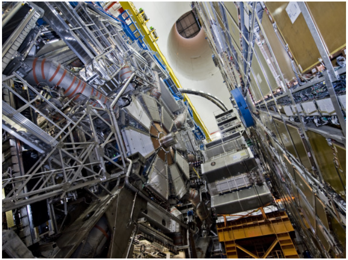
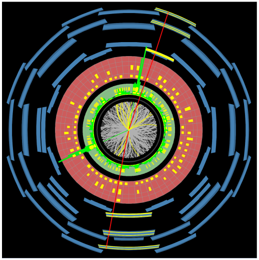
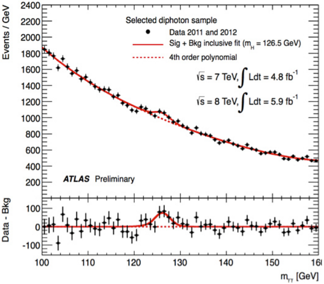
Data analysis in particle physics

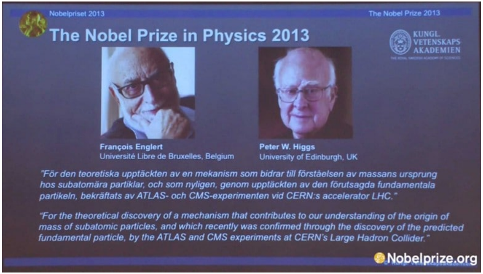
Ina Carli

Charles University, Prague / CERN
ina@cern.ch

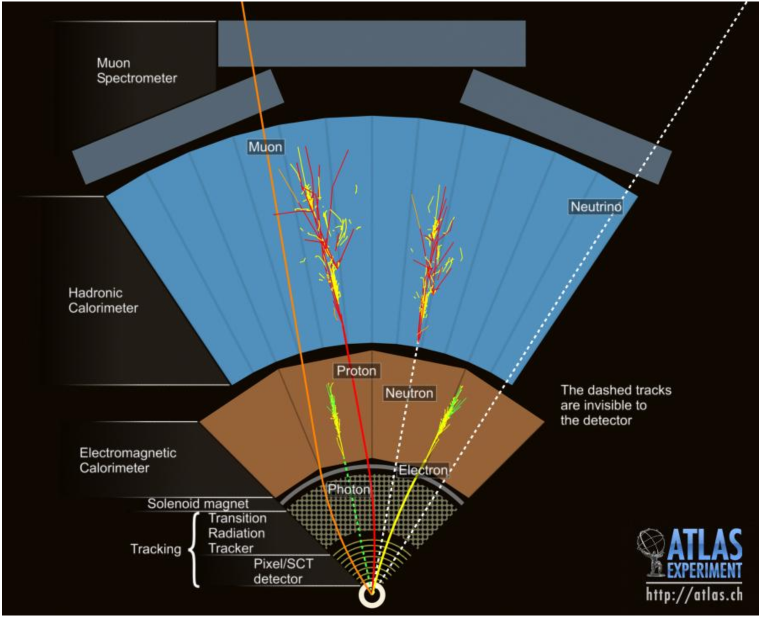
14.8.2017

... or road from  via 

to 

and 

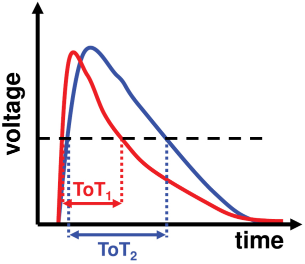
Complex detector systems

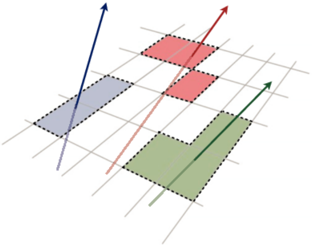


Signal processing on detector

Each detector element (pixel/strip/crystal/wire)

- measure voltage or Σ charge
- pedestal suppression
- analog, binary or Time-over-threshold



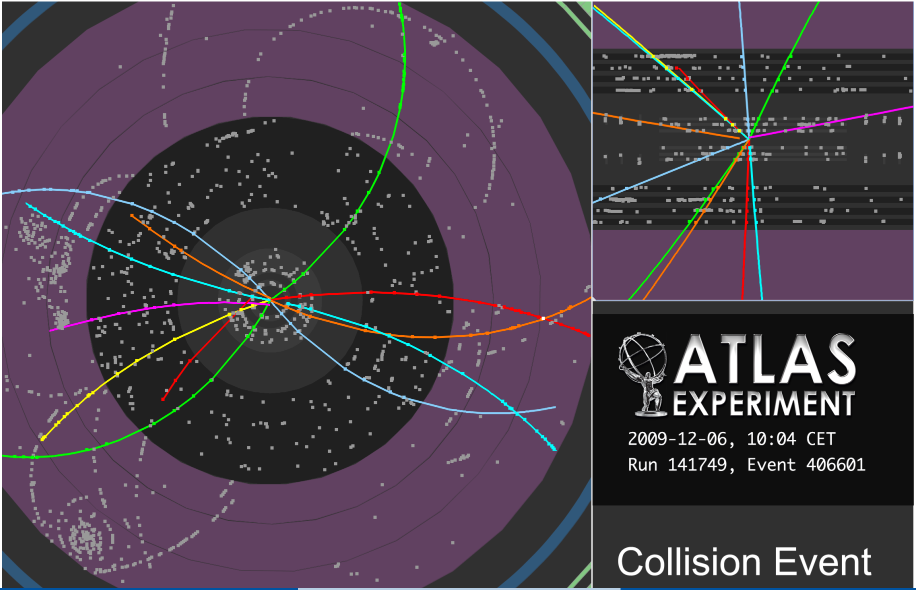


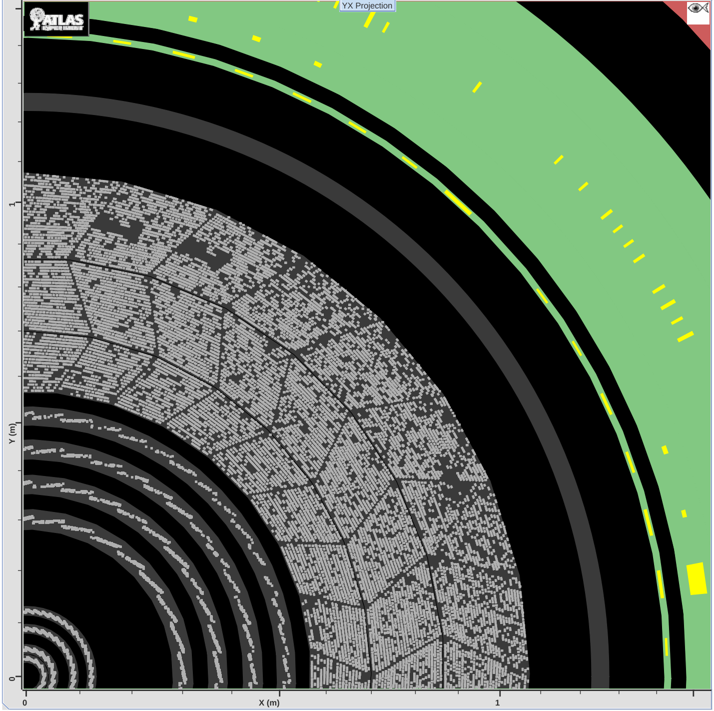
Each structure (module/chamber)

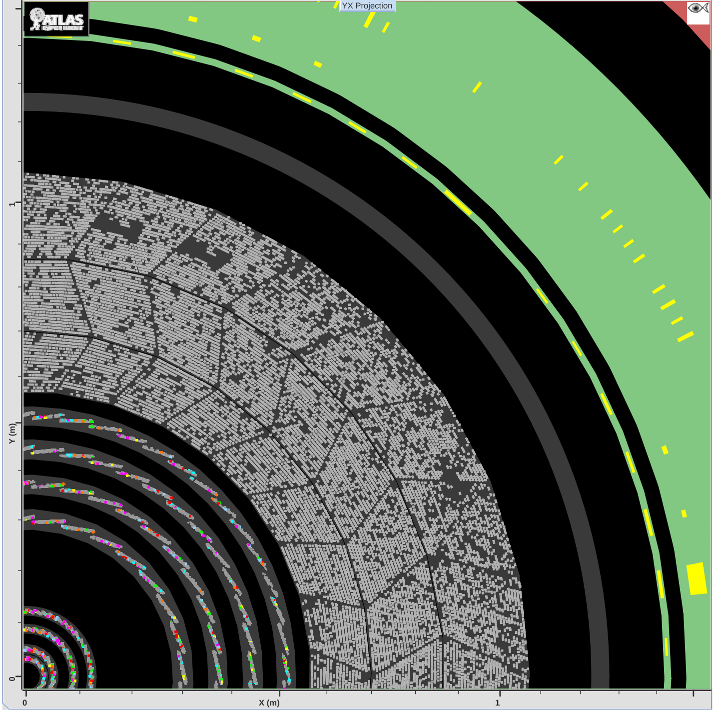
- format and send data
- compression - clustering

Tracking

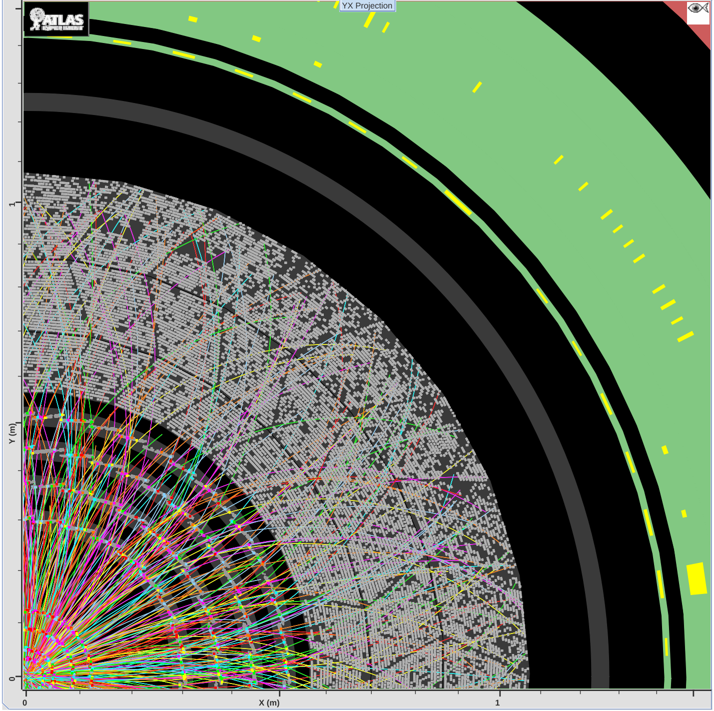
Try to reconstruct a path of charged particles in inner detector:

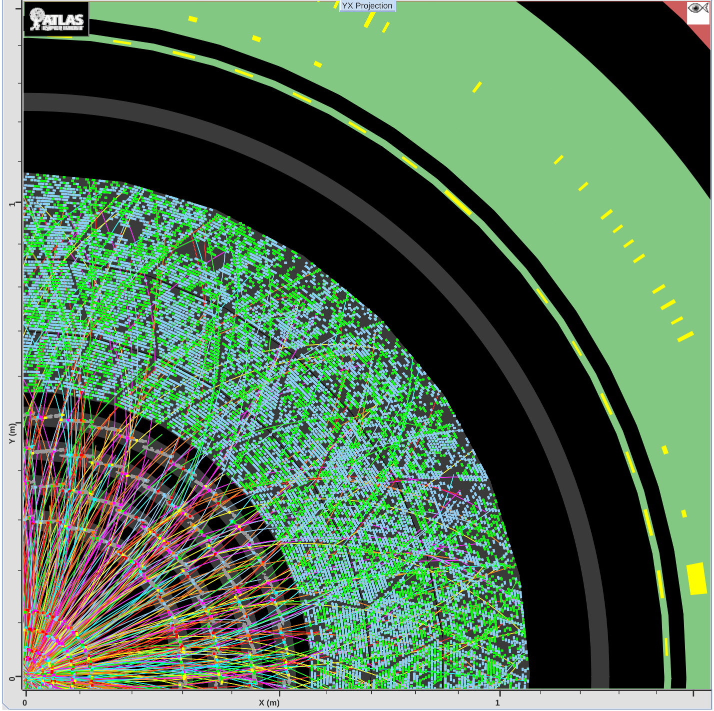


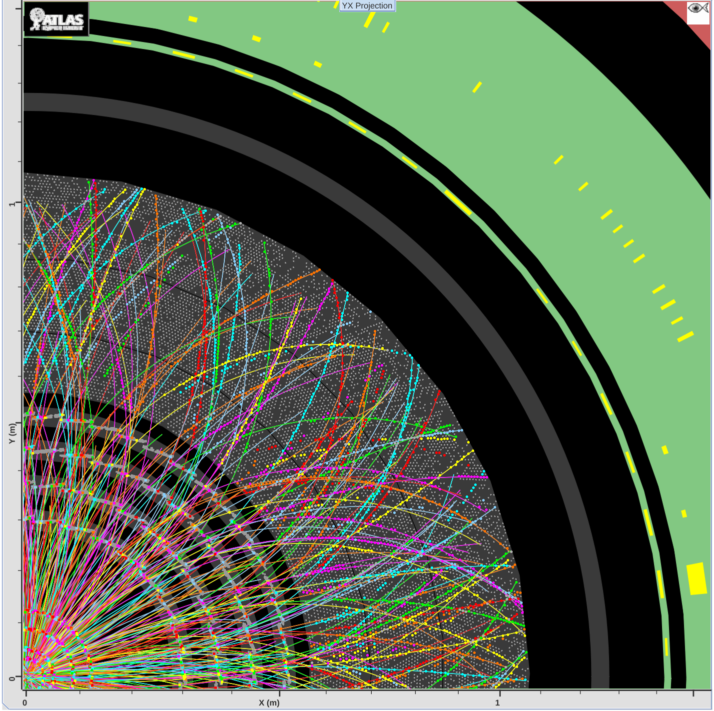

YX Projection
ATLAS
Y (m)
1
0
X (m)
0
1



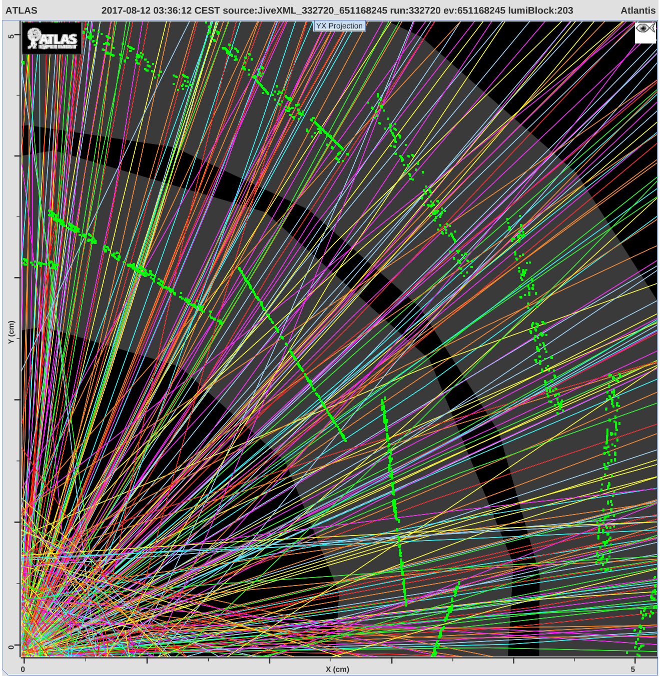
ATLAS
YX Projection
Y (m)
1
0
X (m)
0
1


YX Projection
ATLAS
Y (m)
X (m)
0
1
0
1

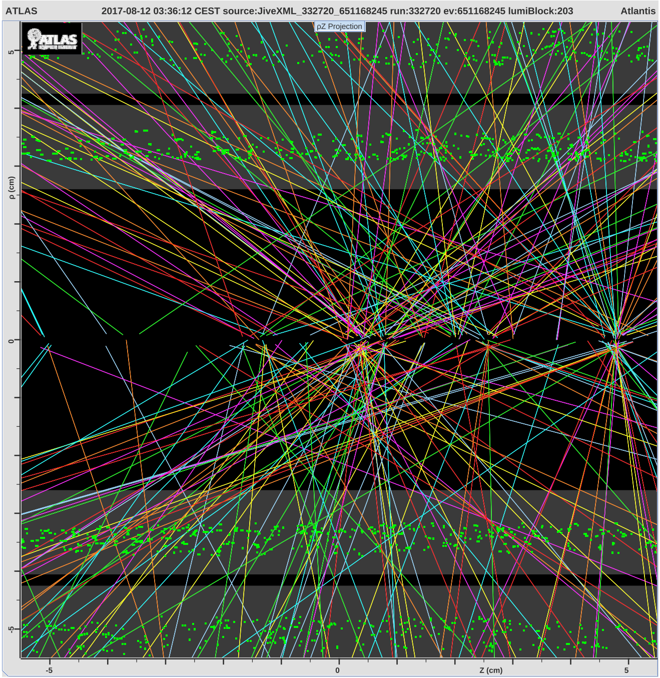

YX Projection
ATLAS
Y (m)
X (m)
0
1
0
1


ATLAS EXPERIMENT
YX Projection
Y (m)
X (m)
0
1
0
1

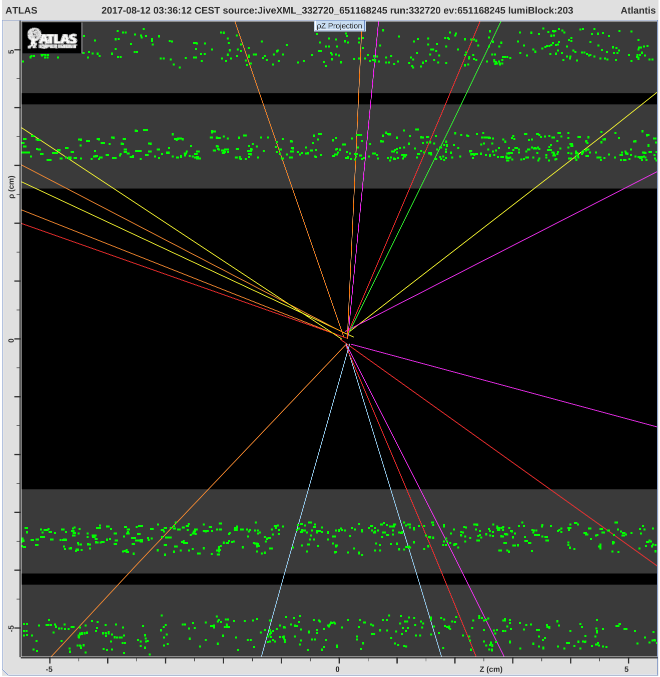
Zoom on interaction point:



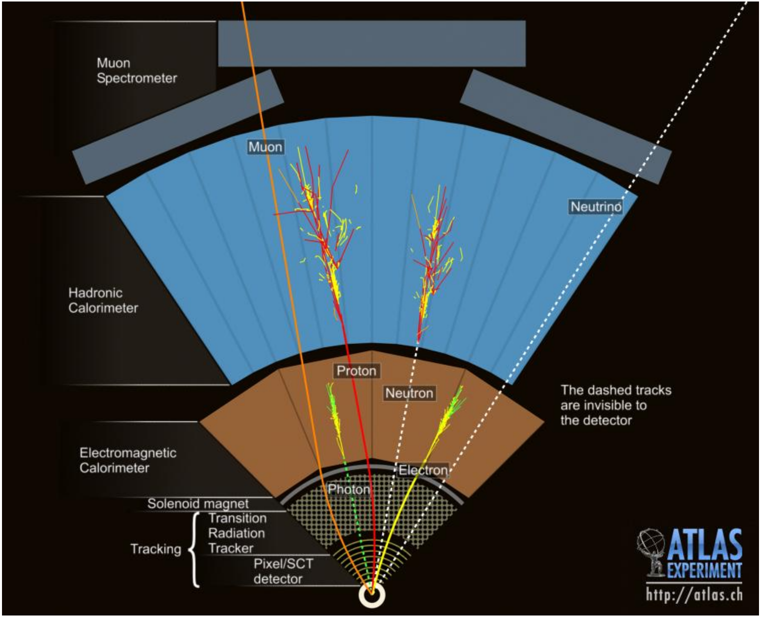
Zoom on interaction point:



Zoom on the primary vertex:



Physics objects



Physics objects

Simple objects:

- track
- calorimeter cluster
- muon segment

→ Combined objects:

- electron = calo cluster + track
- photon = ECAL cluster + no track or e^-e^+ pair - conversion
- proton = HCAL cluster + track

→ Complicated objects:

- jets = big calo cluster + tracks
- b-jets - often contain muon or electron
- τ leptons - decay to hadrons (jets), muons or electrons
- ν - computed from missing energy and momentum

Data format after reconstruction

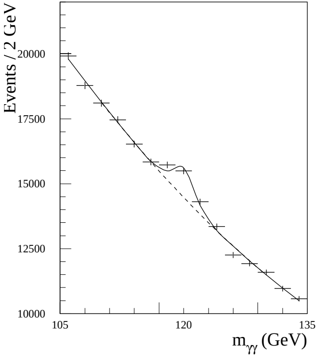
Event = all physics objects reconstructed in 1 LHC bunch crossing:

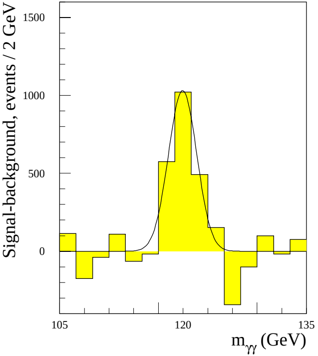
run	332720										
event	651168245										
track0	p_x	p_y	p_z	x_{PV}	y_{PV}	z_{PV}	n_{PIX}	n_{SCT}	n_{TRT}	...	
track1	p_x	p_y	p_z	x_{PV}	y_{PV}	z_{PV}	n_{PIX}	n_{SCT}	n_{TRT}		
⋮											
mu0	p_x	p_y	p_z	x_{PV}	y_{PV}	z_{PV}	n_{PIX}	n_{SCT}	n_{TRT}	MS	n_{MS}
mu1	p_x	p_y	p_z	x_{PV}	y_{PV}	z_{PV}	n_{PIX}	n_{SCT}	n_{TRT}	MS	n_{MS}
⋮											
gamma0	p_x	p_y	p_z	E							
jet0	p_x	p_y	p_z	x_{PV}	y_{PV}	z_{PV}	n_{tracks}				
jet1	p_x	p_y	p_z	x_{PV}	y_{PV}	z_{PV}	n_{tracks}				
p_{T}^{miss}	p_x	p_y	p_z								
PV0	x	y	z	n_{tracks}							
PV1	x	y	z	n_{tracks}							
⋮											

Physics analysis

Example: strategy for $H \to \gamma\gamma$

- run simulations of H signal and background from SM
- think about how to select candidates with high purity - e.g. cuts
  - 2 γ's in central region with $E_\gamma > 20$ GeV, no tracks close to them
  - from simulation: should have at least 15% of real H bosons



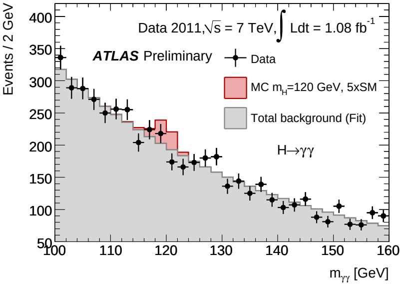


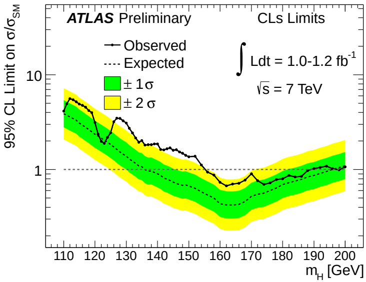
Example: strategy for $H \to \gamma\gamma$

- take data
- remove events with bad quality (subdetectors OFF, unstable conditions)
- select events containing particles in final state you want
- wait and plot what you find...

Example: strategy for $H \to \gamma\gamma$

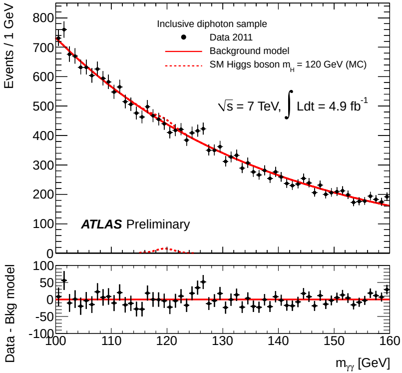
- take data
- remove events with bad quality (subdetectors OFF, unstable conditions)
- select events containing particles in final state you want
- wait and plot what you find...
- ... summer 2011

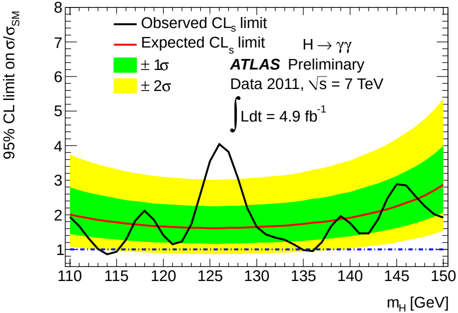




Example: strategy for $H \to \gamma\gamma$

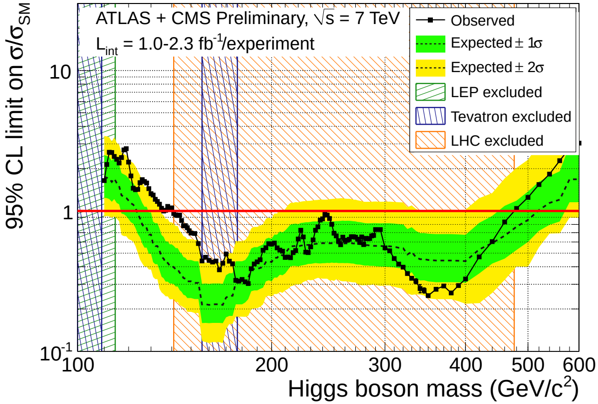
- take data
- remove events with bad quality (subdetectors OFF, unstable conditions)
- select events containing particles in final state you want
- wait and plot what you find...
- ... end of 2011 - is there an excess?





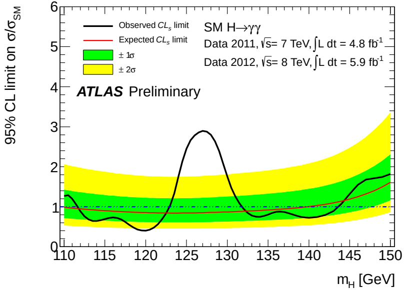
Example: strategy for $H \to \gamma\gamma$

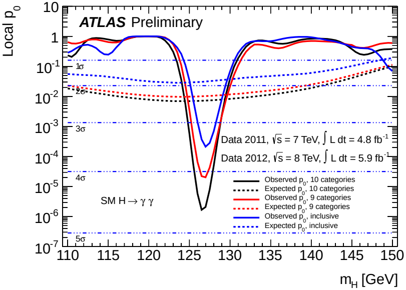
- take data
- remove events with bad quality (subdetectors OFF, unstable conditions)
- select events containing particles in final state you want
- wait and plot what you find...
- ... end of 2011 - ask CMS as well



Example: strategy for $H \to \gamma\gamma$

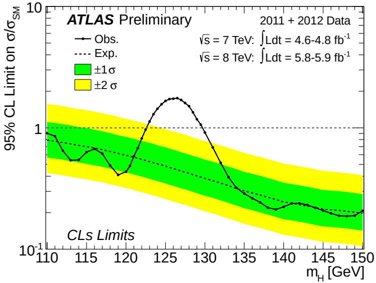
- take data
- remove events with bad quality (subdetectors OFF, unstable conditions)
- select events containing particles in final state you want
- wait and plot what you find...
- ... summer 2012 - add more data

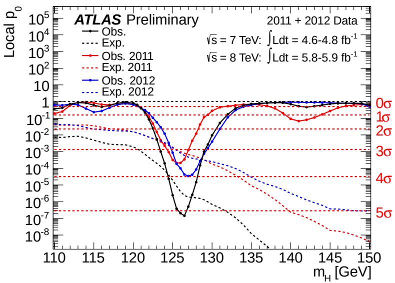




Example: strategy for $H \to \gamma\gamma$

- take data
- remove events with bad quality (subdetectors OFF, unstable conditions)
- select events containing particles in final state you want
- wait and plot what you find...
- ... summer 2012 - combine with $ZZ* \to 4\ell$ and $WW^* \to 2\ell 2\nu$ channel





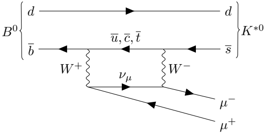
Example: strategy for $H \to \gamma\gamma$

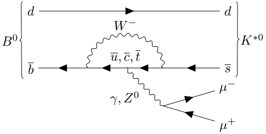
- announce discovery!

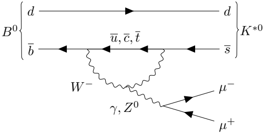
Other analyses

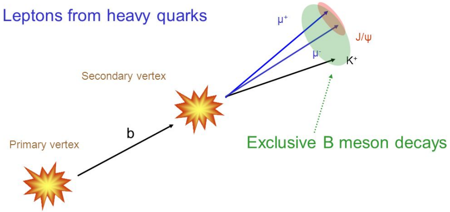
Example: my analysis of $B^0 \to K^*(K\pi)\mu^+\mu^-$ decay

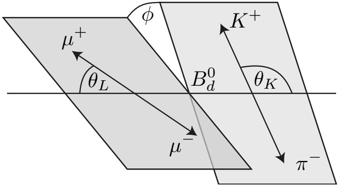
- very rare decay (1 of 10^7 B particles)
- can indirectly observe new particles/interactions in loops











$B^0 \to K^*(K\pi)\mu^+\mu^-$ selection, fitting, checks...

ATLAS analysis of $B_d \to K^*\mu^+\mu^-$

Dataset: ATLAS-CONF-2017-023

- $20.3\ \mathrm{fb}^{-1}$, taken at $\sqrt{s} = 8$ TeV in 2012

Reconstruction and selection:

- preselection: track p_T, ID hits, min. 1 combined muon
- baseline: $|\eta| < 2.5$, $m(K^*) = [846, 946]$ MeV, $m(B) = [5150, 5700]$ MeV $p_\mathrm{T}(\mu) > 3.5$ GeV, $p_\mathrm{T}(\pi, K) > 0.5$ GeV
- final cuts: $\sigma_\tau/\tau > 12.75$, pointing $\cos\theta > 0.999$, $\chi^2/ndf(B) < 2$, $p_\mathrm{T}(K^*) > 3$ GeV, $|(m(B) - m_{PDG}(B)) - (m(\mu\mu) - m_{PDG}(J/\psi))| < 130$ MeV
- trigger - 15 most frequent triggers
- control regions: J/ψ (q^2= [8,11] GeV^2), $\psi(2S)$ (q^2= [12,15] GeV^2)
- signal q^2= [0.04,6] GeV^2 except of ϕ region q^2= [0.98,1.1] GeV^2
- if > 1 candidate/event: candidate with higher $\sigma_m(K^*)/m(K^*)$

Monte Carlo datasets

Signal:

Process	Generator	Dataset	Events
$B_d \to K^*\mu^+\mu^-$	EvtGen, flat	208445	50M
$B_d \to K^*\mu^+\mu^-$	EvtGen, SM	208446	5M
$\overline{B}_d \to K^*\mu^+\mu^-$	EvtGen	208447	5M
$B_d \to K^+\pi^-\mu^+\mu^-$	EvtGen	208451	50M

Inclusive backgrounds:

Process	Generator	Dataset	Events
$b\bar{b} \to \mu^+\mu^- X$	Pythia	208301	20M
$b\bar{b} \to \mu^+\mu^- X$	EvtGen	208303	1M
$b\bar{b} \to \mu^+\mu^- X$ AA	Pythia	208308	40M
$b\bar{b} \to \mu^+\mu^- X$ AB	Pythia	208309	48M
$b\bar{b} \to \mu^+\mu^- X$ BA	Pythia	208310	48M
$b\bar{b} \to \mu^+\mu^- X$ BB	Pythia	208311	130M
$c\bar{c} \to \mu^+\mu^- X$	Pythia	208312	50M

$B_d \to K^*\mu^+\mu^-$ decay in SM

Differential decay rate (optimized):

$$\frac{1}{d\Gamma/dq^2}\frac{d^4\Gamma}{d\cos\theta_\ell d\cos\theta_K\, d\phi dq^2} = \frac{9}{32\pi}\Big[\frac{3(1-F_L)}{4}\sin^2\theta_K + \frac{1-F_L}{4}\sin^2\theta_K\cos 2\theta_\ell$$
$$+ F_L\cos^2\theta_K - F_L\cos^2\theta_K\cos 2\theta_\ell + S_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi$$
$$+ S_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi + S_5\sin 2\theta_K\sin\theta_\ell\cos\phi + S_6\sin^2\theta_K\cos\theta_\ell$$
$$+ S_7\sin 2\theta_K\sin\theta_\ell\sin\phi + S_8\ \sin 2\theta_K\sin 2\theta_\ell\sin\phi + S_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi\Big].$$

$$P_1 = \frac{2S_3}{(1-F_L)}, \quad P_2 = \frac{2}{3}\frac{A_{FB}}{(1-F_L)}, \quad P_3 = -\frac{S_9}{(1-F_L)},$$
$$P'_{4,5,6} = \frac{S_{4,5,6}}{\sqrt{F_L(1-F_L)}}, \quad P'_6 = \frac{S_7}{\sqrt{F_L(1-F_L)}}$$

$B^0 \to K^*(K\pi)\mu^+\mu^-$ selection, fitting, checks...

ATLAS analysis of $B_d \to K^*\mu^+\mu^-$

Dataset: ATLAS-CONF-2017-023

- $20.3\ \text{fb}^{-1}$, taken at $\sqrt{s} = 8$ TeV in 2012

Reconstruction and selection:

- preselection: track p_T, ID hits, min. 1 combined muon
- baseline: $|\eta| < 2.5$, $m(K^*) = [846, 946]$ MeV, $m(B) = [5150, 5700]$ MeV $p_T(\mu) > 3.5$ GeV, $p_T(\pi, K) > 0.5$ GeV
- final cuts: $\sigma_\tau/\tau > 12.75$, pointing $\cos\theta > 0.999$, $\chi^2/ndf(B) < 2$, $p_T(K^*) > 3$ GeV, $|(m(B) - m_{PDG}(B)) - (m(\mu\mu) - m_{PDG}(J/\psi))| < 130$ MeV
- trigger - 15 most frequent triggers
- control regions: J/ψ (q^2= [8,11] GeV^2), $\psi(2S)$ (q^2= [12,15] GeV^2)
- signal q^2= [0.04,6] GeV^2 except of ϕ region q^2= [0.98,1.1] GeV^2
- if > 1 candidate/event: candidate with higher ...

Monte Carlo datasets

Signal:

Process	Generator	Dataset	Events
$B_d \to K^*\mu^+\mu^-$	EvtGen, flat	208445	50M
$\bar{B}_d \to K^*\mu^+\mu^-$	EvtGen, SM	208446	5M
$\bar{B}_d \to K^*\mu^+\mu^-$	EvtGen	208447	5M
$B_d \to K^+\pi^-\mu^+\mu^-$	EvtGen	208451	50M

Inclusive backgrounds:

Process	Generator	Dataset	Events
$b\bar{b} \to \mu^+\mu^- X$	Pythia	208301	20M
$b\bar{b} \to \mu^+\mu^- X$	EvtGen	208303	1M
$b\bar{b} \to \mu^+\mu^- X$ AA	Pythia	208308	40M
$b\bar{b} \to \mu^+\mu^- X$ AB	Pythia	208309	48M
$b\bar{b} \to \mu^+\mu^- X$ BA	Pythia	208310	48M
$b\bar{b} \to \mu^+\mu^- X$ BB	Pythia	208311	130M

$B_d \to K^*\mu^+\mu^-$ decay in SM

Differential decay rate (optimized):

$$\frac{1}{d\Gamma/dq^2}\frac{d^4\Gamma}{d\cos\theta_\ell\, d\cos\theta_K\, d\phi dq^2} = \frac{9}{32\pi}\Big[\frac{3(1-F_L)}{4}\sin^2\theta_K + \frac{1-F_L}{4}\sin^2\theta_K\cos 2\theta_\ell + F_L\cos^2\theta_K - F_L\cos^2\theta_K\cos 2\theta_\ell + S_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi + S_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi + S_5\sin 2\theta_K\sin\theta_\ell\cos\phi + S_6\sin^2\theta_K\cos\theta_\ell + S_7\sin 2\theta_K\sin\theta_\ell\sin\phi + S_8\sin 2\theta_K\sin 2\theta_\ell\sin\phi + S_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi\Big].$$

$$P_1 = \frac{2S_3}{(1-F_L)},\quad P_2 = \frac{2}{3}\frac{A_{FB}}{(1-F_L)},\quad P_3 = -\frac{S_9}{(1-F_L)},$$

$$P'_{4,5,6} = \frac{S_{4,5,6}}{\sqrt{F_L(1-F_L)}},\quad P'_6 = \frac{S_7}{\sqrt{F_L(1-F_L)}}$$

$B_d \to K^*\mu^+\mu^-$ decay in SM

- small number of events → folding of angular distributions

$$P'_4, S_4: \begin{cases}\phi \to -\phi & \text{for } \phi < 0\\ \phi \to \pi - \phi & \text{for } \theta_l > \frac{\pi}{2}\\ \theta_l \to \pi - \theta_l & \text{for } \theta_l > \frac{\pi}{2}\end{cases}\qquad P'_6, S_7: \begin{cases}\phi \to \pi - \phi & \text{for } \phi > \frac{\pi}{2}\\ \phi \to -\pi - \phi & \text{for } \phi < -\frac{\pi}{2}\\ \theta_l \to \pi - \theta_l & \text{for } \theta_l > \frac{\pi}{2}\end{cases}$$

$$P'_5, S_5: \begin{cases}\phi \to -\phi & \text{for } \phi < 0\\ \theta_l \to \pi - \theta_l & \text{for } \theta_l > \frac{\pi}{2}\end{cases}\qquad P'_8, S_8: \begin{cases}\phi \to \pi - \phi & \text{for } \phi > \frac{\pi}{2}\\ \phi \to -\pi - \phi & \text{for } \phi < -\frac{\pi}{2}\\ \theta_l \to \pi - \theta_l & \text{for } \theta_l > \frac{\pi}{2}\\ \theta_k \to \pi - \theta_k & \text{for } \theta_l > \frac{\pi}{2}\end{cases}$$

$$\frac{1}{d\Gamma/dq^2}\frac{d^4\Gamma}{d\cos\theta_\ell\, d\cos\theta_K\, d\phi dq^2} = \frac{9}{8\pi}\Big[\frac{3(1-F_L)}{4}\sin^2\theta_K + F_L\cos^2\theta_K + \frac{1-F_L}{4}\sin^2\theta_K\cos 2\theta_\ell - F_L\cos^2\theta_K\cos 2\theta_\ell + S_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi + S_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi\Big]$$

- loss of sensitivity to S_6, S_8 a A_{FB}

Background: partially reconstructed decays $B \to D \to X$

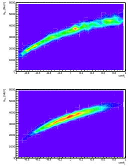
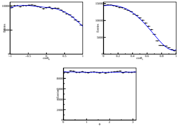


Fig: Example for $M(K\pi\pi)$, data vs. MC

Background

Result:

- D veto 30 MeV and B veto 50 MeV around $m_{D/B}$
- acceptance maps, e.g. q^2= [0.04,6] GeV^2



$B^0 \to K^*(K\pi)\mu^+\mu^-$ selection, fitting, checks...

ATLAS analysis of $B_d \to K^*\mu^+\mu^-$

Dataset: ATLAS-CONF-2017-023

- $20.3\ \mathrm{fb}^{-1}$, taken at $\sqrt{s} = 8$ TeV in 2012

Reconstruction and selection:

- preselection: track p_T, ID hits, min. 1 combined muon
- baseline: $|\eta| < 2.5$, $m(K^*) = [846, 946]$ MeV, $m(B) = [5150, 5700]$ MeV $p_T(\mu) > 3.5$ GeV, $p_T(\pi, K) > 0.5$ GeV
- final cuts: $\sigma_\tau/\tau > 12.75$, pointing $\cos\theta > 0.999$, $\chi^2/ndf(B) < 2$, $p_T(K^*) > 3$ GeV, $|(m(B) - m_{PDG}(B)) - (m(\mu\mu) - m_{PDG}(J/\psi))| < 130$ MeV
- trigger - 15 most frequent triggers
- control regions: J/ψ (q^2= [8,11] GeV^2), $\psi(2S)$ (q^2= [12,15] GeV^2)
- signal q^2= [0.04,6] GeV^2 except of ϕ region q^2= [0.98,1.1] GeV^2
- if > 1 candidate/event: candidate with higher $\sigma_\tau(K^*)/m(K^*)$

Monte Carlo datasets

Signal:

Process	Generator	Dataset	Events
$B_d \to K^*\mu^+\mu^-$	EvtGen, flat	208445	50M
$B_d \to K^*\mu^+\mu^-$	EvtGen, SM	208446	5M
$\bar{B}_d \to K^*\mu^+\mu^-$	EvtGen	208447	5M
$B_d \to K^+\pi^-\mu^+\mu^-$	EvtGen	208451	50M

Inclusive backgrounds:

Process	Generator	Dataset	Events
$b\bar{b} \to \mu^+\mu^- X$	Pythia	208301	20M
$b\bar{b} \to \mu^+\mu^- X$	EvtGen	208303	1M
$b\bar{b} \to \mu^+\mu^- X$ AA	Pythia	208308	40M
$b\bar{b} \to \mu^+\mu^- X$ AB	Pythia	208309	48M
$b\bar{b} \to \mu^+\mu^- X$ BA	Pythia	208310	48M
$b\bar{b} \to \mu^+\mu^- X$ BB	Pythia	208311	130M

$B_d \to K^*\mu^+\mu^-$ decay in SM

Differential decay rate (optimized):

$$\frac{1}{d\Gamma/dq^2}\frac{d^4\Gamma}{d\cos\theta_\ell\, d\cos\theta_K\, d\phi dq^2} = \frac{9}{32\pi}\Big[\frac{3(1-F_L)}{4}\sin^2\theta_K + \frac{1-F_L}{4}\sin^2\theta_K\cos 2\theta_\ell + F_L\cos^2\theta_K - F_L\cos^2\theta_K\cos 2\theta_\ell + S_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi + S_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi + S_5\sin 2\theta_K\sin\theta_\ell\cos\phi + S_6\sin^2\theta_K\cos\theta_\ell + S_7\sin 2\theta_K\sin\theta_\ell\sin\phi + S_8\sin 2\theta_K\sin 2\theta_\ell\sin\phi + S_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi\Big].$$

$$P_1 = \frac{2S_3}{(1-F_L)},\quad P_2 = \frac{2}{3}\frac{A_{FB}}{(1-F_L)},\quad P_3 = -\frac{S_9}{(1-F_L)},$$
$$P'_{4,5,6} = \frac{S_{4,5,6}}{\sqrt{F_L(1-F_L)}},\quad P'_6 = \frac{S_7}{\sqrt{F_L(1-F_L)}}$$

$B_d \to K^*\mu^+\mu^-$ decay in SM

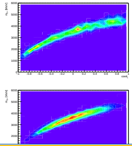
- small number of events → folding of angular distributions

$$P'_4, S_4: \begin{cases}\phi \to -\phi & \text{for } \phi < 0\\ \phi \to \pi - \phi & \text{for } \theta_l > \frac{\pi}{2}\\ \theta_l \to \pi - \theta_l & \text{for } \theta_l > \frac{\pi}{2}\end{cases} \qquad P'_6, S_7: \begin{cases}\phi \to \pi - \phi & \text{for } \phi > \frac{\pi}{2}\\ \phi \to -\pi - \phi & \text{for } \phi < -\frac{\pi}{2}\\ \theta_l \to \pi - \theta_l & \text{for } \theta_l > \frac{\pi}{2}\end{cases}$$

$$P'_5, S_5: \begin{cases}\phi \to -\phi & \text{for } \phi < 0\\ \theta_l \to \pi - \theta_l & \text{for } \theta_l > \frac{\pi}{2}\end{cases} \qquad P'_8, S_8: \begin{cases}\phi \to \pi - \phi & \text{for } \phi > \frac{\pi}{2}\\ \phi \to -\pi - \phi & \text{for } \phi < -\frac{\pi}{2}\\ \theta_l \to \pi - \theta_l & \text{for } \theta_l > \frac{\pi}{2}\\ \theta_K \to \pi - \theta_K & \text{for } \theta_l > \frac{\pi}{2}\end{cases}$$

$$\frac{1}{d\Gamma/dq^2}\frac{d^4\Gamma}{d\cos\theta_\ell\, d\cos\theta_K\, d\phi dq^2} = \frac{9}{8\pi}\Big[\frac{3(1-F_L)}{4}\sin^2\theta_K + F_L\cos^2\theta_K + \frac{1-F_L}{4}\sin^2\theta_K\cos 2\theta_\ell - F_L\cos^2\theta_K\cos 2\theta_\ell$$

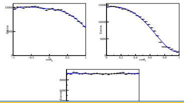
Background: partially reconstructed decays $B \to D \to X$



Background

Result:

- D veto 30 MeV and B veto 50 MeV around $m_{D/B}$
- acceptance maps, e.g. q^2= [0.04,6] GeV^2



Fitting

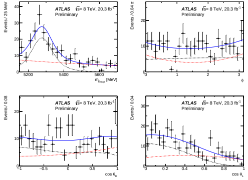
- acceptance maps from MC sample generated with flat angular distributions
- extract nuisance parameters from control regions - $m_B, \sigma_m B$ (Gauss),
- fold distributions → 4 sets of fits
- mass prefit - number of signal a bkg evens
- angular fits

Fitted yields:

q^2[(GeV^2)]	n_{signal}	$n_{\mathrm{background}}$
[0.04, 2.0]	128 ± 22	122 ± 22
[2.0, 4.0]	106 ± 23	113 ± 23
[4.0, 6.0]	114 ± 24	204 ± 26
[0.04, 4.0]	236 ± 31	233 ± 32
[1.1, 6.0]	275 ± 35	363 ± 36
[0.04, 6.0]	342 ± 39	445 ± 40

Ina Carli (Charles University) 19.4.2017 17 / 22

Fit results: S4, bin q^2= [0.04,2] GeV^2

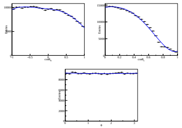


Ina Carli (Charles University) 19.4.2017 18 / 22

Background

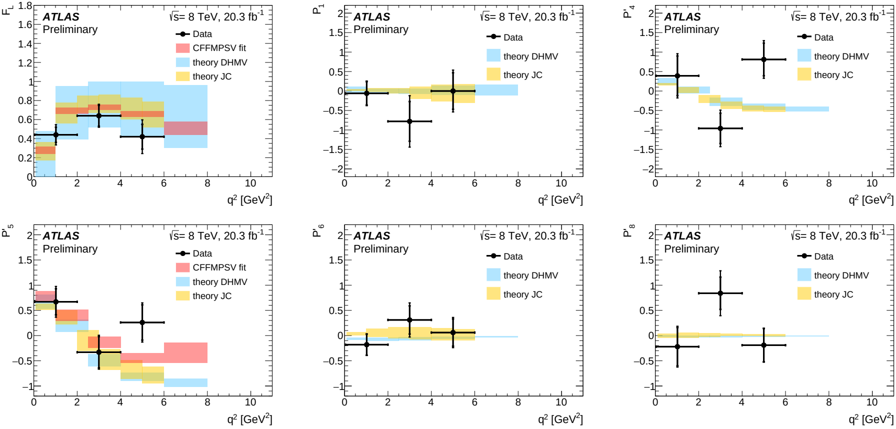
Result:

- D veto 30 MeV and B veto 50 MeV around $m_{D/B}$
- acceptance maps, e.g. q^2= [0.04,6] GeV^2



Ina Carli (Charles University) 19.4.2017 16 / 22

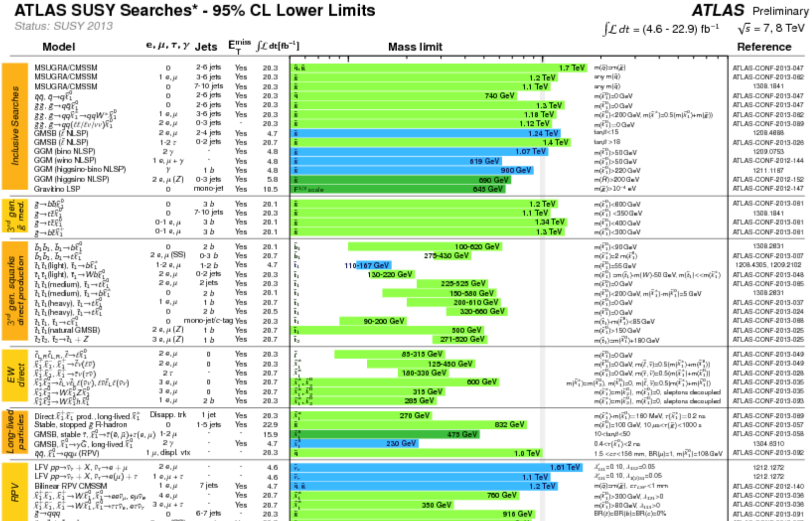
$B^0 \to K^*(K\pi)\mu^+\mu^-$ results



... 3 bins with $\sim 3\sigma$ from SM prediction

Other ATLAS analyses

- some 3000 members of collaboration
- working in 100 sub-groups
- any decay and measurement we can think about



Thank you for your attention!
Questions?

... or write me on ina@cern.ch and we can chat about it over coffee