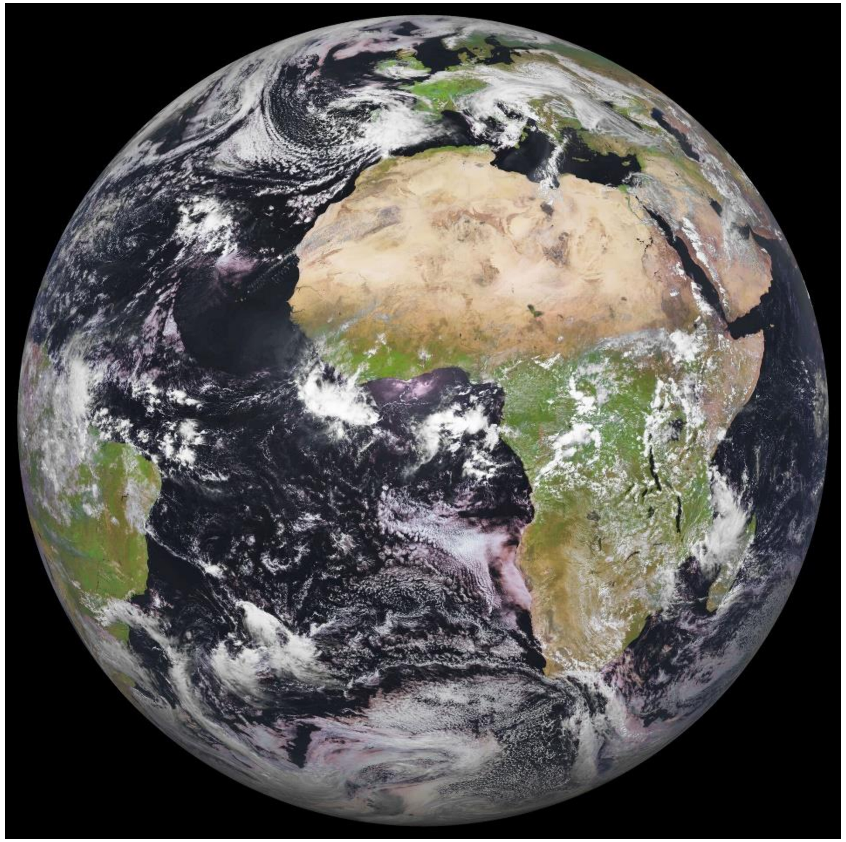
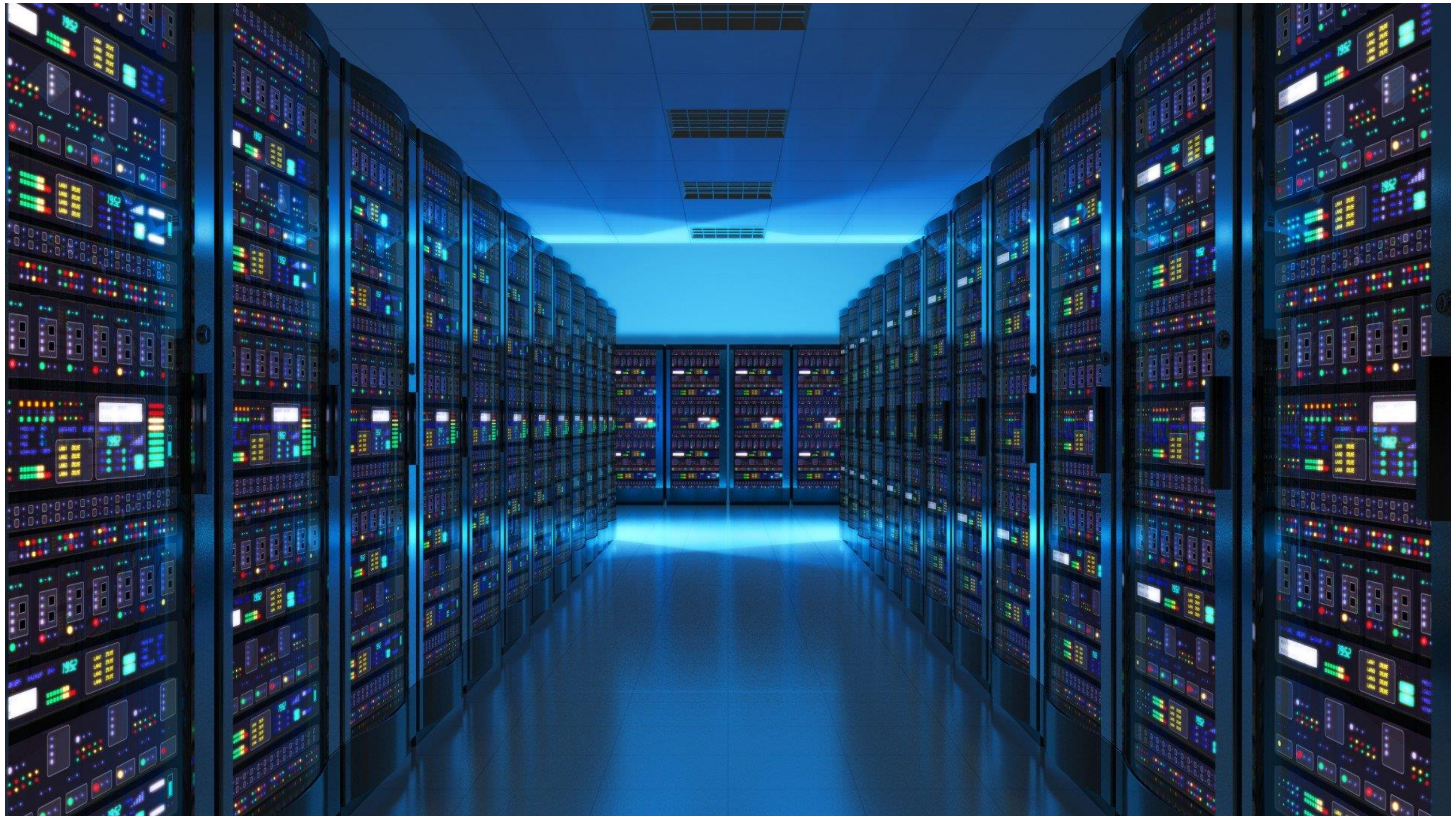
EcoCloud
A center for research on sustainable computing

EcoCloud Town Hall
EPFL

- Welcome and introduction, by Prof. Giovanni De Micheli
- EcoCloud's experimental facility, by Dr. Xavier Ouvrard
- Projects that are already using our facility
    - Post Moore Data Centers: Ali Ansari (PARSA)
    - Heating Bits: Enea Figini (DESL)
    - Urban Twin: Dr. Denisa Constantinescu (ESL)
- Initiatives on sustainability:
    - Methodology to measure carbon impact, by Julia Paolini
- EcoCloud Annual Event, by Prof. Giovanni De Micheli

The Challenge for this decade and beyond



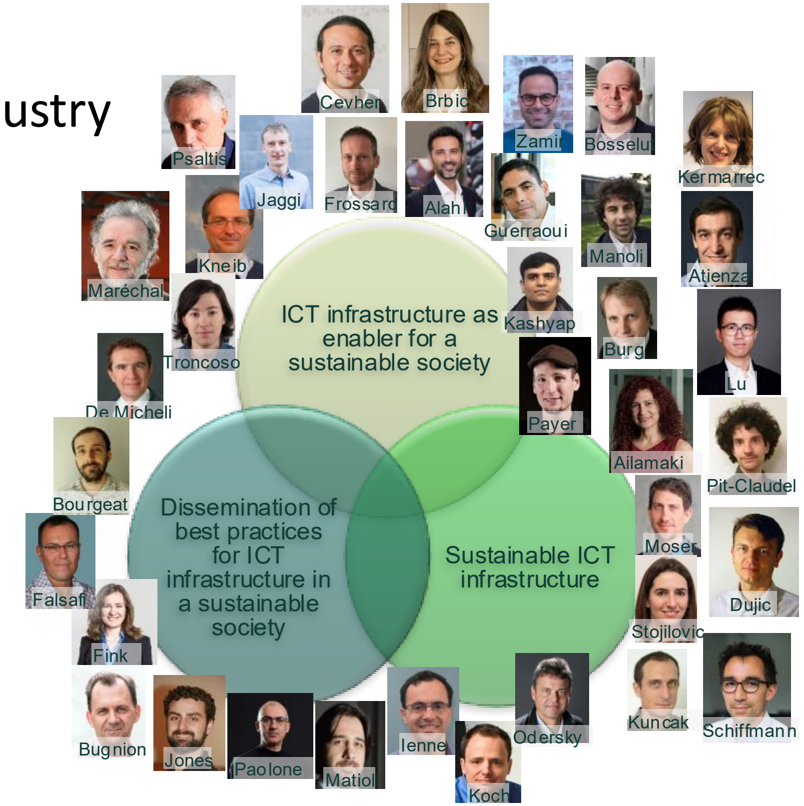


The Challenge for this decade and beyond

- The market is driven by AI/ML applications:
  - Large storage space
  - Large amount of energy for computation
- How do we rethink architectures, circuits and devices?
  - To enable *edge devices* to use AI/ML
  - To curb the energy consumption in servers

EcoCloud: An EPFL research center for sustainable computing

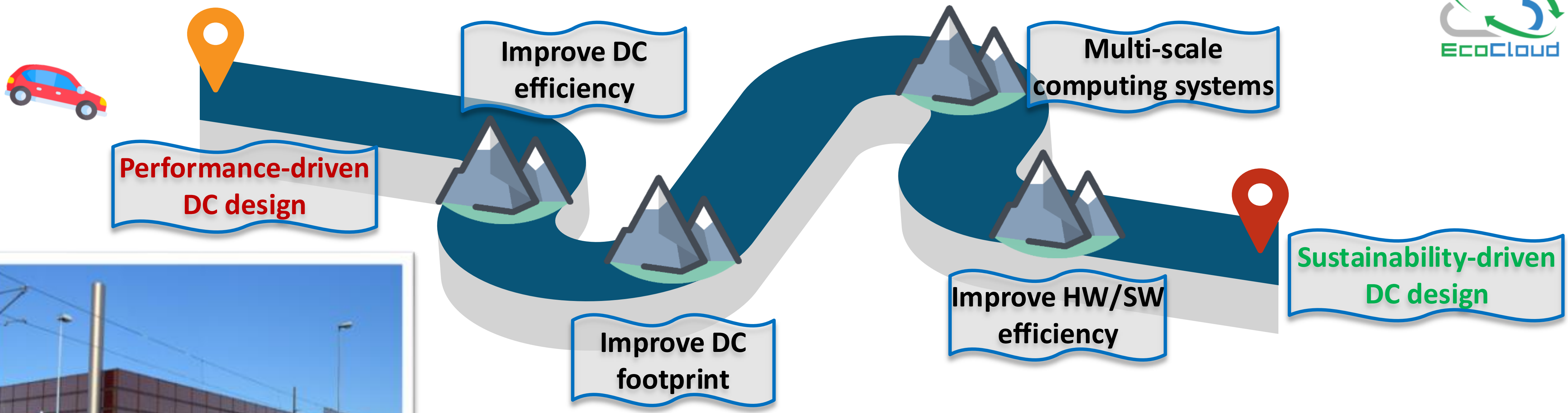
- EcoCloud networks the EPFL research community around sustainability topics
  - IT cross-layer optimization from edge devices to the cloud
  - Promoting large multi-disciplinary projects including EPFL labs and industry
  - 35 faculty affiliated, 4 schools support EcoCloud (IC, STI, ENAC and SB)
- Three main research interests:
  - Research on reducing energy consumption
    - Computing and Communications in Cloud and Edge systems
    - Hardware and Software
  - Run an experimental Data Center
    - Using advanced technologies for cooling
  - Coordinate cooperative projects
- Strong link with local and global industry



HUAWEI · intel · ORACLE · IBM Research Europe · Hewlett Packard Enterprise · EATON

IMO Ventures · AXA Technology Services redefining / service · GEMATEG · corintis · JJ Cooling Innovation · infomaniak

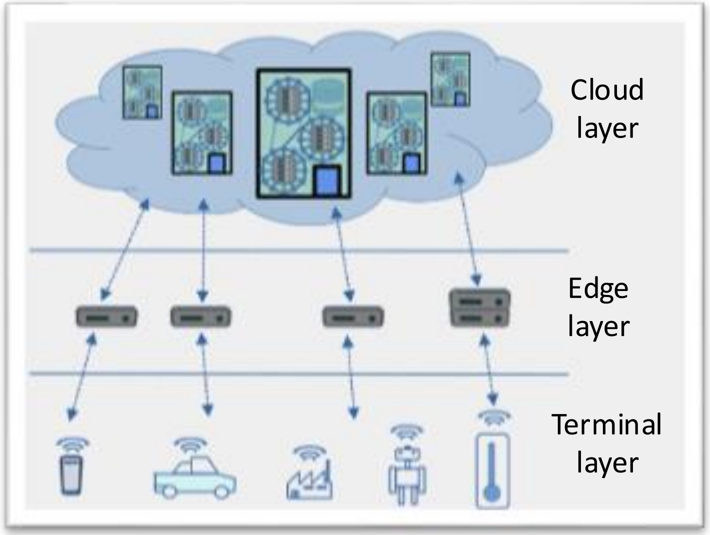
From performance- to sustainability-driven DC design



EPFL's new DC integrating PV generation, water cooling, and heat recovery for campus heating

ECOGreen

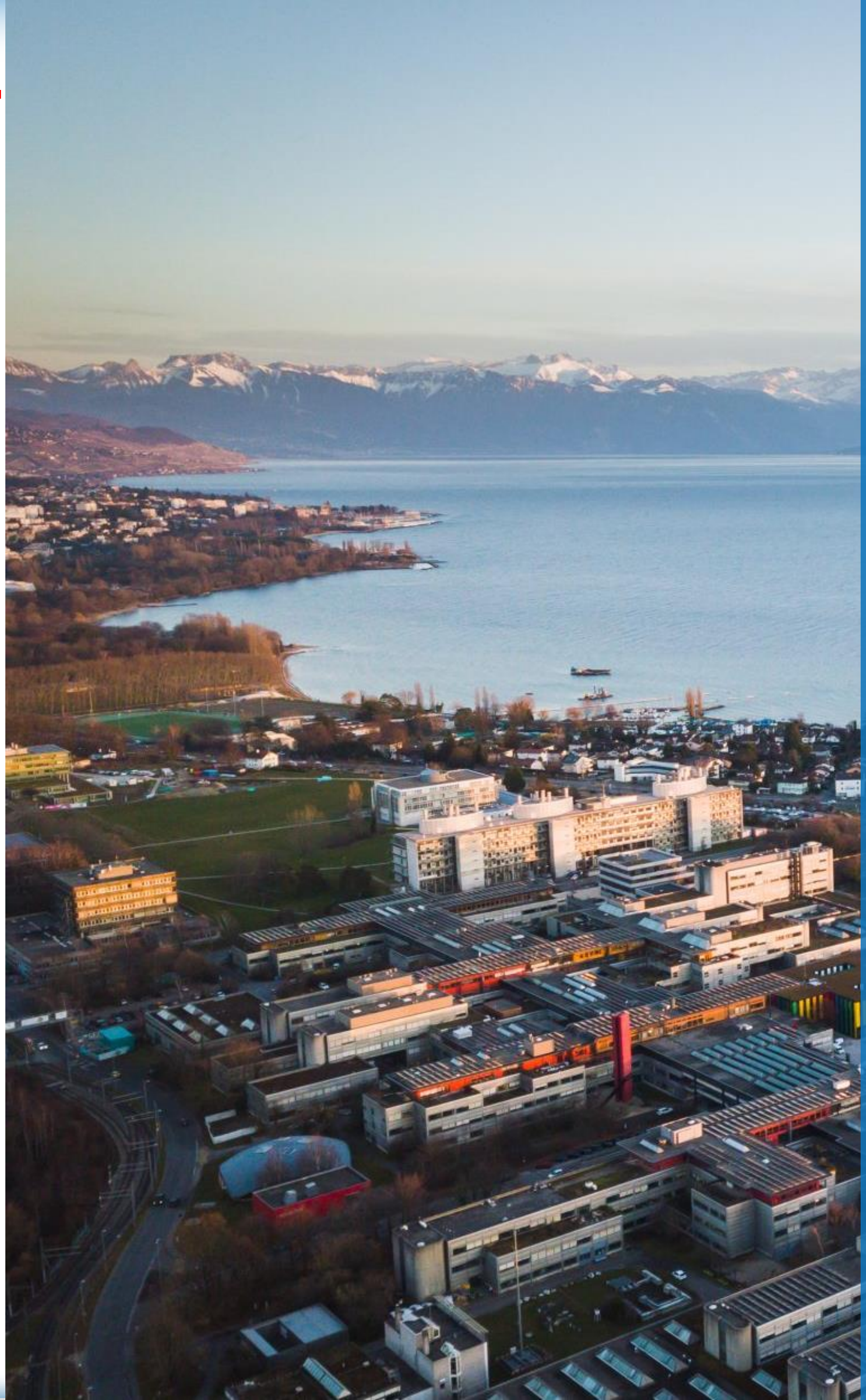
A crystal ball to assign sets of jobs to DCs



Source: Dr. Xavier Ouvrard, EcoCloud[1]

[1] X. Ouvrard, et al., "Special session: Challenges and opportunities for sustainable multi-scale computing systems," 2023.

- Welcome and introduction, by Prof. Giovanni De Micheli.
- EcoCloud's experimental facility, by Dr. Xavier Ouvrard
- Projects that are already using our facility
  - PARSA: Ali Ansari (PARSA)
  - Heating Bits: Enea Figini (DESL)
  - Urban Twin: Dr. Denisa Constantinescu (ESL)
- Initiatives on sustainability:
  - Methodology to measure carbon impact, by Julia Paolini
- EcoCloud Annual Event, by Prof. Giovanni De Micheli



EcoCloud’s experimental area

How can EcoCloud help its community?

- Experimental area
  - EcoCloud has an experimental area of about 100 m^2 + 20 m^2 in CCT
  - Experiments on IT sustainability and IT for sustainability
  - Fully instrumented, data available to researchers

- Characterization of energy efficiency at:
  - Software level
    - Test algorithms on different generation hardware
    - Energy consumption externally instrumented with full monitoring
  - Hardware
    - Different power management techniques
    - Energy-efficiency of FPGA-based accelerators

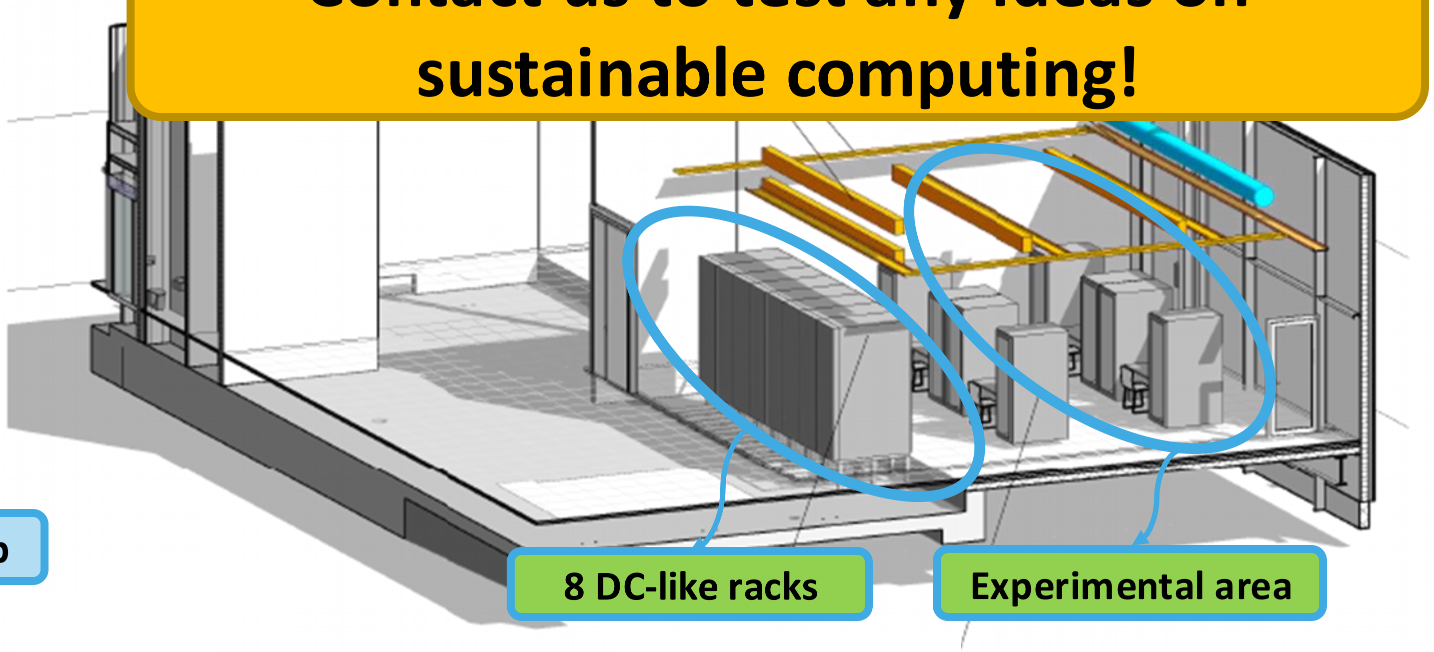
    With possibly tailored configurations
- Experiment:
  - Different liquid cooling solutions
    - Rack, server and chip
  - Experiment heat recovery solutions

EcoCloud's experimental facility for sustainability

EcoCloud

- ~100 m^2 of space for experiments on sustainable computing
  - Recycled clusters
  - SoA servers
  - Donations
- Experimental support: two spaces
  - 35 kW/50 kW passive door air liquid cooling doors
  - Possibility to have soon CDUs on some racks
  - Monitoring: energy, temp., etc.
  - Cooling: air or water cooling

Contact us to test any ideas on sustainable computing!



Racks with air/water passive cooling

Controlled setup

Cooling sub-station

Full supervision in-line with EPFL systems

Virtual tour of EcoCloud's experimental area

Video recorded and produced by Alex Widerski, EcoCloud

Experimental area: What is available as computing resource

- Several state-of-the-art dual socket servers:
  - Intel:
    - 8553Y IceLake 32C,
    - Intel 5418N 20C Sapphire Rapids
    - 6448Y 32C Sapphire Rapids: 2 with air cooling, 2 in DLC
    - 2 Xeon Gold 6240 with 4 Nvidia V100
  - AMD
    - EPYC 7763 64C Zen 3 Milan
    - EPYC 9554 64C Zen 4 Genoa
    - EPYC 9575F 64C Zen 5 Turin
  - Link in 100G by switches or direct links: specific configuration on demand

Experimental area: What is available as computing resource

- 2 clusters:
  - FIDIS nodes :
    - 2 racks are used by Heating Bits and 2 additional available
    - Can be used for computation, better than using, than just generating dummy workloads
  - BlueBrain nodes AMD Zen 1 and Zen 2:
    - 192 GB / 384 GB
    - Slurm cluster is currently put in place on top of it
    - No infiniband, storage limited, not to be used for HPC
- Possibility to have real-time measurements of power and energy consumption of servers

Experimental area: What is available as computing resource

- FPGAs:
  - AMD Virtex UltraScale+ VU19P (in use)
  - AMD Virtex UltraScale+ VU9P (in use)
  - AMD Alveo V80
  - 10 x Zynq UltraScale+
  - Prodigy S7-19PS-2 Logic System with Xilinx Virtex+ 19P

- Possibility to host some VMs for dedicated projects

Experimental area: What is available as computing resource

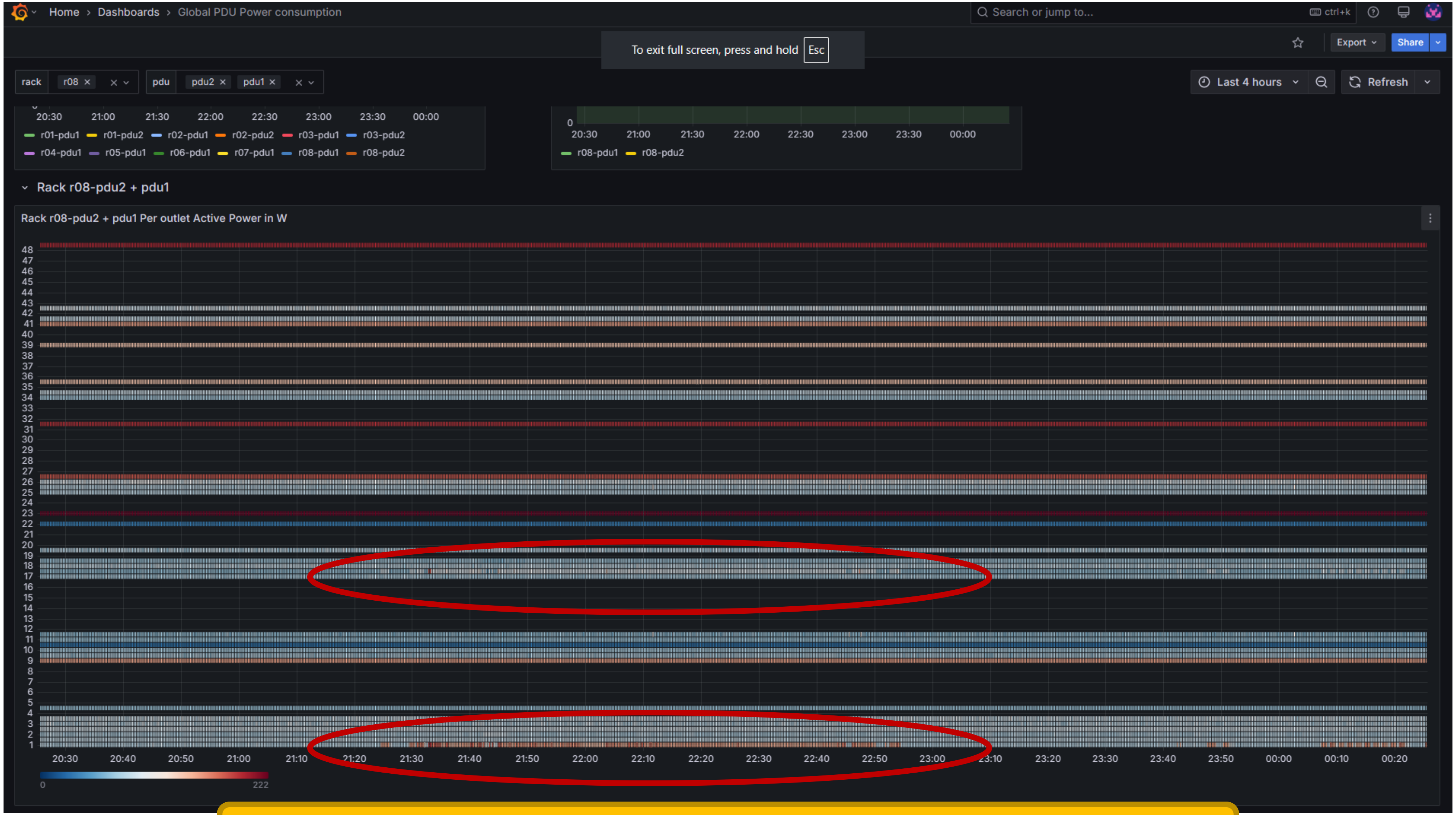
- We acquired different servers with DLC to benchmark heterogeneity and to compare to classical air cooling

Experimental area: What is available as computing resource

- Retention bucket for DLC experiments with 50kW available
- Coming soon: July 2025: Cooling Distribution Unit 200 kW for 2 racks
  - One for testing heterogeneity
  - One for additional experiments on DLC

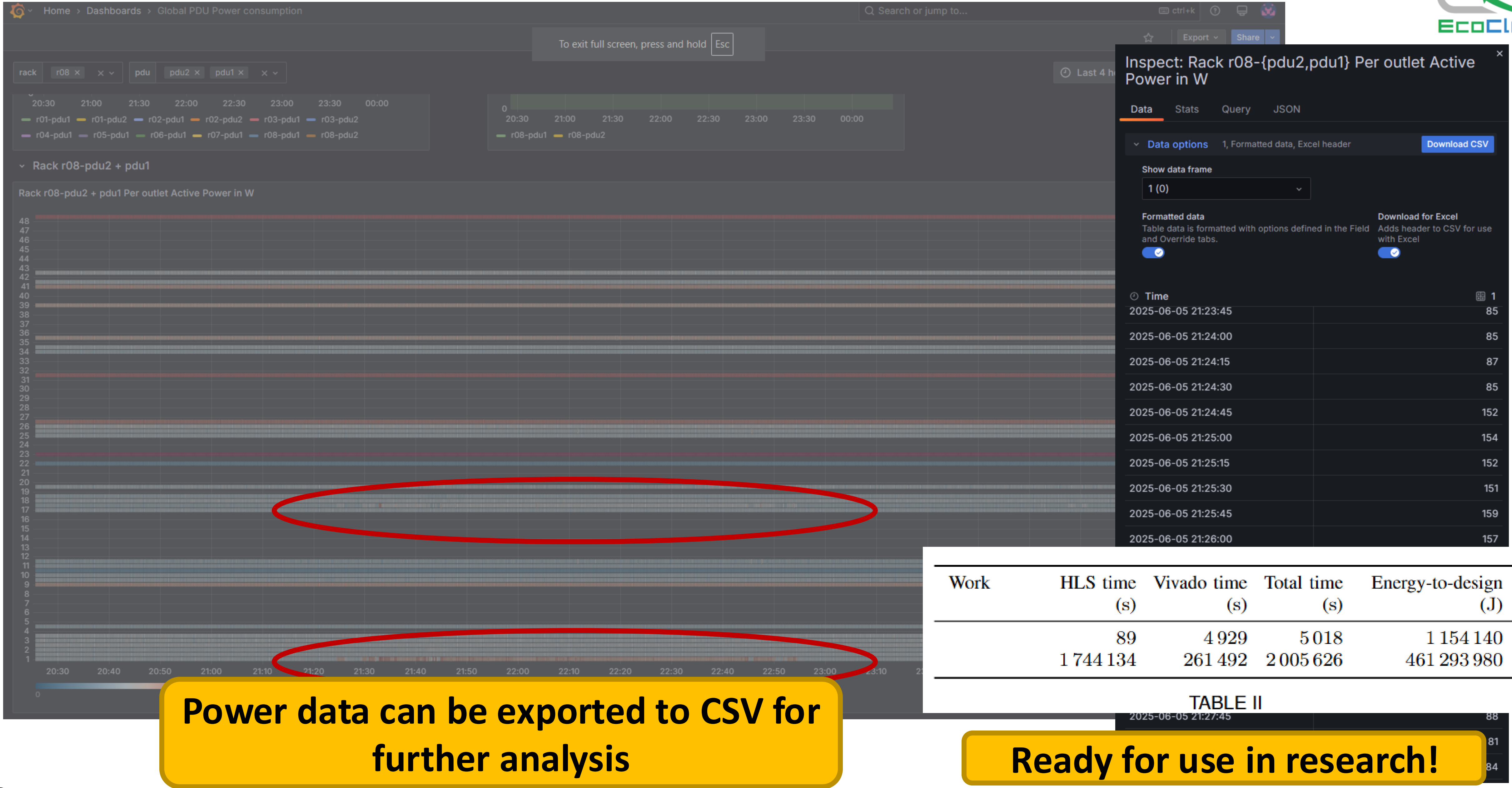
Credits: https://www.coolitsystems.com/

Energy efficiency characterization



Average machine power measured every 15 seconds

Energy efficiency characterization



Work	HLS time (s)	Vivado time (s)	Total time (s)	Energy-to-design (J)
	89	4 929	5 018	1 154 140
	1 744 134	261 492	2 005 626	461 293 980

TABLE II

Power data can be exported to CSV for further analysis

Ready for use in research!

Experimental area: How to?

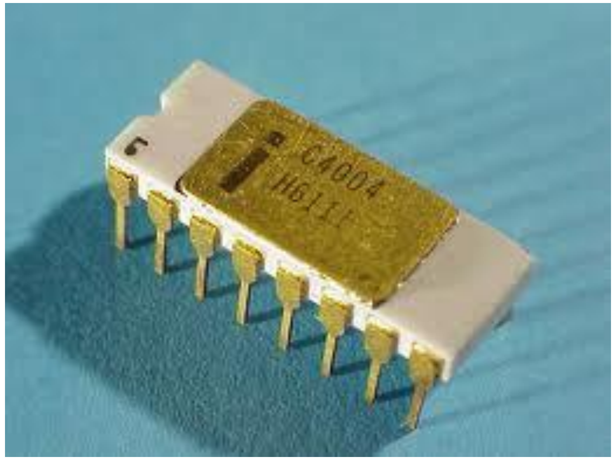
- Project in the experimental space
  - describe in a small project description the research you want to achieve (5-10 lines)
  - ask for specific setups (direct connections between servers, FPGA in the middle, other resources)
  - loan for up to three months, by quarter
  - For larger demand (a rack of server) => requires to check computation power
  - Possibility of DLC servers
  - Other kind of experiment (thermal measurements, power measurements)
  - If additional equipment / means needed might ask for specific funding with contribution to EcoCloud
- Affiliated labs are welcome to make new proposals and ask for space usage
- NB: EcoCloud is not offering computing services, for this you can go to RCP / SCITAS

- Welcome and introduction, by Prof. Giovanni De Micheli
- EcoCloud's experimental facility, by Dr. Xavier Ouvrard
- Projects that are already using our facility
    - Post Moore Data Centers: Ali Ansari (PARSA)
    - Heating Bits: Enea Figini (DESL)
    - Urban Twin: Dr. Denisa Constantinescu (ESL)
- Initiatives on sustainability:
    - Methodology to measure carbon impact, by Julia Paolini
- EcoCloud Annual Event, by Prof. Giovanni De Micheli

MOORE’S LAW

Pillar to all forms of modern computing

1971
Intel 4004



92,000 ops/s

2.5 million times faster!

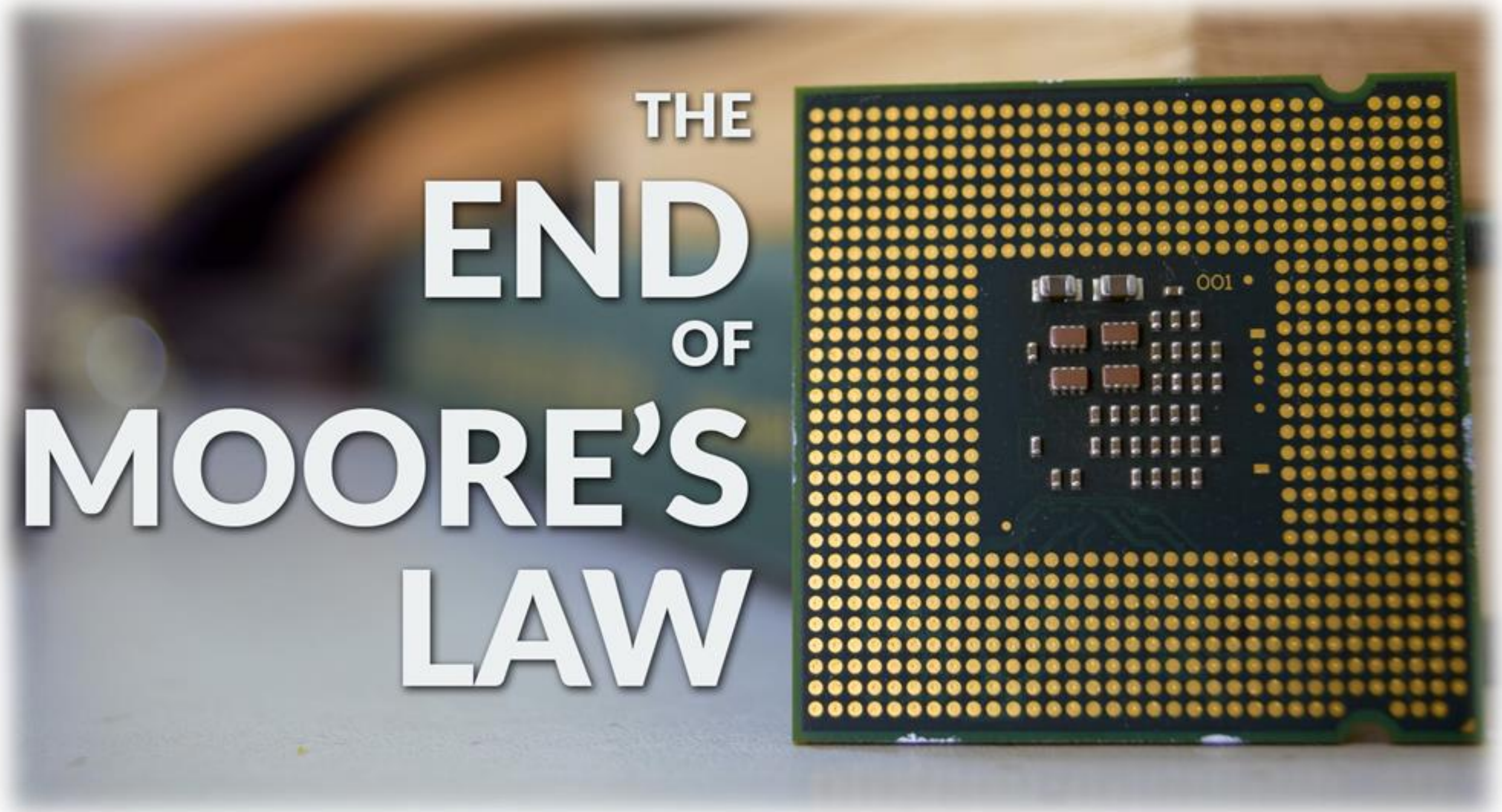
47 YEARS

14000 times less energy/op!

2018
Intel Core i9

218,000,000,000 ops/s

LONG LIVE MOORE'S LAW



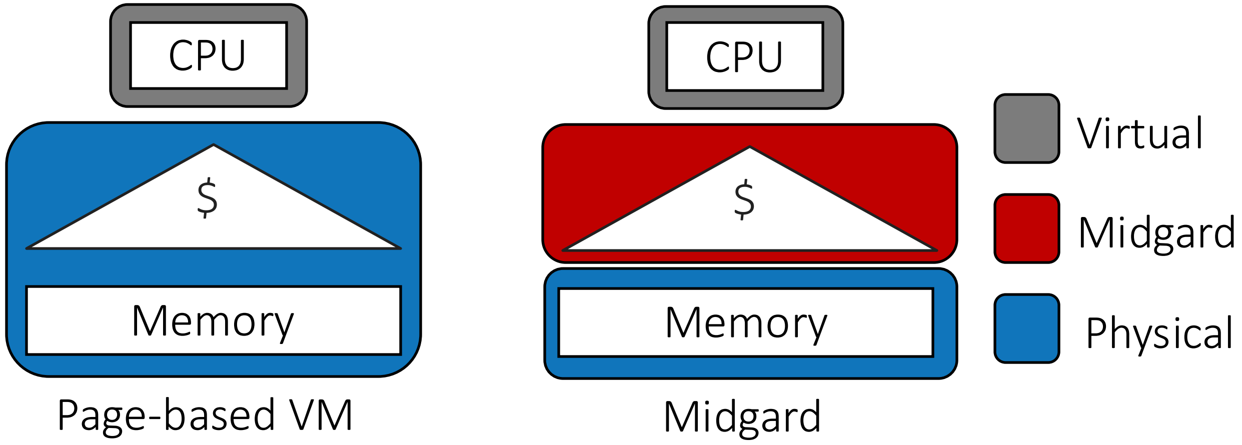
POST-MOORE DATACENTERS

Design for "ISA"

- Integration: minimize data movement
- Specialization: cut resources for analyzing data
- Approximation: compress data & compute

From algorithm to infrastructure

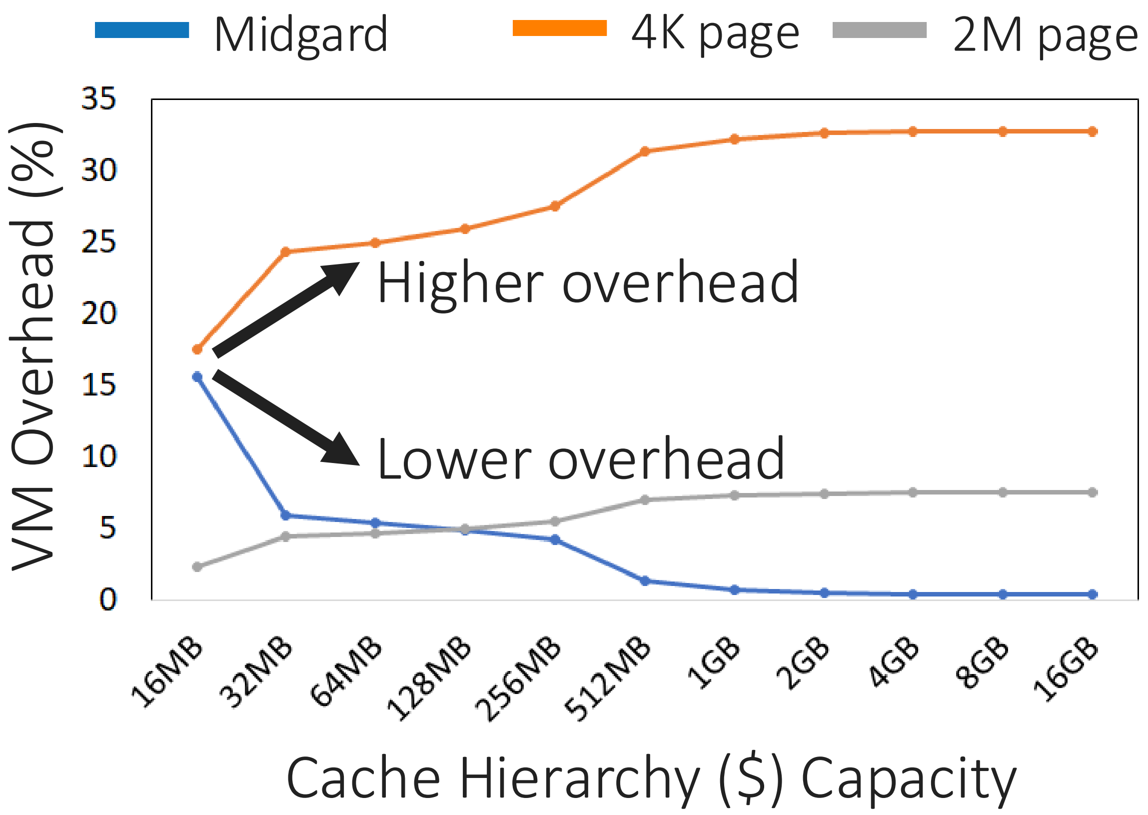
FUTURE-PROOFING VIRTUAL MEMORY



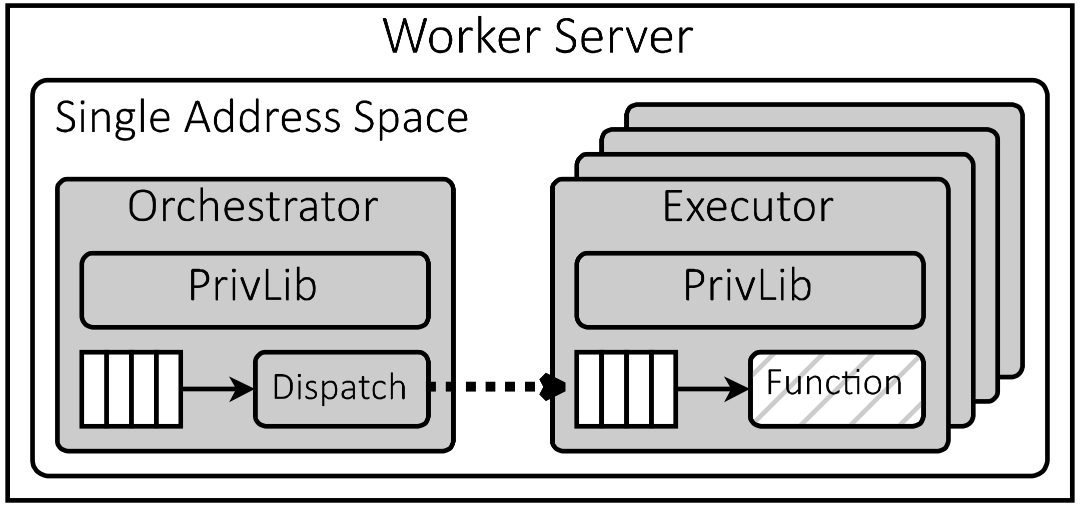
Problem: Memory isolation bottleneck

Solution: Intermediate address space

- Keeps POSIX (VMA) interface to apps
- Eliminate TLBs
- Unclogs virtual memory for virtualization, accelerators, security



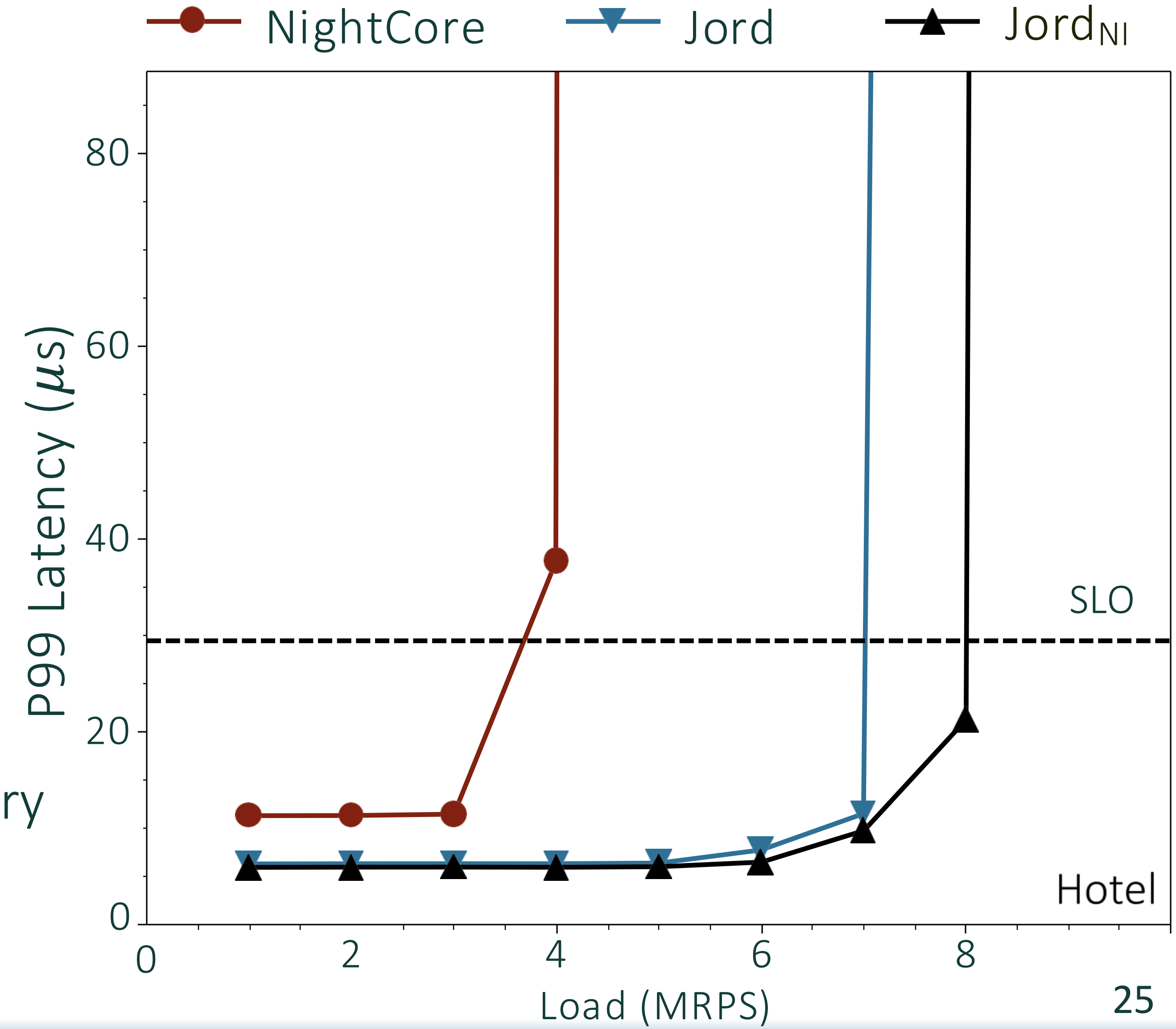
CUTTING DATACENTER TAX W/ SINGLE-ADDRESS SPACE CLOUDS



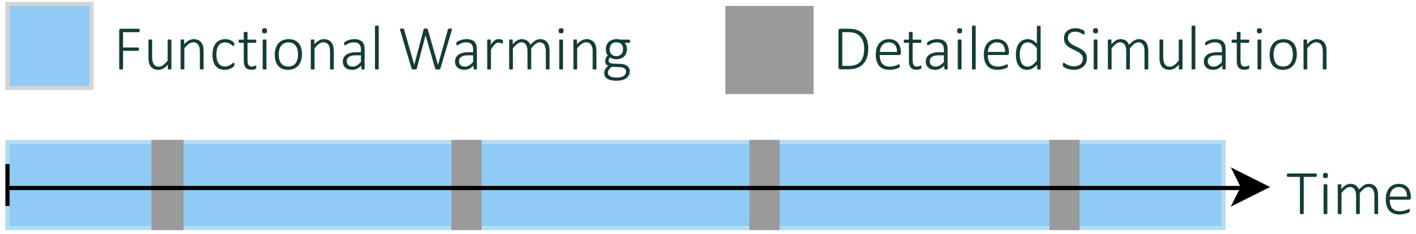
Problem: Today's OS not suitable for datacenters

Solution: Single-address space clouds

- HW/SW co-design for user-level memory isolation, dispatch, communication
- Nanosecond-scale containers



SUB-GIPS SERVER SIMULATION

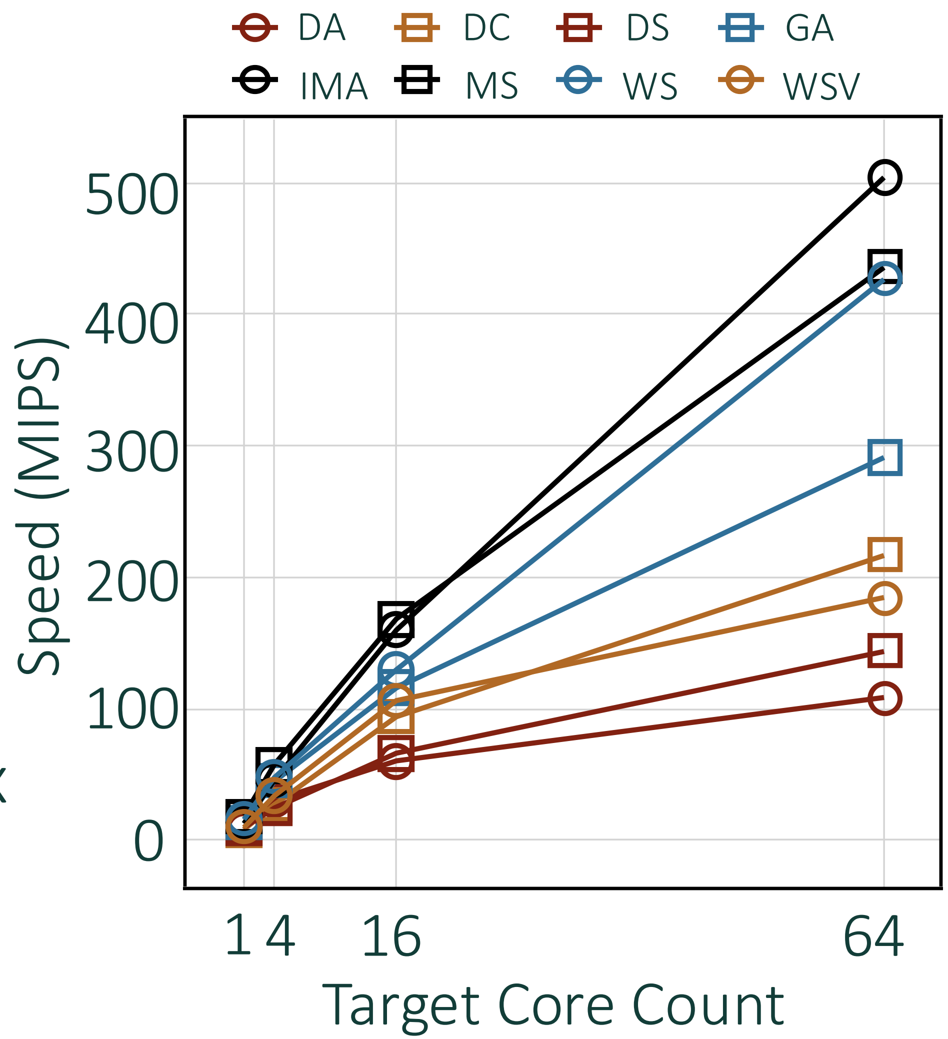


Problem:

- Timing simulators run at 250 KIPS
- Simulation server for 10s → 2 months!

Solution: Parallel sampled simulation

- Simulation at 100-500 MIPS
- Reduces timing requirement by 100x-1000x
- Bounds error with statistical guarantees
- Enables accurate & practical simulation



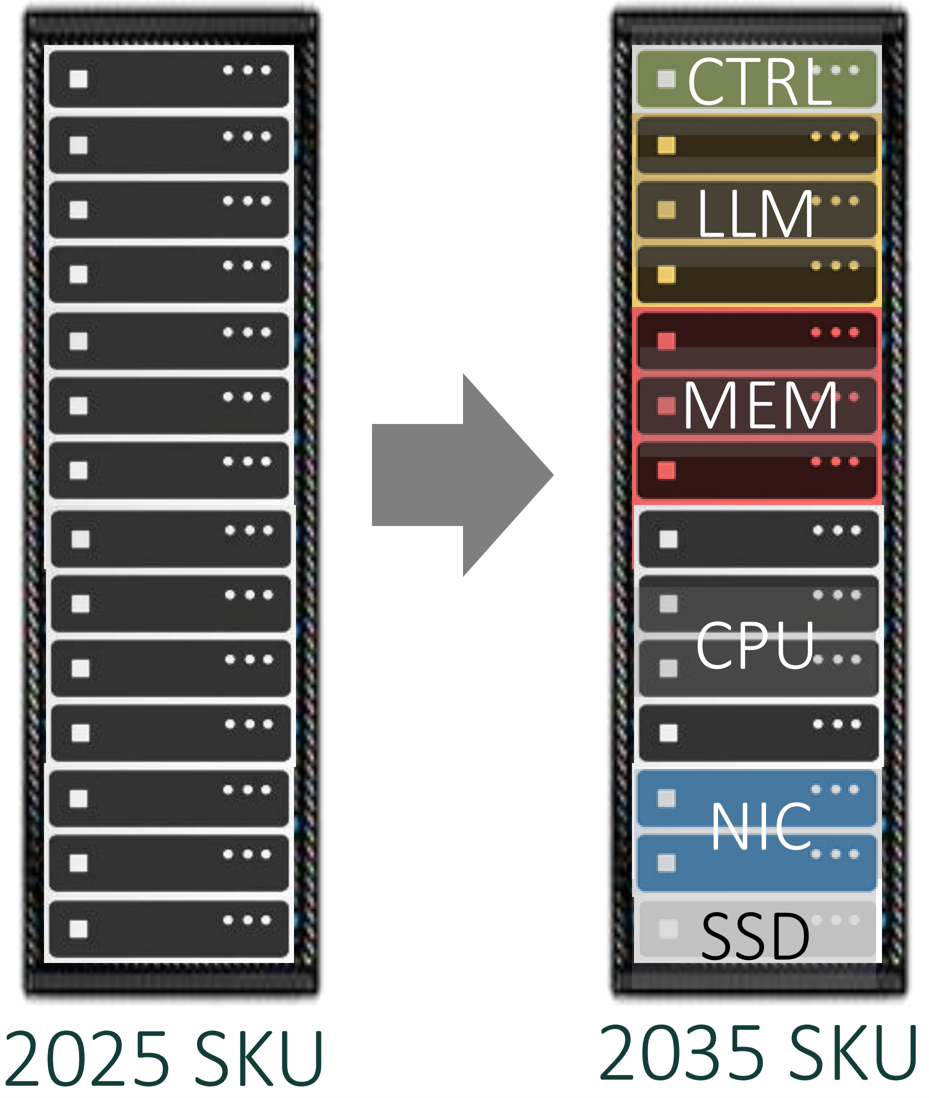
CLOUD-POD: RACK AS A SERVER

Problem:

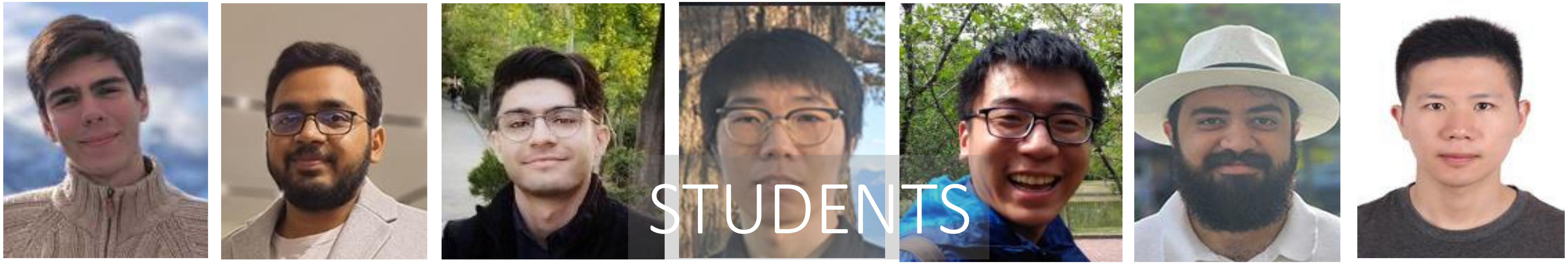
- Workloads have diverse silicon requirements
- Server as SKU strands/fragments silicon

Solution: Disaggregate silicon

- Fabrics at 100s Gb/s/lane
  - NVLink, NeuronLink, CXL
- Pooled HW: CPU, memory, NIC, LLM
  - Manage via OS + fabric
  - Use custom connectivity
- Future SKU for datacenters



EcoCloud



PARSA PARALLEL SYSTEMS ARCHITECTURE LAB

hexhive

HE IG VD

CORNELL

EMBEDDED SYSTEMS LABORATORY

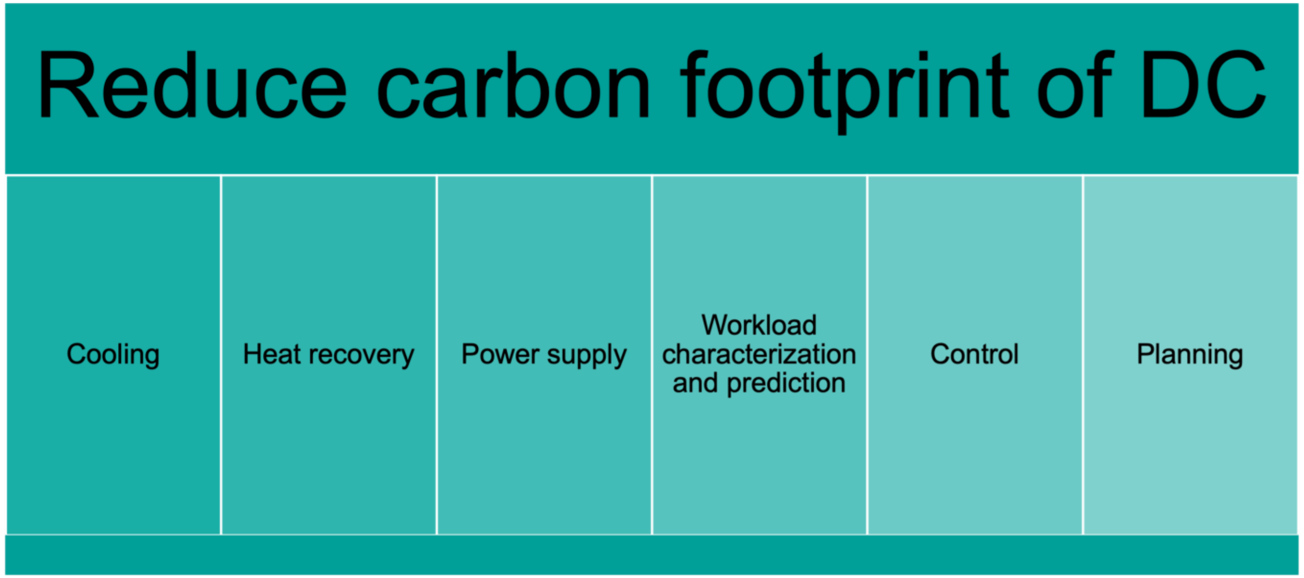
Yale

RS3Lab

- Welcome and introduction, by Prof. Giovanni De Micheli
- EcoCloud's experimental facility, by Dr. Xavier Ouvrard
- Projects that are already using our facility
  - PARSA: Ali Ansari (PARSA)
  - Heating Bits: Enea Figini (DESL)
  - Urban Twin: Dr. Denisa Constantinescu (ESL)
- Initiatives on sustainability:
  - Methodology to measure carbon impact, by Julia Paolini
- EcoCloud Annual Event, by Prof. Giovanni De Micheli

Flagship projects – Heating Bits

Goal: develop technologies to reduce the carbon footprint of data centers

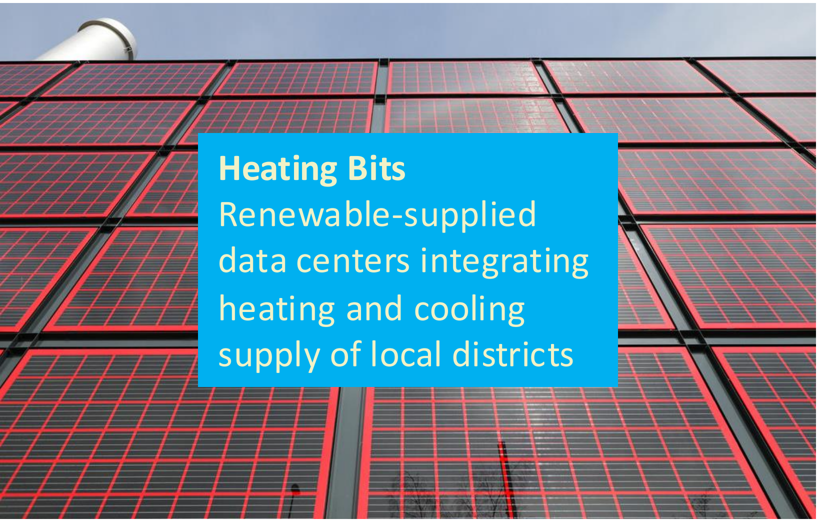


How do we use the facility?

- Small scale experiments
- Compatibility of hardware solutions

Funded by EPFL's S4S initiative — Involves 6 laboratories and EcoCloud

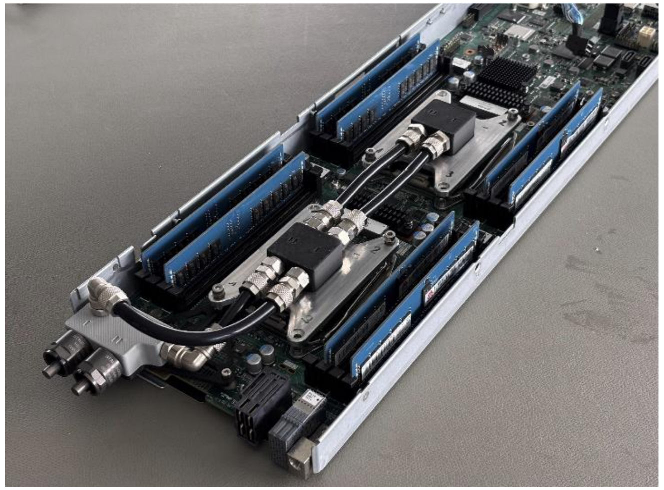
Flagship projects – Heating Bits

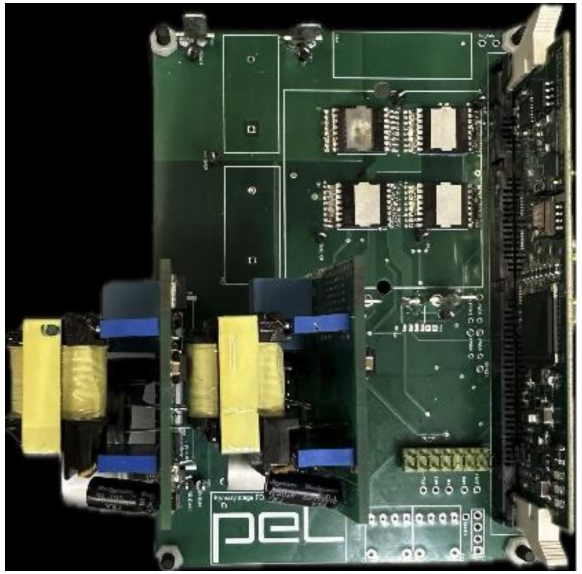


Hardware solutions

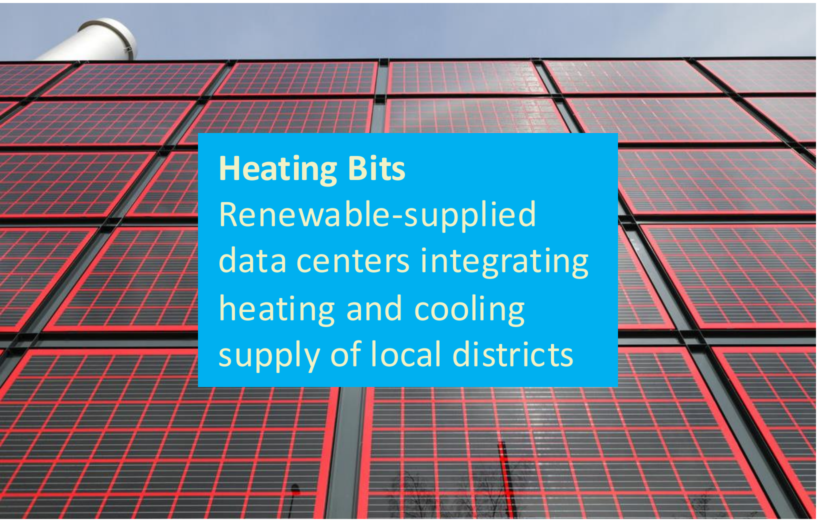
- *Efficient and medium-grade heat recovery*
  - *On-chip high temperature coolant*
- *Valorization of waste heat*
  - Organic Rankine cycle regeneration of electricity
  - Interaction with EPFL district heating
- *Optimized power supply*
  - *AC/DC, DC/DC*
  - *MV & LV*

Being developed for FIDIS nodes compatibility





Flagship projects – Heating Bits

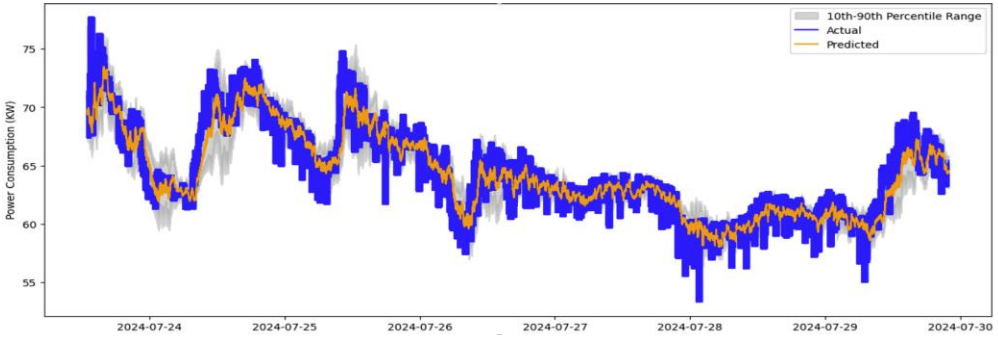


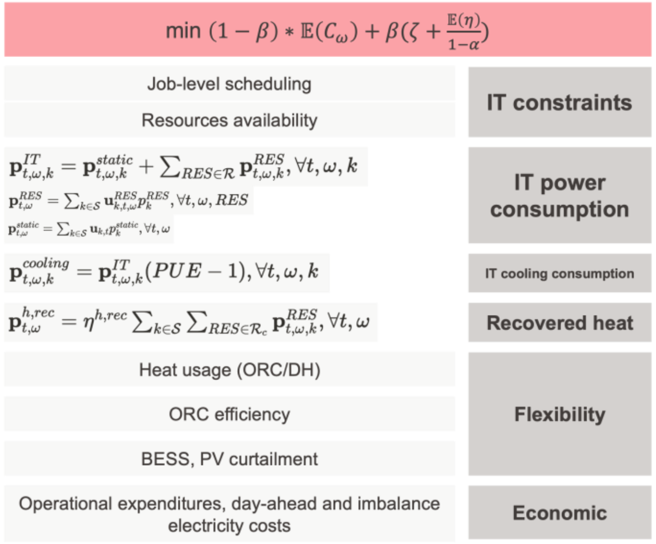
Heating Bits
Renewable-supplied data centers integrating heating and cooling supply of local districts

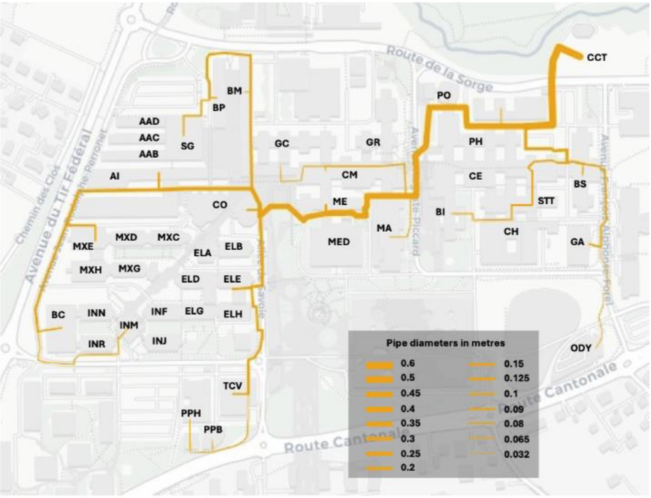
Software solutions

- *Workload*
  - Characterization
  - Prediction
  - Flexibility
- *Control*
  - Aligned with electricity markets
  - Steering all assets, leveraging flexibility
- *Planning*
  - Electricity and heat system-wide models
  - Sizing of assets

Will be **experimentally validated** in the facility, along with the hardware solutions.







- Welcome and introduction, by Prof. Giovanni De Micheli
- EcoCloud's experimental facility, by Dr. Xavier Ouvrard
- Projects that are already using our facility
    - Post Moore Data Centers: Ali Ansari (PARSA)
    - Heating Bits: Enea Figini (DESL)
    - Urban Twin: Dr. Denisa Constantinescu (ESL)
- Initiatives on sustainability:
    - Methodology to measure carbon impact, by Julia Paolini
- EcoCloud Annual Event, by Prof. Giovanni De Micheli

Flagship projects – Urban Twin

Motivation

Urban areas are highly affected and affect climate change

✓responsible for 75% of GHG emissions

✓livability significantly impacted by climate change

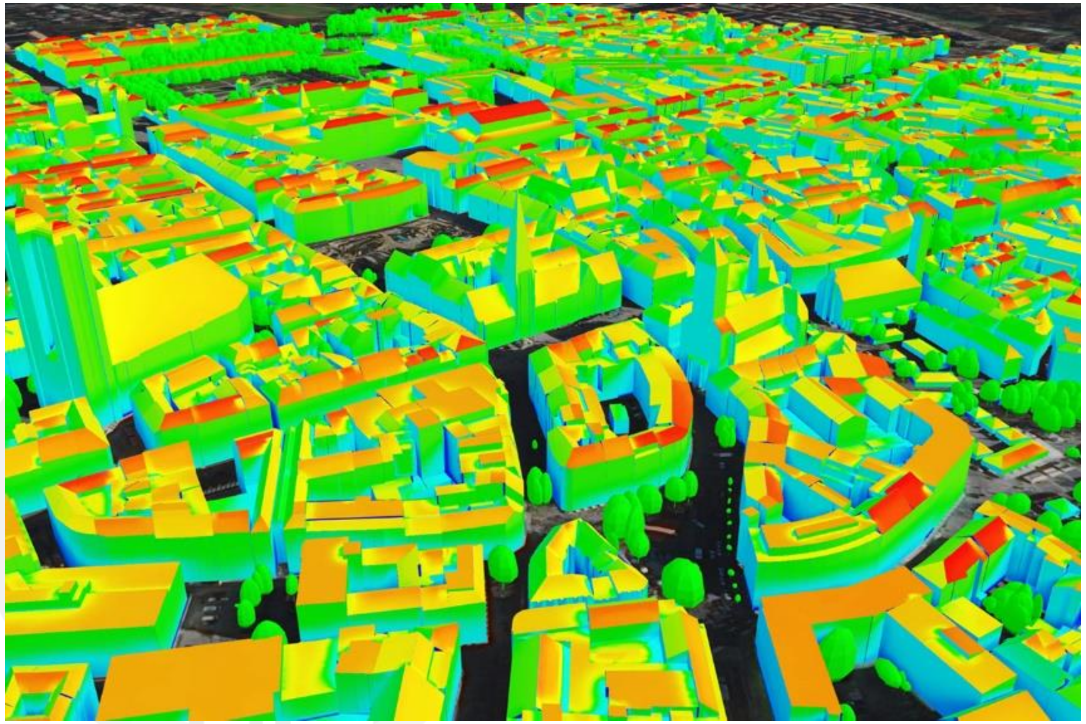
Switzerland is not spared from this situation; extreme events such as floods and heat waves affect its urban areas

Future: Urban-Climate relationship intensified since global urban population is expected to grow (50% -> 70% by 2050)

Aim

- Develop and validate **an integrated tool to support decision-makers** achieve environmental targets at urban scale, such as

Swiss Energy Strategy 2050



Credit: TUM

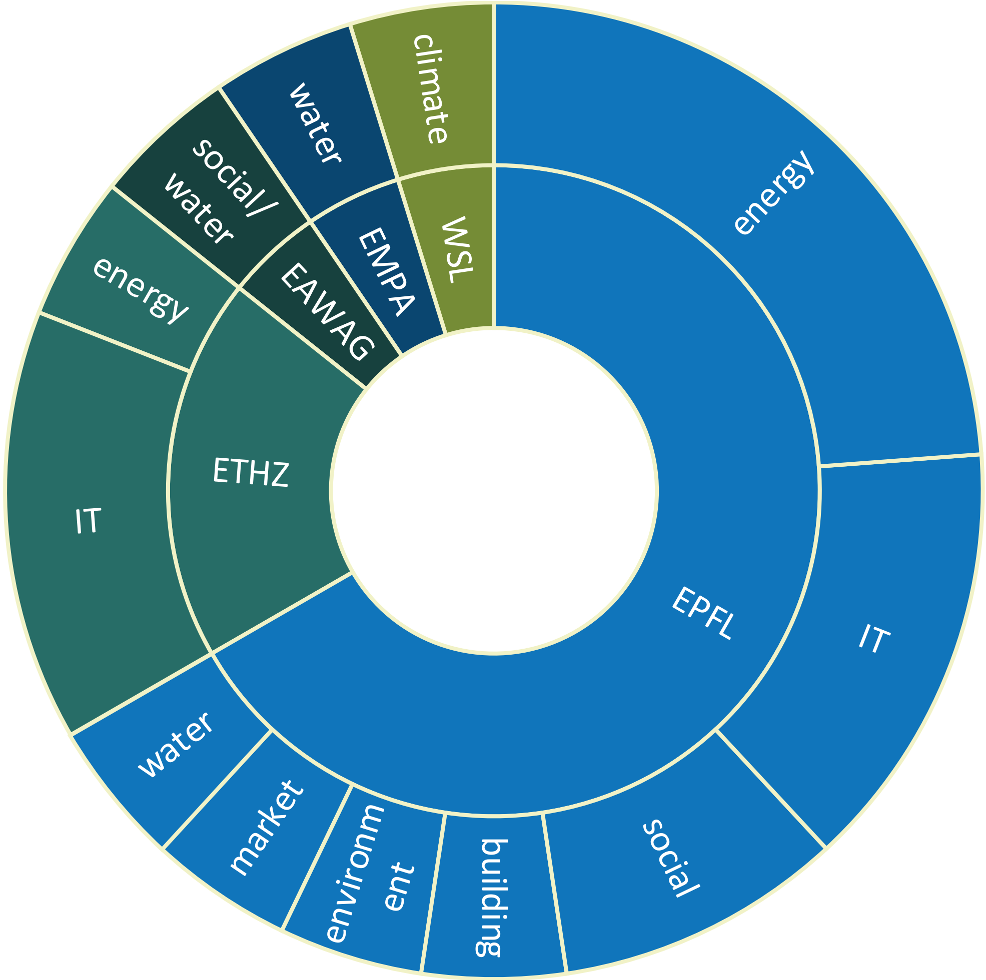
Vision of climate-adaptive "sponge cities"

Credit: Water magazine

Inter-Institutional Collaboration

Each ETH Institution brings a **specific set of capabilities** to complement the others and create the UrbanTwin tool:

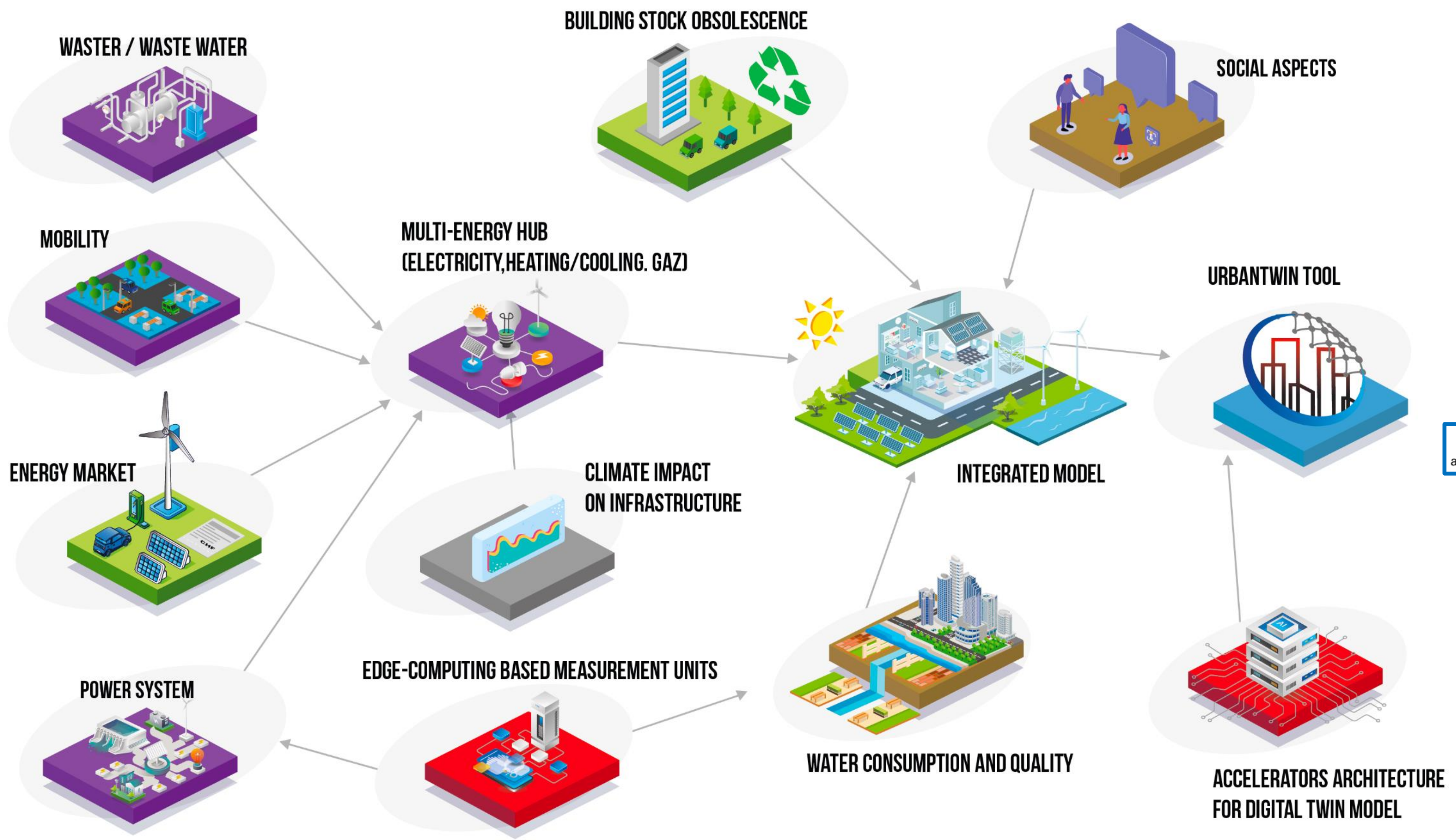
- **EPFL**: urban infrastructure modeling (technological, market aspects), and IT capabilities (energy-efficient distributed computing and cloud computing)
- **ETH**: energy infrastructure modeling (technological and market aspects), and IT capabilities (security and edge computing)
- **EAWAG**: water cycle modeling
- **EMPA**: water quality monitoring
- **WSL**: climate modeling



Joint initiative of the Board of the Swiss Federal Institutes of Technology

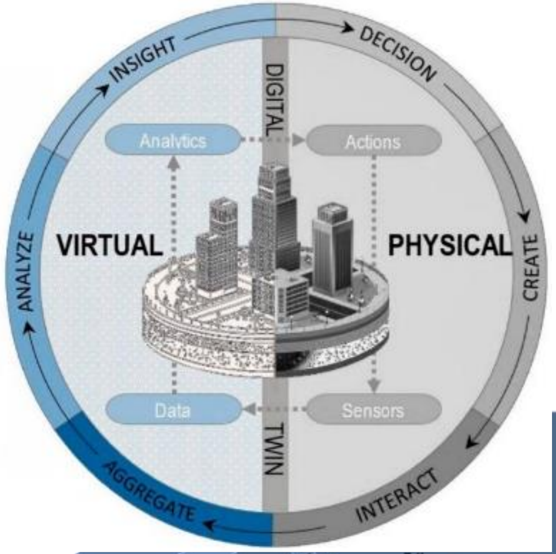
Urban Twin involves 12 EPFL laboratories (from the 4 schools) and 4 centers

UrbanTwin tool architecture

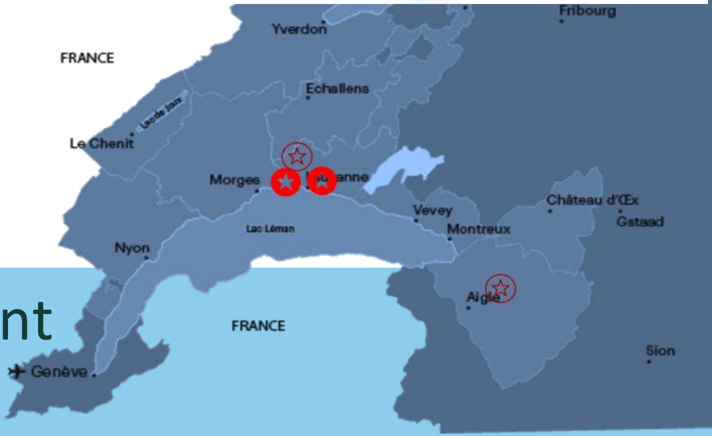


Inter-Institutional Collaboration

EcoCloud support
to UrbanTwin research outcomes



1. **Multi-scale digital twin model** of critical urban infrastructure (energy, water, buildings, mobility, socio-economic and environment)
2. **Energy-efficient multi-level IT computing infrastructure,** including edge-to-cloud computing (and KPI for sustainable big-data processing)
3. **Real-life validation**: four demo sites (different sizes and features) to ensure tool applicability to complete Swiss ecosystem
   - EPFL-EcoCloud facility on digital twin/IT aspects
   - Lausanne (VD) on Energy, water and social aspects + EPFL campus

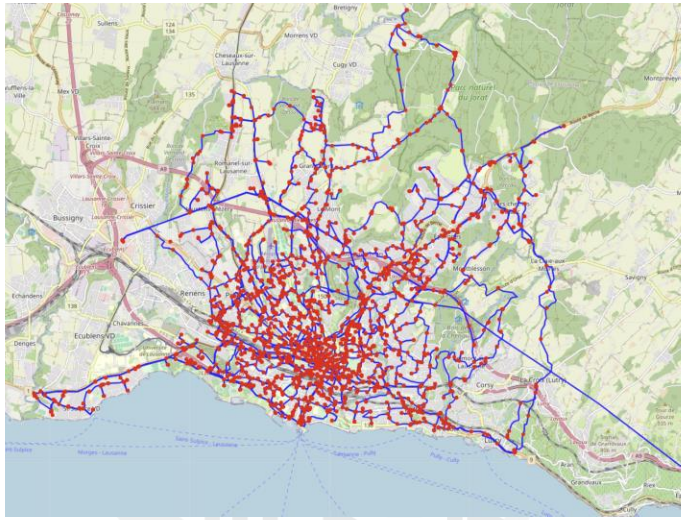


1. Evaluate efficiency of UrbanTwin models and platforms deployment
2. Optimize for sustainability KPIs
3. Demos and prototype of secure and scalable edge-cloud technologies

Key Results

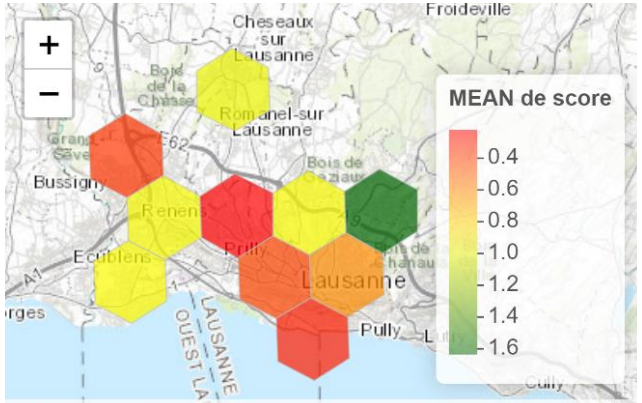
URBANTWIN

UrbanTwin tool v0.1 (platform + database)
Modelling future renewable energy integration infrastructure in cities in collaboration with the Lausanne utility

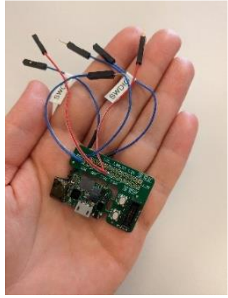


Energy efficient ICT infrastructure
Advances in monitoring and improving efficient use of data centers applied to EPFL EcoCloud

Technology adoption trends
Impact of low carbon policies and behaviours on technology penetration in cities. Application to the city of Lausanne.

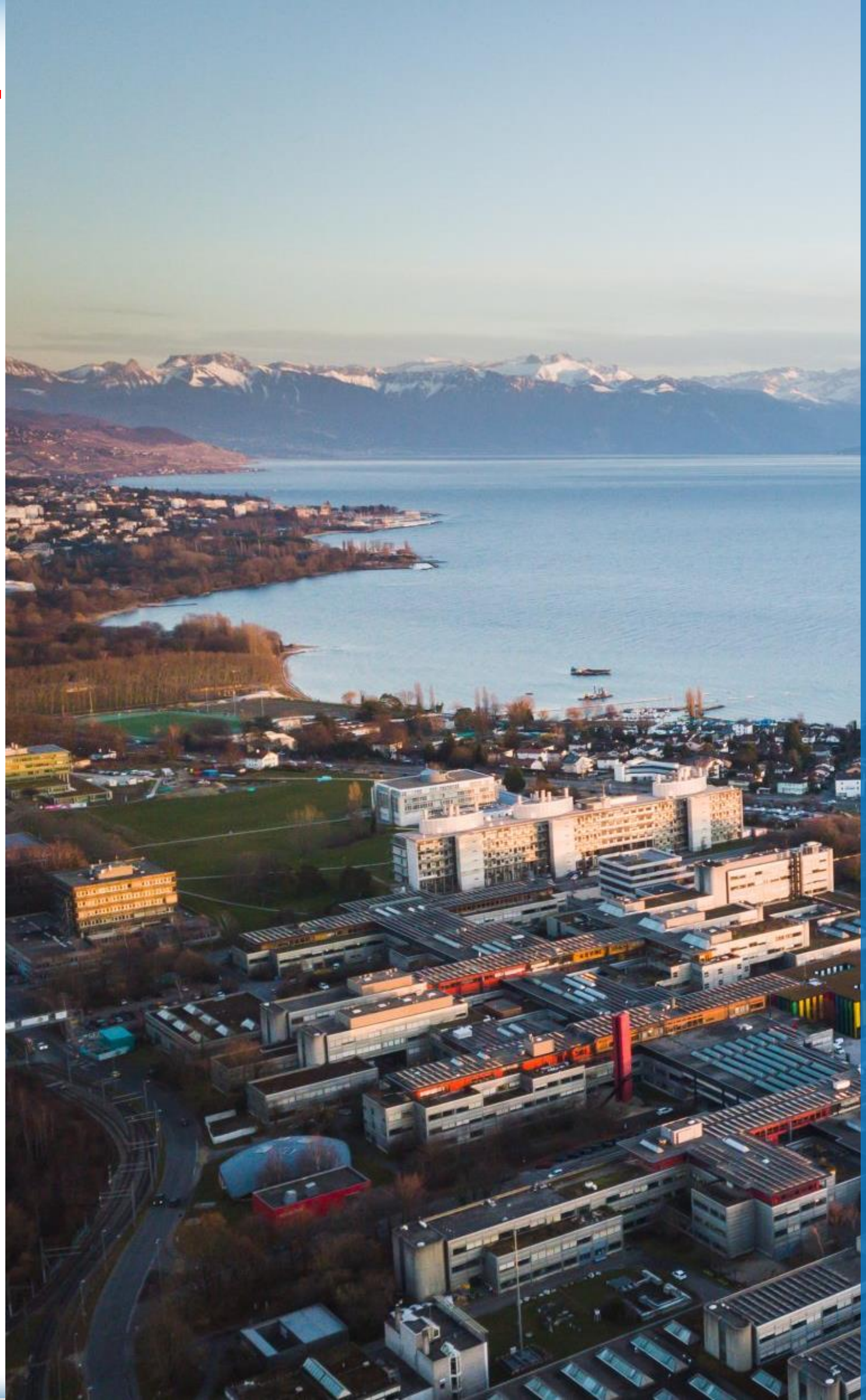


Low-power edge-node prototype
Data sampling and real-time AI-processing with few milliwatts



Water portal
Sensors network for land use scenarios, impact assessment and water volume discharge in urban system, application to the EPFL demo site

- Welcome and introduction, by Prof. Giovanni De Micheli
- EcoCloud's experimental facility, by Dr. Xavier Ouvrard
- Projects that are already using our facility
  - Post Moore Data Centers: Ali Ansari (PARSA)
  - Heating Bits: Enea Figini (DESL)
  - Urban Twin: Dr. Denisa Constantinescu (ESL)
- Initiatives on sustainability:
  - Methodology to measure carbon impact, by Julia Paolini
- EcoCloud Annual Event, by Prof. Giovanni De Micheli



Initiatives on sustainability

Methodology to measure carbon impact

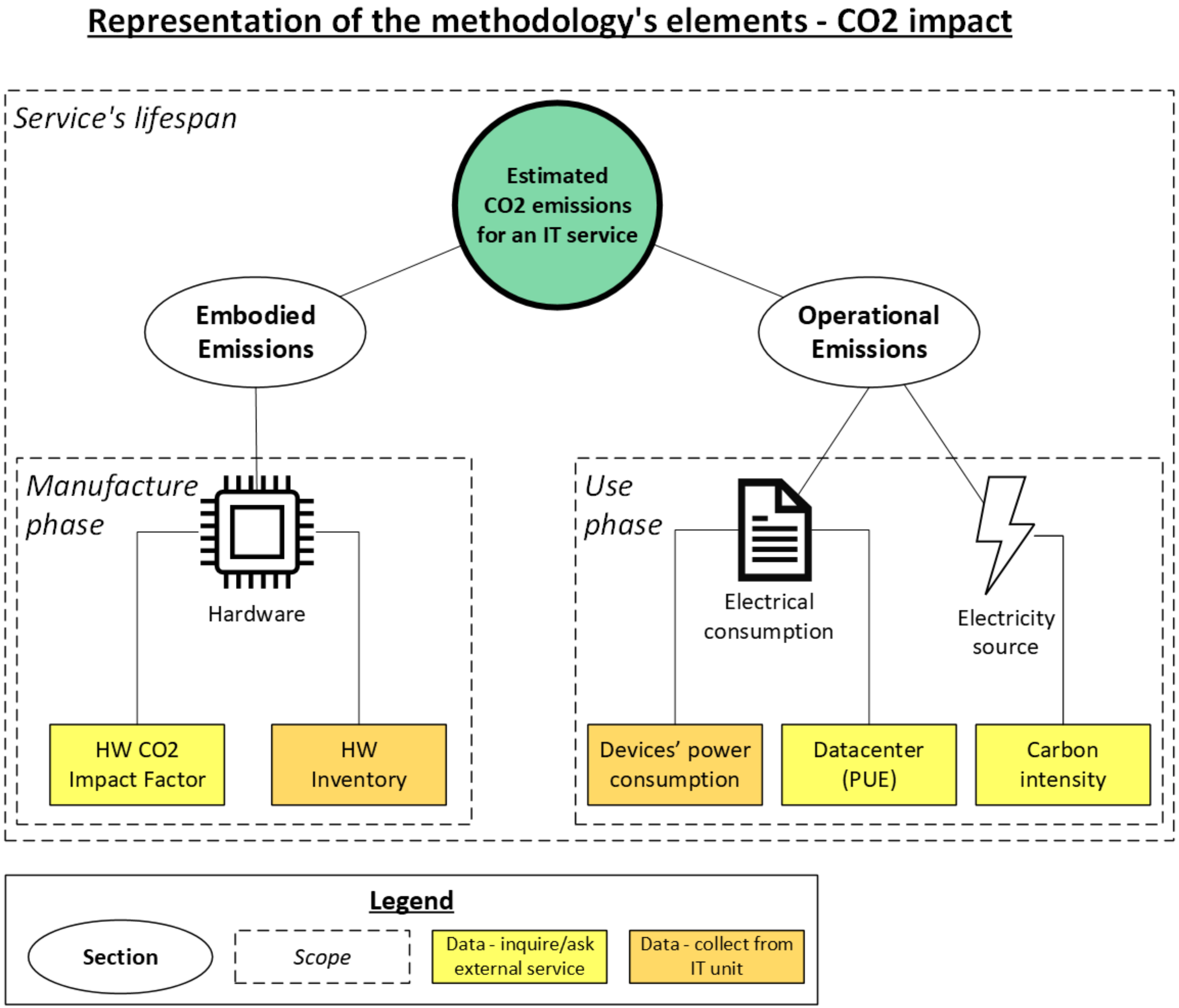
- Context: Shared effort between 3 units: EcoCloud research center, SCITAS scientific platform and Sustainability team

- State of the art of sustainable IT topics from the infrastructure side:
  - Infrastructure's energy efficiency: power, cooling
  - Tools for users to evaluate one's public clouds' environmental impacts in relation to their usage
  - Tools for users to evaluate one's AI's environmental impacts in relation to their usage

- Conclusion:
  - It is very difficult to assess impacts of non-private infrastructure due to the lack of information (quantity and quality)
  - It is possible to have a correct estimation (orders of magnitude) of private infrastructure

- The methodology we are working on will assess the environmental impact of an internal IT service of an institution. The main goal is to be accessible and produce results with realistic orders of magnitude.

Methodology to measure carbon impact

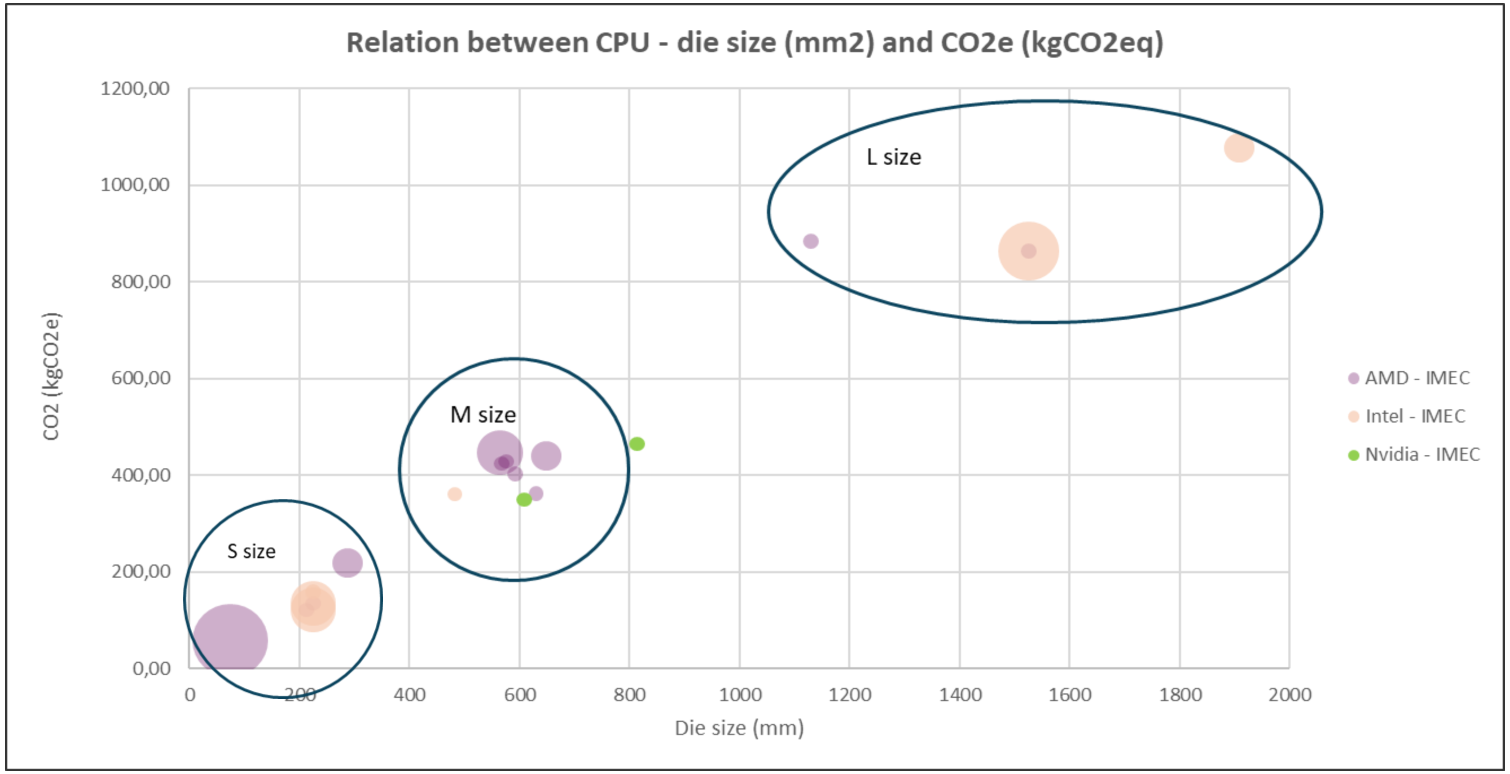


Methodology to measure carbon impact

- Each IT service is composed of different type of servers (computing, administration, etc.)
- Each type of server has several components (CPU, GPU, memory, disk) that can have different specifications (capacity, power, etc.)

To make the job as easy as possible for the future users, we are trying to define impact sizes (S,M,L) and associate corresponding values for each components.

For instance, for CPUs, there is a relation with the die size, lithography and the related CO_2 emissions



Methodology to measure carbon impact - challenges

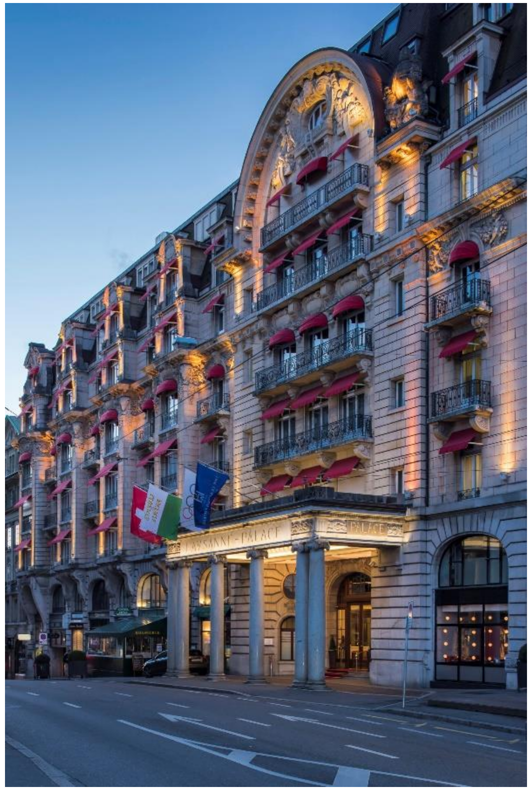
- Embodied emissions:
  - Quality and completeness of inventory
  - Quality, up-to-date and openness of impact factors
  - Choose the right balance between simplicity of use vs precision of the result

- Operational emissions
  - Quality of the in-place monitoring
  - How much effort can be invested by the monitoring team to enhance it?
  - How to calculate the PUE / How reliable the given PUE value is?

- Welcome and introduction, by Prof. Giovanni De Micheli
- EcoCloud's experimental facility, by Dr. Xavier Ouvrard
- Projects that are already using our facility
  - Post Moore Data Centers: Ali Ansari (PARSA)
  - Heating Bits: Enea Figini (DESL)
  - Urban Twin: Dr. Denisa Constantinescu (ESL)
- Initiatives on sustainability:
  - Methodology to measure carbon impact, by Julia Paolini
- EcoCloud Annual Event, by Prof. Giovanni De Micheli

Upcoming EcoCloud Event 2025:

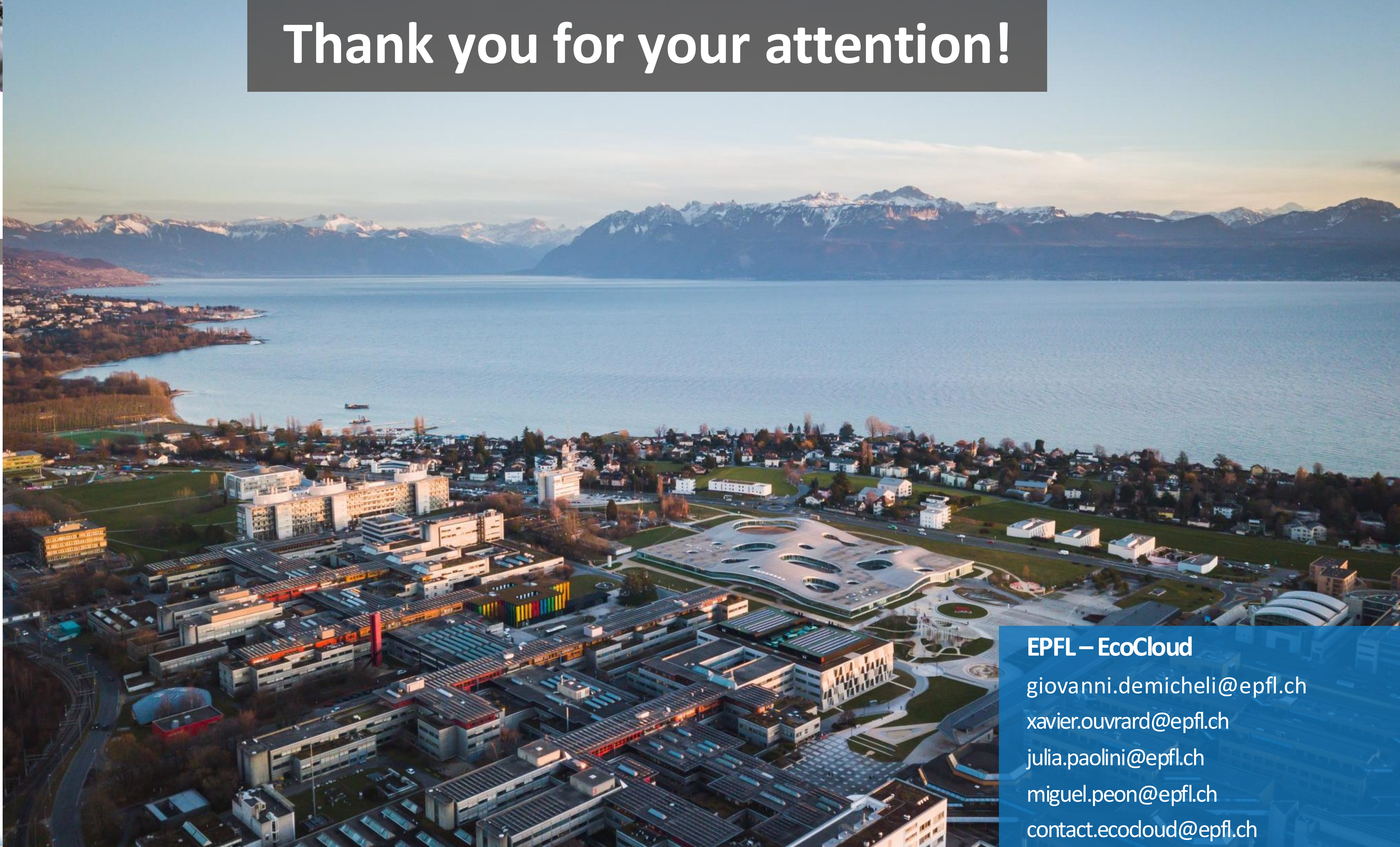
- When? **Tuesday, Oct 8th, 2025**
- Where? Lausanne Palace, Switzerland
- In-Person agenda with networking focus on 4 new multi-center projects and start ups from EPFL:
  - New trends on energy/carbon footprint of Machine Learning
  - New cooling and energy/heat recovery tech.
  - Virtual twins for cities (FUSTIC, CIS and CLIMACT)
- Keynote speaker:
  - Prof. Subhashish Mitra from Standford University



We are looking forward to seeing you there!

Thank you for your attention!

Questions?



EPFL – EcoCloud

giovanni.demicheli@epfl.ch

xavier.ouvrard@epfl.ch

julia.paolini@epfl.ch

miguel.peon@epfl.ch

contact.ecocloud@epfl.ch