Data-driven Estimation of SM Backgrounds for SUSY Searches at the LHC

Takayuki Yamazaki (University of Tokyo)

on behalf of the ATLAS Collaborations & CMS Collaborations

Outline

[1] Introduction

[2] Data-driven Estimation

(1) One Lepton Mode

(2) No Lepton Mode

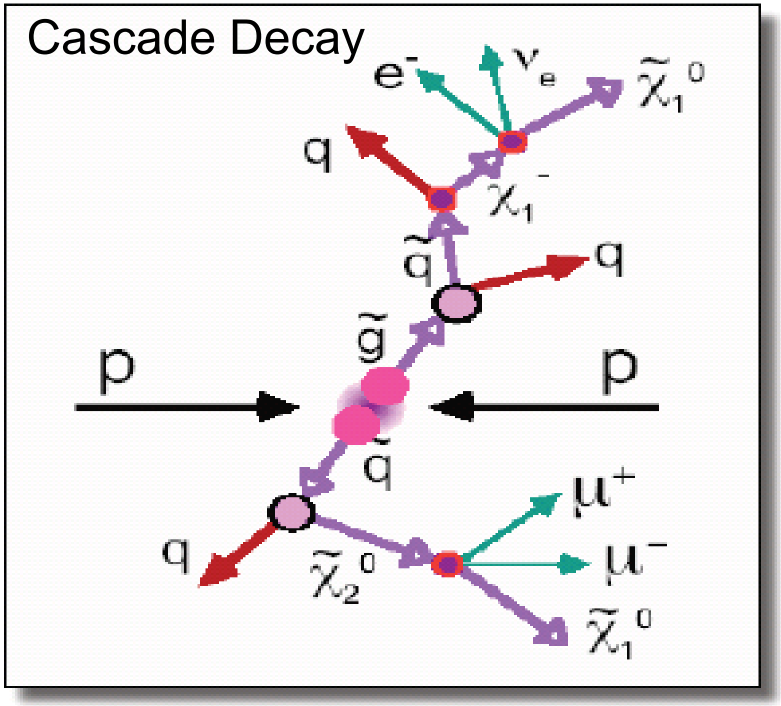
[3] If SUSY exists …

[4] Summary

[1] Introduction

SUSY Particle at the LHC

Gluinos/squarks are produced via strong interactions



R-parity conservation is assumed.

$$R = (-1)^{3(B-L)+2s}$$

→ LSP is stable.

Event Topology

Large missing E_T

High P_T multi-jet

(leptons)

Promising search mode

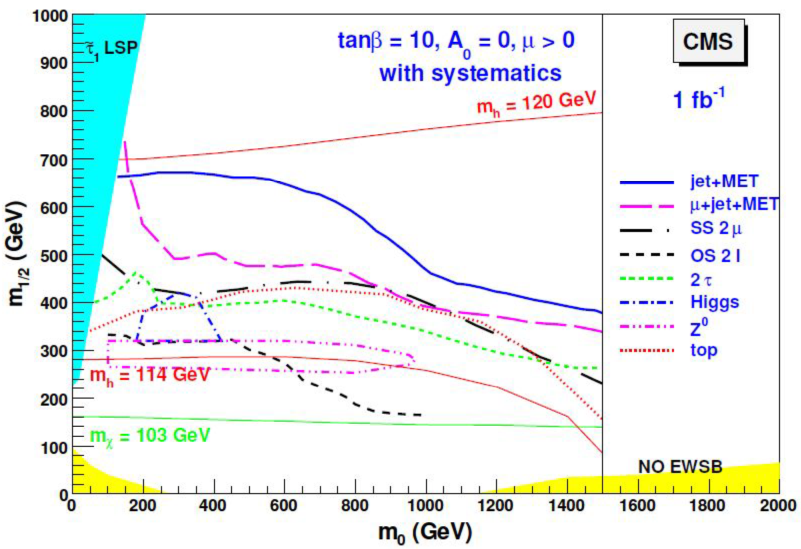
No lepton mode

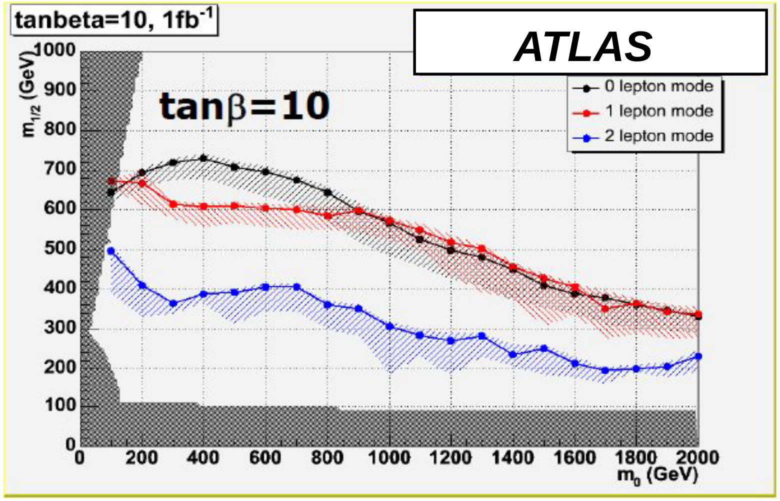
One lepton mode

Opposite-sign dilepton mode

Same-sign dilepton mode

Discovery Potential





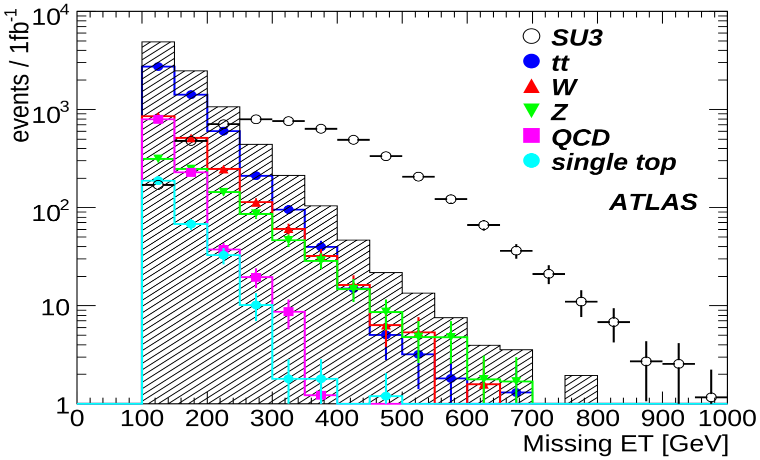
～1TeV in 1fb^{-1}

3TeV in much integrated luminosity

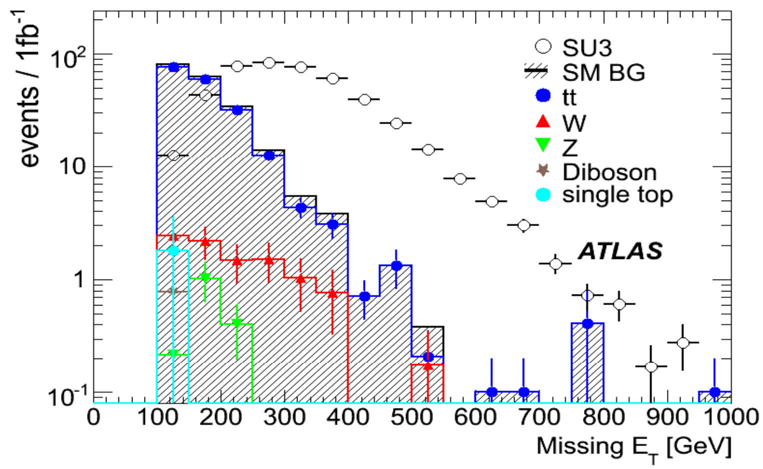
Discovery of SUSY is promising at the LHC.

[2] Data-driven Estimation

No Lepton Mode



One Lepton Mode



If SUSY exists, clear excess can be observed @ $1fb^{-1}$.

Since SUSY signal is observed as an excess from SM background in missing E_T distribution, deep understanding of SM background is essential for the SUSY searches and SM background should be estimated from the data in the early stage of collision.

"SU3 point" : m_0=100GeV, $m_{1/2}$=300GeV

gluino mass～700GeV, squark mass～650GeV

QCD = light flavor, bb, cc

(1) One Lepton Mode

Selection Cuts

Selection Cuts

missing E_T > 100GeV

missing E_T > 0.2Meff

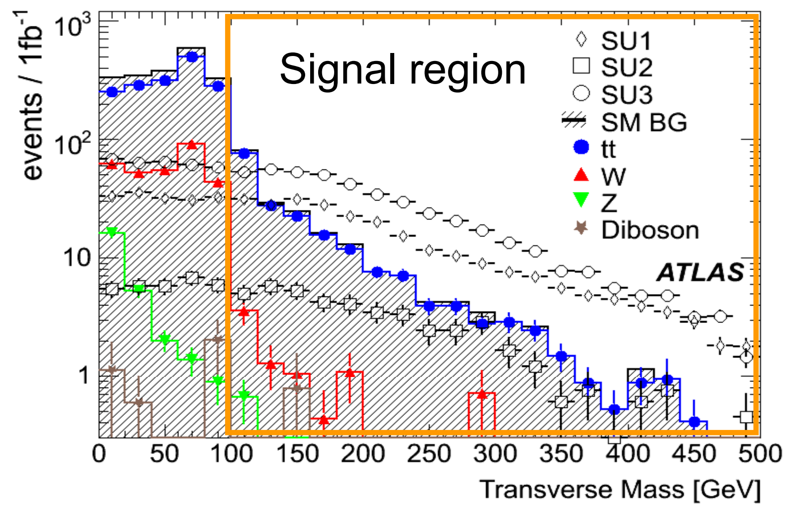
at least 4 jets with P_T>50GeV

at least 1 jet with P_T>100GeV

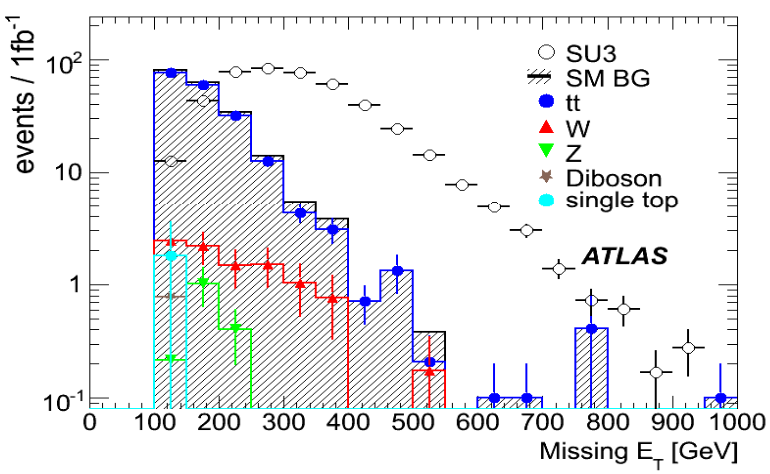
Transverse Sphericity S_T > 0.2

one lepton with P_T>20GeV

M_T(lepton, $\not{E}_T$) > 100GeV



M_T(lepton, $\not{E}_T$) > 100GeV is required to reduce $t\bar{t}$+jets and $W^{\pm}$+jets.



Data-driven

- To estimate the SM BG from data, “control sample” is used.

Control Sample

obtained from data

similar to SM BG in shape

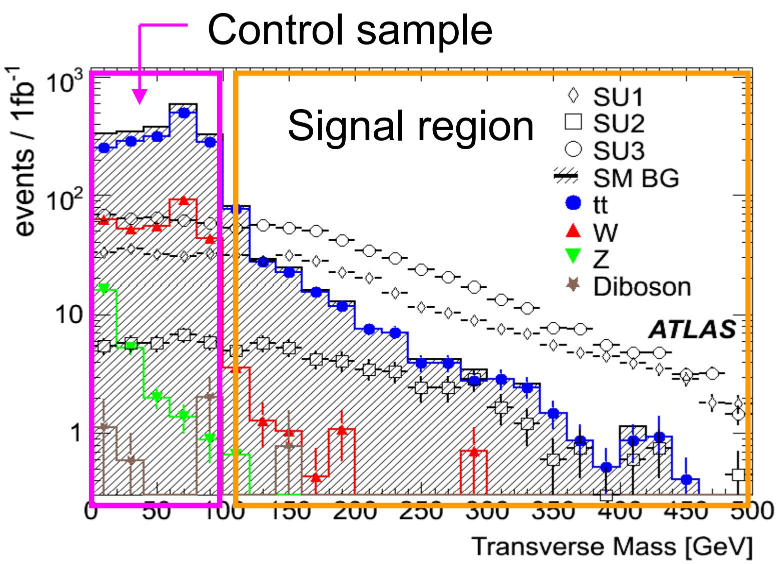
sufficient statistics

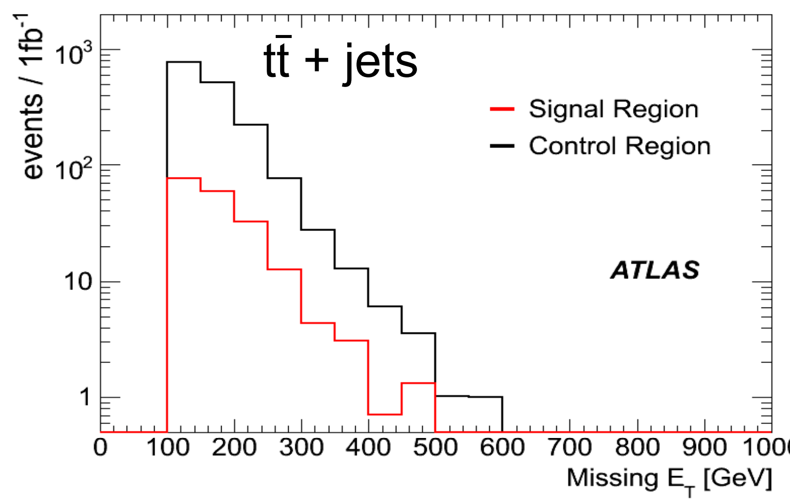
small SUSY contamination

As a control sample for one lepton mode , MT < 100GeV events are selected. (The other conditions are the same as the selection cuts.)

$t\bar{t}$+jets and $W^{\pm}$+jets dominant one lepton control sample is obtained.

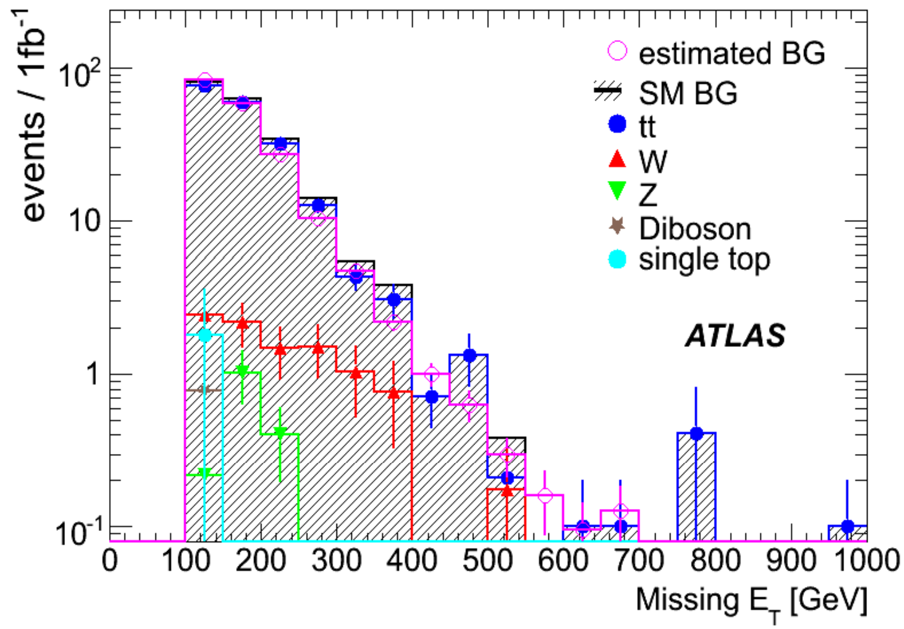
Missing E_T distribution is independent of MT.





Result

The normalization factor is obtained with the event numbers of the signal region and the control sample in $\not{E}_T$=100-200GeV.



In low $\not{E}_T$ region, SUSY contribution is expected to be small.

	mET > 100GeV	mET > 300GeV
True BG	203+/-6	12.4+/-1.6
Estimated BG	190+/-8	9.4+/-0.7
Ratio (Estimated/True)	0.93+/-0.05	0.76+/-0.11

(2) No Lepton Mode

Selection Cuts

Selection Cuts

missing E_T > 100GeV

missing E_T > 0.2Meff

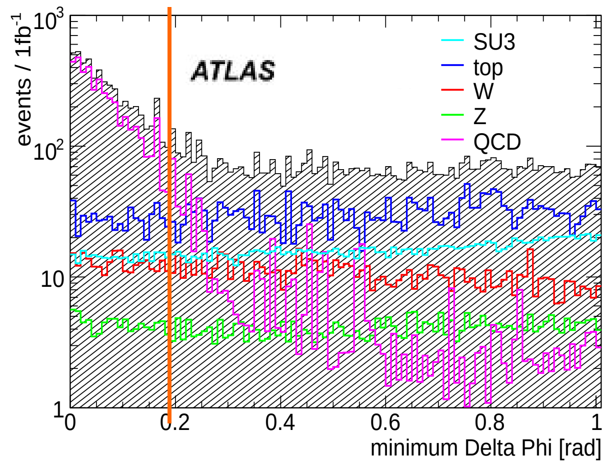
at least 4 jets with P_T>50GeV

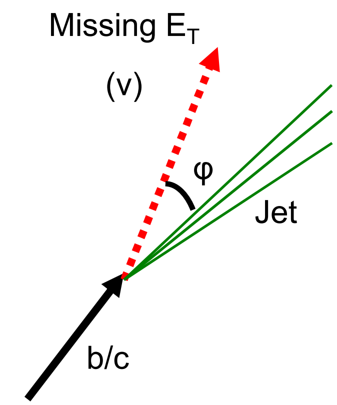
at least 1 jet with P_T>100GeV

Transverse Sphericity S_T > 0.2

no lepton with P_T>20GeV

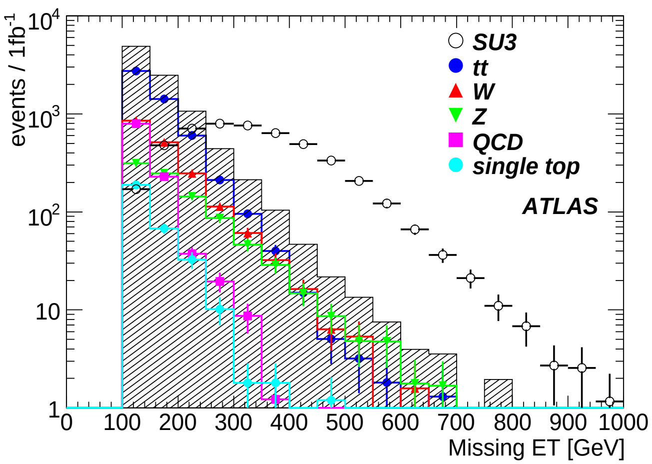
$\Delta\varphi(\not{E}_T$ – jet *i*) > 0.2 (*i* = 1, 2, 3)





semileptonic decay

φ is expected to be small



The reasons of QCD BG are

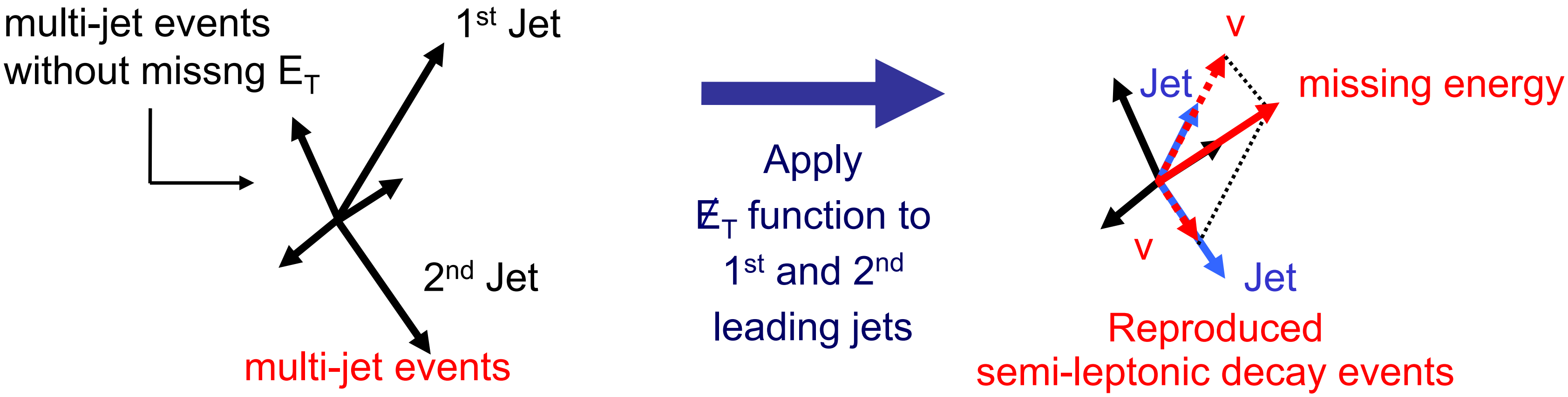
1. neutrinos emitted from semileptonic decays of b/c (real missing $\not{E}_T$)
2. mis-measurement of jet energies (fake missing $\not{E}_T$)

In both cases, the direction of $\not{E}_T$ is approximately pointing to the jet direction.

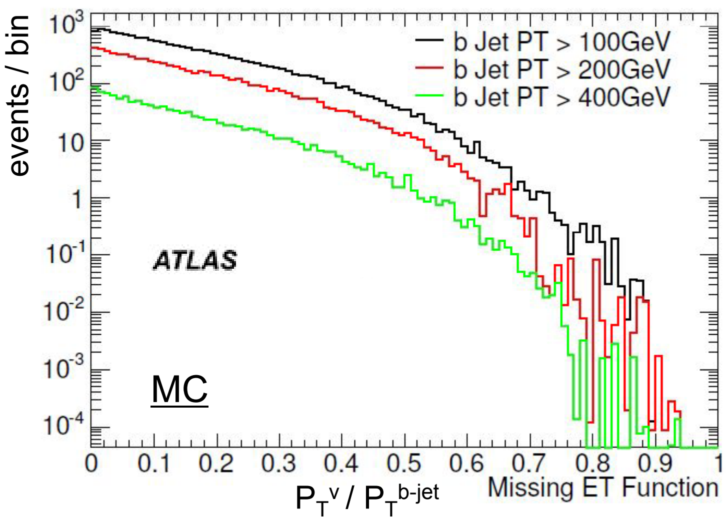
→ $\Delta\varphi(\not{E}_T$ – jet *i*) > 0.2 cut reduces QCD BG by ～80%.

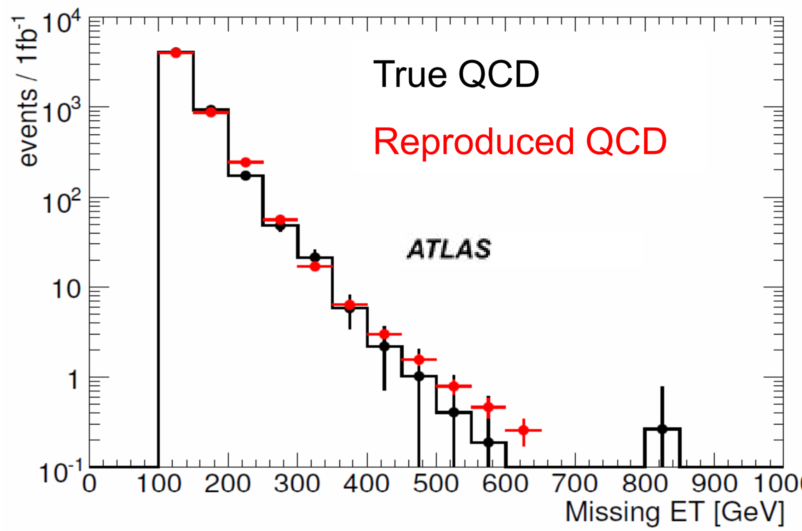
Data-driven (QCD BG)

- QCD BG can be estimated from multi-jet events without missing E_T.



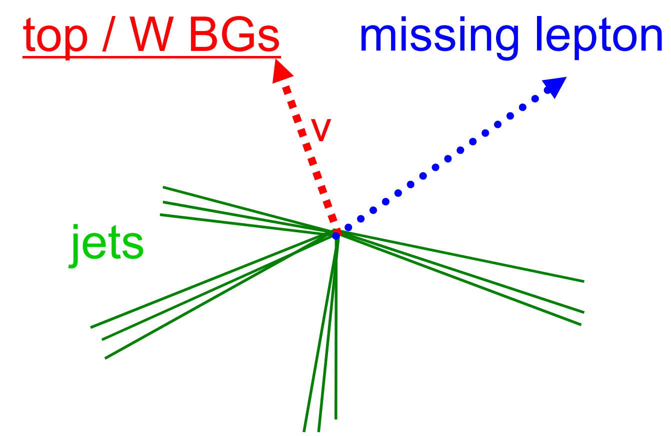
($N_{50GeV} \geq 4$ && $N_{100GeV} \geq 1$ && $\not{E}_T < 100GeV$)





Data-driven (top / W BG)

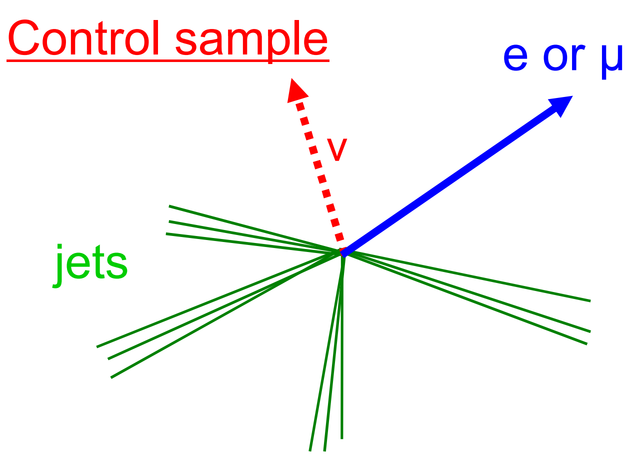
- top / W processes contribute to the background in the no lepton mode when the lepton emitted in the $W^{\pm}\to l\nu$ process is not identified.



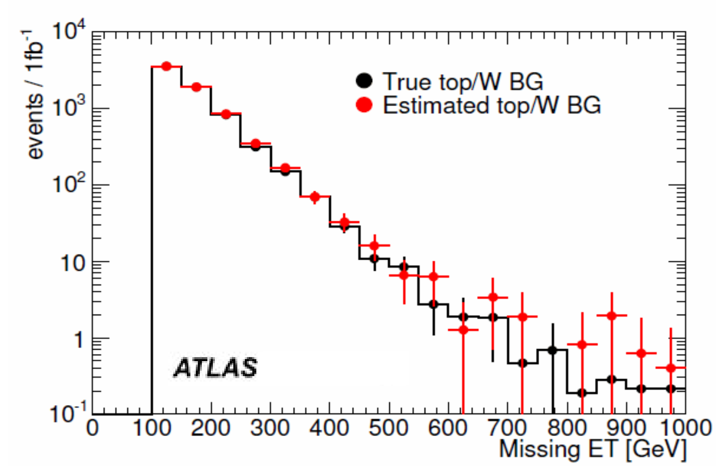
The main reasons of missing lepton are

1. tau->hadrons
2. out of acceptance (P_T < 20GeV)

Since the kinematics of the control sample and that of top / W BGs are almost the same except for the existence of lepton, top / W BGs can be estimated with the control sample.

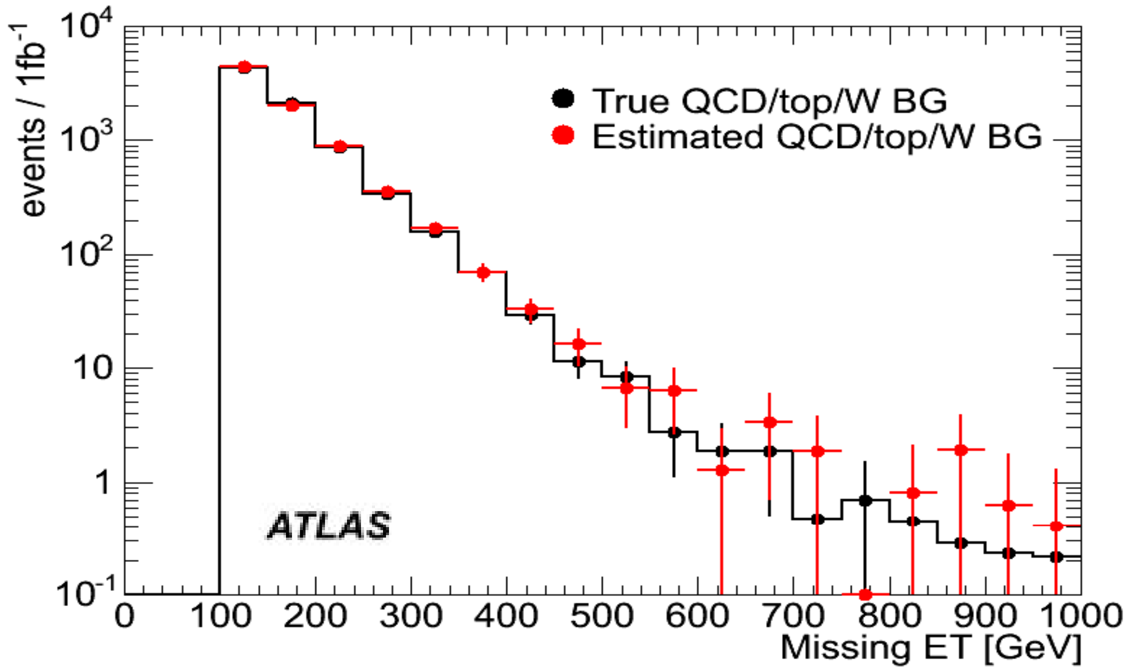


top / W dominant one lepton events



Result (QCD / top / W)

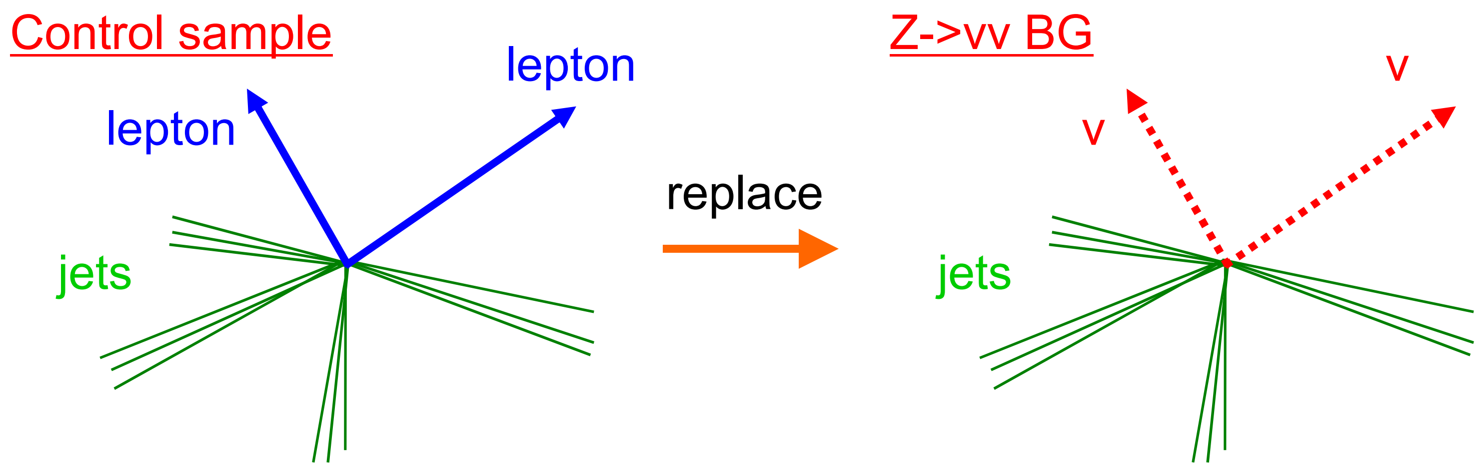
The normalization factor can be obtained from data, but its detail is in the backup slides.



	mET > 100GeV	mET > 300GeV
True QCD/top/W	8077+/-90	300+/-17
Estimated QCD/top/W	8158+/-273	327+/-28
Ratio (Estimated/True)	1.01+/-0.04	1.09+/-0.11

Data-driven (Z BG)

- Z (->νν) BG in the no lepton mode can also be estimated.



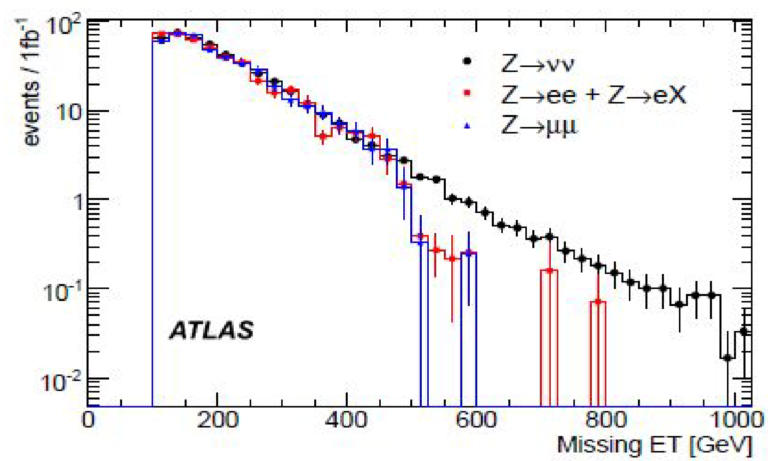
Replace Method

Two opposite-sign same flavor leptons with P_T > 20GeV

$\not{E}_T$ < 30GeV

M_Z - 10GeV < M_{ll} < M_Z + 10GeV

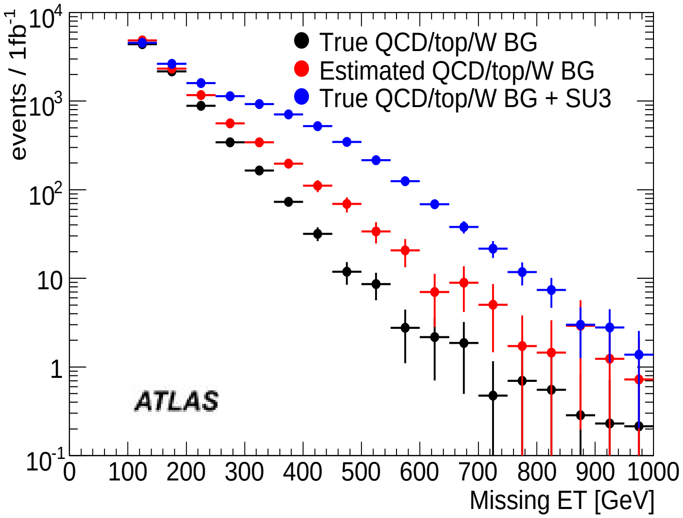
The standard no lepton SUSY cuts are applied after replacing ll with νν.



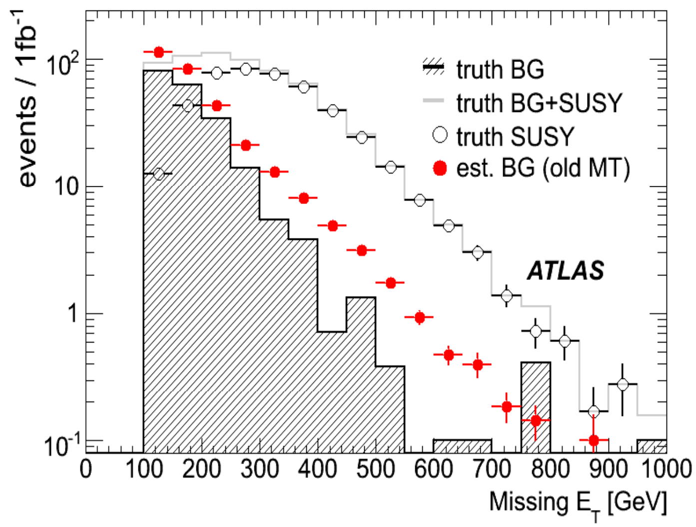
→ uncertainty ～ 15%

[3] If SUSY Exists …

No lepton mode



One lepton mode



	$\not{E}_T$ > 100GeV	$\not{E}_T$ > 300GeV
True BGs	8077+/-90	300+/-17
Estimated BGs	9726+/-312	804+/.-52
True BGs + SUSY	12925+/-114	3000+/-55

	$\not{E}_T$ > 100GeV	$\not{E}_T$ > 300GeV
True BGs	203+/-6	12.4+/-1.6
Estimated BGs	296+/-10	33.3+/-1.4
True BGs + SUSY	653+/-8	245+/-4

If SUSY exists, SM backgrounds will be overestimated because of SUSY contamination in the control samples and its contribution to normalization region ($\not{E}_T$=100-200GeV), but SUSY can be discovered @ 1fb^{-1} since the excess of SUSY signal is much larger.

[4] Summary

- ～1TeV SUSY can be discovered in $1fb^{-1}$ at the LHC
- If SUSY exists, clear excess can be observed in $\not{E}_T$ distribution.
- For the SUSY searches, SM background should be estimated from real data.
- By data-driven estimations, SM background can be estimated.
- If SUSY exists, SM background will be overestimated because of the SUSY contamination, but the over-estimation is much smaller than the excess of the SUSY signal, and clear excess can be observed.

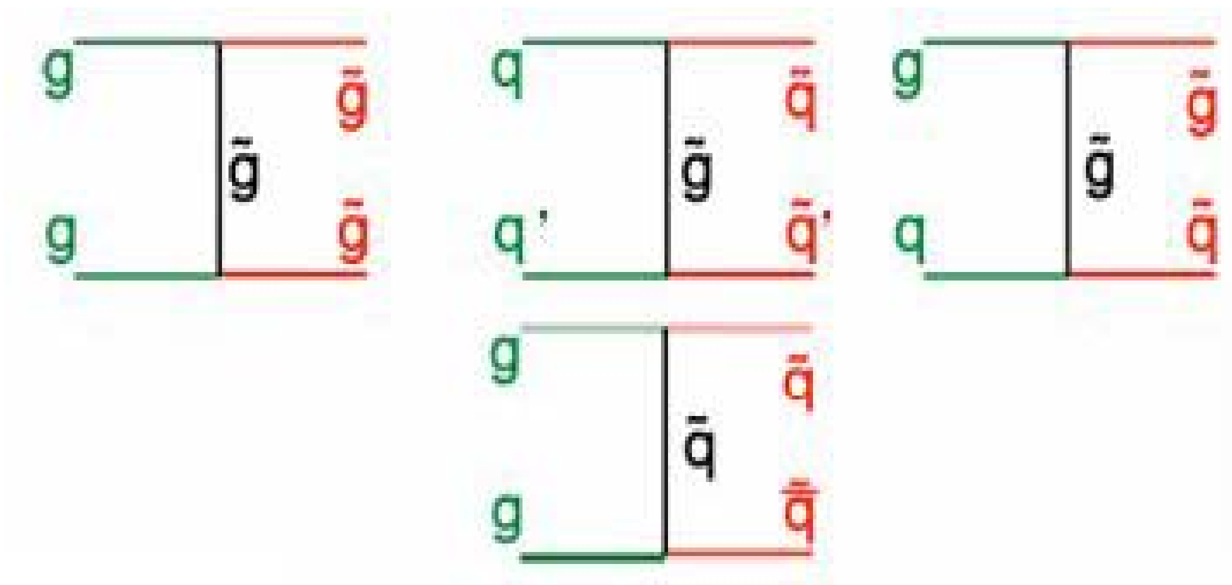
Backup

SUSY Particle at the LHC

LHC : proton-proton collider

Squarks and gluinos are produced via strong interactions.

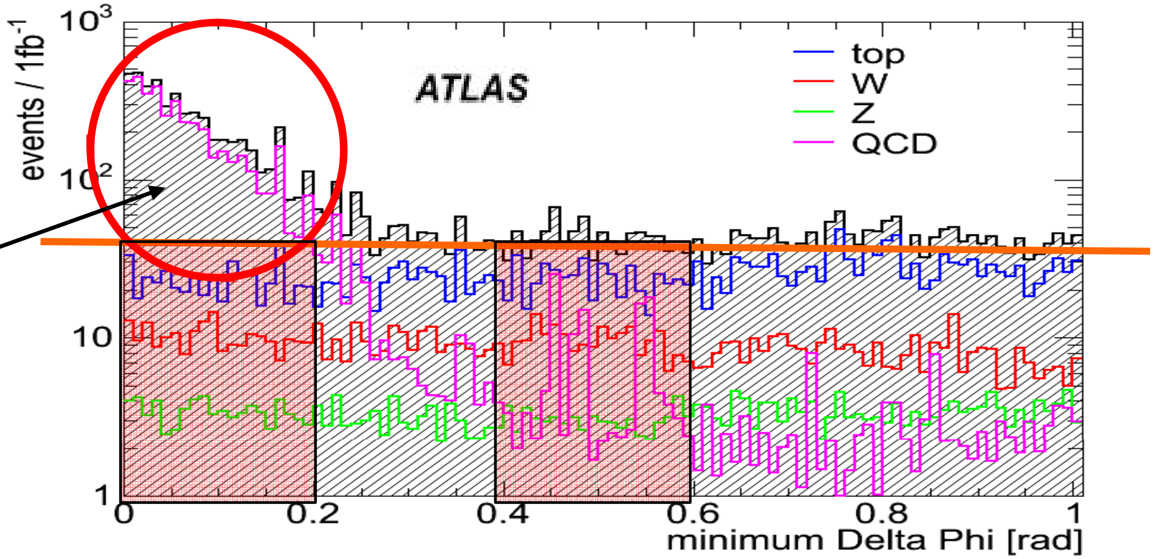
SUSY Particle Production Processes



Normalization of QCD BG

QCD BG is concentrated in min ΔΦ=0.0-0.2. On the other hand, top, W, and Z BGs are *flat*. So, we can estimate the number of top, W, and Z BGs in min Δφ=0.0-0.2 from all BGs in min ΔΦ=0.4-0.6 region. By subtracting the estimated number of top, W and Z BGs from the number of all BGs, the number of QCD BGs in min ΔΦ=0.0-0.2 can be obtained.

We can estimate the number of QCD BG in this region.



- We chose min Δφ=0.4～0.6 to avoid QCD tail effect.
- The number of QCD BG after Δφ cut is obtained from the following equation.

$$\frac{\#\text{ of QCD BG (min }\Delta\varphi = 0.0\text{-}0.2)}{\#\text{ of QCD BG (min }\Delta\varphi = 0.0\text{-}0.2)} \times [\#\text{ of All BG (min }\Delta\varphi = 0.0\text{-}0.2) - \#\text{ of All BG (min }\Delta\varphi = 0.4\text{-}0.6)]$$

We use MC information only to obtain this factor.

Normalization of top and W BG

- We can normalize the control sample using the number of the BG events in mET=100-200GeV. This number is obtained by the following equation.

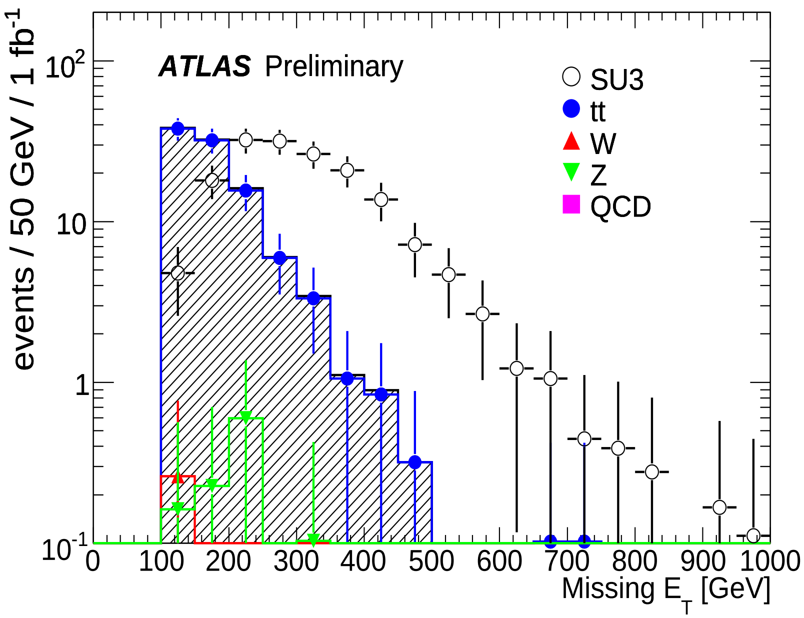
$$\mathrm{N(top\ and\ W\ BG) = N(All\ BG) - N(QCD\ BG) - N(Z\ BG)}$$

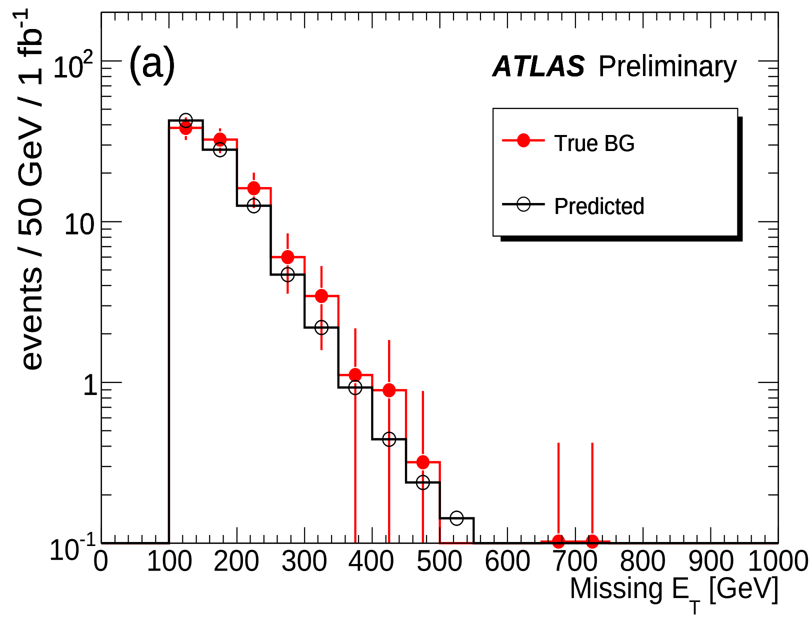
- The number of QCD BG can be obtained from the equation in the last page.
- The number of Z BG can be obtained with “MC method” or “Replace method” and its error is about 20% @ 1fb^{-1}.

- When SUSY exists, N(top and W BG) is overestimated since SUSY signal contributes to N(All BG).

OS Dilepton Mode

In a similar way, the BG in the OS dilepton mode can be estimated from data.





	mET > 100GeV	mET > 300GeV
True BG	98.8+/-9.9	6.0+/-2.5
Estimated BG	92.1+/-10.7	4.2+/-0.5

Systematic Errors

- No lepton mode -

	difference
jet & mET energy scale	< 5%
lepton energy scale	< 5%
lepton efficiency	< 5%
generator (ALPGEN->MC@NLO)	< 5%
ALPGEN generathion parameter	< 5%

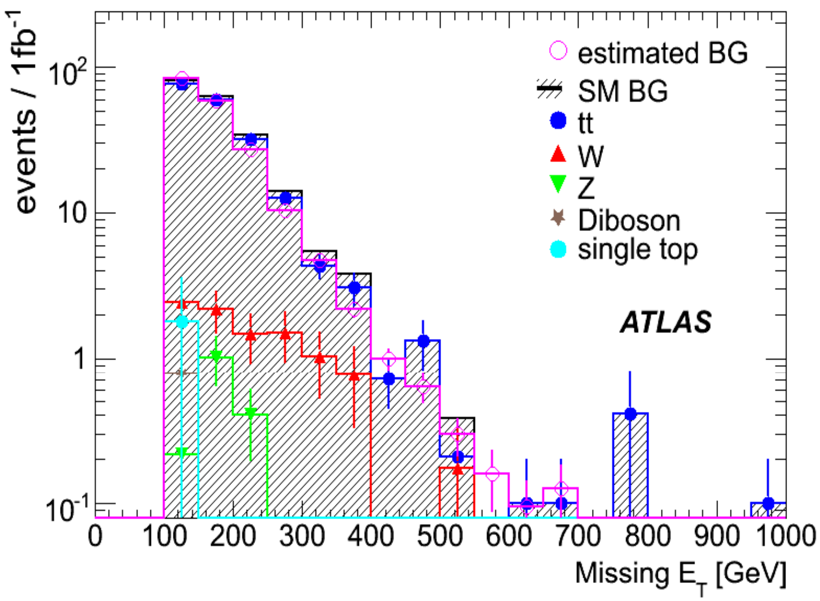
Data-driven estimation is stable against various systematic uncertainties.

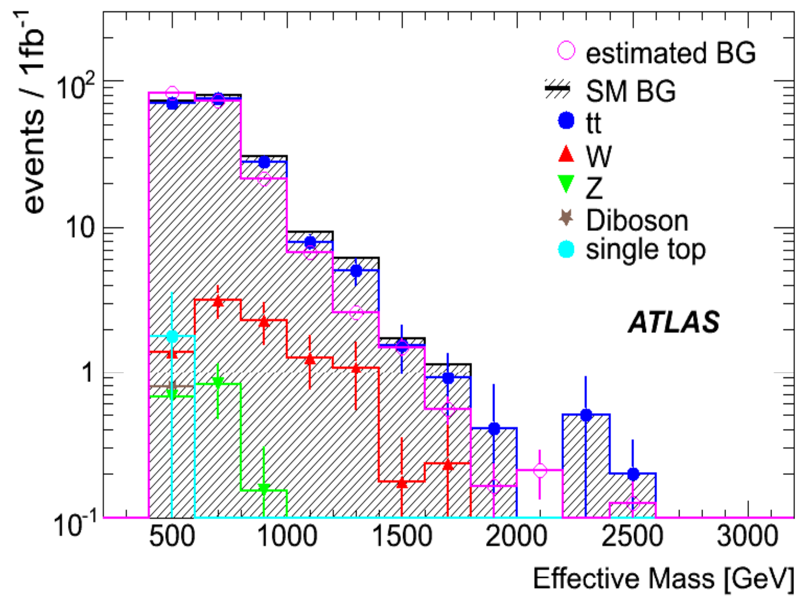
Systematic Error

- OS dilepton mode -

	difference
jet & mET energy scale	< 5%
lepton efficiency	< 5%
generator (ALPGEN->MC@NLO)	< 5%
ALPGEN generation parameter	< 5%

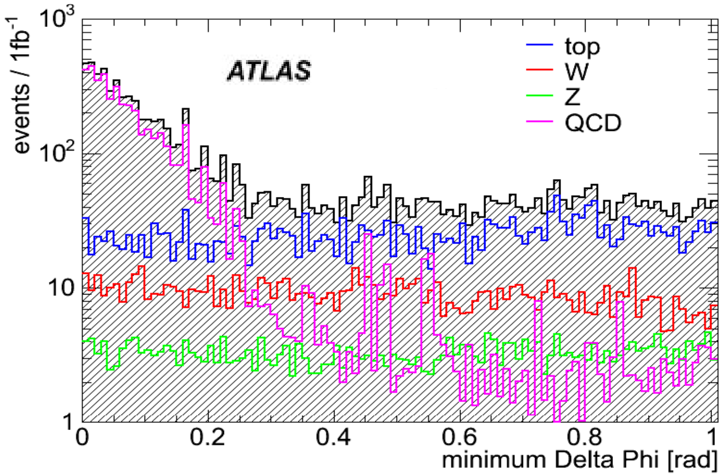
One Lepton Mode





	mET > 100GeV	mET > 300GeV
True BG	203+/-6	12.4+/-1.6
Estimated BG	190+/-8	9.4+/-0.7
Ratio (Estimated/True)	0.93+/-0.05	0.76+/-0.11

No Lepton Mode



min ΔΦ @ mET=100～200GeV

	Total	0.0～0.2	0.2～0.4	0.4～0.6
All BG	11826	4720	1002	838
QCD BG	5025	4002	289	129

Truth QCD (min ΔΦ < 0.2) = 4002+/-63

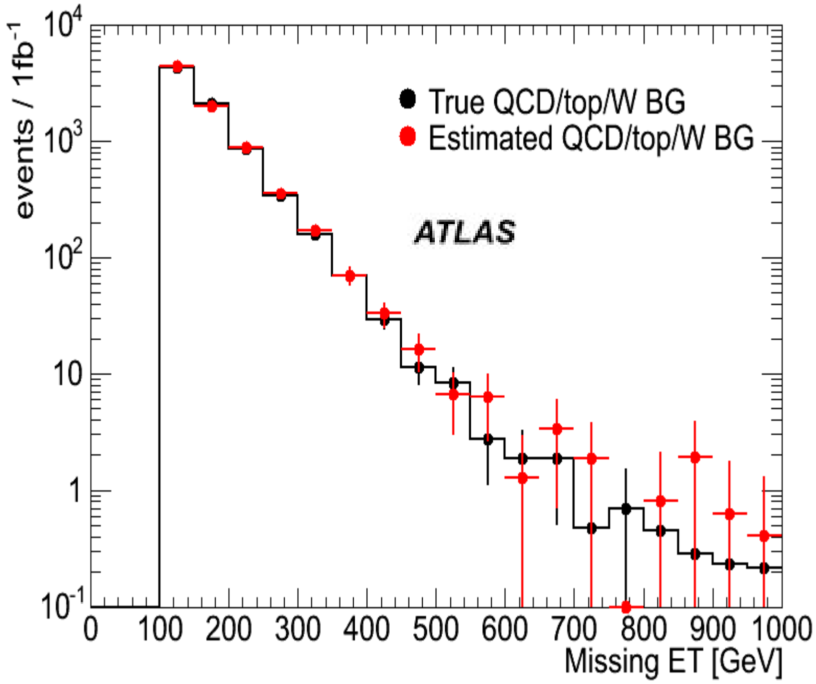
Estimated QCD (min ΔΦ < 0.2) = 3881+/-75

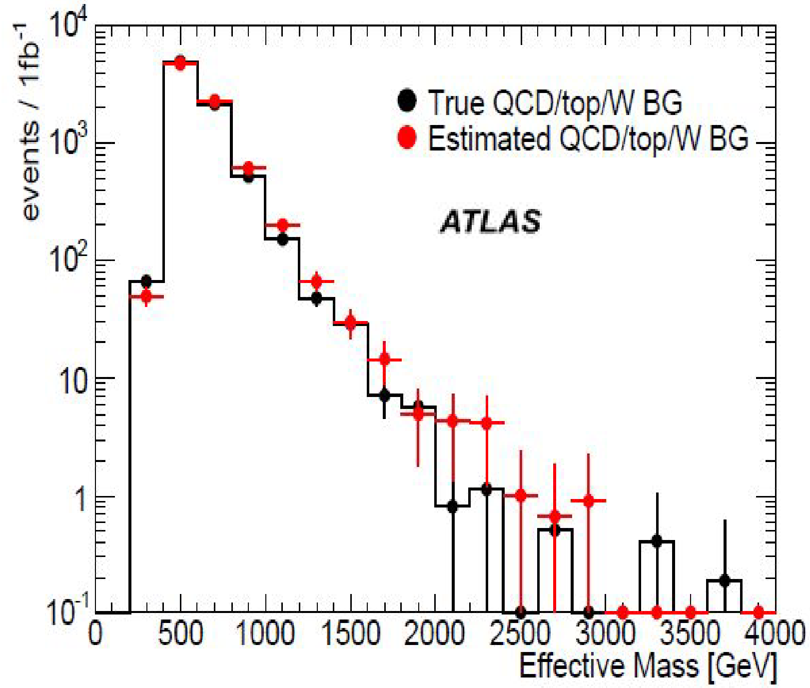
$$\frac{\#\ \text{of QCD BG (min } \Delta\varphi > 0.2)}{\#\ \text{of QCD BG (min } \Delta\varphi < 0.2)} = 0.256 \pm 0.009$$

Numbers @ min ΔΦ > 0.2 && mET=100～200GeV

	All	QCD	Z	top/W
Truth	7106+/-84	1023+/-32	560+/-24	5523+/-74
Estimation	-	992+/-40	560+/-112	5554+/-146

No Lepton Mode

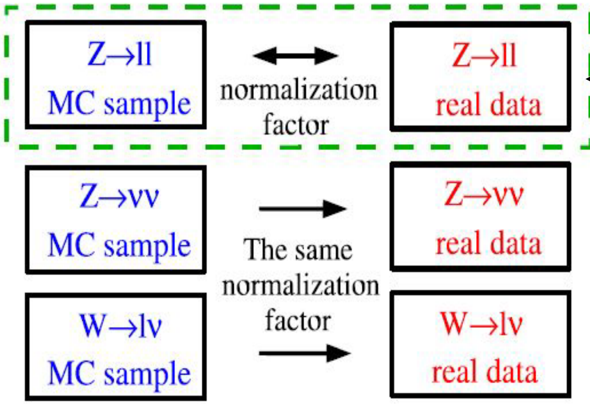




	mET > 100GeV	mET > 300GeV
True QCD/top/W	8077+/-90	300+/-17
Estimated QCD/top/W	8158+/-273	327+/-28
Ratio (Estimated/True)	1.01+/-0.04	1.09+/-0.11

MC Method (No Lepton Z BG)

- Z (->νν) BG in the no lepton mode can also be estimated.



Control sample

Two opposite-sign same flavor leptons with P_T > 20Gev

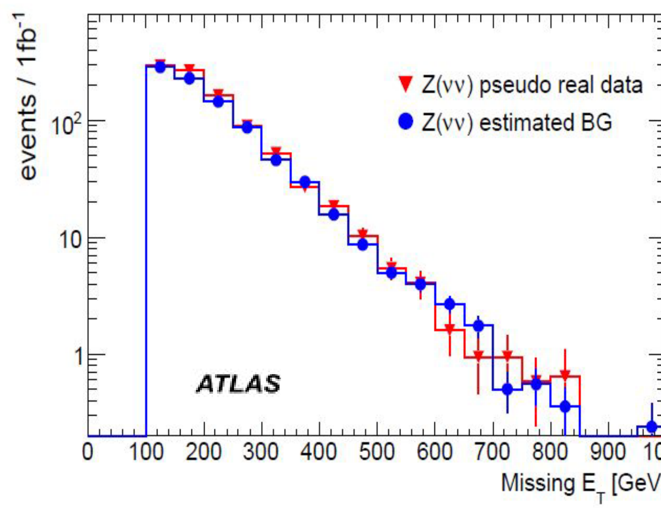
$\not{E}_T$ < 40GeV

M_Z - 10GeV < M_{ll} < M_Z + 10GeV

The standard no lepton SUSY cuts are applied after replacing ll with νν.

The normalization factor is obtained with the event numbers of the MC sample and the control sample.

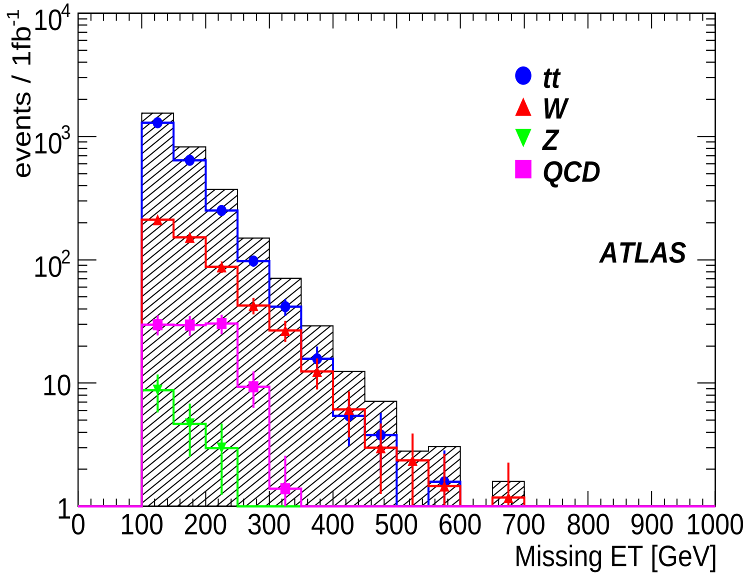
Since Z->ll events are statistically limited, we rely on the MC sample (Z->νν) for the shape of Z->νν BG distributions and just obtain the normalization factor with the event numbers of the MC sample (Z->ll) and the control sample (Z->ll).



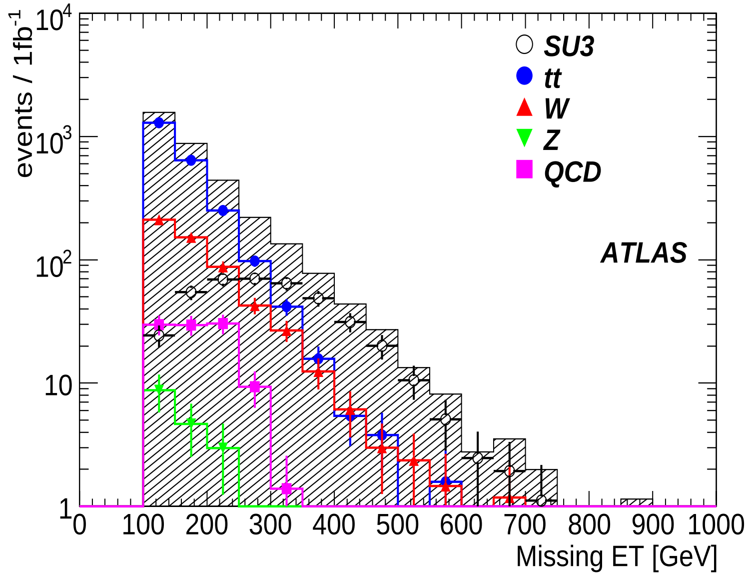
→ uncertainty ～ 20%

Control Sample with SUSY

Without SUSY

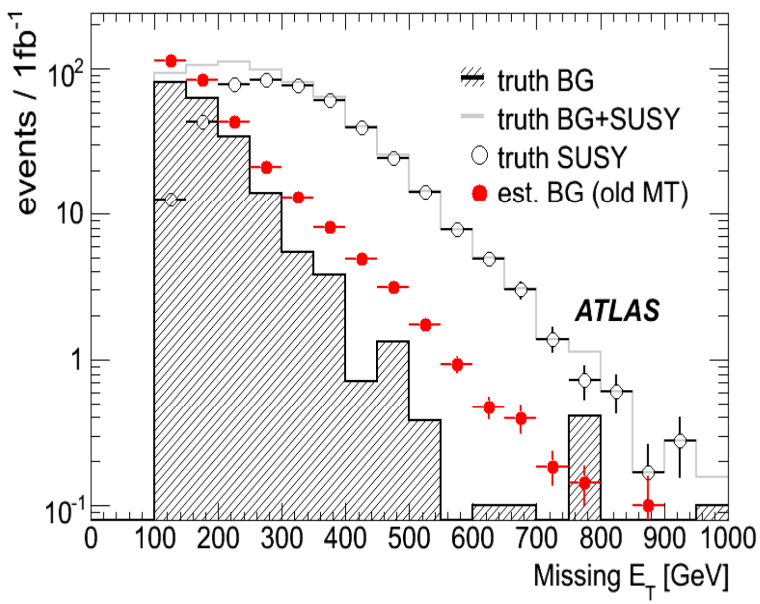


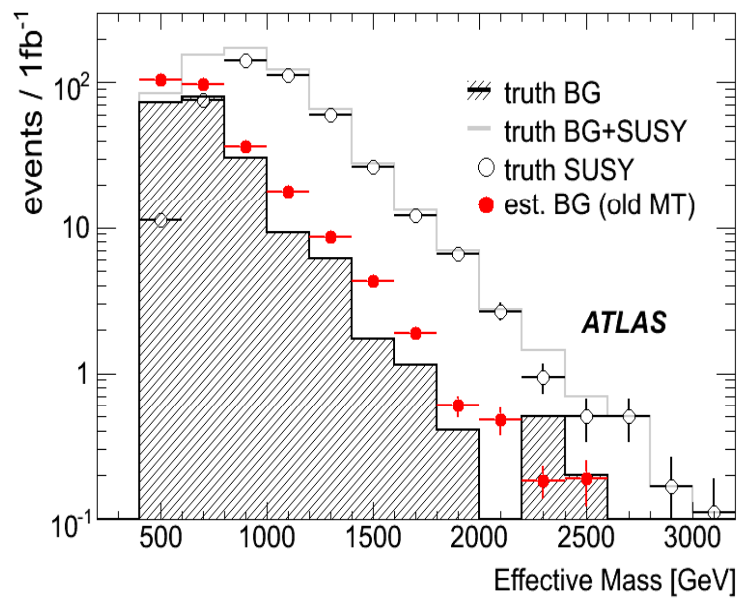
With SUSY



- Because of SUSY contamination, the control sample becomes harder.

One Lepton Mode

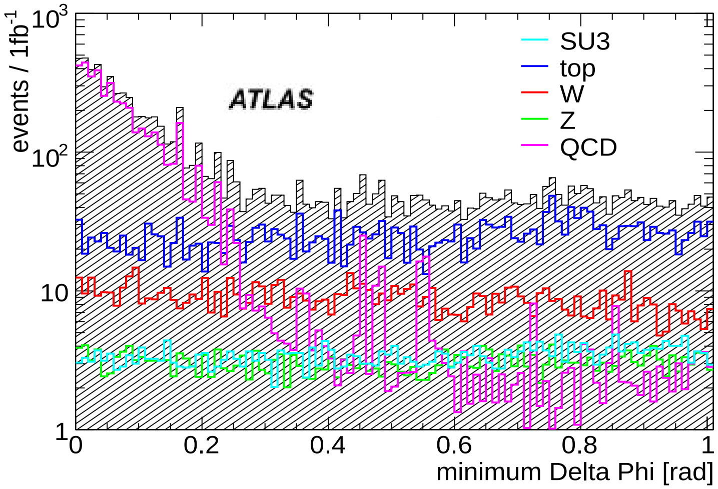




	mET > 100GeV	mET > 300GeV
True QCD/top/W	203+/-6	12.4+/-1.6
Estimated QCD/top/W	296+/-10	33.3+/-1.4
True QCD/top/W + SUSY	653+/-8	245+/-4

-> We can discover SU3 @ $1fb^{-1}$.

No Lepton Mode



min ΔΦ @ mET=100～200GeV

	Total	0.0～0.2	0.2～0.4	0.4～0.6
SUSY + All BG	12540	4785	1068	903
QCD BG	5025	4002	289	129

Truth QCD (min ΔΦ < 0.2) = 4002+/-63

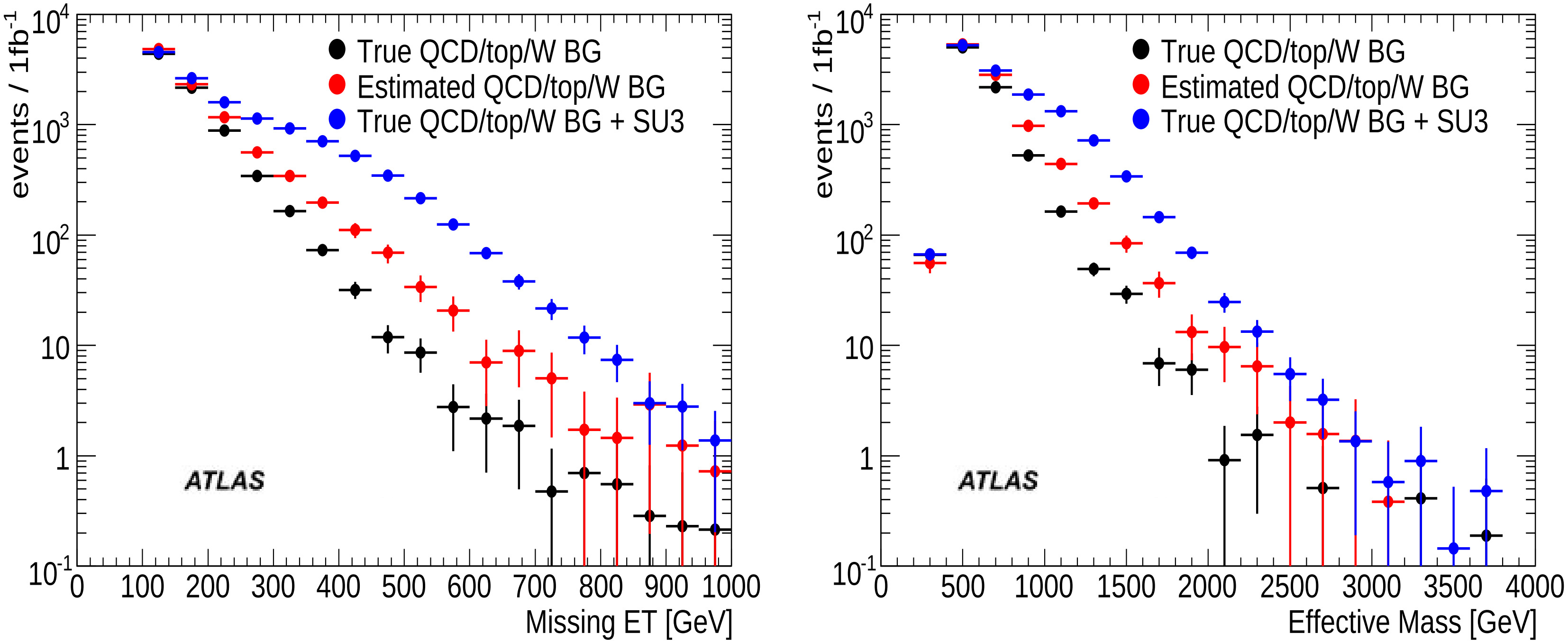
Estimated QCD (min ΔΦ < 0.2) = 3881+/-75

$$\frac{\#\text{ of QCD BG }(\min \Delta\varphi > 0.2)}{\#\text{ of QCD BG }(\min \Delta\varphi < 0.2)} = 0.256 \pm 0.009$$

Numbers @ min ΔΦ > 0.2 && mET=100～200GeV

	All	QCD	Z	top/W	SUSY
Truth	7755+/-88	1023+/-32	560+/-24	5523+/-74	649+/-25
Estimation	-	992+/-40	560+/-112	6203+/-148	(0)

No Lepton Mode



	mET > 100GeV	mET > 300GeV
True QCD/top/W	8077+/-90	300+/-17
Estimated QCD/top/W	9726+/-312	804+/-52
True QCD/top/W + SUSY	12925+/-114	3000+/-55

-> We can discover SU3 @ $1fb^{-1}$.

Improved MT Method

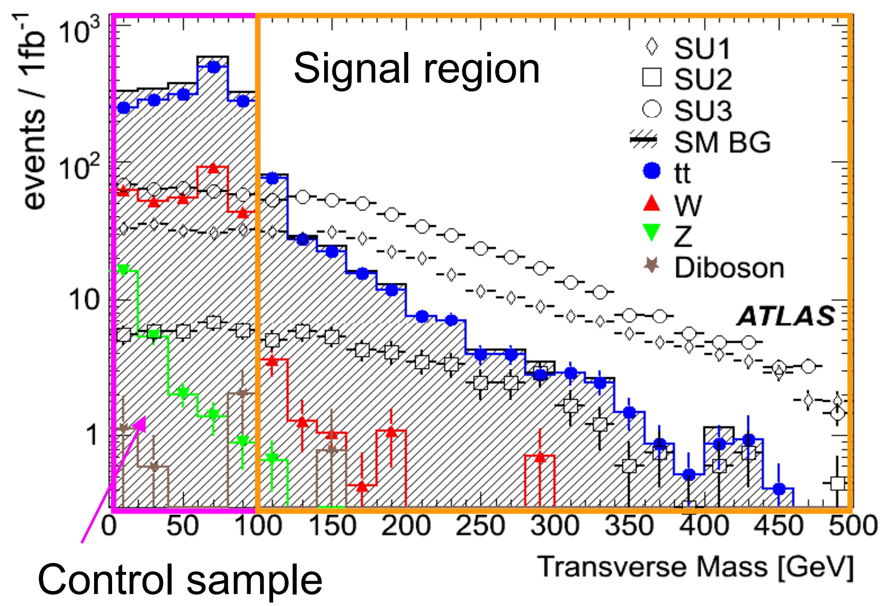
Improved MT Method

- To subtract SUSY contamination from the control sample M_T < 100GeV, signal region M_T > 100GeV times 0.6 is subtracted from the control sample.

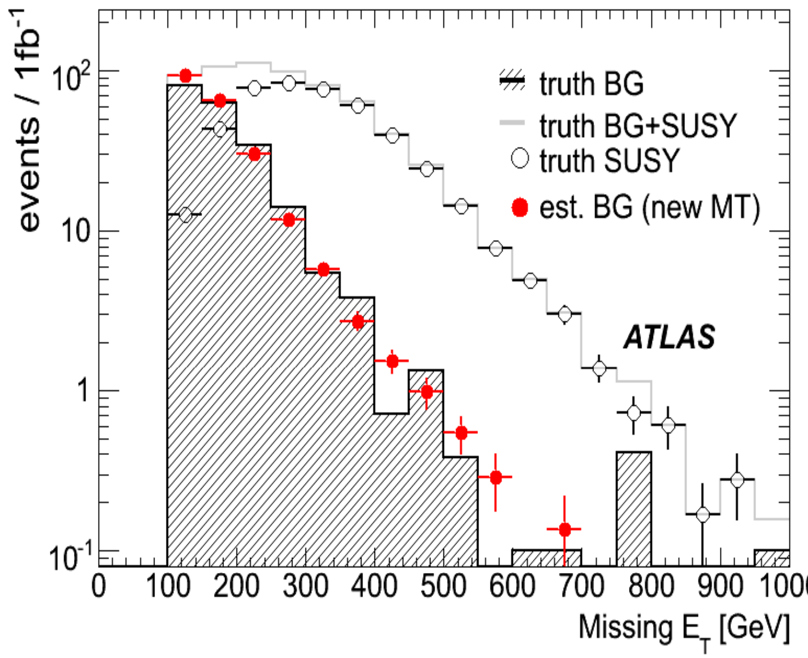
$$\frac{\text{\# of SUSY signal in the control sample } M_T < 100\text{GeV}}{\text{\# of SUSY signal in the signal region } M_T > 100\text{GeV}}$$

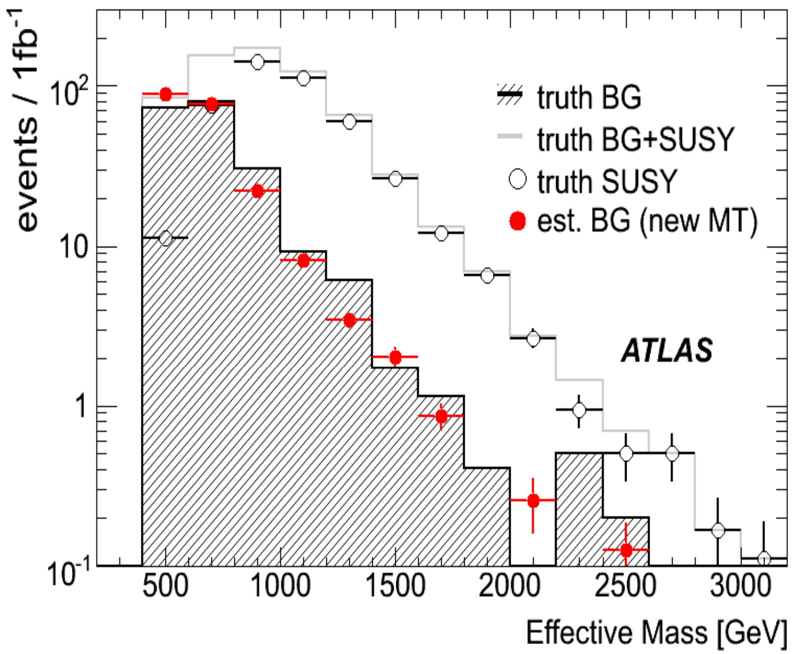
← MC information

- To reduce SUSY contribution in the normalization region, we use the number of events in $\not{E}_T$=100-150GeV.



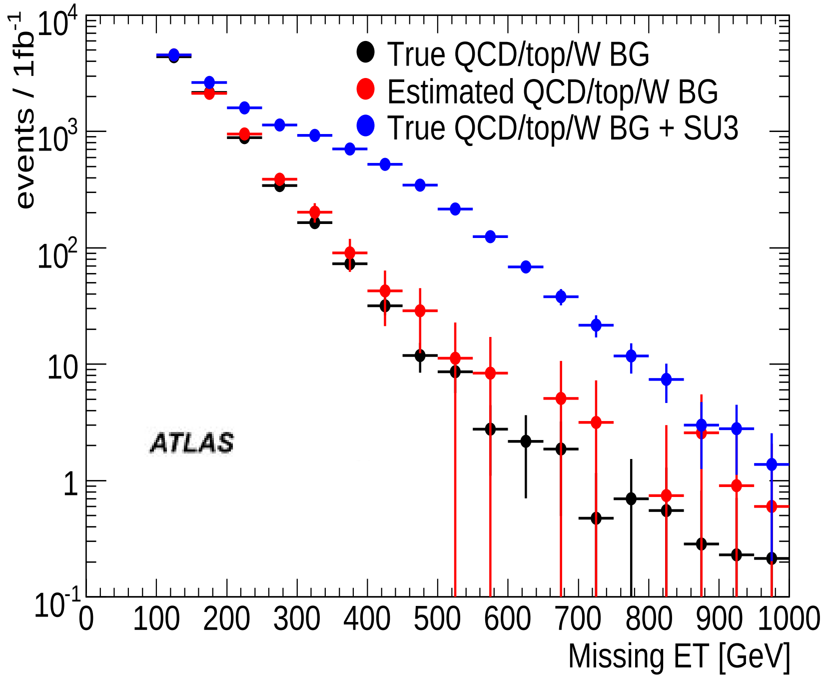
One Lepton Mode – Improved MT-

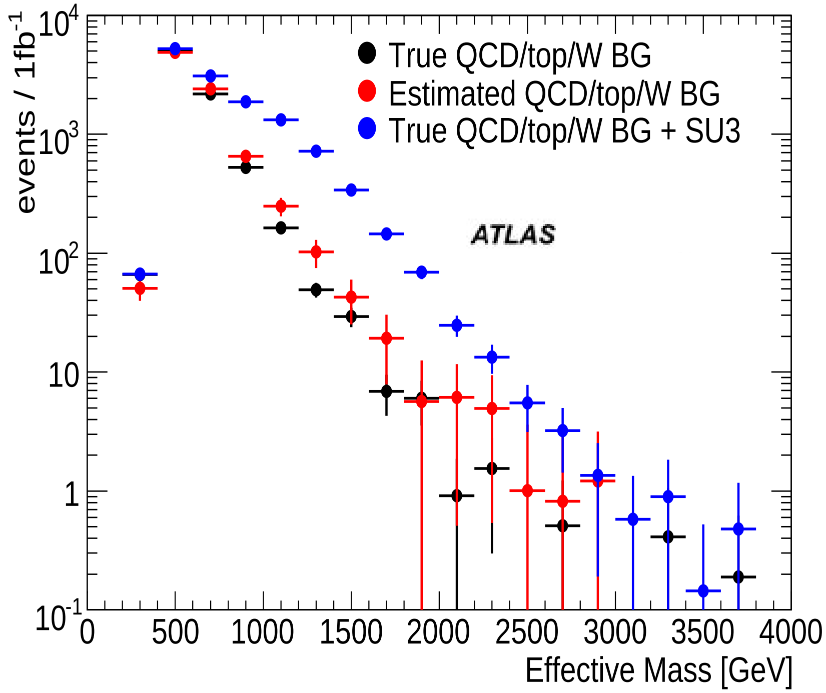




	$\not{E}_T$ > 100GeV	$\not{E}_T$ > 300GeV
True BGs	203+/-6	12.4+/-1.6
Estimated BGs	212+/-11	12.3+/-1.0
True BGs + SUSY	653+/-8	245+/-4

No Lepton Mode - Improved MT -





	mET > 100GeV	mET > 300GeV
True QCD/top/W	8077+/-90	300+/-17
Estimated QCD/top/W	8419+/-405	397+/-67
True QCD/top/W + SUSY	12925+/-114	3000+/-55