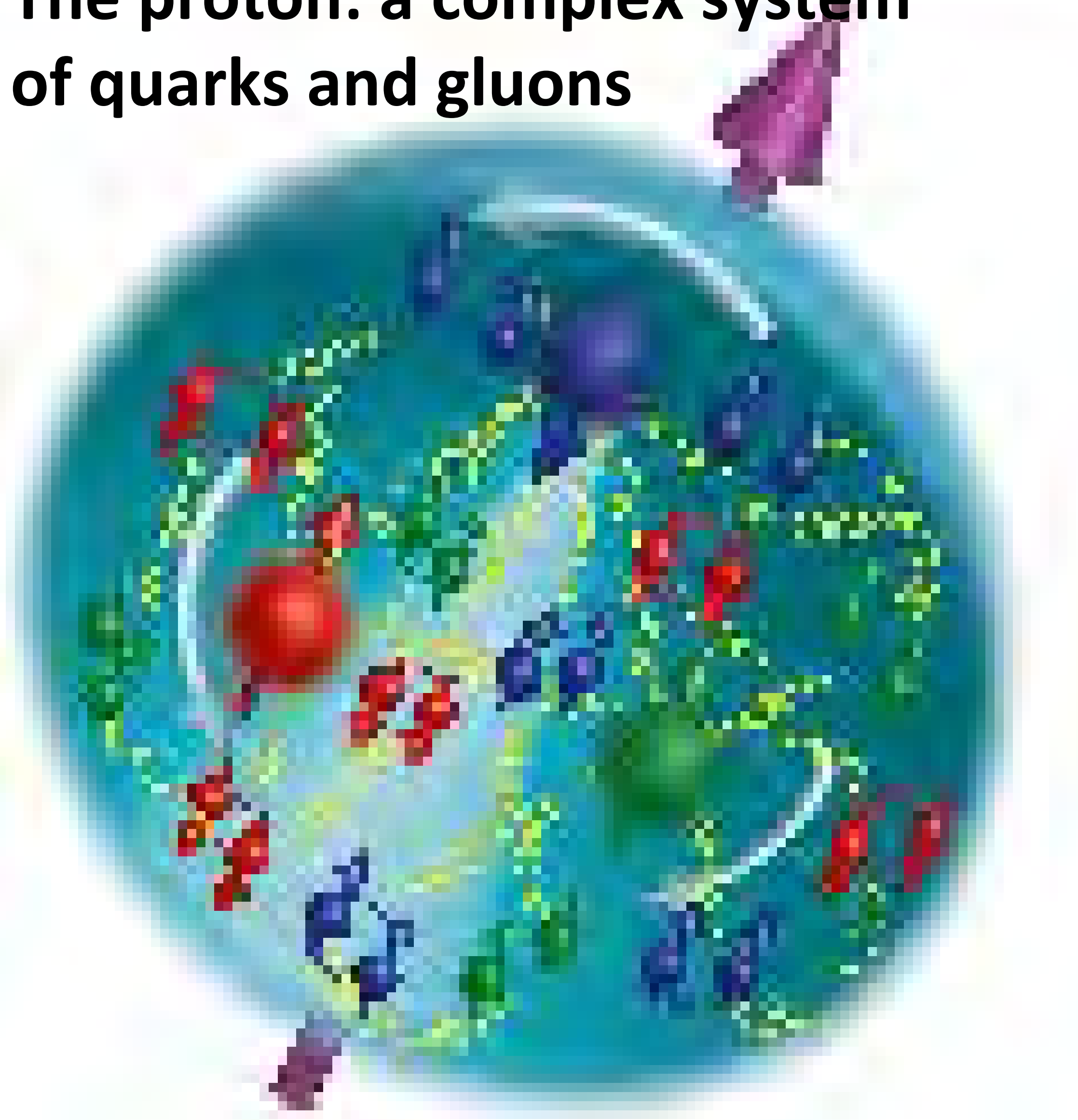
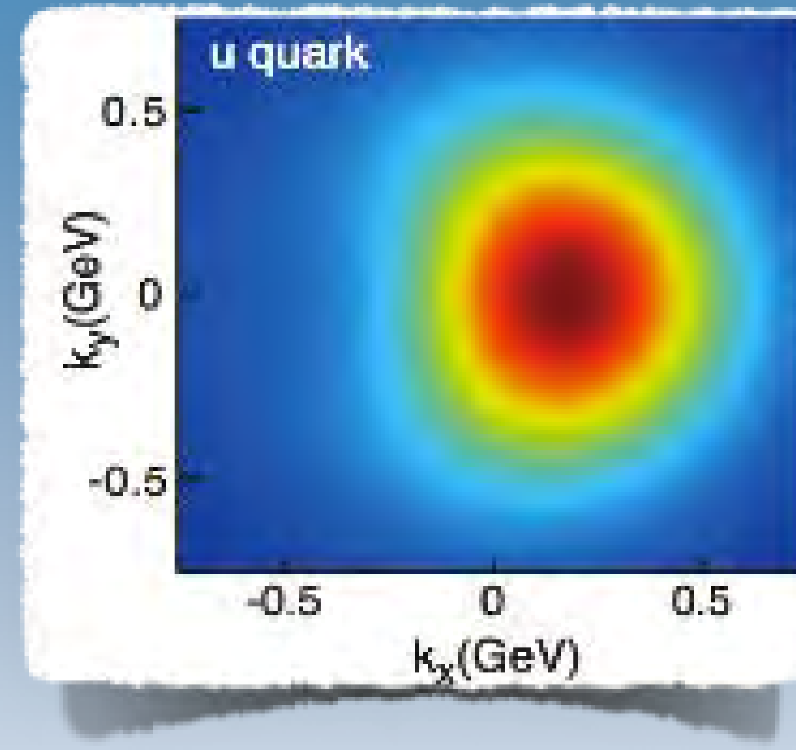
Mapping Proton Quark Structure using Petabytes of COMPASS data

The proton: a complex system of quarks and gluons





Inside the Proton: up-quark transverse momentum distribution

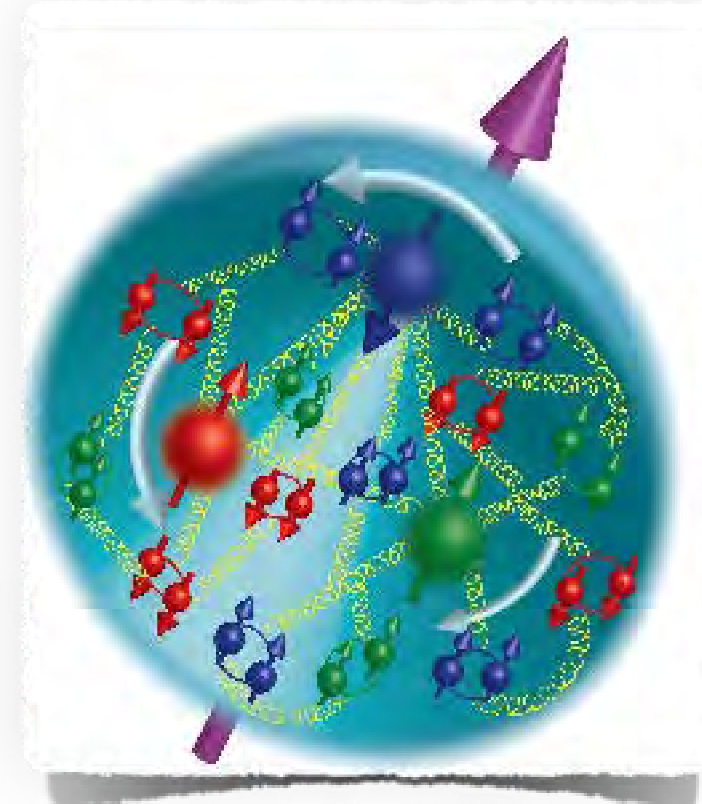
Vincent Andrieux, Robert Heitz, Riccardo Longo, Naomi Makins[2], Marco Meyer, Artem Petrosyan, Matthias Perdekamp[2,3], Caroline Riedl[1] (UIUC)

[1] PI
[2] co-PI
[3] speaker

BLUE WATERS
Symposium 2018

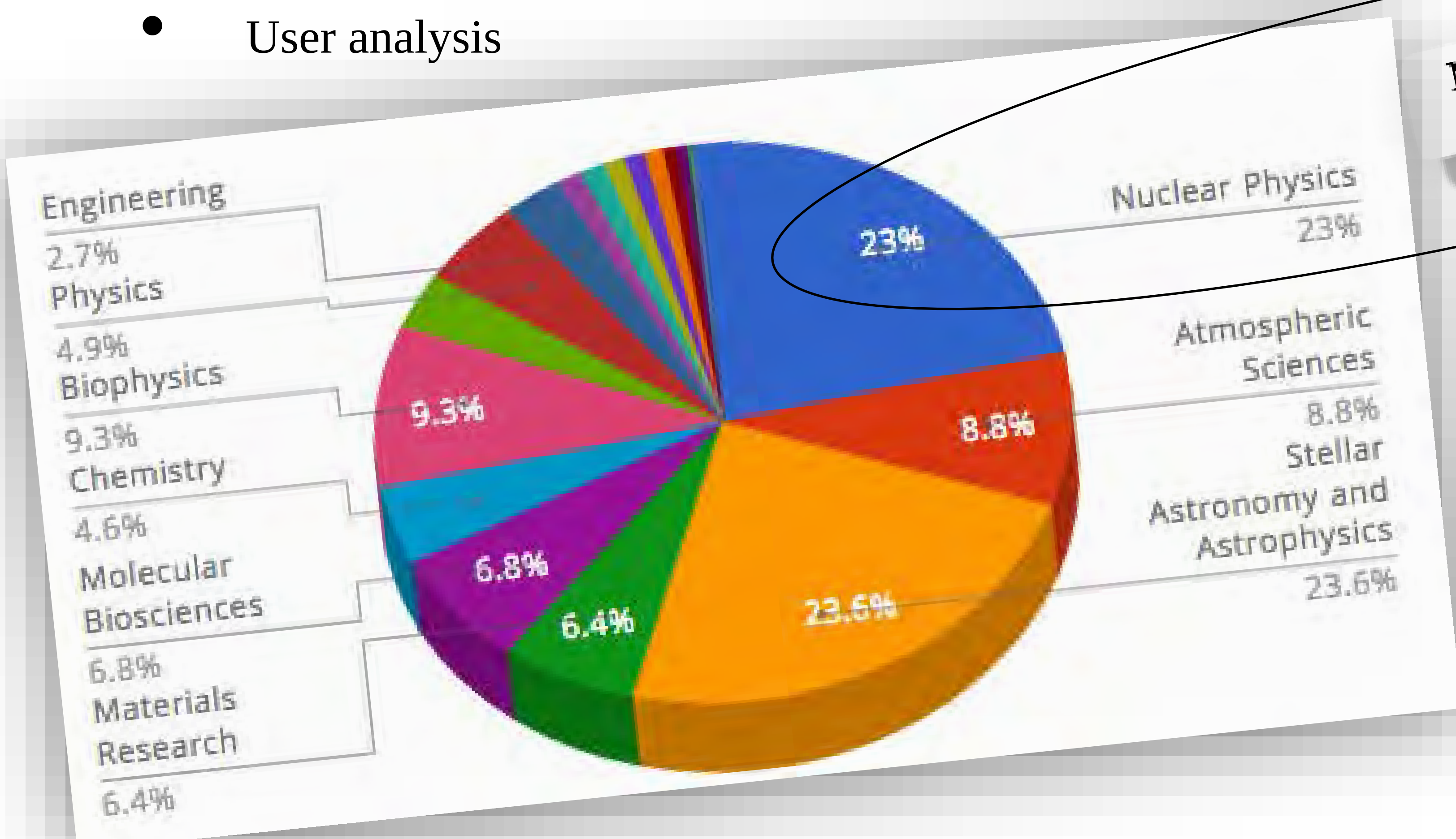
ILLINOIS

Mapping Proton Quark Structure using Petabytes of COMPASS data



- Map quark spin and transverse momentum inside the proton
  ➔ COMPASS Experiment @ CERN
- COMPASS raw data mass production with "PanDA"
- COMPASS Monte Carlo mass productions
- User analysis

NSF grant (award #1713684)
Allocation "PRAC_balh"



Mapping Proton Quark Structure using Petabytes of COMPASS data

Key Challenges

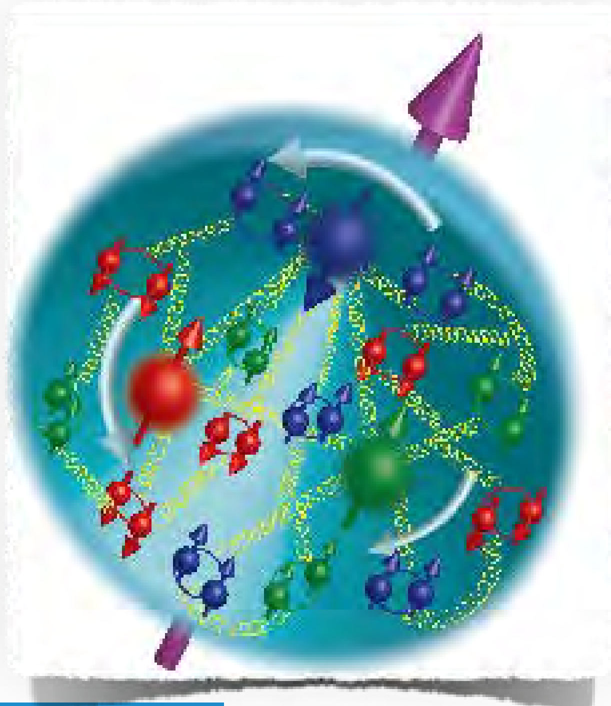
Analysis of large raw data and Monte-Carlo data samples

Why it matters

Exploration of transverse momentum- and spin-distributions of quarks in the proton

Why Blue Waters?

Blue Waters allows for **high-precision** analysis and allows for **novel studies** otherwise too CPU intensive.



Accomplishments

First MC mass productions.
First RD mass production with PanDA about to be launched.
Input to various physics analyses.

Blue Waters team contributions

Strong support in adapting CERN software and data transfer

Broader Impact & Shared Data

Education of students and young postdocs in petascale computing.
Creation of reference data productions shared with ~250 researchers.

Products

Adaption of ATLAS tool “PanDA” to COMPASS.
COMPASS-specific job templates

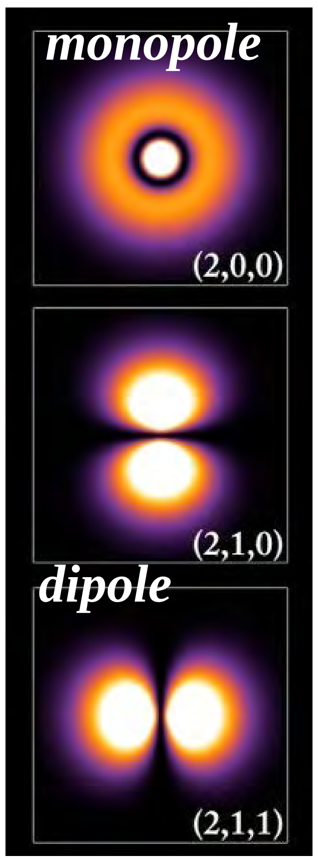
Why it matters

Spin and Angular Momentum Structure: from QED (Hydrogen) to QCD (Proton)

~ 1930's: Quantum Electro Dynamics QED

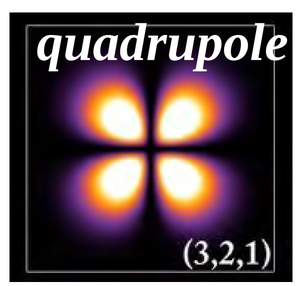
(the theory of the electromagnetic force between charged particles)

Hydrogen atom - "wave function"



$$\psi_{\mathrm{nlm}}(r,\theta,\phi)$$

orbital angular momentum ← → spin



- Understand the bound system and its excitation levels.
- Spin-orbit correlations in QCD similar to those in QED.
- "Proton (hyper)fine structure"

~ 2010's: Quantum Chromo Dynamics QCD

(the theory of the **strong nuclear force between quarks**)

with transverse degrees of freedom

proton \ quark	unpolarized U	long. polarized L →	trans. polarized T ↑
U	f_1		$h_1^\perp$
L		g_1	$h_{1L}^\perp$
T	$f_{1T}^\perp$	g_{1T}	h_1 $h_{1T}^\perp$

Sivers function

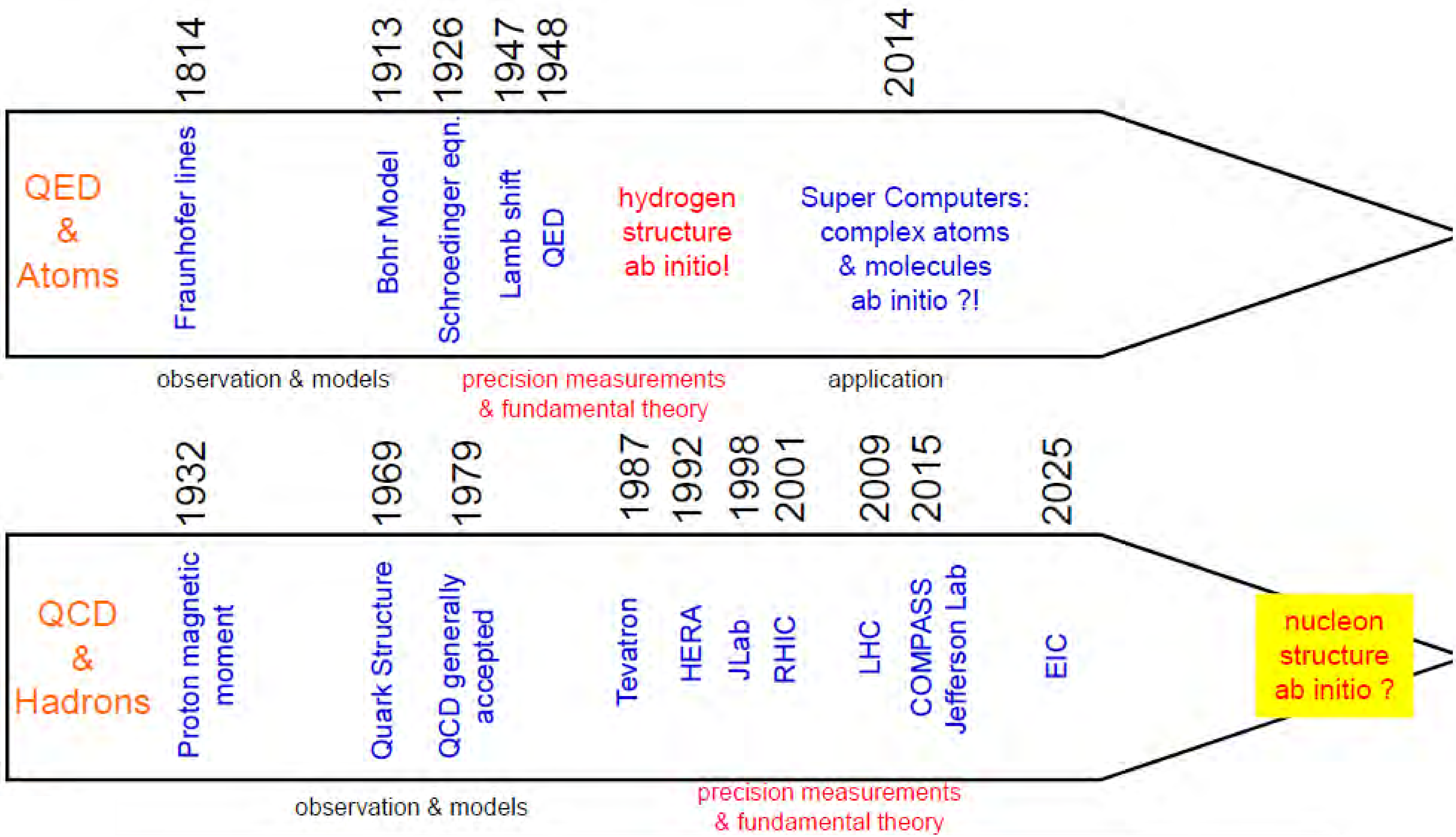
COMPASS:

Correlation between transverse proton spin and intrinsic transverse quark momentum

➔ access to orbital angular momentum?!

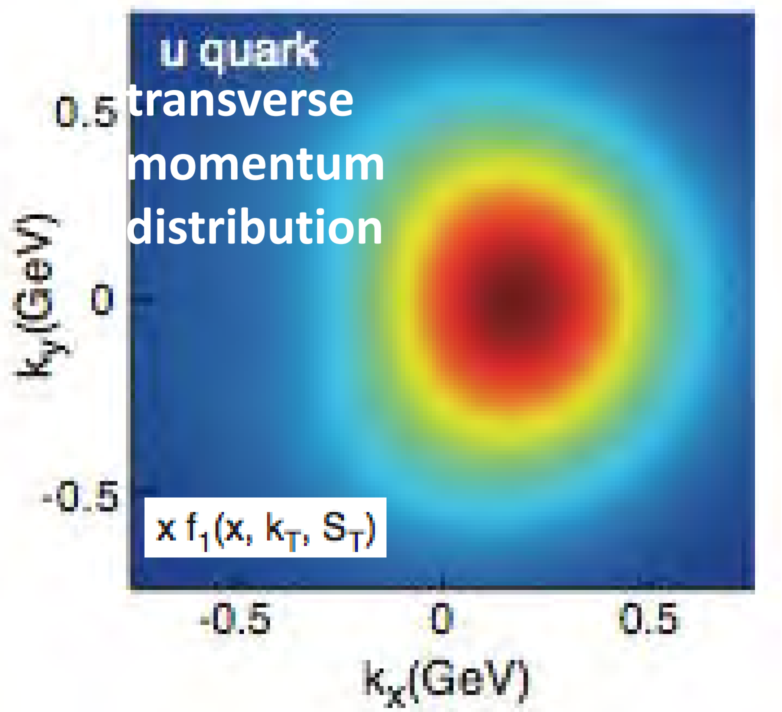
BLUE WATERS

QED vs QCD: The path from early evidence → models → ab initio calculations → applications



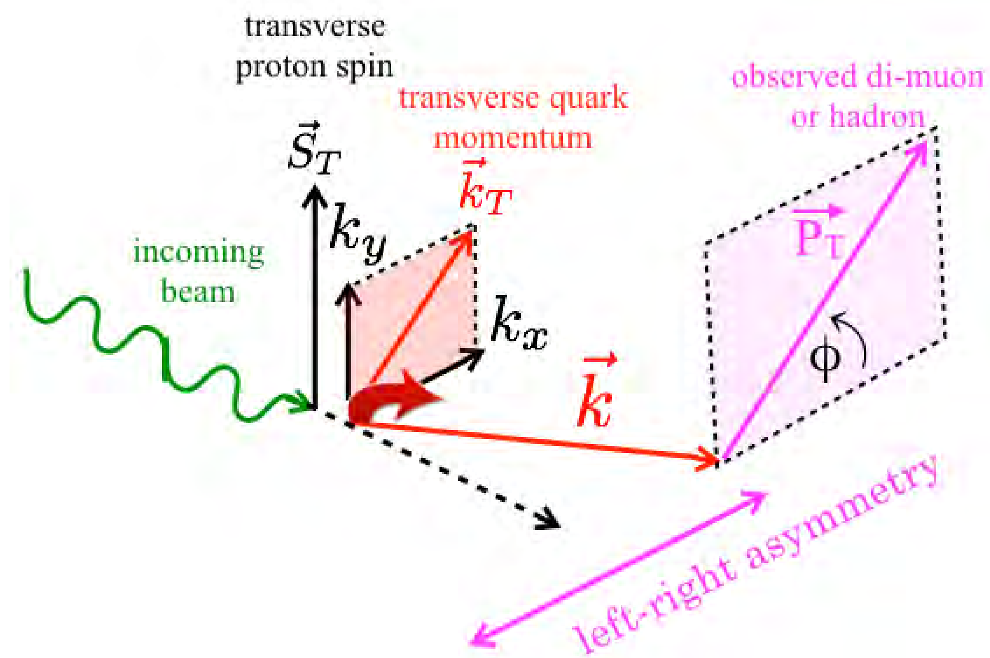
Why it matters

Sivers effect: distortion of quark transverse momentum distributions with respect to proton spin



EIC "White Paper" arXiv:1212.1701, based on M. Anselmino et al., J. Phys. Conf. Ser. 295, 012062 (2011), arXiv:1012.3565

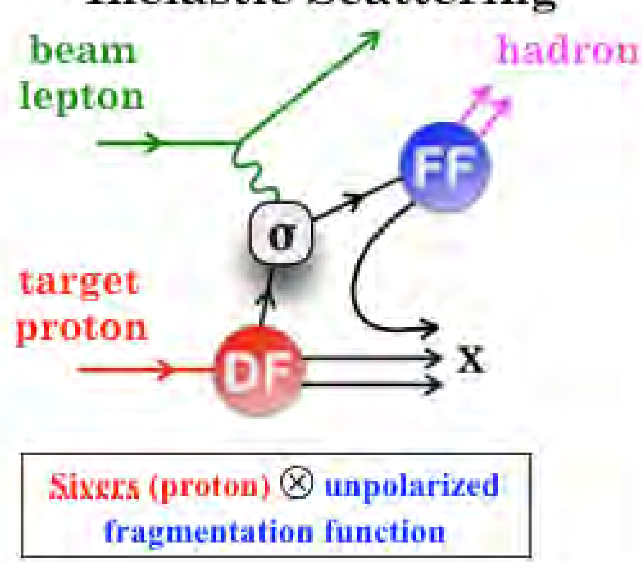
Transverse Quark Momentum Leads to angular asymmetries for final state particle distributions



$$A_{\mathrm{UT}}(\phi) = \frac{1}{f S_T} \frac{N^{\uparrow}(\phi) - N^{\downarrow}(\phi)}{N^{\uparrow}(\phi) + N^{\downarrow}(\phi)}$$

COMPASS Measurements with muon and hadron Beams:

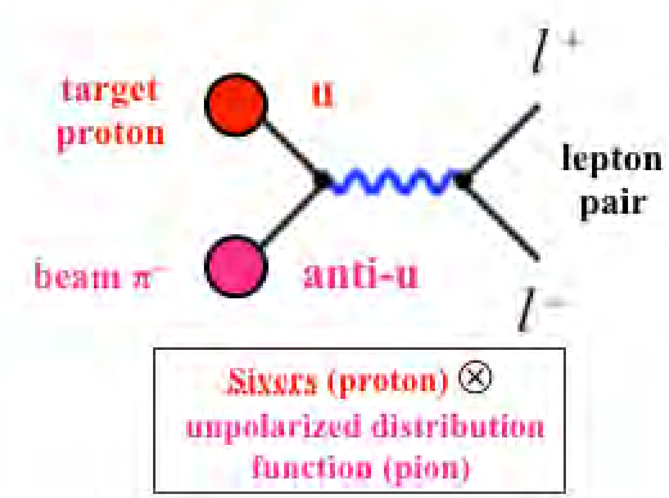
Semi-Inclusive Deep-Inelastic Scattering



The Sivers function is expected to have the **same magnitude** but **opposite sign** in DY. Crucial test of QCD-TMD framework.

Sivers (SIDIS) = −(1)·Sivers (DY)

Drell-Yan

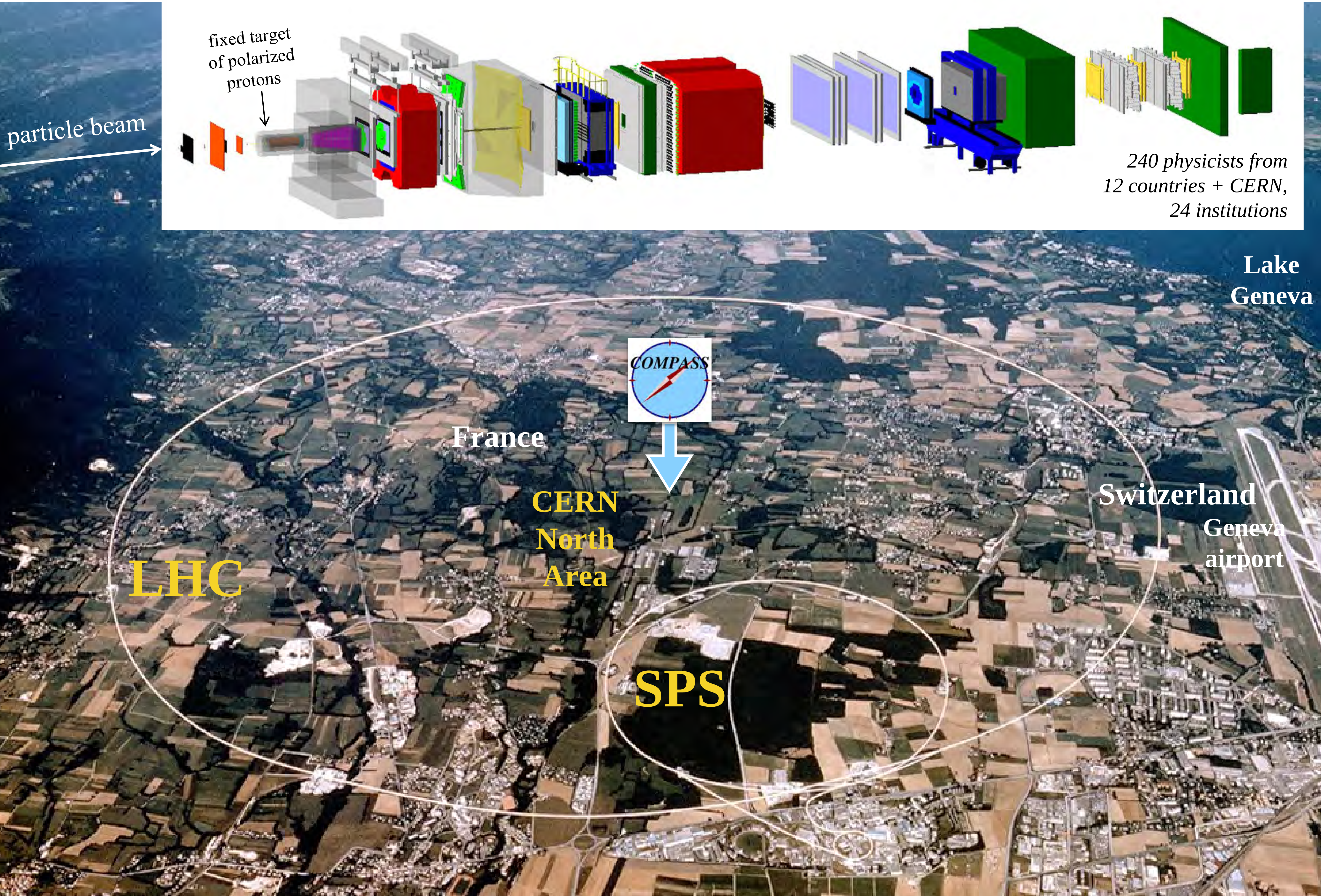


BLUE WATERS

Why it matters

COMPASS @ CERN

CERN = European Center for Nuclear Research
COMPASS = COmmon Muon Proton Apparatus for Structure and Spectroscopy



COMPASS @ CERN

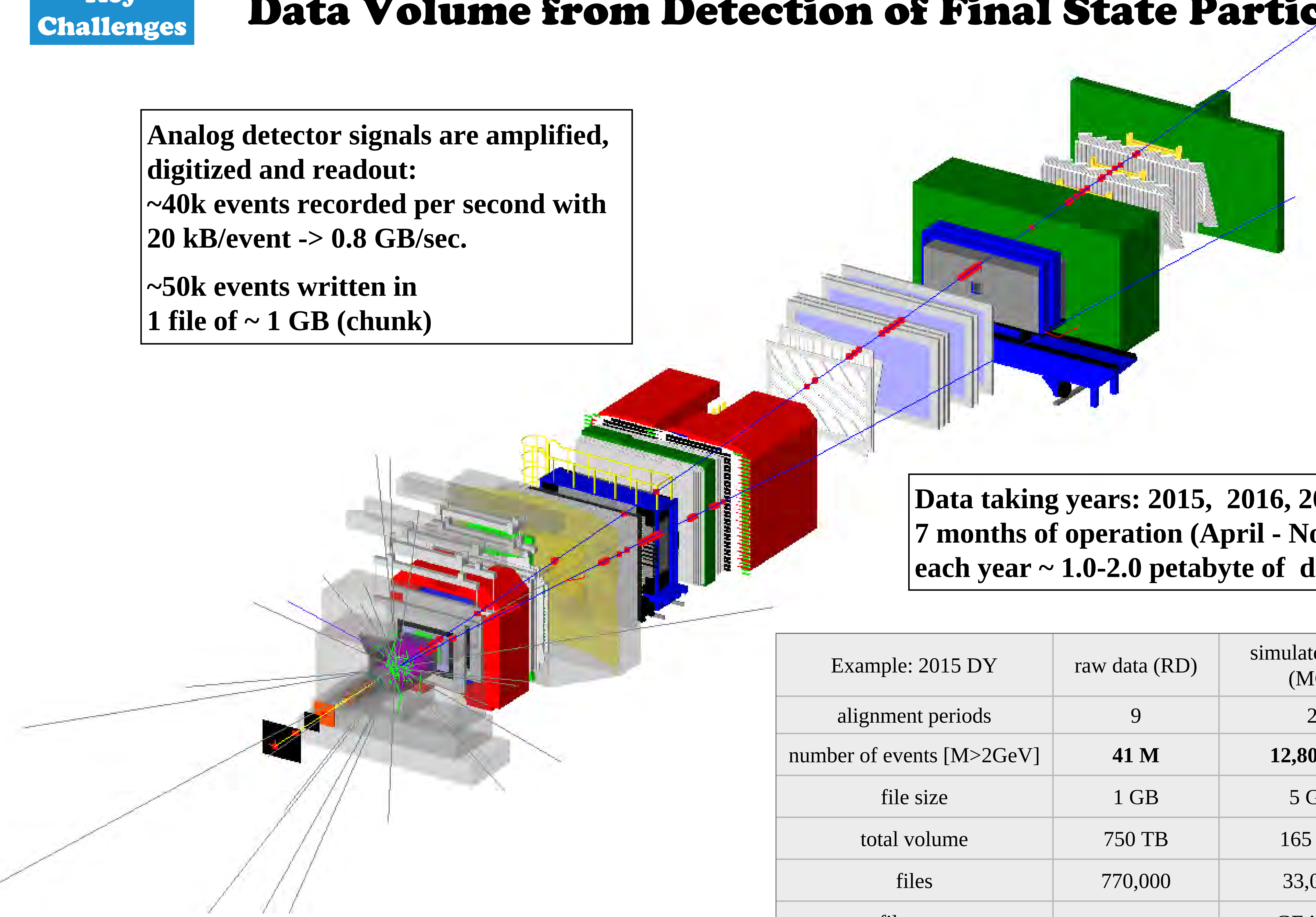
200,000 instrumentation channels readout
40,000 times every second for 220 days/year

Key Challenges

Data Volume from Detection of Final State Particles

Analog detector signals are amplified, digitized and readout:
~40k events recorded per second with 20 kB/event -> 0.8 GB/sec.

~50k events written in 1 file of ~ 1 GB (chunk)



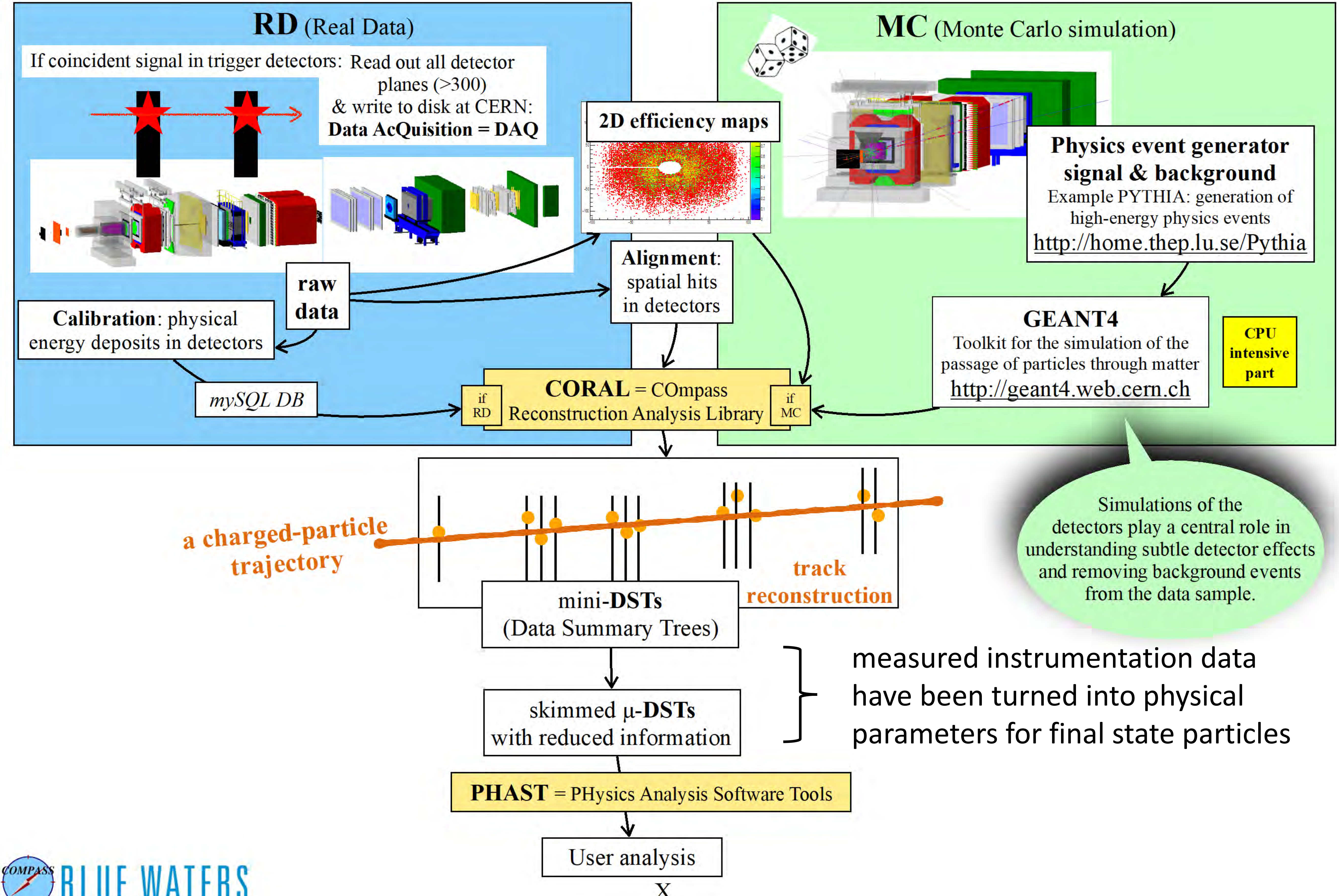
Data taking years: 2015, 2016, 2017, 2018.
7 months of operation (April - November).
each year ~ 1.0-2.0 petabyte of data.

Example: 2015 DY	raw data (RD)	simulated data (MC)
alignment periods	9	2
number of events [M>2GeV]	**41 M**	**12,800 M**
file size	1 GB	5 GB
total volume	750 TB	165 TB
files	770,000	33,000
file type	raw	GEANT

COMPASS BLUE WATERS

Key Challenges

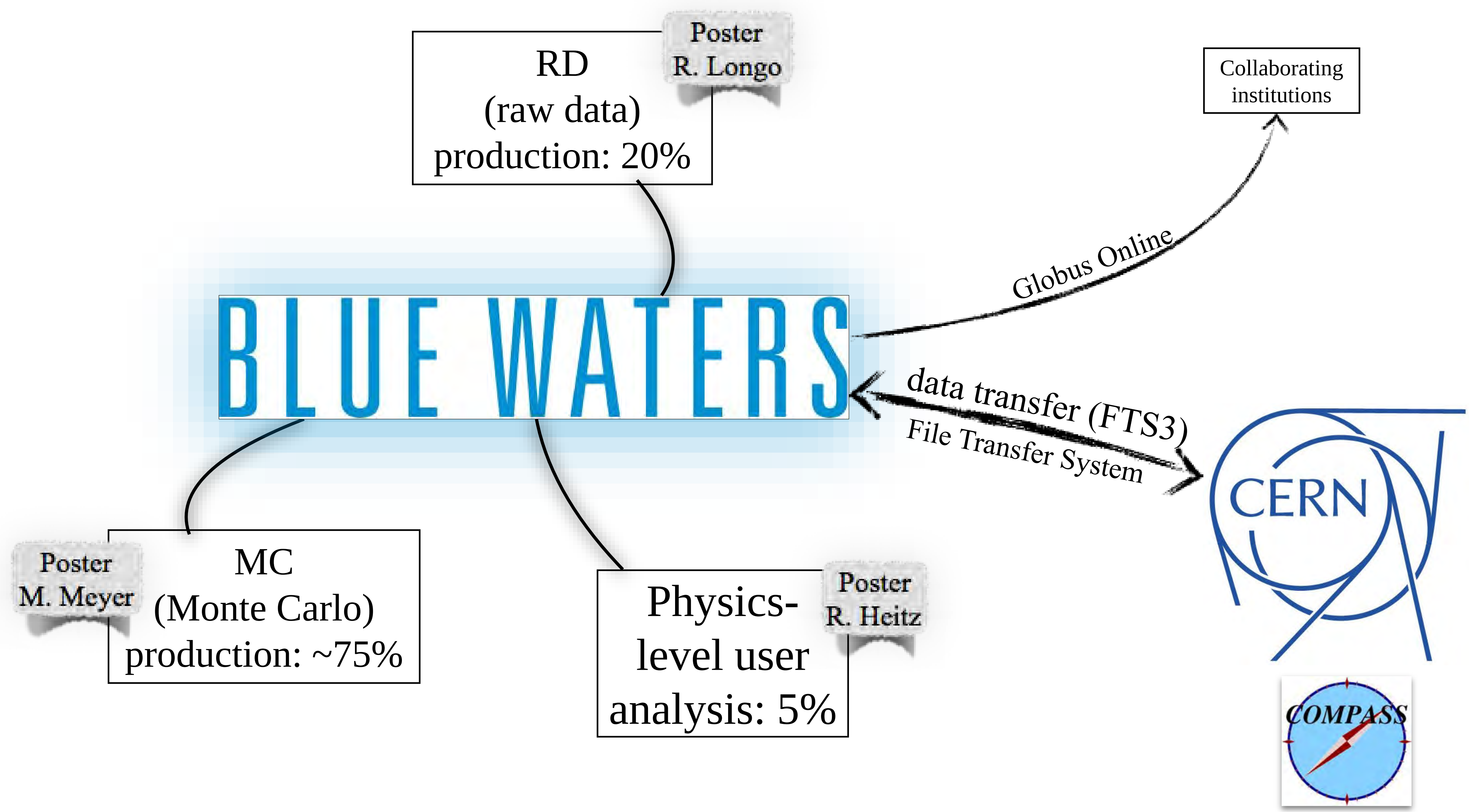
COMPASS data production - the principle



COMPASS BLUE WATERS

Key Challenges

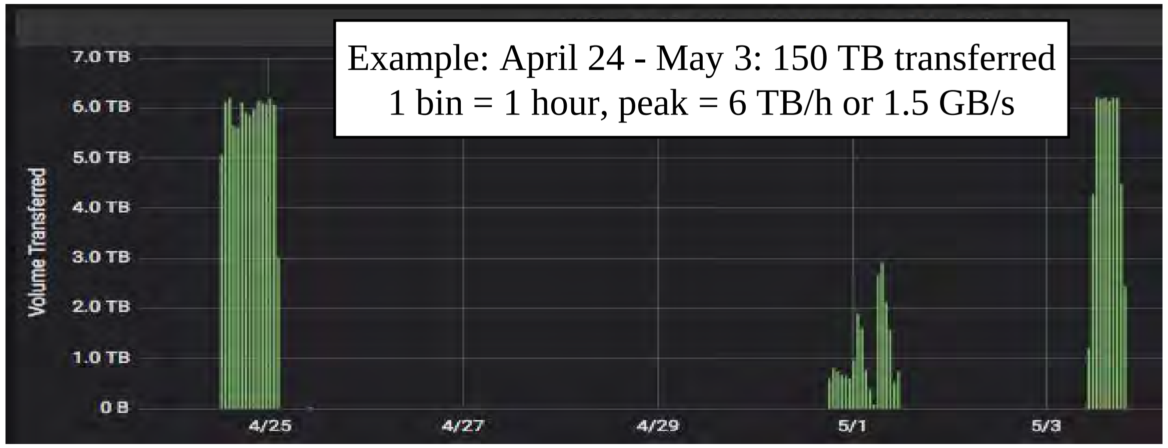
COMPASS on Blue Waters: Breakdown of Analysis Tasks



Accomplishments

Raw Data Mass Production

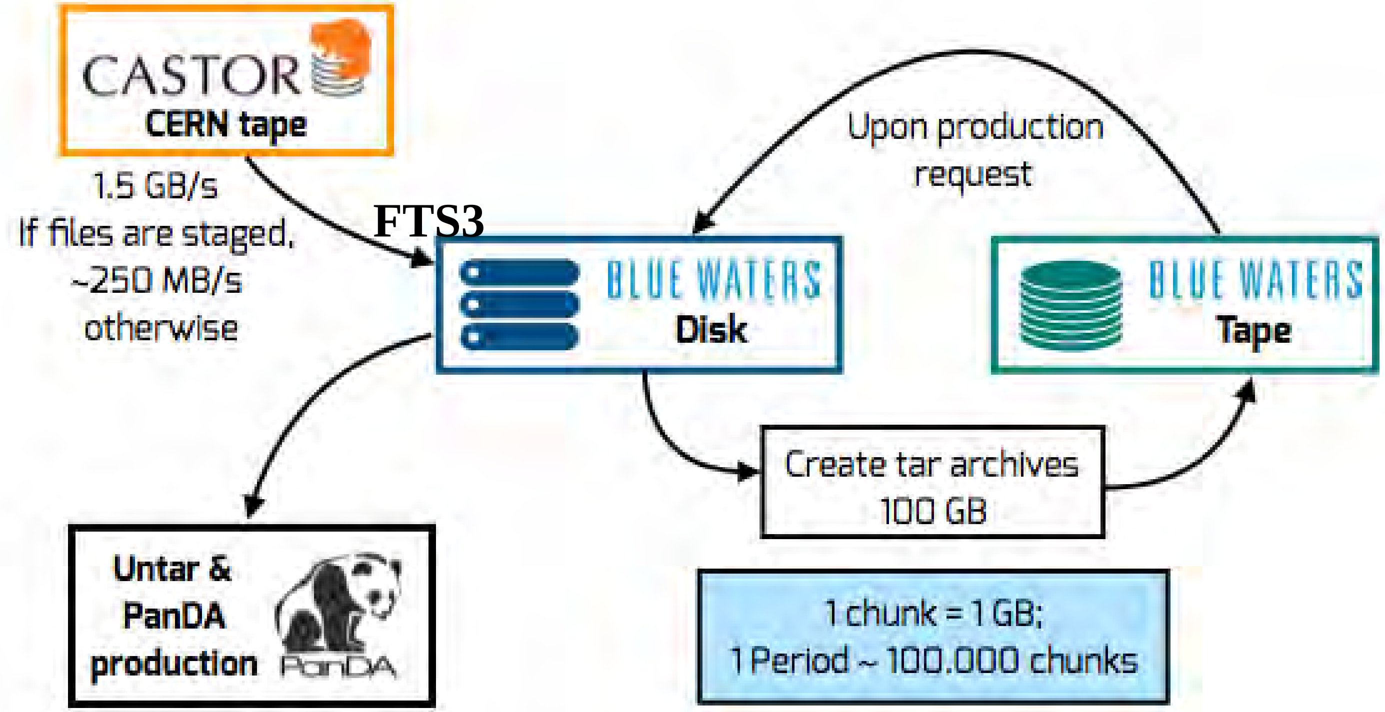
Transfer of raw COMPASS data from CERN to BW



In the past 12 months, 1 petabyte of raw data (mostly 2015) was transferred from CERN to Blue Waters and is stored on Blue Waters Nearline (tape). FTS3 effectively uses globus-url-copy

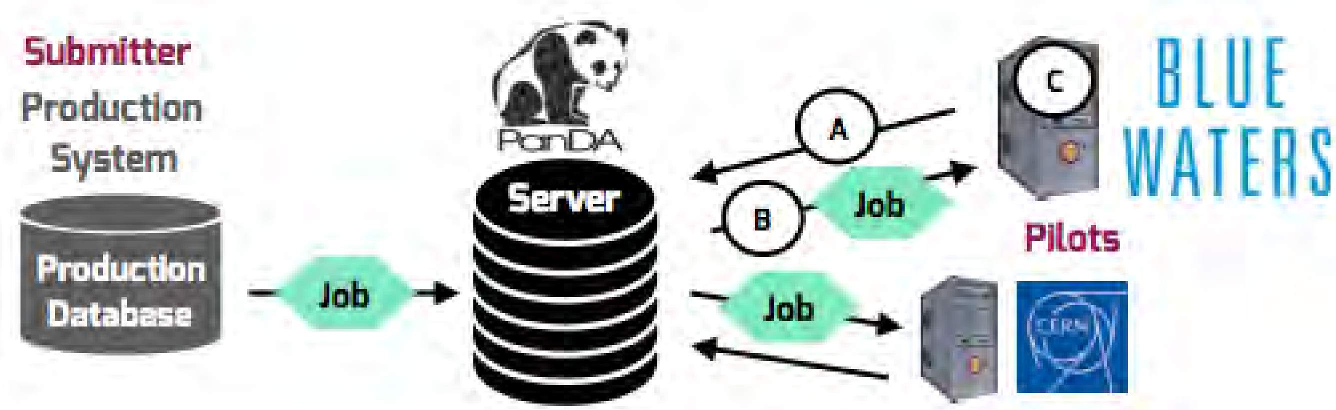
Caroline Riedl

More to come (2016, 2017, 2018)!



Artem Petrosyan & Riccardo Longo

- **PanDA = Production ANd Distributed Analysis** = data production and monitoring system developed for ATLAS-LHC.
- Artem is experienced with running PanDA at CERN and running PanDA on BW-like HPC cluster ANSELM in the Czech Republic.



Raw Data Mass Productions

Artem Petrosyan & Riccardo Longo

1 **pilot** ⇒ 1 qsub

32	32	32	32
32	32	32	32
32	32	32	32
32	32	32	32

total 512 processes
240-320 node hours
~15-20 h wall time, **16** nodes

1 submission campaign

512 processes

x 384

. . .

512 processes

total ~ 192,000 processes
90k-120k node hours
~15-20 h wall time,
~**6,000** nodes

450k-600k node hours

5 submission campaigns of each 20h wall time
= 1 year of COMPASS RD data

→ 1 year of RD data:
~ **5** days on Blue Waters
~ 50 days with CERN computing

Factor **10** in time gain

1 yellow box = 1 node with each 32 CPUs

1 blue box = 1 pilot

1 process = 1 CPU = 1 RD raw data file (chunk)
1 year of data taking ~ 1,000,000 jobs or 10 periods
100,000 jobs = 1 period

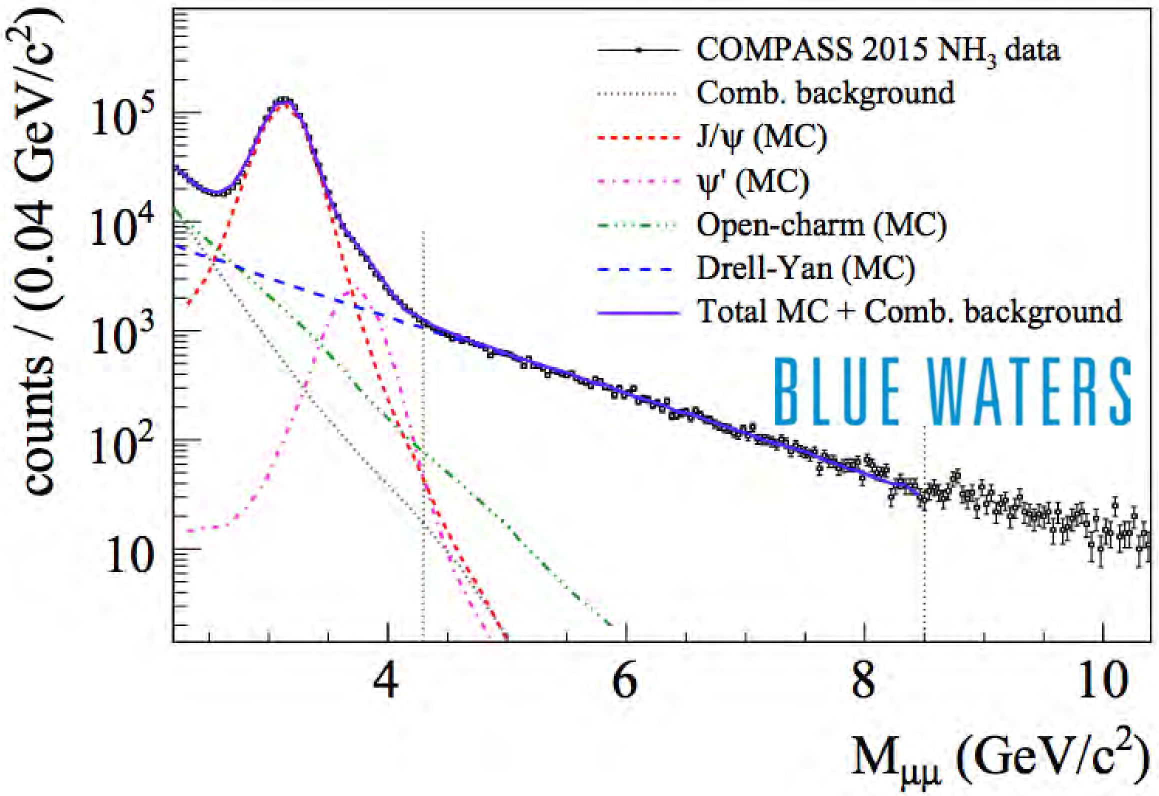
Exploratory productions in progress.
Plan to run up to 3 of these campaigns
(2015, 2016, 2017 COMPASS data
samples) by end of August

Marco Meyer

2016: tested 1/4 green box during exploratory phase (without PanDA)

Monte-Carlo Mass Productions: “DY15”

Invariant mass of 2 oppositely charged muons (= heavy electrons): Real Data vs. Monte Carlo data comparison



*This plot from **PRL** 119, 112002 (2017) contains the data from a 2017 pilot MC production on BW (not the mass production).*

Example: simulation for “Drell-Yan 2015”

MC	W12	W07
Drell Yan <4.2GeV	March 28	April 30
Drell Yan >4.2GeV	March 28	April 30
J/ψ	April 5	May 15
ψ’	April 9	May 15
Open Charm	April 5	May 15

← data periods with different detector and beam alignments

signal & background process types

The dates indicate the 2018 submission date.

Charles Naim & Marco Meyer

- 2-3 boxes are submitted to the BW grid simultaneously.
  - 1 box = 33 jobs or 100,000 node hours
  - 1 job = 100 nodes
- If jobs from 2 boxes run in parallel, they take 2x 3,300 nodes, or 1/4 of BW.
- Using #PBS -l flags=commtransparent

Example of Blue Waters usage on May 16

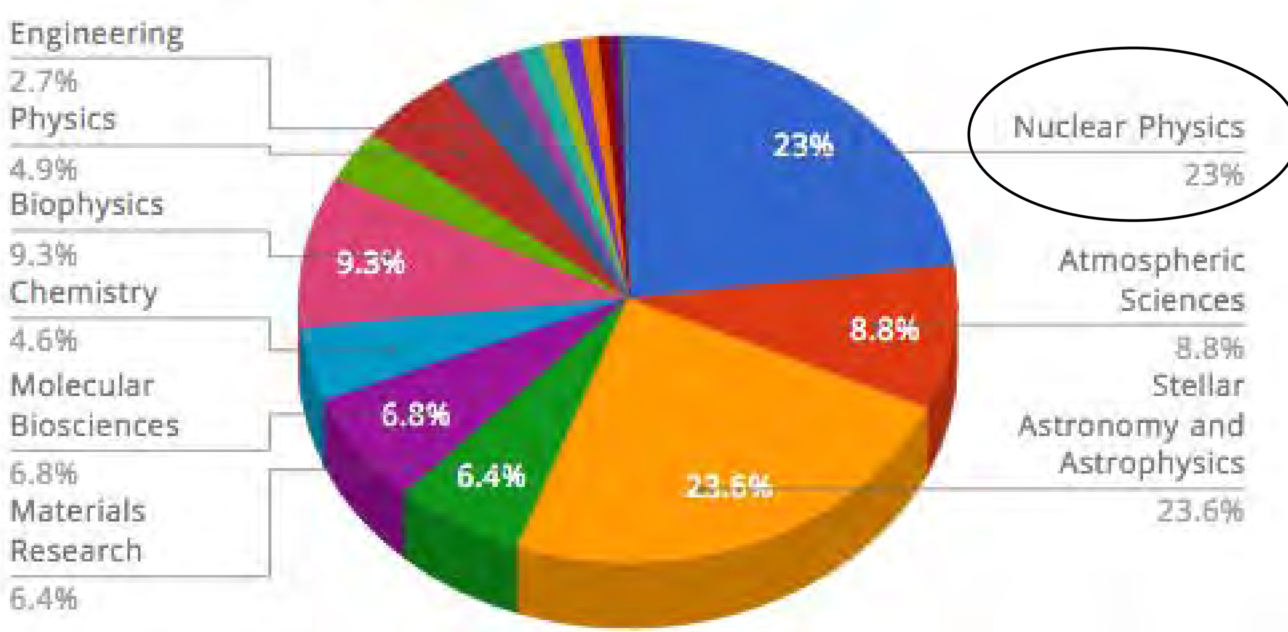
LARGEST JOBS CURRENTLY RUNNING

Jobs are Aggregated by Project

SCIENCE GROUP	NODE TYPE	CORES	CORE HOURS
Mapping Proton Quark Structure using Petabytes of COMPASS data (Jobs:50) PI: Caroline Riedl, University of Illinois at Urbana-Champaign	XE	160000	3,558,984.00
Probing New Physics in Galaxy Formation at Ultra-High Resolution (Jobs:33) PI: Philip Hopkins, California Institute of Technology	XE	55488	984,187.56
Collaborative Research: Predicting the Transient Signals from Galactic Centers: Circumbinary Disks and Tidal Disruptions around Black Holes (Jobs:3) PI: Scott Noble, University of Tulsa	XE	32320	636,325.38
Collaborative Research: Predicting the Transient Signals from Galactic Centers: Circumbinary Disks and Tidal Disruptions around Black Holes (Jobs:61) PI: Scott Noble, University of Tulsa	XK	16000	384,476.62
Collaborative Research: Predicting the Transient Signals from Galactic Centers: Circumbinary Disks and Tidal Disruptions around Black Holes (Jobs:43) PI: Scott Noble, University of Tulsa	XK	12640	219,001.37

3.56M core hours in the past 24h or 3.56M/32 =111k node hours

CURRENT RUNNING JOBS BY SCIENCE AREA



DY15 mass production of MC data: 2 of 10 parts running simultaneously

COMPASS BLUE WATERS

Accomplishments

Drell-Yan cross section

Marco Meyer & Charles Naim

Example of using DY15 MC mass production

Multi-dimensional analysis (M, xF, qT, Zvtx)

Acceptance

$$\sigma_{\mu^+\mu^-} = \frac{N_{\mu^+\mu^-}}{\mathcal{L}} \times \frac{1}{A_{cc}} \times \frac{1}{\mathcal{E}_{tot}}$$

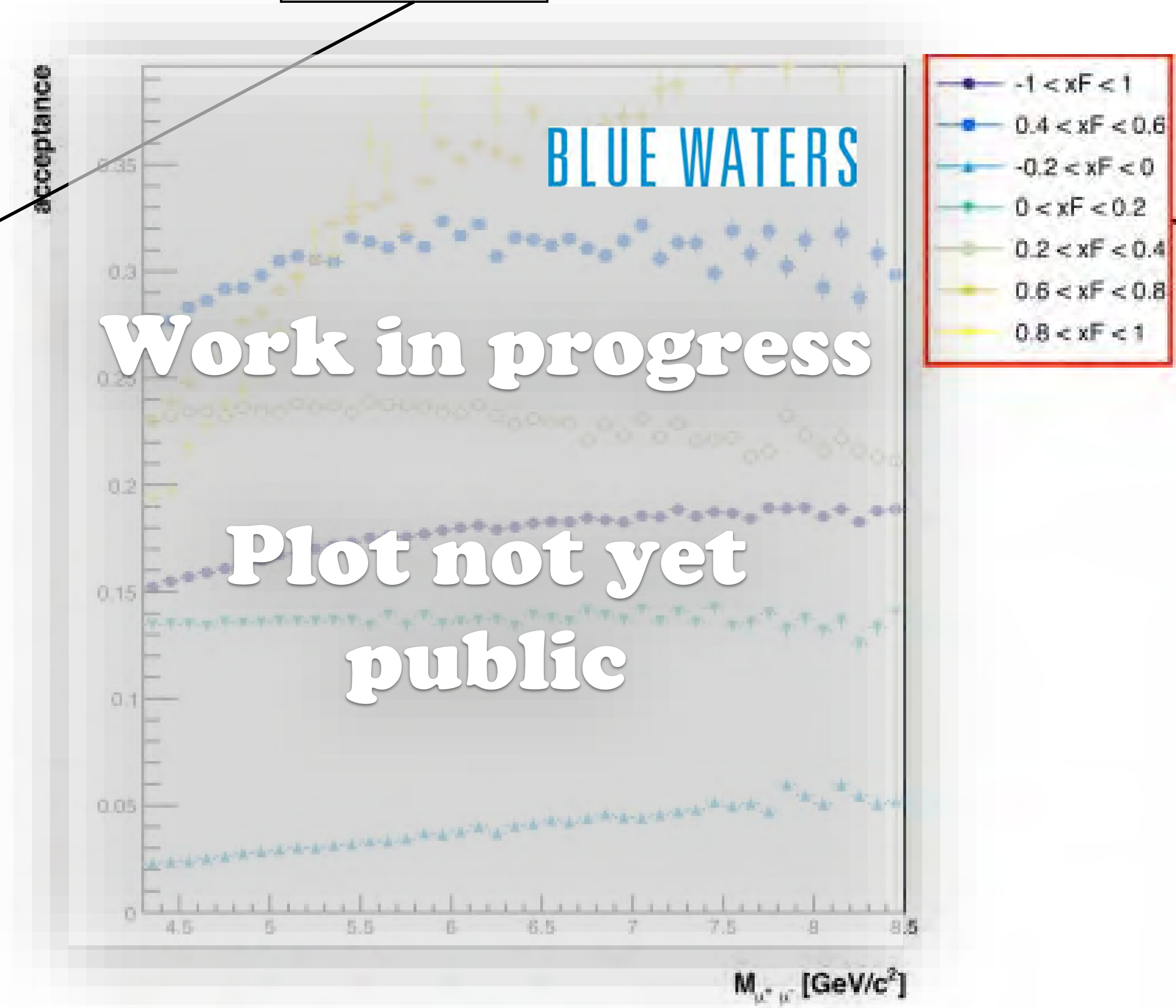
study of flux stability in real data

$$\mathcal{L} = \mathcal{F} \times \varrho_T$$

Luminosity

(trigger, reconstruction, detector) Efficiencies

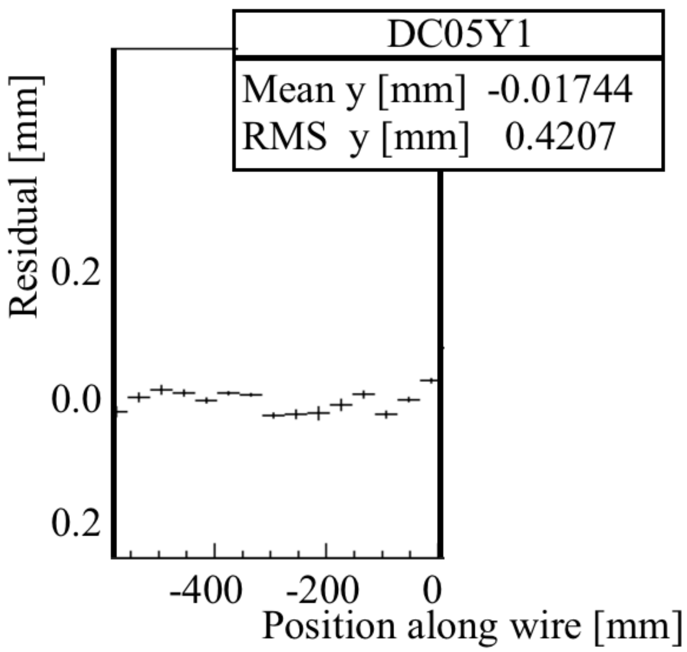
Integrated cross section



Accomplishments

Tuning and analysis of Drell-Yan data

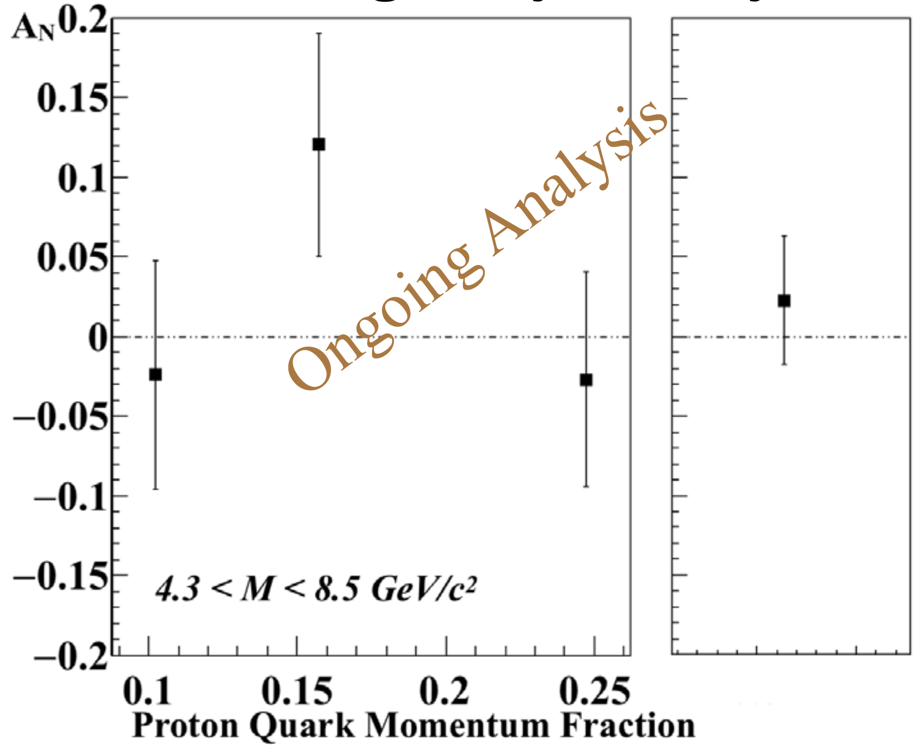
Precision alignment



- Iterative procedure to minimize a χ^2 (track+detector parameters)
- Determines relative position of >300 detector planes

Robert Heitz

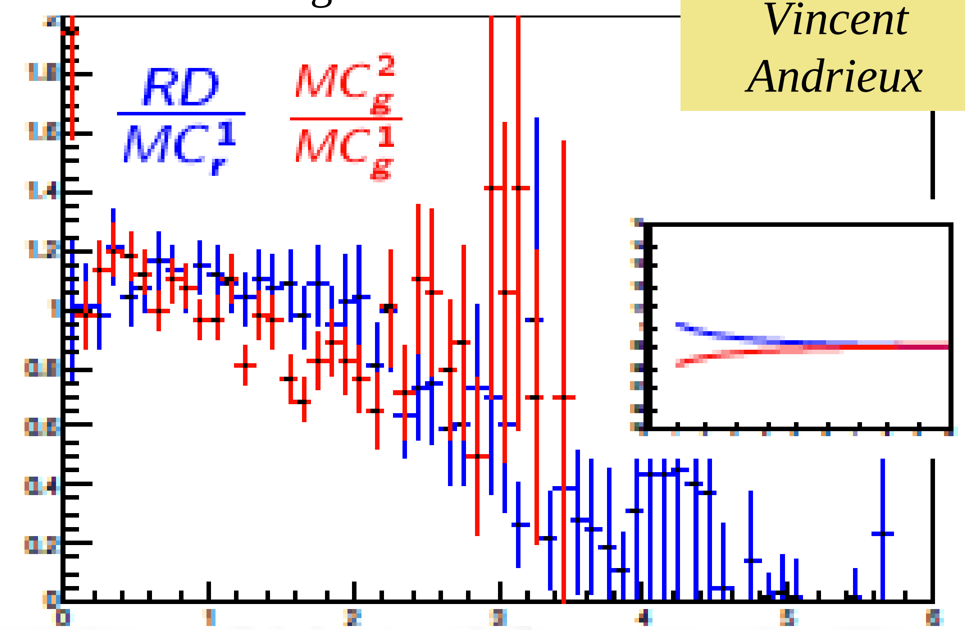
Left-right asymmetry



"kT tuning"

with MC generator PYTHIA

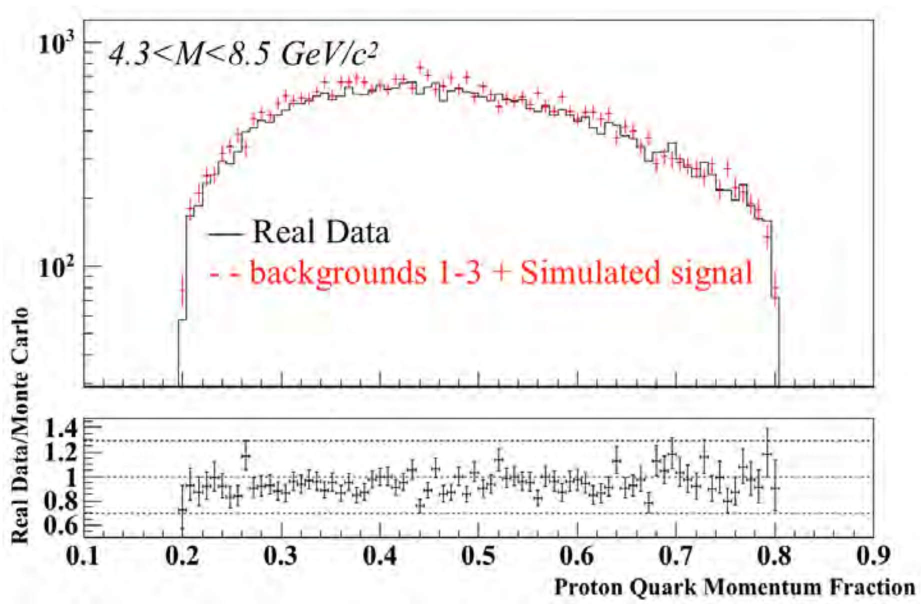
Vincent Andrieux



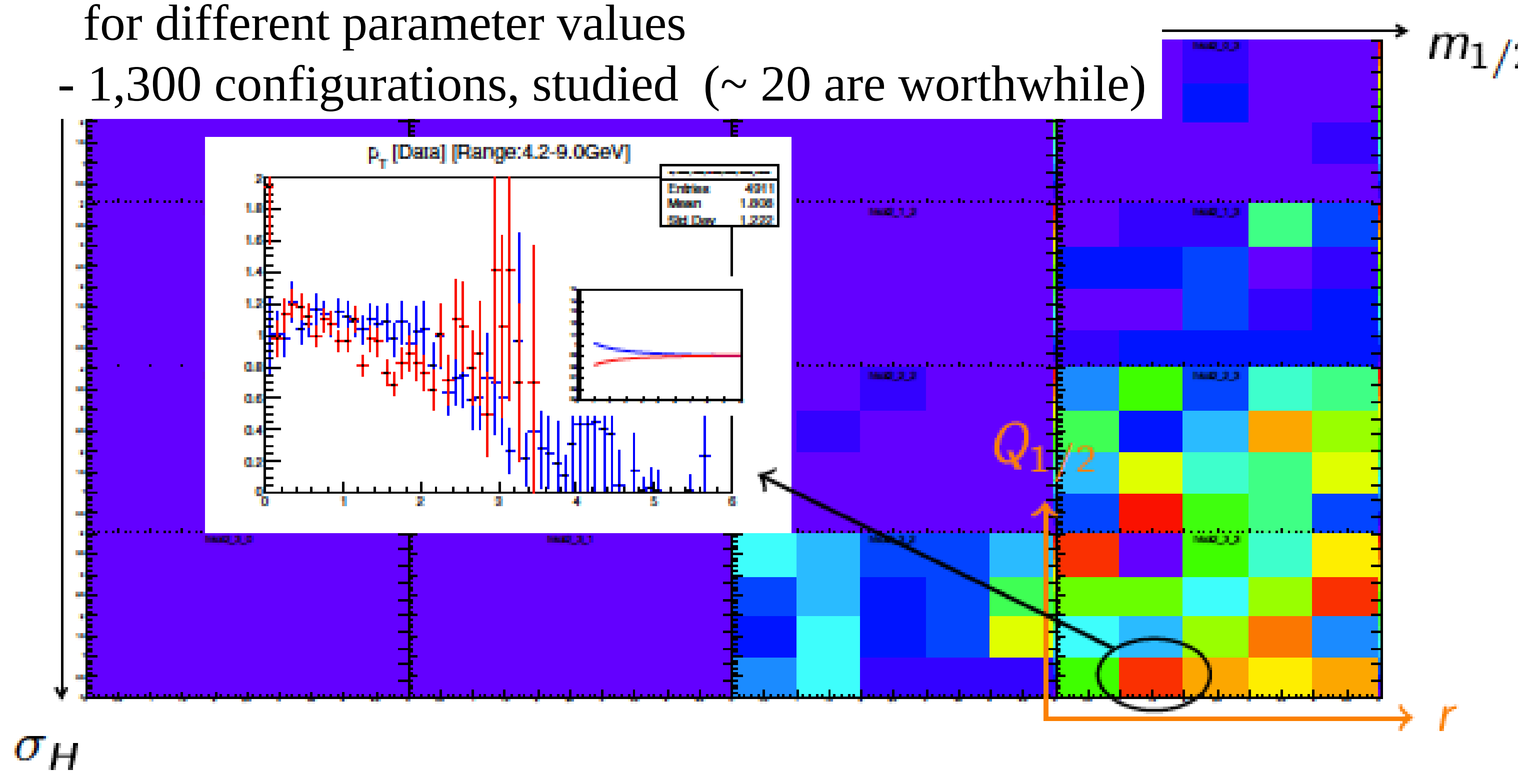
$$\sigma_{k_T}(Q) = \frac{\sigma_S Q_{1/2} + \sigma_H Q}{Q_{1/2} + Q} \times \frac{m}{m + m_{1/2}\left(\frac{E}{m}\right)^r}$$

RD vs. MC comparisons

to understand background contributions



- Kolmogorov test probability of compatibility, for different parameter values
- 1,300 configurations, studied (~ 20 are worthwhile)



COMPASS BLUE WATERS

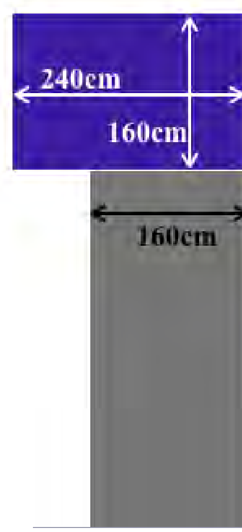
Accomplishments

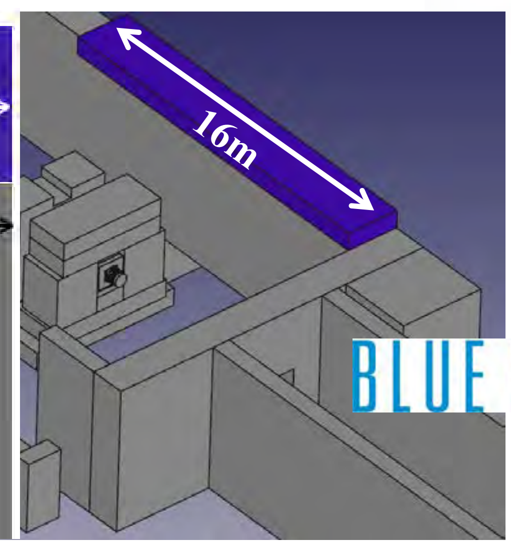
Radio Protection Simulations for COMPASS 2018

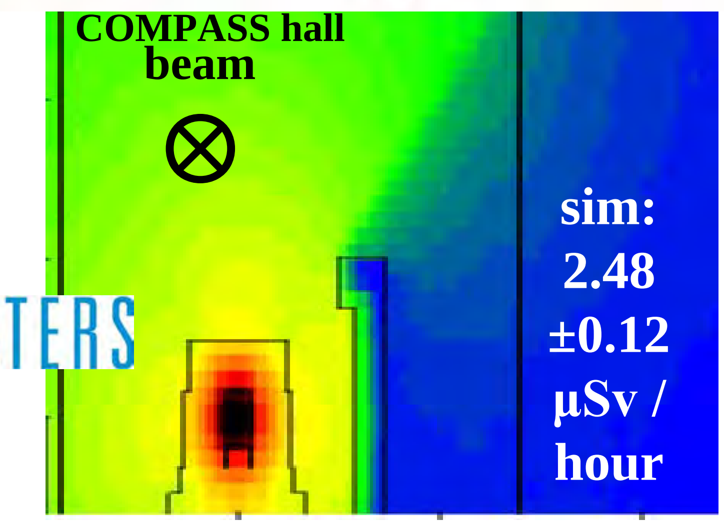
- **Improvement of shielding for better radio protection** at intensity 100 million pions per second on the COMPASS target.
- Simulation with **FLUKA**, a general-purpose tool for calculations of particle transport and interactions with matter. http://www.fluka.org/fluka.php

Angelo Maggiora

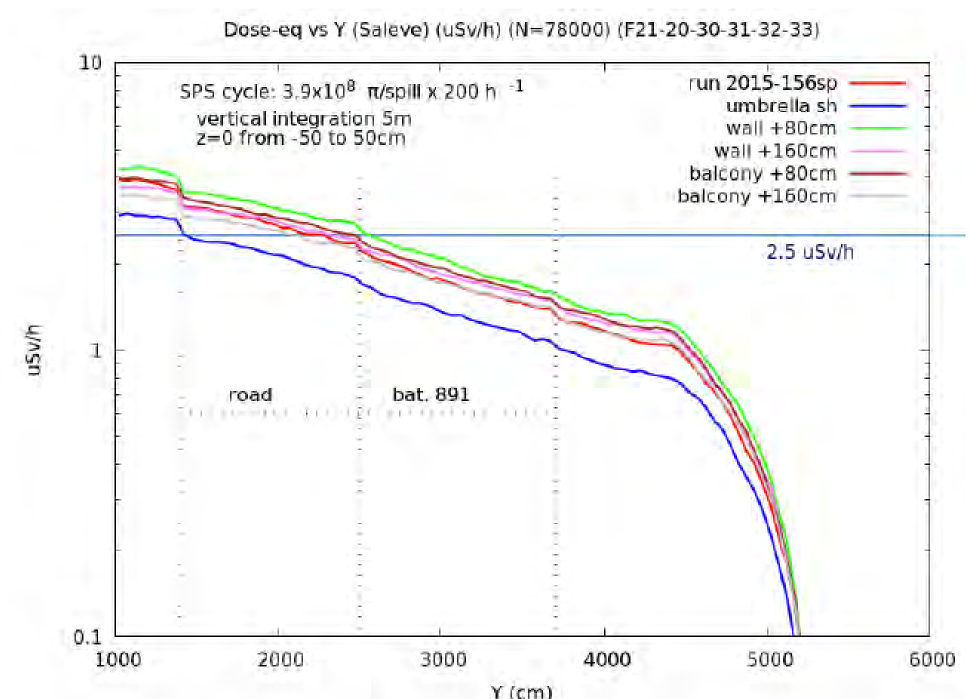
New "balcony" concrete shielding installed at COMPASS after simulations on Blue Waters







- 3M of primary particles reconstructed vs. 0.04M with previously available computing resources - **factor of 75 more**
- This statistical power results in sufficiently small error bars to reveal also small differences between the different configurations of radiation shielding with concrete and polyethylene.



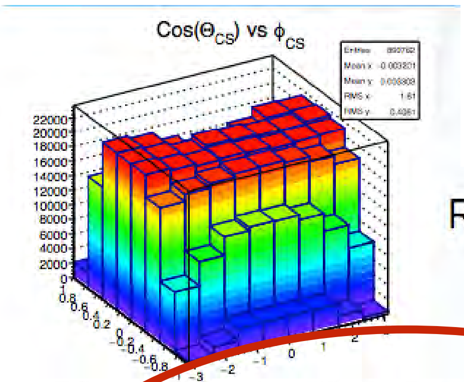
Why it matters

Example of using DY15 MC mass production

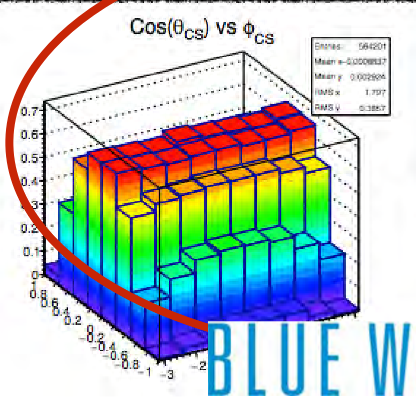
Harmonic decomposition of Drell-Yan cross section (polarization independent)

Riccardo Longo

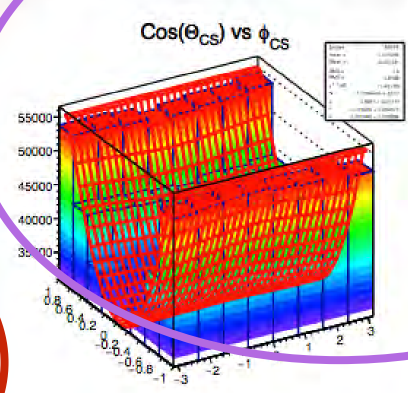
$$\frac{d\sigma}{d\Omega} \propto 1 + \lambda \cos^2\theta + \mu \sin(2\theta)\cos\phi + \frac{\nu}{2}\sin^2\theta\cos(2\phi)$$



Reconstructed data

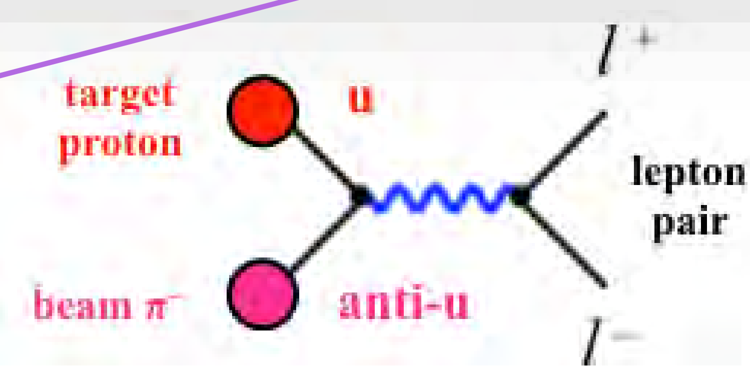


Acceptance extracted from MC



Acceptance-corrected distribution

$$f = N\left[1 + \lambda\cos^2\theta + \mu\sin 2\theta\cos\varphi_{CS} + \frac{\nu}{2}\sin^2\theta\cos 2\varphi_{CS}\right]$$



(θ, Φ)

Lam-Tung sum rule, (Collinear LO pQCD)

(Phys.Rev. D18 (1978) 2447)

$$1 - \lambda - 2\nu = 0 \longrightarrow \lambda = 1, \quad \mu = 0, \quad \nu = 0$$

Intrinsic transverse motion + QCD radiative effects

$$\lambda \neq 1, \quad \mu \neq 0, \quad \nu \neq 0, \quad \text{but} \quad 1 - \lambda - 2\nu = 0$$

Experimental observation: large ν and violation of Lam-Tung relation

NA10, Z.Phys.C 31, 513 (1986) CMS, PLB 750 (2015)

Non-perturbative effects ? $\nu \propto h_{1,\pi}^{\perp q} \otimes h_{1,p}^{\perp q}$

QCD radiative effects? NLO description?

Accomplishments

λ, μ, ν : improvement by Blue Waters

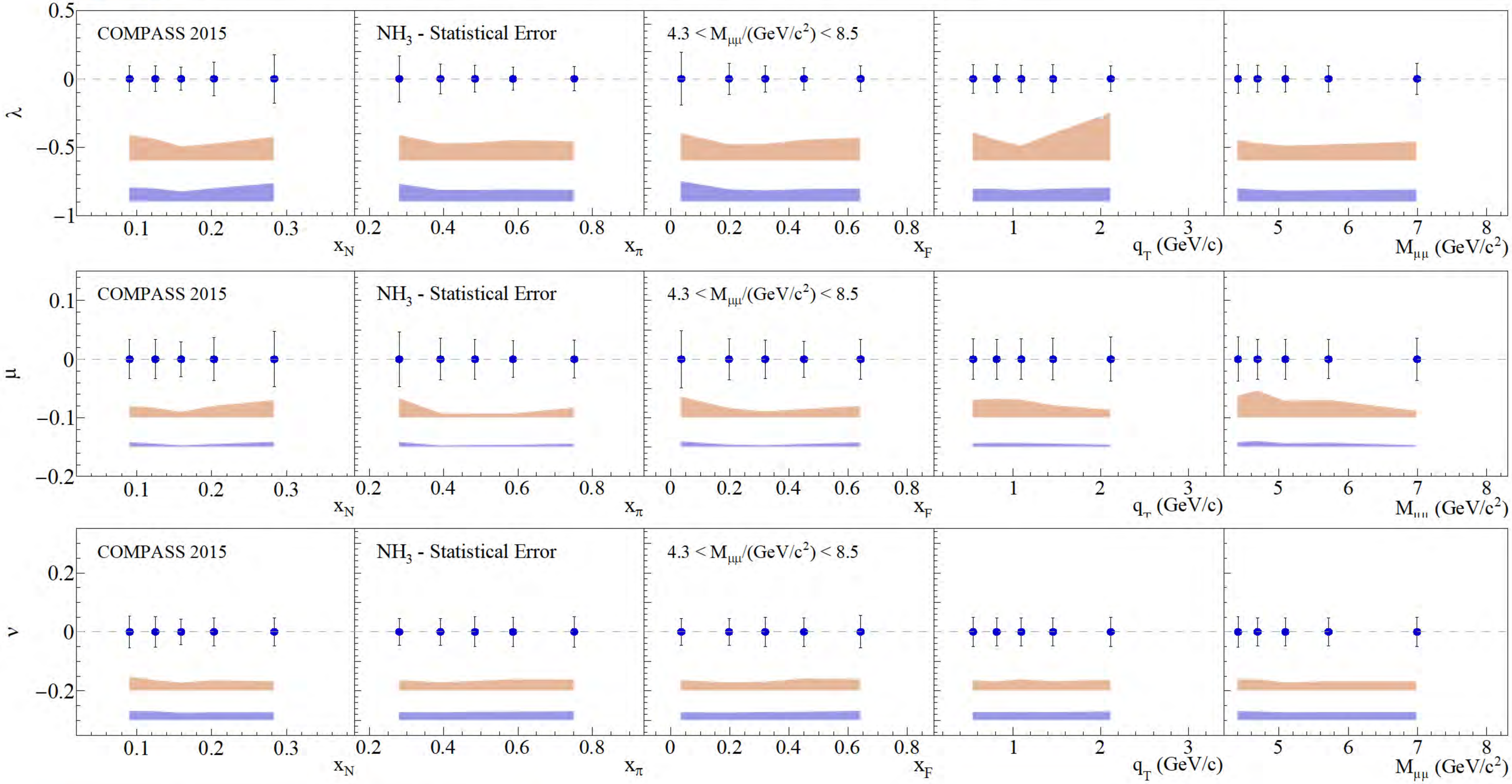
Riccardo Longo

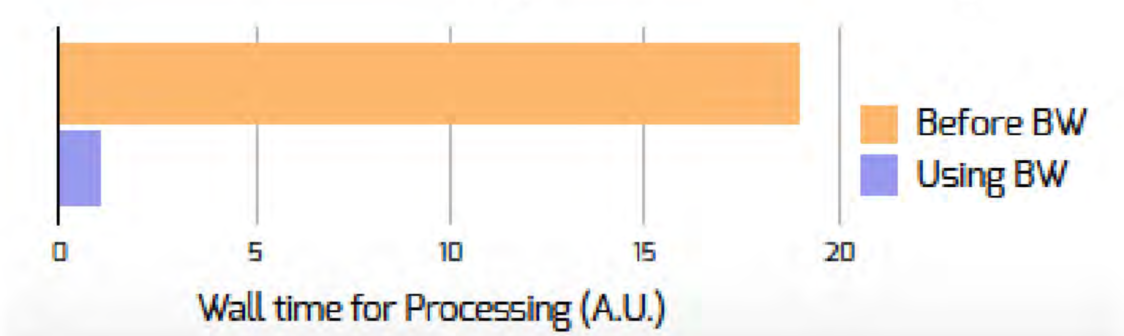
Before BW

Using BW

Results not public, therefore set to 0

Only improvement in systematic uncertainty is shown





	Blue Waters	Lyon + Torino + Taipei Farms
Nodes / CPUs used at the same time	600 nodes 18000 CPUs (running 30 jobs per core + pcp + 1 spare)	Full quota: Lyon 700 + Torino 150 + Taipei 200; Assumption: 700 CPU running at the same time in average (optimistic)
Processing time per sample (2000 events per job, assuming Lyon and Torino processing time is ~ 0.75 BW time)	16.5 hours/sample (gen+rec) 1.5 hours/sample (only rec)	320 hours/sample (gen+rec) 29 hours/sample (only rec)
Total time	~ 90 hours (—> ~ 4 days)	~ 1700 hours (—> ~ 90 days)

COMPASS BLUE WATERS

X

Why Blue Waters?

"The computing resources at Blue Waters are remarkable. For the same amount of MC, I need about 5 days at the previous computing resources, now I just need 2 hours at BW!" (Yu Shiang. L, PhD student, Academia Sinica, Taipei)

"Before Blue Waters I was playing with stones to make fire, now I have my own stove." *(Riccardo L., postdoc at UIUC)*

"Without Blue Waters it would have been impossible to optimize the COMPASS setup for run 2018" (Angelo M., senior researcher at INFN Torino)

- Blue Waters is ~ 10-25 times faster than CERN computing cluster or other computing resources available to COMPASS

o Generate MC samples needed for detailed studies of instrumentation effects on angular correlations

➔ Important for studies of spin-transverse momentum correlations

➔ Reduction of systematic uncertainties to or below the level of statistical errors

- Simulation of realistic pile-up (~100ns) - *increases CPU time by factor of 3*
- Exploration of phase space at very fine granularity: tuning of event generator
- Comparison of different generators (e.g. PYTHIA 6 vs. 8)

o Enables systematic studies previously not possible

o Completion of data analysis within grant periods. Timely completion of Ph.D. and Master thesis.

- Detector resolution in kinematic binning
- 2D efficiency maps - *requires to rerun tracking code for every of the 200 detector planes*

Accomplishments

Complete lists of projects

(*)= shown today, (P) = planned

Drell-Yan

Charles Naim and Marco Meyer (CEA-Saclay, France)
(*) Drell-Yan and Jpsi x-section evaluation, acceptance studies
- Impact of the detectors.dat used (TGEANT or alignment) on the acceptance + SciFI shift
(*) Mass production for the DY group
- Jpsi x-section: Open-Charm contribution in the mass spectrum

Riccardo Longo (UIUC)
- Test of Improvements of TGeant-Drell Yan MC
(*) Systematic studies for DY measurements
(*) MC for DY Unpolarized Analysis
(*) Beam files test
- Test of unbinned maximum likelihood method

Vincent Andrieux (UIUC):
- NH_3 target dilution factor for DY 2015 (NLO & NNLO).
- Check EMC effect for E775 kinematic and targets LO and NLO
(*) Optimization of PYTHIA parameters to match experimental data ("k_T tuning").
- Drell-Yan with RF-separated kaon and anti-protons beams at future experiment.

Robert Heitz (UIUC)
(*) Sivers and left-right asymmetry from DY 2015.

Genki Nukazuka (Yamagata University, Japan)
- J/Psi production: investigation of the dimuon mass peak shift
- Event migration between cells of the polarized target

Yu-Shiang Lien (Academia Sinica, Taiwan)
- Drell-Yan absolute cross-section and DY & J/psi spin-independent angular analysis.
- Drell-Yan Test MC with different Primordial kT tuning

student
postdoc
senior

Ordered by node-hour consumption (balh):

naim
longo
lian1
gridin
meyer1
petrosya
heitz
criedl
nukazuka
pierre1
vidon
andrieux
bressan
chumako
franco
giarra
grube
kaspar1
lin3
maggiora
parsamya
quintans
renz
vasilish
ventura
wagner1
zavertya
zemlyani

Technical

Artem Petrosyan (JINR Dubna, Russia)
(*) Commissioning of PanDA production system

Riccardo Longo (UIUC)
- (p) Running PanDA productions

Angelo Maggiora (INFN Torino, Italy)
(*) Radio Protection simulations for 2018 Drell-Yan run

Robert Heitz (UIUC)
(*) Alignment for DY 2015.

Caroline Riedl (UIUC)
(*) Taring of raw data files for storage on BW Nearline

Hard exclusive reactions

Andrei Gridin (JINR Dubna, Russia)
- Monte Carlo for Double J/psi analysis
- (p) Monte Carlo for GPD analysis

Johannes Giarra (University of Mainz, Germany)
- (p) Analysis of DVCS cross section

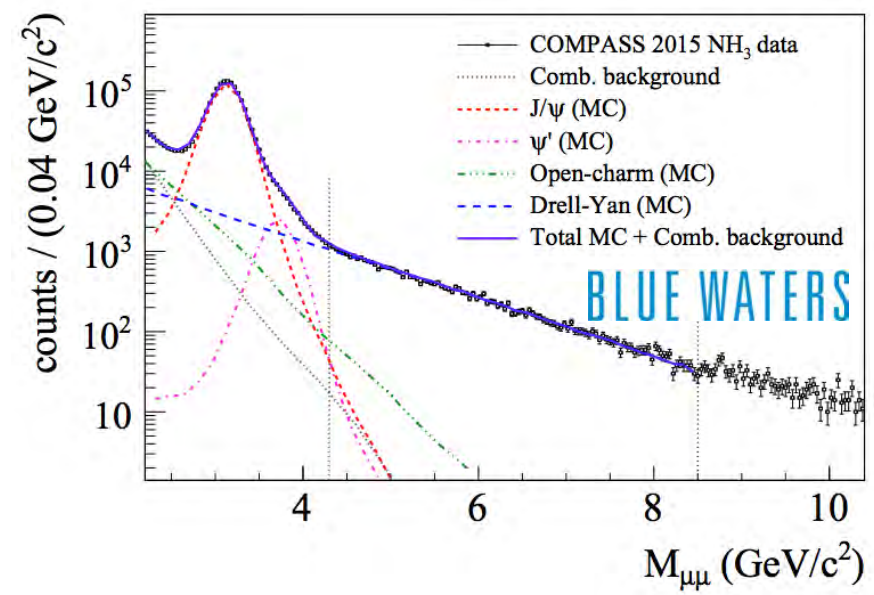
Antoine Vidon, Po-Ju Lin, Brian Ventura, Nicolas Pierre (CEA-Saclay, France)
- (p) Monte Carlo for GPD
- (p) Monte Carlo for radiative corrections

Meson Spectroscopy

Boris Grube, Florian Kaspar (University of Munich, Germany)
Waldemar Renz, Mathias Wagner (University of Bonn, Germany)
Mikhail Zavertyaev (Lebedev Physics Institute Moscow, Russia)
- (p) Meson spectroscopy: Monte Carlo and Analysis

Products from this Blue-Waters project

- **Public results:**
  - Figure 2 of the peer-reviewed publication *COMPASS Collaboration, "First measurement of transverse-spin-dependent azimuthal asymmetries in the Drell-Yan process", **PRL** 119, 112002 (2017)* is based on MC data produced on Blue Waters.
  - ~ 20 talks at international conferences and workshops to disseminate COMPASS DY results, starting with 2 parallel presentations in April 2016 immediately after the results had been released (COMPASS collaborators).



- **Effective changes:**
  - Improvement of radiation shielding for COMPASS 2018 run, and consequently enabling higher beam intensity and higher statistical precision of real data at the end of the day (A. Maggiora & COMPASS technical coordinator with CERN radio protection).
  - Modification of BW firewall policy to allow data transfers of high throughput from CERN (C. Riedl with BW IT).

- **Software and coordination:**
  - Adaption of COMPASS software, in particular PanDA-COMPASS to BW (A. Petrosyan, R. Longo).
  - COMPASS-specific job templates (M. Meyer).
  - Coordination of computing project with ~ 25 accounts (C. Riedl, R. Longo et al.).
    - 2 allocation managers
    - Dedicated mailing list for COMPASS users @ BW
    - Detailed documentation on University-of-Illinois Twiki
    - Regular working meeting and tutorials

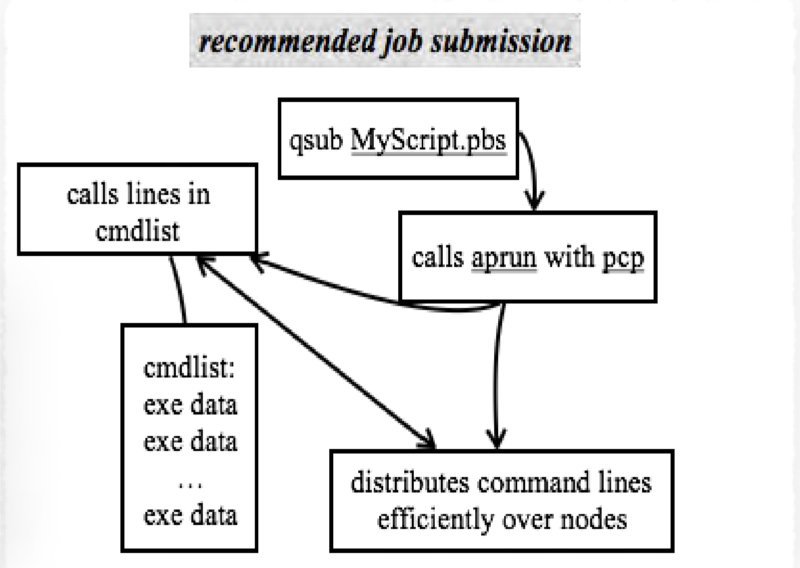
```
M. Meyer: this template is set of scripts to
simplify the parallel job submission using PCP
("Parallel Command Processor")

This submission script executes a 3-step
program :
- Configure the COMPASS job variables
- Prepare the PCP command file
- Submit the job using ".jobwrapper.pbs"
```

common software package (general & COMPASS-specific)

group directory:
/projects/sciteam/balh

setup.sh environment



Broader Impact and Sharing of Data

- **Broader Impact**

  - ***Outstanding educational potential*** for a significant number of students and postdocs towards building a community capable of using petascale computing.

  - ***NSAC milestone***: the Nuclear Science Advisory Committee has designated a measurement of the process dependence of the Sivers functions as one of its few performance milestones for DOE- and NSF-funded research in nuclear physics.

    The Frontiers of Nuclear Science: A Long-Range Plan, The DOE/NSF Nuclear Science Advisory Committee Working Group, December 2007, arXiv:0809.3137.

- **Sharing of Data**

  - Mass productions of RD and MC data are transferred to CERN and collaborating COMPASS institutions.

  - Method: FTS3 (to CERN) and Globus Online (all other sites).

  - The productions will be the reference data to be used by every COMPASS analyzer, in particular for analyses resulting in presentations at international workshops and conferences, and publications in peer-reviewed journals.

Blue Waters team contributions

Thank you to the BW team!

- User support (lost token, login difficulties, wrong e-mail address, etc…)
- Increase of Nearline and Online quota
- Data movement to and from BW-tape (Nearline):
  - Advice to tar archive large number of files
  - Help with stuck files
  - Help with moving old data with the start of a new allocation
- Data transfers from CERN:
  - Solution to gridftp issues on BW side
  - Discussion with IT/FTS3 experts at CERN
- COMPASS calibration mySQL data base: advice how to run it without exceeding login-node CPU limit including a skype conference
- COMPASS data work flow:
  - MC mass production: discussion and optimization of load distribution, usage of commtransparent feature
  - RD mass production: PanDA optimization and nodes with external networking connections

Mapping Proton Quark Structure using Petabytes of COMPASS data

Why it matters

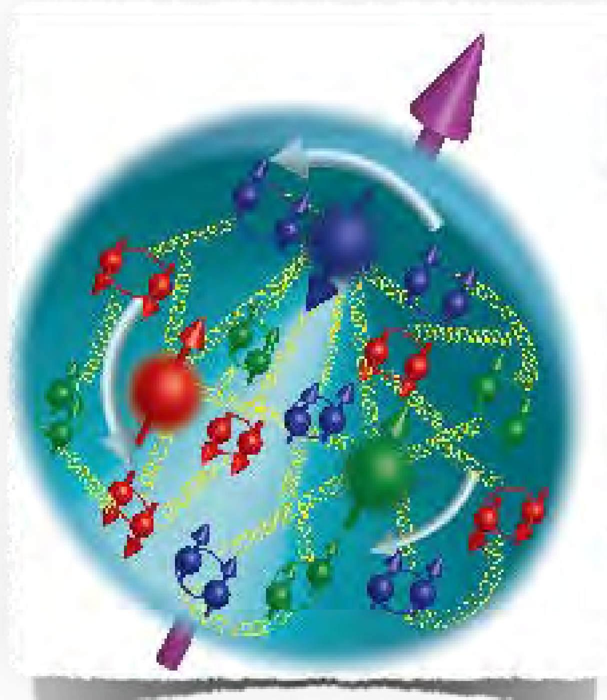
Exploration of transverse momentum- and spin-distributions of quarks in the proton

Key Challenges

Analysis of large raw data and Monte-Carlo data samples

Why Blue Waters?

Blue Waters allows for **high-precision** analysis and allows for **novel studies** otherwise too CPU intensive.



Accomplishments

First MC mass productions.
First RD mass production with PanDA about to be launched.
Input to various physics analyses.

Blue Waters team contributions

Strong support in adapting CERN software and data transfer

Broader Impact & Shared Data

Education of students and young postdocs in petascale computing.
Creation of reference data productions shared with ~250 researchers.

Products

Adaption of ATLAS tool "PanDA" to COMPASS.
COMPASS-specific job templates

Backup

Allocation "balh" (Status May 31)

Node hours used by project team vs. time (end March to end May)

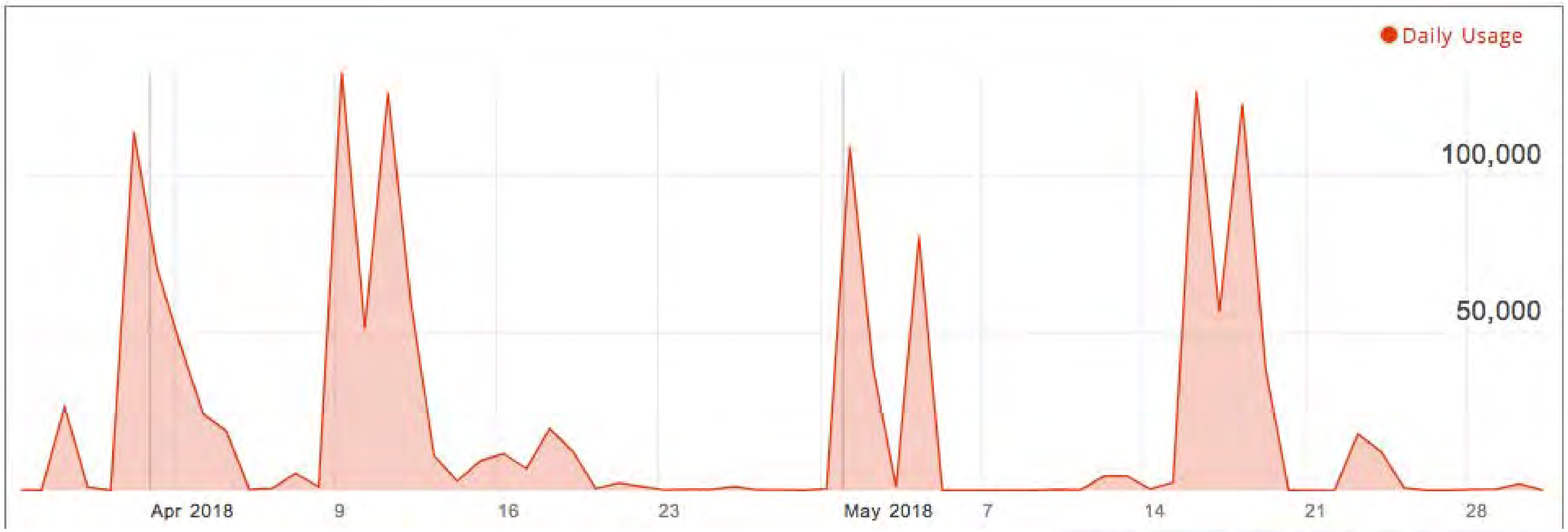
Drell-Yan MC mass production:

W12 - parts 1,2,3

W12 - parts 4,5

W07 - parts 1,2

W07 - parts 3,4,5



Data volumes

TYPE	USAGE	QUOTA
Project Online	705 GiB	5.00 TiB
Project Nearline	1.26 PiB	6.00 PiB
Online Scratch	432 TiB	750 TiB

USER	CURRENT CHARGED USAGE
naim	1.2607 M
longo	134,866
lian1	80,767
gridin	46,208
meyer1	8,420
petrosya	5,413
heitz	3,569

Drell Yan MC

PanDA commissioning

user analysis & alignment

CURRENT ALLOCATION

START DATE: 2017-06-01

EXPIRATION DATE: 2018-08-31

ALLOCATED: 9,440,000

CHARGED USAGE: 1,543,077

requested nodes | cores per node | total CPU per job | wall time | elapsed time | node type | node number

Mass simulation ("DY 2015"), example 2 packages 24h after submission to BW grid.

8564160	naim	-local_proj- (balh)	normal	W07-HMDY_1-6	100	32	3200	35:00	25:27	0:05	R	25085	xe	nid25329
8564161	naim	-local_proj- (balh)	normal	W07-HMDY_2-2	100	32	3200	35:00	25:27	0:04	R	25605	xe	nid25330
8564163	naim	-local_proj- (balh)	normal	W07-HMDY_3-d	100	32	3200	35:00	25:27	0:04	R	21986	xe	nid25331
8564164	naim	-local_proj- (balh)	normal	W07-HMDY_4-2	100	32	3200	35:00	25:22	0:07	R	26592	xe	nid25339
8564165	naim	-local_proj- (balh)	normal	W07-HMDY_5-1	100	32	3200	35:00	25:22	0:06	R	26406	xe	nid25348
8564170	naim	-local_proj- (balh)	normal	W07-HMDY_6-d	100	32	3200	35:00	25:22	0:05	R	5828	xe	nid25349
8564172	naim	-local_proj- (balh)	normal	W07-HMDY_7-d	100	32	3200	35:00	25:22	0:04	R	15023	xe	nid25350
8564178	naim	-local_proj- (balh)	normal	W07-HMDY_8-e	100	32	3200	35:00	25:22	0:03	R	12	xe	nid25351
8564179	naim	-local_proj- (balh)	normal	W07-HMDY_9-9	100	32	3200	35:00	25:18	0:06	R	111	xe	nid25358
8564180	naim	-local_proj- (balh)	normal	W07-HMDY_10-	100	32	3200	35:00	25:18	0:05	R	25011	xe	nid25359
8564182	naim	-local_proj- (balh)	normal	W07-HMDY_12-	100	32	3200	35:00	25:18	0:04	R	20366	xe	nid25425
8564184	naim	-local_proj- (balh)	normal	W07-HMDY_13-	100	32	3200	35:00	25:14	0:07	R	11856	xe	nid25429
8564185	naim	-local_proj- (balh)	normal	W07-HMDY_14-	100	32	3200	35:00	25:14	0:06	R	26198	xe	nid25431
8564186	naim	-local_proj- (balh)	normal	W07-HMDY_15-	100	32	3200	35:00	25:14	0:05	R	26240	xe	nid25432
8564188	naim	-local_proj- (balh)	normal	W07-HMDY_16-	100	32	3200	35:00	25:14	0:04	R	19313	xe	nid25433
8564189	naim	-local_proj- (balh)	normal	W07-HMDY_17-	100	32	3200	35:00	25:14	0:03	R	26719	xe	nid25434
8564190	naim	-local_proj- (balh)	normal	W07-HMDY_18-	100	32	3200	35:00	25:10	0:06	R	26741	xe	nid27562
8564191	naim	-local_proj- (balh)	normal	W07-HMDY_19-	100	32	3200	35:00	25:10	0:05	R	10574	xe	nid27563
8564193	naim	-local_proj- (balh)	normal	W07-HMDY_20-	100	32	3200	35:00	25:10	0:04	R	24342	xe	nid27564
8564197	naim	-local_proj- (balh)	normal	W07-HMDY_21-	100	32	3200	35:00	24:48	0:25	R	25627	xe	nid25260
8564199	naim	-local_proj- (balh)	normal	W07-HMDY_22-	100	32	3200	35:00	24:48	0:24	R	14722	xe	nid25261
8564200	naim	-local_proj- (balh)	normal	W07-HMDY_23-	100	32	3200	35:00	22:13	2:58	R	216	xe	nid25263
8564202	naim	-local_proj- (balh)	normal	W07-HMDY_24-	100	32	3200	35:00	22:13	2:57	R	11288	xe	nid25328
8564204	naim	-local_proj- (balh)	normal	W07-HMDY_25-	100	32	3200	35:00	22:13	2:56	R	27390	xe	nid25329
8564205	naim	-local_proj- (balh)	normal	W07-HMDY_26-	100	32	3200	35:00	21:47	3:21	R	19789	xe	nid25429
8564206	naim	-local_proj- (balh)	normal	W07-HMDY_27-	100	32	3200	35:00	21:47	3:20	R	10234	xe	nid25430
8564207	naim	-local_proj- (balh)	normal	W07-HMDY_28-	100	32	3200	35:00	21:47	3:19	R	25582	xe	nid25431
8564210	naim	-local_proj- (balh)	normal	W07-HMDY_29-	100	32	3200	35:00	21:47	3:18	R	9649	xe	nid25432
8564213	naim	-local_proj- (balh)	normal	W07-HMDY_30-	100	32	3200	35:00	21:34	3:30	R	11759	xe	nid25262
8564214	naim	-local_proj- (balh)	normal	W07-HMDY_31-	100	32	3200	35:00	21:26	3:37	R	526	xe	nid25330
8564216	naim	-local_proj- (balh)	normal	W07-HMDY_32-	100	32	3200	35:00	21:26	3:36	R	17809	xe	nid25331
8564217	naim	-local_proj- (balh)	normal	W07-HMDY_33-	100	32	3200	35:00	19:10	5:51	R	25629	xe	nid25252
8564228	naim	-local_proj- (balh)	normal	W07-LMDY_1-3	100	32	3200	35:00	19:04	5:49	R	823	xe	nid27559
8564232	naim	-local_proj- (balh)	normal	W07-LMDY_2-7	100	32	3200	35:00	17:33	7:19	R	25144	xe	nid25337
8564236	naim	-local_proj- (balh)	normal	W07-LMDY_3-b	100	32	3200	35:00	16:26	8:25	R	25538	xe	nid25428
8564237	naim	-local_proj- (balh)	normal	W07-LMDY_4-7	100	32	3200	35:00	16:17	8:33	R	25607	xe	nid25258
8564239	naim	-local_proj- (balh)	normal	W07-LMDY_5-9	100	32	3200	35:00	14:27	10:22	R	25631	xe	nid25357
8564240	naim	-local_proj- (balh)	normal	W07-LMDY_6-9	100	32	3200	35:00	13:26	11:22	R	19323	xe	nid25329
8564243	naim	-local_proj- (balh)	normal	W07-LMDY_7-e	100	32	3200	35:00	11:46	13:00	R	8347	xe	nid25349
8564247	naim	-local_proj- (balh)	normal	W07-LMDY_8-9	100	32	3200	35:00	9:31	15:14	R	19327	xe	nid27640
8564248	naim	-local_proj- (balh)	normal	W07-LMDY_9-8	100	32	3200	35:00	9:00	15:45	R	8880	xe	nid25257
8564250	naim	-local_proj- (balh)	normal	W07-LMDY_10-	100	32	3200	35:00		24:44	Q		xe	
8564251	naim	-local_proj- (balh)	normal	W07-LMDY_11-	100	32	3200	35:00		24:43	Q		xe	
8564262	naim	-local_proj- (balh)	normal	W07-LMDY_12-	100	32	3200	35:00		24:42	Q		xe	
8564283	naim	-local_proj- (balh)	normal	W07-LMDY_13-	100	32	3200	35:00		24:41	Q		xe	
8564285	naim	-local_proj- (balh)	normal	W07-LMDY_14-	100	32	3200	35:00		24:41	Q		xe	
8564286	naim	-local_proj- (balh)	normal	W07-LMDY_15-	100	32	3200	35:00		24:40	Q		xe	
8564290	naim	-local_proj- (balh)	normal	W07-LMDY_16-	100	32	3200	35:00	8:52	15:46	R	1968	xe	nid25334
8564292	naim	-local_proj- (balh)	normal	W07-LMDY_17-	100	32	3200	35:00		24:38	Q		xe	
8564294	naim	-local_proj- (balh)	normal	W07-LMDY_18-	100	32	3200	35:00		24:37	Q		xe	
8564295	naim	-local_proj- (balh)	normal	W07-LMDY_19-	100	32	3200	35:00		24:36	Q		xe	
8564296	naim	-local_proj- (balh)	normal	W07-LMDY_20-	100	32	3200	35:00		24:35	Q		xe	
8564305	naim	-local_proj- (balh)	normal	W07-LMDY_21-	100	32	3200	35:00		24:34	Q		xe	
8564308	naim	-local_proj- (balh)	normal	W07-LMDY_22-	100	32	3200	35:00		24:33	Q		xe	
8564310	naim	-local_proj- (balh)	normal	W07-LMDY_23-	100	32	3200	35:00		24:32	Q		xe	
8564317	naim	-local_proj- (balh)	normal	W07-LMDY_24-	100	32	3200	35:00		24:31	Q		xe	
8564319	naim	-local_proj- (balh)	normal	W07-LMDY_25-	100	32	3200	35:00		24:30	Q		xe	
8564320	naim	-local_proj- (balh)	normal	W07-LMDY_26-	100	32	3200	35:00		24:29	Q		xe	
8564322	naim	-local_proj- (balh)	normal	W07-LMDY_27-	100	32	3200	35:00		24:28	Q		xe	
8564323	naim	-local_proj- (balh)	normal	W07-LMDY_28-	100	32	3200	35:00		24:27	Q		xe	
8564325	naim	-local_proj- (balh)	normal	W07-LMDY_29-	100	32	3200	35:00		24:26	Q		xe	
8564326	naim	-local_proj- (balh)	normal	W07-LMDY_30-	100	32	3200	35:00		24:25	Q		xe	
8564327	naim	-local_proj- (balh)	normal	W07-LMDY_31-	100	32	3200	35:00		24:24	Q		xe	
8564330	naim	-local_proj- (balh)	normal	W07-LMDY_32-	100	32	3200	35:00		24:23	Q		xe	
8564331	naim	-local_proj- (balh)	normal	W07-LMDY_33-	100	32	3200	35:00		24:22	Q		xe	

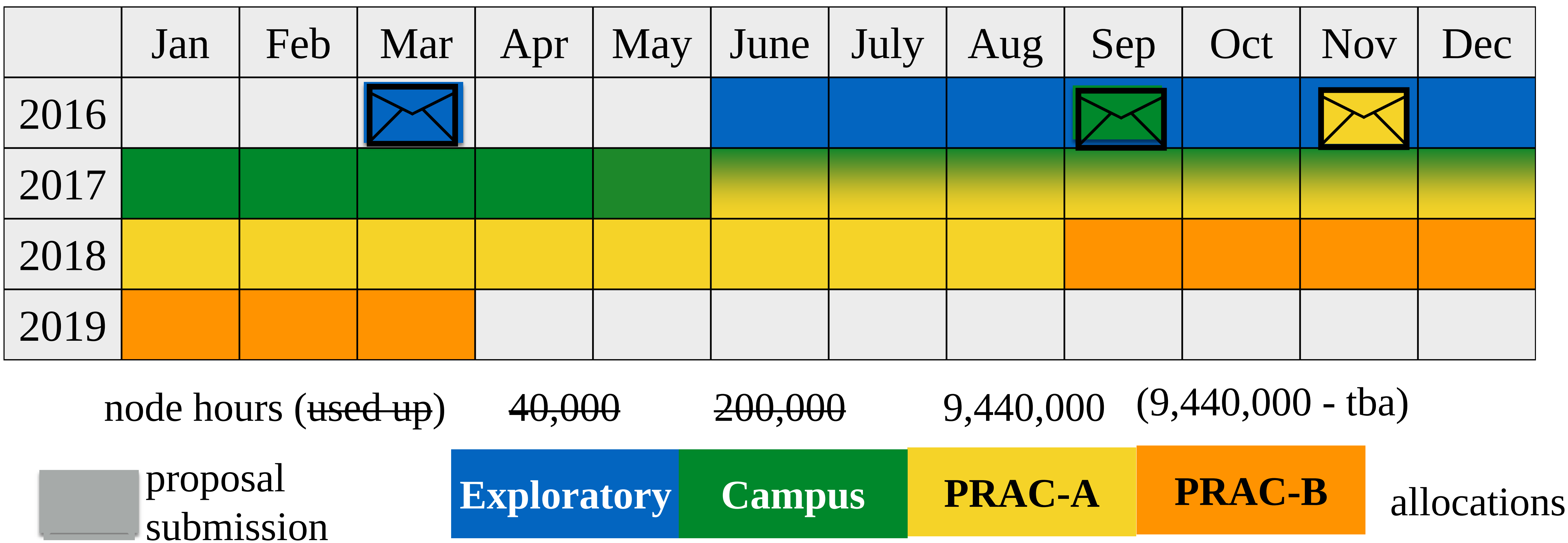
COMPASS Blue Waters: our usage plan 2018

Numbers between 2015 and 2018 refer to the year of COMPASS data taking.
The numbers in [square brackets] are million (M) node hours.

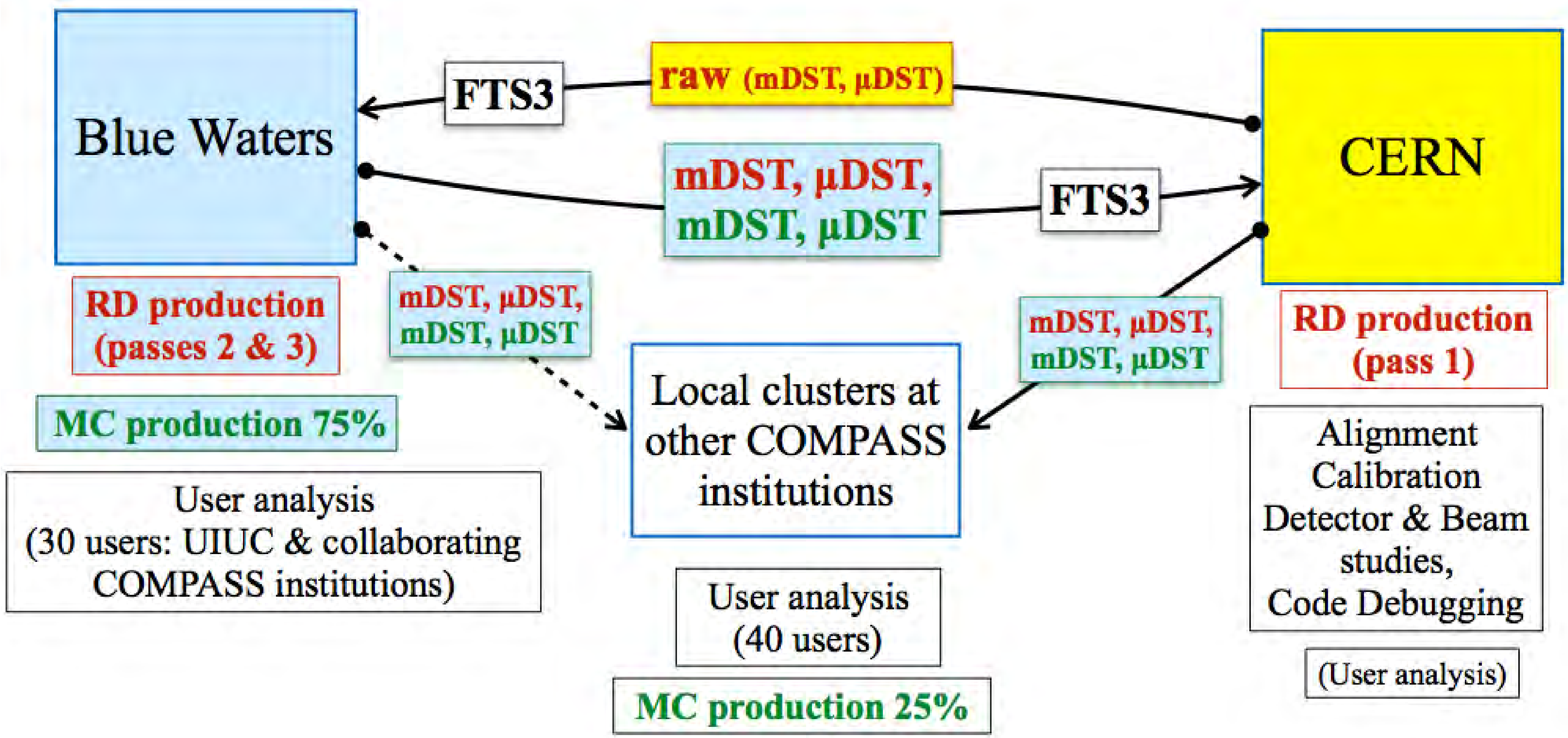
Task	March	April	May	June	July	August	sum [M node hours]
Mass simulations		2015 [1M]	2015 [1M]		other [1M]	other [1M]	4M
Mass prod. real data	PanDA commissioning	PanDA commissioning	PanDA commissioning	2015 [0.5M]	2016 [0.5M]	2017 [0.5M]	1.5M
2D efficiency maps				2015 [1M]	2016 [1M]	2017 [1M]	3M
Online production				2018 [0.1M]	2018 [0.1M]	2018 [0.1M]	0.35M
Other (user analysis, MC)	[0.1M]	[0.1M]	[0.1M]	[0.1M]	[0.1M]	[0.1M]	0.6M
sum							**9.45M**
data transfers to BW	2015	2015	2016 2018	2016 2018	2017 2018	2017 2018	-

… to be continued with part 2 of the PRAC allocation after August 2018

COMPASS Blue Waters: project time line



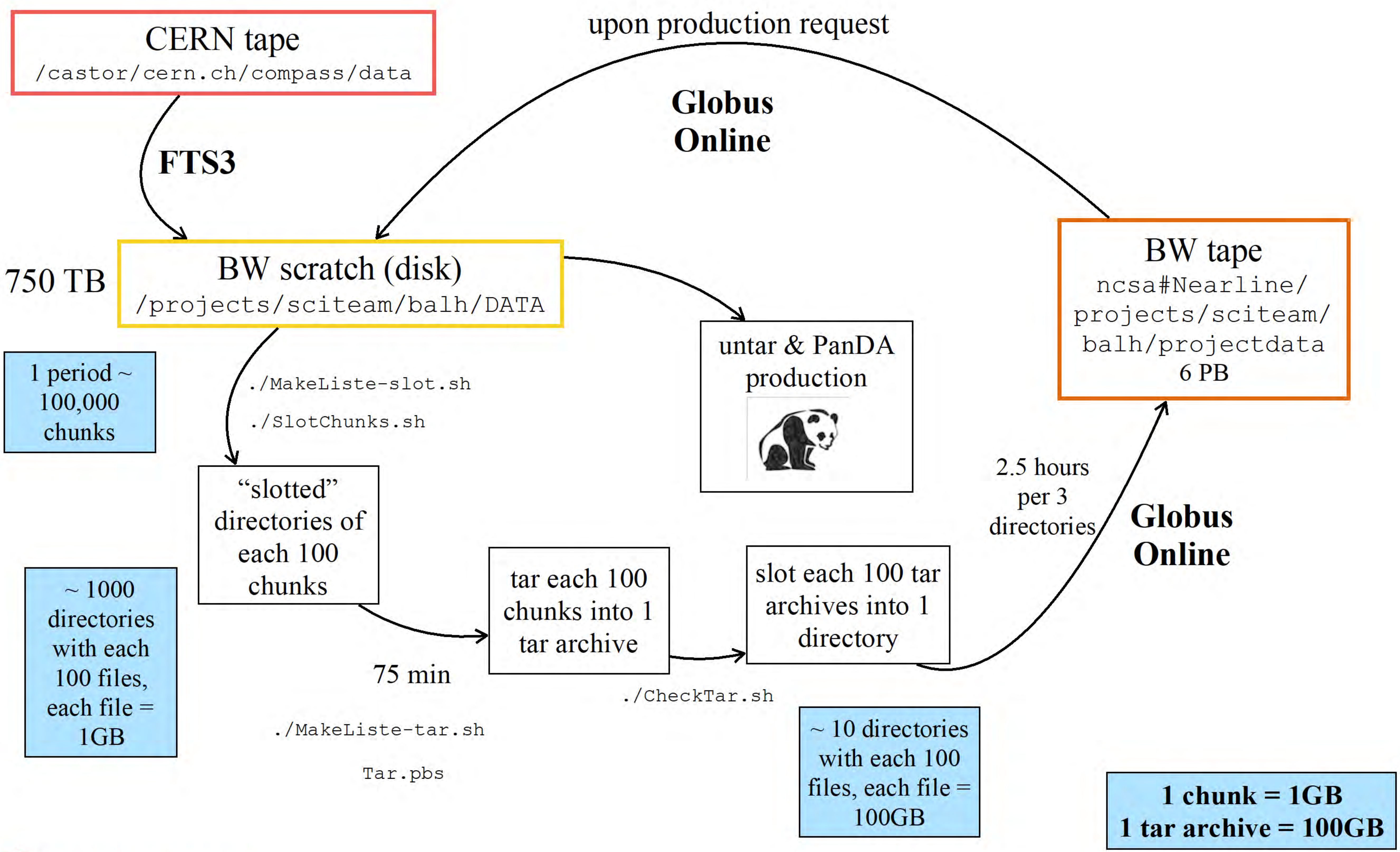
COMPASS computing model with Blue Waters (from the PRAC proposal)



19% experimental data production.
76% Monte-Carlo data production.
5% physics data analysis.

FTS3: bulk data mover created to globally distribute LHC data.

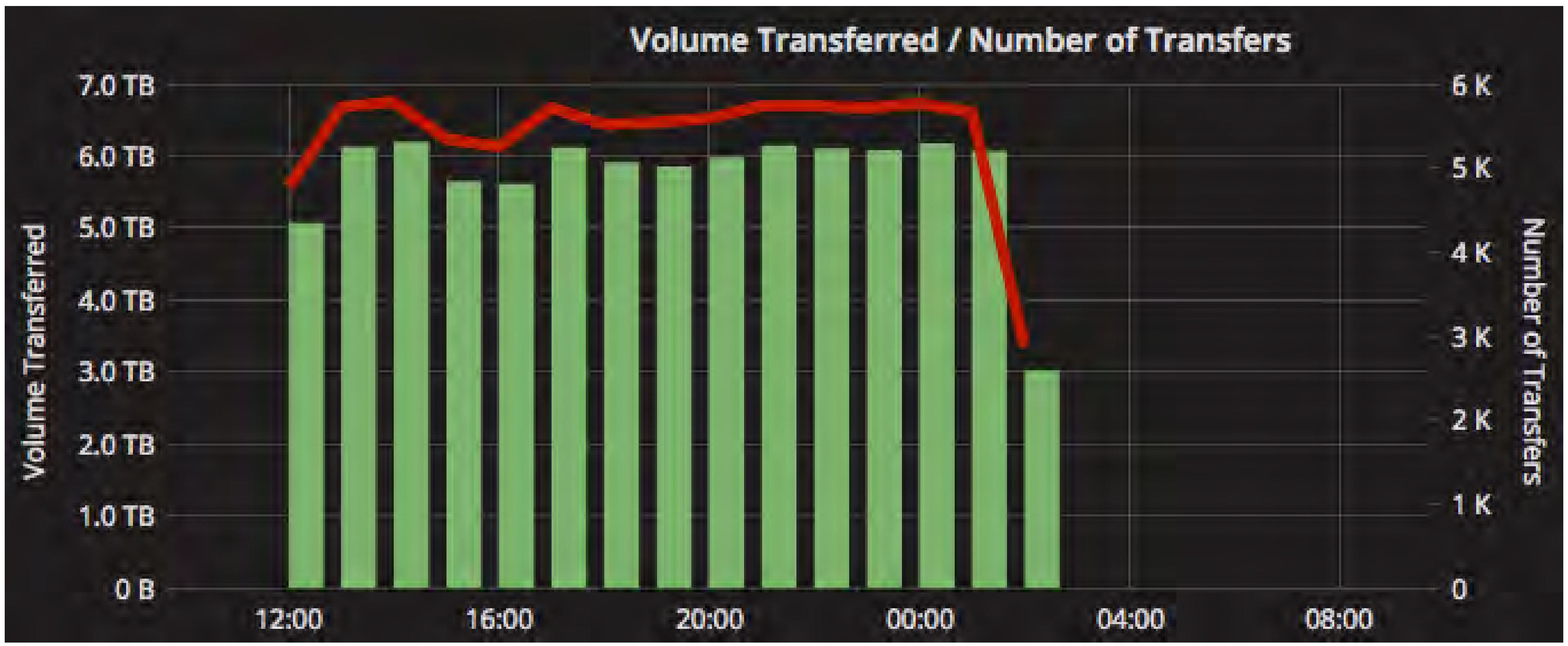
Raw data management on Blue Waters



FTS3 transfers 2018: the ideal case
(files on CERN tape were staged already)

After some exchange with BW team and FTS3 team at CERN:

14 hours 45 min for 80,896 chunks, with the stunning throughput of 1.5 GB/sec or 6TB / hour, 80 TB total
Transferring dy15W11-raw (entire period is 97,766 chunks)



Source	Destination	VO	Submitt	Acti	Staging	S.Acti	Finish	Faile	Cance	(last 1h)	Thr.
+ srm://castorpublic.cern.ch	gsiftp://ie05.ncsa.illinois.edu	vo.cc	13053	125	-	-	67697	25	-	99.96 %	1579.00 MB/s
			13053	125	0	0	67697	25	0	99.96 %	-

PanDA production interface

Cloud, Sites	Status	nJobs	defined	waiting	assigned	throttled	activated	sent	starting	running	holding	merging	transferring	finished	failed	cancelled	% failed
all sites including MCP		77418	0	0	0	0	39	196	16	898	252	0	0	52576	23441	0	30
BW_COMPASS_MCORE	online	3112	0	0	0	0	39	32	16	48	0	0	0	256	2721	0	91
CERN_COMPASS_PROD	online	74306	0	0	0	0	0	164	0	850	252	0	0	52320	20720	0	28

COMPASS PanDA | Dash | Productions | Tasks | Jobs | Errors | Users | Sites | Search | COMPASS | Help

The summary for the **dy2015W12t5BWtest** production started on 03 Apr 2018. The total number of chunks is 968. The average walltime of a finished job is 1234 minutes. Built 10:42

Tasks for dy2015W12t5BWtest | Jobs for dy2015W12t5BWtest | Show settings path for dy2015W12t5BWtest

Show 50 entries Search:

Run	Number of events	Total number of chunks	Staged	Sent	Running	Failed	Finished	Status of mDST merging	X-checked mDST	mDST migration	Status of histogram merging	Histogram migration	Status of event dump merging	X-checked event dump	Event dump migration
263293	22373350	466	-	40	9	-	417	-	no	-	-	-	-	no	-
263330	24204064	502	-	47	39	-	416	-	no	-	-	-	-	no	-

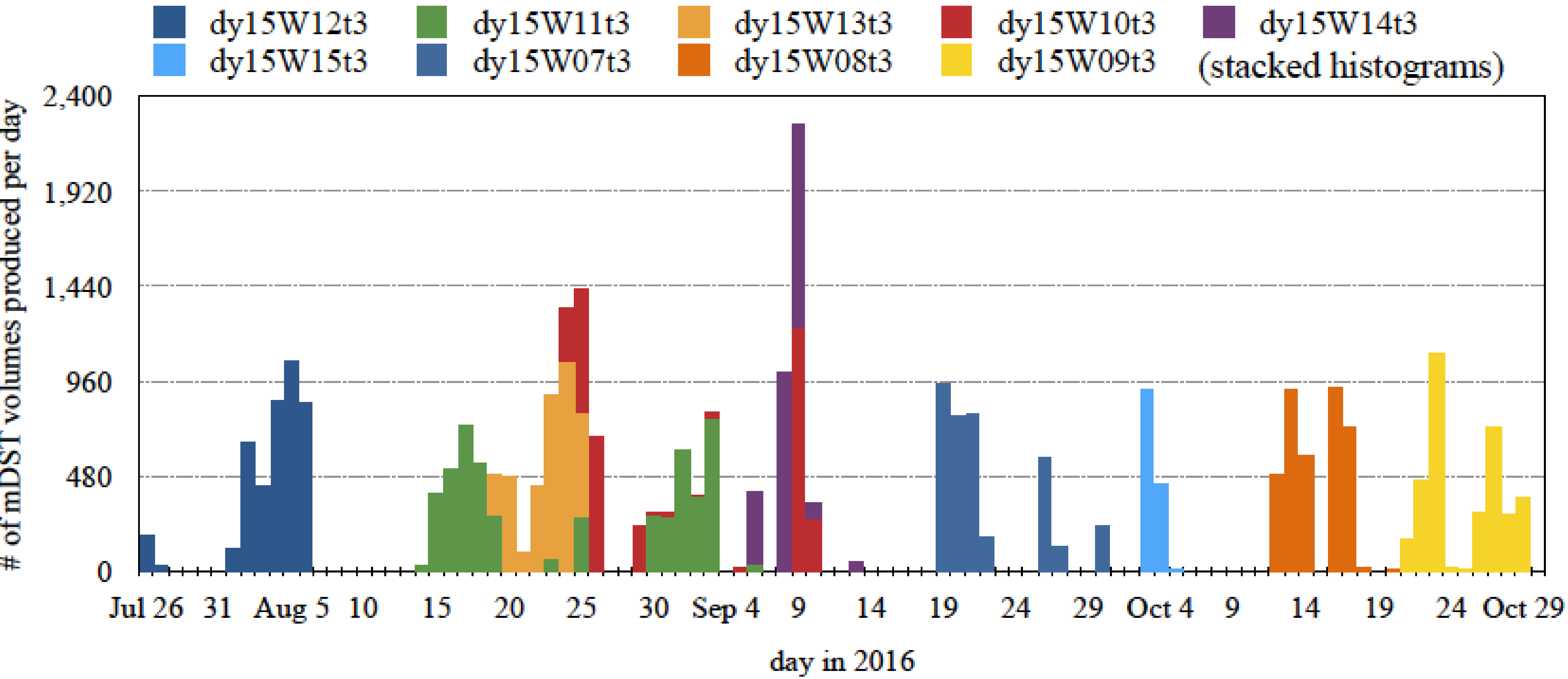
Showing 1 to 2 of 2 entries Previous 1 Next

Brought to you by the PanDA team.
All times are in UTC. Page may be cached; check the build time above.
PanDA home

- The PanDA pilot receives jobs from remote PanDA server, requests jobs by preparing PBS jobs and sending them to PBS MOM nodes
- The pilot checks if jobs are OK, sends monitoring callback to PanDA and, once jobs are finished, performs stage-out of result files to the target directories.

COMPASS BLUE WATERS

dy15 RD mass production at CERN (in 2016 - t3)

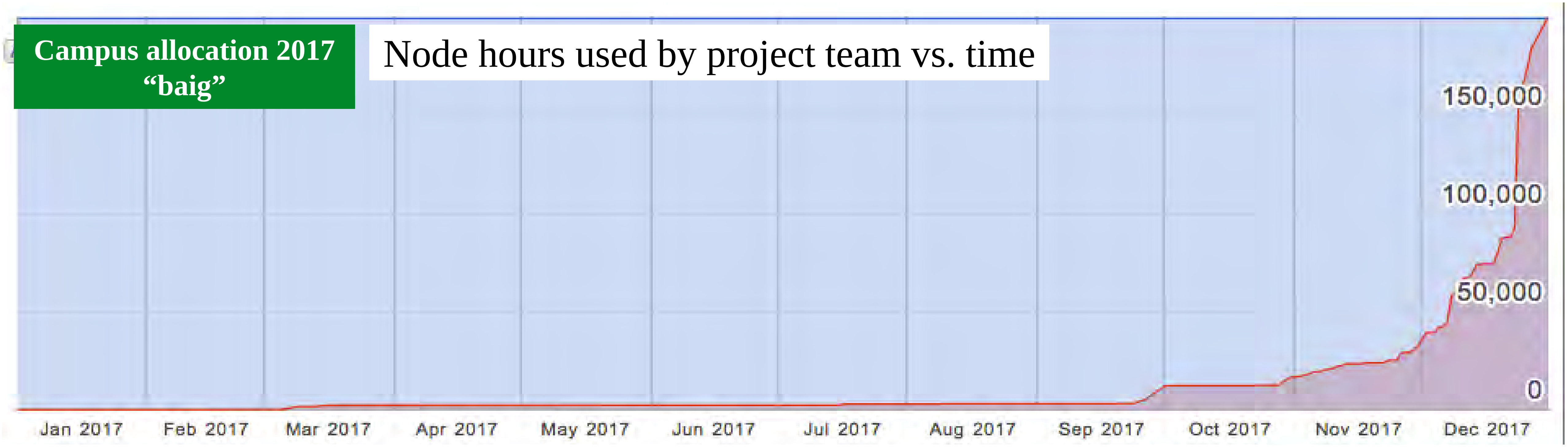


Production date of mDST volumes after merging (CASTOR)
/castor/cern.ch/compass/generalprod/testcoral/dy15W**t3/mDST

Jul 26 - Oct 29 = 5 + 31 + 30 + 29=95
Actual running days (histo counts not 0): ~ **52**

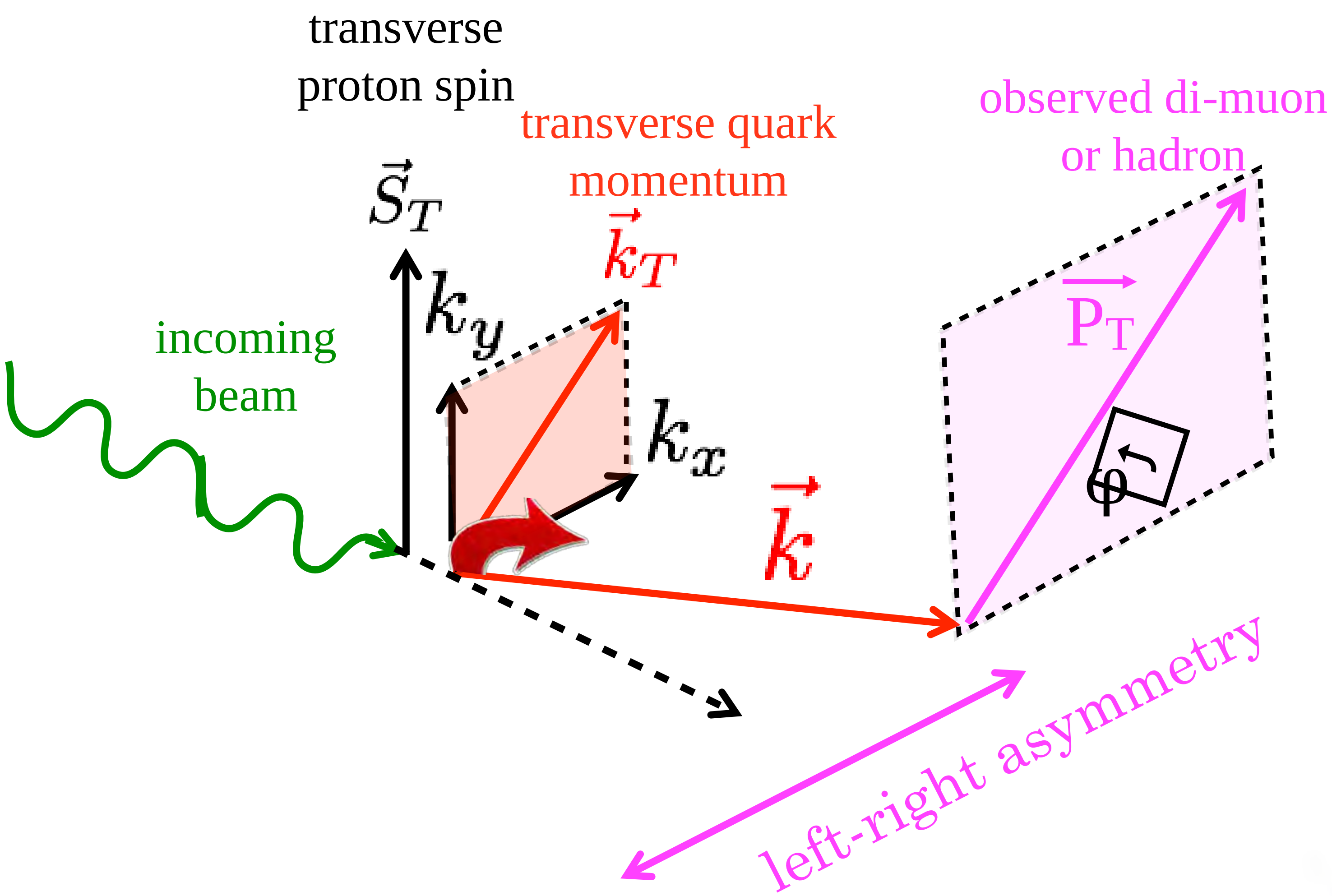
BLUE WATERS

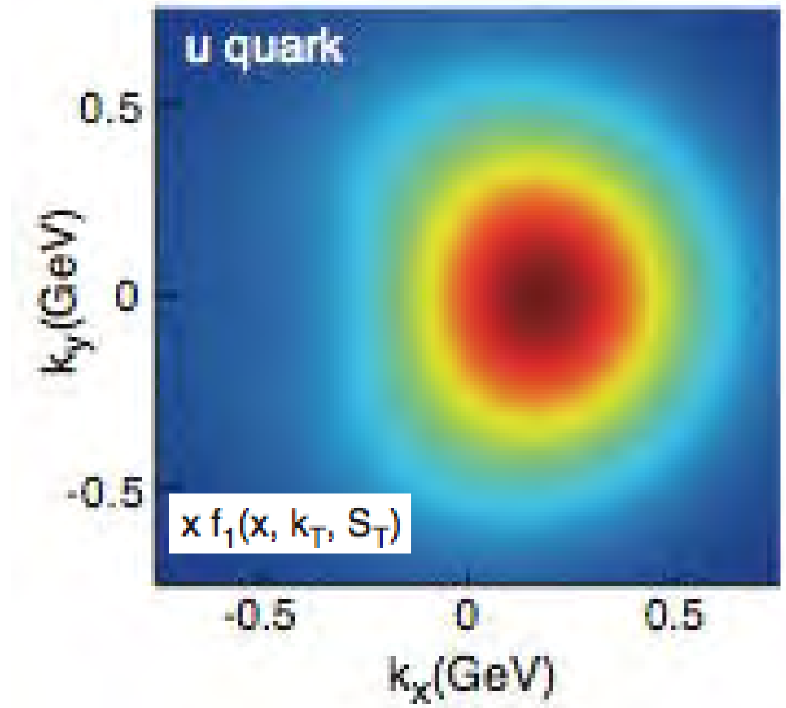
Campus Illinois allocation 2017 ("baig"): 200,000 node hours



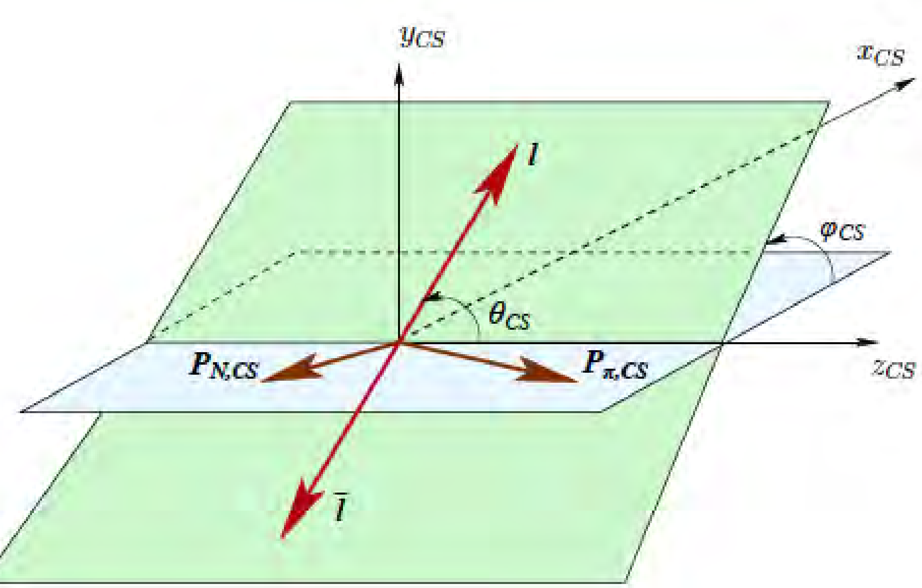
- Marco Meyer and Charles Naim: Multi-dimensional acceptance correction of Drell-Yan cross section from COMPASS 2015 data [37%]
- Riccardo Longo: Acceptance correction of spin-independent asymmetries of 2015 Drell-Yan COMPASS data [28%]
- Robert Heitz: Acceptance studies for left-right asymmetry in Drell-Yan on transversely polarized proton target [20%]
- Angelo Maggiora: Radio Protection simulations for 2018 Drell-Yan run [11%]
- Vincent Andrieux: Drell-Yan with RF-separated kaon and anti-protons beams at COMPASS-like future experiment [1%]

The Sivers effect

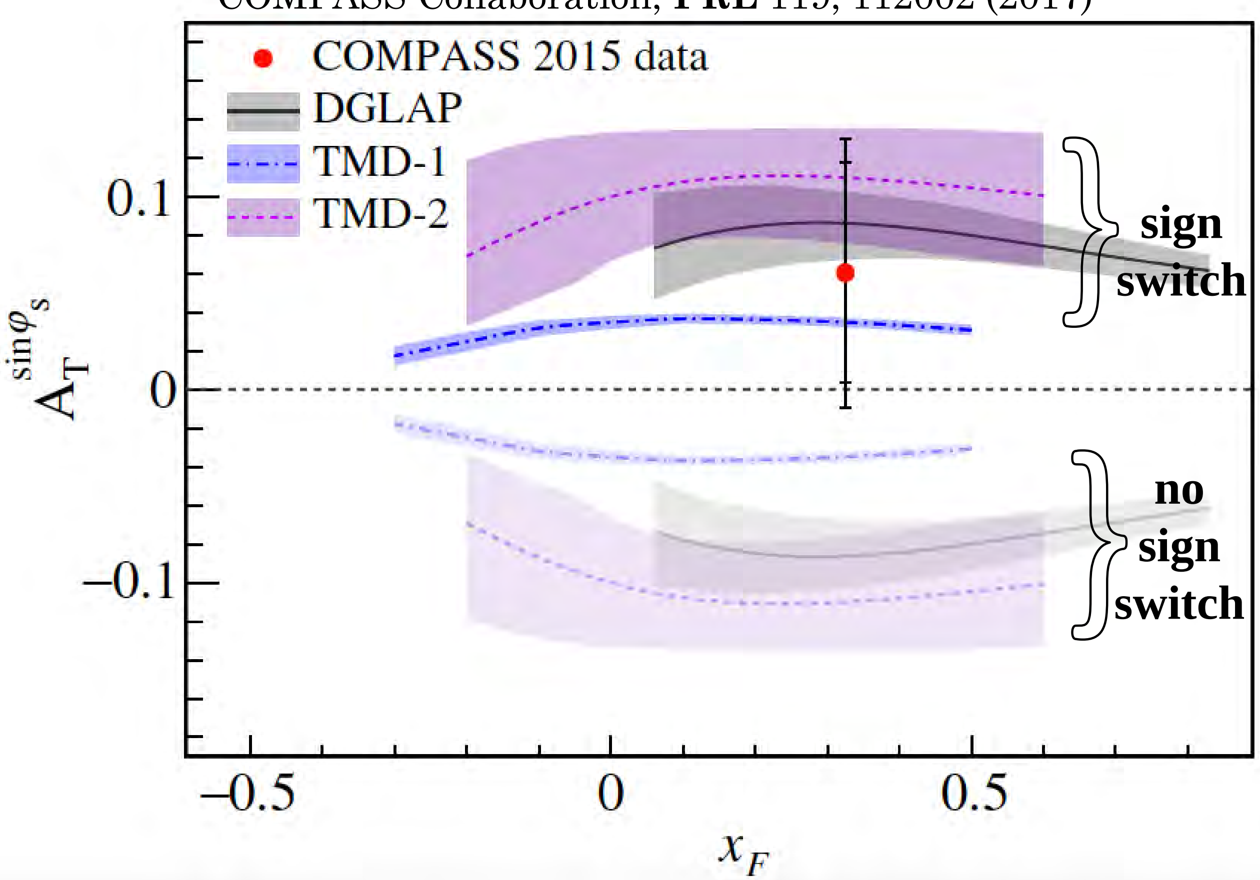




EIC "White Paper" arXiv:1212.1701, based on M. Anselmino et al., J. Phys. Conf. Ser. 295, 012062 (2011), arXiv:1012.3565



COMPASS Collaboration, **PRL** 119, 112002 (2017)



BLUE WATERS

Unpolarized DY asymmetries

Complete DY x-section

$$\frac{d\sigma}{d\Omega} = \frac{d\sigma}{\sin\theta d\theta d\varphi_{CS}} = \frac{\alpha_{em}}{Fq^2}\hat{\sigma_U}\Big\{(1 + A_U^1 \cos^2\theta + \sin 2\theta A_U^{\cos\varphi_{CS}} \cos\varphi_{CS} + \sin^2\theta A_U^{\cos 2\varphi_{CS}} \cos 2\varphi_{CS})$$

$$+ S_L(\sin 2\theta A_L^{\sin\varphi_{CS}} + \sin^2\theta A_L^{\sin 2\varphi_{CS}} \sin 2\varphi_{CS})$$

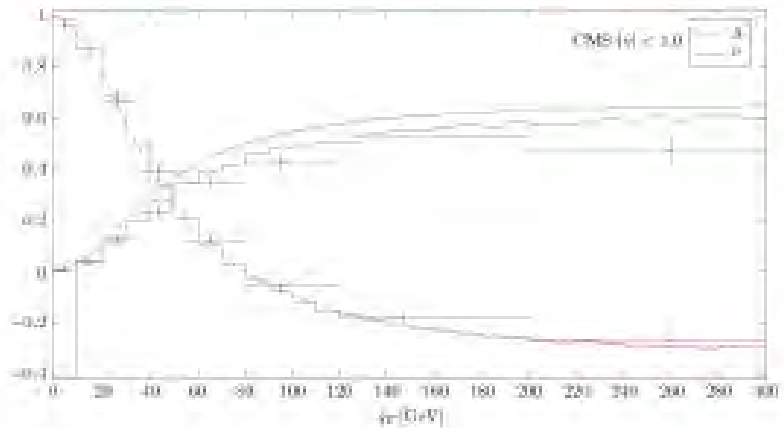
$$+ S_T\Big[(A_T^{\sin\varphi_S} + \cos^2\theta \widetilde{A}_T^{\sin\varphi_S})\sin\varphi_S$$

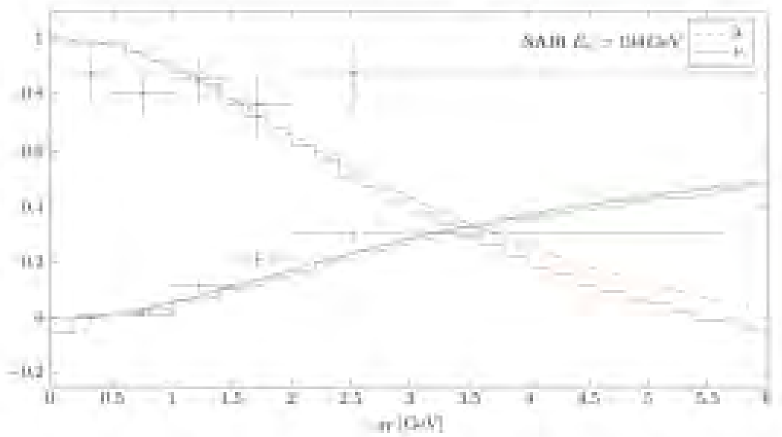
$$+ \sin 2\theta (A_T^{\sin(\varphi_{CS}+\varphi_S)} \sin(\varphi_{CS}+\varphi_S) + A_T^{\sin(\varphi_{CS}-\varphi_S)} \sin(\varphi_{CS}-\varphi_S))$$

$$+ \sin^2\theta (A_T^{\sin(2\varphi_{CS}+\varphi_S)} \sin(2\varphi_{CS}+\varphi_S) + A_T^{\sin(2\varphi_{CS}-\varphi_S)} \sin(2\varphi_{CS}-\varphi_S))\Big\}$$

3 Unpolarized asymmetries

$\lambda = A_U^1 \quad \mu = A_U^{\cos\varphi_{CS}} \quad \nu = 2A_U^{\cos 2\varphi_{CS}}$





Vogelsang and Lambertsen, PRD 93 (2016)

Lam-Tung sum rule, (Collinear LO pQCD)

(Phys.Rev. D18 (1978) 2447)

$$1 - \lambda - 2\nu = 0 \longrightarrow \lambda = 1, \quad \mu = 0, \quad \nu = 0$$

Intrinsic transverse motion +
QCD radiative effects

$$\lambda \neq 1, \quad \mu \neq 0, \quad \nu \neq 0, \quad \text{but} \quad 1 - \lambda - 2\nu = 0$$

Experimental observation:
large ν and violation of Lam-Tung relation

NA10, Z.Phys.C 31, 513 (1986) CMS, PLB 750 (2015)

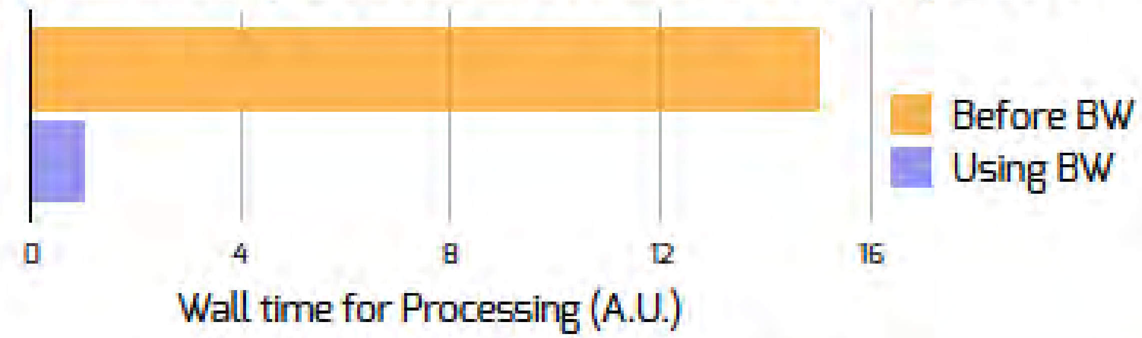
Non-perturbative effects ? $\nu \propto h_{1,\pi}^{\perp q} \otimes h_{1,p}^{\perp q}$

QCD radiative effects? NLO description?

What was produced?

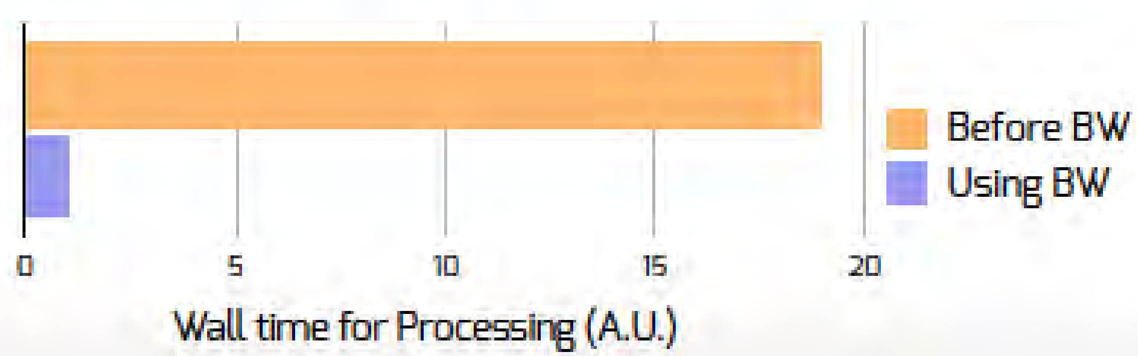
- 5 Different MC samples, ~ 35.000.000 events each;
- Pile up 10^8 π-/s , time window: +- 20 ns,;
- Generation + Reconstruction;
- 1 sample reconstructed in 3 different efficiency conditions;

Preliminary evaluation done for the poster

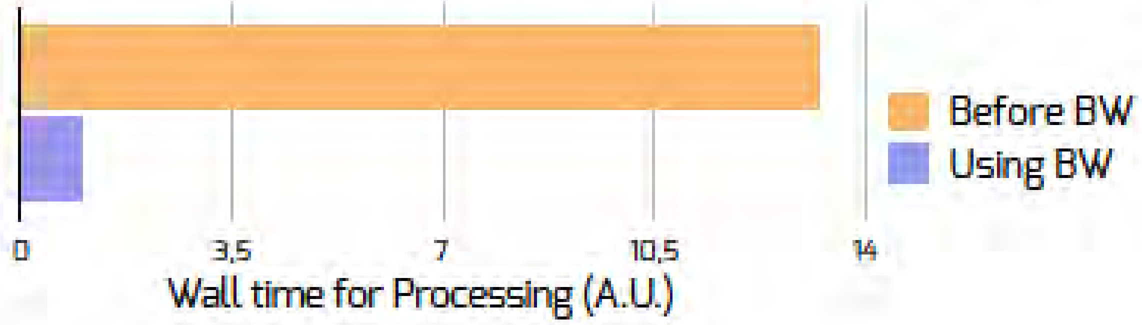


	Blue Waters	Lyon + Torino + Taipei Farms
Nodes / CPUs used at the same time	600 nodes 18000 CPUs (running 30 jobs per core + pcp + 1 spare)	Full quota: Lyon 700 + Torino 150 + Taipei 200; Assumption: 700 CPU running at the same time in average (optimistic)
Processing time per sample (2000 events per job, assuming Lyon and Torino processing time is ~ 0.75 BW time)	16.5 hours/sample (gen+rec) 1.5 hours/sample (only rec)	320 hours/sample (gen+rec) 29 hours/sample (only rec)
Total time	~ 90 hours (—> ~ 4 days)	~ 1700 hours (—> ~ 90 days)

Updated version

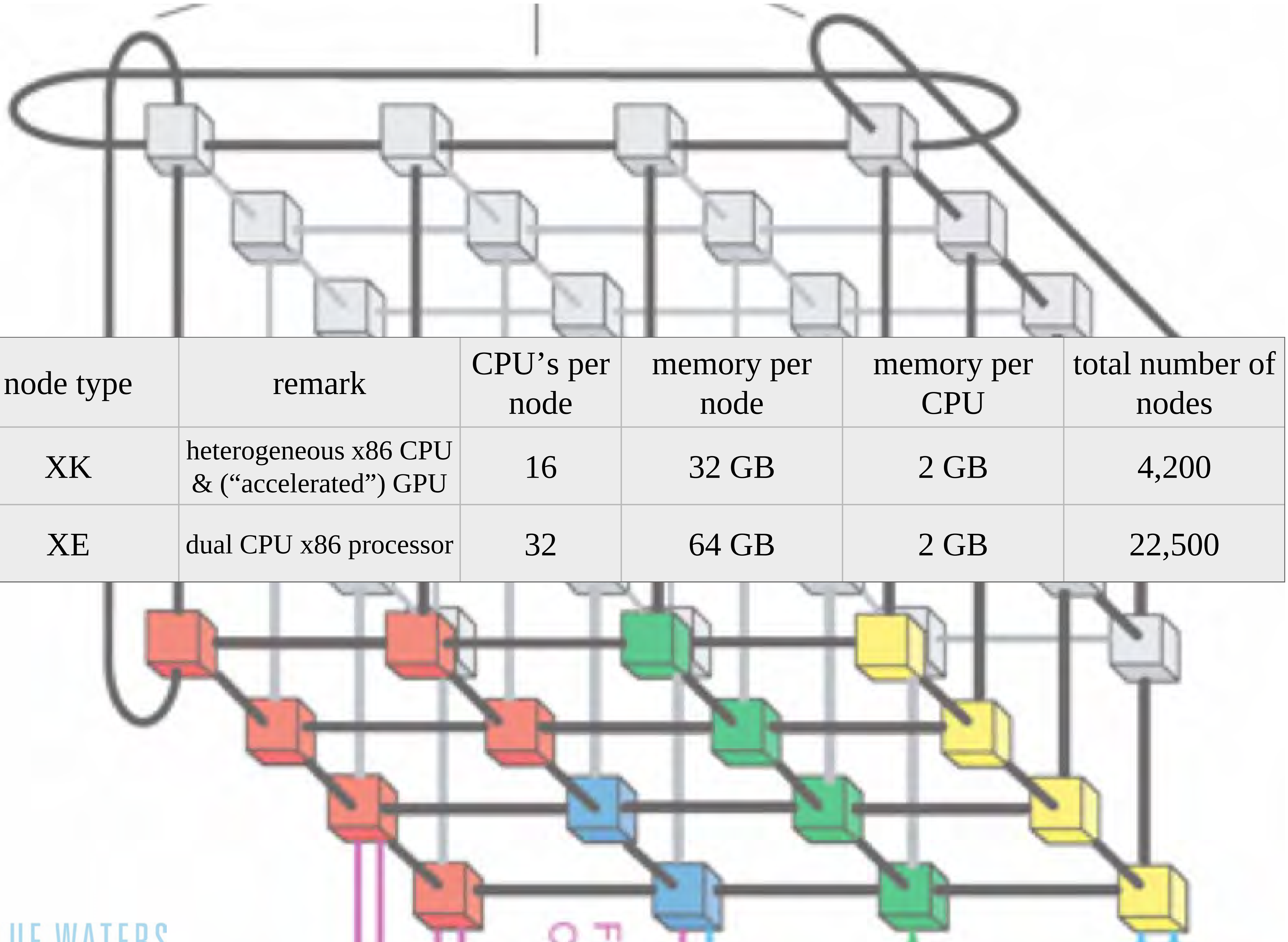


- **Different assumptions / factors neglected:**
  - The number of CPU for old farms was computed as:
    - All the CPU from Torino and Taipei (350) + an average on Lyon, where during the week one can run ~ 100-150 jobs in parallel but may easily reach the upper limit of 700 during the weekend. There it was taken 350 as average.
  - On BW, I stuck to the numbers of CPU I've used (even if I could increase it more);
  - All the farms like Lyon, Torino and Taipei have CPUs from different generations. So their performances may vary significantly from one core to the other. The factor 0.75 assumed wrt Blue Waters is pretty empirical, but it should be not too far from the reality.
  - Queuing time is different: on the old farms, jobs are starting quite promptly (at least on Torino and Taipei, which are devoted to COMPASS only. On Lyon it depends from different factors), while on BW it can take some time. But on BW we can arrange our jobs to backfill and start promptly. Therefore it is hard to account for this factor.
  - Old farms have a limited disk space (For Lyon we have a disk of 30 Tb for all COMPASS members, for Torino 50 Tb), devoted not only to store MC productions but also real data for analysis. Therefore, a big MC production may be not fully recordable in one shot, requiring additional time to move the MC mDST to Castor. This factor was neglected.
  - If we consider **all the CPUs from old farms available at the same time**, we get



  - But, to my mind, this situation is definitively not realistic.

Blue Waters: nodes and CPU's



node type	remark	CPU's per node	memory per node	memory per CPU	total number of nodes
XK	heterogeneous x86 CPU & ("accelerated") GPU	16	32 GB	2 GB	4,200
XE	dual CPU x86 processor	32	64 GB	2 GB	22,500