



CMS TriDAS project

Infrastructure issues for the CMS online farm

`attila.racz@cern.ch`

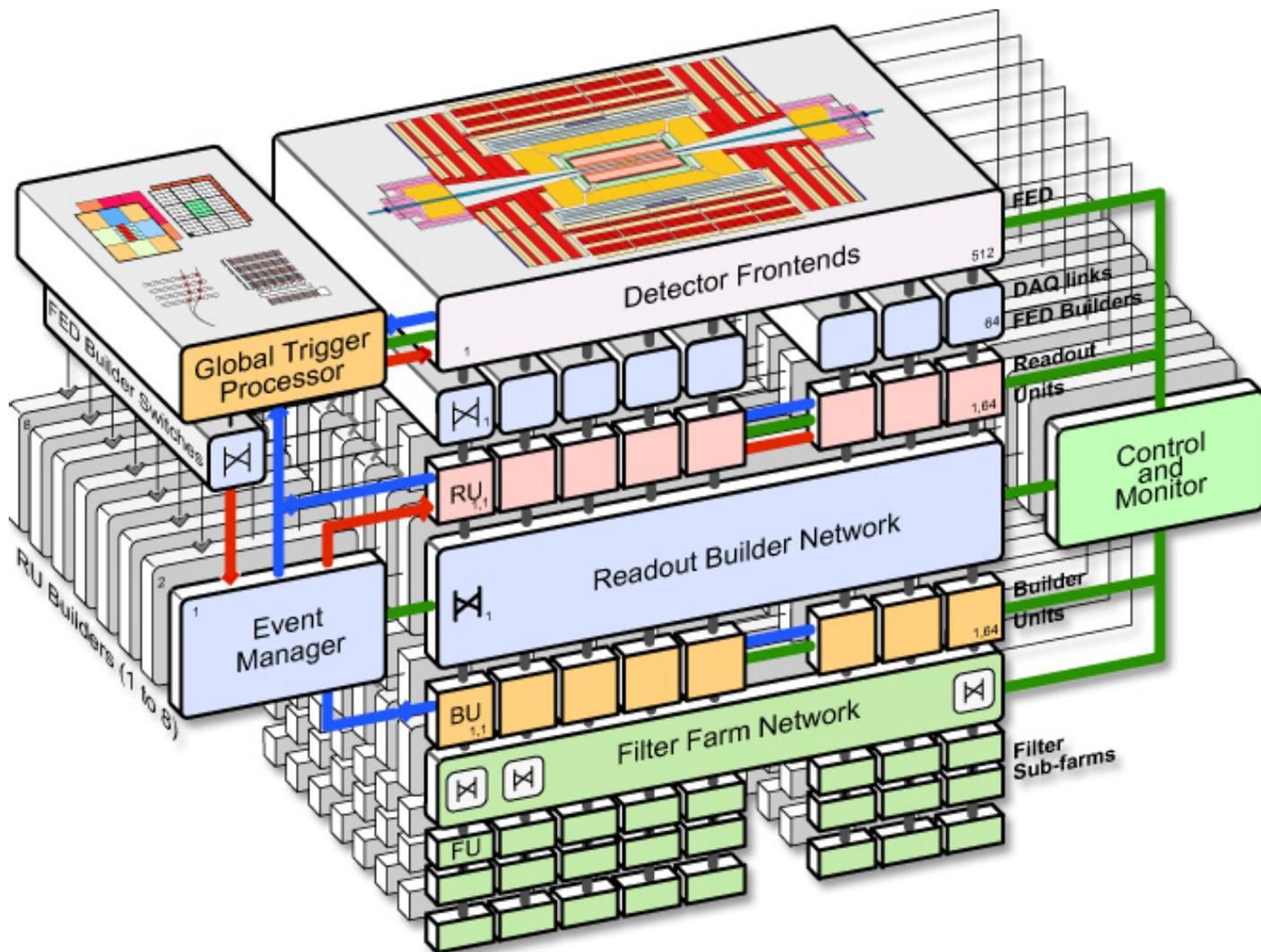


Outline

- Introduction
- Bill of Material
- Cooling issues
- Fire protection
- Power supplies
- Networking
- What next ? / Conclusion

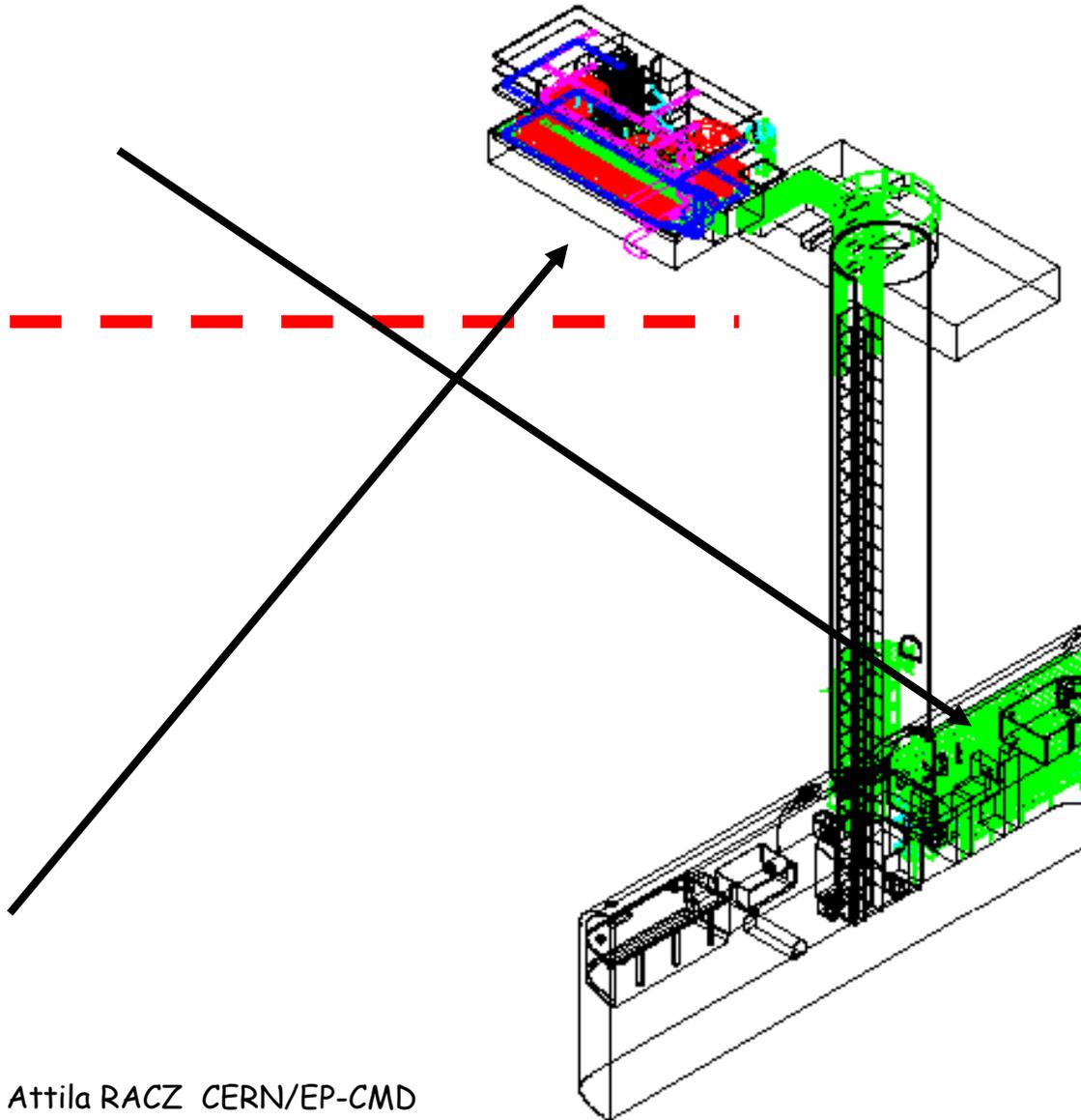
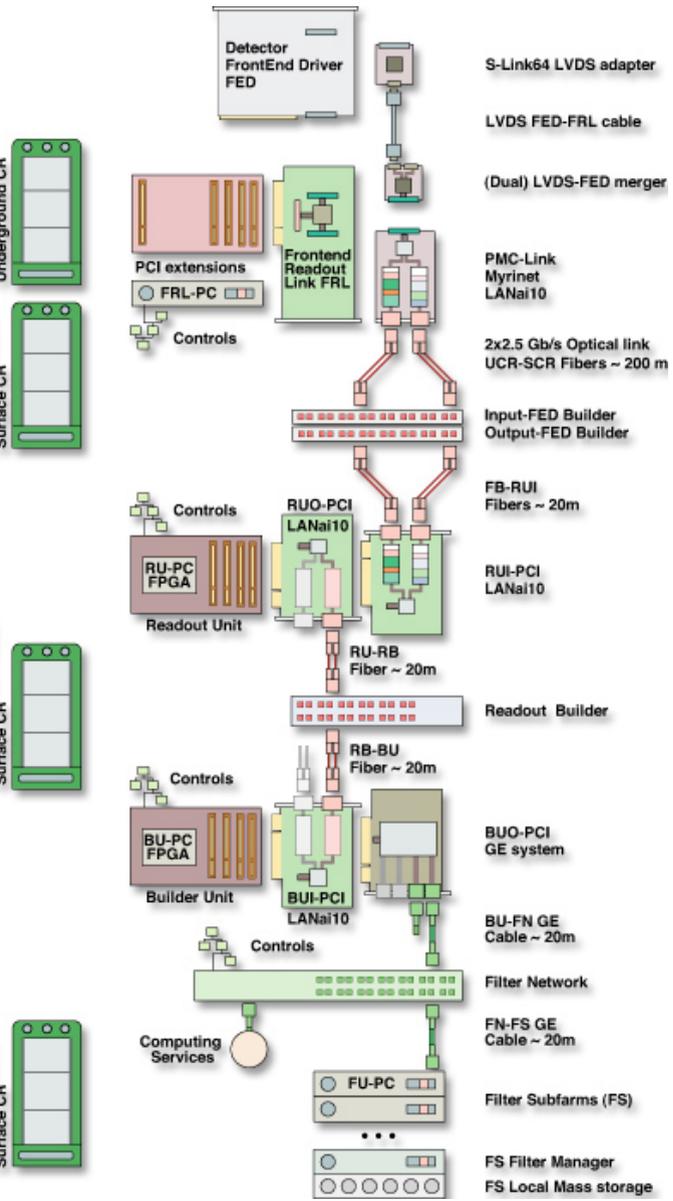


General view



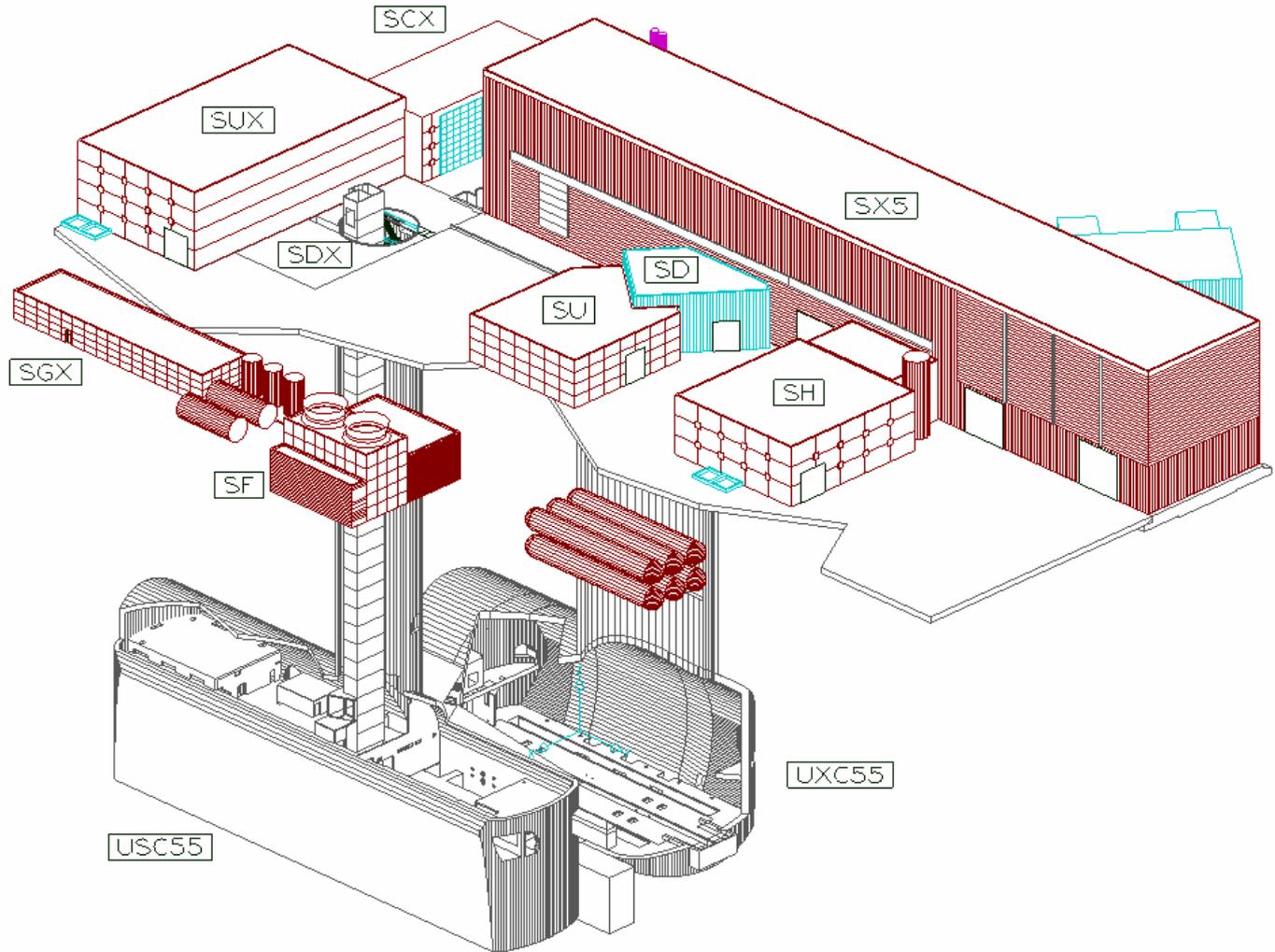


One column... out of 512 !



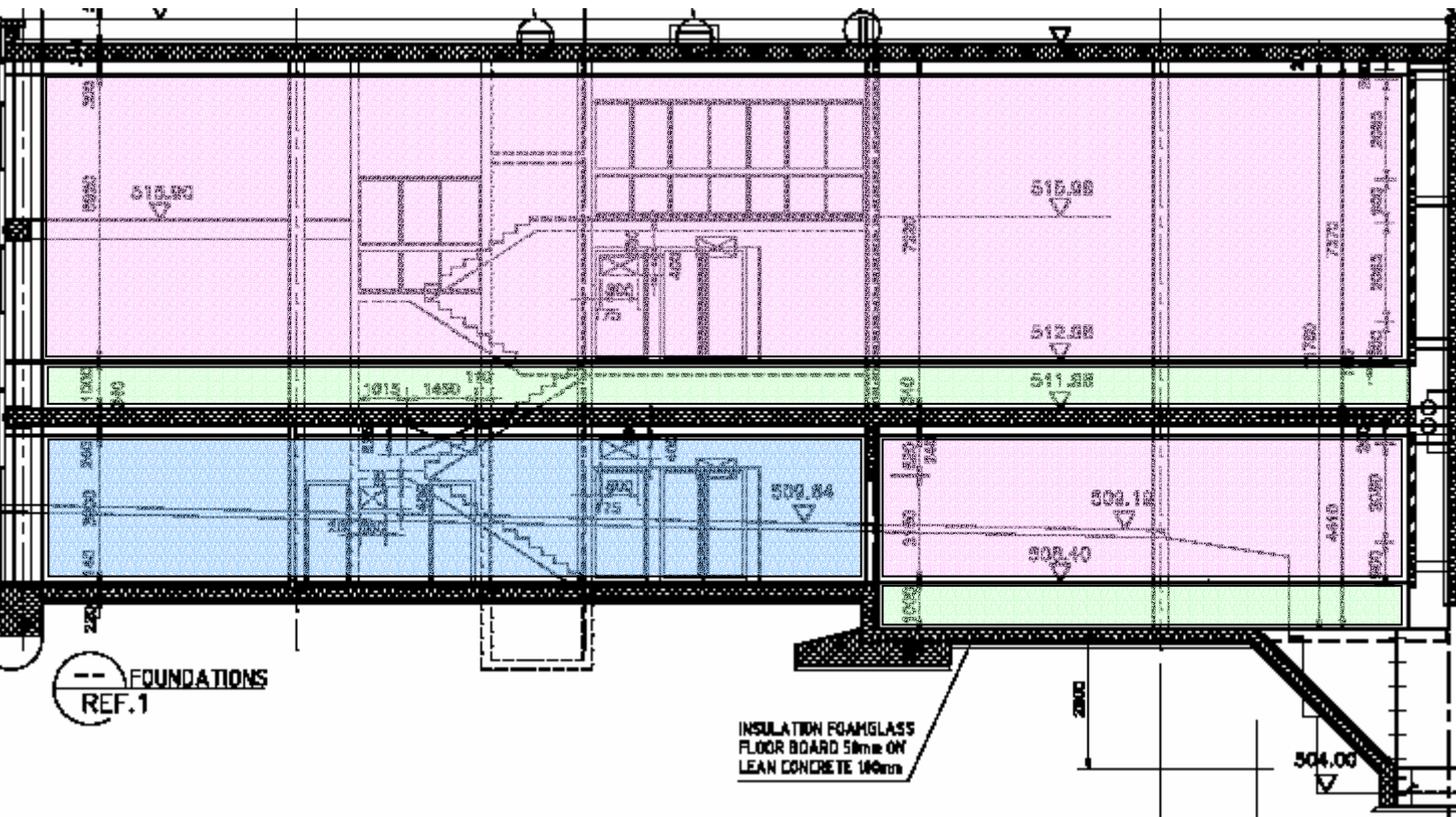


Building names...





DAQ building (SCX)



352 m²,
~6m height

165 m²,
~3m height

←
Data fibers and
commodity network
from the pit

Overall dimension:
~ 14 x 30 m²



Timescales

- Buildings available Q2-Q3 2005
- Start of Detector-DAQ integration: Q2 2006
- First LHC beams: April 1, 2007
- Computing farm is deployed by "slices"
 - Processing power will follow beam luminosity ramping-up and budget time-profile...
 - We want to profit from Moore's law...



SCX BOM...

- FEDB switch: 16 x 128 port switches (2.5 Gb/s)
- RUs: 256 PCs
- RUB switch: 8 x 128 port switches (2.5 Gb/s)
- BUs: 256 PCs
- FUs: 2048 PCs, dual processor
 - Other form factor: blades (see next slide)
- Misc: assuming equivalent of 4 racks

- FU PCs: 1U, 12 kg, 250 W
- RU/BU PCs: 3U, 20 kg, 300 W
- 2.5 Gb/s Switches: 9U, 31 kg, 960 W
- Rack: 42U, 170 kg



1U PC versus Blades

Computer blades allow up to ~170 processors per rack vs ~80 proc. with 1U dual CPU PCs

With such form factor, we could reduce the number of FU racks from 64/52 down to 25

Hardware maintenance operations are optimized

But...power dissipation can go up to ~18kW per rack...

No interchangeability from vendor to vendor as with 1U PCs

Proprietary management software

Farm upgrade/renewal more "problematic"

1U PC for FUs is our baseline today



DAQ racks (1)...

- We don't fill the racks to the max. for keeping the power density to a minimum
 - And some contingency is a good practice...
- **A total of 138 racks:**
 - **4** FEDB racks: 4 * 9U 128 port switch
 - **32** RU racks: 8 * 3U PCs + commodity GB switch
 - **2** RUB racks: 4 * 9U 128 port switch
 - **32** BU racks: 8 * 3U PCs + commodity GB switch
 - **64** FU racks: 32 * 1U PCs + commodity GB switch
 - **4** racks for misc. (servers, local storage, ...)

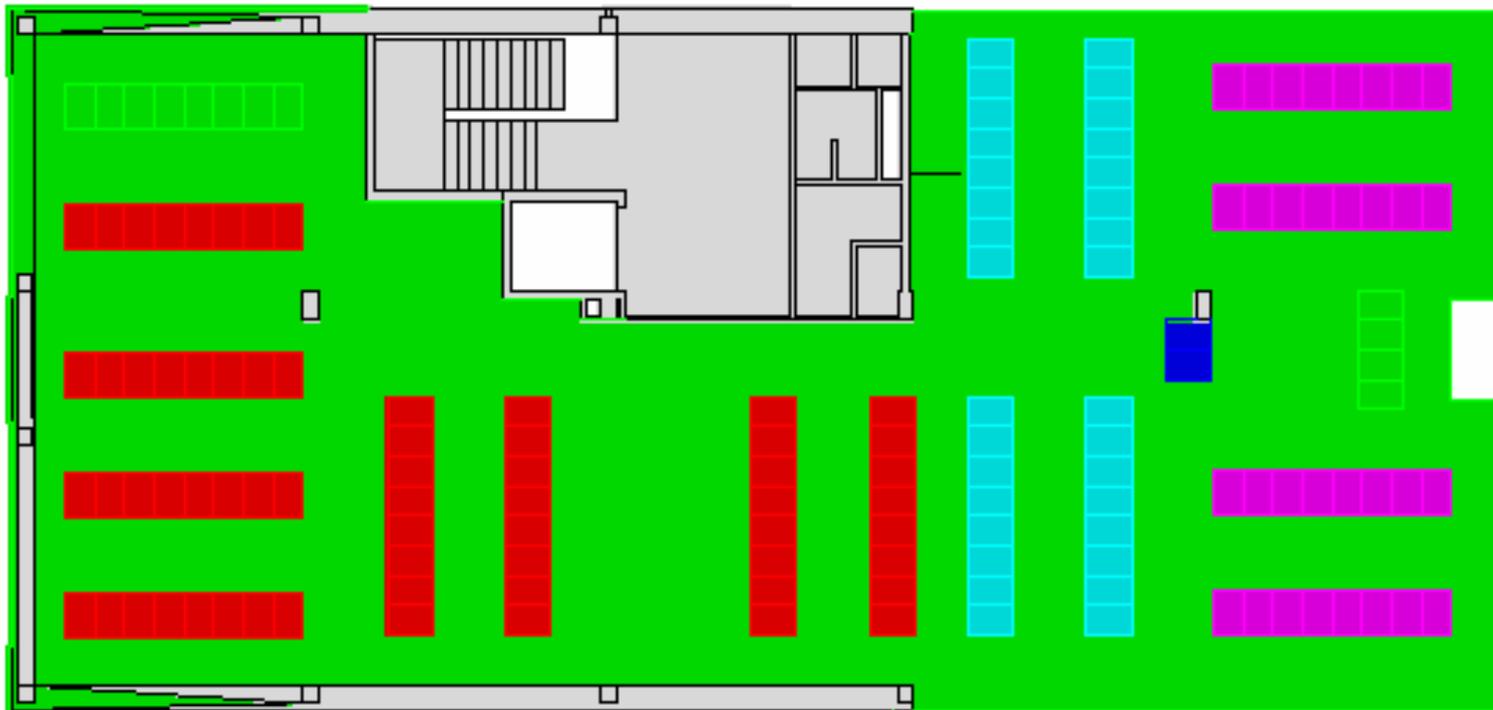


DAQ racks (2)...

Rack type	Weight (kg)	Power per rack (kW)	Rack Quantity	Total power (kW)	At Startup (50 kHz)
FED builder	294	3.84	4	15.36	100%
Readout Unit 256*3U PCs @ 300W	331/411	2.5/3.7	32/22	81.4	50%
Readout Unit builder	294	3.84	2	7.68	50%
Builder Unit 256*3U PCs @ 300W	331/411	2.5/3.7	32/22	81.4	50%
Filter Unit 2k*1U PCs @ 250W	560/655	8.3/10.3	64/52	535.6	50%
Computing services	331/411	4	4	16	100%
Total			138/106	~738	~385

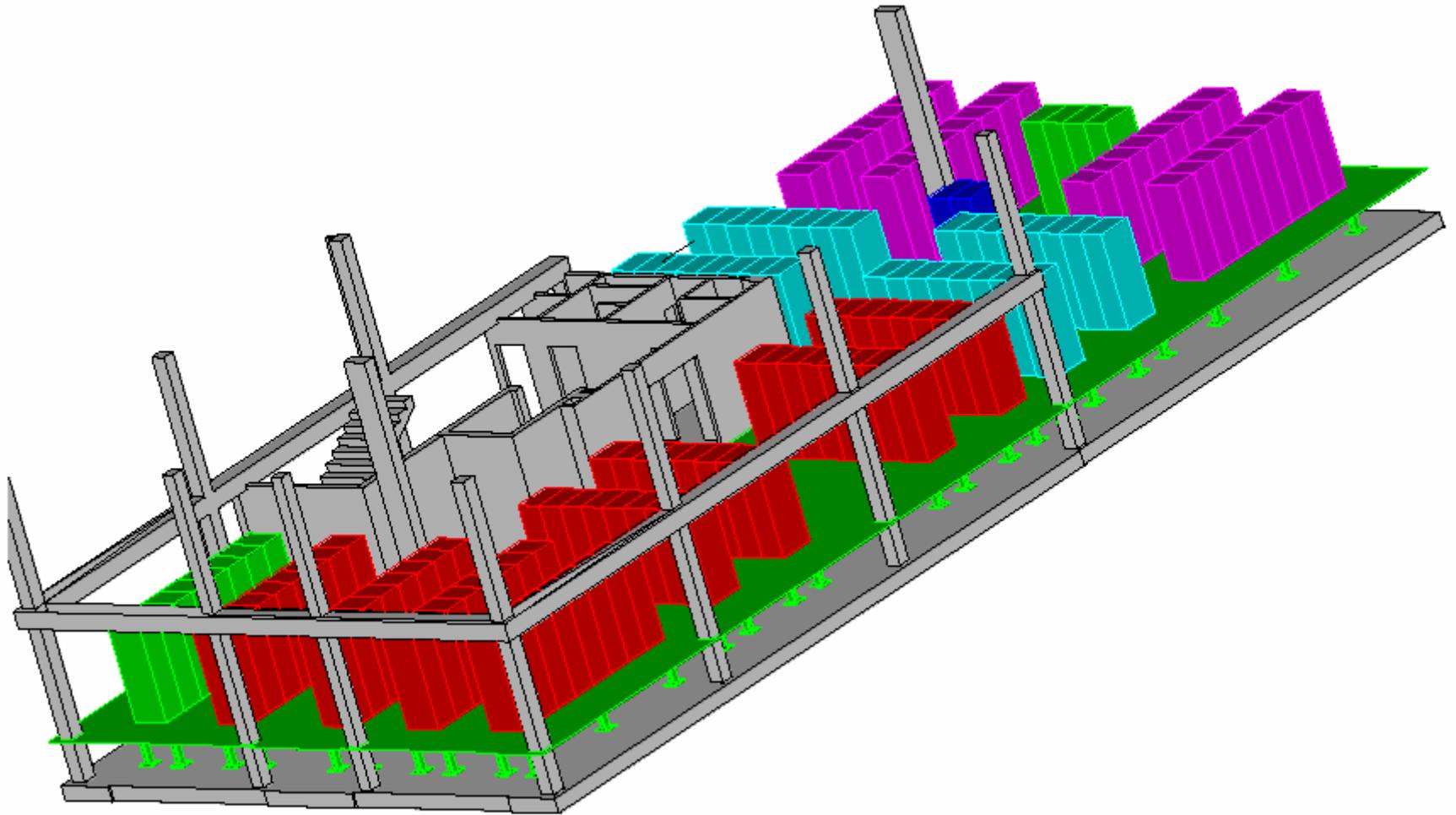


Rack layout (low packing)





3D rack layout





Cooling issues (1)...

- Typical power per rack is 3-4 kW max. and typical power density is 0.5-0.8kW/m²
- We have ~8 kW/rack and 2 kW/m²
 - Traditional plenum floor is not adapted (floor surface, air-flow, efficiency...)
- We intend to use water-cooled racks



Cooling issues (2)...

- We don't want to have water in the PCs
 - Restriction on the PC choice, delicate operations when maintenance or upgrade
- Up to now, very few products on the market
- Water cooling seems to be a solution for future high power density data centers
 - Keep an eye opened on the market...



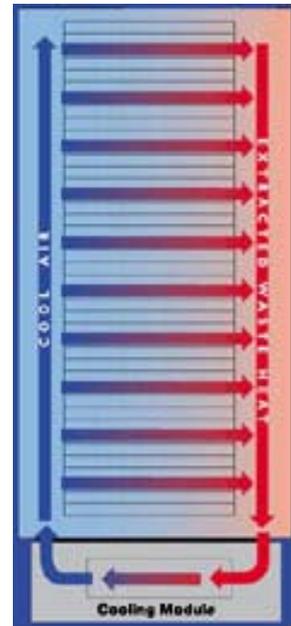
Rack cooling products...



BlackCooler: 8 kW,
cooling module mounted
in standard racks



Knurr: 10 kW
Internal cooling module,
closed air-circuit



Ecobay: 10.5 kW/cab
Internal cooling mod
closed air-circuit



Fire protection

- The experimental zone is at ~30 min away from the fire brigade...
 - Need for a fire inertion system
- Rack type has big influence
 - With open racks, full volume protection (human presence may be a problem...)
 - With closed racks, built-in inertion system and environmental control
- Need for a comparative cost analysis !



Power supplies

- The idea is to have a main switching unit for each rack. Then, within a rack, have a remote controlled Power Distribution Unit
- UPS for strategic machines only
- Too early to go further...



Networking...

- Ideally, separated network for:
 - Data transport
 - Control messages
 - Monitoring and services
- The topology should
 - Minimize the cabling in the building (stay within a rack as much as possible, BU+FUs)
 - Allow easy reconfiguration: no direct intervention on the machines -> usage of starpoints
- Again, too early to go further...



What next ?

- Preserie (1/8th) assembled in 2004 for gaining experience on HW and SW aspects
- Continue to optimize the hardware elements w.r.t. computing requirements
- Look at the technological evolution for Computers/telecom market and its impact on the infrastructures

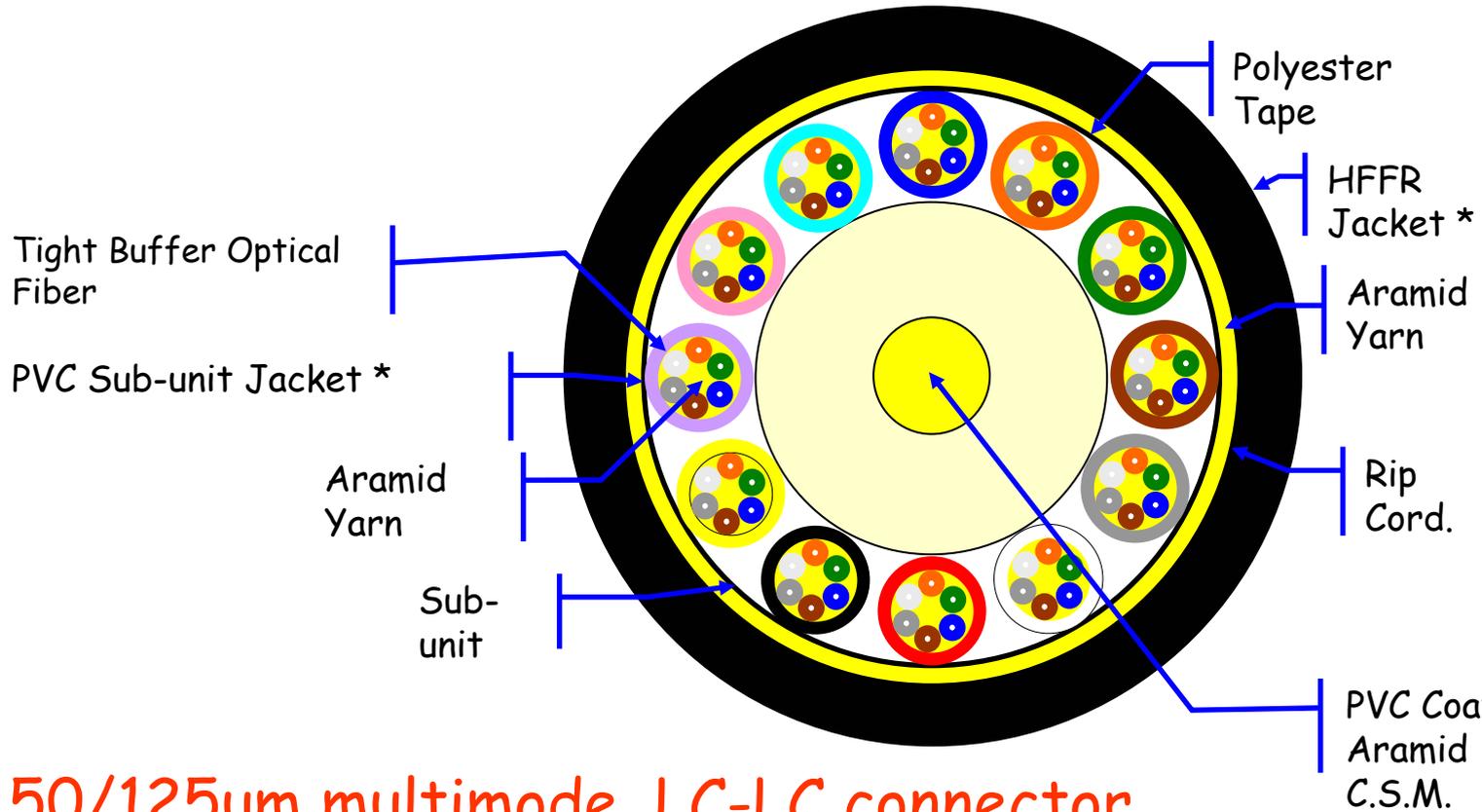


Conclusion

- Most of the DAQ must stay modular...
 - Parallel DAQ (partitions)
 - Sliced deployment
 - Usage of new/appearing technologies
- Mixing of different platforms and different vendors is very likely



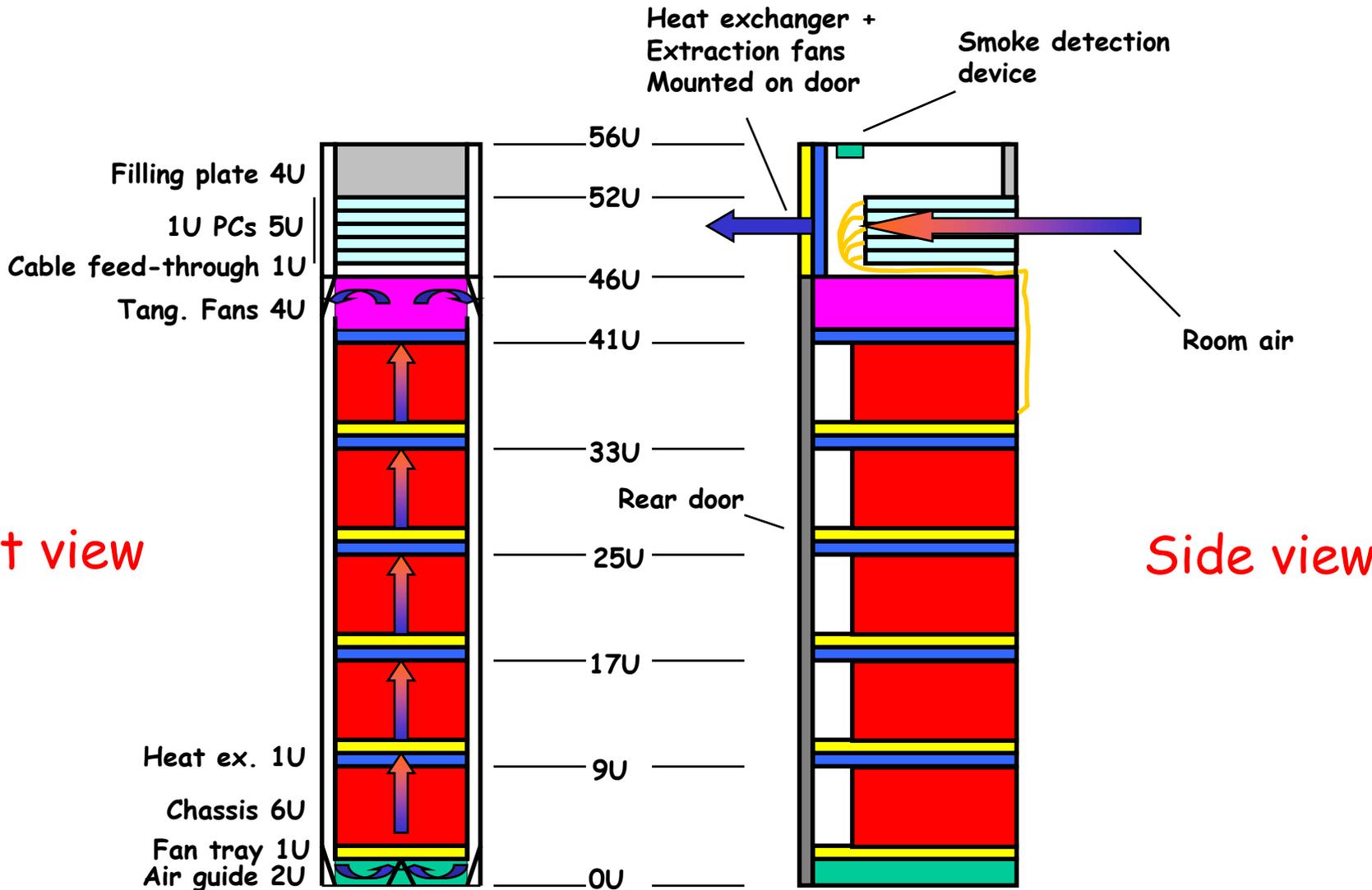
DAQ Opticable



200m long
72 fibers, 50/125um multimode, LC-LC connector
16 FRLs per cable, 30 + 2 cables for whole DAQ
(8 spare cables and 8 spare fibers per cable)



DAQ rack layout in USC



Front view

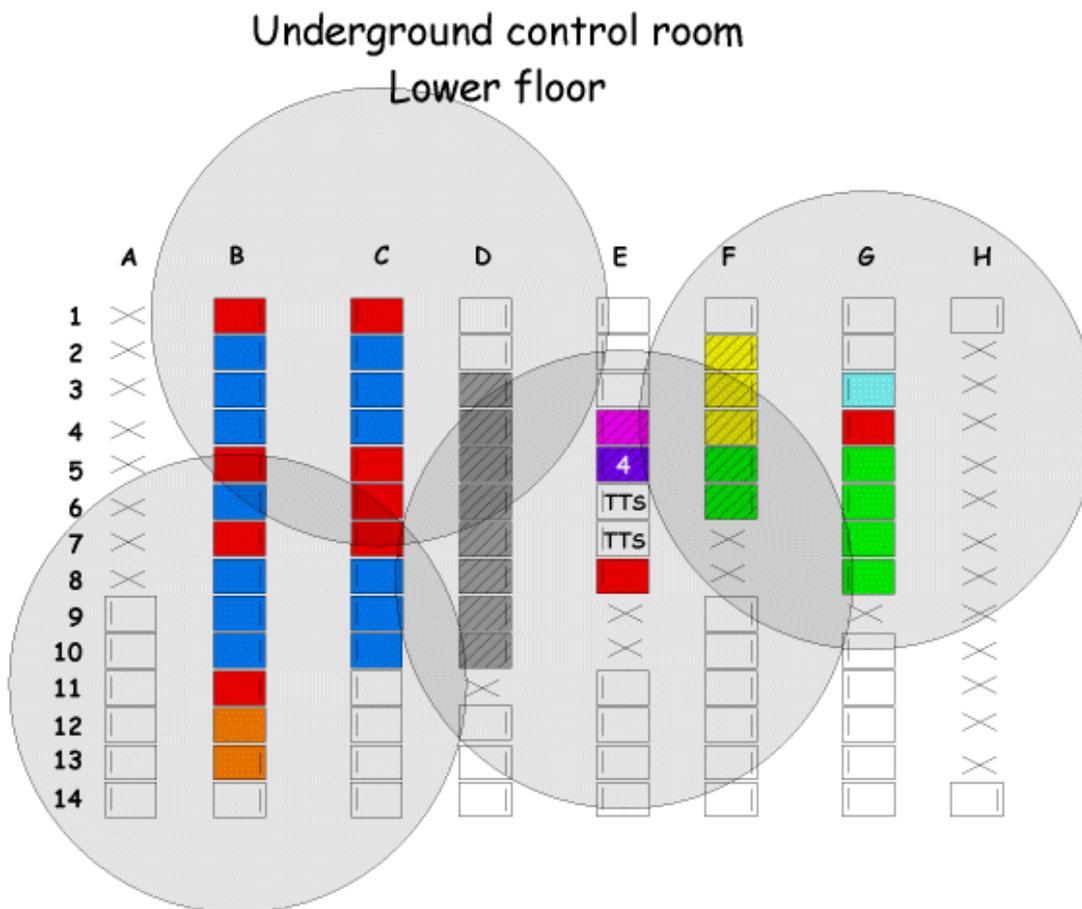
Side view



FE racks and DAQ racks (1)

Data sources

- DAQ
- Tracker
- Preshower
- CSC muons
- ▨ CSC track finder
- ▨ Global trigger
- ▨ Cal. global trigger
- RPC muons barrel
- RPC muons endcap
- ▨ RPC trigger
- DT muons
- ▨ DT track finder
- Pixels
- ✕ Unassigned rack



Last update: March 24, 2003
according to version 3.9 of official layout file

4 meter radius (10 meter cable)



FE racks and DAQ racks (2)

