Quantum Data Management

in the NISQ Era and Beyond

Rihan Hai

Our Research

InfiniData Team

Data systems for AI

Effectiveness
Amalur (ICDE'23, TKDE'24)
AutoFeat (ICDE'24)
TransferGraph (ICDE'24)

Privacy
SiloFuse
(ICDE'24)

Efficiency
Amalur (ICDE'23, TKDE'24)
TranSQL (SIGMOD'26)
Wavestitch (SIGMOD'26)

Energy Efficiency
Work in progress

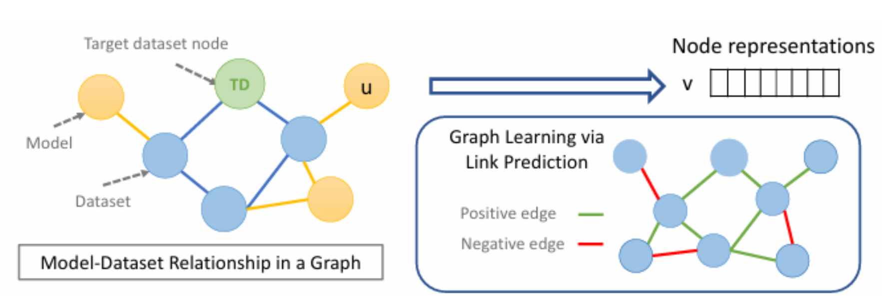
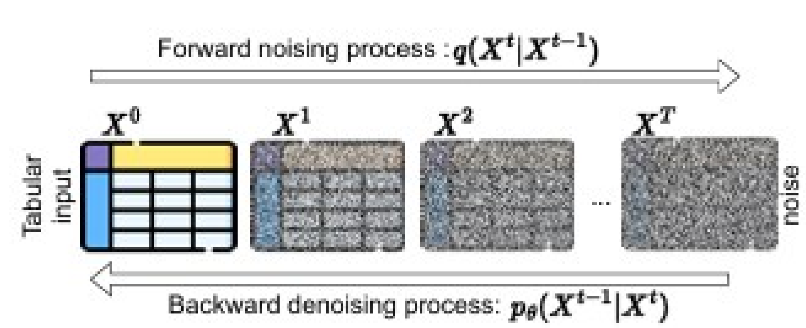
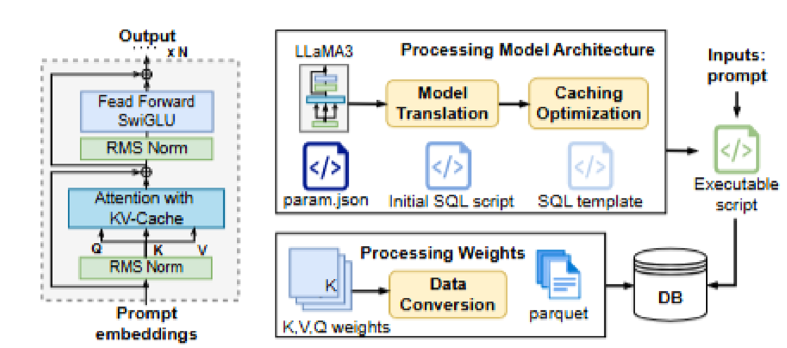


Figure 3: Link prediction in the context of model selection

GNNs

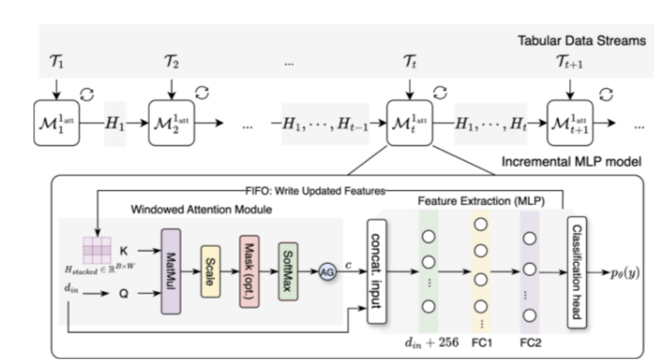


Diffusion models



(a) Transformer unit (b) TranSQL architecture with three major components

LLMs/Diffusion models



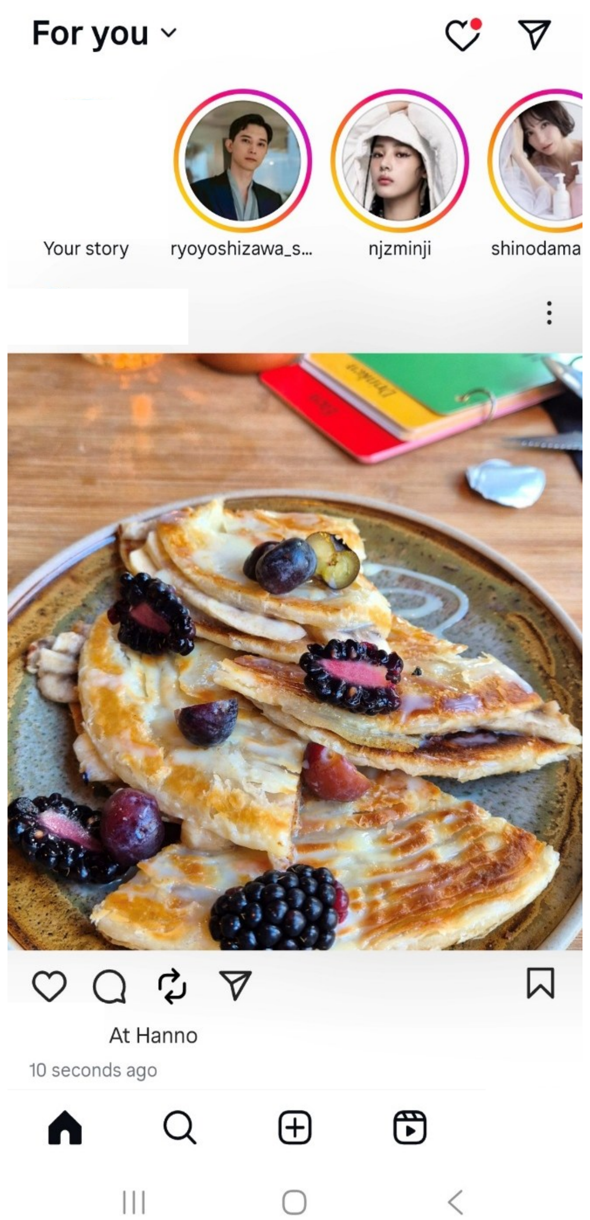
MLP

Learning objective

After this talk, you should have an answer to：

1. What is data?
2. What quantum data systems do you want to build?

Classical data



Full Name	Phone	Email
Anton Johnson	(+31) 123-4567	anton.johnson@gmail.com
Brian Smith	(+42) 234-5678	bsmith@yahoo.com
Cynthia Lee	(+51) 345-6789	cynthia.lee@outlook.com
Daniel Kim	(+64) 456-7890	dkim@icloud.com
Eva Martínez	(+34) 567-8901	eva.martinez@aol.com

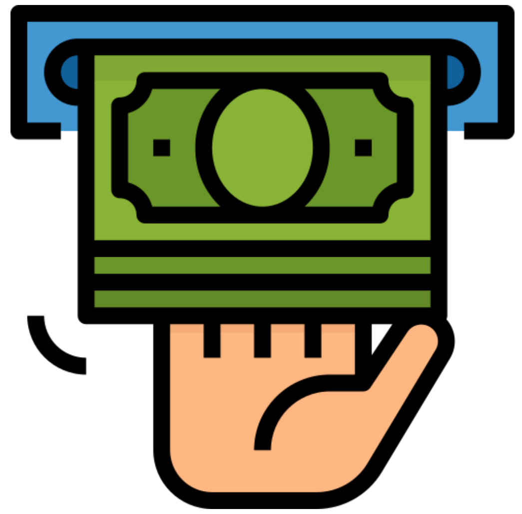
Classical data

- Information that is collected, processed, and stored with traditional computing methods
- Often stored and queried using database management systems (DBMS) such as relational databases, document stores, graph databases, and vector databases

Database management system

Datenbank (DE)

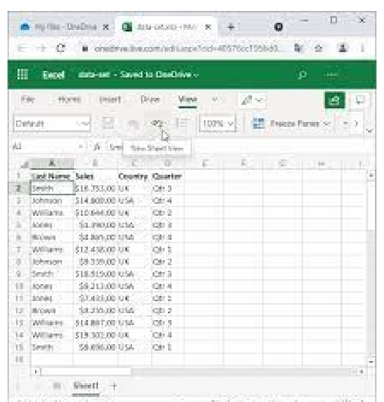
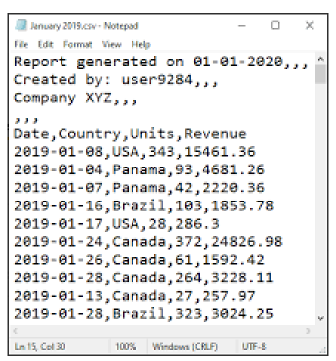
- A database management system, or DBMS, is software designed to assist in maintaining and utilizing large collections of data.

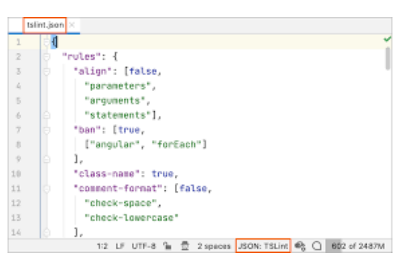


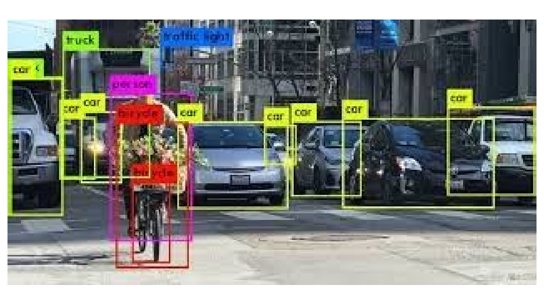
Database management system

- Systems to store, manage and query the data

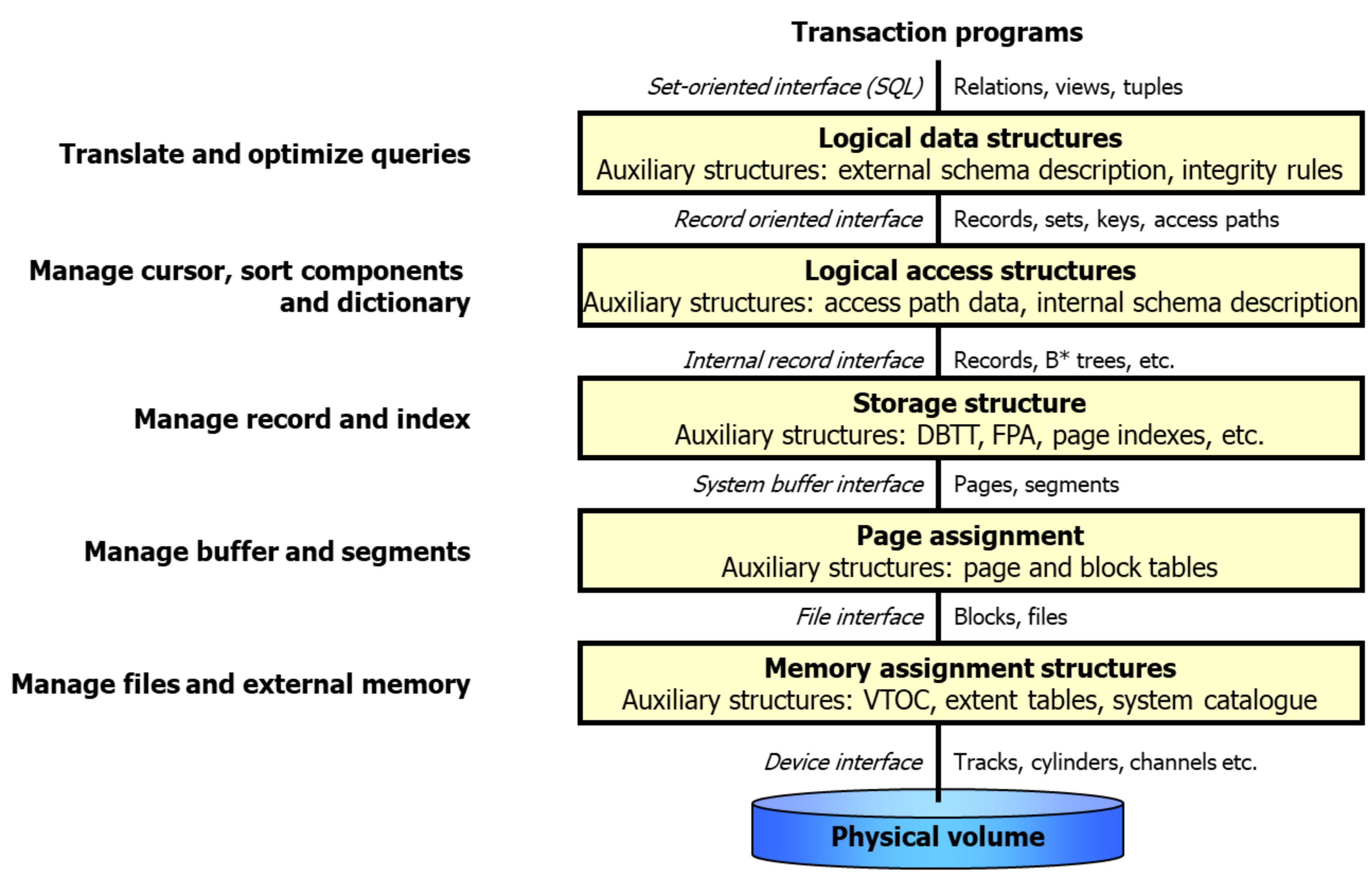
Data





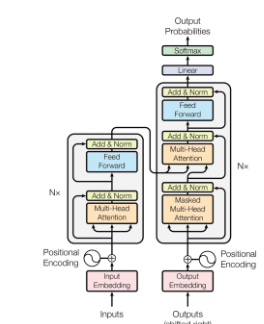
Database management systems

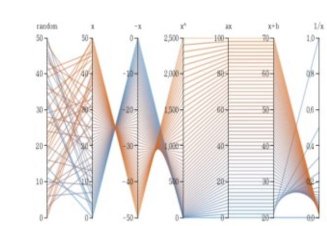


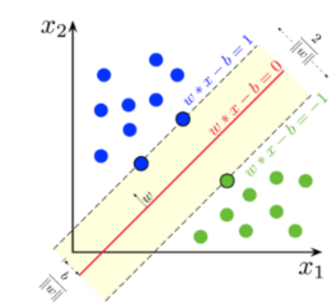
[Härder & Rahm, 2001]

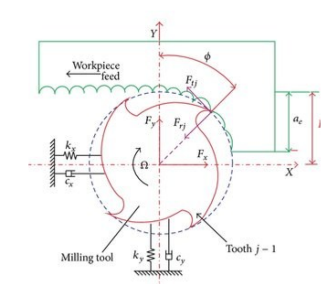
Applications

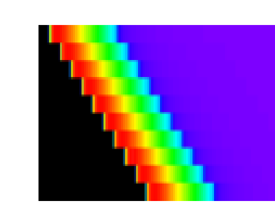

Business report


AI model training


Visual Analytics


Data Analytics

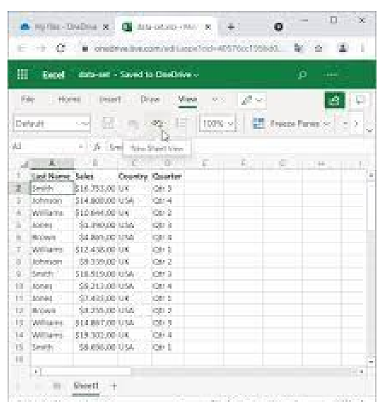

Modeling

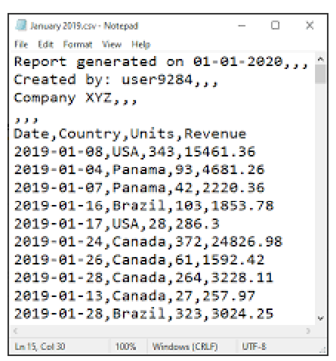

Simulation

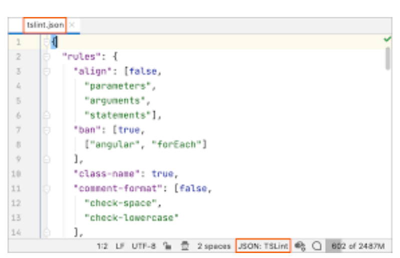
Database management system

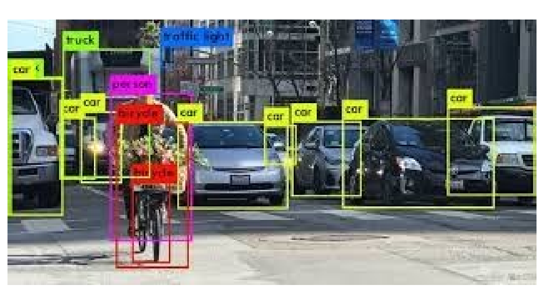
- Systems to store, manage and query the data

Data









Database management systems

ORACLE DATABASE

Microsoft SQL Server

PostgreSQL

IBM DB2

MySQL

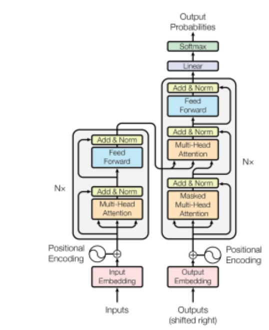
mongoDB

neo4j

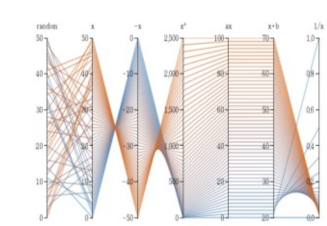
Pinecone

Applications

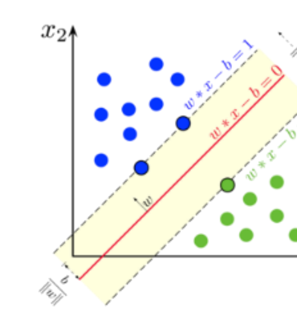
Business report



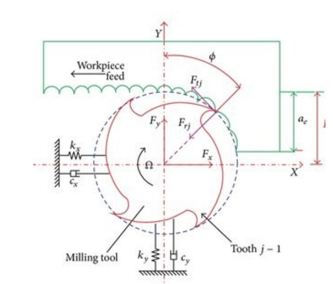
AI model training



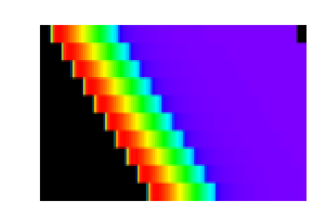
Visual Analytics



Data Analytics



Modeling

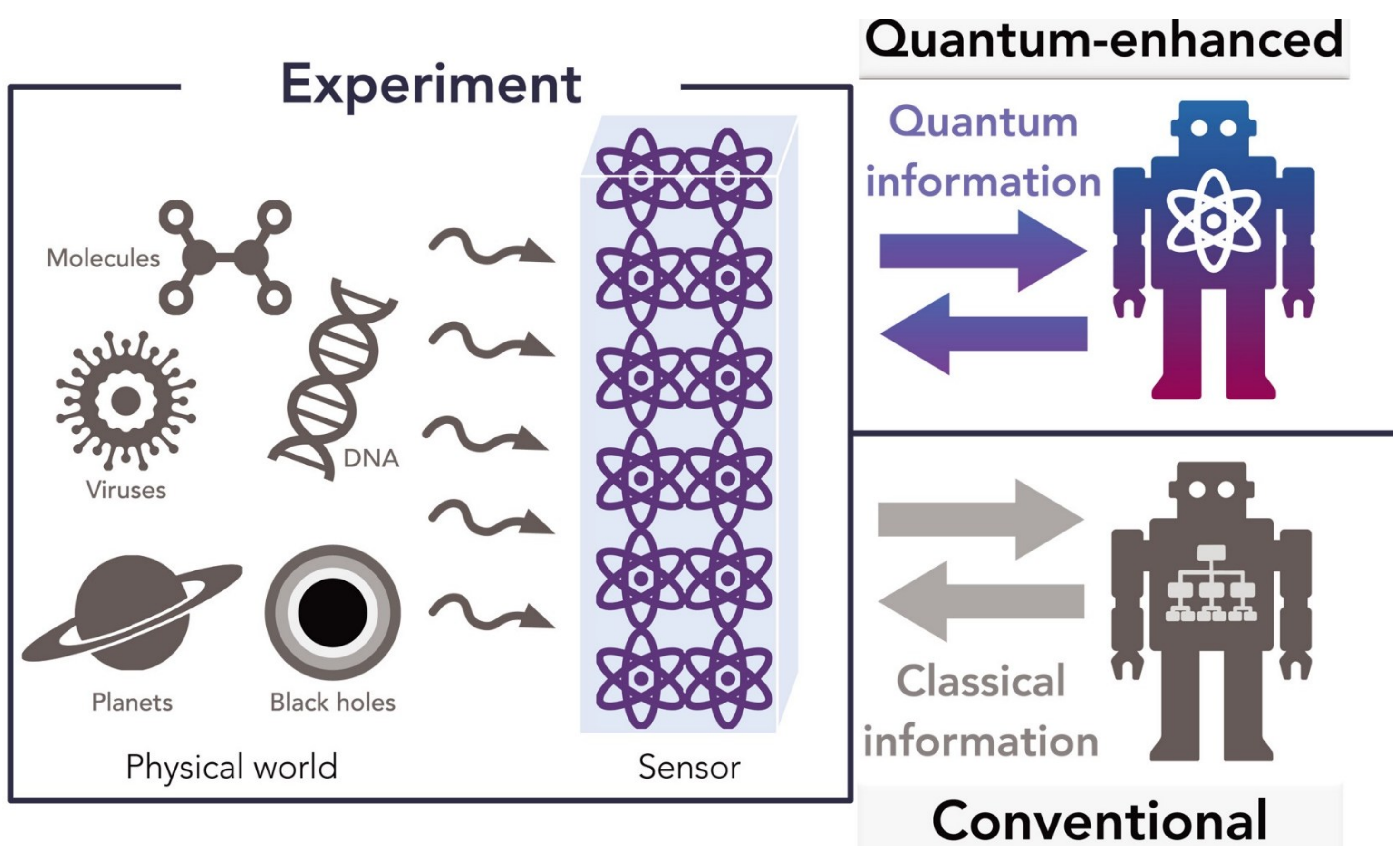


Simulation

From classical data to quantum data

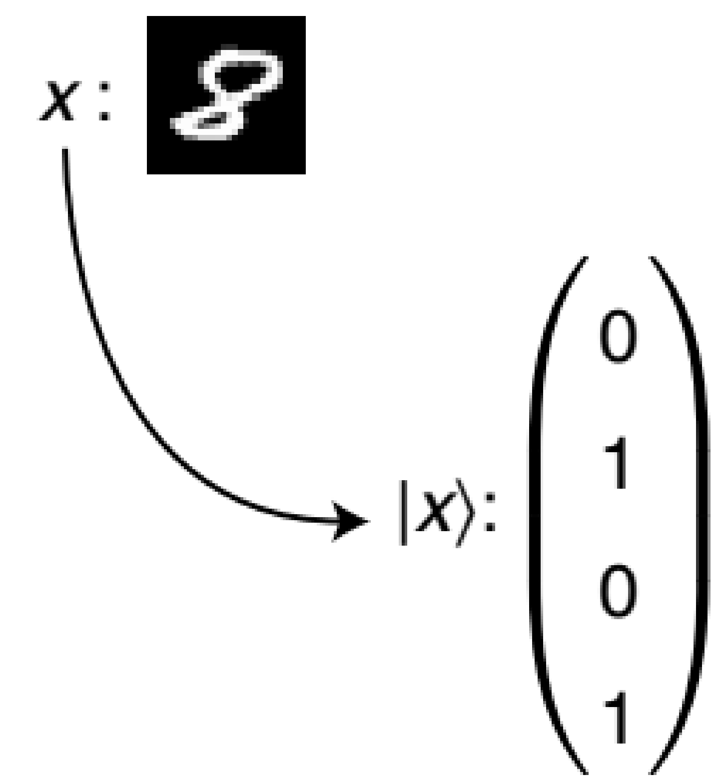
Quantum data

1) Quantum-native



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

2) Classical data encoded as states



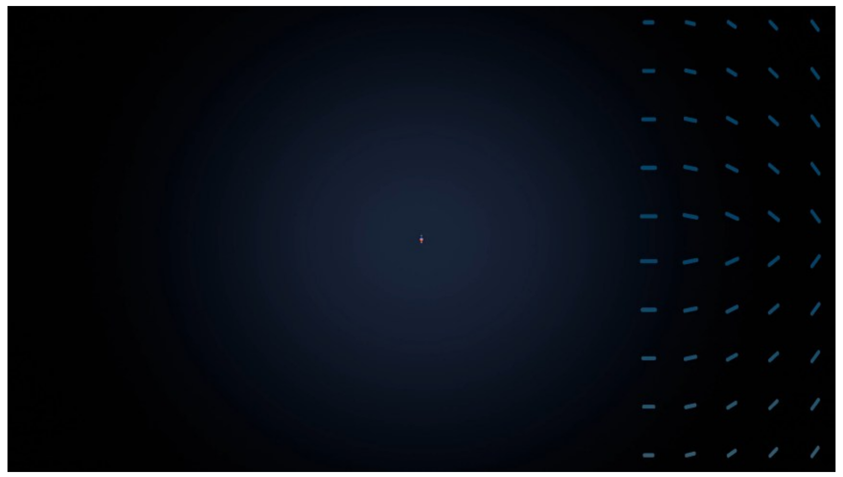
Cerezo, M., Verdon, G., Huang, HY. et al. Challenges and opportunities in quantum machine learning. Nat Comput Sci 2, 567–576 (2022). https://doi.org/10.1038/s43588-022-00311-3

Understanding quantum data

Probabilistic Nature

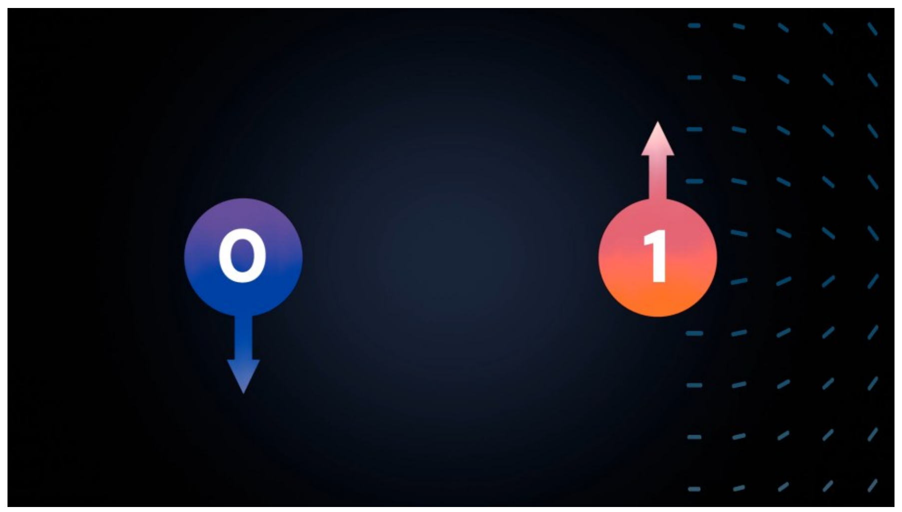
$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

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



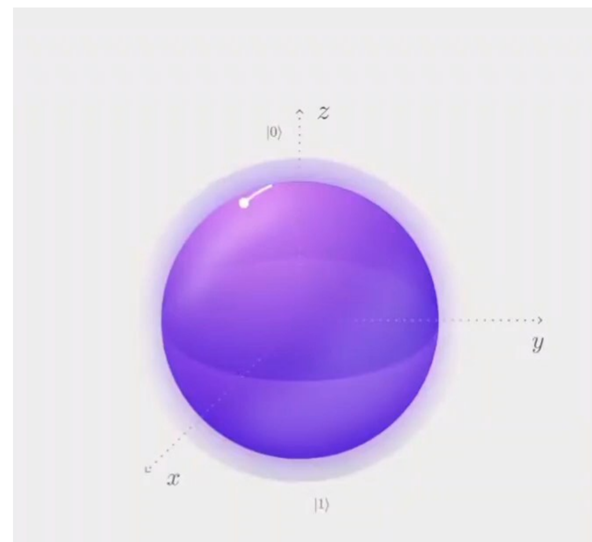
Entanglement

Multiple qubits can be correlated such that measuring one immediately affects others.



Fragility

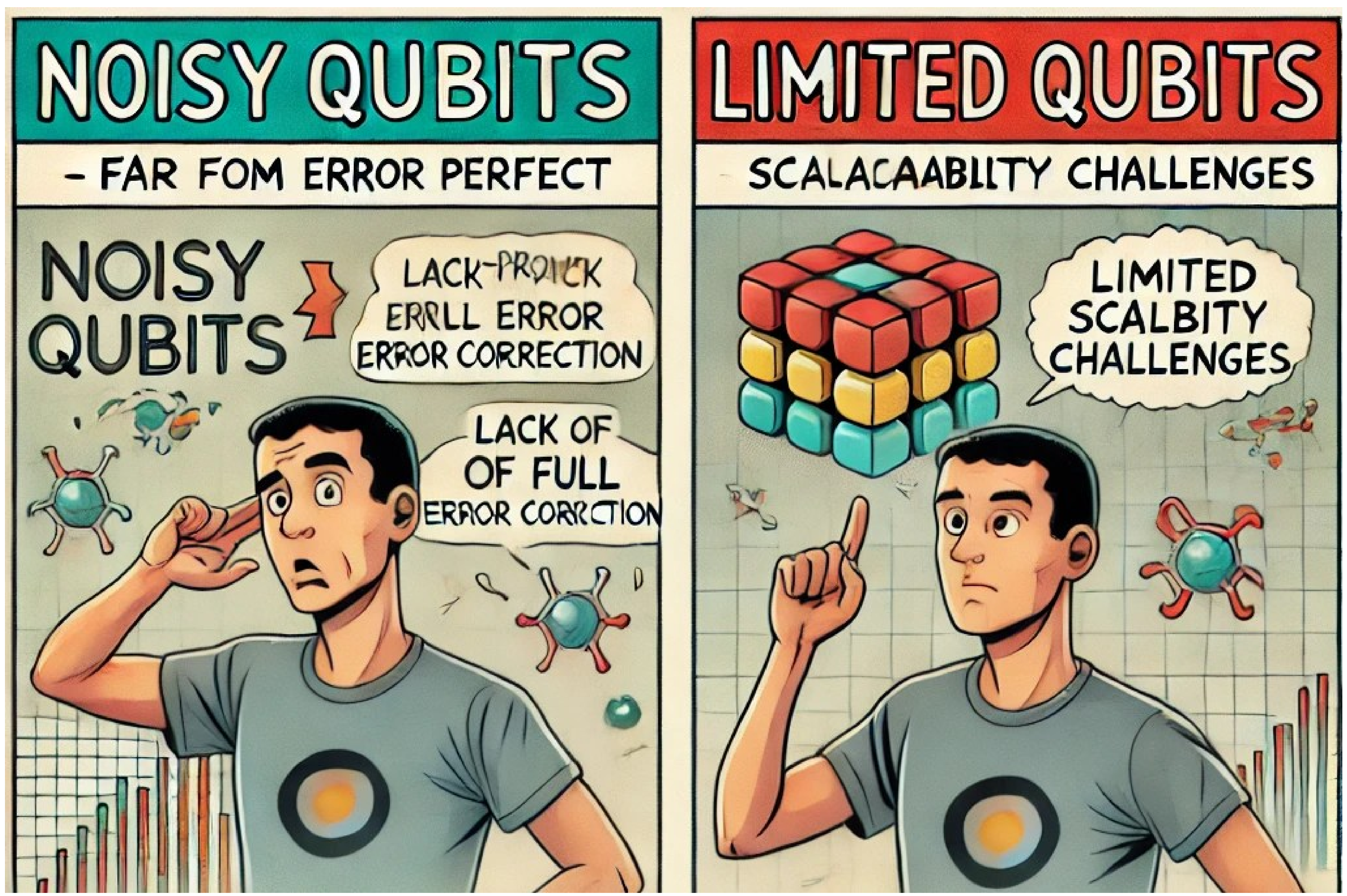
Quantum noise results from unwanted coupling with the environment.



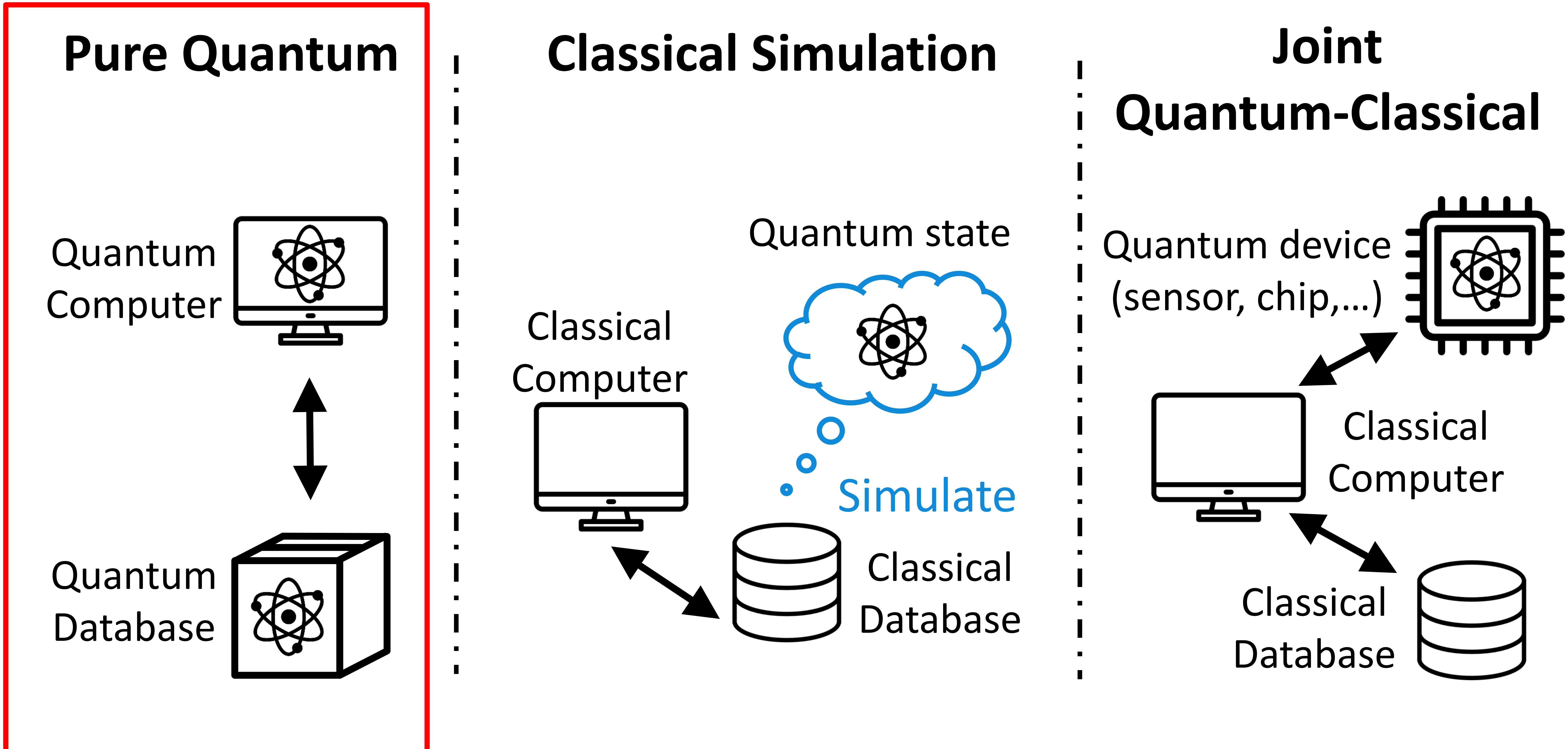
Noisy Intermediate-Scale Quantum (NISQ)

Quantum Computing in the NISQ era and beyond

John Preskill

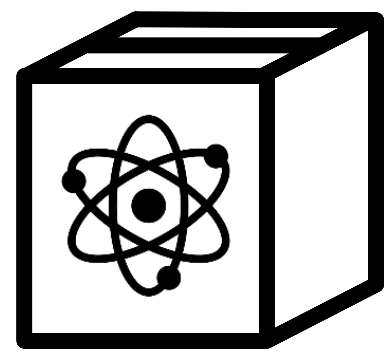


Landscape: data management for quantum computing



Rihan Hai, Shih-Han Hung, Tim Coopmans, Tim Littau, and Floris Geerts. Quantum data management in the NISQ era. PVLDB, 18(6):1720–1729, 2025

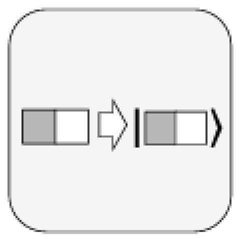
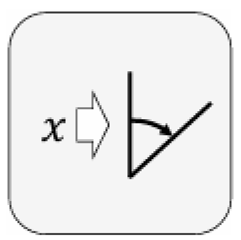
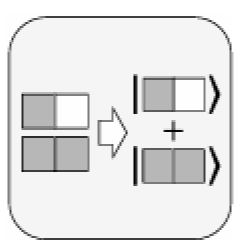
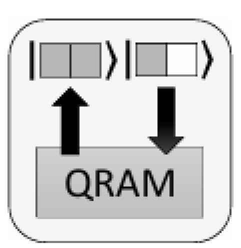
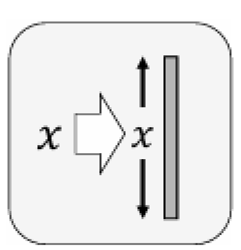
Quantum Database



How to design and implement a quantum database system?

What can a quantum database do?

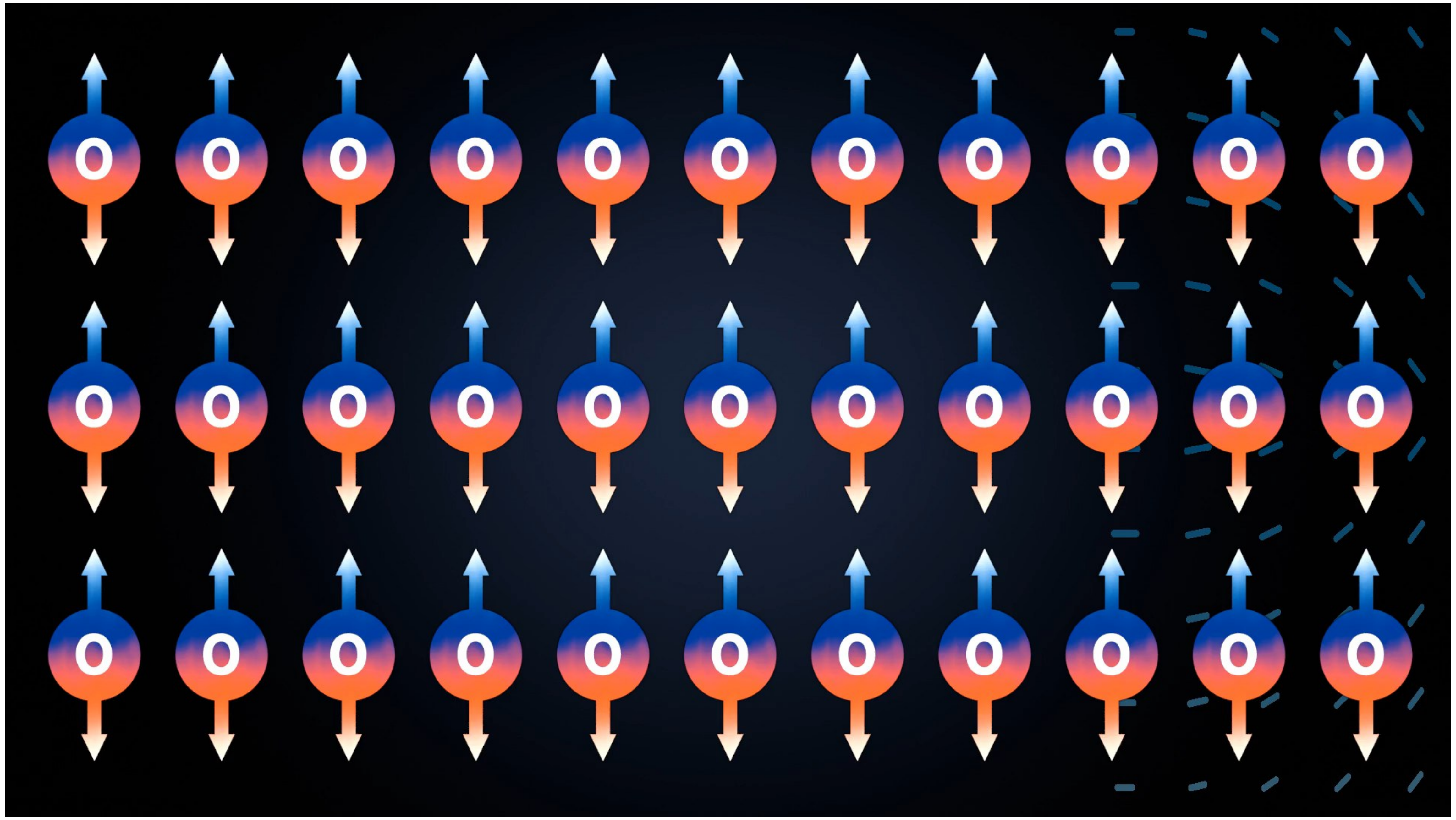
From classical data to quantum data

Encoding Pattern		Encoding	Req. Qubits
	Basis Encoding [1]	$x_i \approx \sum_{i=-k}^{m} b_i 2^i \mapsto \lvert b_m \dots b_{-k} \rangle$	$l = k+m$ per data-point
	**Angle Encoding**	$x_i \mapsto cos(x_i) \lvert 0 \rangle + sin(x_i) \lvert 1 \rangle$	1 per data-point
	QuAM Encoding [1]	$X \mapsto \sum_{i=0}^{n-1} \frac{1}{\sqrt{n}} \lvert x_i \rangle$	l
	**QRAM Encoding**	$X \mapsto \sum_{n=0}^{n-1} \frac{1}{\sqrt{n}} \lvert i \rangle \lvert x_i \rangle$	$\lceil \log n \rceil + l$
	Amplitude Encoding [1]	$X \mapsto \sum_{i=0}^{n-1} x_i \lvert i \rangle$	$\lceil \log n \rceil$

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), 2021, pp. 95-101, doi: 10.1109/ICSA-C52384.2021.00025.

From quantum data to classical data

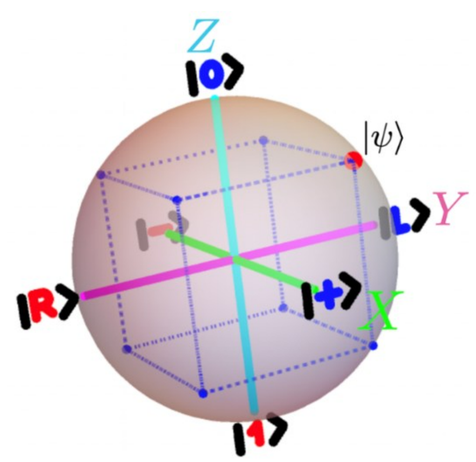
- Qubit fate (0 or 1) determined upon measurement



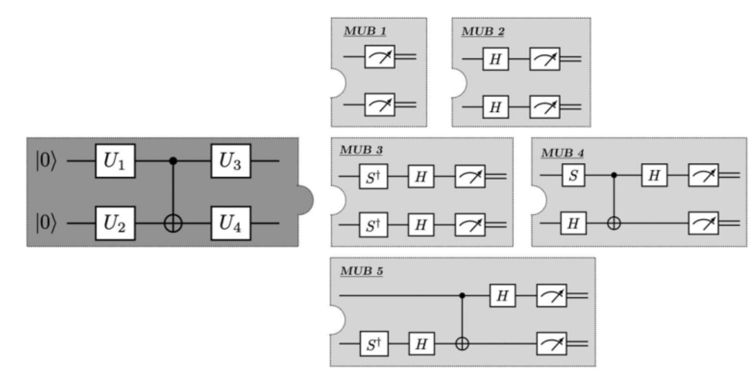
Private Quantum Database

Core quantum technology

Full Name	Phone	Email
Anton Johnson	(+31) 123-4567	anton.johnson@gmail.com
Brian Smith	(+42) 234-5678	bsmith@yahoo.com
Cynthia Lee	(+51) 345-6789	cynthia.lee@outlook.com
Daniel Kim	(+64) 456-7890	dkim@icloud.com
Eva Martínez	(+34) 567-8901	eva.martinez@aol.com



Quantum Random Access Coding



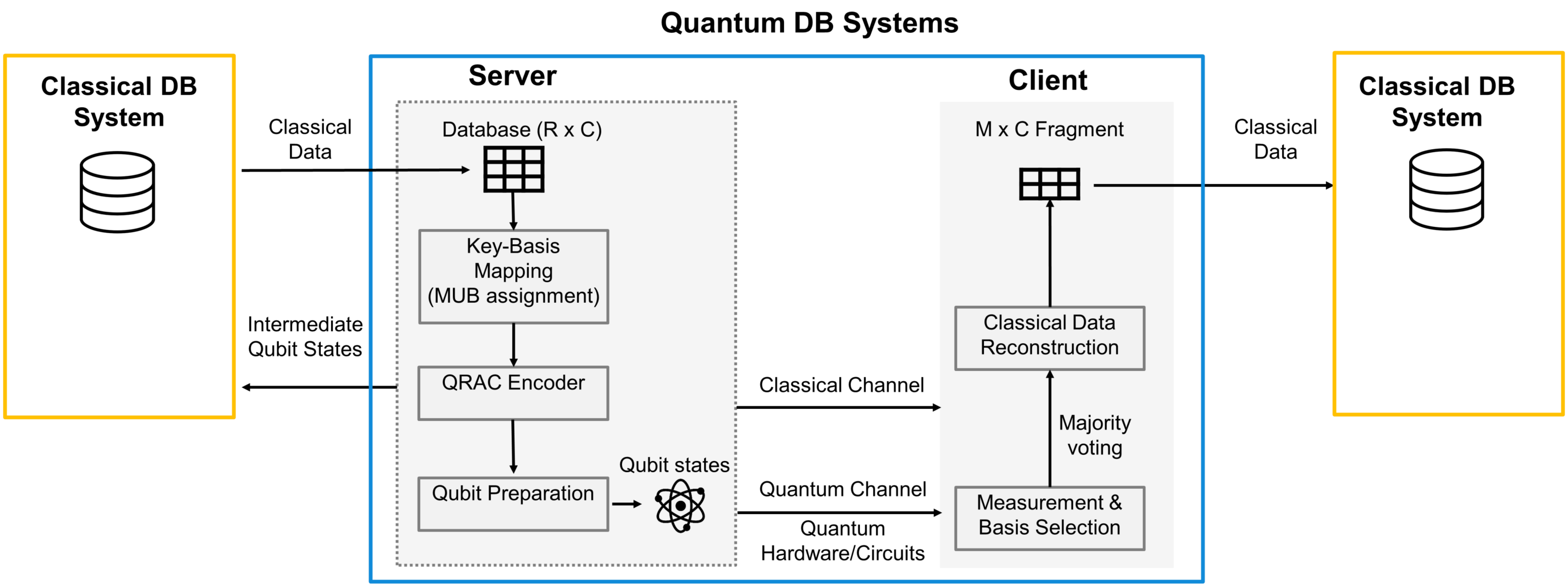
Mutually unbiased bases

Full Name	Phone	Email
Anton Johnson	(+31) 123-4567	anton.johnson@gmail.com

[Gatti, Giancarlo, Floris Geerts, and Rihan Hai. Private Quantum Database. https://arxiv.org/abs/2508.19055.](https://arxiv.org/abs/2508.19055)

Private Quantum Database

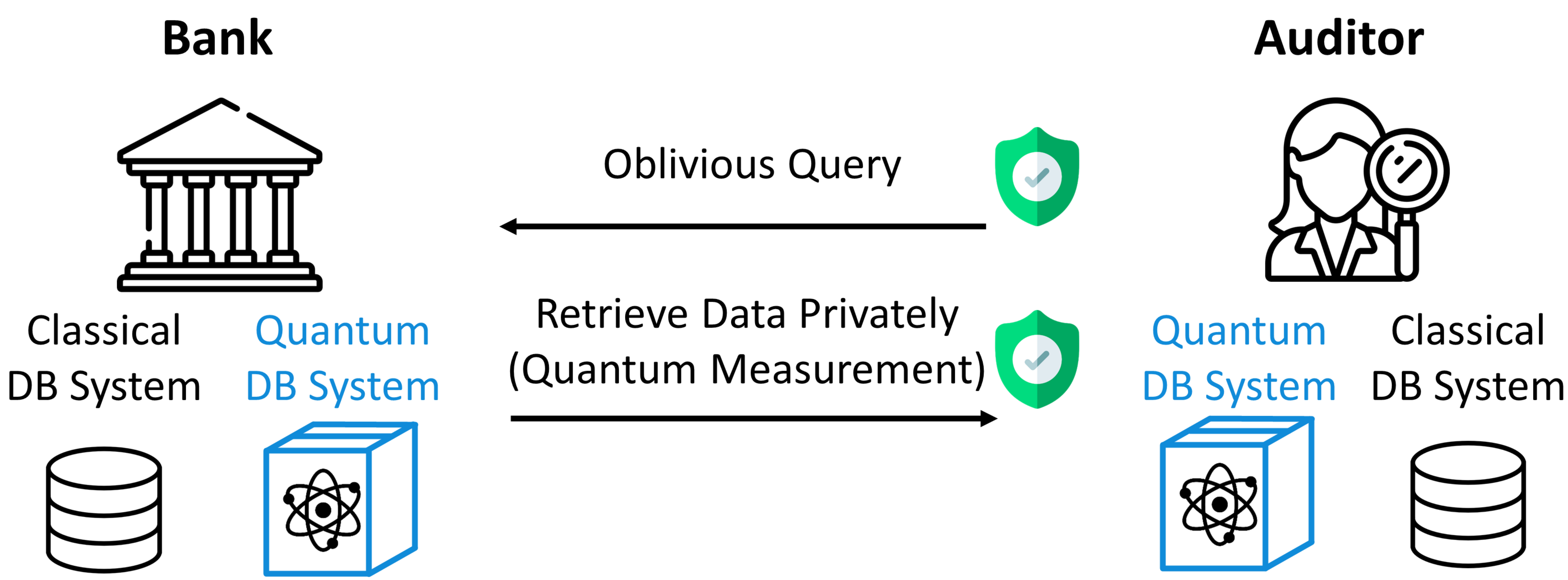
Hybrid architecture



[Gatti, Giancarlo, Floris Geerts, and Rihan Hai. Private Quantum Database. https://arxiv.org/abs/2508.19055.](https://arxiv.org/abs/2508.19055)

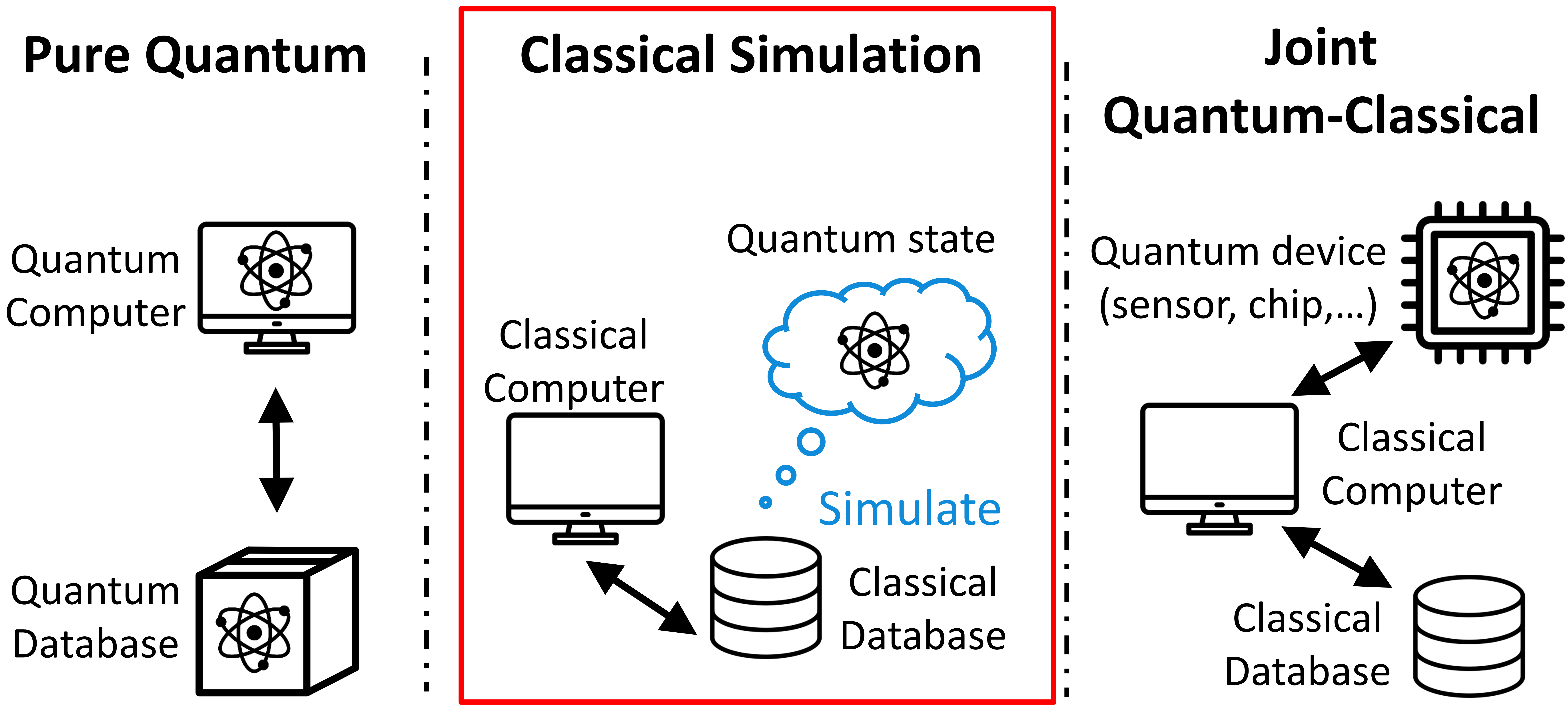
Near-term Application

- User privacy and data privacy



Gatti, Giancarlo, Floris Geerts, and Rihan Hai. Private Quantum Database. https://arxiv.org/abs/2508.19055.

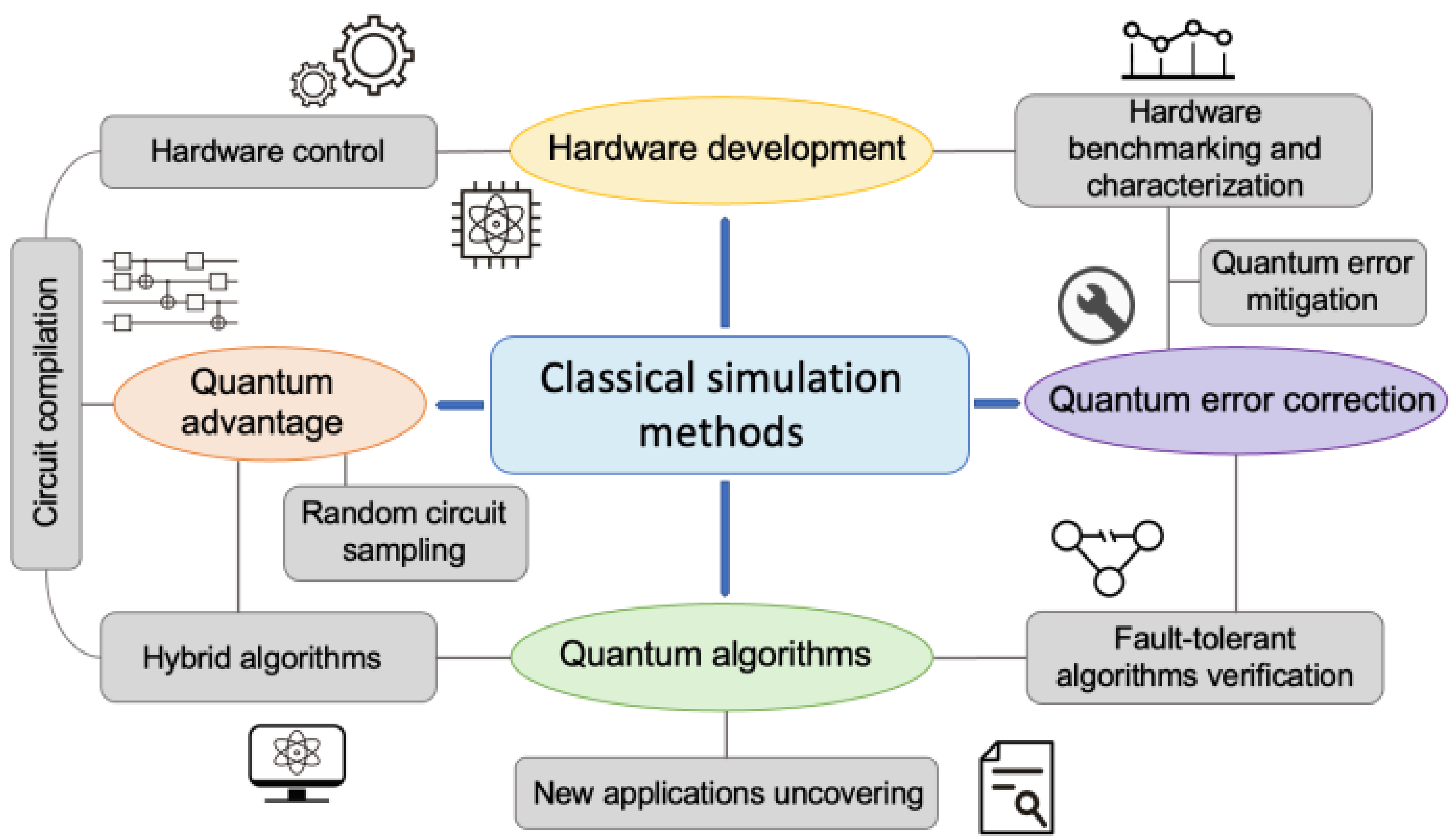
Landscape: data management for quantum computing



Rihan Hai, Shih-Han Hung, Tim Coopmans, Tim Littau, and Floris Geerts. Quantum data management in the NISQ era. PVLDB, 18(6):1720–1729, 2025

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

Simulation problem: scalability

- We can represent an **n**-qubit quantum state as a vector of size $\mathbf{2^n}$

$n = 1$

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\left.|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}\quad\right\} \text{Vector size: 2}$$

Simulation problem: scalability

- We can represent an **n**-qubit quantum state as a vector of size $\mathbf{2^n}$

$n = 2$

$$|00\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad |01\rangle = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad |10\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad |11\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|00\rangle + \alpha_{10}|00\rangle + \alpha_{11}|11\rangle = \left.\begin{bmatrix} \alpha_{00} \\ \alpha_{01} \\ \alpha_{10} \\ \alpha_{11} \end{bmatrix}\right\} \text{Vector size: 4}$$

Simulation problem: scalability

- We can represent an **n**-qubit quantum state as a vector of size $\mathbf{2^n}$

$$|\psi\rangle = \alpha_{0\dots0}|0\dots0\rangle + \alpha_{1\dots1}|1\dots1\rangle = \left.\begin{bmatrix} \alpha_{0\dots0} \\ \alpha_{0\dots1} \\ \\ \cdots \\ \\ \alpha_{1\dots0} \\ \alpha_{1\dots1} \end{bmatrix}\right\} \text{Vector size: } \mathbf{2^n}$$

Simulation problem: scalability

- How much memory in GB do we need?

Number of qubits

Double-precision (16 bytes) for complex numbers

$$2^{n+4-30}$$

1 GiB = 2^{30} bytes

- 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

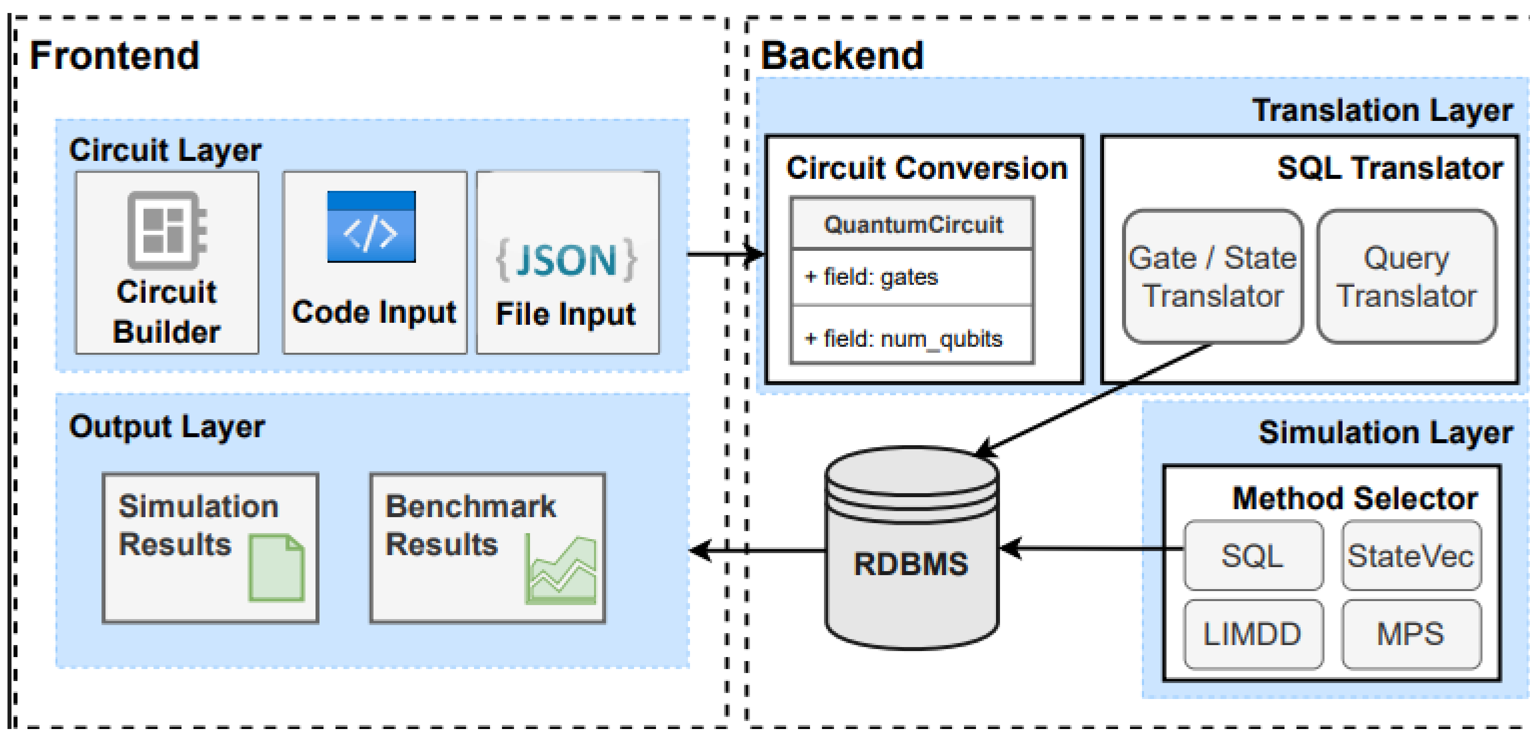
Sergio Boixo[1*], Sergei V. Isakov[2], Vadim N. Smelyanskiy[1], Ryan Babbush[1], Nan Ding[1], Zhang Jiang[3,4], Michael J. Bremner[5], John M. Martinis[6,7] and Hartmut Neven[1]

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

Efficient tensor computation: database to the rescue

Q1: Push the simulation workload to DBMSs?

System



Littau, Tim, and Rihan Hai. "Qymera: Simulating Quantum Circuits using RDBMS." SIGMOD. 2025.

Theory

QC meet CQ: Quantum Conjunctive Queries

Floris Geerts
University of Antwerp
Antwerpen, Belgium
floris.geerts@uantwerpen.be

Rihan Hai
University of Delft
Delft, Netherlands
R.Hai@tudelft.nl

Abstract

We explore how recent methods for evaluating conjunctive queries (CQs) can help to efficiently simulate quantum circuits (QCs), i.e., computing output amplitudes from a given input state.

ACM Reference Format:

Floris Geerts and Rihan Hai. 2025. QC meet CQ: Quantum Conjunctive Queries. In *Workshop on Quantum Computing and Quantum-Inspired Technology for Data-Intensive Systems and Applications (Q-Data '25), June 22–27, 2025, Berlin, Germany.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3736393.3736696

Hypertree width of quantum conjunctive queries. An initial observation is that the treewidth of a quantum CQ [4] aligns with the treewidth of the corresponding QC [6]. Treewidth is defined via the CQ's primal graph, where nodes represent variables and edges connect variables co-occurring in a relation. In QC terms, variables map to qubits and relations to gates, making the primal graph of a quantum CQ the dual of the QC's circuit graph. However, graph-based representations are not always ideal. For instance, acyclic CQs can have arbitrarily large treewidth, despite being evaluable in linear time via the Yannakakis algorithm. To address this,

State $\lvert\psi\rangle$	Order-n tensor (Baseline I)	Relational representation (RDBMS solutions)	MPS (Baseline II)
General state	$O(2^n)$	$O(n \cdot \mathrm{nnz}(\lvert\psi\rangle))$	$O(n\chi^2)$
W_n State	$O(2^n)$	$O(n^2)$	$O(n)$
GHZ_n State	$O(2^n)$	$O(n)$	$O(n)$
QFT_n	$O(2^n)$	$O(n \cdot 2^n)$	$O(n\chi^2)$

Table 1: Space complexity comparison of different representations of state $\lvert\psi\rangle$. Here, n is the number of qubits, $nnz(\lvert\psi\rangle)$ denotes the number of non-zero probability amplitudes in the state $\lvert\psi\rangle$, and the MPS bond dimension χ is a fixed constant that one chooses oneself, potentially making the representation approximate.

Geerts, Floris, and Rihan Hai. "QC meet CQ: Quantum Conjunctive Queries." *Proceedings of the 2nd Workshop on Quantum Computing and Quantum-Inspired Technology for Data-Intensive Systems and Applications*. 2025.

Hai, Rihan, et al. "Quantum Data Management in the NISQ Era: Extended Version." arXiv preprint arXiv:2409.14111 (2024).

Databases to the Rescue: Classical-Quantum Simulation System

Automatic Optimization

Providing the most efficient simulation by selecting optimal data structures and operations based on available resources and circuit properties.

Out-of-Core Operation

Supporting simulation of large circuits that exceed main memory capacity through efficient memory management.

Consistency & Recovery

Preventing data corruption and enabling recovery in the event of large-scale simulation crashes.

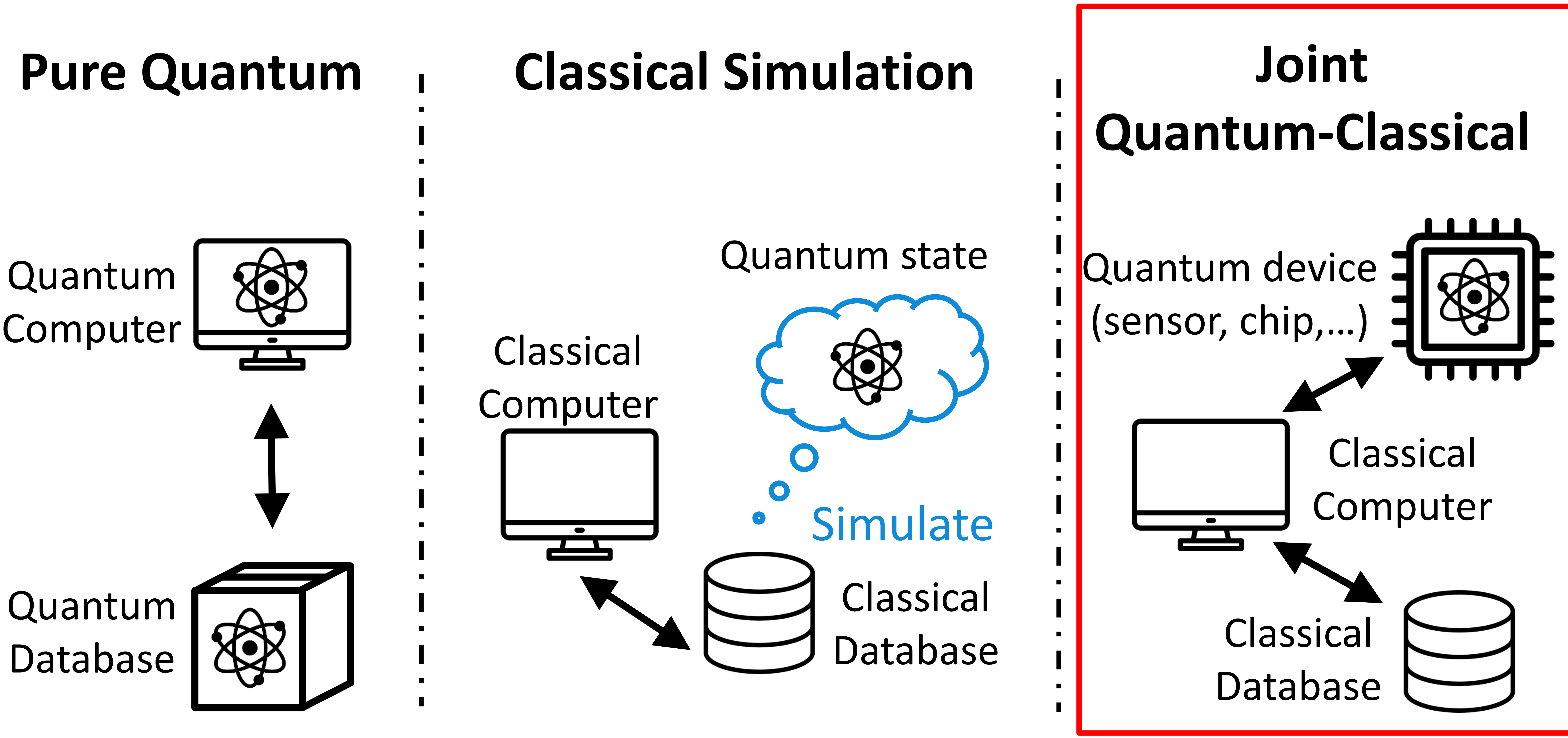
Workflow Improvement

Enhancing the entire simulation process, including parameter tuning, data collection, querying, exploration, and visualization.

At its core, a CQSS must be capable of evaluating quantum circuits, primarily involving tensor network operations.

Rihan Hai, Shih-Han Hung, Tim Coopmans, Tim Littau, and Floris Geerts. Quantum data management in the NISQ era. PVLDB, 18(6):1720–1729, 2025

Landscape: data management for quantum computing



Rihan Hai, Shih-Han Hung, Tim Coopmans, Tim Littau, and Floris Geerts. Quantum data management in the NISQ era. PVLDB, 18(6):1720–1729, 2025

Joint Quantum-Classical

- Can we use AI to solve quantum problems?

Example: Quantum circuit synthesis with diffusion models

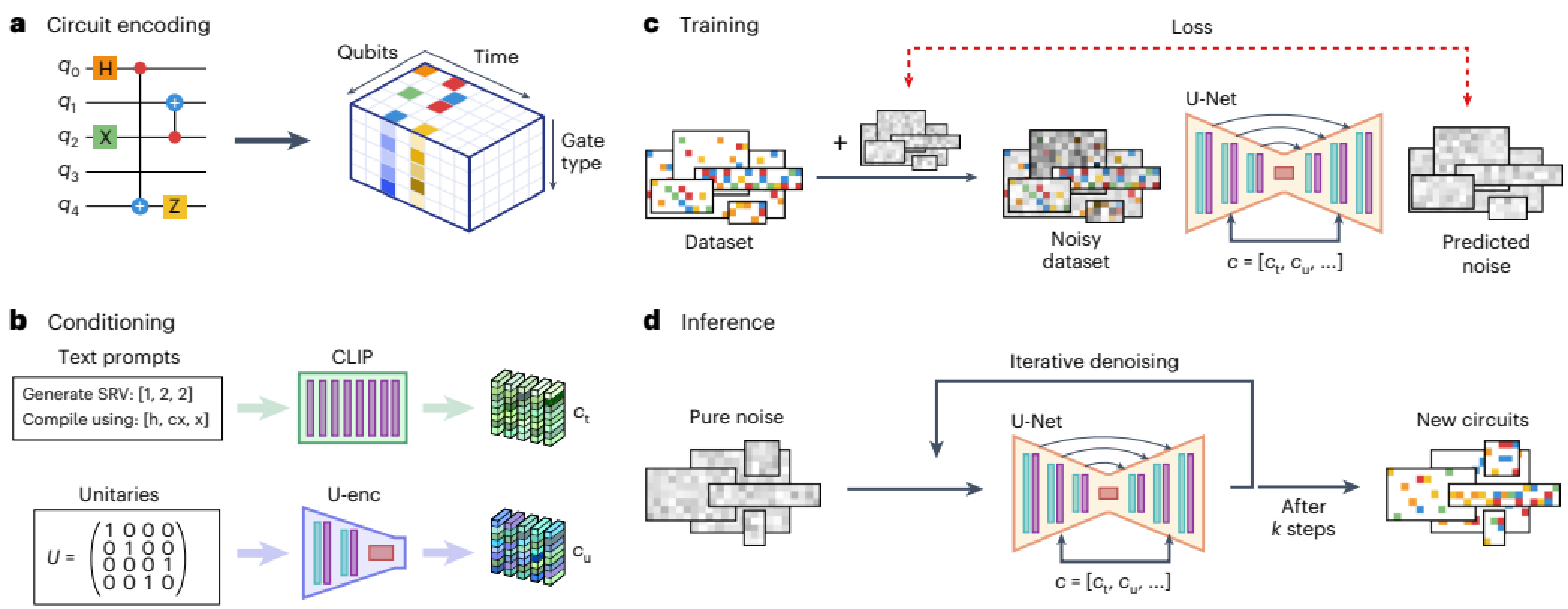


Fig. 1 | Quantum circuit generation pipeline summary. a, Quantum circuits are encoded in a three-dimensional tensor, where each gate is encoded as a continuous vector of certain length (vertical direction), schematically represented here as a colour (upper plane). **b**, Creation of the diffusion model's conditioning (c). Text is transformed into a continuous representation (c_t) by means of a pre-trained CLIP encoder. In other cases, as for unitary compilation, an encoder is trained together with the DM to create the encoding (c_u) of an input unitary. **c**,**d**, Schematic representation of the training of the diffusion model (**c**) and the posterior inference from the trained model (**d**). See text for details.

Fürrutter, Florian, Gorka Muñoz-Gil, and Hans J. Briegel. "Quantum circuit synthesis with diffusion models." Nature Machine Intelligence 6.5 (2024): 515-524.

Development timeline

Quantum Computing

2012
Development of the first single atom transistor

2021
Development of first integrated circuit at the atomic scale

2025-2028
Development of the first commercially useful device

2030+
Error-corrected quantum processor

Classical Computing

1947
Development of the first transistor

1958
Development of the first integrated circuit

1964
Use in the first commercial product

1971
First semi-conductor based industrial computer

Summary: Quantum data management in NISQ era

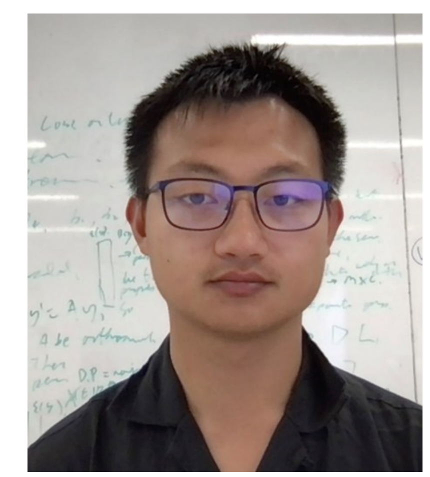
A privileged time in data management,
with many research problems awaiting exploration.

Join us

We are InfiniData Team

Data systems for AI

Data systems for quantum



- How to join us?
  - Thesis projects
  - Honours programme
  - Or simply, write me an email r.hai@tudelft.nl

Q&A: Where do you want to start?

1. What is data?
2. What quantum data systems do you want to build?