SDCC Datacenter Transformation within the Scope of BNL CFR Project and Beyond

Alexandr ZAYTSEV

alezayt@bnl.gov

BROOKHAVEN NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

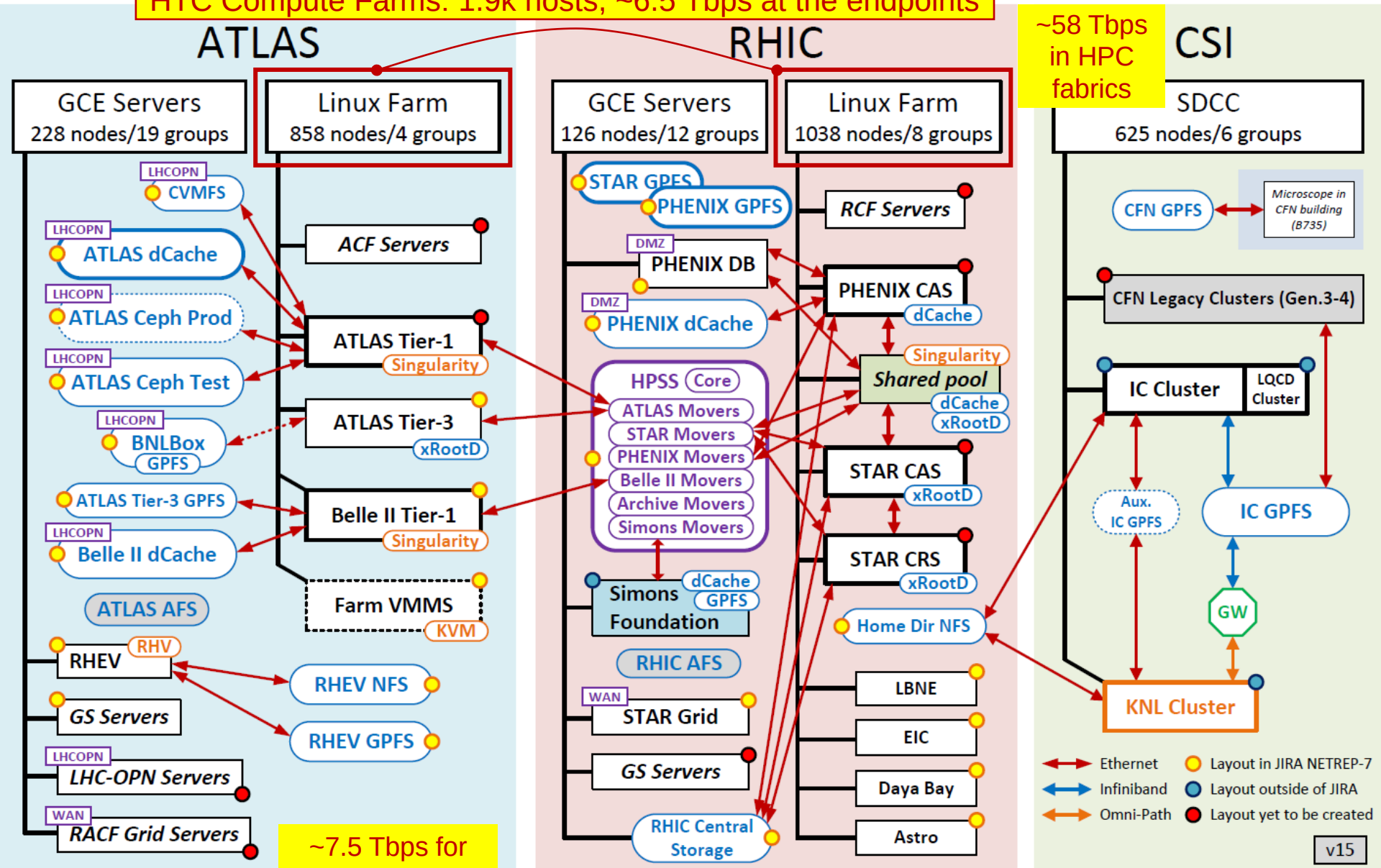
B515 Datacenter

- SDCC Facility B515 datacenter is a 1.5MW scale aircooled universal datacenter operated under the umbrella of BNL CSI (Computational Science Initiative), providing
  - HTC and HPC computing systems
  - Spinning drive based storage
  - Robotic tape storage
  - *Please refer to the BNL Site Report by O.Rind for more details*
- serving multiple international collaborations, research communities based at BNL, and also NY State based research organizations
  - PHENIX and STAR at RHIC (BNL)
  - ATLAS Experiment (ATLAS Tier-1 Site)
  - Belle II Experiment at KEK (Belle II Tier-1 Site)
  - Simons Foundation (SF)
  - NSLS II (BNL) is joining in, potentially 60+ beamline group
  - CFN (BNL) – two groups, may grow higher in connection to NSLS-II
  - CSI research groups and testlabs

SDCC Subsystems (2875 nodes/49 groups on Mar 23, 2019)

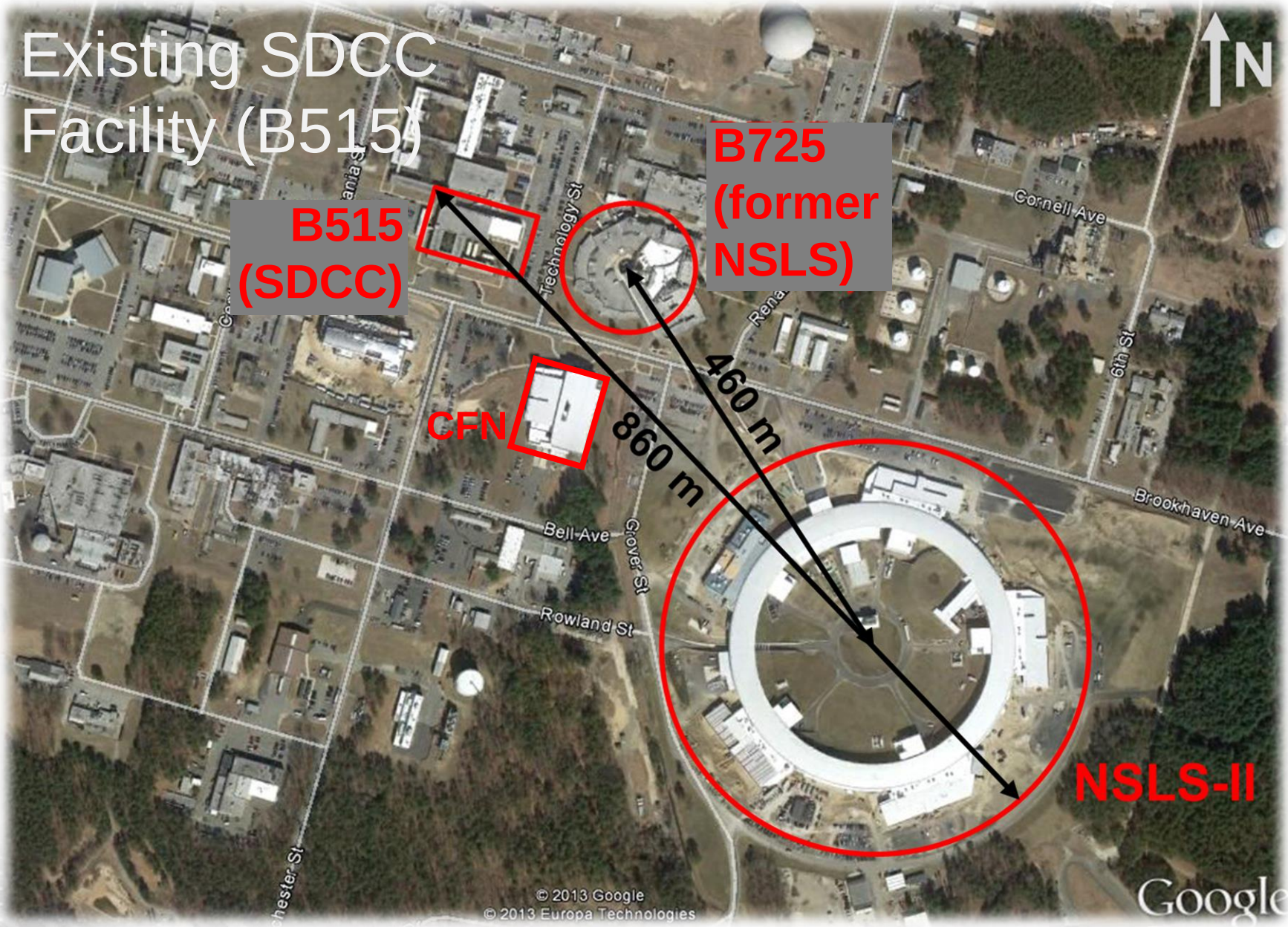
HTC Compute Farms: 1.9k hosts, ~6.5 Tbps at the endpoints

~58 Tbps in HPC fabrics

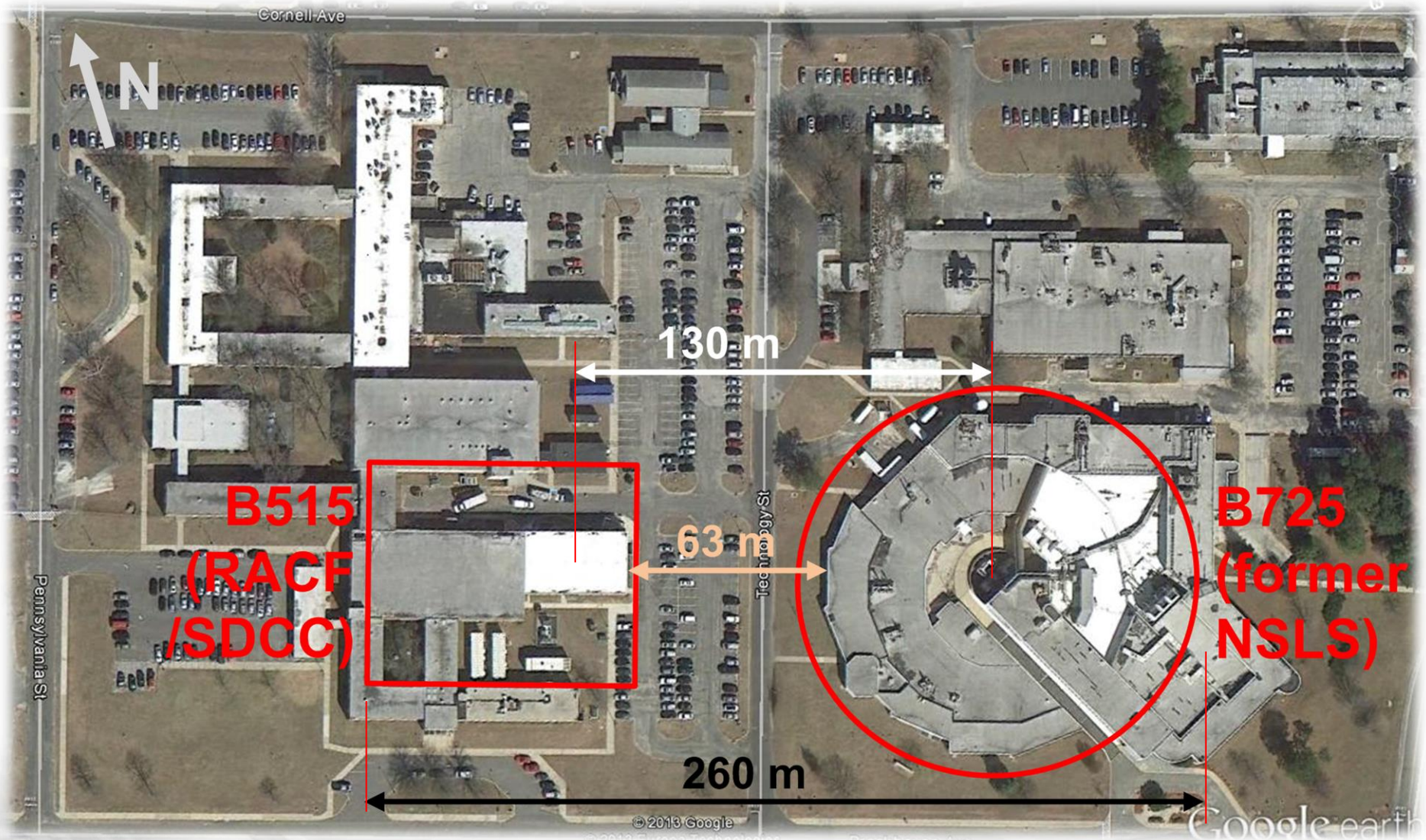


~72 Tbps for aggregate unidirectional endpoint bandwidth in total

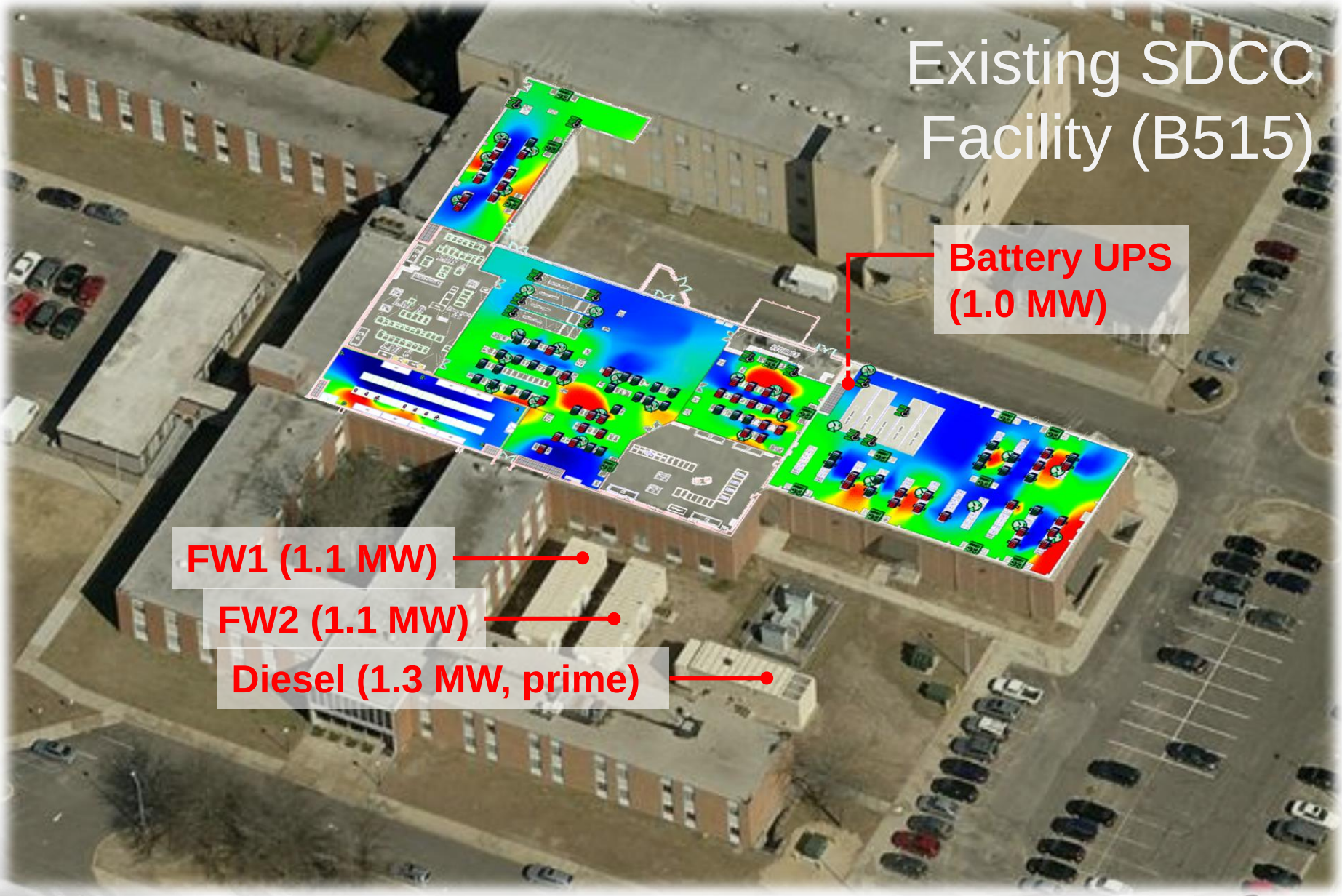
Existing SDCC Facility (B515)



Outlook: B515 & B725



Existing SDCC Facility (B515)

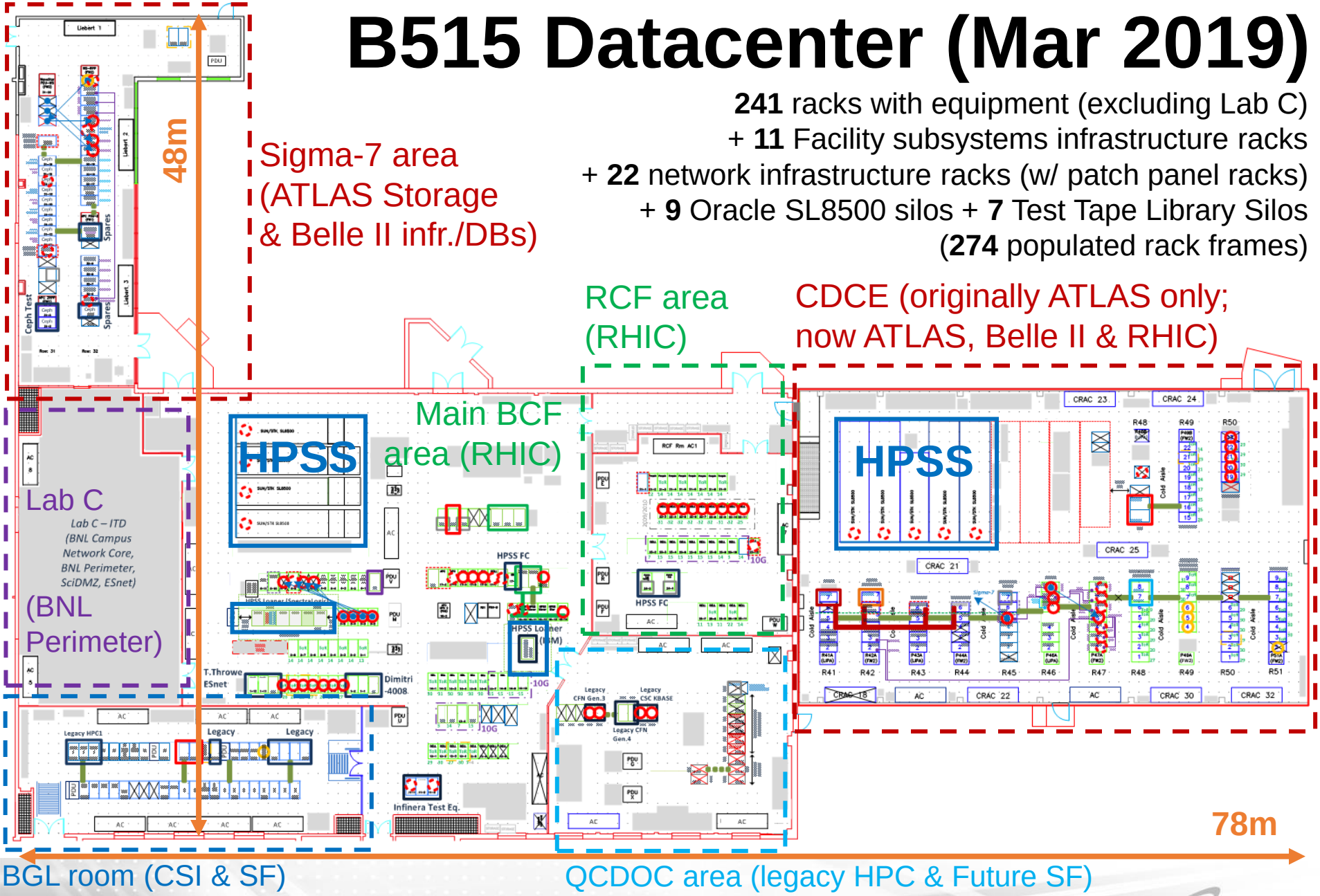


B515 Datacenter Areas

- SDCC B515 datacenter incorporates three IT areas located in the same building each provided with it's own cooling solution yet sharing the power distribution infrastructure
  - **BCF area** (the oldest part of the datacenter with operational history pre-dating the SDCC Facility and going back to the 60s), now subdivided into several physically isolated areas named
    - **RCF** (RHIC Computing, Inergen based fire suppression)
    - **QCDOC** (former BG/Q supercomputer area, now converted to host Simons Foundation resources)
    - **BGL** (former BG/L supercomputer area, now converted to host CSI and part of SF resources)
    - **Lab C** (IT department managed central network resources, BNL perimeter equipment, ESnet equipment, fiber cross for the entire BNL site, Inergen based fire suppression)
    - **Main BCF** (the rest of space left)
  - **Sigma-7 area** (highest power redundancy and cooling infrastructure reliability area; no Inergen fire suppression here though)
  - **CDCE** (the newest addition to B515 datacenter; built in 2009)

B515 Datacenter (Mar 2019)

241 racks with equipment (excluding Lab C)
+ **11** Facility subsystems infrastructure racks
+ **22** network infrastructure racks (w/ patch panel racks)
+ **9** Oracle SL8500 silos + **7** Test Tape Library Silos
(**274** populated rack frames)



B515 Datacenter Mechanical Layout

- SDCC B515 loading bays and ramps:
  - There are 4 loading bays of various heights and state of usefulness, no motorized platforms or embedded lifts on any of them (only two are used for large scale equipment receiving and retiring nowadays)
- SDCC B515 raised floor: 3 distinct configurations
  - 12 in deep; on stacks w/o horizontal stringers using tile type 1 in the main BCF, RCF, QCDOC areas
  - 12 in deep; on stacks w/o horizontal stringers using tile type 2 in Sigma-7
  - 30 in deep; **with** horizontal stringers using tile type 2 in CDCE
  - 18 in deep; on stacks w/o horizontal stringers using tile type 1 in BGL
- SDCC B515 cable management solutions
  - **Main BCF:** underfloor copper and fiber (no trays); overhead cable trays bridging the rows of equipment; copper and fiber along the rows
  - **RCF:** all copper and fiber under the floor and along the rows
  - **Sigma-7:** limited amount of copper under the floor; fiber and copper over overhead cable trays and along the rows
  - **CDCE:** bulk fiber and limited amount of copper on the **underfloor trays**, mostly shared with power receptacles; LC-LC fiber patch cords and most of the copper are on top of the rows and overhead cable trays
  - **BGL:** copper under the floor with cable trays, LC-LC fiber patch cords plus the 4X FDR/EDR IB and OPA Gen.1 fiber over the **dedicated cable tray system**

B515 Datacenter Cooling Systems

- SDCC B515 cooling systems include:
  - 30 Liebert downflow CRAC units in 7 groups (a mix of variable frequency and non-variable frequency drives), cooled by the BNL site chilled water, with pump system located outside the Facility
    - 1 group of 3 CRAC units in Sigma-7 is on diesel backup
    - 27 CRAC units are on unprotected LIPA (2 are in standby mode)
  - 2 building level air handler units in the basement serving the BCF areas through the openings in the concrete slab under the raised floor
  - 2 groups of the overhead Liebert XDV units with refrigerant feed via pipes from above the false ceiling, with chilled water cooled Liebert pumping unit deployed nearby on the floor (not in use anymore, to be removed in FY20-21)
  - Water cooled BG/Q installation in QCDOC with water-to-water heat exchanger located nearby on the floor (removed in 2019Q1)
- Most of the datacenter areas (especially the main BCF area) has the ratio of volume to power dissipation to survive for ~20 min with CRAC units down (sufficient to restart the CRAC units after a short LIPA interruption, or gracefully shutdown the majority of IT payload)
- Most of the areas are arranged into hot & cold aisles without isolation of either one of them

B515 Datacenter Power Distribution

- SDCC B515 main power distribution features:
  - 2x 1.1MW Flywheels, 1.3MW Diesel, 1.0MW Battery UPS (basement)
  - Underfloor power distribution based on floor PDUs and flexible cables with receptacle boxes (mostly 3-phase 30A 208V; many receptacle types/feed types across multiple areas)
  - 1-4 CDUs per rack, mostly 0U vertical CDUs, up to 14.5kW per CDU (newest SF compute racks), while the rest of the compute racks are capped at 12kW per rack
  - Many different power redundancy schemas utilized by IT racks:
    - LIPA only (available in CDCE and Sigma-7)
    - Battery UPS only – compute in BCF
    - Battery UPS + single FW (backed by diesel) – compute and storage in BCF
    - **LIPA + single FW (backed by diesel)** – storage in CDCE
    - **Single FW (backed by diesel)** – compute in CDCE
    - **FW1 + FW2 (both backed by the diesel)** – storage in Sigma-7
  - CRAC units are mostly on LIPA, except the Sigma-7 area which has the CRAC units backed by the Facility diesel, but the pumps for the chilled water are yet to be provided with diesel power on the side of the BNL chilled water plant

B515 Datacenter Equipment Racks

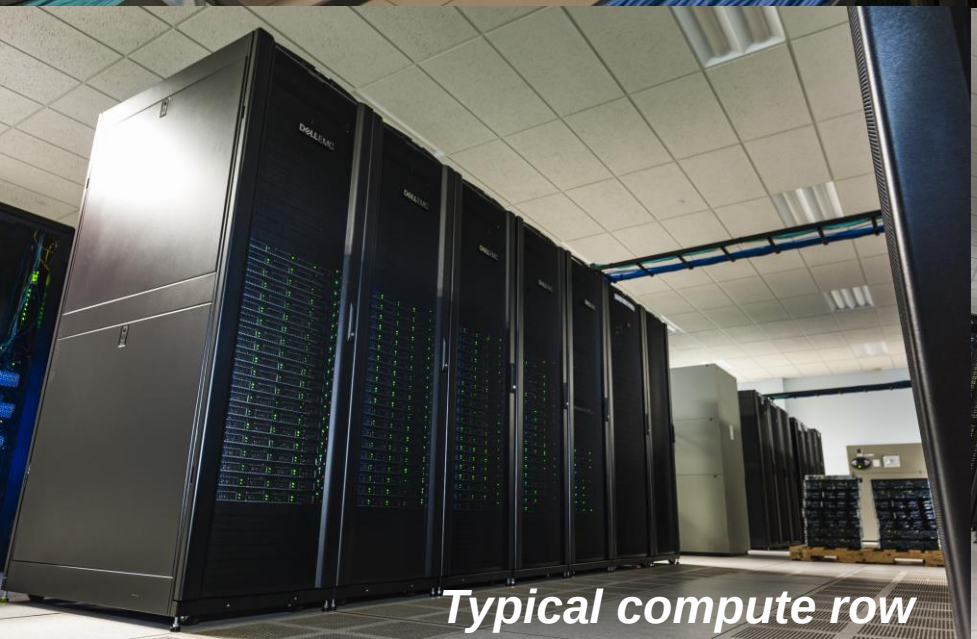
- SDCC B515 IT equipment enclosure fleet (325 rack frames) is dominated by the APC and Dell/APC 42U tall racks of 4 different form-factors:
  - 600mm wide 1070mm deep: most of our compute node racks
  - 600mm wide 1200mm deep: all new JBOD storage require this depth starting from FY18
  - 750mm wide 1070mm deep: legacy network and some legacy storage racks) – getting phased out as the equipment is being replaced
  - 750mm wide 1200mm deep: standard network rack form-factor for both active and passive (patch panel only) network racks
- Compute node rack come into the Facility pre-assembled and tested (in some cases network equipment need to be added to them on site)
- Storage and network racks are assembled and tested on site from components shipped and delivered separately
- 42U rack height seem to be sufficient for all our use cases even for the high power density and is also convenient for maintenance by the Facility personnel (no need for tall ladders)

B515/BCF

B515/BGL


B515/Sigma-7
APC
CAUTION
NOISE HAZARD
HEARING PROTECTION
REQUIRED IN THIS

B515/CDCE

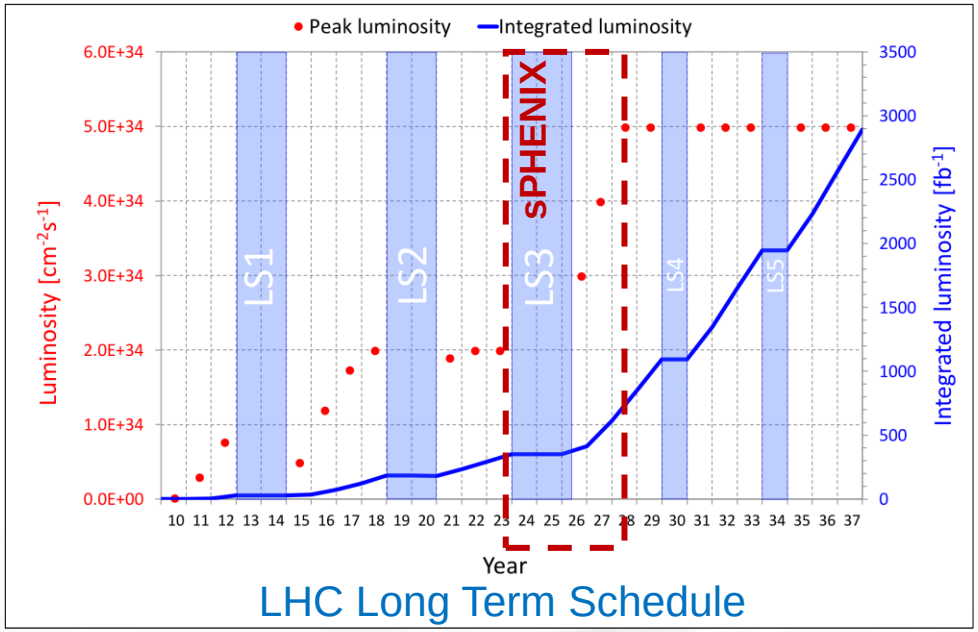


Typical compute row

Typical HW RAID storage row

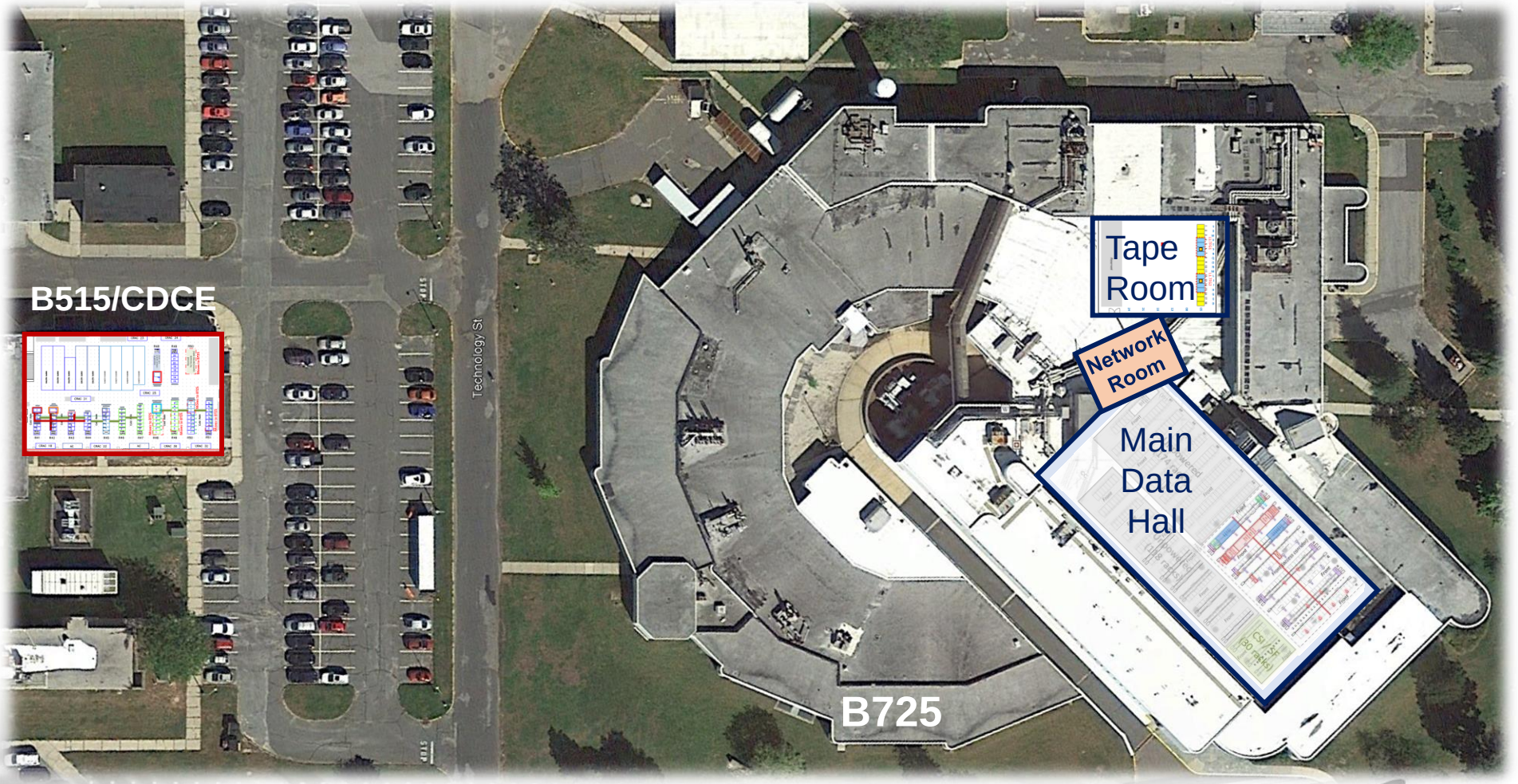
The Path Forward

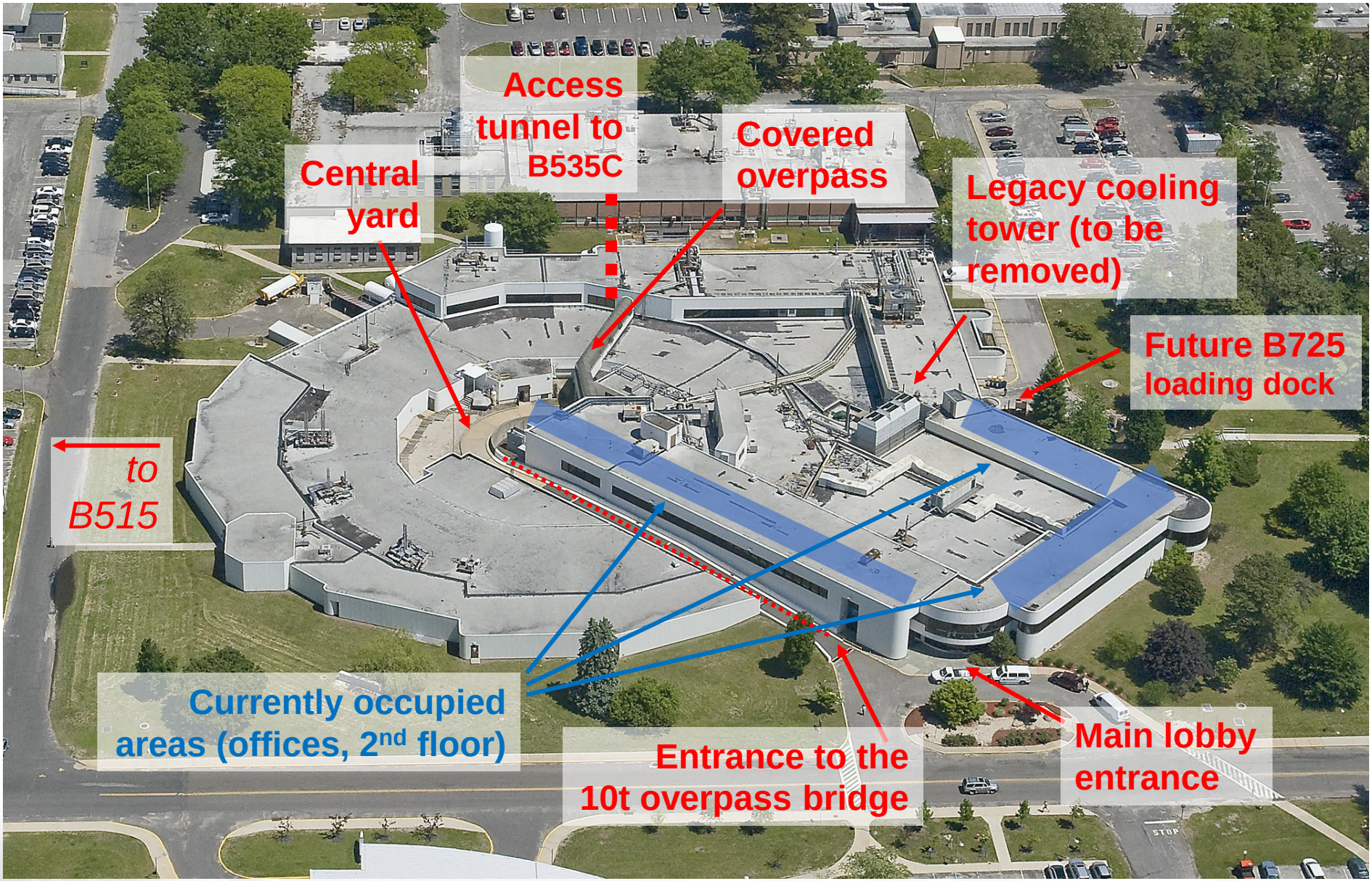
- The existing SDCC B515 datacenter is a highly non-uniform facility aggregating the history of several decades of infrastructure solutions, with none of its areas providing the feature set needed for addressing the future points of growth associated with:
  - sPHENIX Experiment at RHIC starting from FY23-24
  - ATLAS / High Luminosity LHC (HL-LHC), starting from FY26
  - Bringing 60+ beamlines of NSLS-II fully on-board with using our resources
  - Scaling up the CSI HPC systems at BNL and increasing the power density of HPC racks up to 30-50kW/rack (current limit is about 15kW/rack)
  - Scaling the datacenter beyond 2 MW power profile while protecting all its IT payload from the site-wide power outages and allowing the payload to remain operational during the prolonged site-wide power outages



LHC Long Term Schedule

Proposed solution: Constructing the new datacenter in B725 in FY19-21, migrating all spinning disk storage and compute to it in FY21-23; leaving the B515 datacenter reduced to just one area (CDCE) as a tape storage room

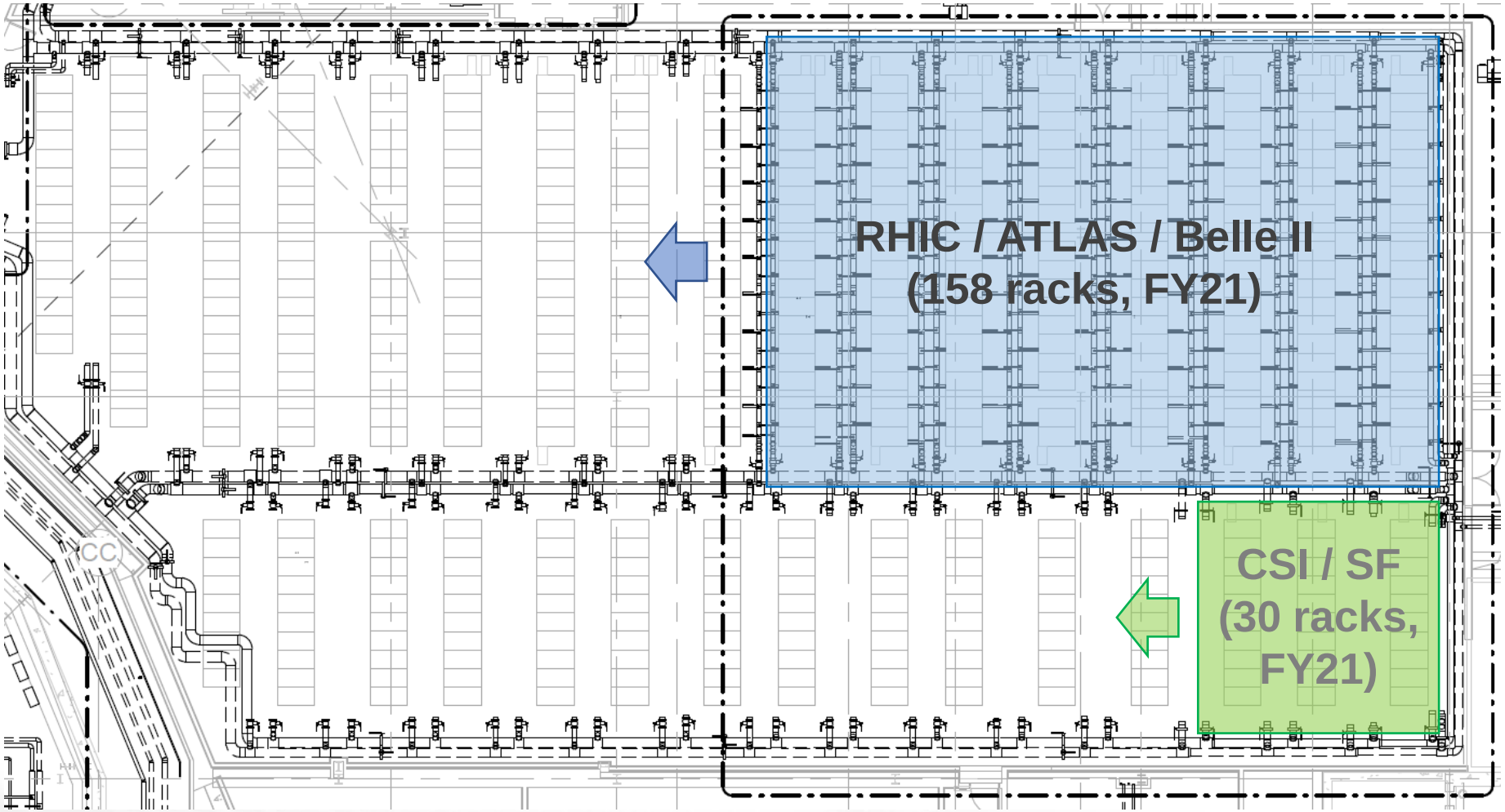



Access tunnel to B535C
Central yard
Covered overpass
Legacy cooling tower (to be removed)
Future B725 loading dock
to B515
Currently occupied areas (offices, 2nd floor)
Entrance to the 10t overpass bridge
Main lobby entrance

BNL Core Facility Revitalization (CFR) Project: Design & Construction of B725 Datacenter

- Main design features of the SDCC B725 datacenter:
  - A single large data hall for the compute and spinning disk based storage equipment **(Main Data Hall)** divided into two aisles
    - RHIC/ATLAS/Belle II aisle: 16 rows of ~20 racks each + one row of 16 racks
    - CSI/SF aisle: 14 rows of 10 racks each plus one row of 8 racks)
    - 480 rack positions in total
  - 188 rack positions to be available starting from 2021
    - 158 for RHIC/ATLAS/Belle II with 2.4MW of power/cooling available
    - 30 for CSI/SF with 900kW of power/cooling available,
    - Unlocking 292 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
  - APC 42U racks (600mm wide, 1070mm deep (HTC compute) or 1200mm deep (JBOD storage)), all equipped with watercooled rear-door heat exchangers with chilled water supplier from under the raised floor (isolation valves on the row-level and individual rack level on the water pipes; zoned drainage system in the concrete floor; nothing but water is distributed under the raised floor).
  - 3 level of overhead cable/power distribution: power/busbar, fiber tray with mini-racks attached, copper tray with RJ-45 patch panels attached. No patch panels inside the racks with equipment.

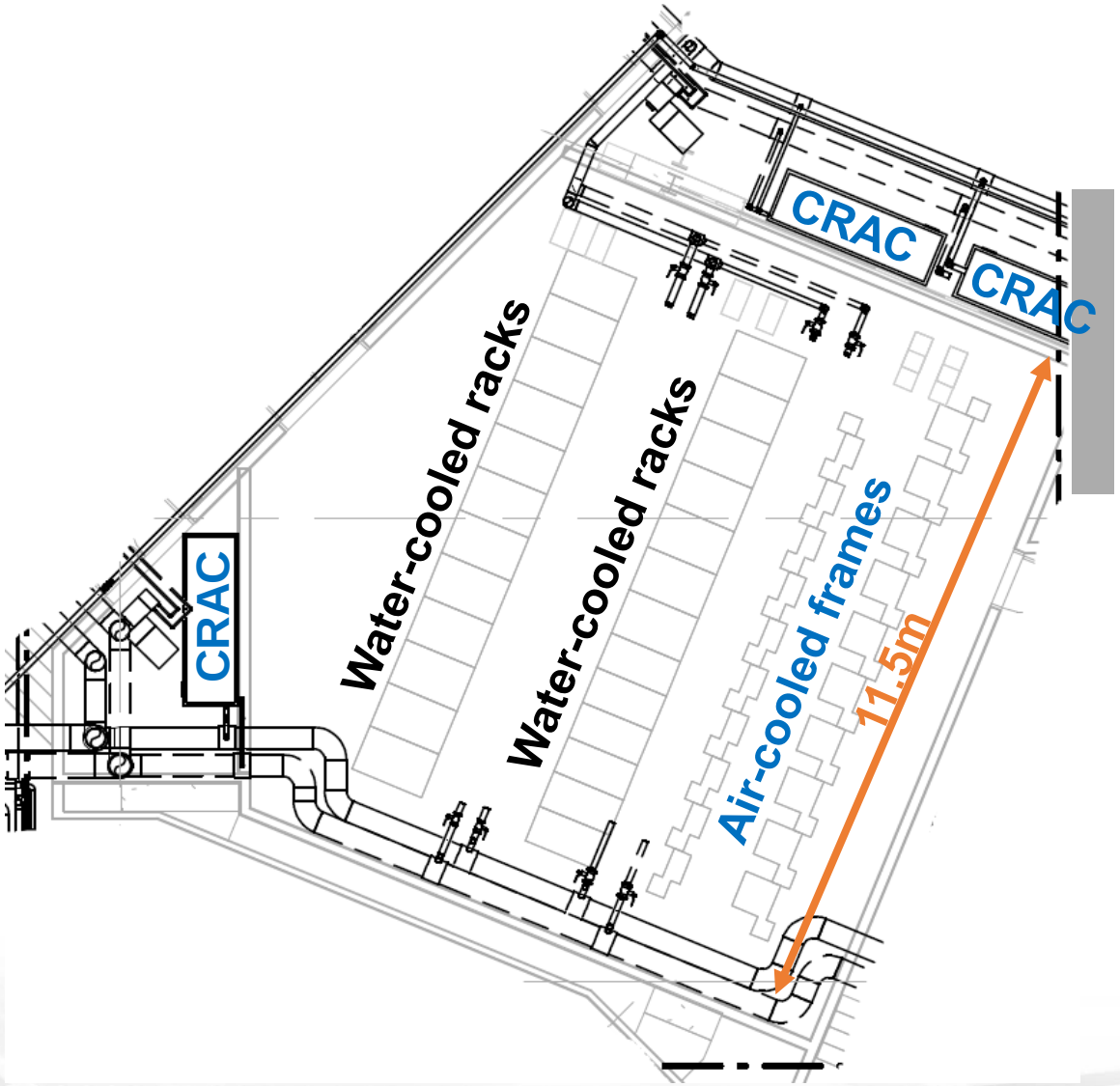
B725 Datacenter: Main Data Hall



BNL Core Facility Revitalization (CFR) Project: Design & Construction of B725 Datacenter

- Main design features of the SDCC B725 datacenter (cont.):
  - Dedicated network room to host all central network equipment, DCIM/management networks requiring direct copper distribution included
    - The B725 network room is going to host only the B725 HTSN equipment starting from early FY21
    - Science DMZ equipment is to migrate here from B515 in early FY22
    - BNL Perimeter and ESnet equipment is expected to migrate here in FY24-25
  - Intergen based fire suppression system for the network room
  - Dedicated 2+1 redundant CRAC group to remove the ambient heat dissipation from the room (located outside the room)
  - 300kW battery UPS (shared with the Tape Room) on the bypass for additional protection during maintenance
  - APC 42U racks (750mm wide or 600mm wide, 1200mm deep) for high power equipment, all provided with water-cooled rear-door heat exchangers with chilled water supplier from under the raised floor similar to the Main Data Hall.
  - 2-post and 4-post open frame racks for fiber patch panels plus copper patch panels and 1 GbE switches supporting (DCIM/management networks) – a dedicated row for each
  - 3 level of overhead cable/power distribution: power/busbar, fiber trays and copper trays optimized for the exact position of the switch equipment and the corresponding patch panels

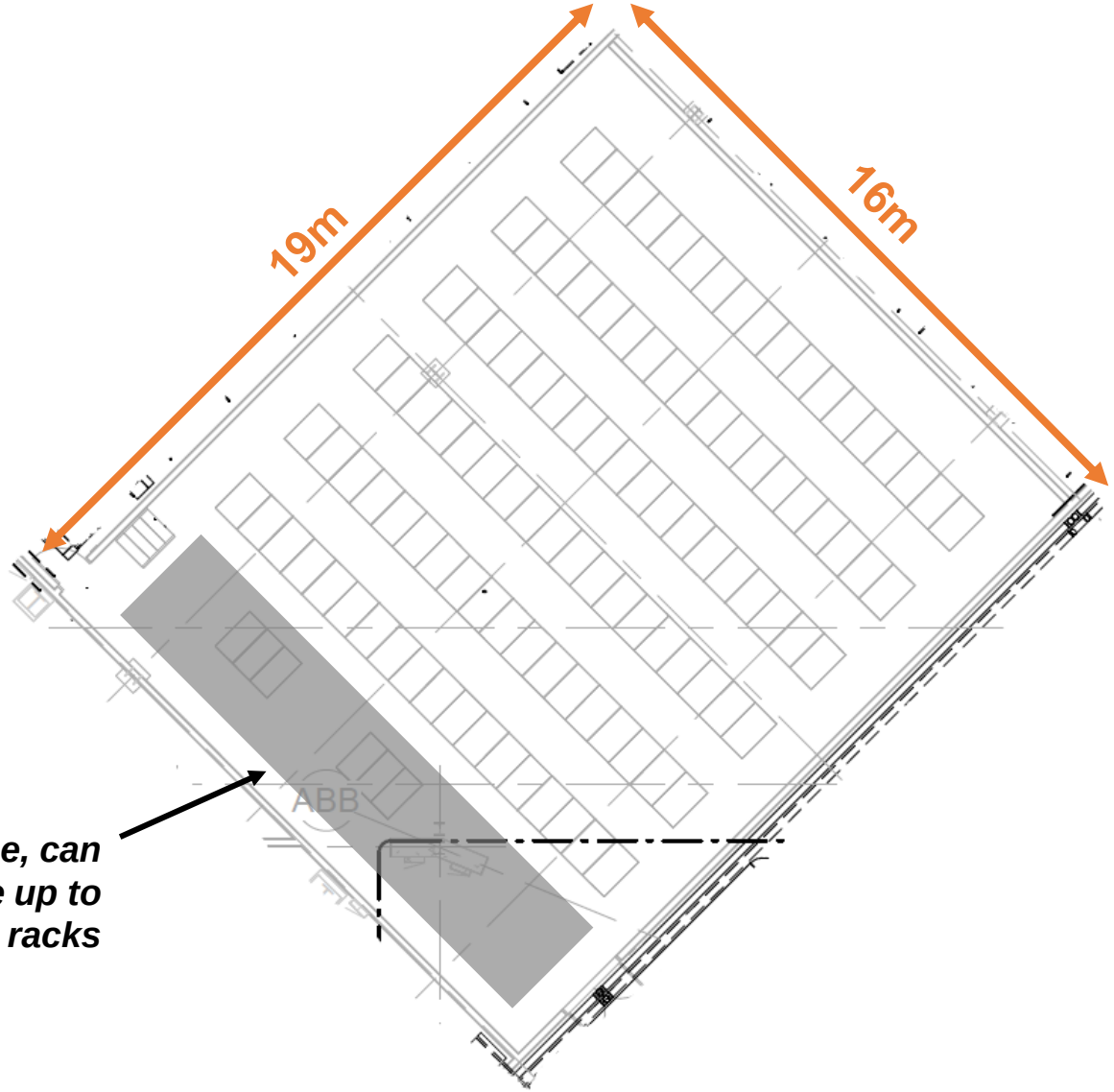
B725 Datacenter: Network Room



BNL Core Facility Revitalization (CFR) Project: Design & Construction of B725 Datacenter

- Main design features of the SDCC B725 datacenter (cont.):
  - Dedicated tape library room to host all B725 robotic tape libraries
    - 90 in of vertical clearance (accommodating two options of SpectraLogic Spectra TFinity and IBM TS4500 modular libraries)
    - Up to 6 individual libraries with 20k slot capacity and up to 96 tape drives each (targeting LTO-8 for the first library deployment in FY21), **120k slots in total**
    - Up to 16 Gbps per tape drive (no FC switches, the HPSS mover systems are to be located in the Main Data Hall in row 8 in FY21-25
    - Unlocking the higher data rates would require putting new HPSS movers in row 17 of the Main Data Hall which would reduce the fiber FC path length from ~ 80m down to ~ 50m
  - Intergen based fire suppression system for the tape room
  - Dedicated 1+1 redundant CRAC group to remove the ambient heat dissipation from the room (located outside the room)
  - 300kW battery UPS (shared with the Network Room) on the bypass for additional protection during maintenance
  - All equipment inside the tape room is air-cooled, no chilled water here
  - 3 level of overhead cable/power distribution: power/busbar, fiber trays with mini-racks attached to the and copper trays with RJ-45 patch panels attached to them optimized for the location of the tape drive and control bays of the libraries

B725 Datacenter: Tape Room



Reserved space, can accommodate up to 6 air-cooled racks

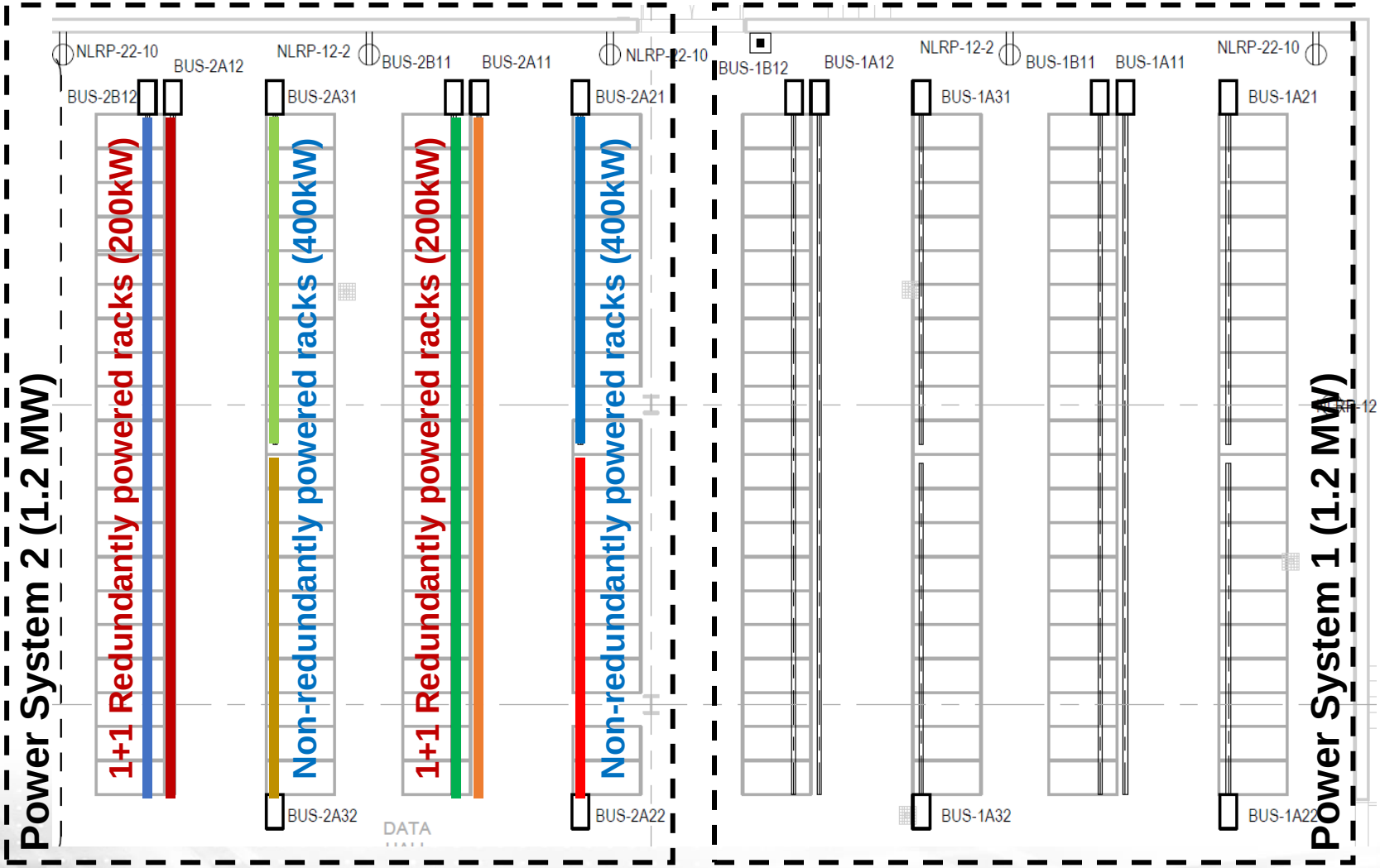
BNL Core Facility Revitalization (CFR) Project: Design & Construction of B725 Datacenter

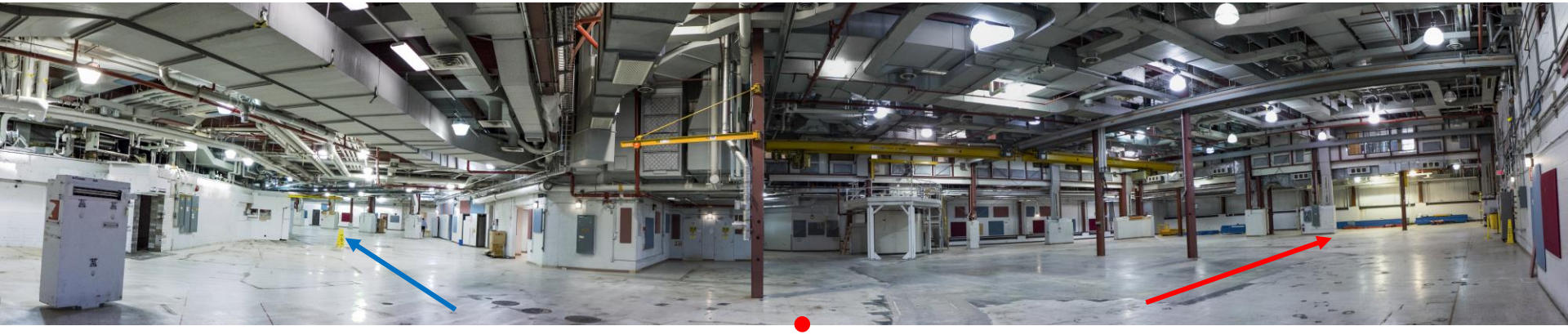
- Main design features of the SDCC B725 datacenter (cont.):
  - The power & cooling infrastructure for the new datacenter is to be provisioned in blocks of 1.2 MW of power feed / heat dissipation capacity, which includes
    - A single Power System consisting of the STS for switching the load to bypass for maintenance, a battery UPS system (up to 20 min runtime at full load), a set of transformers and busbar systems deployed on top of the rows of equipment
    - Chiller unit, cooling tower (on the roof) and the set of associated pumps
    - No CRAC units in the whitespace of the datacenter **– unlike B515**
    - At the minimum we expect to have 3 power systems available starting from FY21 (3.6 MW of total capacity), with an add alternative of getting the 4th power system at the same time (4.8 MW of total capacity)
    - Provisions are expected to be made for adding the remaining 4 or 5 electrical rooms in the future (beyond the scope of CFR project) to reach 9.6 MW capacity
    - Redundantly powered equipment can only draw power from one Power System (necessary cost and power utilization efficiency optimization) – **unlike B515**
  - Each Power system gets its own room inside B725
    - The voltage step down is 13.8kV – (transformer to UPS input) – 480V – (UPS output to the PDU transformers feeding the busbars) 208V (3-phase V)
    - 4 PDUs and 8x 800A busbars per typical Power System (except Power System 3 that has 3x 1200A + 4x 800A + 2x 200A busbars)
    - No PDU units and no 480V in the whitespace of the datacenter **– unlike B515**

BNL Core Facility Revitalization (CFR) Project: Design & Construction of B725 Datacenter

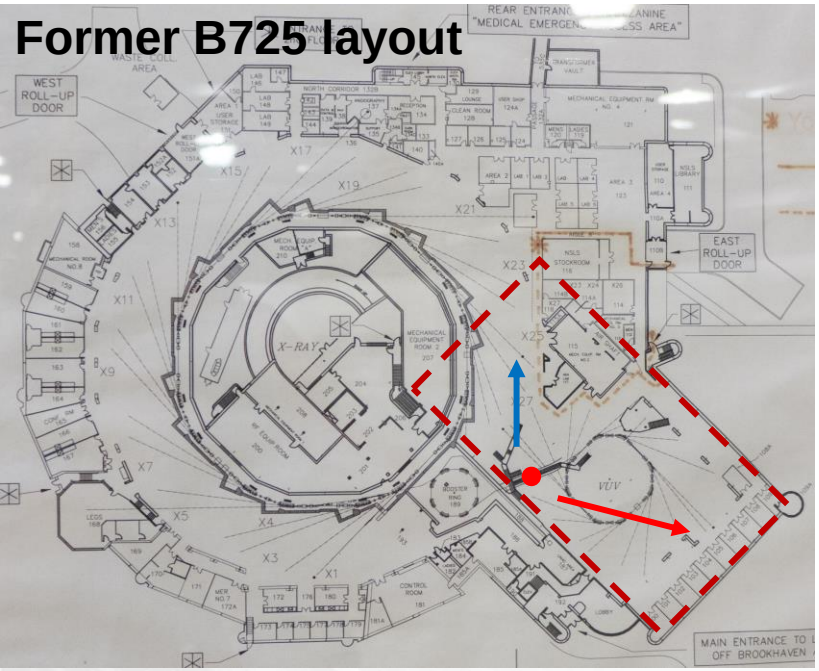
- Main design features of the SDCC B725 datacenter (cont.):
  - Diesel generators come in blocks of 1750kW
    - At the minimum we expect 2 of them available in FY21 which would cover 3.5 MW of backup power; making the 3rd generator available in FY21 is an add alternative (5.3 MW of backup power)
    - The generator complex is built to support 9 generators in total in 3 groups (15.7 MW of backup power) which would allow N+1 redundancy in each group
    - The generator complex is moved ~100m away from the datacenter building **(not the case for B515)**
    - Cooling systems and pumps are all on generator power, so the entire 3.6 MW facility can survive indefinitely on 3 generators, provided that we get 3 generators **(not the case for B515 either)**
    - All generators are live refuelable, with the source of fuel available on site
  - All electrical systems of the new datacenter are provided with a single shared bypass system for maintenance rated at 1.2MW
    - Only one Power System can switch to bypass at a time
    - In the base bid only Power System 3 feeding the Network Room and the Tape Room get partial protection (auxiliary 300kVA Battery UPS), yet there is an add alternative that puts the full size 1.2 MW UPS on the bypass (and removes the auxiliary UPS)
  - Both room level EPO system and the central EPO system to be deployed in B725

B725 Datacenter: Rack Power Allocation Schema (PS#1 & PS#2) Example



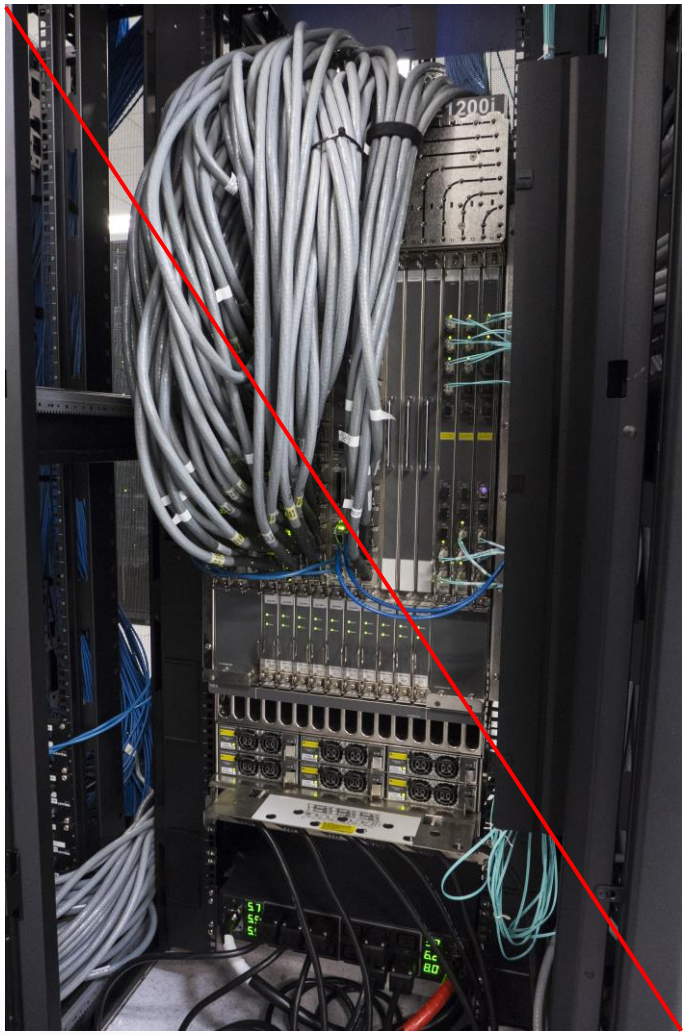


B725/Future Main Data Hall & Power System Areas

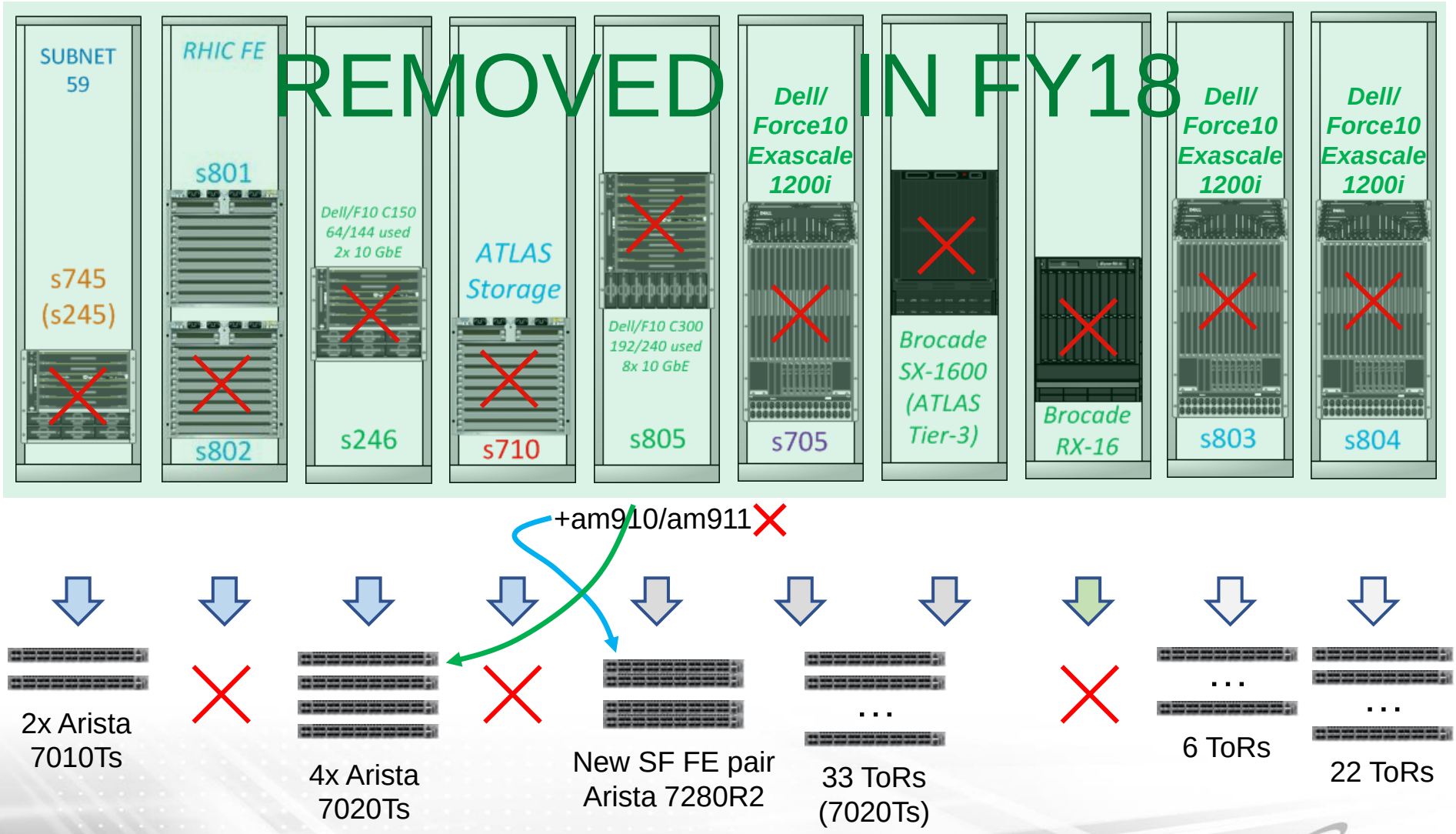


Preparation of B515 Datacenter for the Addition of B725 Datacenter in FY18-20

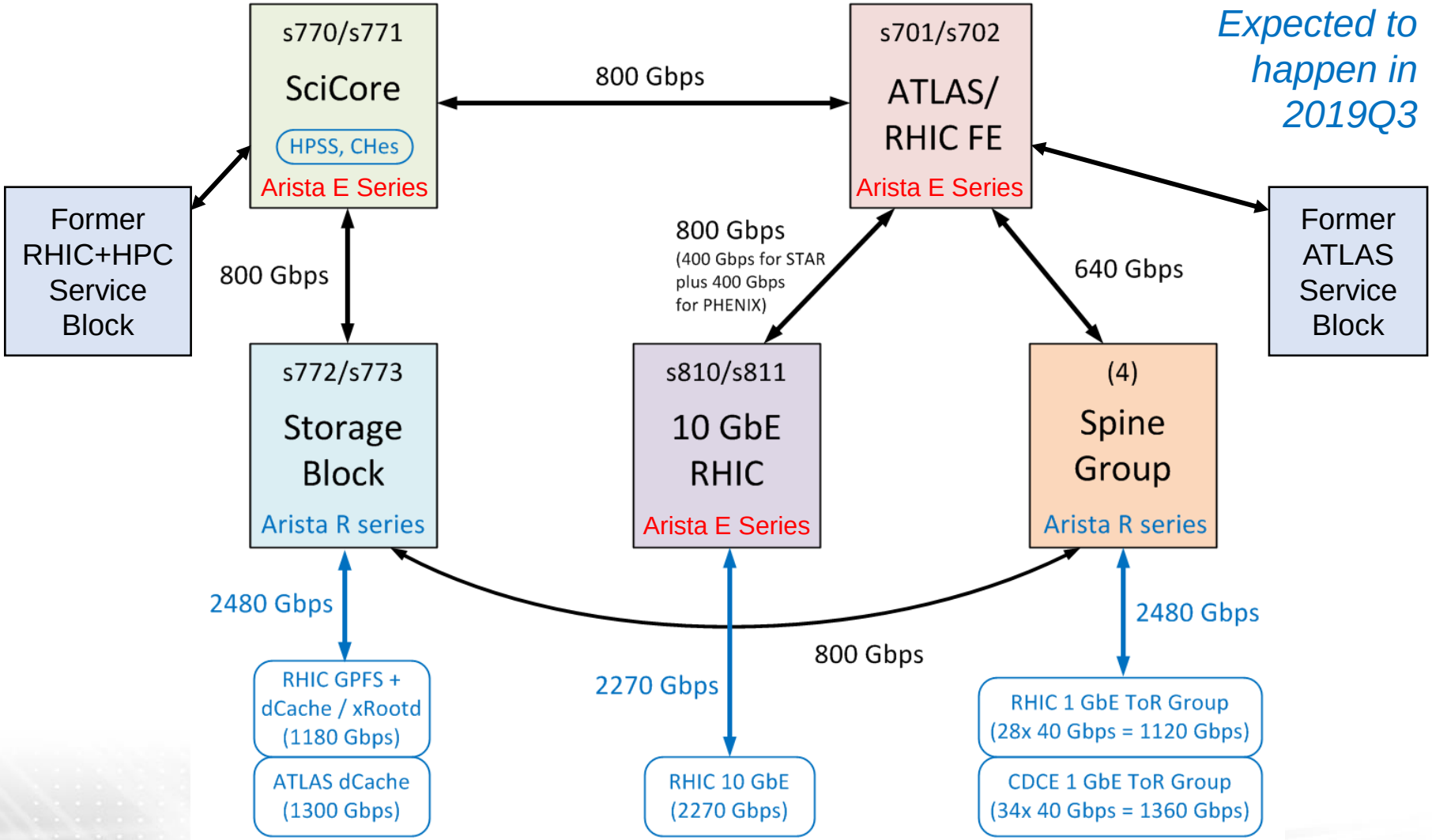
- B515 side network consolidation: retirement of all the legacy central network equipment in FY18 (10 racks worth of w/o taking the passive patch panel racks into account) dropping Brocade and Dell/Force10 equipment and replacing it with Arista equipment
- Dropping the direct copper distribution for 1 GbE attached nodes which was used by the majority of the Farm equipment before FY19 (3x Dell/Force10 Exascale 1200i systems with 6 racks of copper patch panels to break out the HDE connections and **10 miles** of copper cables distributed above the floor in CDCE and under the floor in the main BCF and RCF areas. This simplifies the migration of compute racks from B515 to B725 much easier, as the rack level data uplinks are all fiber
- Rebuilding the HPSS Fibre Channel infrastructure in CDCE paving the way for the existing Oracle SL8500 libraries (9 of them) consolidation into CDCE
- Phasing out Arista 7500E series equipment (where possible) and replacing it with Arista 7500R series equipment



B515 Datacenter: Network Consolidation (1)

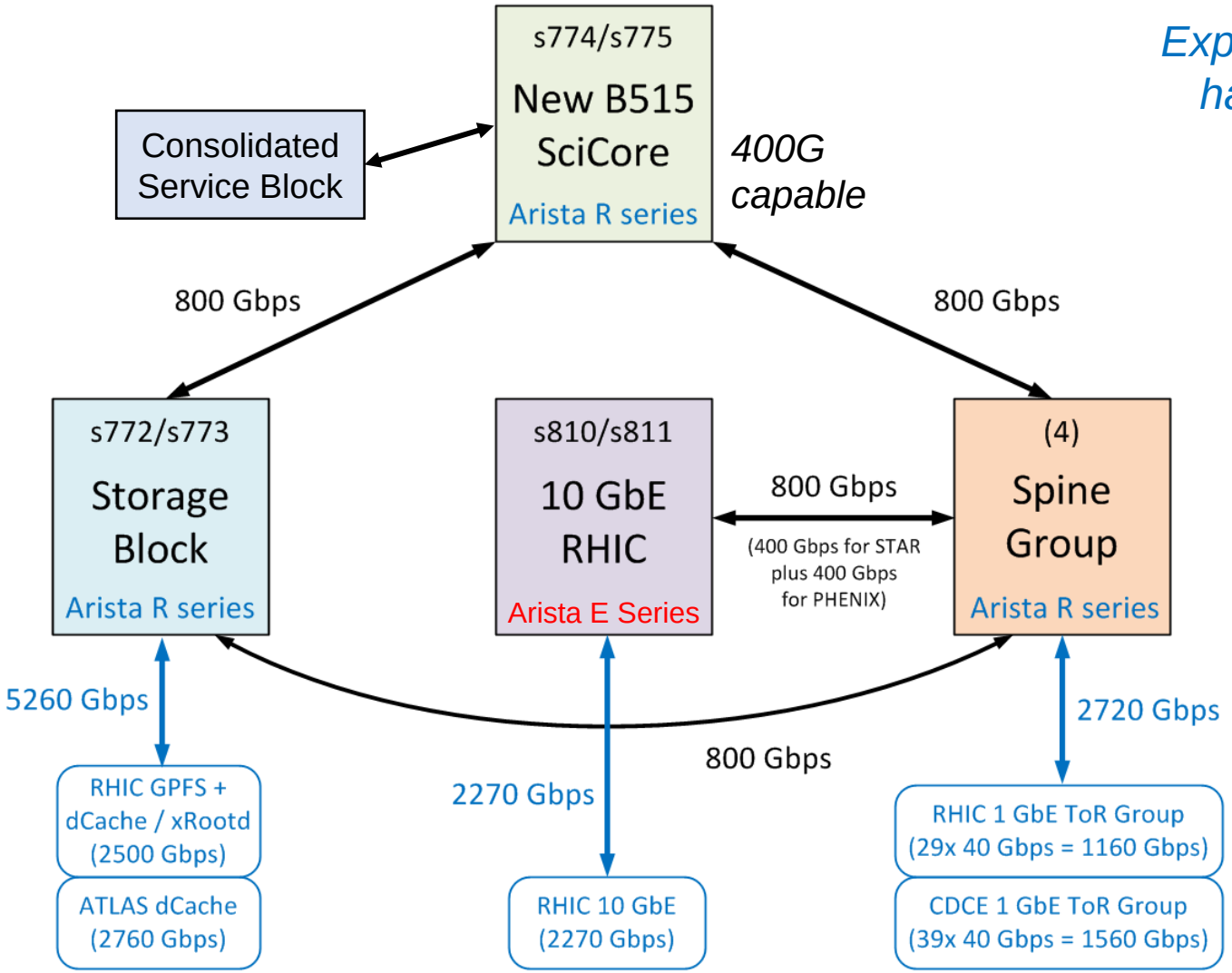


B515 Datacenter: Network Consolidation (2)



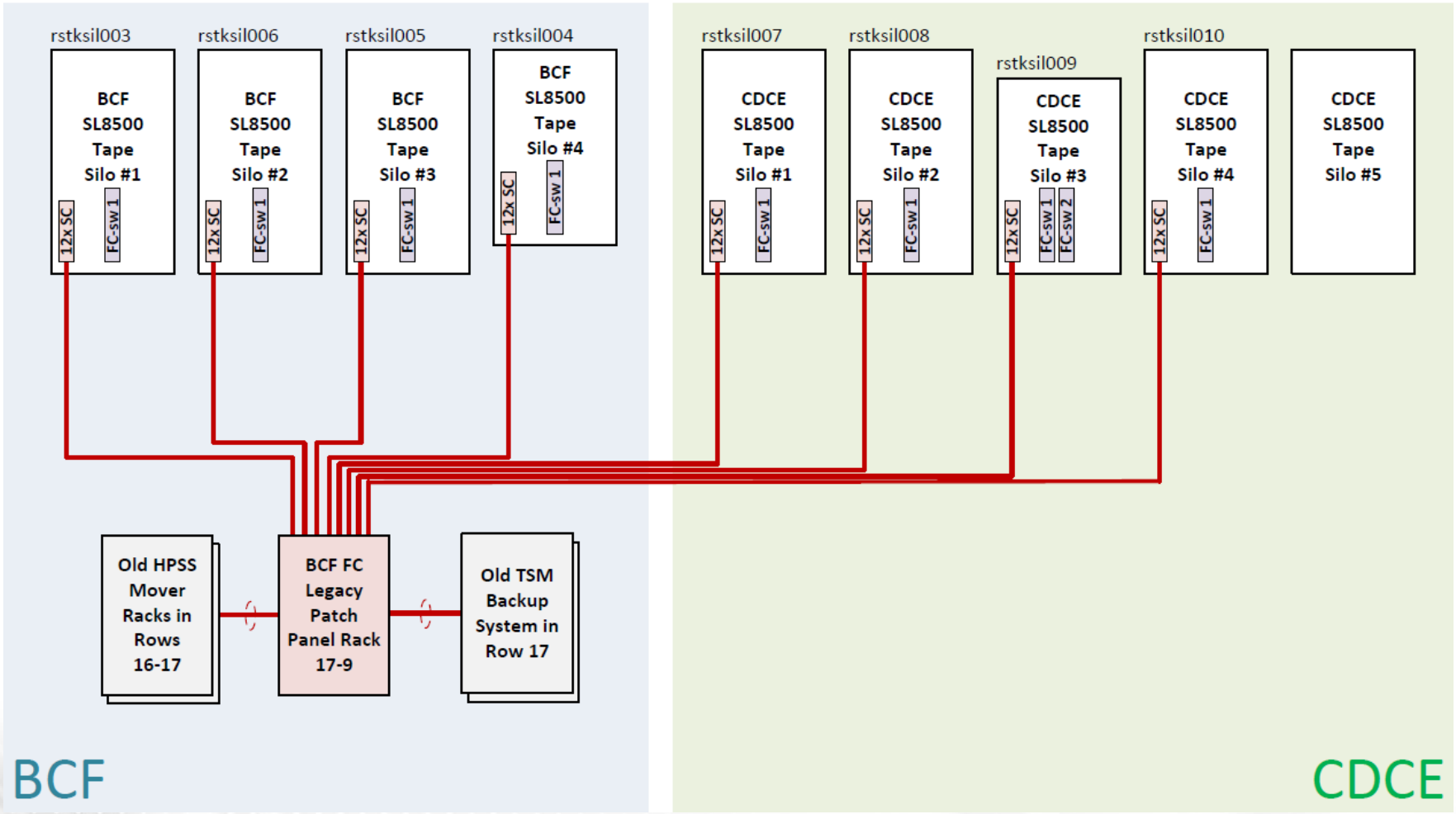
B515 Datacenter: Network Consolidation (3)

Expected to happen in 2019Q3



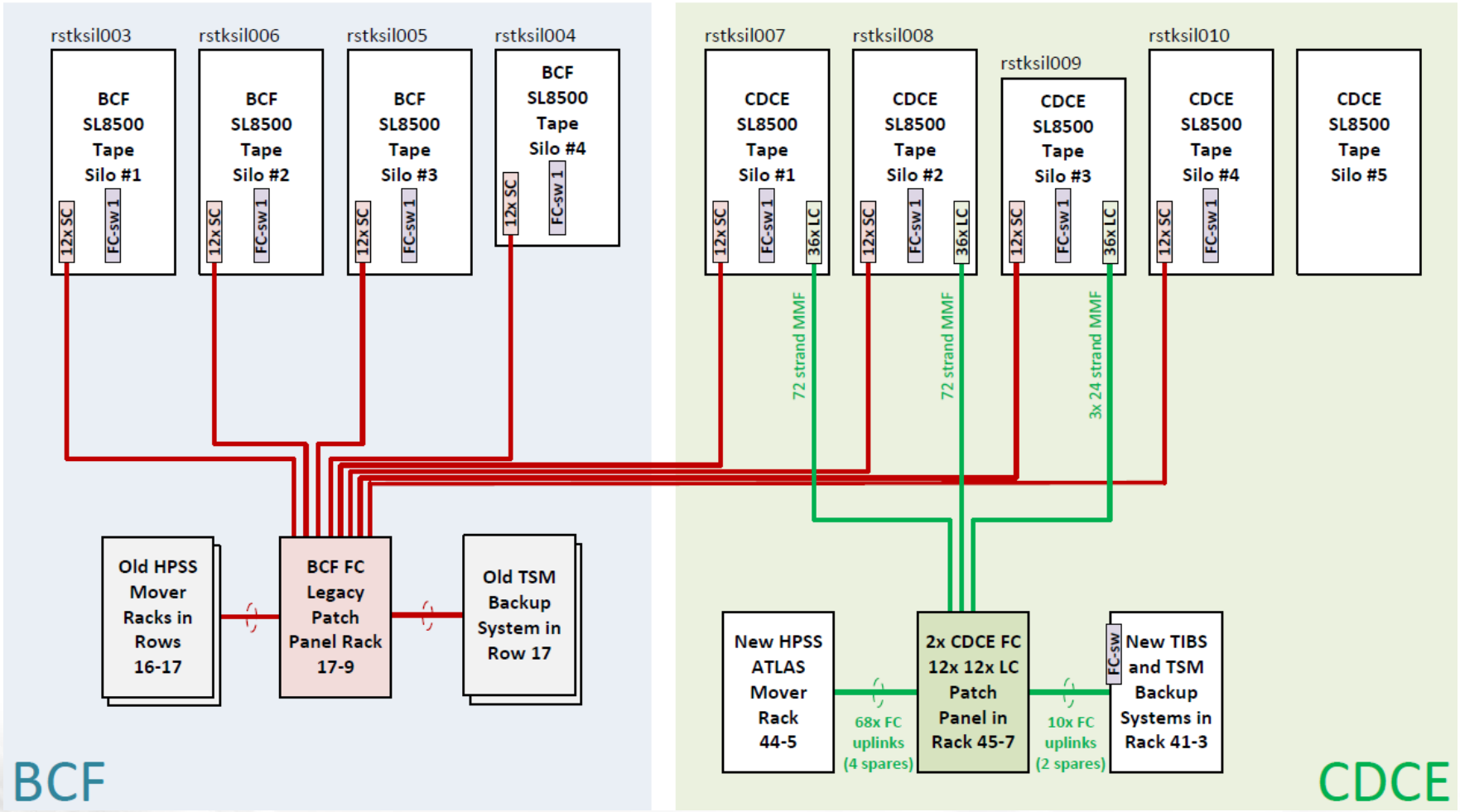
B515 Datacenter: HPSS FC Infrastructure Rebuild (1)

STEP M2
2018Q3



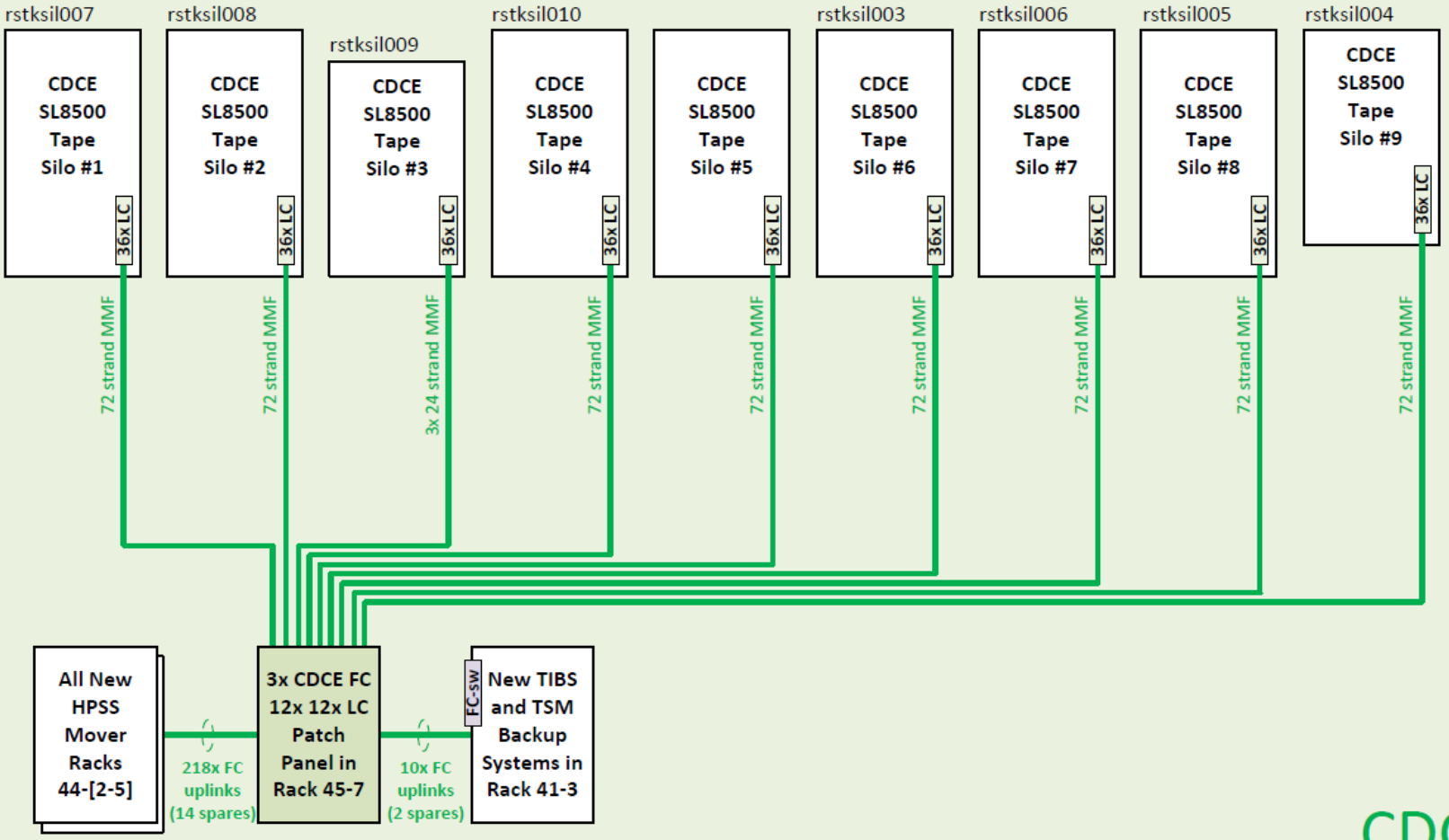
B515 Datacenter: HPSS FC Infrastructure Rebuild (2)

CURRENT

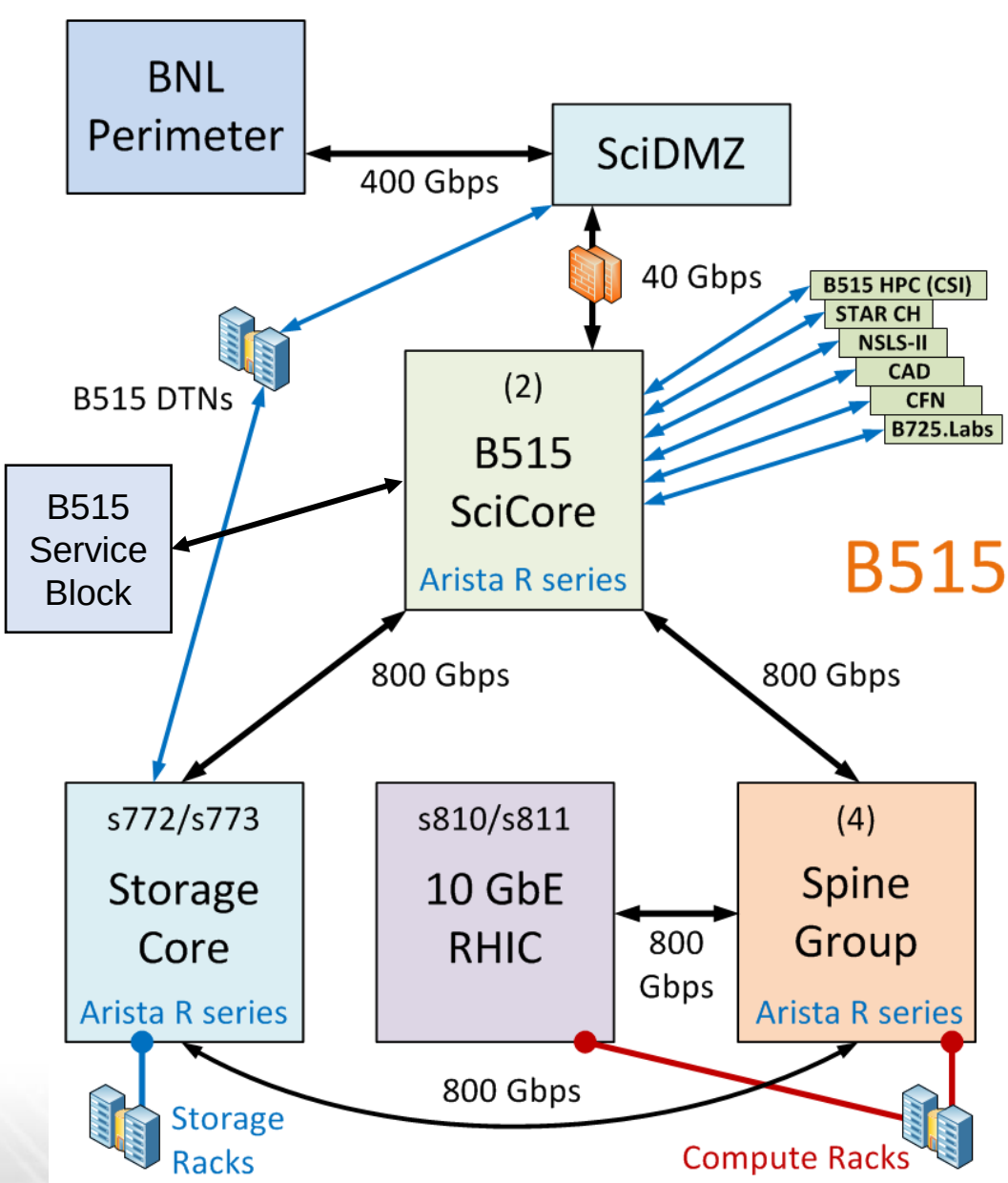


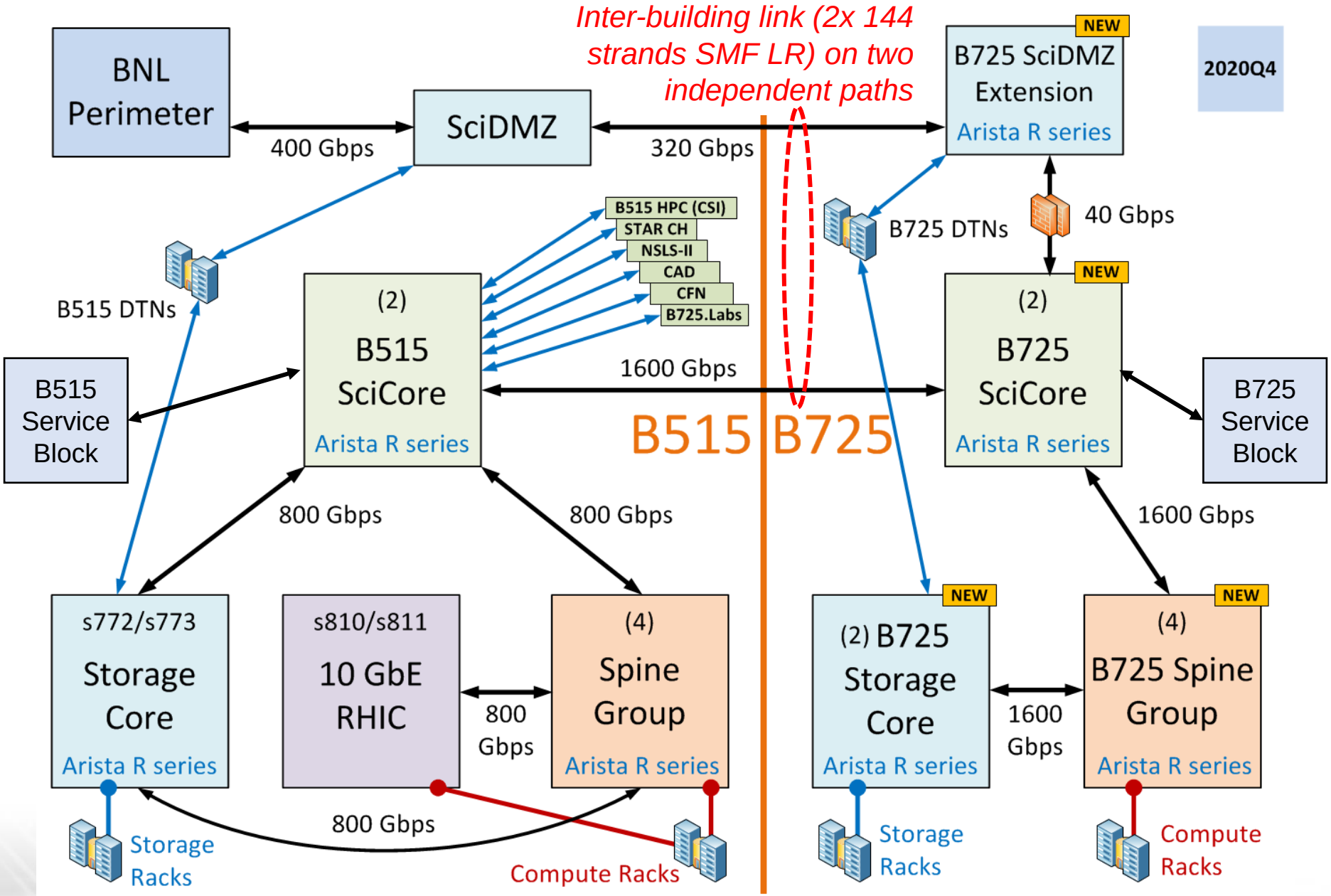
B515 Datacenter: HPSS FC Infrastructure Rebuild (3)

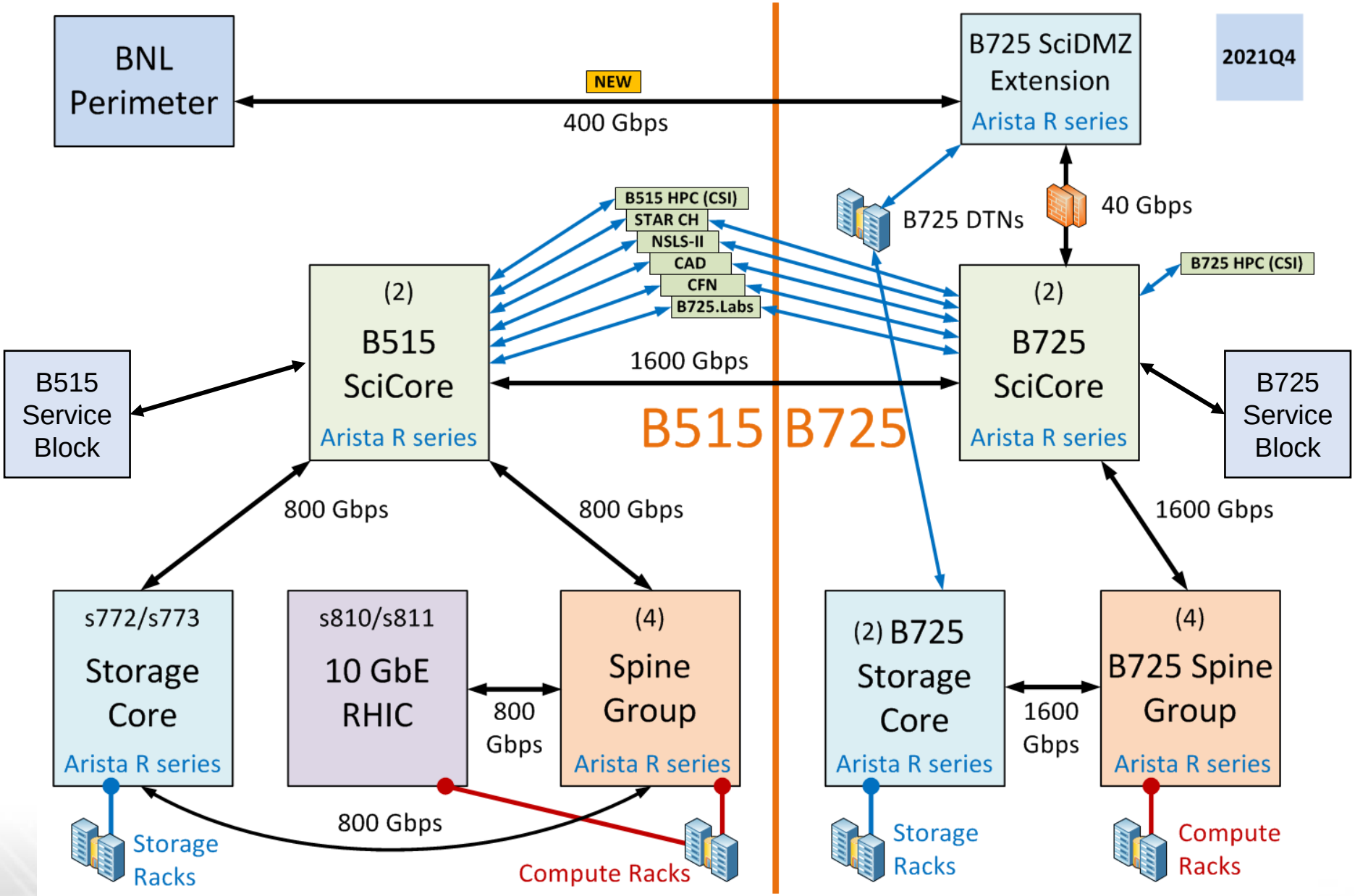
STEP 3
2020Q3

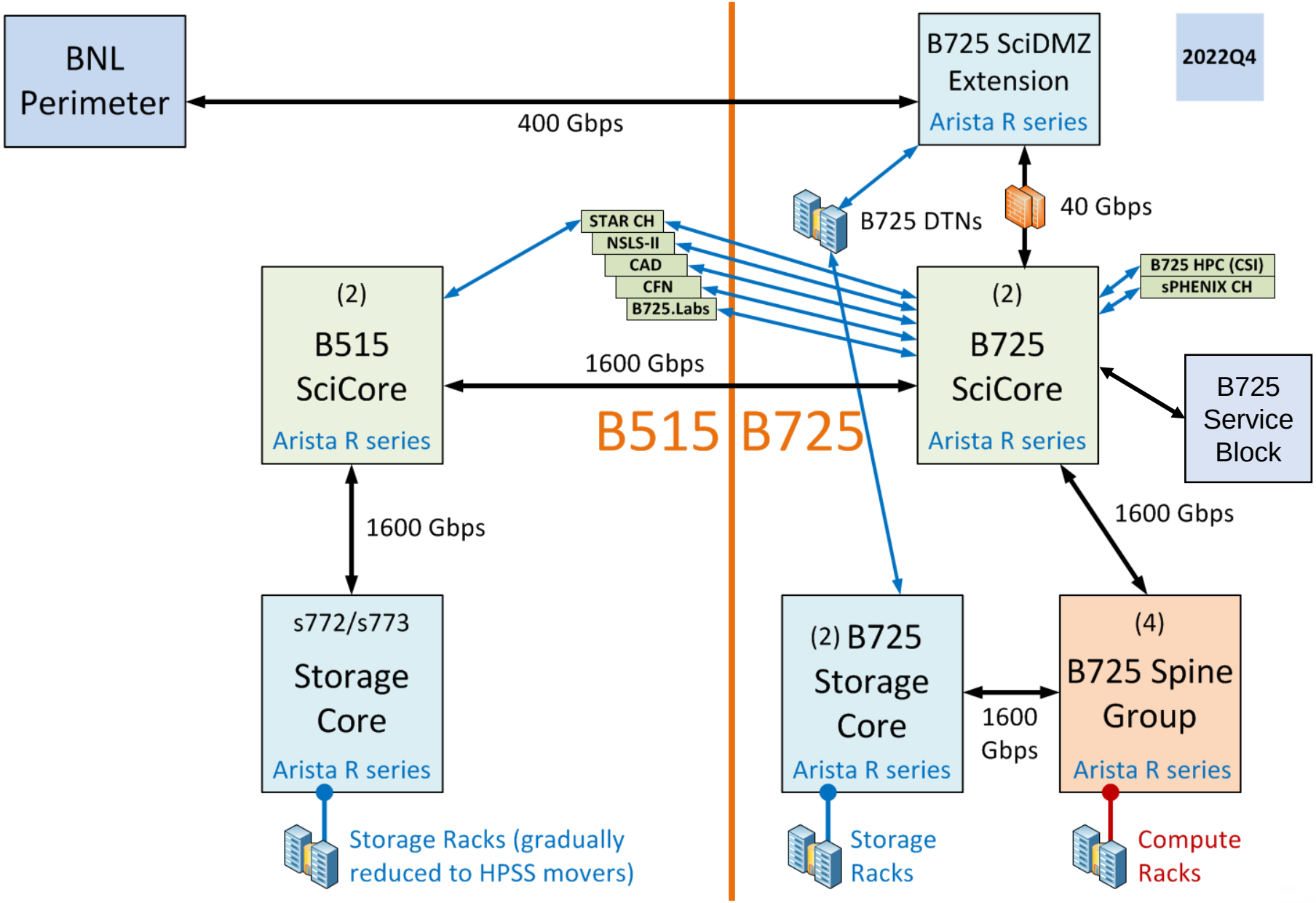


2019Q4


BNL Perimeter
400 Gbps
SciDMZ
40 Gbps
B515 HPC (CSI)
STAR CH
NSLS-II
CAD
CFN
B725.Labs
B515 DTNs
(2)
B515 SciCore
Arista R series
B515 Service Block
B515
800 Gbps
800 Gbps
s772/s773
Storage Core
Arista R series
s810/s811
10 GbE RHIC
800 Gbps
(4)
Spine Group
Arista R series
800 Gbps
Storage Racks
Compute Racks


Inter-building link (2x 144 strands SMF LR) on two independent paths
2020Q4
BNL Perimeter
400 Gbps
SciDMZ
320 Gbps
NEW
B725 SciDMZ Extension
Arista R series
B515 HPC (CSI)
STAR CH
NSLS-II
CAD
CFN
B725.Labs
B515 DTNs
B725 DTNs
40 Gbps
NEW
(2) B515 SciCore
Arista R series
1600 Gbps
(2) B725 SciCore
Arista R series
NEW
B515 Service Block
B725 Service Block
B515 B725
800 Gbps
800 Gbps
1600 Gbps
s772/s773 Storage Core
Arista R series
s810/s811 10 GbE RHIC
800 Gbps
(4) Spine Group
Arista R series
NEW
(2) B725 Storage Core
Arista R series
1600 Gbps
NEW
(4) B725 Spine Group
Arista R series
800 Gbps
Storage Racks
Compute Racks
Storage Racks
Compute Racks


BNL Perimeter
NEW
400 Gbps
B725 SciDMZ Extension
Arista R series
2021Q4
B515 HPC (CSI)
STAR CH
NSLS-II
CAD
CFN
B725.Labs
B725 DTNs
40 Gbps
B725 HPC (CSI)
(2)
B515 SciCore
Arista R series
1600 Gbps
(2)
B725 SciCore
Arista R series
B515 Service Block
B725 Service Block
B515
B725
800 Gbps
800 Gbps
1600 Gbps
s772/s773
Storage Core
Arista R series
s810/s811
10 GbE RHIC
800 Gbps
(4)
Spine Group
Arista R series
(2) B725 Storage Core
Arista R series
1600 Gbps
(4)
B725 Spine Group
Arista R series
800 Gbps
Storage Racks
Compute Racks
Storage Racks
Compute Racks


BNL Perimeter
400 Gbps
B725 SciDMZ Extension
Arista R series
2022Q4
B725 DTNs
40 Gbps
STAR CH
NSLS-II
CAD
CFN
B725.Labs
B725 HPC (CSI)
sPHENIX CH
(2)
B515 SciCore
Arista R series
1600 Gbps
B515 B725
(2)
B725 SciCore
Arista R series
B725 Service Block
1600 Gbps
1600 Gbps
s772/s773
Storage Core
Arista R series
(2) B725 Storage Core
Arista R series
1600 Gbps
(4)
B725 Spine Group
Arista R series
Storage Racks (gradually reduced to HPSS movers)
Storage Racks
Compute Racks

Transition into B515 & B725 Datacenter Operations Period in early FY21 (1)

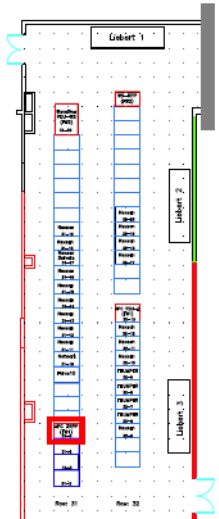
- Compute resources (HTC):
  - All of the 1 GbE connected Farm racks in B515 are ToR switch based as of now.
  - All new compute node racks are placed in CDCE starting from FY18.
  - The compute node racks that still have at least 2 year of vendor supported life as of FY21 are going to be moved to B515 and connected to B725 Spine Group (all such racks are going to be located in CDCE; the number of racks expected to move is 16-18 depending on the results of the FY19-20 compute node purchases for ATLAS, RHIC and Belle II).
  - The 10 GbE connected part of the RHIC Farm is frozen in its state as of FY18 and not going to be extended anymore and will be retired as it goes off support in B515 along with it Arista 7508E switch pair, and will be gradually replaced by the new ToR switch connected 1 GbE attached compute nodes in B725.
  - 1 GbE connected part of the RHIC and ATLAS Farm that is to remain in B515 as of FY21 is going to be gradually retired and replaced with new compute node racks in B725 in FY21-23 period.
  - No compute node racks are expected to be hosted in B515 after FY23.
  - SF compute node racks located in B515 as of FY21 are expected to be either retired or moved to B725 by FY23.
- Compute resources (HPC)
  - The compute racks of the HPC clusters located in the BGL room are expected to be retired or moved to B725 by FY23 (they are likely to be replaced by the new HPC clusters to be built in B725).

Transition into B515 & B725 Datacenter Operations Period in early FY21 (2)

- Storage (dCache/xRootD/Ceph):
  - The purchase of BNL ATLAS dCache storage in FY20 is intentionally suppressed, such that the large purchase of JBODs for ATLAS made in FY21 can be deployed directly in B725. The purchase of JBODs for ATLAS in FY19 was scaled up in order to meet the pledge of FY19 and FY20.
  - We are trying to avoid moving storage racks between B515 and B725 at all, thus preferring the storage backends to be gradually replaced by new equipment deployed directly in B725. This goal is expected to be reached by FY23.
  - The retirement of the existing BNL ATLAS dCache storage in FY19 was planned in such a way that the new ATLAS Ceph cluster with 10 PB of usable space (erasure code 12+3+1 based) can be built using the best portion of this retired storage (all HWRAID based). The main purpose of this cluster is to serve as storage backed for the ATLAS dCache and it going to remain on B515 side (part in CDCE and part in Sigma-7) until FY23. This cluster can scale up in storage capacity in FY21 using the new set of retired ATLAS dCache storage made available in Sigma-7.
- Storage (GPFS/Central NFS)
  - New RHIC GPFS installations is expected to be deployed in B725 in FY23 (likely to be sPHENIX related); the existing RHIC GPFS installations in CDCE are to be retired by FY24.
  - New central NFS appliance is to be deployed in B725 in FY23, and the existing one in CDCE is to be retired shortly after that.

Transition into B515 & B725 Datacenter Operations Period in early FY21 (3)

- Virtualization platforms (RHEV & Farm VMMS):
  - RHEV instance is to be replaced with the new installation in B725 in FY21 already. The retired RHEV installation may still be used for supporting test VMs until FY22.
  - Farm VMMS system is just getting deployed in FY19, so it is likely to survive in B515 until FY24 and to be replaced along with the Farm infrastructure components by the new servers deployed in the infrastructure racks of B725 in FY24.
- HPSS
  - One 18-frame ATLAS tape library (IBM TS4500 or SpectraLogic Tfinity, both with ~20k tape slots per library) and one new HPSS mover rack is to be deployed in B725 in early FY21.
  - The HPSS installation in B725 is expected to grow up to 6 libraries (576 tape drives) and 9 mover racks by FY25 (depending on sPHENIX and ATLAS plans).
  - The exiting 9x Oracle SL8500 libraries are to be consolidated to CDCE on the B515 side by the end of FY20; provided with 4 racks of new movers, out of which 2 are already deployed.
  - The HPSS presence in main BCF and RCF rooms is to be eliminated in FY21 (all legacy mover and disk cache systems, including ATLAS dCache HPSS read cache are to be retired).
- Infrastructure racks
  - 8 infrastructure racks distributed across 2 Power Systems are to be available in B725 starting from FY21. All B515 infrastructure racks are to be removed by FY25.

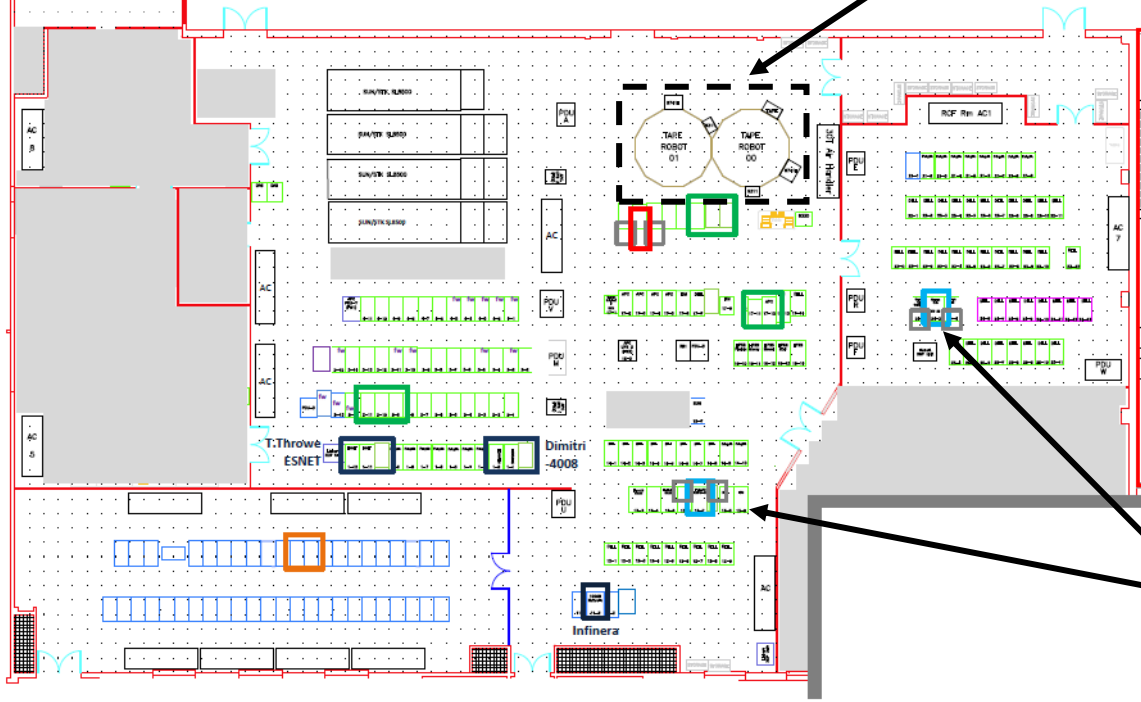


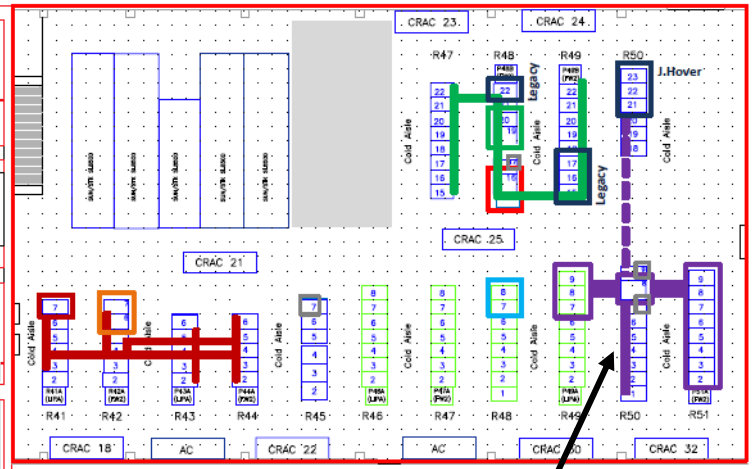
Legend

- Rack is absent
- Rack is empty
- Infrastructure rack
- Critical nodes in the Farm & isolated racks
- Rack is empty, yet power needs to be provisioned
- Rack being retired
- Rack being commissioned

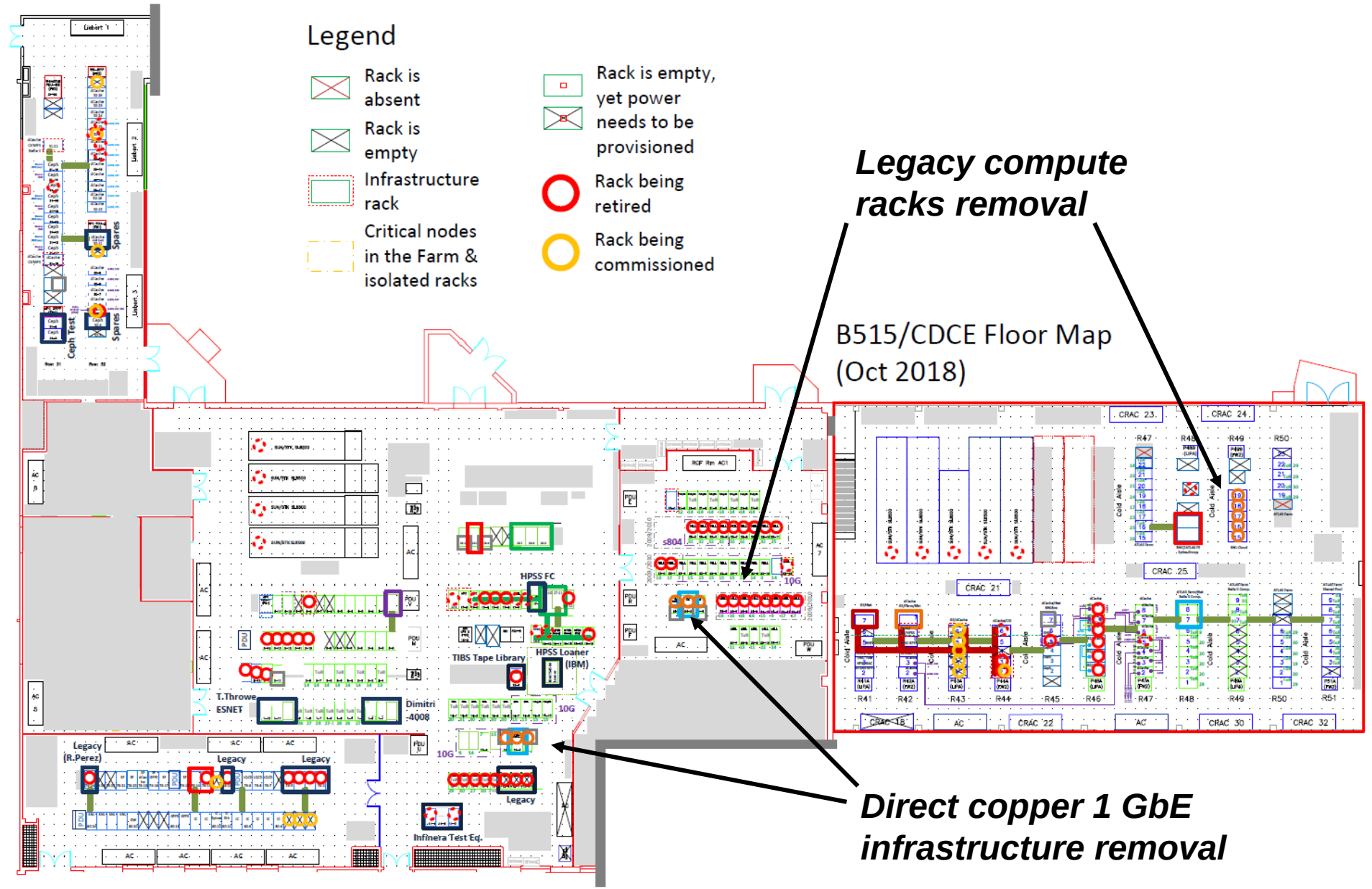
Round silos removal

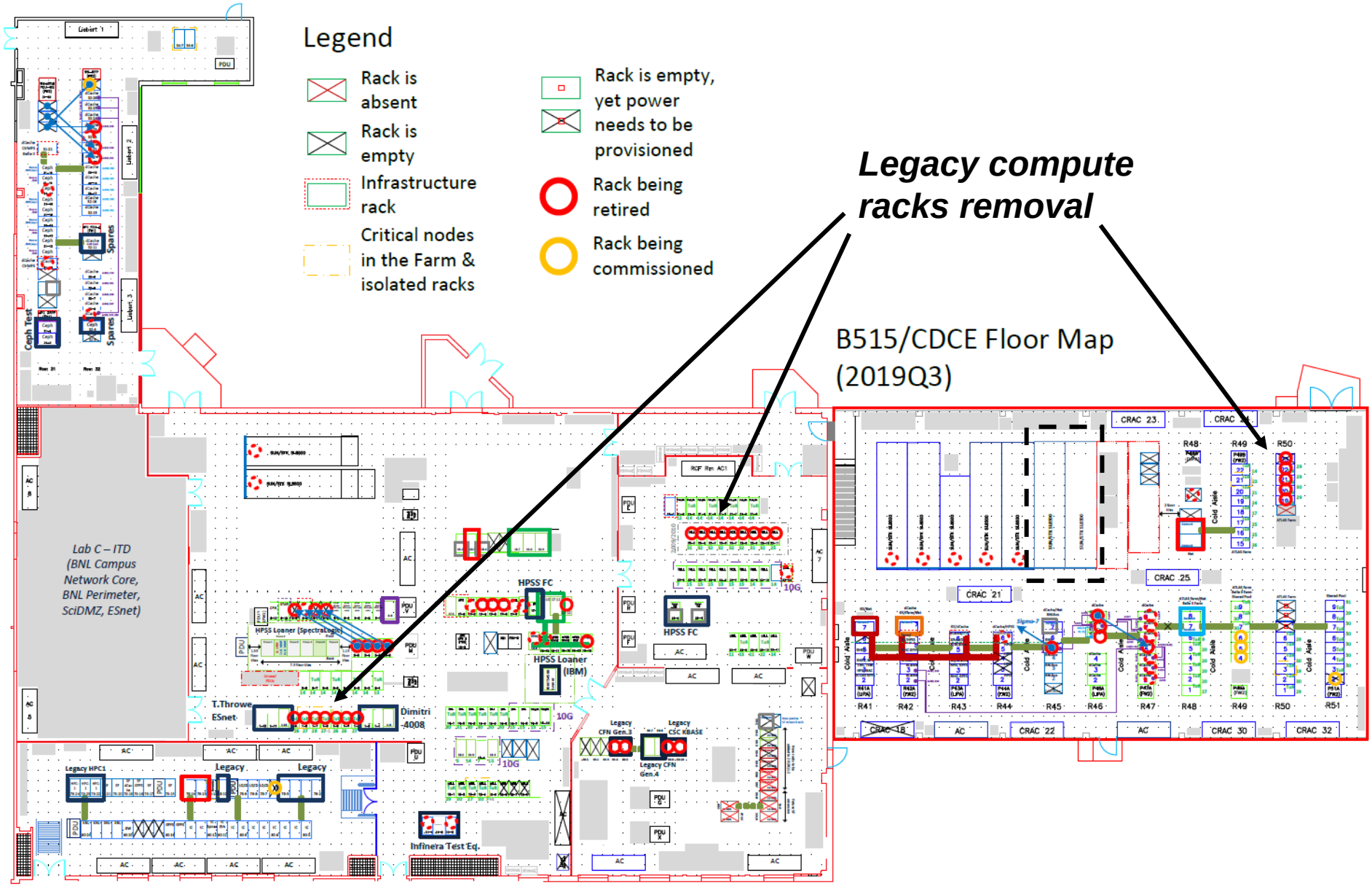
B515/CDCE Floor Map (Dec 2017)

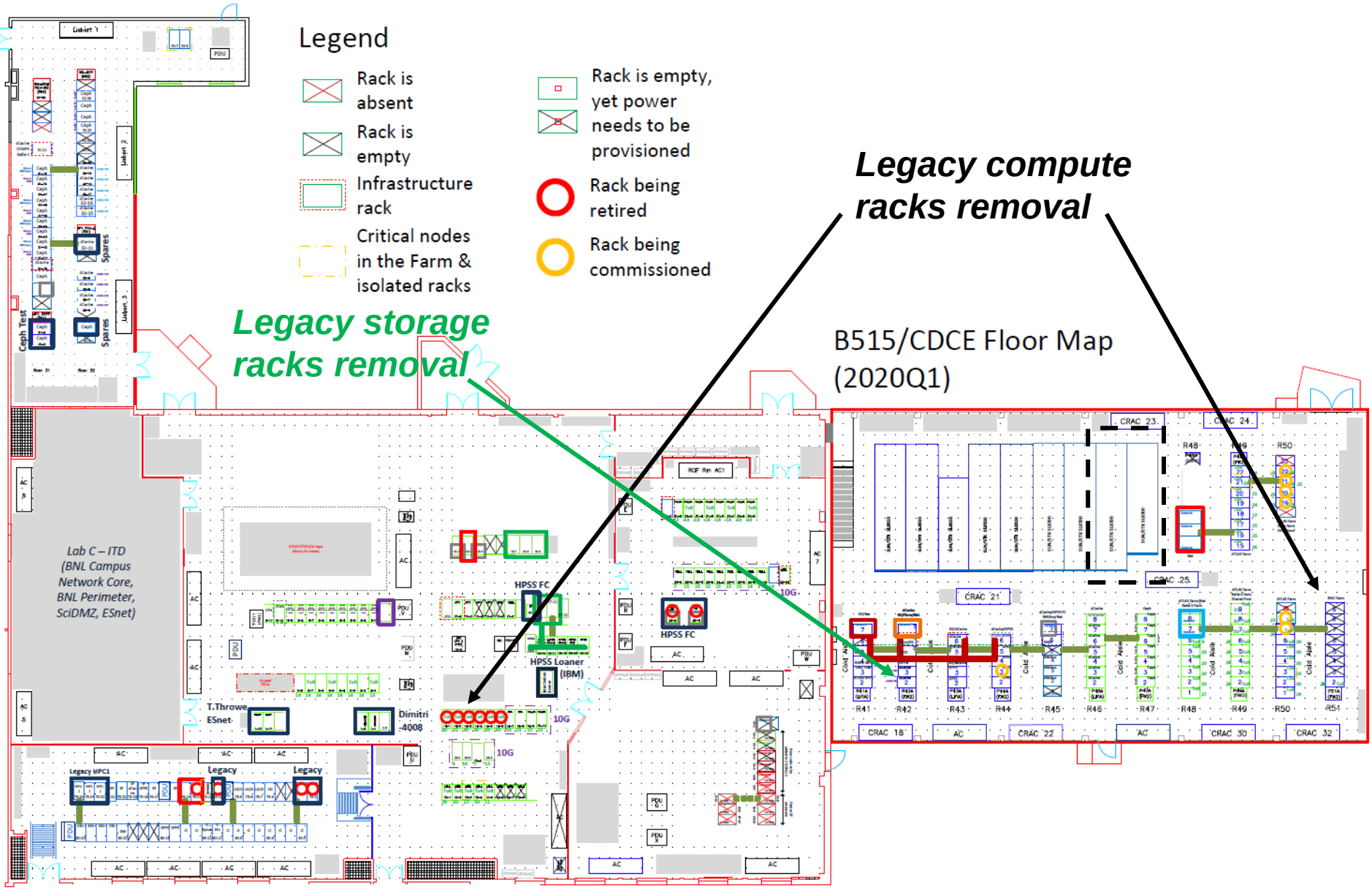


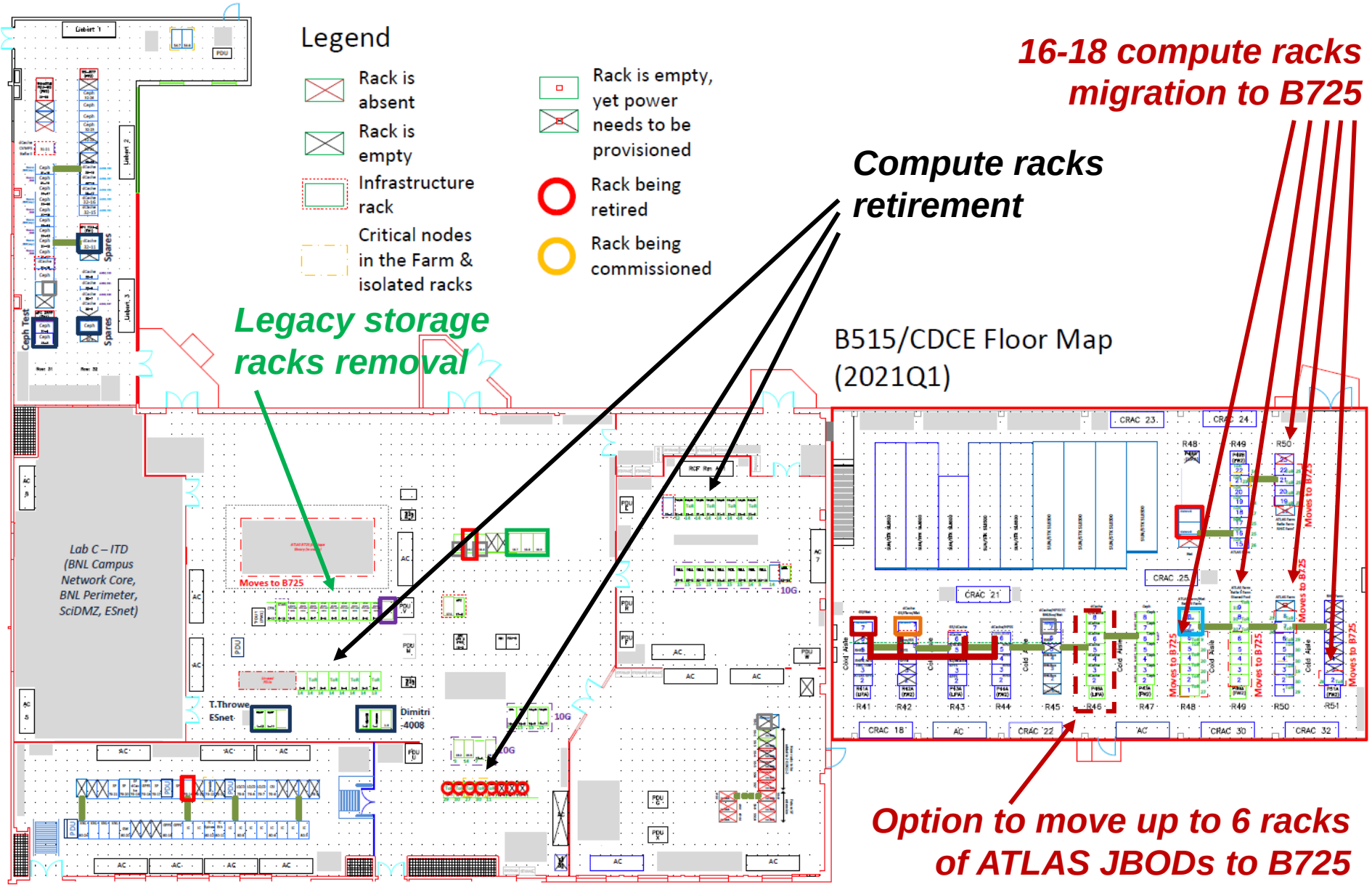


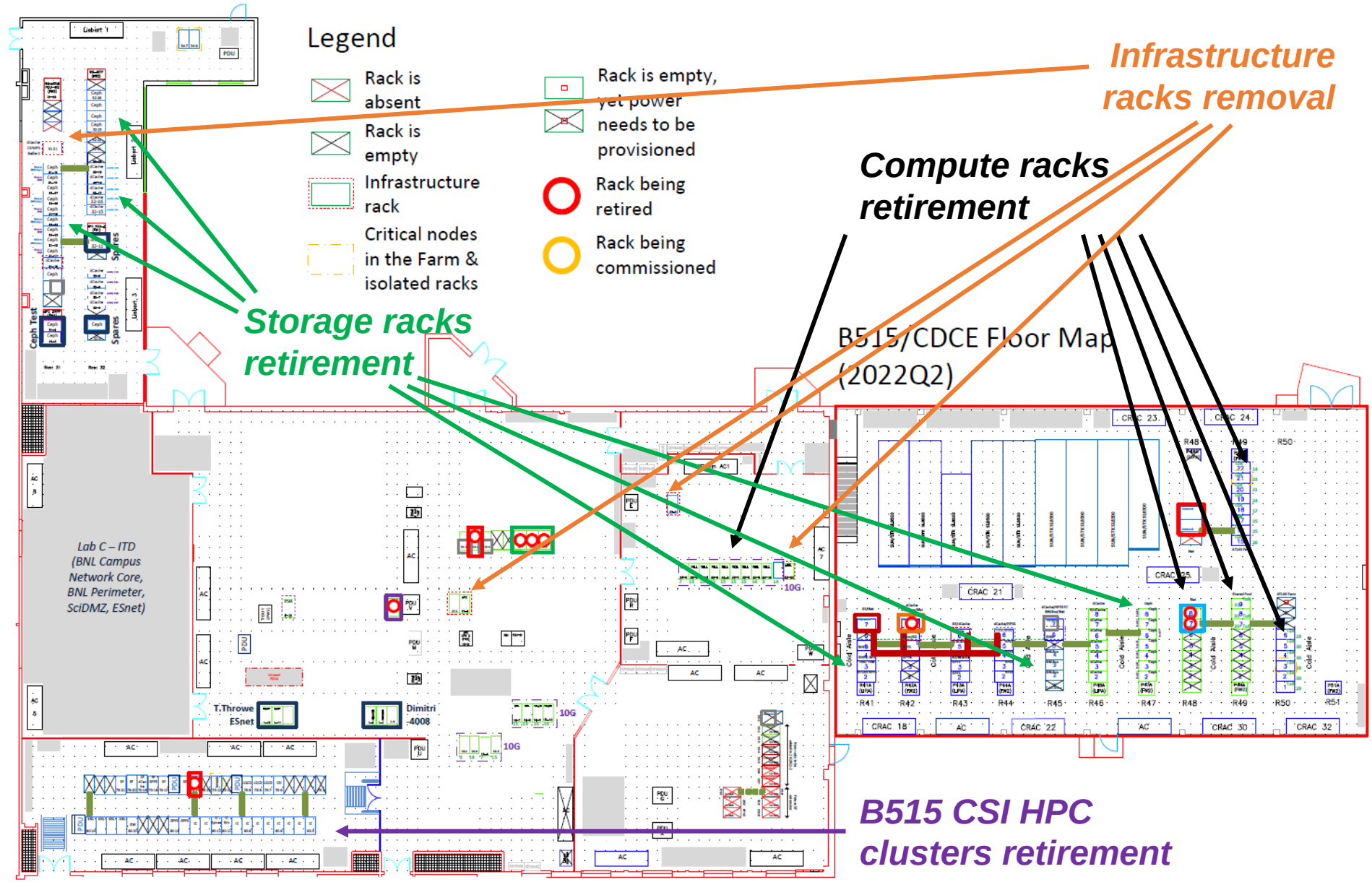
Direct copper 1 GbE infrastructure removal

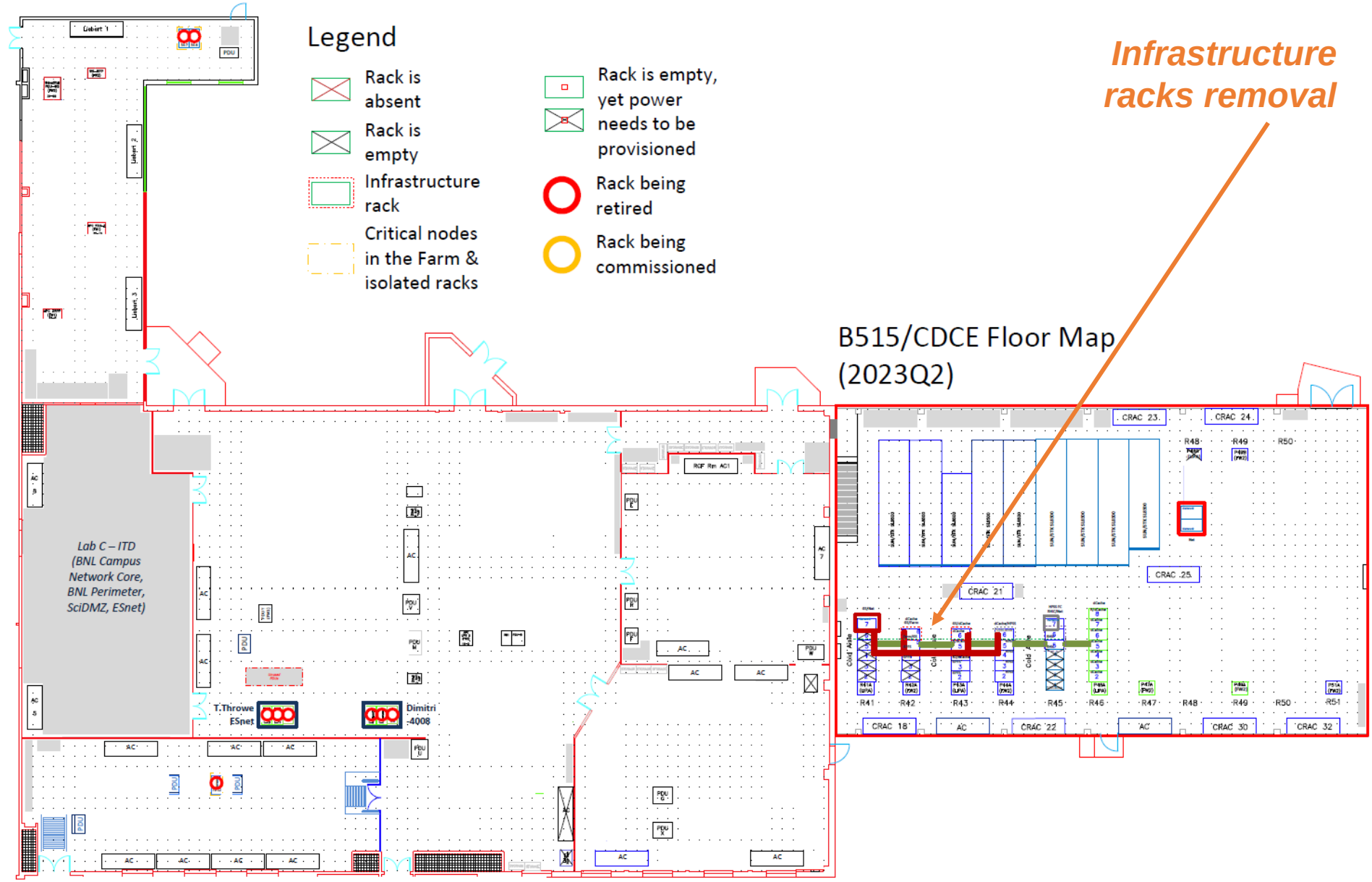

Legend
Rack is absent
Rack is empty
Infrastructure rack
Critical nodes in the Farm & isolated racks
Rack is empty, yet power needs to be provisioned
Rack being retired
Rack being commissioned
Legacy compute racks removal
B515/CDCE Floor Map (Oct 2018)
Direct copper 1 GbE infrastructure removal
HPSS FC
TIBS Tape Library
HPSS Loaner (IBM)
T.Throwe ESNET
Dimitri 4008
Legacy (R.Perez)
Legacy
Legacy
Legacy
Infinera Test Eq.
Ceph Test
Spares
Spares
CRAC 23
CRAC 24
CRAC 25
CRAC 21
CRAC 18
CRAC 22
CRAC 30
CRAC 32

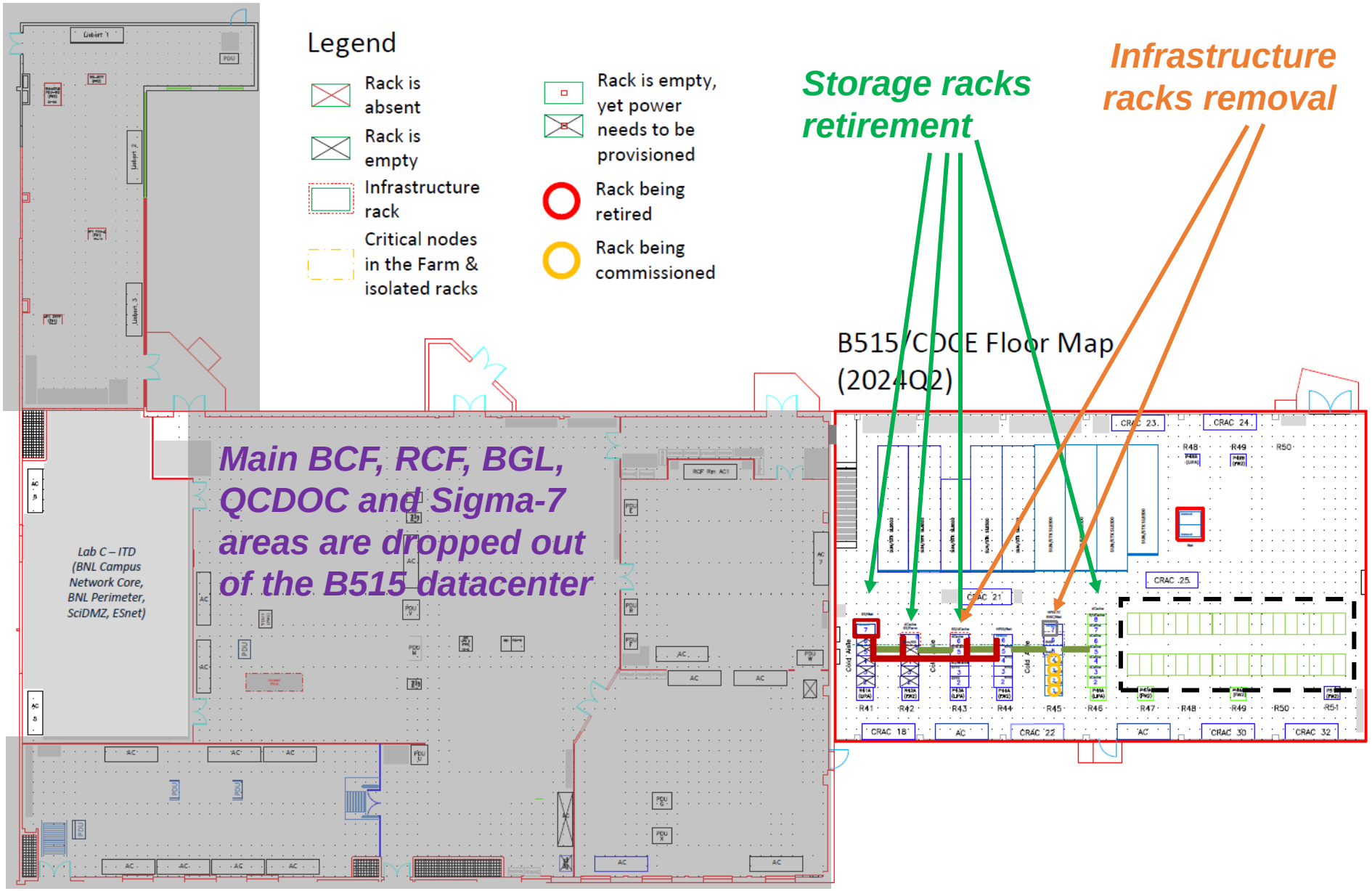

Legend
Rack is absent
Rack is empty
Infrastructure rack
Critical nodes in the Farm & isolated racks
Rack is empty, yet power needs to be provisioned
Rack being retired
Rack being commissioned
Legacy compute racks removal
B515/CDCE Floor Map (2019Q3)
Lab C – ITD (BNL Campus Network Core, BNL Perimeter, SciDMZ, ESnet)
HPSS FC
HPSS Loaner (SpectraLogic)
HPSS Loaner (IBM)
T.Throwe ESnet
Dimitri 4008
Legacy HPC1
Legacy
Legacy
Legacy CFN Gen.3
Legacy CSC KBASE
Legacy CFN Gen.4
Infinera Test Eq.
10G
CRAC 23
CRAC 24
CRAC 21
CRAC 25
CRAC 18
CRAC 22
CRAC 30
CRAC 32
R41
R42
R43
R44
R45
R46
R47
R48
R49
R50
R51


Legend
Rack is absent
Rack is empty
Infrastructure rack
Critical nodes in the Farm & isolated racks
Rack is empty, yet power needs to be provisioned
Rack being retired
Rack being commissioned
Legacy storage racks removal
Legacy compute racks removal
B515/CDCE Floor Map (2020Q1)
Lab C – ITD (BNL Campus Network Core, BNL Perimeter, SciDMZ, ESnet)
HPSS FC
HPSS Loaner (IBM)
T.Thrower ESnet
Dimitri 4008
Legacy
Legacy HPC1
10G
CRAC 18
CRAC 21
CRAC 22
CRAC 23
CRAC 24
CRAC 25
CRAC 30
CRAC 32


Legend
Rack is absent
Rack is empty
Infrastructure rack
Critical nodes in the Farm & isolated racks
Rack is empty, yet power needs to be provisioned
Rack being retired
Rack being commissioned
16-18 compute racks migration to B725
Compute racks retirement
Legacy storage racks removal
B515/CDCE Floor Map (2021Q1)
Lab C – ITD (BNL Campus Network Core, BNL Perimeter, SciDMZ, ESnet)
Moves to B725
T.Throwe ESnet
Dimitri -4008
10G
CRAC 21
CRAC 23
CRAC 24
CRAC 25
CRAC 18
CRAC 22
CRAC 30
CRAC 32
Option to move up to 6 racks of ATLAS JBODs to B725


Legend
Rack is absent
Rack is empty
Infrastructure rack
Critical nodes in the Farm & isolated racks
Rack is empty, yet power needs to be provisioned
Rack being retired
Rack being commissioned
Infrastructure racks removal
Compute racks retirement
Storage racks retirement
B515/CDCE Floor Map (2022Q2)
Lab C – ITD (BNL Campus Network Core, BNL Perimeter, SciDMZ, ESnet)
B515 CSI HPC clusters retirement


Legend
Rack is absent
Rack is empty
Infrastructure rack
Critical nodes in the Farm & isolated racks
Rack is empty, yet power needs to be provisioned
Rack being retired
Rack being commissioned
Infrastructure racks removal
B515/CDCE Floor Map (2023Q2)
Lab C – ITD (BNL Campus Network Core, BNL Perimeter, SciDMZ, ESnet)
T.Throwe ESnet
Dimitri 4008
CRAC 23
CRAC 24
CRAC 25
CRAC 21
CRAC 18
CRAC 22
CRAC 30
CRAC 32

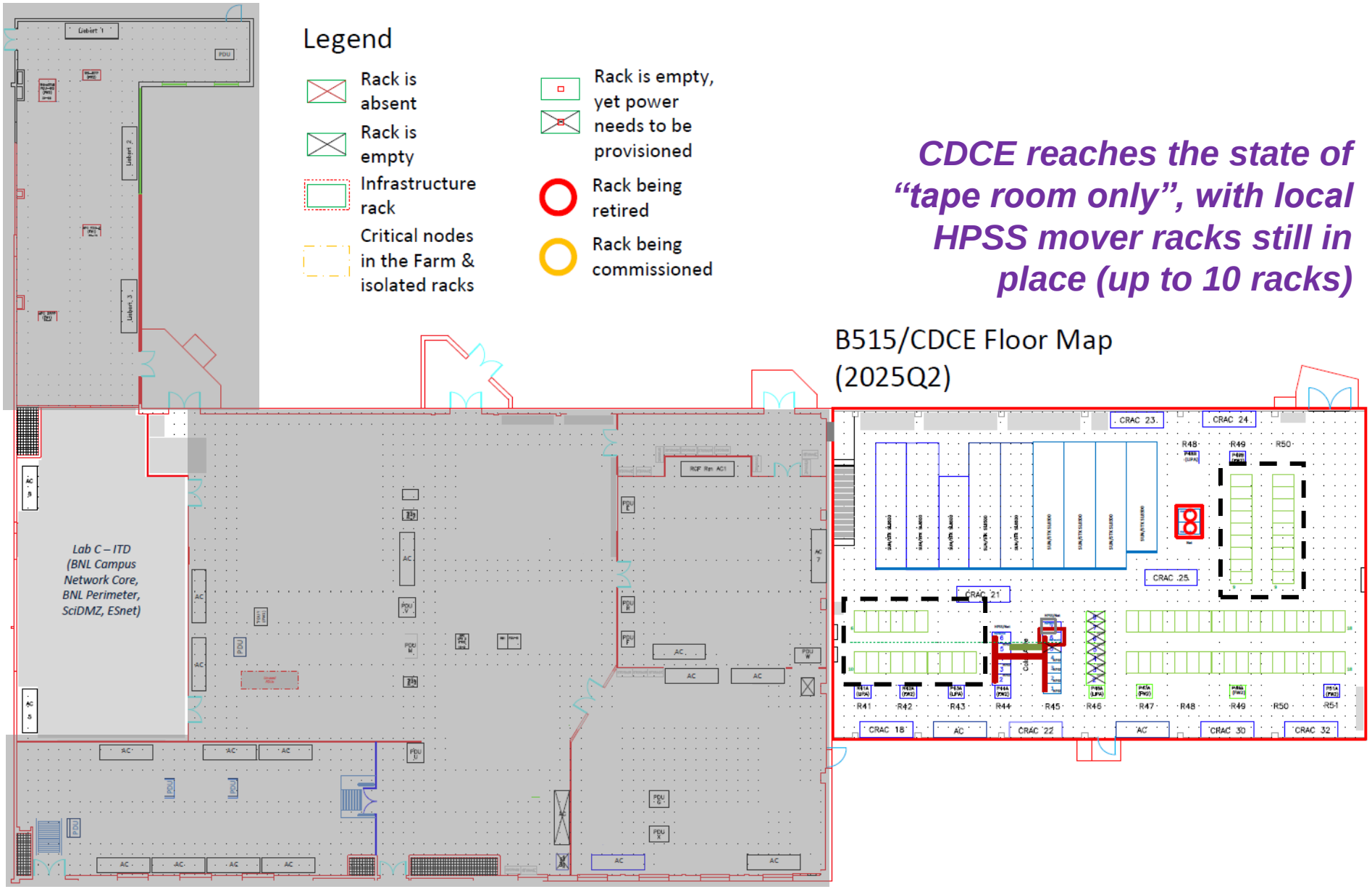

Legend
Rack is absent
Rack is empty
Infrastructure rack
Critical nodes in the Farm & isolated racks
Rack is empty, yet power needs to be provisioned
Rack being retired
Rack being commissioned
Storage racks retirement
Infrastructure racks removal
B515/CDCE Floor Map (2024Q2)
Main BCF, RCF, BGL, QCDOC and Sigma-7 areas are dropped out of the B515 datacenter
Lab C – ITD (BNL Campus Network Core, BNL Perimeter, SciDMZ, ESnet)
CRAC 23
CRAC 24
CRAC 25
CRAC 21
CRAC 18
CRAC 22
CRAC 30
CRAC 32
R41
R42
R43
R44
R45
R46
R47
R48
R49
R50
R51

Legend

- Rack is absent
- Rack is empty
- Infrastructure rack
- Critical nodes in the Farm & isolated racks
- Rack is empty, yet power needs to be provisioned
- Rack being retired
- Rack being commissioned

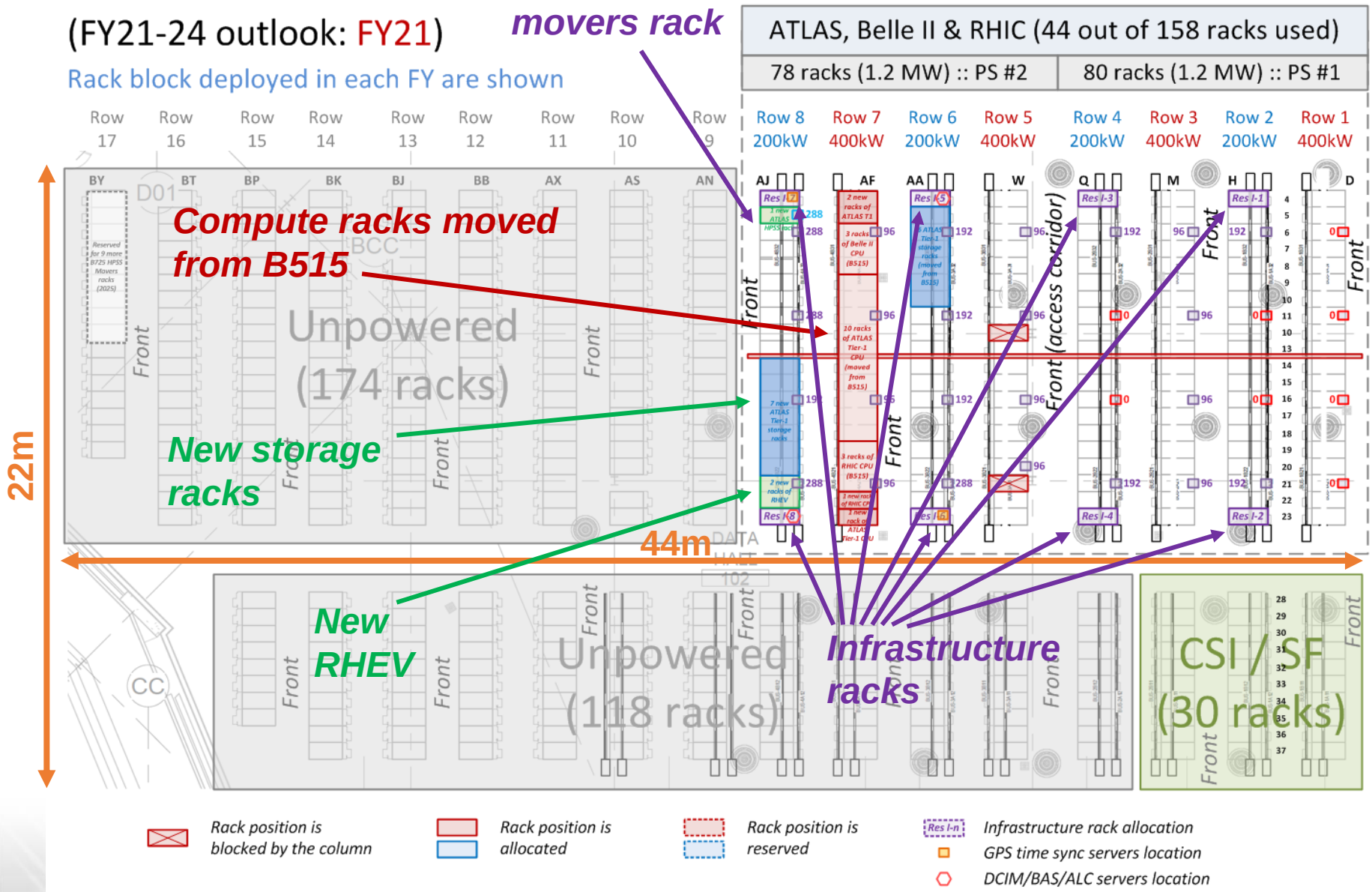
CDCE reaches the state of "tape room only", with local HPSS mover racks still in place (up to 10 racks)

B515/CDCE Floor Map (2025Q2)



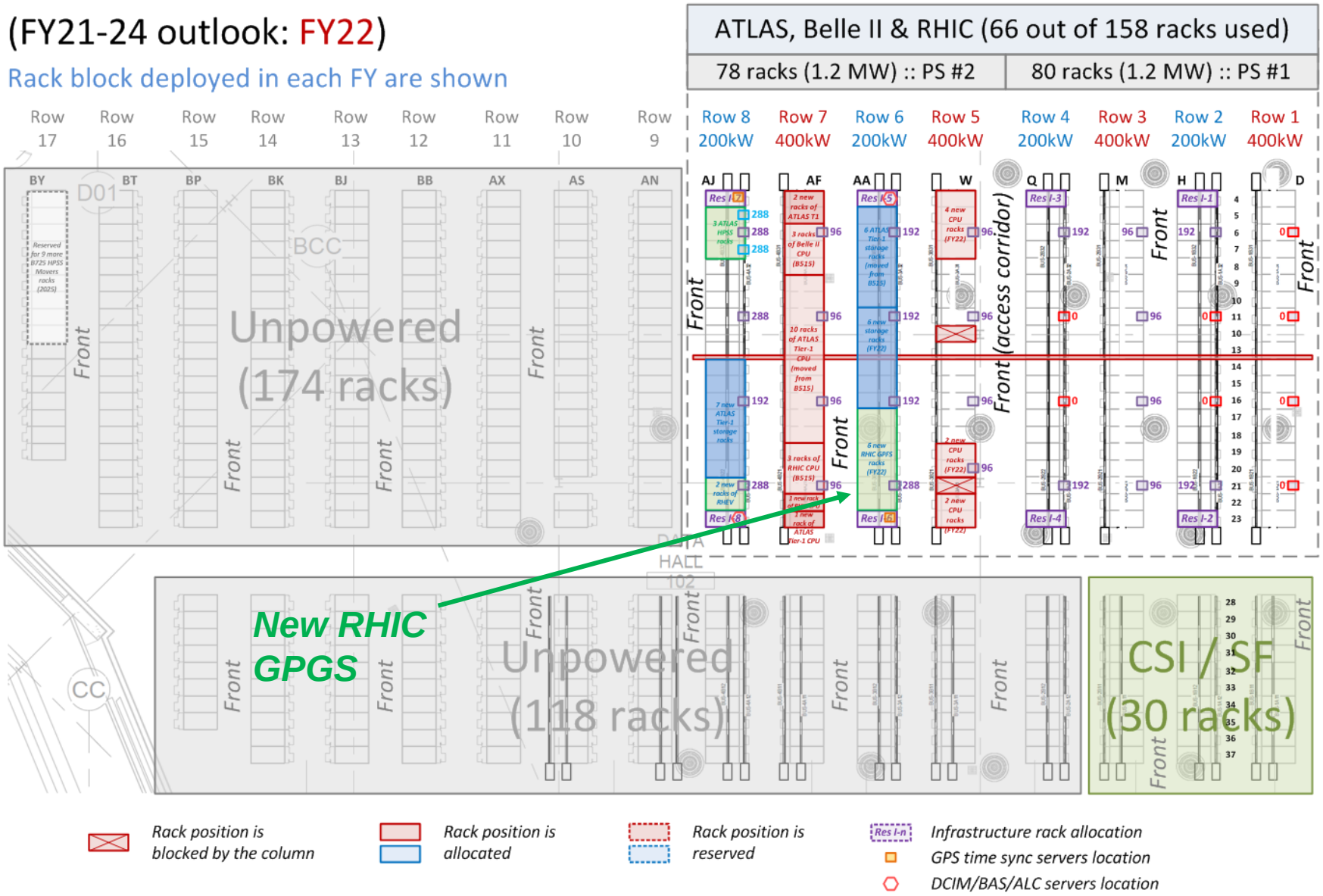
B725 Main Data Hall
(FY21-24 outlook: FY21)

Rack block deployed in each FY are shown



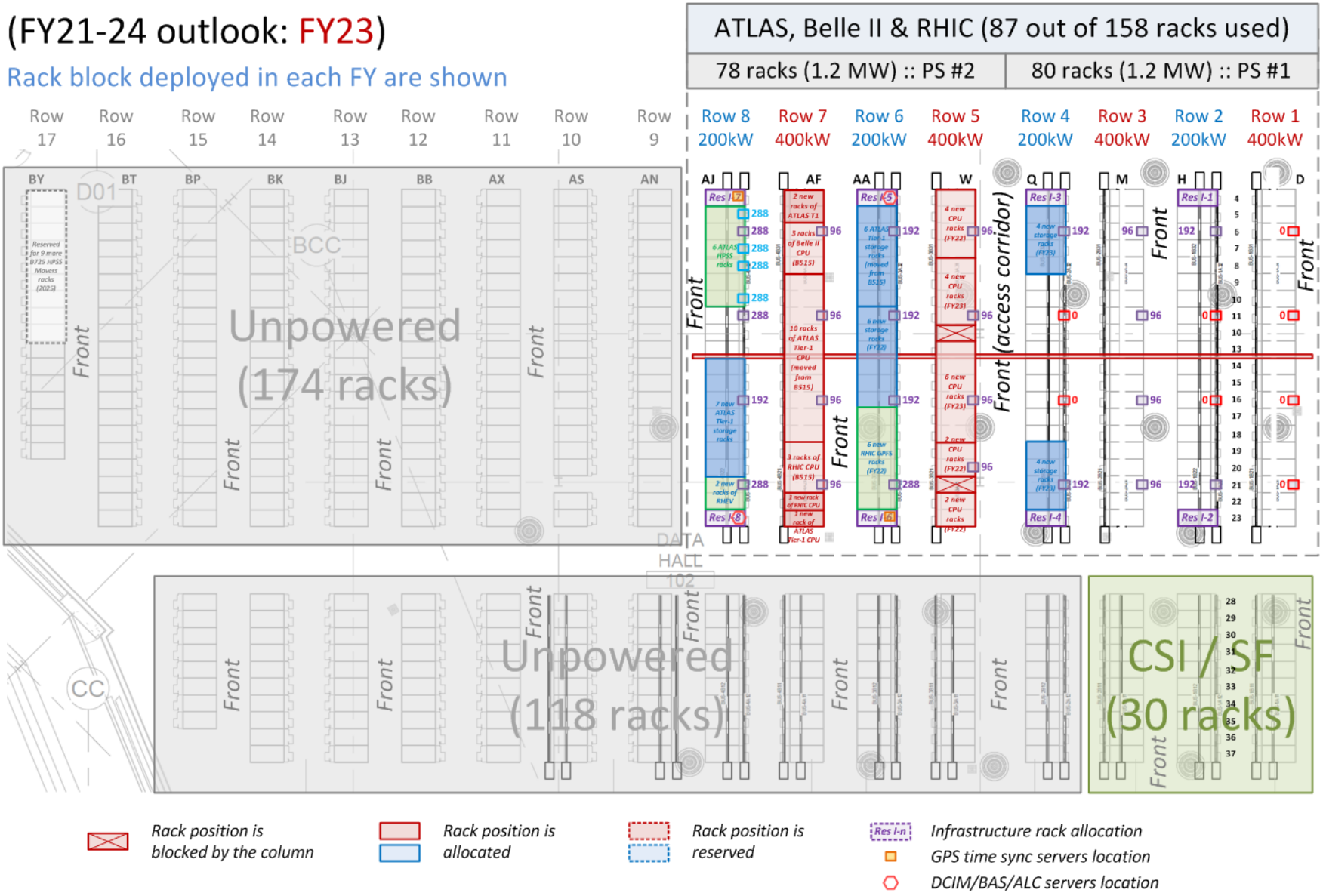
B725 Main Data Hall
(FY21-24 outlook: FY22)

Rack block deployed in each FY are shown



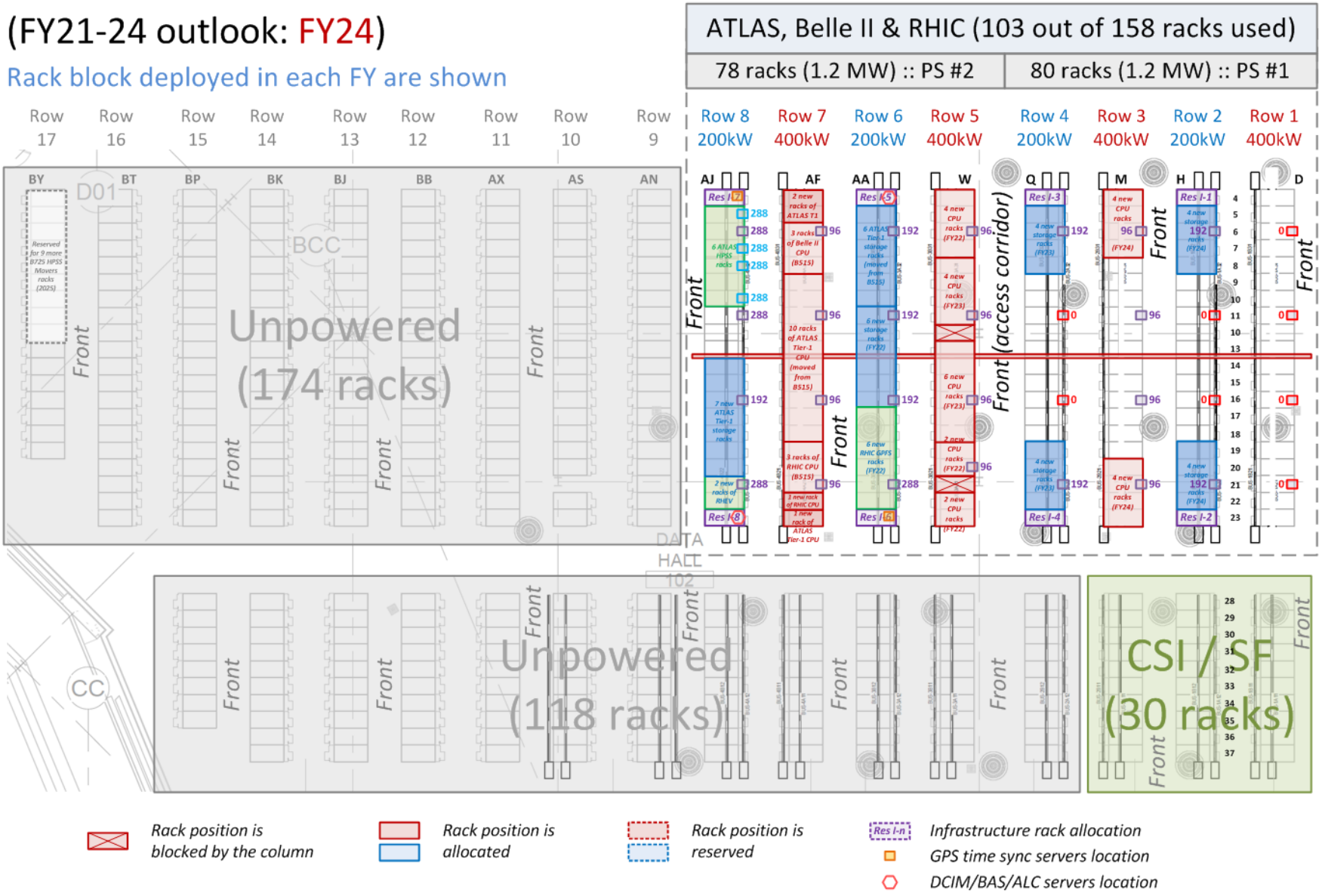
B725 Main Data Hall
(FY21-24 outlook: FY23)

Rack block deployed in each FY are shown



B725 Main Data Hall
(FY21-24 outlook: FY24)

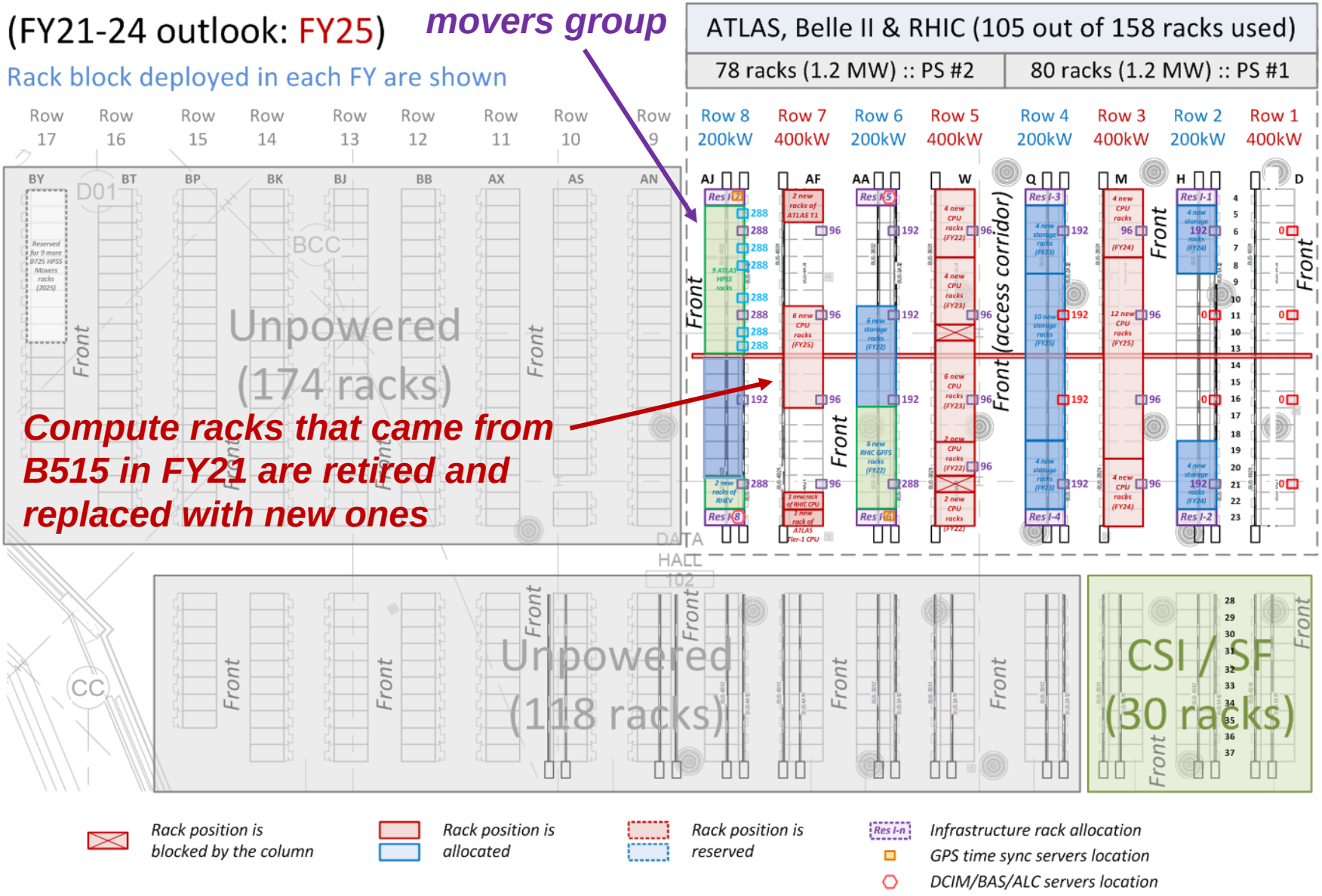
Rack block deployed in each FY are shown



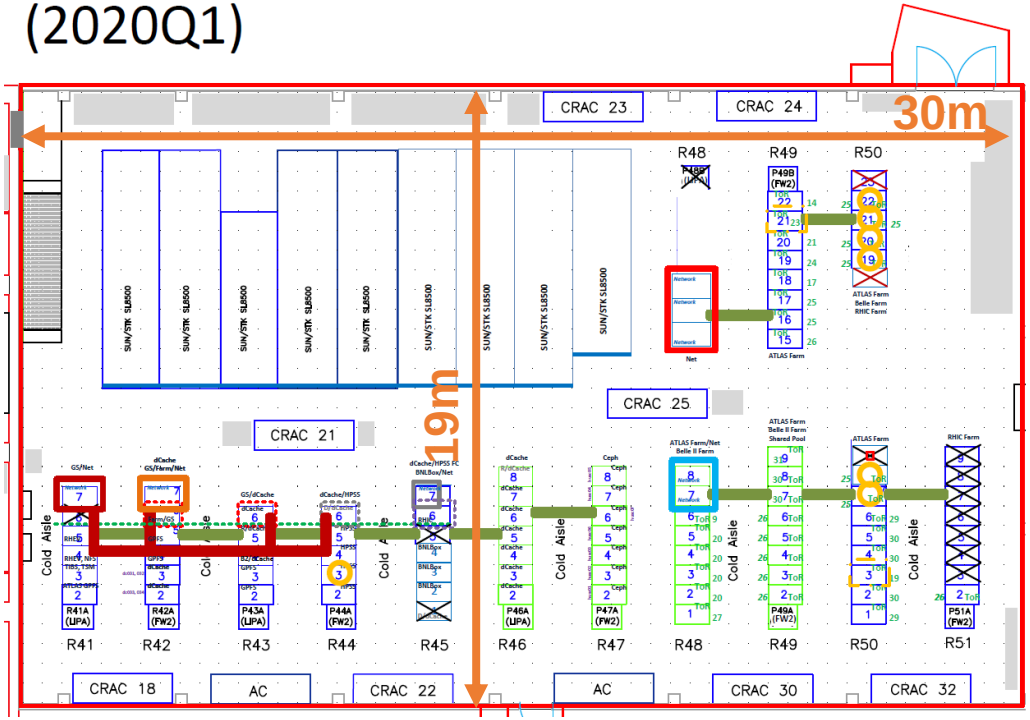
B725 Main Data Hall (FY21-24 outlook: FY25)

Rack block deployed in each FY are shown

Maxed out B725 HPSS movers group

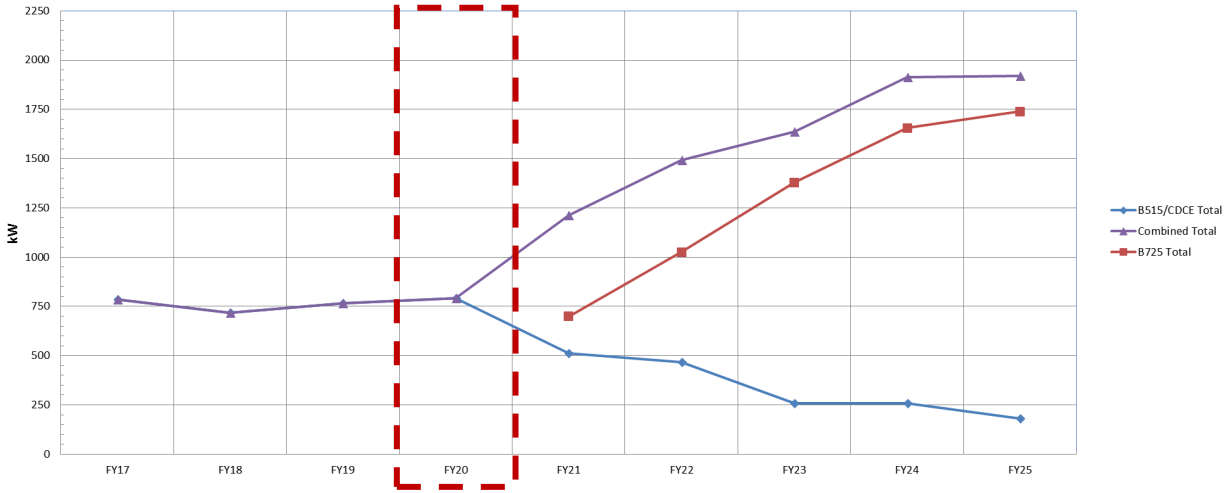


B515/CDCE Floor Map (2020Q1)



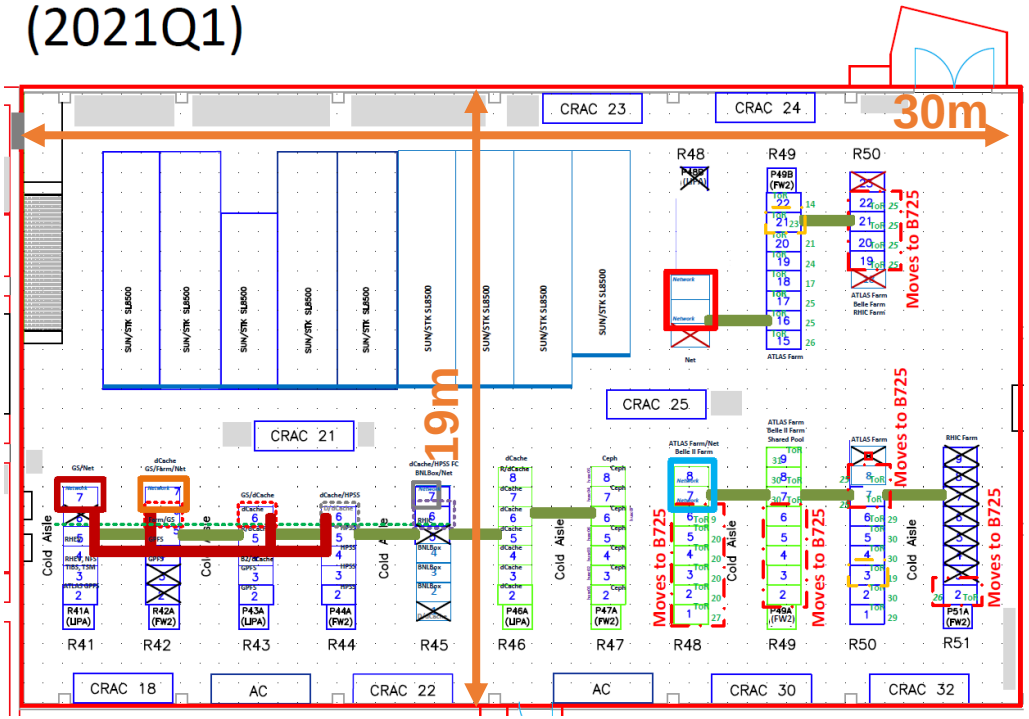
B725 datacenter is under construction

B515/CDCE and B725 Data Center Power Allocated to RHIC, ATLAS and Belle II IT Payload vs Time

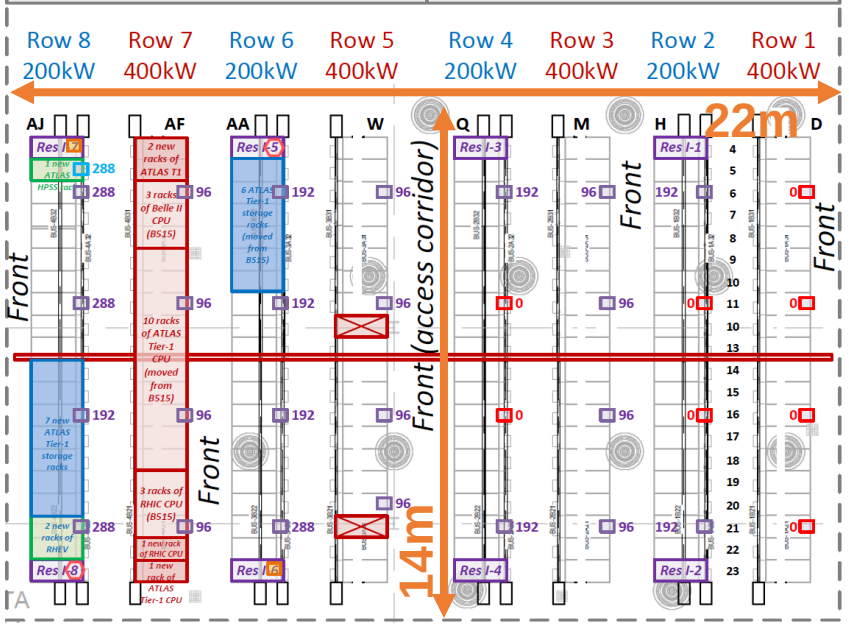


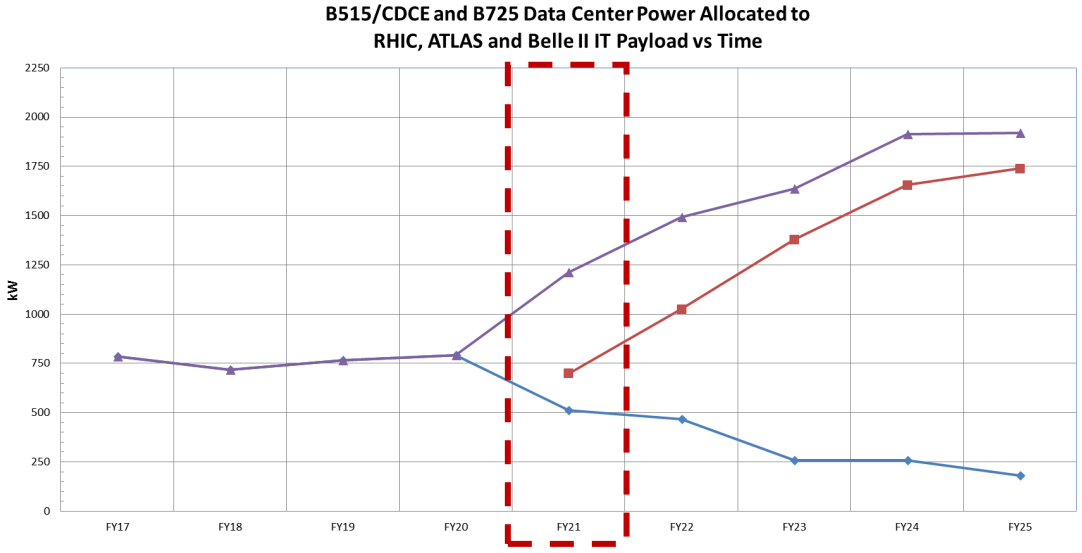
BROOKHAVEN NATIONAL LABORATORY

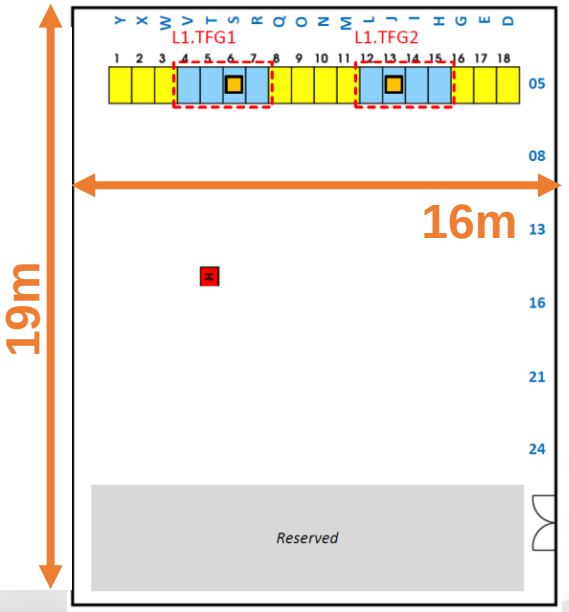
B515/CDCE Floor Map (2021Q1)



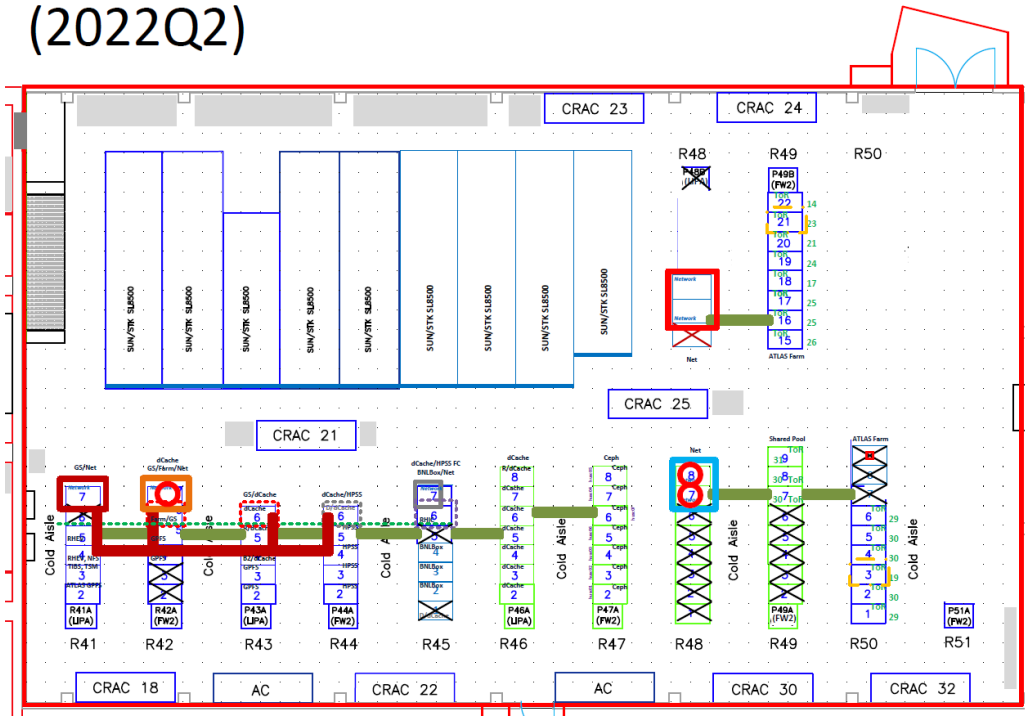
ATLAS, Belle II & RHIC (44 out of 158 racks used)	
78 racks (1.2 MW) :: PS #2	80 racks (1.2 MW) :: PS #1



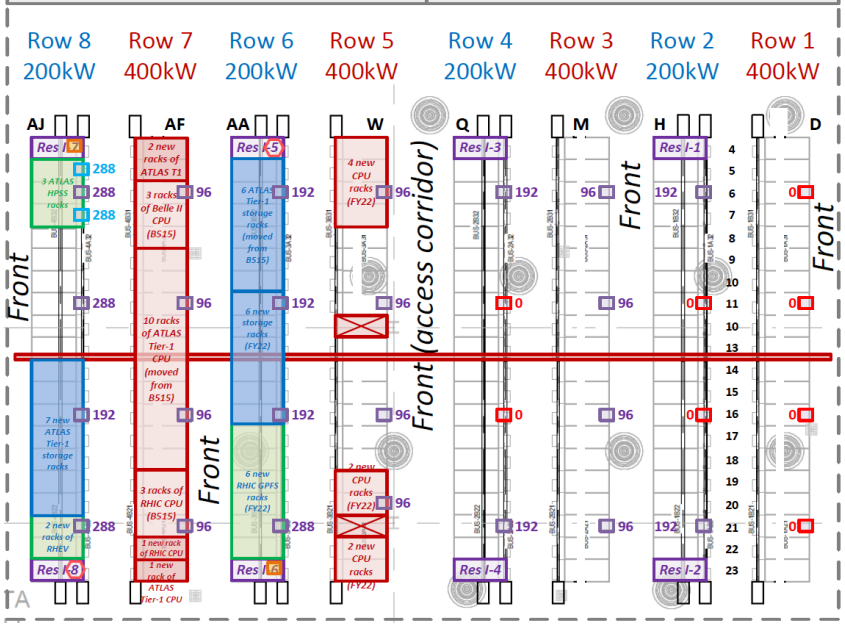


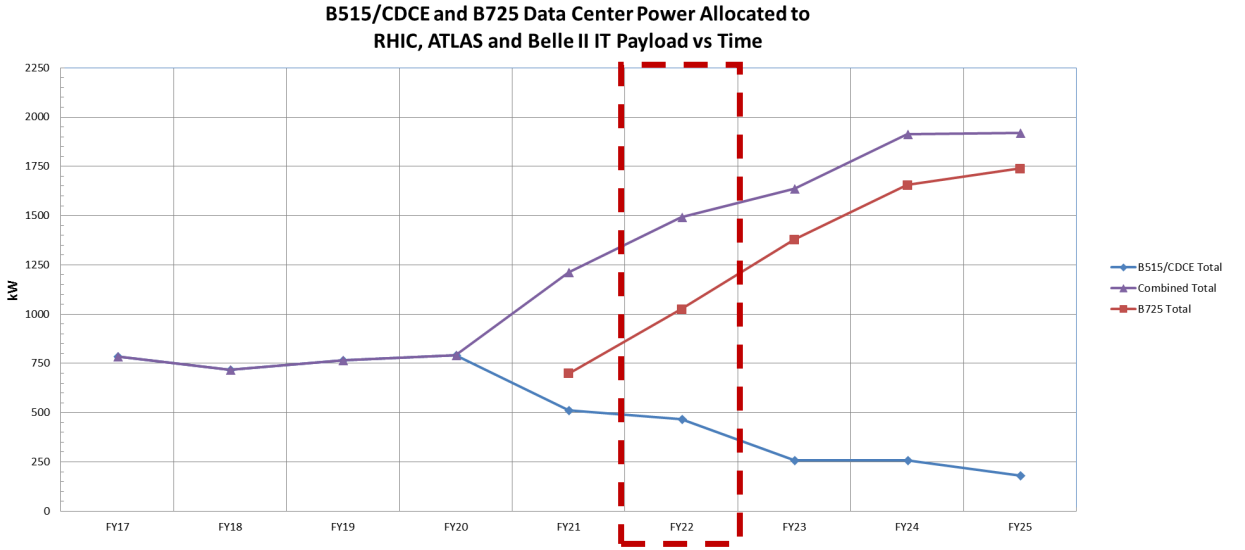


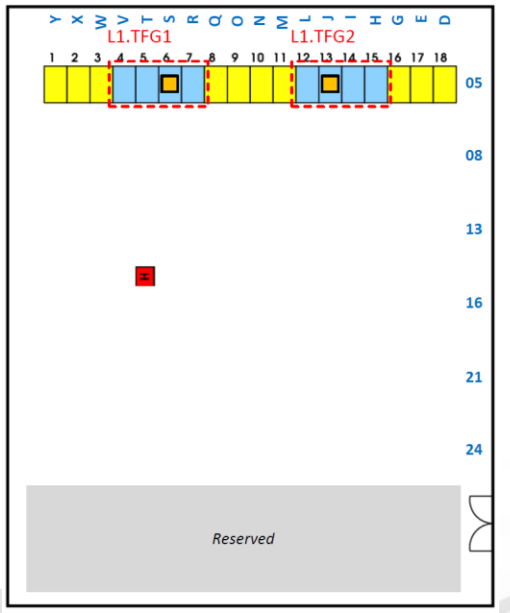
B515/CDCE Floor Map (2022Q2)



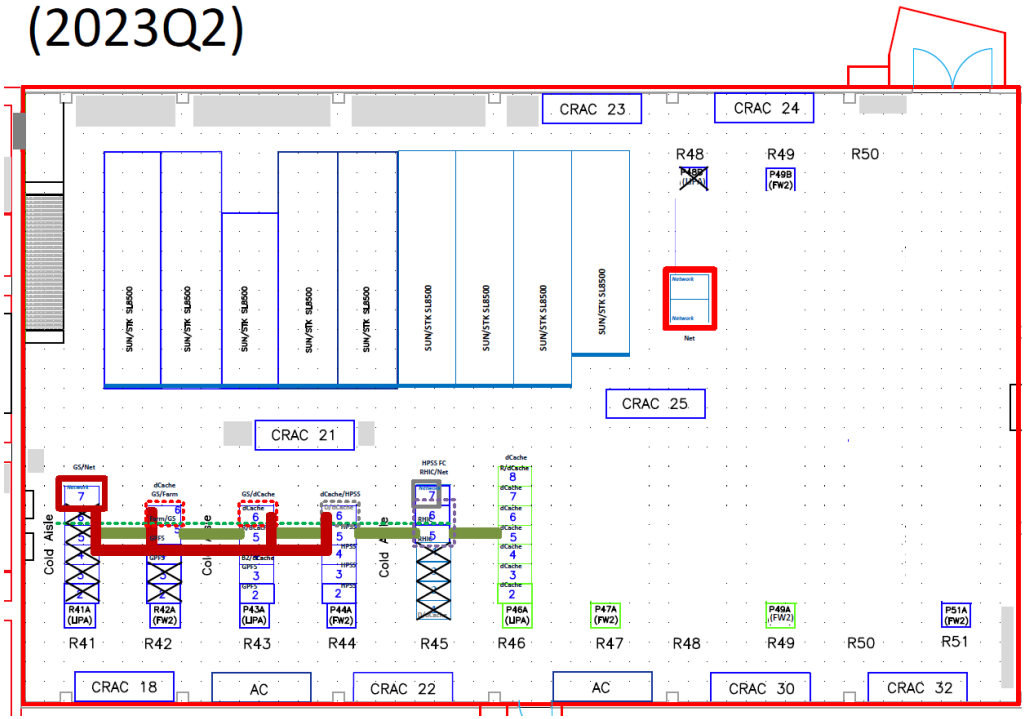
ATLAS, Belle II & RHIC (66 out of 158 racks used)	
78 racks (1.2 MW) :: PS #2	80 racks (1.2 MW) :: PS #1



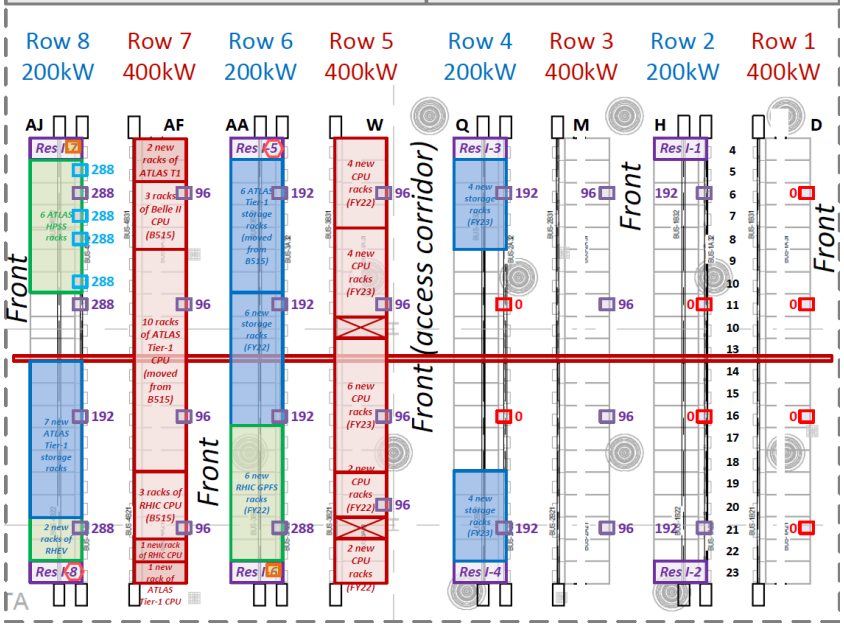


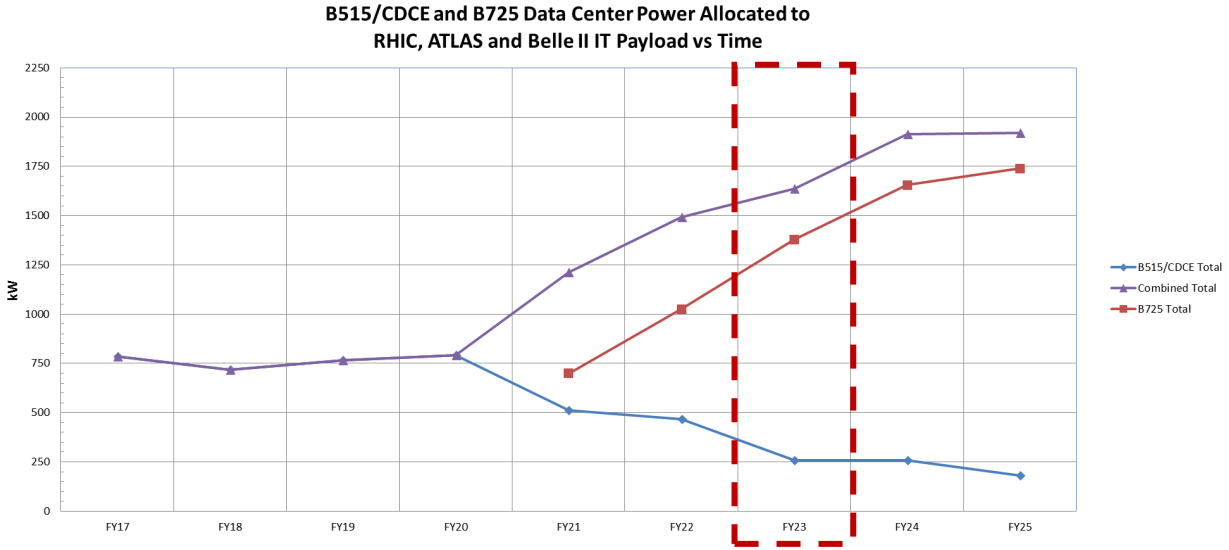


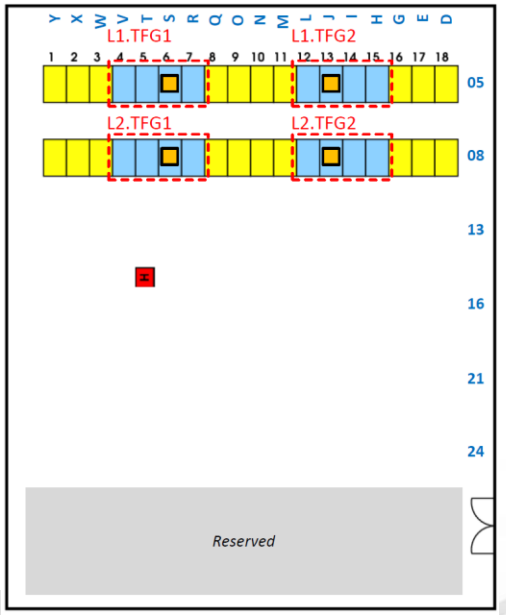
B515/CDCE Floor Map (2023Q2)



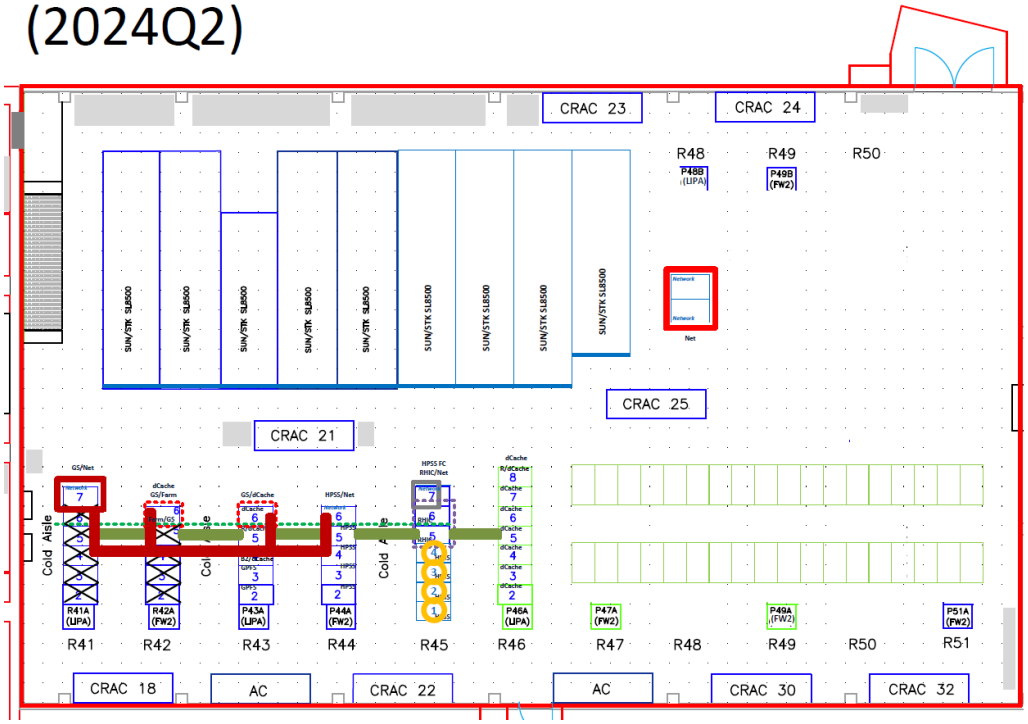
ATLAS, Belle II & RHIC (87 out of 158 racks used)	
78 racks (1.2 MW) :: PS #2	80 racks (1.2 MW) :: PS #1



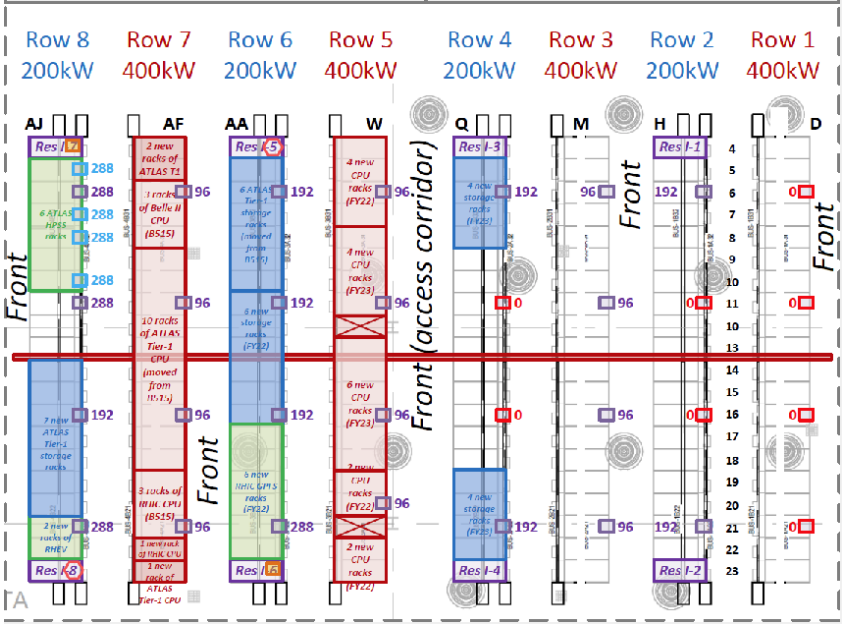


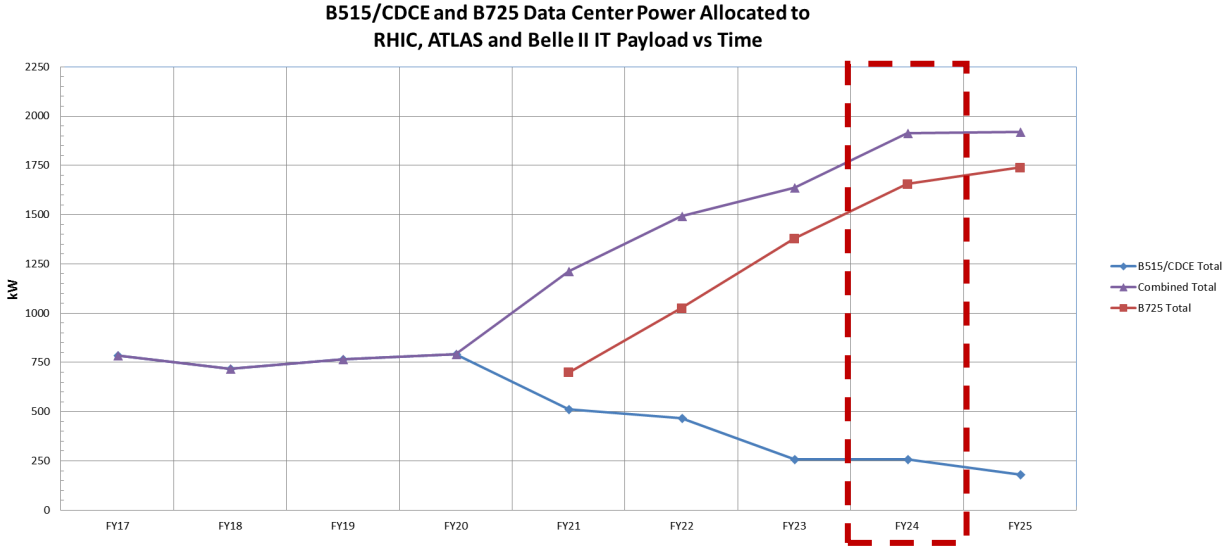


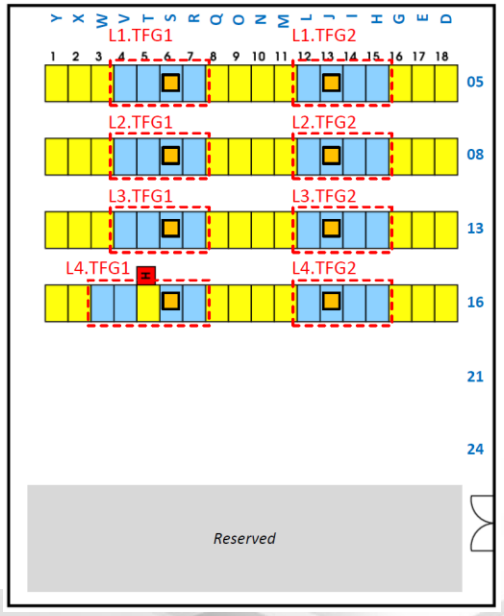
B515/CDCE Floor Map (2024Q2)



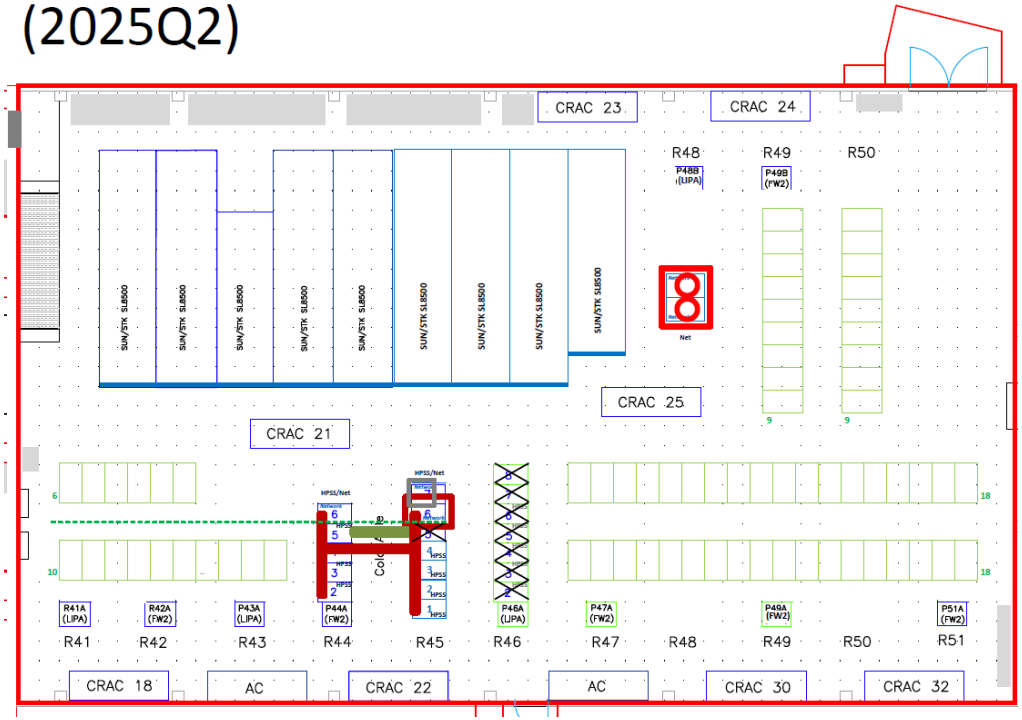
ATLAS, Belle II & RHIC (87 out of 158 racks used)	
78 racks (1.2 MW) :: PS #2	80 racks (1.2 MW) :: PS #1

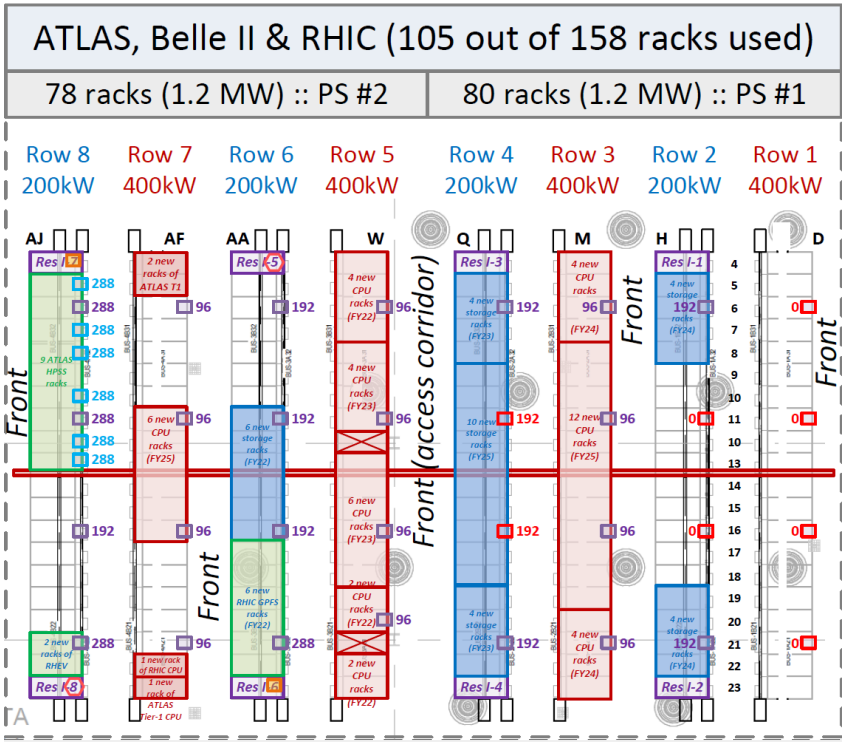


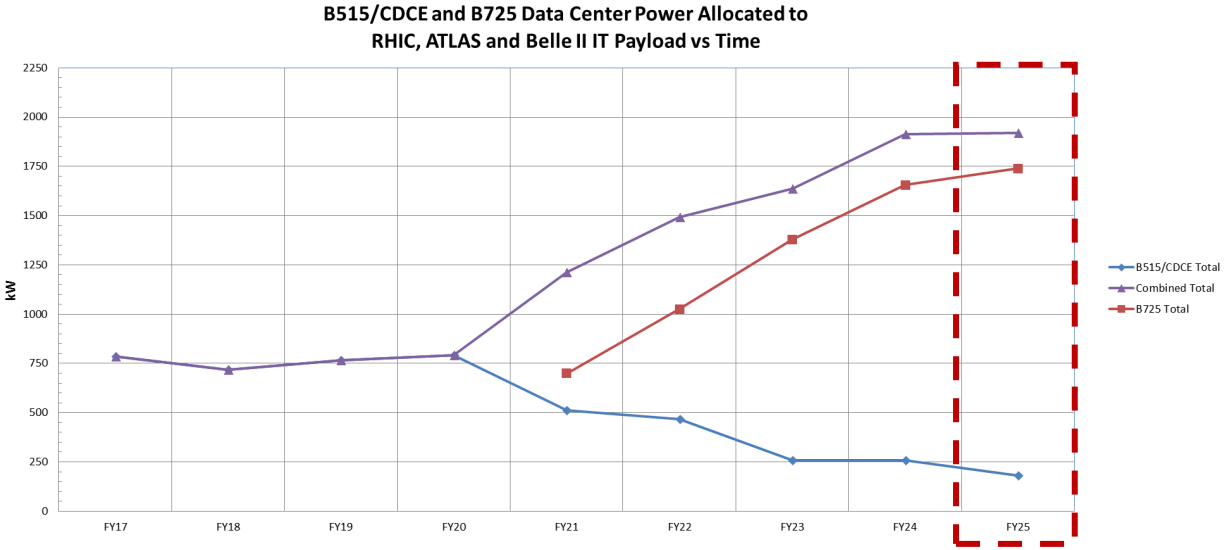


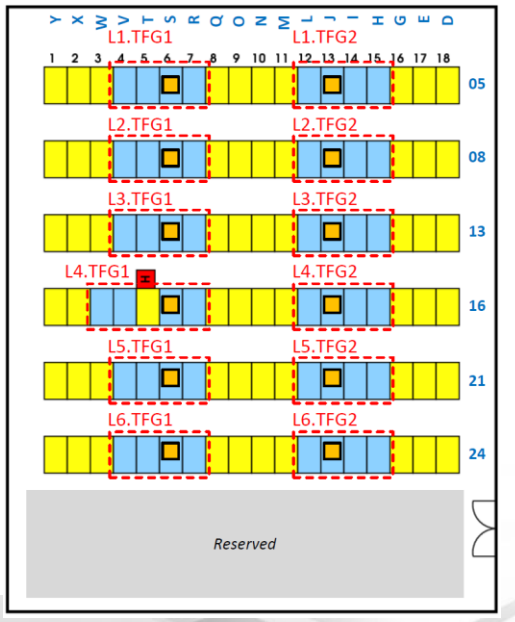


B515/CDCE Floor Map (2025Q2)

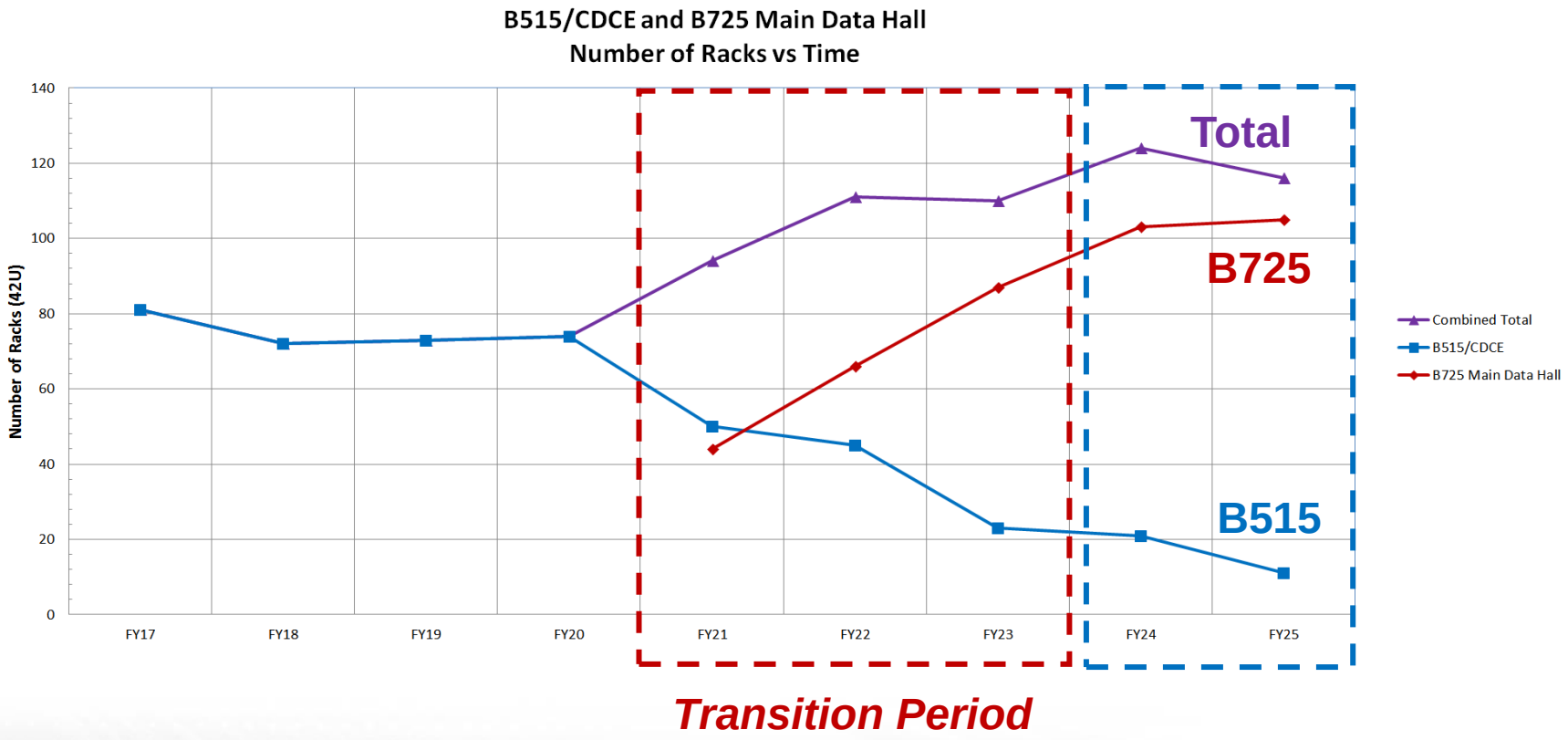




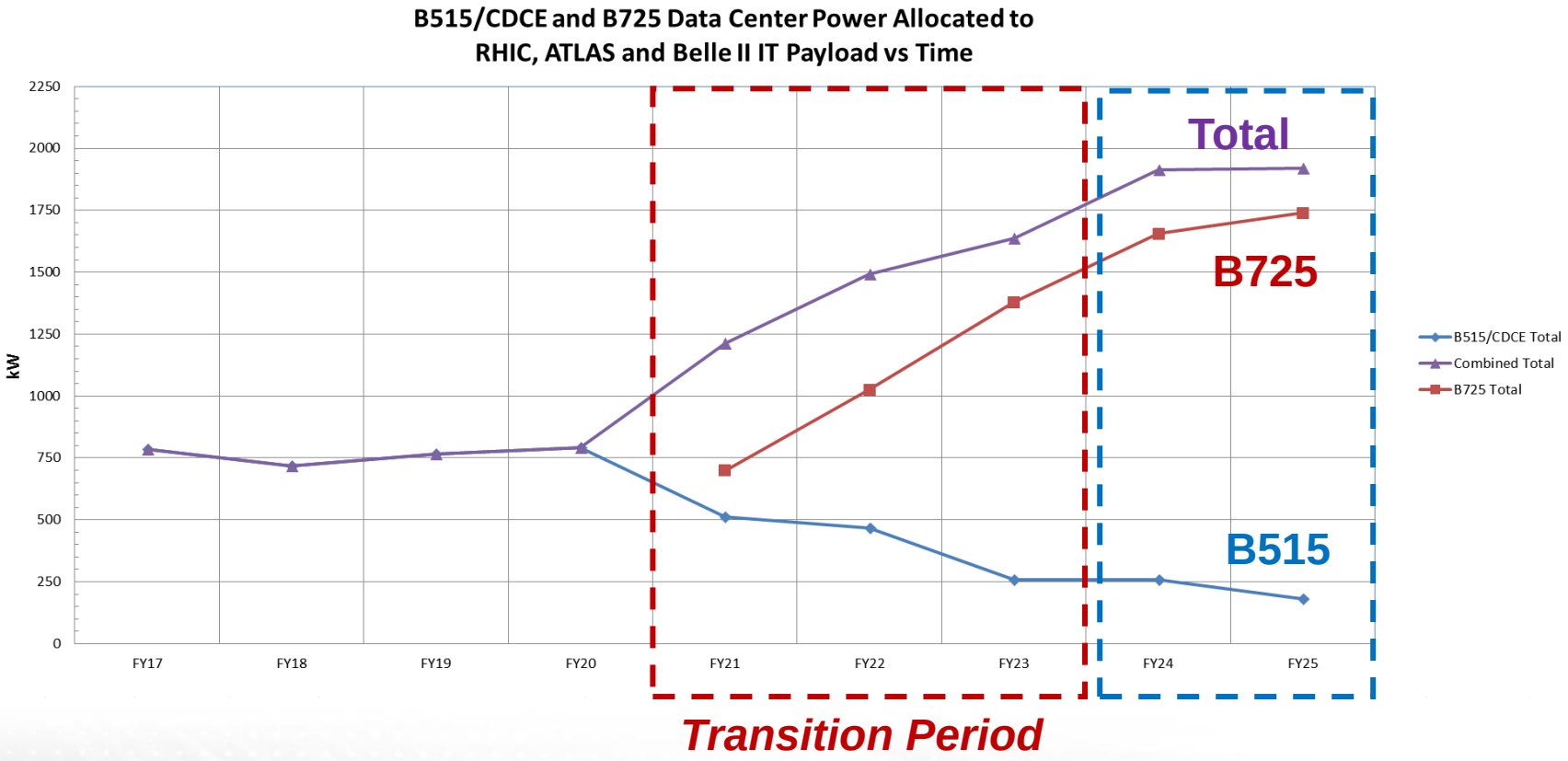




B515 & B725: Projected Number of Racks vs Time in early FY17-25 (only RHIC, ATLAS and Belle II for B725)

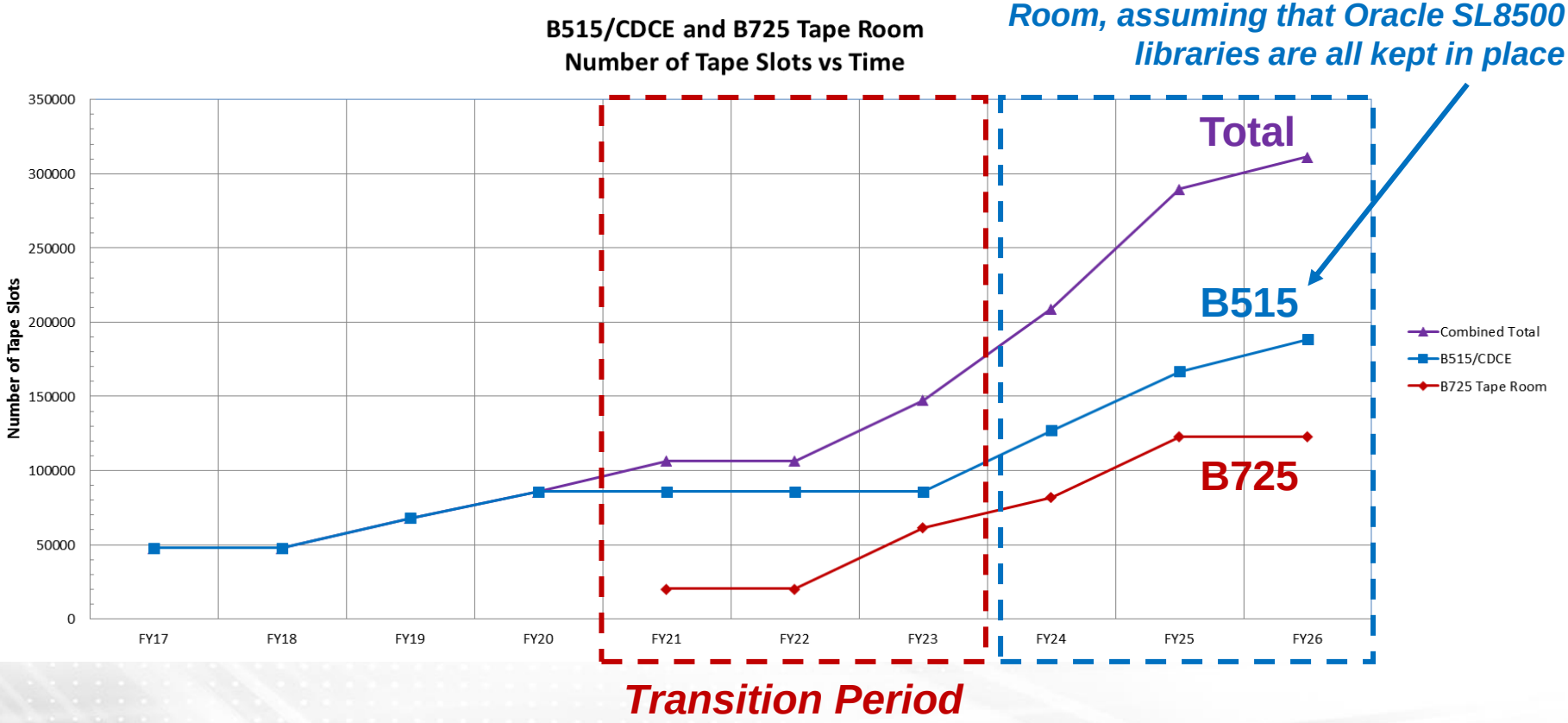


B515 & B725: Projected Power Allocation vs Time in early FY17-25 (only RHIC, ATLAS and Belle II for B725)



B515 & B725: Projected Tape Slot Count vs Time in early FY17-25 (only RHIC, ATLAS and Belle II for B725)

Under the scenario of populating the freed space in CDCE with the same type of tape libraries to be installed in B725 Tape Room, assuming that Oracle SL8500 libraries are all kept in place



Summary

- The design phase for the B725 datacenter as part of the BNL CFR project is nearly over and construction is expected to begin in Apr 2019
- B515 datacenter is already on track for consolidation into the CDCE and Lab C areas since FY18, yet there is much left to do in FY19-20 to achieve that goal before the new datacenter becomes available
- The gradual migration of compute and storage equipment between the B515 and B725 datacenters is expected to occur in FY21-23 period, with the first ATLAS equipment to be operational in Feb 2021
- B515 datacenter reduction to the CDCE and Lab C areas is expected to be completed by the end of FY23 by which time the B725 datacenter is expected to reach the level of 1.8 MW of combined IT power for the RHIC, ATLAS and Belle II equipment (the sPHENIX, CSI HPC, SF, CFN and NSLS II contribution to this process is yet to be quantified and added to the plan)

Questions & Comments