

# Finalizing Transition to the New Data Center at BNL

Imran Latif<sup>1</sup>, Shigeki Misawa<sup>1</sup>, Alexandr Zaytsev<sup>1,\*</sup>

<sup>1</sup> Brookhaven National Laboratory, Upton, NY 11973-5000, U.S.A.

\*alezayt@bnl.gov

May 9, 2023



@BrookhavenLab

# SDCC Data Centers at BNL

Scientific Data and Computing Center (SDCC) Facility is operating two data centers at BNL:

- The old (B515) data center – a 1.5 MW scale air-cooled facility currently hosting 180 racks of IT equipment and 10 tape silos
- The new (B725) data center – a 3.6 MW scale water-cooled facility capable of hosting 210 racks of IT equipment and 6 tape silos in its present state (500 racks and 12 tape silos in the full build-out configuration)

Both data centers combined are currently hosting:

- HTC and HPC computing systems (on the scale of 2.4k nodes)
- DISK storage (on the scale of 130 PB)
- Robotic tape storage (on the scale of 220 PB)

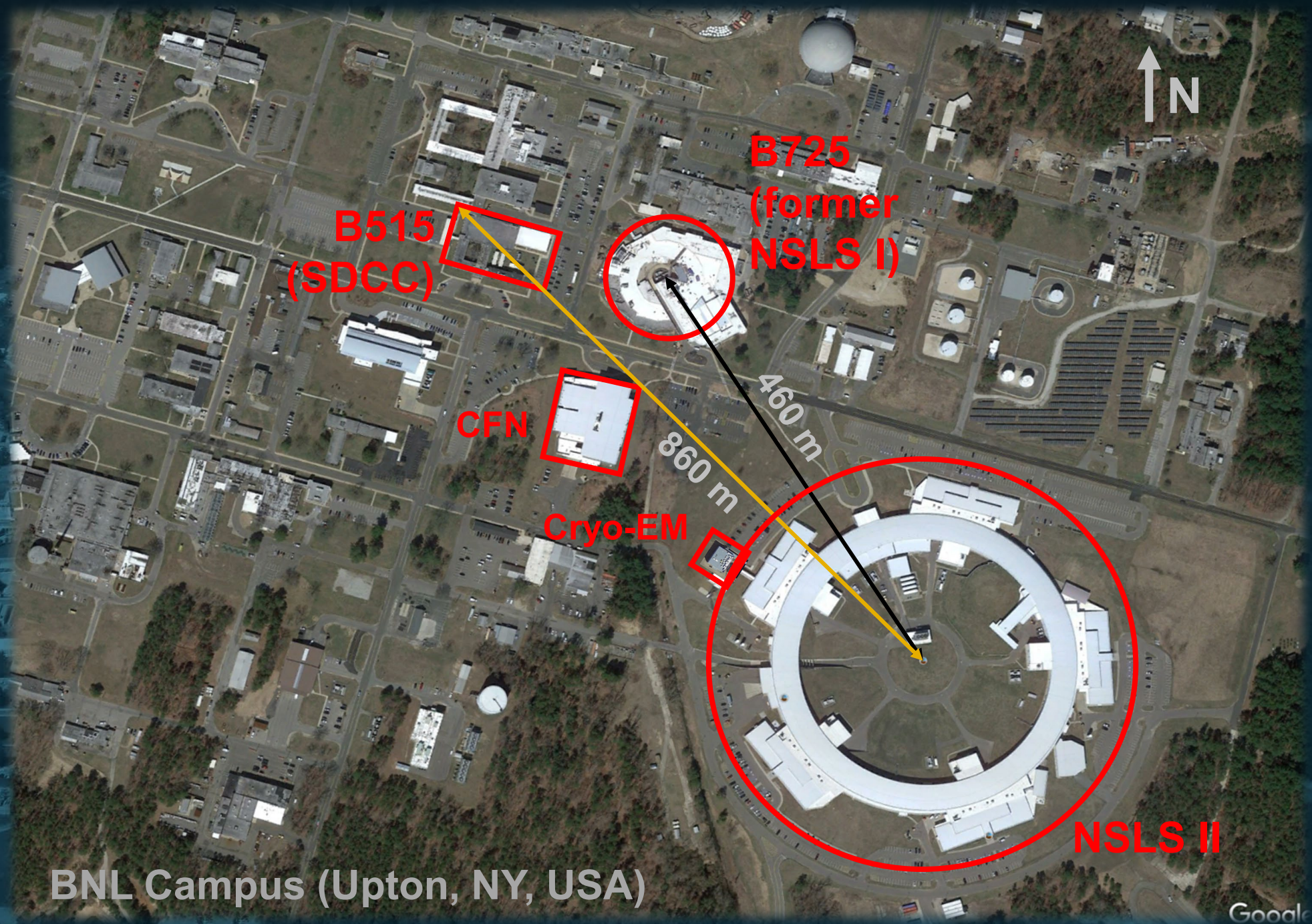
Serving multiple international collaborations and BNL research communities:

- STAR, PHENIX and sPHENIX at RHIC (BNL) with two on-site Counting Houses
- ATLAS Experiment at the LHC (ATLAS Tier-1 Site)
- Belle II Experiment at KEK (Belle II Tier-1 Site)
- National Synchrotron Light Source (NSLS) II at BNL & Cryo-EM Lab in the Laboratory for BioMolecular Structure (LBMS) at BNL
- Center for Functional Nanomaterials (CFN) at BNL
- BNL Computational Science Initiative (CSI) research groups and test labs

Main drivers for future growth:

- sPHENIX Experiment at RHIC – currently starting its first run & continuing taking data in FY24-25
- STAR Experiment at RHIC continuing taking data in FY23-25
- Future detectors at Electron-Ion Collider (EIC) at BNL – data taking anticipated in early 2030s
- ATLAS Experiment at the LHC (LHC) – resuming taking data under Run 3 (FY22-25)
- ATLAS Experiment at High Luminosity LHC (HL-LHC) – starting data taking in FY29
- Scaling of the NSLS-II Facility to 60 active beamlines





B515  
(SDCC)

B725  
(former  
NSLS I)

CFN

Cryo-EM

NSLS II

BNL Campus (Upton, NY, USA)



# BNL Core Facility Revitalization (CFR) Project: Design & Construction of B725 Data Center

Main design features of the new (B725) data center:

- A single large data hall for CPU & DISK resources (Main Data Hall) divided into two aisles:
  - Low Density Area (HTC): 16 rows of 20 racks each + one row of 16 racks
  - High Density Area (HPC): 14 rows of 10 racks each + one row of 8 racks
  - 478 rack positions in total with up to 9.6 MW of IT load combined in a fully built out configuration
- 188 rack positions are available in the MDH as delivered by the CFR project in 2021:
  - 158 rack positions with 2.4 MW of power/cooling available on the Low Density Area of MDH
  - 30 rack positions with 900 kW of power available on the High Density Area of the MDH
  - 1.2 MW of diesel generator backup power (IT load) with upgrade to a 1+1 redundant generator pair performed in 2022Q1 (two more generators to be added in FY24-25)
  - Unlocking 290 remaining rack positions will require construction of additional electrical rooms, installation of additional power distribution and UPS equipment, chillers, cooling towers and diesel generators
- Dry-pipe/pre-action double interlock sprinkler system for fire suppression in the MDH
- Fire suppression in the Tape Room and Network Room is 3M Novec 1230 based
- Standard APC 42U racks 600mm wide, 1070 mm deep (HTC compute) or 1200 mm deep (storage/infrastructure, HPC compute) are to be used across the entire floor of B725 MDH:
  - All equipped with water-cooled rear-door heat exchangers (RDHx units)
  - Chilled water supplied from under the raised floor: nothing but water is distributed under the raised floor in the B725 MDH
- 3 level of overhead cable/power distribution:
  - Power/busbars, cable trays with mini-racks attached for fiber (LC) & copper (RJ-45) patch panels
  - No patch panels are placed inside the racks with equipment



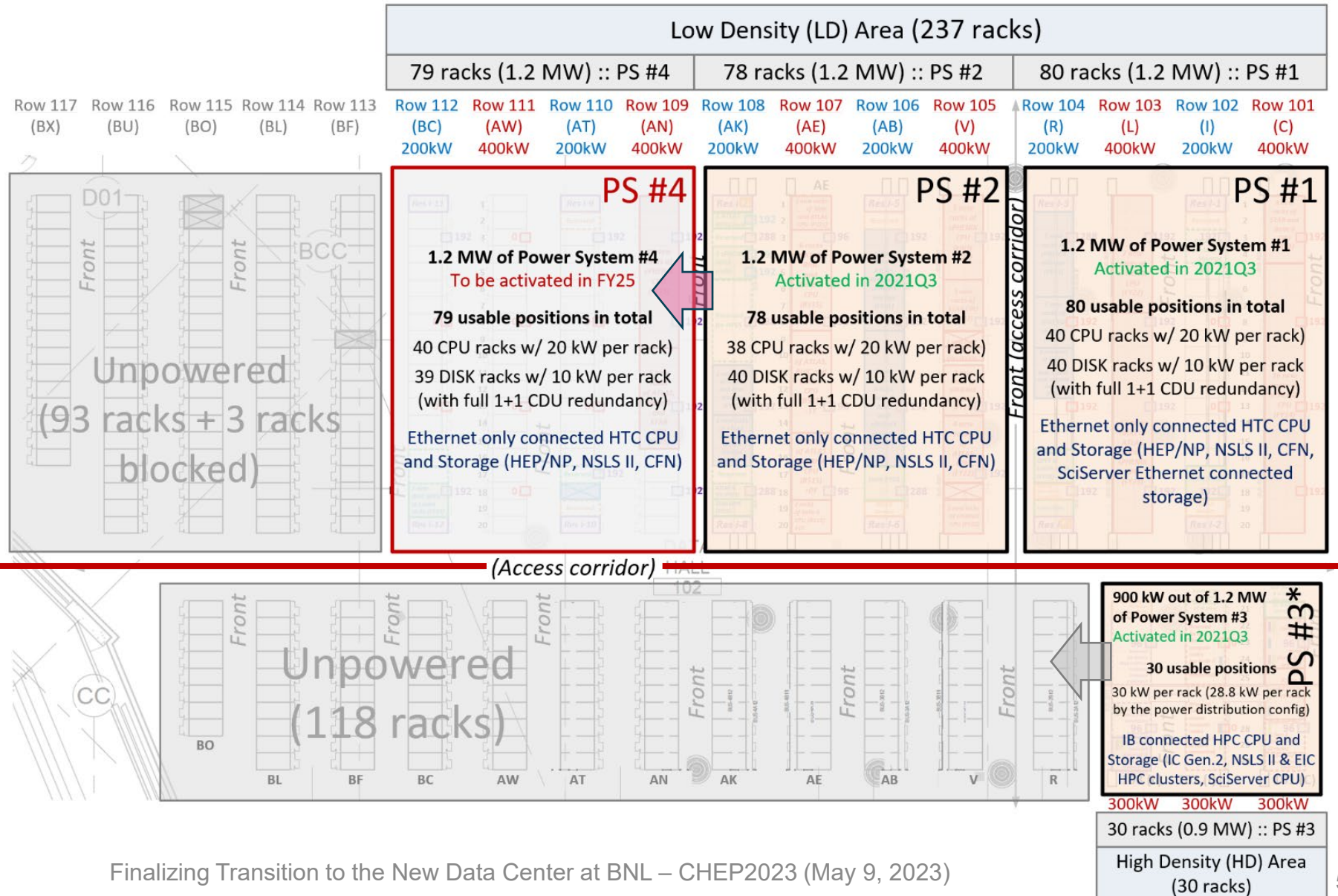
# B725 Data Center: Main Data Hall

Engineering study is in progress for addition of (1.2 MW of IT payload capable) PS #4 to B725 data center as of FY23 (3.6 MW → 4.8 MW upgrade)

Procurement & installation of infrastructure equipment is expected to occur in FY24-25

**Low Density Area (330 racks max: 162 CPU + 168 storage/infrastructure racks)**

**High Density Area (148 racks max)**





# BNL Core Facility Revitalization (CFR) Project Timeline

The main stages of the project went as follows:

- CFR project finished the design phase in the first half of 2019 and completed the construction phase by the end of FY21
  - COVID-19 pandemic caused the delay of construction completion for the project by about 9 months (mainly due to a period of construction work being suspended in 2020)
  - The global supply chain issues for IT components triggered by the pandemic in 2021 are still ongoing (mostly affecting network and CPU purchases) and are likely to continue into FY24 imposing 6-9 months delay with data center IT payload scaling compared to the pre-pandemic time (2019-early 2020)
- The occupancy of the B725 data center for production CPU and DISK resources for all programs started in 2021Q4 and ramped up in 2023Q2 to the level of **78** racks populated with equipment in the B725 Main Data Hall (MDH)
  - Pre-pandemic plans implied deployment of all of the FY21 purchase equipment in the new data center, which had to be corrected due to the delay in construction
  - The bulk of FY21 equipment purchases was affected by the global supply chain issues resulting in equipment deliveries delayed up to 6-9 months, so we were able to place that equipment directly into B725 as originally planned, regardless of the construction delays
  - Central network equipment of B725 was activated Aug-Sep 2021
  - 20 existing CPU and CPU/GPU racks were moved from the old data center (B515) to the new one (B725) in 3 interventions carried out in Nov 2021, Jan 2022 and Jan 2023
  - 14 new CPU and CPU/GPU racks, 18 storage/infrastructure racks, and 26 high capacity JBOD racks were deployed directly in B725 in 2021Q4-2023Q2
- We are now 1.5+ years into B515 and B725 inter-operation phase of the SDCC transition to the new data center with the next major milestone coming on Sep 30, 2023 (the end of FY23) at which stage all the areas of B515 data center except the two newest ones (CDCE and Sigma-7) are expected to be permanently retired



# B725 Data Center: Main Data Hall (MDH)

- 117 rack frames are already deployed in B725 MDH so far, out of which 34 are CPU/GPU racks, 44 are active storage and infrastructure racks, and 39 rack frames are provisioned for the future growth
- 49 more rack places are available for placing more CPU racks in the Low Density Area
- 22 more rack places are available for placing more racks in the High Density Area



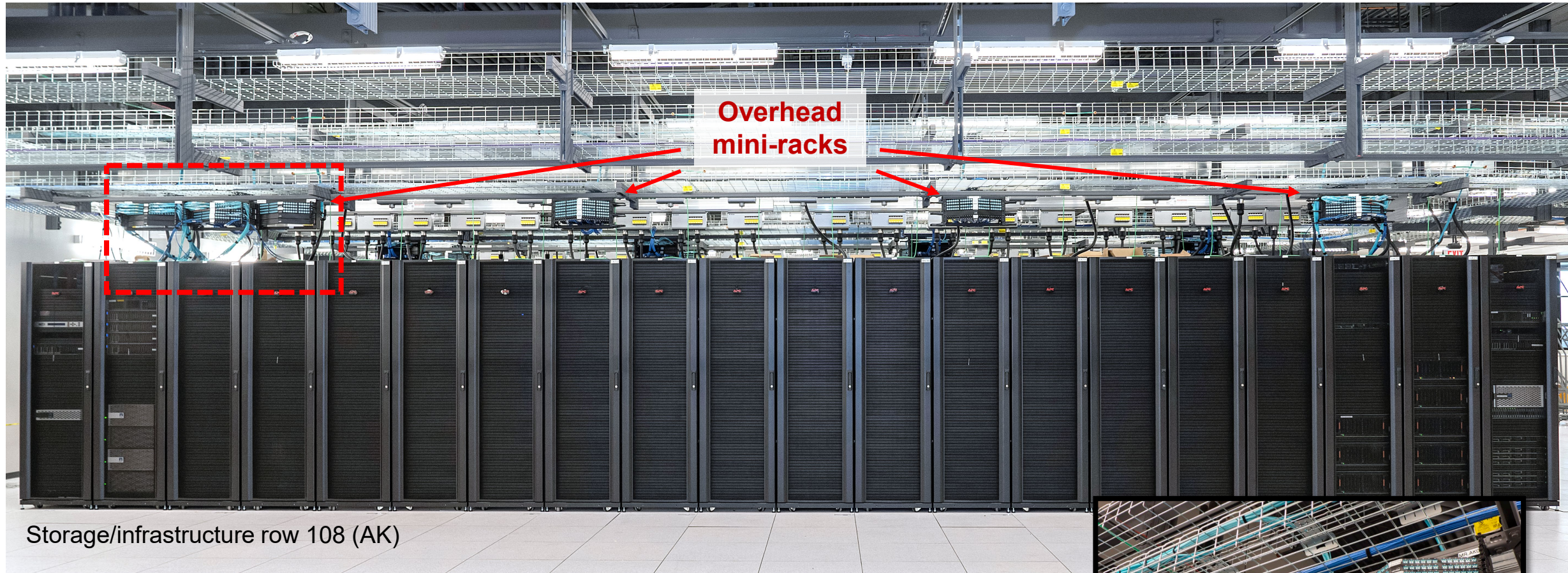
(Low Density Side)

1<sup>st</sup> fully populated CPU row (rack frames are delivered pre-configured and retired with equipment in them)

Storage/infrastructure rows  
(all rack frames are pre-deployed)



# B725 Data Center: Main Data Hall (MDH)



Storage/infrastructure row 108 (AK)

- Power distribution, and copper / fiber connectivity distribution is done from the overhead busbars and cable trays with mini-racks attached to them
- Very limited air cooling capability in the room itself – all heat removal is done by chilled water circulating through the RDHx units, with water distribution done under the raised floor (no temperature gradient in the room during normal operations, and the design setpoint is 75F)







# B725 Main Data Hall (MDH)

## Power & Cooling

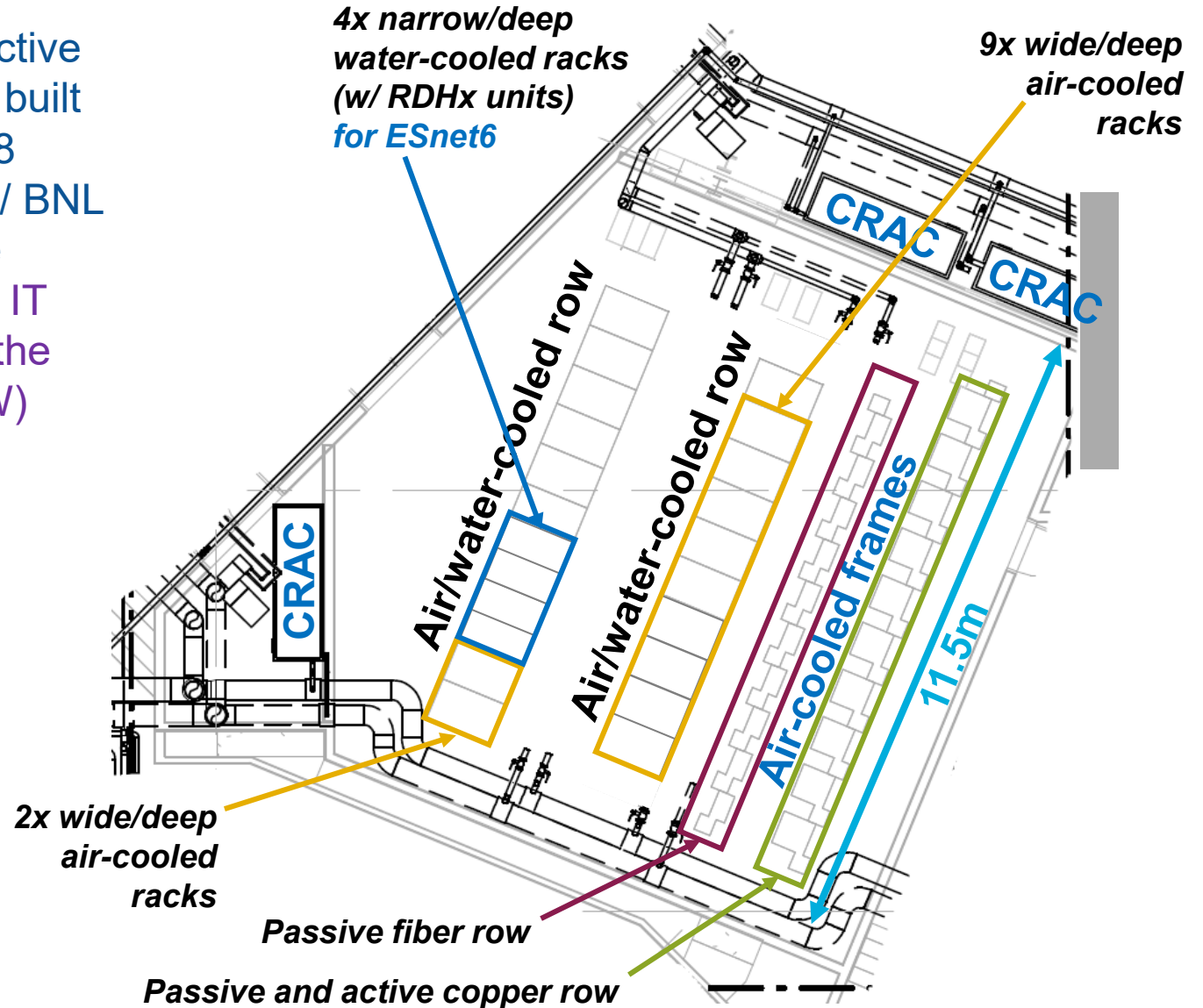
- Each rack is provided with 14.4kW capable (208V) zero-U per-outlet-sensing ServerTech CDUs fed from the overhead busbars: *two separate busbars for 1+1 redundantly powered storage and infrastructure racks (9 kW of IT power per rack) and a single busbar for non-redundantly powered CPU racks (18 kW or 27 kW of IT power per rack)*
- Each active rack on the floor of MDH is provided with an RDHx unit (Usystems ColdLogic [CL20-C12] 5-fan model that can cover the range of rack power density of 10-30 kW per rack)



# B725 Data Center: Network Room

Capable of hosting all passive and active network equipment to sever the fully built out B725 data center in 9.6 MW / 478 racks configuration, including ESnet / BNL Network Perimeter serving the entire BNL site if required (up to 0.5 MW of IT load can be deployed in this area in the future; the initial power limit is 250kW)

*Both primary power source and bypass power are battery UPS protected in this room for maximum resiliency*





# B725 Network Room

*20x 100G line cards are deployed right now (72 Tbps)  
2x more 100G line cards and the first set of 4x 400G line cards are to be added in FY23-24 to get B725 to 118 Tbps of switching capacity*

Room is cooled by a mix of CRAC units sitting in the dedicated adjacent rooms and the RDHx units deployed on a subset of rack. The cold air is supplied through perforated raised floor tiles.

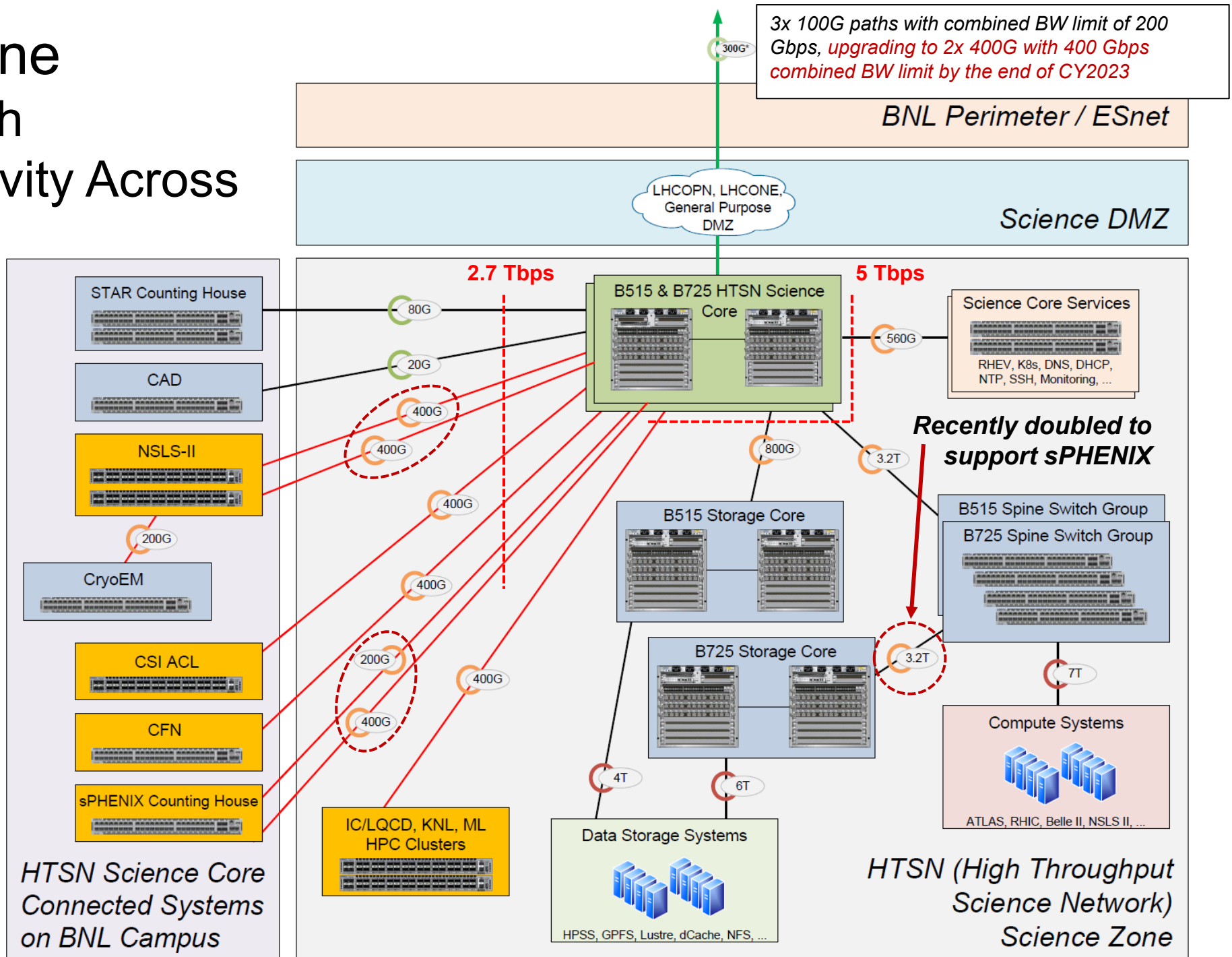
Water supply and return for RDHx units is also located under the raised floor.





# SDCC SciZone

## High Bandwidth Fiber Connectivity Across BNL Campus as of 2023Q2

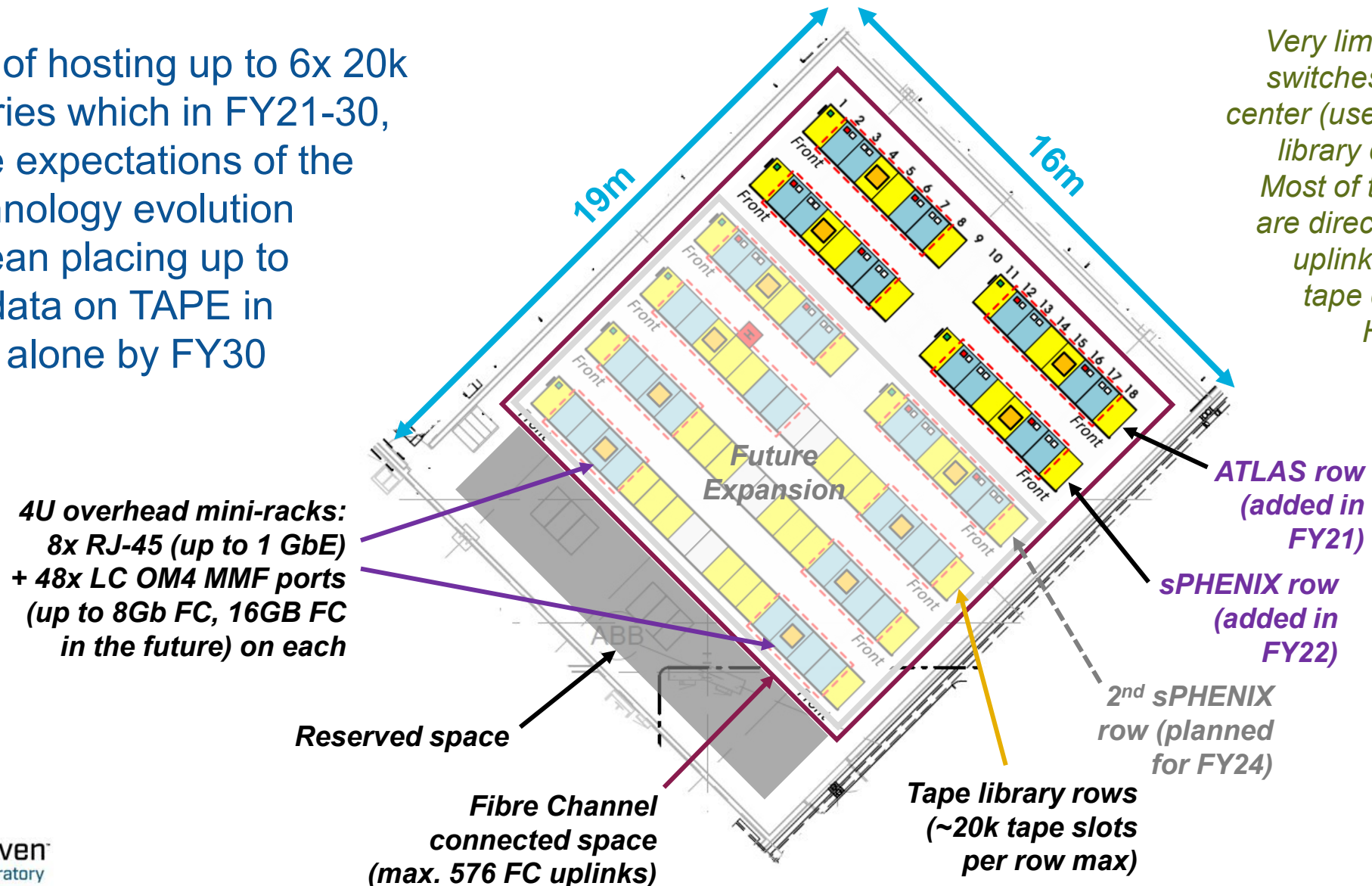




# B725 Data Center: Tape Room

Capable of hosting up to 6x 20k slot libraries which in FY21-30, given the expectations of the LTO technology evolution could mean placing up to 3 EB of data on TAPE in this area alone by FY30

*Very limited use of FC switches in B725 data center (used primarily for library control paths). Most of the FC uplinks are direct passive fiber uplinks between the tape drives and the HPSS movers.*





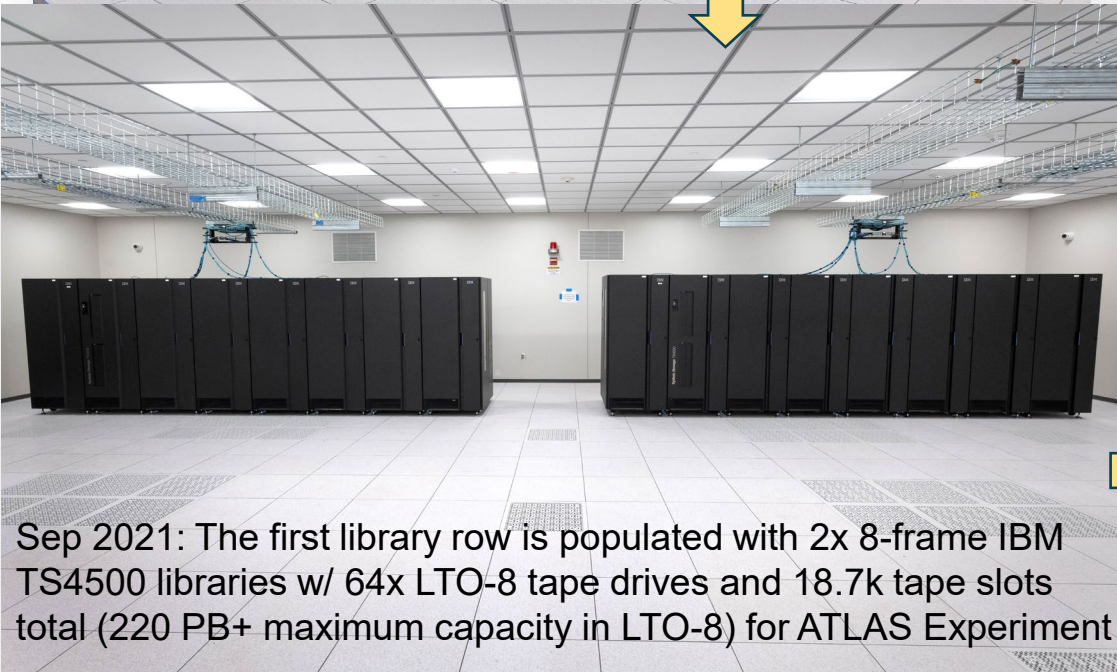
Jul 2021: B725 Tape Room is Mechanically & Electrically Ready for Equipment Deployment

Overhead cable trays for all fiber & copper connectivity

Room is air-cooled by the CRAC units sitting in the dedicated adjacent room. The cold air is supplied through perforated raised floor tiles.

Hot air return is through these vents

Power distribution is all under the raised floor (lowest power density in the entire data center)



Sep 2021: The first library row is populated with 2x 8-frame IBM TS4500 libraries w/ 64x LTO-8 tape drives and 18.7k tape slots total (220 PB+ maximum capacity in LTO-8) for ATLAS Experiment

# B725 Tape Room

Mar 2022: The second library row is populated with 2x 8-frame IBM TS4500 libraries w/ 64x LTO-9 tape drives and 18.7k tape slots total (330 PB+ maximum capacity in LTO-9) for sPHENIX Experiment





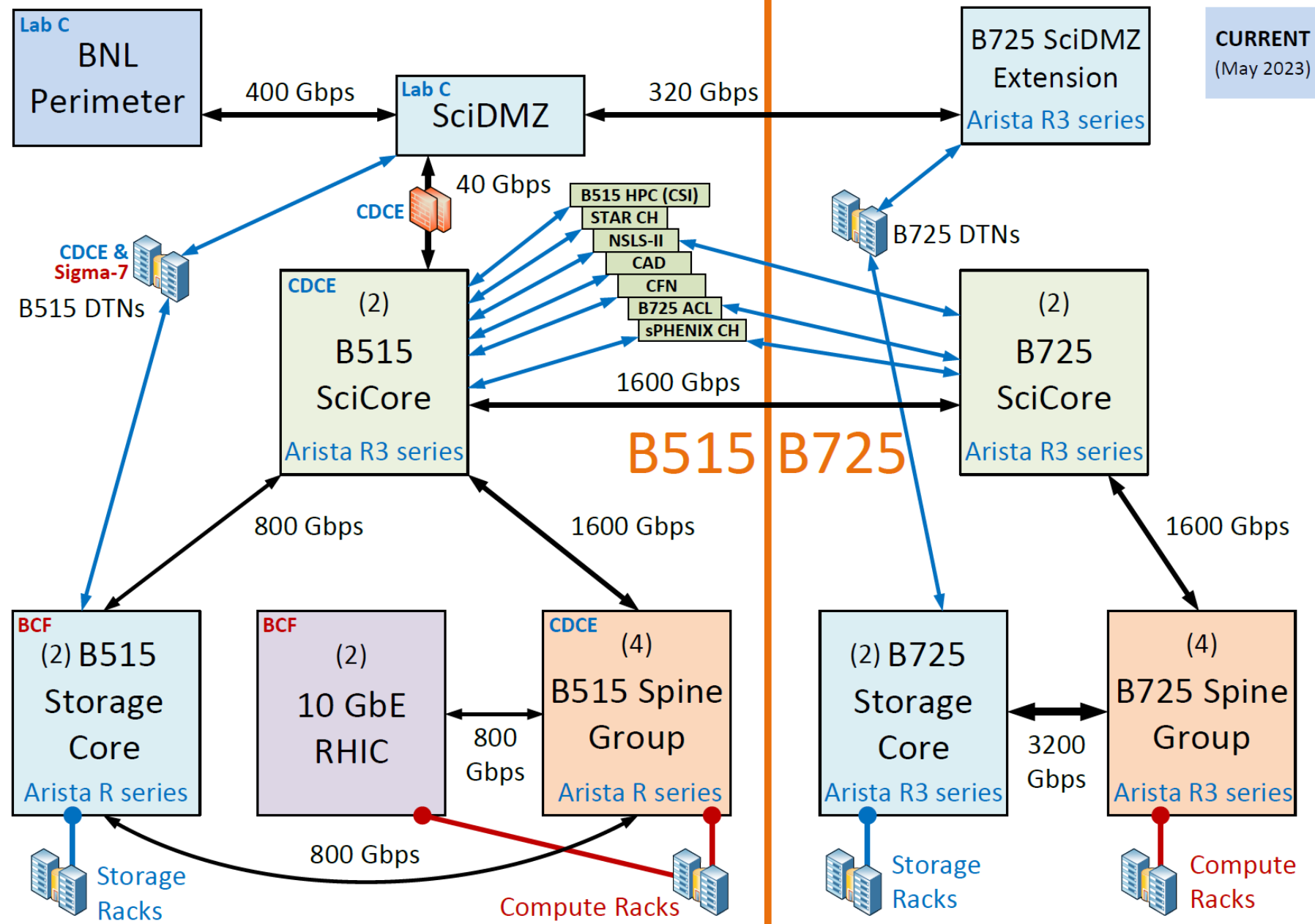
# B515 to B725 Transition

The main goals of the transition area as follows:

- Place all new equipment purchased starting from mid-FY21 to the new (B725) data center with very few exceptions (Oracle SL8500 tape library complex specific systems, Campus connected test / evaluation systems)
- Retire the oldest and least efficient (from PUE standpoint) areas of the old (B515) data center by the end of FY23 (Sep 30, 2023).
  - The list of areas includes Main BCF, RCF, BGL and QCDOC areas
  - The power distribution, air-cooling infrastructure and all underfloor fiber & copper infrastructure is expected to be retired in these areas and later repurposed as non-data center space
  - Most of the equipment being retired in these areas belonging to HEP/NP programs are getting hardware refreshed into CDCE area since 2018 and subsequently into B725 starting from 2022
  - A subset of equipment belonging to BES programs that was deployed in the QCDOC area earlier is getting moved to B725 MDH in 2023
- Convert Sigma-7 area of the old (B515) data center into BNL Campus only connected area by end of FY23
  - Consolidating all BNL Campus connected systems that used to be hosted in Main BCF and QCDOC areas that are needed to be kept operational in a longer term into Sigma-7 area by that date
- Several network interventions to be performed in the upcoming 7 months are needed to achieve these goals:
  - Migrate BNL perimeter equipment from being deployed in B515 Lab C) into a redundant configuration using both B515 Lab C and B725 Network Room before the end of FY23
  - Migrate the B515 Storage Core network system from Main BCF to CDCE in a transparent fashion before the end of FY23 (all systems connected to it are redundantly attached using LACP which allows us to do so)
  - Migrate the primary source of Science DMZ (WAN) connectivity from B515 (Lab C) to B725 Network Room in 2023Q4 – a scheduled disruptive network intervention is going to be needed in 2023Q4 (likely in Dec 2023)



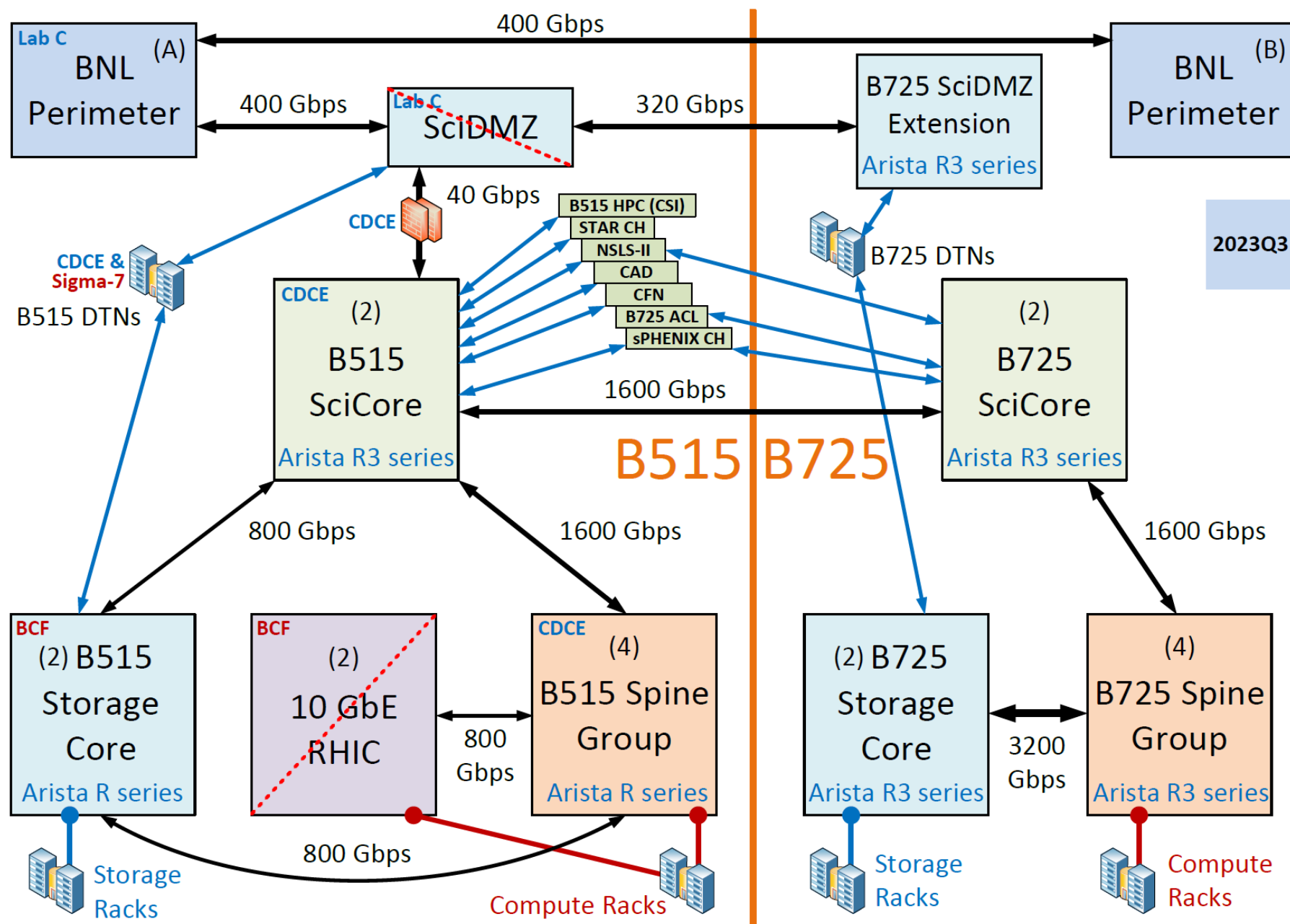
# SDCC SciZone & SciDMZ Network Systems





# SDCC SciZone & SciDMZ Network Systems

Migration to new  
(400 GbE enabled)  
BNL Perimeter  
later this year

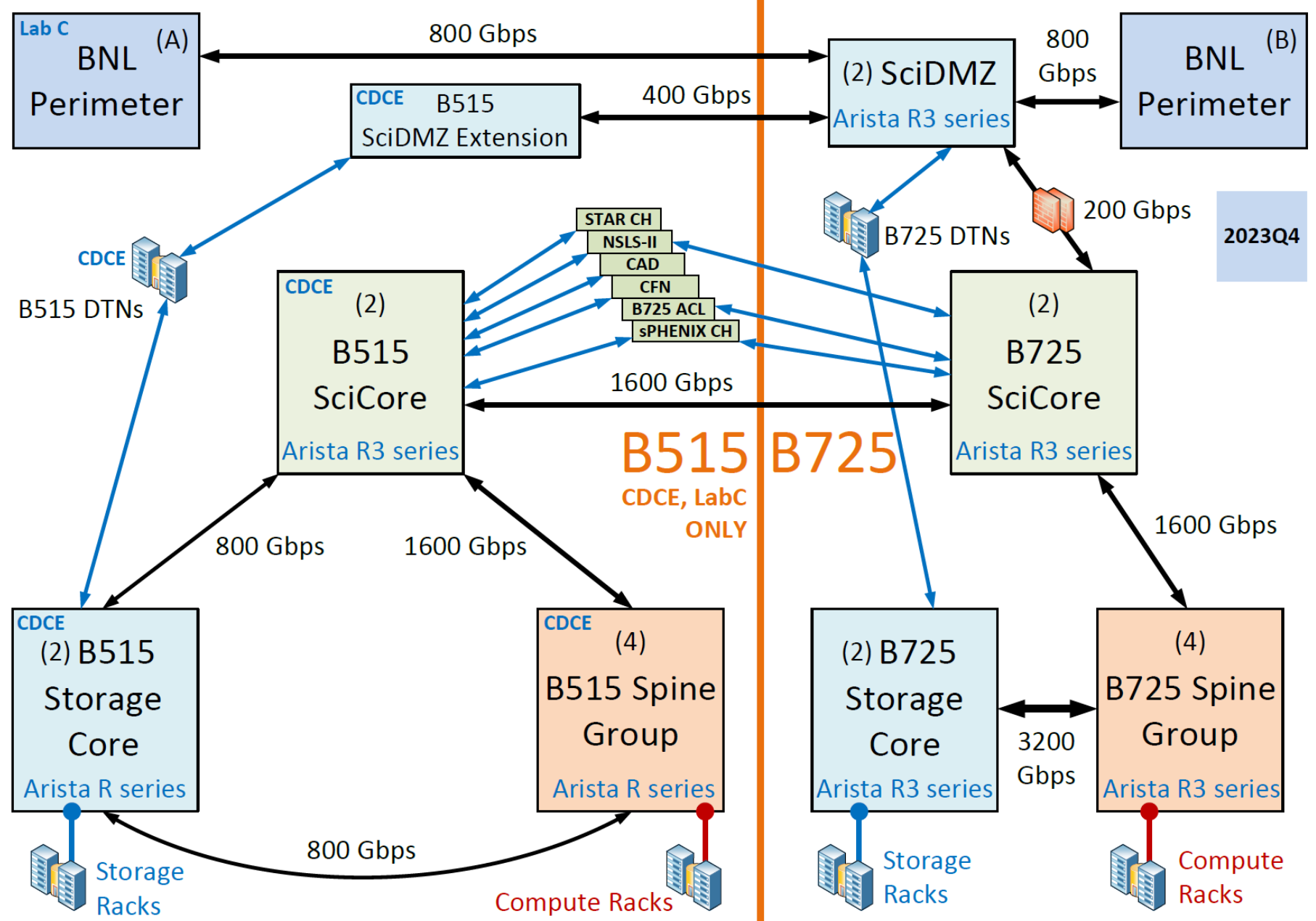




# SDCC SciZone & SciDMZ Network Systems

Migration of B515  
Storage Core switch  
pair to CDCE area  
out of the way of  
BCF retirement by  
the end of FY23

Reaching the  
steady state for  
FY24-25





# B515 Data Center in 2023Q2-3

End of FY23  
(projection)

108 rack frames  
+ 10 tape silos  
remaining

Sigma-7 with  
BNL Campus  
connectivity only

Current State  
(May 2023)

180 rack frames  
+ 10 tape silos  
across all areas

70 racks left to retire +  
16 racks to migrate in  
May-Sep 2023

CDCE with full  
SciZone & SciDMZ  
connectivity

# Summary

- CFR project finished the design phase in the first half of 2019 and completed the construction phase by the end of FY21
- The occupancy of the B725 data center for production CPU and DISK resources for all programs started in 2021Q4 and ramped to the level of 78 racks populated with equipment in the Main Data Hall in 2023Q2
- Two library rows in B725 Tape Room were populated with IBM TS4500 tape libraries so far to serve ATLAS and sPHENIX experiments, and one more library row is expected to be added for sPHENIX in FY24
- The combined IT payload deployed in B725 data center is expected to exceed 1.2 MW in 2023Q3-4
- A series of network interventions is to be performed in 2023Q2-4 period to achieve a building level redundancy for BNL WAN perimeter and minimize the dependency of SDCC central network systems on B515 infrastructure
- Transition of the majority of CPU and DISK resources deployed in SDCC environment to the B725 data center is expected by the end of FY23 and the oldest areas in B515 data center are planned to be retired at that stage
- This would signify the completion of 6 years long process of preparation and execution of B515 to B725 data center transition for SDCC Facility (2018-2023)



# Questions & Comments

