Advancing a more energy-efficient and sustainable SDCC at BNL

Naveed Anwer, Costin Caramarcu, Matt Cowan, Kevin Casella, Ricky Garcia, Imran, Latif, Al Maruf, John McCarthy,Tom Smith, Tony Wong

WLCG Environmental Sustainability Workshop

December 11, 2024

@BrookhavenLab

A brief history of scientific computing at BNL

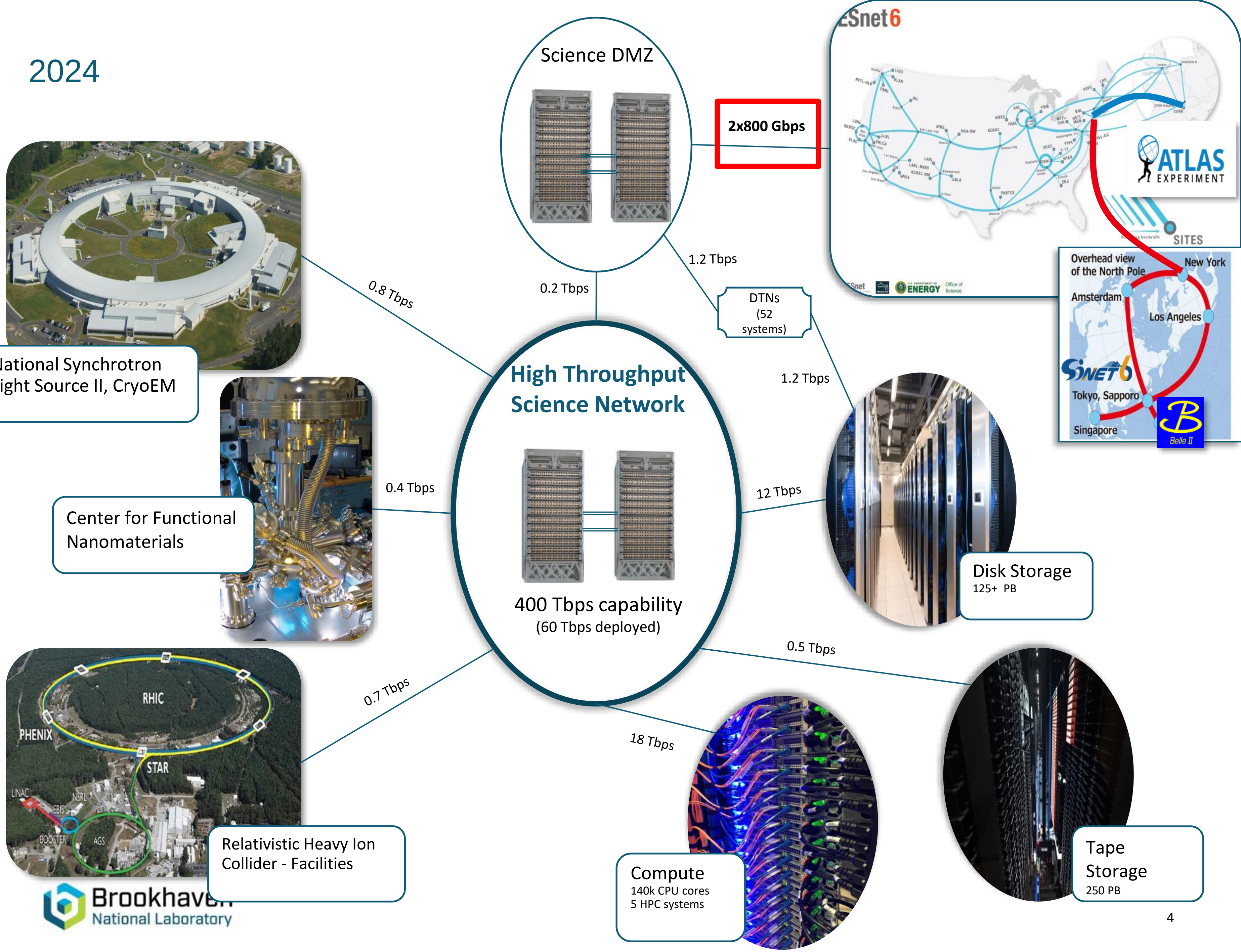
- The Physics Dept. created the RHIC Computing Facility (RCF) (circa 1997) to support its nascent experimental program.
- In the subsequent years, it was tasked with supporting the USATLAS Tier-1 computing (~2004) needs, other physics programs and experiments (neutrino, astrophysics) in ~2006, other BNL departments and programs (CFN, NSLS-II, Nuclear Science, CSI, etc) in ~2016 and Belle-II in 2018.
- The RACF was renamed the Scientific Data & Computing Center (SDCC) in ~2016.
- The SDCC outgrew its old data ~14,000 ft^2 center and migrated into a new and larger (~40,000 ft^2) data center. The SDCC is the sole source of scientific computing services at BNL.

Brookhaven National Laboratory

Scientific Data and Computing Center Overview

- Tier-0 computing center for the current RHIC experiments
  - sPHENIX and STAR
  - Support for earlier experiments (BRAHMS, PHENIX, PHOBOS)
- US Tier-1 Computing facility for the ATLAS experiment at the LHC
  - Also one of the three ATLAS shared analysis (Tier-3) facilities in the US
- Tier -1 and RAW Data Center for Belle II at KEK
- Providing computing and storage for proto-DUNE/DUNE together w/ FNAL serving data to all DUNE OSG sites
- Providing computing resources for smaler programs in NP and HEP
- Computing support for NSLS-II, CSI, CFN and other programs
- Serving ~2,000 users from >20 projects with ~40 staff
- Active R&D on future IT technologies
- Active contributor to workshops (HEPIX, HTCondor, etc) and conferences (CHEP, ACAT, etc) with scientific computing tracks
- Contributing resources to OSG at BNL
- Computing support for Electron-Ion Collider (EIC) activities

2024



Brookhaven National Laboratory

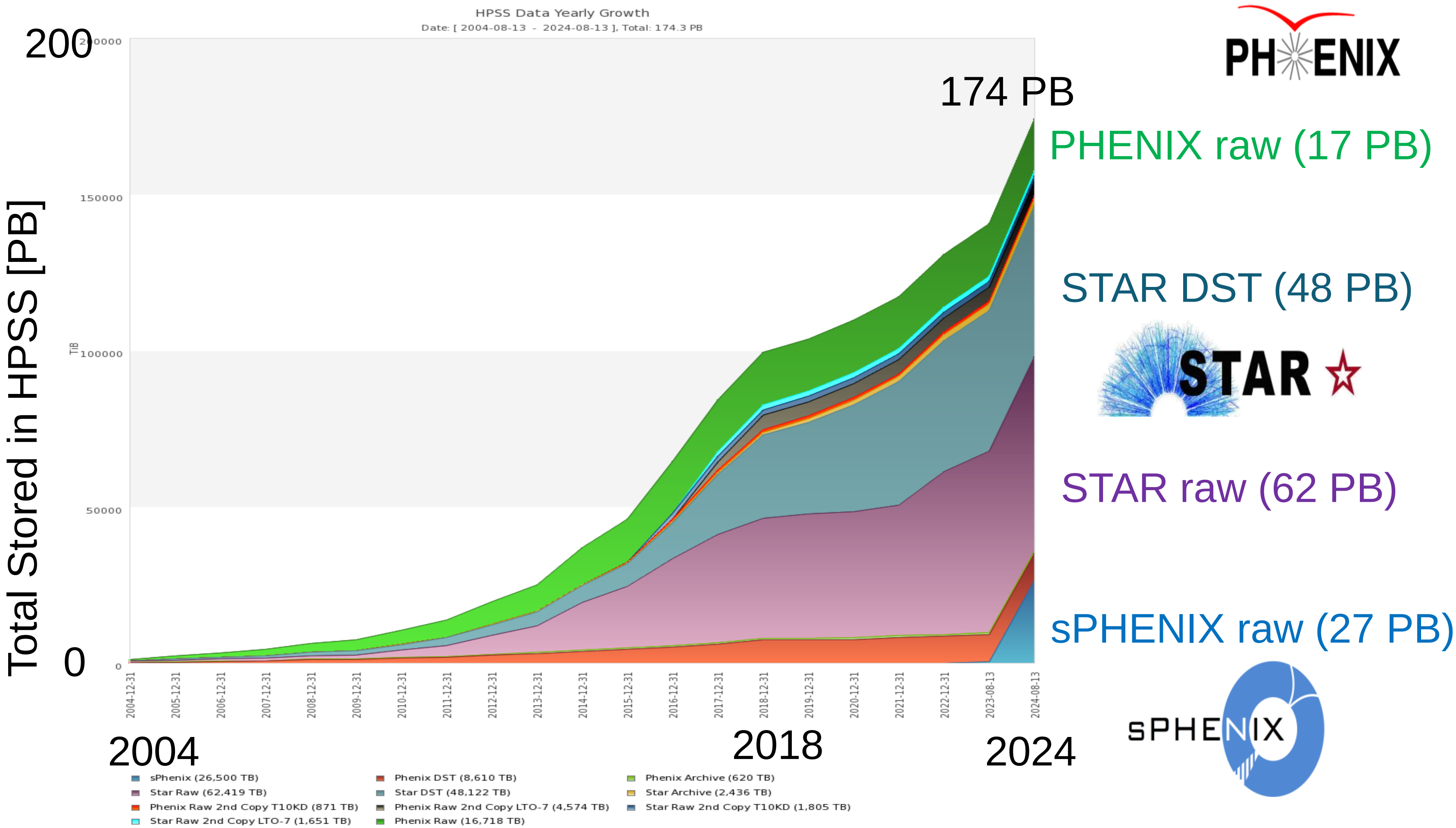
SDCC current activities

Current Priorities

- Support for sPHENIX computing model
  - Two, concurrent DAQ streams to tape (archival storage) and disk cache (for prompt processing)
  - Good experience for EIC streaming DAQ models
- Continued support for existing HEP experimental programs
  - ATLAS run3 at CERN (and run4 and beyond in the HL-LHC era)
  - Belle-II at KEK
- Continued support for RHIC data processing and analysis beyond 2025
  - Plan for constant effort over ~5 years after data taking ends
  - RHIC data, knowledge and workforce preservation
- Support for EIC
  - Kick-start activities with Program Development funds
  - Evolution of SDCC services
  - Prototyping of new computing architectures
    - non-x86 platforms
    - GPU's
  - Coordinated effort with JLAB on EIC computing being developed
    - Formation of EIC Computing and Software Joint Institute (ECSJI)

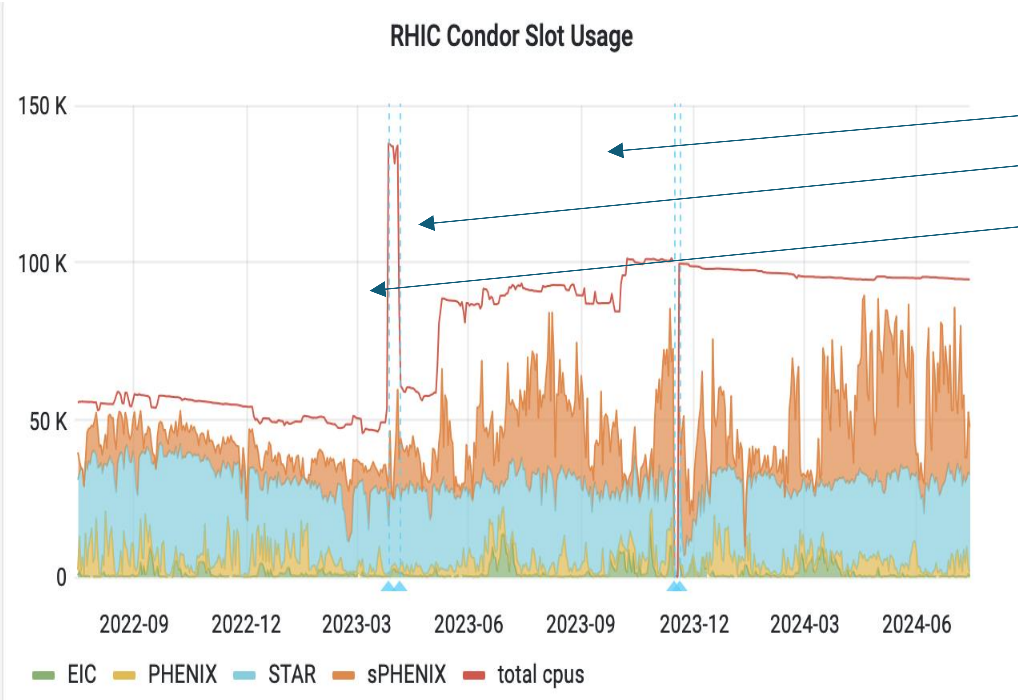
Brookhaven National Laboratory

Large datasets stored in HPSS



- SDCC just crossed 300 PB stored in HPSS tape system
  - Largest in Nuclear and Particle Physics (in U.S.)
  - 3rd largest in U.S. after NERSC (355 PB) and NOAA (350 PB)

Large Computing Deployments for RHIC



Major purchases since 2022:

•CPU

- 22 Racks, 616 servers - Purchase in progress
- 5 Racks, 158 servers - Q1 FY24
- 16 Racks, 495 servers - Q4 FY22
- 5 Racks, 150 servers - Q1 FY22

• SDCC projected to archive ~50 PB for sPHENIX in 2024 run

• Estimate to archive another ~500 PB for sPHENIX in 2025 run

Brookhaven National Laboratory

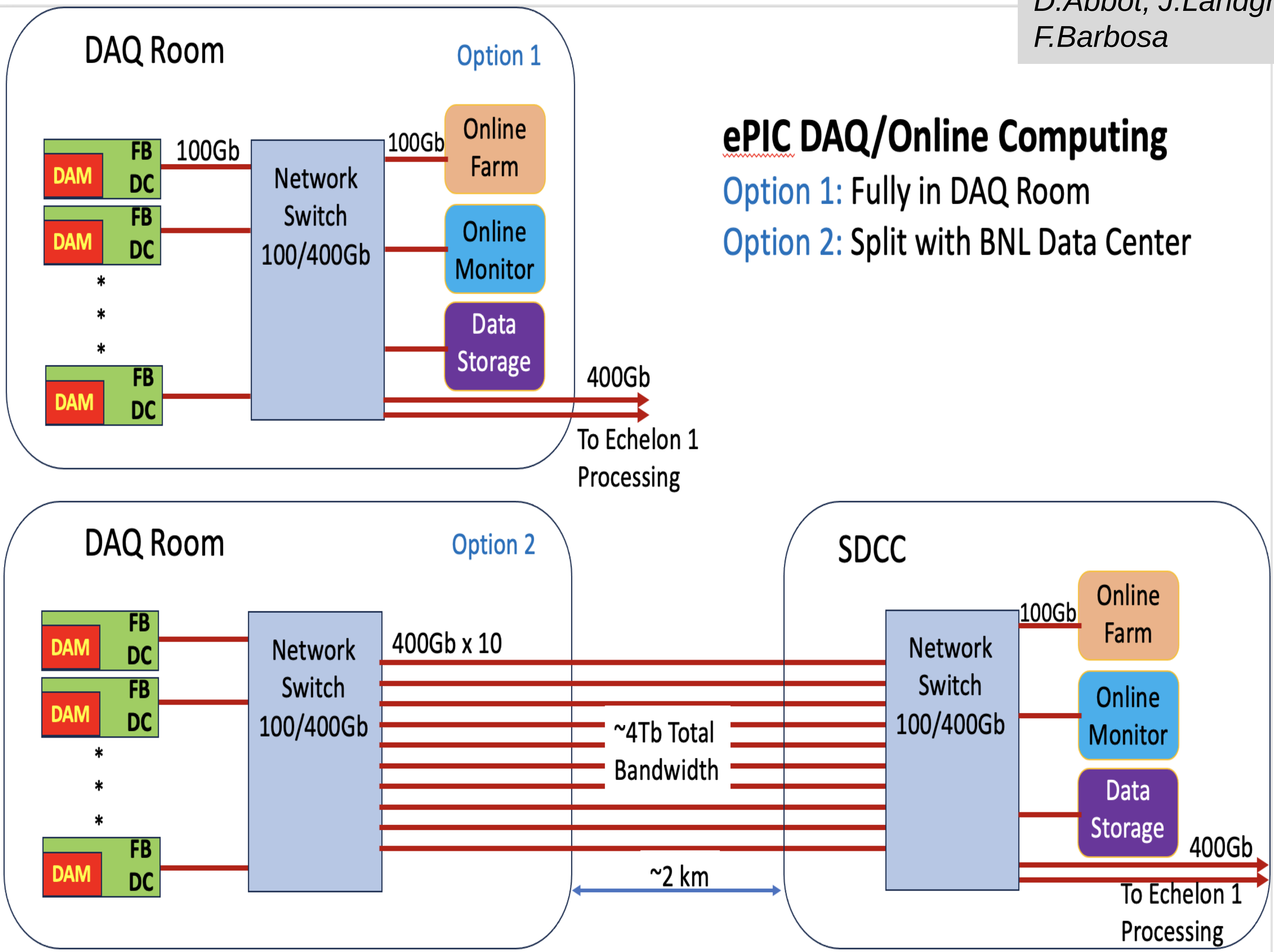
Impact of EIC Computing Activities

Plans for DAQ/Electronics Integration/Testing/Installation and needs for Off-Project Support D.Abbot, J.Landgraf, F.Barbosa

ePIC DAQ/Online Computing

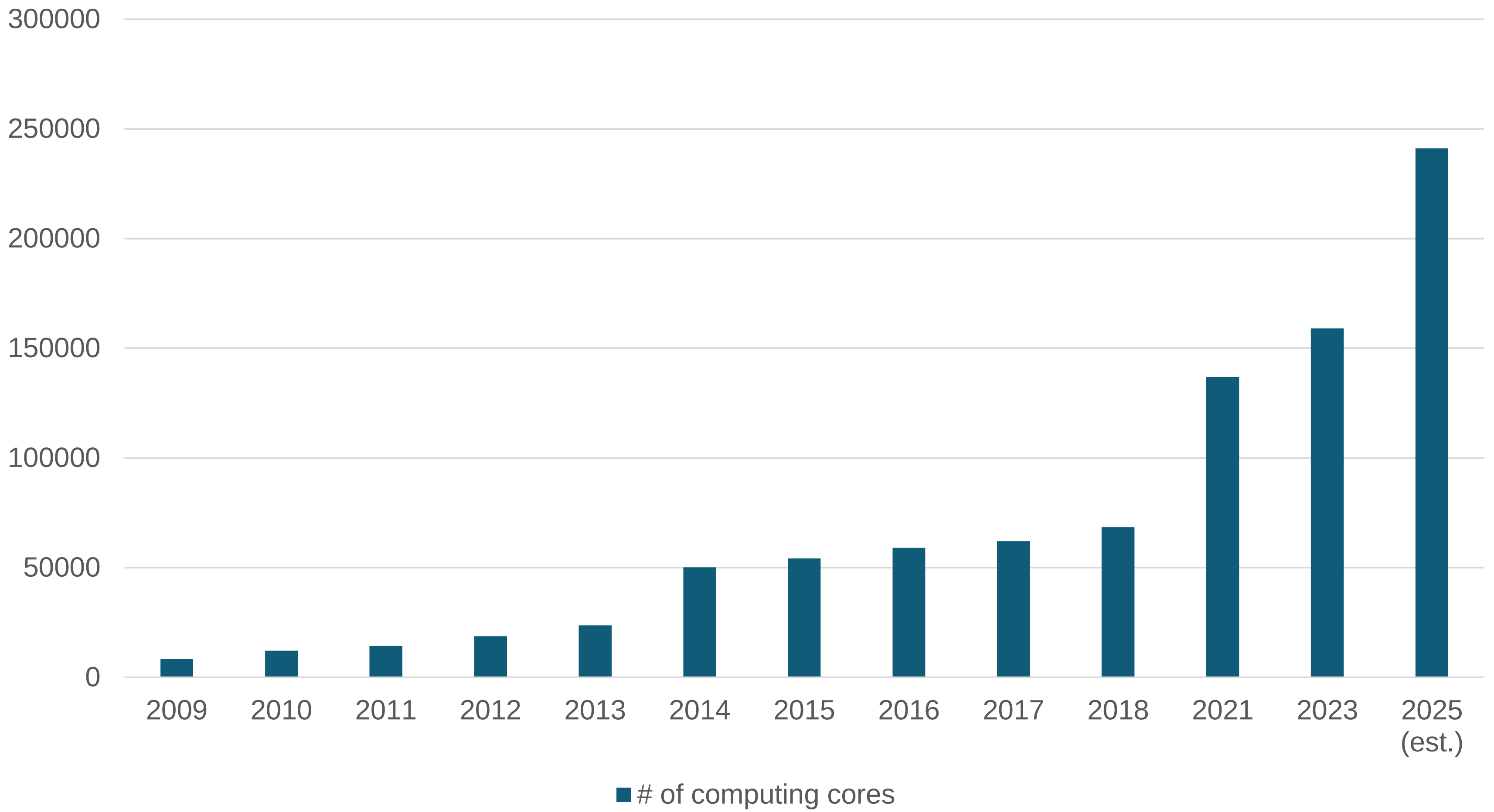
Option 1: Fully in DAQ Room

Option 2: Split with BNL Data Center

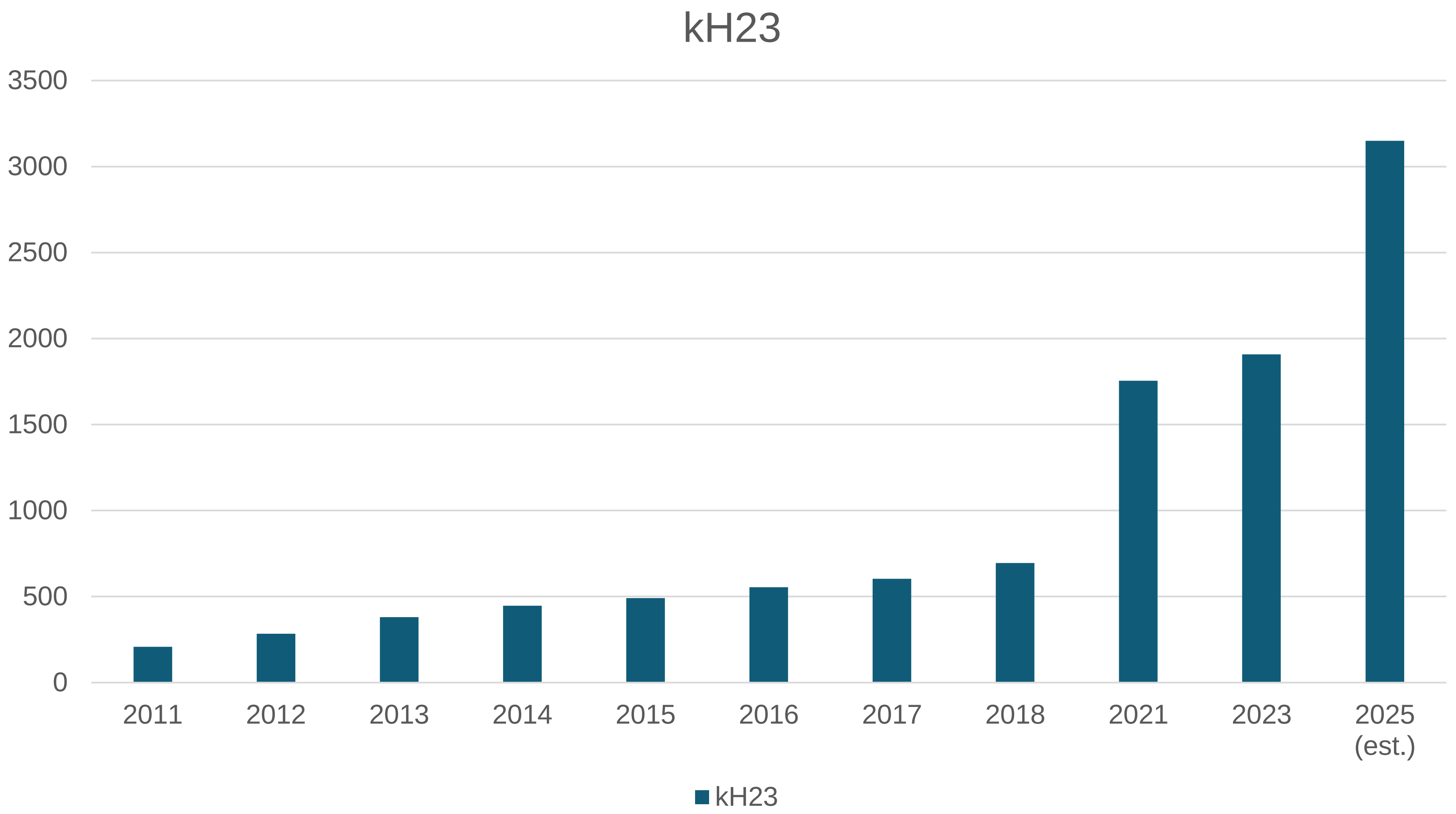


SDCC trends and evolution

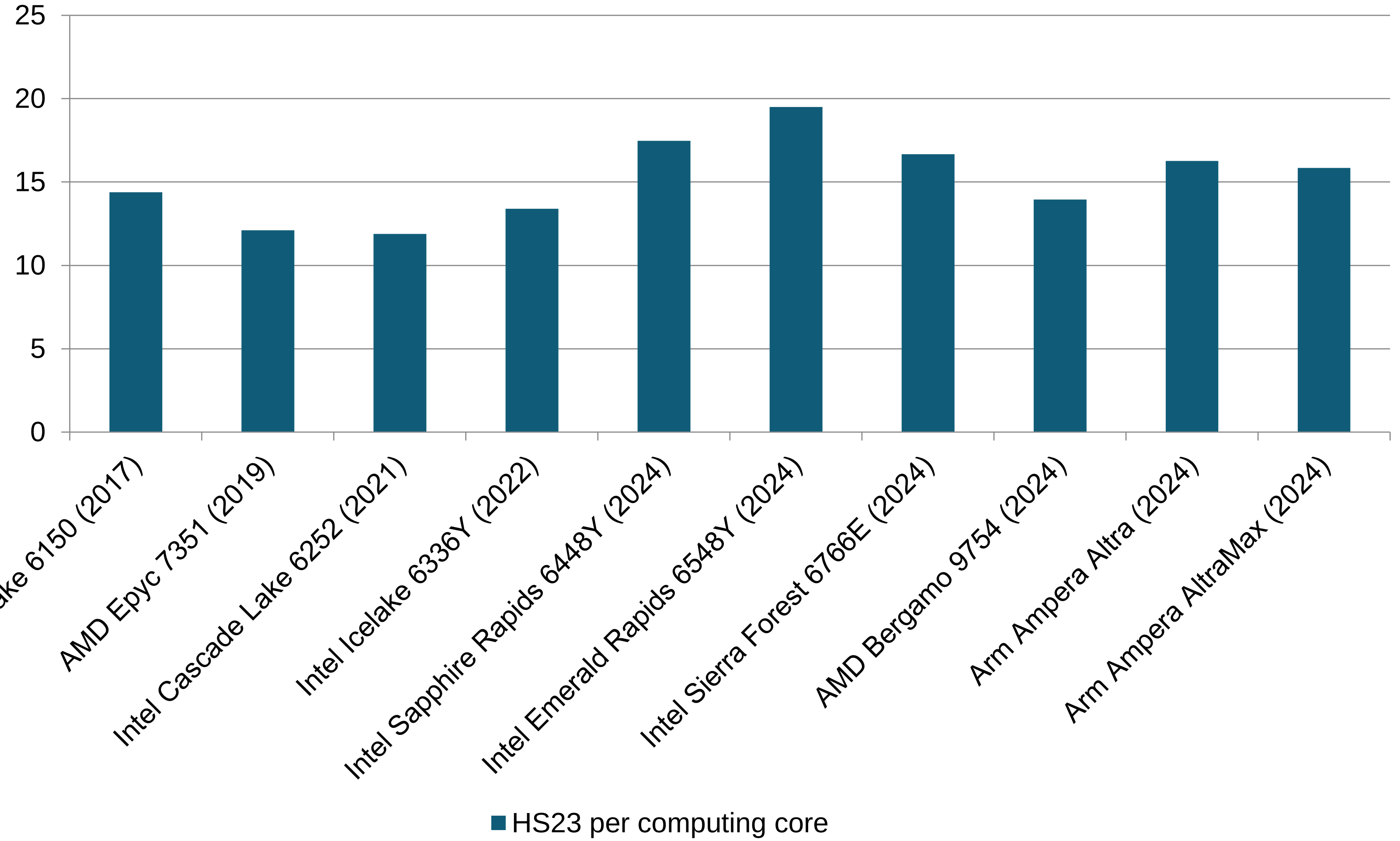
Facility Growth (computing cores)



Facility Growth

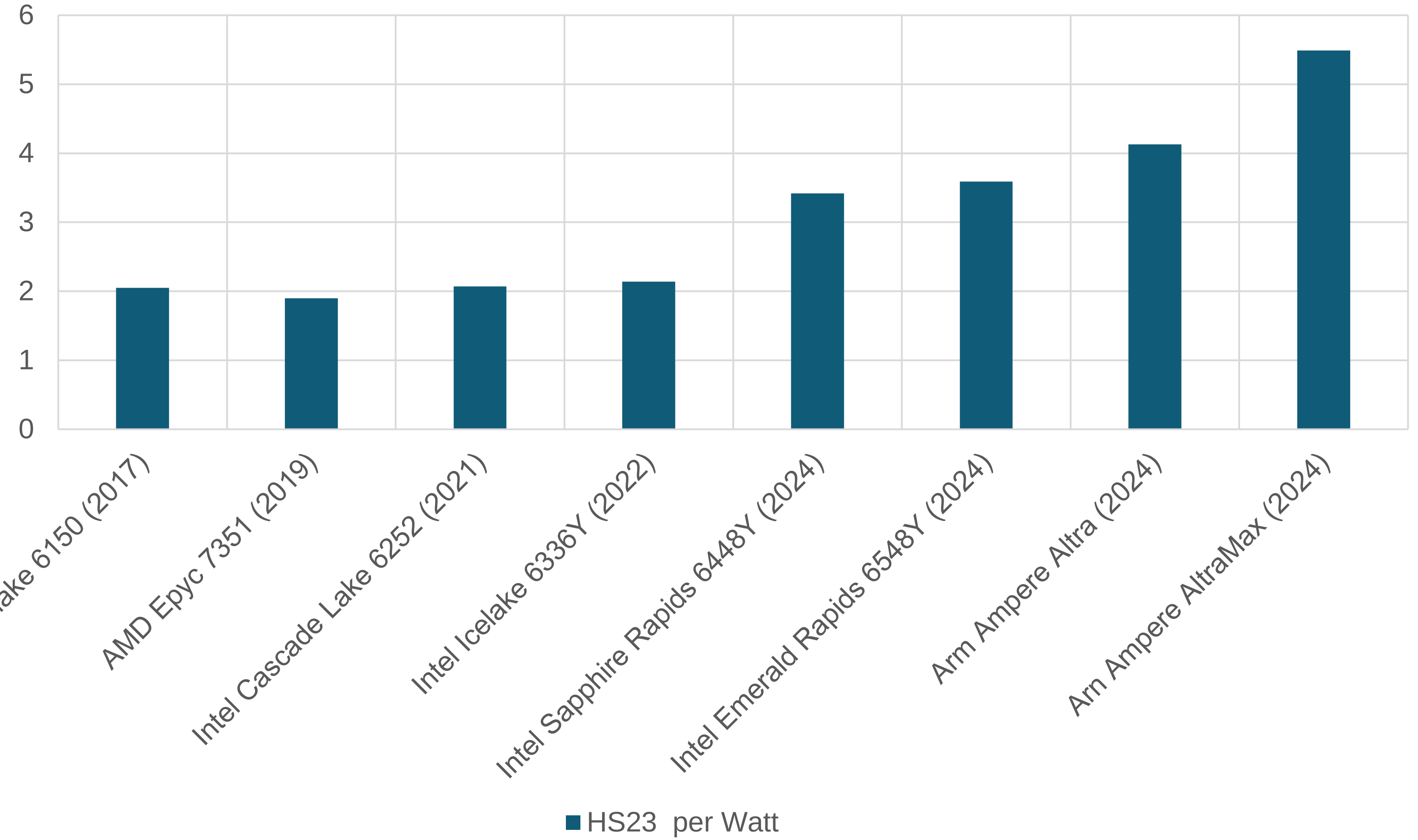


HS23 per computing core

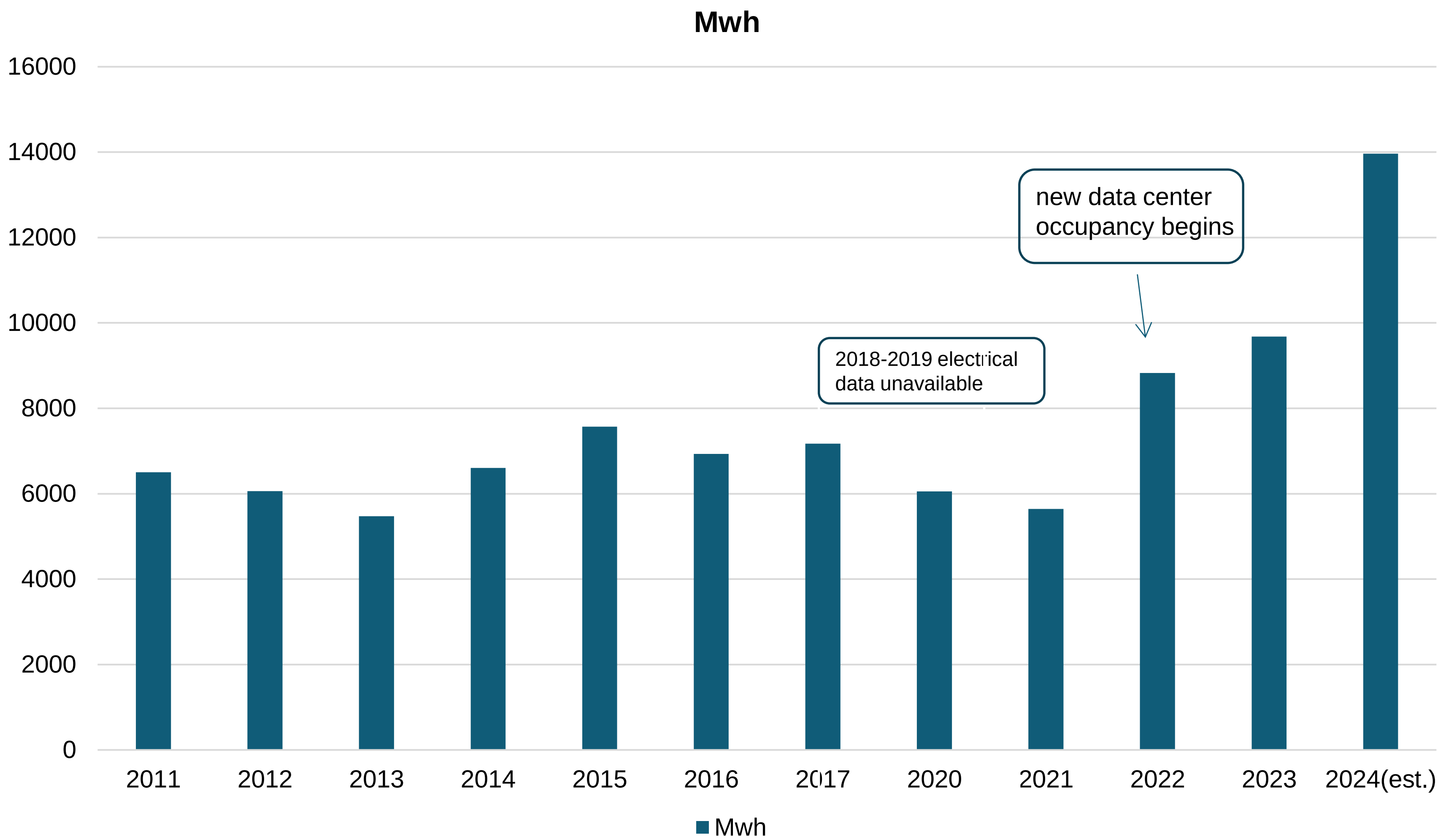


Brookhaven National Laboratory

Power efficiency of cpu's



Growth of electrical power usage



Some observations

- Continued growth of facility resources for existing and future programs led to the construction of a new data center (~3x larger than old one) in 2021
  - Resource growth rate may exceed data center capacity
- Continued usage of older resources beyond nominal retirement age (~5 years)
  - Needed to meet demand, but not ideal for carbon emissions
- Are improvements in HS23/Watt and HS23/core ratios sufficient to keep up with growth in demand?
- On-going evaluation of alternative cpu architectures
  - Arm
  - Nvidia Grace

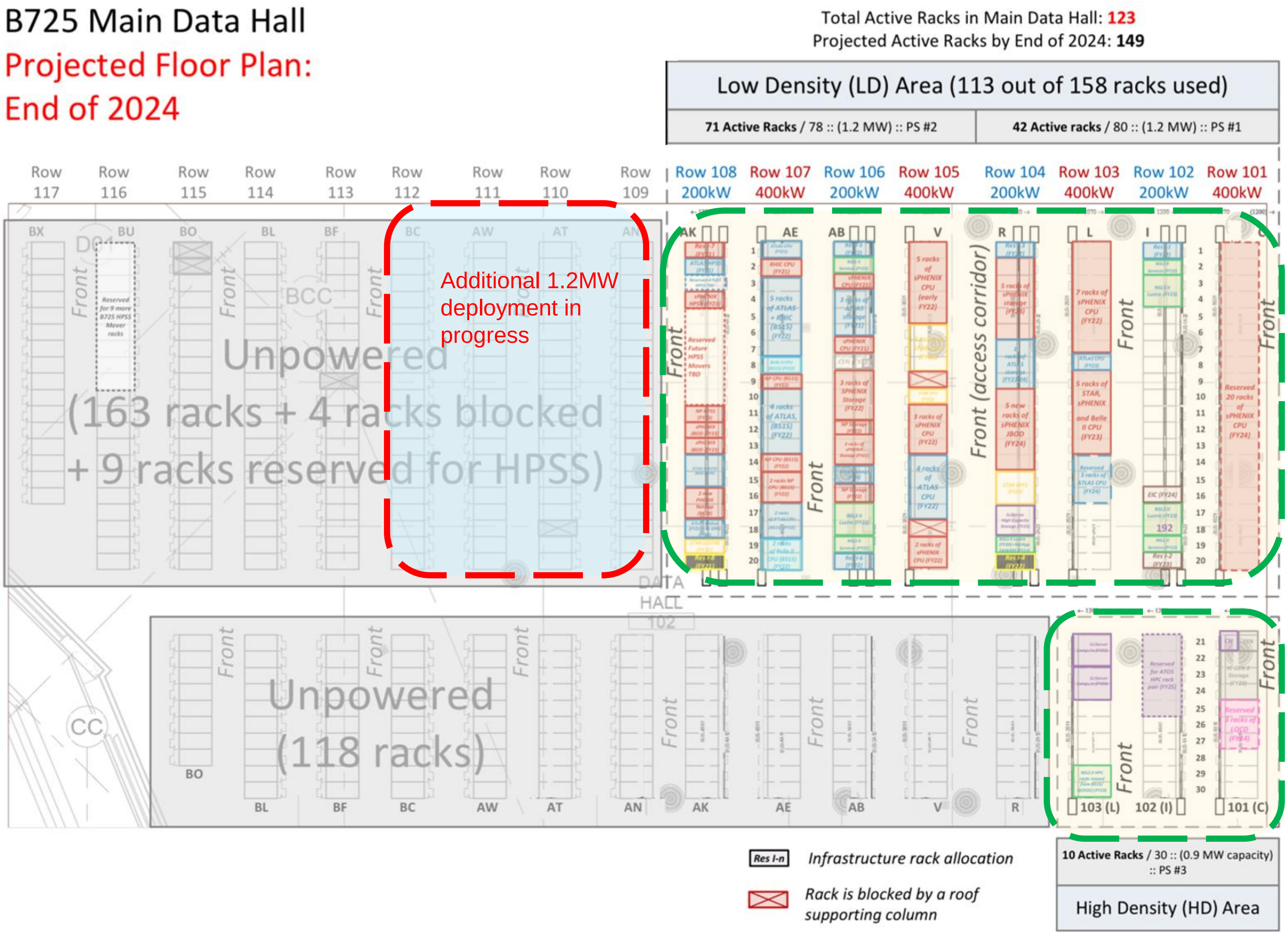
The New Data Center

State-of-the-Art Scientific Data and Computing Center (SDCC)

- 59,000 sq-ft^2 Data Center
- Opened for operations in September 2021
- Running community services for more than 20 projects and experiments:
  - RHIC (STAR and sPHENIX), ATLAS, Belle II, NSLS-II, DUNE, EIC, CFN, LQCD, IBM-Q Hub

Brookhaven National Laboratory

Floor Plan: Main Data Hall



Data Center Key Features

- Modular, Scalable and Robust design– 9.6MW of ultimate IT payload capacity
- Currently, 3.6MW (UPS/Generator) backed up power available
- Additional 1.2MW power block to be available by end CY-25
- Cooling with high efficiency chillers and Rear Door Heat Exchangers
- Liquid (Direct to Chip/Immersion) cooling ready for latest GPU based IT hardware deployment for AI, HPC and Digital Twin applications
- Energy efficient data center with 1.3 current PUE, aimed for 1.2 with full IT payload deployment
- Streamlined operations through Data Center Infrastructure Management (DCIM) including node level electric billing, asset management, environmental monitoring and capacity planning

Sustainability Activities

Case Study: Empirical Measurements of AI Training Carbon Footprint on a GPU-Accelerated Node

- **Background and Motivation**: Increased AI/HPC (High-Performance Computing) workload demand is driving up **scope 3** carbon emissions at main data hall. Quantifying the energy footprint of computational infrastructure requires models parameterized by the power demand of AI hardware during training.

- Empirical measurement of the instantaneous power draw of an 8-GPU NVIDIA H100 HGX node obtained during the training of open-source **image classifier (ResNet)**, **large-language models (Llama2-13b)** and **GPU Burn stress test**

- **Case Study Goal**: Benchmark carbon footprint of AI/HPC workloads on Supermicro HGX nodes with Nvidia's H-100 **8** GPUs
  - Optimize 'Stranded Power' in the SDCC data center by understanding the energy footprint of the AI training workloads for power optimization for additional ATLAS payload deployment
  - Reduce the embodied carbon - Scope 3

- **Current Findings**: (All Air-cooled)
  - Air-cooled systems struggle to reach maximum TDP – under utilized in power
  - Smaller batches require more time, decreasing efficiency
  - Larger batches require less time increasing significant efficiency at full GPU utilization

- **Proposed Solution**: Convert from **air-cooled** to **liquid-cooled** by:
  - 20-30% power savings – Scope 1
  - Reduced CapEx, OpEx and TCO

Brookhaven National Laboratory

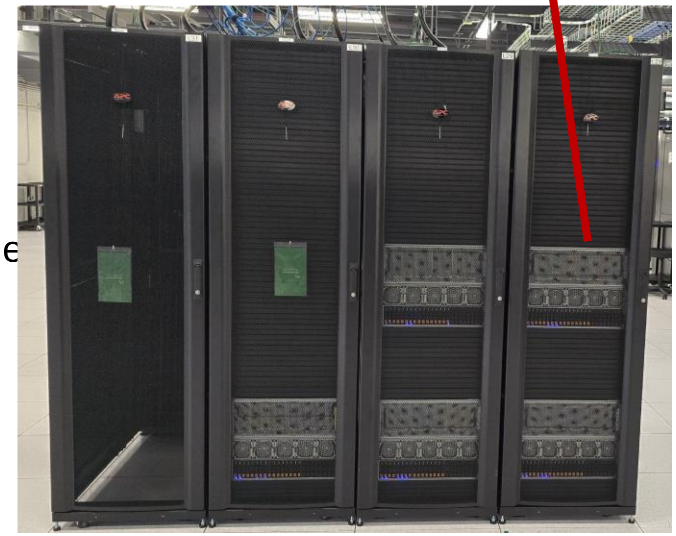
Hardware Information

- Supermicro HGX 8U air cooled node with (8) H100 GPU’s
- CPU : AMD EPYC 9354 32-Core Processor with 1.5 TB memory
- GPU: NVIDIA H100 with 80 GB memory
- Rated (node) TDP 10.2kW

Power and Cooling

- Standard 19" wide, 42U rack deployment
- Cooling provided by Rear-Door Heat Exchangers (RDHx)
- The RDHx is supplied with 15.5 °C facility chilled water
- 23.8 °C ambient cooling temperature maintained
- Power to the HGX node is provided via rack mounted power
- distribution units (PDU’s) using six standard C13 outlets
- Power monitoring available via DCIM system by NLYTE
- Provisioned Power is 14kW per High-Performance (HPC) node

HGX Node



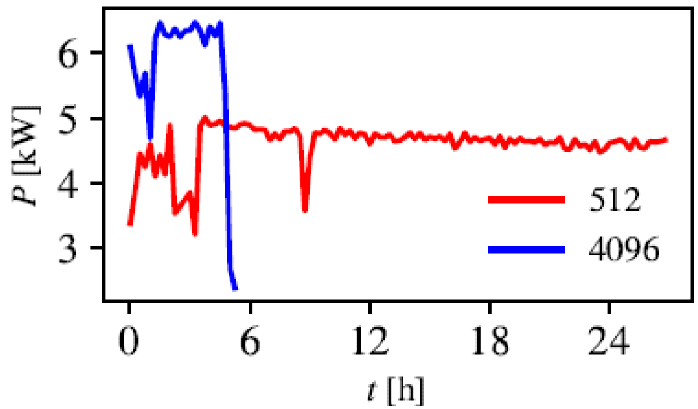


FIGURE 1. Instantaneous pre-training power demand measured in kW. Data corresponds to training ResNet with a batch size of 512 and 4096 images. Total energy usage for these two cases were computed by integrating the curves and are 123 and 30 kWh, respectively.

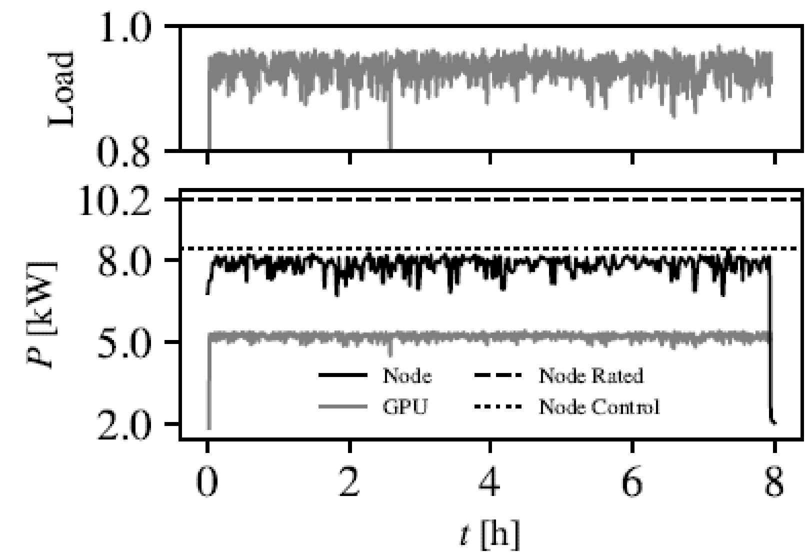


FIGURE 2. Instantaneous pre-training power demand measured in kW, and average total GPU load, for the Llama2-13b parameter model. The rated power of the system (10.2 kW) is shown as a horizontal dashed line. The median node power draw from the GPU+CPU burn control experiment is shown as a horizontal dotted line (at 8.43 kW).

TABLE 1. Power was sampled at 1-minute intervals over the duration of training. Batch size is measured in images and tokens for the ResNet and Llama models, respectively.

Training Workload	Model Task	Median Power (kW)	Batch Size	Duration (min)	Total Energy (kWh)	Average GPU Utilization (%)
Idle Node	N/A	1.86	N/A	N/A	N/A	N/A
Resnet-512	Image	4.68	512	1,605	123	36
Resnet-4096	Image	6.26	4,096	315	30	77
Llama2-13b	VQA	7.92	128	480	62	93
GPU Burn (no CPU)	Stress test	7.97	N/A	60	8	100
GPU Burn (+CPU)	Stress test	8.43	N/A	77	10.5	100

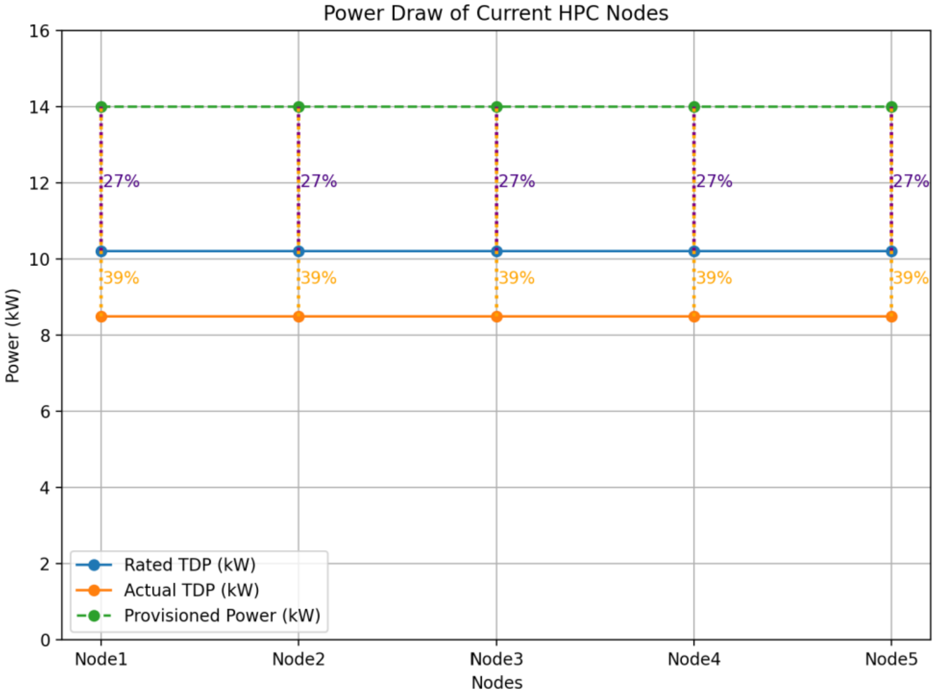


Figure 3: Power draw of currently deployed AI/HPC nodes

Current AI/HPC H100 Nodes

- 27% excess power is available from the provisioned power assuming the rated TDP
- Actual power does not exceed 8.5 kW, allowing a 39% excess power

Node Availability per Power Metric

- 63 nodes using provisioned power
- 87 nodes using rated power
- 105 using actual power
- 67% increase in additional node (42) deployment on the basis of actual power consumption
- 38% increase in additional node (24) deployment on the basis of rated (TDP) consumption

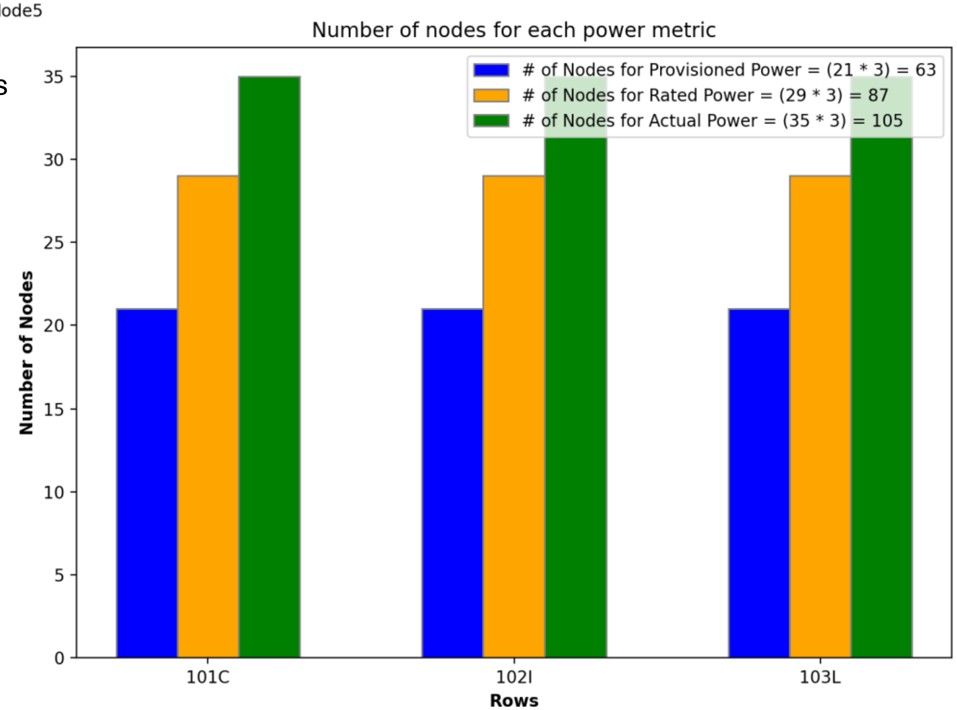


Figure 4: Number of nodes for each row by each power metric

Brookhaven National Laboratory

Potential Number of Nodes with Air-Cooled Machines

- 20%-25% more power efficient
- 33% increase in number of nodes from Air-Cooled to Liquid-Cooled
- Increased in performance with the same power draw

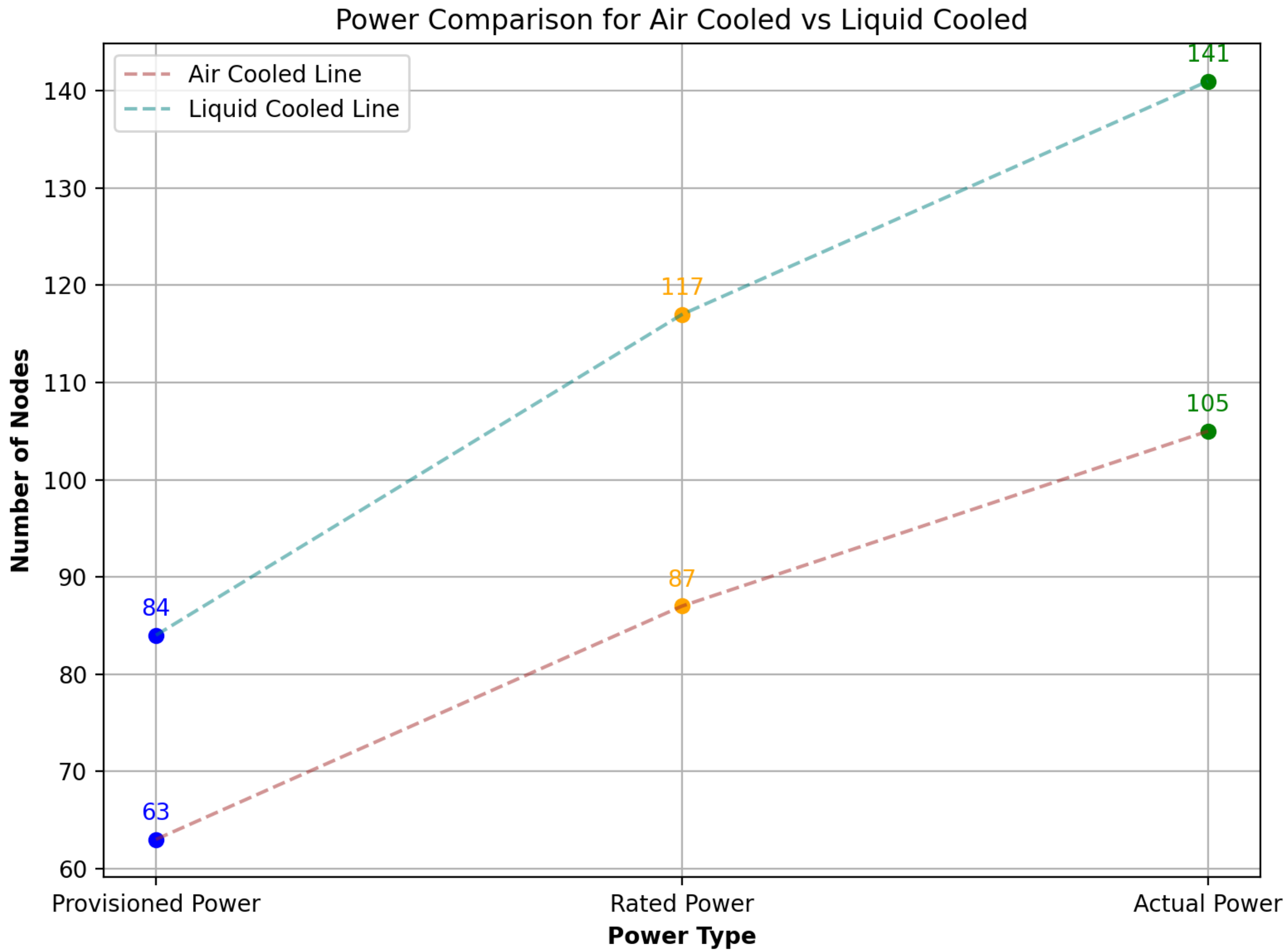


Figure 5: Node Availability in Air-Cooled V.S. Power-Cooled system

Brookh
National Laboratory

What is CodeCarbon? + Data Collection Process

CodeCarbon:

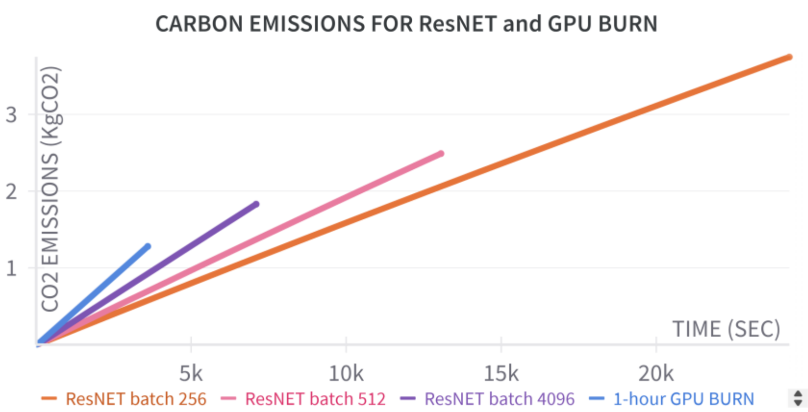
- Open-source python package used to track and estimate CO_2 emissions from computational workloads.
- Monitors hardware usage (CPU, GPU, RAM) and energy consumption
- The AI workloads were **distributed** over all 8 GPUs this time to take less time per batch run using **distributed data parallel** (DDP).

Region: New York (NY):

- CodeCarbon uses data from local electricity grids to estimate emissions.
- NY primarily uses an energy mix consisting of:
    - **Natural Gas**, **Nuclear**, **Hydropower**, and **Other Renewables**.
- Regional carbon intensity (approx.): ~0.464 $kgCO_2/kWh$
- Carbon emissions formula:
    - **Power consumed (kW) x time (hrs) x 0.464 $kgCO_2/kWh$**

Integration with Weights & Biases (W&B):

- A tool for tracking machine learning experiments, enabling visualization, collaboration, and optimization in real-time.
- Logs carbon emissions data alongside model performance metrics.
- Enables seamless visualization of environmental impact in W&B dashboards.
- Allows for data-driven approaches for debugging actively running models in real-time.



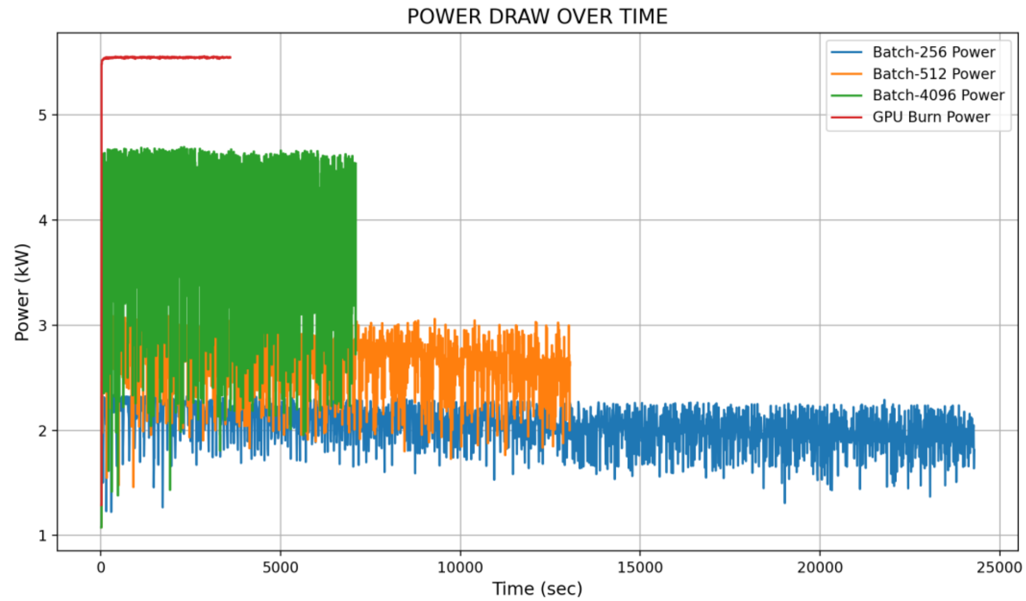


Figure 6: The figure demonstrates that as batch size increases, carbon emissions decrease. Larger batches consume more power but complete the same task in significantly less time, leading to lower overall emissions.

Figure 7. The instantaneous power draw for each run over time (in seconds) is shown in the figure. Larger batch sizes, while requiring more power, complete the tasks in less time. This reduction in runtime results in lower overall carbon emissions, despite the higher power consumption per run.

Table 2. Carbon Emissions were sampled at 1 second intervals over the time. Batch size is measured in

Training Workload (with DDP)	Model Task	Median Power (kW)	Batch Size	Duration (hrs)	Total Energy (kWh)	Average GPU Utilization (%)	Carbon Emissions (kgCO2)
ResNET-256	Image	2.14	256	6.7	14.338	52	3.75
ResNET-512	Image	2.78	512	3.6	10.008	68	2.49
ResNET-4096	Image	4.29	4096	1.97	8.4513	77	1.82
GPU Burn (no CPU)	Stress Test	5.55	N/A	1	5.55	100	1.29

Results and Conclusion

- The maximum observed power draw during a GPU Burn stress was approximately 8.4 kW, 18% lower than the manufacturer-rated 10.2 kW at 100% GPU utilization
- Holding model architecture constant, increasing batch size from 512 to 4096 images for ResNet used 1kW higher power on average but reduced total training energy consumption by a factor of 4
- Llama training GPU load was 93% on average, with a median power draw of 7.9 kW, indicating, also well below the rated maximum
- The minimal difference in power consumption between the GPU stress test and the real-world Llama training workload demonstrates that our empirical measurements offer a reliable characterization of the power requirements for computationally demanding workloads on this hardware
- The rack-level power distribution at the SDCC supplies 70 kW (14kwx5) of power to the racks which contain (5) Nvidia HGX nodes
- Use existing ‘stranded’ power capacity in low-utilized areas within the data center instead of adding new infrastructure for expansion of ATLAS computing deployment at BNL.

Results and Conclusion

- Hardware Consolidation may reduce CapEx, OpEx, and overall TCO of the data center. Also reduces Embodied Carbon (Scope 3) by reusing server chassis during server refresh by reusing physical infrastructure instead of purchasing entirely new equipment.
- Current data only shows air-cooled nodes. Ongoing research is focused on understanding the energy gain and further **carbon footprint reduction** by running similar benchmarks on a liquid-cooled HPC node.
- Through the application of a DTC cooling system, a potential power savings of **20-25%** can be used for scalability by introducing more space for high density HPC racks, as it reduces the power draw for fans.
- Smaller batch sizes require more iterations per epoch, leading to extended runtime and higher carbon emissions, whereas larger batch sizes, despite higher power consumption per step, complete tasks more efficiently and result in reduced overall emissions.

Future Research

- **Capacity Planning and Energy Estimation:** These findings can guide SDCC data center capacity planning, precise energy estimation and allow the use of existing power to expand ATLAS computing infrastructure at BNL.

- **Liquid cooling Direct-to-Chip (DTC)/Immersion Cooling Impact:** Efficient cooling technologies, like DTC liquid cooling, could maintain high computational utilization while significantly lowering node-level energy consumption, potentially by 30-40%.

- **Heat Reuse Potential (Circular Economy):** Energy absorbed from the chilled water system that uses liquid cooling can be effectively reused for heating applications such, space heating, water desalination, fish farming, etc.

- **Component-Level Power Visibility and Optimization:** Detailed sub-metering of individual components within nodes could reveal the sensitivity of power draw across workloads, allowing for optimized hardware configurations.

- **Future Research on Cooling and Scheduling**: Further studies will investigate how advanced cooling technologies and carbon-aware scheduling can impact the energy consumption of AI workloads.

References

A. de Vries, “The growing energy footprint of artificial intelligence,” Joule, Oct. 2023. [Online]. https://www.sciencedirect.com/science/article/pii/S2542435123003653

A. S. Luccioni, S. Viguier, and A.-L. Ligozat, “Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model,” Journal of Machine Learning Research, vol. 24, no. 253, pp. 1–15, 2023. [Online]. http://jmlr.org/papers/v24/23-0069.html

Mohammad Azarifar a, et al. “Liquid Cooling of Data Centers: A Necessity Facing Challenges.” *Applied Thermal Engineering*, Pergamon, 5 Apr. 2024, www.sciencedirect.com/science/article/abs/pii/S1359431124007804

Wilicc. “WILICC/GPU-Burn: Multi-GPU Cuda Stress Test.” *GitHub*, https://github.com/wilicc/gpu-burn

Brookhaven National Laboratory