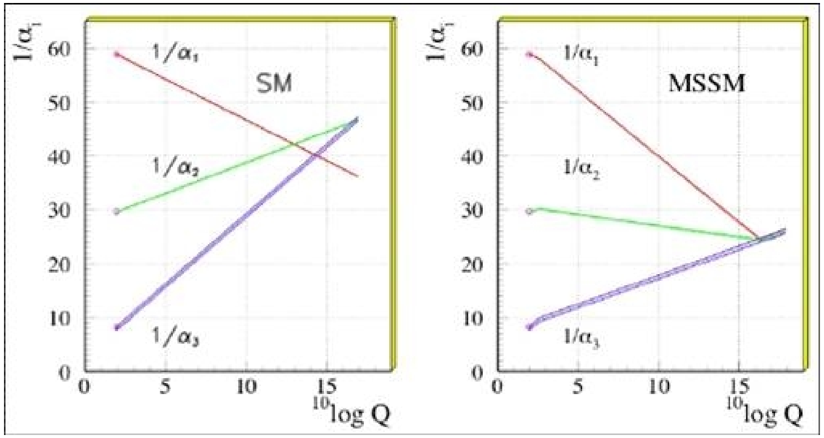
A Data-Driven Method to Estimate the Top Pair Background for SUSY Searches

ROYAL INSTITUTE OF TECHNOLOGY

Per Hansson (KTH)
Jörgen Sjölin, Christophe Clement (Stockholm University),
Vadym Zhuravlov (Munich)

Backgrounds to New Physics?

- TeV-scale SUSY – one attractive extension to the Standard Model
  - ✔ Natural explanation for M_h~100 GeV – addresses the hierarchy problem
  - ✔ Natural candidate for Dark Matter – lightest SUSY particle stable
  - ✔ Consistency with EW-data – coupling unification at GUT scale



- SM particles have **s**particle partners with Δ(spin)=±1/2 (min. extension MSSM)
- Gauge boson- and Higgs partners mix (gauginos and Higgsinos)
  - ⇨ four neutralinos (χ^0_i) and two charginos($\chi^\pm_i$)
  - – Sfermions also mix (according to their mass ie. 3rd generation different)

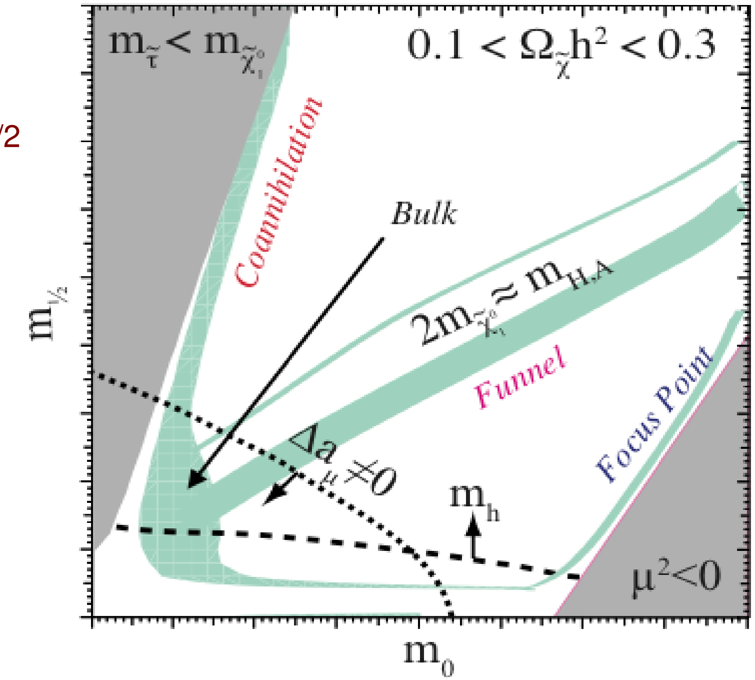
"If new physics <10^6 GeV is not SUSY, it quacks very much like it"
John Ellis, CERN*

*hep-ph/0203114

SUSY a signal scenario

- MSSM is simplest general SUSY model (but >100 free parameters)
  - Arbitrary choice gives FCNC, proton decay, etc.
  - mSUGRA: gravitational interactions -> unified couplings and masses at GUT scale $[m_0, m_{1/2}, A_0, \text{sign}(\mu), \tan(\beta)]$
- Typically good discovery reach in m_0-$m_{1/2}$ plane
- Only parts of parameter space give correct relic density
- Don't "trust" mSUGRA
  - Use it for benchmarking
  - Choose a few representable points for detailed studies

Qualitative plot – no mass scale



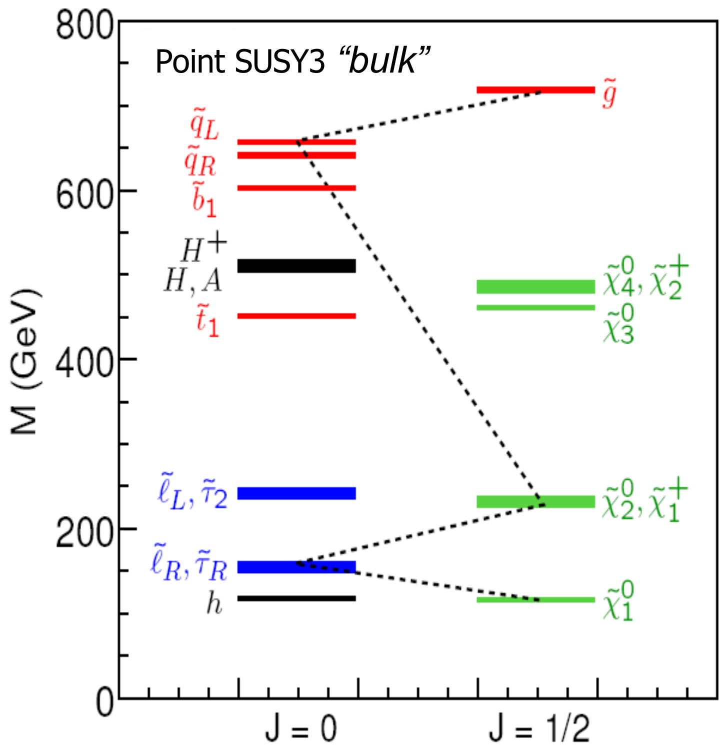
SUSY event signatures

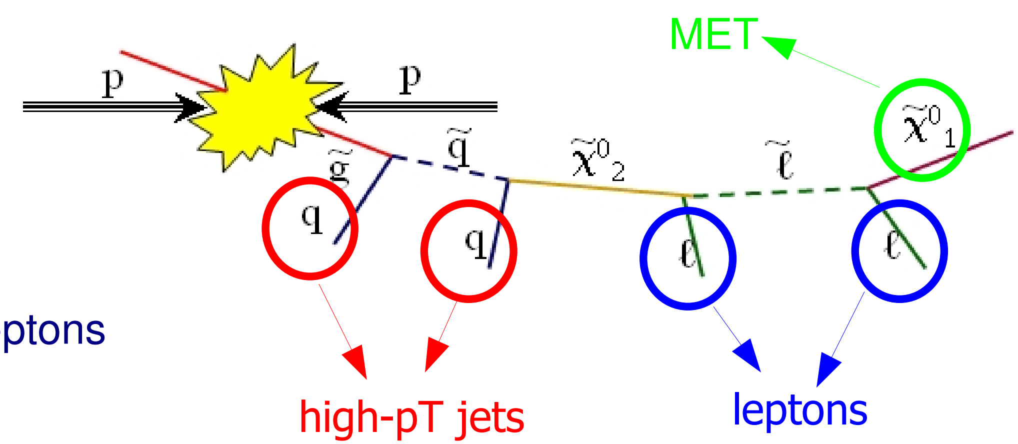
- Squark and gluino production dominates at LHC

$$R = \left(-1\right)^{3B-3L-2S}$$

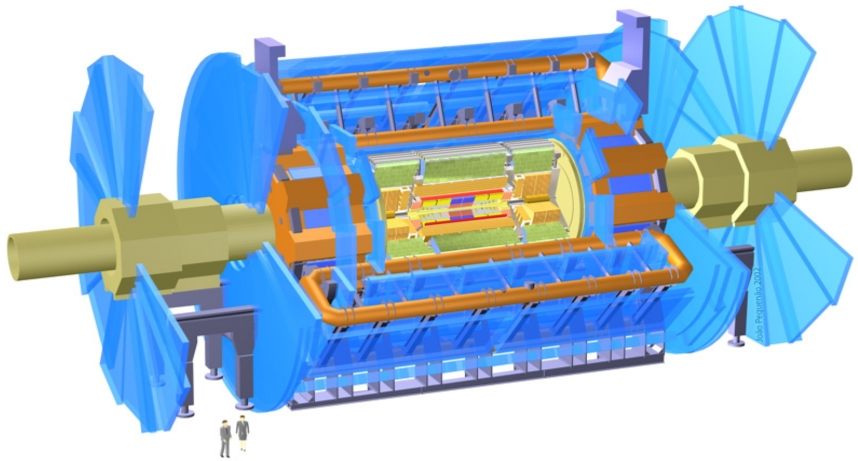
gluino & squark pair production | lightest SUSY particle stable

- Complex SUSY phenomenology
  - High mass objects cascade decay to $\chi^0{}_1$
- Smoking gun of SUSY
  - Large missing tr. energy (MET)
  - High transverse momentum jets
- # isolated leptons gives natural separation between analyzes
  - Focusing on **signatures** with two leptons (e or μ) + jets + MET





ATLAS a multi-purpose detector



Overall length: 44 m
height: 25 m
weight: 7000 tons

- LHC operation goals
  - Beam commissioning in May 2008
  - 14 TeV collisions in July 2008
  - Push towards $10^{32}cm^{-2}s^{-1}$ end of 2008
  - Integrated luminosity $1fb^{-1}$ -> 1-2 months of data-taking

- Muon system
  - Toroid magnets
  - 3 layers of drift tubes, strip chambers and resistive plate chambers
- Calorimetry
  - LAr-lead EM calorimeter
  - Steel-tile (/LAr-Cu) hadronic calorimeter
- Inner detector
  - 3 layers of silicon pixels (r=5-24 cm)
  - 4 layers microstrip-tracker (r<55 cm)
  - Straw tube tracker (r<108 cm)

- ATLAS commissioning
  - Cosmic data taken daily w/ full DAQ
  - Ready for collisions in May 2008

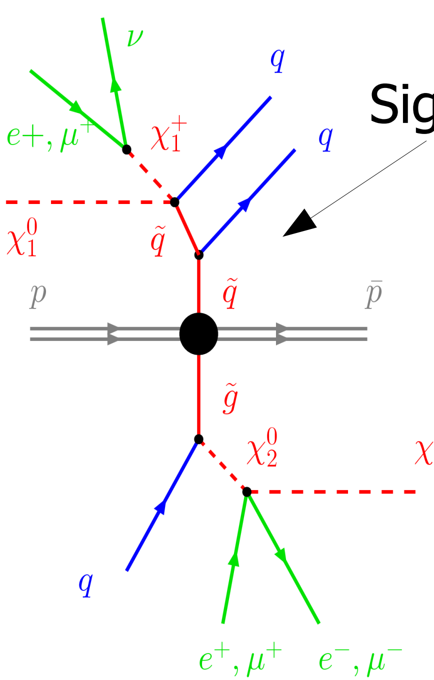
Real='real isolated'

SM di-lepton signatures

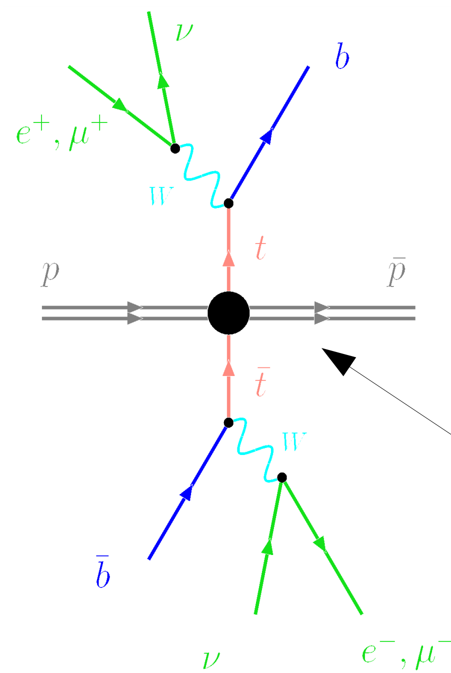
Process	+	-
Z→ll	Cross-section	No real MET
Di-boson	Real MET	Low cross-section
tt→ll	**Real MET,high mass**	
W→lν	Cross-section, real MET	No 2nd real lepton
QCD	Cross-section	No real MET/leptons
tt→l	Real MET,high mass	No 2nd real lepton

Fake leptons (W→lν, QCD, tt→l)

- Top pair production has similar signature
  - Largest background
  - Expectations from simulation
- Experience of analysis techniques from Tevatron valuable



Signal



Background

VETENSKAP OCH KONST KTH

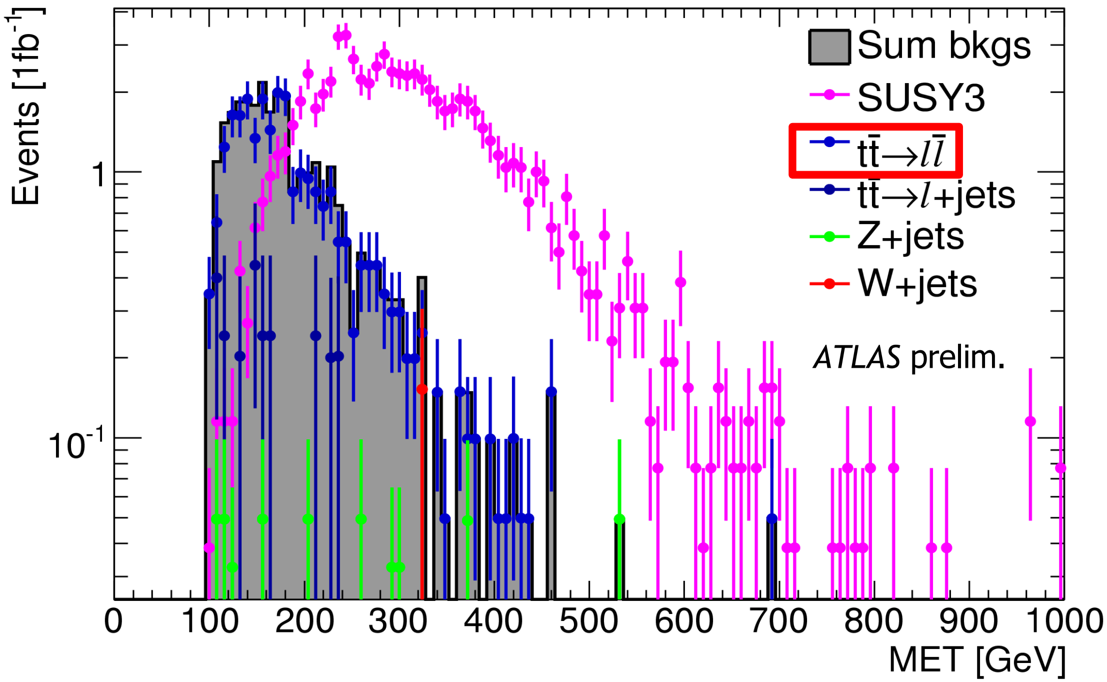
ATLAS preliminary results

Signal region selection

- Select events with high mass scale
  - Clean signature -> one dominating background in opposite lepton charge sign mode

"Standard SUSY" selection

lepton(s) OR jets+MET trigger
2 OS leptons (e or μ) pT>20 GeV
>3 jets pT>(100,50,50,50) GeV
Missing energy>max(100,0.2Meff)



Source	Events (%)
Top ll+jets	24
Top l+jets	2
W+jets	0
Z+jets	0
SUSY3	73
Total # events	109

Question

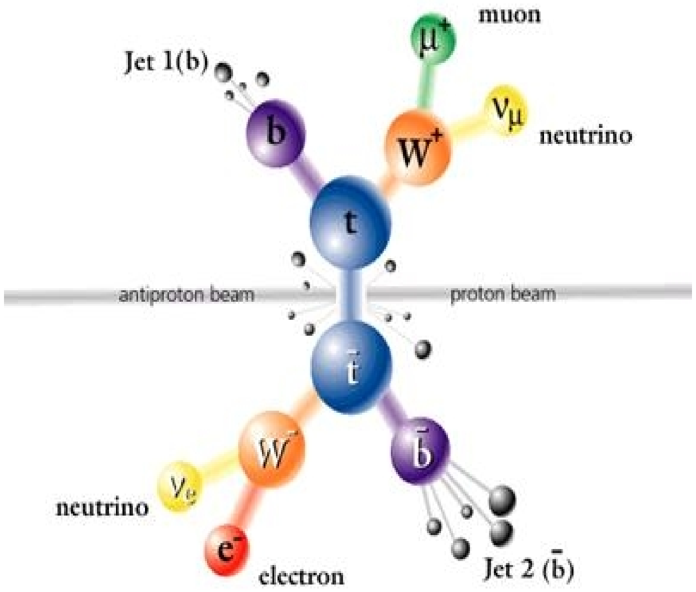
For a given integrated luminosity, how can we estimate the tt→ di-lepton background from data?

VETENSKAP OCH KONST KTH

tt clean control sample

- Strategy based on selecting a sample dominated by the background process

1. Apply "standard" SUSY selection
2. Form b-jet pairs
   - Consider all combination of 4 leading jets as input to the following equations



quartic equation
2,4-fold ambiguity
(sets of ν-momentum)

$$p^i_{\nu_1}, p^i_{\nu_2}$$

$$\begin{cases} m^2_{W_1} = \left(p_{l_1} + p_{\nu_1}\right)^2 \\ m^2_{W_2} = \left(p_{l_2} + p_{\nu_2}\right)^2 \\ m^2_{t_1} = \left(p_{l_1} + p_{\nu_1} + p_{b_1}\right)^2 \\ m^2_{t_2} = \left(p_{l_2} + p_{\nu_2} + p_{b_2}\right)^2 \\ MET_x = \left(p_{\nu_{1x}} + p_{\nu_{2x}}\right) \\ MET_y = \left(p_{\nu_{1y}} + p_{\nu_{2y}}\right) \end{cases}$$

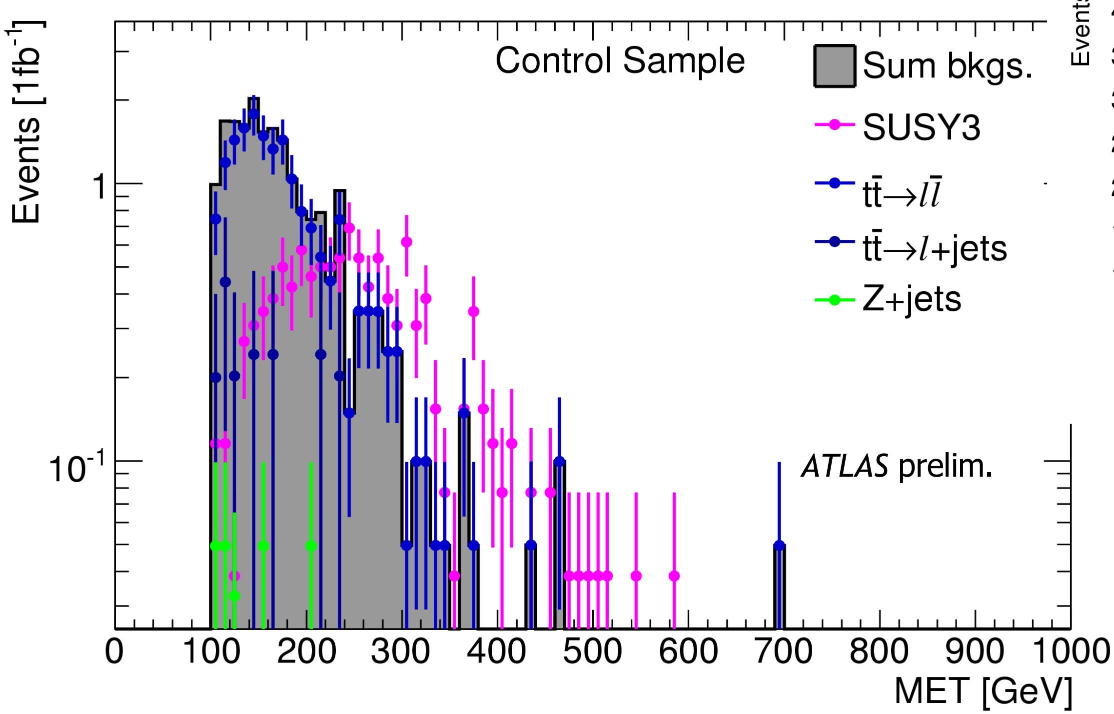
3. If it exist a pair of jets giving a real solution: consider event as a "tt like"
4. **Number of (b-)jet pairs** discriminates between tt and other processes

VETENSKAP OCH KONST KTH

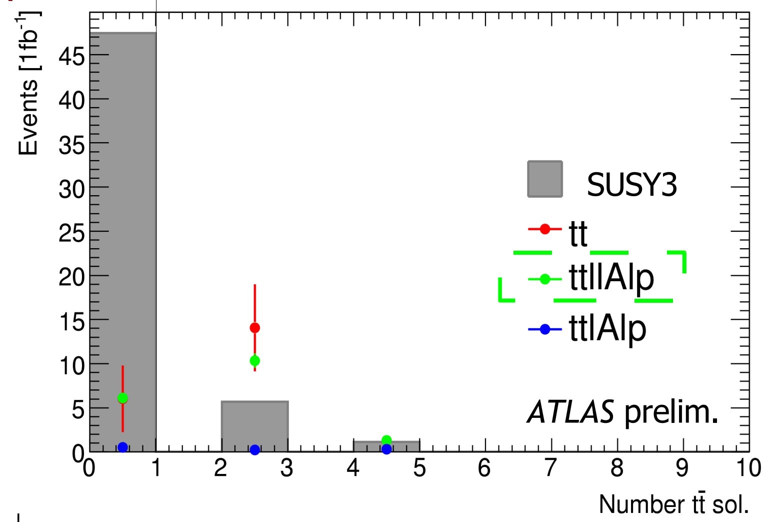
ATLAS preliminary results

tt control sample composition

- Control sample is dominated by tt→di-lepton events



> # b-jet pairs >0 (⇔#sol>0)



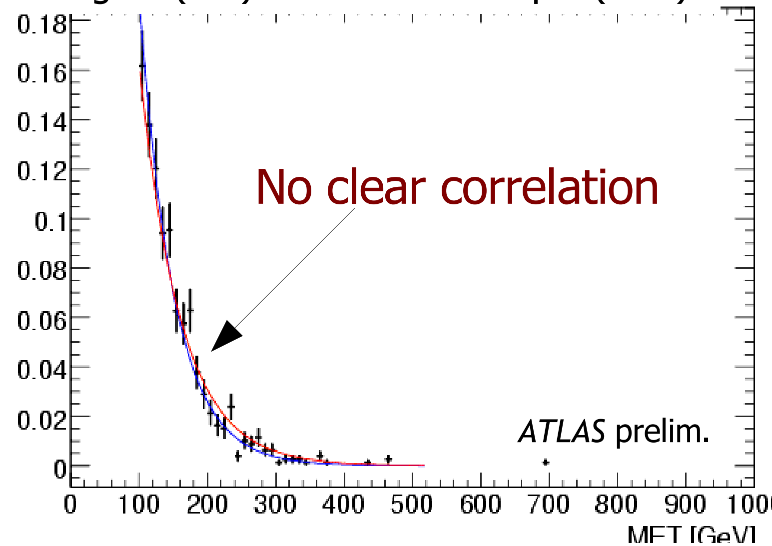
Source	Events (%)
Top ll+jets	58
Top l+jets	6
W+jets	0
Z+jets	1
SUSY3	36
Total # events	30.6

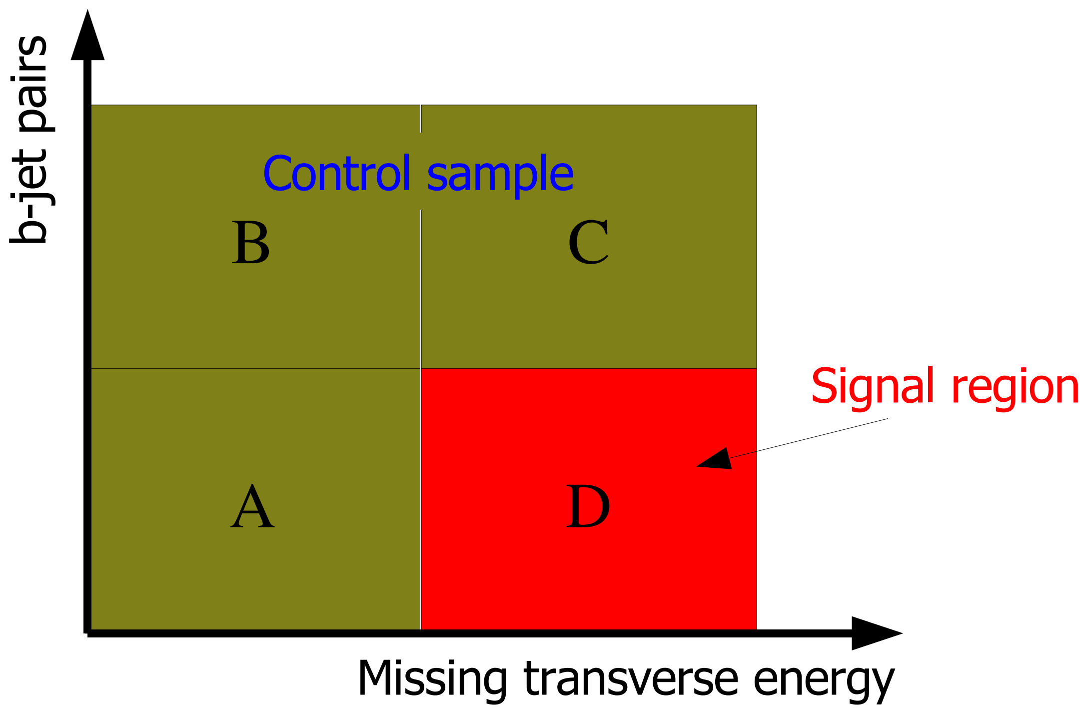
- Potentially large signal contamination
- tt dominated sample obtained from collider data
  - But how much is expected in the signal region?

tt normalization

- Normalize background prediction using uncorrelated variable
  - Here: missing transverse energy
  - Simple scaling into signal region

Comparison of MET in signal (red) and control sample (blue)





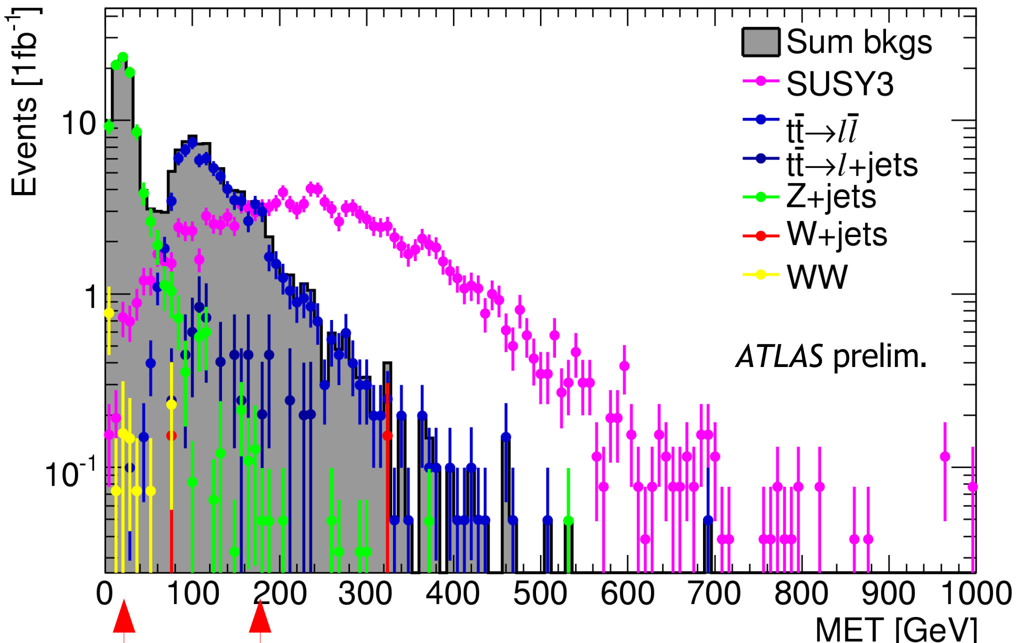
Normalization factor

$$D = C \times \frac{A}{B}$$

ATLAS preliminary results

tt normalization region

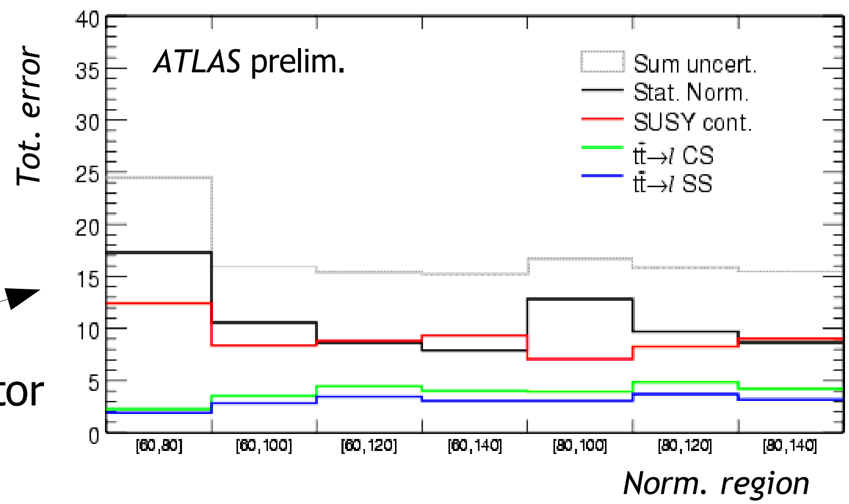
- Normalize background prediction in region with low expected signal contribution
  - Subtract other backgrounds



Norm. region

- Understanding low MET region
  - Z background heavily suppressed vs. signal contamination
  - Perhaps an iterative approach once data exist

Sys. uncertainty:
total error on the normalization factor



Note: some MC samples have gen. cuts

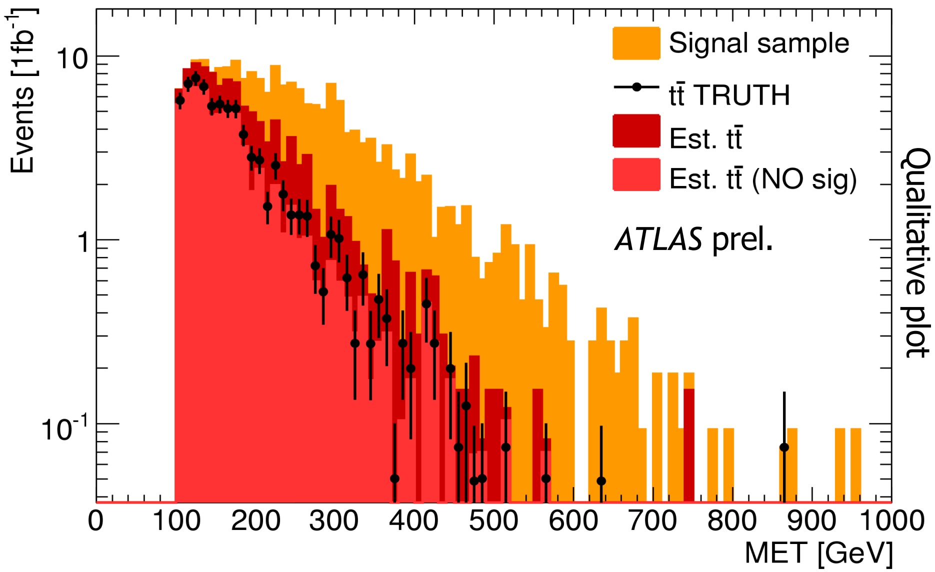
ATLAS preliminary results

tt estimation expected performance

- Estimate tt in “data cocktail” of events
  - Compare with MC truth

Process	Yield (%)
“true tt”	**100** (by def.)
Estimated tt	**90**
Including SUSY3	**122**

- Signal contamination
  - Clearly very model dependent
  - Inverse proportional to mass scale
  - Important -> **result is conservative**
- Systematics also input to searches
  - Result is expected uncertainty of tt
  - More “realistic” reach predictions



Sys. Uncertainty	(%)
Stat. Control Sample	20
Correlation of MET and # b-jet pairs	20
Semi-leptonic tt contribution	4
Jet Energy Scale	9
Jet Energy Resolution	1
MET Scale	1
Leptons via τ	1

Conclusion

- A method to estimate the tt → di-lepton background to SUSY-like signatures
  - Data-driven background estimations are crucial for discovery claims
  - Based on kinematic reconstruction of top pair events
  - Feasible after ~1fb^{-1} of collected data
  - Mass scale dependent signal contamination -> but method is conservative!
- Generic method for any BSM search
  - With same signature
- Part of a larger group-wide effort to provide methods/ideas for early searches
  - Important to have many different cross-checks for each source background
- Many systematic uncertainties are studied (tot. uncertainty ~30%)
  - Dominating source: statistical uncertainty of the control sample
  - Detector resolutions appears to have less effect

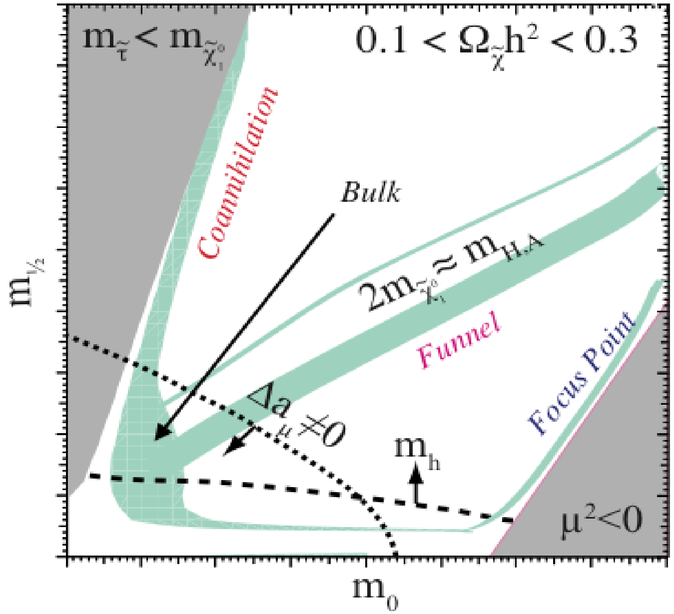
Backup slides

Benchmark Points mSUGRA

- Benchmark points in mSUGRA space
- Different annihilation processes leads to correct relic density

Point	m_0 (GeV)	$m_{1/2}$ (GeV)	A_0	$\tan(\beta)$	$\mathrm{sign}(\mu)$	σ_{LO} (pb)
SU1 – Coannihilation	70	350	0	10	+	6.8
SU2 – Focus	3550	300	0	10	+	4.9
SU3 – Bulk	100	300	-300	10	+	19.3
SU4 – Low Mass	200	160	-400	10	+	280
SU6 – Funnel	320	375	0	50	+	4.5

Qualitative plot – no mass scale



- Wide range of signatures
- Important for generic discovery potential

IsaJet 7.71

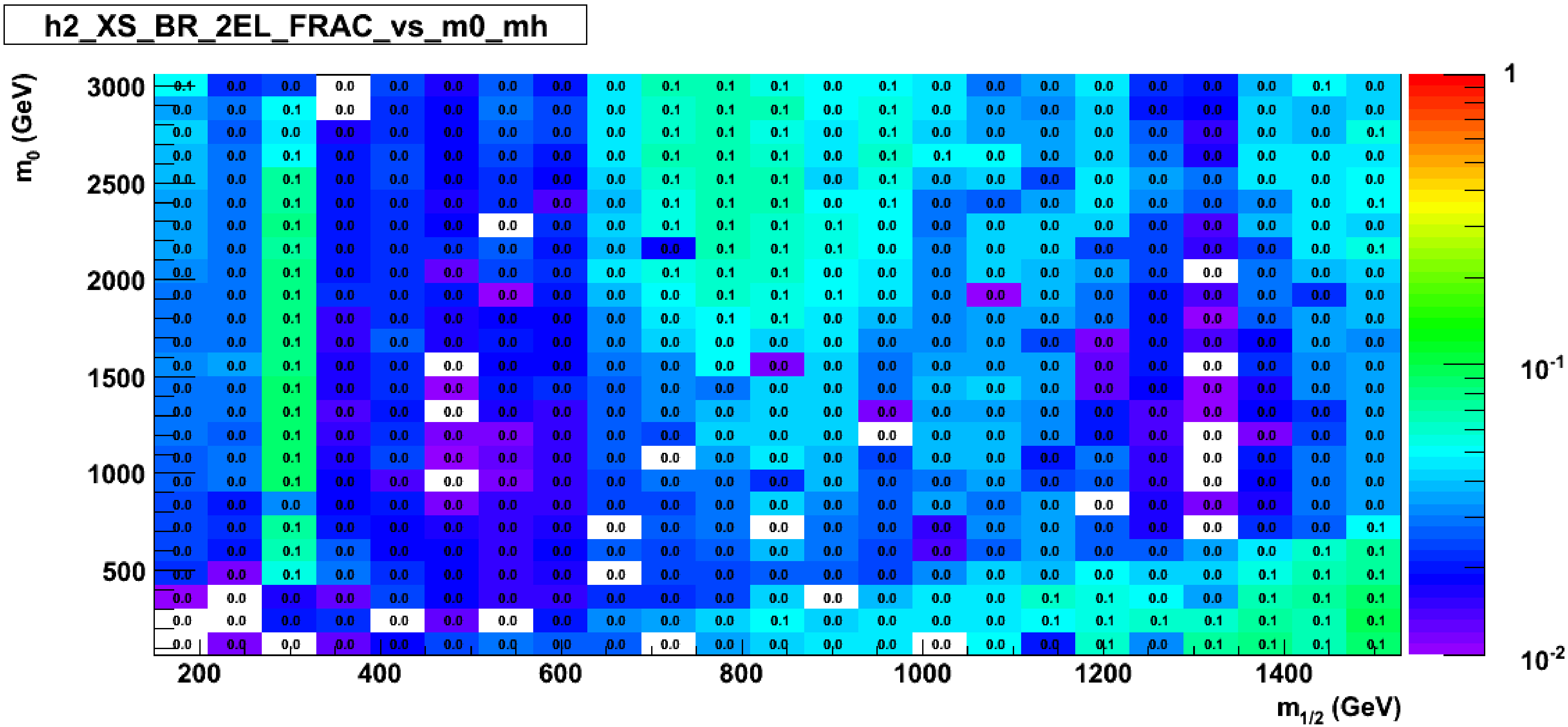
http://atlahttp://www.usatlas.bnl.gov/twiki/bin/view/Projects/AtlasSusyPoints

Systematics

- Uncertainties from Monte Carlo specific sources are not applicable (1^{st} order)
  - Ex. cross-sections cancels -> using data
  - From Monte Carlo
  - Sample composition from data (Z and QCD must be estimated from data)
- Correlation of MET and number of b-jet pairs
  - Fit the MET distributions and take nominal difference
- Detector performance related systematics
  - Nominal uncertainty is (very) luminosity dependent
  - Jet Energy Scale and resolution
  - Electron/Muon Scale and resolution
  - MET resolution

mSugra dilepton signal

- Overview of the SUSY di-electron production cross-section



$$\frac{\sigma\left(SUSY \rightarrow ee + X\right)}{\sigma\left(SUSY\right)}$$

2 electrons pT>X GeV
IsaJet 7.74

mSUGRA typical reach