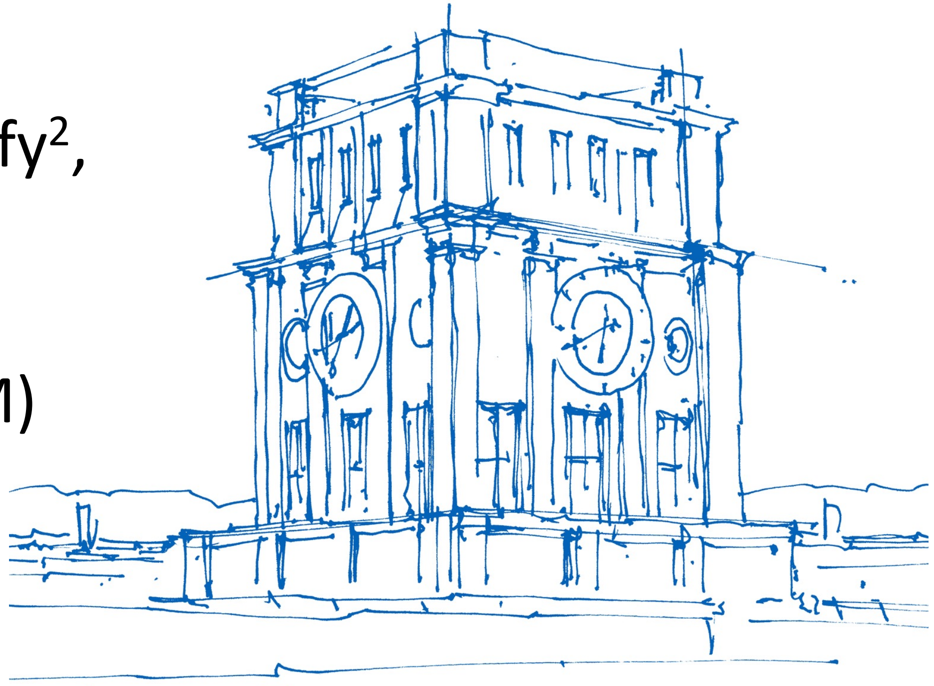
Sustainability in HPC: Vision and Opportunities

Mohak Chadha[1], **Eishi Arima**[1], Amir Raoofy[2],
Michael Gerndt[1], Martin Schulz[1,2]

[1] Technische Universität München (TUM)
[2] Leibniz-Rechenzentrum (LRZ)



Uhrenturm der TUM

Sustainable Supercomputing
SC'23, Nov 12, 2023

Executive Summary

We will discuss potential solutions to reduce the carbon footprint of supercomputers from various perspectives (this is a position paper)

- Carbon Emission = ~~Direct (Scope 1)~~ + Indirect (Scope 2/3)
- Scope 2: Operational Carbon
- Scope 3: Embodied Carbon

Carbon Reduction

Scope 2

Scope 3

Dynamic Power Bound Scaling

Dynamic Resource Scaling

Scheduling & Checkpointing

Making HPC Users Greener

Component (or Chip) Design

System Design & Procurement

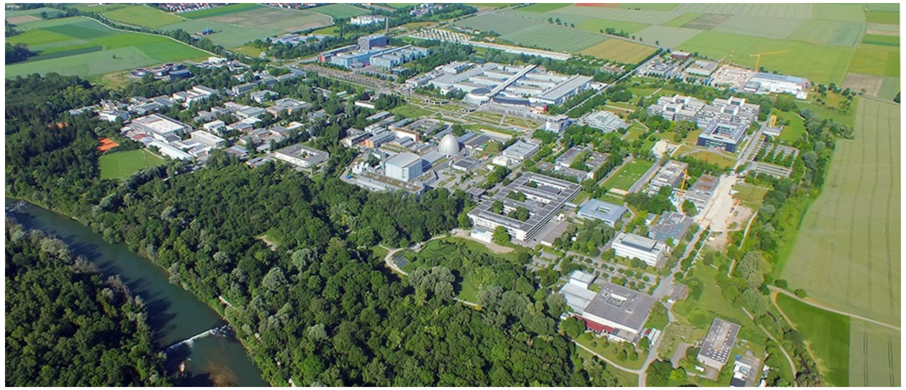
System Lifetime, Reuse, & Recycle

DEEP-SEA

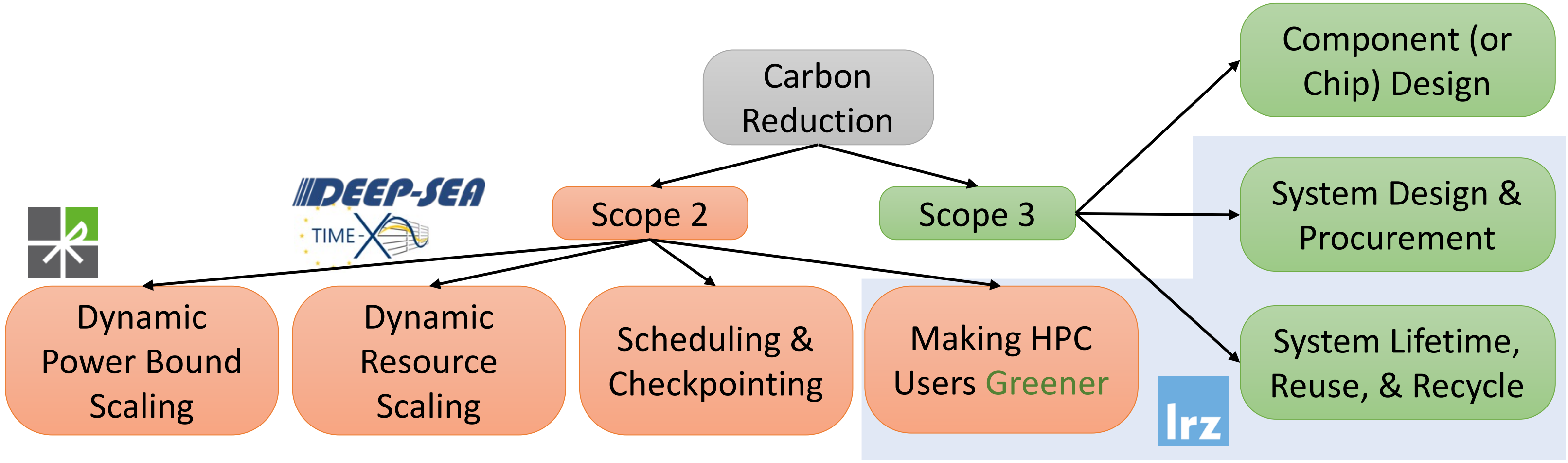
TIME-X

lrz

Outline

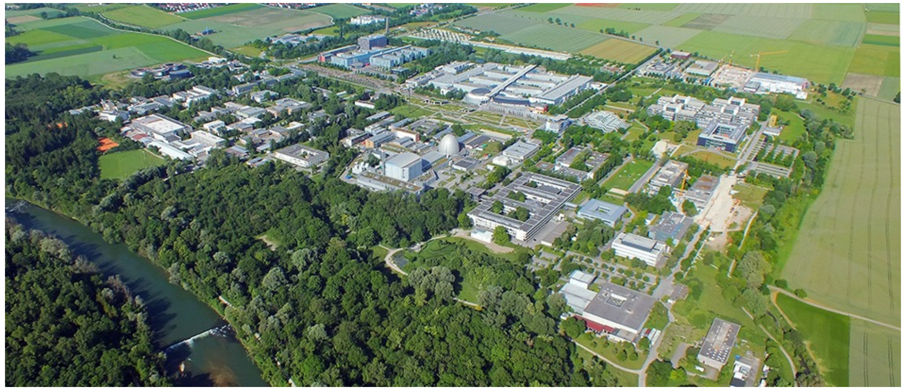


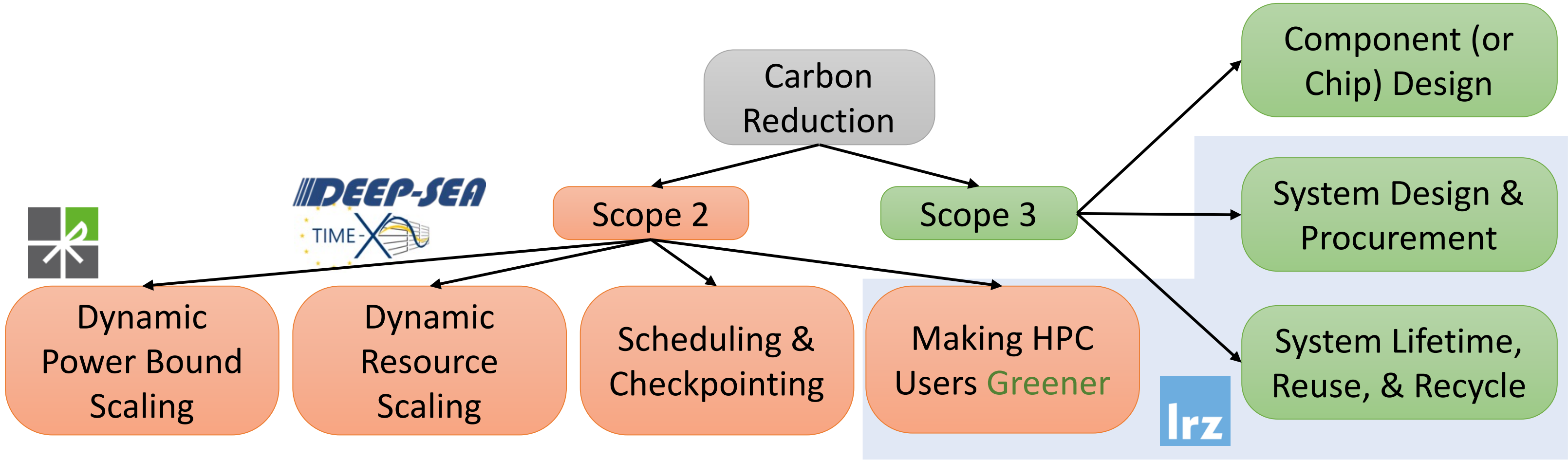
- Background
- Reducing Embodied Carbon (Scope 3)
- Reducing Operational Carbon (Scope 2)
- Conclusion & Acknowledgement



Outline

- Background
- Reducing Embodied Carbon (Scope 3)
- Reducing Operational Carbon (Scope 2)
- Conclusion & Acknowledgement





Needs for Carbon-awareness in Supercomputing

- Supercomputers indirectly account for a tremendous amount of carbon emissions
  - Mainly due to the scale: **(1)** consist of over 100K of components; **(2)** consume few 10s MW of power
- Need to limit their carbon emissions while following the GHG (Green House Gas) protocol

GREENHOUSE GAS PROTOCOL

Carbon Awareness

Fugaku
0.44Exa @30MW

Frontier
1.2Exa @23MW

Aurora
2Exa? @60MW?

Future Green Supercomputers
(Image generated by deep AI)

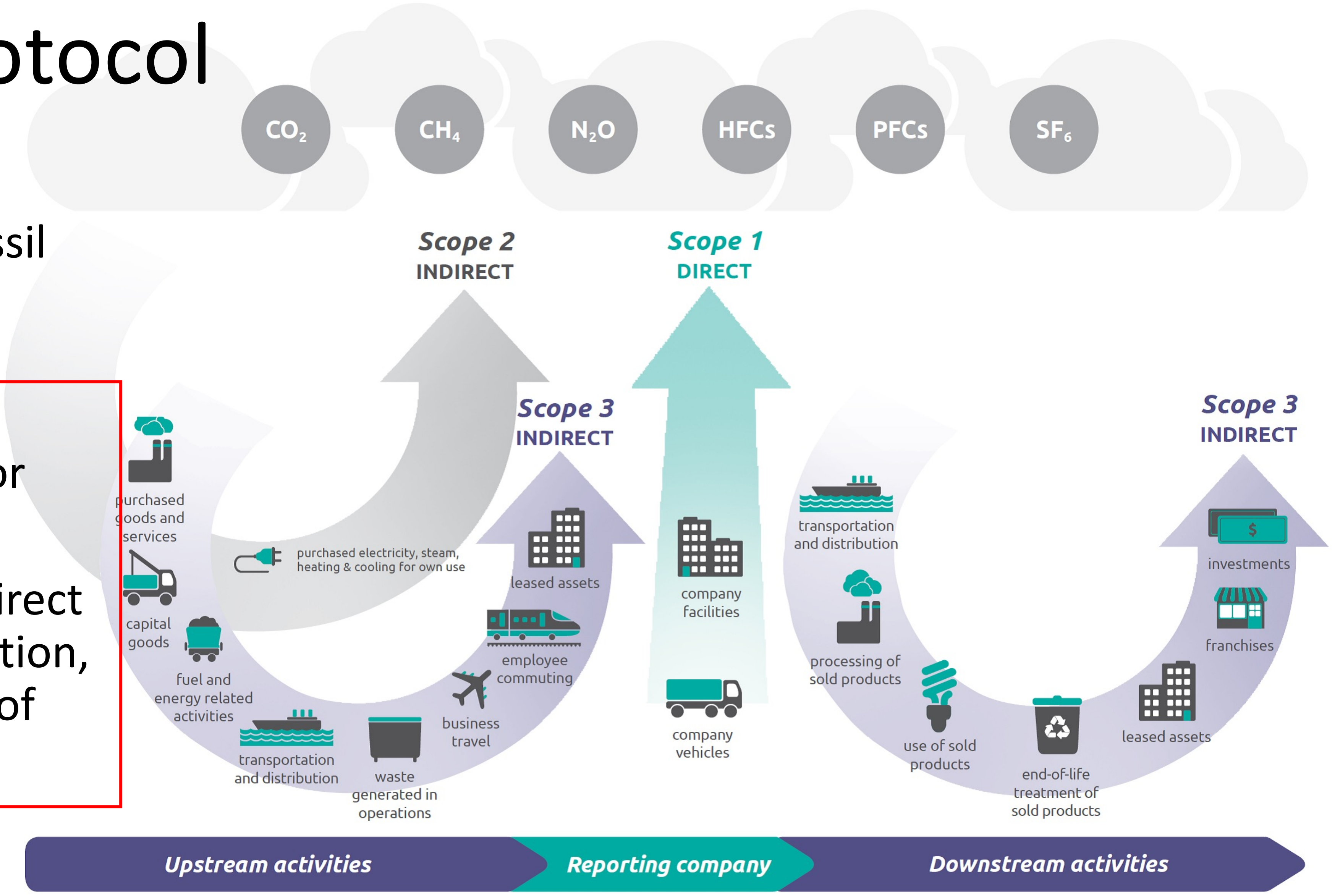
The GHG Protocol

Scope 1: Relevant if fossil fuels are burnt on site (usually not the case)

Scope 2: Electricity, heating, cooling, etc. for the site operation

Scope 3: The other indirect emissions, e.g., production, shipping, and disposal of system components

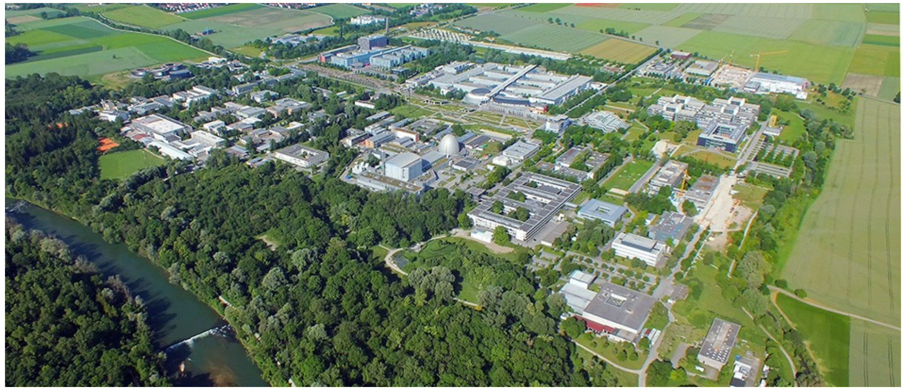
Our Focus

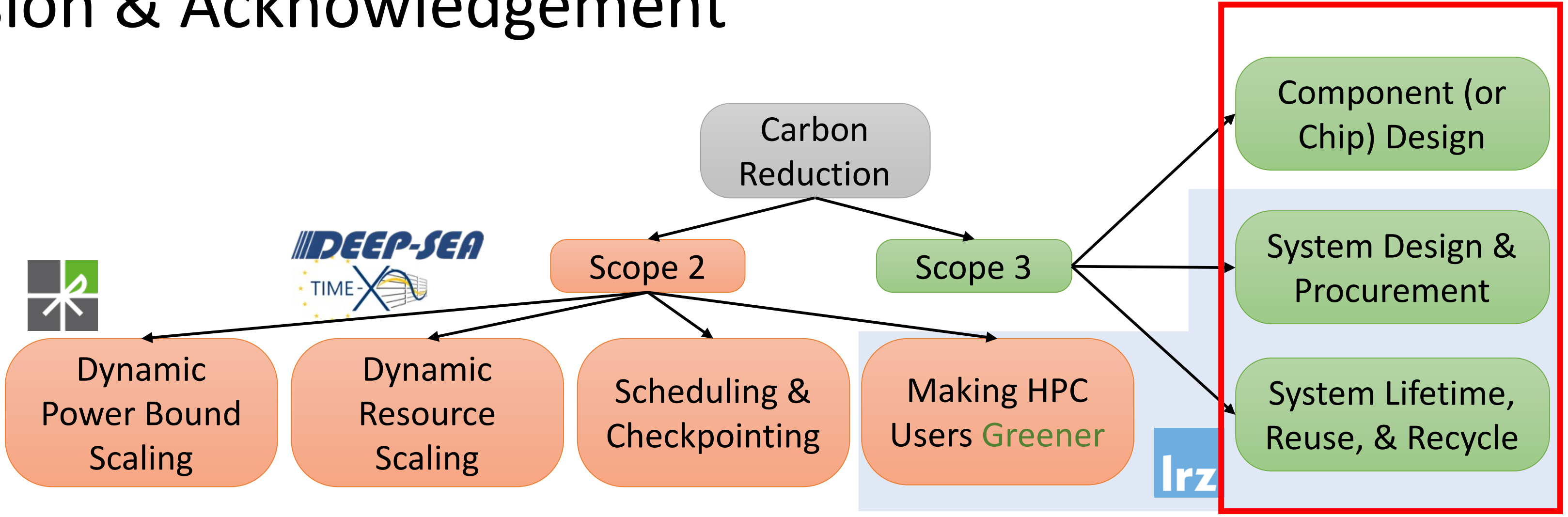


Source: https://ghgprotocol.org/sites/default/files/standards/Corporate-Value-Chain-Accounting-Reporing-Standard_041613_2.pdf

Outline

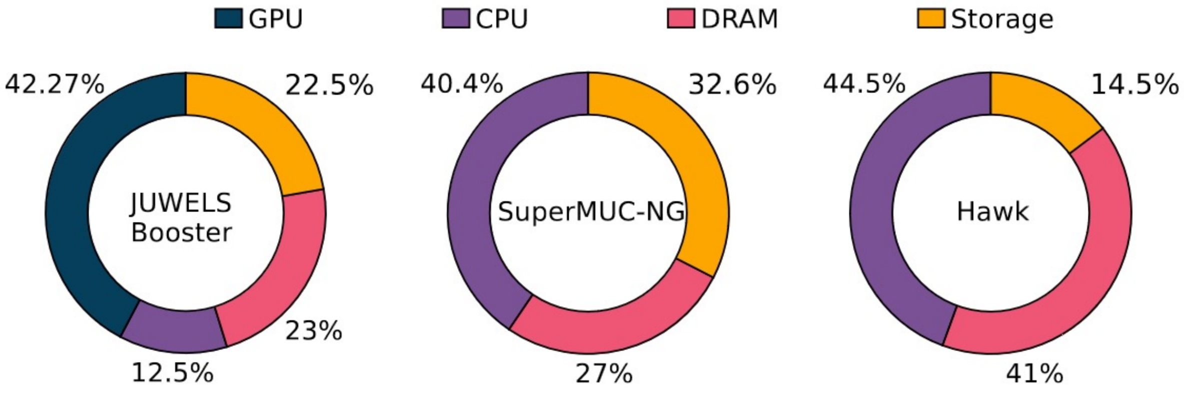
- Background
- Reducing Embodied Carbon (Scope 3)
- Reducing Operational Carbon (Scope 2)
- Conclusion & Acknowledgement





Reducing Embodied Carbon (Scope 3)

- Embodied carbon consists of production, transportation, & disposal
  - The production part (manufacturing + packaging) is dominant [U. Gupta+ ISCA'22]
  - The embodied carbon trend differs across different systems
    - The system architecture & procurement (incl. product choices for components) matter
- **Promising Approaches for Embodied Carbon Reduction:**
  1. Carbon-aware Component Design
  2. Carbon-aware System Architecture & Procurement
  3. System Lifetime Extension & Reuse/Recycling



Estimated Embodied Carbon Footprint Breakdowns
Calculated based on a prior study [B. Li + SC'23]

Designing Carbon-efficient HPC Components

Conventional Chip Design:

max Perf(W, d)
 s.t. Pow(W, d) ≤ Pow_{Target}
  Cost(d) ≤ $Cost_{Target}$
  d ∈ D

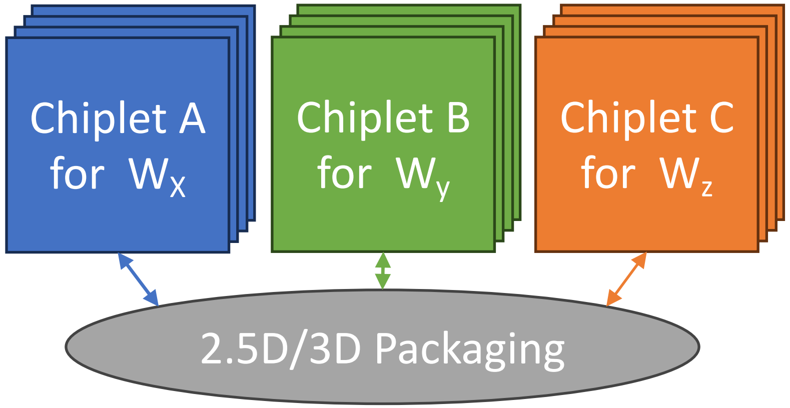
In: W, Out: d_{opt} (∈ D)

Carbon-aware Chip Design:

max Perf(W, d)
 s.t. Pow(W, d) ≤ Pow_{Target}
  Cost(d) ≤ $Cost_{Target}$
  EmCrbn(d) ≤ $EmCrbn_{Target}$
  OpCrbn(W, d, CI) ≤ $OpCrbn_{Target}$
  d ∈ D

In: W, CI, Out: d_{opt} (∈ D)

Hierarchical Optimization for 2.5D/3D Integrations:



W: Target workloads or proxy apps
D: A set of design points to explore
d (∈ D): A set of design parameters
X_{Target}: Target value for metric X
EmCrbn/OpCrbn: Embodied/operational carbon
CI: Typical carbon intensity of the target system

Alternative design goals [U. Gupta+ ISCA'22]:
Carbon-delay product
Carbon-$delay^2$ product

Embodied Carbon in Production ≃ Manufacturing + Packaging

Both inter- and intra-chiplet optimizations matter!

* inter: manufacturing part, intra: packaging part

System Architecture and Procurement

System Optimization

Goal: Max SystemThroughput(Workloads)

Constraints:

Embodied Carbon Constraint

$$\sum EmbodiedCarbon_i \leq Target$$

Best in Performance

The optimal system design will change by taking carbon into account!

The decision will be even harder due to the emerging architectures/devices!

Embodied Carbon Budget + Operational Carbon Budget

Trade-off

∝ System's Energy Budget

B

B

Perf: 1.5, Pow: 1
Cost: 1, EC: 1

The same procedure goes for all the other components

Best in Embodied Carbon

System Lifetime, Reuse, and Recycling

System Lifetime: Typically 4-6 years

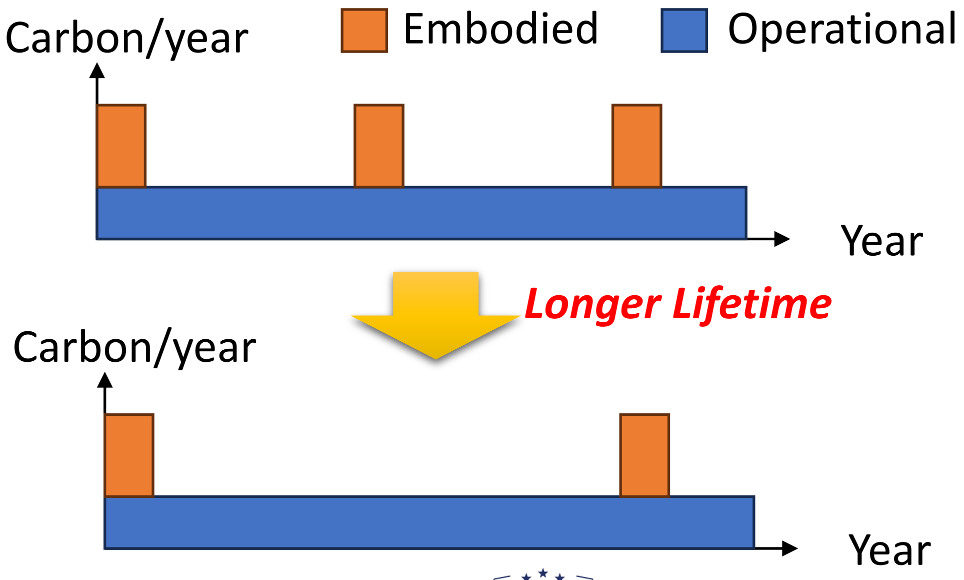
- Extending the lifetime will contribute to the embodied carbon reduction (right fig)

Reuse & Recycling: for reducing the carbon emissions in disposal & production

- **Reuse:** LRZ offers decommissioned machines to other public institutions for free
- **Recycling:** DRAM chips (DDR4=>DDR5), heat pipes, etc.

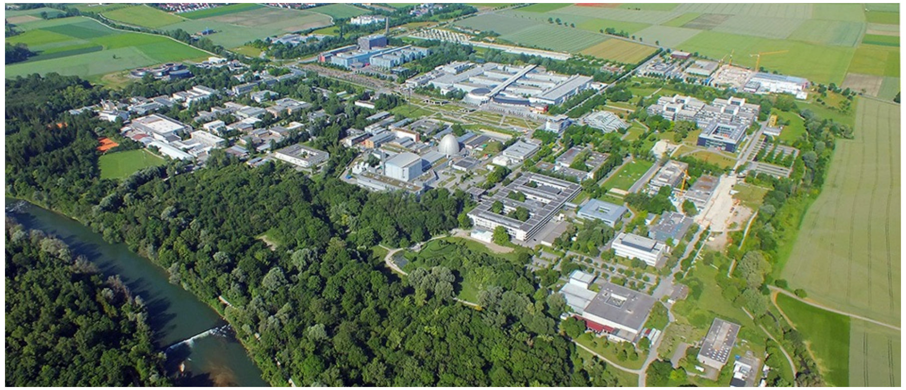
HPC System	Start of Operation	Decommissioned
SuperMUC	2012	2018
SuperMUC Phase2	2015	2019
SuperMUC-NG	2019	2024
SuperMUC-NG Phase2	2023	–

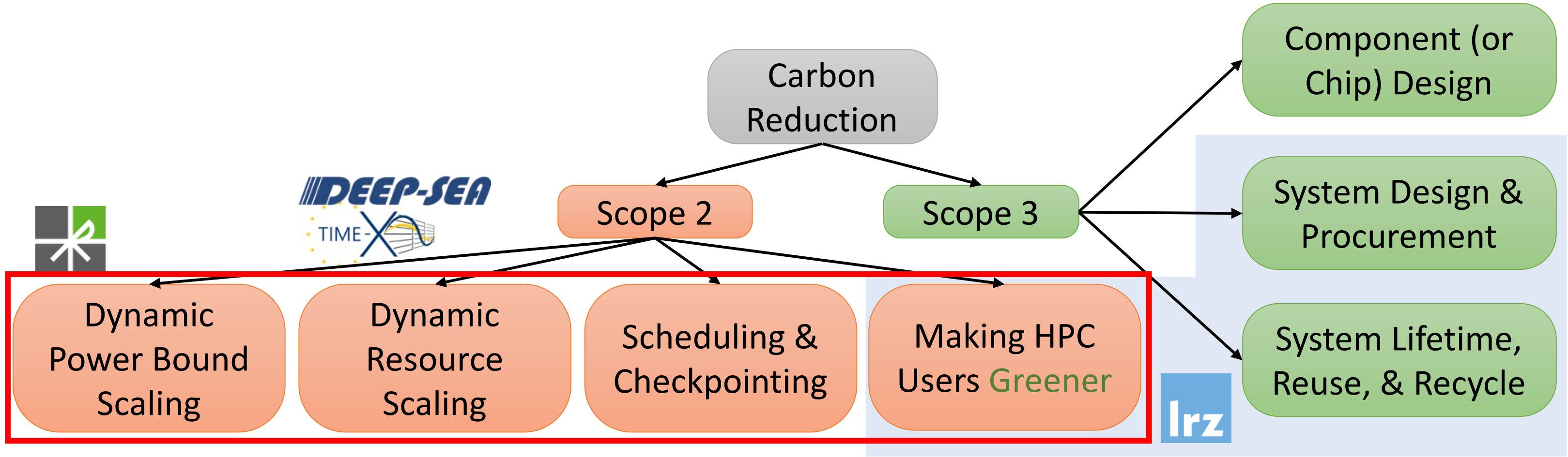
4-6yrs



Outline

- Background
- Reducing Embodied Carbon (Scope 3)
- Reducing Operational Carbon (Scope 2)
- Conclusion & Acknowledgement

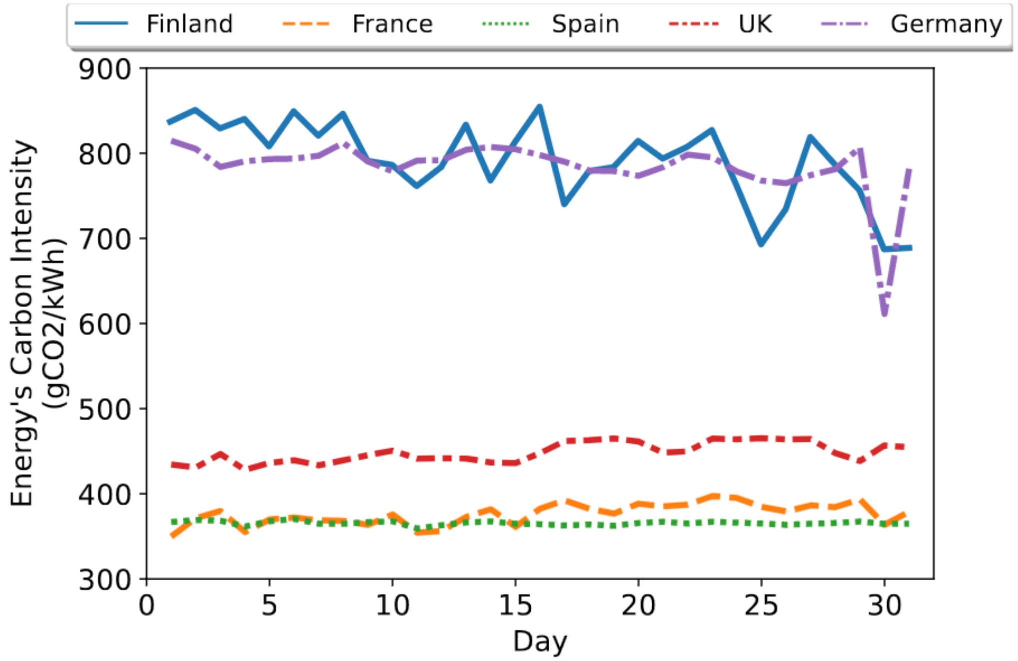




Reducing Operational Carbon (Scope 2)

- Operational Carbon Footprint = $\int CarbonIntensity * PowerConsumption\, dt$
- Carbon intensity depends on the location and can ***change over time***
- **Promising Approaches:** Scaling or shifting the following properties in accordance the carbon intensity
  1. System Power Bound
  2. System Scale (# of nodes)
  3. Peak Load

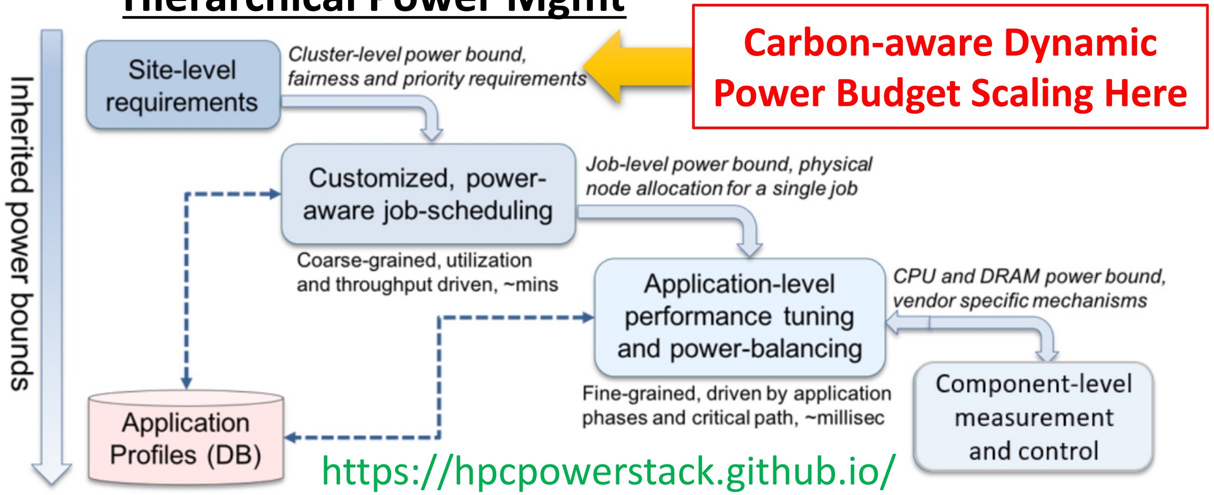
  * Existing energy/power reduction techniques generally ignore the carbon intensity



Carbon-aware Dynamic Power Budget Scaling

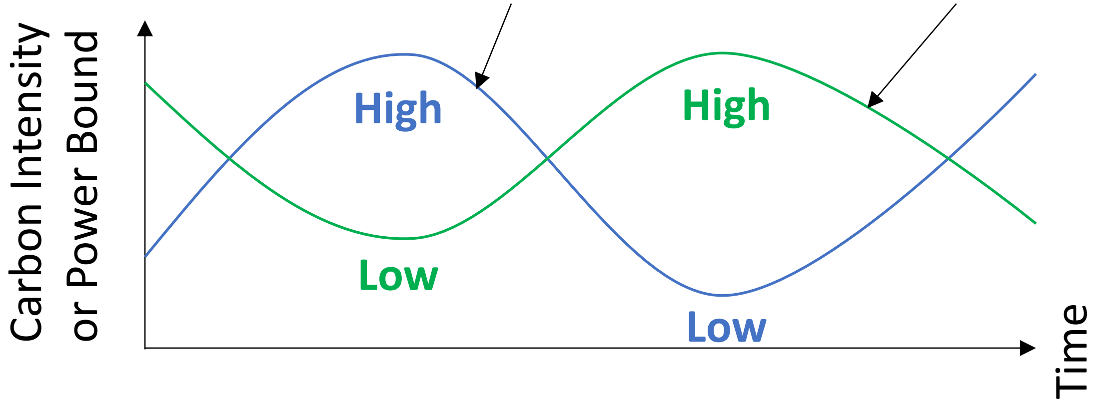
- A new use case of HPC PowerStack
- Scaling up/down the total system power bound based on the carbon intensity
  - Goal: to limit the operational carbon footprint within a given target

Hierarchical Power Mgmt

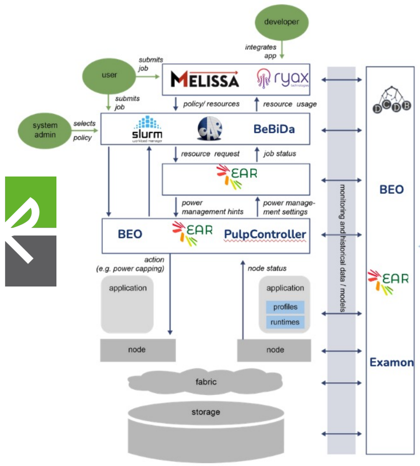


https://hpcpowerstack.github.io/

The Concept



Promising Solution



REGALE PowerStack

+

e.g., watttime

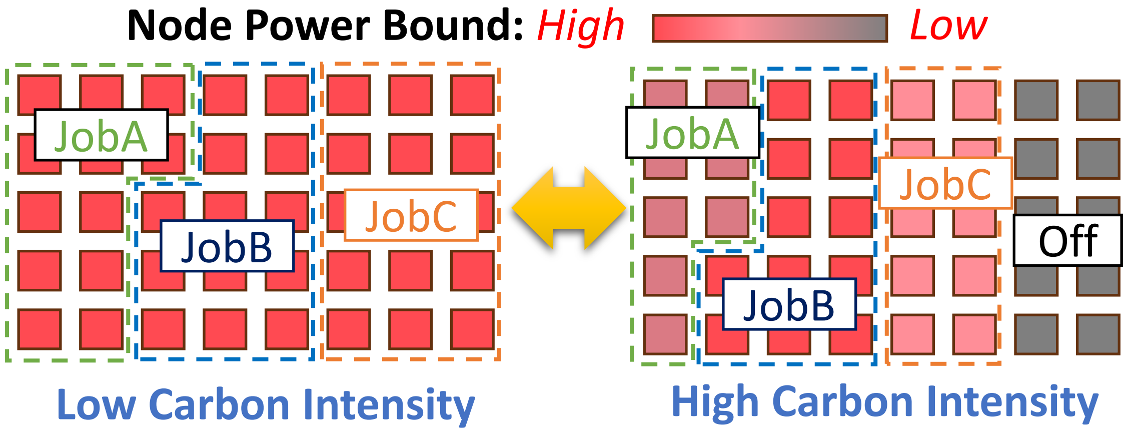
https://www.watttime.org/

Carbon Intensity Estimation Tools

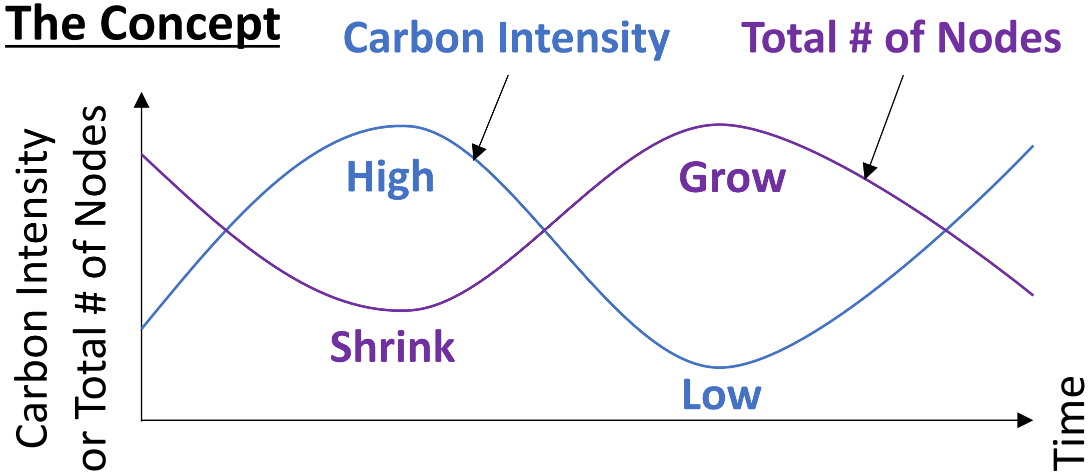
Carbon-aware Dynamic Resource Scaling

- Dynamically adapt the system scale in accordance with the carbon intensity
- Change the resource assignments to running jobs – known as **malleability**
  - Should cooperate with **power budgeting**

Hierarchical/Dynamic Res. & Pow Mgmt



The Concept



Ongoing Efforts toward Malleability Support

Closs Layer Support:

MPI + PMIx + slurm workload manager flux

Our Projects:

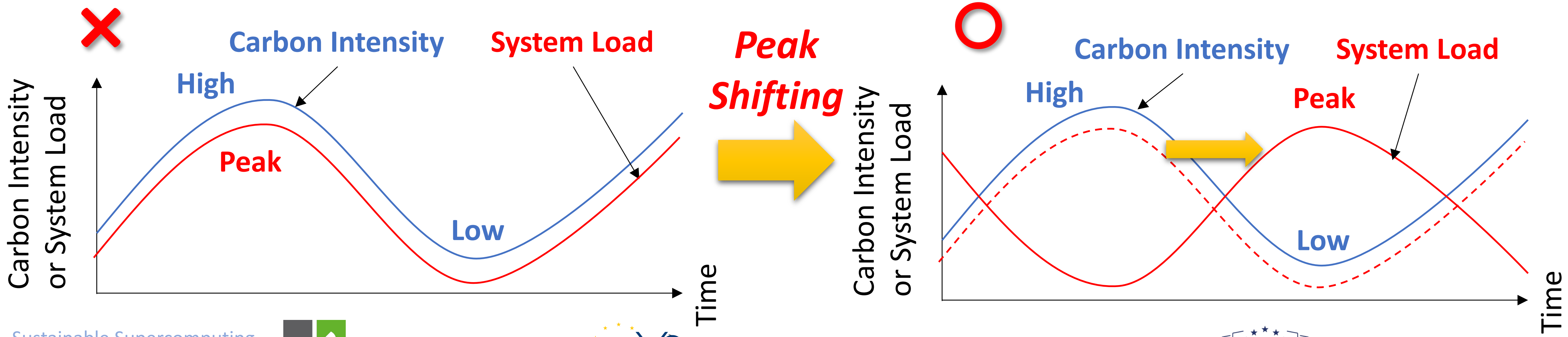
DEEP-SEA TIME-X

Carbon-aware Scheduling and Checkpointing

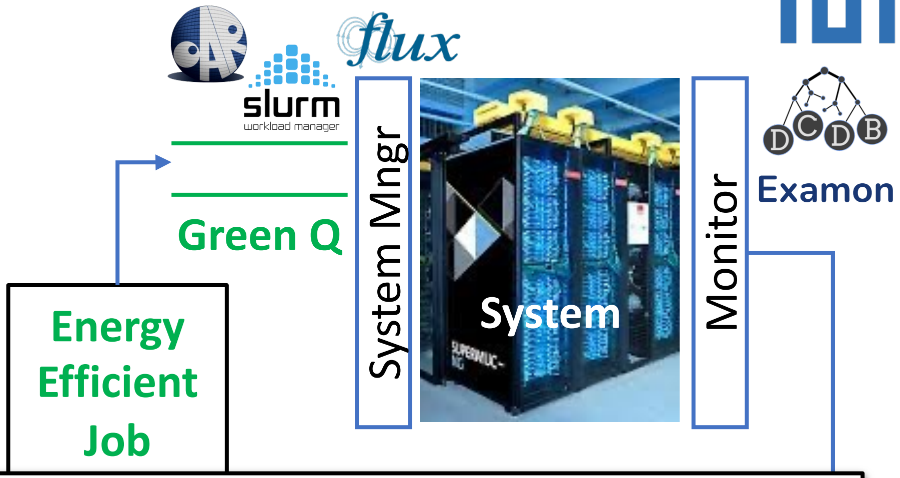
Concept: Adapt the system loads so as to have less jobs/loads when the carbon intensity is higher

Promising Solution: ***Peak Load Shifting***

- **Proactive:** ***predict*** the duration of ***green (low carbon intensity) period*** and optimize the scheduling decisions (e.g., green period-aware backfilling)
- **Reactive:** suspending, checkpointing, and restarting (large) jobs



Making HPC Users Greener

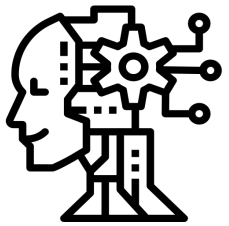


- Users can also contribute to ***peak shifting*** & ***energy footprint reduction***

Current HPC: Time is money and money is time

Future HPC: Carbon is money and money is carbon

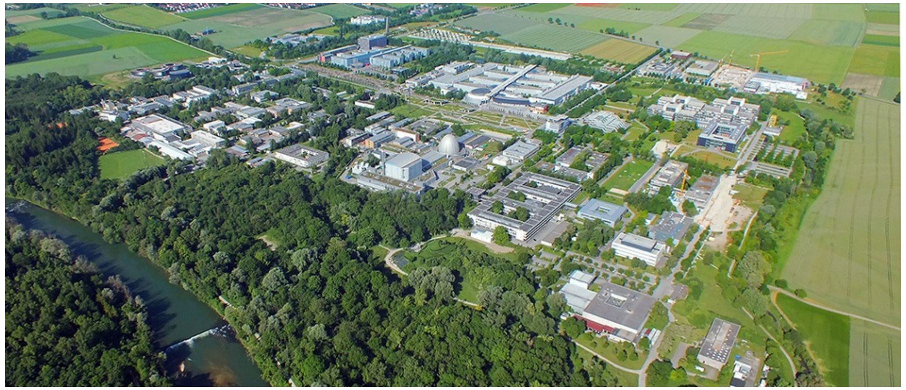
Your job accounted for the same amount of CO2 emissions as xxxxx, and as a result it consumed yy% of quota assigned to your project! You could save zz% of the quota consumption by simply following the steps below…

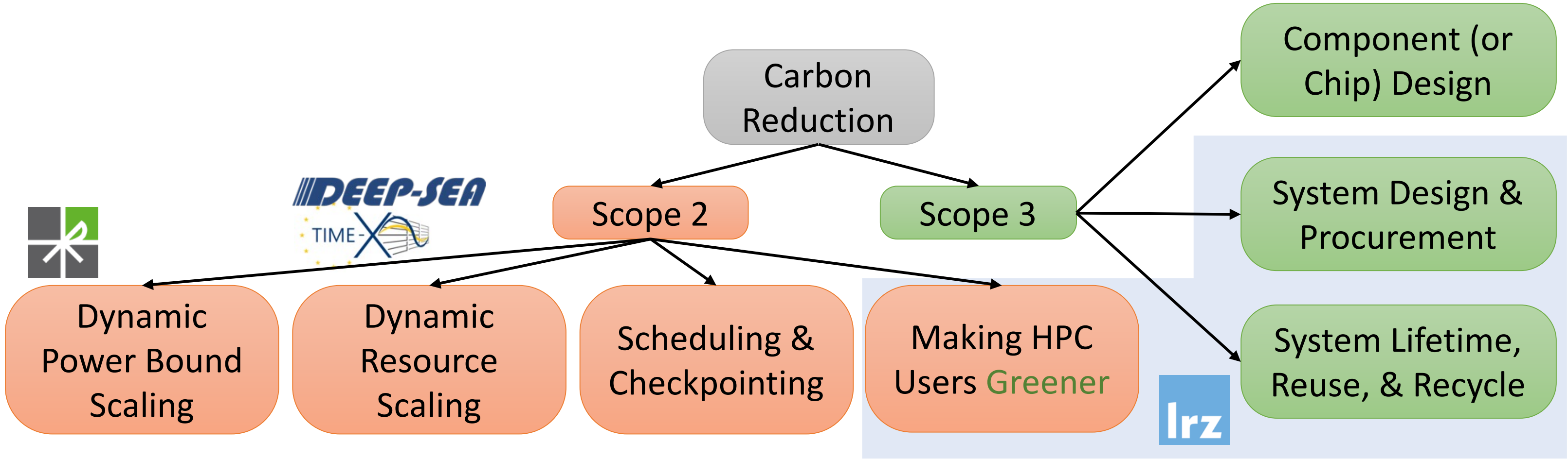


incentives to users to make them green

Outline

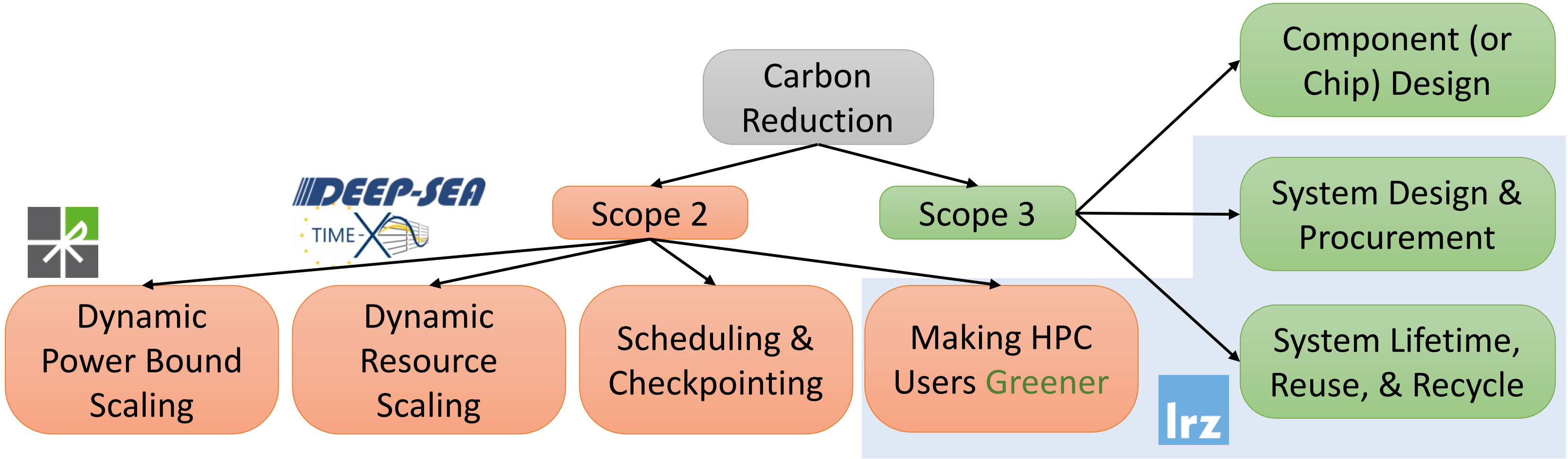
- Background
- Reducing Embodied Carbon (Scope 3)
- Reducing Operational Carbon (Scope 2)
- **Conclusion & Acknowledgement**





Concluding Remarks

- Carbon-aware supercomputing ≠ power-/energy-aware supercomputing!
- **Embodied Carbon Reduction (Scope 3):** Needs an end-to-end carbon-aware optimization from chip to system, and even procurement
- **Operational Carbon Reduction (Scope 2):** Needs (1) adaptive power, resource, & job management, and also (2) **green users**



Acknowledgement

This work has received funding under the European Commission's EuroHPC and Horizon 2020 programmes under grant agreements no. 956560 (REGALE), no. 955606 (DEEP-SEA), and no. 955701 (TIME-X).

REGALE

TIME-X

Visit our booth (#1311)!

Feel free to contact us!

DEEP-SEA

CAPS TUM