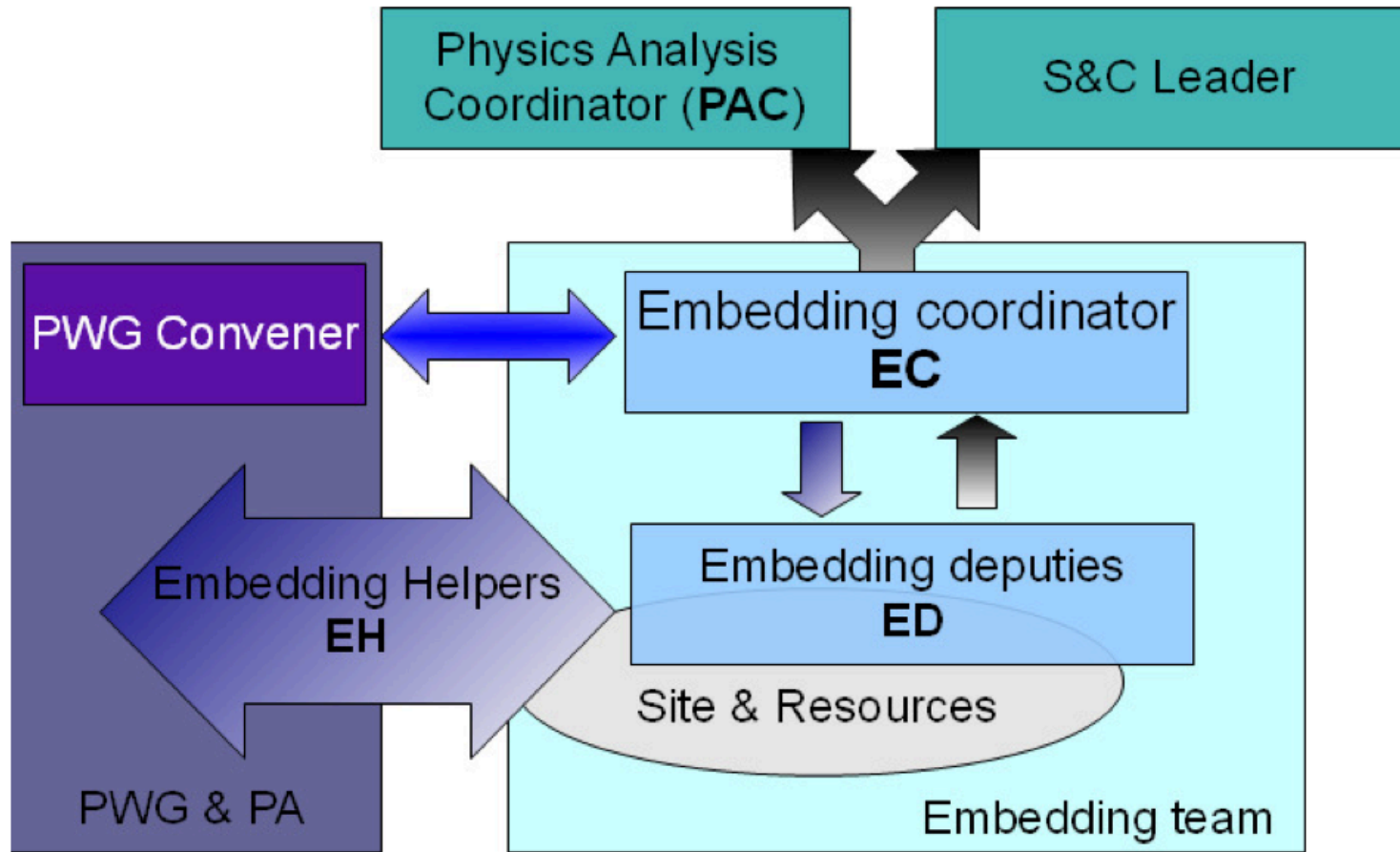# Embedding status

Xianglei Zhu
for the Embedding Team
PWGC meeting
2016/07/19

Embedding organization

Embedding Team members

- Embedding Coordinator (EC): Xianglei Zhu (Tsinghua)
- Embedding Deputy (ED): Kunsu Oh (Pusan)
- Embedding Helpers (EH):
    - Spin: Jinlong Zhang (Shandong), Kevin Adkins (Kentucky)
    - Heavy Flavor: Zachariah Miller (UIC), David Tlusty (Rice)
    - LF Spectra: Yifei Xu (SINAP), Muhammad Usman Ashraf (Tsinghua)
    - Jet-corr: Prabhat Bhattarai (UTA), Zillay Khan (UIC), Derek Anderson (TAMU)
    - UPC: Leszek Adamczyk (AGH)
    - Bulk-corr: Ning Yu (CCNU), Neha Shah (UCLA)
- PDSF support: Jeff Porter, Jochen Thaeder, Jan Balewski
- Plus invaluable support from the core S&C members! (Lidia, Yuri, Jason, Levente, Dmitry, Gene, Jerome…)
- Many thanks to former team members! (Kefeng, Chanaka, Josh…)

Current status

- The full list of STAR embedding requests: http://drupal.star.bnl.gov/STAR/starsimrequest

- **ALL requests from LFS, Bulkcorr and Jetcorr PWG have been processed!
No new requests till now!**

- **Existing requests from Spin PWG are very close to be done!**

- One request from UPC has been setup properly, code has been handed to EH

Current status (continued)

- The **remaining** requests: ALL from HF PWG!
  - Upsilon to dimuon in run14 AuAu200 (MTD)
  - Jpsi to dielectron in run14 AuAu200 (BHT1&BHT2)
  - Upsilon to dielectron in run14 AuAu200 (BHT2)
  - Electron/positron in run14 AuAu200 (BHT1&BHT2&BHT3)
  - Gamma in run14 AuAu200 (BHT1&BHT2&BHT3)
  - Pi0Dalitz in run14 AuAu200 (BHT1&BHT2&BHT3)
  - EtaDalitz in run14 AuAu200 (BHT1&BHT2&BHT3)
- For BHT requests, old test samples produced, but
  - Little statistics for one FSET (~1-2K), need 1M hours CPU time (Kunsu), impossible to be fully produced before HP2016!
  - No constraint on P.V. with vpd vertex
  - DAQ files did not cover runs before 15110001
- Then we do the following
  - **6 times MORE DAQ files** have been prepared (including runs before 15110001), Statistics in each FSET has been boosted up to ~22 K bht1 triggered events.
  - DAQ files have been chopped to keep BHT triggered events only, to significantly save computing time & disk space !
  - Productions for BHT requests are on-going right now…

Current status (continued)

- Potential issues in sector 20

Rongrong Ma,
http://www.star.bnl.gov/protected/lfspectra/marr/Analysis/Jpsi/Run14_AuAu200/20160526_HF_FollowUp.pdf

data

embedding

Xianglei
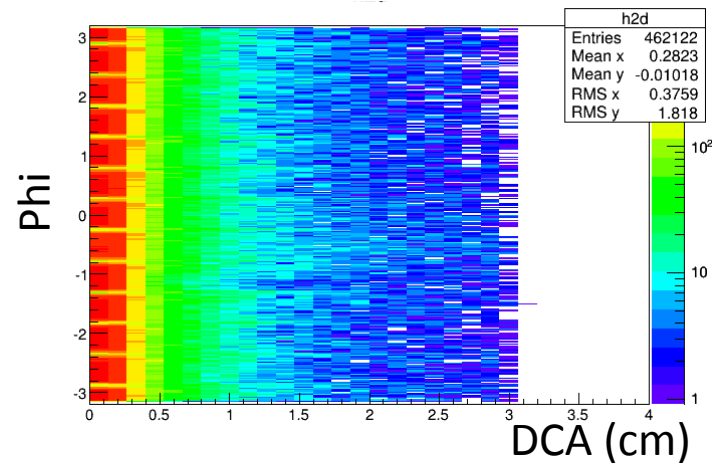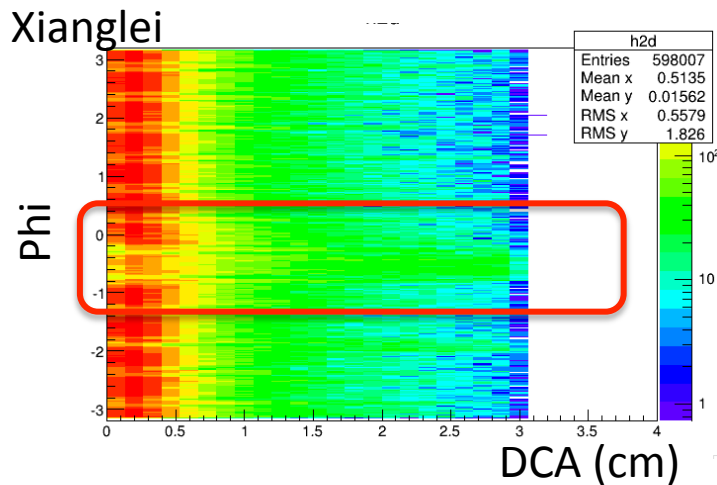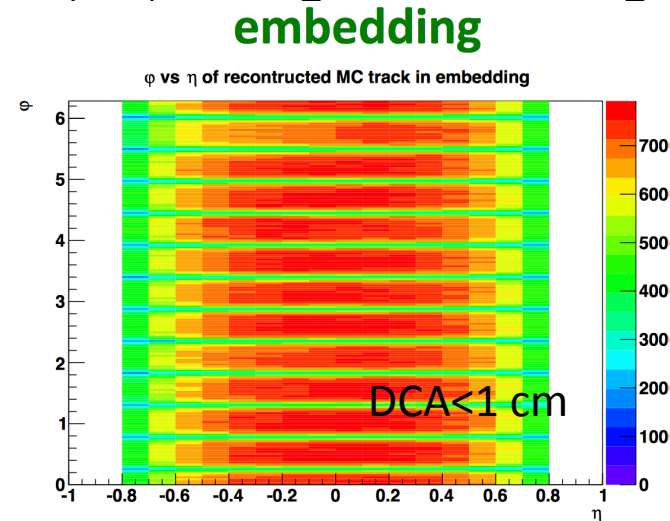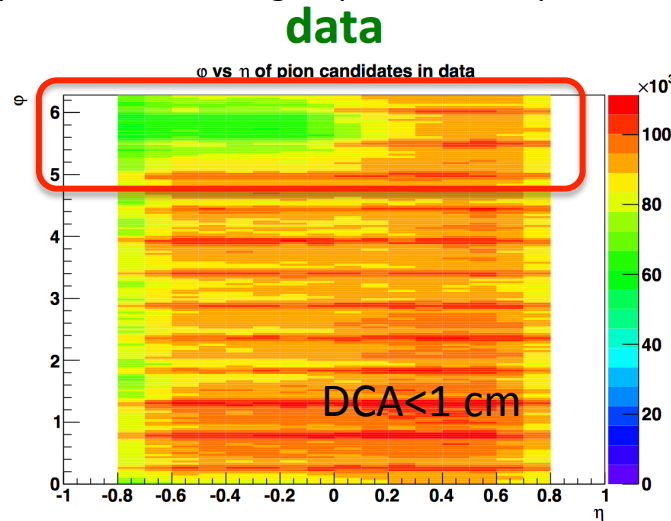
How embedding works, PWG, EC

- PA submit requests to PWG mailing list
- PWG conveners review/approve the requests, input the details of requests to the STAR simulation request page: https://drupal.star.bnl.gov/STAR/starsimrequest
- EC open the requests according to the priority list provided by PWG/PWGC, if no priority list provided, the requests will be opened according to the availability of resources, and request submission time (easy & early requests have higher priority).
- EC will contact with PA & conveners if further information is needed to setup the request
- EC will then assign the request to ED

How embedding works, ED

- ED will prepare the resources according to the request details (do it once for similar requests, for example, electron and positron in run14 auau200, bht1)
  - DAQ & tag files, (from HPSS)
    sample the daq files according to the request Run range, trgsetupname, triggers
  - MuDST files (corresponding to the daq files above, from HPSS)
    for embedding QA, and possible moretags files.
  - moretags files (corresponding to the daq files above, by analyzing the mudst files)
    provide additional information for embedding production, for example, the VpdVz constraint on P.V.
  - Chop the daq files if necessary
    not all the events in daq files are used for embedding, for example the BHT events are only a small fraction of total min. bias events in st_physics_*.daq, reading those events in the embedding production is big waste of computing power, so we chop the original daq file with "daqFileChopper" to keep only the BHT triggered events.
  - Transfer the daq,tags,moretags,mudst files to PDSF with grid tools

How embedding works, ED

- ED will prepare the embedding code for each request according to the request details.
  - Provide inputs for $STAR/StRoot/macros/embedding/ get_embedding_xml.pl
  - Check out and modify $STAR/StRoot/macros/embedding/ bfcMixer_Tpx.C if necessary
    Verify the chain3 setting with Lidia, add the chain setting to this macro, geometry tag, tune the switches for StPrepEmbedMaker
  - If necessary, check out $STAR/StRoot/St_geant_Maker, to put by hand the tune-ups for StPrepEmbedMaker.
  - Cons to compile the local code under StRoot if necessary.
  - Run get_embedding_xml.pl the get the xml and test.csh file for job submission,
  - Test the setup on terminal, watch the log carefully, if everything goes well, hand out the code to EH, finally ready for launch a test sample!

How embedding works, EH

- EH submit the jobs to the PDSF computing farm
  - Submit jobs and watch the running status of the jobs if there is a problem, the team will work with PDSF POC/ED/EC to solve the problem.
  - Send the location of test samples to embedding list
  - Provide base QA plots for the test sample
    If a problem is identified, ED/EC will check the setup. If necessary, the team will work with S&C core team to solve the potential issue in STAR library.
  - If test samples are OK, proceed to full productions.

How embedding works, PWG, EC

- PA will provide PWG QA on the test sample or full sample
  - If a problem is identified, notify embedding team. If necessary, the team will work with S&C core team to solve the issue
  - If PA confirm the samples are OK, request will be closed.
- EC supervises and changes the status of the embedding requests in the simulation request page
- EC backup the embedding data to HPSS, maintain the disks of embedding area. Restage old embedding data to disks if there are requests.

How to speed up further?

- The whole embedding process is complicated
  - several people across different time-zones get involved, communications mainly via emails.
  - EC, ED, EH are service works. The whole process can be interrupted at any stage, because the assigned team members may have other priorities.
  - EC/ED takes major responsibilities, but need to invest a lot of time in embedding, training a new ED will take a lot of time too. We are still lack of a Deputy on the maintenance of embedding base QA code.
  - PA might not be familiar with embedding data production and analysis. Hence problems (or new requirements) can be found at very late stage of the process.

How to speed up further? (continued)

- Computing resources are always limited
  - Limited bandwidth between RCF and PDSF, sometimes need to wait several days for daq files transfer to PDSF
  - PDSF farm is busy, there are competition between different experiments, and between STAR collaborators. Embedding jobs have lower priorities in the queue some-times.

```
[pdsf11] /<1>projectdirs/starprod/embedding/AuAu200_production_2011/> sgeusers
r      qw      Eqw   hr    hqw   Ehqw  dr    dt    jobIDs User-Project
--------------------------------------------------------------------------
767    100     0     0     0     0     0     0     867    alicesgm alice
3      0       0     0     0     0     0     0     3      aschmah star
3      0       0     0     0     0     0     0     3      bxyzhu cuore
0      10      0     0     0     0     0     0     10     carels lz
355    538     0     0     0     0     0     0     893    ehuang7 dayabay
57     0       0     0     0     0     0     0     57     hack dayabay
1004   3110    0     0     0     0     0     0     4114   huangbc star
256    1850    0     0     0     0     0     0     2106   jennetd atlas
212    1197    0     0     0     0     0     0     1409   jthaeder star
62     1792    0     0     0     0     0     0     1854   lscbob dayabay
277    4598    0     0     0     0     0     0     4875   mustafa star
39     0       0     0     0     0     0     0     39     nehrkorn lux
0      0       26    0     0     0     0     0     26     rexwg star
10     3786    0     0     0     0     0     0     3796   roliesha star
27     2846    0     0     0     0     0     0     2873   staremb star
0      10      0     0     0     0     0     0     0      starefl star
1      0       0     0     0     0     0     0     1      wisecg majorana
69     226     0     0     0     0     0     0     295    xlchen star
3      1476    0     0     0     0     0     0     1479   zhux star
0      109     0     0     0     0     0     0     109    zye20 star
--------------------------------------------------------------------------
3153   21716   26    0     0     0     0     0     24878  Totals

r      qw      Eqw   hr    hqw   Ehqw  dr    dt    jobIDs Project
--------------------------------------------------------------------------
767    100     0     0     0     0     0     0     867    alice
256    1850    0     0     0     0     0     0     2106   atlas
3      0       0     0     0     0     0     0     3      cuore
474    2330    0     0     0     0     0     0     2804   dayabay
39     0       0     0     0     0     0     0     39     lux
0      10      0     0     0     0     0     0     10     lz
1      0       0     0     0     0     0     0     1      majorana
1613   17426   26    0     0     0     0     0     19048  star
--------------------------------------------------------------------------
3153   21716   26    0     0     0     0     0     24878  Totals
```

Key information for an embedding requests

- Embedding is to embed simulated MC signals to **real data.**

- Detailed information about the real data should be provided
  - list of real data
    http://www.star.bnl.gov/public/comp/prod/DataSummary.html
  - Trigger sets and file type, for example, *"AuAu_200_production_mid_2014", "st_physics"*
  - Production Tag, for example, *"P15ic"*
  - Run range, with the list of bad runs. (or a list of good runs), for example, *"15076101 – 15167014"*
  - Triggers, for example, *HT1 "450201, 450211"*
  - Vertex cut, and vertex selection method, *"|Vertex_z| < 30 cm", "vertex is constrained by VPD vertex" or "default highest ranked TPC vertex"*
  - Other possible event cuts in data analysis*, for example, refmult > 250.*

- Details for simulation & reconstruction
  - Particle type (and decay mode), *"Jpsi to di-electron"*
  - pT range, *"pT from 0 to 20 GeV/c"*, and the distribution, *"flat"* or *"exponential"*
  - pseudo-rapidity or rapidity range (usually flat), *"pseudo-rapidity η, from -1 to 1"* or *"rapidity y, from -1 to 1", **please make this clear it is η or y in the request!***
  - Number of MC particles per event, *usually "5% of refmult", can also be "5% of global refmult for run14 AuAu200 GeV data"*, or a fixed number for all events.
  - Special requirement for production chain, for example, *"P15ic + NoPxlIT, NoIstIT, NoSstIT"*
  - For Event generator type requests, like Pythia in zero-bias, or StarLight in zero-bias, please provide the generator version at least. For example, *"Pythia 8.1.62".*

Please provide all above information in the simulation request!

It is OK to write all above in the descriptions.

Summary

- Almost all embedding requests from LFS, Bulkcorr, Jetcorr, Spin, UPC have been processed!
- The remaining HF requests have been setup and speeded-up, now under production.
- Towards QM17, more HF requests will come along with the new productions, more EH from HF PWG would be helpful.
- Thanks to all the embedding team members, S&C team members and PDSF experts for contributions and support!