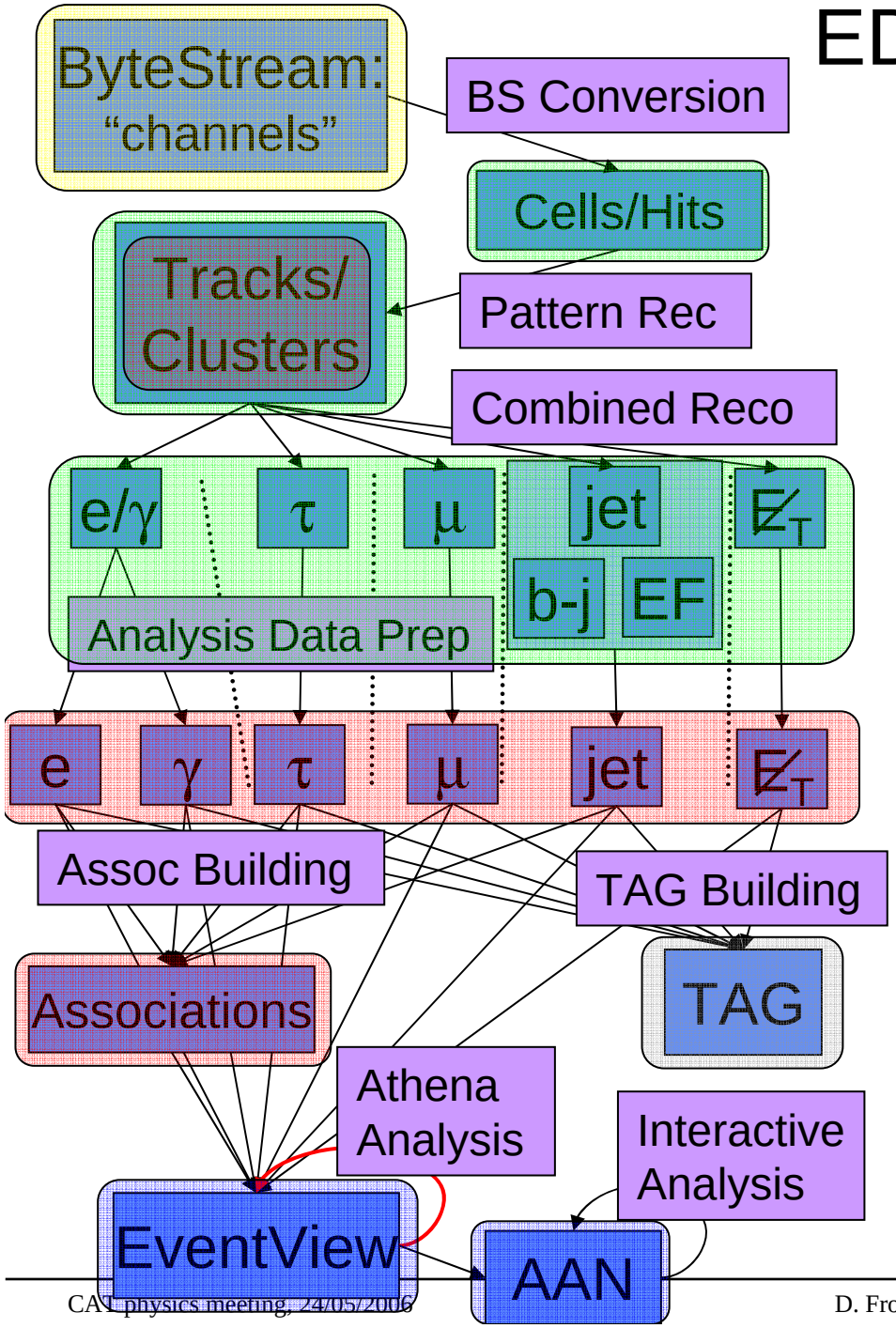
Possible evolution of AOD Event Data Model

Recent Physics Analysis Tools workshop in Japan had a whole discussion session on this with input from physicist "developers", technical discussions in the a very open and animated atmosphere

PAT is chaired by Ketevi Assamagan and has a user forum called Analysis Model chaired by Amir Farbin. Session had a number of presentations:

- **Transient-Persistent Separation (D. Malon)**
- **Analysis on ESD/AOD - Maintaining Same Interface (A. Farbin)**
- **Saving Complex Objects (AOD) into AAN (K. Assamagan)**
- **Proposed Additions/Modifications to ESD/AOD (D. Froidevaux)**
- **Difficulties with Our Current EDM (K. Cranmer)**
- **INavigable4Momentum Interface Tree (P. Loch)**

EDM And Computing Model



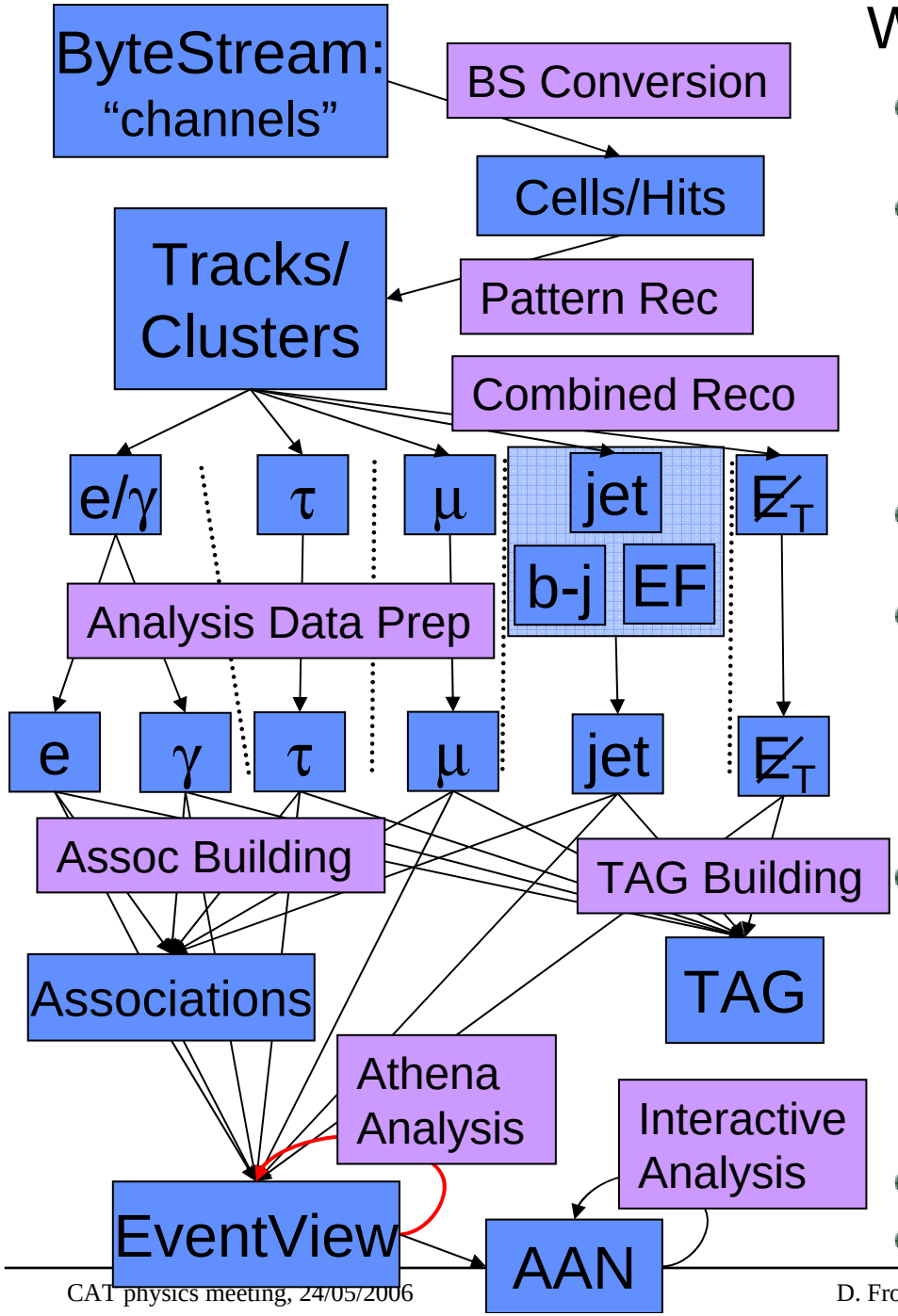
Object	BS	ESD	AOD	TAG	DPD
Raw Object Data	P	T			
Cells, Hits		PT	PT		
Clusters, tracks		PT	PT		?
Combined Reco		PT			?
Analysis Data			PT	P	?
Associations			PT		?
EventView					PT
AA Ntuple					P

New possibility: Cells for some clusters in AOD

Different rep. in AOD/ESD.

Fundamental difference between AOD/ESD: Particle object representation

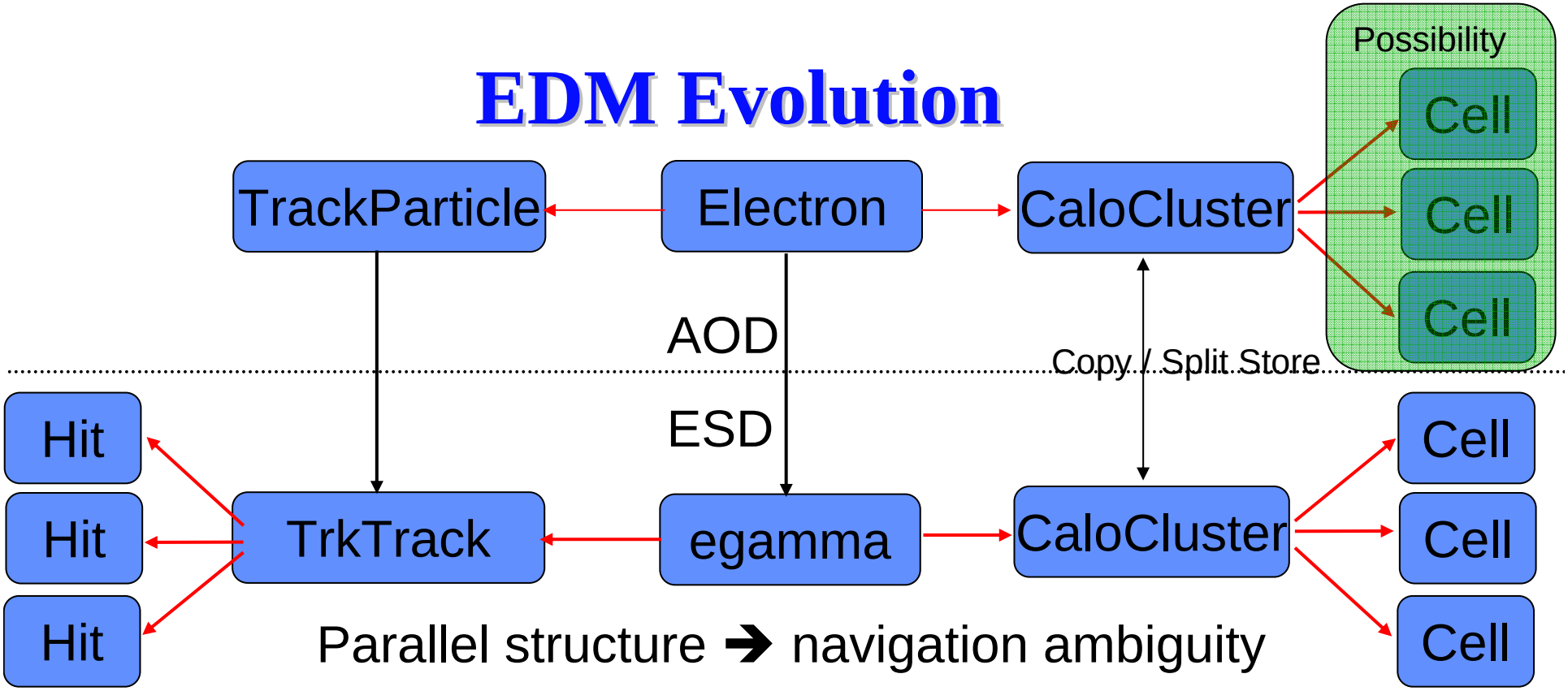
- **Obviously we have to be mindful of what data is accessible when, how fast, and where.**
- **CM cannot support everyone running all analysis on ESD.**
- **We have therefore think hard about what information is available on AOD so there is sufficient info to extract (imperfect) measurements.**



What can be done at each level?

- **Some "calibrations" applied before BS, so cannot be redone offline. Ex: OFCs**
- **Most (all?) calibrations/alignment corrections applied in reconstruction**
  - ➔ **ideally ESD (comb reco) objects should be properly calibrated.**
  - ➔ **ESD contains nearly all information to redo pattern rec and calib**
- **A few selection/decisions made in Analysis Data Prep.**
- **TAGs keep limited information:**
  - ➢ **Save summary info about N particles of specific type**
  - ➔ **probably need additional selections beyond std AOD to get useful level of eff/fake rate.**
- **Analysis on AOD**
  - ➢ **Some reco algs moving. Ex: Jet reco on clusters.**
  - ➢ **Some calibrations can be done here Ex: sampling/cluster based calibrations of electrons/jets.**
- **Assocations can keep "quality" of overlap.**
- **AAN allows going back into athena in interactive session.**

EDM Evolution



- **Where do you do start ESD analysis?**
- ➔ **Start w/ AOD, navigate to ESD.**
- ➔ **So we will build AOD objects when reading ESD.**
- ➔ **Allows you to keep common code for AOD/ESD based analysis.**
- **Provocative question: Do we need to persistify all of the Comb. Reco objects into ESD if AOD gives you essentially the same info?**
- **Particle obect redesign?**

Possible evolution of AOD Event Data Model

Presentation based on many discussions over the past months, often prompted by colleagues from running experiments (BaBar, CDF, D0)

Many thanks to K. Assamagan, A. Dell'Acqua, A. Farbin, A. Hoecker, S. Laplace and E. Moyse for their insight and technical help

Goal of this presentation is to a certain extent to respond to criticisms of current implementation:

- **"AOD not important for first years" See J. Tanaka's presentation this morning. This was also my personal belief until recently**
- **AOD quantities should "all" be recomputable from AOD itself. Main reason is that they will be wrong initially and therefore AODs will not be used if users cannot at least correct things up to a point**

Possible evolution of AOD Event Data Model

ATLAS computing model and scientific policy dictate some boundary conditions on AOD scope, size and universality:

1. **Size should not exceed 100 kBytes per event**
2. **AODs will be starting point for all physics analyses**
3. **Centrally produced AODs will be common to all high-p_T physics streams (at least)**

For some reason, alarming size was reported at last software week (T. LeCompte): about 150 kBytes/event!

Current size of AOD (release 11.3.0)

Total size is 66 kBytes per event, of which:

- **30 kBytes are McEventCollection**
- **11 kBytes are new entry since Rome: CaloTopoCluster (this is good: all jet algos can run off from AODs!)**
- **11 kBytes are TrackParticle**
- **14 kBytes are the rest (depends to quite an extent on type of event)**

Clearly need to balance number of e.g. electrons, muons, photons, taus saved to AODs against amount of information per object!

What are current limitations of AODs?

Assessment based on some feedback from users and on comparison between degree of maturity of various AOD components (not always related to degree of maturity of underlying software)

1. **Jets and b-jets quite mature, vertexing mature by 12.0.0**

2. **Electrons, photons and muons not mature yet, missing software functionality partially the reason**

3. **E_T^{miss} not at all mature, other global event quantities even missing altogether**

AOD mature contents: vertexing (4)

A. Wildauer

ESD/AOD Collections + Keys

Jets	ParticleJets	ParticleJets
Cone4TowerJets	Cone4TowerParticleJets	Cone4TowerParticleJets
ConeTowerJets	ConeTowerParticleJets	ConeTowerParticleJets
KtTowerJets	KtTowerParticleJets	KtTowerParticleJets
	all b-tagged!	retagged e.g.:
		ConeTowerParticleJetsAOD

User can do following on AOD today:

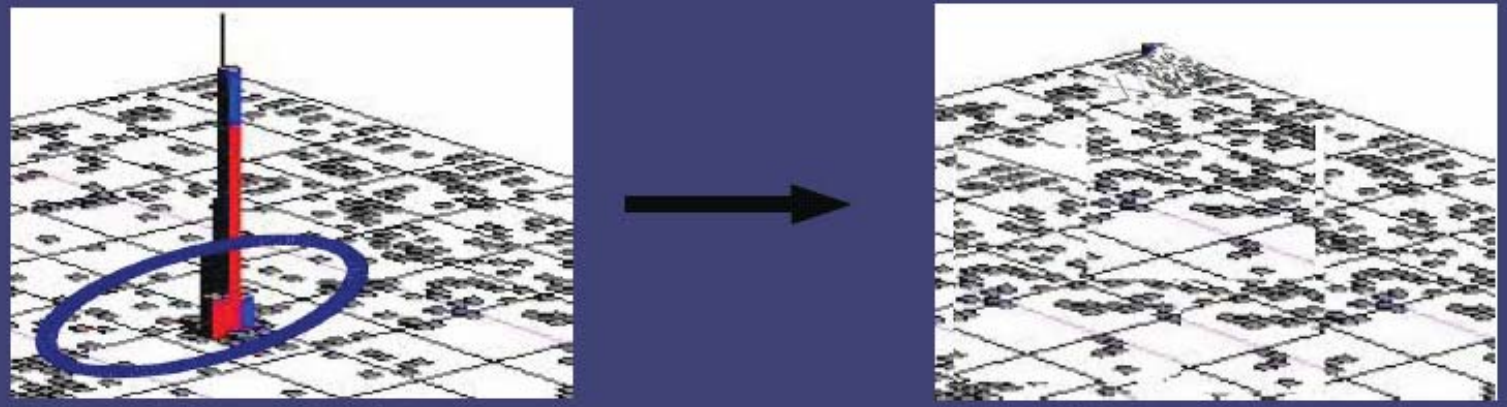
1) Scale up perigee errors of TrackParticles
2) Redo full vertexing
3) Rerun his/her favourite b-tagger
4) Write the result out to AOD (if wanted)

naming convention in the standard BTagging_standAlone.py jobO.

Electrons/photons in AODs (2)

Transform electron from Z decay into neutrino to assess systematics on W mass measurement: need cells! E. Klinkby

A better approach is the following



- After reconstruction: find Z's using the standard selection methods.
- Retrieve the egamma object of one electron, and backnavigate to obtain the cells from which it is build.
- Replace the energy deposit in these cells by noise evaluated from neighbouring cells (or in a band of the same η).
- Recalculate EtMiss and compare with truth as before.

Current status(=struggle): In order to allow for chances, the entire "AllCalo" container must be copied due to ownership issues.

Electrons/photons in AODs (3)

There are many other use-cases, especially for initial physics, requiring cell and track-hit information for electrons and photons on AODs:

- **correction of wrong calibrations (to validate reprocessing)**
- **bremsstrahlung refits, especially those combining track and calorimeter position**
- **photon vertex finding and fitting, photon direction corrections leading to improvements on e.g. $H \rightarrow \gamma\gamma$ mass resolution**

Estimates of size increases have been done (thanks to E. Moyse and S. Laplace for their help!):

1) **4 kBytes/electron for 3x7 cell matrix**
2) **4 kBytes/electron for hits on track (conservative)**

Muons in AODs (1)

Use-cases are:

- **energy radiated by muons in calorimeter for similar number of cells as electrons (less cells in ϕ, some hadronic cells)**
- **refits, e.g. those involving low-p_T segments in the muon spectrometer**

Estimates of size increases extrapolated from electrons

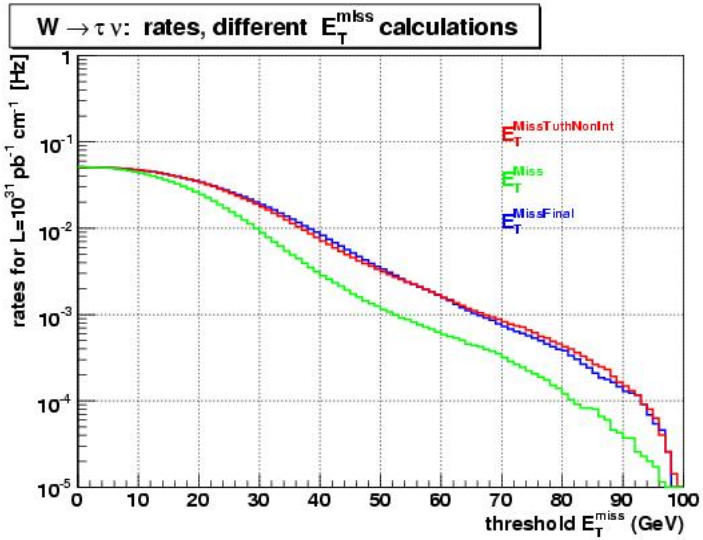
1) **4 kBytes/muon for 3x7 cell matrix**
2) **8 kBytes/muon for hits on track (conservative)**

Preliminary conclusions:

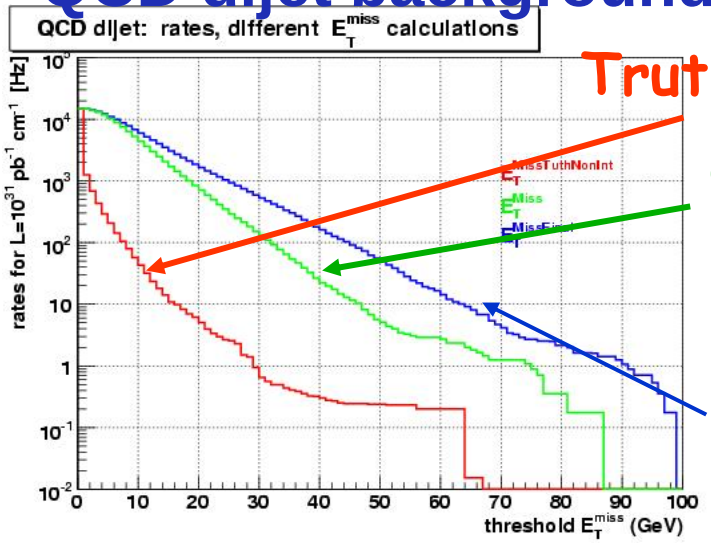
1) **Size implications may be tolerable (needs more precise work on optimal size achievable and more quantitative estimates for each of the major physics streams**
2) **Need to limit number of electrons/muons in AODs!**

E_T^{miss} and global event variables in AODs (1)

$W \rightarrow \tau\nu$ signal



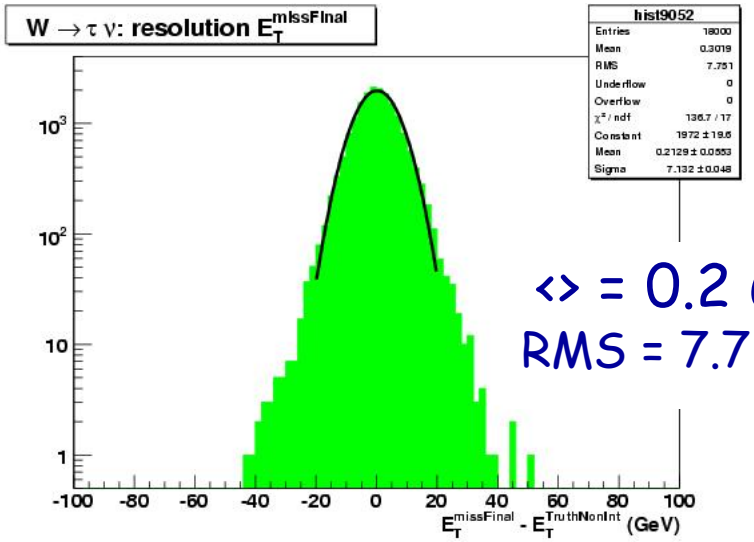
QCD dijet background



Truth E_T^{miss}

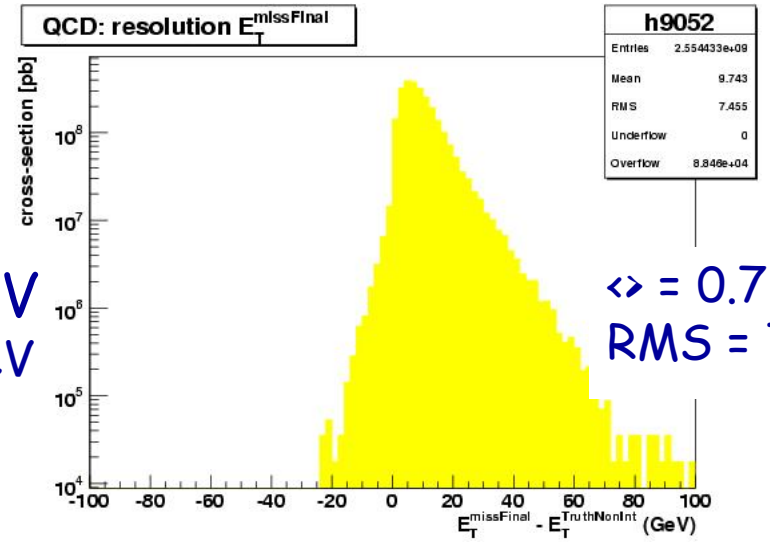
Offline E_T^{miss}

Final offline E_T^{miss}



<> = 0.2 GeV

RMS = 7.7 GeV



<> = 0.7 GeV

RMS = 7.4 GeV

E. Richter-Was and D.F.

E_T^{miss} and global event variables in AODs (2)

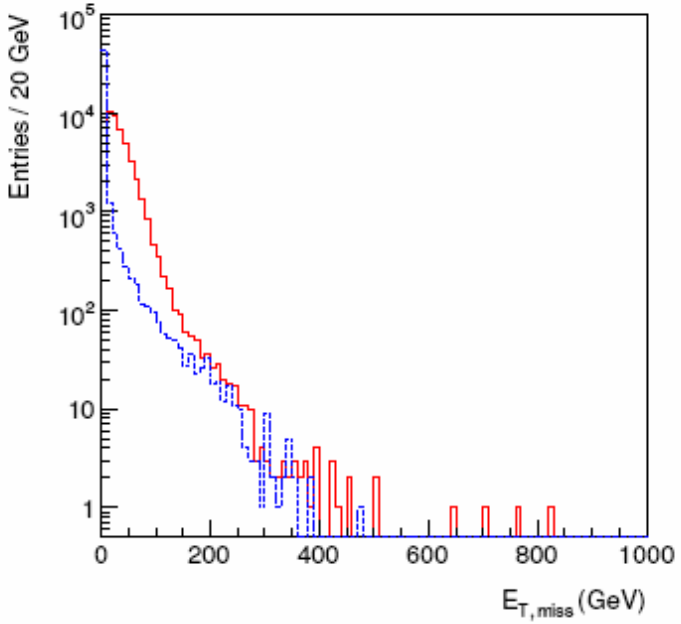
What about E_T^{miss}, the most complicated event attribute?

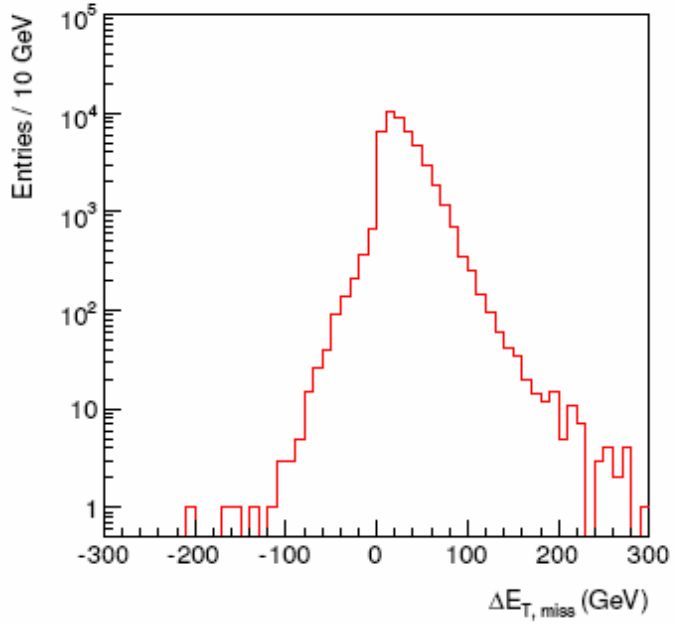
Look at dijet events with high p_T. Sample corresponds to about 130 pb^{-1}!

$\not\!\!E_T$ Tails with 11.0.41

Frank Paige, BNL

Have examined 46488 `J6_pythia_jetjet` events simulated and reconstructed with 11.0.41. Reconstructed (solid) and Monte Carlo (dash) $\not\!\!E_T$ (left) and difference $\Delta\not\!\!E_T = \not\!\!E_{T,\mathrm{reco}} - \not\!\!E_{T,\mathrm{MC}}$ (right):





E_T^{miss} and global event variables in AODs (3)

Have typically $\Delta \not\!\!E_T \lesssim 100\,\text{GeV}$ but with long tail. Find 26/46488 `J6_pythia_jetjet` events with $\Delta \not\!\!E_T > 250\,\text{GeV}$:

F. Paige

Event	$\Delta \not\!\!E_T$	$\not\!\!E_T$	Event	$\Delta \not\!\!E_T$	$\not\!\!E_T$
102	253541	same	675	257227	same
1146	368858	same	1321	271024	same
3784	317299	507243	6515	279061	same
9184	277658	352156	14855	253218	350078
16013	760036	same	19391	261944	264038
20412	301773	456709	23003	507829	508221
23452	828827	same	23816	707013	same
25679	317695	same	30743	389279	same
33012	291657	same	35225	326488	436776
44739	252732	same	44816	277436	same
44961	1.14125e+06	same	45309	320881	same
46234	577608	643522	47097	369405	same
49004	266090	266405	49842	1.26602e+06	same

E_T^{miss} and global event variables in AODs (4)

Have rerun reconstruction on 16/26 events with RDO files available at BNL. Classify as follows:

Class	Number
Jet leakage from TileBar/TileExt crack	4
Fake muons from TileBar/TileExt crack	1
Jet Leakage from TileBar/HEC crack	1
Fake muons from TileBar/HEC crack	4
Jet punchthrough	5
Other	1

F. Paige

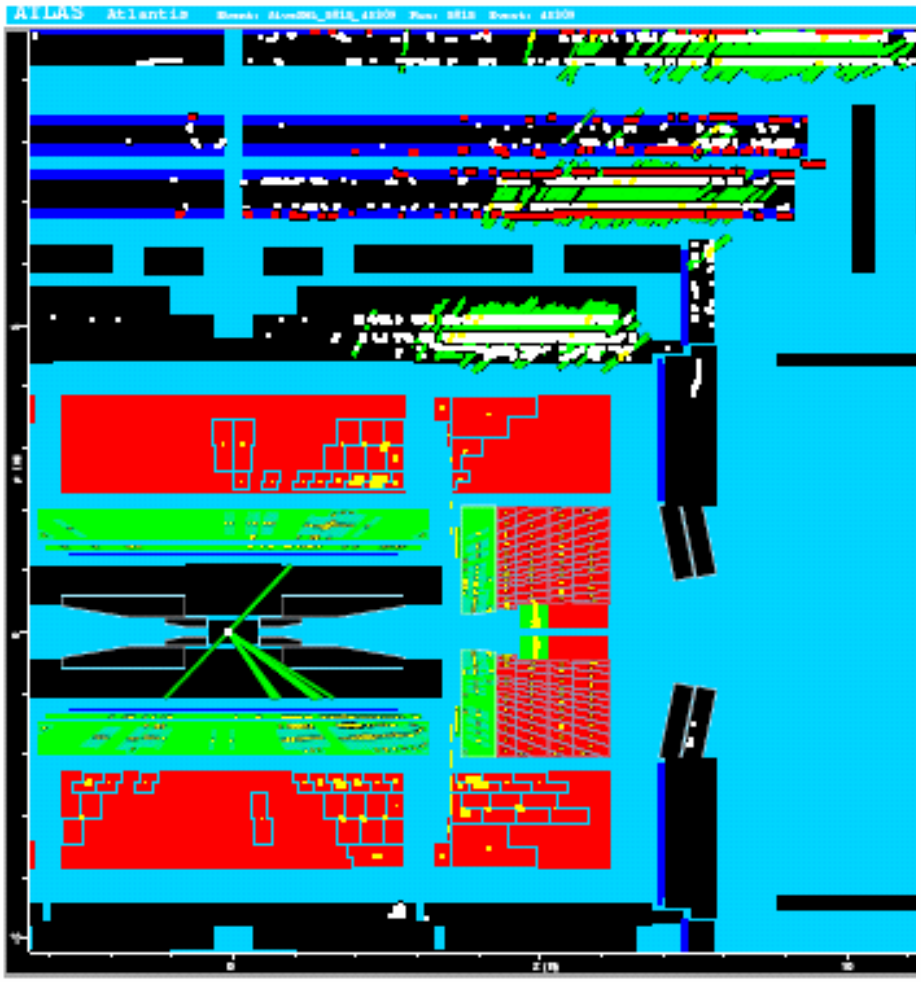
Limited statistics, but no single class dominates.

Display events with calorimeter rotated so that $\not\!\!E_T$ points up. Use 10 GeV cut on tracks.

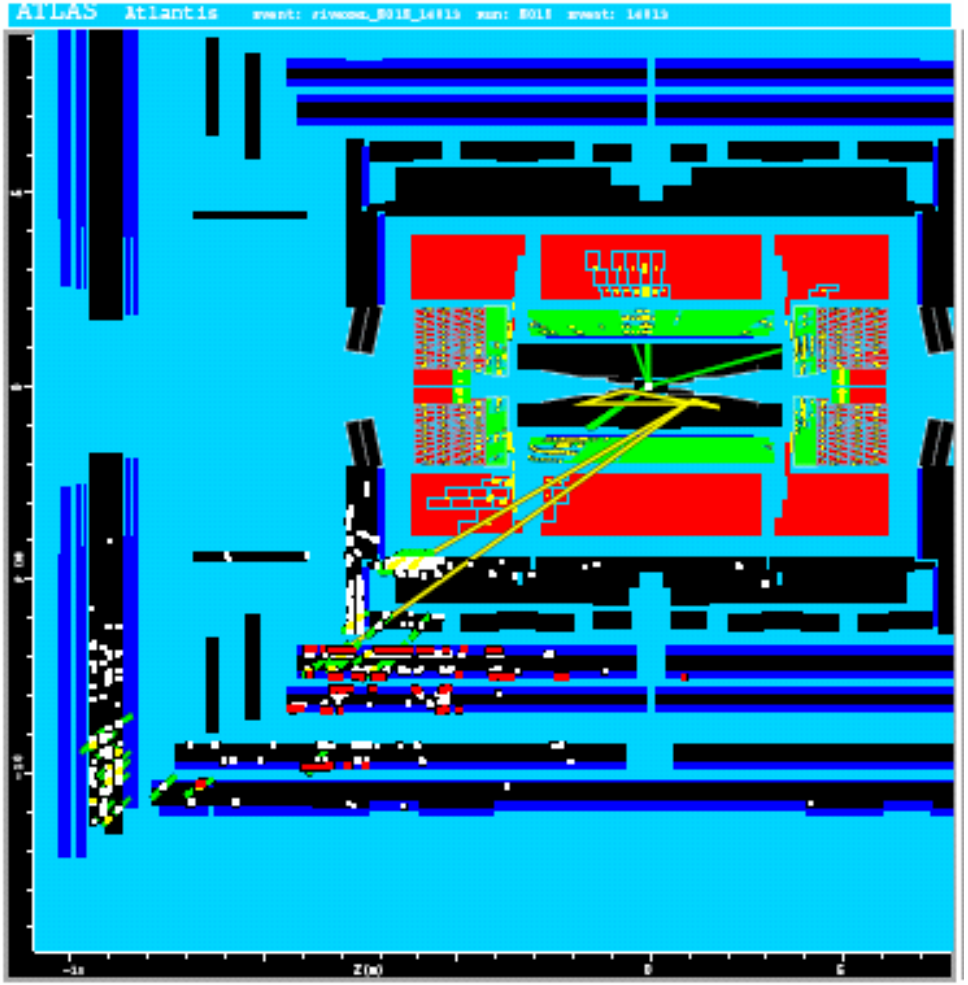
E_T^{miss} and global event variables in AODs (5)

F. Paige

Tile-tile crack



Tile-HEC crack:
fake high-p_T muons

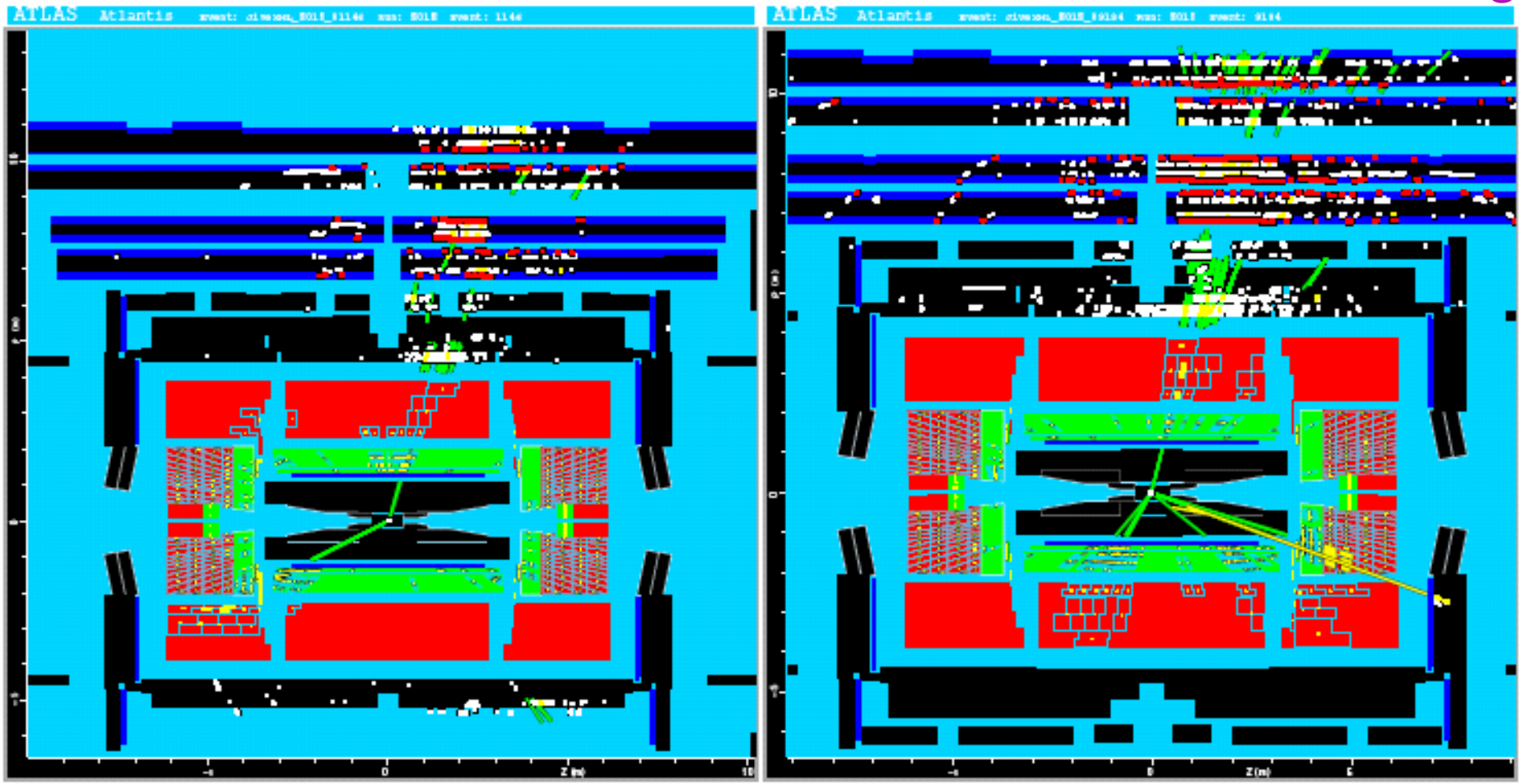


E_T^{miss} and global event variables in AODs (6)

Real worry is this: will we see what G4 tells us?

Five events with punchthrough (1146, 9184, 30743, 44961, 47097):

F. Paige



E_T^{miss} and global event variables in AODs (8)

- The previous examples illustrate the fact that users are now beginning to implement complex analysis use-cases into AODs.
- Calculation of E_T^{miss} is perhaps the most complex and analysis-dependent use-case:
  - need to understand direction of E_T^{miss} in transverse plane
  - need to assign cell energies properly for electrons and muons
  - need to optimise use of full calorimeter information
- There is also clearly a need for global event variables to minimise the use of event displays for analysis tasks
- Most of these use-cases do require list of cells and of hits on tracks for rare objects in AODs (electrons and muons for sure, perhaps even hadronic τ-candidates). Issue to be addressed globally by ESRAT, Event Management Board and future RIG group.

Implications beyond AODs: ESDs, core software

- Arguments on previous slides can be translated as technical requirement from physics towards AOD EDM: please provide tools whereby the AODs can be extended or curtailed on demand without the need for long and painful discussions and slow release cycles

- What about calibration and alignment tasks?

This issue is an even harder one!

These tasks will require intensive and repetitive analysis cycles with multiple access read/write to CondDB and other Athena services not normally used on the AODs. What model shall we adopt here?

Conclusion: AM, PAT, EDM, ESRAT need to look at this across board (object types, ESD to AOD to AANtuple) in a controlled but flexible way