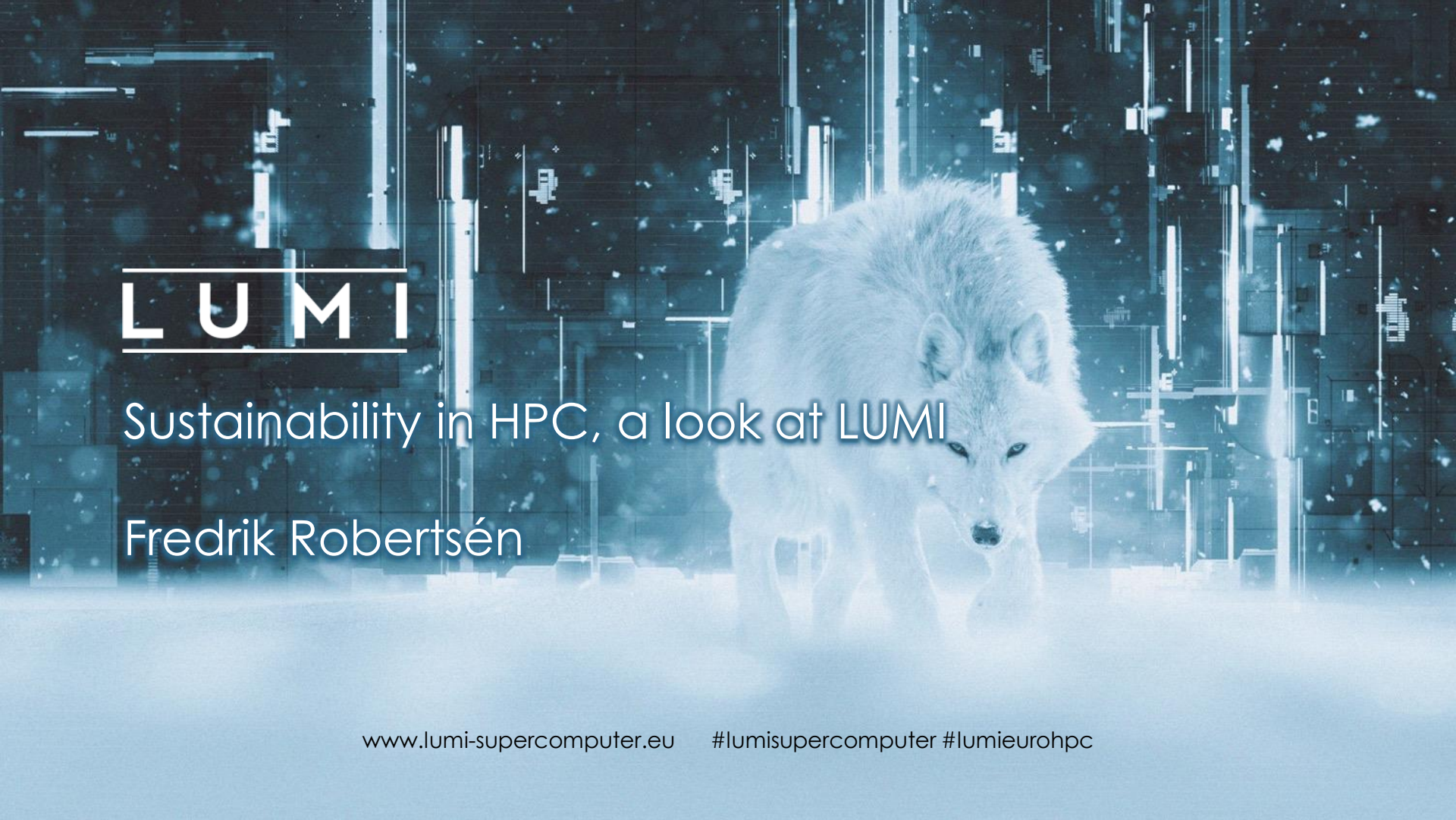


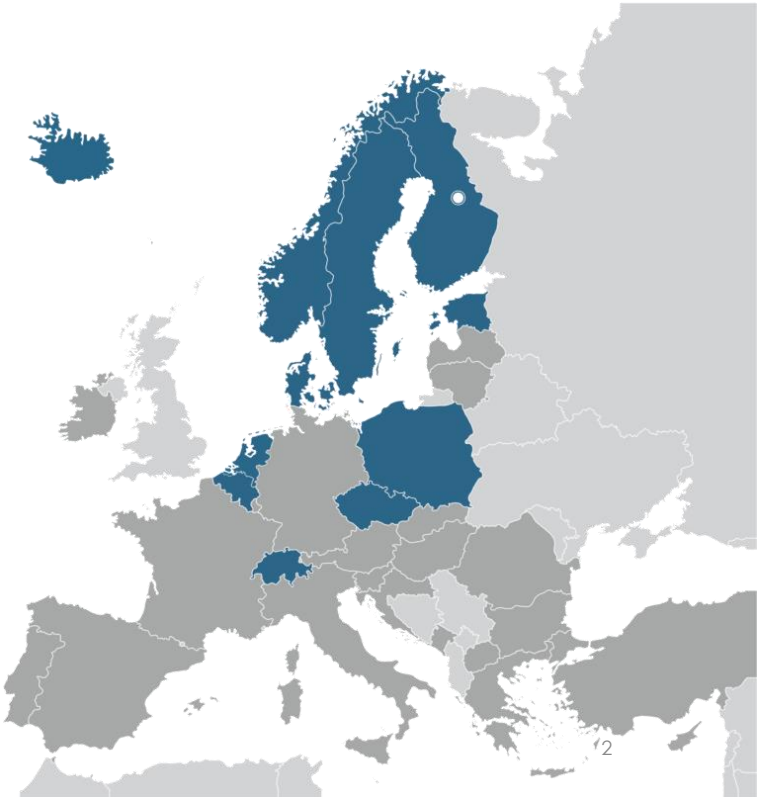
Sustainability in HPC, a look at LUMI

Fredrik Robertsén

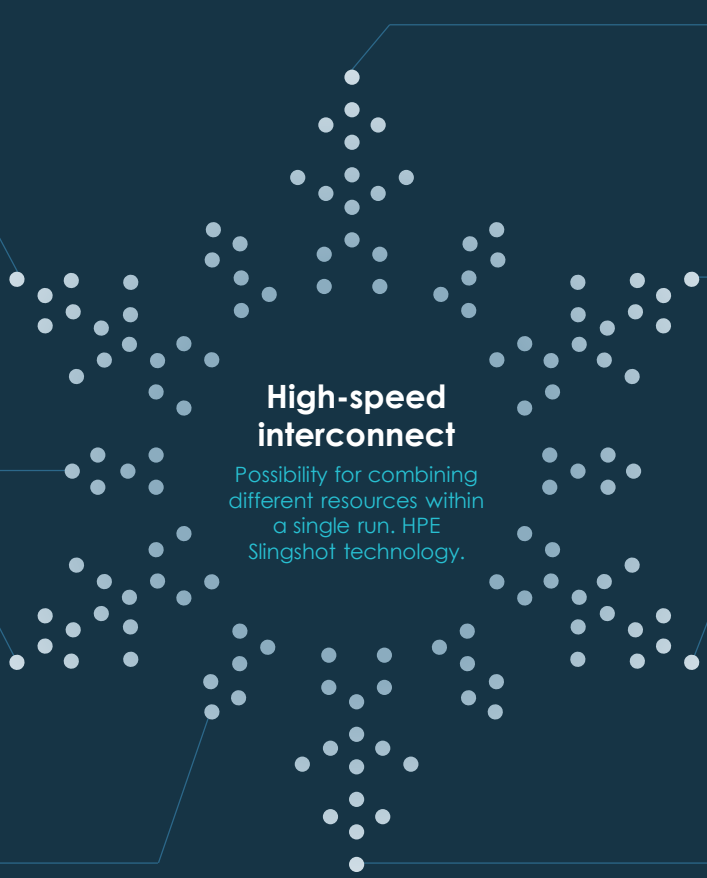
www.lumi-supercomputer.eu #lumisupercomputer #lumieurohpc

LUMI and EuroHPC intro

- EuroHPC is an EU initiative to increase the unions HPC capacity, started 2018
- EU funds part of the machines, and gets a share of the resources to grant to European researchers
- LUMI is one of the pre-exascale machines
- Based on a consortium of 11 countries, all paying some share of the machine
- The machine is hosted by CSC in Kajaani Finland



LUMI architecture



LUMI-C:
x86 Partition

Supplementary CPU partition:
over 262,000
AMD EPYC CPU cores.

LUMI-K:
Container Cloud Service

LUMI-O:
Object Storage Service

30 PB
encrypted object storage (Ceph) for storing, sharing and staging data.

LUMI-Q:
Quantum Computing

High-speed interconnect

Possibility for combining different resources within a single run. HPE Slingshot technology.

LUMI-G:
GPU Partition

Sustained performance
380
Pflop/s powered by AMD Radeon Instinct™ MI250X GPUs.

LUMI-D:
Data Analytics Partition

Interactive partition with
32 TB
of memory and graphics GPUs for data analytics and visualization.

LUMI-F:
Accelerated Storage

10 PB
Flash-based storage layer with extreme I/O bandwidth of 2 TB/s and IOPS capability.

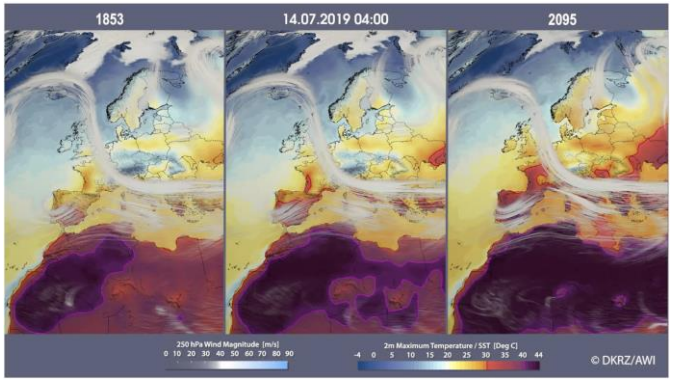
LUMI-P:
Lustre Storage

80 PB
parallel file system.

A wide user base

- Users from our 11 consortium countries
- EuroHPC users from all over Europe
- Currently:
  3500 users
  1600 projects
- Largest 2023 projects:
  - The continuum limit of lattice QCD for high-precision tests of the standard model of particle physics … (JU, 5M GPUh)
  - Deep Generative Language Modeling for Finnish (FI, 5M GPUh)
  - Destination Earth / DestinE climate DT and associated use (JU, 4M GPUh)
  - SISSI: Studying Small-Scale dynamo action in the Sun (FI, 3,3M GPUh)
  - Open source large language models (FI, 2,4M GPUh)
  - LUMI-G benchmarks of Grid and GPT (JU, 2,9M GPUh)
  - LumiLLM - LUMI Large Language Models (FI, 2,3M GPUh)
  - GRaCoS Gravitational radiation from Cosmic Strings (FI, 2,2M GPUh

GPUh = 1x MI250x utilized in 1h

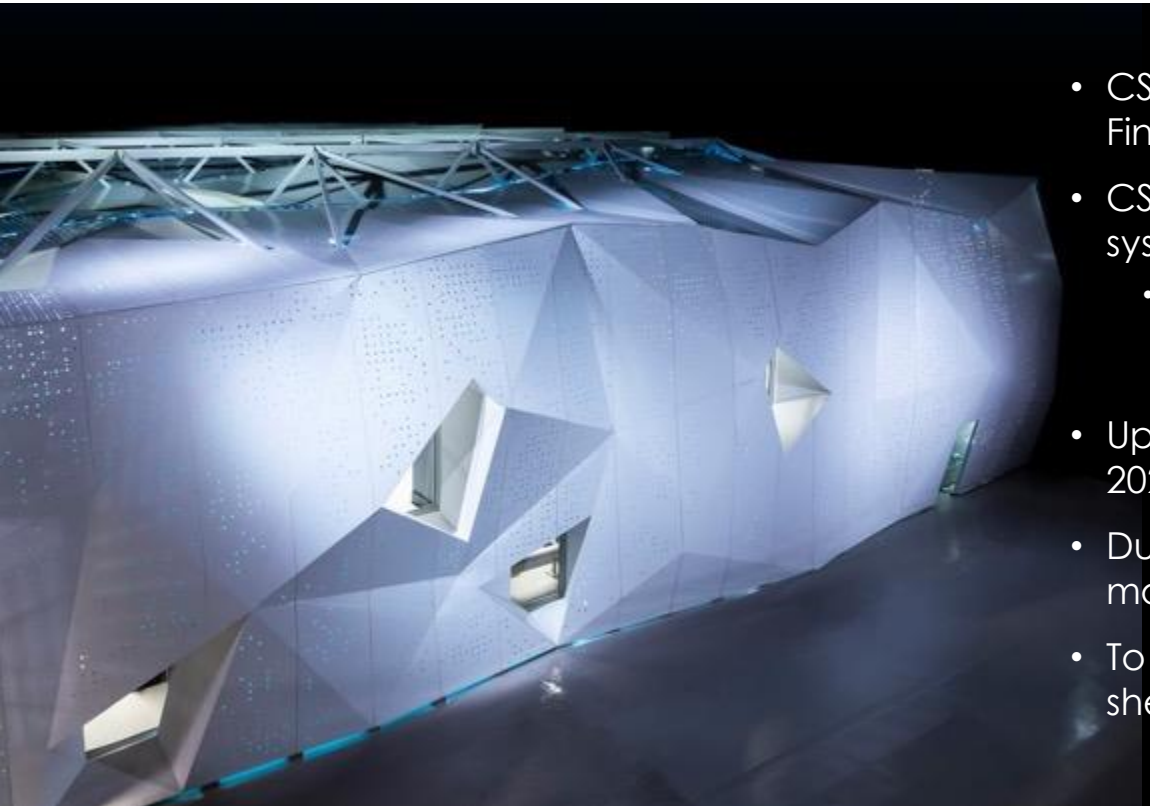


Picture: T. Jung (AWI)

Renforsin Ranta Business Park, Kajaani, Finland

- Old paper mill moved out 2008, left behind large empty halls and a robust electricity infrastructure, 200MW available on site, redundant feeds from 3 directions
- The climate in Kajaani is not warm, daily mean temperature of 2.6c, 36f
- Low energy costs, and transfer pricing, power availability

Kajaani DC evolution



- CSC hosts and maintains HPC systems for Finnish academia
- CSC installed new Finnish academic systems in Kajaani in 2012
  - At that point hosted in one of the storage buildings, and cooling was with just dry (ish) coolers, or air cooling
- Updated the national systems in 2019 to 2020
- Due to the space and power needed we moved into one of the main machine halls.
- To house the machine we constructed a shell inside the machine hall

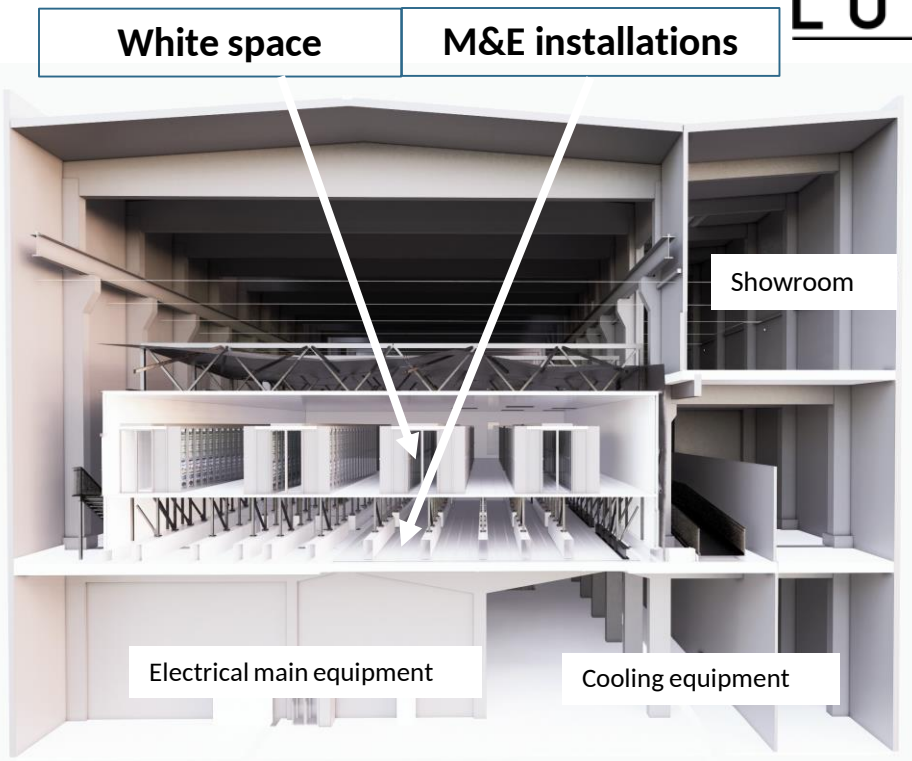
LUMI data center

LUMI facility overview

- 5800m² in three floors
- 800m² whitespace for IT devices
- Power capacity 15MW at full buildout
- 14 000m² free space for future expansions

Data center cooling

- Mechanical cooling with 3 heat pumps with 7.2MW total cooling capacity and 9MW of total heating capacity
- 32°C inlet for DLC HPC cabinet CDUs
- Free cooling, total capacity 10MW, dry coolers no water consumption
- Year around free cooling possible



Heat reuse

Intro into district heating

LUMI

- Centralized heat source, heating water that is then piped to residential and commercial spaces
  - Used for heating and warm water production
  - 50% of the total heating market in Finland, 90% of apartment blocks and half of all terraced houses have a district heating connection
- Heat is usually captured from waste heat from industry and power generation
- The papermill already produced heat and the current district heating boiler is in the same industrial park as our DC
- Just pipe our cooling water into that
  - The problem is that our outlet temperature is around 40c, district heating is far above that
- Need heat pumps to raise the water temperature

LUMI data center statistics 2023

Metric	
Free cooling PuE	<1.05
PuE with heat re-use	1.32
Annual average PuE	1.24
Heat re-use COP	4
ERE (Energy reuse efficiency, lower better)	0.52 (0.20*)
ERF (Energy reuse factor, higher better)	0.58 (0.84*)
Annual heat production	26,72 GWh
Reduced co2 emissions	2895 co2 tonnes
Source of electricity	100% hydro power

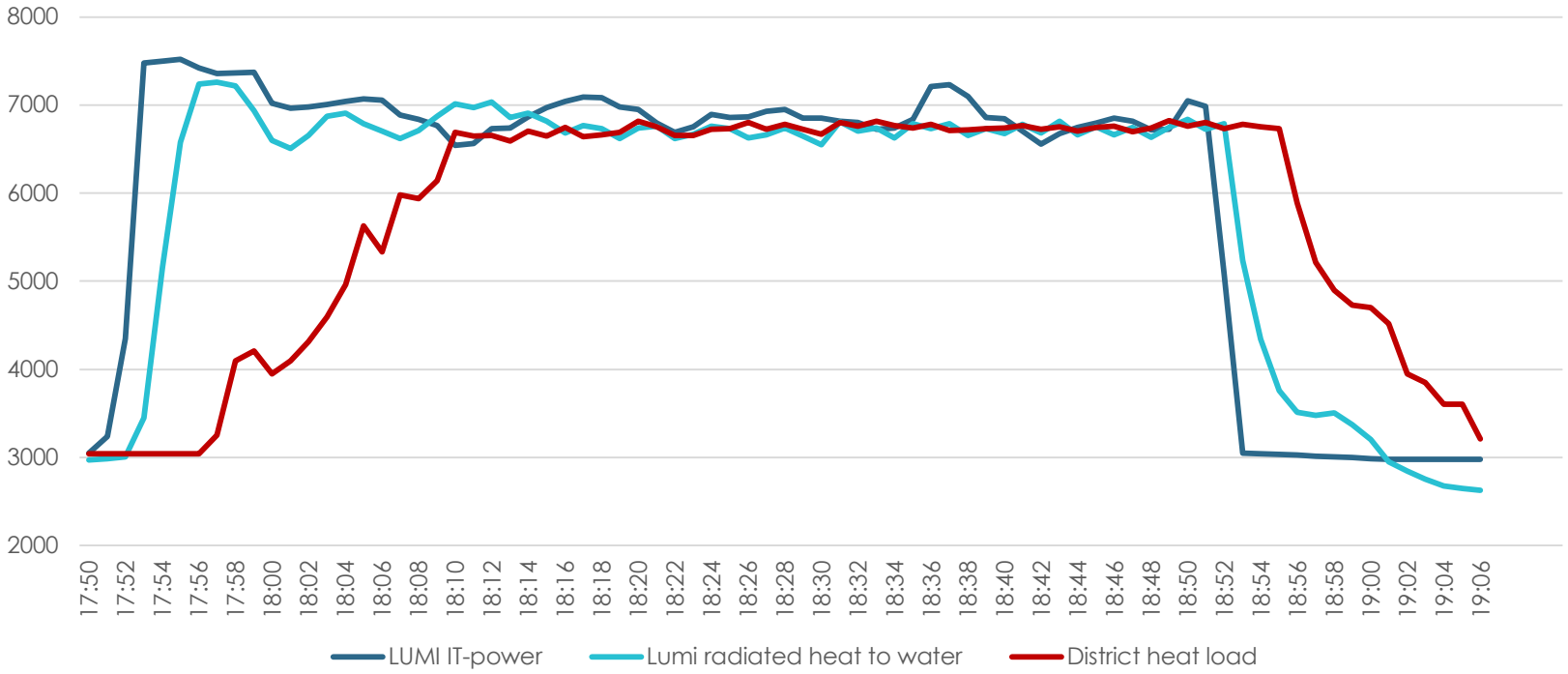
*Best monthly ERE and ERF on 2023

Waste heat utilization

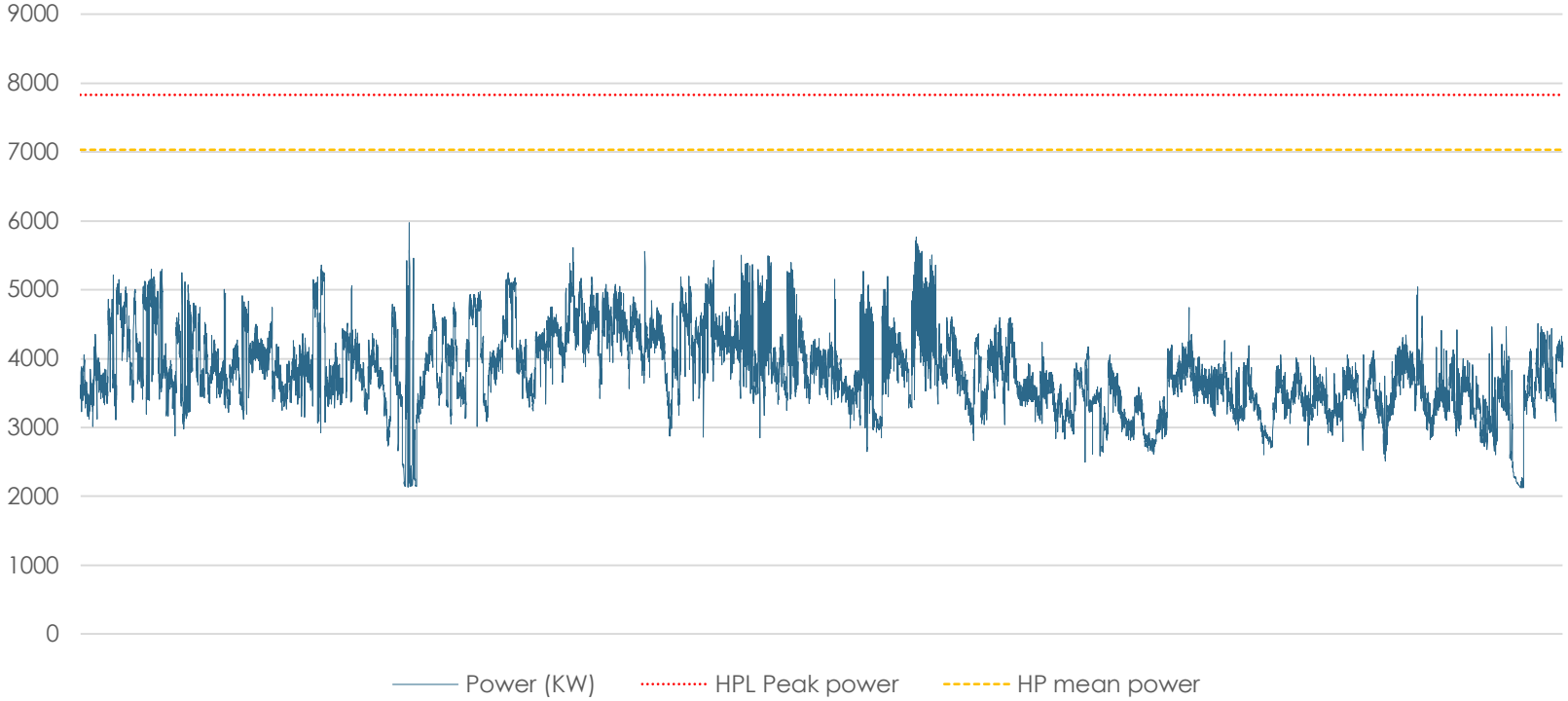
LUMI

- Heat pumps has required lot of maintenance and adjustment to get efficiency on reasonable level
- Heat pumps running most of time with partial load 60..85% on max, i.e. less efficient
- Heat pumps are difficult to operate on HPC load swings
  - Applies mostly to GPU partition where load swings are larger based on system utilization
  - Buffer tank or another mechanism to CDU outlet side to align heat load variation to heat pumps
- We have 3 heat pumps, turning off/on one is a large change in capacity
  - It might be more efficient to run just N and radiate some heat through dry coolers instead of turning on one more
  - Many smaller heat pumps would give more flexibility but it is questionable if it would be cost efficient
- Outlet temperature varies based on IT-load of HPC system -> hard to maintain optimal circumstances for heat production
- Many interesting modeling challenges
  - How to run the heat pumps
  - Modeling of system load is hard, too many users with too much of a spread in what they run

Example of power usage and heat rejection delay

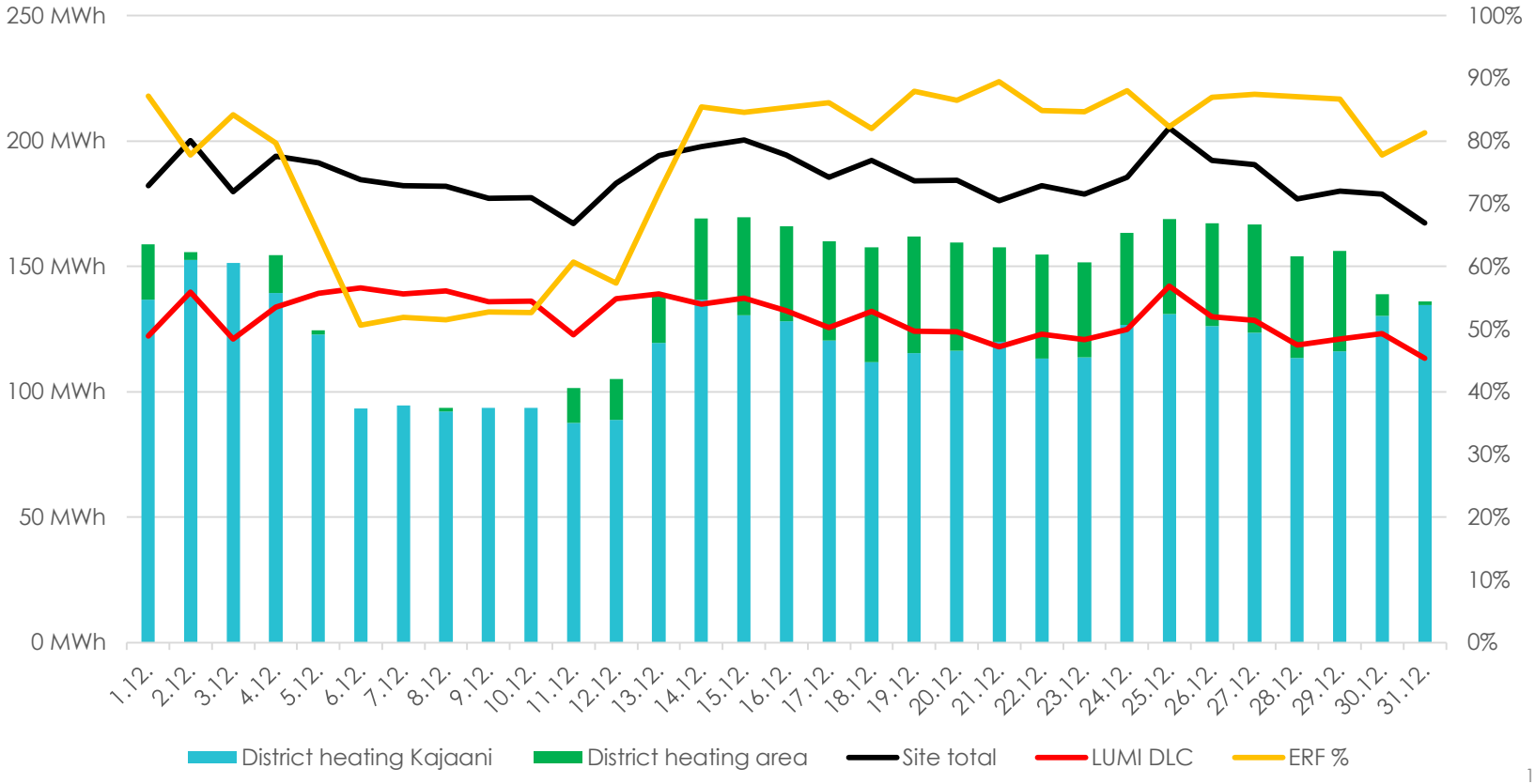


GPU power in HPL compared to real life power usage



Real workloads ~40..60% from HPL power usage on GPU partition, maximum power consuption after HPL 5,808MW
4 months of production usage

LUMI data center energy consumption and district heating



Future

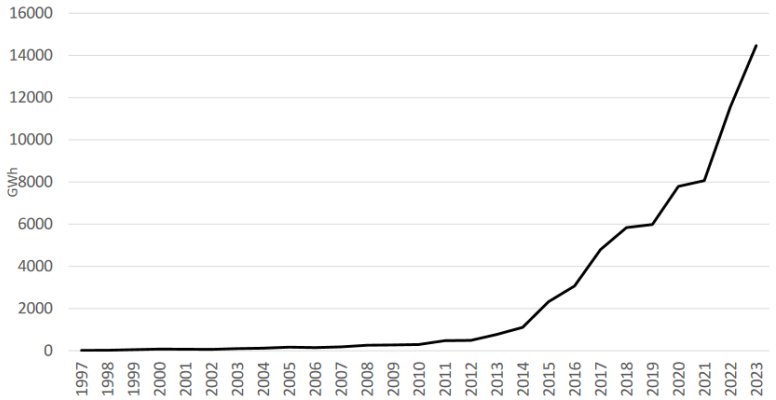
Heat re-use in future and drivers to implement heat re-use

- Energy taxation classes for data centers drives more energy efficient data centers with lower Opex
  - Energy taxation Class 1, 22.53 €/MWh, Class 2, 6.3 €/MWh
- ERE
  - Data center with 0.5…5MW IT-power calendar year average, ERE < 0.90
  - Data center with 5…10MW IT-power calendar year average, ERE < 1.00
- PuE
  - Data center with 0.5…10MW IT-power calendar year average PuE <1.25
  - From 2026 Data center with 0.5…10MW IT-power calendar year average PuE <1.20
- Drives investment in datacenters in Finland, Microsoft moving next to Helsinki
  - Will do waste heat reuse and heat the city
  - If nothing else it is a good image for them
- The energy company in Kajaani announced in July 2023 that they are investing in new carbon neutral heat production
  - New electrical boiler and heat pumps will be installed
  - We won't need to operate heat pumps
  - The energy company can take waste heat form multiple sources and smooth out our swing loads

Solar and Wind power

- We are far enough north that with current tech solar is out of the questions
  - Shortest day in Kajaani during the winter is about 4h, and then you hope its not cloudy…
- But there is a lot of wind power being constructed and even more planned in the north of Finland
  - 7.3 GW in production now, expected to hit about 10 GW in 2025,
  - 130GW "planned", different stage of planning but even with a fraction of that it is still a lot
- Switching to buying wind energy would in practice mean we need to pay more when there is no wind
- 24h lead on the electricity price
- Not economic to shut the machine down
  - Full shutdown wastes about 12-24h of production
- Maybe possible to down clock, but that has other issues
  - Slower execution, need to adjust time limits for jobs

Wind energy production in Finland



Issues from the machine side

LUMI

- We want higher inlet and outlet, 40c inlet for free cooling in larger areas of the world
  - Even with dry coolers
- Tcase (the temperature at the top of the IHS/chip) is dropping
- One reason is due to stacked silicon, HBM stacks, stacked cache, future stacked logic
- Allowed temperature is dropping current discussion seems to be to about 50c, currently we are at around 70c
- Can be mitigated by high flow rates in the systems requiring large pumps, energy wasted on pumping
  - Or more exotic cooling solutions, phase change, new cold plates
- Rack delta T will be low, currently we are at about 10c-12c on lumi
  - Lower delta T will affect heat pump efficiency, or we may even end up in a situation where its just not doable efficiently

Conclusions

LUMI

- Geographic location of datacenters will be more important in the future
  - Cheap, sustainable electricity
  - Efficient cooling, and how that will evolve as the climate changes
- Heat reuse
  - Actually do something with the heat produced
    - Don't just radiate it to the ambient air, or evaporate a bunch of water
  - Not trivial, lots of processes that needs to be optimized
  - Very hard to model the actual load
- New evolutions in chip manufacturing is making things harder
  - We want high inlet temperatures, but the chips effective operating temperatures are dropping
  - New cooling methods

LUMI

Fredrik Robertsen

CSC – IT Center for Science

Follow us

X: @LUMIhpc

LinkedIn: LUMI supercomputer

YouTube: LUMI supercomputer

www.lumi-supercomputer.eu

contact@lumi-supercomputer.eu

EuroHPC Joint Undertaking

The acquisition and operation of the EuroHPC supercomputer is funded jointly by the EuroHPC Joint Undertaking, through the European Union's Connecting Europe Facility and the Horizon 2020 research and innovation programme, as well as the of Participating States FI, BE, CH, CZ, DK, EE, IS, NO, PL, SE.

Leverage from the EU 2014–2020

European Union
European Regional Development Fund

REGIONAL COUNCIL OF KAINUU

EURO