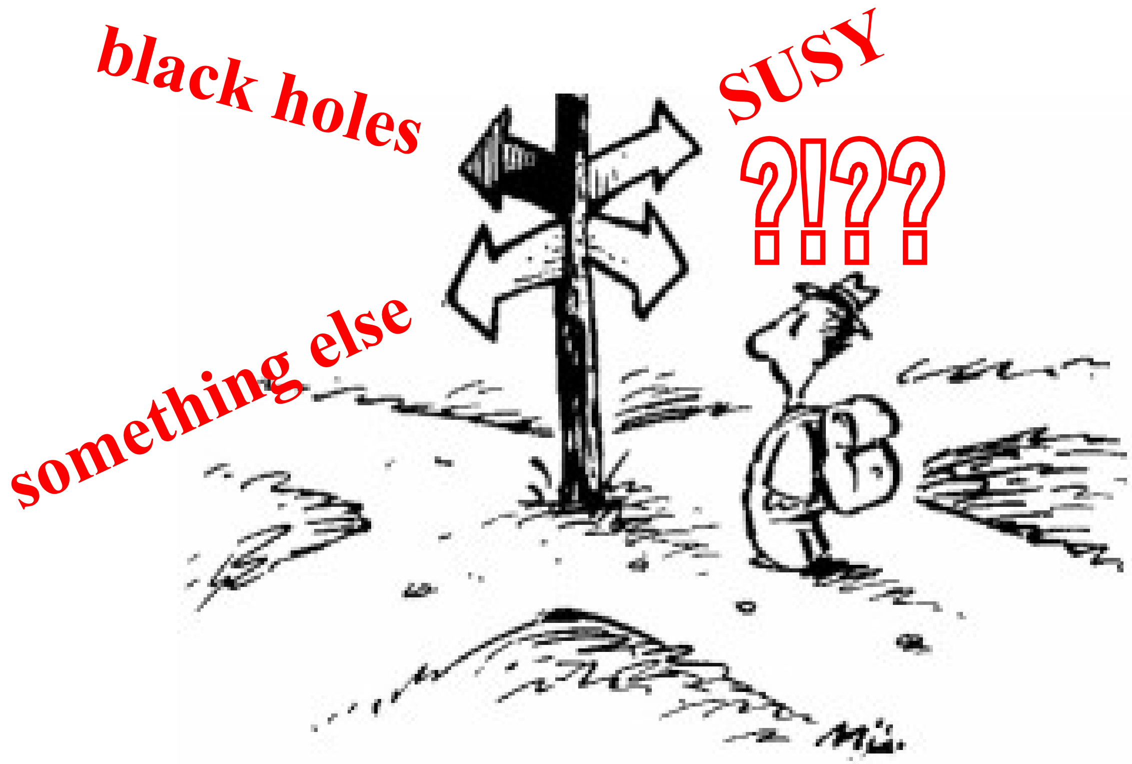
MUSiC – The Model-Unspecific Search in CMS



P. Biallass, T. Hebbeker, Carsten Hof, A. Meyer,
S. Schmitz, K. Padeken, H. Pieta, E. Dietz-Laursonn
RWTH Aachen

Terascale Workshop, Hamburg, 2009

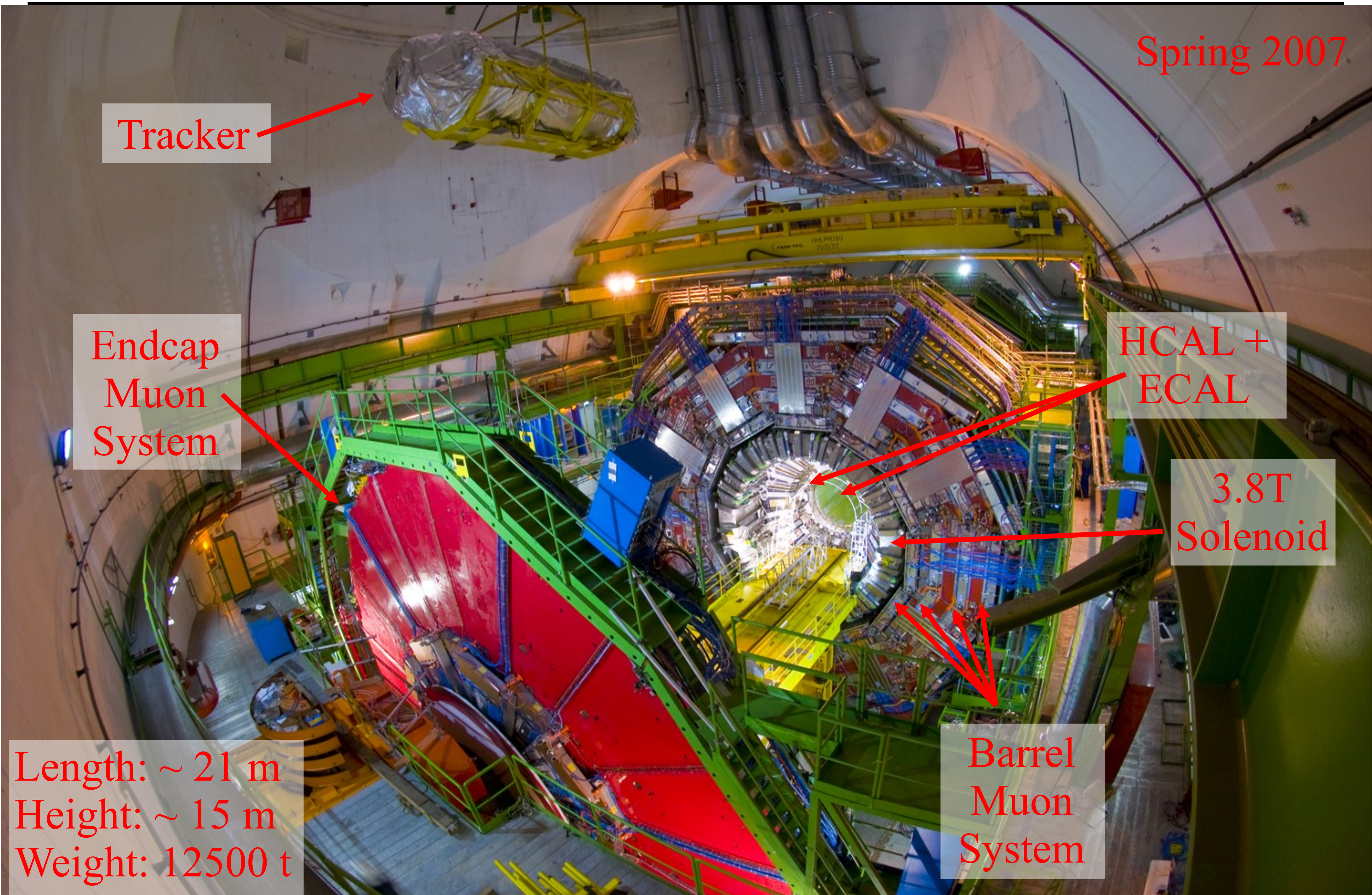
Outline

- Why Model-Independent?
- The Concept
- Probing MUSiC with Benchmarks
- Results
- Conclusion & Outlook

Updates of the study approved as CMS PAS EXO-08-005

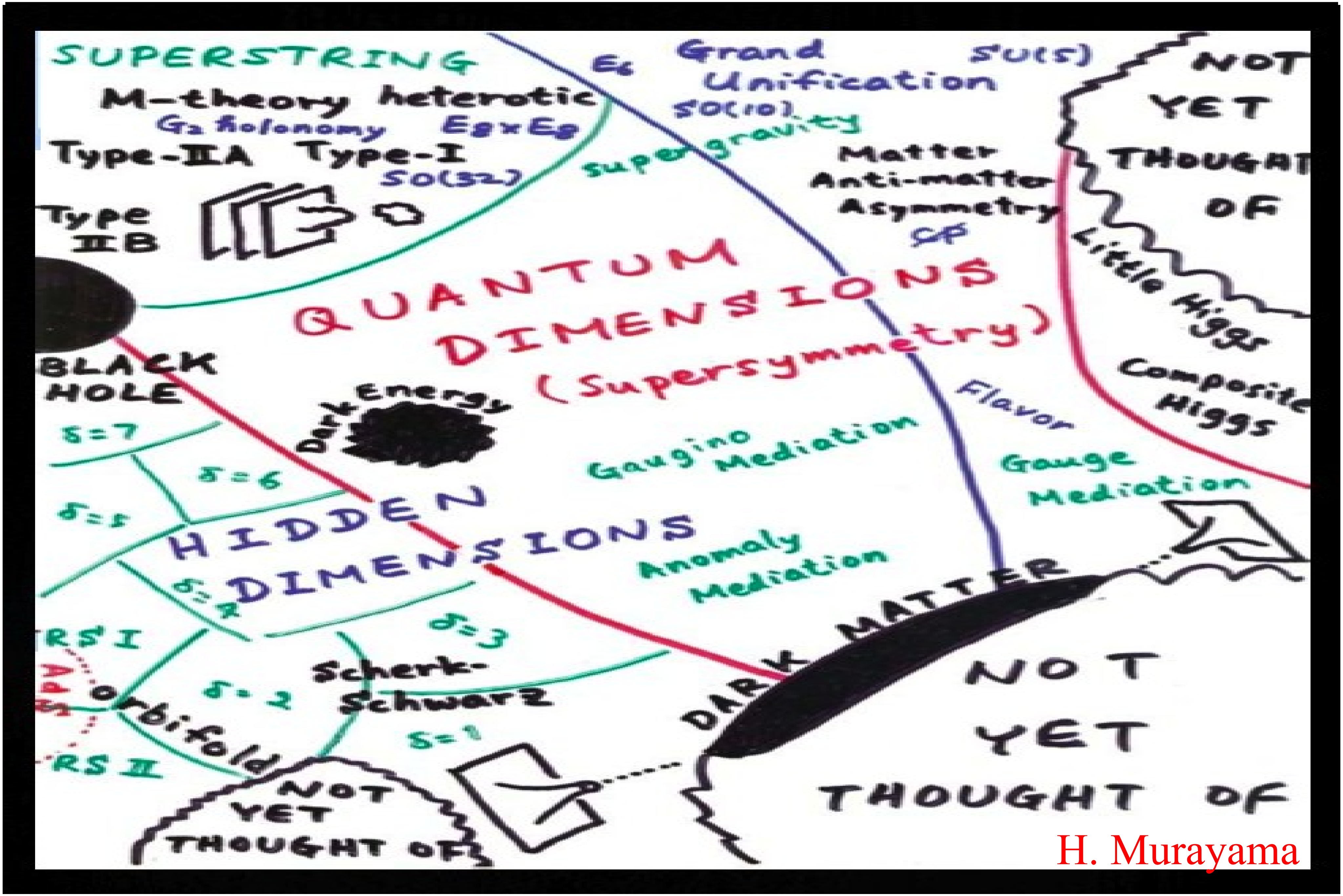
https://twiki.cern.ch/twiki/bin/view/CMS/PhysicsResults

The Compact Muon Solenoid



CMS Ready for Collisions

Theory Landscape



Experimental Challenge

How do we ensure not to miss anything?!?

(especially the "not yet thought of")

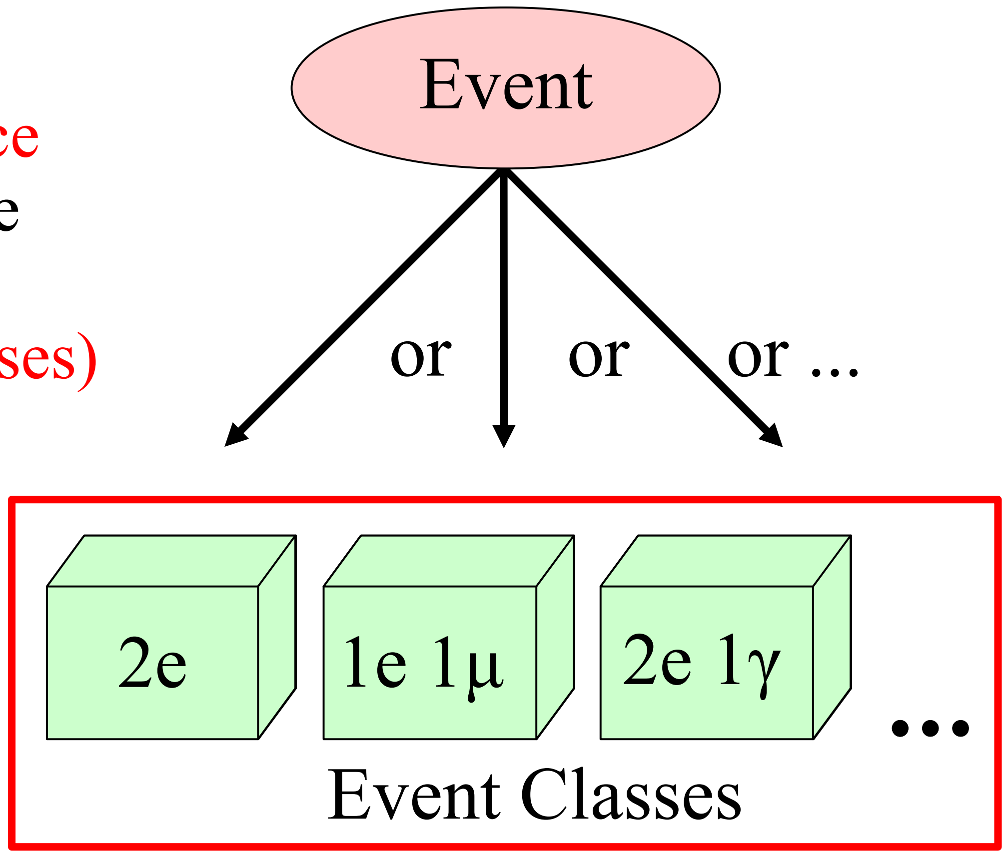
One idea:

Model-Independent Analysis

The Model Independent Approach

Idea:

- Assume only the Standard Model
- Look at each event without prejudice
- Classify events according to particle content (e, μ, γ, jet, MET)
- Perform broad data scan (~300 classes)



➡ Detects significant deviations from the Standard Model

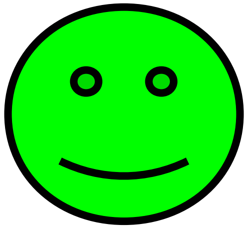
➡ New Physics or detector effects

Strategy already successfully performed at L3, DØ, H1, CDF, ...

Advantages & Disadvantages

- Independent of theoretical prejudice
- All topologies are investigated (e.g. $pp \rightarrow 7\mu$)
- Fits well into landscape of LHC start-up

➡ no theorist nor experimentalist can tell what we will see!

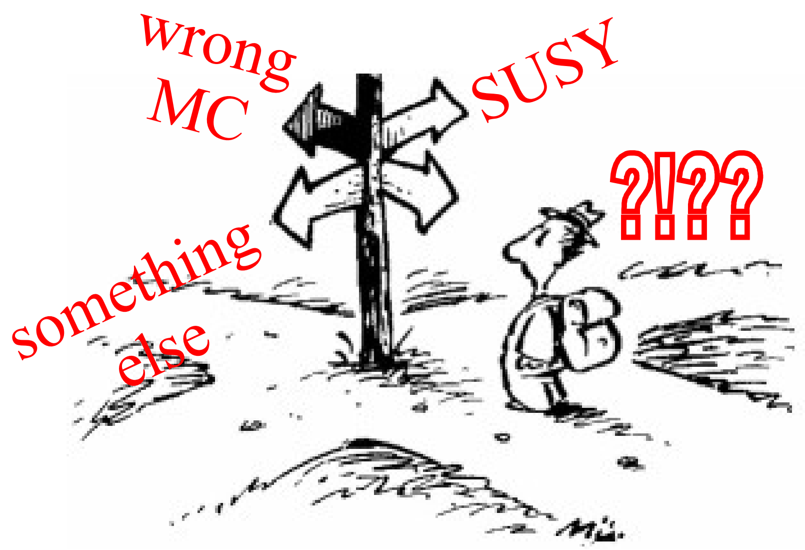


- Selection not optimized for special channel

➡ Lower sensitivity expected (e.g. $H \rightarrow \mu\mu\ ee$)

- Statistical penalty factor for “looking at many places”

The Scope of MUSiC

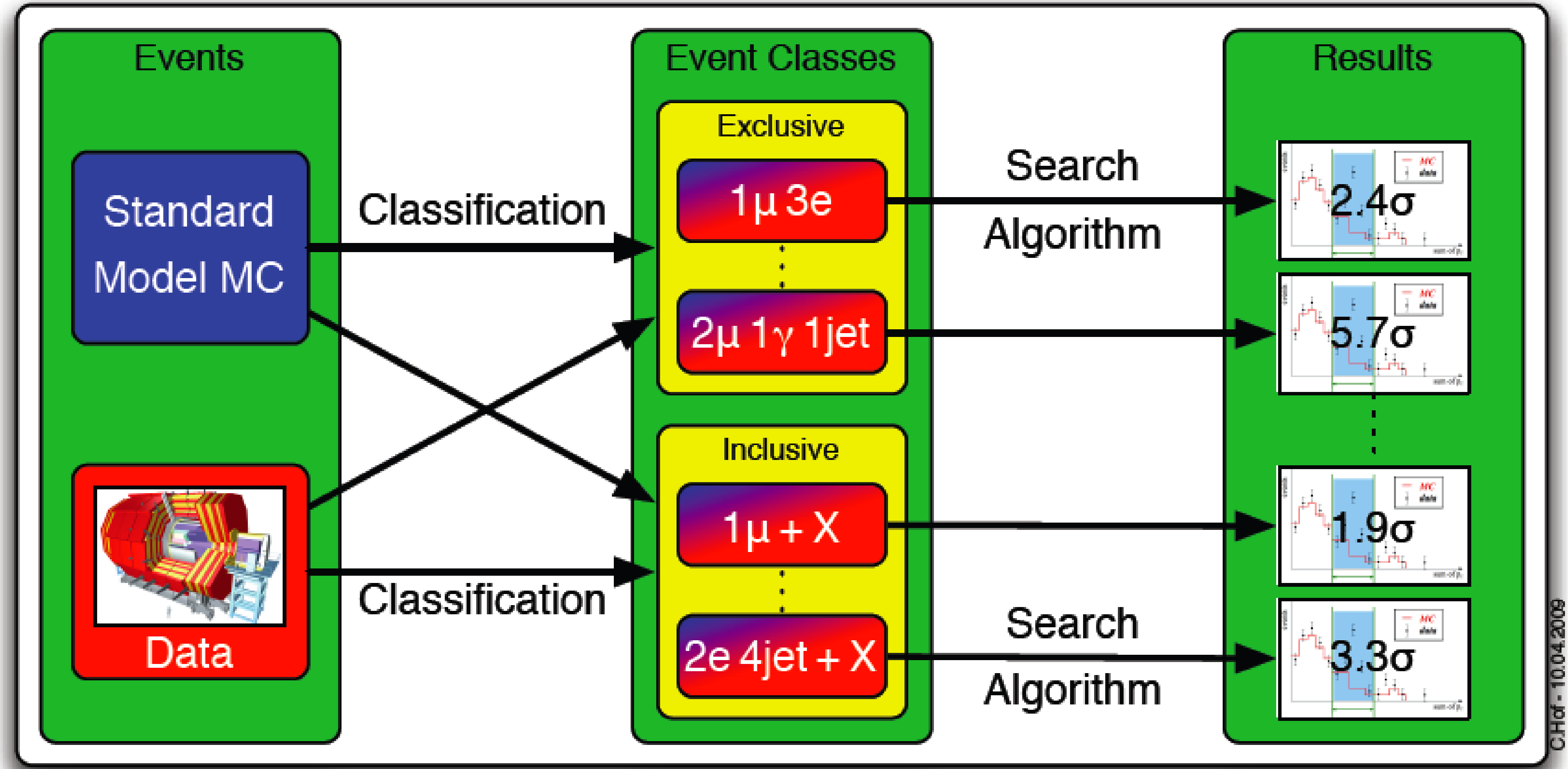


What is MUSiC?

- A global physics monitor
- Alarm system for discrepancies
- Complementary approach &
  cross check to dedicated analyses
- Not an automated discovery tool
- Might help to spot detector effects, quantify initial understanding of SM
- Participate in physics commissioning at CMS start-up
- After SM rediscovery → sensitive to New Physics

➡ **Deviations need to be carefully investigated!**

Analysis Flow of MUSiC



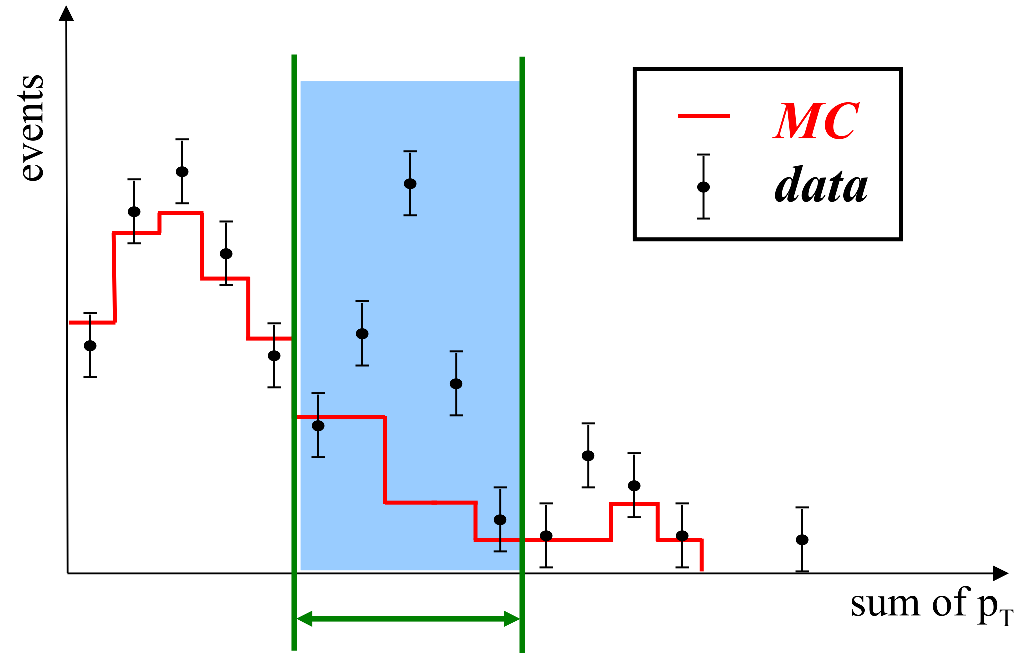
Physics Objects

- Focus on well-understood objects first, even if statistics is lost
- Define simple acceptance cuts (high p_T and central η)
- Ensure quality of measurement (e.g. N_{Hits}, χ^2 of fit)

	e/γ	μ	Jet (SiSCone)	MET
p_T cut	> 30 GeV	> 30 GeV	> 60 GeV	>100 GeV
$\|\eta\|$ cut	< 2.5	< 2.1	< 2.5	-
isolation	calorimeter/ track based	track based	-	-

The Search Algorithm (following H1 analysis)

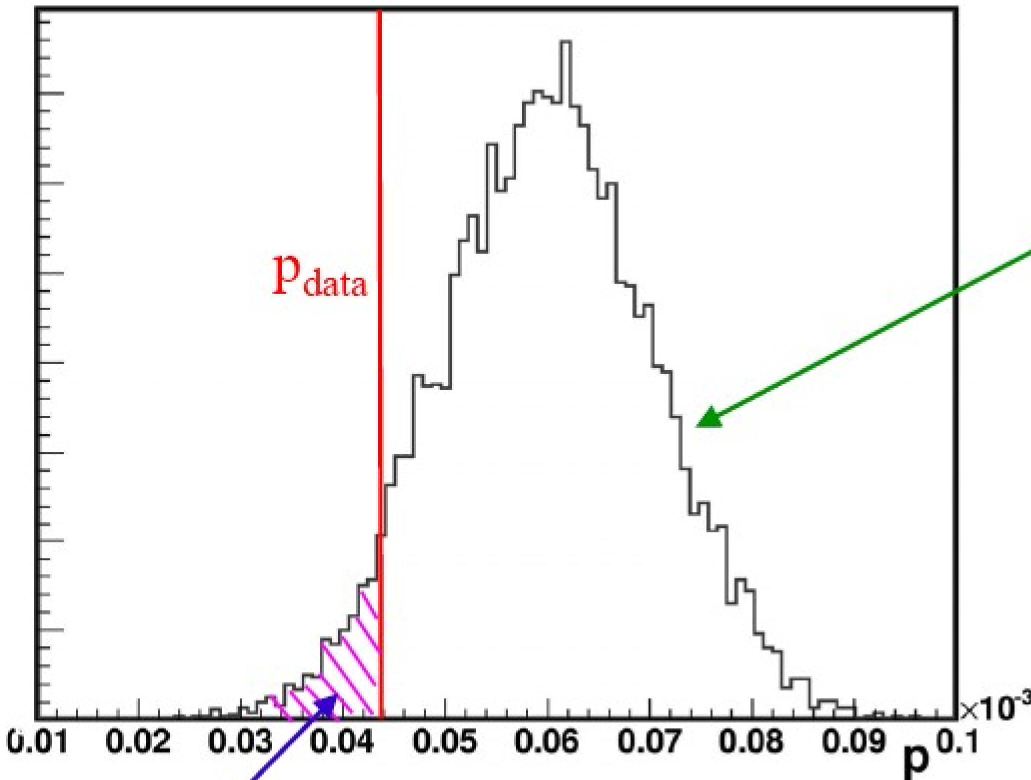
- Define:
  all possible *connected* regions
- For every region:
  count N_{data} and $N_{MC} \pm \Delta N_{MC}$



- **First Step:**
  identify region where "probability" for N_{MC} to fluctuate to N_{data} is smallest ➡ Region of Interest ➡ p_{data}
- **Second Step:** Account for "look-elsewhere-effect"
  repeat "experiment" to determine probability $\tilde{P}$ for finding value $p \leq p_{data}$

Algorithm takes systematic uncertainties into account!

From p to $\tilde{P}$



From many MC experiments (SM only)

$\tilde{P}$ = fraction of MC experiments with p less than p_{data}

Example: $p_{data} = 10^{-5}$ could lead to $\tilde{P}$ = 10%

Feasibility of MUSiC

How to prove the concept?

- No LHC collision data available
- No signal to search for (no exclusion/discovery limits)

Way out

- Use benchmark deviations (detector/MC effect, new physics, ..) to check feasibility
- Of course a bit contradictionary to the MUSiC concept
- Benchmarks reflect only the 'idea' and should be seen in a much broader concept

MUSiC – Timeline

Focus with first data

- Understand the detector (efficiencies, noise, ...)
- Tune the Monte Carlo generators & the detector simulation

After initial difficulties

- Re-establish the Standard Model
- Higher order effects in the tails (LO vs NLO, k-factors, ...)

Confidence in Detector and MC

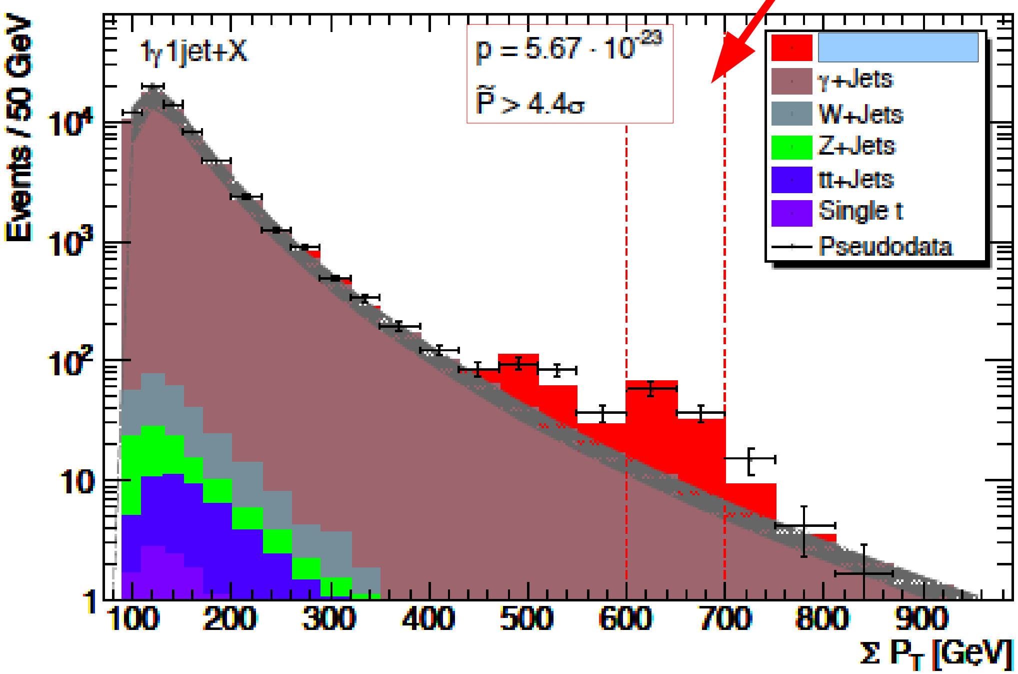
- Look for deviations from the Standard Model
- Especially interesting: physics not covered by dedicated analysis

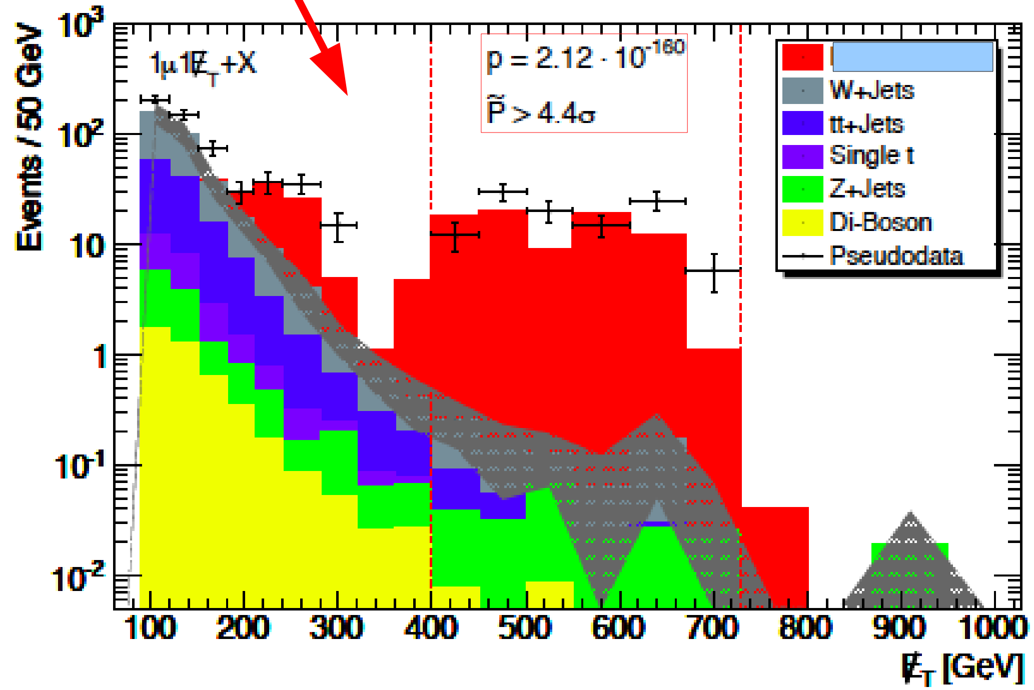
→ MUSiC can contribute in all experiment phases

New Physics?!?

Deviations visible in many jet & missing transverse energy distributions

100pb⁻¹

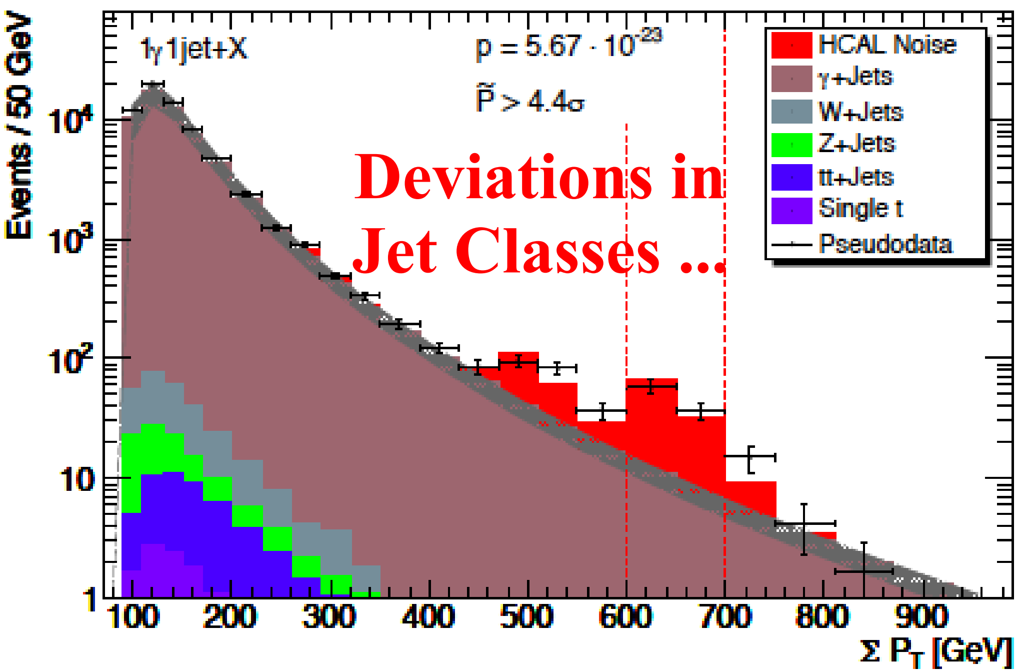


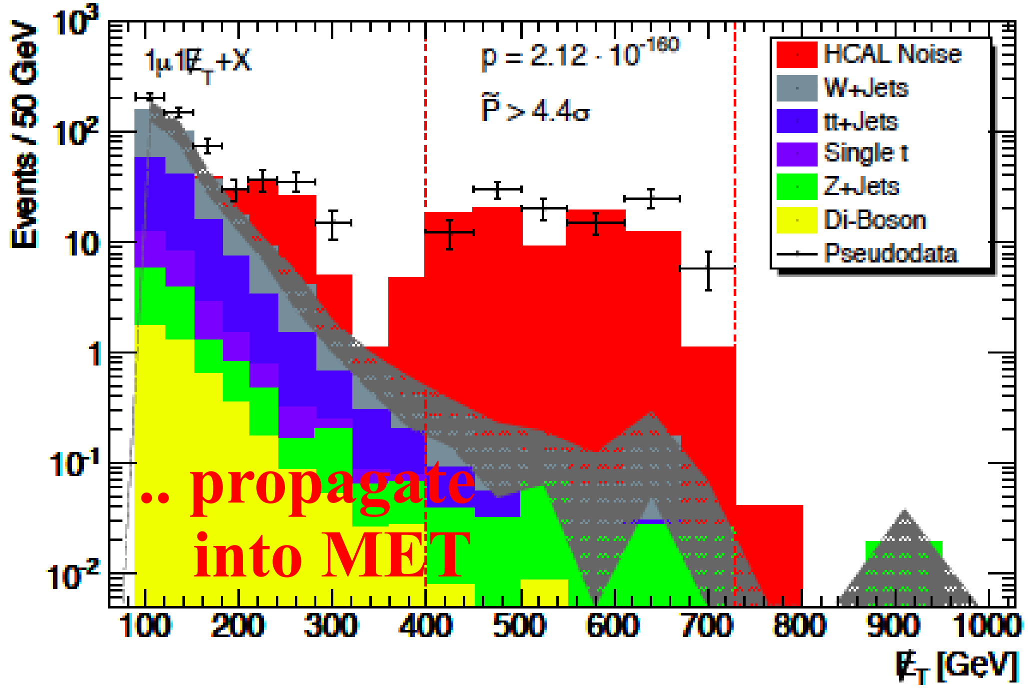


No! Noise in the Detector

"Warm" Calorimeter Cells

- Added randomly in 1/1000 of the events an additional calo deposit
- Average energy: 600 GeV at η = -0.1, 0.8, -1.6



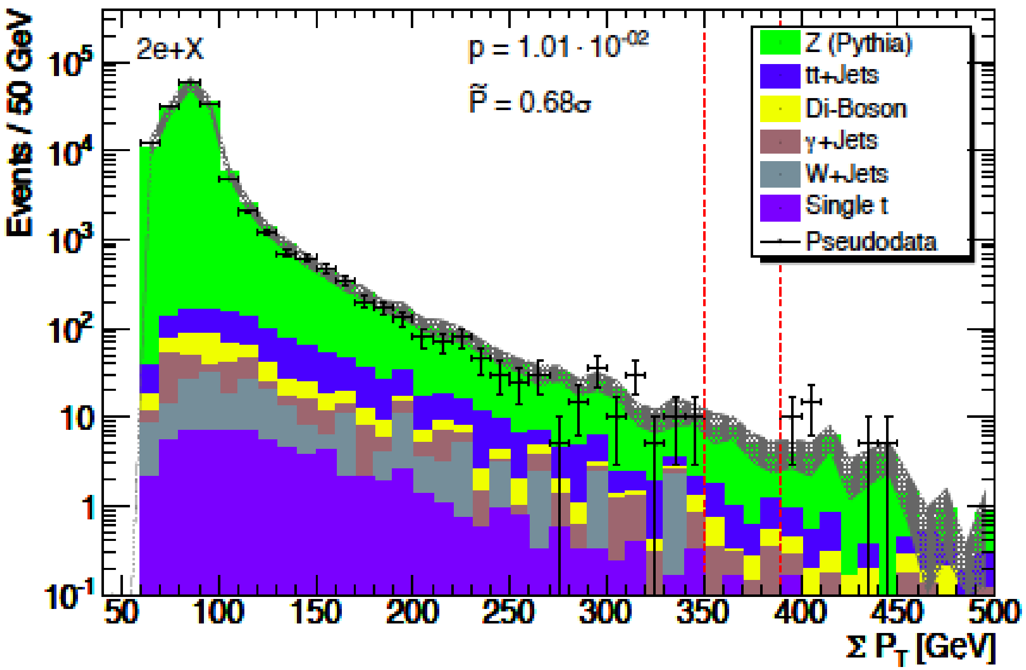


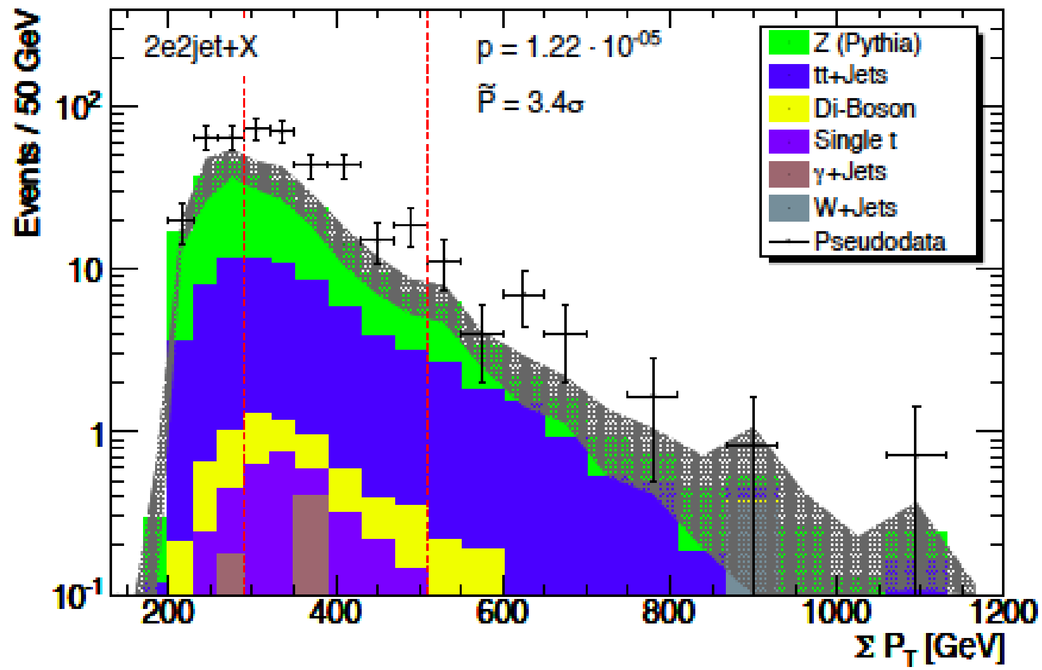
Peaks ≠ New Physics. Missing transverse energy heavily affected!

Monte Carlo Tuning

Toy Example: Pythia vs MadGraph

- Assume: Data follow MadGraph, MC also except Drell-Yan (Pythia)



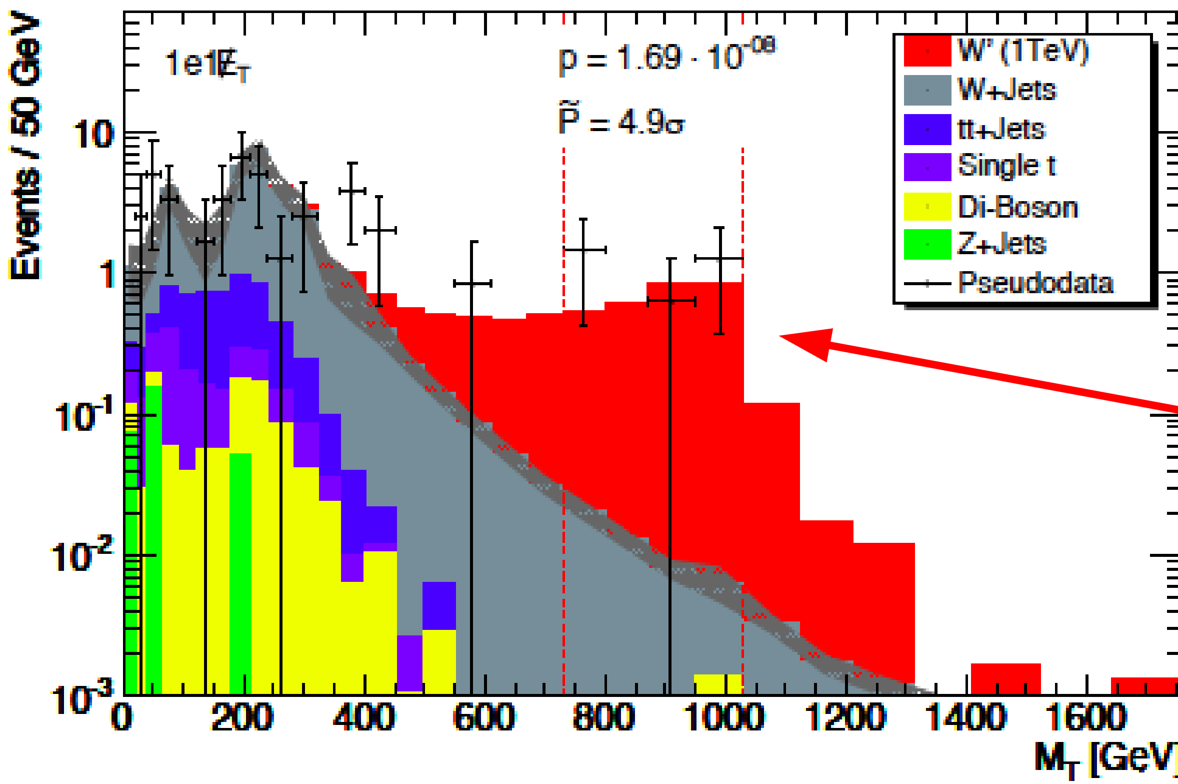


➡ Agreement in 0jet class, increasing discrepancy with jet multiplicity

MadGraph produces more/harder jets. Pythia's parton shower softer

First Day Physics: W'

- Search for heavy brother of the W
- Free parameter: mass (1 TeV here), σ (W' $\rightarrow$ eν) = 1230 fb
- Most significant: M_T in 1e + MET class (also used in dedicated search!)



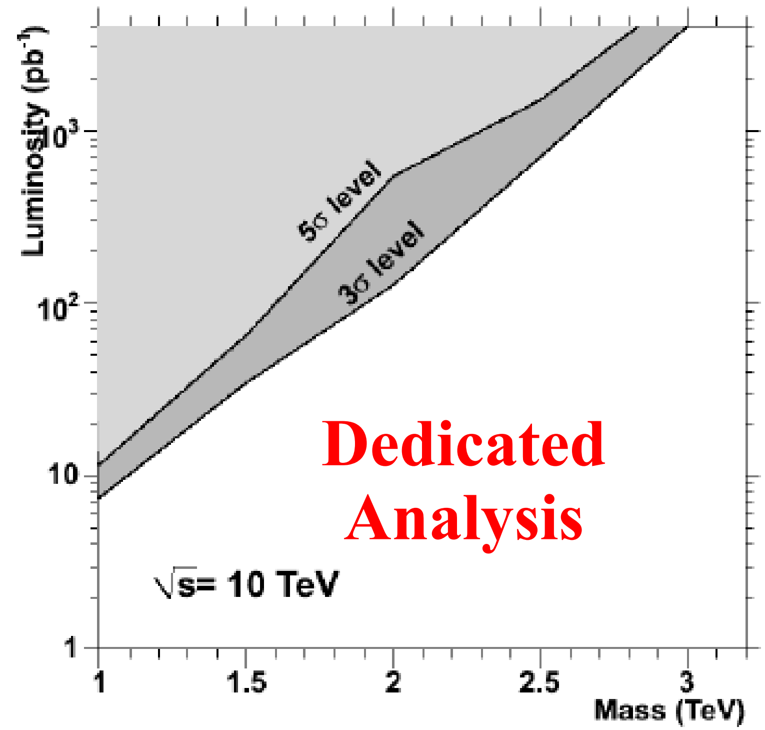
W' Jacobian peak picked by algorithm

Comparison to Dedicated Analysis

- Choose luminosity in MUSiC according to 5σ contour of dedicated analysis

Result

- Quite comparable despite trial factor!
- Explanation:
  - Algorithm looks for largest discrepancy
  - Trial factor relatively small

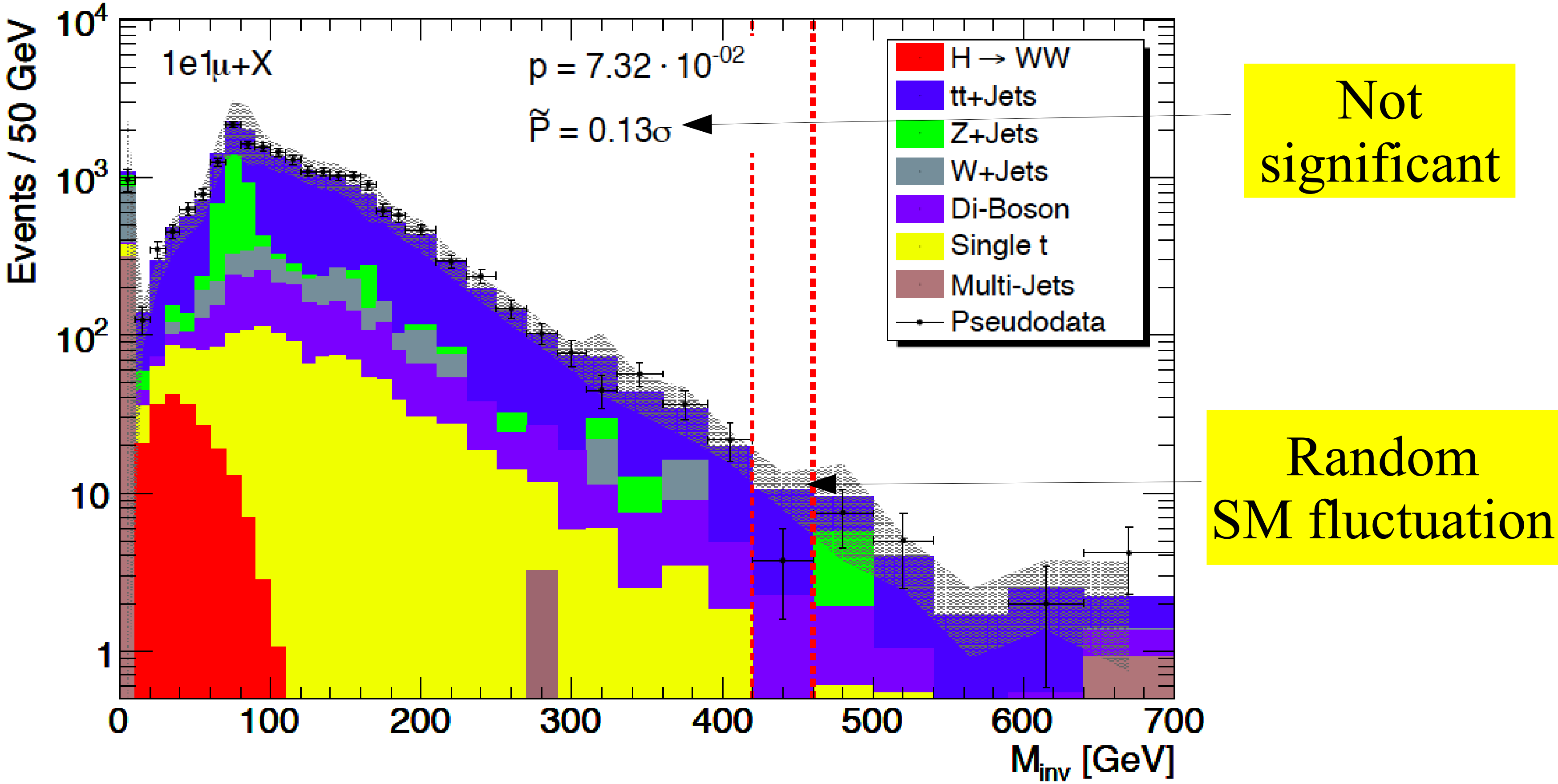


W' mass	$\mathcal{L}_{\text{int}}$	W' Analysis	MUSiC Z_N	MUSiC Z_{LN}	$p_{\text{data}}^{\text{min}}$ (expected) Z_N	$p_{\text{data}}^{\text{min}}$ (expected) Z_{LN}
1 TeV	10 pb^{-1}	$\approx 5\sigma$	$(5.04 \pm 0.08)\sigma$	$(5.12 \pm 0.06)\sigma$	$7.8 \cdot 10^{-9}$	$1.1 \cdot 10^{-8}$
1.5 TeV	65 pb^{-1}	$\approx 5\sigma$	$(5.09 \pm 0.08)\sigma$	$(5.5 \pm 0.3)\sigma$	$3.6 \cdot 10^{-9}$	$4.9 \cdot 10^{-9}$
2 TeV	325 pb^{-1}	$\approx 5\sigma$	$(5.11 \pm 0.08)\sigma$	$(5.3 \pm 0.1)\sigma$	$2.9 \cdot 10^{-9}$	$5.0 \cdot 10^{-9}$

$\widetilde{P}$

“Negative” Example: SM Higgs @ 1fb^{-1}

- SM Higgs with m = 160 GeV, $\sigma(H \rightarrow WW)$ = 710 fb



- Tiny signal in front of huge SM background
- No signal → No deviation → Consistent result!

Lepton Charges

Motivation

- Standard Model: only few processes with same sign leptons

Concept

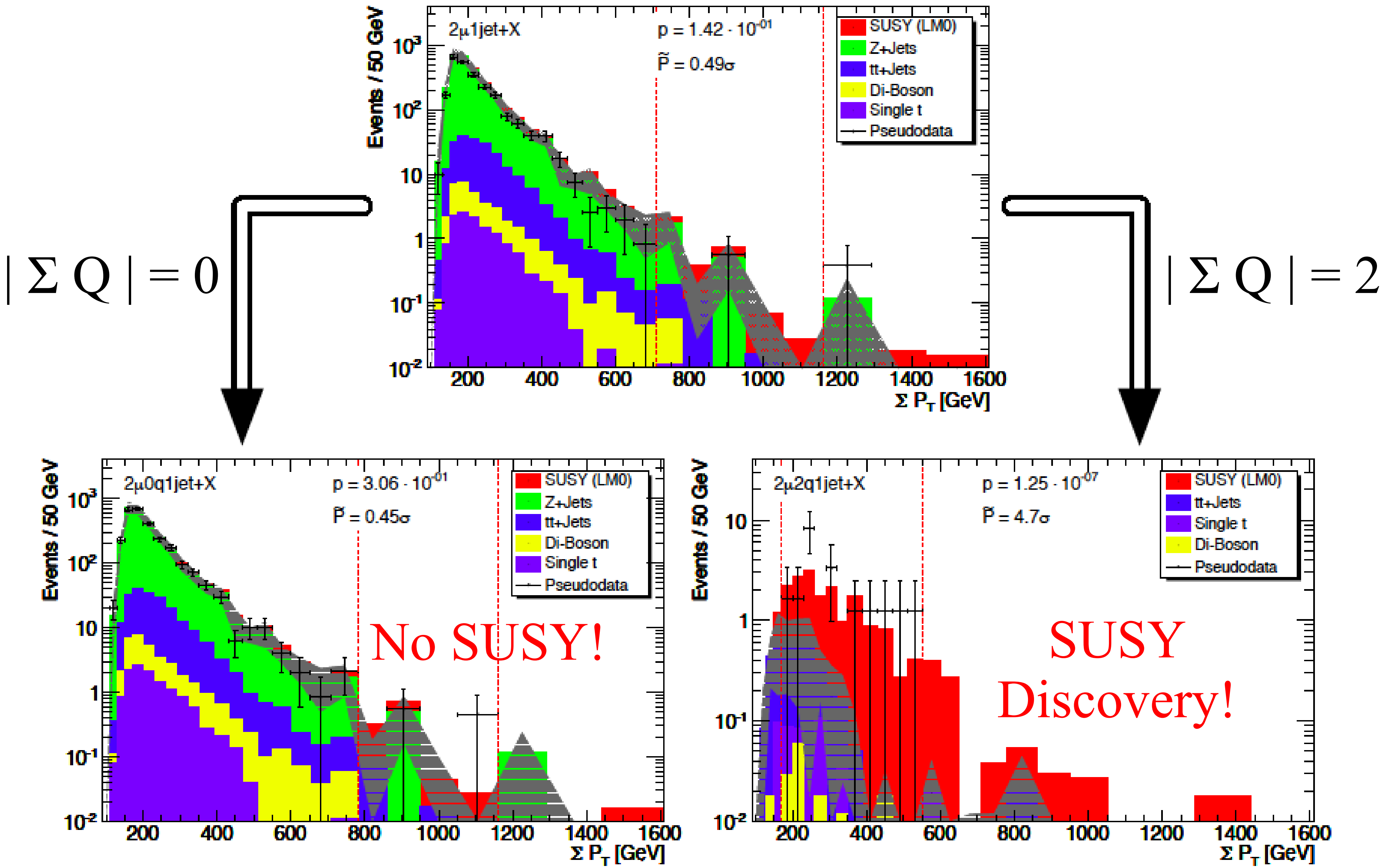
- Addition of all possible lepton charge combinations (e.g. 1e+ 1μ- 1μ+ 3jets) would increase number of classes by a huge factor :-(
- Alternative: only look at absolute value of sum of lepton charges: $|\Sigma Q|$

Benchmark: CMS Supersymmetry Point LM0 $\sigma_{LO} = 110\text{pb}$

$$m_{1/2} = 160\ \text{GeV},\ m_0 = 200\ \text{GeV},\ A_0 = -400\ \text{GeV},\ \mu > 0,\ \tan\beta = 10$$

Charge Example: 2μ 1jet+X @ 100 pb^{-1}

No Charges! No SUSY!



MUSiC was my first love ...

Today:

- MUSiC allows to generally look for deviations from the SM
- Complementary alternative to „conventional searches“
- Deviations need to be interpreted by physicicists!

  Detector effect, MC feature, New Physics, something else...?!?
- Method works as examples demonstrate

Future:

- Analyze first LHC pp-data!!!

Be alert to all possibilities!

Backup Slides

Definition of p-value

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

- Convolution of Gaussian (systematics) and Poisson (statistics)
- This is a Baysian-frequentist hybrid method, has reasonable coverage
- Since N_{data}, N_{SM} and δN_{SM} are always stated one can easliy 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 !

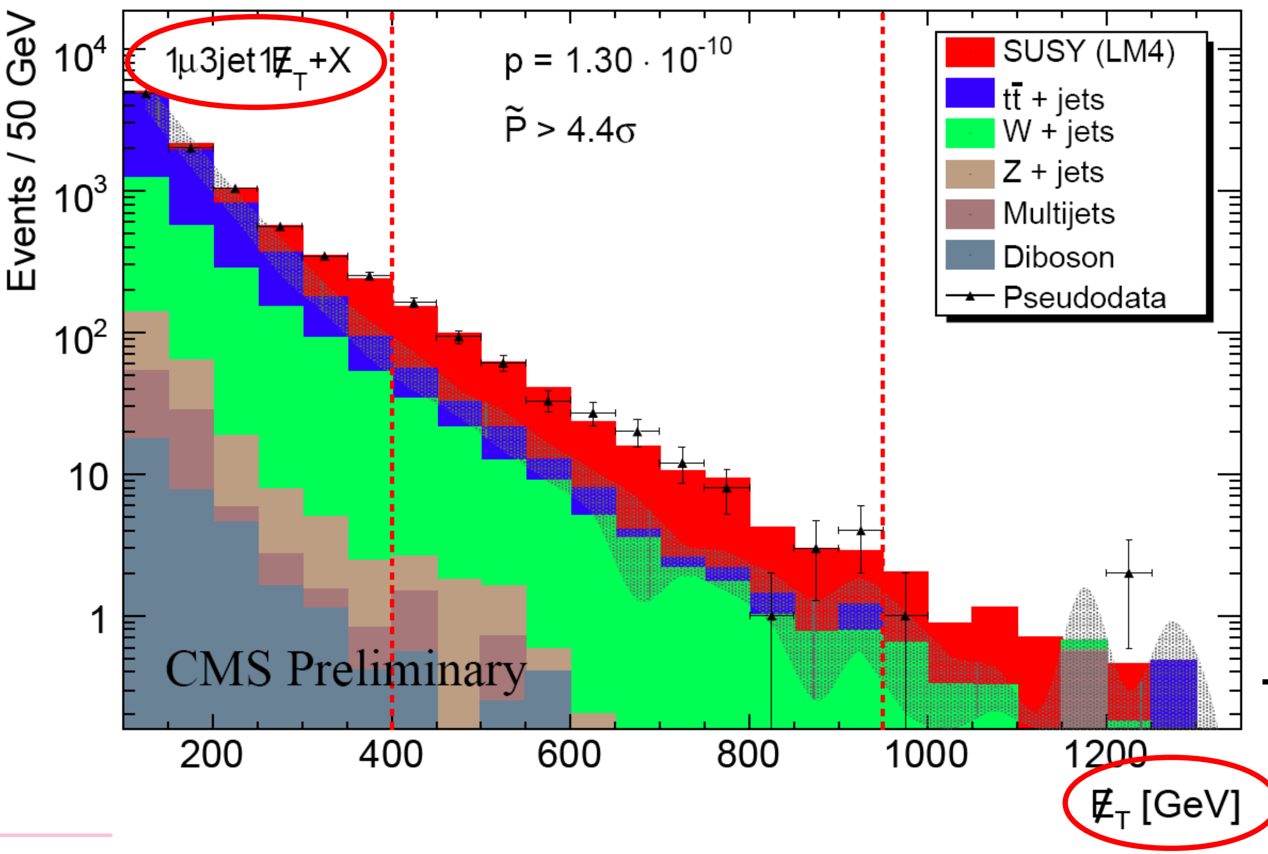
Systematic Uncertainties

- 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
  - 10% 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,μ,γ,jet)
  - 100% error on MC based misidendification-probability
- Developed infrastructure to include various errors, can be extended easily
- Used flat k-factors for W/Z/tt 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

SUSY Results

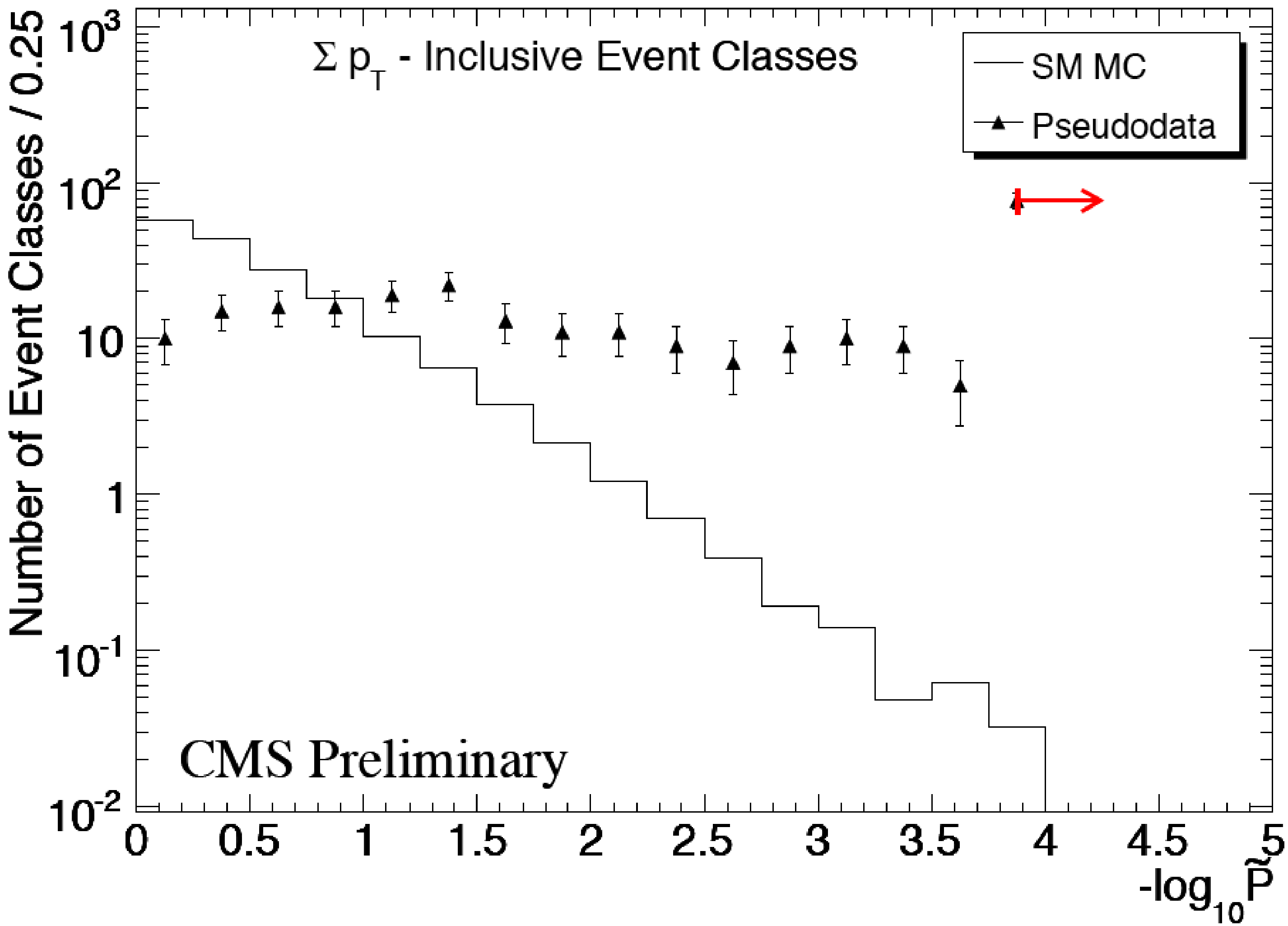
- In total 375 inclusive and 315 exclusive classes are populated
  - LM4 contributes to 160 (260) *exclusive* (*inclusive*) classes, 94 (170) classes with E_T^{miss}:
    15% (36%) show significant deviations with $\tilde{P}$ (expected) $< 1 \cdot 10^{-3}$ in $\sum p_T$
    38% (59%) show significant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

single lepton + jets + MET:



SUSY Global Picture – Inclusive Case

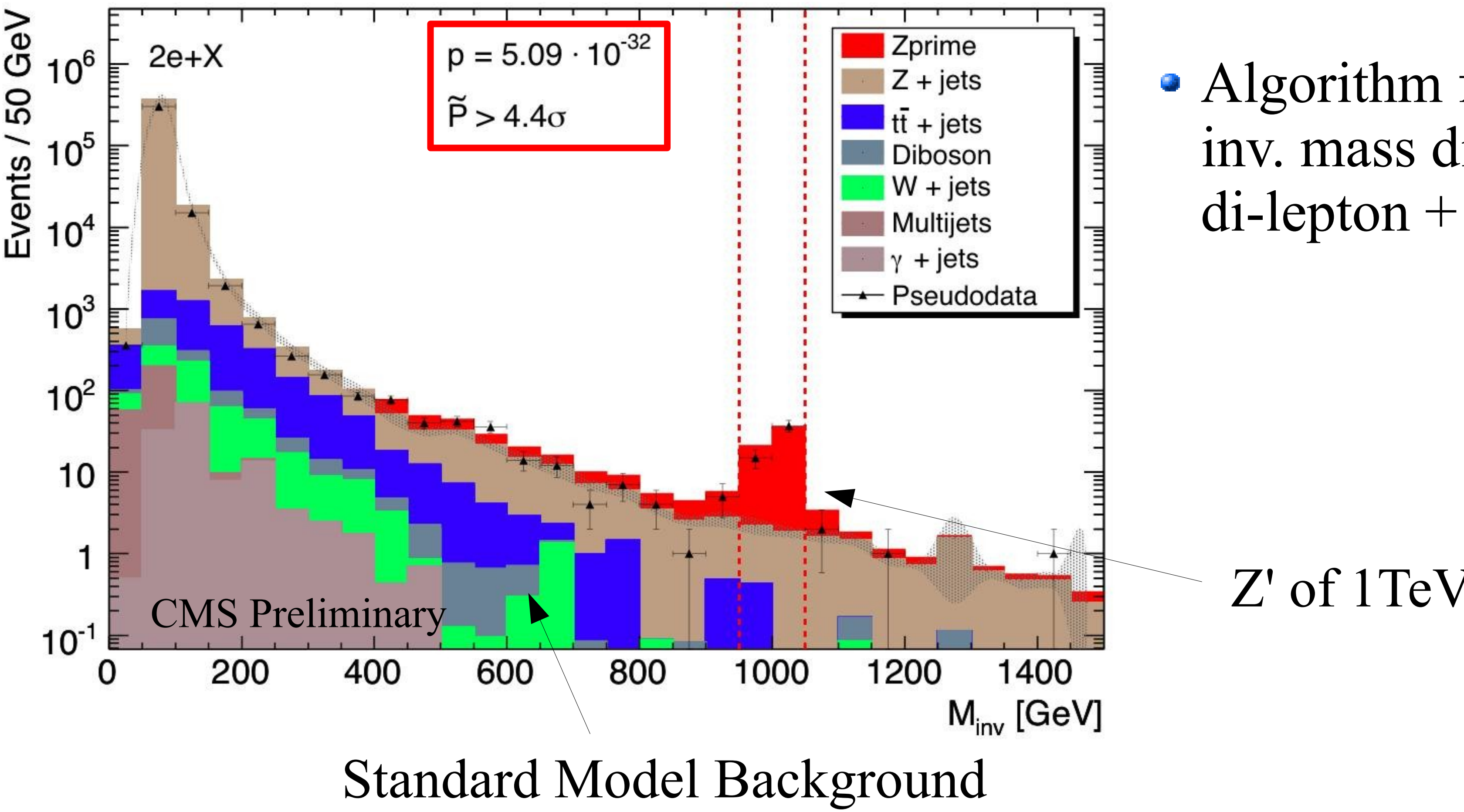
- Inclusive classes look more promising for SUSY



- But:

Proof of Principle: Z'

- Assume: $1fb^{-1}$ data @14TeV with Z' of m = 1TeV (σ = 365pb)



- Algorithm finds peak in inv. mass distribution of di-lepton + X class

Global Trial Factor

- MUSiC is scanning many distributions O(100)

  ➔ apply global penalty factor (trial factor)

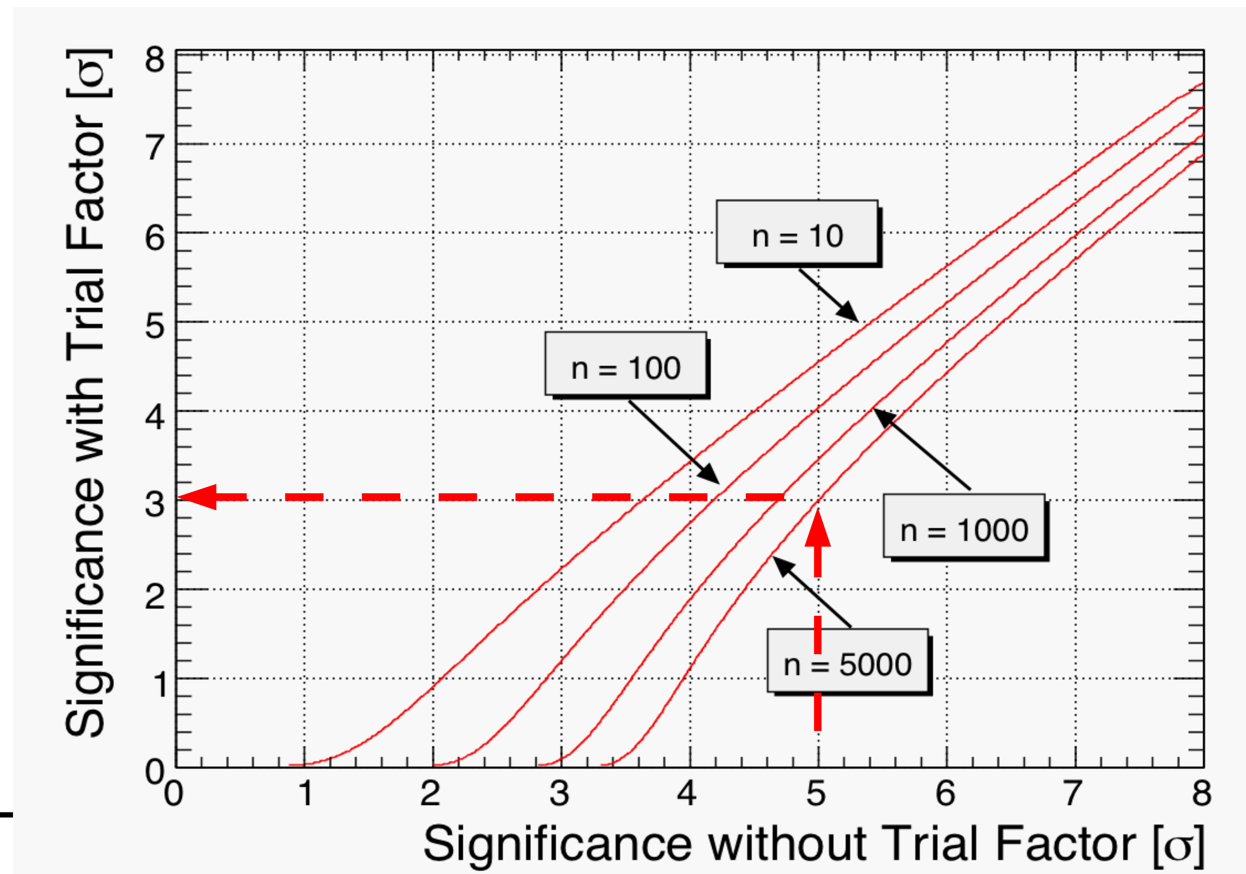
$$\tilde{P}_{CMS} = 1 - (1 - \tilde{P})^n$$

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

- 5 become 3 when looking at 5000 plots

- Not specific to MUSiC: Remember, we have several hundreds of PHDs in CMS doing analysis too ....

☺☺☺

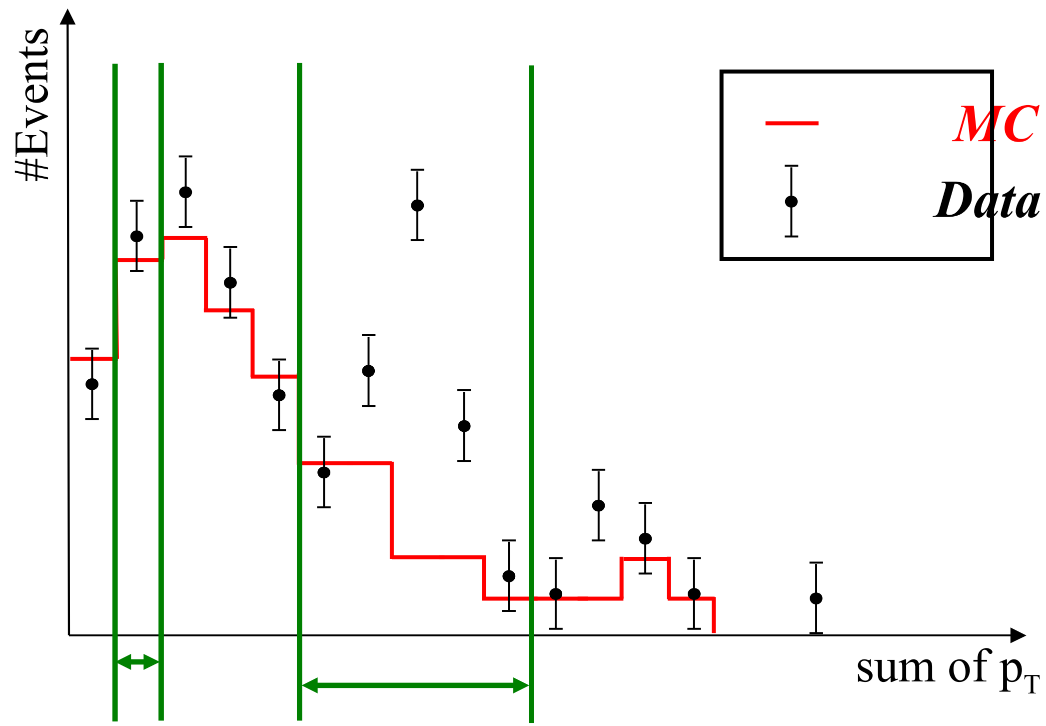


Search Algorithm Part I

First step: Find the most interesting region in $\sum p_T$ distribution

- Define:
  all possible *connected* regions
- For every region:
  count N_{data} and N_{MC}
- Identify region where the probability p_{min}^{data} of N_{MC} fluctuating to N_{data} is the smallest (following H1 analysis)

➡ Region of Interest



No meaningful significance (discovery) in this step!!!
Just a Region of Interest!!!

Search Algorithm Part II

Second step: Determine the Significance

1. Dice the SM repeatedly (taking errors into account)
2. Perform step 1 with diced SM and SM MC as Input
   i.e. test Background Only Hypothesis BOH (again for all regions)
3. Define Significance as:

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

- Takes the look-elsewhere effect into account
- „If you try to find a signal everywhere in data, then you also have to scan the background for signal-like fluctuations everywhere!"

Mind: In a MIS you don't know where the signal is!

Searches without Data

- „Advantage" of a MC study: can repeat CMS experiment several times

➡ Dice also data (signal+background MC) e.g. SM + SUSY

➡ Probe discovery for optimal/average/worse case

Interpretation of $\tilde{P}$:

- Statistical estimator for agreement between data and MC
- $\tilde{P}$ is a so called p-value
- Comparable to likelihood ratio -2lnQ (often used for CL_S/CL_B)

➡ Discovery possible if two curves well seperated

➡ $\tilde{P}$ small