

# Searches for SUSY with leptons, jets and missing $E_T$

Carsten Hof

III. Phys. Inst. A, RWTH Aachen  
for the CMS Collaboration

- Introduction
- Discovery Channels
- Parameter Determination
- Conclusion

SPONSORED BY THE

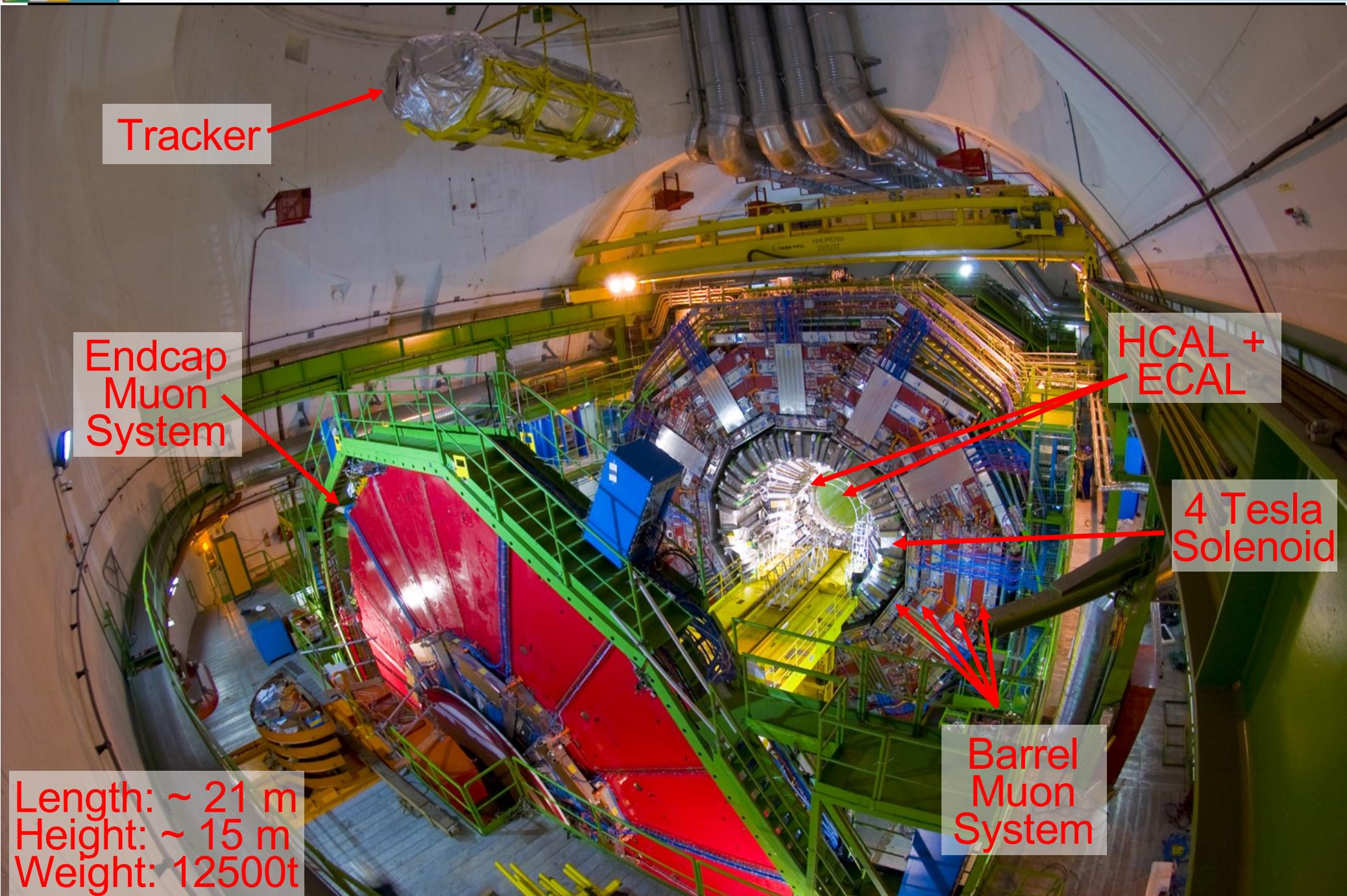


Federal Ministry  
of Education  
and Research

Carsten Hof  
Physics@LHC  
Split, October 2008

RWTHAACHEN  
UNIVERSITY

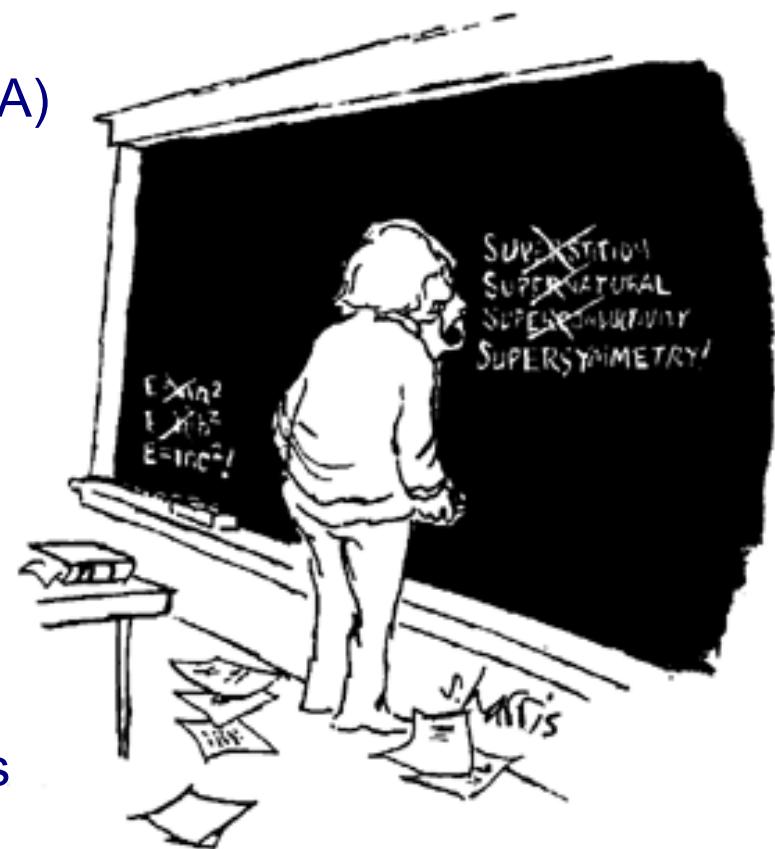
# The CMS Detector



# Setting the Scope ...

## “Standard” Supersymmetry (SUSY)

- ▶ Restrict to minimal supergravity model (mSUGRA)
- ▶ Determined by 5 parameters
  - ◆ Common scalar / gaugino mass  $m_0 / m_{1/2}$
  - ◆ Trilinear coupling  $A_0$
  - ◆ Sign of Higgsino mixing parameter  $\text{sign}(\mu)$
  - ◆ Ratio of the Higgs vacuum expectation values  $\tan \beta$
- ▶ Consider R-parity conservation only  
→ pair-production of sparticles
- ▶ Lightest SUSY particle (LSP) stable  
→ dark matter candidate
- ▶ Hadron collider: squark/gluino production dominant (if not too heavy)



Early analyses ( $L \sim 1\text{fb}^{-1}$  assuming  $E_{\text{CMS}} = 14\text{ TeV}$ )

- ▶ Discovery channels (inclusive)
- ▶ Data-driven background estimation
- ▶ SUSY parameter determination (exclusive)

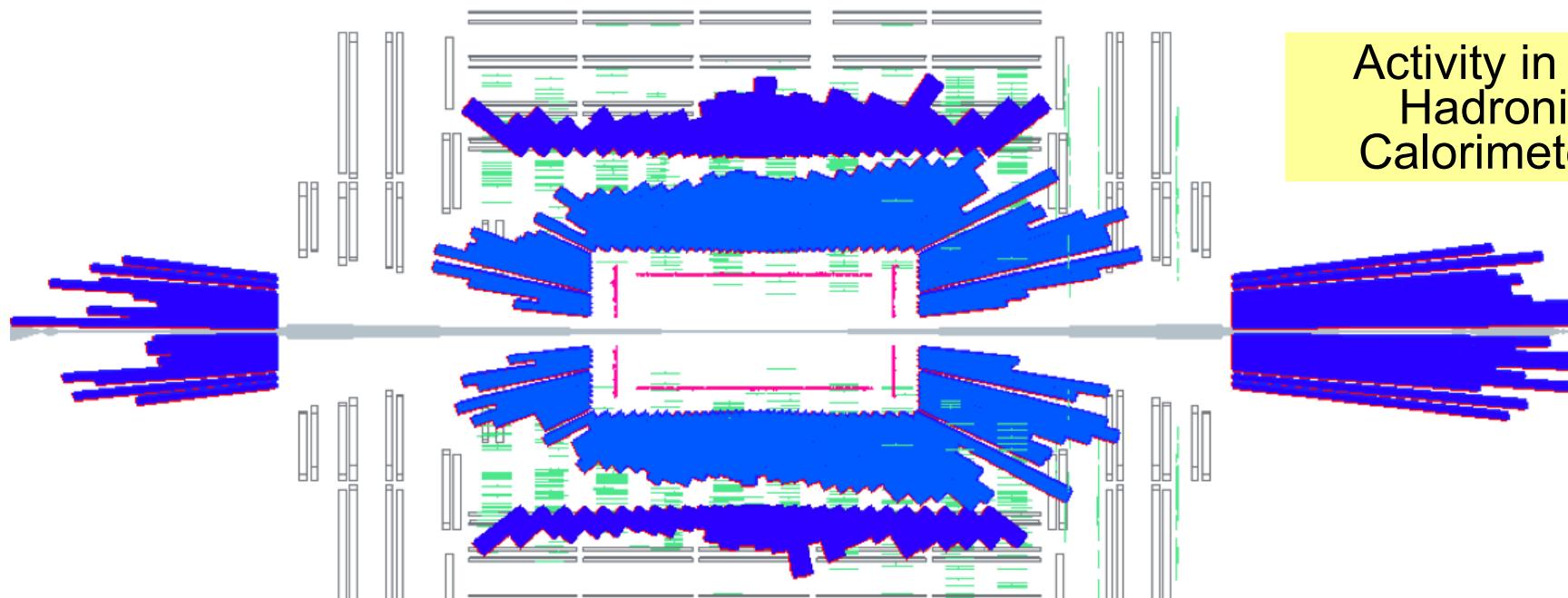


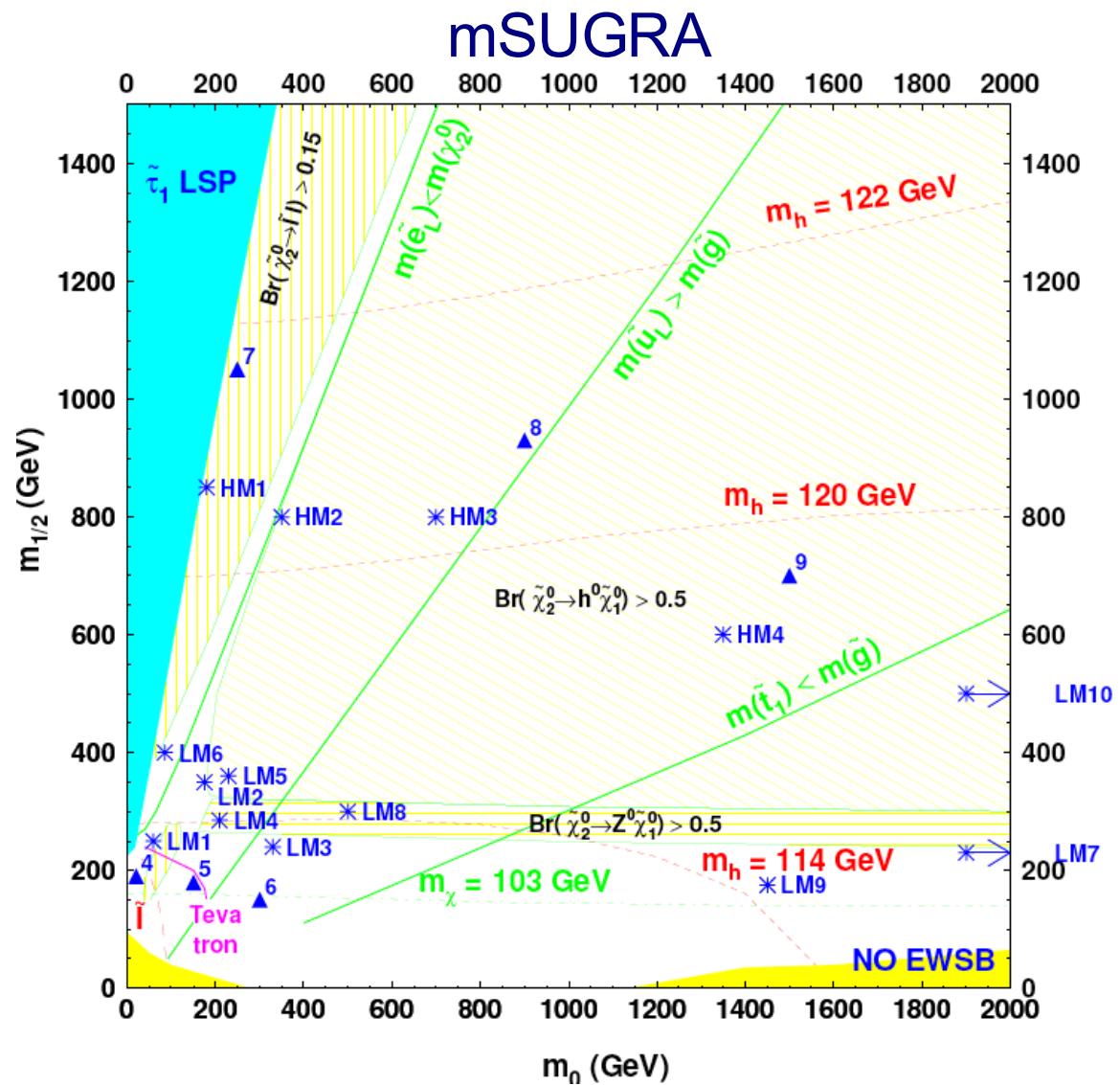
Many Analyses!!!

I'll present a personal selection ...

First LHC Events: Shot on a Collimator 150 m in Front of CMS

Run 62063, Event 1534





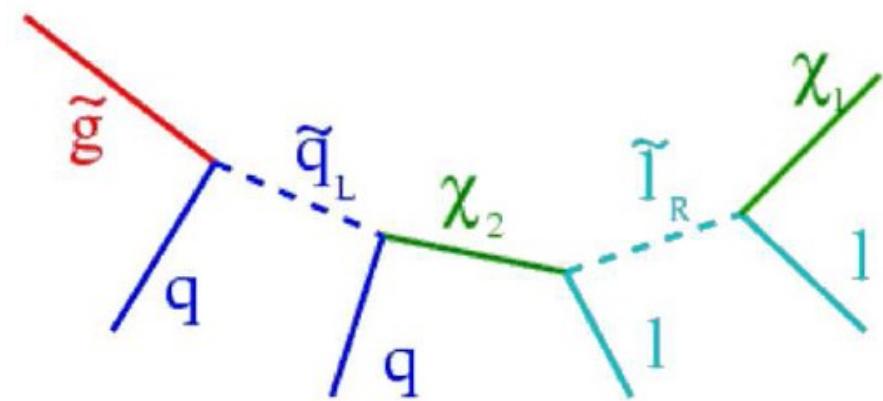
- ▶ LM/HM = Low/High Mass
- ▶ All points are beyond current Tevatron limits
- ▶ High mass points are close to the ultimate LHC reach
- ▶ Cross section depends on masses/parameter space

Point	$m_0$	$m_{1/2}$	$\tan \beta$	$\text{sgn}(\mu)$	$A_0$
LM1	60	250	10	+	0
LM2	185	350	35	+	0
LM3	330	240	20	+	0
LM4	210	285	10	+	0
LM5	230	360	10	+	0
LM6	85	400	10	+	0
LM7	3000	230	10	+	0
LM8	500	300	10	+	-300
LM9	1450	175	50	+	0
LM10	3000	500	10	+	0
HM1	180	850	10	+	0
HM2	350	800	35	+	0
HM3	700	800	10	+	0
HM4	1350	600	10	+	0

# A typical SUSY Event Selection

## Event Selection (Optimize Signal over Background)

- ▶ Large missing  $E_T$  (MET):  $O(> 200\text{GeV})$  ( $\rightarrow$  LSP)  
difficult to measure precisely at start-up
- ▶  $\geq 3$  hard Jets  $O(>100\text{-}200\text{GeV})$  ( $\rightarrow$  decay chains)
- ▶ N isolated leptons (according to investigated topology)  
reduced QCD with growing N
- ▶ Angular cuts for background reduction
- ▶ Further cleanup against e.g. beam halo, cosmics



- ▶ Expected background: W+jets, Z+jets,  $t\bar{t}$ +jets, di-boson, single t, multi-jets (QCD)

# Inclusive Analysis: MET + Jets

## Event Selection

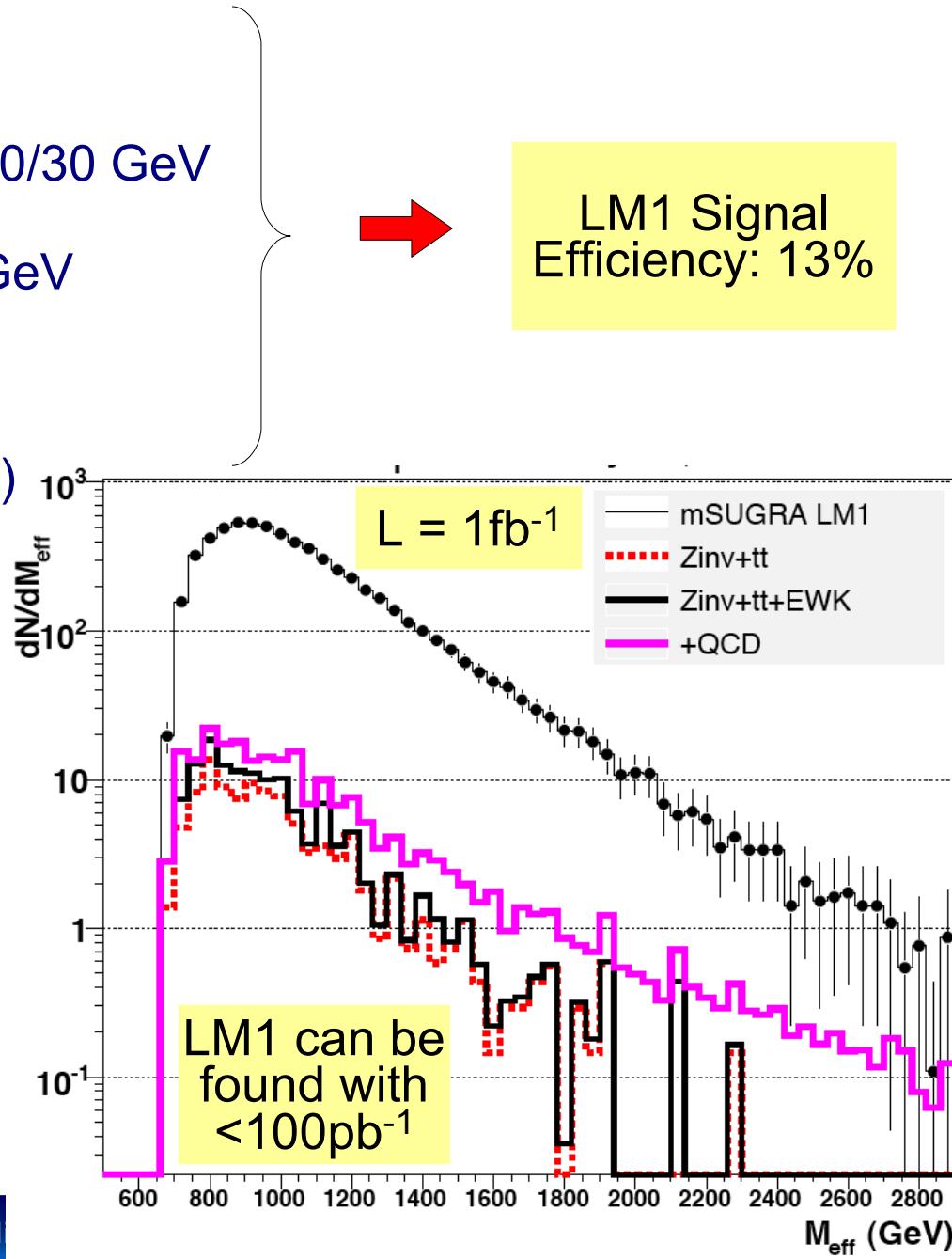
- ▶ MET > 200 GeV
- ▶  $\geq 3$  jets ( $|\eta| < 1.7/3/3$ ) with  $E_T > 180/110/30$  GeV
- ▶  $H_T := E_{T,j2} + E_{T,j3} + E_{T,j4} + \text{MET} > 500$  GeV
- ▶ Indirect lepton veto
- ▶ Cleanup and QCD rejection (next slide)

LM1 Signal  
Efficiency: 13%

## Main Backgrounds

- ▶ QCD: MET via mismeasurements & jet resolution
- ▶  $Z \rightarrow \nu\nu$ : irreducible
- ▶  $t\bar{t}$ : hadronic or misidentified lepton(s)
- ▶  $W$ : hadronic or misidentified lepton

Determine background from data!!!



## QCD Reduction

- ▶ MET in QCD: mismeasured jets
- ▶ Suppressed via topological cuts:

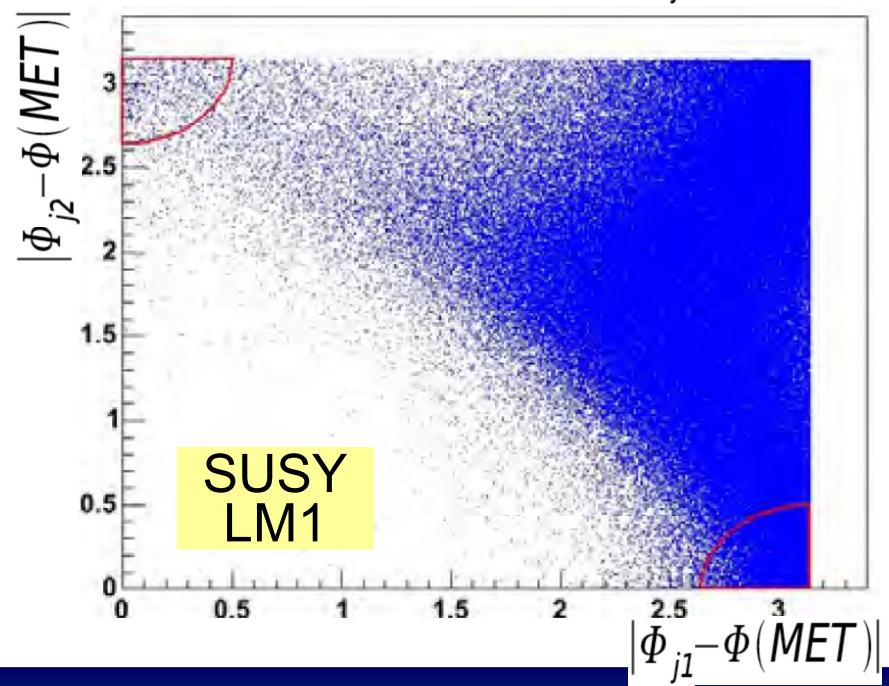
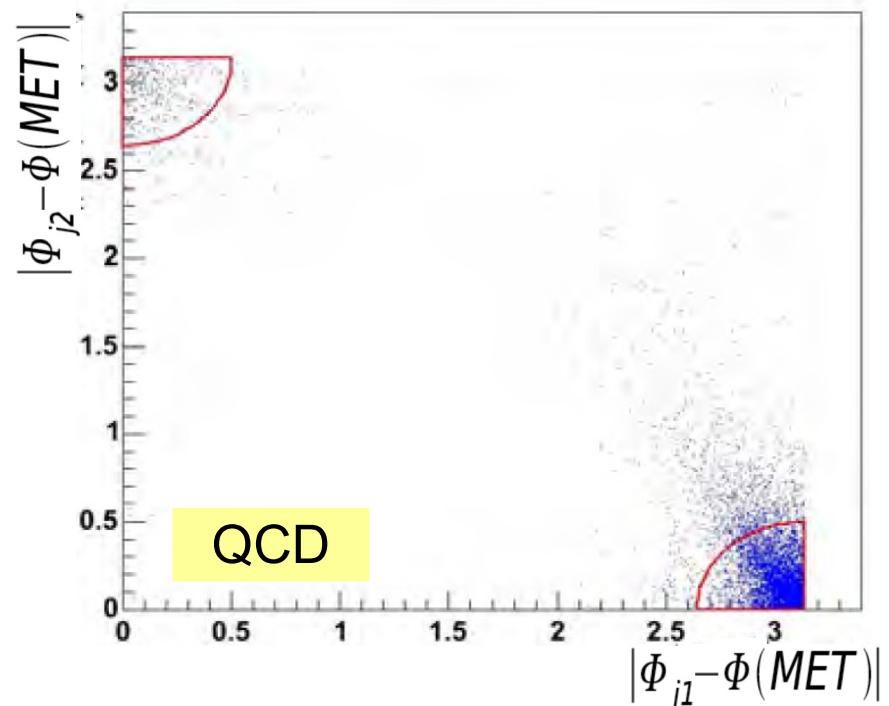
$$R_{1,2} = \sqrt{\Delta\Phi_{1,2}^2 + (\pi - \Phi_{2,1})^2} > 0.5$$

with  $\Delta\Phi_{1,2} = |\Phi_{j1,j2} - \Phi(MET)|$

- ▶ MET back-to-back to jet

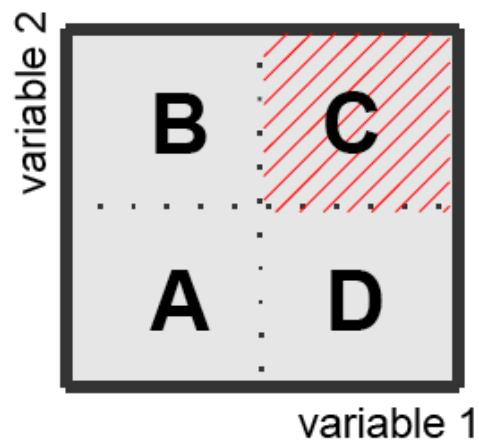
## MET Cleaning, Beam Halo, Noise, Cosmics

- ▶  $\geq 1$  primary vertex
- ▶ Event electro-magnetic fraction: 
$$\frac{\sum_{jets} p_T^j \cdot EMF^j}{\sum_{jets} p_T^j} > 0.1$$
- ▶ Event charged fraction: 
$$\frac{\sum_{tracks} p_T^t}{p_T^{jet}} > 0.175$$



- ▶ Previous studies: Estimate mainly via MC e.g. cut factorisation
- ▶ Currently: Many data-driven methods are explored/developed
- ▶ Leptonic channels: relax/invert isolation ( $e, \mu$ ) or ID variables ( $e$ )
- ▶ Hadronic channels:

### “ABCD” method



- ▶ Idea:  
IF variable 1 and variable 2 are uncorrelated:

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

- ▶ Avoid signal contamination in A,B,D
- ▶ Further studies: Correlated variables

### Jet smearing

- ▶ Idea:  
MET in QCD = mismeasured jets
- ▶ Use well-measured QCD spectrum (e.g. low MET,  $\gamma + \text{jet}$ )
- ▶ Use smearing function (data or MC) to extrapolate well-measured QCD to QCD with high MET

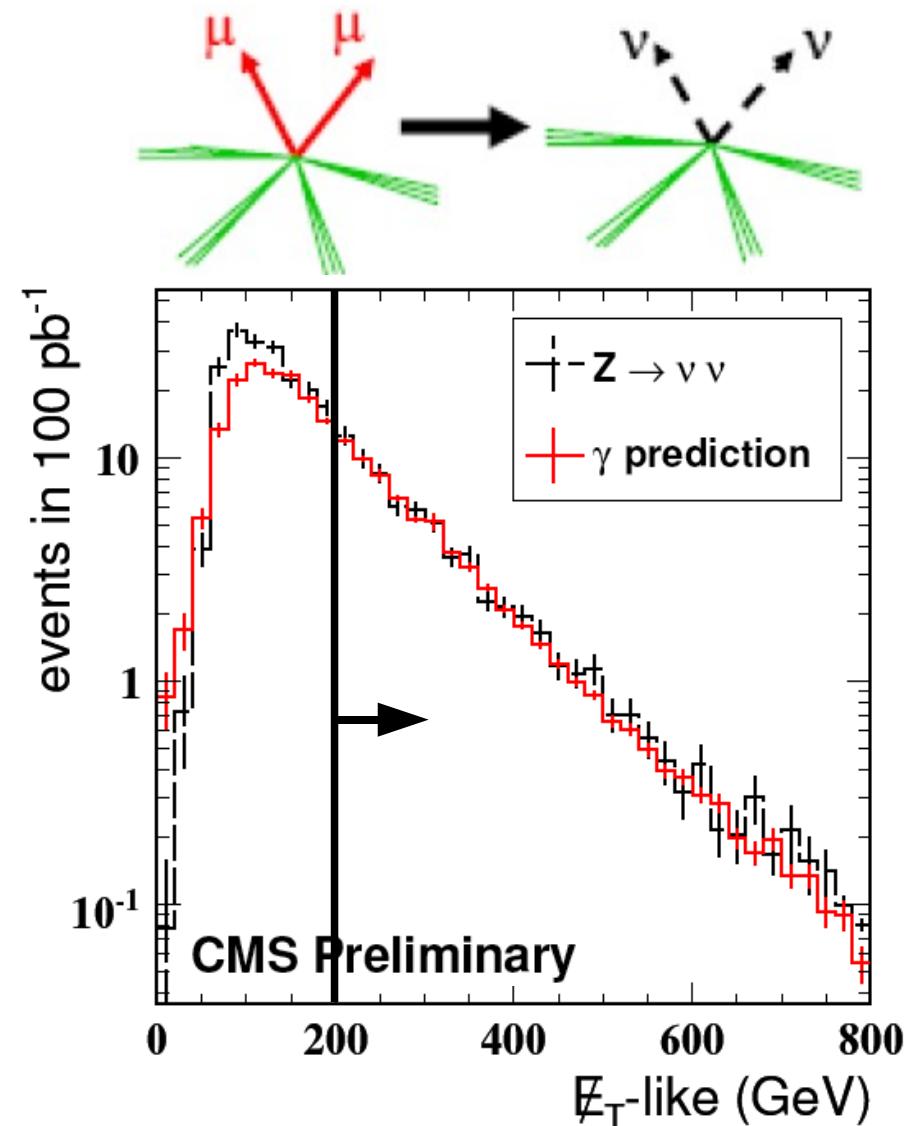
# Irreducible Background: $Z \rightarrow \nu\nu$

## Data-driven Estimation I ( $Z+jets$ )

- ▶ “Standard Candle”: use  $Z \rightarrow \mu\mu$  data
  - ◆ Replace leptons by neutrinos
  - ◆ Total uncertainty  $\sim 20\%$  for  $1\text{fb}^{-1}$   
stat. limited  $\text{Br}(Z \rightarrow \mu\mu) = 1/6 \text{ Br}(Z \rightarrow \nu\nu)$

## Data-driven Estimation II ( $W,\gamma+jets$ )

- ▶ Approximation:  
bosonic events at high  $p_T$  have  
similar event shapes  $\rightarrow$  use  $\gamma, W+jets$
- ▶ Gain in statistics ( $\rightarrow 100\text{pb}^{-1}$  analysis)  
 $\sigma(Z+2\text{jets}) \approx 3\sigma(W+2\text{jets}) \approx 0.8\sigma(\gamma+2\text{jets})$
- ▶ Complementary
  - ◆ Other triggers
  - ◆ Different bkg/signal contamination



$Z \rightarrow \nu\nu$ background estimate ( $100 \text{ pb}^{-1}$ )	
MC-truth	35
From $\gamma+jets$	$29 \pm 3 \text{ (stat)} \pm 5 \text{ (sys)}$
From $W+jets$	$35 \pm 10 \text{ (stat)} \pm 8 \text{ (sys)} \pm 3 \text{ (theory)}$

# Di-Jet Analysis

Robust extension of full-hadronic search

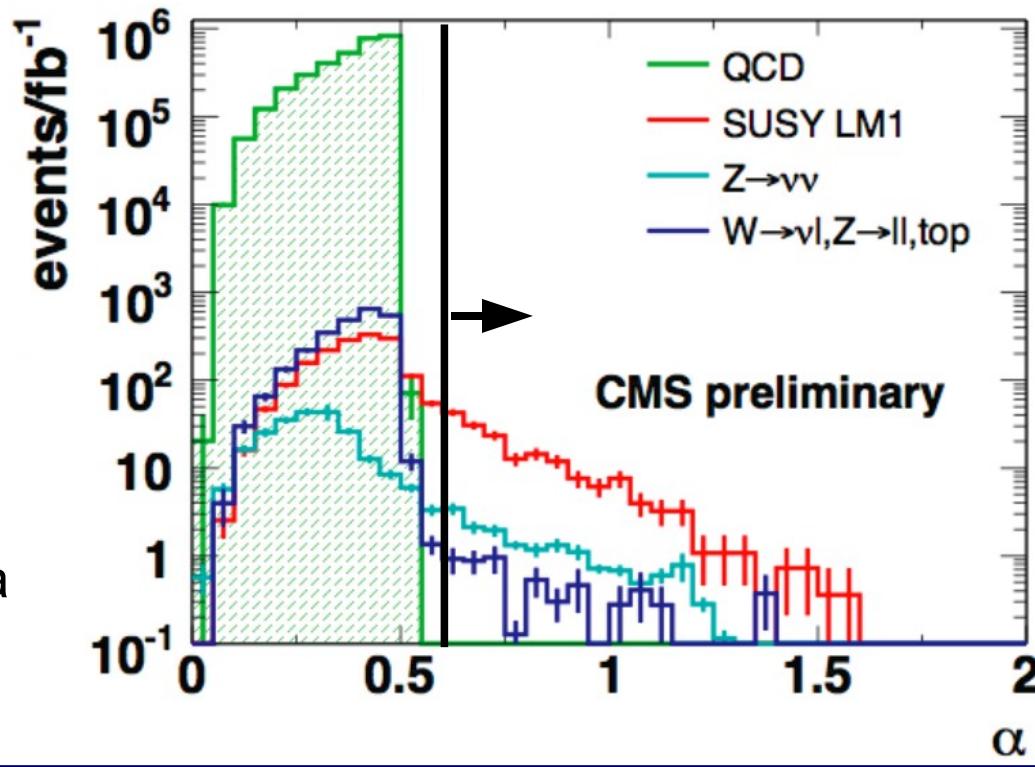
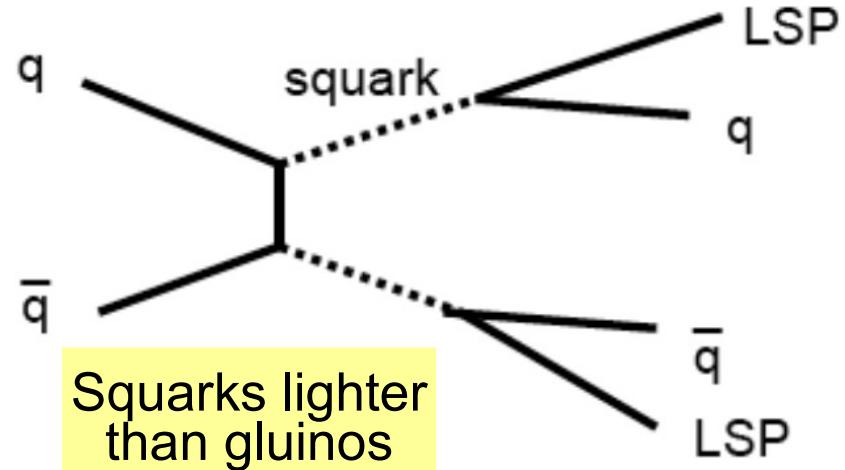
## Event Selection

- ▶ = 2 jets with  $P_T > 50$  GeV, lepton veto
- ▶  $H_T := P_{T,j1} + P_{T,j2} > 500$  GeV
- ▶ Angular/acceptance cuts for cleaning
- ▶ New variable (Randall/Tucker-Smith):

$$\alpha = \frac{E_{T,j2}}{M_{j1j2}} = \frac{E_{T,j2}}{\sqrt{2E_1E_2(1-\cos\theta)}} > 0.55$$

## Features:

- ▶ Robust: calorimetric MET not used
- ▶ Dominant Backgrounds
  - ◆  $Z$  invisible, QCD: estimate from data
- ▶ **LM1 discovery within  $100\text{pb}^{-1}$**



Event Selection

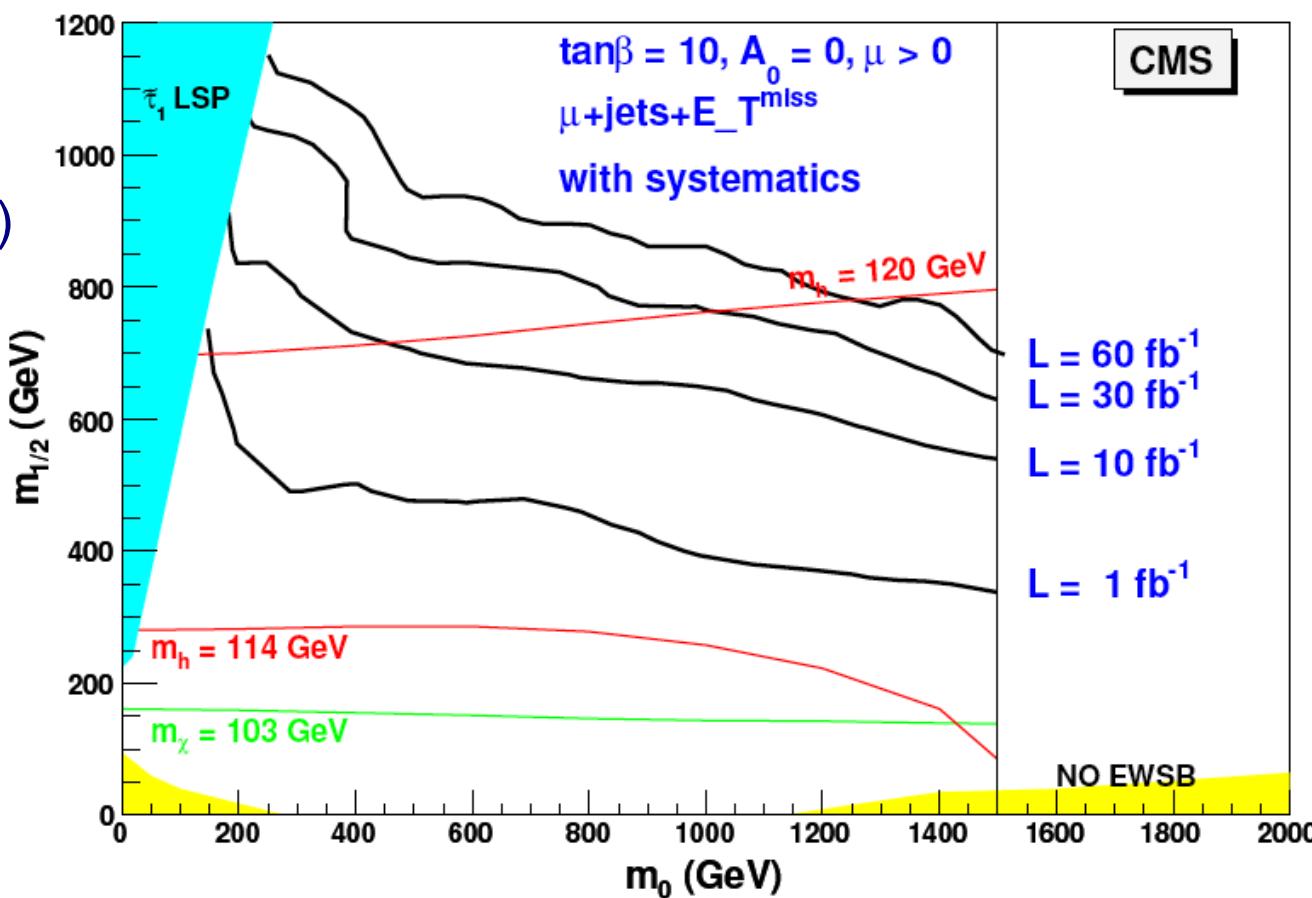
- ▶  $\geq 1$  isolated muon with  $P_T > 30$  GeV
- ▶ 1<sup>st</sup>, 2<sup>nd</sup> jet  $E_T > 440$  GeV, 3<sup>rd</sup> jet  $E_T > 50$  GeV  
within optimized pseudorapidity
- ▶ MET  $> 130$  GeV
- ▶ single/di-muon trigger
- ▶ Angular cuts  $\Delta\phi(\text{jets, MET})$   
for MET cleaning

Cut optimization  
via genetic  
algorithm on 10fb<sup>-1</sup>

Efficiency

- ▶ LM1: 300 events (0.07%)
- ▶ Bkg: 3 events  
(uncertainty:  $\sim 20\%$ )

→ Early SUSY  
discovery possible!



# Same sign di- $\mu$ + MET + Jets

## Event Selection

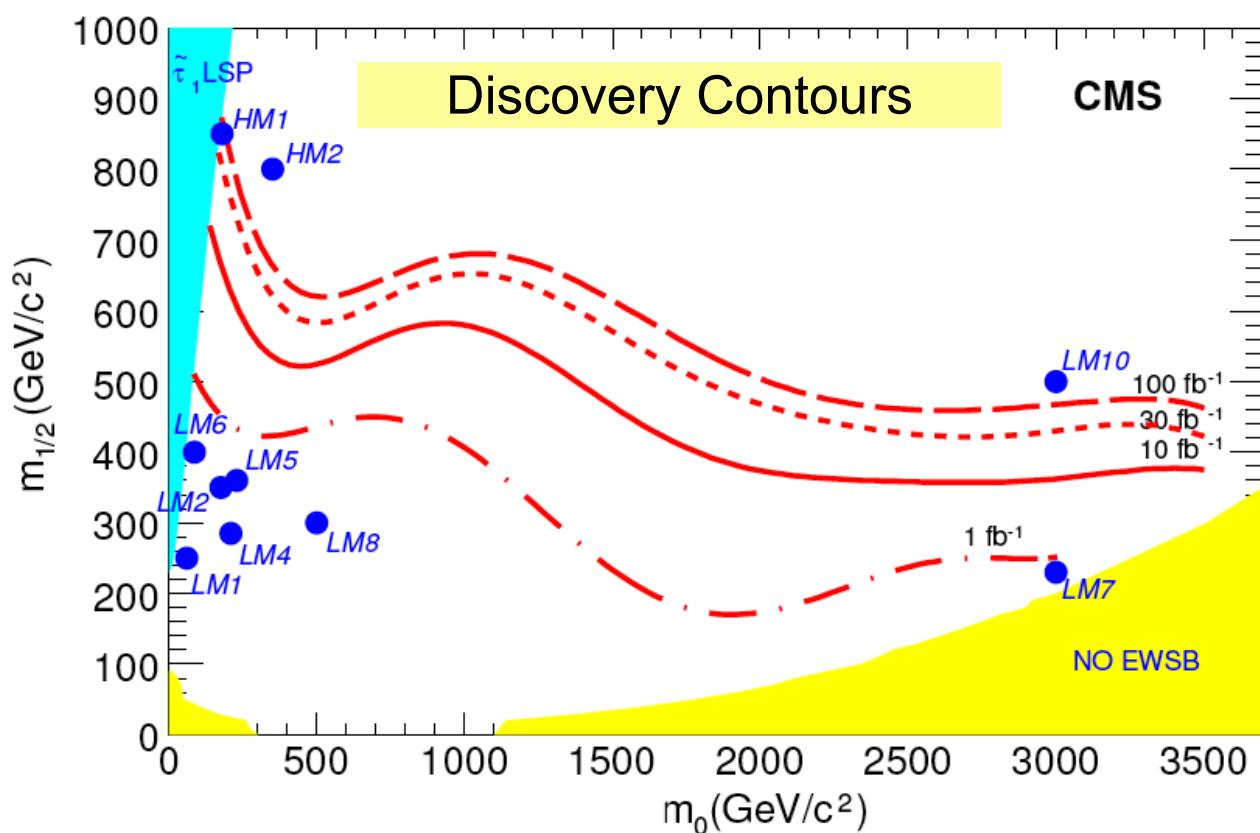
- ▶  $\geq 2$  isolated muons ( $P_T > 10$  GeV, same sign)
- ▶  $\geq 3$  jets with  $E_T > 175/130/55$  GeV
- ▶ MET  $> 200$  GeV

Cut optimization  
via genetic  
algorithm on  $10\text{fb}^{-1}$

## Result

- ▶ Almost no SM background expected
- ▶ Utilization of muon trigger:
  - ◆ Expected to be robust at LHC start-up

→ Complementary to hadronic channel



# SUSY Parameter Estimation

No mass peaks due to LSP/MET → measure shapes and endpoints!

Important example:  $\tilde{\chi}_2^0 \rightarrow \ell^\pm \ell^\mp \tilde{\chi}_1^0$

Opposite Sign, Same flavor

► Endpoint:  $m_{\ell\ell}^{max} = m_{\tilde{\chi}_2^0} \sqrt{1 - \frac{m_{\tilde{\ell}_R}^2}{m_{\tilde{\chi}_2^0}^2}} \sqrt{1 - \frac{m_{\tilde{\chi}_1^0}^2}{m_{\tilde{\ell}_R}^2}}$

## Data-driven background estimate

- tt and diboson background from eμ data ( $Br(ee) = \frac{1}{2} Br(\mu e)$ )

## Unbinned fit to data (7 parameter)

$$F(m) = N_{sig}S(m) + N_{bkg}B(m) + N_Z Z(m)$$

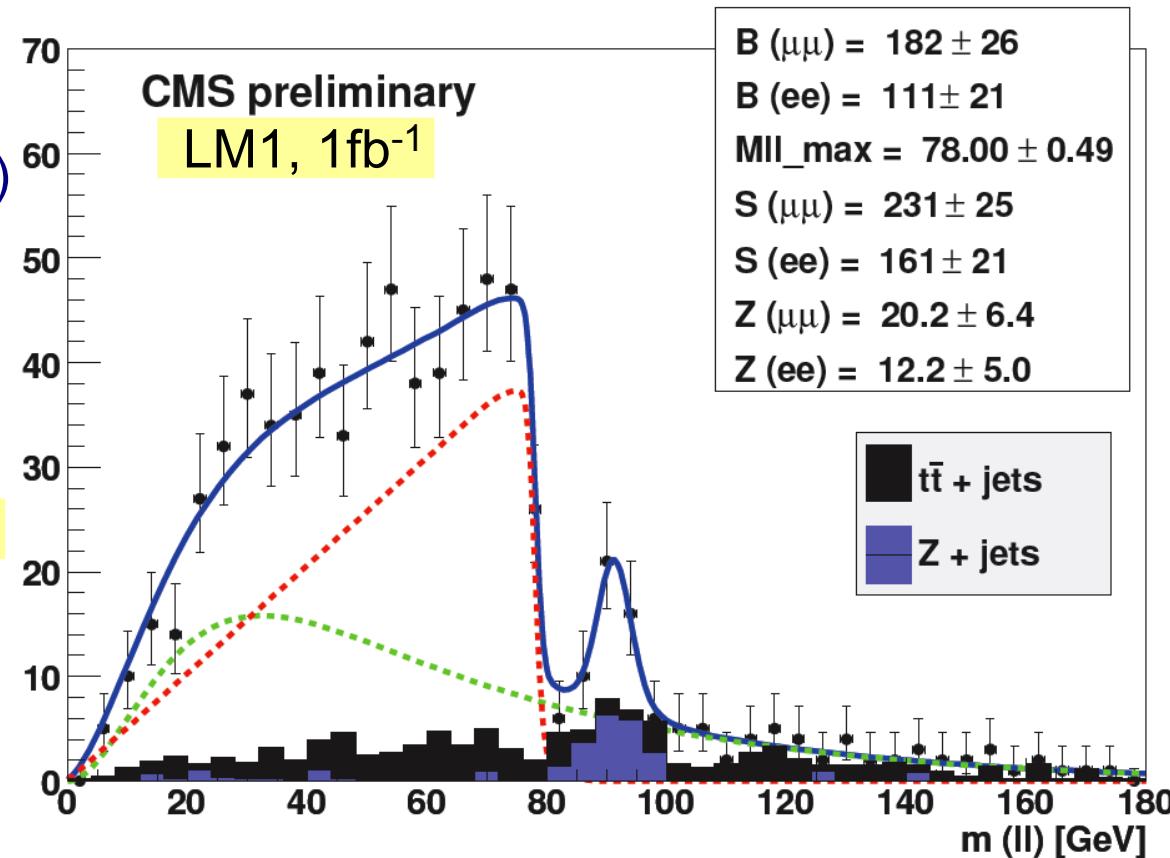
Signal Model

Bkg from data

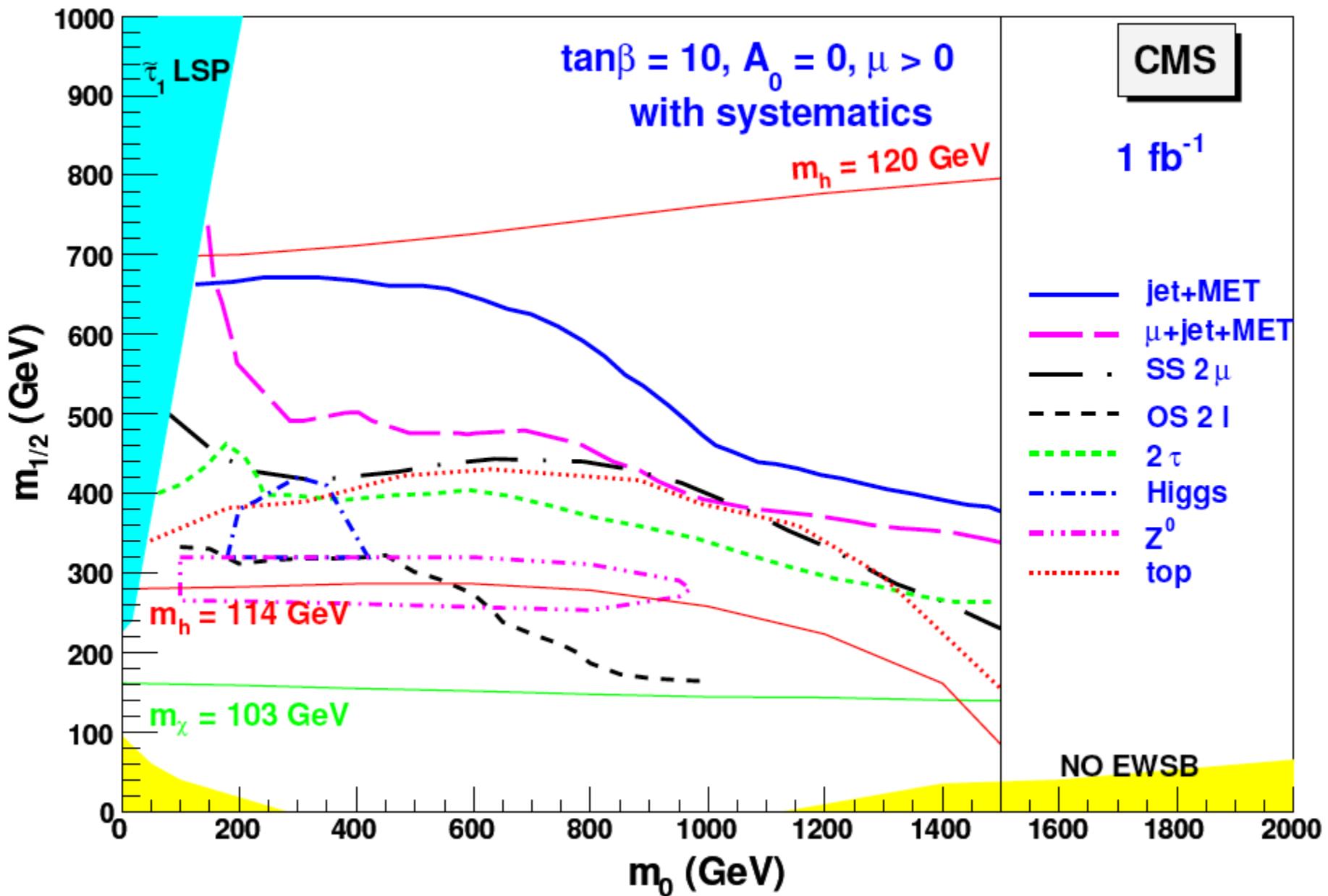
Z peak

→  $\Delta m_{ee}^{max} = \pm 1.07(stat.) \pm 0.36(syst.) GeV$

$\Delta m_{\mu\mu}^{max} = \pm 0.75(stat.) \pm 0.18(syst.) GeV$

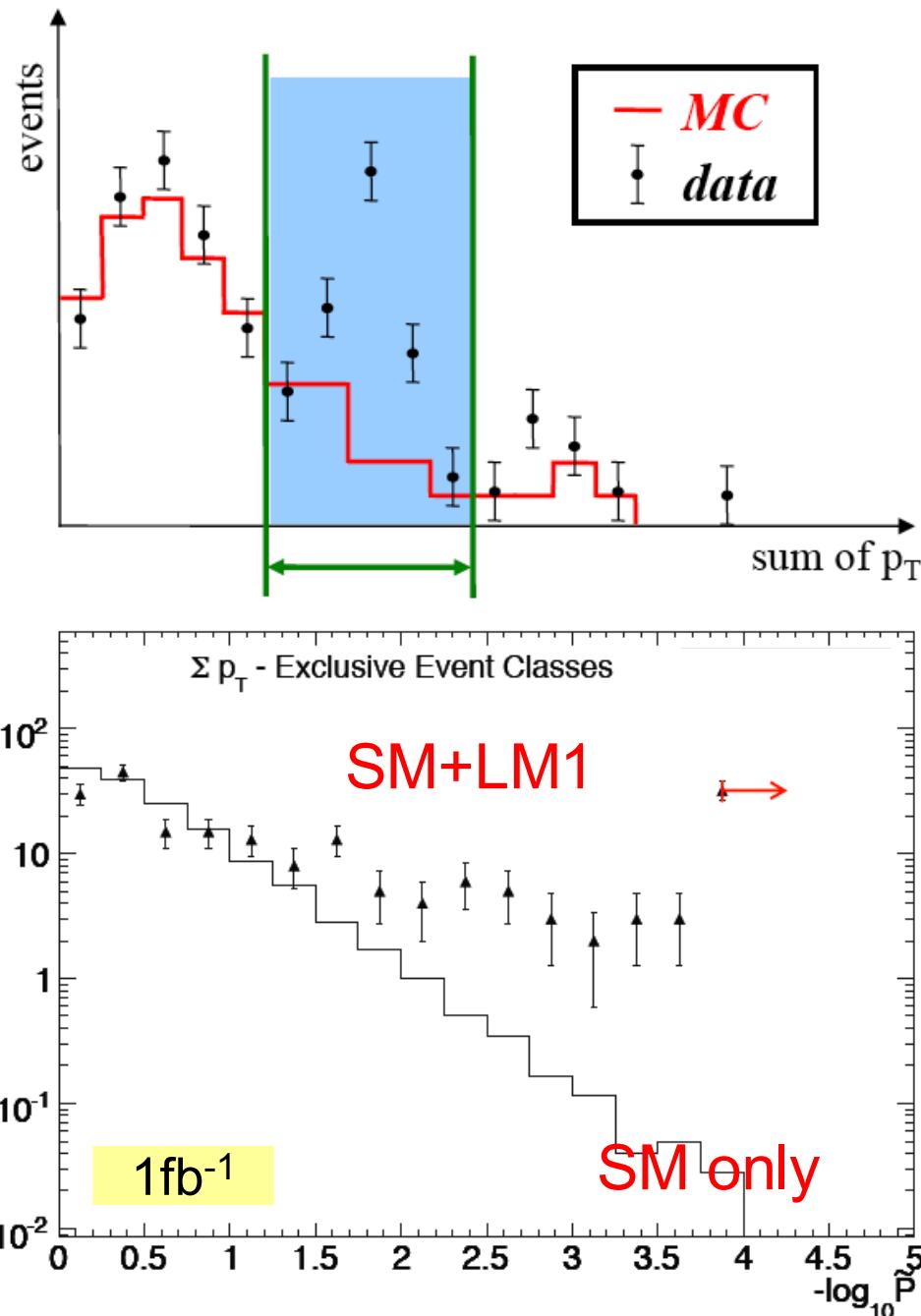


# Overview: Expected Discovery Reach



# Alternative Approach: MUSiC

- ▶ Automated global data/MC comparison
- ▶ Alarm system for deviations
- ▶ Classify events by particle content
  - ◆ Exclusive & inclusive final states (~ 300 classes each)
  - ◆  $e, \mu, \gamma, \text{jet}, \text{MET}$
- ▶ Scan distributions ( $\sum p_T, M_{\text{inv}}, \text{MET}$ ) for statistically significant deviations
  - ◆ Dedicated algorithm searching biggest discrepancy
  - ◆ Includes systematic uncertainties
- ▶ A priori sensitive to MC shortcomings, detector effects, later: new physics
- ▶ If SUSY is realised: many alarms!



- ▶ Signatures of gluinos and squarks ( $m < 1\text{TeV}$ ) can be discovered within  $1\text{fb}^{-1}$
- ▶ If and only if the backgrounds are understood
- ▶ Data-driven methods to ensure this are currently in the analysis focus
- ▶ Many exclusive measurements needed to determine SUSY parameters

## „Master Plan“:

- ▶ Get first collisions
- ▶ Commission & understand detector (already started!)
- ▶ Rediscover Standard Model
- ▶ Discover SUSY
- ▶ Determine its parameters

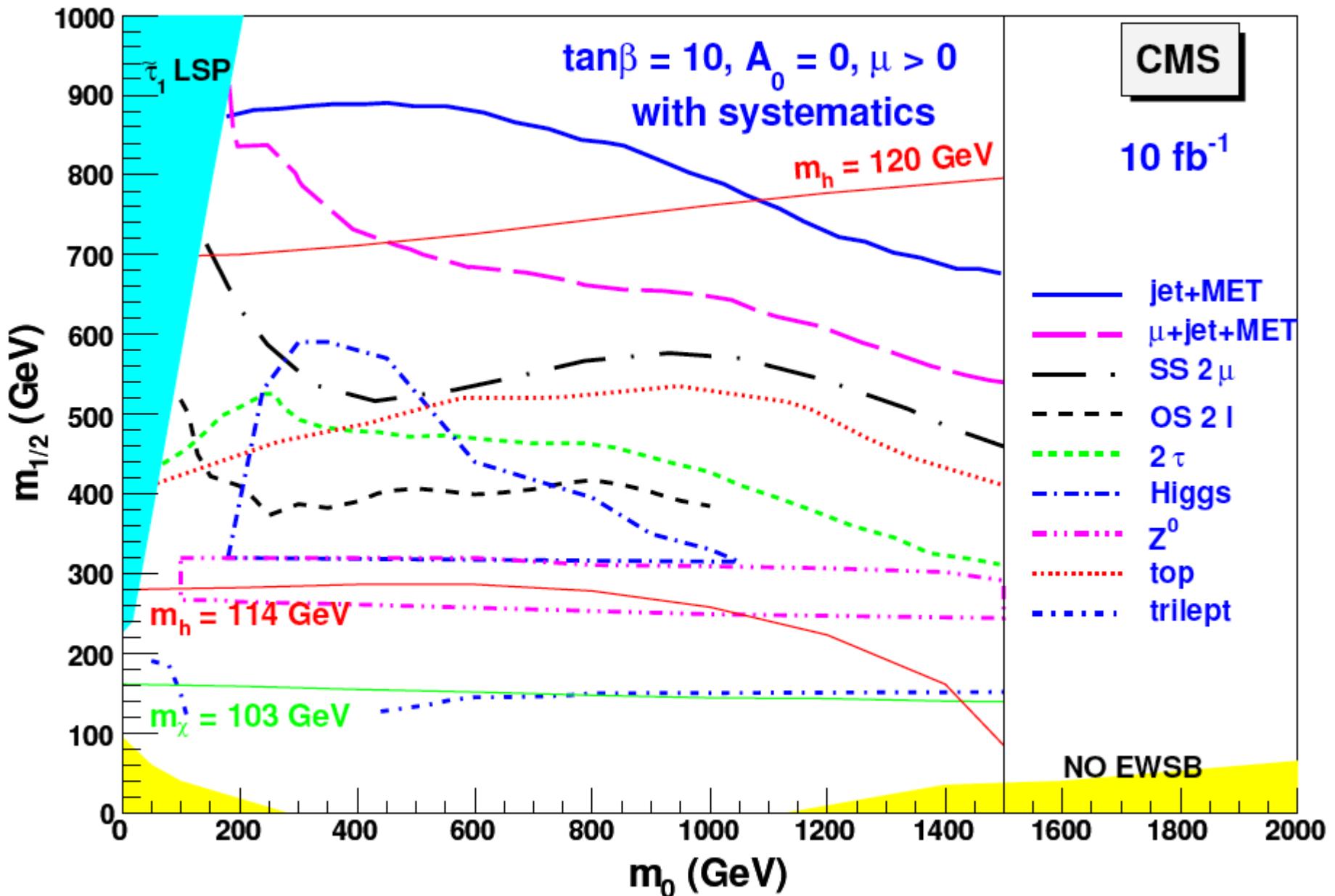


**Thanks for your attention!**

*Backup*

- ▶ GARCON = Genetic Algorithm for Rectangular Cut Optimisation
- ▶ Problem: optimization of signal vs background in multi-dimensional cut space
  - ◆ CPU intensive if done in great detail
- ▶ GARCON: genetic algorithm
  - ◆ Optimization by trying of effectively  $\sim 10^{50}$  cut set parameters
  - ◆ Timing: some hours with order of million of input events
  - ◆ Verification of cut stability
- ▶ Idea:
  - ◆ Define upper/lower bound on variable to be cut on (=Individual), all variables with lower/upper bound = Community
  - ◆ Define quality function (e.g.  $S/\sqrt{B}$ ) --> good/bad evolution
  - ◆ Each individual involved in evolution:
    - New individuals created by combining two parent individuals
    - High quality individuals survive while others die
    - Mutation & Cataclysmic updates to avoid local high quality communities

# Overview: Expected Discovery Reach



# Details: LM Points

LM Point	LM1	LM2	LM3	LM4	LM5	LM6	LM7	LM8	LM9	LM10	LM11	LM9p	LM9t175
m0	60	185	330	210	230	85	3000	500	1450	3000	250	1450	1450
m12	250	350	240	285	360	400	230	300	175	500	325	230	175
tanb	10	35	20	10	10	10	10	10	50	10	35	50	50
sm	1	1	1	1	1	1	1	1	1	1	1	1	1
A0	0	0	0	0	0	0	0	-300	0	0	0	0	0
LO (pb)	46	7,8	35	21	6,4	4,3	7,4	9,4	32	0,16	10	10	31
NLO (pb)	61	11	49	28	8,7	5,7	11	13	52	0,23	14	16	51
h	109	113	110	110	112	113	118	113	114	119	112	114	115
H	372	434	450	466	567	579	2937	689	488	2999	427	534	616
A	372	434	450	466	567	579	2937	689	488	2999	427	534	616
H+	381	442	457	473	573	585	2939	694	495	3001	435	540	622
d L	560	777	626	660	807	857	2997	817	1478	3119	748	1510	1478
d R	536	746	607	635	776	820	3000	796	1477	3112	720	1506	1477
u L	552	770	619	653	800	850	2990	811	1474	3111	741	1506	1475
u R	540	752	610	640	782	828	2999	800	1477	3114	725	1507	1477
s L	560	777	626	660	807	857	2997	817	1478	3119	748	1510	1478
s R	536	746	607	635	776	820	3000	796	1477	3112	720	1506	1477
c L	552	770	619	653	800	850	2990	811	1474	3111	741	1506	1475
c R	540	752	610	640	782	828	2999	800	1477	3114	725	1507	1477
b 1	510	671	548	598	734	785	2448	710	1008	2577	642	1052	1009
b 2	536	724	596	632	771	816	2976	789	1124	3087	695	1168	1148
t 1	407	580	444	481	599	647	1788	544	882	1918	551	918	859
t 2	580	748	606	658	787	839	2453	760	1024	2582	717	1070	1026

Crosssection: Prospino2 (CTEQ6), Mass spectrum: SoftSusy

# Details: LM Points

LM Point	LM1	LM2	LM3	LM4	LM5	LM6	LM7	LM8	LM9	LM10	LM11	LM9p	LM9t175
m0	60	185	330	210	230	85	3000	500	1450	3000	250	1450	1450
m12	250	350	240	285	360	400	230	300	175	500	325	230	175
tanb	10	35	20	10	10	10	10	10	50	10	35	50	50
sm	1	1	1	1	1	1	1	1	1	1	1	1	1
A0	0	0	0	0	0	0	0	-300	0	0	0	0	0
LO (pb)	46	7,8	35	21	6,4	4,3	7,4	9,4	32	0,16	10	10	31
NLO (pb)	61	11	49	28	8,7	5,7	11	13	52	0,23	14	16	51
e L	186	304	370	289	338	287	2994	539	1450	3009	336	1454	1451
e R	120	231	344	239	270	178	2998	514	1450	3002	281	1451	1450
nu e	167	292	360	276	327	275	2992	533	1447	3006	325	1451	1447
mu L	186	304	370	289	338	287	2994	539	1450	3009	336	1454	1451
mu R	120	231	344	239	270	178	2998	514	1450	3002	281	1451	1450
nu mu	167	292	360	276	327	275	2992	533	1447	3006	325	1451	1447
tau 1	111	156	323	233	264	171	2973	506	1054	2978	213	1054	1055
tau 2	190	314	370	291	339	289	2982	539	1267	2996	339	1270	1268
nu tau	167	279	354	275	326	274	2980	530	1265	2994	309	1269	1266
g	603	827	597	687	851	932	637	738	488	1260	776	618	488
neu 1	96	141	94	112	144	161	94	120	65	209	130	91	70
neu 2	178	264	173	208	271	303	177	228	110	359	244	159	133
neu 3	344	455	330	387	475	521	348	458	184	421	426	247	316
neu 4	363	468	347	405	490	535	366	470	225	481	440	281	328
chg 1	179	267	174	210	273	305	179	230	107	360	246	158	134
chg 2	360	465	345	402	487	531	364	467	223	479	437	279	328

Crosssection: Prospino2 (CTEQ6), Mass spectrum: SoftSusy



# MUSiC Timeline



## Focus with first data:

- ▶ Understand the detector, tune the MC generators, re-establish the SM

## After initial problems:

- ▶ Higher order effects in tails, compare e.g. PYTHIA  $\leftrightarrow$  ALPGEN
- ▶ MUSiC can contribute to all these points, gives global picture of data-MC comparison
- ▶ MUSiC as global physics monitor, synergy with specific analyses

## Confidence in detector and MC:

- ▶ Start looking for deviations from the SM, possible signals not covered by specific analyses yet

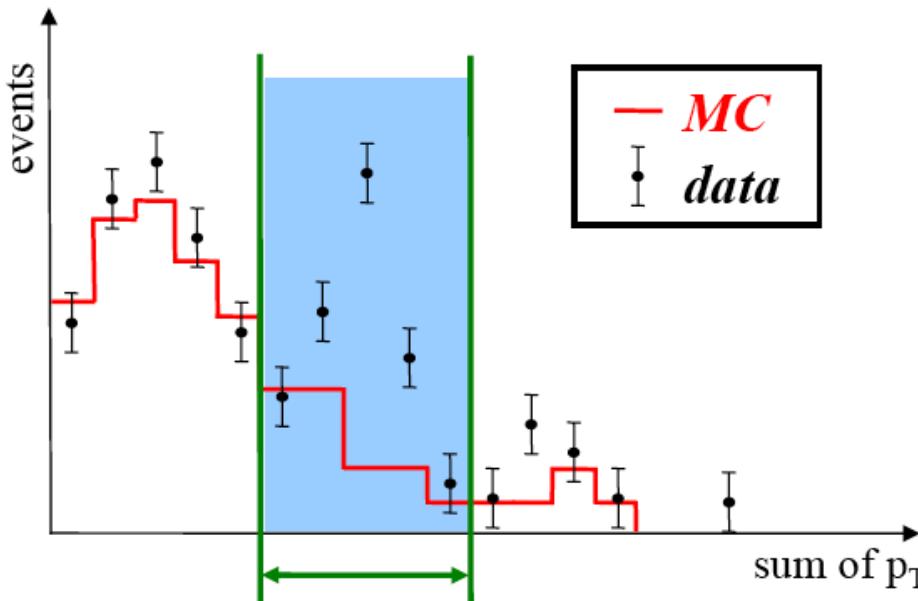
# Setup and Selection

- ▶ “1 fb<sup>-1</sup>” analysis: using CSA07 MC samples with 100 pb<sup>-1</sup> align/calib
- ▶ Standard Model backgrounds:
  - ◆ CSA07 soups, SUSYBSM HLT skims (electron and muon)
  - ◆ Plus high  $\hat{p}_T > 300$  GeV bins for W+jets and Z+jets, WW/WZ/ZZ
- ▶ General strategy: Robustness
  - ◆ Official reco-tools (ElectronID, isolation-tool, JES,...)
  - ◆ Standard objects with standard efficiencies, cuts etc. (→synergy)
  - ◆ Focus on well-understood objects, even if statistics lost

	e/γ	μ	ItCone Jet (ΔR = 0.5)	MET
p <sub>T</sub> cut	30 GeV	30 GeV	60 GeV	100 GeV
η  cut	2.5	2.1	2.5	-
isolation	tracks	tracks	-	-

→ high p<sub>T</sub>, central η  
plus several  
quality criteria  
(N<sub>hits</sub>, χ<sup>2</sup>, ...)

- ▶ Define all possible connected regions in every distribution

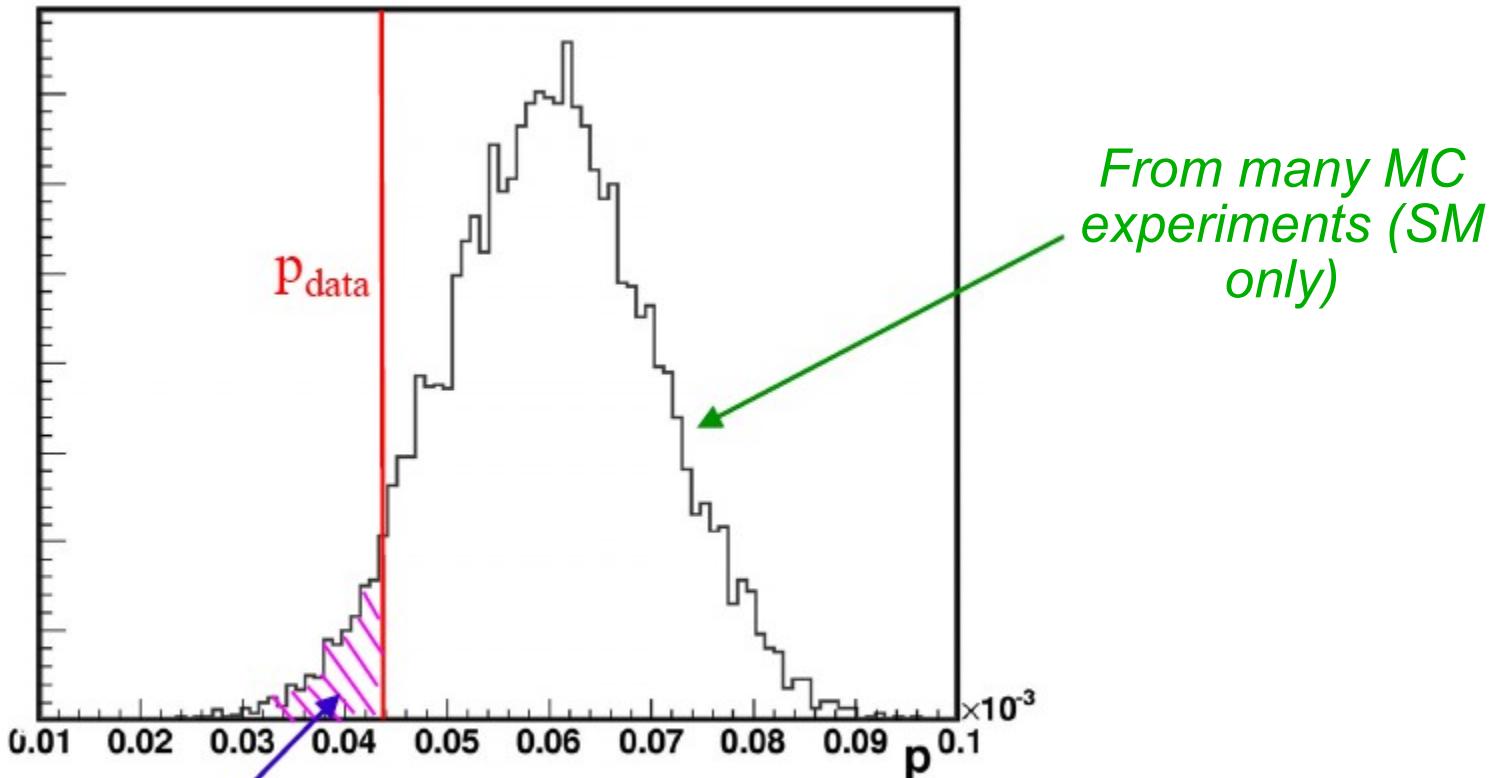


- ▶ For each region count  $N_{\text{data}}$  and  $N_{\text{MC}}$
- ▶ LHC has not started yet:  
Dice data according to errors

- ▶ First step:  
identify region where “probability” for  $N_{\text{MC}}$  to fluctuate to  $N_{\text{data}}$  is smallest  
→ Region of Interest →  $p_{\text{data}}$
- ▶ Second step: Account for “look-elsewhere-effect”
  - ◆ repeat “experiment” with bkg-only hypothesis many times (scan all regions)
  - ◆ determine probability  $\tilde{P}$  for finding value  $p \leq p_{\text{data}}$

Example:  $p_{\text{data}} = 10^{-6}$  could lead to  $\tilde{P} = 10\%$

# MUSiC: From $p$ to $\tilde{P}$

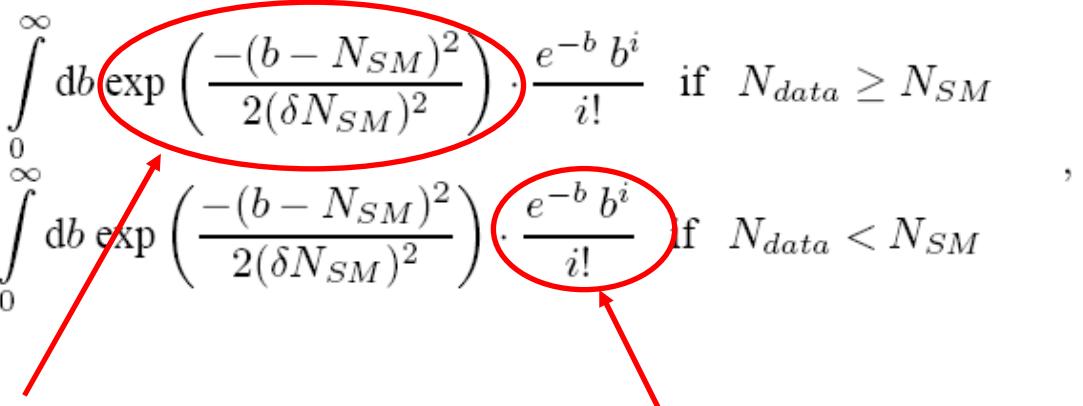


$$\tilde{P} = \frac{\text{number of HDH with } p_{\text{min}}^{\text{SM}} \leq p_{\text{min}}^{\text{data}}}{\text{total number of HDH}}$$

( HDH = Hypothetical Data Histogram )

- Logic very similar to  $CL_s$  method,  $\tilde{P}$  can be interpreted as  $CL_B$

# Definition of p-value

$$p = \begin{cases} \sum_{i=N_{data}}^{\infty} A \cdot \int_0^{\infty} db \exp\left(\frac{-(b - N_{SM})^2}{2(\delta N_{SM})^2}\right) \cdot \frac{e^{-b} b^i}{i!} & \text{if } N_{data} \geq N_{SM} \\ \sum_{i=0}^{N_{data}} A \cdot \int_0^{\infty} db \exp\left(\frac{-(b - N_{SM})^2}{2(\delta N_{SM})^2}\right) \cdot \frac{e^{-b} b^i}{i!} & \text{if } N_{data} < N_{SM} \end{cases},$$


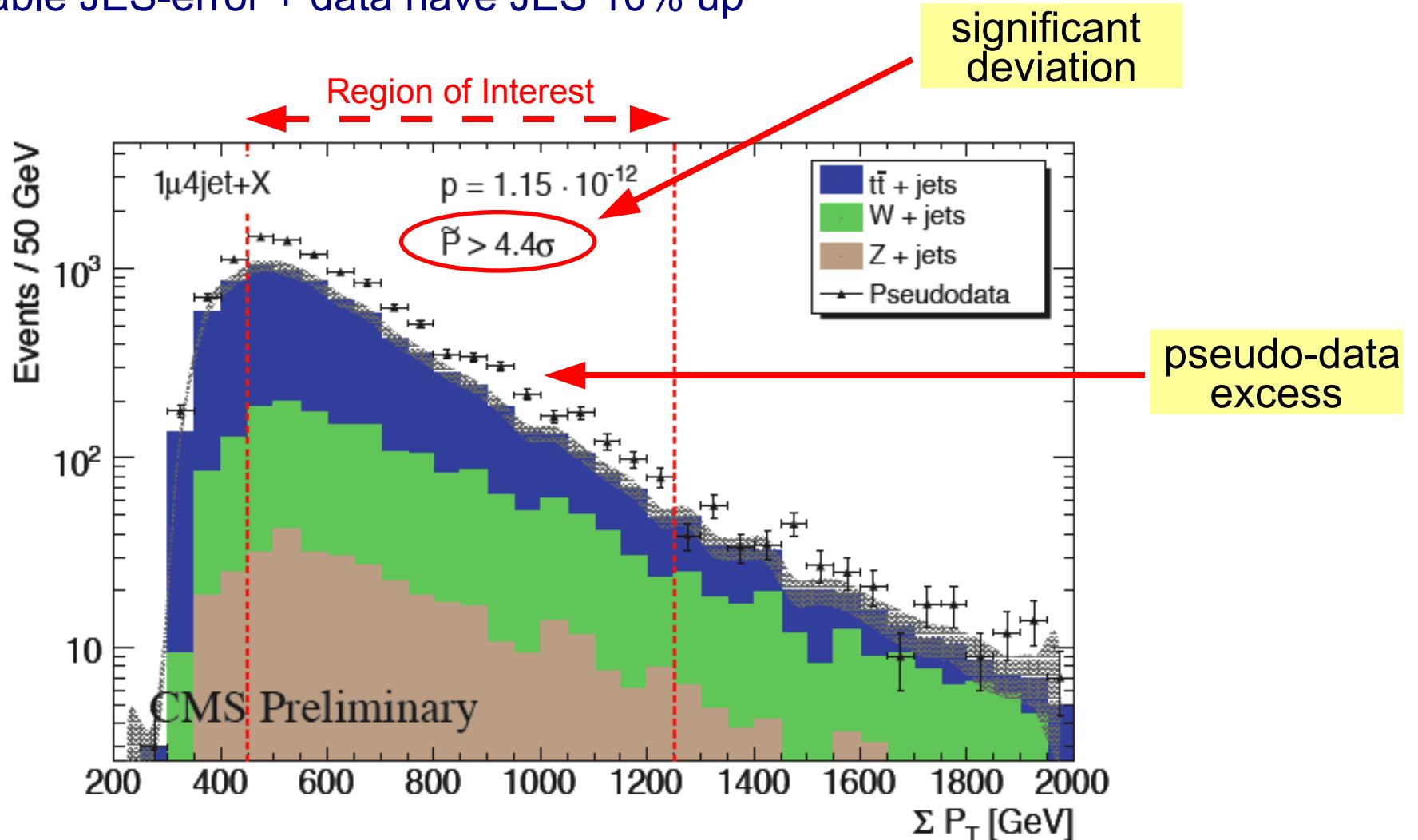
- ▶ Convolution of Gaussian (systematics) and Poisson (statistics)
- ▶ This is a Bayesian-frequentist hybrid method, has reasonable coverage
- ▶ Since  $N_{data}$ ,  $N_{SM}$  and  $\delta N_{SM}$  are always stated one can easily check using alternative statistical methods
- ▶ Including syst. errors in statistical estimator long discussed problem, see e.g. R.D. Cousins et al., arXiv:physics/0702156v3
- ▶ MUSiC is an **alarm-system** for interesting deviations, precise value of p not of major importance !

# Systematic Uncertainties

- ▶ Crucial to include them in algorithm to tell detector effect apart from signal
- ▶ Lack of detector/MC-understanding should be absorbed by systematics
- ▶ Various systematic uncertainties, respecting correlations
  - ◆ 5% luminosity
  - ◆ 10% cross sections (e.g. detailed PDF variation studies yield 2% - 8%)
  - ◆ 5% jet energy scale
  - ◆ 1-2% on possible efficiency correction factors ( $e, \mu, \gamma, \text{jet}$ )
  - ◆ 100% error on MC based misidentification-probability
- ▶ Developed infrastructure to include various errors, can be extended easily
- ▶ Used flat k-factors for  $W/Z/t\bar{t}$  NLO estimate (maybe better in the future...)
- ▶ More uncertainties with data: Smearing corrections, cosmics/beam halo, ...
- ▶ General philosophy: Assumed certain errors and check global data-MC agreement → learn from the result

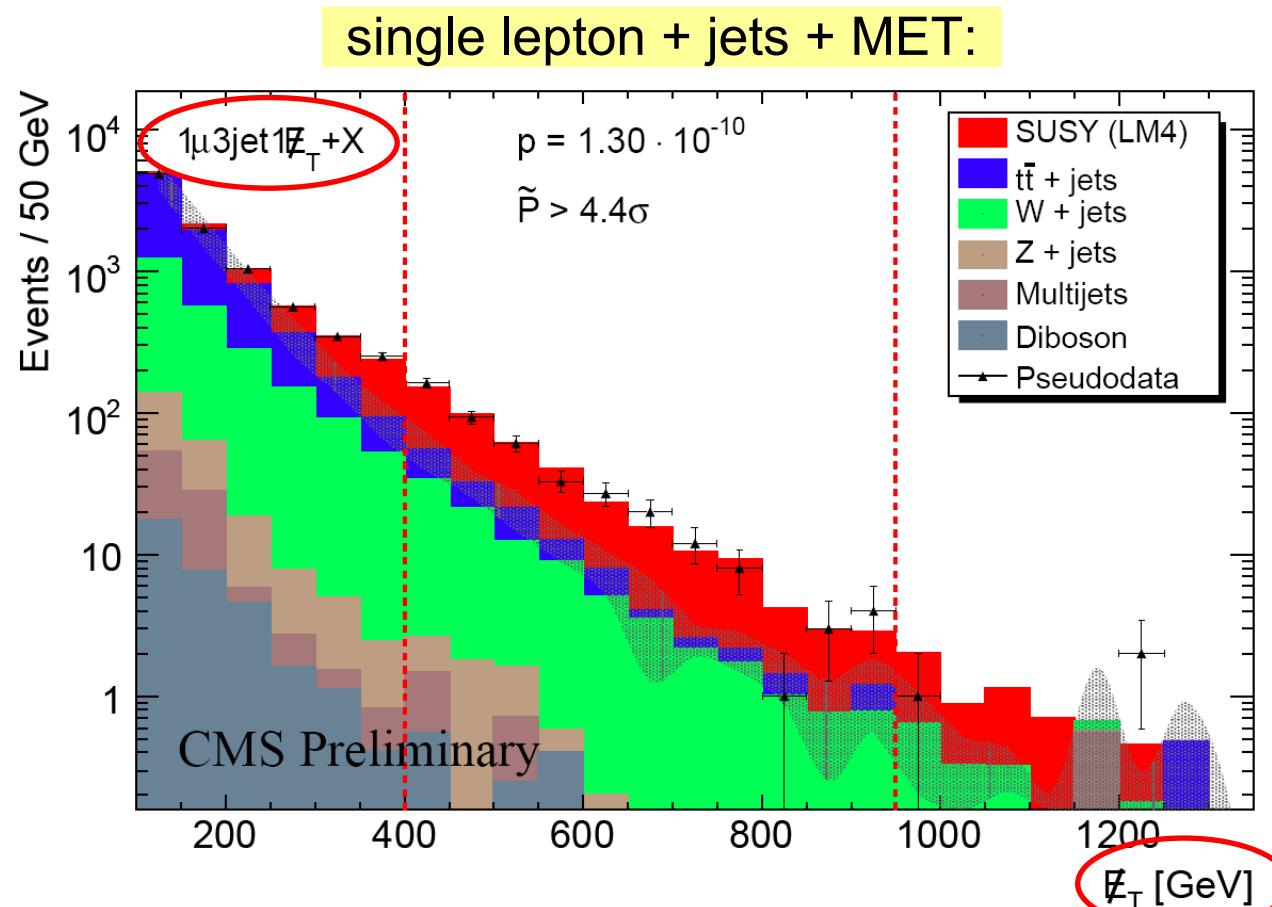
# Detector Effect

- ▶ Assume unknown detector effect:  
Disable JES-error + data have JES 10% up

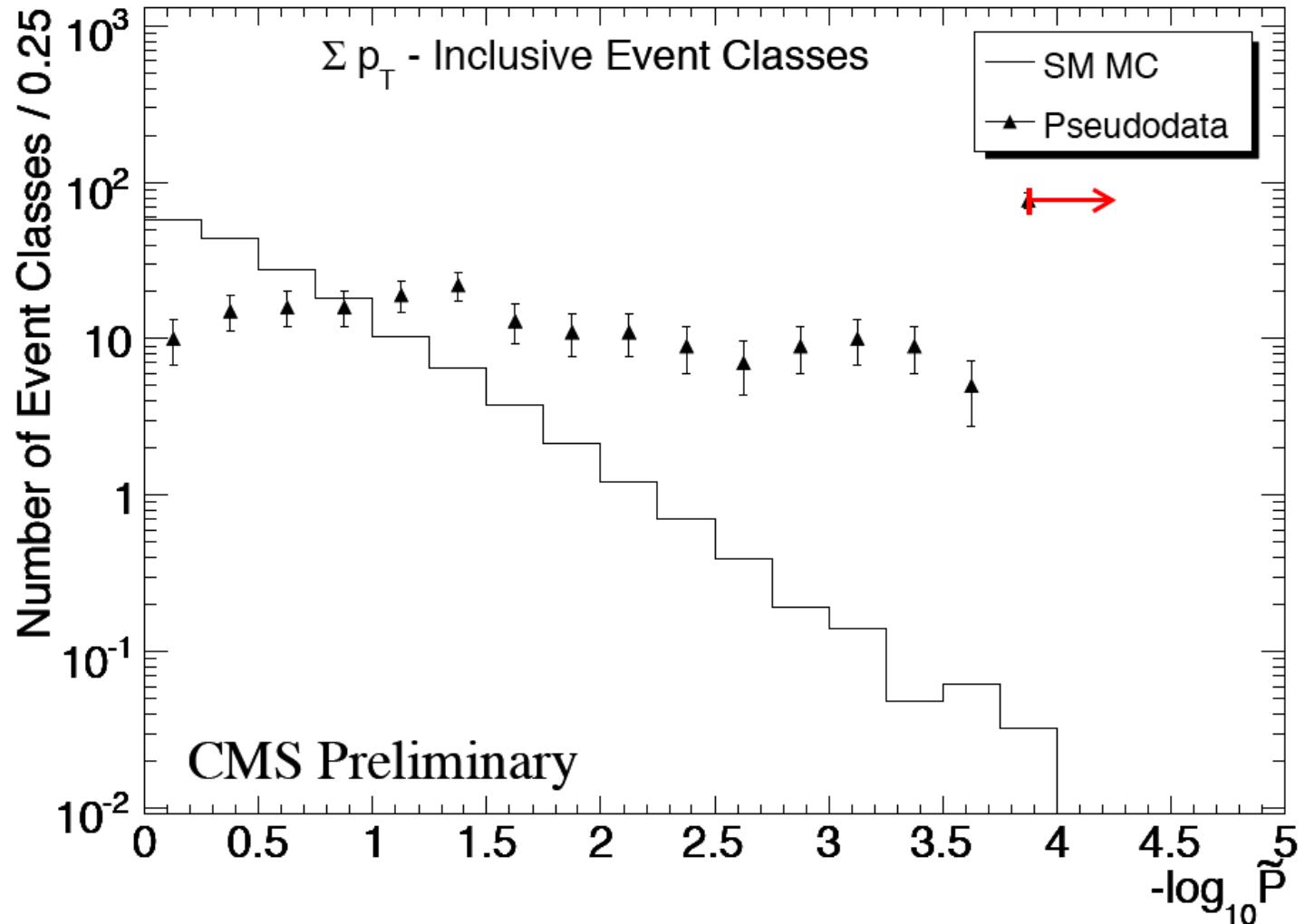


- ▶ Possible to spot problem in many classes with jets → consistent picture
- ▶ Re-enable 5% JES-error: Only  $1.6\sigma$  effect left

- ▶ In total 375 inclusive and 315 exclusive classes are populated
  - LM4 contributes to 160 (260) exclusive (inclusive) classes, 94 (170) classes with  $E_T^{\text{miss}}$ :  
15% (36%) show significant deviations with  $\tilde{P}$  (expected)  $< 1 \cdot 10^{-3}$  in  $\sum p_T$   
38% (59%) show significant deviations with  $\tilde{P}$  (expected)  $< 1 \cdot 10^{-3}$  in  $E_T^{\text{miss}}$
- ▶ Deviations ( $>3\sigma$ ) found in many classes, two examples:

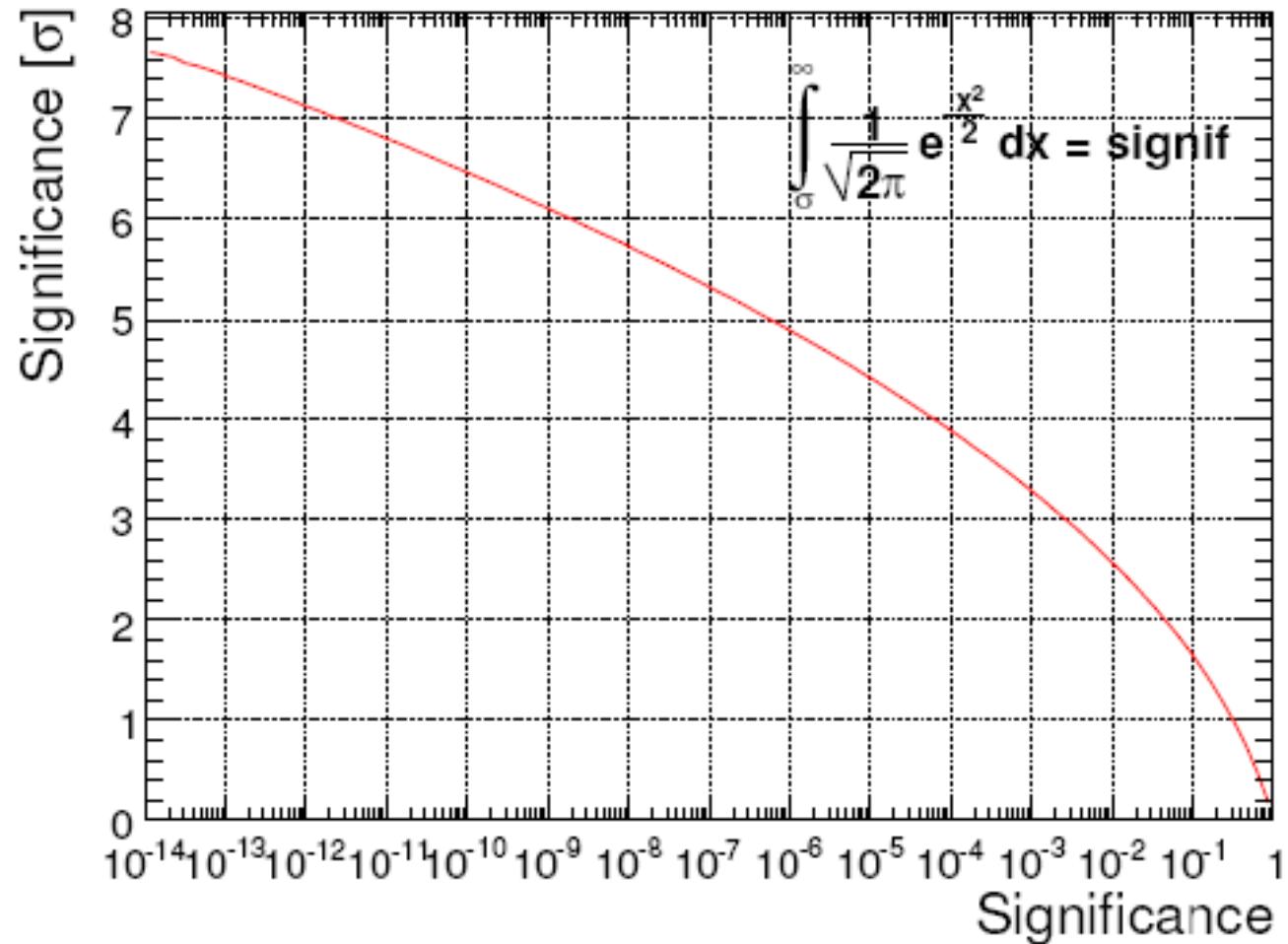


- Inclusive classes look more promising for SUSY



- But:
  - Inclusive classes are not disjunct!!! (Double-counting)
  - Statistical combination not possible in a generic simple way

# Significance in Standard Deviations



- ▶ “Two-sided Gaussian” since MUSiC is looking for excess AND deficit

# Global Trial Factor

- ▶ MUSiC is scanning many distributions  $O(100)$   
→ apply global penalty factor (trial factor)

$$\tilde{P}_{\text{CMS}} = 1 - (1 - \tilde{P})^n$$

- ▶ Significance of a distribution in the context of  $n$  distributions:

- ▶  $5\sigma$  become  $3\sigma$  when looking at 5000 plots
- ▶ Not specific to MUSiC:  
Remember, we have several hundreds of PhDs in CMS doing analysis too ....

