Highlights of the 3rd LPCC Detector Simulation Workshop CERN, 26-27 June 2017

F. Hariri, W. Pokorski, A. Ribon

CERN PH/SFT

PH/SFT group meeting, 3 July 2017

LPCC Detector Simulation

- *The goal is to update the assessment of the status of LHC detector simulations, including the validation against data, the definition of the future needs for analysis work and for upgrade R&D studies, and the progress in the development of the tools. The workshop will bring together the current experience of the users from the experiments' physics performance groups, and the developers of the major simulation codes*
- This was the 3rd of these workshops:
  - 1st workshop in October 2011
  - 2nd workshop in March 2014
- About ~30 people following the workshop on average
  - of which ~10 remotely (via Vidyo)
- Two full days
  - 9h – 18h

Sessions

- *1st Day*
  1. Tracking
  2. **Electromagnetic Physics**
  3. **Hadronic Physics**
  4. Upgrades & Test-beams

- *2nd Day*
  1. Pileup & Overlay
  2. Geant4 Status and Plans
  3. **Fast Simulation**
  4. GeantV

Main Needs for Detector Simulations (1/2)

- More accurate physics modelling
  - For LHC experiments, there is currently a general satisfaction for the Geant4 physics performance, however there are known deviations between simulation and data that need to be addressed
  - With more recorded experimental data (e.g. LHC-HL), smaller systematic uncertainties due to the inaccuracies of physics modeling will be needed (otherwise limiting physics analyses)
  - Design of detectors for future, higher energies experiments (e.g. FCC-hh), requires extensions of the physics models to higher energies, including rare processes
  - An important common effort between developers and experiments is in physics validation, in order to use more effectively the test-beam and collider data to provide prompt and clear feedback on the quality of the simulation
    - Further discussed in the HSF meeting in Annecy: agreed to collaborate between the experimentalists of a test-beam and Geant developers to prepare a stand-alone simulation to be maintained by Geant team

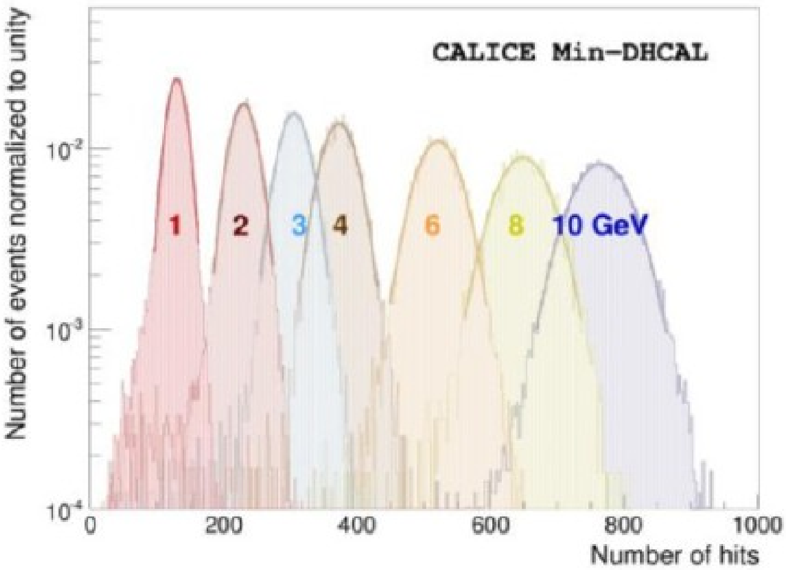
Main Needs for Detector Simulations (2/2)

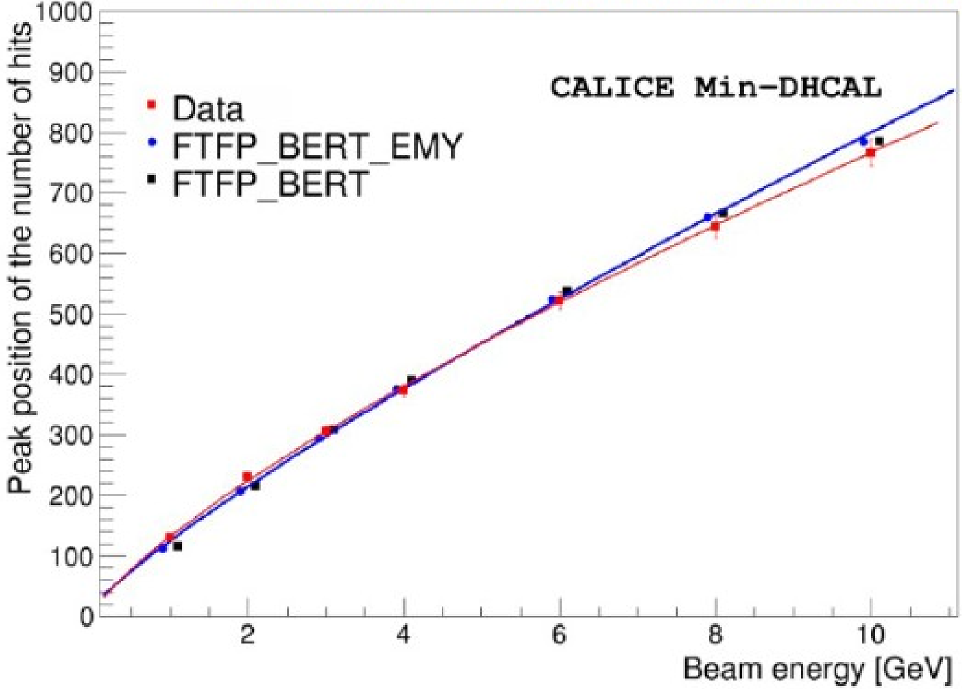
- Faster Simulations
  - Full simulation is and will be crucial for the experiments, and any effort to speed it up is mandatory
    - We are reviewing the algorithms to gain performances
    - We are replacing rejction sampling with better sampling algorithms in key areas
    - We are exploring vectorization and locality in GeantV
  - Fast simulation is becoming increasingly important
    - Intense work activities in each experiments, as traditionally, but there is a growing interest for common efforts
      - Machine-learning approaches to fast simulation
      - Generic tools that can streamline the tuning / training of fast-simulation based on the full simulation
    - Need to be able to mix full and fast simulation in the same event, so the simulation framework should allow it
      - Geant4 and GeantV allow this

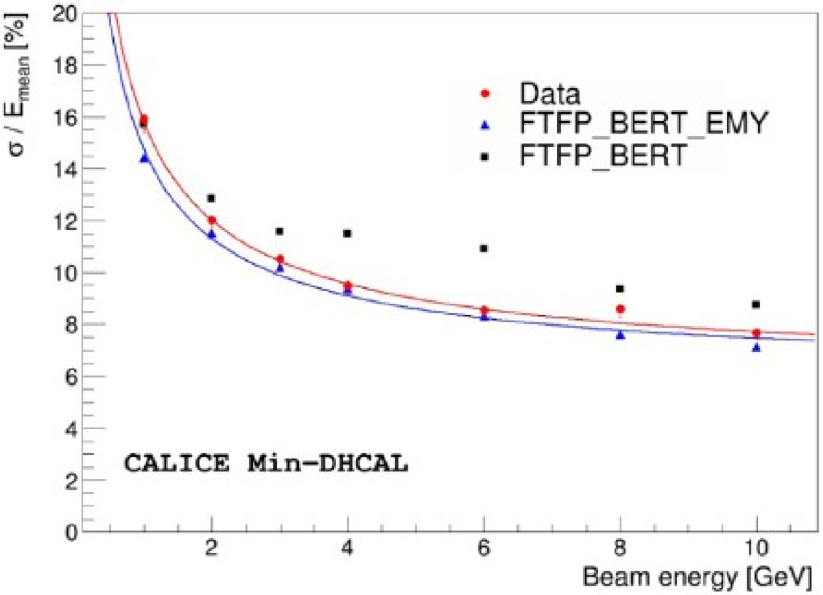
ElectroMagnetic (EM) Physics

DHCAL – Electromagnetic response

2016 JINST 11 P05008; e-print: arXiv:1603.01652

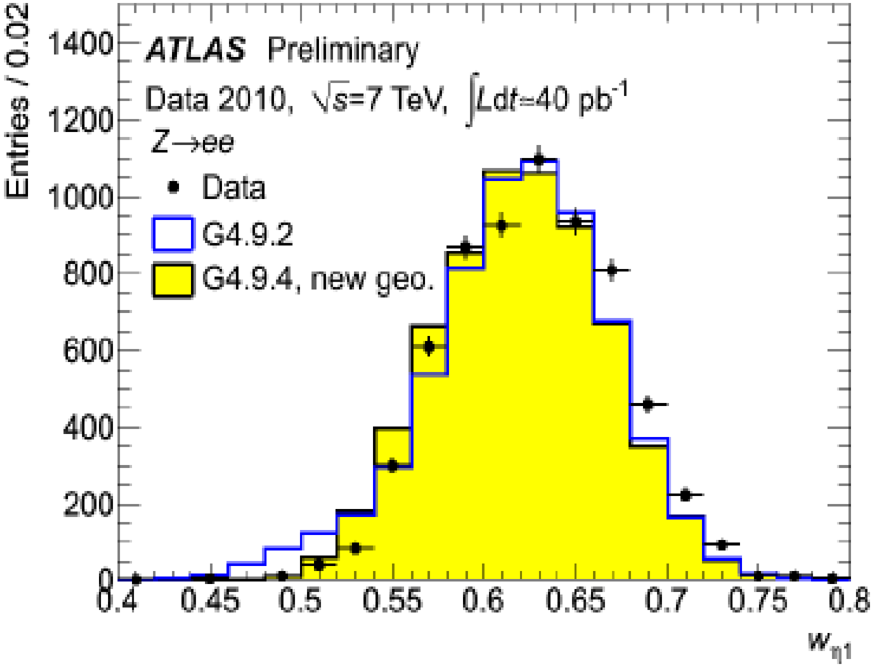


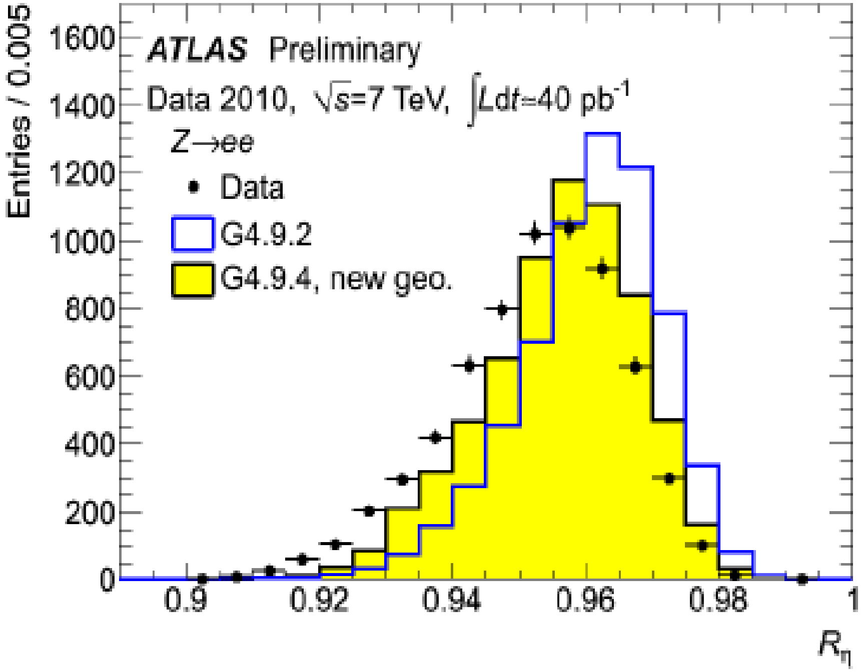




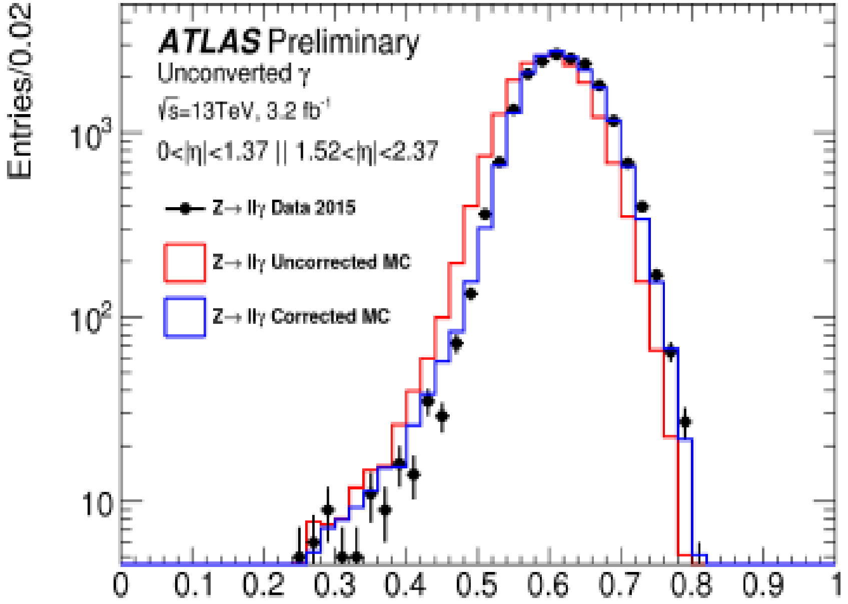
- Number of hits scales with energy
- Satisfactory description of mean hit number
- Resolution requires EMY option for electromagnetic cascade
  - Smaller step size => less fluctuation

ATLAS experiment: *slide from G. UNAL*





ATL-PHYS-PUB-2016-014

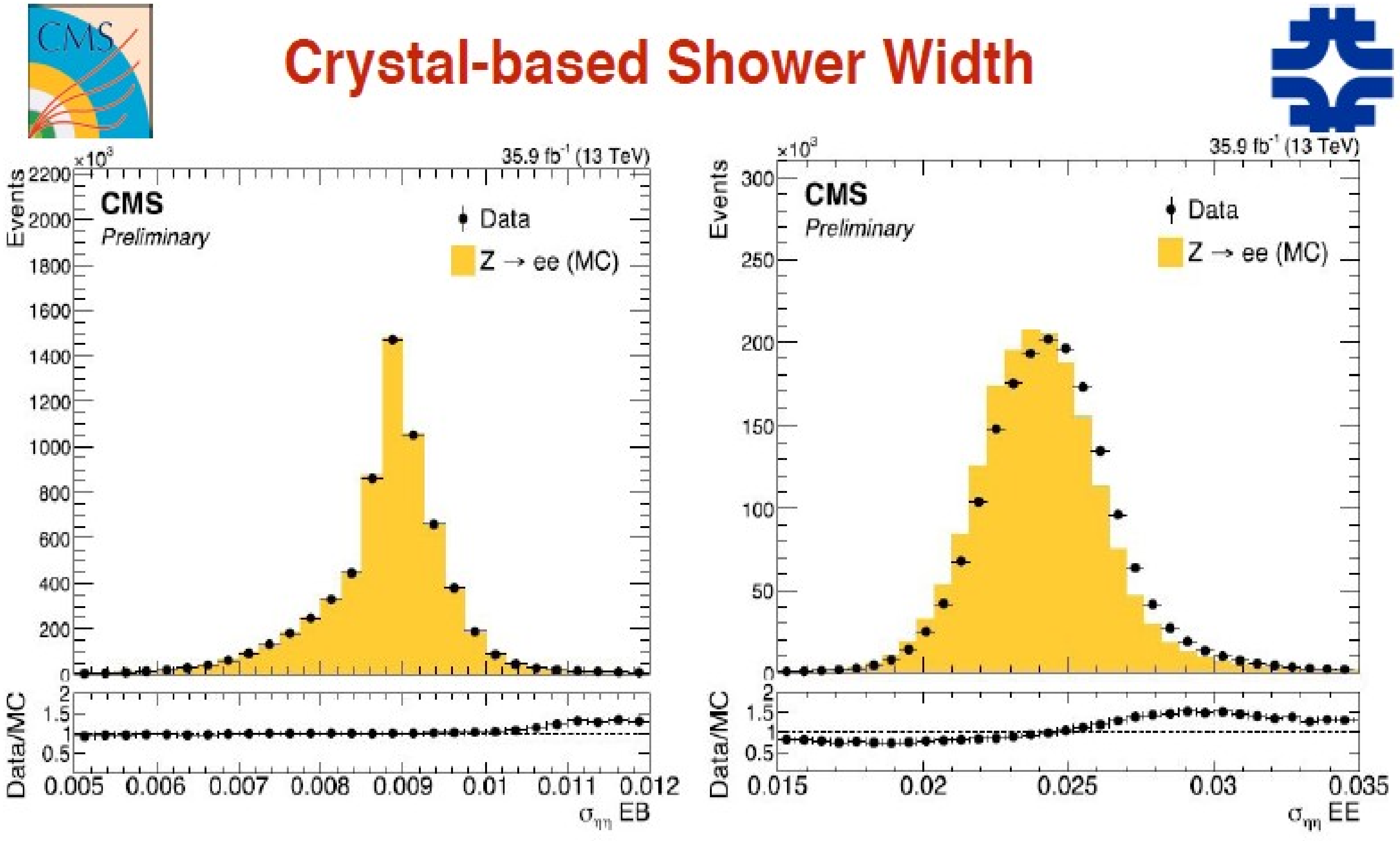


Shower shapes in eta-direction consistently wider in data than simulation both for electrons and photons

Original discrepancy observed beginning of run 1 was reduced using G4.9.4 instead of G4.9.2 (problem with blended material).

Keep using more detailed absorber geometry description

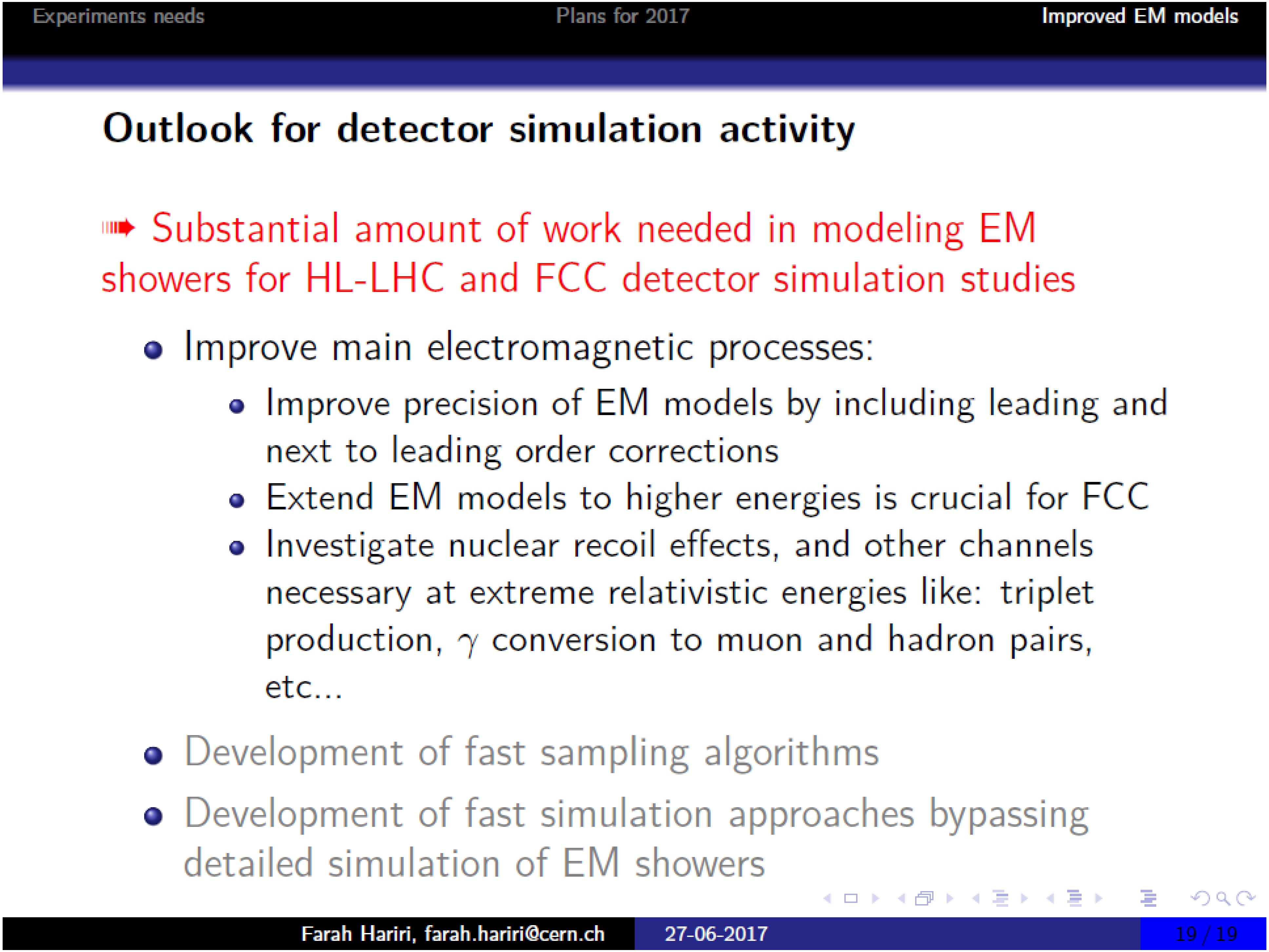
CMS experiment: *slide from S. Banerjee*



Crystal-based Shower Width

$\sigma_{\eta\eta} = \sum_{5x5} (\eta_i - \eta_{seed})^2 (E_i/E_{seed})$

- Crystal based shower width in the η direction for EB and EE
  - Level of agreement worse in the endcap region (matreial effect?)



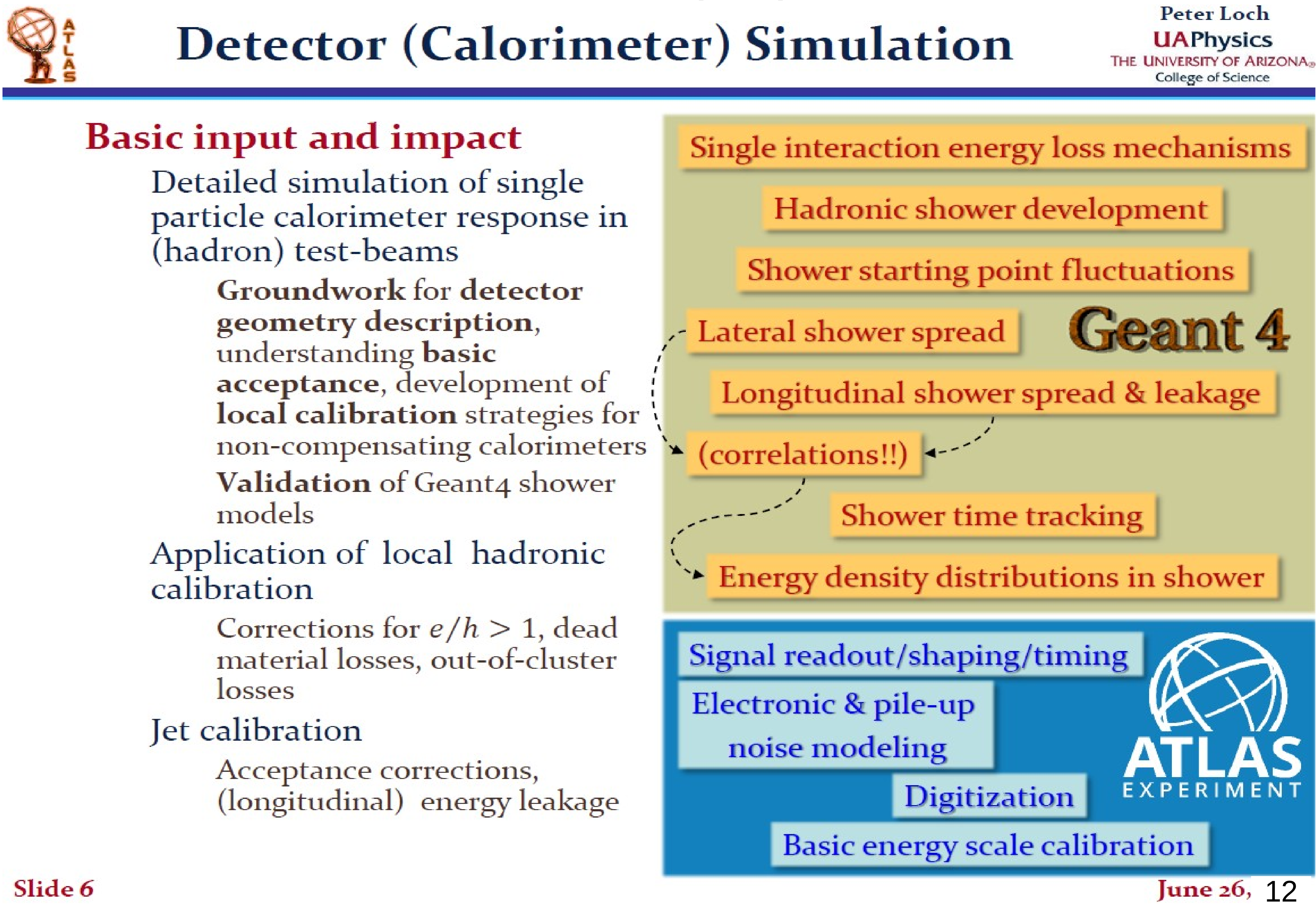
Outlook for detector simulation activity

➠ Substantial amount of work needed in modeling EM showers for HL-LHC and FCC detector simulation studies

- Improve main electromagnetic processes:
    - Improve precision of EM models by including leading and next to leading order corrections
    - Extend EM models to higher energies is crucial for FCC
    - Investigate nuclear recoil effects, and other channels necessary at extreme relativistic energies like: triplet production, γ conversion to muon and hadron pairs, etc...
- Development of fast sampling algorithms
- Development of fast simulation approaches bypassing detailed simulation of EM showers

Hadronic (HAD) Physics

ATLAS (1/3)



Peter Loch
UAPhysics
The University of Arizona
College of Science

Detector (Calorimeter) Simulation

Basic input and impact

- Detailed simulation of single particle calorimeter response in (hadron) test-beams
  - **Groundwork** for **detector geometry description**, understanding **basic acceptance**, development of **local calibration** strategies for non-compensating calorimeters
  - **Validation** of Geant4 shower models
- Application of local hadronic calibration
  - Corrections for $e/h > 1$, dead material losses, out-of-cluster losses
- Jet calibration
  - Acceptance corrections, (longitudinal) energy leakage

Single interaction energy loss mechanisms

Hadronic shower development

Shower starting point fluctuations

Lateral shower spread

Geant 4

Longitudinal shower spread & leakage

(correlations!!)

Shower time tracking

Energy density distributions in shower

Signal readout/shaping/timing

Electronic & pile-up noise modeling

Digitization

Basic energy scale calibration

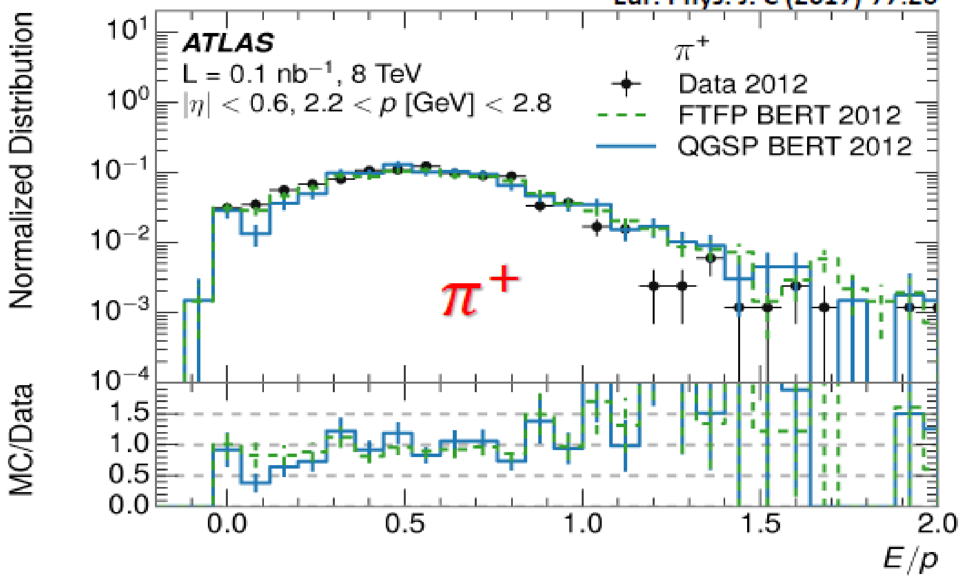
Slide 6

June 26,

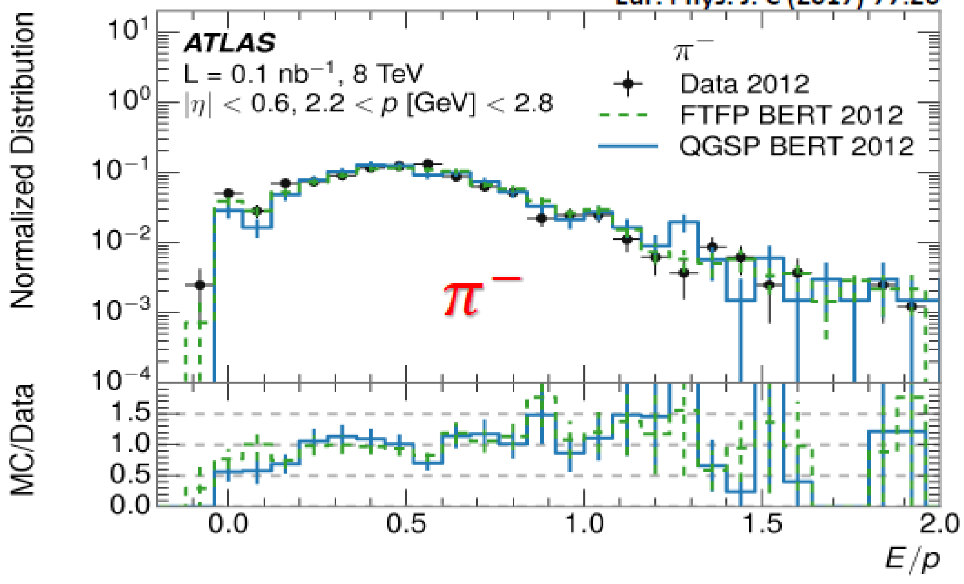
ATLAS (2/3)

E/p for Identified Particles

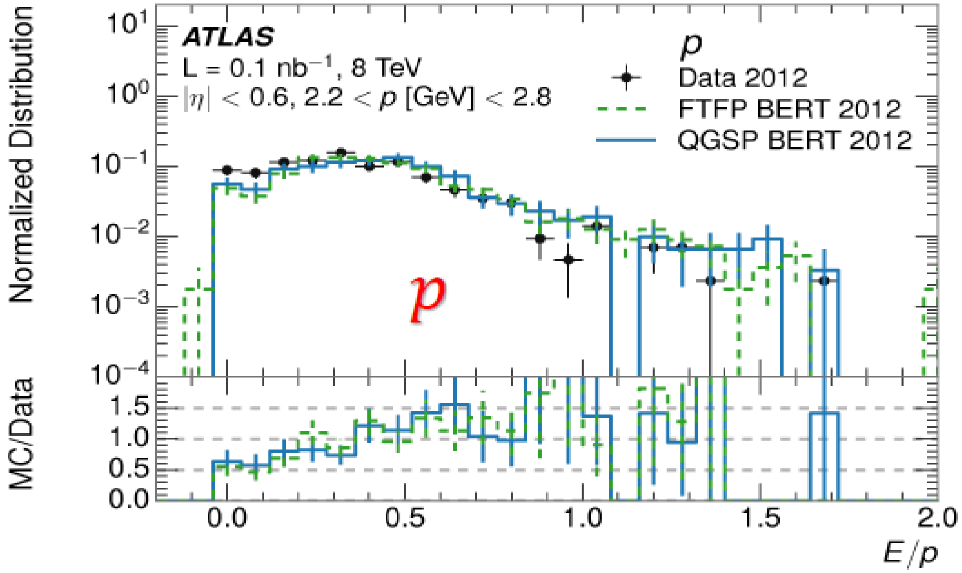
Eur. Phys. J. C (2017) 77:26



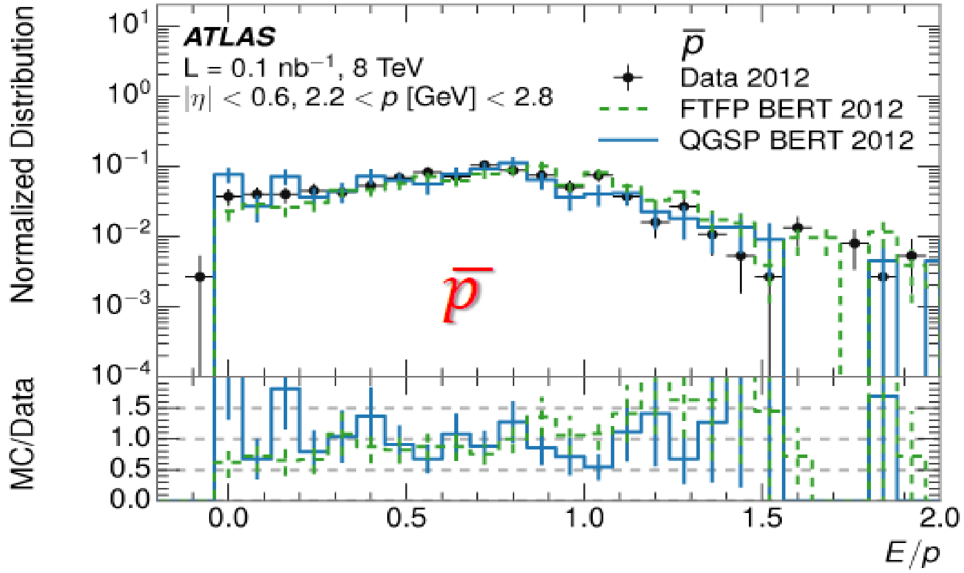
Eur. Phys. J. C (2017) 77:26



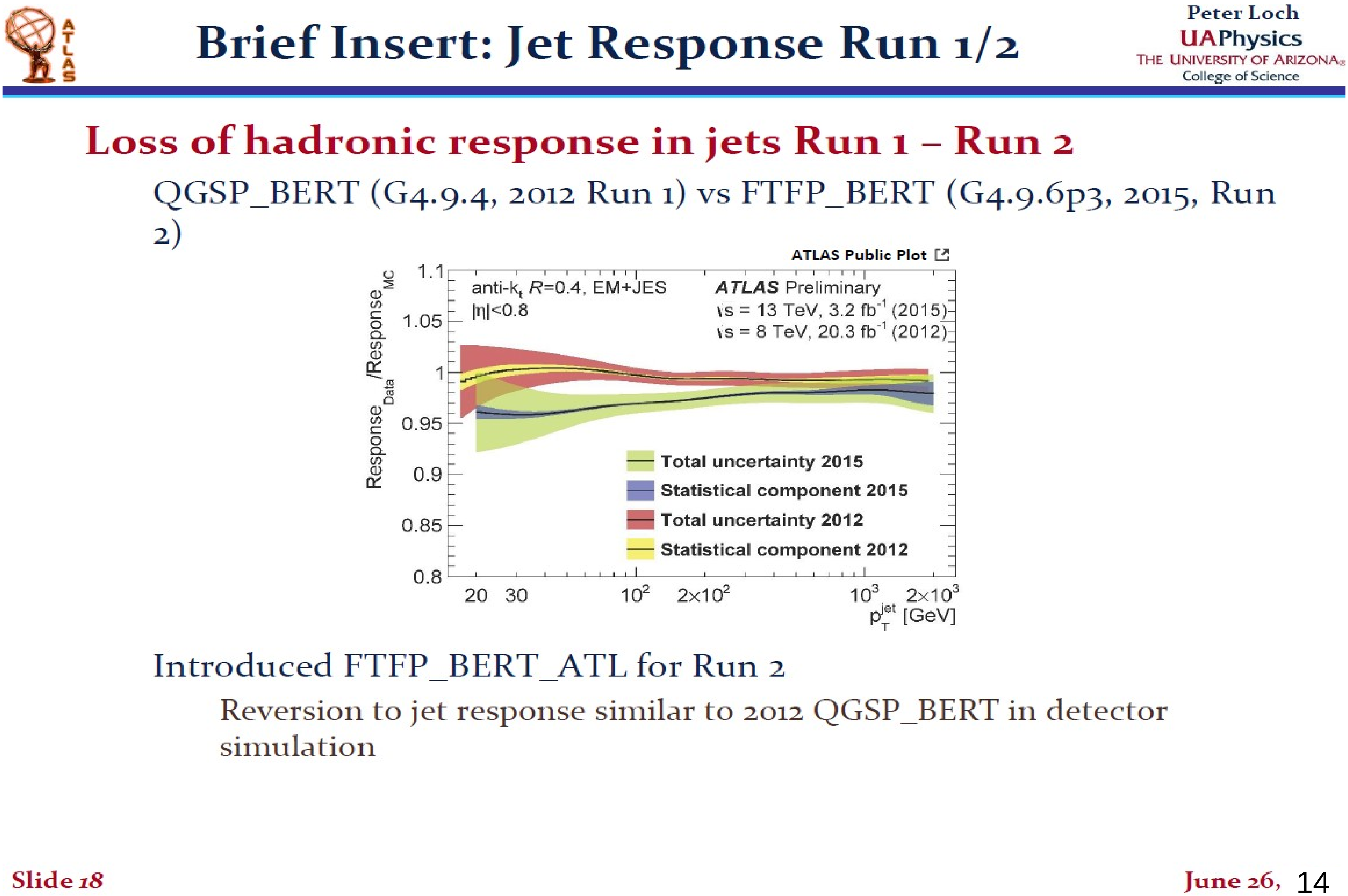
Eur. Phys. J. C (2017) 77:26



Eur. Phys. J. C (2017) 77:26



ATLAS (3/3)



Brief Insert: Jet Response Run 1/2

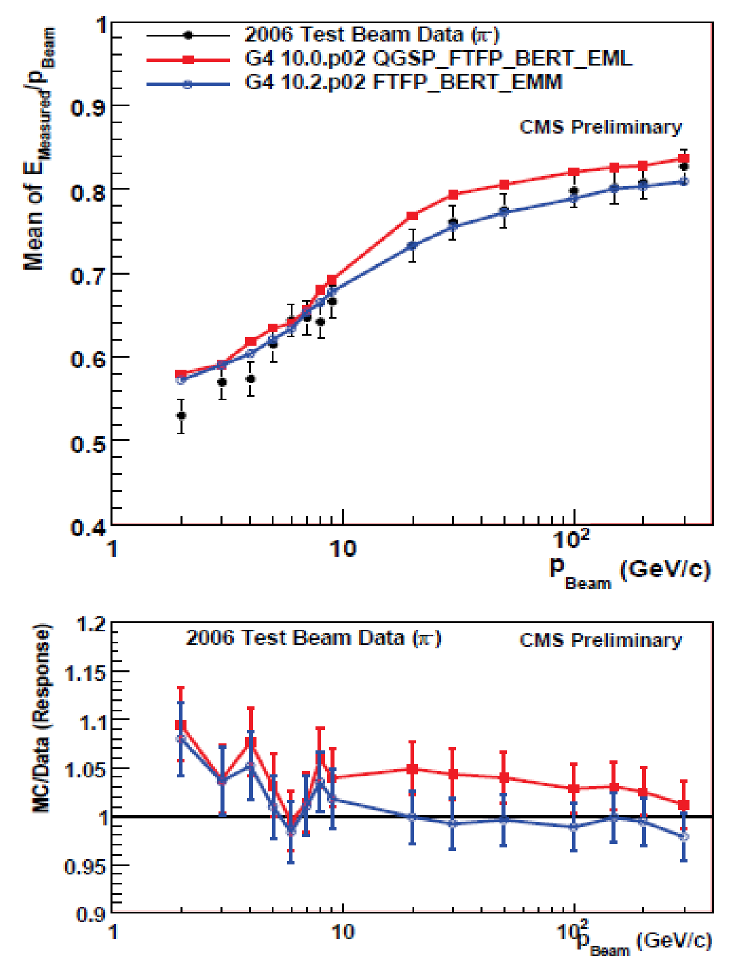
Loss of hadronic response in jets Run 1 – Run 2

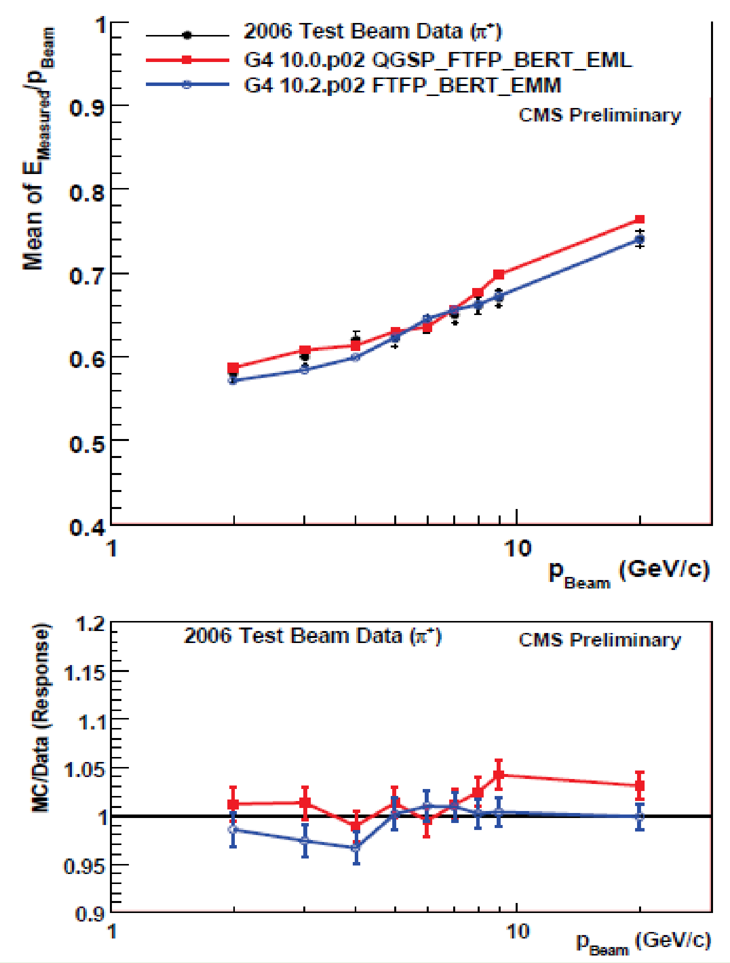
QGSP_BERT (G4.9.4, 2012 Run 1) vs FTFP_BERT (G4.9.6p3, 2015, Run 2)

Introduced FTFP_BERT_ATL for Run 2

- Reversion to jet response similar to 2012 QGSP_BERT in detector simulation

Mean response with pions





Main improvement for E > 10 GeV due to use FTFP and EMM

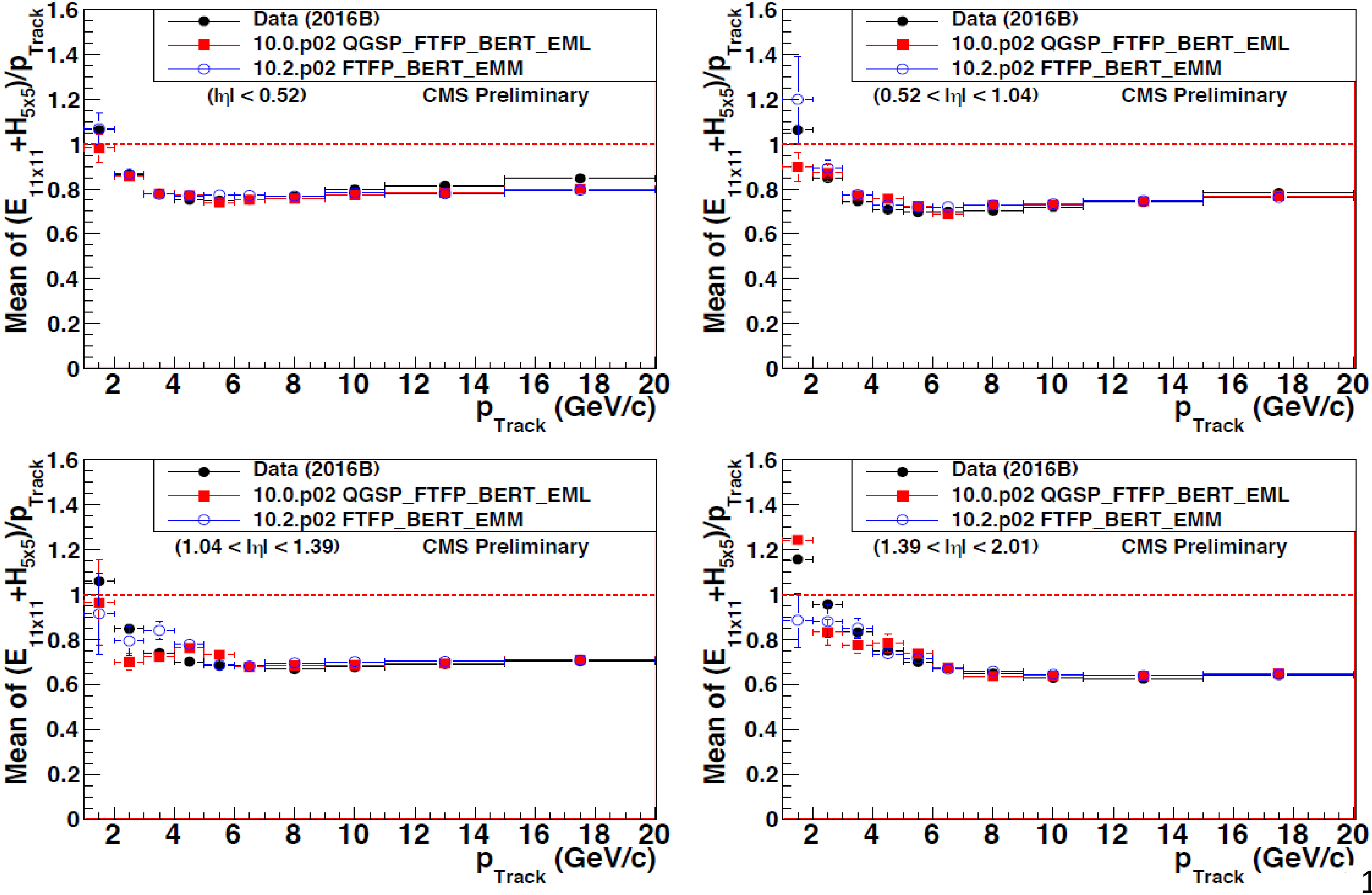
Summary from mean response

- The level of agreement between data and MC improve in the new model for pions
  - More discrepancies for pbar and kaons
- pp collisions at high energies produce mostly pions
  - one expects to have a better agreement between data and MC with the new physics list and 10.2.p02

	Pi-	Pi+	p	pbar
G4 10.0p02 QGSP_FTFP_BERT_EML	(3.6 ± 0.6)%	(1.9 ± 0.5)%	(4.3 ± 1.0)%	(3.5 ± 0.8)%
G4 10.2p02 FTFP_BERT_EMM	(1.8 ± 0.7)%	(1.0 ± 0.5)%	(2.2 ± 1.1)%	(3.1 ± 0.8)%

CMS (3/4)

Combined Calorimeter Energy (11x11+5x5 matrix)



CMS (4/4)

Collision Data

- The level of disagreement between data and MC is between 2 to 5% depending on the region of the detector as well as the physics list used

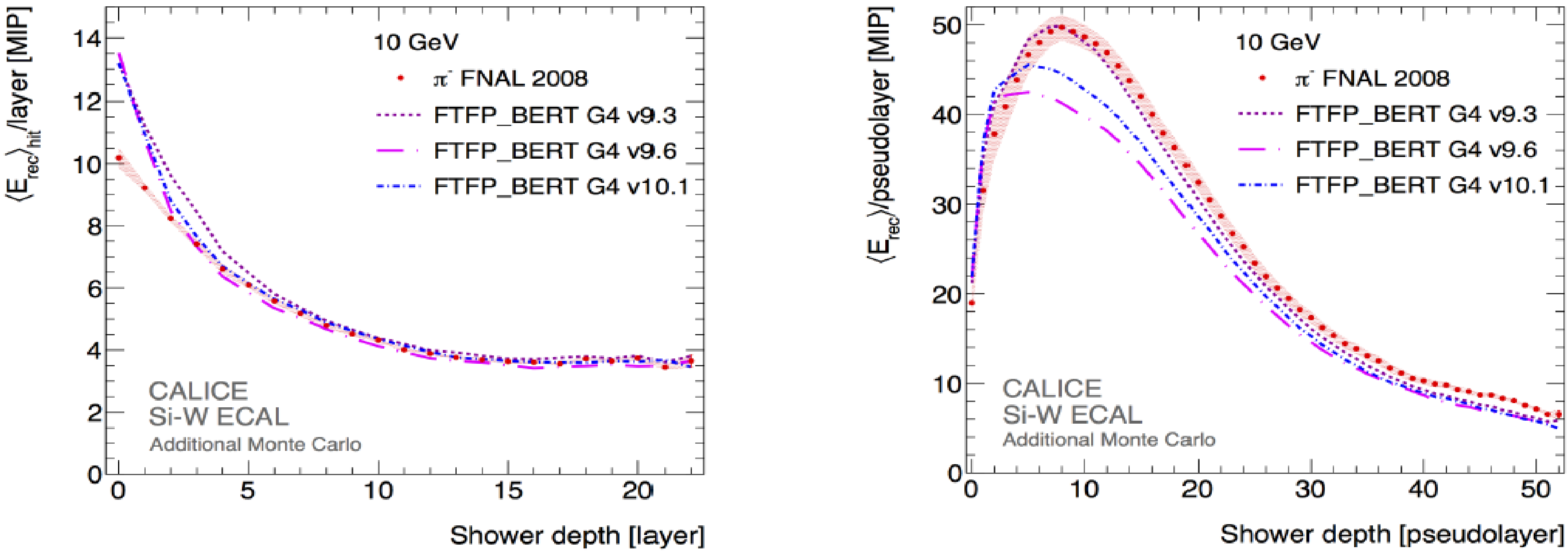
Mean level of disagreement between MC and data

	$(E_{7x7}+H_{3x3})/p$ 10.0.p02	$(E_{7x7}+H_{3x3})/p$ 10.2.p02	$(E_{11x11}+H_{5x5})/p$ 10.0.p02	$(E_{11x11}+H_{5x5})/p$ 10.2.p02
Barrel 1	(1.1±0.4)%	(2.4±0.4)%	(2.5±0.4)%	(2.6±0.4)%
Barrel 2	(3.4±0.4)%	(3.6±0.4)%	(1.9±0.4)%	(2.2±0.4)%
Transition	(3.7±0.5)%	(4.9±0.5)%	(1.6±0.5)%	(2.2±0.5)%
Endcap	(1.1±0.3)%	(4.1±0.5)%	(4.7±0.4)%	(1.6±0.5)%

CALICE (1/6)

Si ECAL: Longitudinal shower profiles

NIM A794 (2015) 240-254; e-print: arXiv:1411.7215, CAN-050a

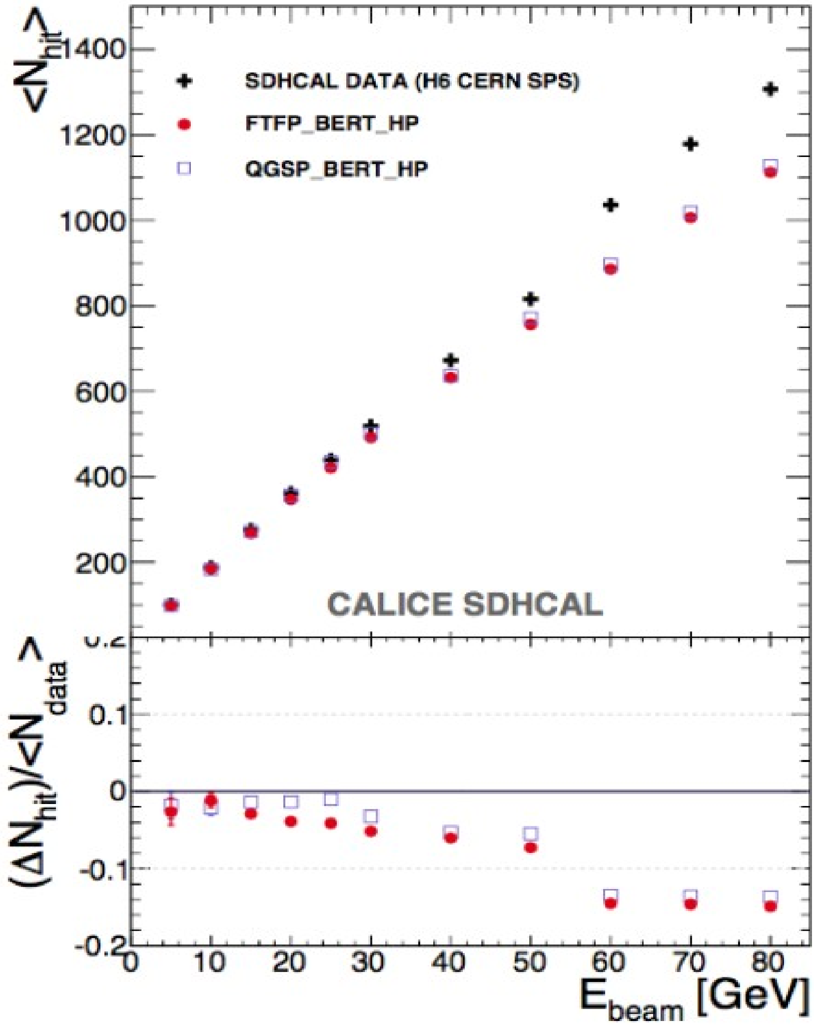


- Big change observed in FTFP_BERT observed between Geant4 versions 9.3 and 9.6
- Only observed in silicon, not for scintillator prototypes; tuning of Geant4 parameters on thin target scintillator data?
- Bug in Geant4 v9.6, fixed in v10.0, however still insufficient energy in v10.1
- Disagreement in individual hit energies between data and Geant4 affects longitudinal profile

CALICE (2/6)

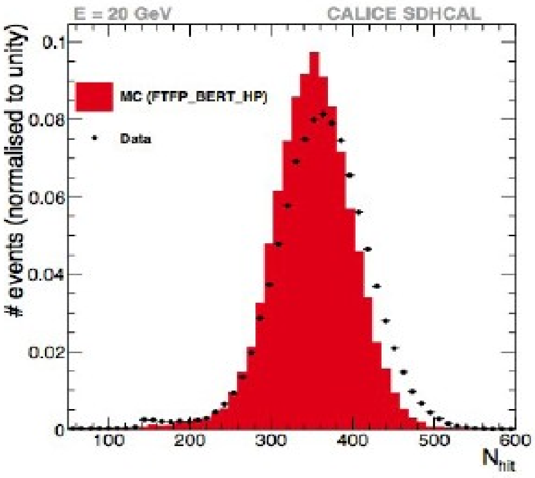
SDHCAL – Number of hits

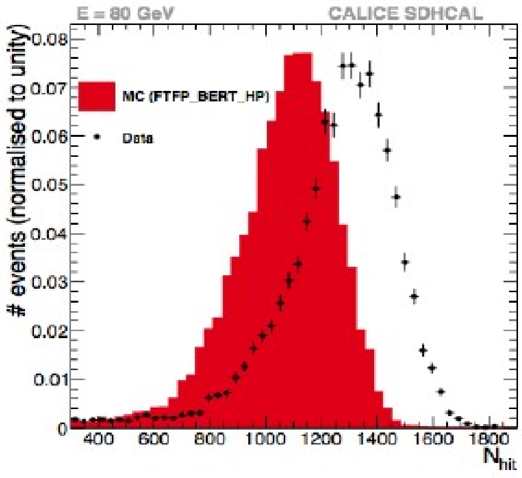
JINST 11 (2016) P04001; e-print: arXiv:1602.02276



- active material: gas
- absorber: steel
- Geant4 9.6

- Number of hits underestimated at high beam energies
- Cross checks for beam composition or difference in elm. Response negative
- Same observation in CALICE DHCAL (also gaseous)

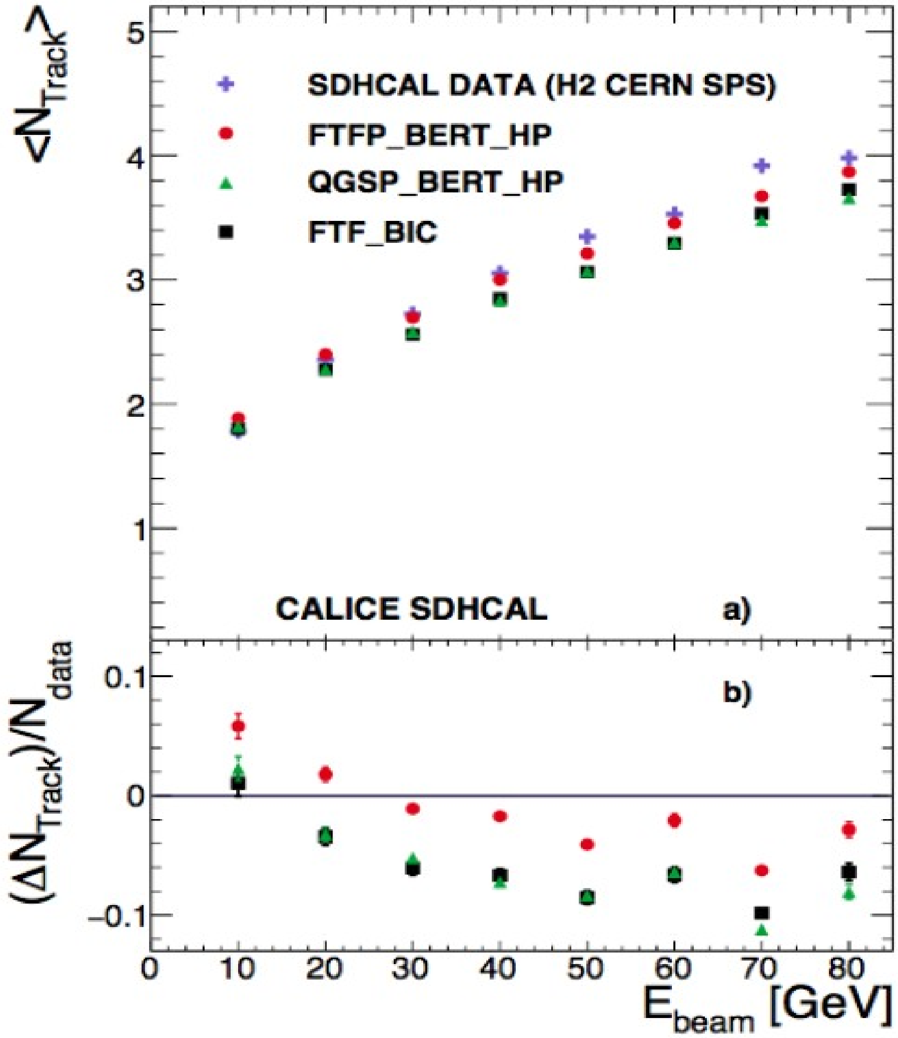




SDHCAL – Number of tracks II

JINST 12 (2017) no.05, P05009,arXiv:1702.08082

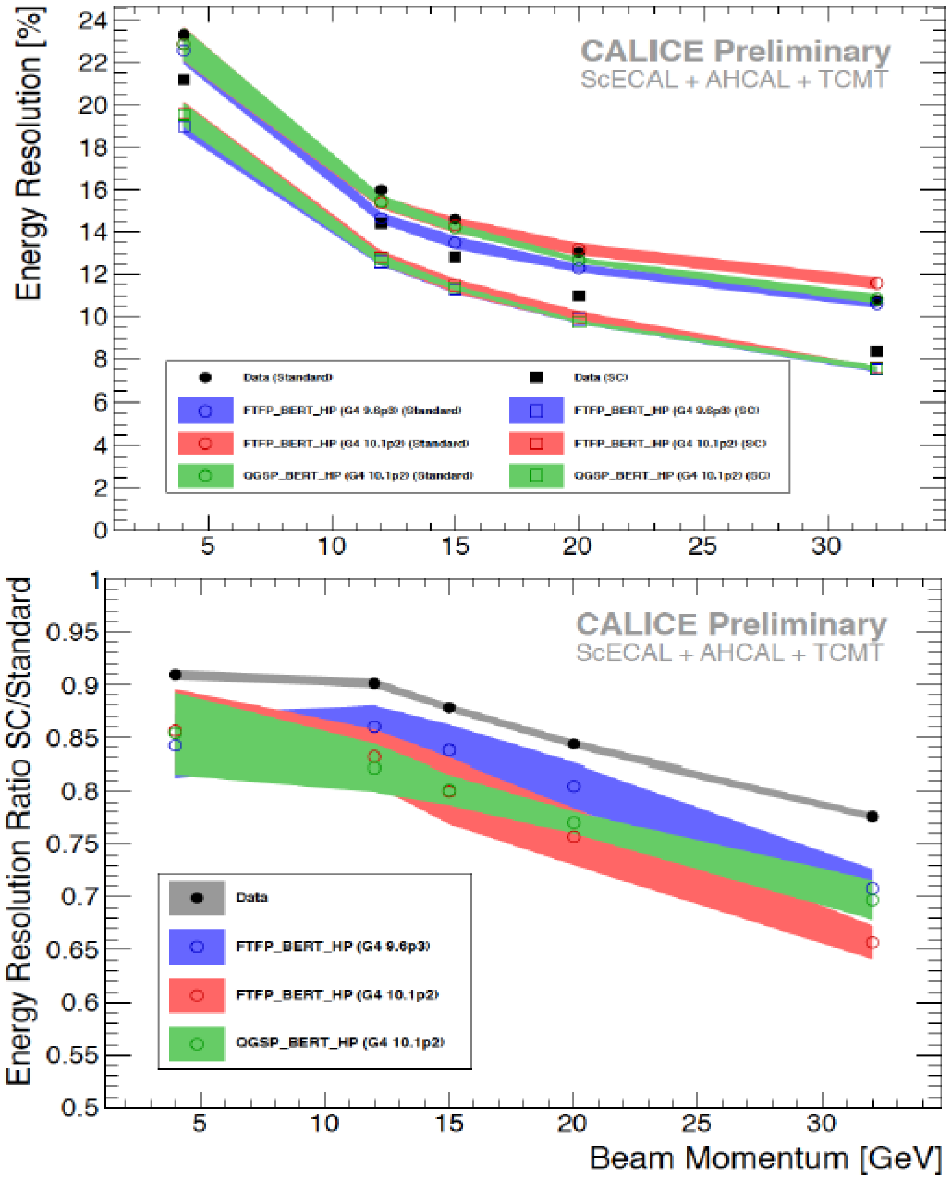
- active material: gas
- absorber: steel
- Geant4 9.6



- Clearer picture in summary plot
- FTFP_BERT_HP superior to other physics lists
- Towards higher energies MC predictions systematically below data
- However, no “catastrophy” discrepancies max. 10%

CALICE (4/6)

ScECAL + AHCAL – reconstruction methods



CAN-056

- active material: scintillator
- absorber:
  - tungsten in ScECAL
  - steel in AHCAL

Two methods for energy reconstruction:

- “standard”: sum of energy depositions corrected for absorber thickness
- software compensation (“SC”): energy depositions weighted by weight depending on hit energy density
→ sensitive to shower structure

- energy resolution for standard rec. reasonably well described

- resolution improvement due to software compensation overestimated by ~5%

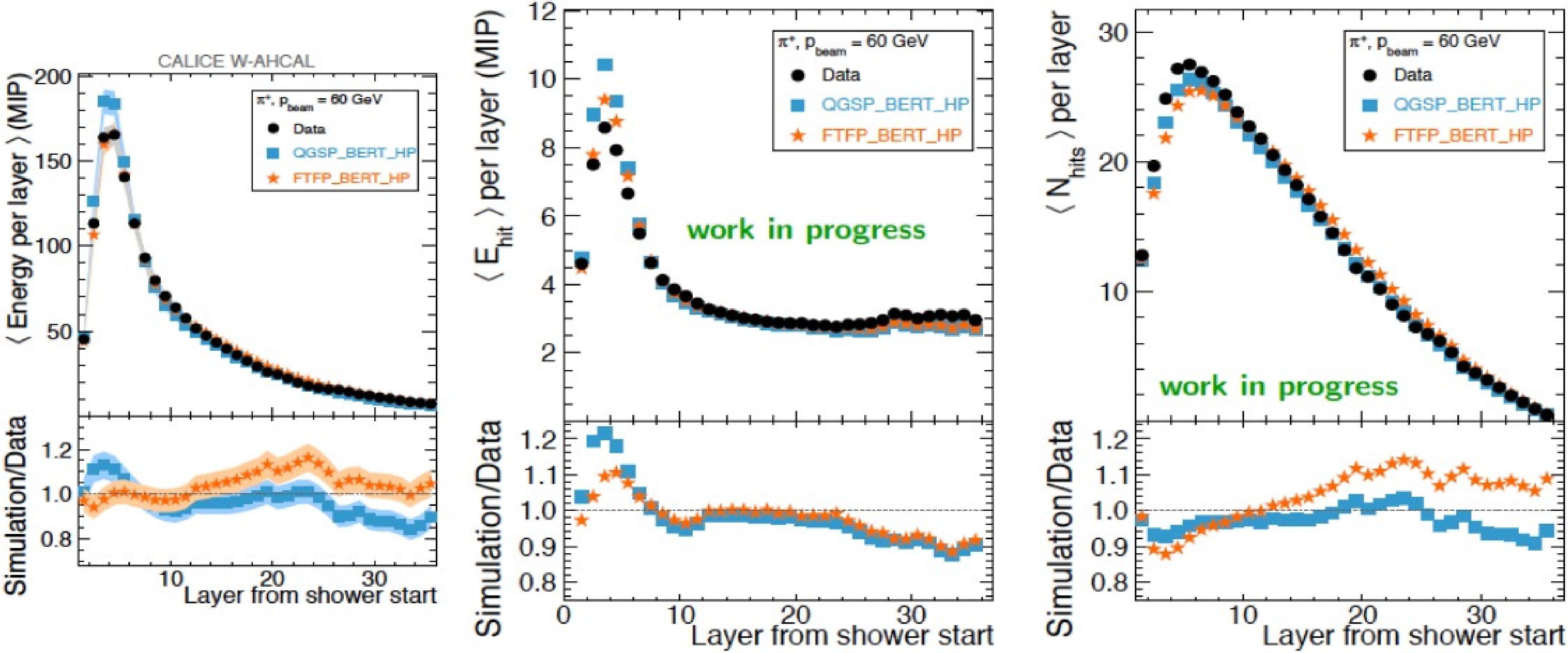
- overall: scintillator ECAL + HCAL system reasonably well described, QGSP_BERT_HP in details a bit better than FTFP_BERT_HP

CALICE (5/6)

W-AHCAL: pions

- similar observation as for electrons: energy per layer reasonably well described, but a tendency for too few hits with too high energy in first layers

- active material: scintillator
- absorber: tungsten
- Geant4 9.6p02



CALICE (6/6)

“Quality Matrix”

Electromagnetic showers: generally well described
Exceptions:

- detailed shower mode (option EMY) needed for gaseous calorimeters
- number of hits and hit sizes at shower start for scintillator/tungsten

Hadronic showers

	W	Fe
Scintillator	✓ energy (✓) longitudinal profile ✗ shower start: # of hits, hit size	✓ energy ✓ longitudinal profile ✓ tracks
Silicon	✓ energy ✗ longitudinal profile ✓ tracks, angles	n.A.
Gas	?	✗ number of hits (✓) tracks

CAVEAT: sometimes significant changes between Geant4 versions

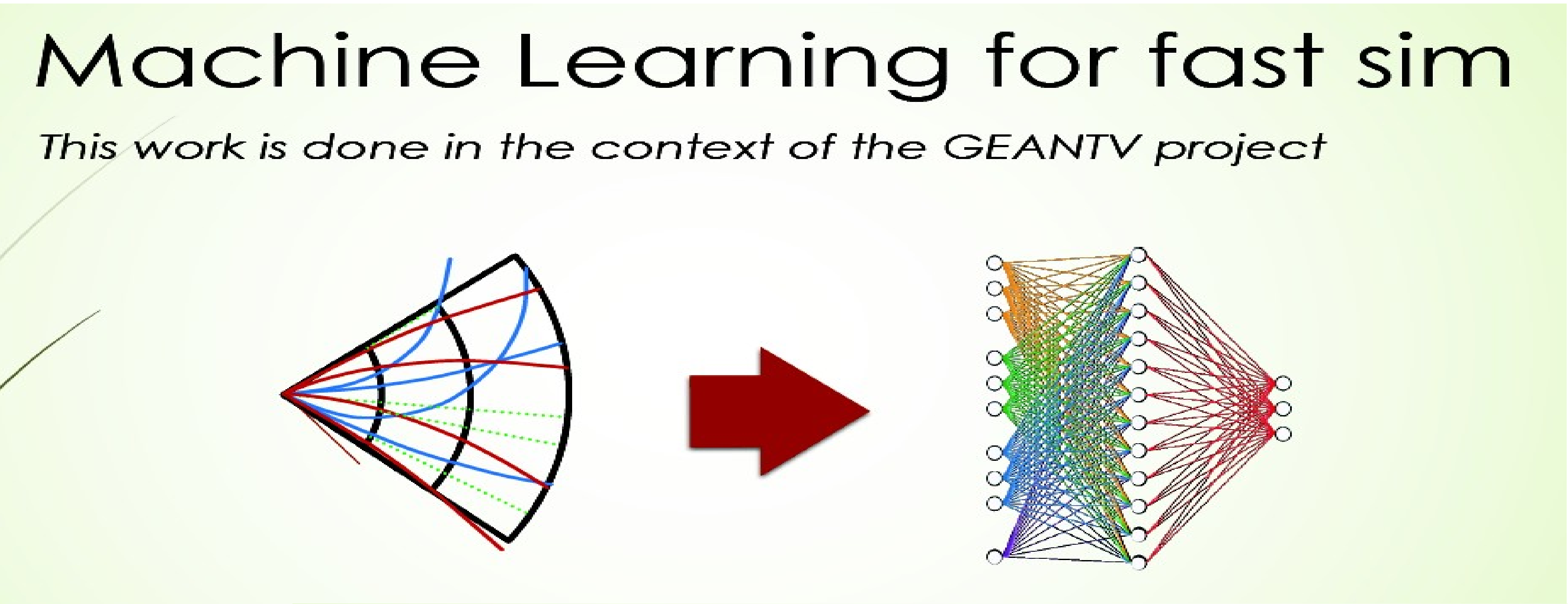
Summary & Outlook of Hadronic Physics

- On-going **consolidation** of Geant4 hadronics in terms of: **physics** & **algorithms**, and **validation testing suite**
  - Bottom-up review of physics models (theory, approximations and implementation, including rewriting from scratch) very useful ! Prefer a more theory-based model than a phenomenological one...

- For HEP applications, in particular for hadronic showers, the main G4 hadronic models – **FTF** , **BERT** , **Preco** – seem to be reaching their “potential limits”...
  - More promising developments instead in low-energy hadronics, driven by medical and nuclear physics...

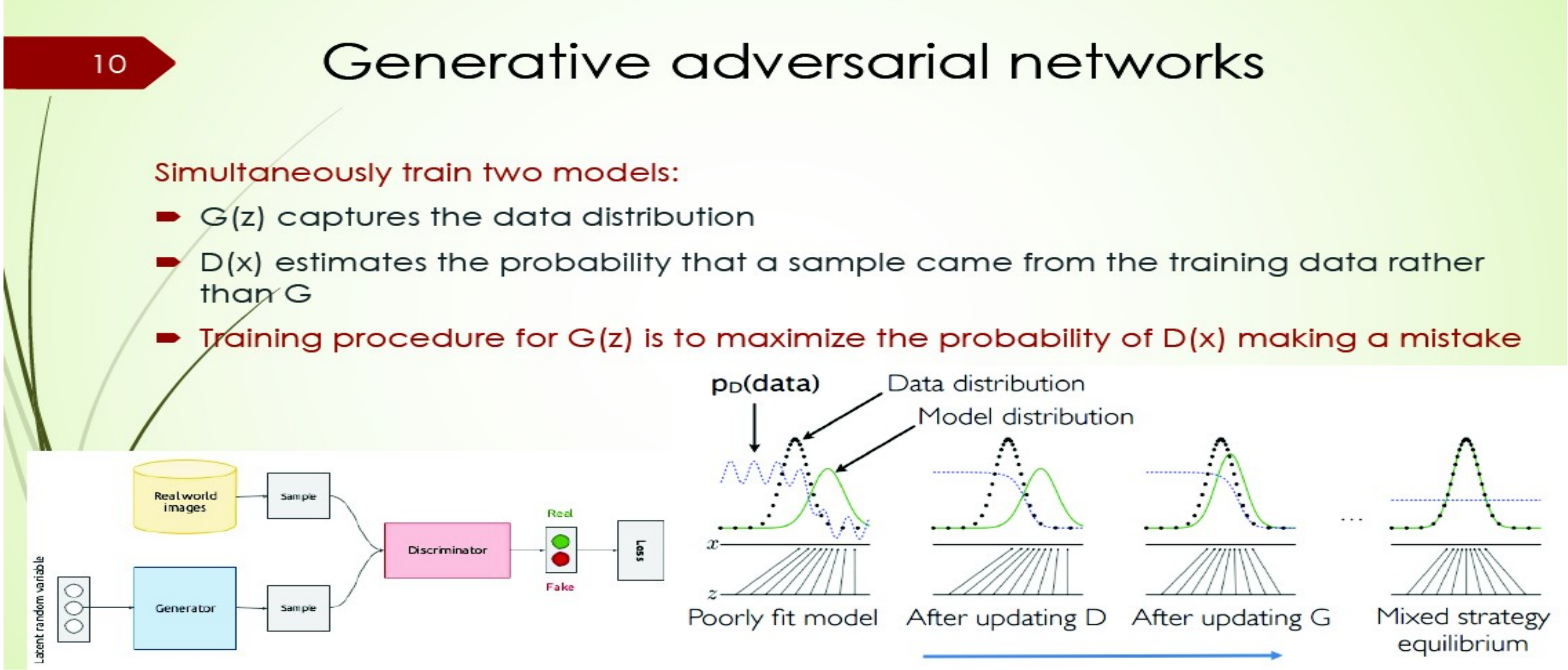
- Future major developments for HEP applications are likely to come from alternative string models:
  - **QGS** in the short & medium term
  - **EPOS** in the long term

Fast Simulation (FastSim)



Machine Learning for fast sim

This work is done in the context of the GEANTV project

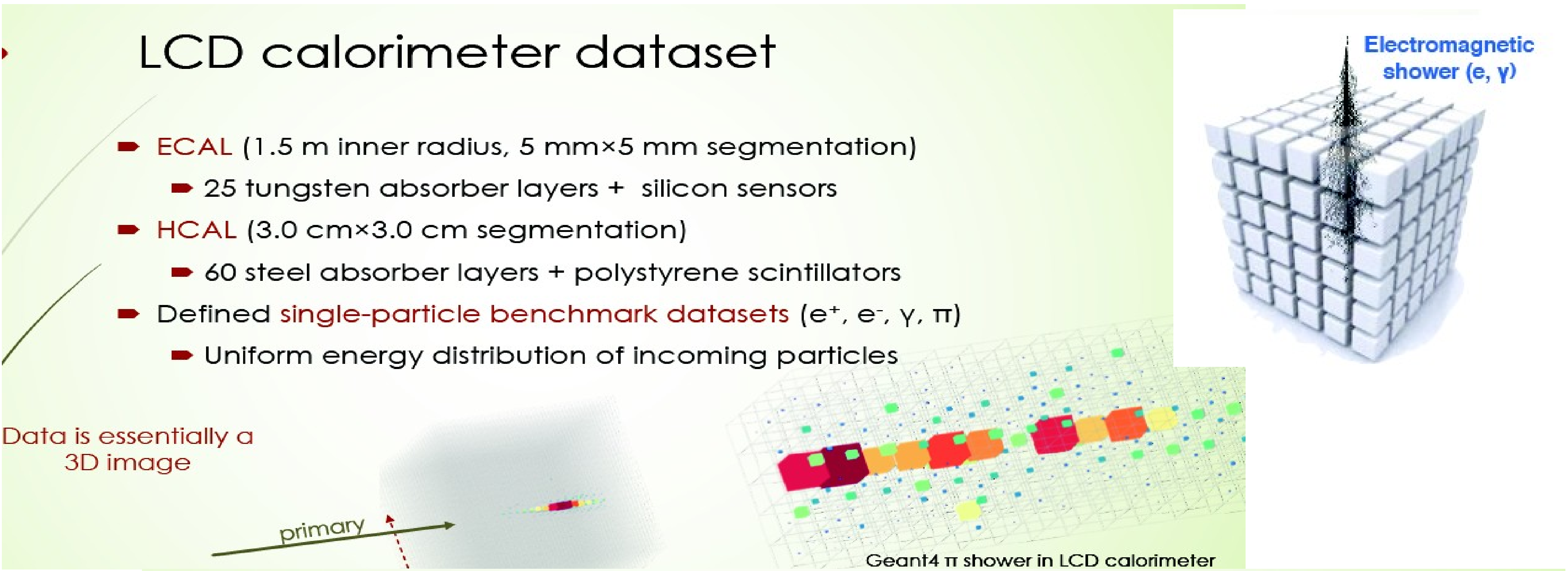


Generative adversarial networks

Simultaneously train two models:

- G(z) captures the data distribution
- D(x) estimates the probability that a sample came from the training data rather than G
- Training procedure for G(z) is to maximize the probability of D(x) making a mistake

Sofia Vallecorsa “Machine Learning-based fast simulation for GeantV”



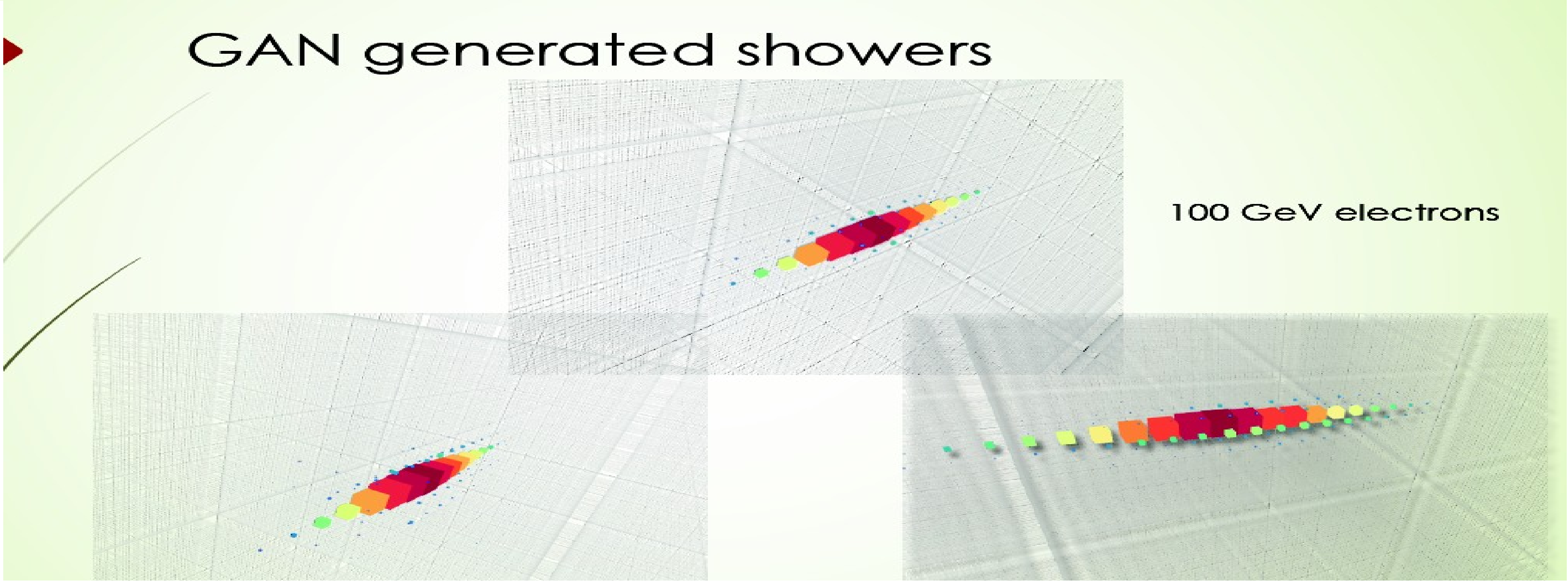
LCD calorimeter dataset

- ECAL (1.5 m inner radius, 5 mm×5 mm segmentation)
  - 25 tungsten absorber layers + silicon sensors
- HCAL (3.0 cm×3.0 cm segmentation)
  - 60 steel absorber layers + polystyrene scintillators
- Defined single-particle benchmark datasets (e^+, e^-, γ, π)
  - Uniform energy distribution of incoming particles

Data is essentially a 3D image

Geant4 π shower in LCD calorimeter

Electromagnetic shower (e, γ)

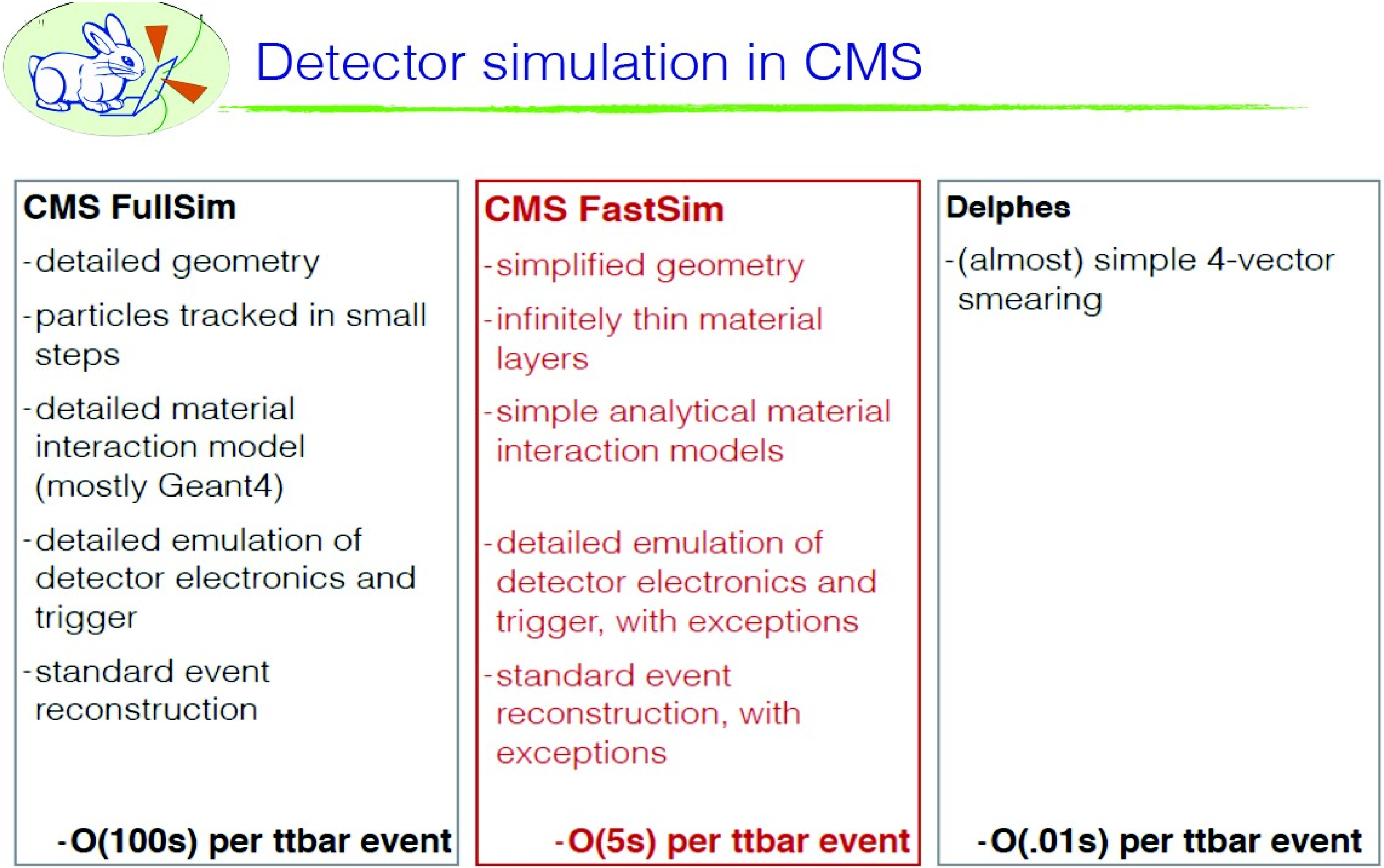


GAN generated showers

100 GeV electrons

Sofia Vallecorsa “Machine Learning-based fast simulation for GeantV”

CMS Fast Sim (1/2)



Detector simulation in CMS

CMS FullSim

- detailed geometry
- particles tracked in small steps
- detailed material interaction model (mostly Geant4)
- detailed emulation of detector electronics and trigger
- standard event reconstruction

-O(100s) per ttbar event

CMS FastSim

- simplified geometry
- infinitely thin material layers
- simple analytical material interaction models
- detailed emulation of detector electronics and trigger, with exceptions
- standard event reconstruction, with exceptions

-O(5s) per ttbar event

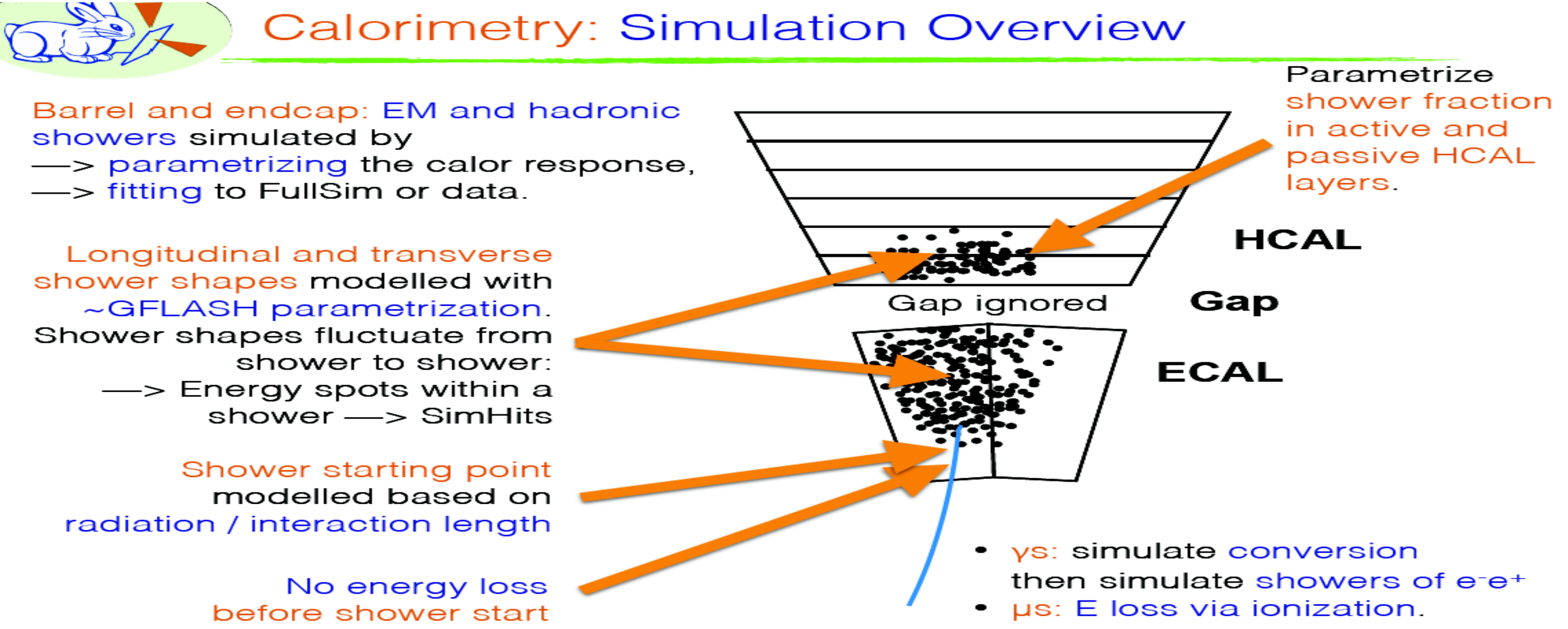
Delphes

- (almost) simple 4-vector smearing

-O(.01s) per ttbar event

Sezen Sekmen “Fast Simulation in CMS”

CMS Fast Sim (2/2)



Calorimetry: Simulation Overview

Barrel and endcap: EM and hadronic showers simulated by
—> parametrizing the calor response,
—> fitting to FullSim or data.

Longitudinal and transverse shower shapes modelled with ~GFLASH parametrization.
Shower shapes fluctuate from shower to shower:
—> Energy spots within a shower —> SimHits

Shower starting point modelled based on radiation / interaction length

No energy loss before shower start

Parametrize shower fraction in active and passive HCAL layers.

- γs: simulate conversion then simulate showers of e^-e^+
- μs: E loss via ionization.

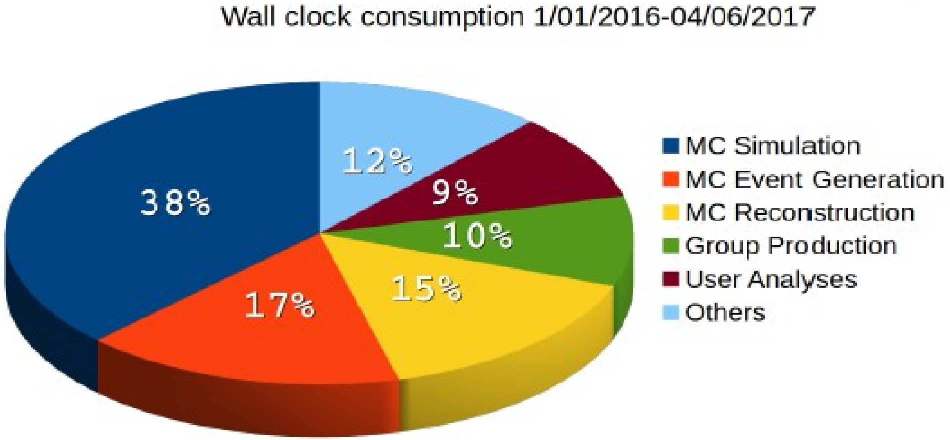
Calorimetry: HF showers

- We use a shower library simulated using GEANT4 to simulate the showers in the hadron forward calorimeter (HF) region.
- Showers are classified according to particle type, η and energy.
- Material in front of the HF is not modelled. We recently applied FastSim-specific correction factors to correct the incomplete modelling.

Sezen Sekmen “Fast Simulation in CMS”

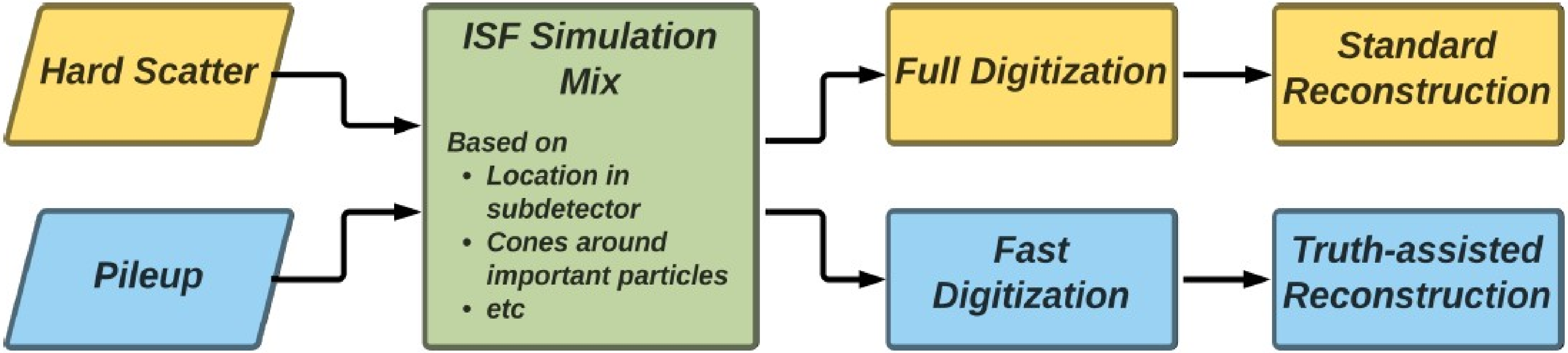
ATLAS Fast Sim (1/3)

- High CPU requirement for MC generation, detector simulation and reconstruction.
- ~70% of the CPU usage in 2016 and 2017 (till now) for MC simulation.
- Not possible to sustain this CPU requirement at high luminosity and pile-up expected in next LHC runs.



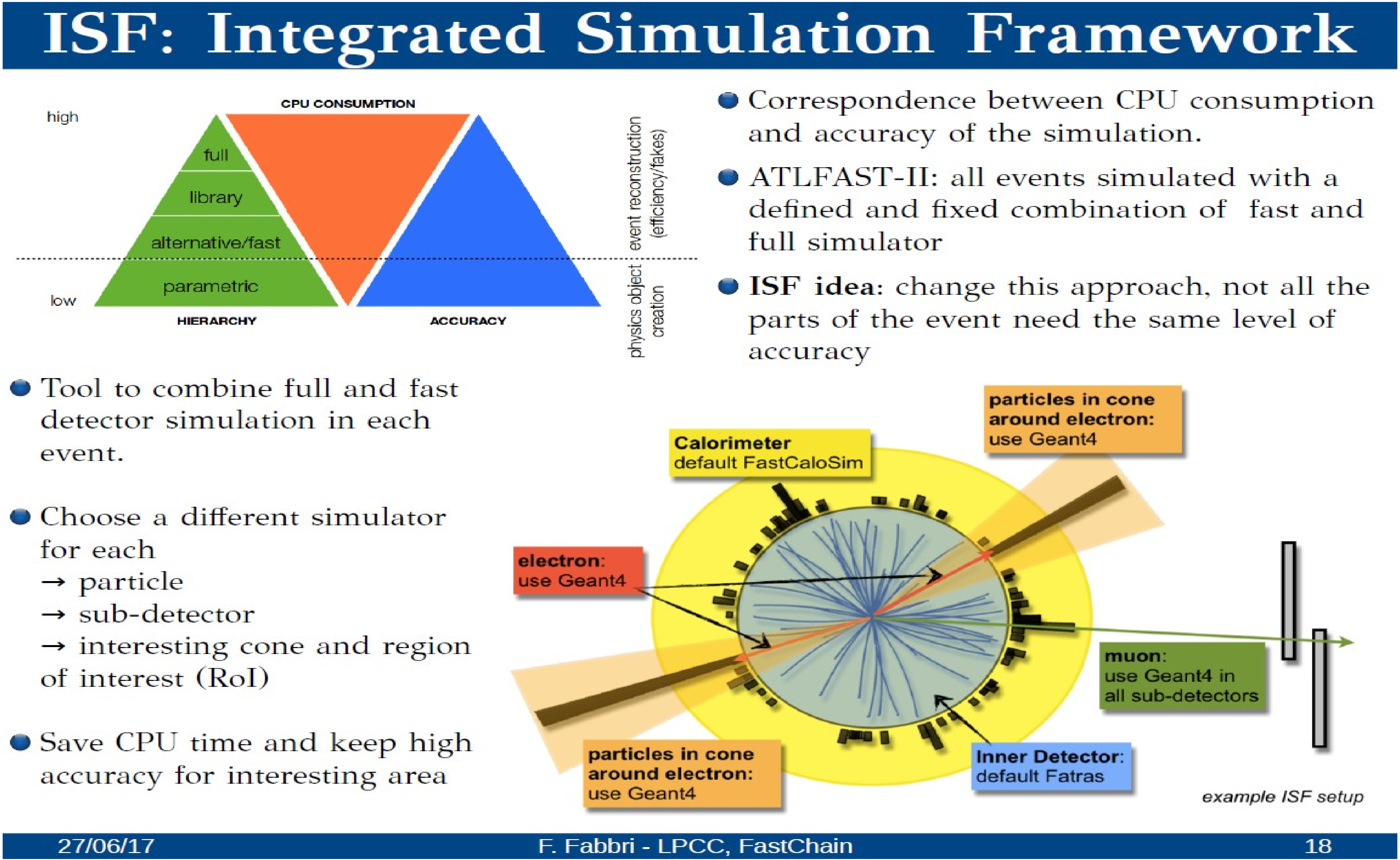
Fast chain idea:

- Speed-up all the MC simulation steps
- Combine the fast and full simulation to maintain the highest accuracy in interesting events/part of the events



Federica Fabbri “Fast Simulation Chain in ATLAS”

ATLAS Fast Sim (2/3)



ISF: Integrated Simulation Framework

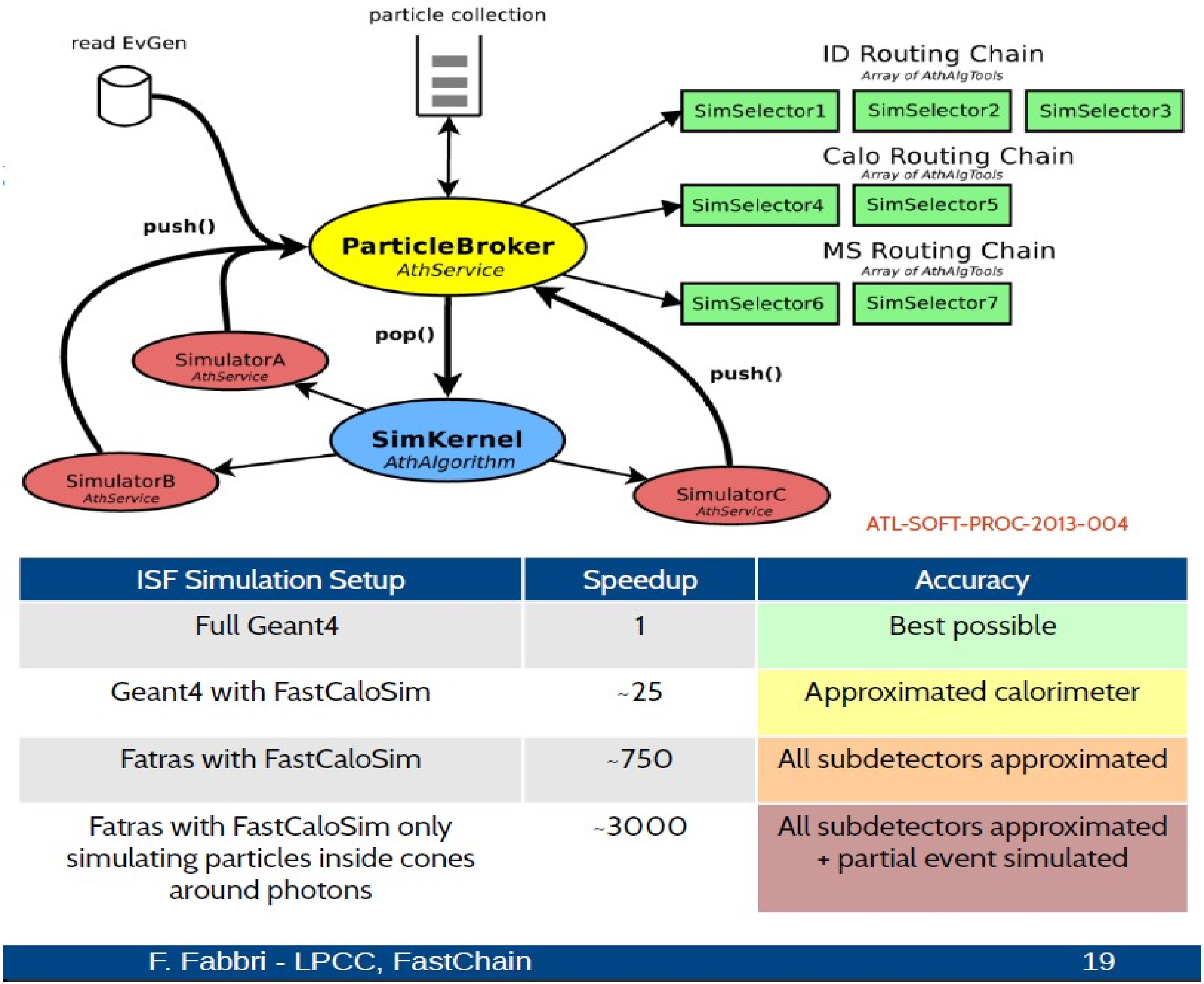
- Correspondence between CPU consumption and accuracy of the simulation.
- ATLFAST-II: all events simulated with a defined and fixed combination of fast and full simulator
- **ISF idea**: change this approach, not all the parts of the event need the same level of accuracy

- Tool to combine full and fast detector simulation in each event.
- Choose a different simulator for each
  → particle
  → sub-detector
  → interesting cone and region of interest (RoI)
- Save CPU time and keep high accuracy for interesting area

27/06/17 F. Fabbri - LPCC, FastChain 18

Federica Fabbri “Fast Simulation Chain in ATLAS”

ATLAS Fast Sim (3/3)



ATL-SOFT-PROC-2013-004

ISF Simulation Setup	Speedup	Accuracy
Full Geant4	1	Best possible
Geant4 with FastCaloSim	~25	Approximated calorimeter
Fatras with FastCaloSim	~750	All subdetectors approximated
Fatras with FastCaloSim only simulating particles inside cones around photons	~3000	All subdetectors approximated + partial event simulated

F. Fabbri - LPCC, FastChain 19

Federica Fabbri “Fast Simulation Chain in ATLAS”

Summary & Outlook for FastSim

- Fast simulation is playing an increasingly important role
  - Only way to cope with the increasing statistics of LHC data
- Lot of independent activities in each experiment
  - Fast simulation is very much detector-specific
- However, there are opportunities and interest for common activities between the experiments and Geant developers
  - Machine-learning approach to fast simulation
  - Generic tools that streamline tuning/training FastSim on FullSim

- Could it be useful to set-up a forum for discussing & meeting FastSim between the experiments and Geant developers ?

Conclusions

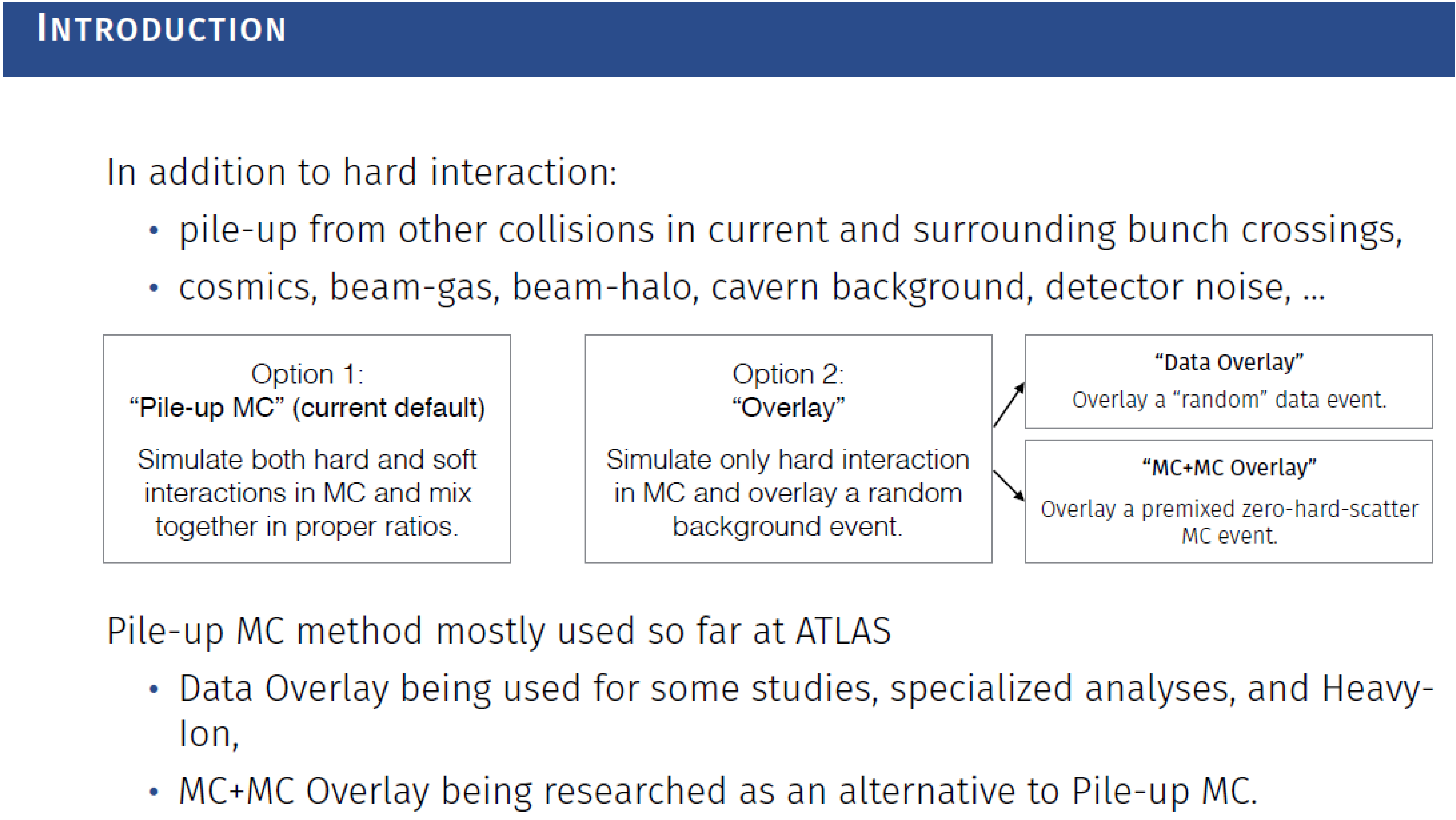
Conclusions

- Positive and useful workshop !
- We still need to “digest” all the presentations, and eventually adjust our work plan according to the experiments' needs and priorities
- Keeping the 3-year gap between LPCC Detector Simulation workshops seems resonable
  - It takes a lot of effort for all parties (experiments and developers) to prepare and attend it !
  - There are other venues (e.g. Geant4 Technical Forum, a few times per year) where issues from the experiments can be reported and discussed with the developers
- Also the duration of the Workshop, 2 full days, seems a reasonable compromise
  - To have enough time to cover different aspects and to discuss
  - To ease the participation of people, in particular experimentalists

Back-up

Pileup & Overlay

ATLAS Pile-up & Overlay



INTRODUCTION

In addition to hard interaction:

- pile-up from other collisions in current and surrounding bunch crossings,
- cosmics, beam-gas, beam-halo, cavern background, detector noise, ...

Option 1:
“Pile-up MC” (current default)

Simulate both hard and soft interactions in MC and mix together in proper ratios.

Option 2:
“Overlay”

Simulate only hard interaction in MC and overlay a random background event.

“Data Overlay”

Overlay a “random” data event.

“MC+MC Overlay”

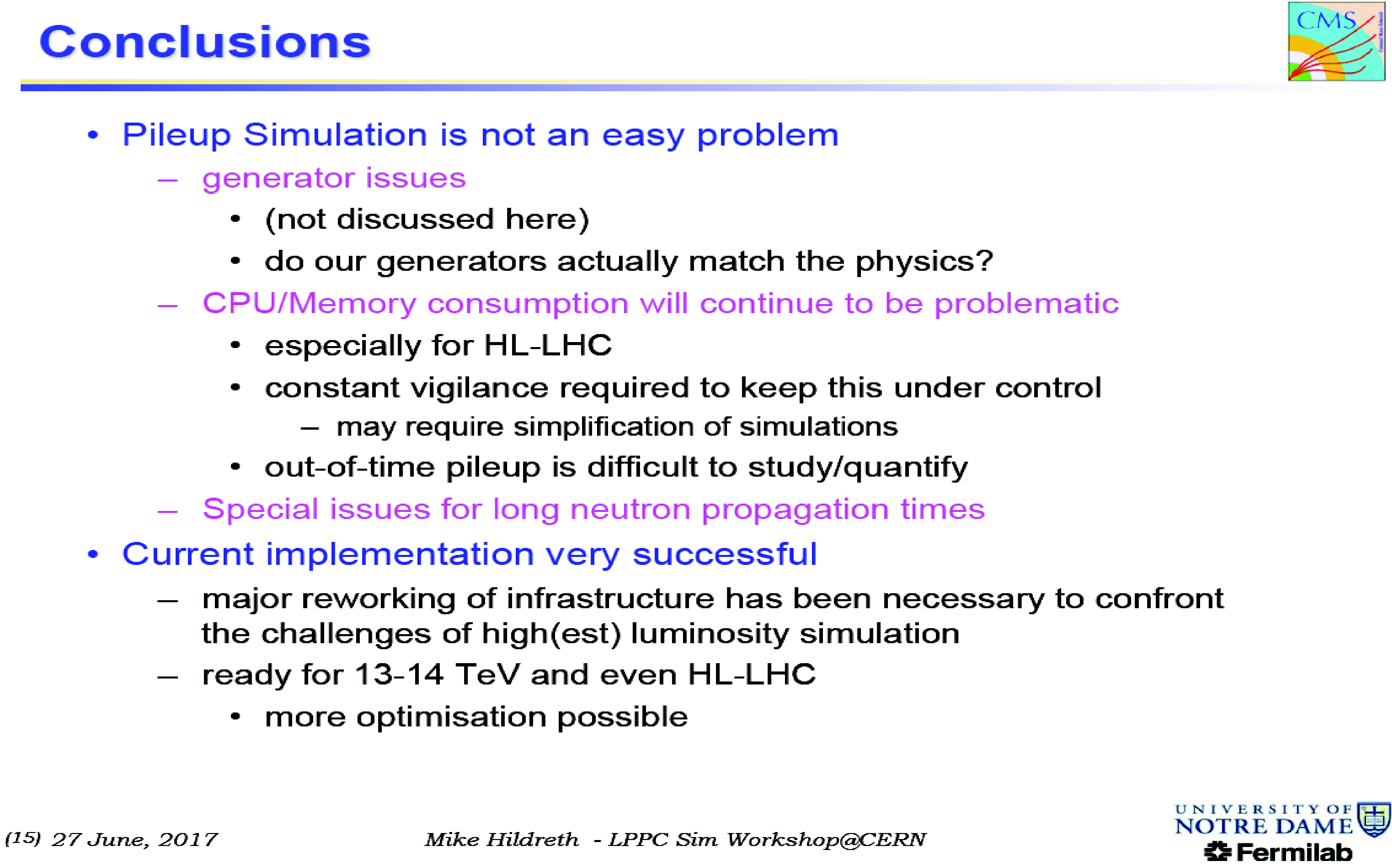
Overlay a premixed zero-hard-scatter MC event.

Pile-up MC method mostly used so far at ATLAS

- Data Overlay being used for some studies, specialized analyses, and Heavy-Ion,
- MC+MC Overlay being researched as an alternative to Pile-up MC.

Tadej Novak “ATLAS Pile-up and Overlay simulation”

CMS Pile-up & Overlay



Conclusions

- Pileup Simulation is not an easy problem
  - generator issues
    - (not discussed here)
    - do our generators actually match the physics?
  - CPU/Memory consumption will continue to be problematic
    - especially for HL-LHC
    - constant vigilance required to keep this under control
      - may require simplification of simulations
    - out-of-time pileup is difficult to study/quantify
  - Special issues for long neutron propagation times
- Current implementation very successful
  - major reworking of infrastructure has been necessary to confront the challenges of high(est) luminosity simulation
  - ready for 13-14 TeV and even HL-LHC
    - more optimisation possible

(15) 27 June, 2017 *Mike Hildreth - LPPC Sim Workshop@CERN*

UNIVERSITY OF NOTRE DAME

Fermilab

Mike Hildreth "Strategies for Modeling Extreme Luminosities in the CMS Simulation"

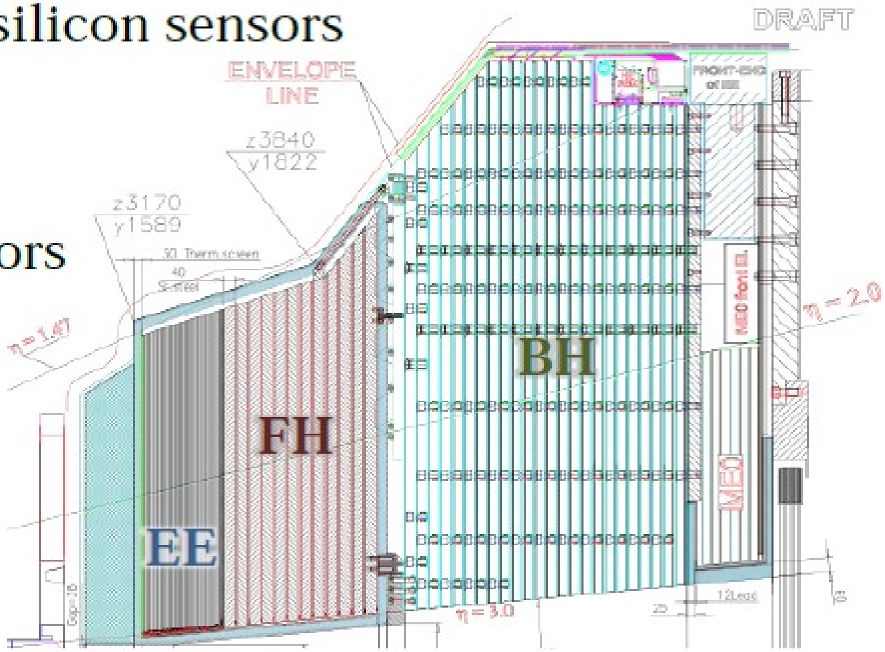
Upgrades & Test-Beams

CMS Upgrades (1/2)

Phase 2: Endcap Calorimeter

Replace entire endcap calorimeter system w/ High Granularity Calorimeter

- **EE: Endcap ECAL**
  - 28 layers of tungsten/copper absorber and silicon sensors
  - ~26 X_0 / 1.5 λ_0 thick, 4.3M channels
- **FH: Front HCAL**
  - 12 layers of steel absorber and silicon sensors
  - 3.5 λ_0 thick, 1.8M channels
- **BH: Back HCAL**
  - 12 layers of steel absorber and (radiation-hard) plastic scintillator
  - 5 λ_0 thick, ~700K channels



Aging:

- Expected to be negligible for silicon components
- For plastic component, depends on material, operating temperature, shielding, etc. (design decisions still in progress)

CMS Upgrades (2/2)

Phase 2: Computing – Considerations (1)

- Complex geometries lead to notable increases in simulation time
  - ➢ VecGeom can make a difference (~15% improvement in scalar mode)
  - ➢ GeantV will make even bigger difference (maybe factor ~5)
- Precise detectors provide opportunities
  - o e.g. improve EM and hadronic shower models based on validation w/ HGCal test beam data
  - o More precise physics lists could use more CPU (up to 25%[†])
  - ➢ Need to optimize physics list algorithms
- Fast simulation may become more important for HL-LHC
  - o Increase in particle multiplicities per event (200PU)
  - o Increase in collected data (3–4.5 ab^{-1})
  - ➢ Need to mirror full sim developments (eventually)

[†] based on standalone CMS simulation w/ CALICE physics list

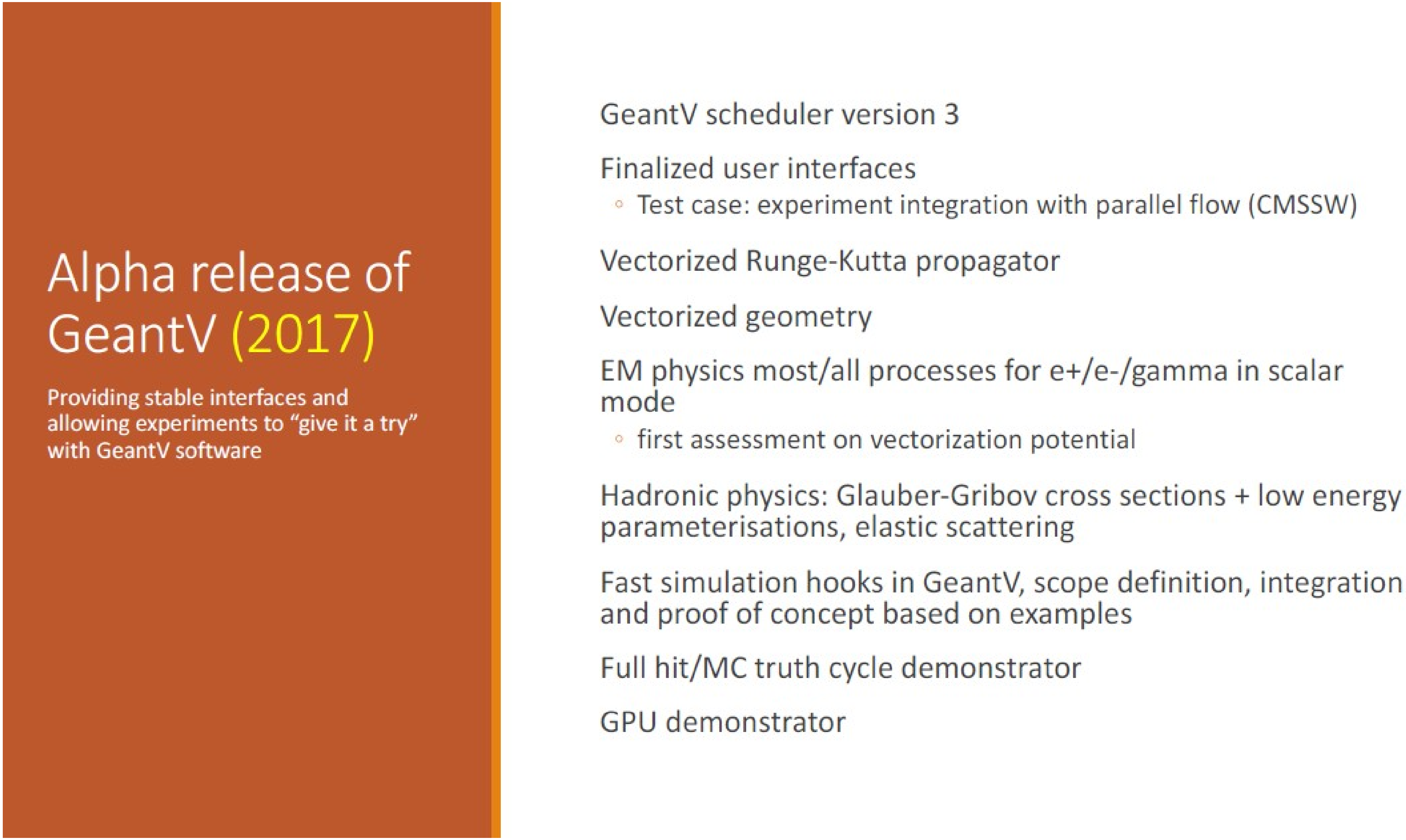
ATLAS Test-beams

TileCal test beam measurements planned or in progress

- **Signal-to-noise measurements using μ data**
  - Layer-by-layer measurements
  - dE/dx predictions from Monte Carlo simulation critical
- **Electromagnetic (EM) scale validation and comparison with Cs calibration system using μ, e data**
  - μ, e measurements critical
  - Comparison to
- **Assessments of differences in detector response to low energy $p/\pi^{\pm}/K^{\pm}$ to tune GEANT4 modeling**
  - Requires several days of data taking in 2017 run, and very delicate beam setting
  - Containment, and particle ID using Cerenkov counters are critical
- **Analysis of detector response to high energy $\pi^{\pm}$ data**
  - $\pi^{\pm}$ beam data: $E^{\pi^{\pm}}_{\text{beam}} = 300$ GeV from June 2016
  - Containment (lateral and longitudinal) measurements for testbeam simulation

GeantV

GeantV (1/2)



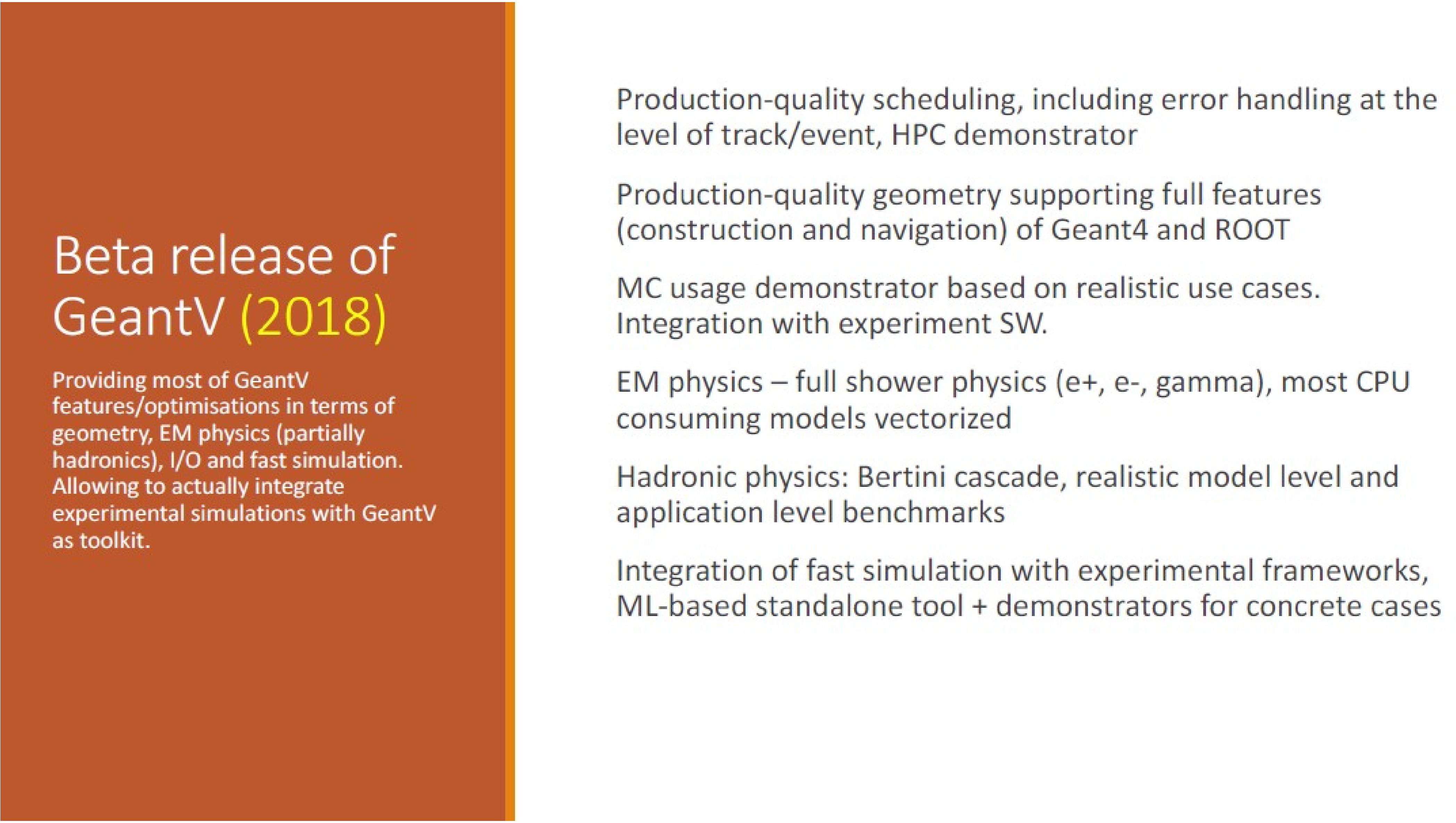
Alpha release of GeantV (2017)

Providing stable interfaces and allowing experiments to “give it a try” with GeantV software

- GeantV scheduler version 3
- Finalized user interfaces
  - Test case: experiment integration with parallel flow (CMSSW)
- Vectorized Runge-Kutta propagator
- Vectorized geometry
- EM physics most/all processes for e+/e-/gamma in scalar mode
  - first assessment on vectorization potential
- Hadronic physics: Glauber-Gribov cross sections + low energy parameterisations, elastic scattering
- Fast simulation hooks in GeantV, scope definition, integration and proof of concept based on examples
- Full hit/MC truth cycle demonstrator
- GPU demonstrator

Andrei Gheata “GeantV current status and plans”

GeantV (2/2)



Beta release of GeantV (2018)

Providing most of GeantV features/optimisations in terms of geometry, EM physics (partially hadronics), I/O and fast simulation. Allowing to actually integrate experimental simulations with GeantV as toolkit.

Production-quality scheduling, including error handling at the level of track/event, HPC demonstrator

Production-quality geometry supporting full features (construction and navigation) of Geant4 and ROOT

MC usage demonstrator based on realistic use cases. Integration with experiment SW.

EM physics – full shower physics (e+, e-, gamma), most CPU consuming models vectorized

Hadronic physics: Bertini cascade, realistic model level and application level benchmarks

Integration of fast simulation with experimental frameworks, ML-based standalone tool + demonstrators for concrete cases

Andrei Gheata “GeantV current status and plans”