

SUSY searches with b -jets in CMS

Joshua Thompson

On behalf of the CMS Collaboration

Cornell University

Berkeley Workshop on SUSY at the LHC

20 Oct 2011

Introduction and motivation

- CMS emphasizes multiple, complementary SUSY searches
 - Want to confirm any discovery in multiple analyses with different approaches and systematics
 - In inclusive hadronic, have different kinematic approaches:
 - α_T , MHT (MET), MT2, Razor
 - b-tagging adds another dimension to this
 - Different proportions of SM backgrounds
 - Suppressed background generally allow for looser selections than inclusive searches
 - For b-rich models, better S/B
- This talk focuses on 2 analyses (1.1 fb^{-1} of data)
 - MT2+b and MET+b

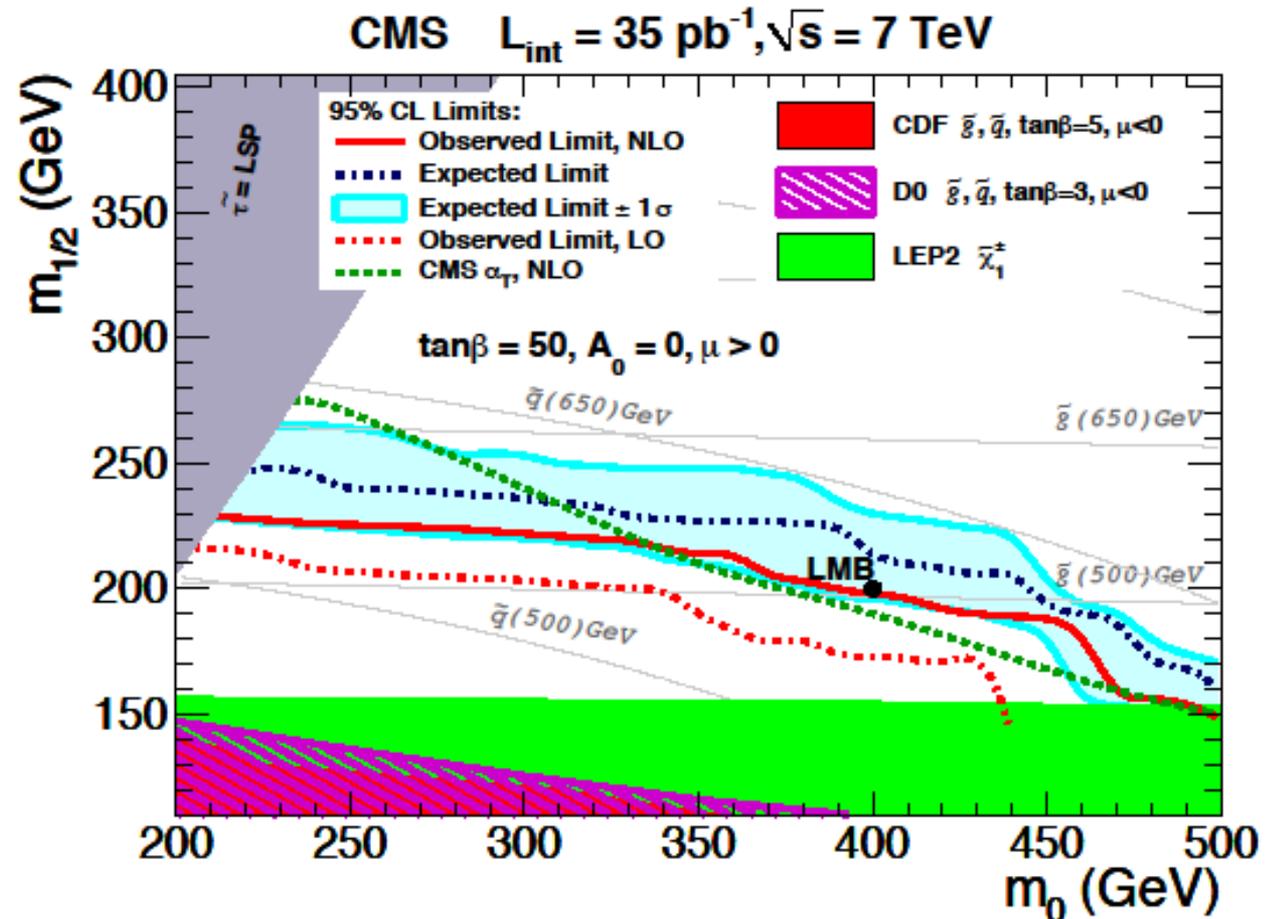
Overview of signals and backgrounds

- Example signals
 - CMSSM
 - High $\tan \beta$, high m_0 , low m_{12}
 - Light gluinos and heavy squarks
 - Production: gluino-gluino
 - Decay: gluino \rightarrow qqX~
 - 3rd gen are lightest squarks \rightarrow q=b,t favored
 - Simplified models:
 - e.g. $g\tilde{g} \rightarrow bbX\tilde{b}$
- Backgrounds (after b tag)
 - ttbar
 - $t \rightarrow bW \rightarrow b\ell\nu$ gives both b jets and real MET
 - Lost lepton or $\tau \rightarrow$ hadrons survive lepton veto
 - Smaller contributions, reduced by b tag:
 - QCD
 - Z+Jets, with $Z \rightarrow \nu\nu$
 - W+Jets

NB: searches not, in general, optimized for a particular model. Stay inclusive!

Setting the stage: 2010 results

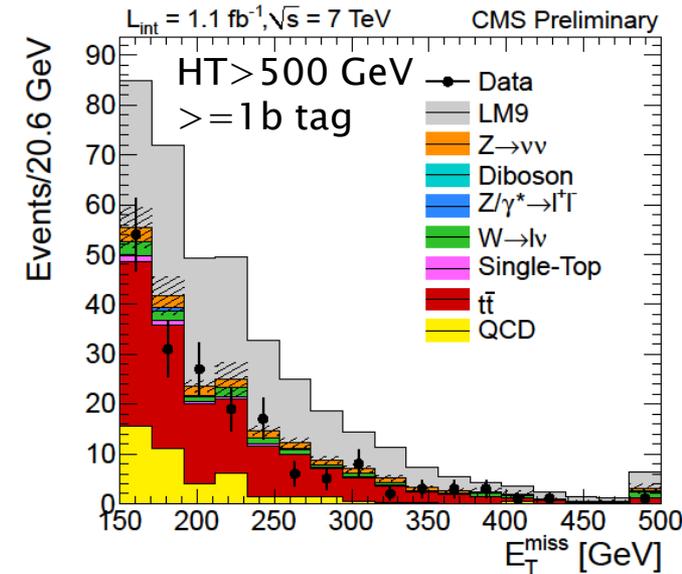
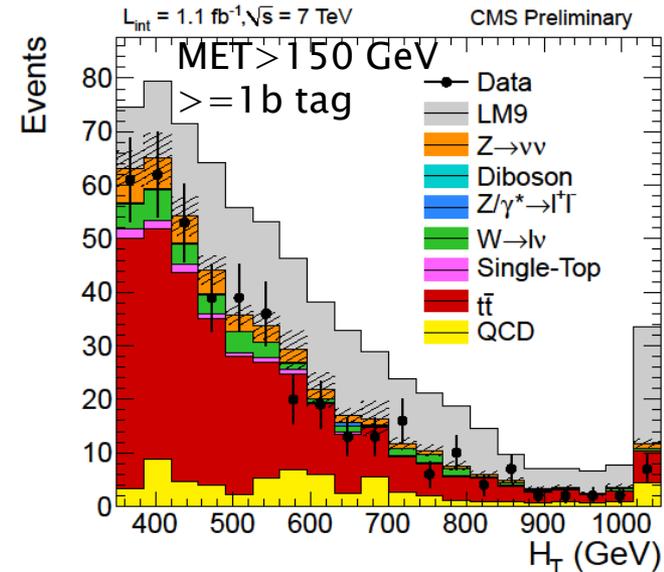
- b-tagged extension of inclusive hadronic search using the α_T kinematic variable
- Improves on the untagged result at high m_0



Basic event selection

- Large hadronic activity
 - Multiple hard jets
 - ≥ 4 for MT2 analysis
 - ≥ 3 for MET analysis
 - Large $H_T = \sum_{\text{jets}} |p_T|$
 - > 650 GeV for MT2 analysis
 - > 350 (500) GeV for Loose (Tight) branch of MET analysis
- Veto isolated leptons (e, μ)
- Veto events with small $\Delta\phi(\text{jet}, \text{MET})$
- Large missing energy
 - Either directly as MET or recast as MT2
 - MT2 analysis: $MT2 > 150$ GeV
 - MET analysis: $\text{MET} > 200$ (300) GeV for Loose (Tight)

What is MT2? $(M_{T2})^2 = 2A_T = 2p_T^{\text{vis}(1)} p_T^{\text{vis}(2)} (1 + \cos\phi_{12})$
An extension of the transverse mass concept (commonly used for $W \rightarrow l\nu$ decays) to decay chains with 2 unobserved particles.
 \rightarrow Largely correlated with MET, but gives better rejection of backgrounds

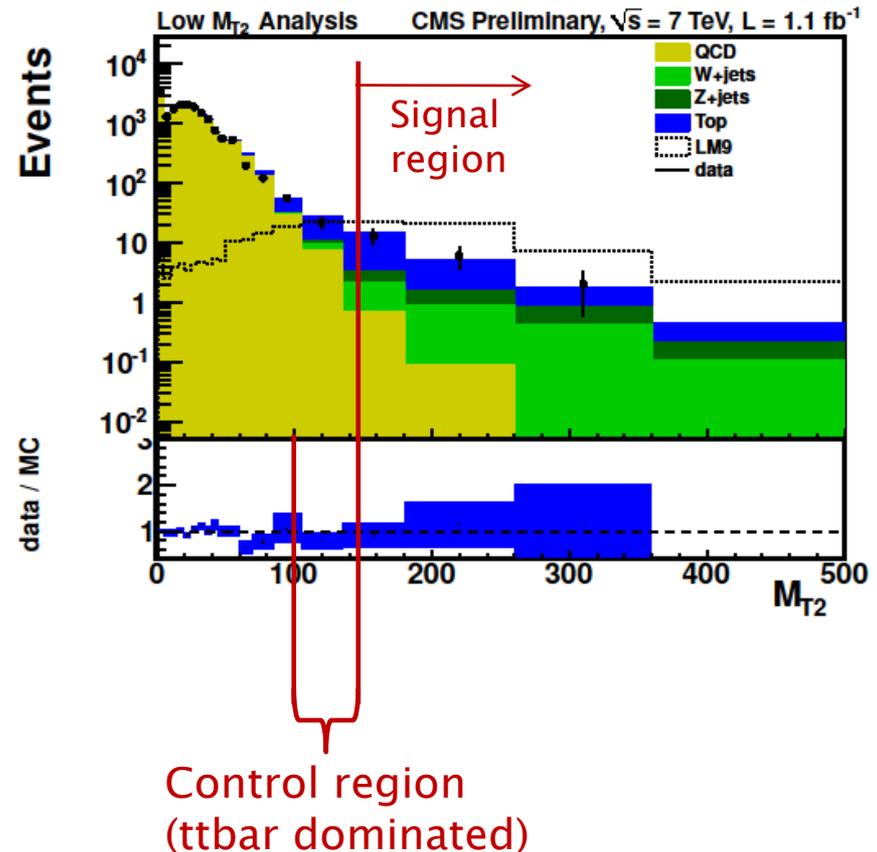


b tagging

- Both of these analyses use the Simple Secondary Vertex High Purity algorithm
 - Find a secondary vertex in a jet with at least 3 tracks
 - Make a tight selection on the discriminator value with ~50% efficiency and ~0.1% mistag
- For signal efficiency evaluation, use data-driven scale factors to correct MC b-tag efficiency
 - Up to p_T of 240 GeV, these are centrally provided by the CMS b-tag group
 - From 240–350 GeV, the MET+b analysis performed an evaluation using the ratio of double b-tagged events to single b-tagged events using a 1 lepton (~ttbar) control sample
 - Found scale factors to be the same, with a larger uncertainty
 - Above 350 GeV, MET+b analysis uses a scale factor of 0 for signal efficiency (conservative for a limit)
 - Not enough statistics (yet) for a proper evaluation of the scale factor in data
- Both analyses use ≥ 1 b tag selections
 - MET+b also uses selections with ≥ 2 b tags

MT2 plus b-tag

- b-tagging allows for looser MT2 selection
 - $MT2 > 150$ GeV (400 GeV in inclusive analysis)
 - Looser cut enhances sensitivity to some models
 - e.g. CMSSM test point LM9 has relatively soft missing energy distribution
- ttbar background estimate:
 - Use ttbar-dominated sample with 1 electron or 1 muon
 - Use MC efficiency numbers to move from 1 lepton \rightarrow 0 lepton sample
 - Perform this method in control region $100 < MT2 < 150$ GeV
 - Compare prediction for 0 lepton sample to MC for 0 lepton sample; level of agreement quantified in the uncertainty
 - Scale from control region to signal region using MC, propagating uncertainties
- Result
 - Background = $10.6 \pm 1.9 \pm 4.8$ events
 - Observed = 19 events
 - (LM9 signal = 42.9 events)

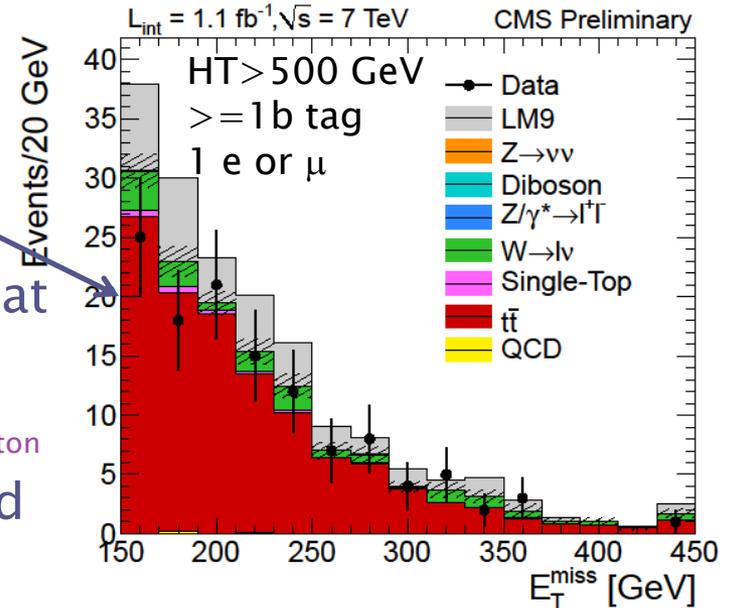


MET+b tag: background methods

All methods data-driven!

▫ $t\bar{t} + W + t$

- Find MET shape in 1 lepton control sample
- Normalize to $t\bar{t}$ -dominated region at medium MET ($150 < \text{MET} < 200$ GeV)
 - $(N_{\text{high MET}})^{0 \text{ lepton}} = (N_{\text{medium MET}})^{0 \text{ lepton}} (N_{\text{high MET}} / N_{\text{medium MET}})^{1 \text{ lepton}}$
- Cross-check with independent method



MET+b tag: background methods

All methods data-driven!

ttbar+W+t

- Find MET shape in 1 lepton control sample
- Normalize to ttbar-dominated region at medium MET (150 < MET < 200 GeV)

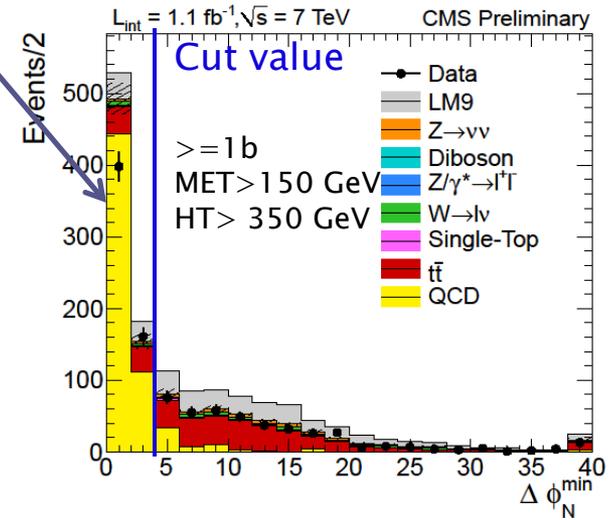
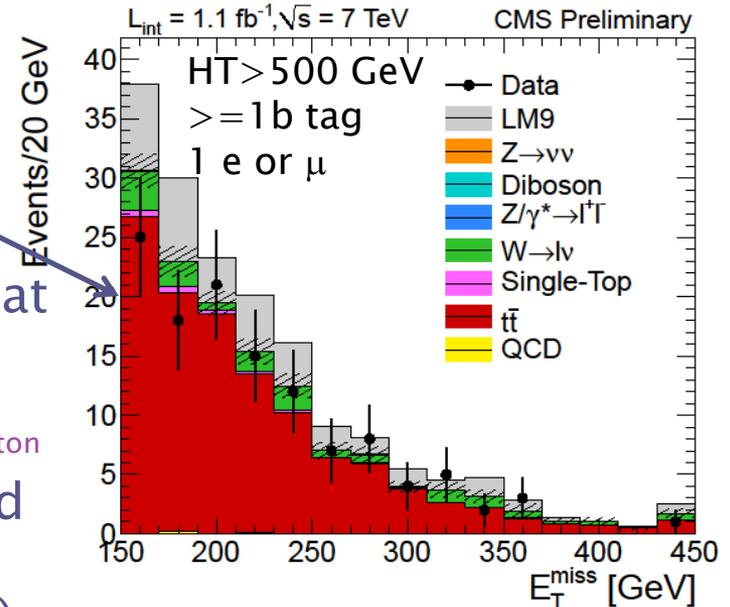
$$(N_{\text{high MET}})^{0 \text{ lepton}} = (N_{\text{medium MET}})^{0 \text{ lepton}} (N_{\text{high MET}}/N_{\text{medium MET}})^{1 \text{ lepton}}$$

- Cross-check with independent method

QCD

- Novel resolution-normalized $\Delta\phi(j, \text{MET})$ variable and MET are uncorrelated

$$(N_{\text{pass}})^{\text{high MET}} = (N_{\text{fail}})^{\text{high MET}} (N_{\text{pass}}/N_{\text{fail}})^{\text{low MET}}$$



MET+b tag: background methods

All methods data-driven!

ttbar+W+t

- Find MET shape in 1 lepton control sample
- Normalize to ttbar-dominated region at medium MET (150 < MET < 200 GeV)

$$\left(\frac{N_{\text{high MET}}}{N_{\text{medium MET}}} \right)^{0 \text{ lepton}} = \left(\frac{N_{\text{high MET}}}{N_{\text{medium MET}}} \right)^{1 \text{ lepton}}$$

- Cross-check with independent method

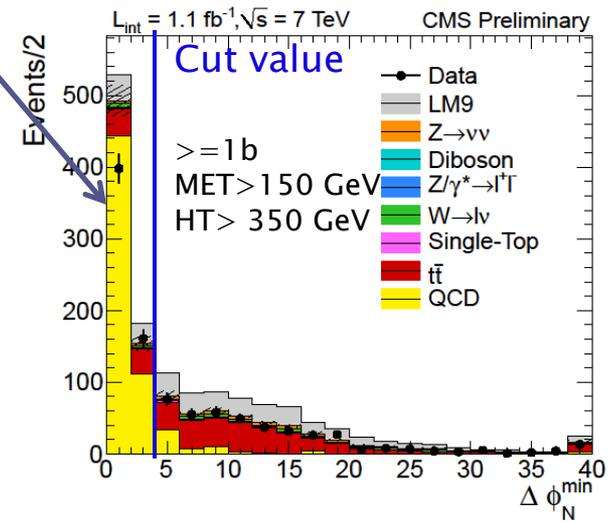
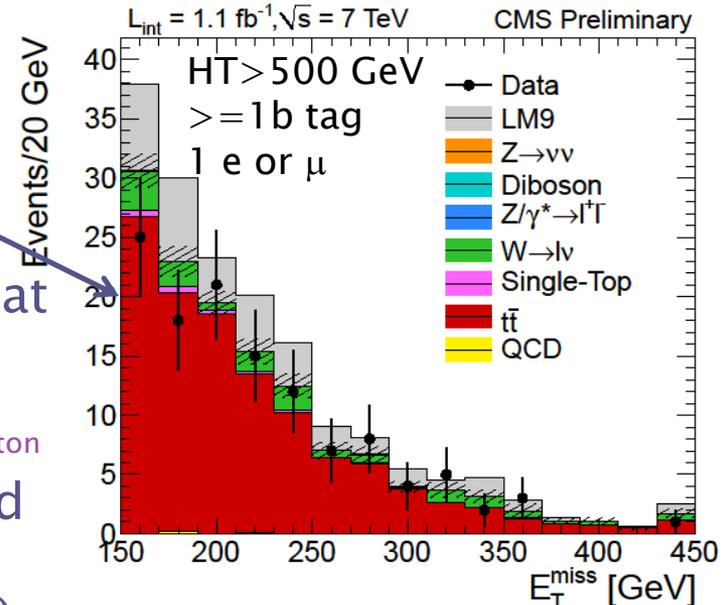
QCD

- Novel resolution-normalized $\Delta\phi(j, \text{MET})$ variable and MET are uncorrelated

$$\left(\frac{N_{\text{pass}}}{N_{\text{fail}}} \right)^{\text{high MET}} = \left(\frac{N_{\text{pass}}}{N_{\text{fail}}} \right)^{\text{low MET}}$$

Z → νν

- Data-driven translation of Z → ll control samples



MET+b: results

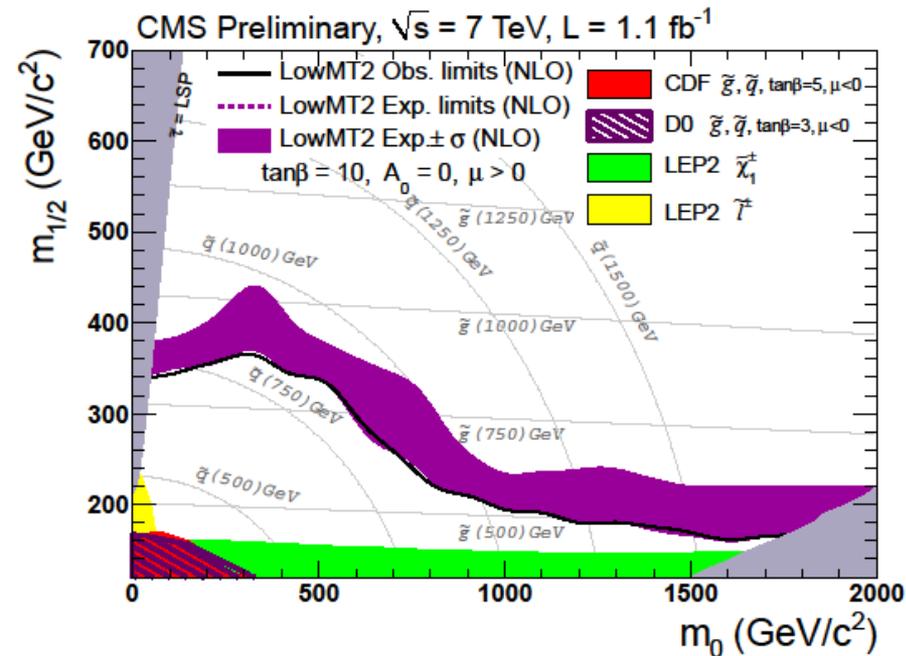
- Background predictions agree with data

	<u>"$\geq 2b$ Loose"</u> HT > 350 GeV MET > 200 GeV	<u>"$\geq 1b$ Tight"</u> HT > 500 GeV MET > 300 GeV
	$\geq 2b$	$\geq 1b$
QCD	$0.0 \pm 0.4^{+5.8}_{-0.0}$	$0.2 \pm 0.2^{+0.5}_{-0.2}$
top and W+jets	$24 \pm 7 \pm 5$	$13 \pm 5 \pm 4$
top and W+jets cross-check	—	$17.0 \pm 5.7 \pm 2.1$
$Z \rightarrow \nu\bar{\nu}$	$2.6 \pm 2.9 \pm 2.0$	$5.0 \pm 1.6 \pm 2.0$
Total SM	$25.8 \pm 7.4^{+7.8}_{-5.2}$	$18.2 \pm 5.3 \pm 4.5$
Data	30	20
SM MC prediction	35.7 ± 1.3	25.1 ± 1.6
LM9 signal	60.0 ± 2.5	27.7 ± 2.2

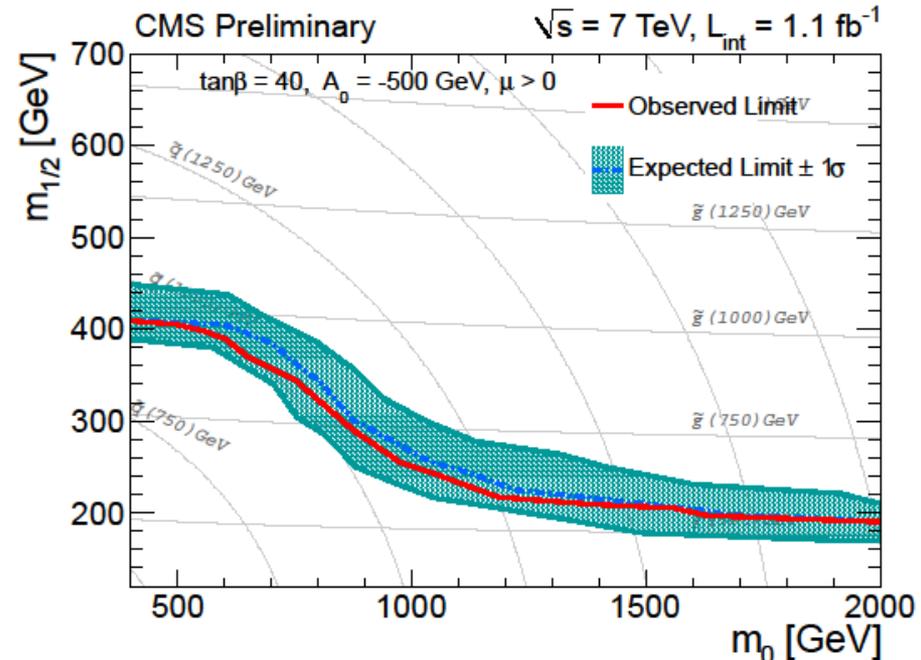
Not shown here: results of " $\geq 1b$ Loose" and " $\geq 2b$ Tight" selections.
 → Also good agreement between SM prediction and data.

Comparison of results in the CMSSM

MT2+b



MET+b



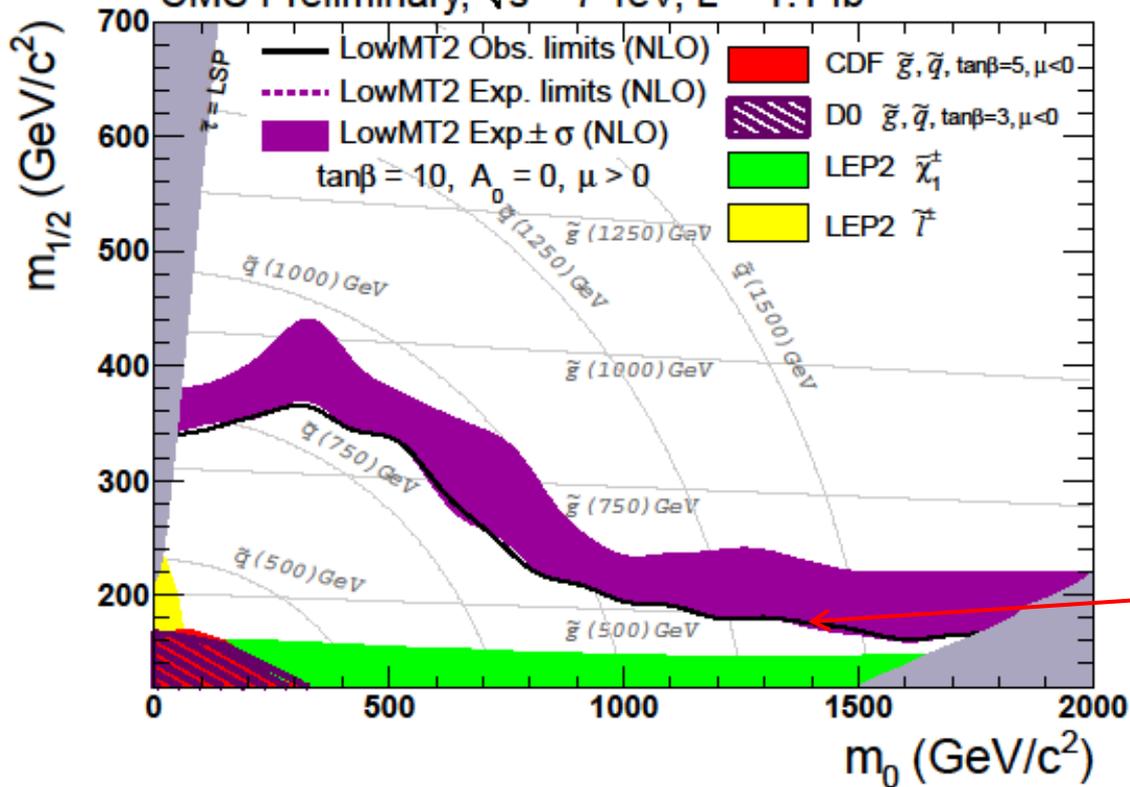
Note: $\geq 1\text{b}$ “Tight” selection gives best *expected* limit everywhere in CMSSM, so we focus on that result

Note: MT2+b is $\tan\beta=10$ while MET+b is $\tan\beta=40$
 \rightarrow ignoring this difference, limits are similar

More on MT2+b results in CMSSM

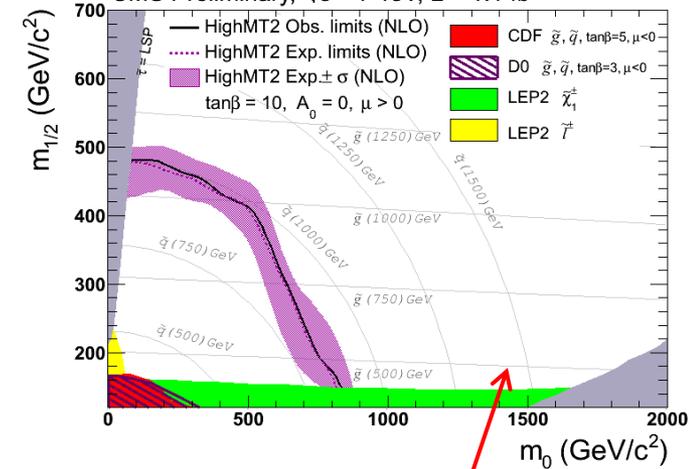
With b tag, looser MT2 cut

CMS Preliminary, $\sqrt{s} = 7$ TeV, $L = 1.1 \text{ fb}^{-1}$



No b tag, tighter MT2 cut

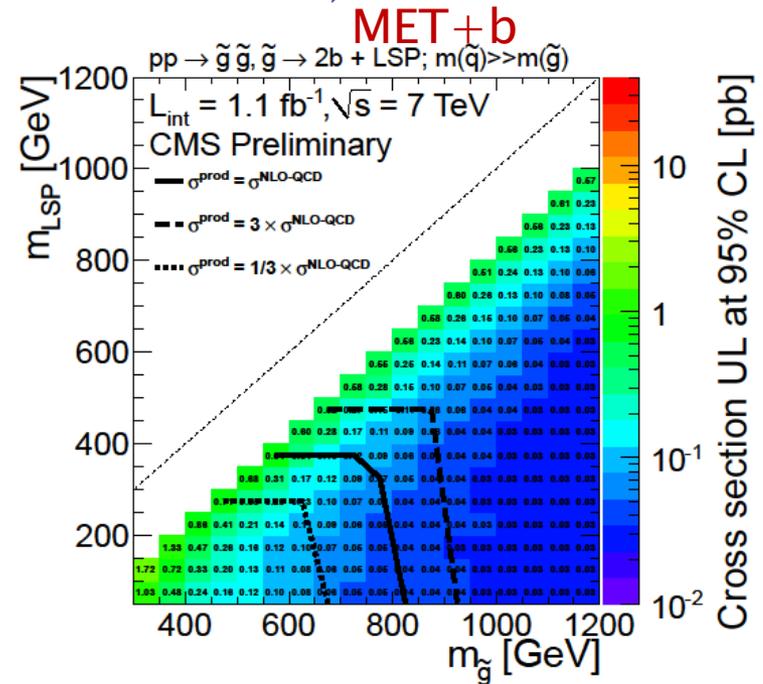
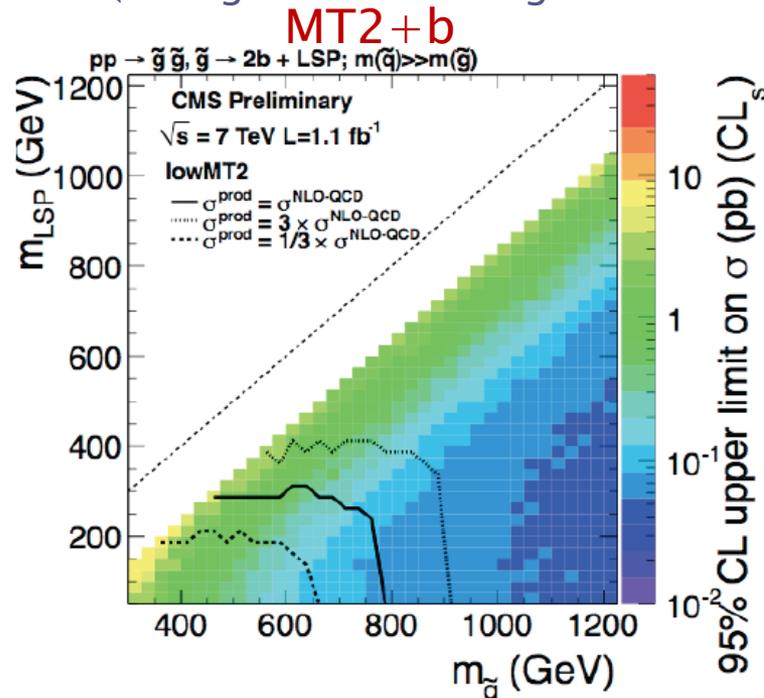
CMS Preliminary, $\sqrt{s} = 7$ TeV, $L = 1.1 \text{ fb}^{-1}$



b-tagged analysis
does better at high m_0

Interpretation in Simplified Models

- Hard to generalize results in full models like CMSSM
 - Instead look at a simplified model, which is easier for a theorist to use when building new models
 - In our case: $g\tilde{g} \rightarrow bbX\tilde{}$
 - Exclusive production and decay
 - Set an upper limit on the cross section as function of $m_{g\tilde{}}$, $m_{X\tilde{}}$
 - (Also get excluded region based on NLO cross section)

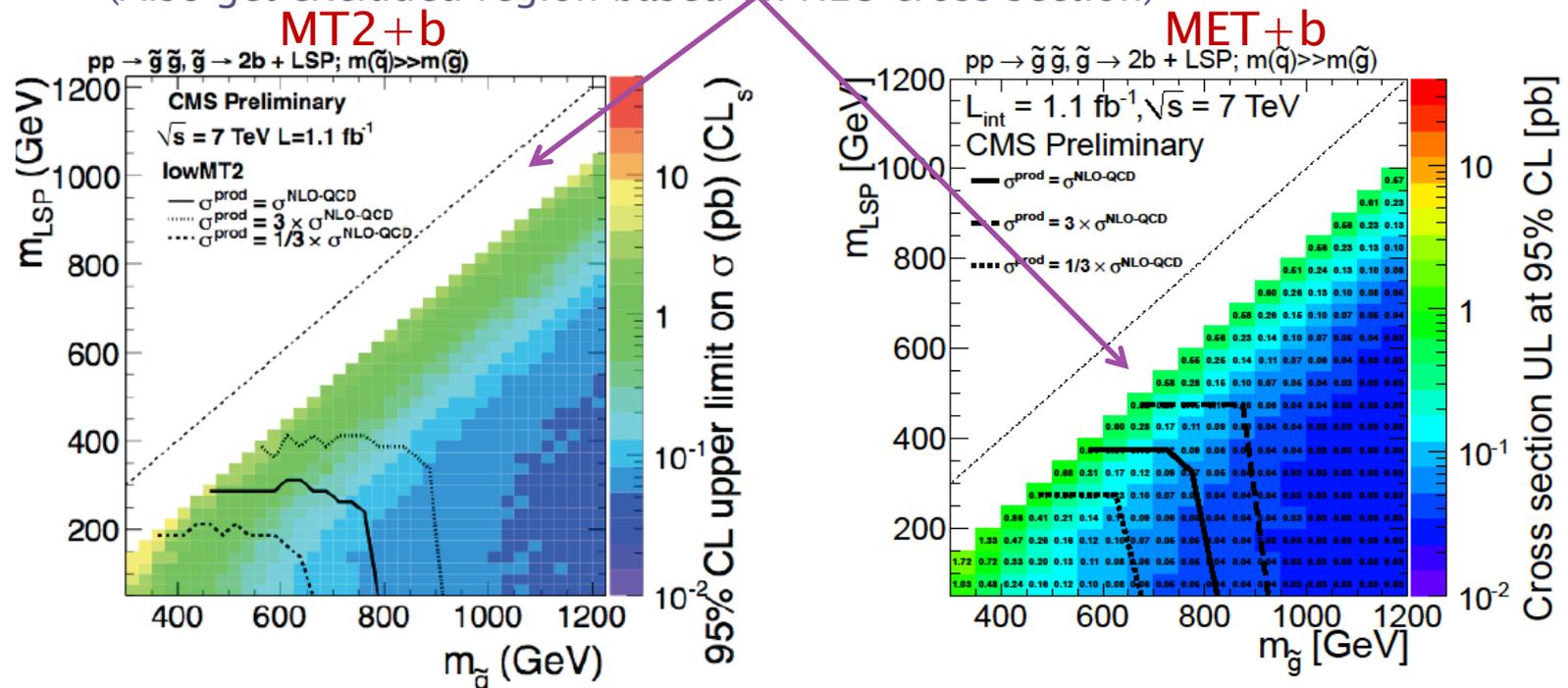


→ Similar sensitivity; MET + b does better in regions closer to the diagonal

Interpretation in Simplified Models

- Note: Region very near the diagonal is very sensitive to ISR.

At the moment we do not consider a systematic uncertainty due to ISR in these analyses, so we do not show results in this region.

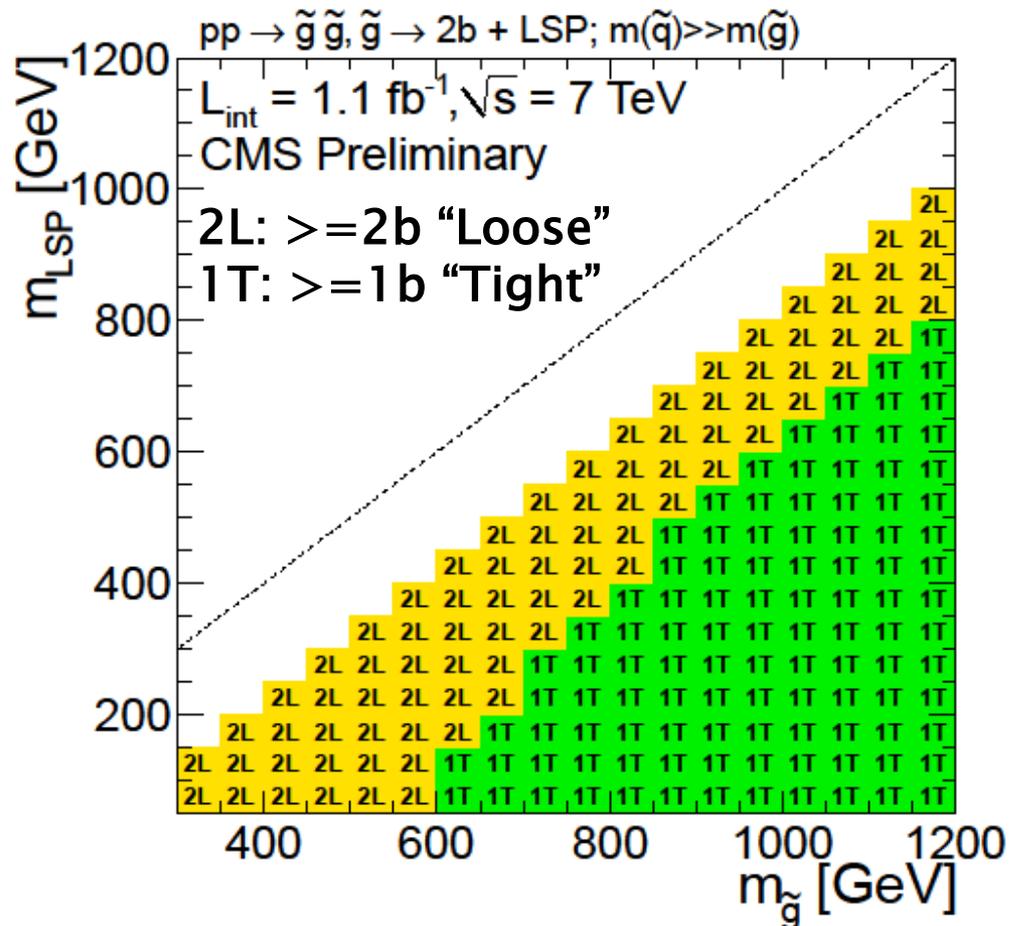


→ Similar sensitivity; MET + b does better in regions closer to the diagonal

Note on kinematics and selections

- Simplified models have widely varying kinematics by construction
 - Heavy gluino, light LSP gives high p_T daughters \rightarrow hard jets and lots of MET
 - Nearly degenerate gluino, LSP \rightarrow soft jets and little MET
 - Challenging! Favors looser selections
- In MET+b, choose to show the limit at each point as determined by the best expected limit
 - “expected” limit is derived from data-driven background estimates, but without using the observed data counts in the signal region
 - The limit you would expect if your observed data exactly matched your background estimate

MET+b: which selection is best



Conclusion

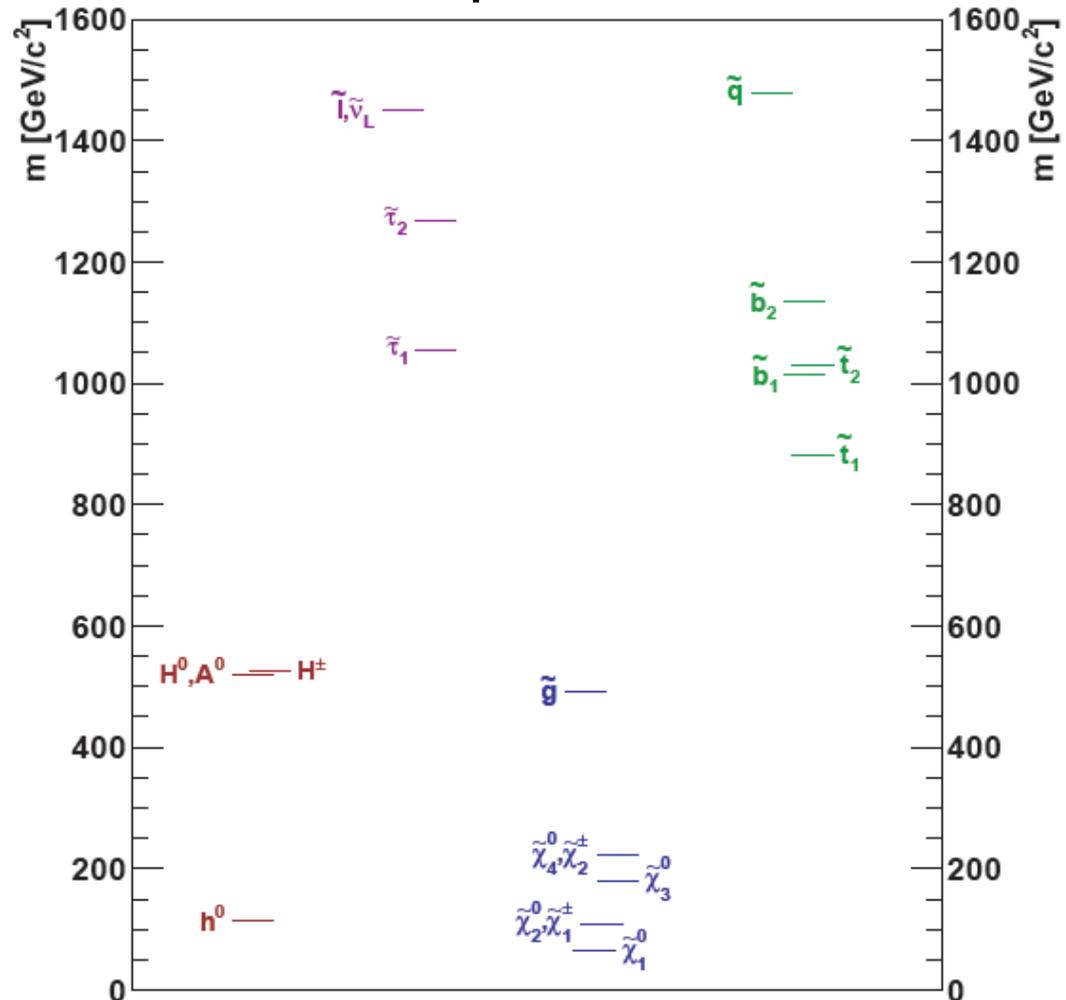
- CMS has two complementary searches for SUSY with b-jets
 - Different kinematic variable, different background methods, different choices of cuts
 - Similar results: no SUSY yet
- Results interpreted in CMSSM and 4b simplified model
 - Watch for more interpretation in simplified models in the future
- Further information:
 - <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>
 - MT2+b: CMS PAS SUS-11-005
 - <http://cms-physics.web.cern.ch/cms-physics/public/SUS-11-005-pas.pdf>
 - MET+b: CMS PAS SUS-11-006
 - <http://cms-physics.web.cern.ch/cms-physics/public/SUS-11-006-pas.pdf>

Extra slides

Details of signal point “LM9”

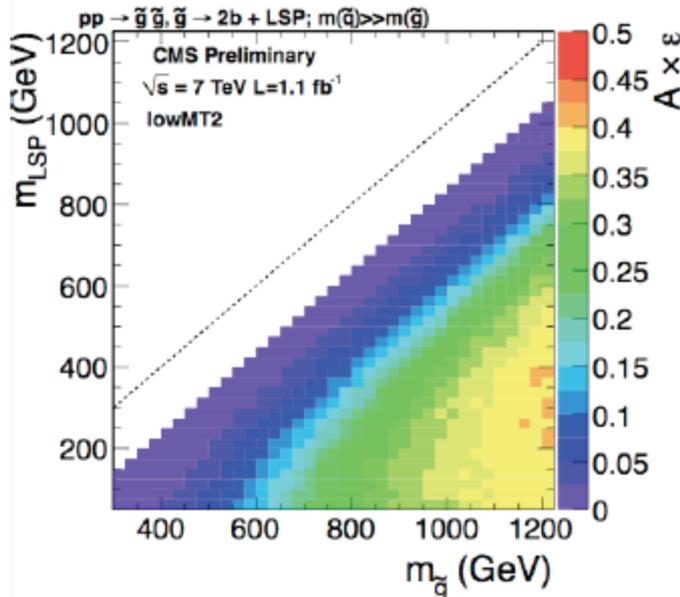
- High m_0 , low $m_{1/2}$, high $\tan \beta$
 - $m_0 = 1450$ GeV
 - $m_{1/2} = 175$ GeV
 - $A_0 = 0$ GeV
 - $\tan \beta = 50$
 - $\mu > 0$
- Light gluino, heavy squarks
 - 3rd generation SM from decays of gluinos

Mass spectrum

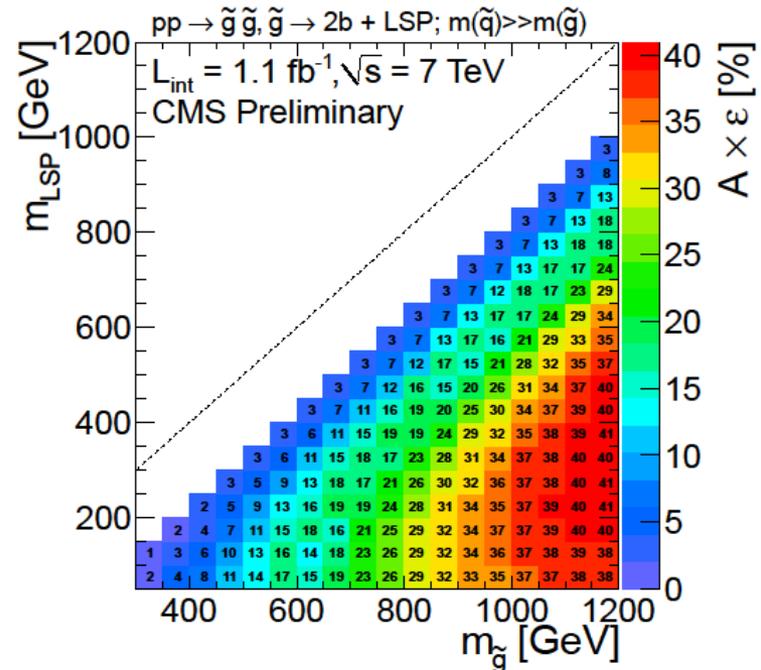


Signal efficiency in 4b model

MT2+b



MET+b



Shows efficiency of the selection used to make UL plot (best expected limit)

MC background expectations

Table 4: Expected background event yields and observed number of events in data for all relaxed cuts after preselection for events with at least one b-tagged jet.

Process	QCD	W+jets	Z+jets	Top	Other	Total Bkg.	data
After full selection	16857.5	27.8	14.8	445.3	24.6	17370	17370
MT2 > 80 GeV	58.8	7.5	5.5	61.4	0.0	133.3	131
MT2 > 100 GeV	10.1	5.2	4.6	36.9	0.0	56.9	49
MT2 > 120 GeV	3.0	3.6	3.9	23.3	0.0	33.8	26
MT2 > 135 GeV	0.8	2.7	2.6	15.8	0.0	21.9	21
MT2 > 150 GeV	0.2	2.2	1.8	10.8	0.0	15.0	19

Table 1: Number of data events and corresponding MC predictions for the loose ($H_T > 350$ GeV, $E_T^{\text{miss}} > 200$ GeV) and tight ($H_T > 500$ GeV, $E_T^{\text{miss}} > 300$ GeV) signal selections. MC results for the CMSSM test point LM9 are also shown. The MC uncertainties are statistical. The normalization is to 1143 pb^{-1} .

MET+b

	$(H_T, E_T^{\text{miss}}) > (350, 200) \text{ GeV}$		$(H_T, E_T^{\text{miss}}) > (500, 300) \text{ GeV}$	
	$\geq 1 \text{ b-jets}$	$\geq 2 \text{ b-jets}$	$\geq 1 \text{ b-jets}$	$\geq 2 \text{ b-jets}$
Data	155	30	20	5
Total SM	183 ± 5	35.7 ± 1.3	25.1 ± 1.6	4.54 ± 0.37
t \bar{t}	122 ± 2	28.9 ± 0.7	14.7 ± 0.8	3.49 ± 0.24
Single top	4.54 ± 0.38	0.77 ± 0.09	0.59 ± 0.15	0.12 ± 0.04
W+jets	17.0 ± 2.1	1.21 ± 0.45	4.20 ± 1.28	0.42 ± 0.28
$Z \rightarrow \nu\bar{\nu}$	22.5 ± 0.5	2.23 ± 0.10	4.25 ± 0.20	0.43 ± 0.04
$Z/\gamma^* \rightarrow \ell^+\ell^-$	0.17 ± 0.17	0.01 ± 0.01	0	0
Diboson	0.69 ± 0.07	0.10 ± 0.02	0.10 ± 0.02	0.006 ± 0.002
QCD	16.4 ± 3.9	2.5 ± 0.9	1.28 ± 0.40	0.08 ± 0.01
SUSY LM9	147 ± 5	60.0 ± 2.5	27.7 ± 2.2	10.1 ± 1.0

MET+b analysis

Signal efficiency systematics

Table 17: Systematic uncertainties, in percent, on the efficiency of the LM9 signal. The “Other” category includes the trigger efficiency, the lepton veto, and the anomalous E_T^{miss} terms.

Source	Loose search region		Tight search region	
	≥ 1 b	≥ 2 b	≥ 1 b	≥ 2 b
Jet energy scale	7.7	8.6	12.1	13.7
Jet energy resolution	0.1	0.3	3.0	4.2
Unclustered energy	2.0	1.6	5.7	7.5
Pileup	3.4	3.1	4.3	4.2
b-tagging efficiency	6.5	15.8	7.1	17.2
Parton distribution functions	11.1	11.2	11.8	12.1
Other	3.5	3.5	3.5	3.5
Luminosity	4.5	4.5	4.5	4.5
Total uncertainty	16.5	22.2	20.7	27.5

→JES, unclustered energy, b-tag eff, PDF are evaluated point-by-point across the CMSSM and simplified model planes

→Other uncertainties are fixed to LM9 values.

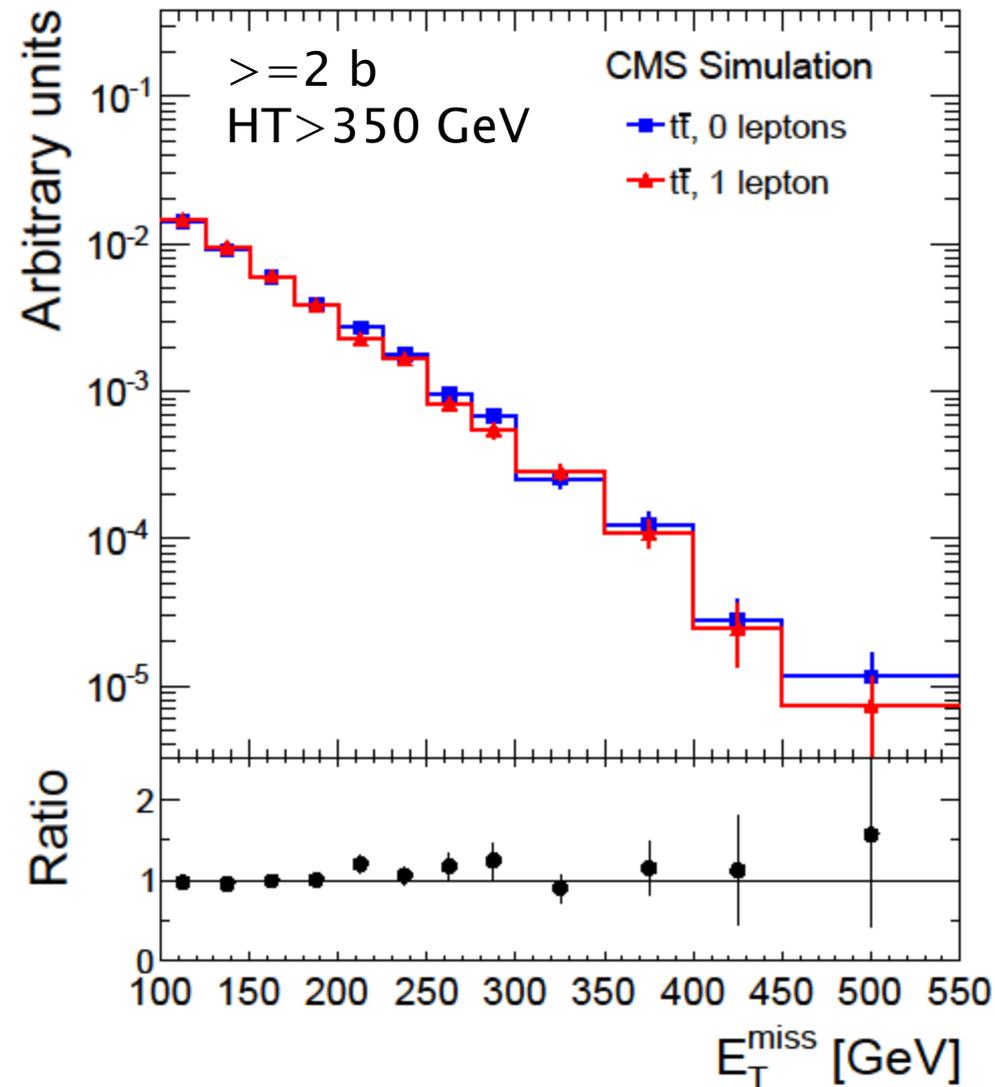
Event selection details

- Jets: in both cases, particle flow R=0.5 anti-kT jets
 - **MT2+b:**
 - $p_T > 20$ GeV, $|\eta| < 2.4$, passing quality criteria
 - Note that HT is calculated with calorimeter-only jets while all other quantities use particle flow
 - p_T cuts on lead jets
 - **MET+b:**
 - $p_T > 50$ GeV, $|\eta| < 2.4$, passing quality criteria
 - HT is calculated using all jets passing the above requirements
 - For b jets, use $p_T > 30$ GeV
- Leptons (particle flow):
 - $p_T > 10$ GeV
 - $|\eta| < 2.4$ (plus veto of barrel/endcap transition for electrons)
 - Various quality and isolation requirements
- $\Delta\phi_{(N)}^{\min}(\text{jet}, \text{MET})$
 - **MT2+b:** $\Delta\phi^{\min} > 0.3$ for all jets $p_T > 20$ GeV, $|\eta| < 5$
 - **MET+b:** $\Delta\phi_N^{\min} > 4$ for lead 3 jets passing criteria given above

ttbar background method details

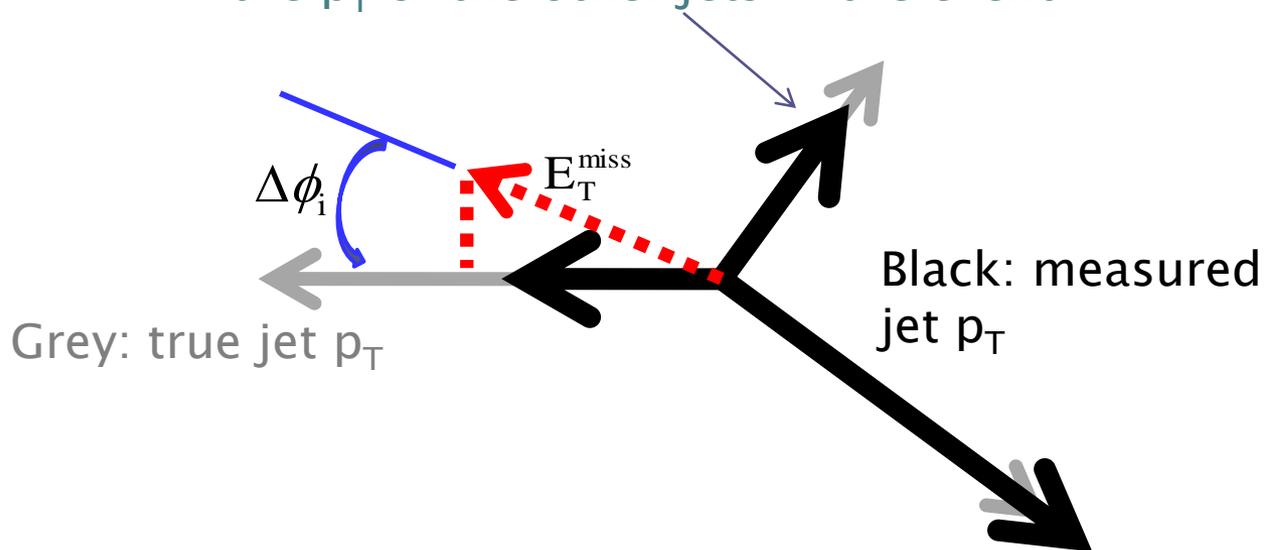
- Method depends on MET spectrum being the same in 1 lepton and 0 lepton samples
 - Checked in MC; with the present statistics it works well

MET+b analysis

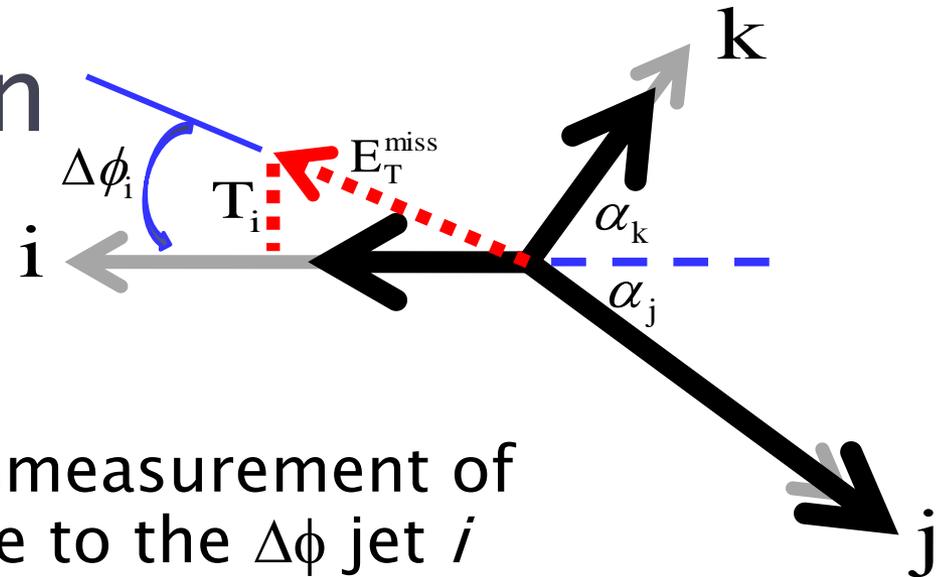


Motivation for $\Delta\phi_N(\text{jet}, \text{MET})$

- The standard $\Delta\phi(\text{jet}, \text{MET})$ variable is great for rejecting QCD at high MET
 - But it is also highly correlated with MET (and MT_2)
- For an event with a very badly measured jet, why is the angle $\Delta\phi(\text{jet}, \text{MET})$ non-zero?
 - The MET direction is smeared by the small mismeasurements of the p_T of the other jets in the event



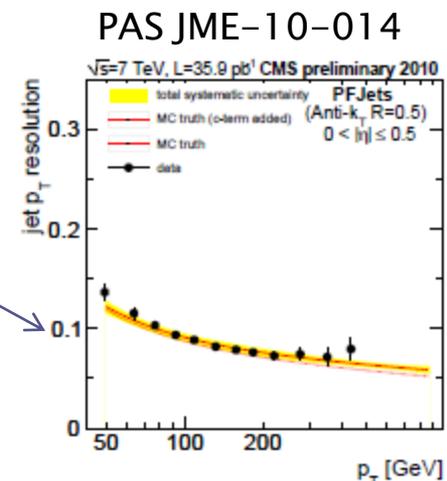
- This smearing becomes less important as the big mismeasurement (hence MET) increases \rightarrow MET and $\Delta\phi(\text{jet}, \text{MET})$ are correlated
- we try to model this and construct an uncorrelated variable

$\Delta\phi_N$ construction

- T_i is the component of mismeasurement of other jets that is transverse to the $\Delta\phi$ jet i

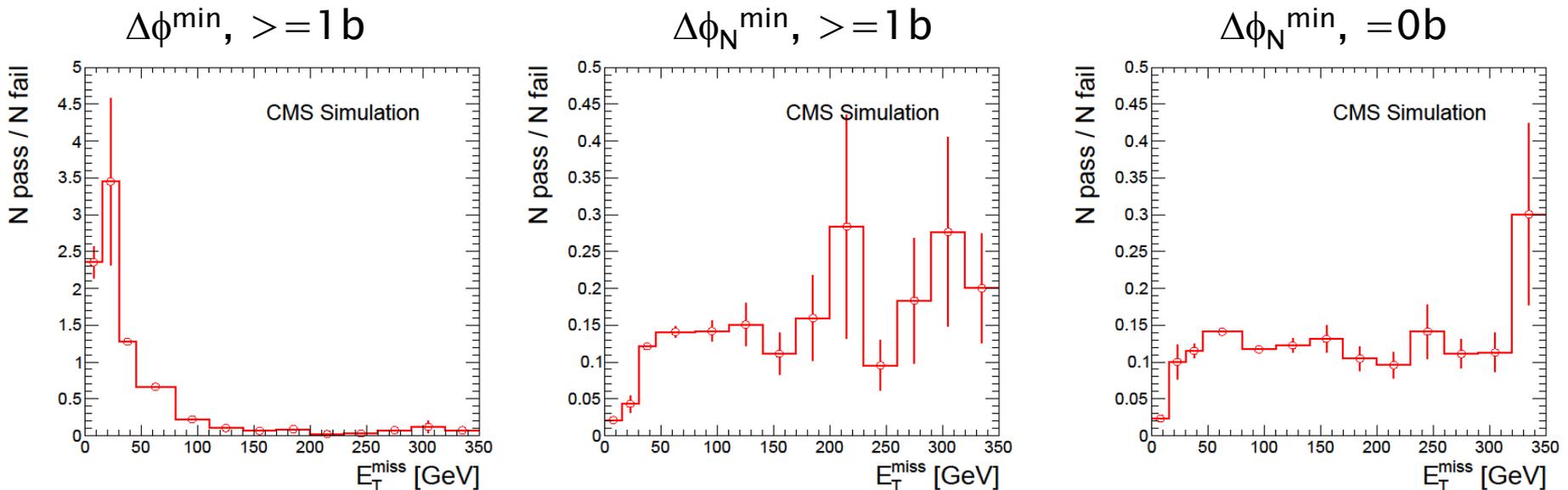
$$T_i^2 \approx \left(\sum_n \sigma_{p_{T,n}} \sin \alpha_n \right)^2$$

- Use 10% for jet p_T resolution $\sigma_{p_{T,n}}$
 - Cross-checks done to show we are not sensitive to this choice
- $\Delta\phi_{N,i} = \Delta\phi_i / \tan^{-1}(T_i / \text{MET})$
- This new variable is $\Delta\phi_i$ normalized by its resolution



$\Delta\phi$ versus $\Delta\phi_N$

- Plot the ratio of events passing the $\Delta\phi$ cut to the ratio failing it, as a function of MET
 - This is a good way to judge the correlation
 - (flat means uncorrelated)



→ pass/fail ratio for $\Delta\phi_N^{\text{min}}$ is ~constant for MET > ~30 GeV and independent of b tagging.

Can do an "ABCD" background estimation and use 0 b control sample

$Z \rightarrow \nu\nu$ method notes

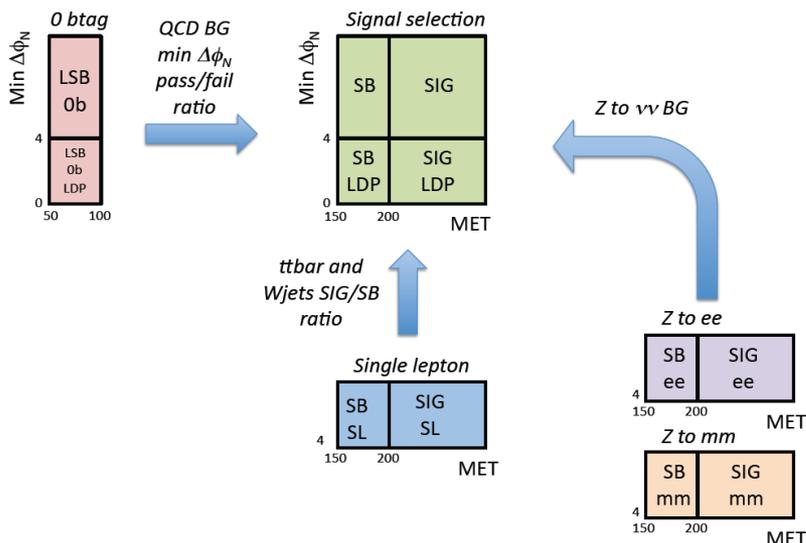
- $Z \rightarrow ll, l=e,\mu$ is simple (efficiency factors mostly straightforward to extract from data) but statistics-limited
 - In loosest selection (> 1 b, Loose), can directly apply signal region cuts to $Z \rightarrow ll$ samples
 - In other cases, need to loosen kinematic selections and then scale final estimate using MC
 - This MC scaling has been checked in several ways, including a data-driven method where the nominal MET, HT cuts are used but the b-tagging is loosened, and the (nominal b tagger)/(loose b tagger) factor is taken from a data control sample

MET+b analysis

Likelihood treatment (for limits)

- Combine background estimates into a RooStats framework that incorporates uncertainties and SUSY contamination
 - Event counts in data get Poisson uncertainties
 - 12 numbers total (11 control regions + signal box)
 - Note that the 5 of the control boxes can be “contaminated” by SUSY and this is treated in a consistent way in the likelihood
 - Other parameters get log normal uncertainties
 - 95% CL upper limits are evaluated using CLs tools built into RooStats

Data observables



Other Parameters

- systematics on the background estimation methods
 - e.g. closure test results, $Z \rightarrow \nu\nu$ efficiency factors, ...
- statistical and systematic uncertainty on signal efficiency