From Sustainable IT to the IT for the Sustainable World

Ivona Brandić

Vienna University of Technology

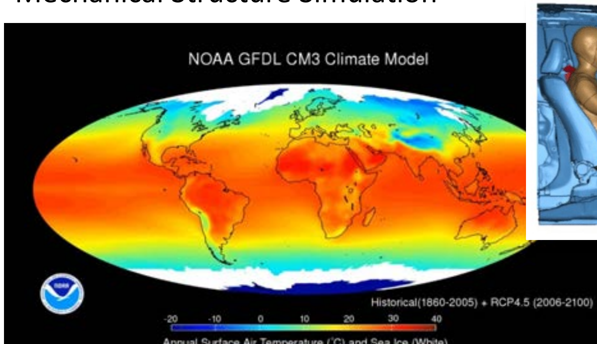
ivona.brandic@tuwien.ac.at

http://hpc.ec.tuwien.ac.at

Computational Power

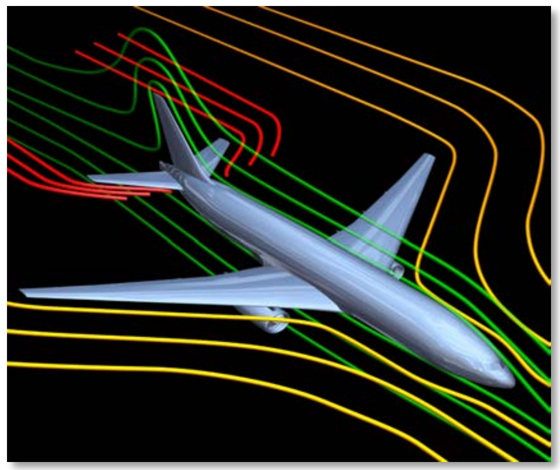
Simulation

Mechanical Structure Simulation

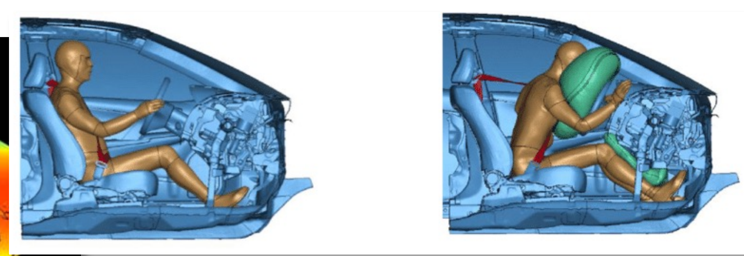


Climate Prediction

Optimization

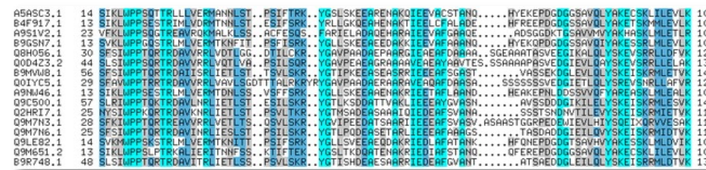


Airflow Optimization



Finite Element Simulation
Hyper Parameter Optimization

Today: Analytics, AI

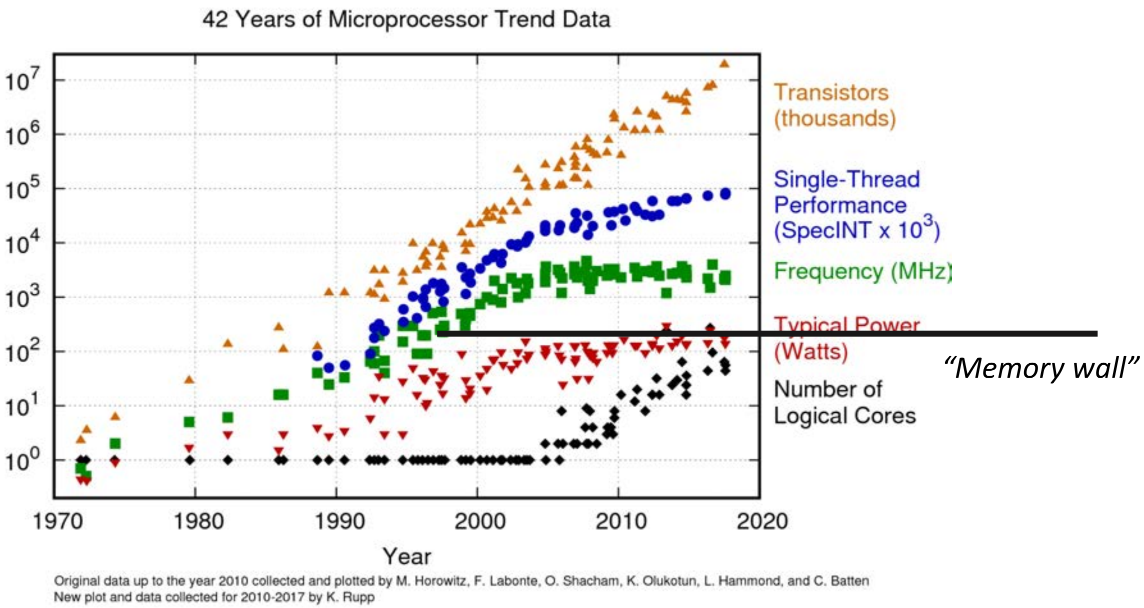


DNA Sequence Analysis
(e.g., Genomic sequencing of SARS-CoV-2)

Social Network Analysis

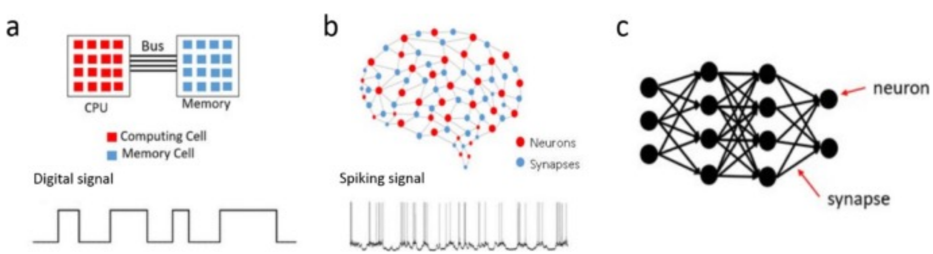
Recommendation Engines

Problem 1: Practical Limitations and new Challenges



- #cores per chip doubles every 18 months instead of clock
- CPU-memory communication is becoming a bottleneck
- Too much heat is produced
- As transistors get smaller, power density increases because these do no scale with size anymore

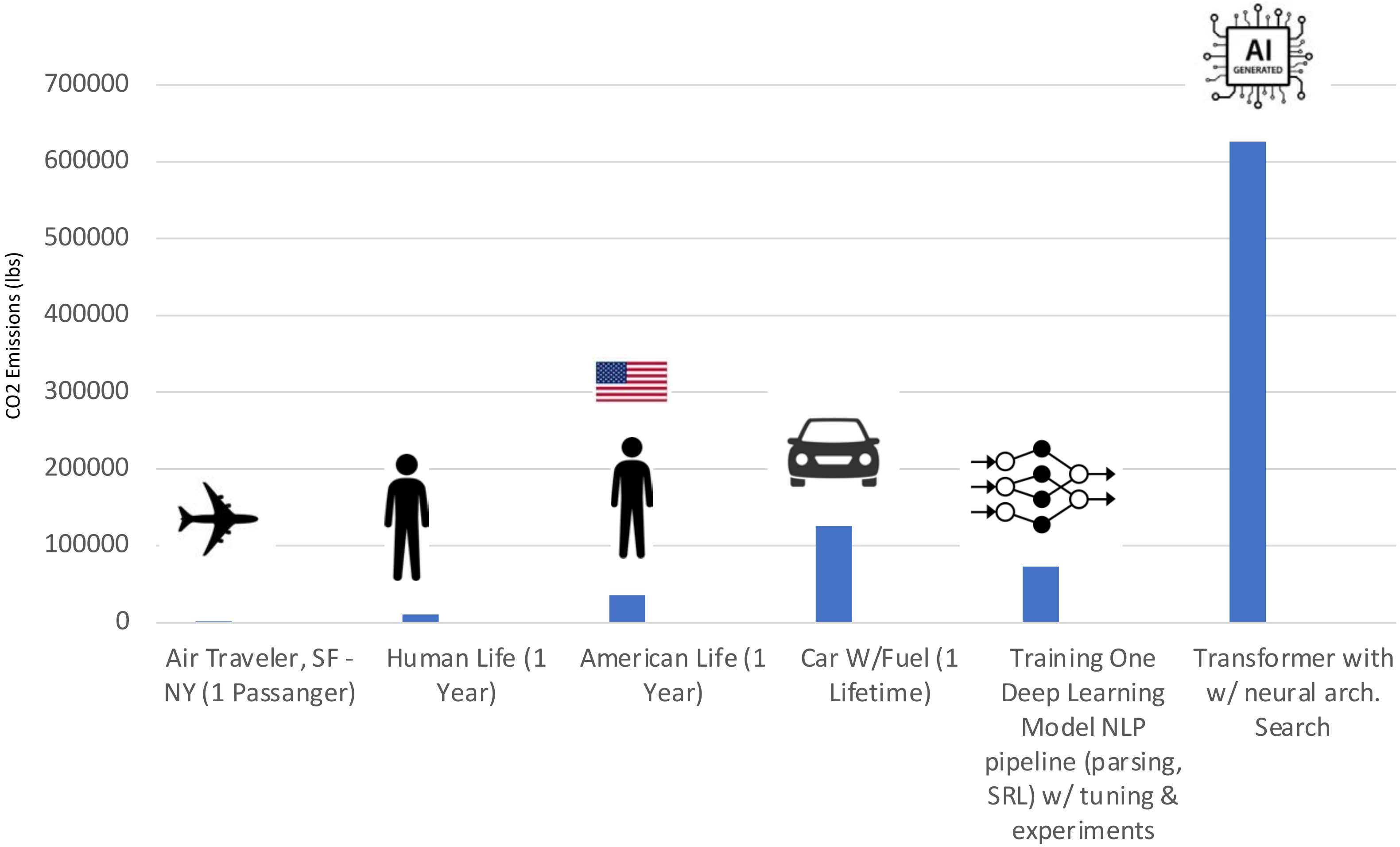
→ practical limitations to processor frequency to around 4 GHz since 2006



In Memory Computing, Neuromorphic Computing

Goi, E., Zhang, Q., Chen, X. et al. Perspective on photonic memristive neuromorphic computing. PhotoniX ***1,*** *3 (2020). https://doi.org/10.1186/s43074-020-0001-6*

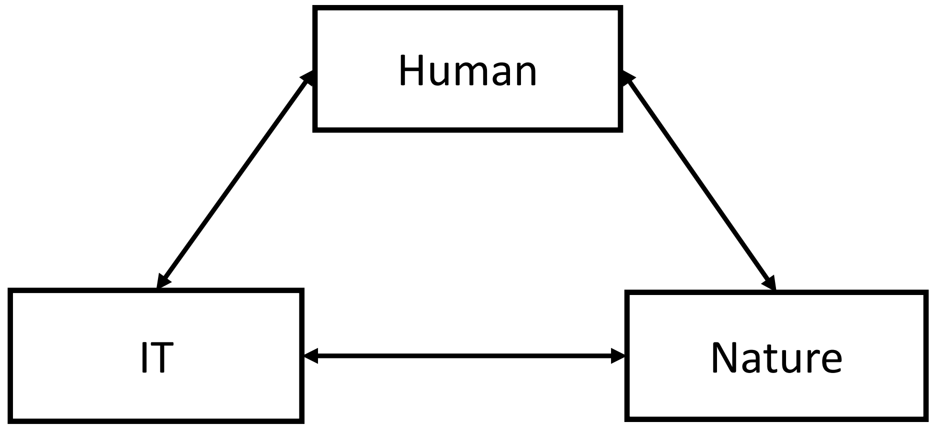
Problem 2: CO_2 Footprint of (generative) AI



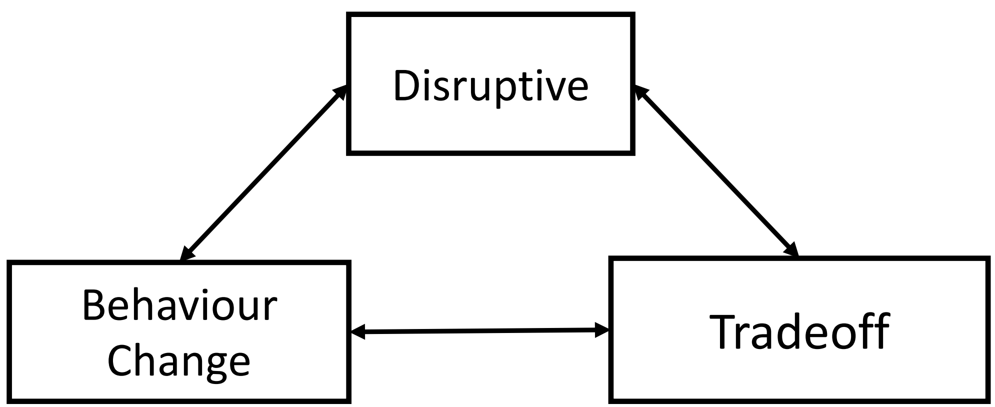
Source: Emma Strubell, Ananya Ganesh, Andrew McCallum: Energy and Policy Considerations for Deep Learning in NLP, ACL (1) 2019: 3645-3650
Inspiration for Visualisation: Keren Bergman, Multicore World 2023 https://2019multicoreworld.files.wordpress.com/2023/02/bergman-keren-23.pdf

Computational Sustainability

Actors:

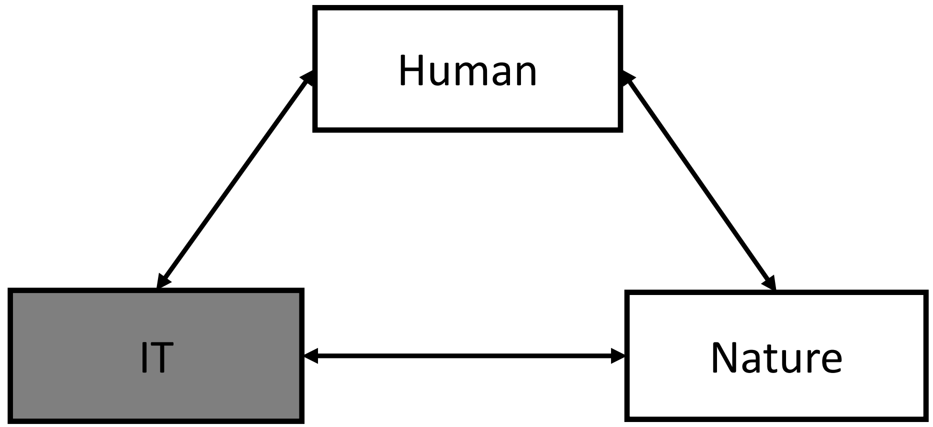


Methods:

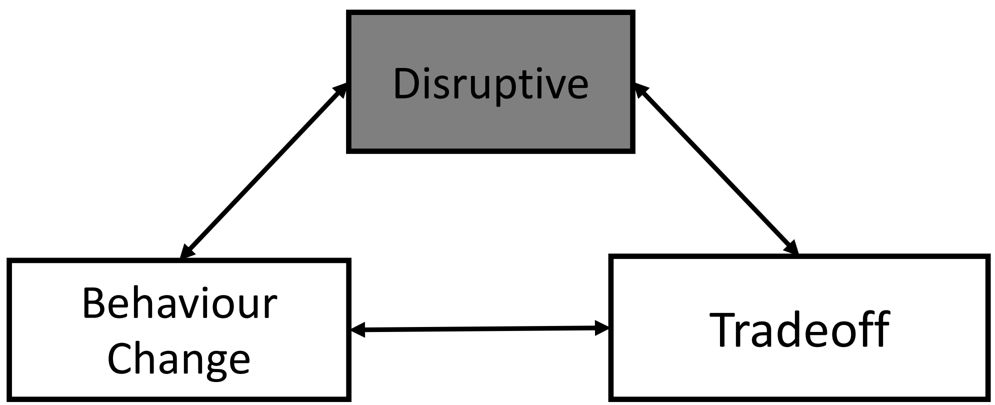


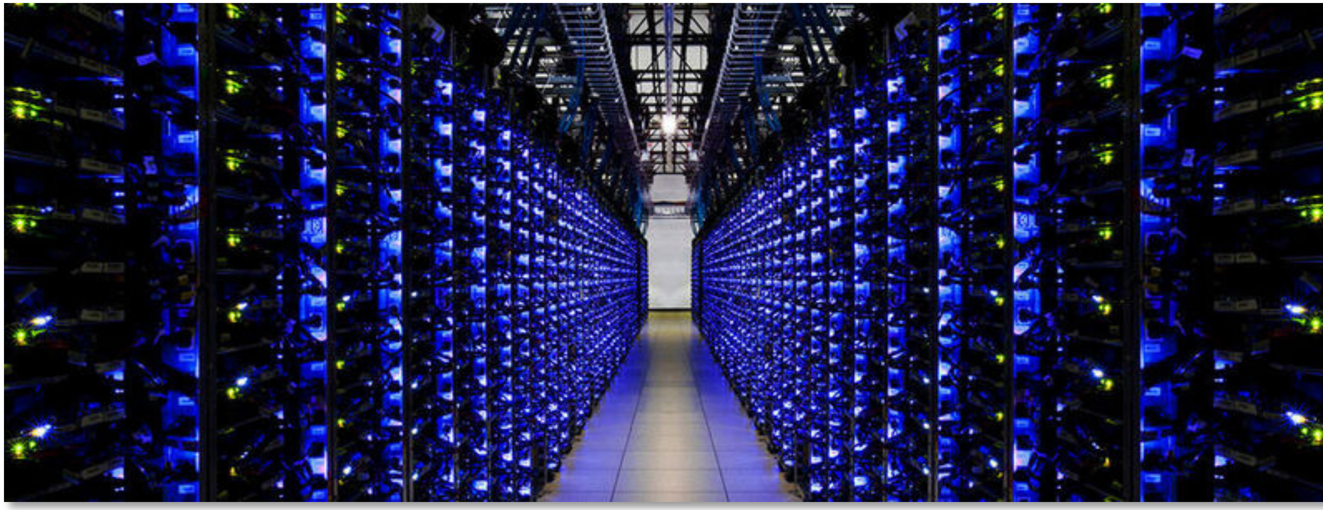
Computational Sustainability

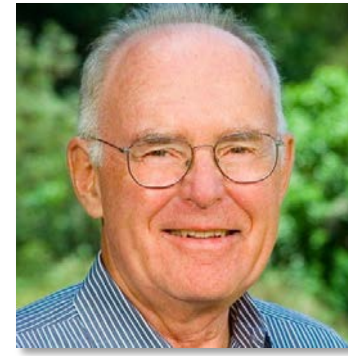
Actors:



Methods:







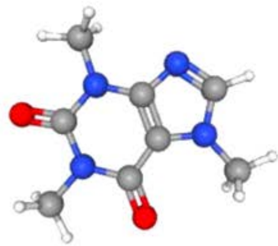
Gordon Moore:
Moore's Law
(1929 - 2023)

Data volumes are growing faster than the processing power

Alternatives:

- Neuromorphic Computing
- Quantum Computing

A cup of coffee?



- Representing the energy configuration of a single caffeine molecule at a single instant requires approximately 10^{48} bits in a classical computer
- Can be done using 160 logical qubits on a quantum machine

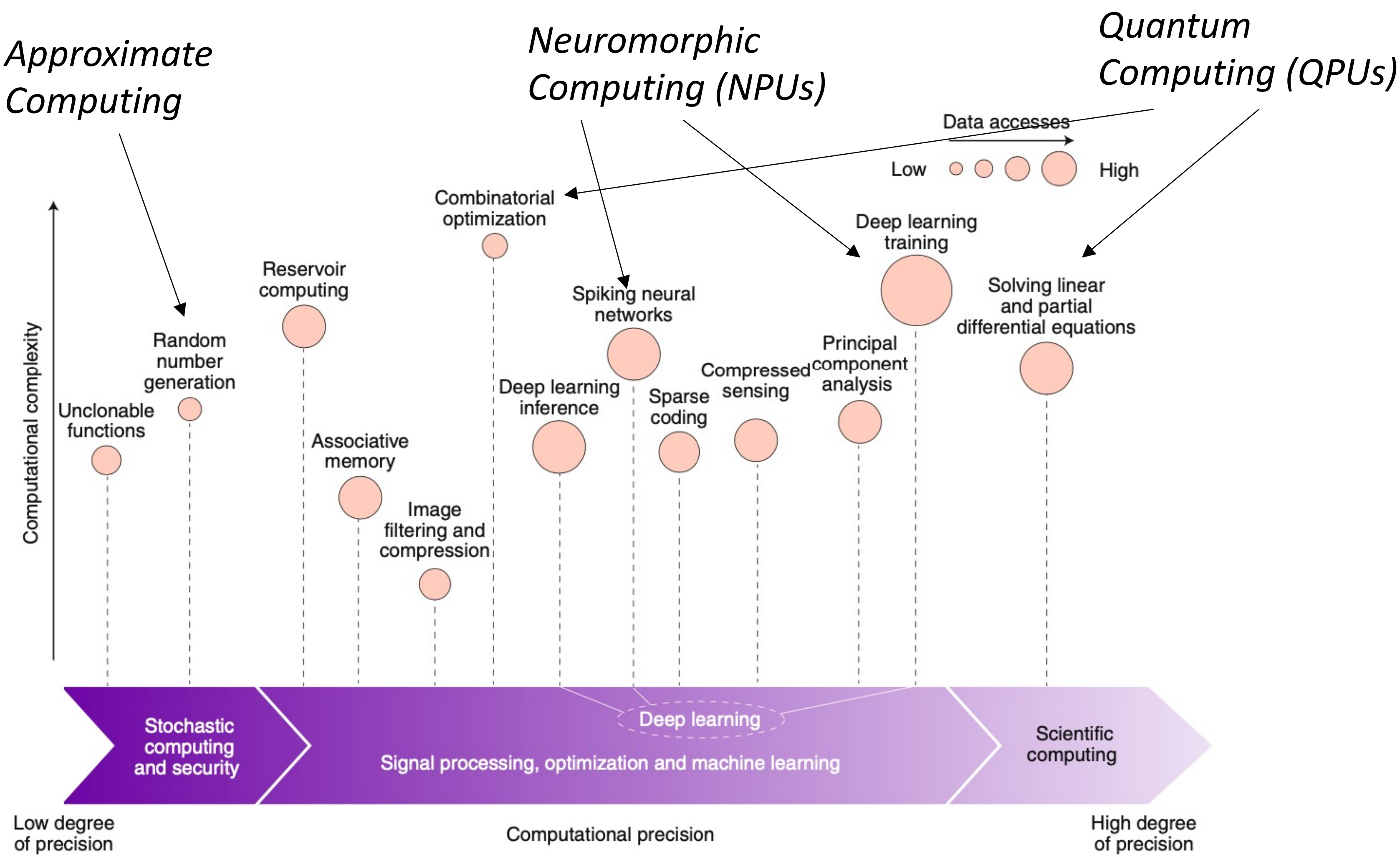
“Every time you add a qubit, you double your possible outcomes,
With 20 qubits there are a million possible outcomes.
With 100 qubits, you have more possibilities than there are bits in all the hard drives in the world. With 300 qubits—that’s more possibilities than there are particles in the universe.”

https://quantum.duke.edu/2020/10/16/more-possibilities-than-there-are-particles-in-the-universe/

1 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 vs. 160

Courtesy Nico Einsiedler (IBM), TU Vienna Quantum Fall Fest 2022

Future: Hyper-Heterogeneous Architectures



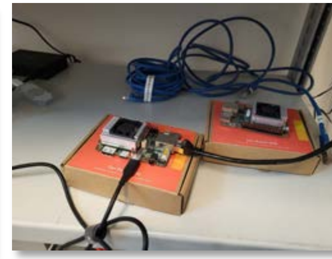
Hyper-heterogeneity, heavy geographical distribution

- ***More time is spent on communication!***

Source: Sebastian, A., Le Gallo, M., Khaddam-Aljameh, R. et al. Memory devices and applications for in-memory computing. Nat. Nanotechnol. ***15,*** *529–544 (2020). https://doi.org/10.1038/s41565-020-0655-z*

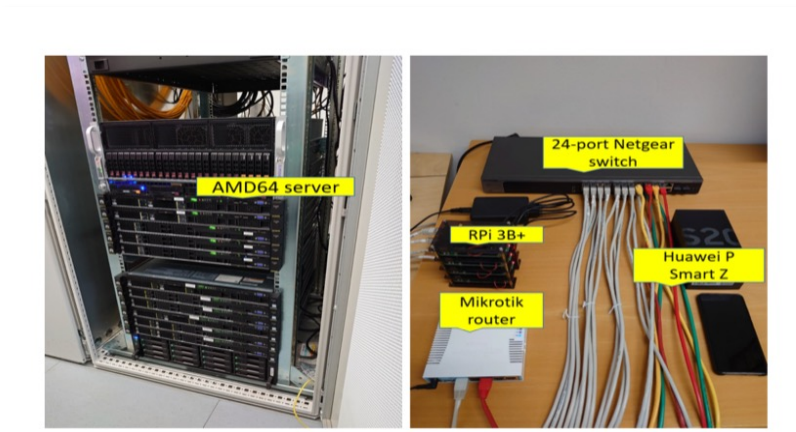
HPC lab @ TU Wien

Edge Testbed



1-tier Edge nodes

2-tier Edge nodes

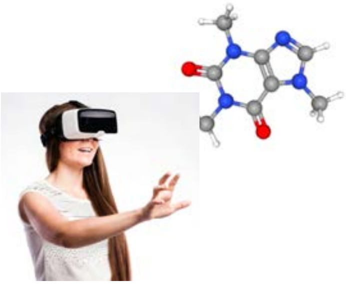


Mobile Testbed

Energy measurement infrastructure

RasQberry (courtesy IBM)

Computational Continuum



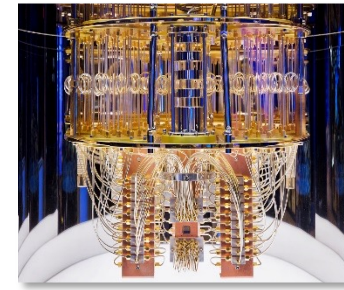
Measurement interpretation

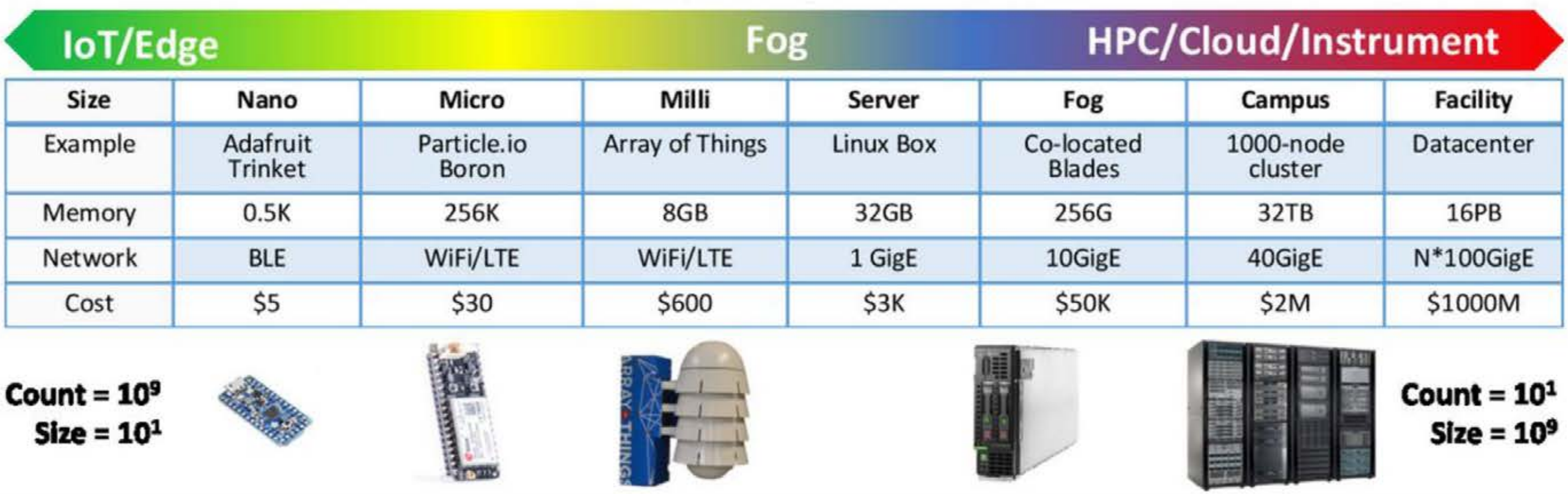
Non-determinism

...

Transpilation

Data transforamtion





Size	Nano	Micro	Milli	Server	Fog	Campus	Facility
Example	Adafruit Trinket	Particle.io Boron	Array of Things	Linux Box	Co-located Blades	1000-node cluster	Datacenter
Memory	0.5K	256K	8GB	32GB	256G	32TB	16PB
Network	BLE	WiFi/LTE	WiFi/LTE	1 GigE	10GigE	40GigE	N*100GigE
Cost	$5	$30	$600	$3K	$50K	$2M	$1000M

Source: Beckman, P.H., Dongarra, J.J., Ferrier, N.J., Fox, G., Moore, T.L., Reed, D.A., & Beck, M. (2020). Harnessing the Computing Continuum for Programming Our World.

Known Quantum Speedup

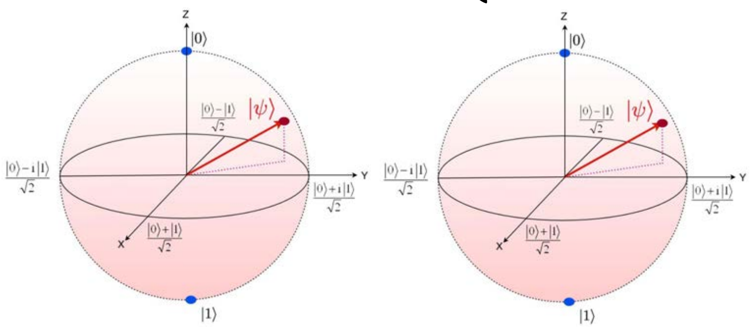
- **Grover's algorithm** (unstructured search): $O(\sqrt{n})$ vs $O(n)$, developed 1996
- **Shor's algorithm** (finding the prime factors in integer): Polynomial vs Exponential, developed 1994
- Quantum ML
  - Bayesian Inference: quadratic
  - SVM: exponential
  - Reinforcement Learning: quadratic
- In reality
  - Lack of standardization
  - Data transformation / quantum state preparation
  - Decoherence
  - Noise

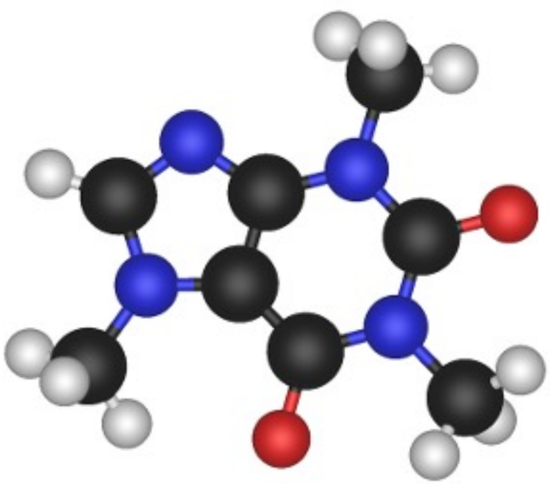
Source: Machine Learning: Quantum vs Classical, Tariq M. Khan et al., IEEE Access, November 2020

Beyond 0 and 1

Von Neumann

Quantum





Bottom up approach

- *Variational Quantum Linear Solver (VQLS)*
- *Quantum Eigenvalues* –> **Native 3d modeling of scientific applications**

Problem: Currently quantum systems can be used by quantum researchers only!

Quantum Computing

- **Computational basis**

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

- **Superposition**

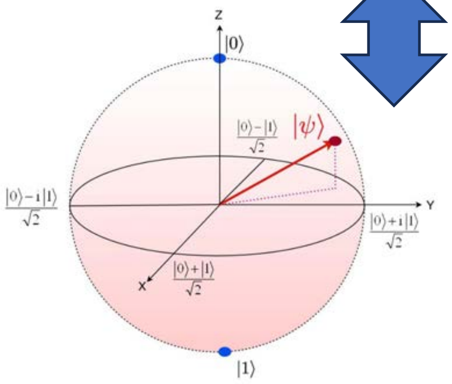
$$|\phi\rangle = c_1|0\rangle + c_2|1\rangle \quad c_1, c_2 \in \mathbb{C} \quad |c_1|^2 + |c_2|^2 = 1$$

$$|\psi\rangle_2 = c_1|00\rangle + c_2|01\rangle + c_3|10\rangle + c_4|11\rangle \quad |c_1|^2 + |c_2|^2 + |c_3|^2 + |c_4|^2 = 1$$

- **Entanglement**
  - Theoretical speedup
  - Space efficiency
  - Quantum advantage

Data Encoding

- Data coming from "classic" world have to be encoded in a quantum state
- Important to decide whether using quantum systems can bring benefits
- The number of operations to prepare the quantum state must be small as qubits decay fast and quantum gates are error-prone
- Trade-off between two major forces:
  - the number of required qubits and
  - the runtime complexity for the loading process
- Data encoding can "kill" quantum speedup!
  - Worst case exponential time
  - Low latency
  - Data streaming
  - Qubits decay over time
  - You cannot copy qubits state

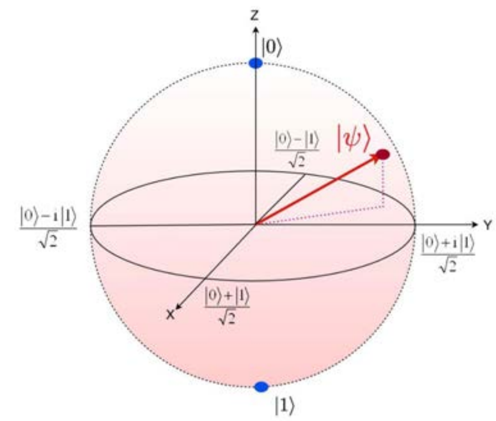


Data Encoding Examples

- Basis encoding (**digital encoding**): $x = 1011 \rightarrow |1011\rangle$ (n bit -> n qubits)

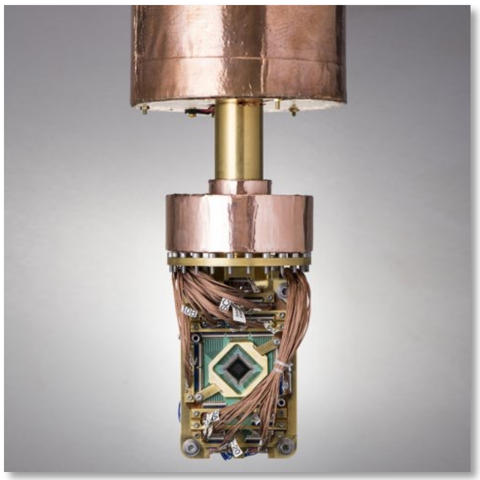
$$5 \rightarrow 101 \rightarrow |1\rangle|0\rangle|1\rangle \rightarrow |101\rangle$$

- Amplitude encoding (**analog encoding**):
  $|x\rangle = \sum_i^N x_i \, |i\rangle$ (n bits -> $\lceil \log_2(N) \rceil$ qubits)



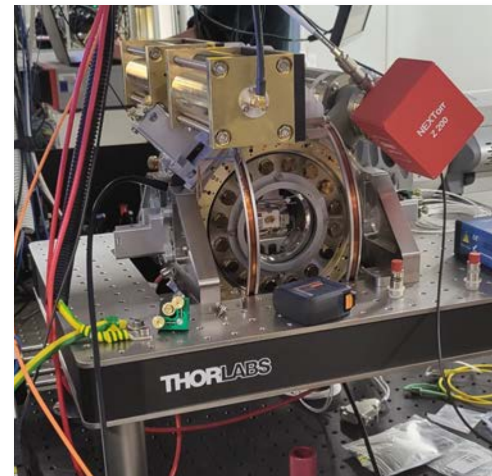
- Hamiltonian Evolution Ansatz Encoding
- …

Quantum Hardware



D-WAVE

SUPERCONDUCTING



TRAPPED ION*

- No “best” technology at the moment
- No standards
- Every machine different architecture
- Integration →

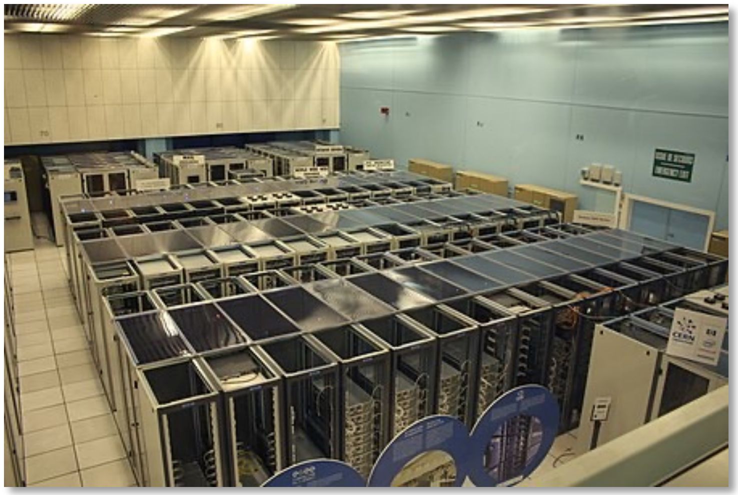
**Courtesy of University of Innsbruck, department of experimental physics*

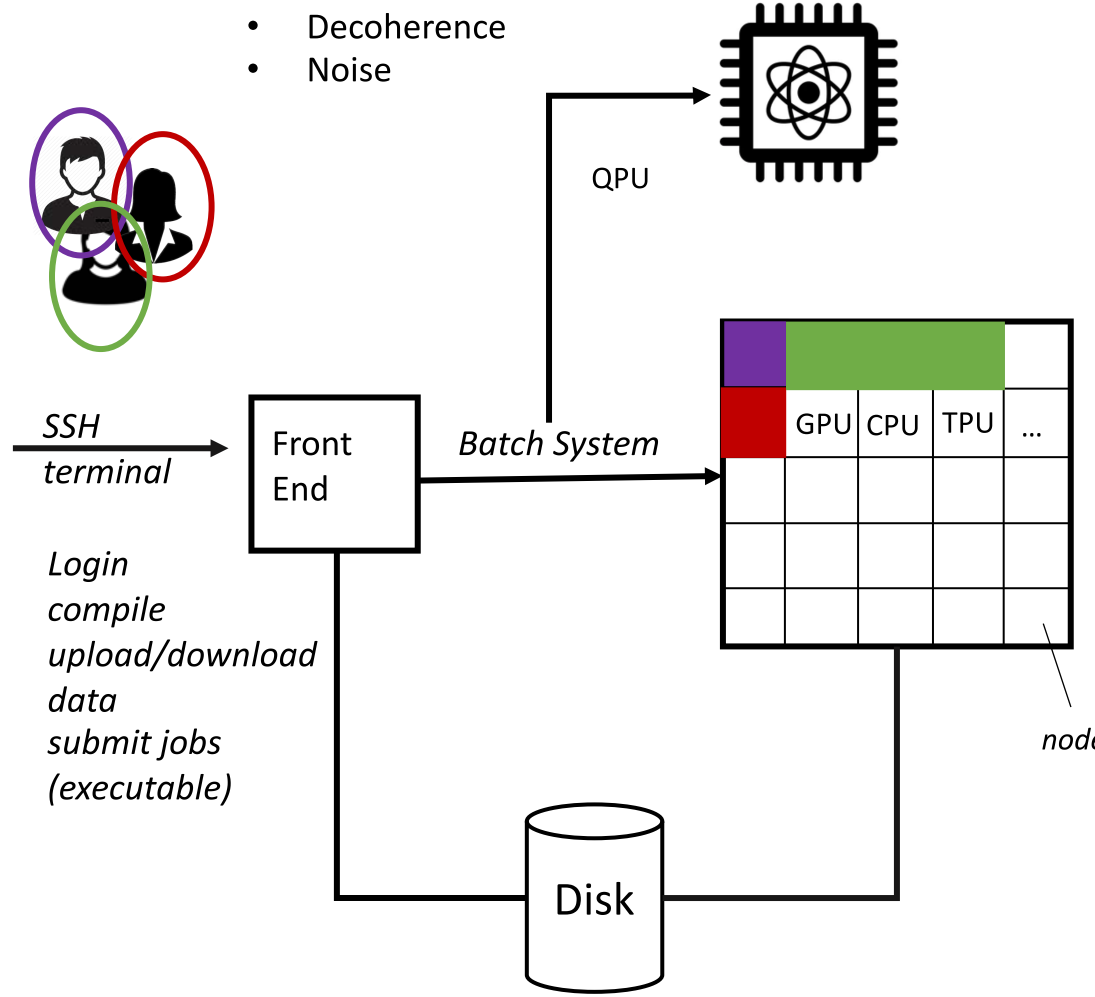
FFG Flagship Project High Performance Integrated Quantum Computing (HPQC)

HPQC
HIGH-PERFORMANCE QUANTUM COMPUTING

HPC Clusters

- Lack of standardization
- Data transformation / quantum state preparation
- Decoherence
- Noise

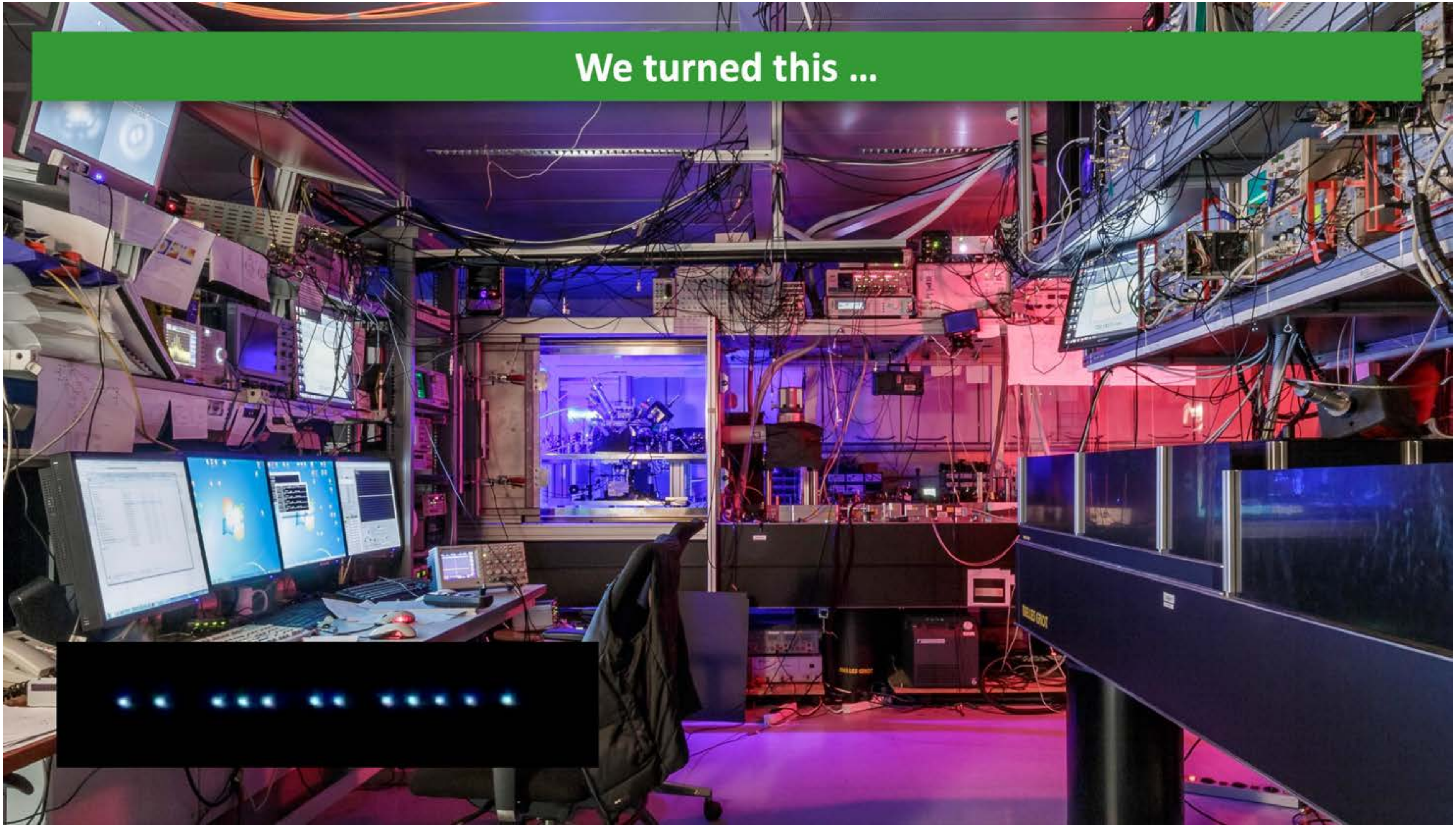




- Each “node” has its own operating system
- Nodes are interconnected with a network cable
- Higher performance demand more processors
- Accessed via front-end node/computer
- Shared with may users

HPQC
HIGH-PERFORMANCE QUANTUM COMPUTING

We turned this ...



Slide: courtesy Thomas Monz, Uni Innsbruck & AQT

AQT

... into this

AQT DEMONSTRATED:

- 50+ ion-qubits
- 24-qubit entanglement
- Shor's algorithm
- Quantum Error Correction
- Fault-tolerant performance
- Demo'd finance applications
- Demo'd security applications
- Demo'd chemistry applications
- ...

WITH OUR SYSTEM BEING:

- Rack-mounted
- Cloud-accessible
- Data-center compatible

Slide: courtesy Thomas Monz, Uni Innsbruck & AQT

HPQC Cluster

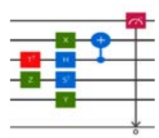
- Improve QC performance
- Complement QEC capabilities
- Extend QC research with HPC

- Implement Hybrid Q-Libs
- Benchmark PoC use-cases in hybrid infrastructure

Direct access

Login console

HPC node



Circuit API

QC system

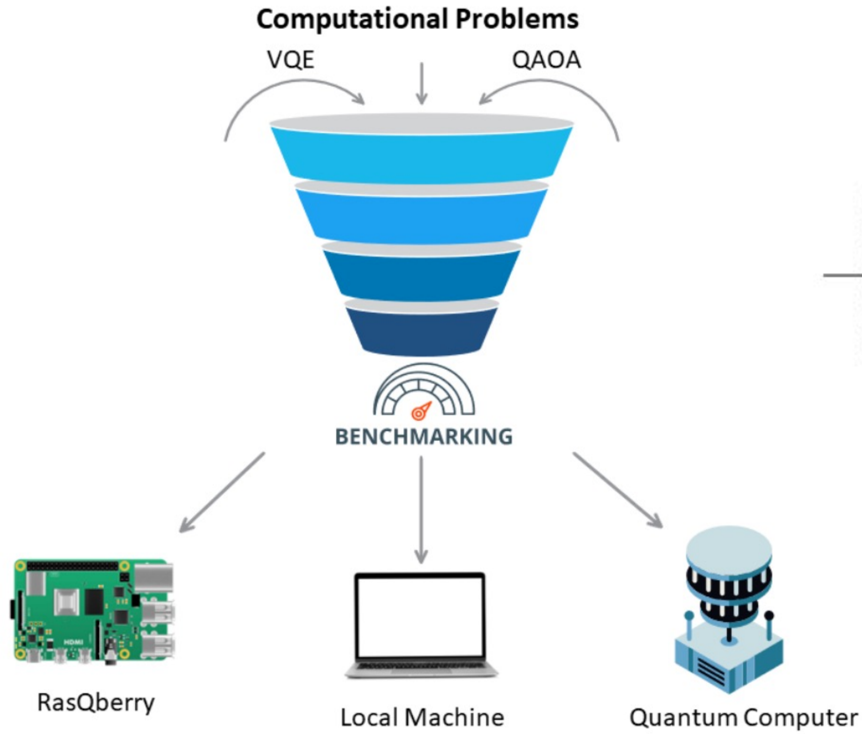
Offloading API

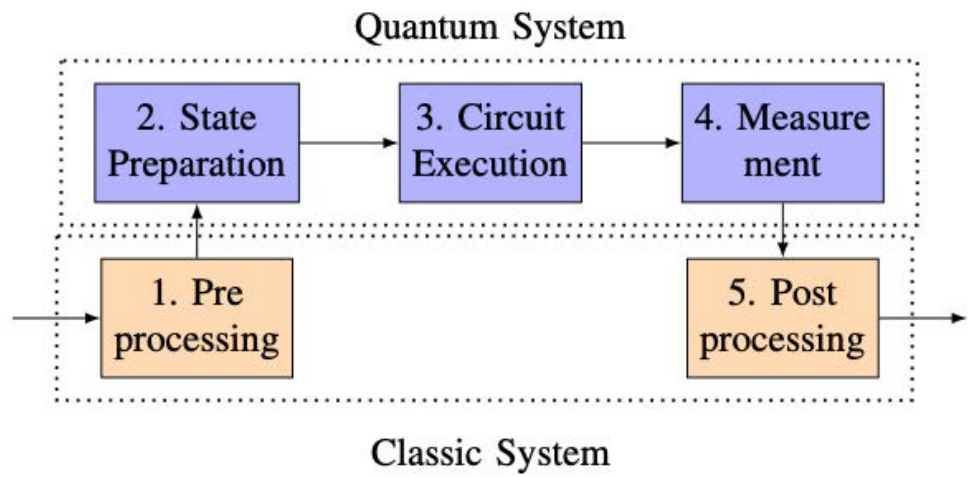
"string": "Hi",
"number": 2.5,
"boolean": true,
"null": null,
"object": { "name": "Kyle", "age": 24 },
"array": ["Hello", 5, false, null, { "key":
"arrayOfObjects": [
{ "name": "Jerry", "age": 28 },
{ "name": "Sally", "age": 26 }
]

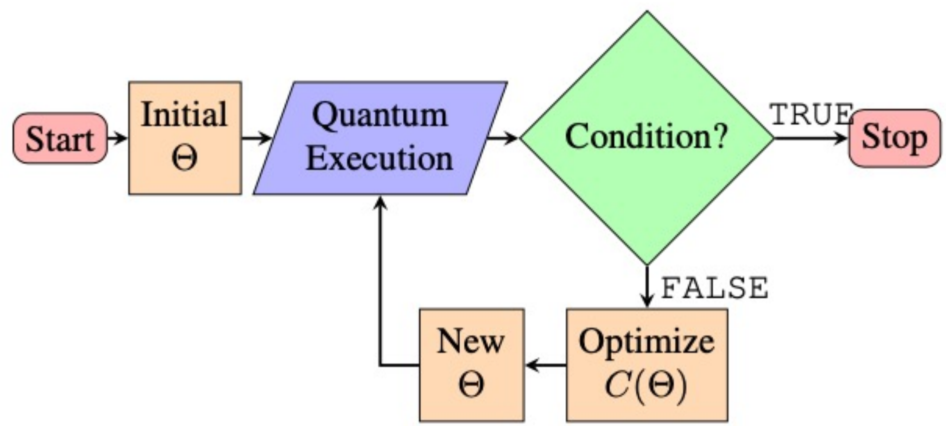
aconet

AQT JKU JOHANNES KEPLER UNIVERSITÄT LINZ universität innsbruck IBM MATH+TEC TU WIEN FFG Forschung wirkt. HPQC HIGH-PERFORMANCE QUANTUM COMPUTING

Hybrid Classic/Quantum Benchmarking







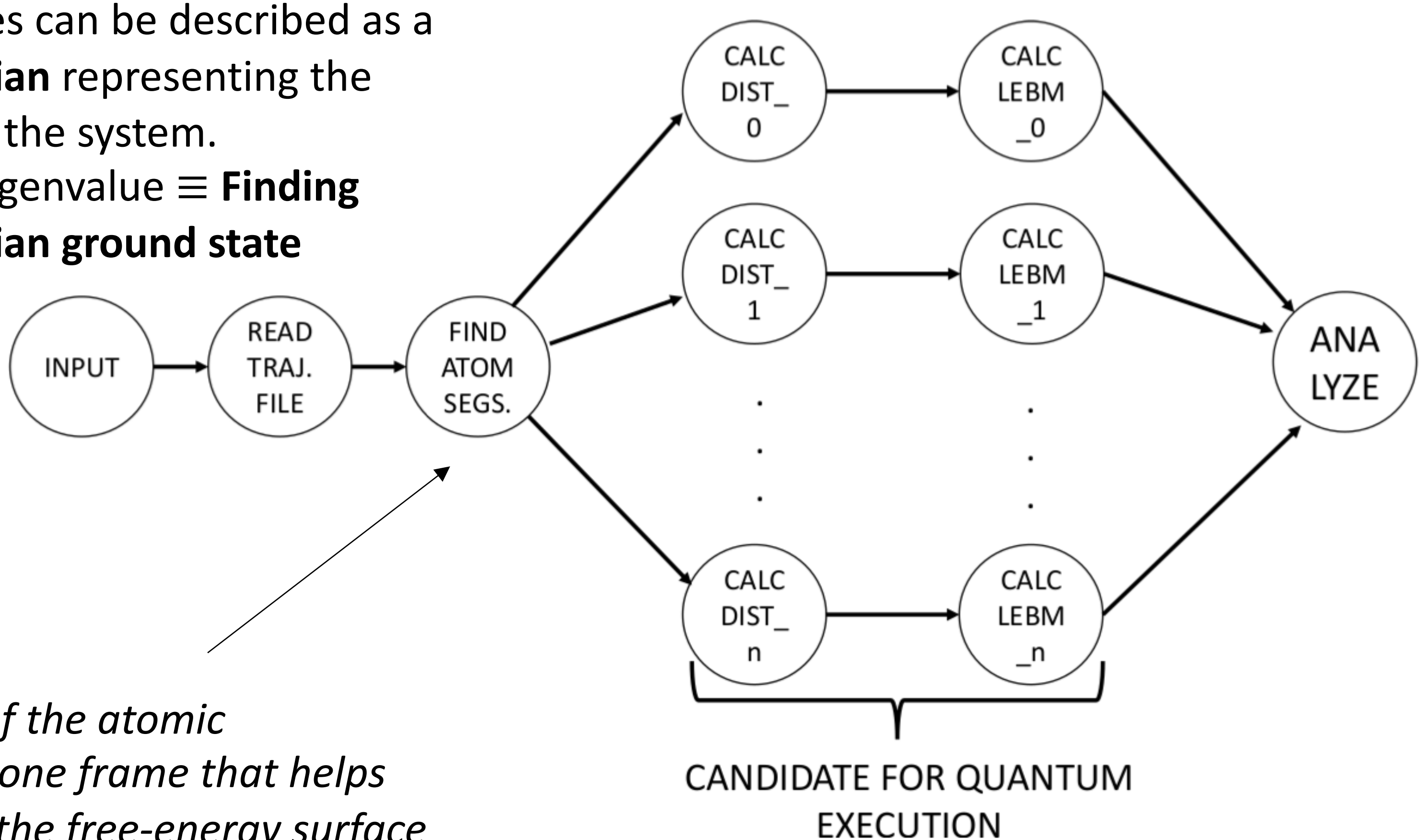
Variational Quantum Algorithm

Teaching purposes:
- Simulators first
- Mature prototypes on the quantum machines

Source: courtesy Dominik Jandl (TU Wien)

Benchmarking Molecular Dynamic

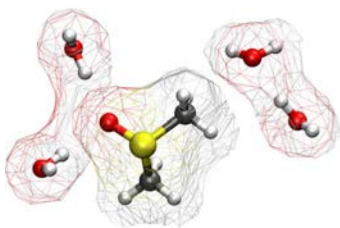
- In quantum mechanics, a system of particles can be described as a **Hamiltonian** representing the energy of the system.
- Finding eigenvalue ≡ **Finding Hamiltonian ground state**



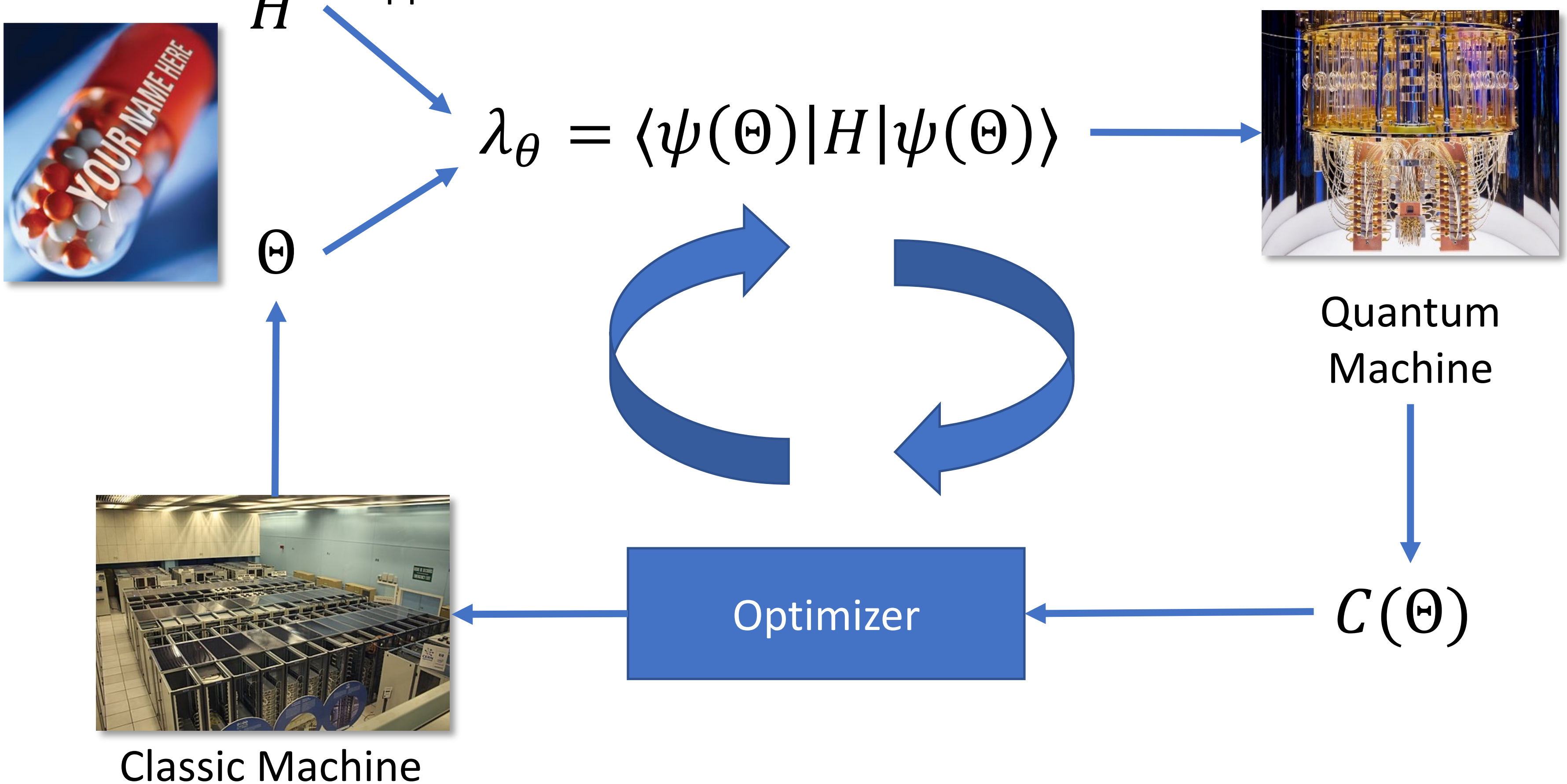
*CVs: function of the atomic coordinates in one frame that helps to reconstruct the free-energy surface for enhanced sampling, e.g.**, distance between two atoms, largest eigenvalues of bipartite matrix (LEBM)***

Source: Sandeep Suresh Cranganore, Vincenzo De Maio, Ivona Brandic, Tu Mai Anh Do, Ewa Deelman. Molecular Dynamics Workflow Decomposition for Hybrid Classic/Quantum Systems. IEEE eScience 2022, October 11-14, 2022 Salt Lake City, Utah, USA.

Benchmarking on Hybrid Systems



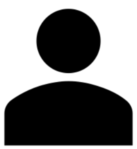
Molecular Dynamic Application



Source: Sandeep Suresh Cranganore, Vincenzo De Maio, Ivona Brandic, Tu Mai Anh Do, Ewa Deelman. Molecular Dynamics Workflow Decomposition for Hybrid Classic/Quantum Systems. IEEE eScience 2022, October 11-14, 2022 Salt Lake City, Utah, USA.

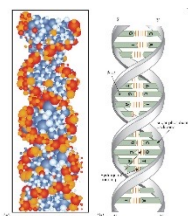
Molecular Dynamics

- Analyzing trajectories of backbone $C\alpha$ atoms of amino-acids segments
- Identifying collective variables capturing molecular motions in a region of interest



User input

Read trajectory file



Atom segments

$$D = \begin{bmatrix} 0 & \cdots & D_{IJ} \\ \vdots & \ddots & \vdots \\ D_{IJ}^T & \cdots & 0 \end{bmatrix}$$

Distance matrix

$$Dv = \lambda v$$

Find maximum eigenvalue

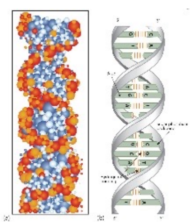
Question: are there application parts that could benefit from quantum execution?

Quantum Decomposition

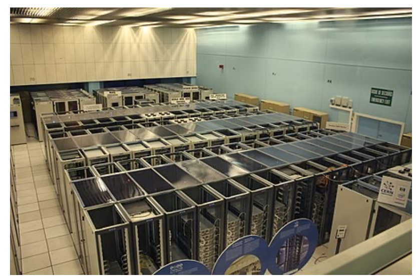
User input

Read trajectory file

End Device



Atom segments



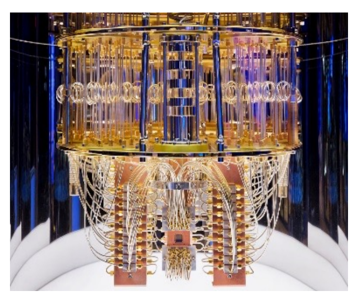
Classic HPC

$$\begin{bmatrix} 0 & \cdots & D_{IJ} \\ \vdots & \ddots & \vdots \\ D_{IJ}^T & \cdots & 0 \end{bmatrix}$$

Distance matrix

$$Dv = \lambda v$$

Find largest eigenvalue



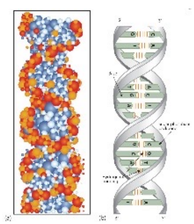
Quantum machine

Quantum Decomposition

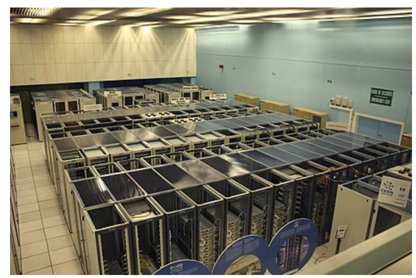
User input

Read trajectory file

End Device



Atom segments



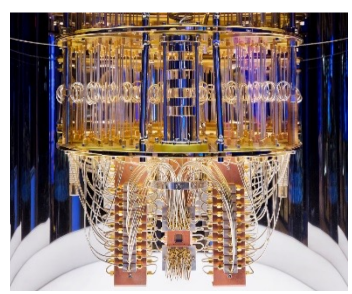
Classic HPC

$$\begin{bmatrix} 0 & \cdots & D_{IJ} \\ \vdots & \ddots & \vdots \\ D_{IJ}^T & \cdots & 0 \end{bmatrix}$$

Distance matrix

$$Dv = \lambda v$$

Find largest eigenvalue



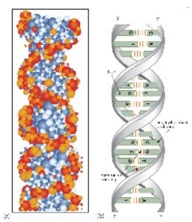
Quantum machine

Quantum Decomposition

User input

Read trajectory file

End Device



Atom segments

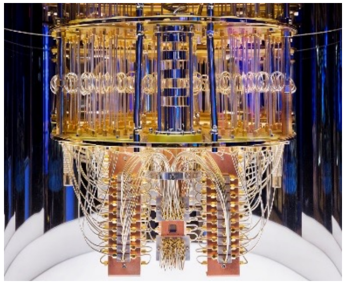
Classic HPC

$$\begin{bmatrix} 0 & \cdots & D_{IJ} \\ \vdots & \ddots & \vdots \\ D_{IJ}^T & \cdots & 0 \end{bmatrix}$$

Distance matrix

$$Dv = \lambda v$$

Find largest eigenvalue



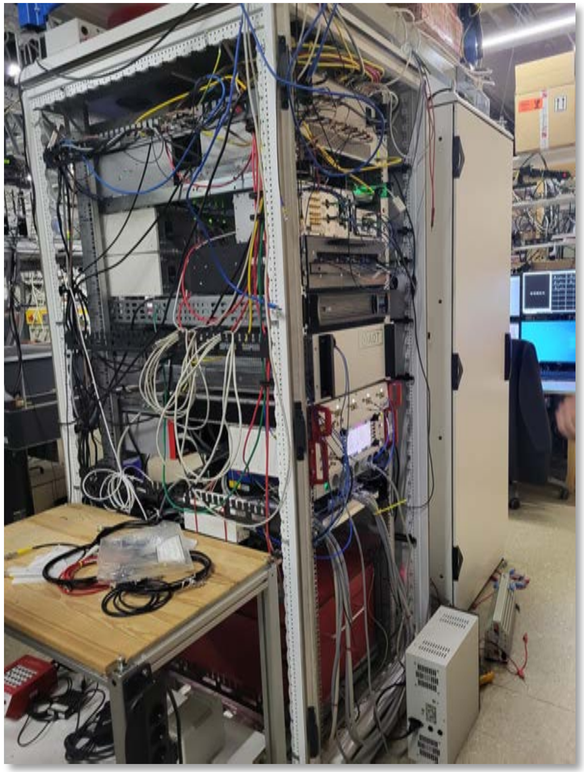
Quantum machine

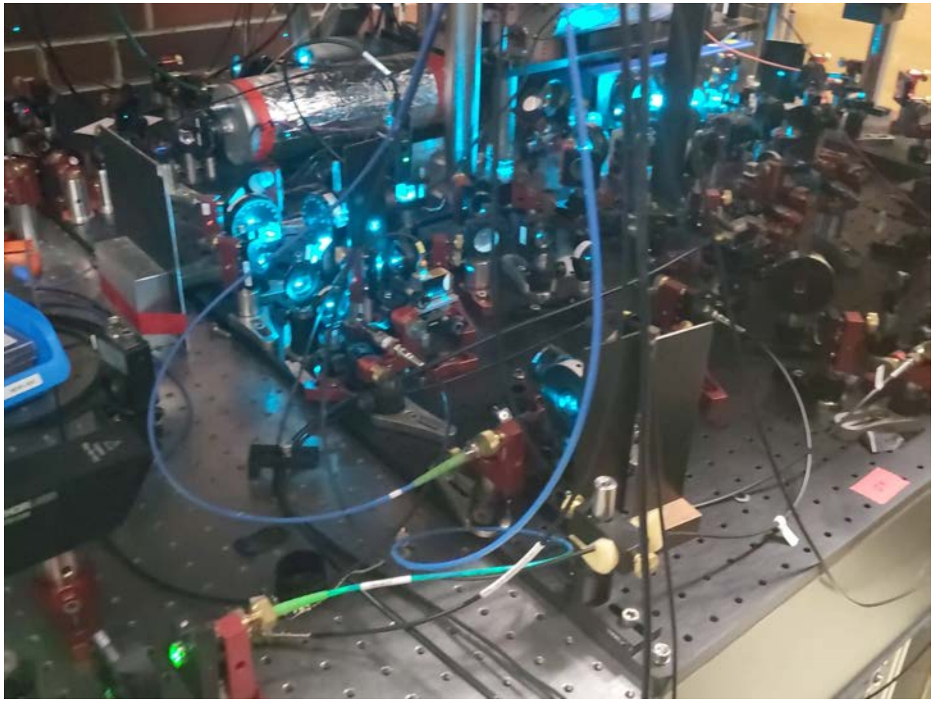
Mechanical Installation of HPQC Cluster

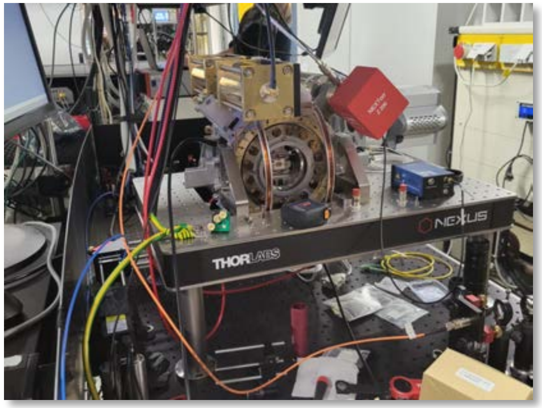
- Move from Ca+ to Ba+
- New system with
  stage 1: 10x higher T_1
  stage 2: infinit. Higher T_1
- 2q error rate:
  legacy: $< 10^{-2}$
  target: $< 10^{-3}$
- Init error
  legacy: $< 10^{-3}$
  target: $< 10^{-4}$
- Readout error
  legacy: $< 10^{-3}$ -> ~ 10^{-4}
  target: $< 10^{-5}$

HPQC HIGH-PERFORMANCE QUANTUM COMPUTING

Courtesy Experimentalphysik, Univ. Innsbruck







Ion-trap quantum computer

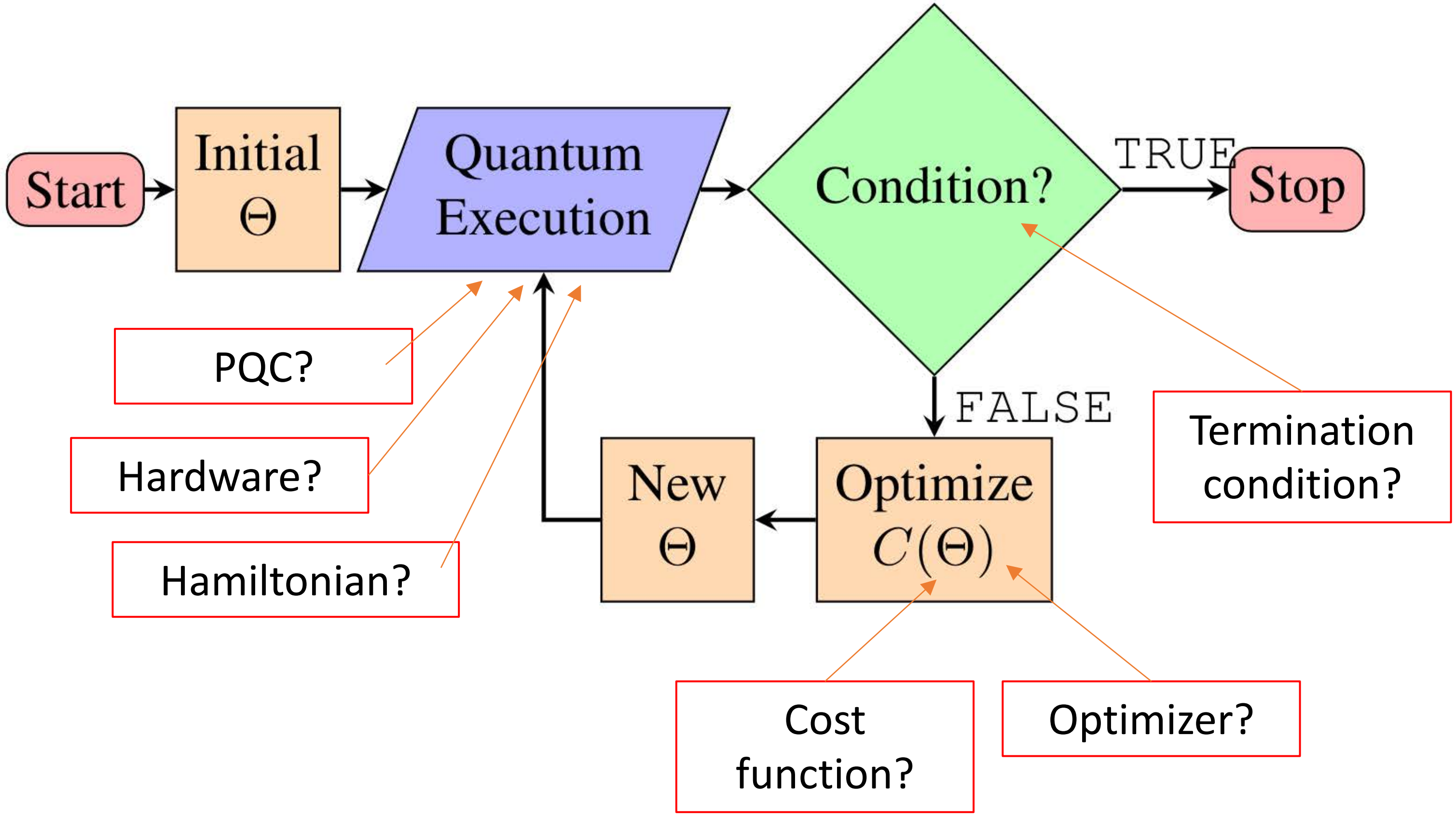
Courtesy Experimentalphysik, Univ. Innsbruck
Pictures, video: courtesy Vincenzo de Maio

universität innsbruck

HPQC
HIGH-PERFORMANCE QUANTUM COMPUTING

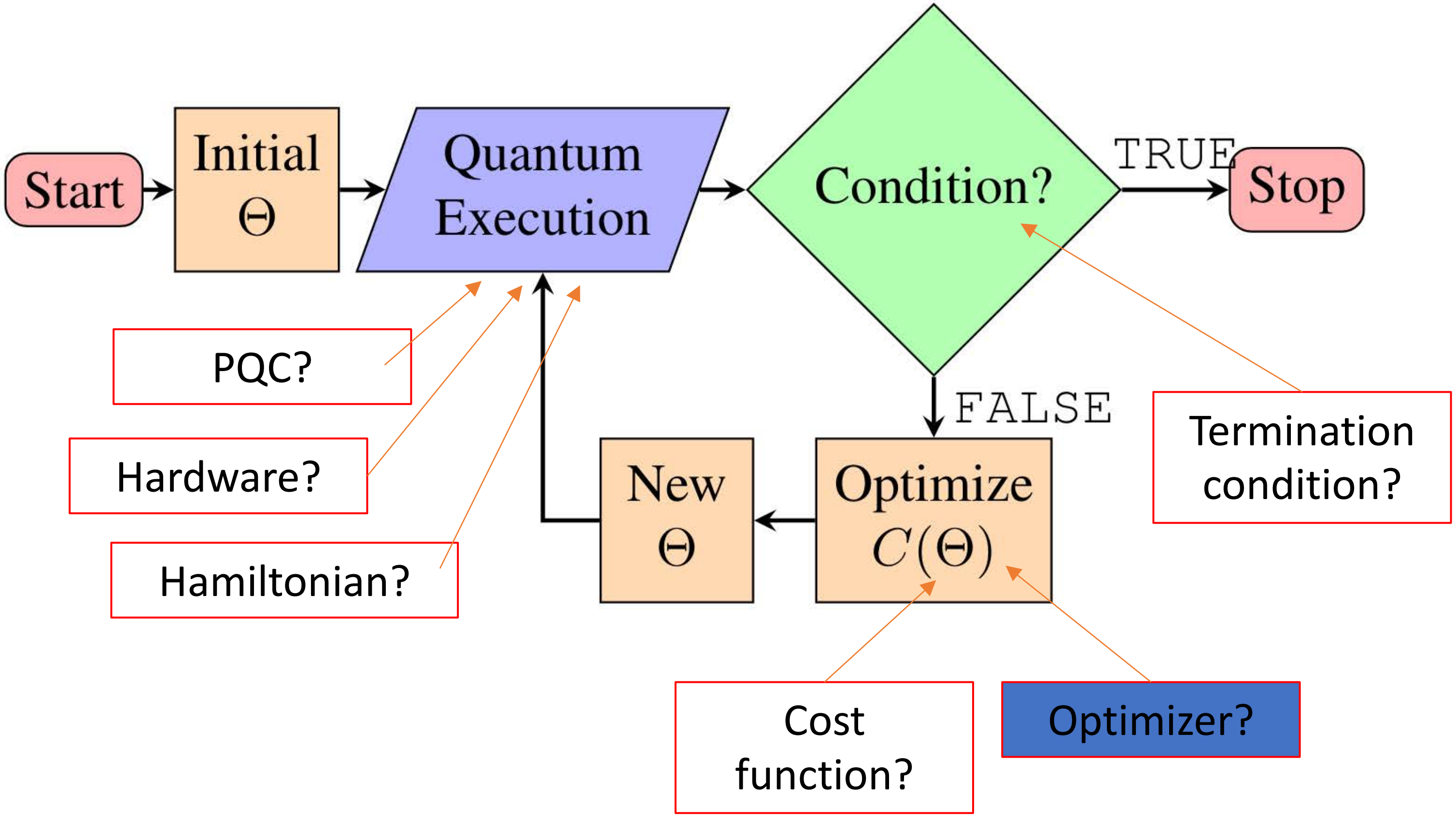
Benchmarking Variational Quantum Algorithms

- Defined by a set of **hyperparameters**
- Main candidates to achieve Quantum Supremacy (Cerezo et al., 2021)



Benchmarking Variational Quantum Algorithms

- Defined by a set of **hyperparameters**
- Main candidates to achieve Quantum Supremacy (Cerezo et al., 2021)



Calculation of LEBM (Least Eigenvalue Bipartite Matrix)

Classic maschine

Quntum maschine

$$Err(\hat{B}, \Pi) = \sum_{i \in [0, |\hat{B}|]} \frac{(\Lambda_c(B^i_{IJ}) - \Lambda_{vqe}(B^i_{IJ}, \Pi))^2}{|\hat{B}|}$$

Benchmarking the error to the execution on the classic machine

Π … can be PQC, optimizer, classic hardware node, etc.

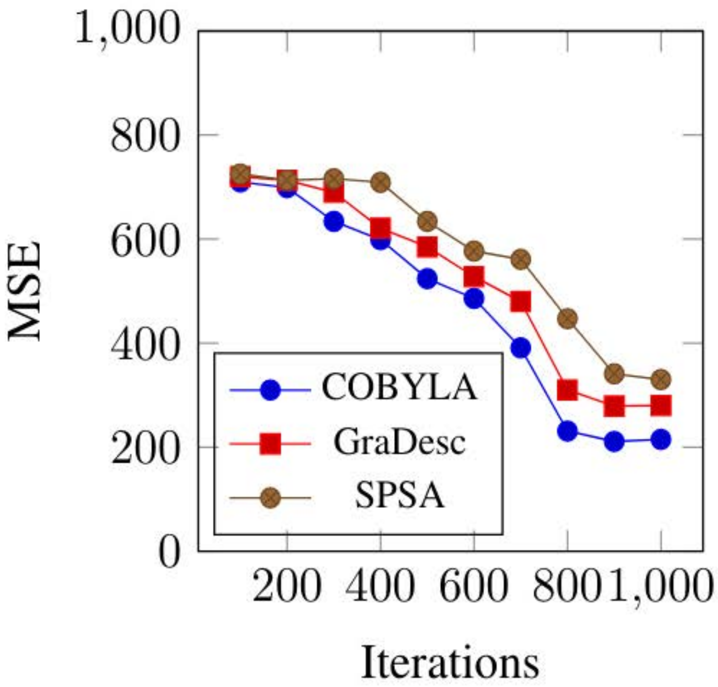
Π^*_{err} … Minimizing error compared to classic execution

PQC	Width	Repetitions	Entanglement
SU2	$[1,5]$	$[1,5]$	{full, linear, SCA, circular}
RealAmplitudes	$[1,5]$	$[1,5]$	{full, linear, SCA}
PauliTwo	$[1,5]$	$[1,5]$	{linear}
ExcitationPreserving	$[1,5]$	$[1,5]$	{full, linear, SCA}

Node id	Version	Processor	Qubits
ibmq_manila	1.0.29	Falcon r5.11L	5
ibmq_santiago	1.4.1	Falcon r4L	5
ibm_lagos	1.0.27	Falcon r5.11H	7
ibm_jakarta	1.0.33	Falcon r5.11H	7

Results

1. ***Constrained Optimization BY Linear Approximations (COBYLA):*** COBYLA performs linear approximations of both target and constraints function, aiming at optimizing a simplex within a trust region of the parameter space.
2. ***SPSA:*** SPSA is a stochastic optimization methods focusing on measurement of objective function.
3. ***Gradient Descent:*** Minimizes target objective function f by iteratively moving in the direction of steepest descent, defined by the negative of the gradient.

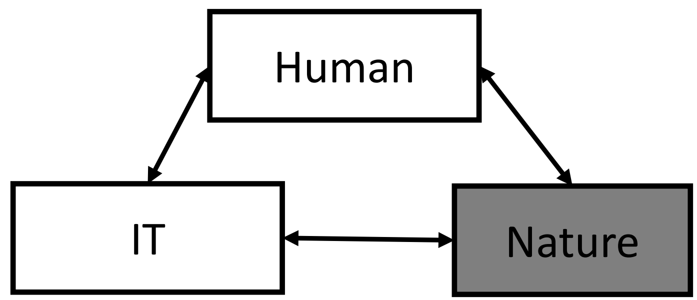


- VQE calculation using different hyperparameters
- Benchmarking data collected on different machines

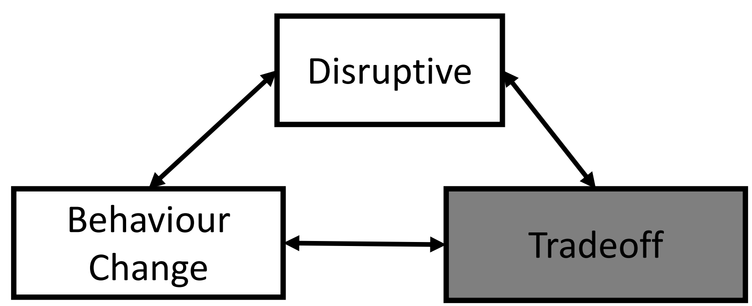
Source: Sandeep Suresh Cranganore, Vincenzo De Maio, Ivona Brandic, Tu Mai Anh Do, Ewa Deelman. Molecular Dynamics Workflow Decomposition for Hybrid Classic/Quantum Systems. IEEE eScience 2022, October 11-14, 2022 Salt Lake City, Utah, USA.

Computational Sustainability

Actors:

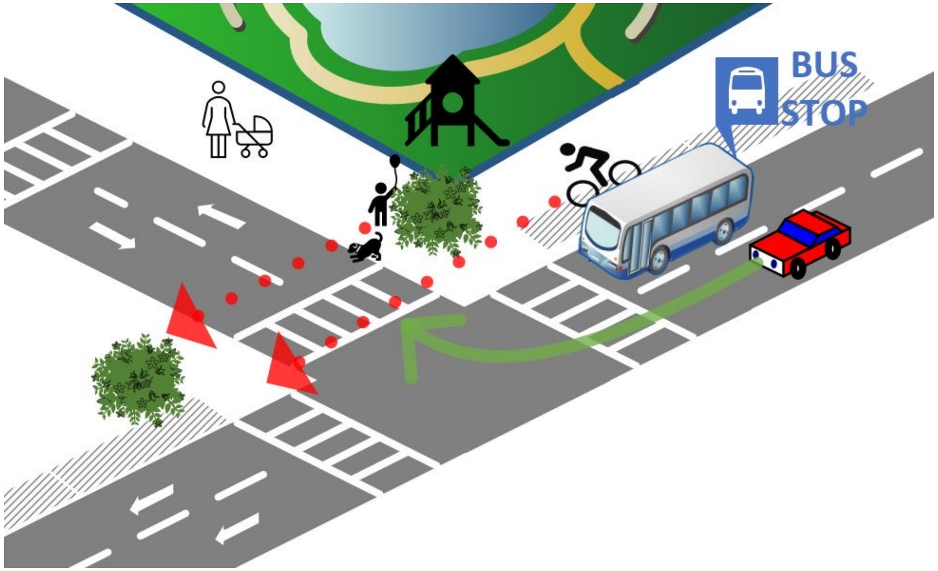


Methods:



Edge Computing in Action: Smart City

- Traffic accidents
  - causing injuries and deaths
- Distractions, poor visibility (e.g., bad road and weather conditions), …
- Drivers' brake reaction time
  - 1500ms on average



Deaths among pedestrians and cyclists:
29% of all EU road deaths
ETSC (European Transport Safety Council) PIN Flash Report 38

Slide: courtesy Ivan Lujic (Ericsson Nikola Tesla d.d.)

Contemporary Systems: Smart City

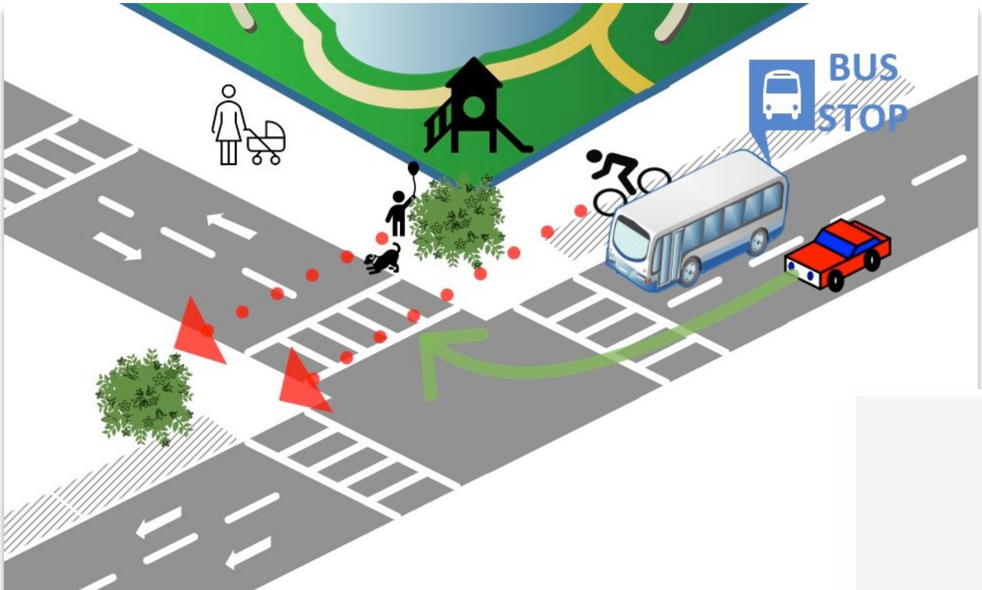


Fig 1: Smart Traffic System

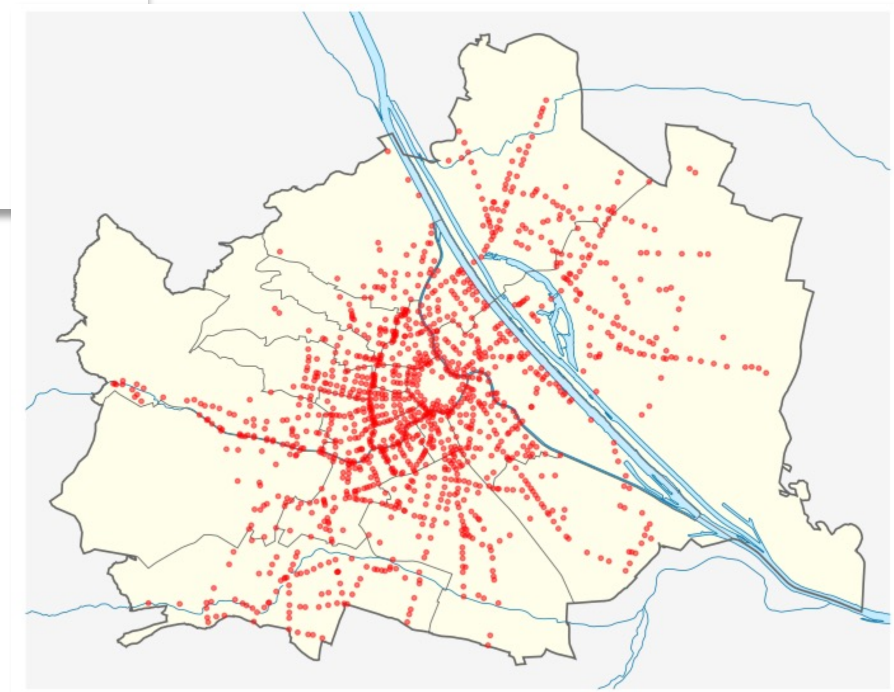


Fig 2: Distribution of Smart Traffic Lights in Vienna

Slide: courtesy Ivan Lujic (Ericsson Nikola Tesla d.d.)

Source: Vienna Municipal Department 33, "Traffic lights with/without audible signal devices in Vienna," https://www.data.gv.at/, 2019, OpenData Österreich.

Smart City I

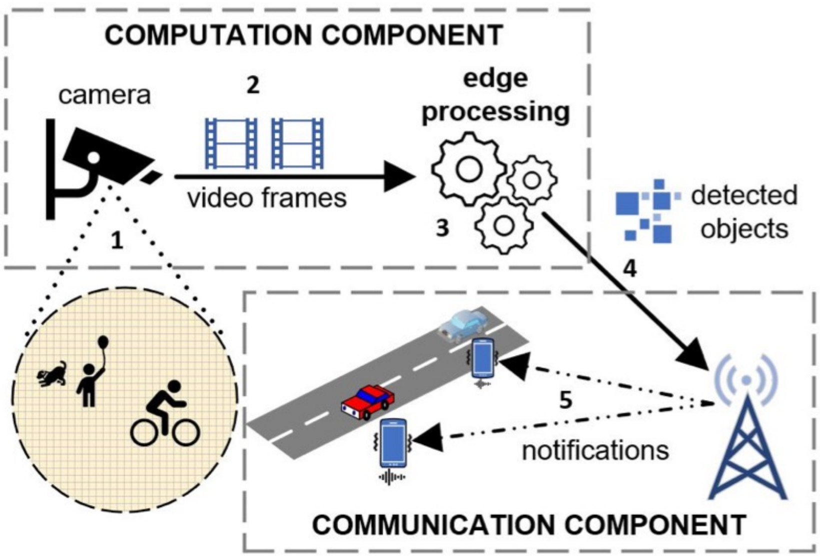


Fig 1. Architecture View

Fig 2 and 3: Intrasafed 5G installation

Fig 5: App

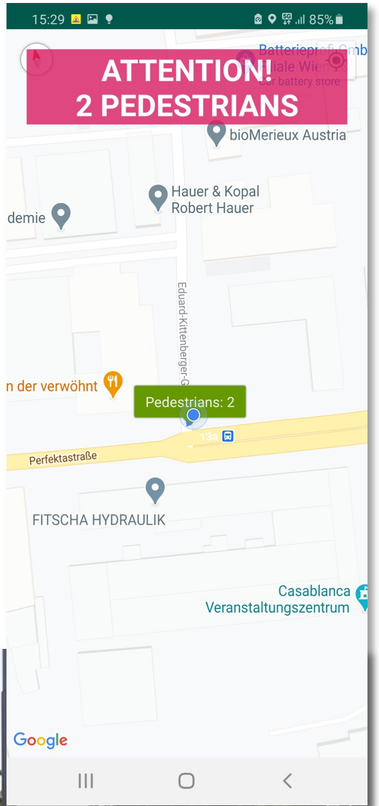
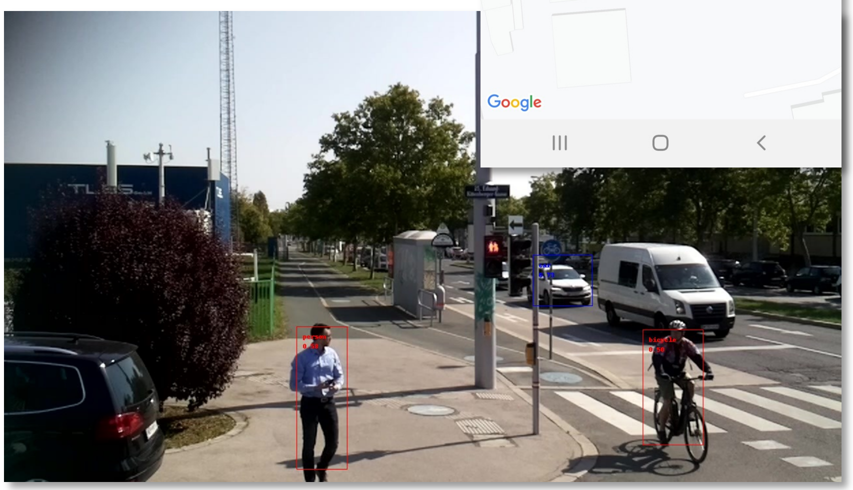


Fig 4: Object detection



Source: Lujic, De Maio, Pollhammer, Bodrozic, Lasic, and Brandic, "Increasing Traffic Safety with Real-Time Edge Analytics and 5G," EdgeSys, pp. 19-24, 2021.

Slide: courtesy Ivan Lujic (Ericsson Nikola Tesla d.d.)

Smart City II

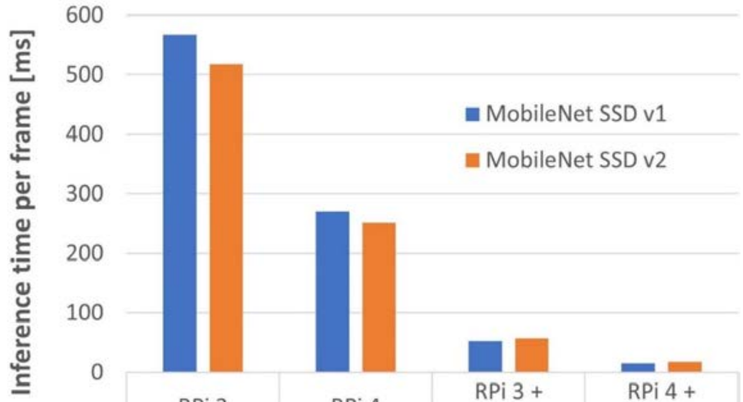
- App for visually impaired people
- App for aggregated information (Asfinag, City of Vienna, etc.)

Magenta swarco Stadt Wien TU Wien

http://intrasafed.ec.tuwien.ac.at

Design Choices: Computation Part

- Edge real-time processing of video frames
- Object detection workflows
  - Pre-trained neural network models (TensorFlow Lite)

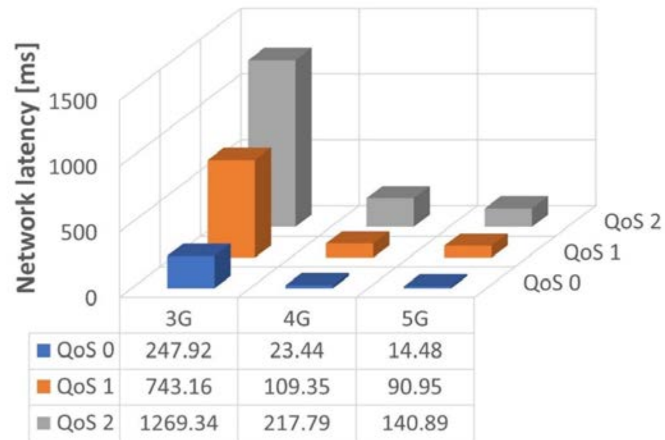


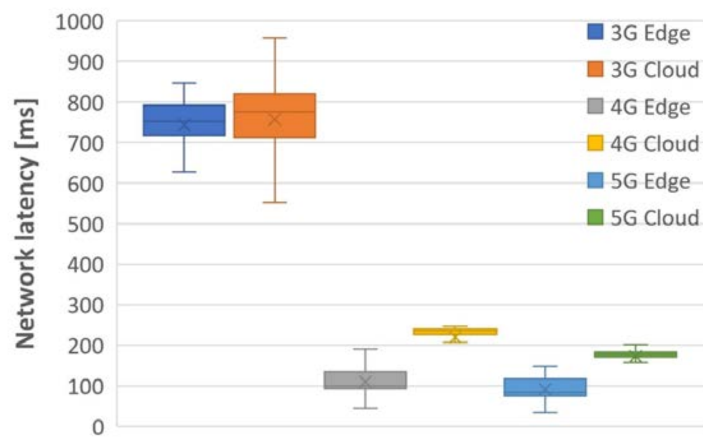
	RPi 3	RPi 4	RPi 3 + Edge TPU	RPi 4 + Edge TPU
MobileNet SSD v1	566.84	270.34	52.41	15.32
MobileNet SSD v2	517.63	251.44	57.44	17.78

- Traffic light chamber configuration:
  - Raspberry Pi model 4
  - Edge TPU
  - 8MP Camera Module V2

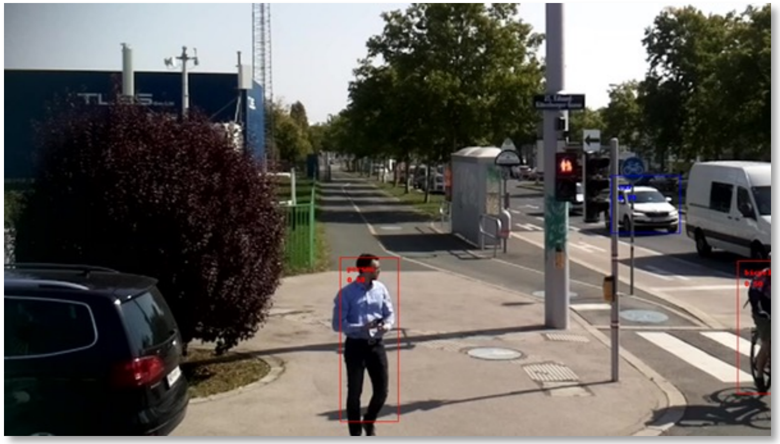
Design Choices: Communication Part

- Requirements:
  - Low latency
  - Guaranteed delivery of messages
- MQTT
  - Publish/Subscribe messaging
  - Quality of service (0, 1, 2)
- MQTT broker placement
  - Edge (TU Wien's infrastructure)
  - Cloud service
- Reduced network latency (edge vs. cloud) around
  - 50.20% (4G)
  - 47.18% (5G)





System Deployment



- Edge nodes setup
- Critical intersection

- Detection of pedestrians and cyclists
- 108.73ms on average (5G)

- Notifications sent to mobile client
- Location based on mobile phone's GPS

1 → 2 → 3

Digital Twins

Siberia's gateway to the underworld

Batagaika crater

Source: https://www.science.org/content/article/siberia-s-gateway-underworld-grows-record-heat-wave-thaws-permafrost

Ambarnaya river, Norilsk, Siberia

Source: https://www.brusselstimes.com/news/115566/u-s-offers-to-help-russia-following-arctic-oil-spill/

SWAIN - Sustainable Watershed Management Through IoT-Driven Artificial Intelligence

- In-silico digital representation of the river
- viscoelasticity
- **Data from IoT sources + expert knowledge = digital model of the river**

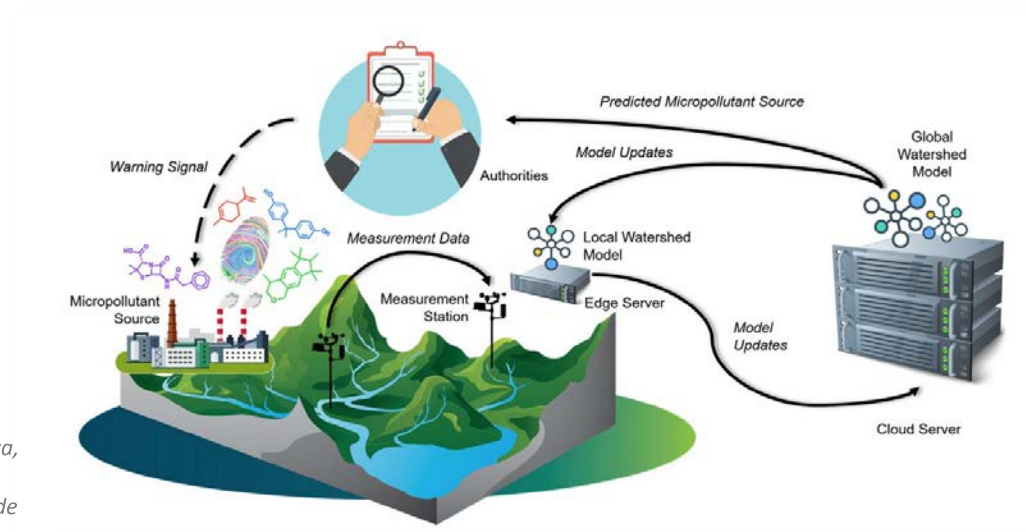
TU WIEN

Boğaziçi Üniversitesi 1863

SYKE
Finnish Environment Institute

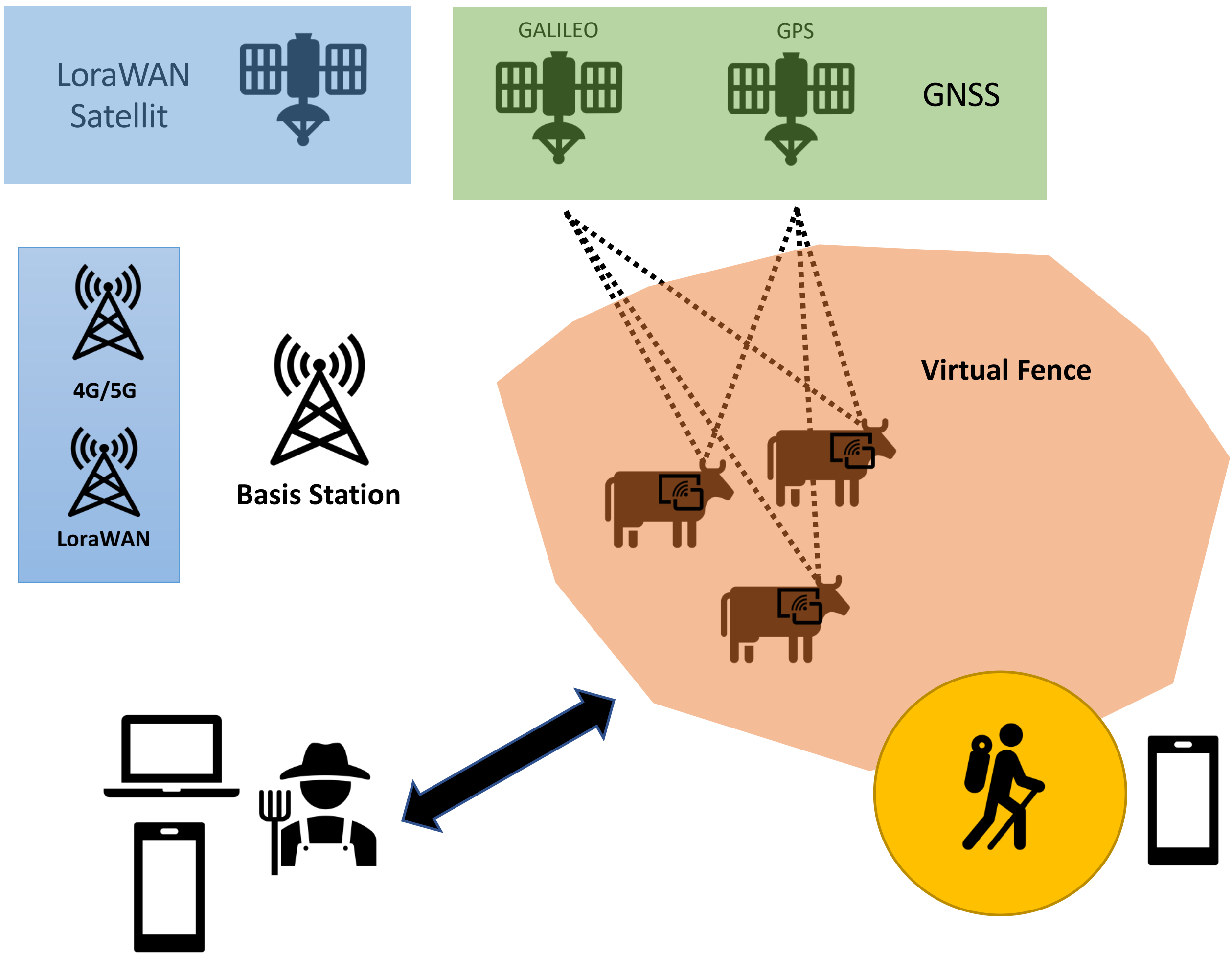
universität wien

chist-era



Source: S. Ahmad, H. Uyanık, T. Ovatman, M. T. Sandıkkaya, V. De Maio, I. Brandic, A. Aral. Sustainable Environmental Monitoring via Energy and Information Efficient Multi-Node Placement. IEEE Internet of Things Journal

Virtual Shepherd



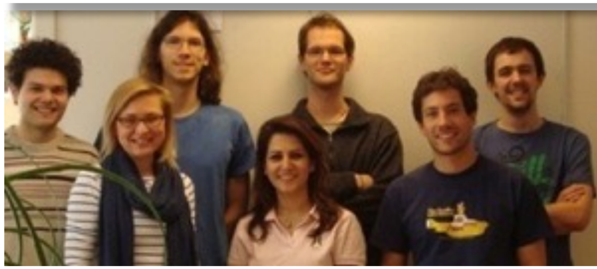
FFG
Forschung wirkt.
MOVING LAYERS
VIEHFINDER
TU WIEN
TECHNISCHE UNIVERSITÄT WIEN
TU Graz

Conclusion

- QPU for very specific operations
  - Chemistry
  - ML
- Challenge of integrating hybrid systems
- Mindset and education
- Non von Neuman architectures
- Limited hardware availability
  - Importance of simulators for teaching and engineering
  - Importance of benchmarking on real machines
- Focus on telescope technology
- Digital Twin technologies
  - Depends on the power/simplicity of the model

Thanks to funding agencies and my team

2010



2013

European Commission | Horizon 2020 European Union funding for Research & Innovation

cost EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

chist-era

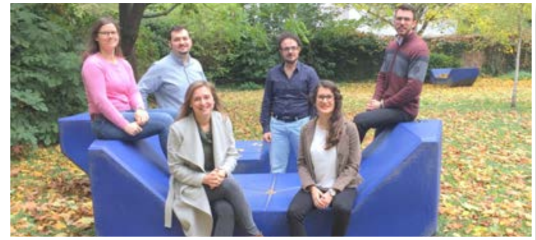
QUANTUM AUSTRIA

W|W|T|F

Stadt Wien

MARIE CURIE ACTIONS

2017



2021

2023

FWF

FFG Forschung wirkt.