



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN



# Grundlagen der Analyse

Bachelor-Einführungskurs 2017

**Sascha Mehlhase** (LMU München)

# Überblick

- Typischer Aufbau einer Analyse
  - Datenauswahl (event selection I)
  - Objektdefinition (object definition/preselection)
  - Auflösen von Mehrdeutigkeiten (overlap removal)
  - Objektselektion (object selection)
  - Berechnung kinematischer Observablen (kinematic observables)
  - Datenauswahl (event selection II)
  - Kontroll-, Validierungs- und Signalregionen
  - Statistische Analyse



# Überblick

## ▸ Typischer Aufbau einer Analyse

- Datenauswahl (event selection I)
- Objektdefinition (object definition/preselection)
- Auflösen von Mehrdeutigkeiten (overlap removal)
- Objektselektion (object selection)
- Berechnung kinematischer Observablen (kinematic observables)
- Datenauswahl (event selection II)
- Kontroll-, Validierungs- und Signalregionen
- Statistische Analyse



Beispielanalyse I:

**Search for electroweak production of supersymmetric particles in  $\sqrt{s} = 8$  TeV pp collisions with the ATLAS detector**

<http://arxiv.org/abs/1509.07152>

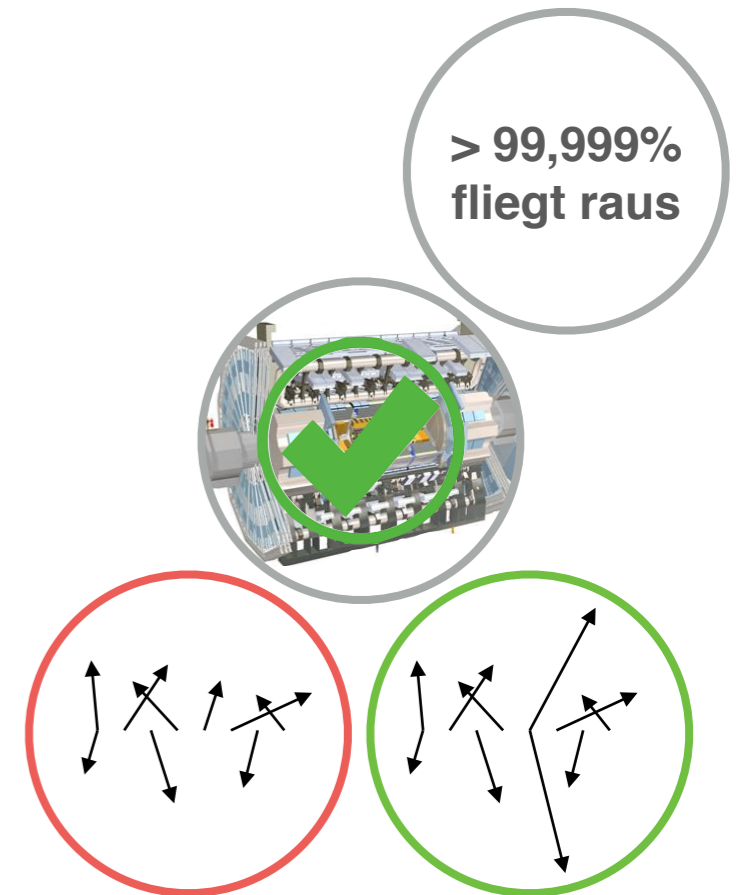
Beispielanalyse II:

**Search for heavy long-lived charged R-hadrons with the ATLAS detector in 3.2/fb of proton-proton collision data at  $\sqrt{s} = 13$  TeV**

<http://arxiv.org/abs/1606.05129>

# Typischer Aufbau einer Analyse

- Datenauswahl (event selection I)
  - ATLAS und CMS sind Vielzweck-Detektoren
    - Daten werden für eine Vielzahl verschiedener Messungen/Analysen gesammelt
    - je nach Analyse können unterschiedliche Ereignisse interessant sein
  - Abfrage eines geeigneten Triggers
    - i.d.R. anhand markanter/signifikanter Objekte im Ereignis
  - Abfrage notwendiger Detektorkomponenten
    - Good Runs List (GRL)
  - Abfrage grundlegender Ereignisparameter
    - z.B. gibt es einen klar rekonstruierten primären Vertex



# Typischer Aufbau einer Analyse

## ▸ Datenauswahl (event selection I)

Search for electroweak production of supersymmetric particles in  $\sqrt{s} = 8$  TeV pp collisions with the ATLAS detector

“Events recorded during stable data-taking conditions are analyzed if the reconstructed primary vertex has five or more tracks with transverse momentum  $p_T > 400$  MeV associated with it. The primary vertex of an event is identified as the vertex with the highest  $\Sigma p_T^2$  of associated tracks.”

“After the application of beam, detector and data-quality requirements ...”

Trigger	$p_T$ threshold [ GeV]	Analysis
Single $\tau$ Double $\tau$	150 40,25	Direct stau production
Single Isolated $e$ Single Isolated $\mu$	25 25	Compressed spectra $\ell^+\ell^-$ , $3\ell$
Double $e$	14,14 25,10	Compressed spectra $\ell^+\ell^-$ , $\ell^\pm\ell^\pm$ , $3\ell$
Double $\mu$	14,14 18,10	Compressed spectra $\ell^+\ell^-$ , $\ell^\pm\ell^\pm$ , $3\ell$
Triple $e$	20,9,9	Compressed spectra $3\ell$
Triple $\mu$	7,7,7 19,5,5	Compressed spectra $3\ell$
Combined $e\mu$	14( $e$ ),10( $\mu$ ) 18( $\mu$ ),10( $e$ ) 9( $e$ ),9( $e$ ),7( $\mu$ ) 9( $e$ ),7( $\mu$ ),7( $\mu$ )	Compressed spectra $3\ell$
$E_T^{\text{miss}}$	120	Chargino production via VBF

Search for heavy long-lived charged R-hadrons with the ATLAS detector in 3.2/fb of proton-proton collision data at  $\sqrt{s} = 13$  TeV

“Events are selected online via a trigger based on the magnitude of the missing transverse momentum ...”

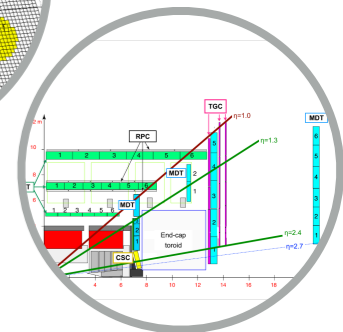
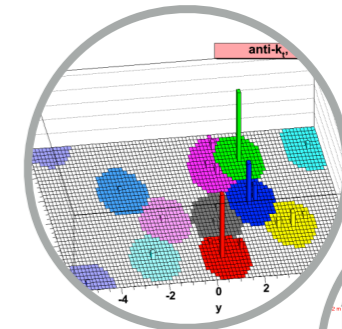
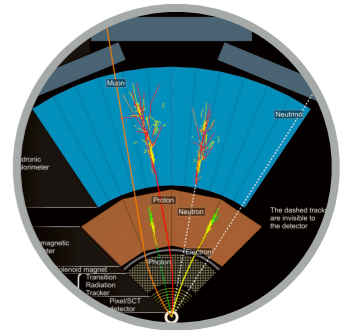
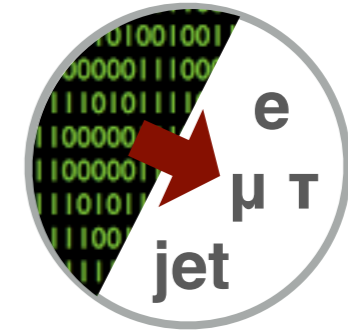
“... offline event selection requires all relevant detector components to be fully operational ...”

“... primary vertex (PV) built from at least two well-reconstructed charged-particle tracks, each with a transverse momentum,  $p_T$ , above 400 MeV ...”



# Typischer Aufbau einer Analyse

- Objektdefinition (object definition / preselection)
  - Rohdaten werden zu *physikalischen* Objekten
- Objektidentifikation (object identification)
  - Vielzahl verschiedener Objekte
    - Leptonen
    - Jets
    - Fehlende transversale Energie
    - ...
- Objektrekonstruktion (object reconstruction)
  - verschiedene Algorithmen zur Rekonstruktion eines Objekts vorhanden
    - Jetalgorithmen (anti- $k_t$ ,  $k_t$ , Cambridge-Aachen, ...)
    - Myonen (combined, stand-alone, extrapolated, ...)
    - ...
  - verschiedene Qualitätsklassen (baseline, loose, medium, tight, ...)
    - definieren wie gut ein Objekt rekonstruiert ist und wie frei von *Überlagerungen* es ist



# Typischer Aufbau einer Analyse

## ▸ Objektdefinition (object definition / preselection)

Search for electroweak production of supersymmetric particles in  $\sqrt{s} = 8$  TeV pp collisions with the ATLAS detector

*“Electrons must satisfy “medium” identification criteria ...”*

*“Muon candidates are reconstructed by combining tracks in the ID and tracks in the MS ...”*

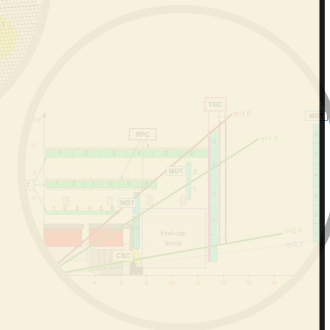
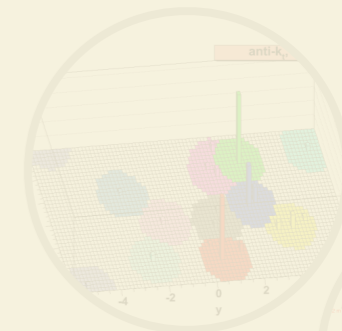
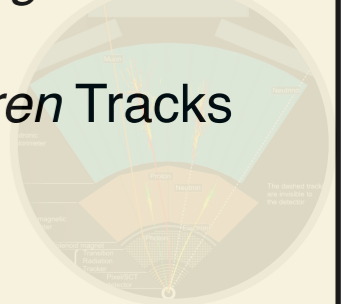
*“Jets are reconstructed with the anti-kt algorithm ... Three-dimensional calorimeter energy clusters are used as input to the jet reconstruction. The clusters are calibrated using the local hadronic calibration ... The final jet energy calibration corrects the calorimeter response to the particle-level jet energy ... Corrections for in-time and out-of-time pileup are also applied based on the jet area method ...”*

...

Search for heavy long-lived charged R-hadrons with the ATLAS detector in 3.2/fb of proton-proton collision data at  $\sqrt{s} = 13$  TeV

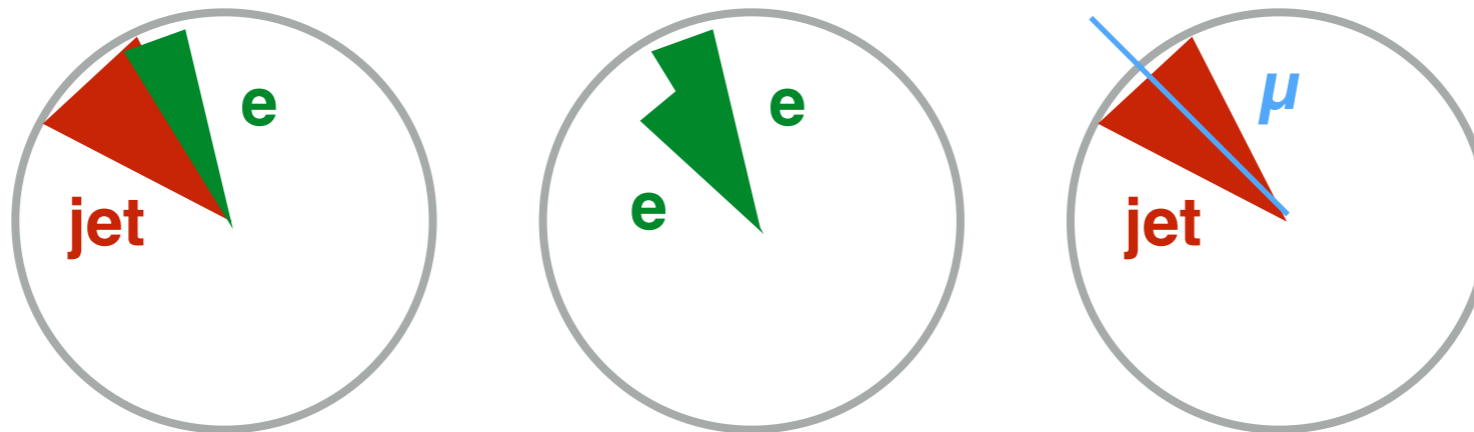
*“... reconstructed using the anti-kt jet algorithm ...”*

spezielle Analyse sucht nach *besonderen* Tracks



# Typischer Aufbau einer Analyse

- Auflösen von Mehrdeutigkeiten (overlap removal)
  - unterschiedliche Objekte können auf der selben Detektorinformation basieren
    - Unterobjekte werden als eigenständige Objekte fehlidentifiziert (z.B. Myonen in b-Jets)
    - Sekundäreffekte führen zu *falschen* Objekte (Jets basierend auf Brems/Punch-through von Leptonen)
  - Objekte unterschiedlichem Typs nach *Rekonstruktionsqualität* gegeneinander ausgespielt (mehr oder weniger)
    - z.B. Jet rekonstruiert nahe Elektron nicht als solches betrachtet (Brems/punch-through/...)
  - Objekte gleichen Typs bei zu großem Überlapp gegeneinander priorisiert
    - z.B. Vernachlässigung des niederenergetischen Elektrons bei (nahezu) Überlapp der Objekte





# Typischer Aufbau einer Analyse

## ▸ Auflösen von Mehrdeutigkeiten (overlap removal)

Search for electroweak production of supersymmetric particles in  $\sqrt{s} = 8$  TeV pp collisions with the ATLAS detector

*“To avoid potential ambiguities among objects, “tagged” leptons are candidate leptons separated from each other and from jets in the following order:*

- 1. If two electron candidates are reconstructed with  $\Delta R < 0.1$ , the lower energy candidate is discarded.*
- 2. Jets within  $\Delta R = 0.2$  of an electron candidate, and  $\tau$  had candidates within  $\Delta R = 0.2$  of an electron or muon, are discarded.*
- 3. Electron and muon candidates are discarded if found within  $\Delta R = 0.4$  of a remaining jet to suppress leptons from semileptonic decays of  $c$ - and  $b$ -hadrons.*
- 4. To reject bremsstrahlung from muons,  $e\mu$  ( $\mu\mu$ ) pairs are discarded if the two leptons are within  $\Delta R = 0.01$  (0.05) of one another.*
- 5. Jets found within  $\Delta R = 0.2$  of a “signal”  $\tau$  lepton (see below) are discarded”*

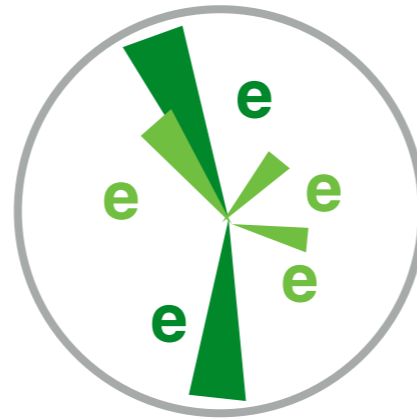
Search for heavy long-lived charged R-hadrons with the ATLAS detector in 3.2/fb of proton-proton collision data at  $\sqrt{s} = 13$  TeV

*“Candidates must not be within  $\Delta R = 0.3$  of any jet with  $p_T > 50$  GeV ... must not have any additional nearby ( $\Delta R < 0.2$ ) tracks with  $p_T > 10$  GeV.”*



# Typischer Aufbau einer Analyse

- Objektselektion (object selection)
  - Auswahl von Objekten mit bestimmten kinematischen Eigenschaften
    - minimaler transversaler Impuls
    - bestimmter Detektorbereich (z.B. zentral in Pseudorapidität)
    - physik-motivierte Anforderung an Isolation
    - ...



# Typischer Aufbau einer Analyse

## ▸ Objektselektion (object selection)

Search for electroweak production of supersymmetric particles in  $\sqrt{s} = 8$  TeV pp collisions with the ATLAS detector

“Electron candidates are required to have  $|\eta| < 2.47$  and  $p_T > 7$  GeV ...”

“Muon candidates ... are required to have  $|\eta| < 2.5$  and  $p_T > 5$  GeV.”

“... jets must have  $|\eta| < 2.4$  and  $p_T > 20$  GeV ...”

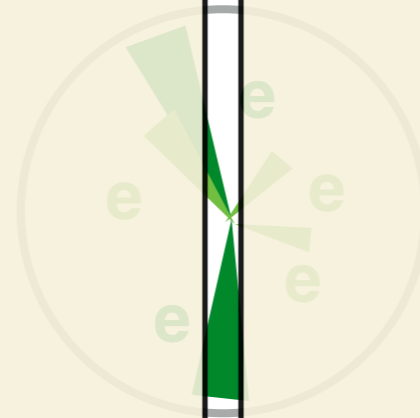
...

Search for heavy long-lived charged R-hadrons with the ATLAS detector in 3.2/fb of proton-proton collision data at  $\sqrt{s} = 13$  TeV

“... candidates are based on ID tracks with  $p_T > 50$  GeV and  $|\eta| < 1.65$  ...”

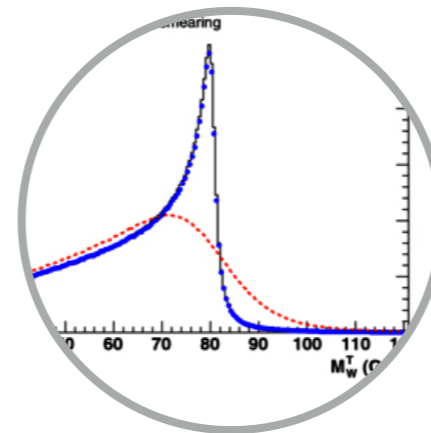
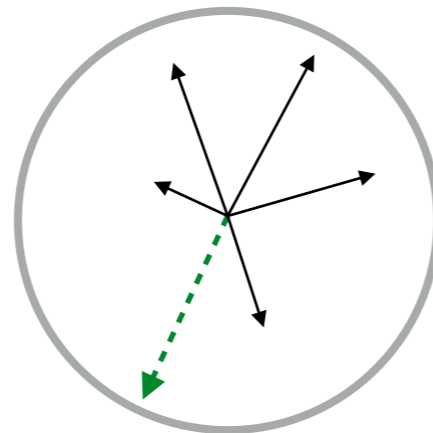
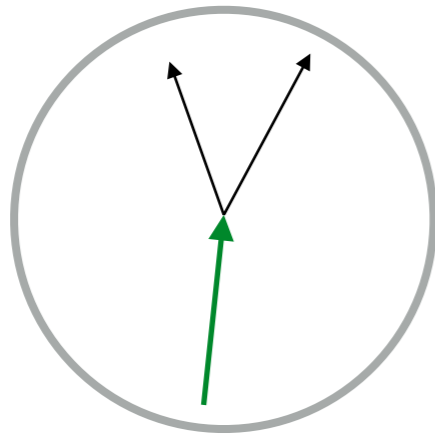
“Candidates with  $|z_0^{PV} \sin(\theta)| > 0.5$  mm or  $|d_0| > 2.0$  mm are removed ...”

...



# Typischer Aufbau einer Analyse

- Berechnung kinematischer Observablen (kinematic observables)
  - mit Hilfe der selektierten Objekt Ereignis-Observablen bauen
    - invariante Masse
    - fehlende transversale Masse (ggf. neu berechnet)
    - $\Sigma p_T$  (skalare Summe *aller* transversalen Impulse)
    - $H_T$  (skalare Summe *aller* transversalen Impulse und Energieablagerungen)
    - $m_T$  (transversale Masse  $m_T = \sqrt{(|p_{T,1}| |p_{T,2}| (1 - \cos \phi_{12}))}$  )
    - $m_{T2}$  ('stransverse mass')
    - ...



event variable used to bound the masses of an unseen pair of particles which are presumed to have decayed semi-invisibly into particles which were seen. it is a function of the momenta of two visible particles and missing transverse momentum

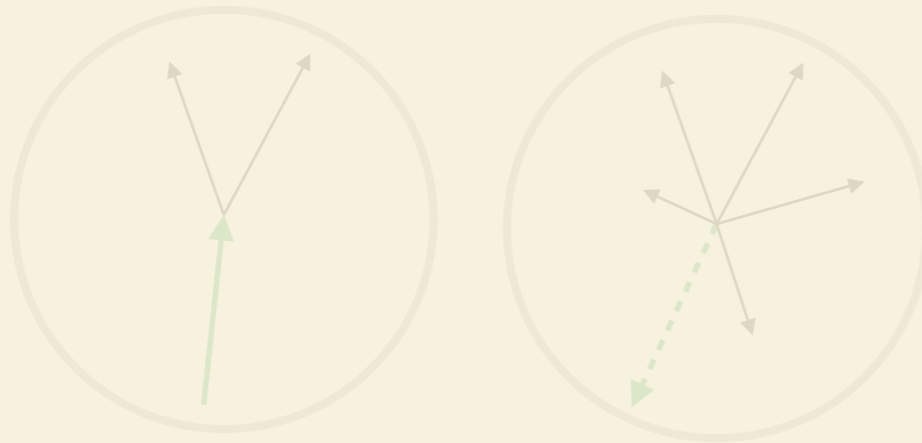
# Typischer Aufbau einer Analyse

- Berechnung kinematischer Observablen (kinematic observables)

Search for electroweak production of supersymmetric particles in  $\sqrt{s} = 8$  TeV pp collisions with the ATLAS detector

invariante Masse  
 fehlende transversale Masse (ganz allgemein)  
 $\Delta\phi(E_T^{miss}, \ell) \dots |\Delta\eta_{jj}| \dots E_T^{miss} \dots E_T^{miss,rel} \dots p_{T\ell\ell}$   
 $\dots H_T \dots m_T \dots m_{SFOS} \dots m_{SFOS}^{min} \dots m_{\ell\ell\ell} \dots m_{\tau\tau} \dots$   
 $m_{T2} \dots m_{eff} \dots R_2 \dots$

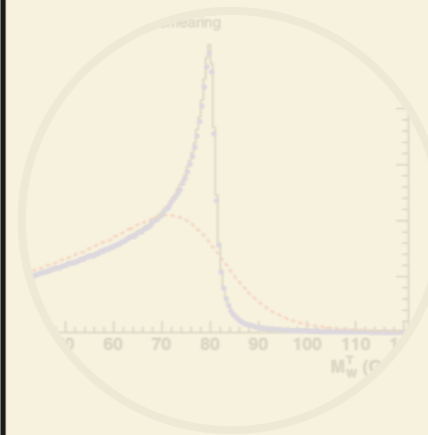
Analyse nutzt viele Observablen in multivariater Methode



Search for heavy long-lived charged R-hadrons with the ATLAS detector in 3.2/fb of proton-proton collision data at  $\sqrt{s} = 13$  TeV

Observables können auch offline kalibriert werden

*“A series of calibration techniques is applied to achieve optimal performance ...”*



event variable used to bound the masses of an unseen pair of particles which are presumed to have decayed semi-invisibly into particles which were seen. it is a function of the momenta of two visible particles and missing transverse momentum

# Typischer Aufbau einer Analyse

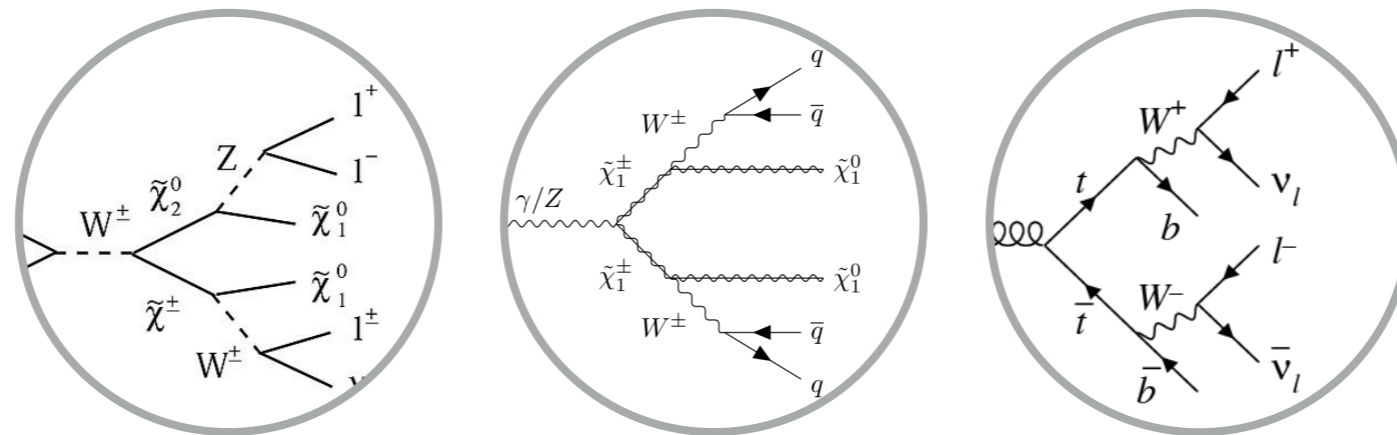
- Datenauswahl (event selection II)

- Auswahl von Signalereignissen basierend auf erwarteter Topologie

- Anforderung an Anzahl bestimmter Objekte (Multiplizität)
- Mindestwerte für einzelne kinematischen Ereignisgrößen

- Optimierung der Auswahl auf bestmögliches Signal-zu-Untergrund-Verhältnis

- beste Signifikanz ( $S/B$ ,  $S/\sqrt{B}$ ,  $S/(S+B)$ , ...)



# Typischer Aufbau einer Analyse

## ▸ Datenauswahl (event selection II)

Search for electroweak production of supersymmetric particles in  $\sqrt{s} = 8$  TeV pp collisions with the ATLAS detector

“if taus form an SFOS pair with  $m_{\text{SFOS}} < 12$  GeV, the event is rejected. Events with exactly two hadronically decaying tau candidates are selected, ... required to have opposite-sign (OS) charge. ... To suppress events from Z boson decays, events are rejected if the invariant mass of the tau pair lies within  $\pm 10$  GeV of the peak value of 81 GeV for Z boson candidates. To suppress background from events containing a top quark, events with b-tagged jets are vetoed. To further select SUSY events from direct stau production and suppress WW and  $t\bar{t}$  production,  $m_{\text{T2}}$  is calculated using the two taus and the E miss T in the event. The additional requirement of  $m_{\text{T2}} > 30$  GeV is applied to select events for the training and optimization of the multivariate analysis ...”

Beispiel für eine der Topologien, es existieren mehrere Signalregionen

...

Search for heavy long-lived charged R-hadrons with the ATLAS detector in 3.2/fb of proton-proton collision data at  $\sqrt{s} = 13$  TeV

“... at least one R-hadron candidate track that meets the criteria specified ...”

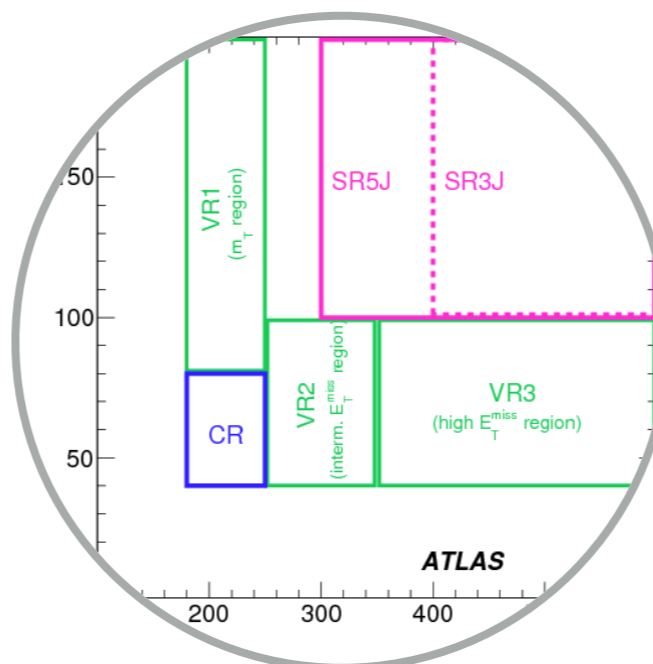
implizite Voraussetzung an  $E_{\text{T}}^{\text{miss}}$  durch entsprechen Trigger (normalerweise Offline-Anforderung im *Plateau*)

“... candidates ... are rejected if an oppositely-charged track with almost specular direction ... is identified on the opposite side of the detector.”



# Typischer Aufbau einer Analyse

- Kontroll-, Validierungs- und Signalregionen
  - Signalregion(en)
    - reich an Signal, arm an Untergründen (so gut wie möglich)
    - beste Signifikanz ( $S/B$ ,  $S/\sqrt{B}$ ,  $S/(S+B)$ , ...)
  - Kontrollregion(en) zur datenbasierten Abschätzung von Untergründen
    - arm an Signal, reich an Untergründen (ggf. eine Region pro Untergrund)
    - orthogonal zu Signalregionen
  - Validierungsregion(en) zur Kontrolle der Abschätzungen
    - nah an Signalregionen, wenig Signal (idealerweise)
    - orthogonal zu Kontroll- und Signalregionen





# Typischer Aufbau einer Analyse

## ► Kontroll-, Validierungs- und Signalregionen

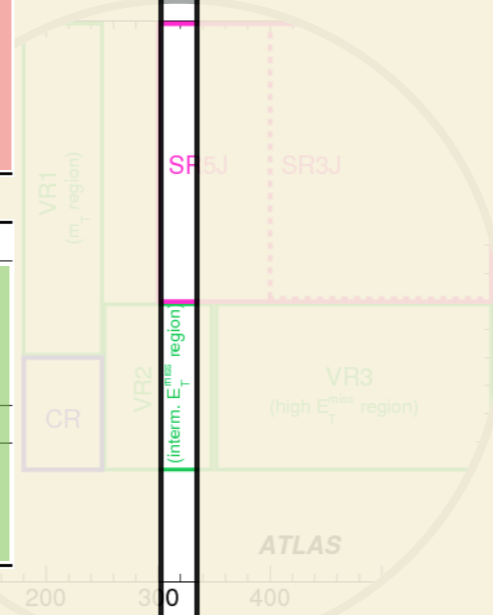
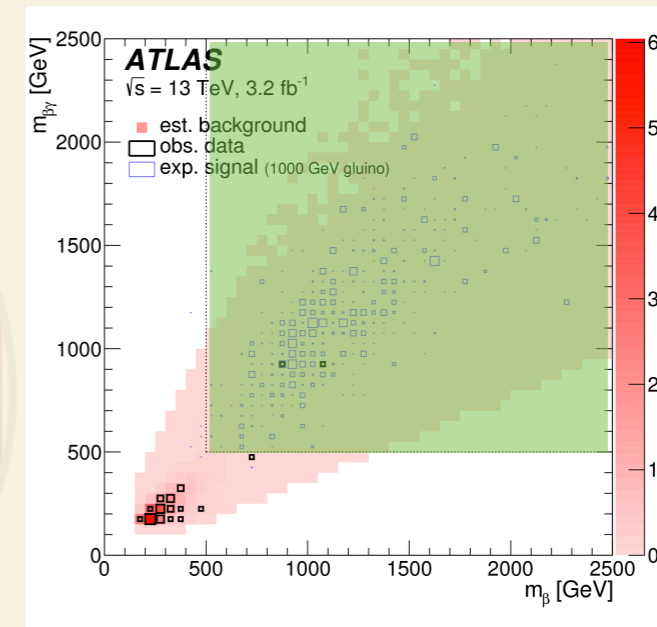
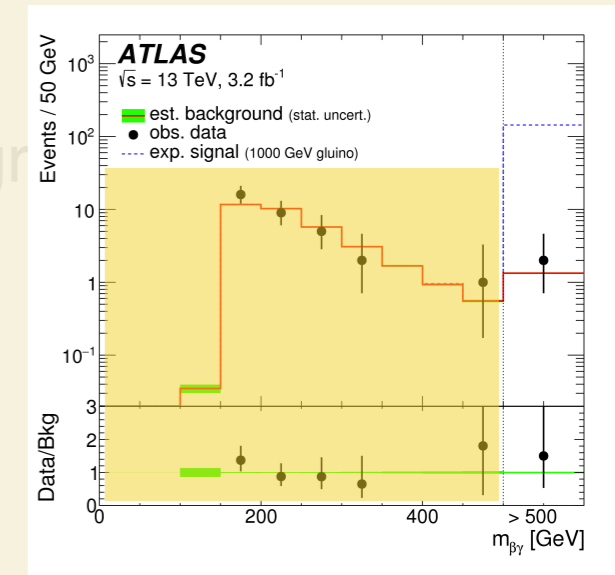
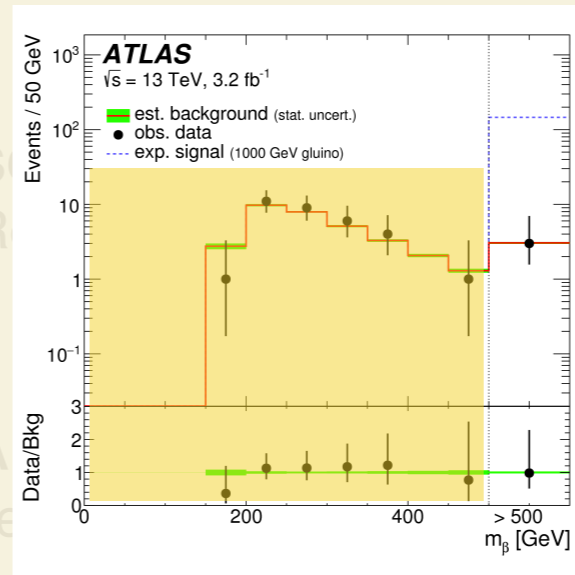
Search for electroweak production of supersymmetric particles in  $\sqrt{s} = 8$  TeV pp collisions with the ATLAS detector

	Common				
	exactly 2 medium OS taus ≥ 1 tight tau tagged $\ell$ veto $b$ -jet veto Z-veto				
	Signal region SR	Multi-jet VR1	Multi-jet VR2	W-VR1	W-VR2
$m_{T2}$	> 30 GeV	30–50 GeV	50–80 GeV	> 30 GeV	> 30 GeV
$E_T^{\text{miss}}$	–	–	–	> 100 GeV	> 90 GeV
$t_{\text{cut}}$	> 0.07	< 0.07	< 0.07	–0.2–0.07	–0.2–0.07

	Common					
Central light-flavor jets Forward jets $M_{\Delta}^R$ [GeV]	=1 veto > 20					
	SR2 $\ell$ -1a	SR2 $\ell$ -1b	CR2 $\ell$ -Top	CR2 $\ell$ -WW	CR2 $\ell$ -ZV	
$\ell$ flavor/sign	$\ell^{\pm}\ell^{\mp}$	$\ell^{\pm}\ell^{\mp}$	$\ell^{\pm}\ell^{\mp}$	$\ell^{\pm}\ell^{\mp}$	$\ell^{\pm}\ell^{\mp}$	$\ell^{\pm}\ell^{\mp}$
Central $b$ -tagged jets	veto					
$m_{\text{SFOS}}$ [GeV]	veto 81.2–101.2					
$p_T^{\ell\ell}$ [GeV]	–	< 40	< 40	–	> 70	> 70
$p_T^{\text{jet}}$ [GeV]	> 80	> 80	> 60	> 80	–	–
$R_2$	> 0.5	> 0.7	> 0.65	> 0.75	–	–
$\Delta\phi_{\beta}^{\beta}$ [rad]	> 2	> 2.5	> 2	> 2.5	–	< 2
$p_T^{\text{central light jet}}$ [GeV]	–	–	–	–	> 80	–

	Common			
$\ell$ flavor/sign	$\ell^{\pm}\ell^{\mp}\ell, \ell^{\pm}\ell^{\mp}\ell'$			
$p_T^{\text{lep 1}}$	< 30 GeV			
$b$ -jet	veto			
$E_T^{\text{miss}}$	> 50 GeV			
$m_{\text{SFOS}}$	veto 8.4–10.4 GeV			
SR	SR3 $\ell$ -0a	SR3 $\ell$ -0b	SR3 $\ell$ -1a	SR3 $\ell$ -1b
Central jets	no jets $p_T > 50$ GeV		≥ 1 jet $p_T > 50$ GeV	
$m_{\text{SFOS}}^{\text{min}}$	4–15 GeV	15–25 GeV	5–15 GeV	15–25 GeV
Other	$30 < m_{\ell\ell} < 60$ GeV $m_T < 20$ GeV	$30 < m_{\ell\ell} < 60$ GeV	$\Delta\phi(E_T^{\text{miss}}, \text{jet 1}) > 2.7$ rad $p_T^{\text{lep 1}}/p_T^{\text{jet 1}} < 0.2$	$\Delta\phi(E_T^{\text{miss}}, 3\ell) > 0.7\pi$ rad

Search for heavy long-lived charged R-hadrons with the ATLAS detector in 3.2/fb of proton-proton collision data at  $\sqrt{s} = 13$  TeV



# Typischer Aufbau einer Analyse

- Statistische Analyse
  - Zählen der erwarteten und beobachteten Ereignisse in einer Signalregion
    - ‘single-bin cut-and-count’
  - analog mit mehreren Signalregionen
    - ‘multi-bin cut-and-count’
  - Verteilungs-Fits in einer oder mehreren Signalregionen
    - ‘shape fit’ (effektiv ein ‘multi-bin’-Fit der jeden Bin einer Verteilung berücksichtigt)
  - analog mit Normalisierung der Untergründe als Teil des Fits
    - ‘multi-bin shape fit including control and validation regions’
    - ggf. werden nur ausgewählte Untergründe in Fits (durch Daten) normalisiert und andere direkt aus Monte-Carlo-Simulationen entnommen
  - ...

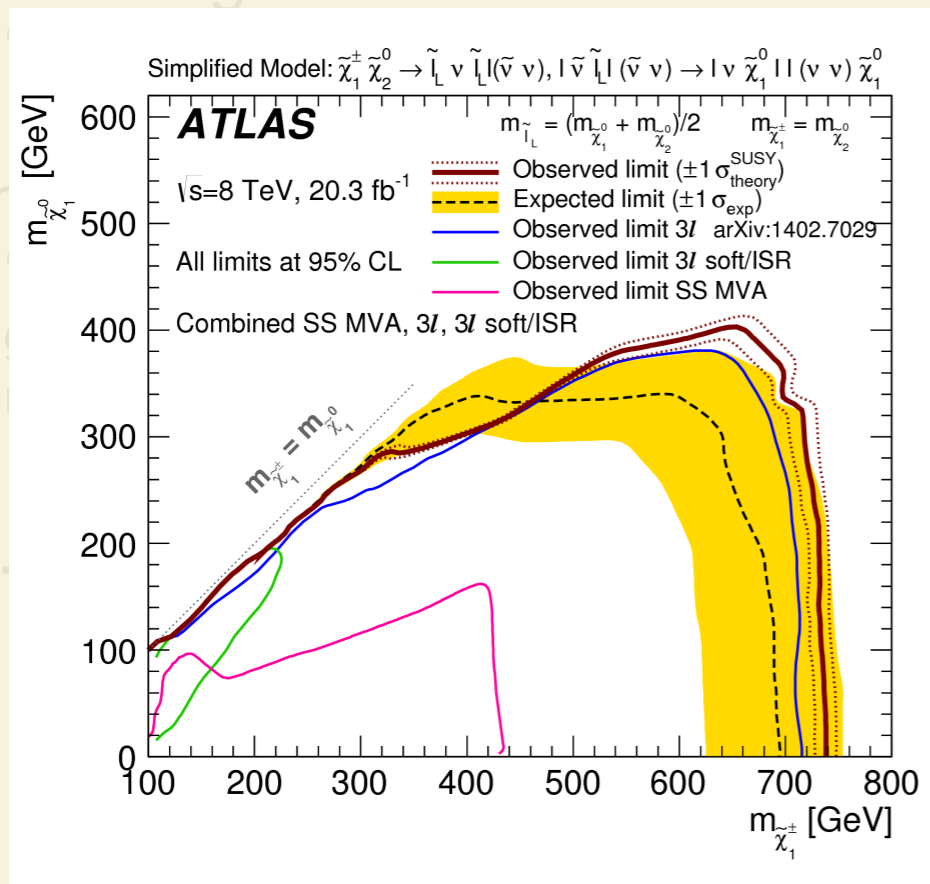


# Typischer Aufbau einer Analyse

## ► Statistische Analyse

Search for electroweak production of supersymmetric particles in  $\sqrt{s} = 8$  TeV pp collisions with the ATLAS detector

“The HistFitter ... framework is used in all analyses for constraining the background normalizations and the statistical interpretation of the results.”



Search for heavy long-lived charged R-hadrons with the ATLAS detector in 3.2/fb of proton-proton collision data at  $\sqrt{s} = 13$  TeV

“... limits are obtained from the expected signal and the estimated background in the signal region and using a one-bin counting experiment applying the CLs prescription ...”

