Bachelor-/Master-Forum 2024

Quantum and Intelligent Computing (QIC)

Institute of Information Systems (IFIS)

29.10.2024

Professor Dr. rer. nat. habil. Sven Groppe
https://www.ifis.uni-luebeck.de/index.php?id=groppe

Quantum and Intelligent Computing (**QIC**)

- Head: Prof. Dr. rer. nat. habil. Sven Groppe
- Projects and Research Assistants
  - **QC4DB**: Accelerating Relational Database Management Systems via Quantum Computing (BMBF)
    - Umut Çalikyilmaz
    - Tobias Winker
    - Nitin Nayak
  - **QualityOnt**: High Quality Knowledge Graphs from recent English, French and German Emergent Trends with the example of COVID-19 (DFG/ANR)
    - Hanieh Khorashadizadeh
  - **Semantic Data** Integration and Analysis (Bosch)
    - Simon Paasche (External PhD Student)
  - **Hybrid**^{2}-Index Structures for Main Memory Databasess (DFG)
    - Tobias Groth

Supervision of Bachelor/Master Thesis & Result

- Often co-supervision of Prof. Groppe together with PhD student
  - meetings regularly and on request, typical:
    - weekly meetings with PhD student
    - monthly meetings with Prof. Groppe
- Experience
  - 99 supervised thesis (bachelor/master/student research project/Diploma/PhD)
- Publications based on bachelor/master thesis
  - improves visibility of student's contribution
  - improves chances for good job (in academia and industry)
  - 47 (out of 186) publications (of QIC lab) are co-authored by a bachelor/master student (being a student at time of writing)
    - 25% of the publications

Typical Outline of Bachelor/Master Thesis

1. Introduction/Einführung
   1.1. Motivation
   1.2. Tasks of the Thesis/Aufgabenstellung
   1.3. Organization/Organisation der Arbeit
2. Basics/Grundlagen
   2.1. ...
   2.2. Further Related Work/Weitere wissenschaftliche Literatur
3. Concept/Konzept
4. Realization/Realisierung
5. Evaluation
6. Summary and Conclusions/Zusammenfassung und Ausblick

- Latex Template available, e.g.: **In Moodle**
- FAQs on Bachelor's and Master's Theses (from examination board for MINT): in **English**/**German**

DVD in addition to Thesis & for IFIS Archive

Please do not forget to burn DVDs for each of the thesis, content:

- Readme-file with installation instructions
- source code with documentation
  - additionally push to thesis repository in IFIS-Gitlab
- all necessary third-party-libraries
- all data sets for reproducing evaluation in thesis
- PDF of the thesis
- source files of the thesis (Word-file/latex-folder)

On the day of defense please deliver DVD for the IFIS-archive:

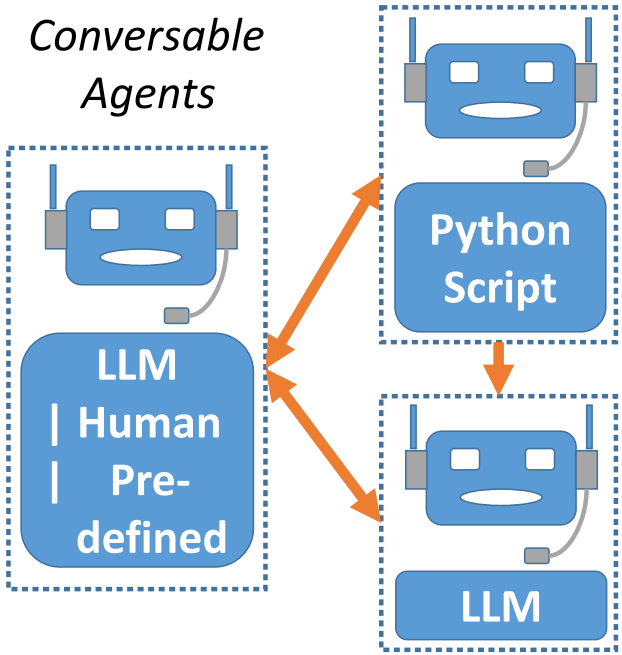
- Content as above
- PDF of the presentation for the defense
- source files of the presentation for the defense (Powerpoint file/latex-folder etc.)

Typical Defense of Thesis

- 20-25 minutes presentation
  - often similar structure like thesis, but without 4. Realization
- maybe with succeeding short demonstration ($\approx$ 5 minutes) of developed software (dependent on thesis)
- Afterwards questions of reviewers and listeners
- Reviewers discuss alone in room about result
- Reviewers announce score to student and explain the reasons for the score
- In total: up to 1 hour

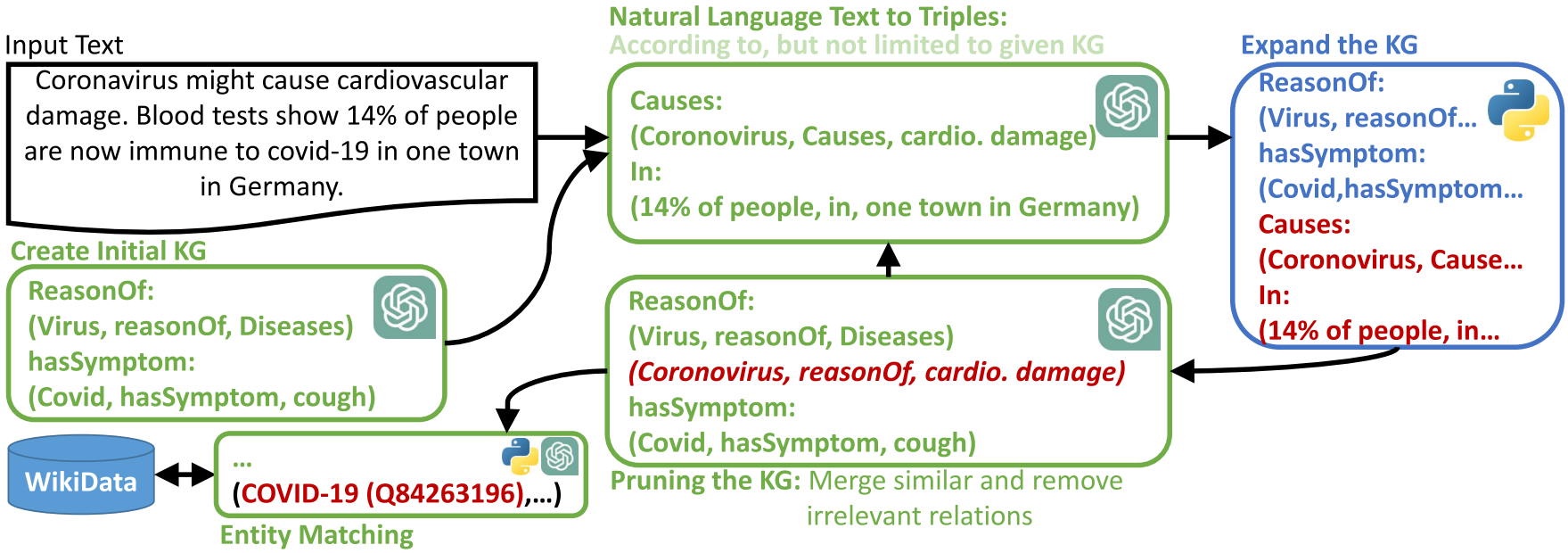
Large Language Model (LLM) Applications

- New kind of application in form of a network of (LLM, human, code) agents
  - Applying LLMs as new paradigm for software development of applications close to natural language/human users
  - Development of new LLM applications like
    - modern chatbots
    - extraction from natural languages text (e.g., for knowledge graph construction)
    - new types of applications with reasoning capabilities (self-reflection etc. for human-like generation of content)



  - Framework for platform-independent LLM application
    - Utilizing modern programming languages like Kotlin to run code across JVM, binary targets and browsers

Construct Knowledge Graph (KG)/LLMs

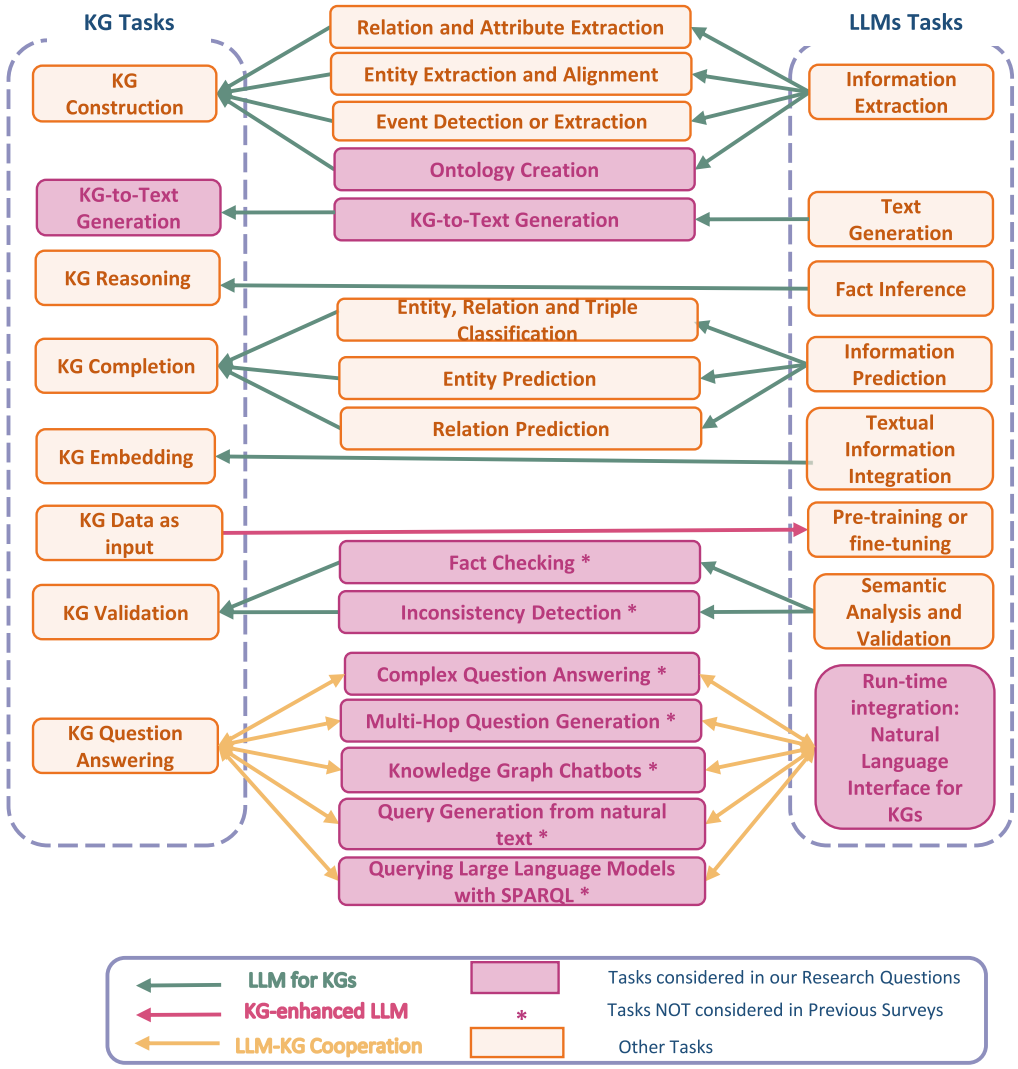


- Result (comparison to manual labeling)

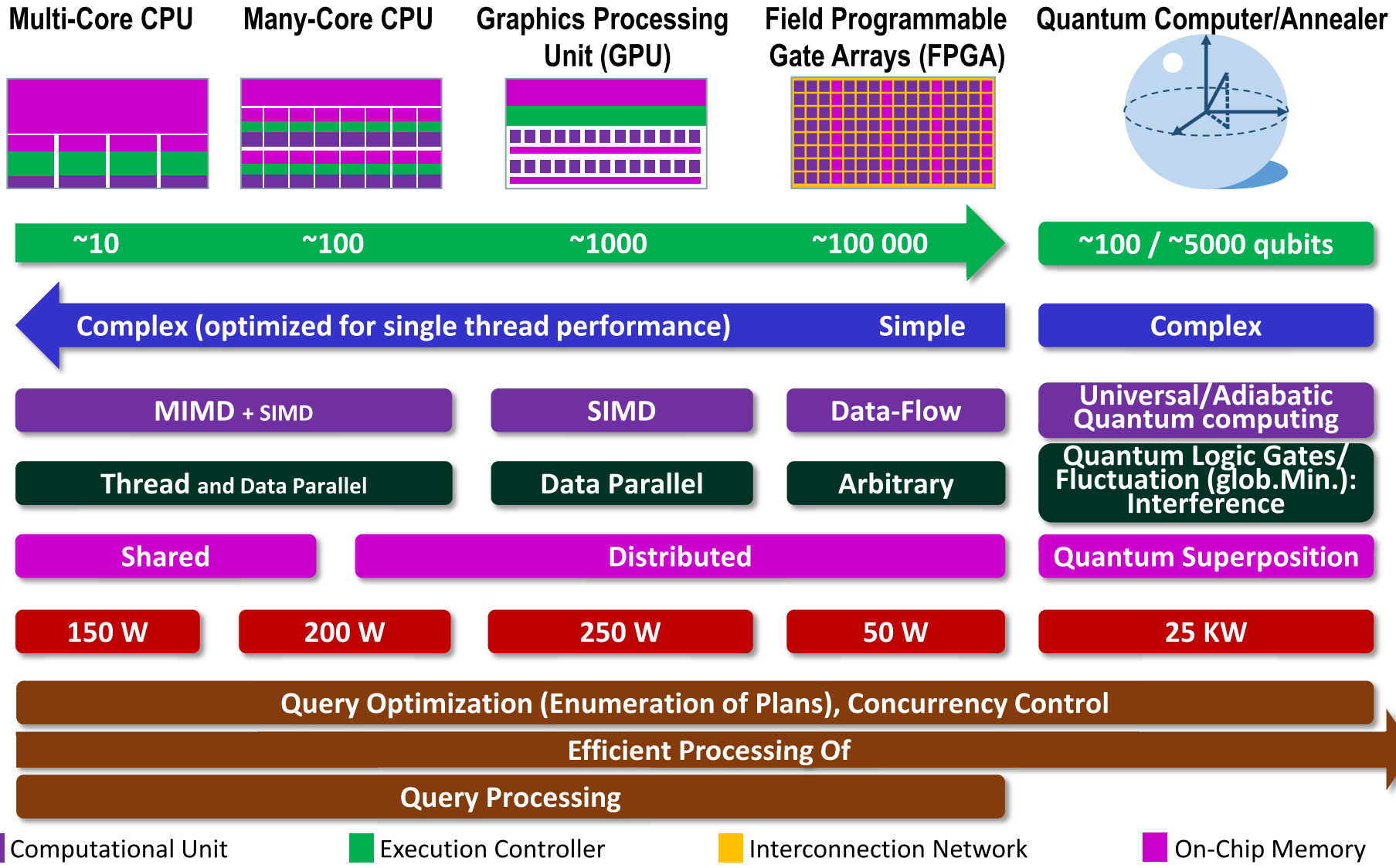
	Cov	Gen	Db
Macro F1-score	66.60	**76.93**	72.85

  - COV: Represents the original triples extracted using GPT-4
  - GEN: COV with augmented data generated by GPT-4
  - DB: Merges the original triples from COV with triples from an external data source (random set of 308 tuples from a publicly available structured database about the daily statistical updates and policy actions)

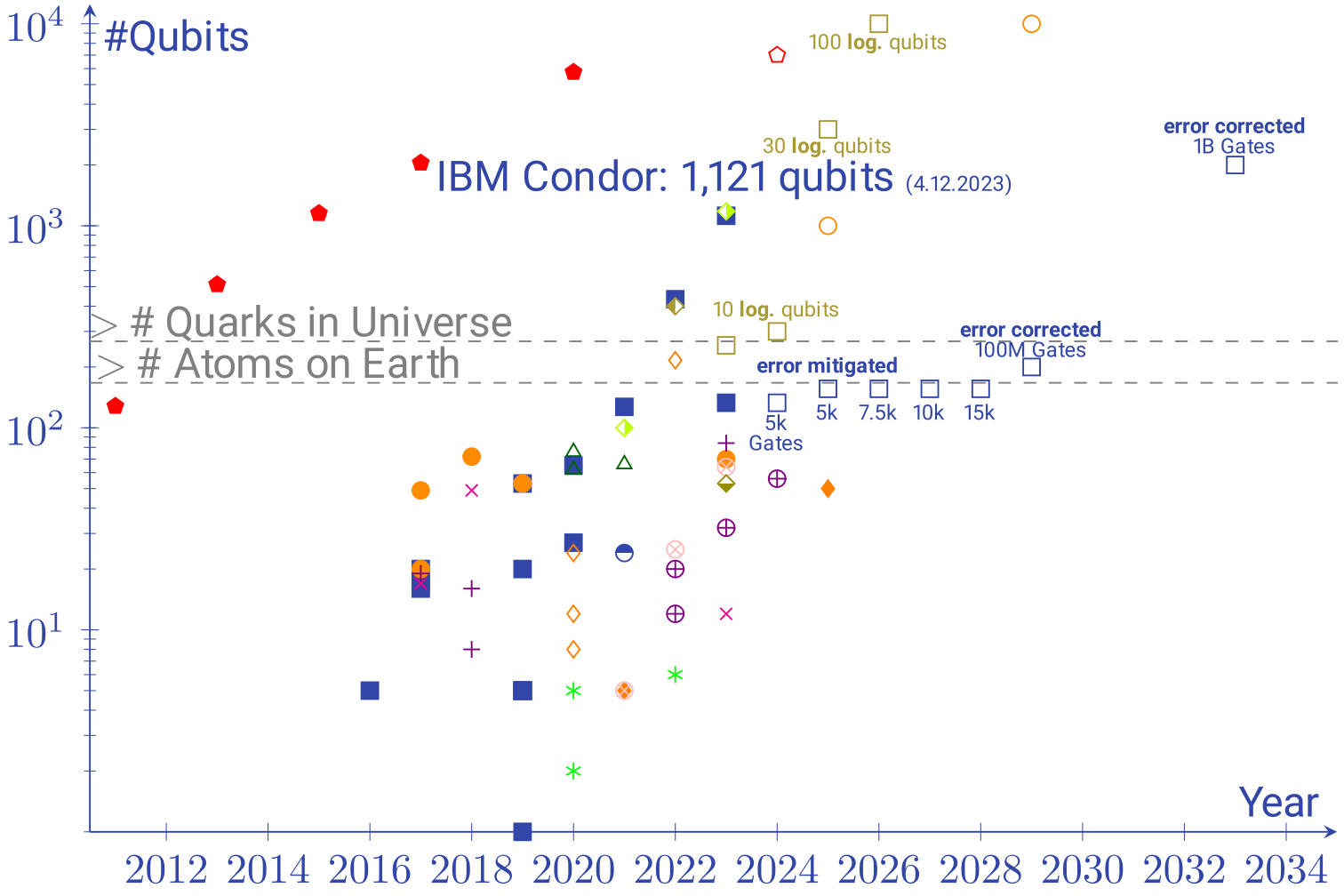
Interplay of Large Language Models (LLM) and Knowledge Graphs (KG)



Architectures of Emergent Hardware

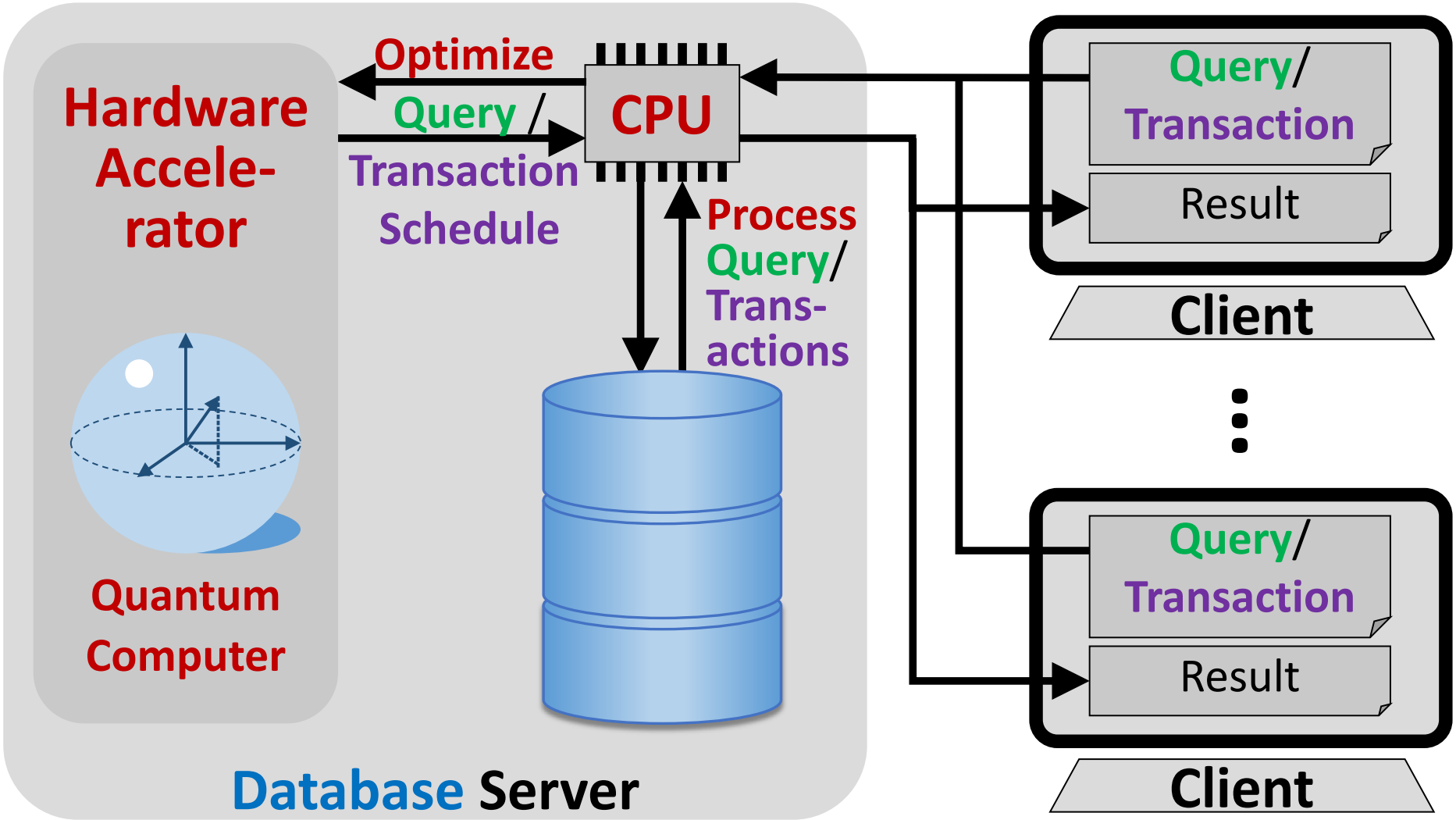


Timeline of Quantum Computers



- IBM
- Google
- Intel
- Rigetti
- QuTech
- USTC
- Xanadu
- Quantum Brilliance
- Quantinuum
- Quantware
- RIKEN
- AQT
- Atom Computing
- M Squared Lasers
- |QuEra⟩
- D-Wave *(Quantum Annealing)*

Using Hardware Accelerator for optimizing Queries / Transaction Schedules



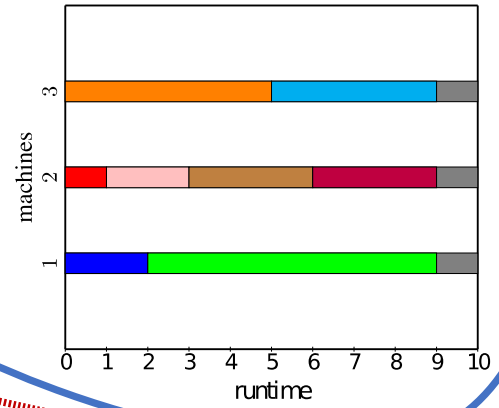
Approaches for Query/Transaction Schedule Optimization

Query Optimization:

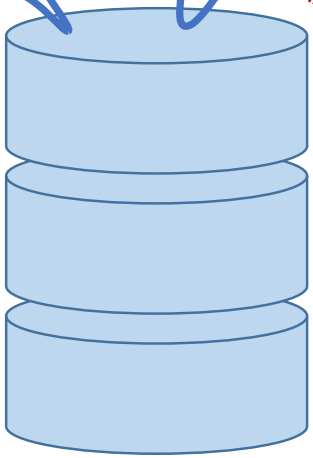
$$\bowtie_{i=1}^{n} R_i \;?\; \begin{matrix} (R_1 \bowtie R_2) \cdots \bowtie R_n \\ \vdots \\ (R_1 \bowtie R_n) \bowtie (\cdots) \end{matrix}$$

Transaction Schedule Optimization:

$$\{T_1, \ldots, T_m\} \;?$$



Open Source Relational Database Management System (RDBMS), e.g. PostgreSQL, MySQL



Dynamic Programming
Random Walk
Linear Program-ming
...
Machine Learning
Genetic Algorithm
Simulated Annealing

Algorithms (used e.g. in Query Optimization) and their Quantum Counterparts

Query Optimization Approach	Basic Algorithm	Quantum Computing Counterpart
[S+79]	Dynamic Programming [E04]	[R19] [A+19]
[IW87], QA: [TK16]	Simulated Annealing [KGV83]	[J+11]
[MP18] [Y+20] [W+19] [O+19]	Reinforcement Learning [BSB81]	[S+21] [DCC05]
[GPK94]	Random Walk [BN70]	[ADZ93] [A+01]
[BFI91]	Genetic Algorithm [H92]	[W+13]
[TC19]	Ant Colony Optimization [CDM91] [DBS06]	[WNF07] [G+20]
[TK17]	Mixed Integer Linear Programming [BGG+71] [D02]	[HHL09] [A12] [CKS17] [SSO19] [AL22] [AL22]

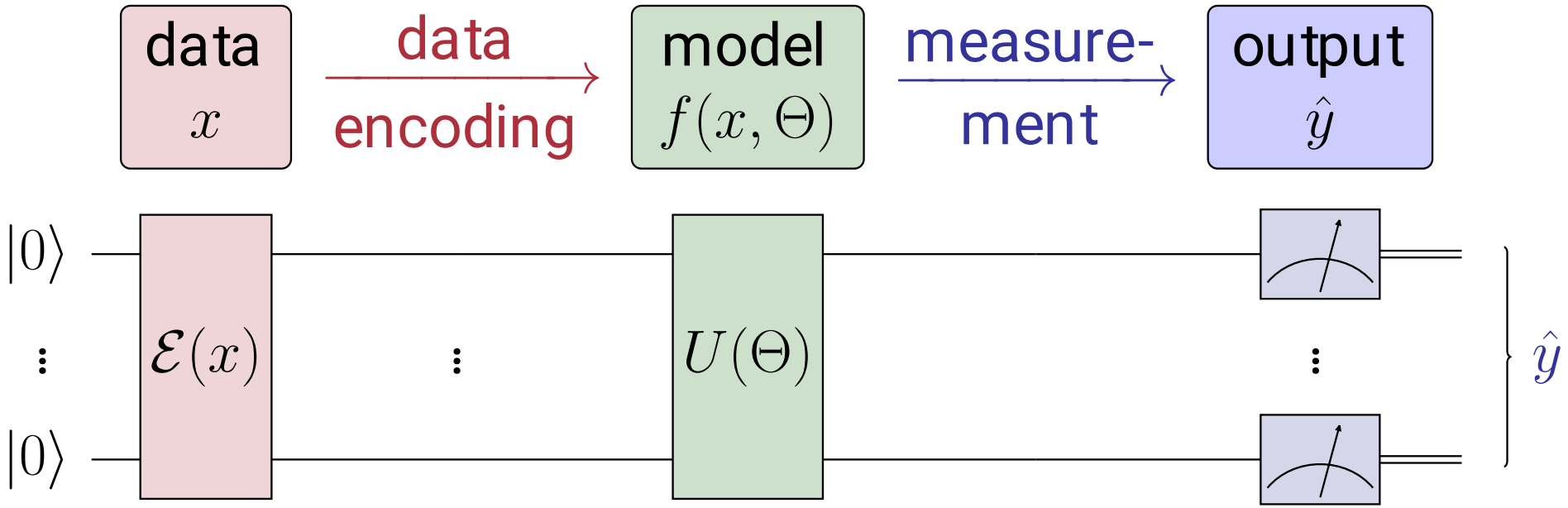
This list is not complete...

- Please check my lecture about quantum computing: https://www.ifis.uni-luebeck.de/~groppe/lectures/qc

Quantum ML for Query Optimization

(Project with Quantum Brilliance)

- Compared to ML: **Learning with fewer data, higher accuracy**



Input: Relations

$R_{x_1} \bowtie ... \bowtie R_{x_k} \rightarrow$
$(cos(x_1) \cdot |0\rangle + sin(x_1) \cdot |1\rangle) \otimes ... \otimes (cos(x_k) \cdot |0\rangle + sin(x_k) \cdot |1\rangle)$
with $x_1, ..., x_k$ normalized to $[0, 2 \cdot \pi]$

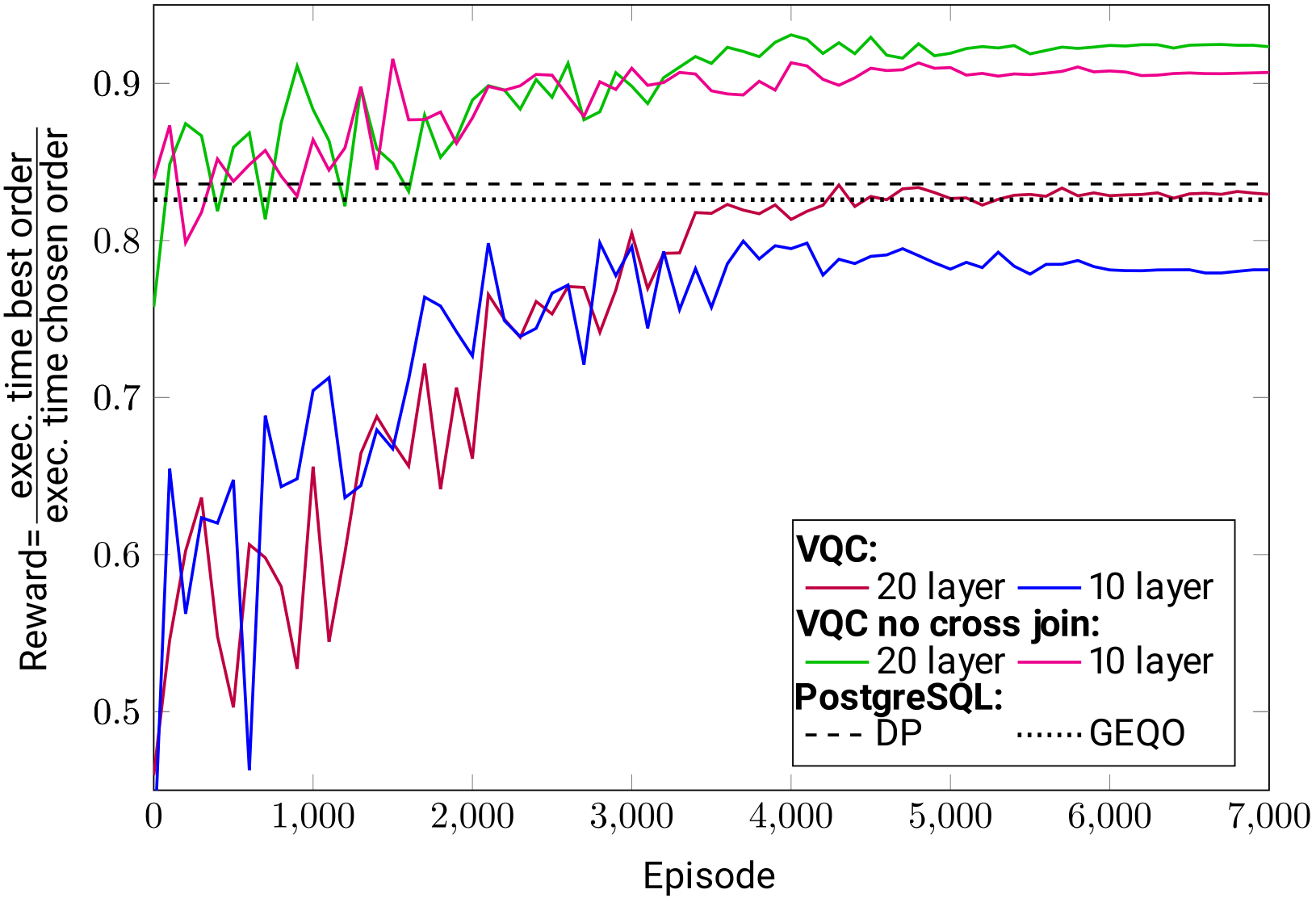
Trained Model

Output: Join Order

$o \in \{0, ..., 2^k - 1\}$: measured value with highest probility
Choose join order $o \mod k$
with k number of valid join orders

Quantum ML for Query Optimization

(Project with Quantum Brilliance)



Quantum Annealing for Query Optimization

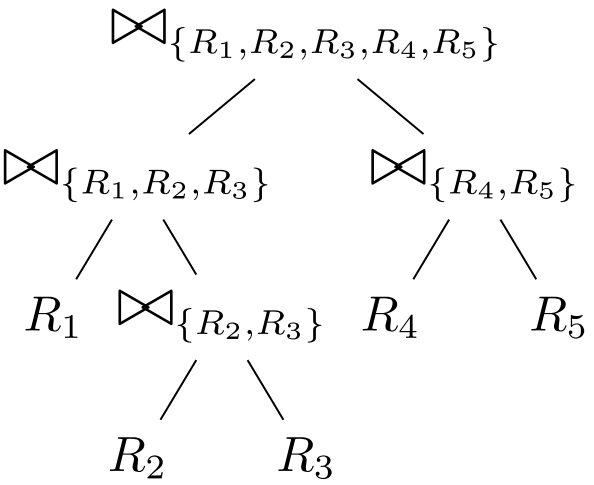
(Project with Quantum Brilliance)

- **Quantum Annealing** solves quadratic unconstrained binary optimization (**QUBO**) problems

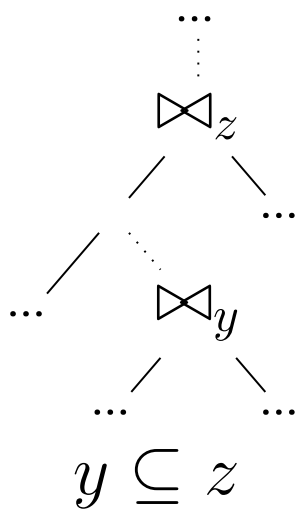
Example Join Ordering:

- P: power set of relations to join
- P_k: set of elements representing joins of k relations, i.e., $P_k = \{a | a \in P \wedge |a| = k\} \subseteq P$
- $w_{max} = max(w_i) + c$, where $c > 0$ and w_i represents the cost of join i
- **Rewarding joins with lowest costs:** $S_1 = \sum_{j=2}^{m} \sum_{i \in P_j} x_i \cdot (w_i - w_{max})$

Example



Constraint:



Punish other combinations:

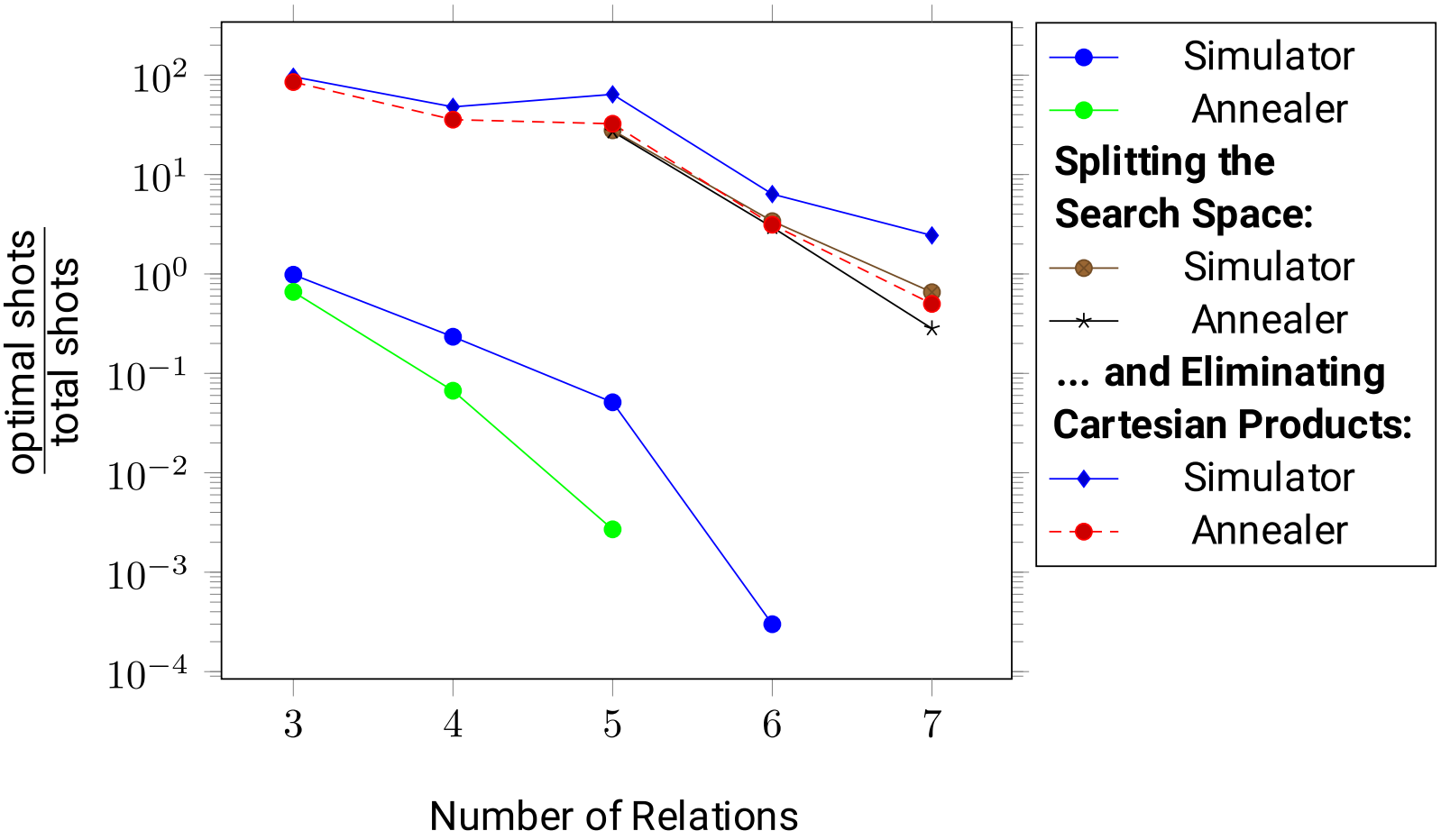
$$\Rightarrow S_2 = \sum_{i=2}^{m-1} \sum_{j=i}^{m} \sum_{\substack{y \in P_i \\ \wedge z \in P_j \\ \wedge y \cap z \neq \emptyset \\ \wedge y \not\subseteq z}} x_y * x_z * w_{max}$$

Minimize $S = S_1 + S_2$

Quantum Annealing for Query Optimization

(Project with Quantum Brilliance)

- Real-world queries of the ErgastF1 Benchmark with PostgreSQL



- QPU access time approx. 100ms

Optimizing Transaction Schedules via Quantum Annealing

- Experiments on real Quantum Annealer (D-Wave 2000Q cloud service)
  - first minute free
    (afterwards too much for our budget)
- Versus Simulated Annealing on CPU
- Preprocessing time/Number of QuBits:
  $O((n \cdot k \cdot R)^2)$

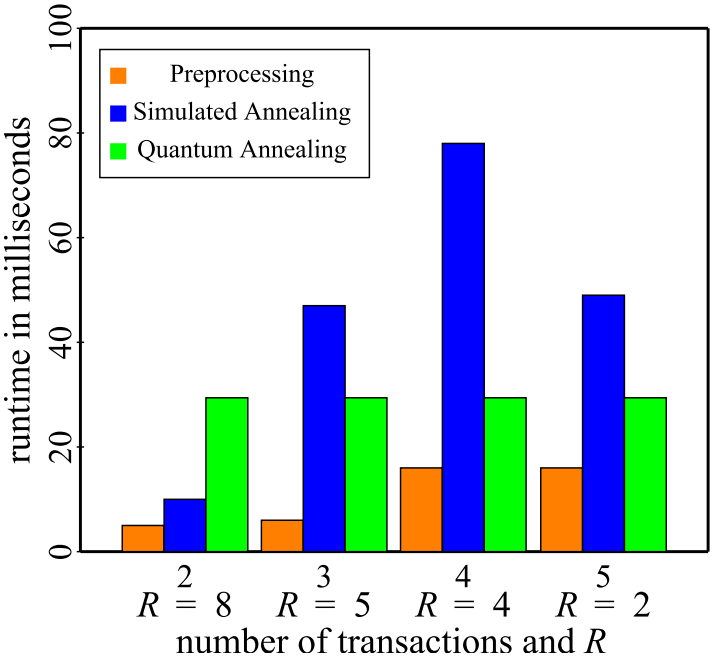


Fig.	k	n	R	O	$l_1, \ldots, l_n$	$r_1, \ldots, r_n$	req. var.
11	2	2	8	$\{\}$	8, 4	0, 4	8
		3	5	$\{(t_1, t_3)\}$	4, 5, 1	1, 0, 4	10
		4	4	$\{(t_2, t_4)\}$	3, 2, 1, 2	1, 2, 3, 2	16
		5	2	$\{(t_1, t_2), (t_4, t_5)\}$	1, 1, 1, 1, 1	1, 1, 1, 1, 1	10

Open Challenges for QC for Databases/ Topics for Thesis

- Replacing basic algorithms with their QC counterparts in query optimizations for speeding up databases
  - Query Optimization (QO): Tobias Winker
  - Transaction Schedule (TS) Optimization: Umut Çalikyilmaz
  - Quantum Annealing (QO/TS): Nitin Nayak
- What should be the properties of a quantum computer (e.g. #qubits, latencies of gates) to achieve certain speedups?
- How to combine classical and quantum computing algorithms to achieve good speedups with few qubits?
  (...for running database optimizations on current available quantum computers...)
- What other (database) domains besides query and transaction schedule optimizations benefit from quantum computers?

QC4DB: Accelerating Relational Database Management Systems via Quantum Computing

Name:	QC4DB: Accelerating Relational Database Management Systems via Quantum Computing	
Proj. Web:	**Project Website@Quantentechnologien**	
Funded by:	BMBF, Fördermaßnahme Anwendungsnetzwerk für das Quantencomputing	
Duration:	3 years	
Volume:	1.8M Euros	
Topics:	Optimizing an open source relational database management system • Queries • Transaction Schedules	
Partners:	UNIVERSITÄT ZU LÜBECK INSTITUT FÜR INFORMATIONSSYSTEME (Coord.)	QUANTUM BRILLIANCE
Expertises:	Hardware-Acceleration of Databases	Room Temperature Diamond Quantum Accelerators/qbOS
Website:	https://www.ifis.uni-luebeck.de/~groppe/	https://quantumbrilliance.com/

Quantum Computing for Chip Design

- FPGA routing
  - Eliminating cross-talks of connections between components in an FPGA design and find an optimal "wiring" (see e.g. [NSR02])
- Logic Synthesis
  - Minimizing Boolean Formulas
    - Classical Software: **Logic Friday** and **Espresso-IISOJS**

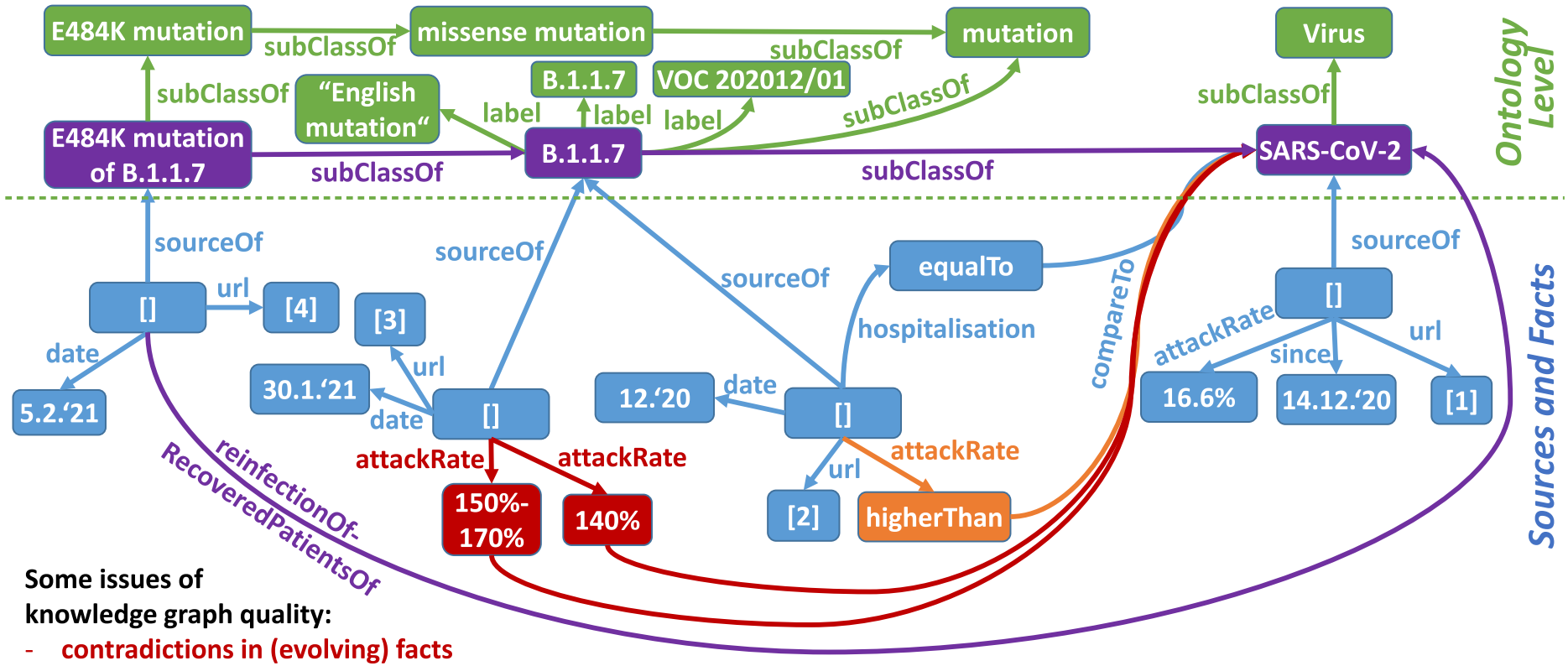
References:

[NSR02] G. Nam, K. Sakallah and R. Rutenbar, A New FPGA Detailed Routing Approach Via Search-Based Boolean Satisfiability, IEEE Transactions on computer-aided design of integrated circuits and systems, vol. 21, no. 6, 2002. **PDF@RG**

Quantum Compting for...

- astrophyiscs, e.g.,
  - estimating the weight of black holes
  - galaxy clustering
- image processing, e.g.,
  - classification of images
- graph neural networks, e.g.,
  - software vulnerability detection
  - graph pattern matching)
- many more use cases

COVID-19 Knowledge Graph (KG)



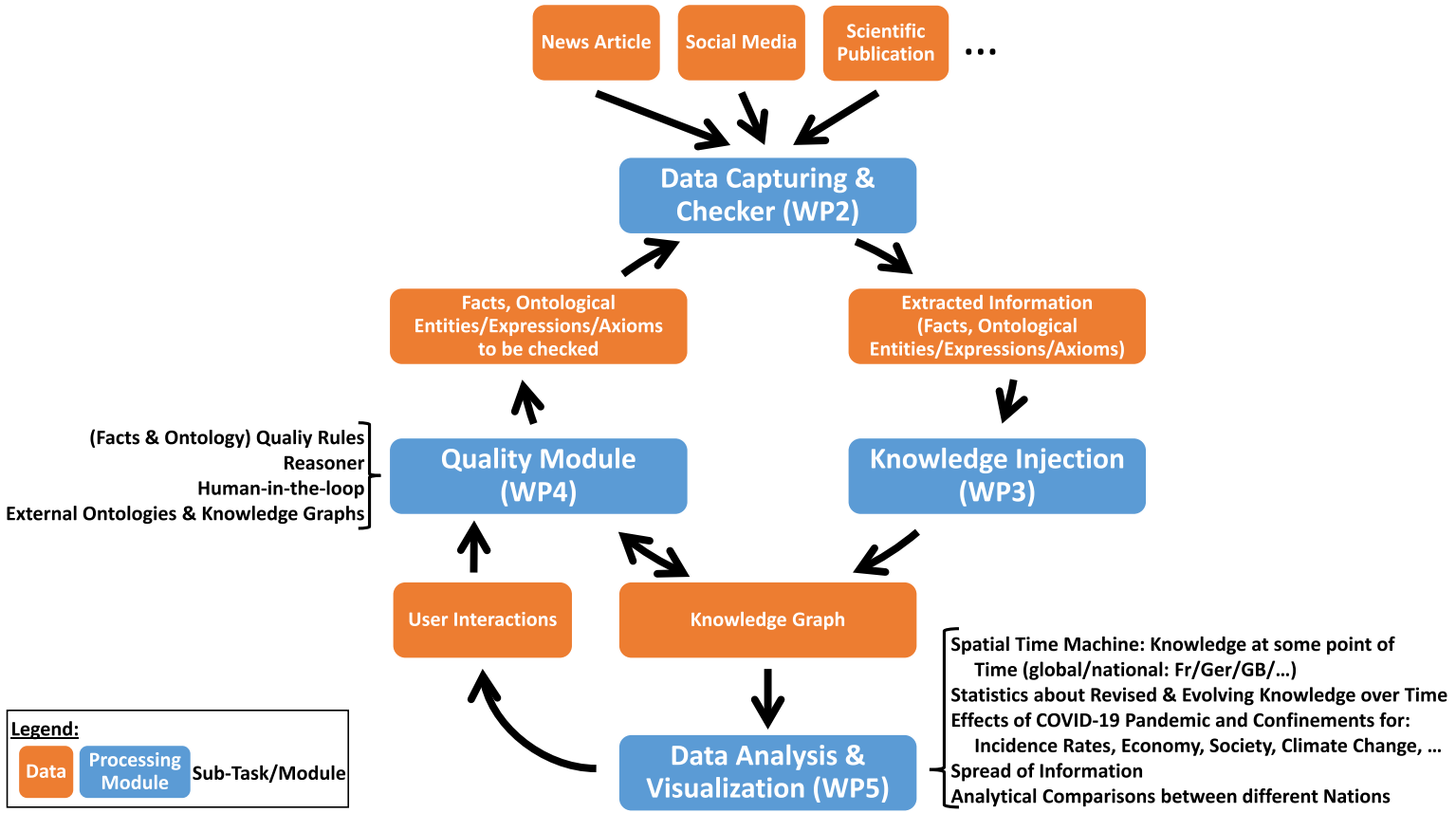
Some issues of knowledge graph quality:

- **contradictions in (evolving) facts**
- **checking vague formulations and compare them with other given information**
- **errors with ambiguity: E484K mutation of B.1.1.7 is a SARS-CoV-2 virus, but reinfection with E484K is only possible for patients recovered from SARS-CoV-2 virus not mutating E484K (→ introduce class "SARS-CoV-2 without E484K mutation")**

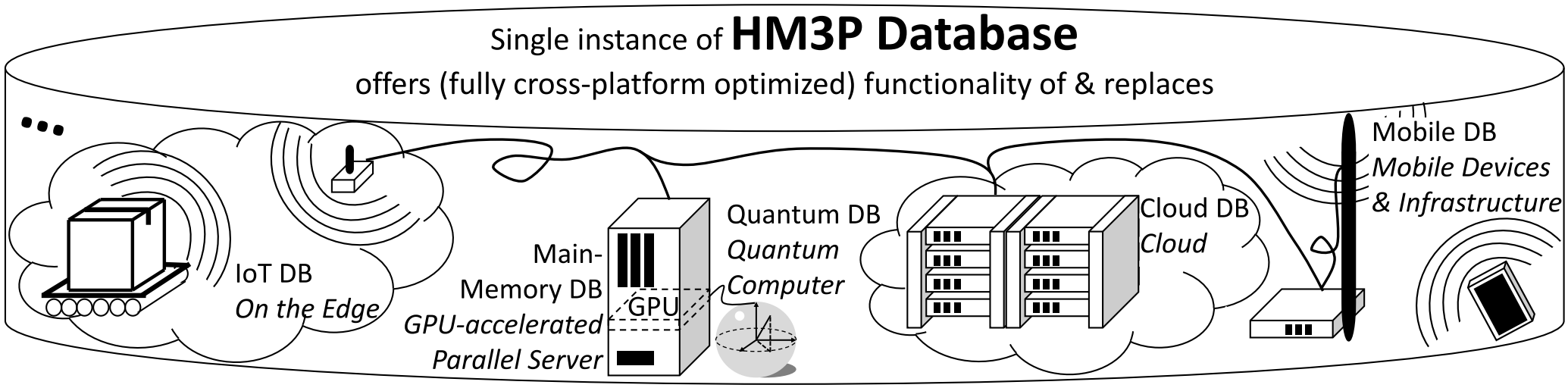
[1] https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2774102
[2] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959361/Technical_Briefing_VOC202012-2_Briefing_2.pdf
[3] https://www.ruhr24.de/service/corona-britische-mutation-neue-studie-mutante-virus-toedlicher-sterberate-deutschland-90184403.html
[4] https://www.bmj.com/content/372/bmj.n359

High Quality KGs from recent English, French & German Emergent Trends with the example of COVID-19

- Project with partners in Paris & Toulouse (Hanieh Khorashadizadeh)
- Thesis in the areas of
  - data capturing,
  - visualization & analysis,
  - detection of contradictions in KG, ...

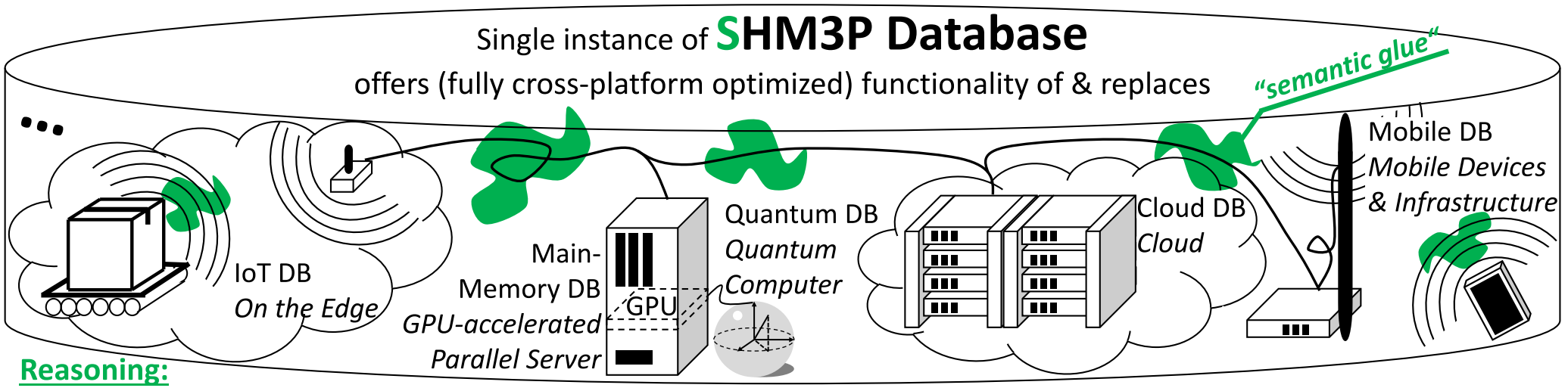


Hybrid Multi-Model Multi-Platform (HM3P) Database



- full and uniform data integration at database level
- performance: fully optimized across different data models
- transparent fault-tolerance
- SQL standards: relational ('87), XML ('03), temporal ('11), JSON ('16), Multi-dimensional Arrays ('19), schemaless ('19), streams ('20?), property graphs ('21?)
- **features of different types of databases running on different platforms can be used**

Variant: Semantic HM3P (SHM3P) DB



Reasoning:

Lightweight reasoning on large data sizes of IoT devices

Heavyweight reasoning on moderate data sizes

Heavyweight reasoning on large data sizes

Reasoning on small data sizes of mobile devices

How to integrate the different reasoning capabilities and requirements into one transparent global reasoner?

- Semantic Layer as glue between other models and platforms
- new challenges like integrating different types of reasoners in a transparent global reasoner

\+ Features of HM3P databases

\+ Easier data integration

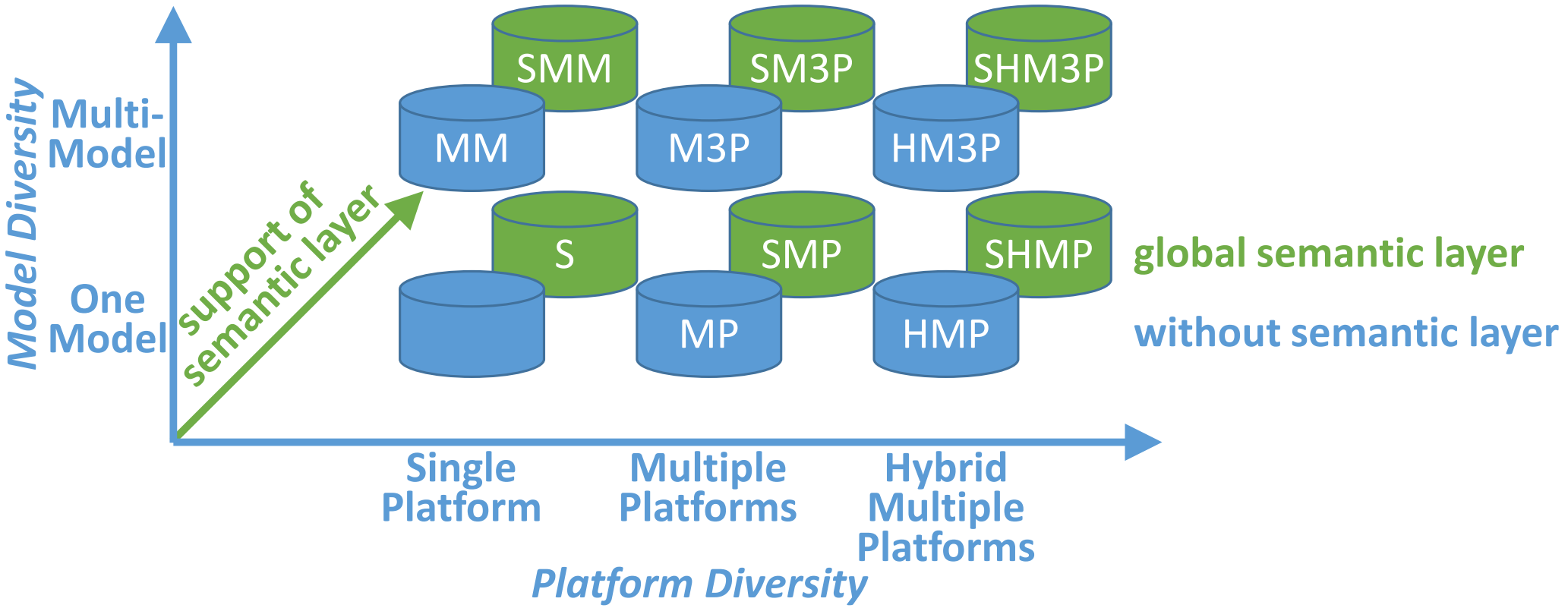
\- Performance issues may occur due to semantic layer

Types of DBMS

state-of-the-art

partly/ rudimentarily addressed

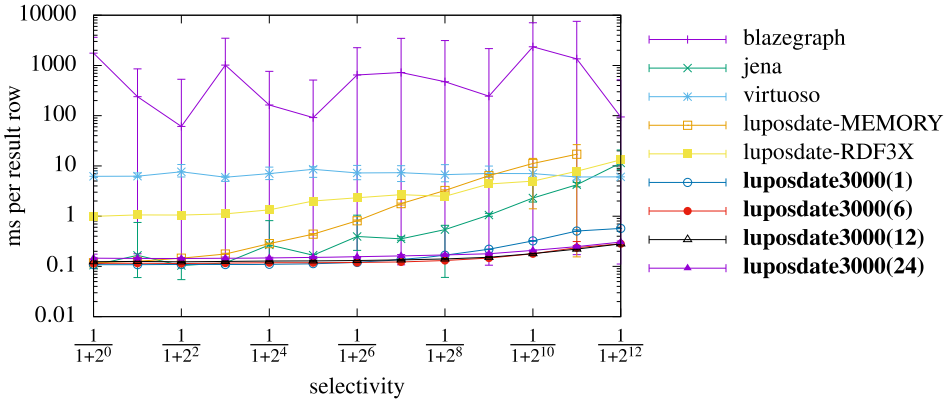
visionary/ single attempts (e.g. hybrid cloud)



Legend: S: Semantic MP: Multi-Platform MM: Multi-Model M3P: MM MP H: Hybrid

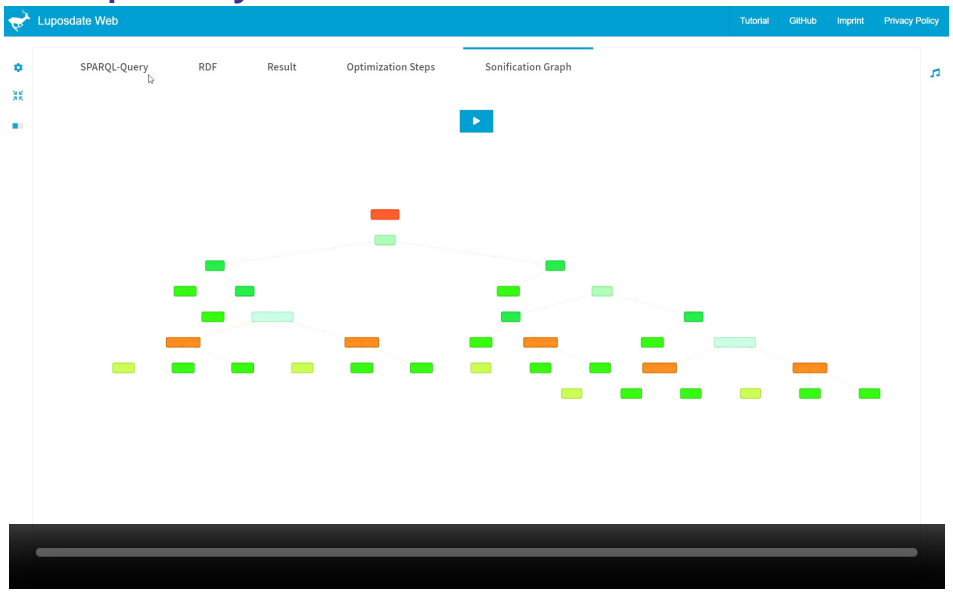
The Power of Multi-Platform: LUPOSDATE3000

- ultra-fast in jvm...



B. Warnke, M.W. Rehan, S. Fischer, S. Groppe: Flexible data partitioning schemes for parallel merge joins in semantic web queries in: BTW'21

- ...but also enabling web demos running completely in the browser!



S. Groppe, R. Klinckenberg, B. Warnke. Sound of Databases: Sonification of a Semantic Web Database Engine. PVLDB, 14(12), 2021

Quantum and Intelligent Computing (**QIC**)

- Head: Prof. Dr. rer. nat. habil. Sven Groppe
- Projects and Research Assistants
  - **QC4DB**: Accelerating Relational Database Management Systems via Quantum Computing (BMBF)
    - Umut Çalikyilmaz
    - Tobias Winker
    - Nitin Nayak
  - **QualityOnt**: High Quality Knowledge Graphs from recent English, French and German Emergent Trends with the example of COVID-19 (DFG/ANR)
    - Hanieh Khorashadizadeh
  - **Semantic Data** Integration and Analysis (Bosch)
    - Simon Paasche (External PhD Student)
  - **Hybrid**^{2}-Index Structures for Main Memory Databasess (DFG)
    - Tobias Groth