Learning how to count

A high multiplicity search for the LHC

Sonia El Hedri

with

Anson Hook, Martin Jankowiak and Jay Wacker

November 12, 2013

arXiv:1302.1870v1

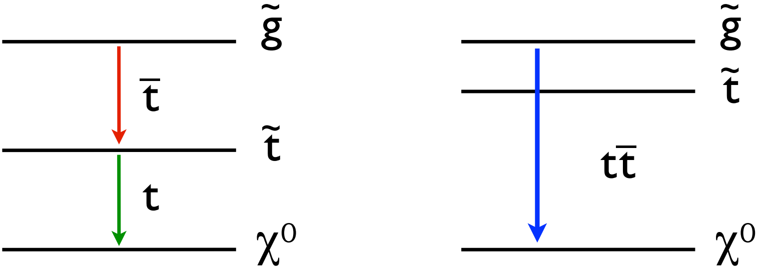
Traditional searches

Handles to look at new physics signals:

- Leptons
- Heavy flavor jets (b-tagging)
- Kinematic reconstruction (m_T, MT2, ...)
- Boosted jets: W or top tagging using jet substructure
- High p_T jets, radius $R = 0.4, 0.5$
- Missing E_T

One target: natural SUSY

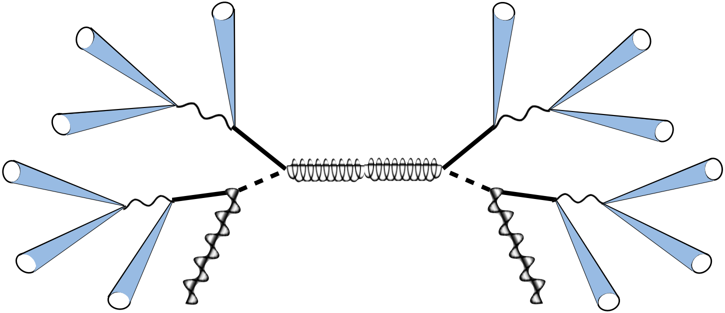
Decouple all particles not cancelling the top quadratic divergences



$$\tilde{g} \rightarrow t\bar{t}\chi^0$$

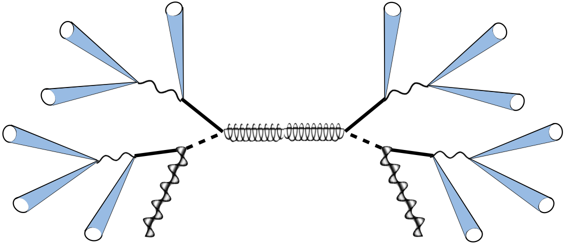
High multiplicity signals

>12 jet signals from natural SUSY



Other signals: RPV, strong dynamics, cascade decays, ...

High multiplicity signals



- Dominating if the light particles are hard to see
- Low production rate
- Signatures distributed across many channels
  - Exclusive searches are low efficiency
  - Inclusive searches are high background

Traditional approaches

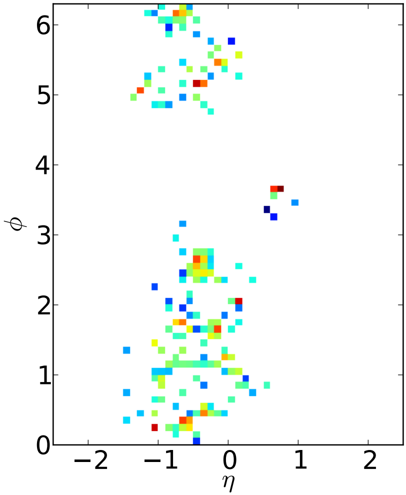
- Cluster thin jets, $R = 0.4 - 0.5$, $p_T > 50$ GeV
- Cut on the number of jets
- Cut on $\not{E}_T$

But

- Soft jets, $p_T \sim 50$ GeV
- Low $\not{E}_T$
- Discriminate hard structure from parton shower
- Complicated phase space (3^{N_j})
- High-multiplicity backgrounds hard to model

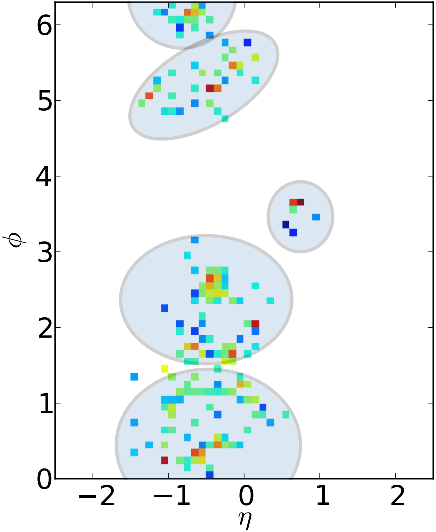
High multiplicity signals

Jets hard to resolve individually...



High multiplicity signals

...or accidental boost!

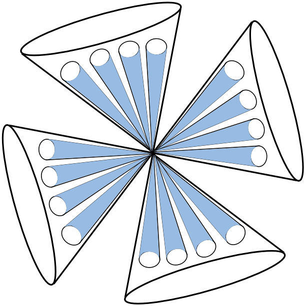


Using fat jets: an organizational principle

>12 low p_T thin jets $\Rightarrow$ four high p_T fat jets

- Lower phase space dimensionality
- Four hard objects, comparable p_T
- QCD fat jets weakly correlated

$\Rightarrow$ **Data-driven backgrounds**



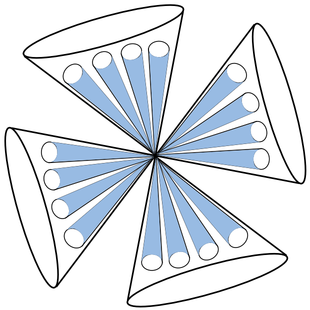
Using fat jets: an organizational principle

$>$12 low p_T thin jets $\Rightarrow$ four high p_T fat jets

- Lower phase space dimensionality
- Four hard objects, comparable p_T
- QCD fat jets weakly correlated

  $\Rightarrow$ **Data-driven backgrounds**
- Find new discriminating variables

  $\Rightarrow$ **Jet substructure techniques**

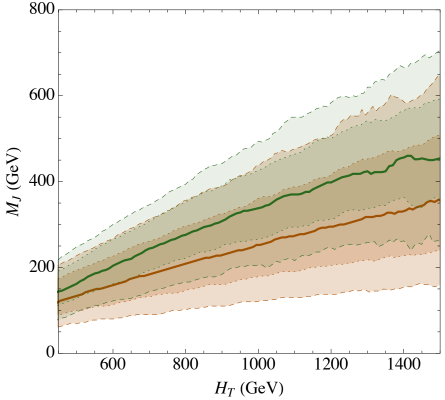


New observables: Jet mass

$$M_J = \sum_i m_i$$

$$H_t = \frac{\sqrt{1 + (\kappa R)^2}}{\kappa R} M_J$$

QCD $(\kappa = \sqrt{\alpha_s})$, $t\bar{t}$ $(\kappa = 1)$

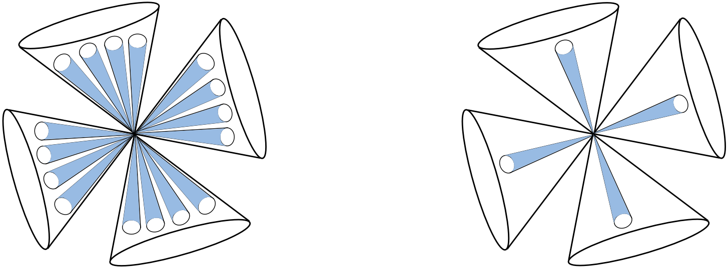


A. Hook et. al., arXiv:1202.0558v3

New observables: subjet counting

"Count" the number of subjets using jet substructure techniques

Signal Background



$$N = \sum_i N_i^{subjets}$$

Data-driven background estimates

- From low multiplicity to high multiplicity
- Evaluate 4-jet QCD backgrounds using 2-jet samples
- Model each jet/MET using templates

$$\rho_{\rm jet}(m_{\rm jet}, n_{\rm jet}, p_T)$$

- Combine templates, account for jet correlations

$$\sigma(p_{Ti}, M_J, N_J, \not{E}_T) = \sigma_{4J}(p_{Ti}) \otimes \mathcal{P}(\not{E}_T) \otimes \rho_1(m_1, n_1, p_{T1}) \otimes \dots$$

Preliminary analysis

Reduce the dimensionality – Assumptions

- Quark-gluon ratio similar in all jets
- Jet properties independent on the environment

$$m_i, n_i, p_{Ti} \Rightarrow \rho_J \left(\frac{m_i}{p_{Ti}}, n_i; p_{Ti} \right)$$

- Jets independent from each other

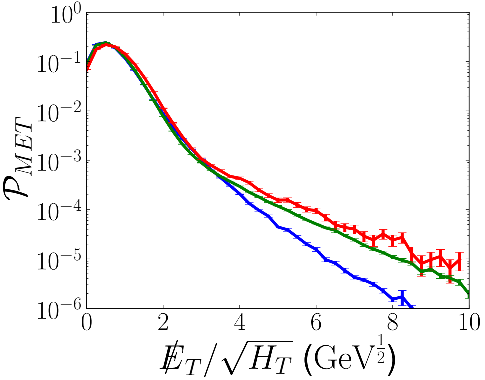
$$\sigma \left(\not{E}_T, M_J, N_J \right) = \sigma_{4J}(\not{E}_T, p_{Ti}) \prod_i \rho_J \left(x_i, n_i, p_{Ti} \right)$$

- $\not{E}_T$ depends only on H_t

$$\sigma \left(\not{E}_T, M_J, N_J \right) = \sigma_{4J}(p_{Ti}) \, \mathcal{P} \left(\not{E}_T, H_t \right) \prod_i \rho_J \left(x_i, n_i, p_{Ti} \right)$$

Missing E_T templates

- Detector smearing effects
- Scales as $\sqrt{H_t}$
- Orthogonal to jet substructure properties

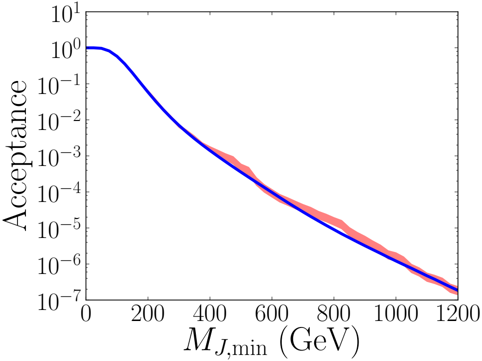


$$\mathcal{P}_{MET} = \mathcal{P}\left(\frac{\not{E}_T}{\sqrt{H_t}}, H_t\right)$$

Results

Preliminary: find $\rho_J(x, n; p_T)$ using 4-jet Monte-Carlo samples

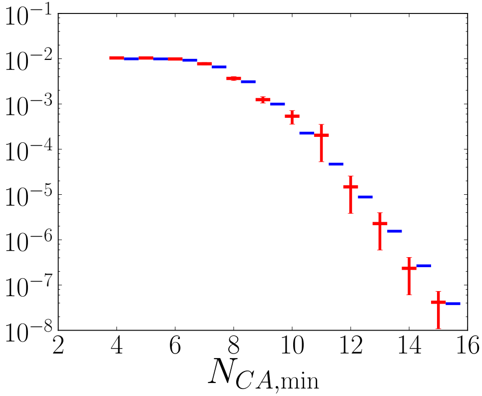
- SHERPA sample: 4.8 million weighted events
- H_t binning for $\not{E}_T$ templates
- Construct $3D$ binned $\rho_J(x, n; p_T)$
- M_J distribution for $\not{E}_T > 150$ GeV



Results

Preliminary: find $\rho_J(x, n; p_T)$ using 4-jet Monte-Carlo samples

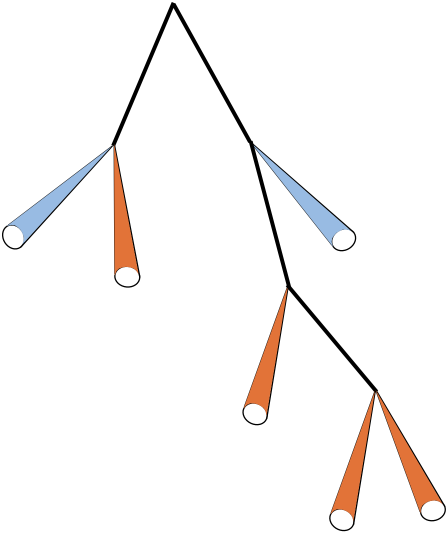
- SHERPA sample: 4.8 million weighted events
- H_t binning for $\not{E}_T$ templates
- Construct $3D$ binned $\rho_J(x, n; p_T)$
- N_J distribution for
  $\not{E}_T > 150$ GeV
  $M_J > 200$ GeV



Outlook

- Good agreement between reconstructed and real distributions
- Need more statistics/data
- Take quark-gluon content into account
- Dependance of ρ to the environment
  $\Rightarrow$ Less than 10%
- Take jet correlations into account, pile up effects
- Test data-driven methods on other topologies (γ+jets, etc...)
- More elaborated analysis: T. Cohen, M. Lisanti, T. Lou

Counting with the Cambridge-Aachen algorithm

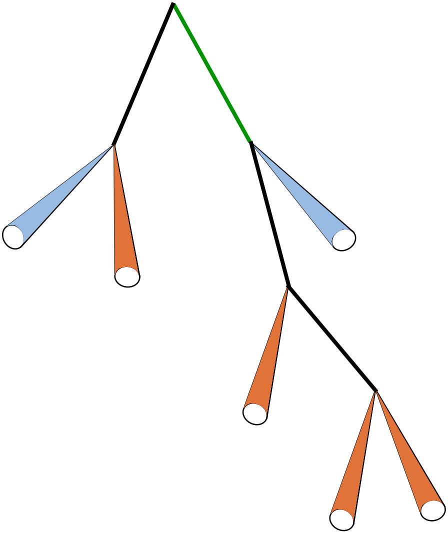


$$d_{ij} = \Delta R_{ij}^2 \qquad (1)$$

Cluster the jet with CA and go down the clustering tree

- Uncluster j into j_1 and j_2

Counting with the Cambridge-Aachen algorithm

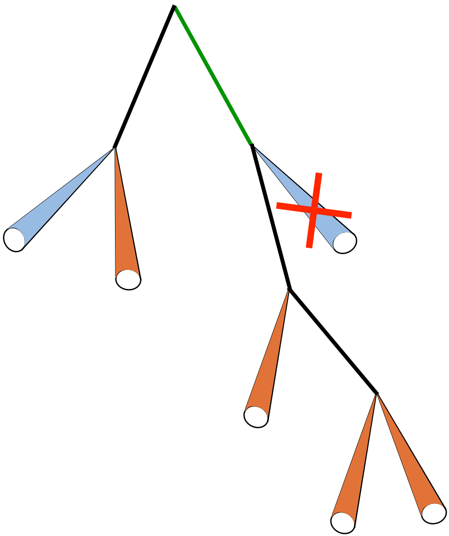


$$d_{ij} = \Delta R_{ij}^2 \qquad (2)$$

Cluster the jet with CA and go down the clustering tree

- Uncluster j into j_1 and j_2
- If p_Ts are imbalanced, remove soft jet

Counting with the Cambridge-Aachen algorithm

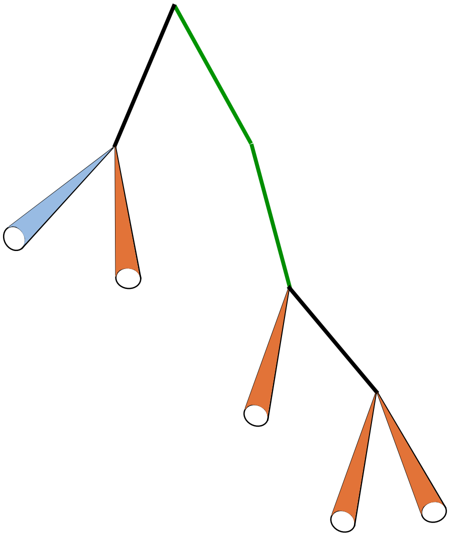


$$d_{ij} = \Delta R_{ij}^2 \qquad (3)$$

Cluster the jet with CA and go down the clustering tree

- Uncluster j into j_1 and j_2
- If p_Ts are imbalanced, remove soft jet
- If $m_j < m_{cut}$ or $d_{12} < R_{min}$, j is a subjet
- Keep subjets with $p_T > p_{Tcut}$

Counting with the Cambridge-Aachen algorithm

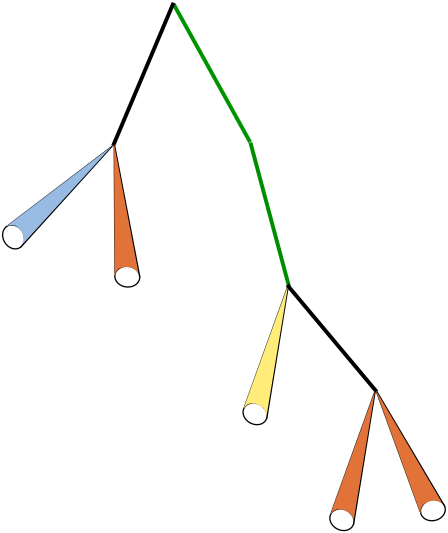


$$d_{ij} = \Delta R_{ij}^2 \qquad (4)$$

Cluster the jet with CA and go down the clustering tree

- Uncluster j into j_1 and j_2
- If p_Ts are imbalanced, remove soft jet
- If $m_j < m_{cut}$ or $d_{12} < R_{min}$, j is a subjet
- Keep subjets with $p_T > p_{Tcut}$

Counting with the Cambridge-Aachen algorithm

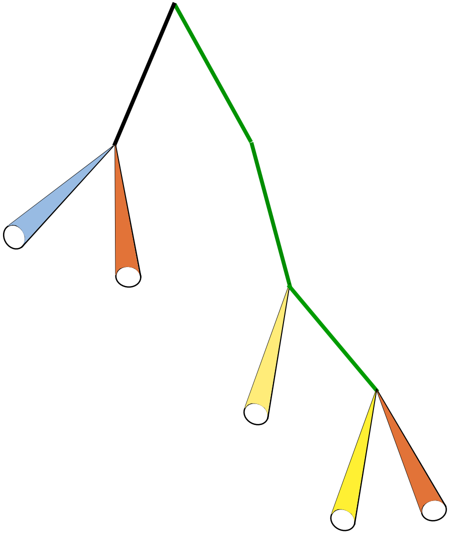


$$d_{ij} = \Delta R_{ij}^2 \qquad (5)$$

Cluster the jet with CA and go down the clustering tree

- Uncluster j into j_1 and j_2
- If p_Ts are imbalanced, remove soft jet
- If $m_j < m_{cut}$ or $d_{12} < R_{min}$, j is a subjet
- Keep subjets with $p_T > p_{Tcut}$

Counting with the Cambridge-Aachen algorithm



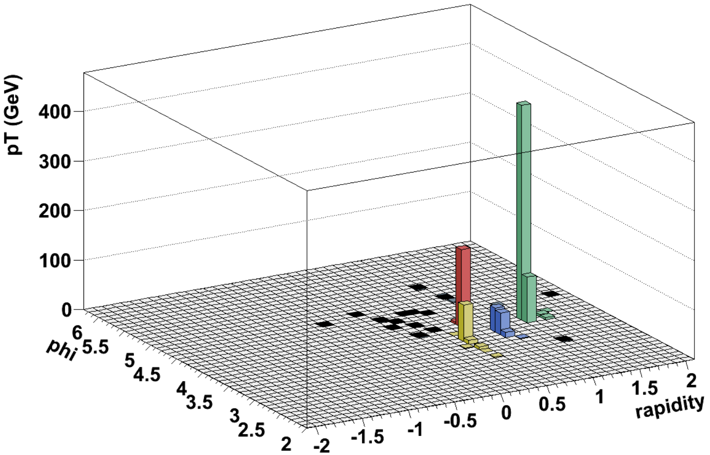
$$d_{ij} = \Delta R_{ij}^2 \qquad (6)$$

Cluster the jet with CA and go down the clustering tree

- Uncluster j into j_1 and j_2
- If p_Ts are imbalanced, remove soft jet
- If $m_j < m_{cut}$ or $d_{12} < R_{min}$, j is a subjet
- Keep subjets with $p_T > p_{Tcut}$

Counting with CA

- Subjets consistent with the decay of a massive particle
- Soft radiation discarded
- $m_{cut} = 30$ GeV, $y_{cut} = 0.10$, $R_{min} = 0.15$, $p_{Tcut} = 30$ GeV

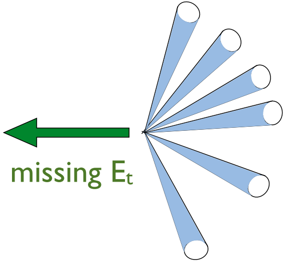


Fastjet plugin

**fastjet.hepforge.org/trac/browser/contrib/
contribs#SubjetCounting**

ATLAS high multiplicity search

ATLAS-CONF-2012-103



- 8 TeV, 5.8 fb^{-1}
- Anti-k_t algorithm with $R = 0.4$
- 7, 8 or 9 jets with $p_T > 55$ GeV
- 6, 7 or 8 jets with $p_T > 80$ GeV
- $\frac{\not{E}_t}{\sqrt{H_t}} > 4\,\mathrm{GeV}^{1/2}$

Searches

$N_{\rm jets}$ cut $+$ $\not\!\!E_T$ cut (ATLAS)

vs

M_J cut $+$ $\not\!\!E_T$ cut

vs

M_J cut $+$ $\not\!\!E_T$ cut $+$ $N_{\rm subjets}$ cut

Benchmark models

		Tops jets?	Cascade decay?	RPV?
$\tilde{g}$	$\rightarrow$	$t\bar{t}\chi_i^0$	$\chi_i^0 \rightarrow VV\chi_1^0$	$\chi_1^0 \rightarrow jjj$
		+12 jets	+8 jets	+6 jets

- 8 possible topologies
- from 4 to 26 jets
- signals with and without $\not{E}_T$

Benchmark models and searches

Optimal cuts depend on :

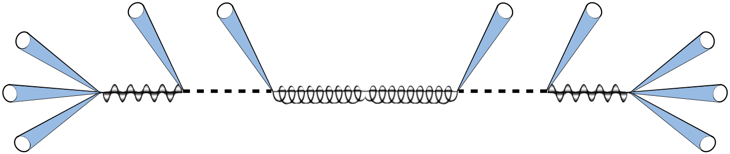
- Jet multiplicity
- $\not{E}_T$
- Presence of leptons
- Mass of the initial particle $m_{\tilde{g}}$

Inclusive search:

- Leptons clustered in jets (no lepton cuts)
- Find **minimal** number of cuts on $M_J + \not{E}_T + \ldots$ so that the bounds are close to optimal
  - For each signal
  - For each mass

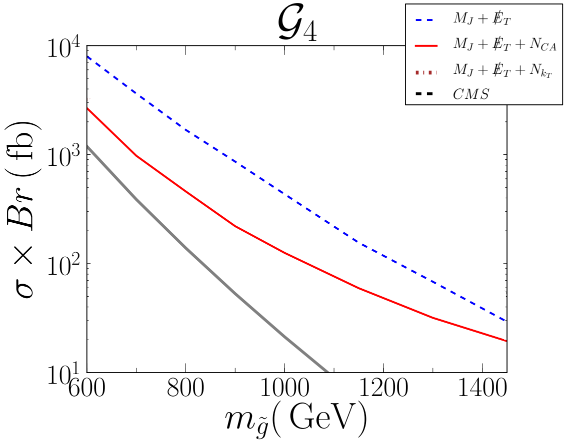
Gluino decay to light quarks, RPV

$\tilde{g} \rightarrow jj\chi_1^0$, $\chi_1^0 \rightarrow jjj$



10 jets, no $\not{E}_T$

Gluino decay to light quarks, RPV – 8 TeV, 30 fb^{-1}



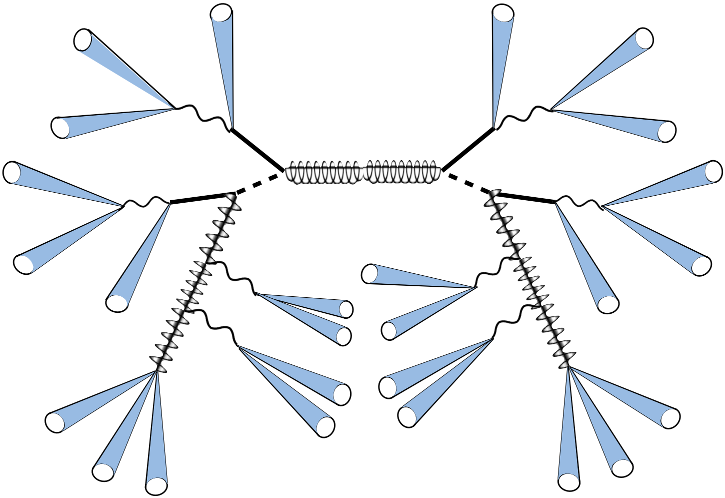
$M_J \geq 1.3$ TeV

$M_J \geq 1$ TeV
$N_{CA} \geq 13$

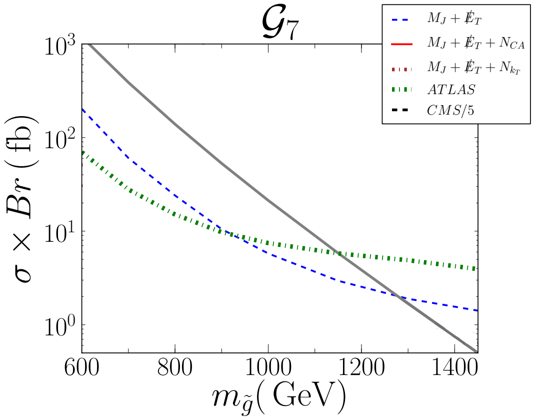
- Factor of 2 to 4 improvement over $M_J + \not{E}_T$
- M_J cut loosened

Gluino 2 step decay, RPV

$\tilde{g} \to t\bar{t}\chi_2^0,\ \chi_2^0 \to VV'\chi_1^0,\ \chi_1^0 \to jjj$

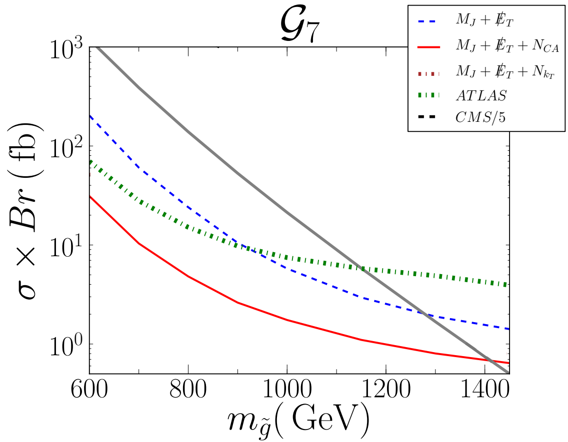


Gluino 2 step decay, RPV – 8 TeV, 30 fb^{-1}



- $M_J > 725$ GeV, $\not{E}_T > 175$ GeV
- $M_J + \not{E}_T$ search better at high mass

Gluino 2 step decay, RPV – 8 TeV, 30 fb^{-1}



$M_J \geq 425$ GeV
$\not{E}_T \geq 125$ GeV
$N_{CA} \geq 14$

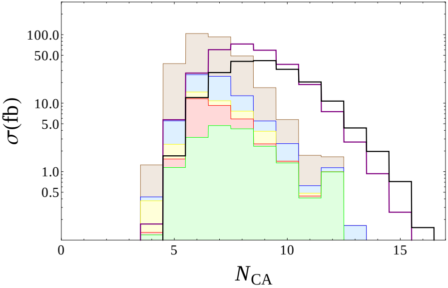
- Factor of ~ 4 improvement over M_J + MET
- Factor of ~ 5 improvement over ATLAS at high mass
- M_J and $\not{E}_T$ cuts significantly looser

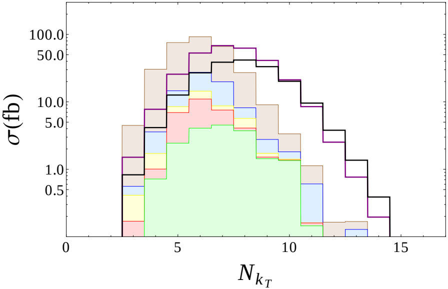
Summary

- Common new physics scenarios predict events with very high multiplicity
- Standard handles not appropriate (not boosted, complicated kinematics, low energy)
- QCD high multiplicity backgrounds hard to model
- Fat jet techniques: new organizational principle, but requires finding new variables
- Lower dimensionality makes data-driven estimates of QCD background easier
- Counting subjets in an event provides good discriminating power
- M_J and $\not{E}_T$ cuts loosened, could be used to probe $\not{E}_T$-less signals

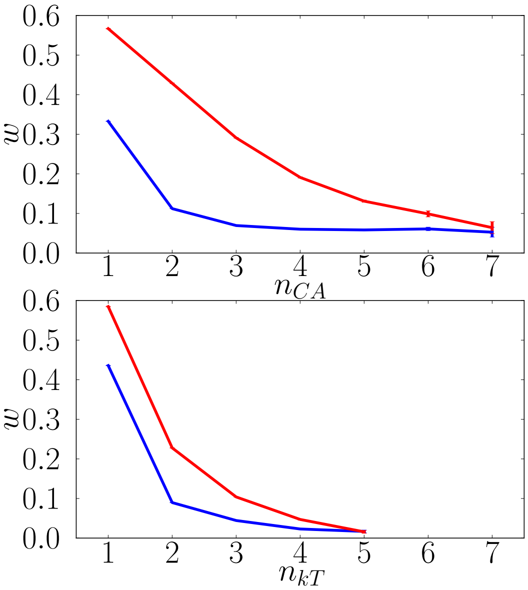
Backup

Signal and background distributions

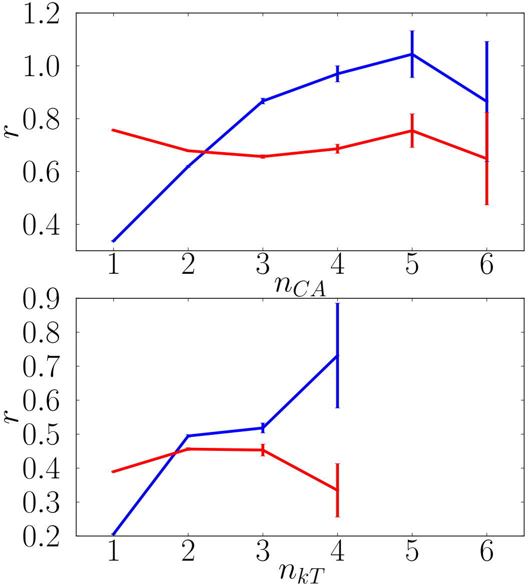




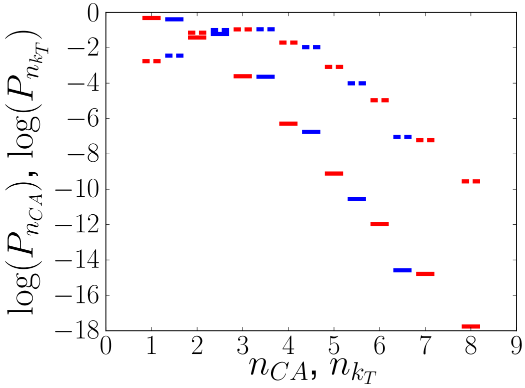
Scaling patterns



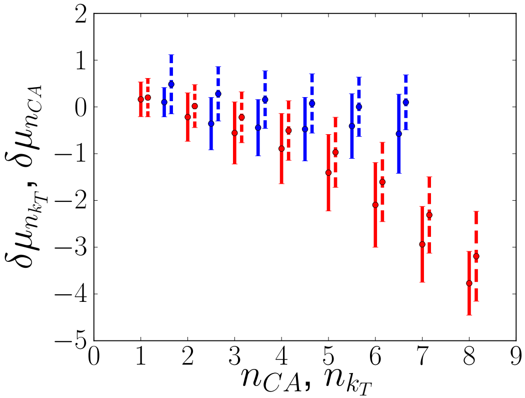
Scaling patterns



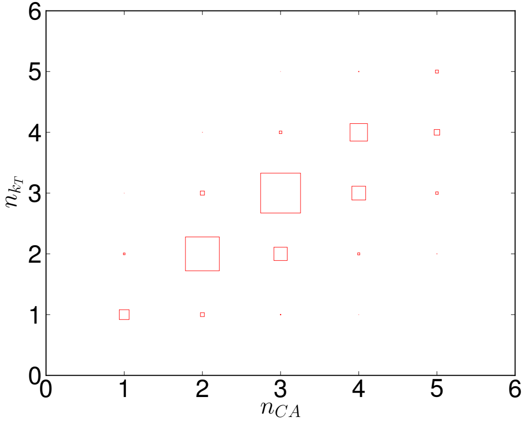
Correlations between N_{CA} and N_{k_T}



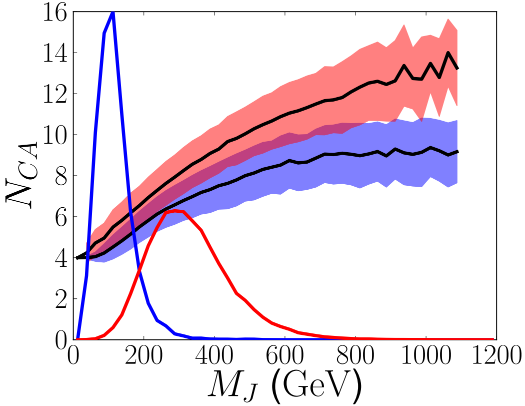
Correlations between N_{CA} and N_{k_T}



Correlations between N_{CA} and N_{k_T}



N_{CA} vs M_J



N_{CA} vs M_J