JOHANNES GUTENBERG UNIVERSITÄT MAINZ

July 2021, Offshell 2021

Studies of the H->bb process at the LHeC using a full simulation

S. Behera, M. Schott, B. Brickwedde

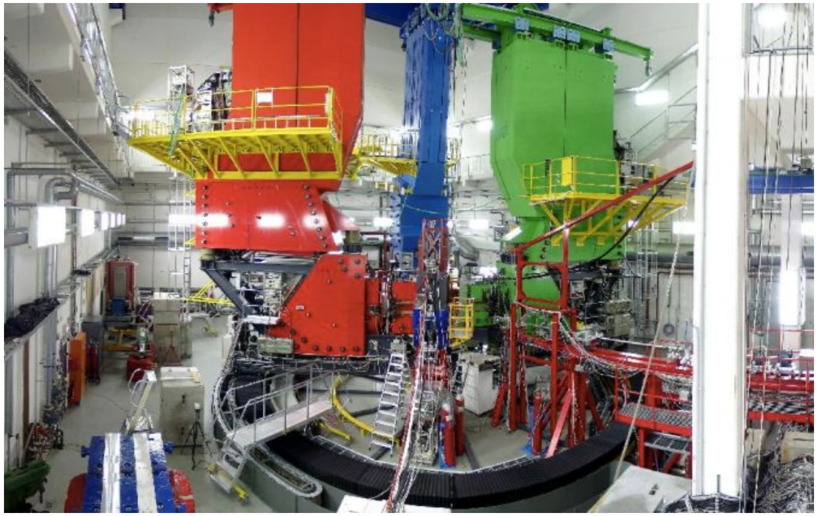
Prof. Dr. Matthias Schott,
Dr. S. Behera

Some facts about Mainz

- Mainz is small town, but capital of Rhineland-Palatinate
  - Next to the river Rhine (with some quite nice castles)
  - 20 Minutes from Frankfurt International Airport
  - Founded by romans 2K years ago
  - The cathedral is only 1000 years old (and burnt down several times)
- Time-Magazine's man of the millennium:
  - Johannes Gutenberg, who invented the printing press in Mainz

The Johannes Gutenberg University

- Founded in 1477 and reopened by the French occupation forces in 1946
  - 37.000 students for all subjects (bachelor, master, PhD) and Postdoc
  - German cluster of excellence PRISMA for fundamental physics
    - Since September 2018: PRISMA+
- Own electron accelerator MAMI and research reactor
  - 60 physics professors and research groups: LHC, IceCube, Xenon, SOX, NA62, JUNO, ALPS,

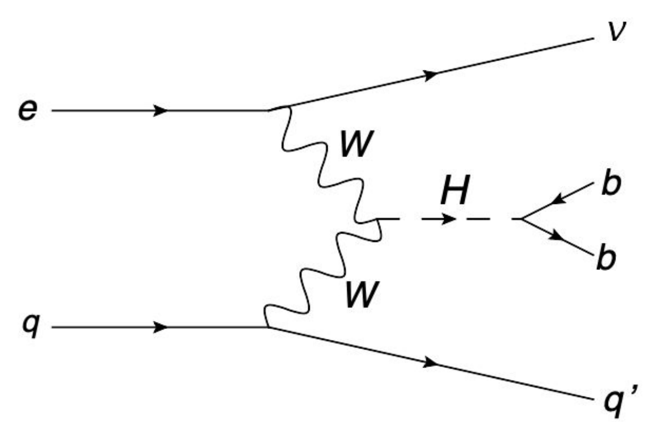


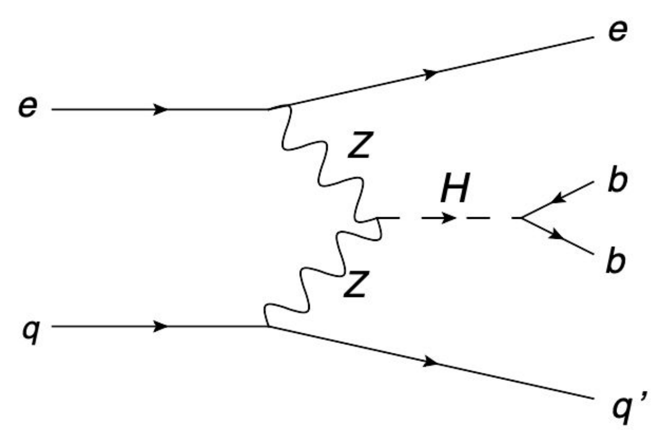
Higgs Production in proton electron collisions

Higgs Production

- Higgs production in electron proton collisions via vector boson fusion
  - Charged Current (CC): Neutrino + forward jet + Higgs-decay products
  - Neutral Current (NC): Electron + forward jet + Higgs-decay production

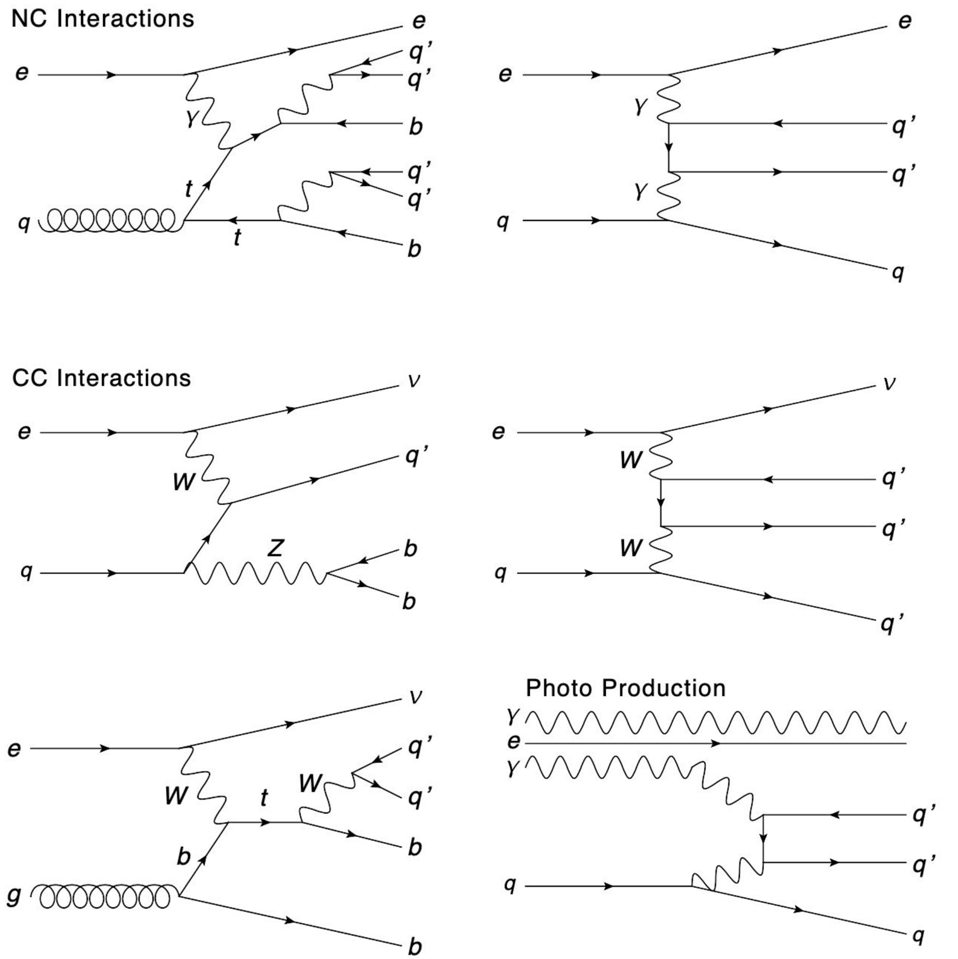
- Cross-Section Calculation has been in several previous studies
  - Leading order: Scale-dependencies in the order of 5-10%
  - Inclusive NLO QCD corrections can induce shape dependencies
    - [J. Blumlein, G.J. van Oldenborgh , R. Ruckl, Nucl.Phys.B395:35- 59,1993]
    - [B.Jager, arXiv:1001.3789]
    - [LHeC CDR: arXiv: 2007.14491]





Background processes for H->bb

- Neutral Current Interactions
  - Dijet (2 b) production
  - Top-quark pair production
  - Z-Boson Production
- Charged Current Interactions
  - Three parton final states
  - Z boson production
  - Single Top
- Photon-Production
  - Pair production of two quarks, induced by virtual photon from electron
  - Theoretical uncertainty of this process is rather high

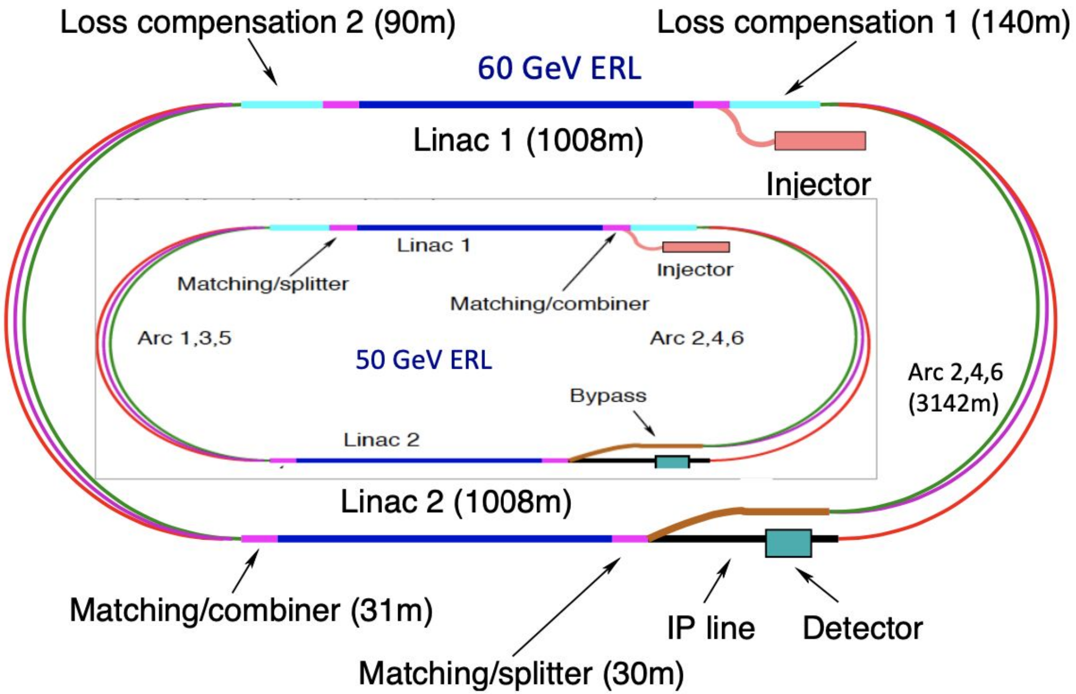


JOHANNES GUTENBERG UNIVERSITÄT MAINZ

The LHeC Collider and Previous Studies

The LHeC Project

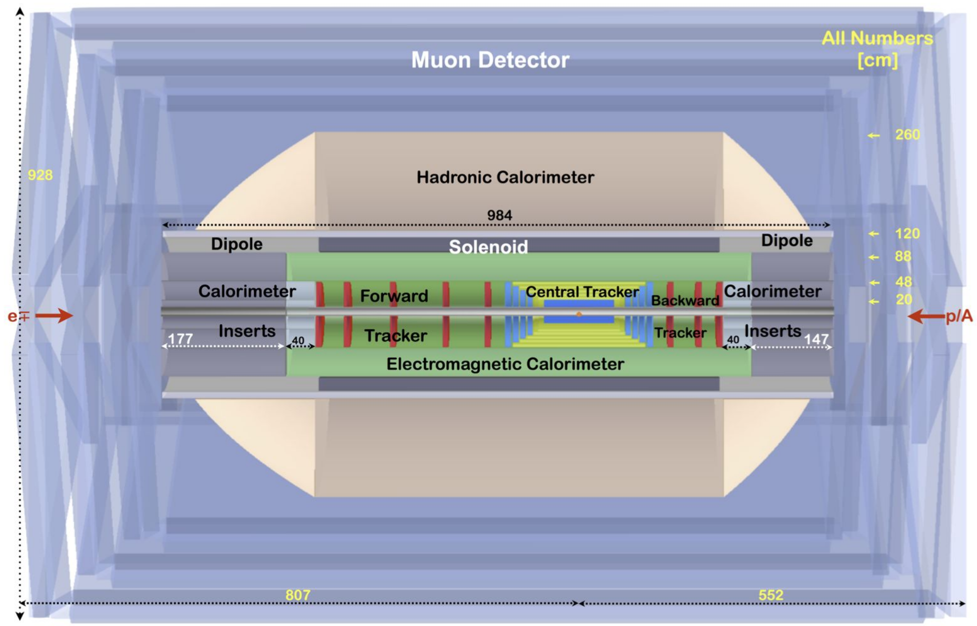
- Idea: Collide Protons from the LHC with electrons from an ERC electron accelerator
  - Two 802 MHz Electron LINACs + return arcs: using energy recovery technology
  - E_e = 60 GeV, E_p = 7 TeV,
  - CME: ~ 1.3 TeV
  - ep peak lumi 10^{34} cms^{-2} s^{-1}
  - Possibility to polarize electrons P=-80%
  - Expected integrated luminosity: $1ab^{-1}$
- Expected Higgs Production Cross Sections
  - σ_{CC} =196 fb (109 fb w.o. pol.)
  - σ_{NC} =25 fb (21 fb w.o. pol.)



[LHeC CDR: arXiv: 2007.14491]

The LHeC Detector

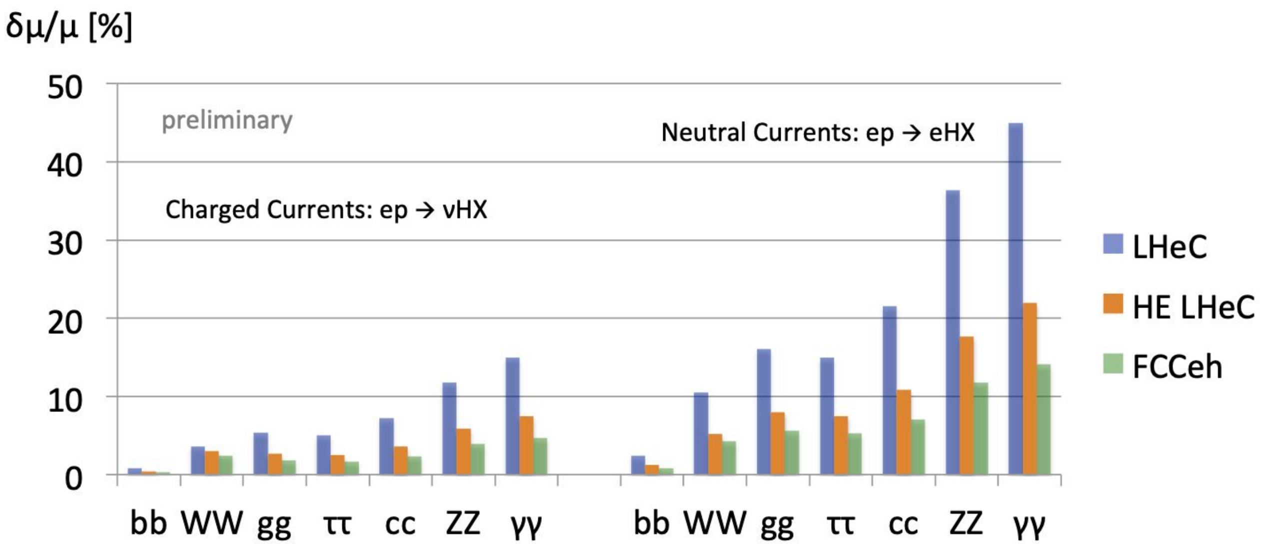
- Tracking Detector
  - Resolution 10^{-3} GeV^{-1}
  - Coverage $|\eta| < 3.3$
- Electromagnetic and Hadronic Calorimeter
  - Coverage $-4.3 < \eta < 4.9$
  - Dedicated Calorimeters at the end cap for the measurements of scattered electrons (not simulated)
- Magnet System: 3.5 T
- Muon Tracking System
  - Coverage $|\eta| < 4.7$



[LHeC CDR: arXiv: 1211.4831]

Prospects of the Higgs at the LHeC

- Lots of work already done in the past years: Full performance can be only reached with MVA based analyses
  - [LHeC CDR: arXiv: 2007.14491], [U. Klein: ICHEP 2020, July 29th, 2020], …
  - A precision on H->bb below 1% reachable at the LHeC, but very difficult channels at the LHC (Similar for charm-quark decay mode)



- Caveat: All these studies are based on Fast Simulations
  - Our goal: Confirm these prospects with a full detector simulation and state-of-the-art reconstruction algorithms

The Cut-Based Analysis

- While the full sensitivity of the LHeC requires a MVA type analysis, a simple cut-based analysis was also published within arXiv: 2007.14491
  - The cut-based approach will be used for our validation of the fast simulation as it allows to test directly the impact of different kinematic observables
- We assume that all NC background processes will be effectively reduced by 90% using a forward electron tagger and hence become negligible.
  - The case when no electron tagger is present will be also discussed

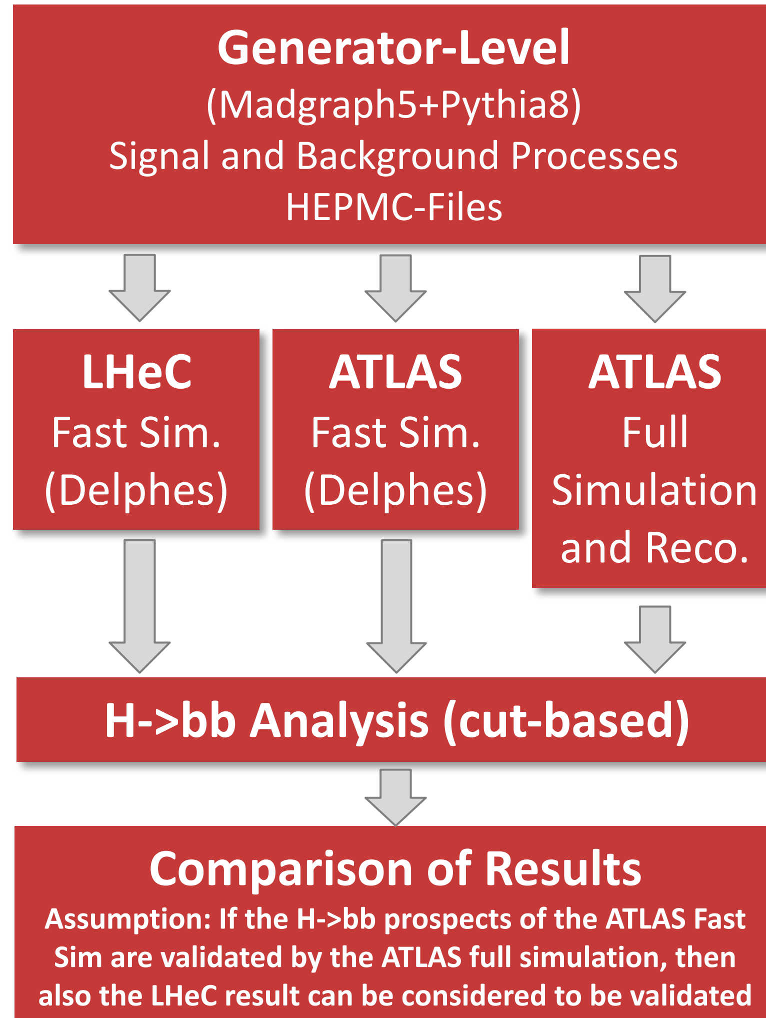
- Observables
  - Jets, reconstructed with anti-kT algorithm with cone size of R = 0.4
- Selection Cuts
  - Kinematic Requirements (DIS)
    - E_T^{Miss}>30 GeV, y_h<0.9, Q_h^2>500 GeV
  - Jet Requirements
    - p_T>25 GeV
    - 2 b-tagged jets
  - Top-background veto
    - m(bbj)>250 GeV, m(bj)>130 GeV
  - At least one forward jet (η>2.0)
  - $\Delta\varphi(E_T^{Miss}$, b-jet)>0.2

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

How to validate these results using a full simulation?

Basic Idea

- The development of a full detector simulation is a huge task which takes several years
  - Not even speaking of the development of state-of-the-art reconstruction algorithms
- Observation: The kinematic constraints of the future LHeC detector and the current ATLAS at the LHC are not so different
  - The minimal p_T requirement of the b-quarks forces them in the more central rapidity region
- Idea: Compare a simple H->bb study based on a fast ATLAS simulation using Delphes with the full ATLAS Simulation



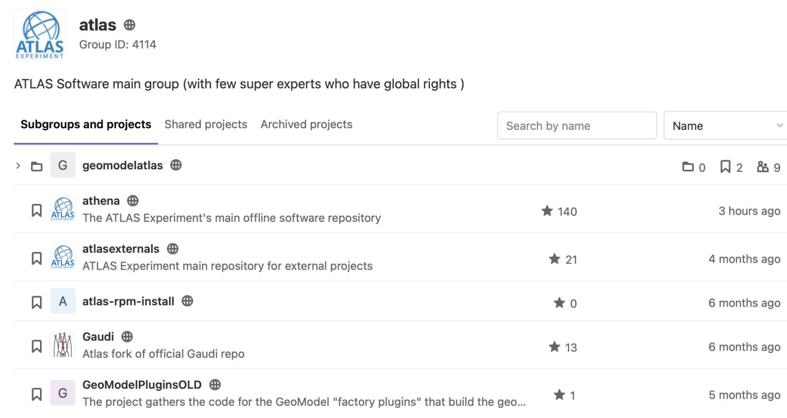
MC Samples and Fast Simulations

- All samples have been generated with MadGraph5
- Showering and Hadronization with Pythia8
  - Cross-Checks with Pythia6
- All cross sections which have been previously reported could be validated
  - reduce in photo-production CS is possible due to the fact that we demand 2b -jets in the final state.
- Fast Detector Simulation with the DELPHES Framework
  - Standard LHeC card available from LHeC study Group.

Short Description	Process	Generator-Level Cuts	Generator	Cross Section [pb]	Number Of Events $\times 10^3$
Signal	$pe^- \to \nu_e hq \to \nu_e b\bar{b}q$	$\|\eta\| < 10,\ m_{jj} > 60$	MadGraph5+ Pythia8	0.09997^3	150
CC-qqq	$pe^- \to \nu_\ell qqq$ (w.o. Z, H, t, $\bar{t}$)	$p_T^q > 10\ GeV, p_T^b > 10\ GeV,$ $\|\eta\| < 10,\ m_{jj} > 70,\ m_{bb} > 70$	MadGraph5+ Pythia8	5.49074	214
CC-top	$pe^- \to \nu_\ell \bar{t}q$	$p_T^q > 10\ GeV, p_T^b > 10\ GeV,$ $\|\eta\| < 10,\ m_{jj} > 70,\ m_{bb} > 70$	MadGraph5+ Pythia8	0.36820	137
CC-Z	$pe^- \to zq\nu_\ell, \to b\bar{b}q\nu_\ell$	$p_T^q > 10\ GeV, p_T^b > 12\ GeV,$ $\|\eta^q\| < 5.5, \|\eta^b\| < 5.5, \|\eta^\gamma\| < 5,$ $\|\eta^\ell\| < 5,\ m_{jj} > 60,\ m_{bb} > 60$	MadGraph5+ Pythia8	0.13631	107
Photo-qqq	$p\gamma \to b\bar{b}q$	$p_T^q > 12\ GeV, p_T^b > 19\ GeV,$ $\|\eta^q\| < 5.2, \|\eta^b\| < 3.5, \|\eta^\gamma\| < 10,$ $\|\eta^\ell\| < 10,\ m_{jj} > 115,\ m_{bb} > 115$	MadGraph5+ Pythia8	0.90876	100
PA-tt	$p\gamma \to t\bar{t} \to bqq\bar{b}qq$	$p_T^q > 10\ GeV, p_T^b > 12\ GeV,$ $\|\eta^q\| < 5.5, \|\eta^b\| < 4, \|\eta^\gamma\| < 10$ $\|\eta^\ell\| < 10,\ m_{jj} > 80,\ m_{bb} > 80$	MadGraph5+ Pythia8	0.00876	100
NC-Z	$pe^- \to Zqe^- \to b\bar{b}qe^-$	$p_T^q > 10\ GeV, p_T^b > 12\ GeV$ $, p_T(\ell) > 0.01\ GeV, \|\eta^q\| < 5.5, \|\eta^b\| < 5.5,$ $\|\eta^\gamma\| < 10, 4 < \|\eta^\ell\| < 10,\ m_{jj} > 60,\ m_{bb} > 60$	MadGraph5+ Pythia8	0.02246	100
NC-bbq	$pe^- \to e^- b\bar{b}q$	$p_T^q > 10\ GeV, p_T^b > 12\ GeV,$ $p_T(\ell) > 0.01\ GeV, \|\eta^q\| < 5.5, \|\eta^b\| < 4$ $, \|\eta^\gamma\| < 10, 4 < \|\eta^\ell\| < 10,\ m_{jj} > 80,\ m_{bb} > 80$	MadGraph5+ Pythia8	2.37302	100
NC-tt	$pe^- \to e^- t\bar{t} \to e^- bqq\bar{b}qq$	$p_T^q > 10\ GeV, p_T^b > 12\ GeV,$ $p_T(\ell) > 0.01\ GeV, \|\eta^q\| < 5.5, \|\eta^b\| < 4,$ $\|\eta^\gamma\| < 10, 4 < \|\eta^\ell\| < 10,\ m_{jj} > 80,\ m_{bb} > 80$	MadGraph5+ Pythia8	0.81091	100

Full Simulation: Using the ATLAS Softwareframework Athena

- The major task of this study was to operate the ATLAS Software framework ATHENA without any resources from the ATLAS Collaboration
  - Software is open source and available under
    - https:// gitlab.cern.ch/atlas
  - Simulation is based on Geant4
  - Digitalization and state of the art reconstruction algorithms are developed by the ATLAS Collaboration
- Several modifications had to be performed in order to run ATHENA on our University Cluster Mogon
  - e.g. some steps required a connection to databases at CERN, which had to be emulated (We cannot connect from Mogon to the outside world)

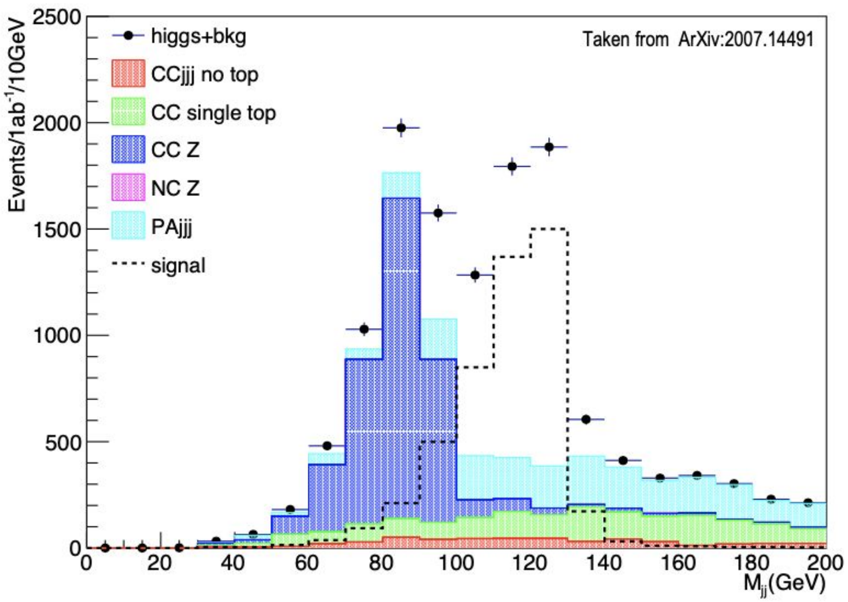


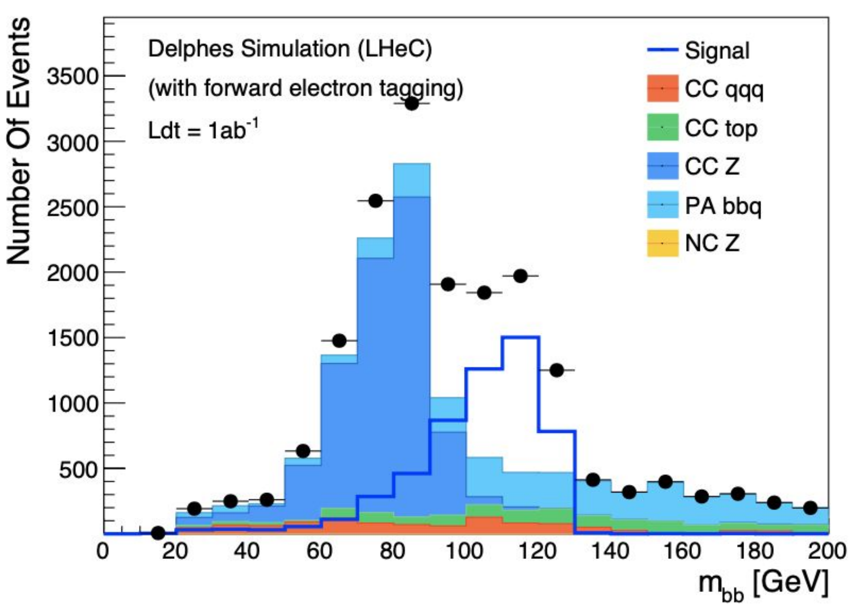
Results

Validation of the official LHeC cut-based analysis

- CDR selection cuts have been applied
  - Good agreement in expected signal and background yields within the stat. uncertainties
  - Shape Comparison also reasonable
  - Exception: CC-Z, where we see 1.8 more events in total and a shape difference

Process	LHeC CDR (Delphes)	This Study (Delphes)
Signal	3720	3540±50
CC-qqq	157	200±70
CC-top	329	310±30
CC-Z	173	90±10
Photo-qqq	606	840±90

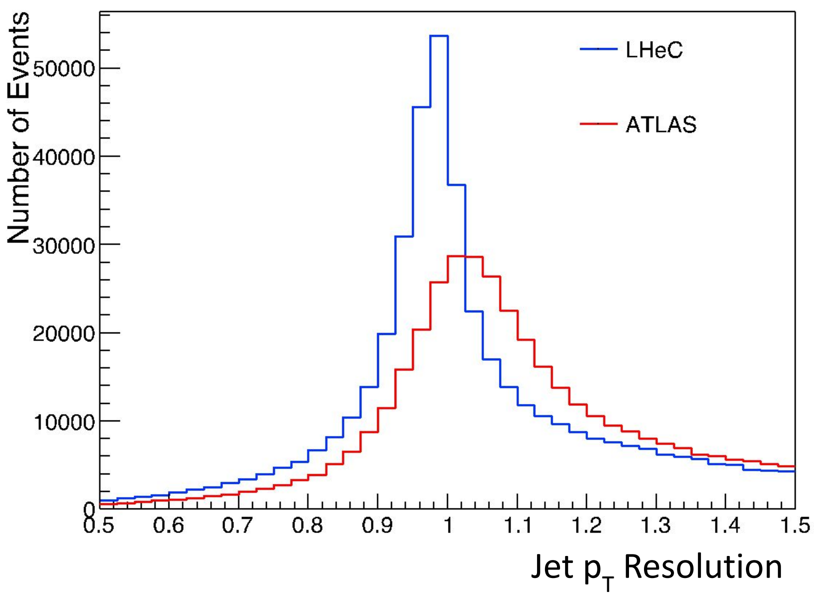




Comparison between ATLAS and LHeC using a fast simulation

- Moving one step further and compare the H->bb expectations of the simple cut-based analysis if we would perform it at the current ATLAS detector (still fast simulation)
  - We observe a rather consistent cut-flow between ATLAS and LHeC, however differences in the order of 20% are introduced by the jet-requirements
    - Differences in the b-tagging efficiency (i.e. coverage)
    - Differences in the jet-energy resolution

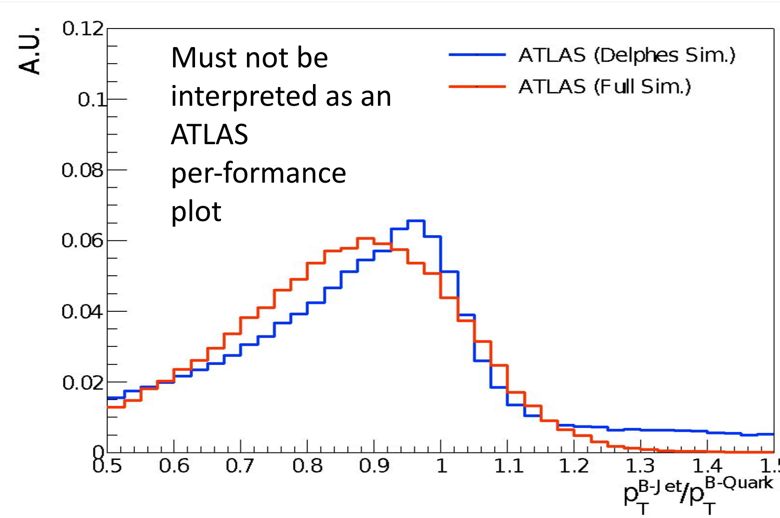
Signal Selection	LHeC (Delphes)	ATLAS (Delphes)
All Events	99970	99970
No Electron	97100	99941
Kinematic Cuts $(\not{E}_T, Y, Q^2)$	70356	70971
Jet Requirements	18325	13373
Top Rejection	6809	6003
Forward Jet	6745	5878
$\Delta\Phi(B_{1/2}, \not{E}_T) > 0.2$	5438	4662
$100 < m_{bb} < 130$	3544	3160



Comparison between fast and full simulations (1/4)

- 25% less signal events pass in the full detector simulation compared to the fast detector simulastion
  - The MET resolution is significantly worse in the full simulation compared to Delphes (significantly less events pass the MET cut)
  - Jet-energy resolution is worse, yielding a broader m_{bb} spectrum
    - less events pass the m_{bb} requirement

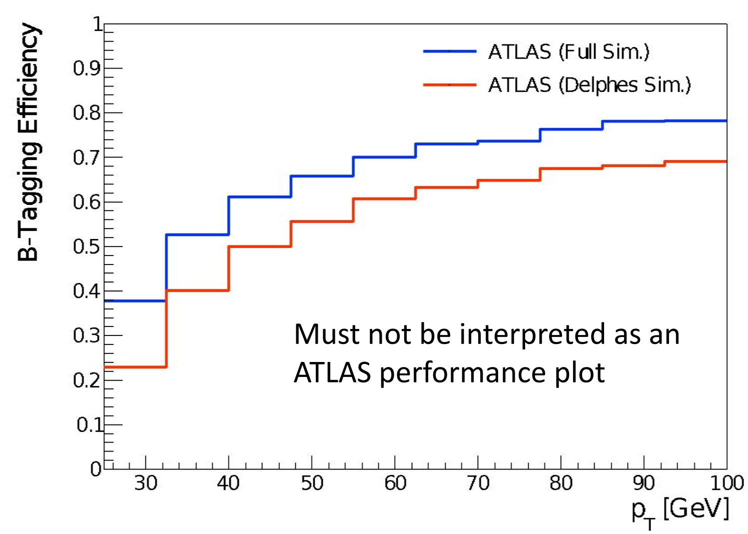
Signal Selection	ATLAS (Delphes)	ATLAS (Full Sim.)
All Events	99970	99970
No Electron	99941	98311
Kinematic Cuts ($\not{E}_T, Y, Q^2$)	70971	65236
Jet Requirements	13373	12738
Top Rejection	6003	5753
Forward Jet	5878	5643
$\Delta\Phi(B_{1/2}, \not{E}_T) > 0.2$	4662	4267
$100 < m_{bb} < 130$	3160	2480



Comparison between fast and full simulations (2/4)

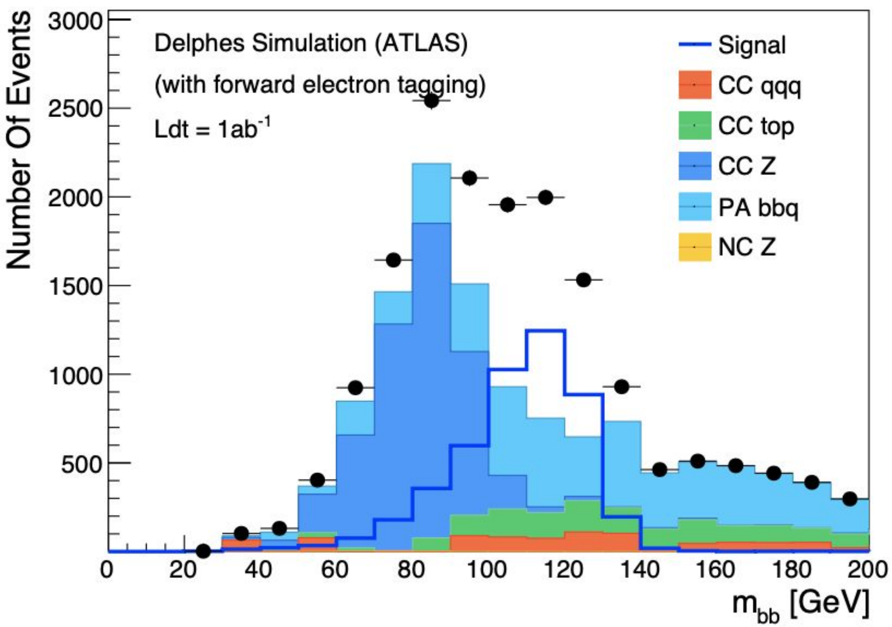
- Background is largely consistent between full and fast simulation
- Exception: CC-qqq due to higher fake-rates in the full simulation
  - a b-tagging efficiency different in DELPHES and full simulation
    - B-tagging algorithms are not optimized for e-p running in ATLAS
    - Photo-production is not affected too much since we required already 2 b-quarks on MC generator level
- Given that no optimization has been performed, we conclude that an agreement can be reached for the backgrounds

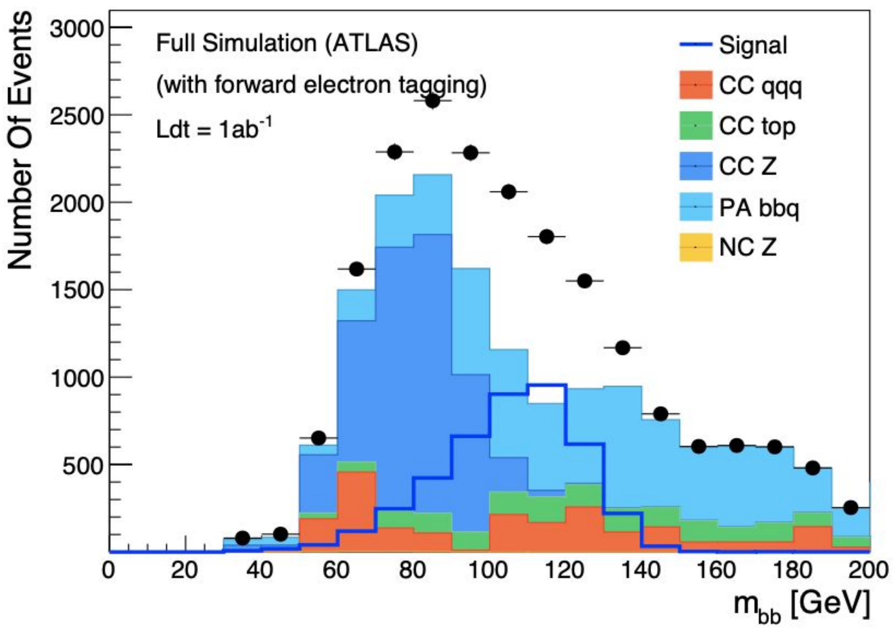
Process	ATLAS (Delphes)	ATLAS (Full Simulation)
Signal	3160±50	2480±50
CC-qqq	430±110	880±160
CC-top	480±40	410±35
CC-Z	240±20	240±20
Photo-qqq	1340±110	1660±130
NC-Z	1± 1	0 ±0
NC-bbq	700 ±130	950 ±160
NC-tt	1200±100	900±90



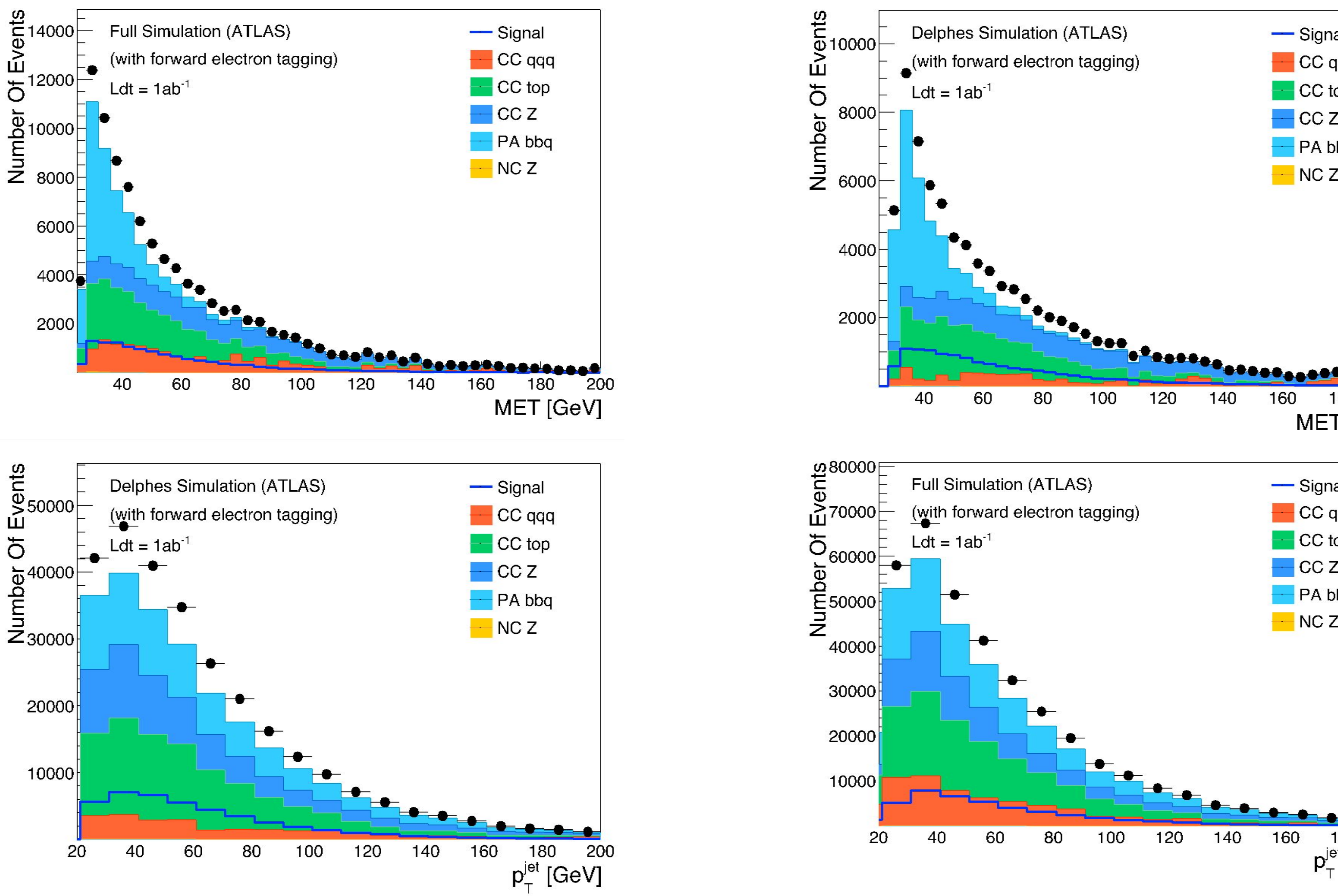
Comparison between fast and full simulations (3/4)

- Shapes between full and fast simulation are comparable, however, a broader m_{bb} spectrum is observed





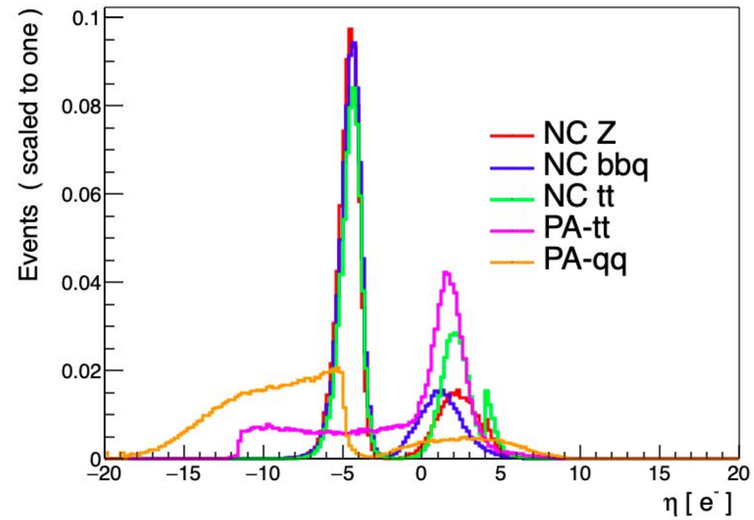
Comparison between fast and full simulations (4/4)

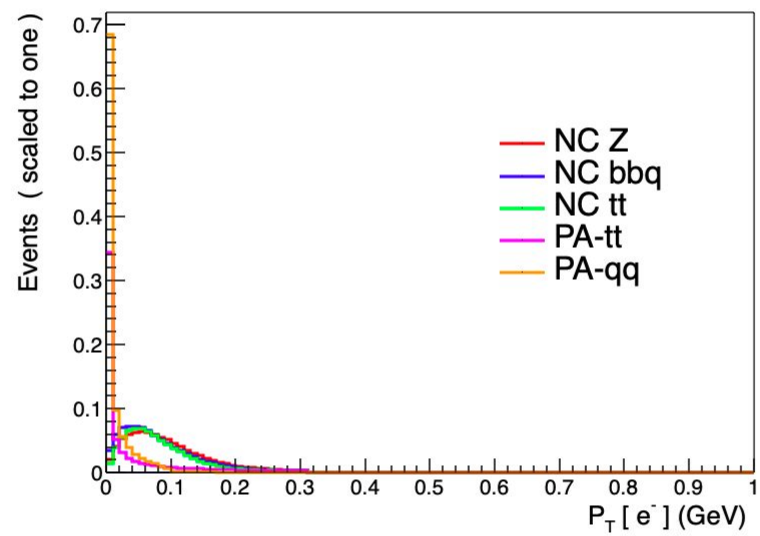


Forward Electron Tagging

Why is forward Electron Tagging Important

- No simulation for a forward electron tagger is available in Delphes
- Forward electron tagging would allow to reduce NC backgrounds
  - Eta-distribution of electrons for NC-Z, NC-bbq and NC-tt processes are reaching up to |η|<10
  - How efficient can be forward electron tagger?
    - In any case: The normalization of those processes could be studied
- The Photo-Production process peaks even more forward
  - Tagging of the full spectrum would be impossible

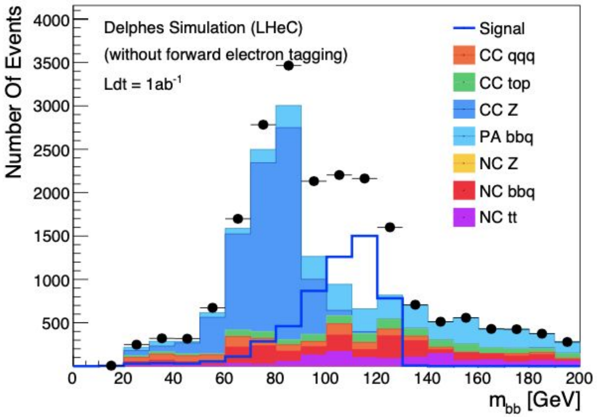


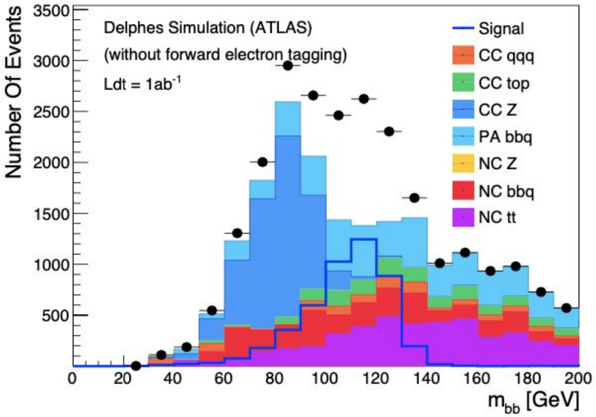


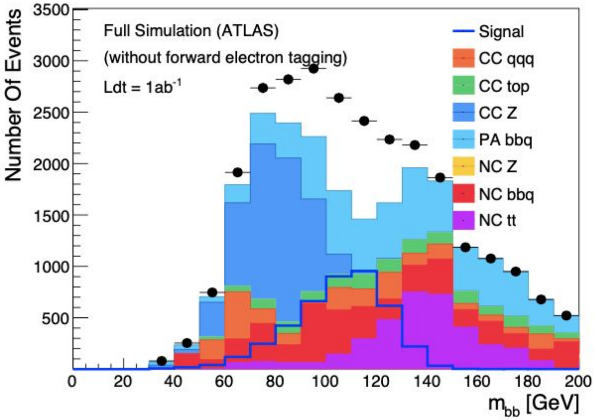
Results without Forward Electron Tagging

- The expected background contribution at the LHeC from NC processes increases by less than 50%, while the background yields nearly double for the fast and full simulation of ATLAS
  - Main reason is that we do not used forward electrons in our electron-veto selection. Hence the situation is expected to be more similar to the case of the LHeC

- Preliminary conclusion
  - Even without any forward tagging, the Higgs signal stays visible and dedicated selections cuts will enable a significant reduction





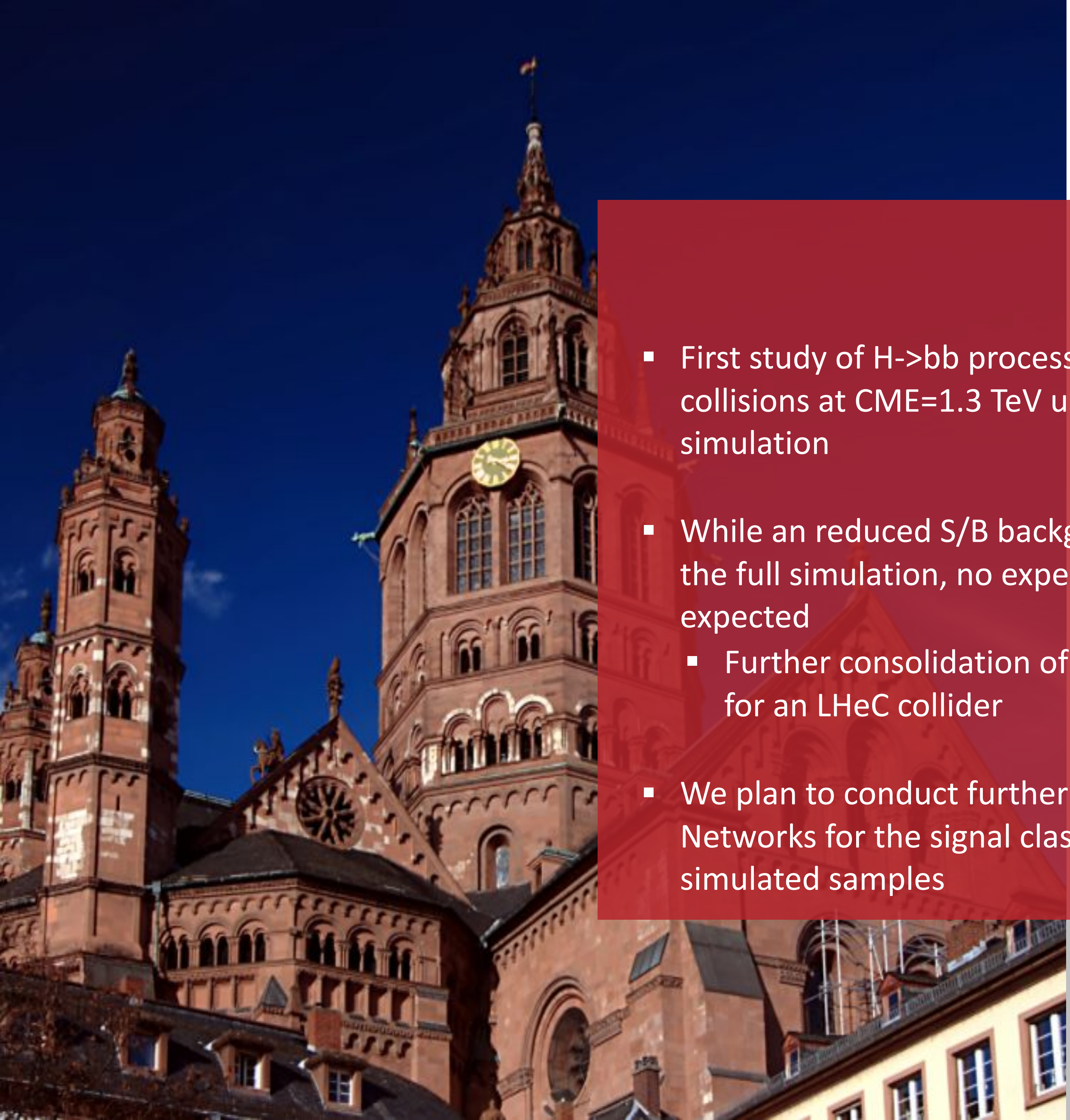


JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Conclusions

What have we learnt?

- How to run the ATHENA-framework without ATLAS resources
- We could (largely) reproduce the cut-based H->bb results of the LHeC CDR
  - Gives us confidence that we don't mess up any important aspect of the analysis
    - Differences in the predicted cross-sections for photon-induced processes and CC-Z production
- When comparing the ATLAS full (Athena) and the ATLAS fast simulation (DELPHES) we see
  - a reduction of signal efficiency by 25%
  - An increase of background yields by 20%
  - A decrease of the signal/background ratio from 1.3 to 0.8
- Assuming reasonable values for systematic uncertainties, we can expect a cross-section uncertainty for H->bb of <2%
  - With more advanced classifiers the signal selection efficiency can be increased by factors, while reducing the background yields
- Caveats
  - Did we forget a background process?
    - In particular photon-production
  - How reliable is the forward electron tagging?
- Is it really impossible to built an LHeC interaction point at ATLAS?
  - … it seems quite well suited (and will be even better after the upgrade)



JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Summary

- First study of H->bb process in electron-proton collisions at CME=1.3 TeV using a full detector simulation
- While an reduced S/B background ratio is observed in the full simulation, no experimental show-stoppers are expected
  - Further consolidation of the excellent physics case for an LHeC collider
- We plan to conduct further studies using Deep Neural Networks for the signal classification based on the fully simulated samples