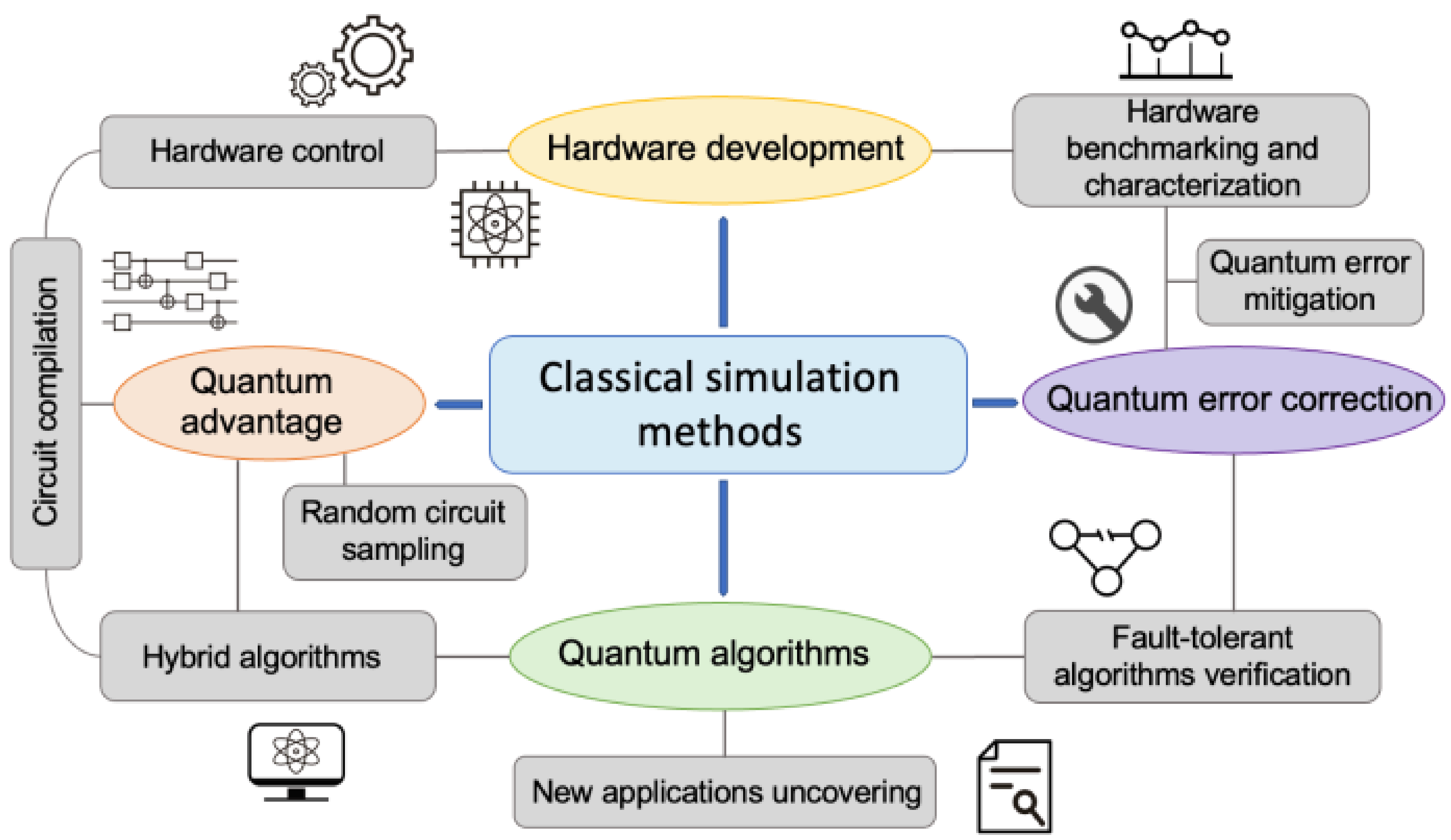
Many possibilities

Can DB technologies boost the development of quantum computing?

In-DB Tensor Computation	→	Classical simulation
Graph analytics	→	Error correction
Distributed databases	→	Quantum network
Scientific data management	→	Quantum experiments

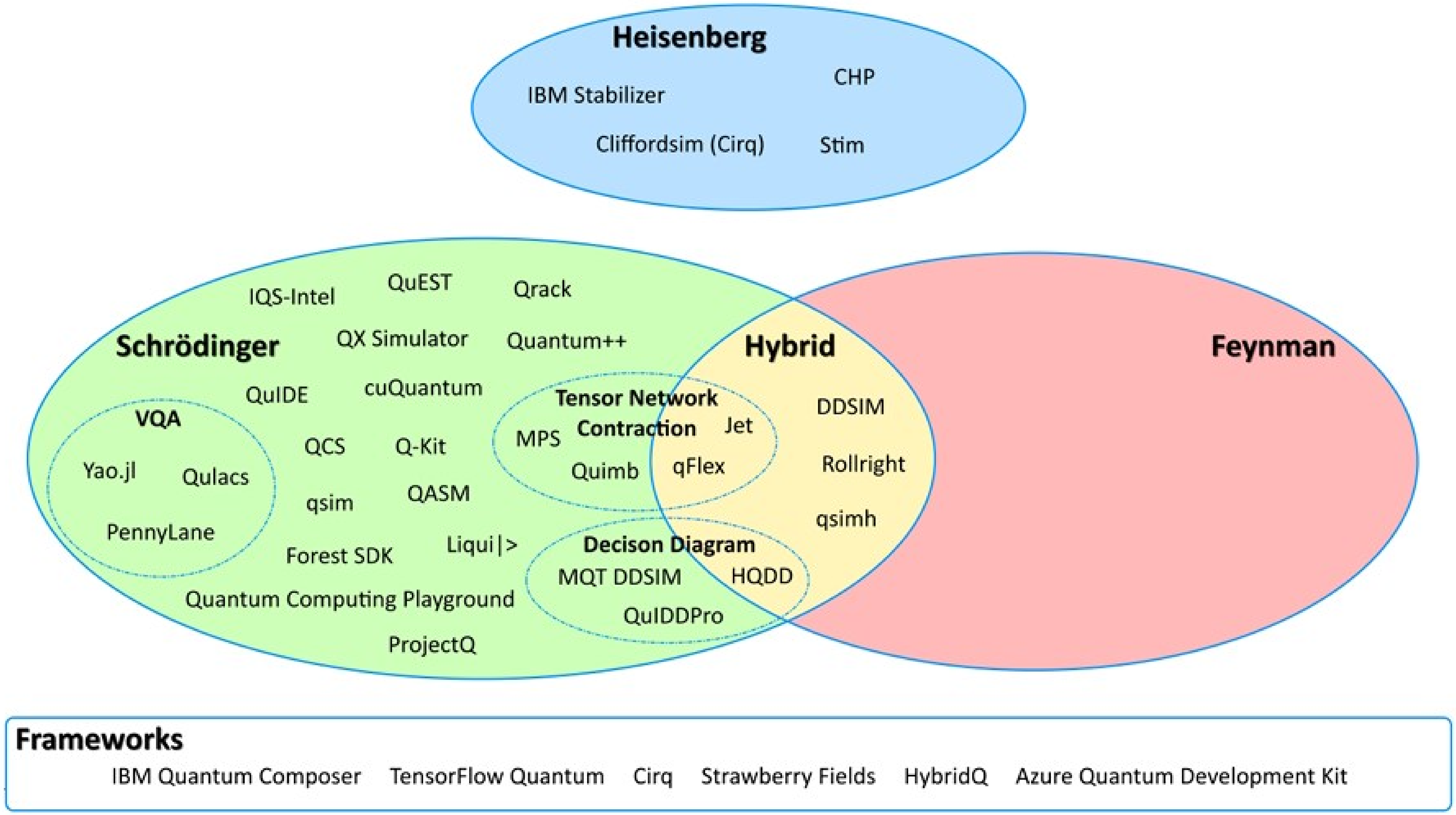
Classical simulation

- The process of emulating quantum computation, enabling researchers to model and analyze quantum processes as if they were operating on actual quantum hardware
- A **powerful, foundational** tool



Xiaosi Xu, Simon Benjamin, Jinzhao Sun, Xiao Yuan, and Pan Zhang. 2023. A Herculean task: Classical simulation of quantum computers. https://arxiv.org/abs/2302.08880

Categorization of existing methods



Kieran Young, Marcus Scese, and Ali Ebnenasir. 2023. Simulating Quantum Computations on Classical Machines: A Survey. arXiv:2311.16505

Strong simulation vs. weak simulation

- Strong simulation
  - Compute the output

Theorem 1 (Gottesman-Knill) *Every (uniform family of) Clifford circuit(s), when applied to the input state* $|\mathbf{0}\rangle \equiv |0\rangle^{\otimes N}$ *and when followed by a* Z *measurement of the first qubit, can be efficiently simulated classically in the strong sense.*

- Weak simulation
  - Sample from the output

Theorem 2 *Let* $\mathcal{C}$ *be an arbitrary poly-sized Clifford circuit. Then there exists a poly-sized Clifford circuit* $\mathcal{C}'$ *satisfying* $\mathcal{C}|\mathbf{0}\rangle = \mathcal{C}'|\mathbf{0}\rangle$ *such that* $\mathcal{C}'$ *can be decomposed into three "rounds": (ROUND 1) apply Hadamard gates to an arbitrary subset of qubits; (ROUND 2) apply a poly-sized circuit of NOTs and CNOTs; (ROUND 3) apply a poly-size circuit of PHASEs and CPHASEs. The circuit* $\mathcal{C}'$ *can be efficiently determined.*

Theorem 3 *Let* $\mathcal{C}$ *be an arbitrary* n*-qubit Clifford operation. Then there exist: (a) poly-size circuits* M_1 *and* M_2 *composed of* CNOT, PHASE *and* CPHASE *gates and (b) a tensor product of* HADAMARD *gates and identities* $\mathcal{H} = H^S \otimes I$ *acting nontrivially on a subset* S *of the qubits, such that* $\mathcal{C} \propto M_2 \mathcal{H} M_1$*. Moreover,* M_1*,* M_2 *and* $\mathcal{H}$ *can be determined efficiently.*

Simulation problem: scalability

- 3-qubit GHZ state

$$\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$$

The basis states for 3 qubits are $|000\rangle$, $|001\rangle$, $|010\rangle$, $|011\rangle$, $|100\rangle$, $|101\rangle$, $|110\rangle$, $|111\rangle$.

- GHZ state as a vector

$$2^3 \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right.$$

Amplitude of the vector $\mathbf{2^n}$, where n is the number of qubits

Simulation problem: scalability

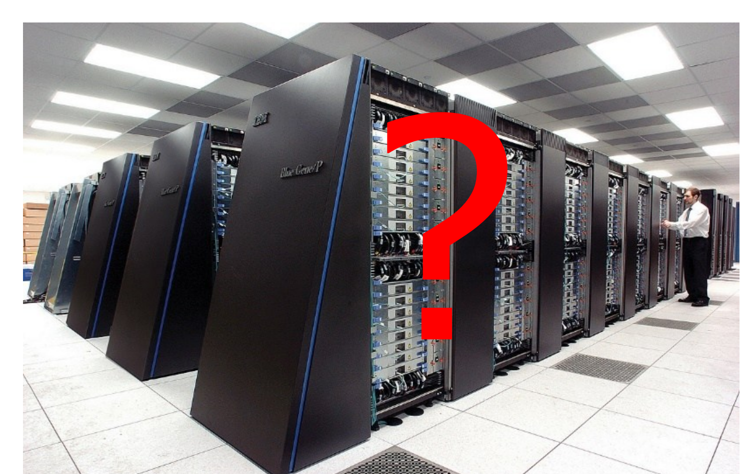
- Amplitude of the vector 2^{n+4}
  - n is the number of qubits
  - 2^4 for the double-precision complex numbers
- Reaching the memory limits of today's supercomputers

nature physics — ARTICLES
https://doi.org/10.1038/s41567-018-0124-x

Characterizing quantum supremacy in near-term devices

Sergio Boixo[1*], Sergei V. Isakov[2], Vadim N. Smelyanskiy[1], Ryan Babbush[1], Nan Ding[1], Zhang Jiang[3,4], Michael J. Bremner[5], John M. Martinis[6,7] and Hartmut Neven[1]

2.25 **petabytes** for **48** qubits (single precision)



Solutions to overcome the memory restriction

- Approximation
- Data compression
- Parallelization
- Distributed computing

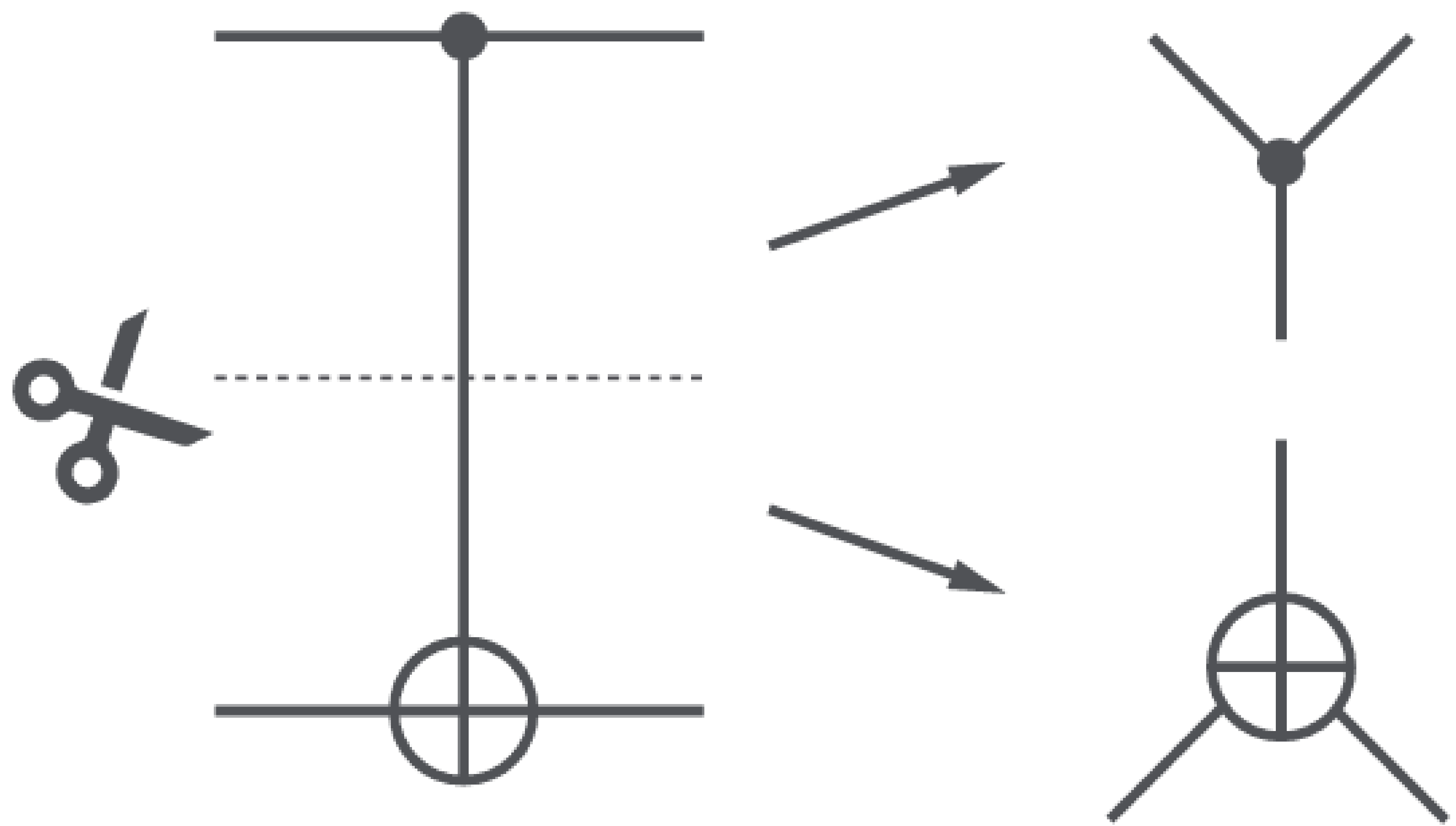
- Unexplored direction: database technologies

Databases to the rescue

We envision a classical-quantum simulation system (CQSS) with the following capabilities:

(i) automatically providing the most efficient simulation of the input circuit by selecting optimal data structures and operations based on available resources and circuit properties;
(ii) operating inherently out-of-core to support the simulation of large circuits that exceed main memory capacity;
(iii) ensuring consistency to prevent data corruption and enabling recovery in the event of large-scale simulation crashes; and
(iv) improving the entire simulation workflow, including parameter tuning, data collection, and querying, exploration, and visualization.

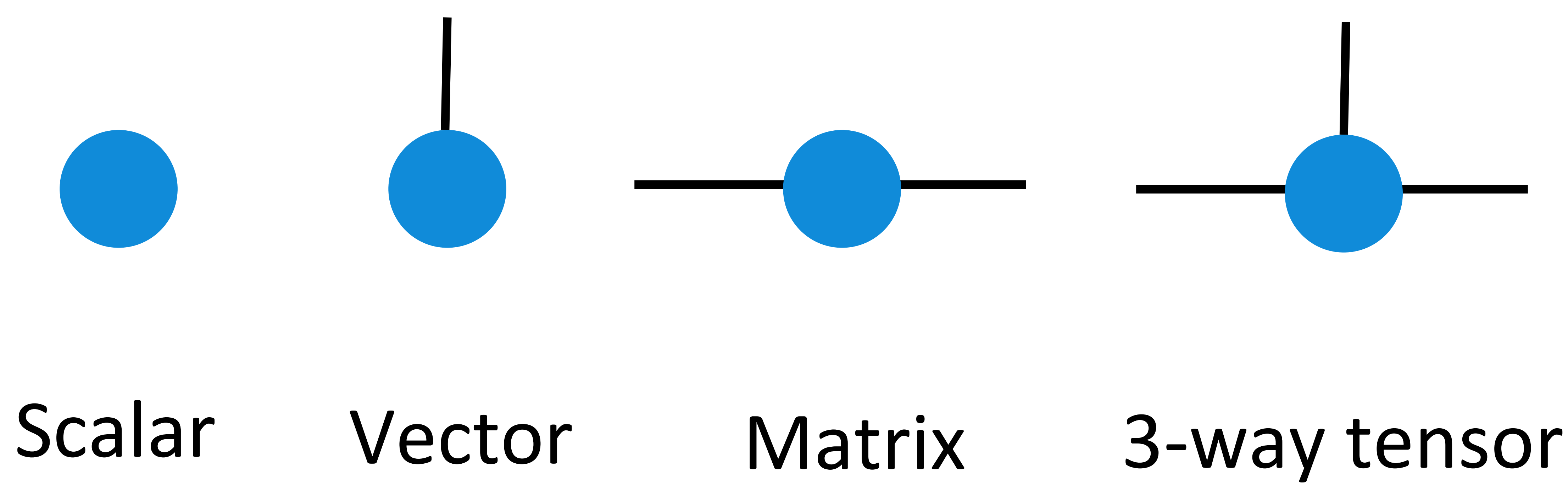
Qubit states & gates represented as tensor networks



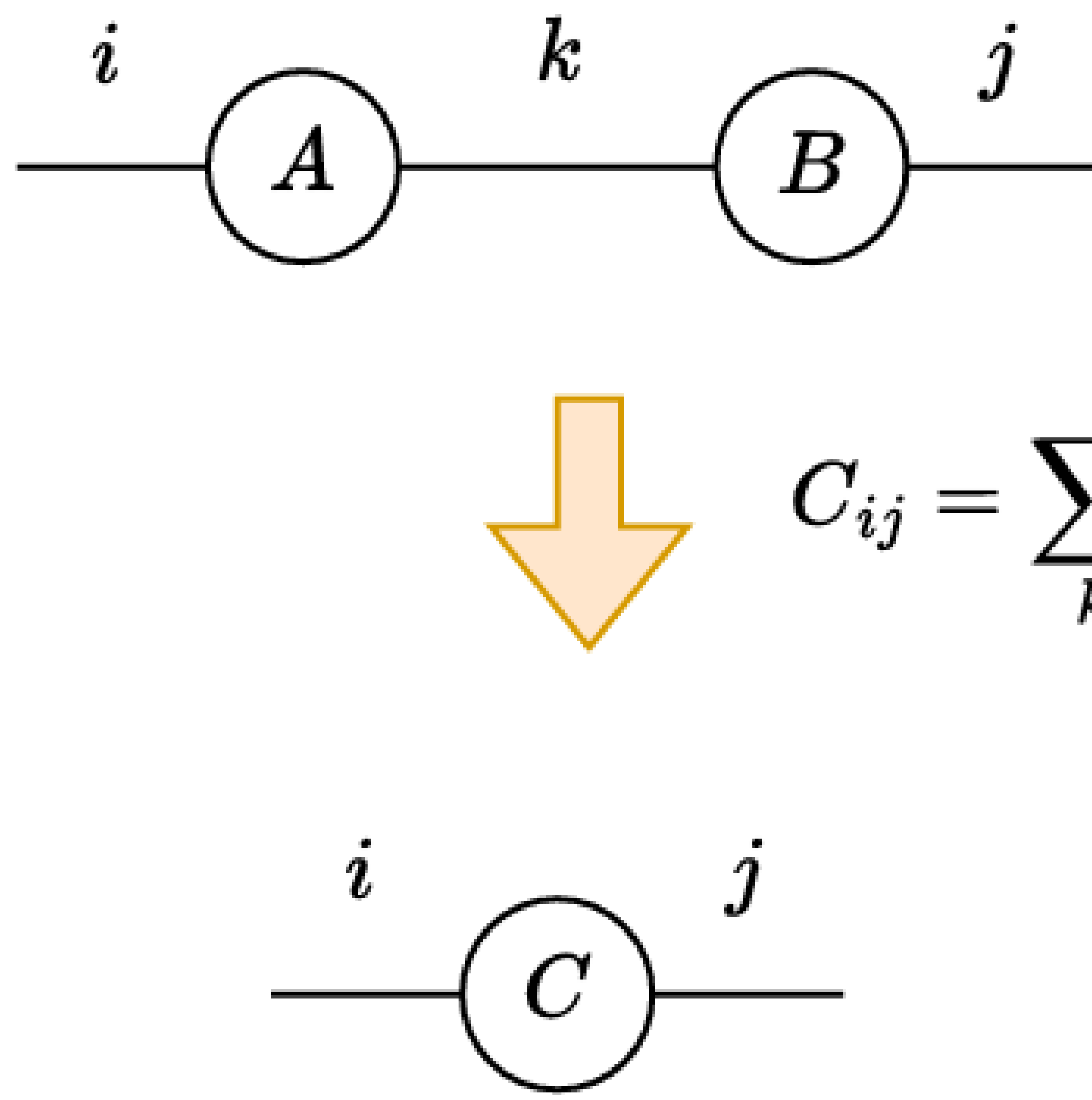
Tensor

- Multidimensional array

Tensor network diagram



Tensor contraction



$$C_{ij} = \sum_k A_{ik} B_{kj}$$

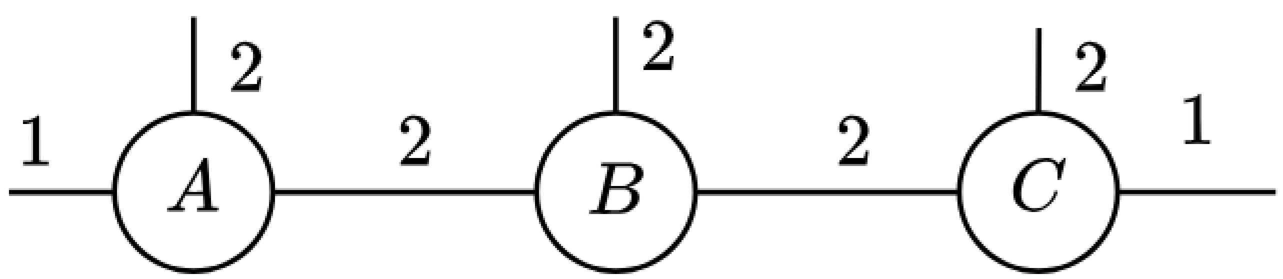
Quantum state as tensors

- 3-qubit GHZ state

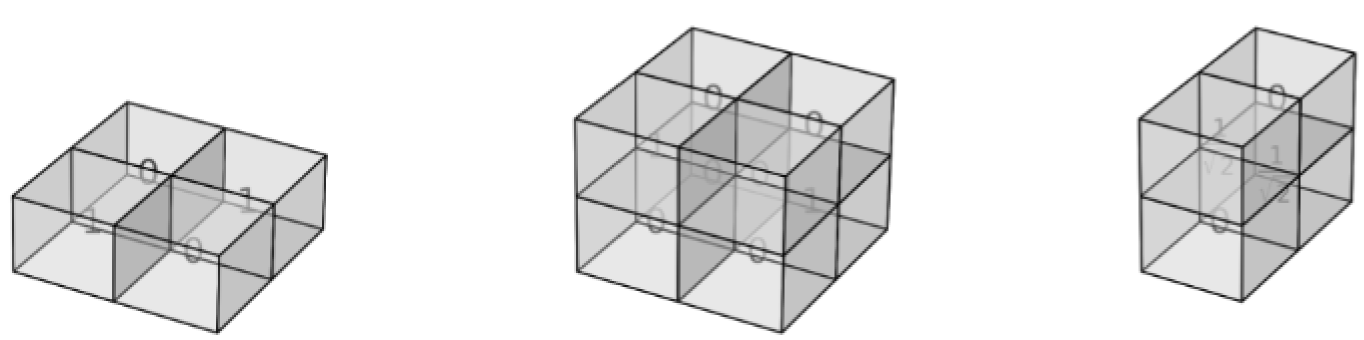
$$\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$$

$$2^3 \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right.$$

First qubit Second qubit Third qubit



Tensor A (1x2x2) Tensor B (2x2x2) Tensor C (2x2x1)



Matrix product state (MPS)

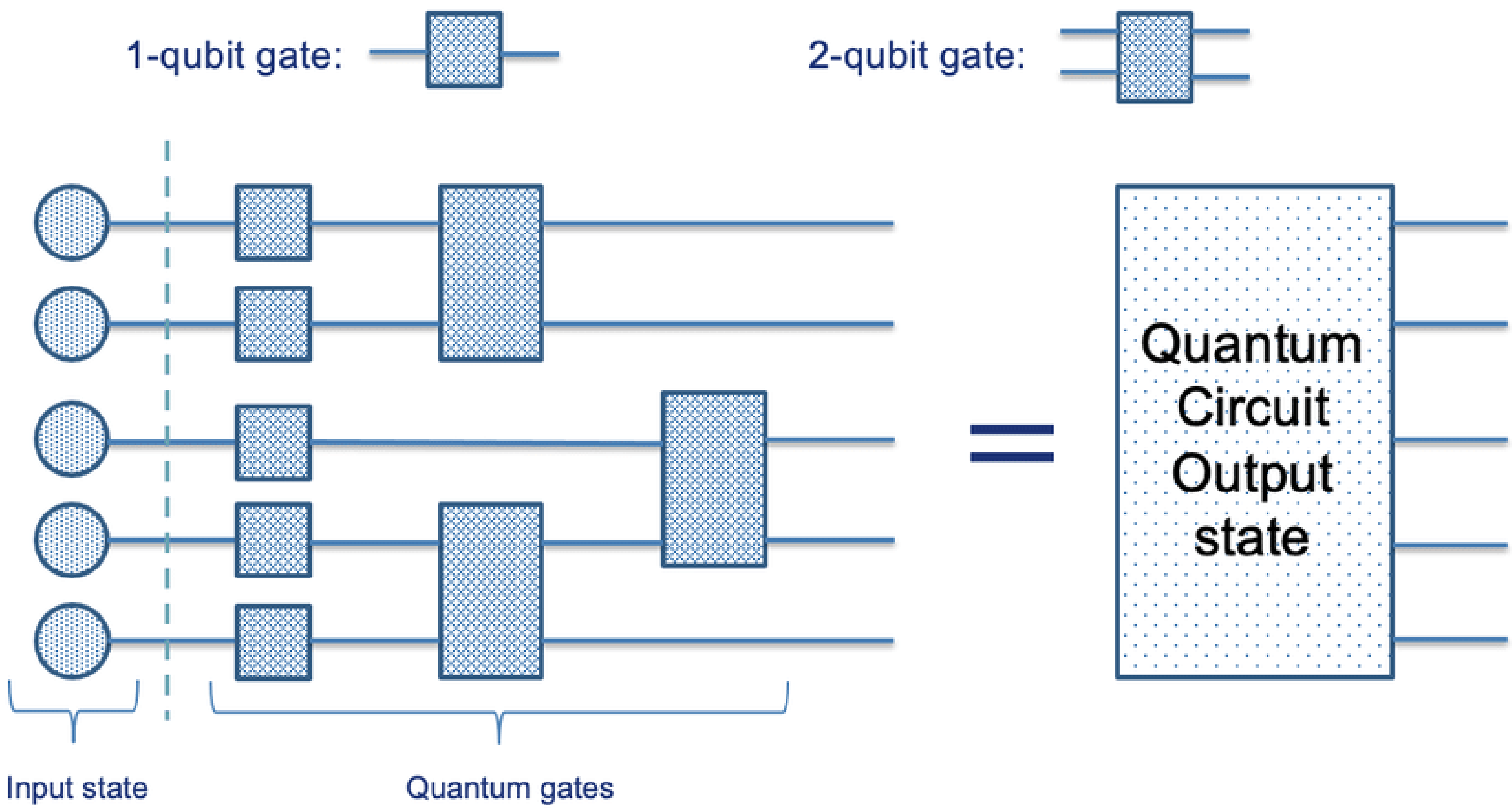
$$A^0 = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad A^1 = \begin{bmatrix} 0 & 1 \end{bmatrix}$$

$$B^0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad B^1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$C^0 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \quad C^1 = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Gates as tensors

Matrix product operator (MPO)



Guo, Yuchen, and Shuo Yang. "Quantum error mitigation via matrix product operators." *PRX Quantum* 3.4 (2022): 040313.

Noise as MPO

- Consider noise modeled as gates (one example)

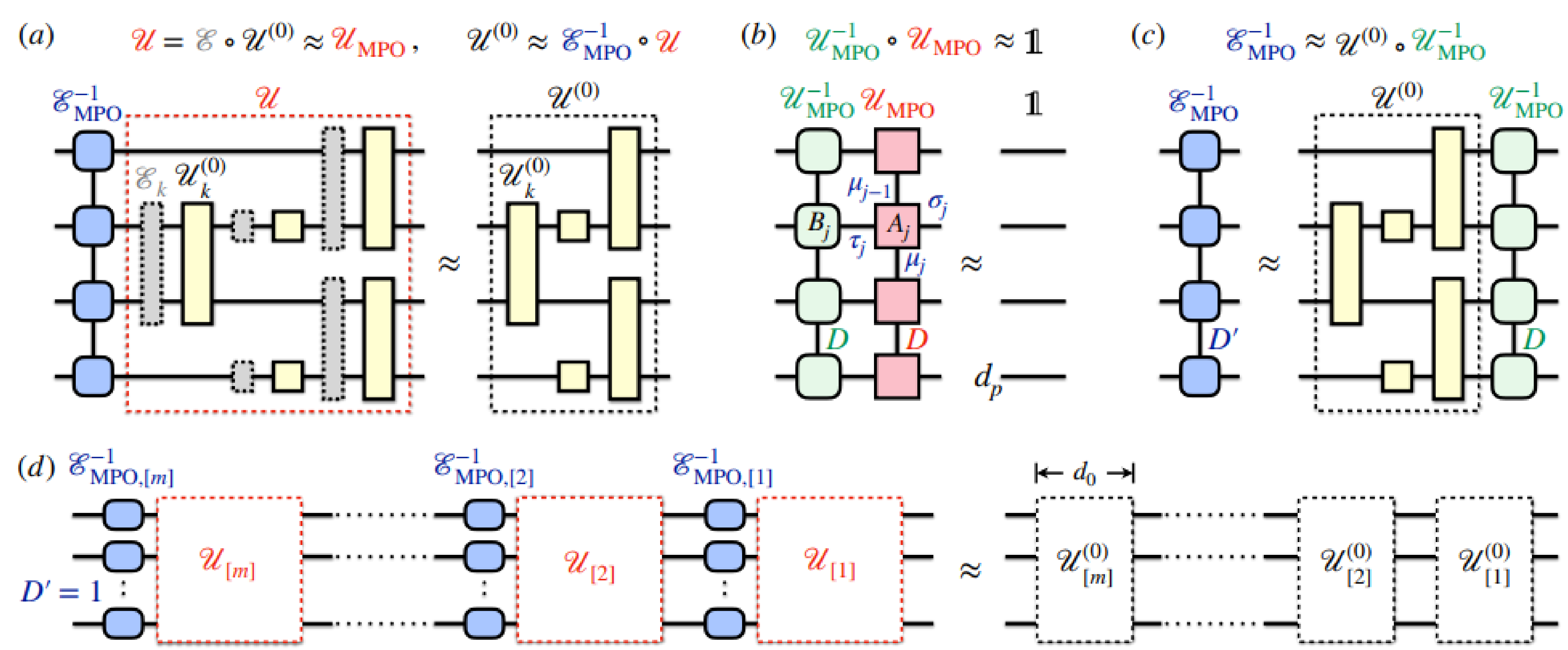
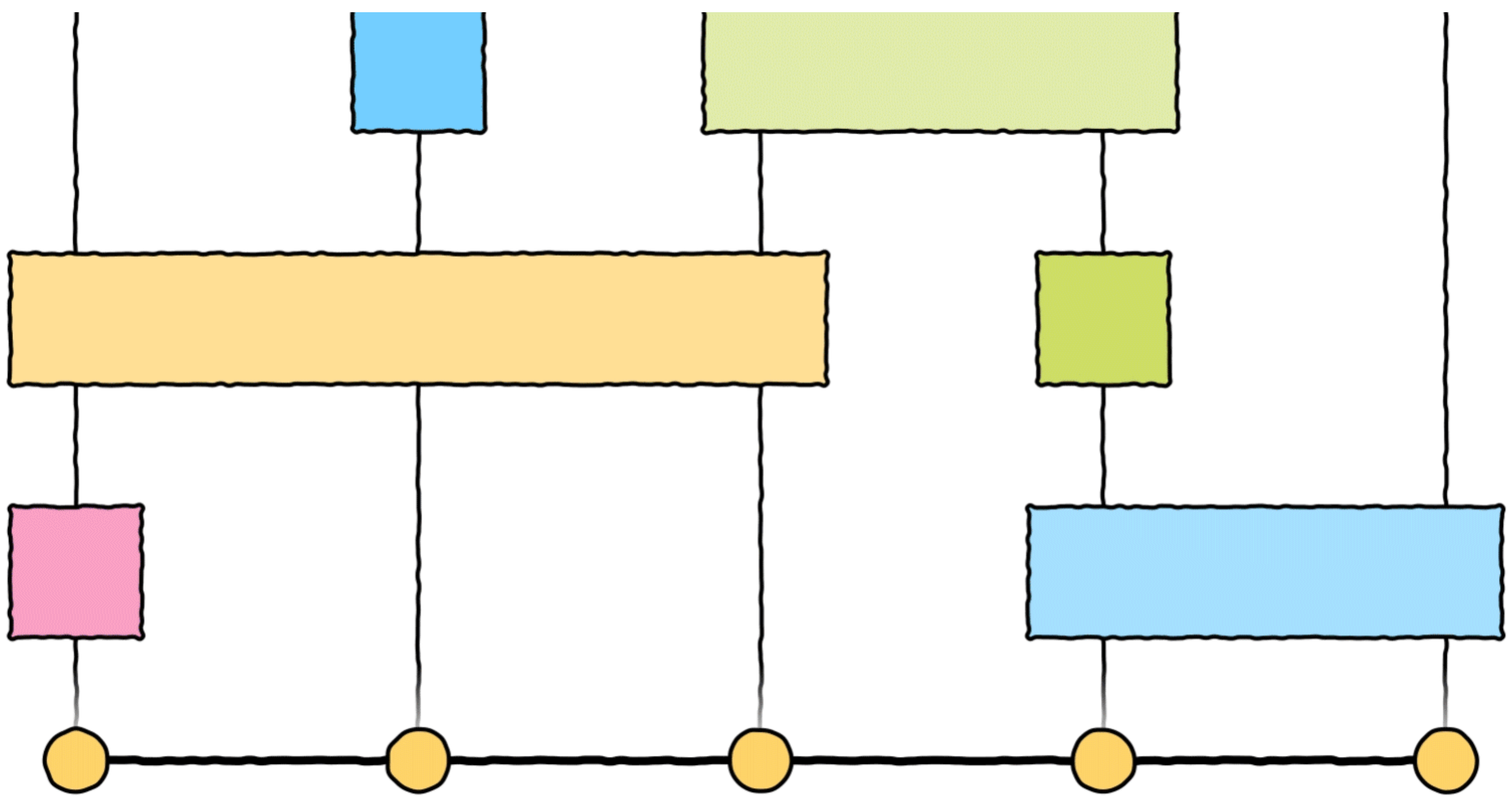


Figure 1. (Color online) (a) The schematic diagram of our QEM method based on MPO. We first use an MPO to represent the noisy quantum circuit $\mathcal{U}_{\mathrm{MPO}}$. Then we calculate the inverse noise channel $\mathcal{E}^{-1}_{\mathrm{MPO}}$, which is applied after $\mathcal{U}$ to compensate for the error and to restore the ideal circuit $\mathcal{U}^{(0)}$. (b) Our variational MPO-inverse method. We calculate the inverse of an MPO-represented quantum channel $\mathcal{U}_{\mathrm{MPO}}$, which is parameterized as an MPO $\mathcal{U}^{-1}_{\mathrm{MPO}}$ with the same bond dimension D. (c) Calculation of the inverse noise channel $\mathcal{E}^{-1}_{\mathrm{MPO}}$ via MPO contraction and truncation methods, whose bond dimension is D'. (d) A deep circuit is divided into m parts, each with d_0 layers. One may apply our QEM method on each part, where $\mathcal{E}^{-1}_{\mathrm{MPO},[k]}$ is truncated to $D' = 1$ and simulated by single-qubit gates.

Simulating quantum circuits with tensor network

- The state after executing the circuits is obtained by **contracting** all the tensors



Efficient tensor computation: database to the rescue

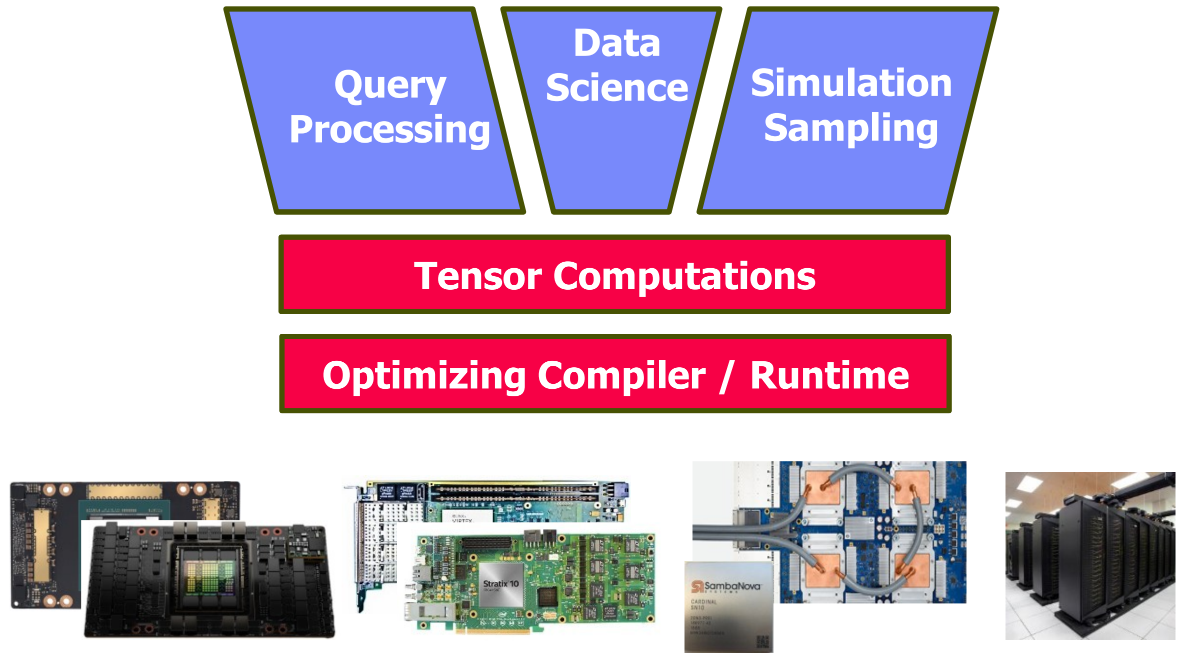
Push the simulation workload to DBMSs

```
CREATE TABLE States(qubits text, areal float, aimg float);
```

```
WITH (SELECT A.quibits as Aqubits, B.qubits as Bqubits,
 A.areal as Aareal, A.aimg as Aaimg,
 B.areal as Bareal, B.aimg as Baimg
 FROM States A, States B WHERE
 SUBSTRING(A.qubits,1,k-1)=SUBSTRING(A.qubits,1,k-1)
 AND SUBSTRING(A.qubits,k+1,n-k)=SUBSTRING(A.qubits,k+1,n-k)
 AND A.qubits < B.qubits) AS J
(SELECT Aqubits, q11*Aareal+q12*Bareal, q11*Aaimg+q12*Baimg)
UNION ALL (SELECT Bqubits, q21*Aareal+q22*Bareal,
 q21*Aaimg+q22*Baimg)
```

Immanuel Trummer. 2024. Towards Out-of-Core Simulators for Quantum Computing. In Proceedings of the 1st Workshop on Quantum Computing and Quantum-Inspired Technology for Data-Intensive Systems and Applications (Q Data ’24). https://doi.org/10.1145/3665225.3665441

Compilation and Runtime



Matthias Boehm, Matteo Interlandi, and Chris Jermaine. 2023. Optimizing Tensor Computations: From Applications to Compilation and Runtime Techniques. In Companion of the 2023 International Conference on Management of Data. 53–59.

Many possibilities

Can DB technologies boost the development of quantum computing?

- In-DB Tensor Computation → Classical simulation
- Graph analytics → Error correction
- Distributed databases → Quantum network
- Scientific data management → Quantum experiments

Q&A

Can DB technologies boost the development of quantum computing?

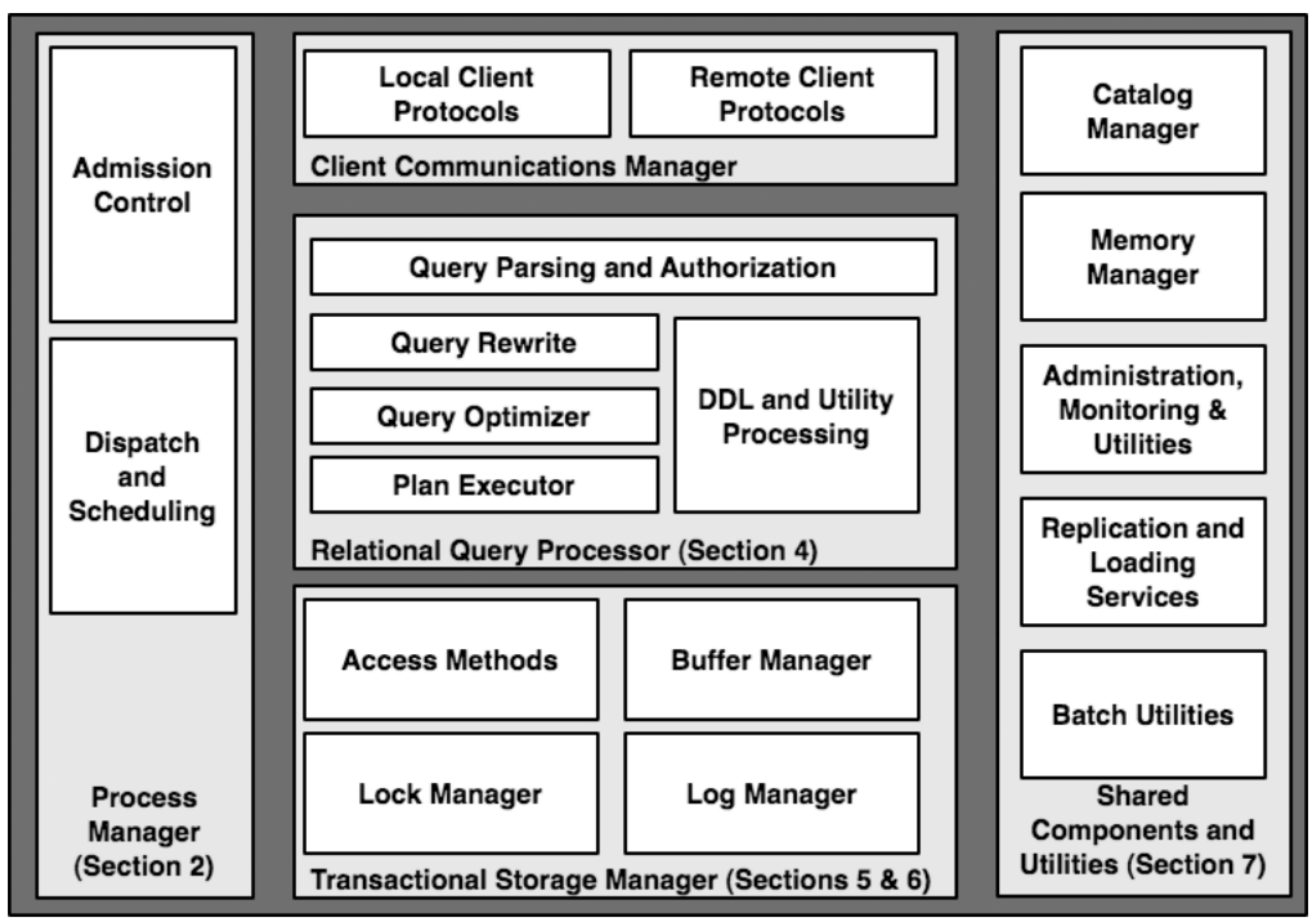
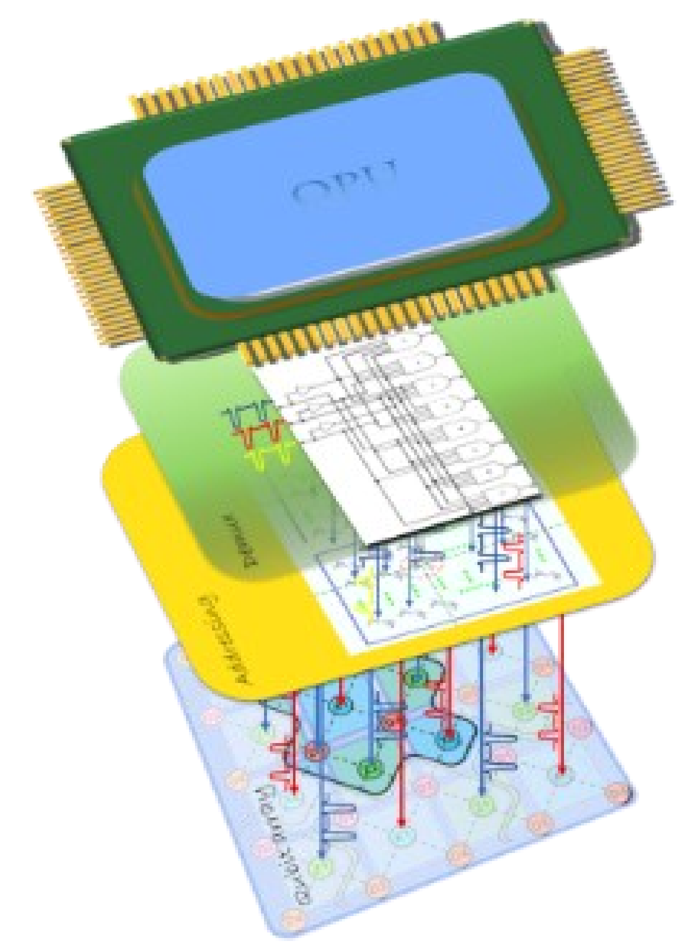


Fig. 1.1 Main components of a DBMS.



Quantum computing