Identification of hadronic Tau decays in CMS

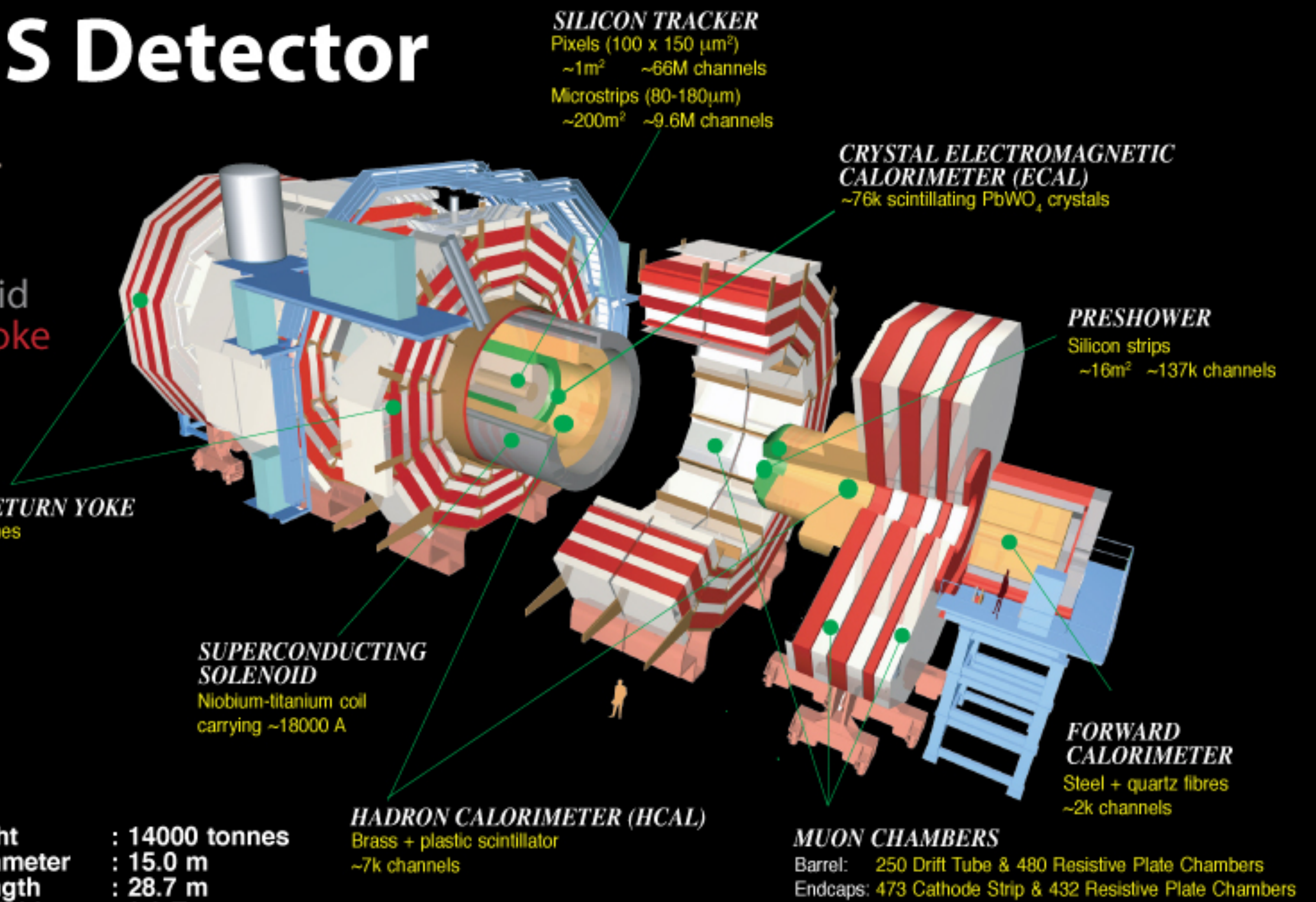
Christian Veelken

for the CMS Collaboration

Tau2014 Conference, September 18th 2014

CMS Detector

Pixels
Tracker
ECAL
HCAL
Solenoid
Steel Yoke
Muons

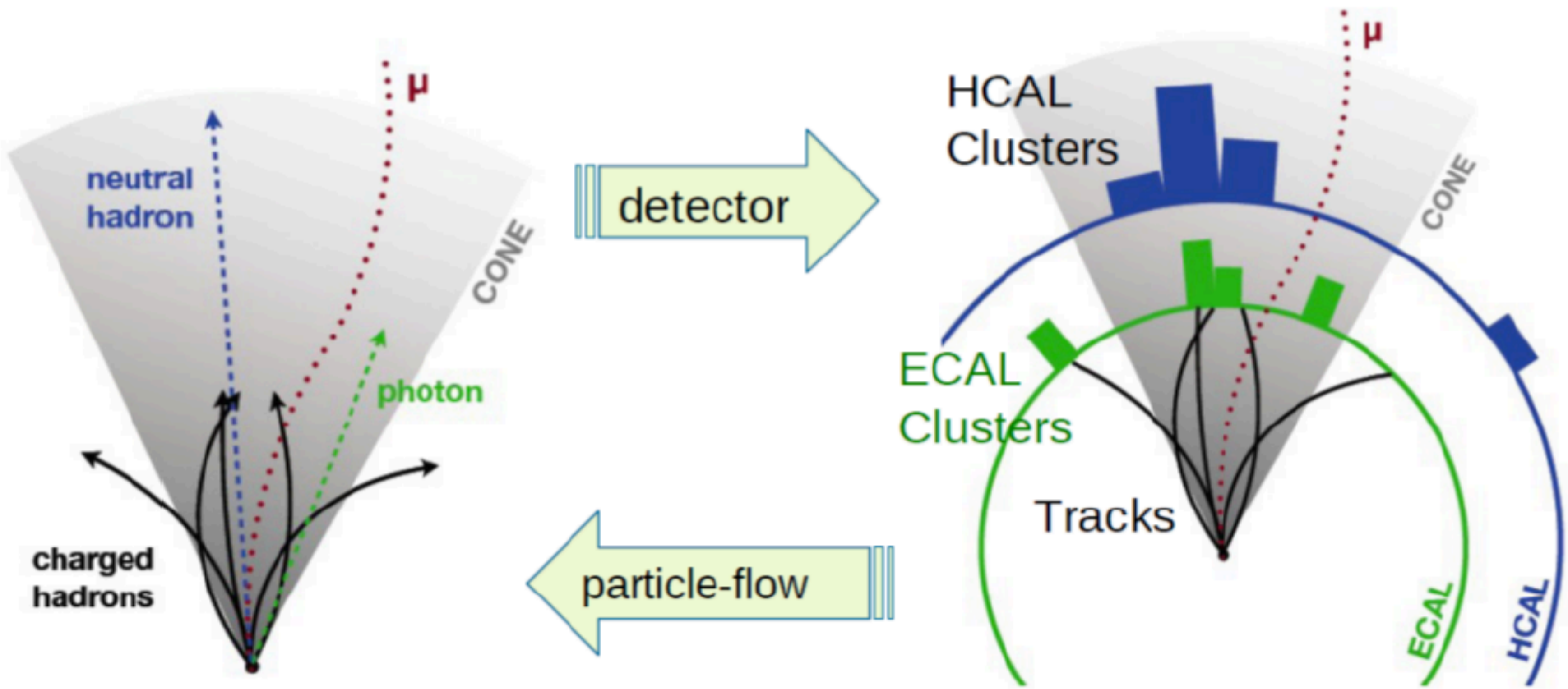


Total weight	**: 14000 tonnes**
Overall diameter	**: 15.0 m**
Overall length	**: 28.7 m**
Magnetic field	**: 3.8 T**

Particle-Flow Algorithm

Consistent Interpretation of all detector signal in terms of individual particles:

- e, μ, photons, charged hadrons, neutral hadrons



Higher level objects are reconstructed using individual particles as input:

- τ_h, Jets (incl. b-tagging), E_T^{miss}

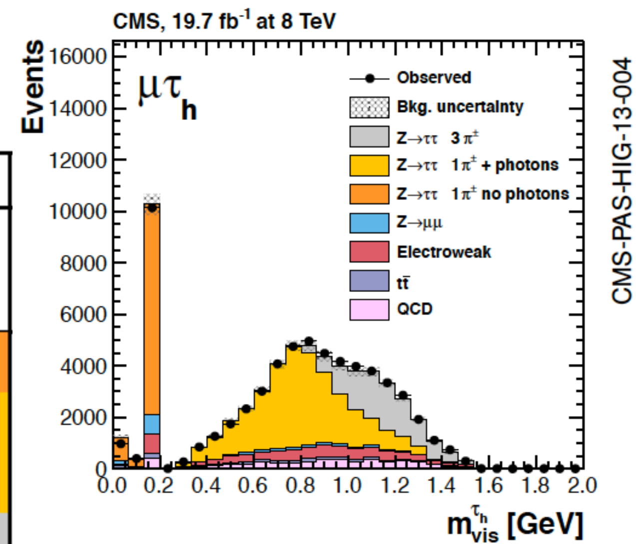
Tau Decays

Mass m_τ= 1.78 GeV

Lifetime c • τ = 87 µm

Decay Mode	Resonance	BR [%]
$\tau^- \rightarrow e^- \overline{\nu}_e \nu_\tau$		17.8
$\tau^- \rightarrow \mu^- \overline{\nu}_\mu \nu_\tau$		17.4
$\tau^- \rightarrow \pi^- \nu_\tau$	$\pi(140)$	11.6
$\tau^- \rightarrow \pi^- \pi^0 \nu_\tau$	$\rho(770)$	26.0
$\tau^- \rightarrow \pi^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	10.8
$\tau^- \rightarrow \pi^- \pi^+ \pi^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \rightarrow \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$		4.8
Other hadronic modes		1.7
All hadronic modes		64.8

PDG



Electrons/muons from $\tau \rightarrow e/\mu$ decays reconstructed by standard CMS electron/muon reconstruction

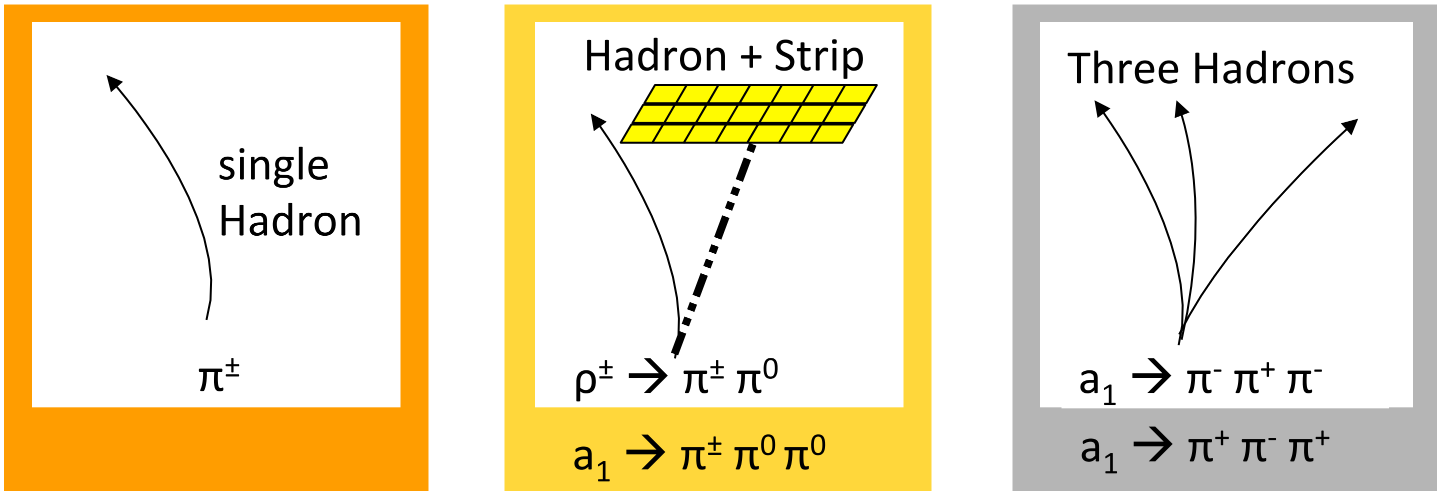
τ_h Identification ≅ reconstruction of $\pi^\pm$, $\rho^\pm$, $a_1^\pm$ signatures

Acceptance of τ_h reconstruction: $P_T^\tau \gtrsim 20$ GeV and $|\eta_\tau| < 2.3$
typically used by physics analyses in CMS

"Hadron plus Strips" (HPS) Algorithm

① Seeded by anti-k_T (R = 0.5) jets build from particle-flow output

② Build combination of charged Hadrons + Strips (= τ_h candidate)



③ Select combinations passing mass window cuts for $\pi^\pm$, $\rho^\pm$, a_1

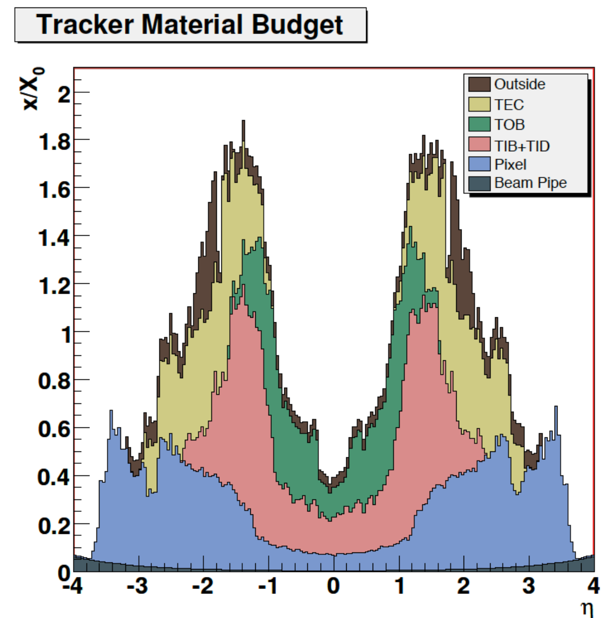
④ Most isolated τ_h kept in case multiple combinations pass mass cuts

⑤ Apply cut-based isolation or MVA based τ_h identification

⑥ Compute discriminators against e and μ

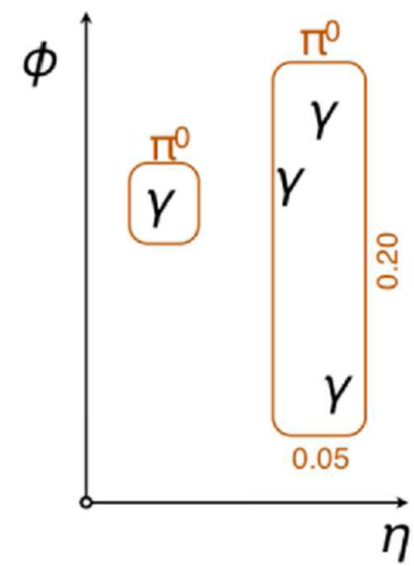
HPS π^0 Reconstruction

Neutral pions decay via $\pi^0 \rightarrow \gamma\gamma$ almost instantaneously



CMS has all silicon tracking detector

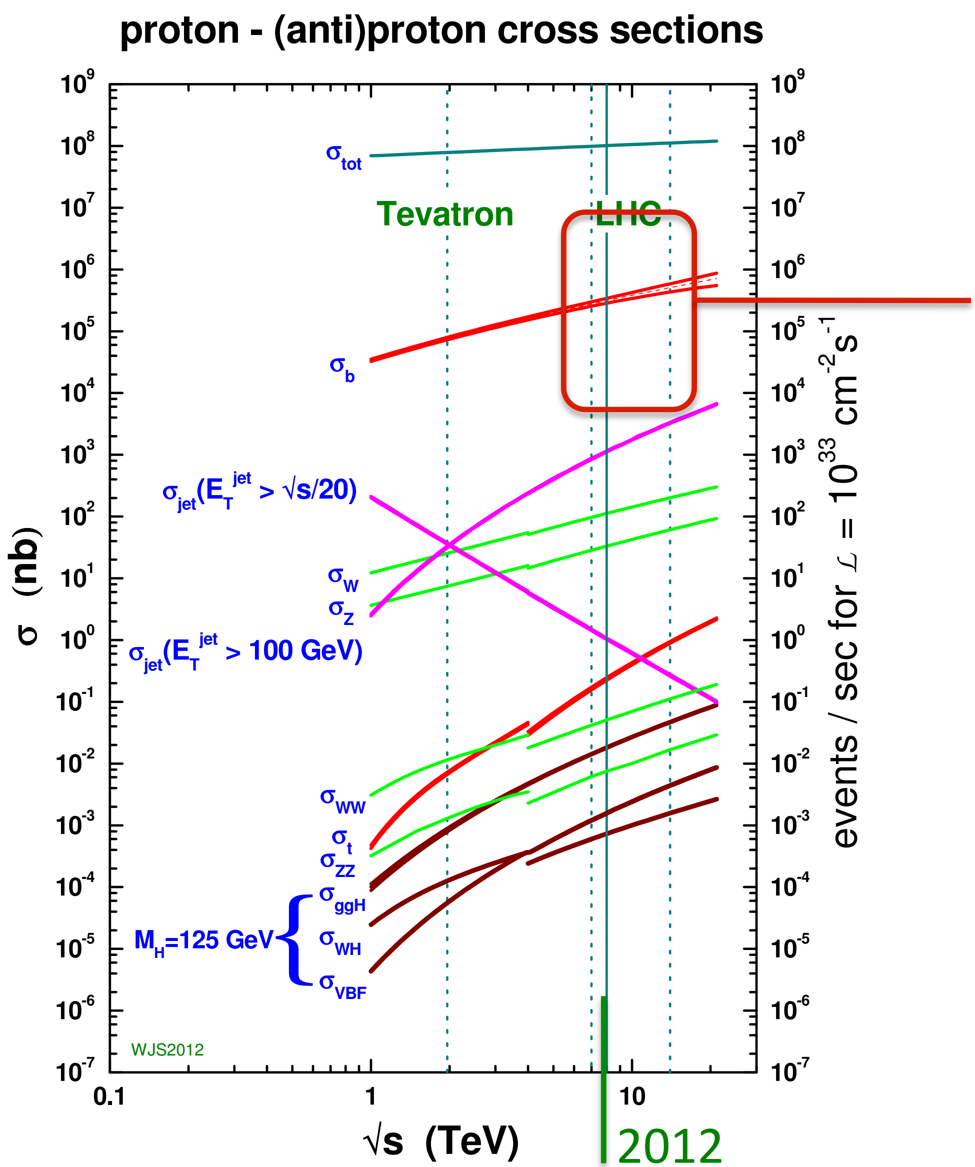
- ➔ High probability for photons to convert
- ➔ 3.8 T magnetic field separates e^+e^- pair in ϕ



- ➔ Reconstruct (converted) photons in η-ϕ strips

Strips may contain either both photons from π^0 decay or just one

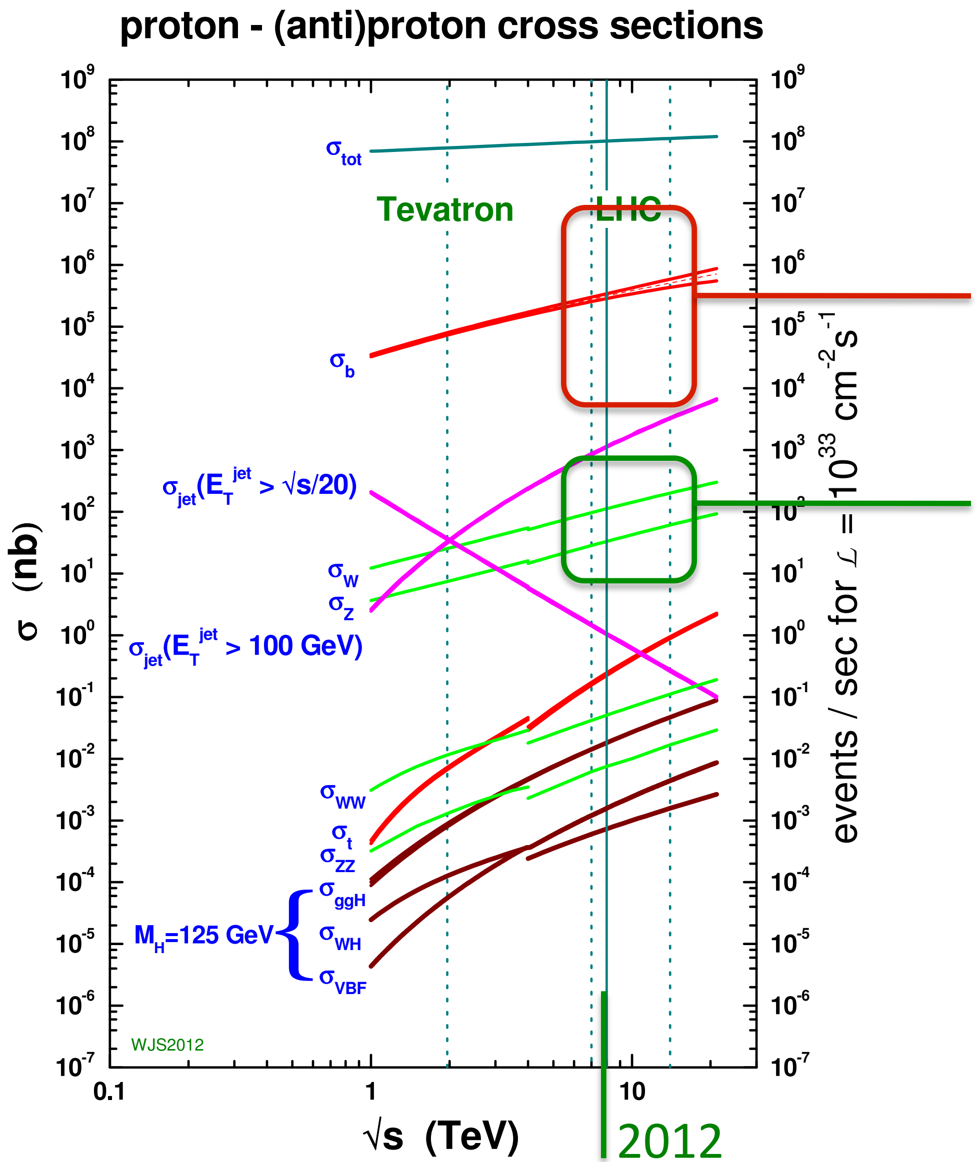
Taus and Backgrounds at the LHC



Large background from QCD multi-jet production

http://www.hep.ph.ic.ac.uk/~wstirlin/plots/crosssections2012_v5.pdf

Taus and Backgrounds at the LHC

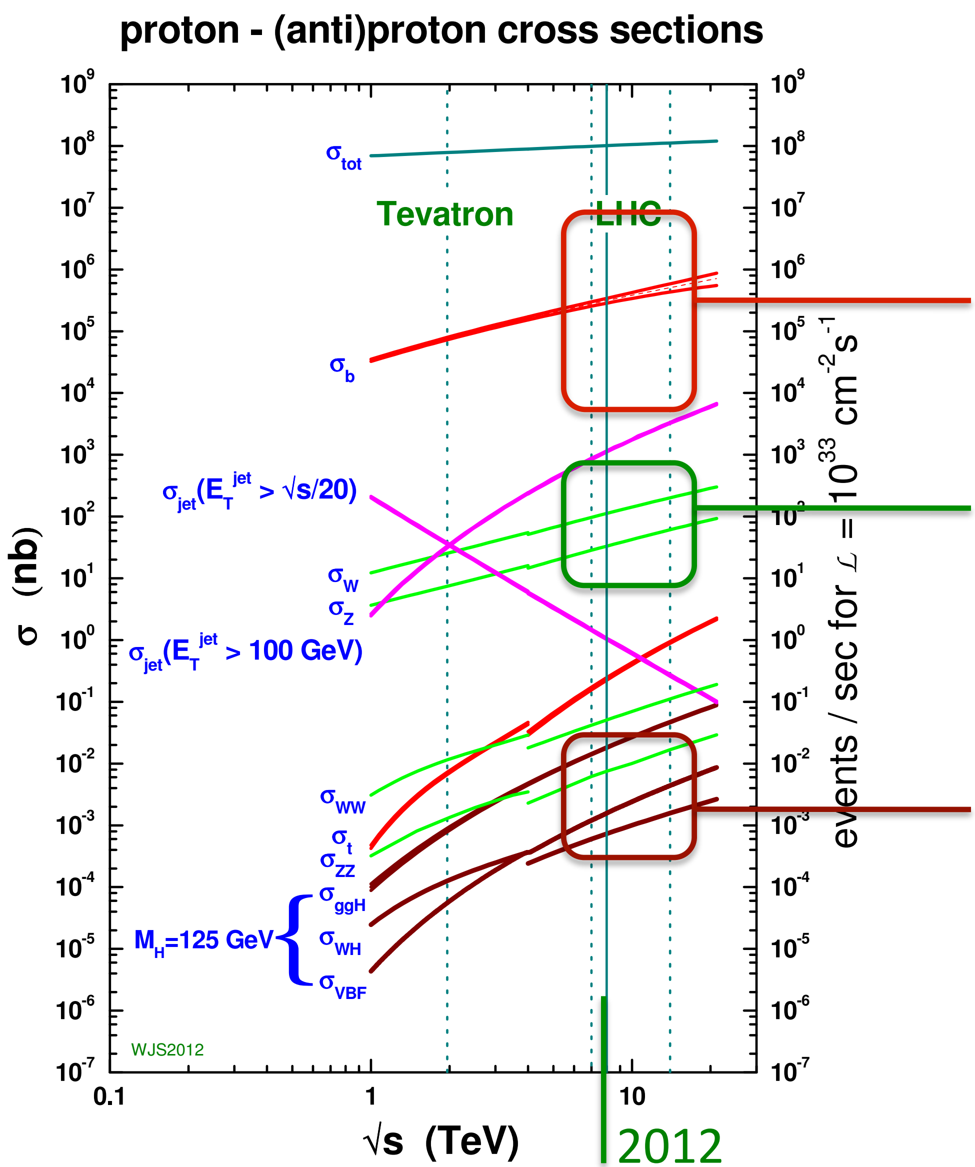


Large background from QCD multi-jet production

W → τν and Z/γ* → ττ are the dominant sources for taus in the Standard Model: "Standard Candle" for tau ID commissioning

http://www.hep.ph.ic.ac.uk/~wstirlin/plots/crosssections2012_v5.pdf

Taus and Backgrounds at the LHC



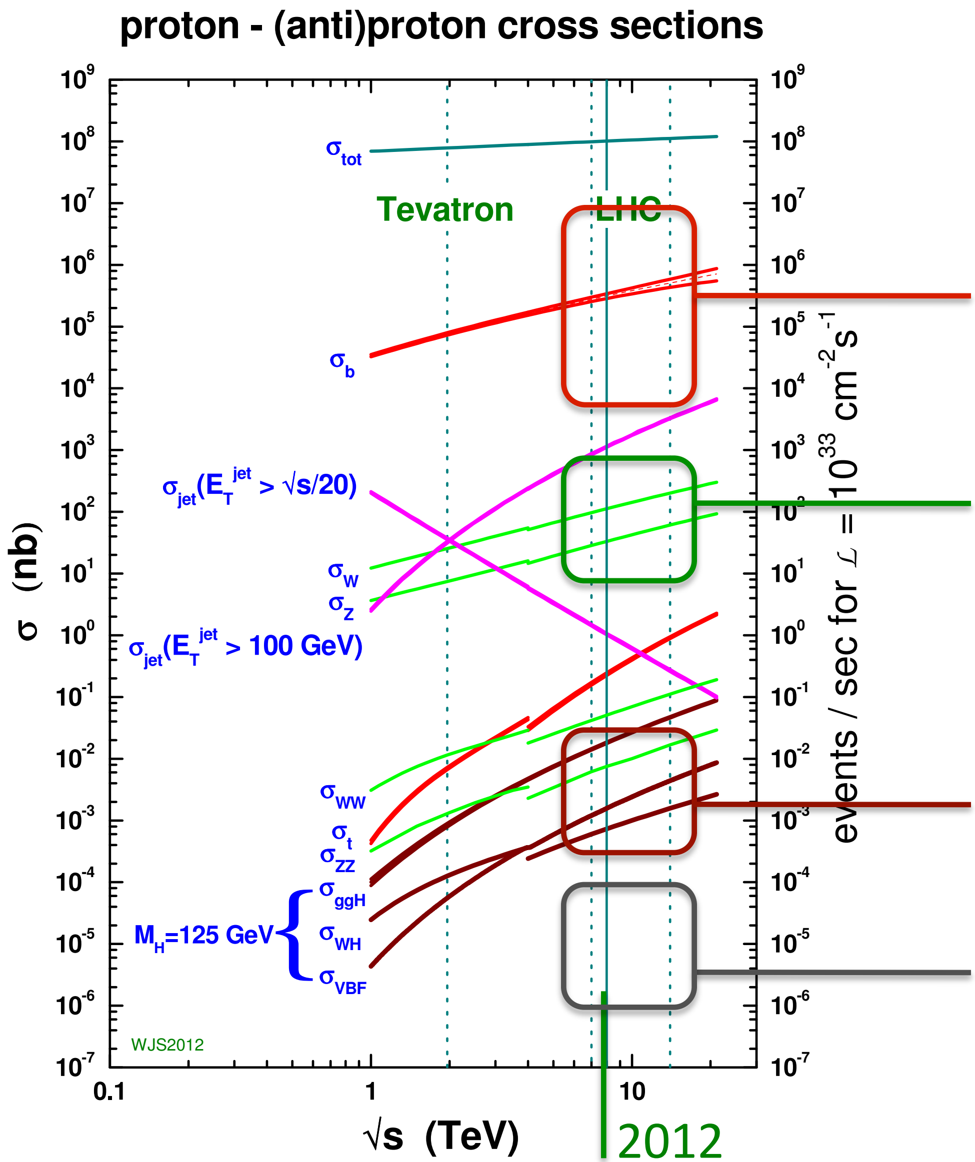
Large background from QCD multi-jet production

W → τν and Z/γ* → ττ are the dominant sources for taus in the Standard Model: "Standard Candle" for tau ID commissioning

Higgs → ττ

http://www.hep.ph.ic.ac.uk/~wstirlin/plots/crosssections2012_v5.pdf

Taus and Backgrounds at the LHC



Large background from QCD multi-jet production

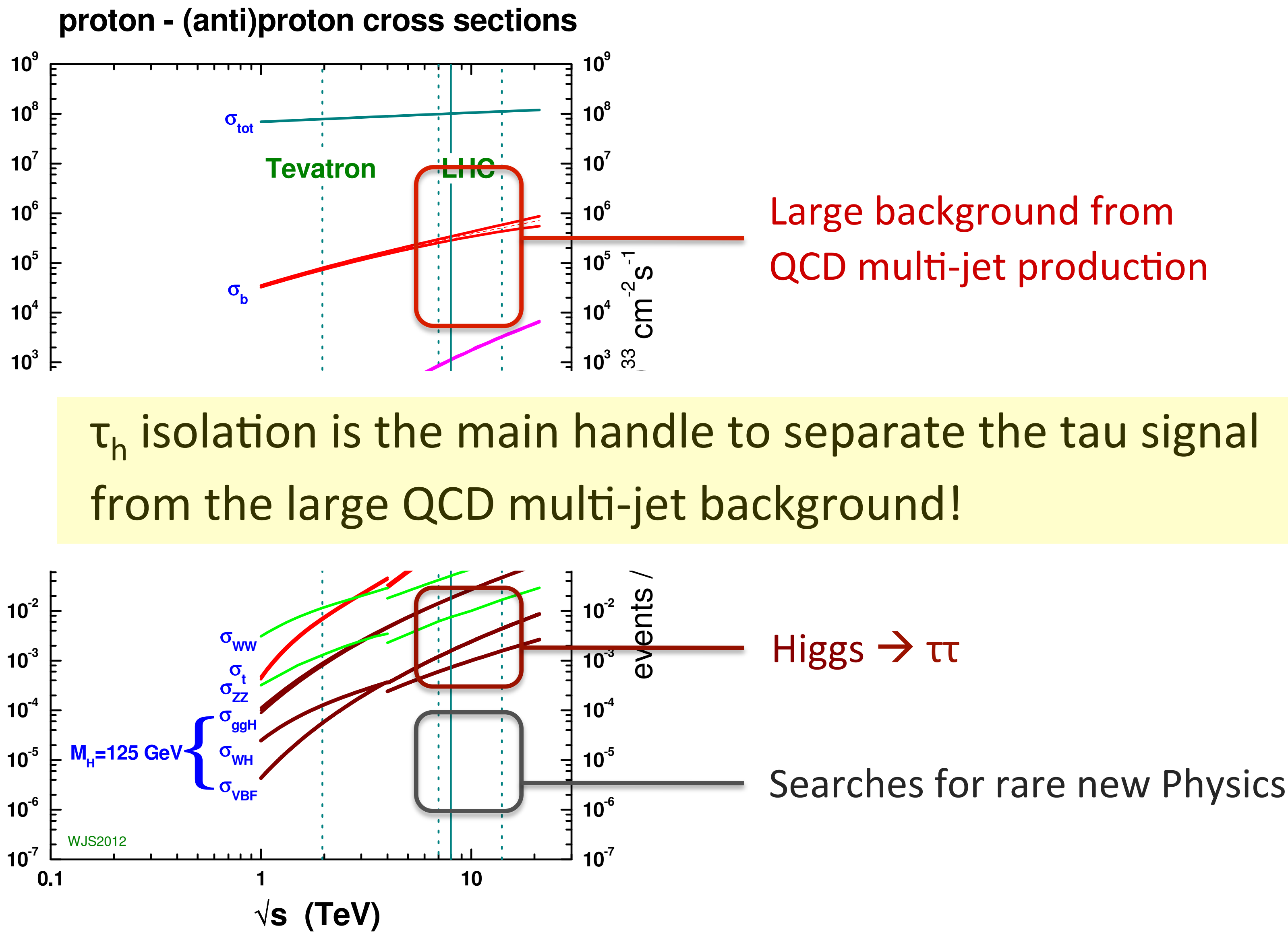
$W \rightarrow \tau\nu$ and $Z/\gamma^* \rightarrow \tau\tau$ are the dominant sources for taus in the Standard Model: "Standard Candle" for tau ID commissioning

Higgs $\rightarrow \tau\tau$

Searches for rare new Physics

http://www.hep.ph.ic.ac.uk/~wstirlin/plots/crosssections2012_v5.pdf

Taus and Backgrounds at the LHC



τ_h isolation is the main handle to separate the tau signal from the large QCD multi-jet background!

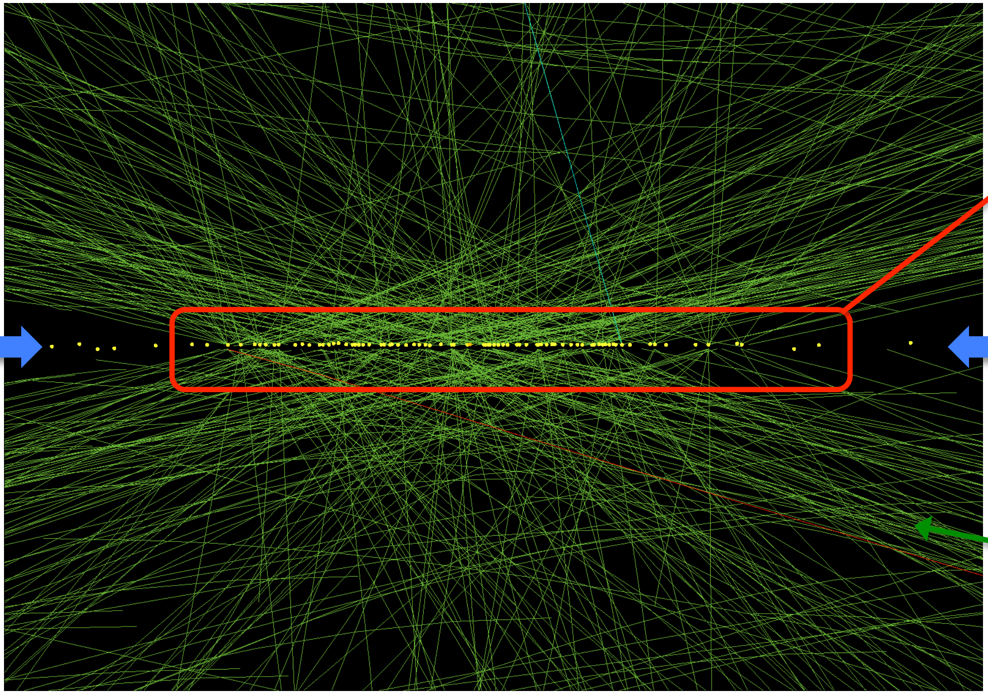
http://www.hep.ph.ic.ac.uk/~wstirlin/plots/crosssections2012_v5.pdf

The Challenge: Pile-up

pp inelastic cross-section @ 8 TeV: 69.4 mb

Typical instantaneous luminosity in 2012: $5 \cdot 10^{33}\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}$

➔ O(25) simultaneous pp interactions expected per bunch-crossing

p →  ← p

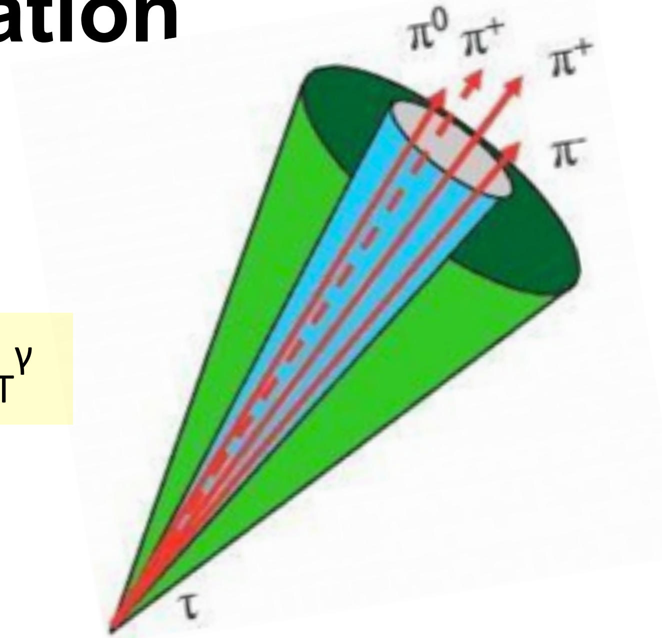
Each yellow dot represents 1 pp interaction vertex

Each green line represents 1 reconstructed track

CMS detector is hardly ever "empty"!

➔ Sophisticated methods are necessary to use particle isolation at the LHC

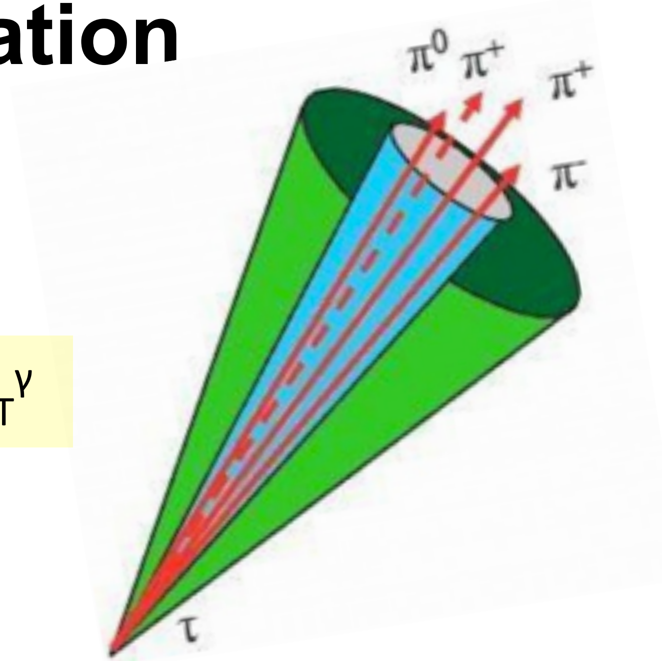
Cut-based Tau Isolation



Sum transverse momenta of charged hadrons plus photons within isolation cone of size $\Delta R = 0.5$

Isolation P_T-sum: $Iso = \sum P_T^{h\pm}(dZ < 2mm) + P_T^{\gamma}$

Cut-based Tau Isolation



Sum transverse momenta of charged hadrons
plus photons within isolation cone of size $\Delta R = 0.5$

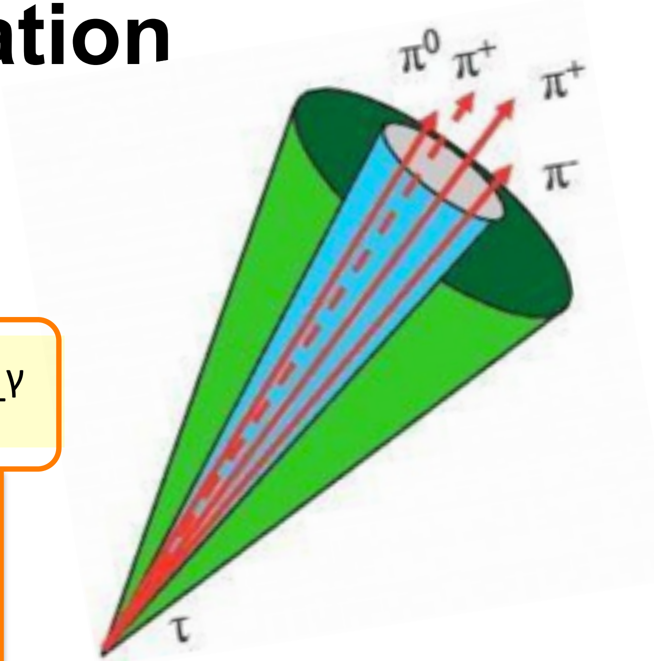
Isolation P_T-sum: $Iso = \sum P_T^{h\pm}(dZ < 2mm) + P_T^{\gamma}$

Charged hadrons of $P_T > 0.5$ GeV
within $dZ < 2mm$ of τ_h production vertex [1]

- Effect of PU on charged hadron isolation minimal due to $dZ < 2mm$ cut

[1] τ_h production vertex taken to be the vertex
that is closest to the highest P_T ("leading") track of the τ_h

Cut-based Tau Isolation



Sum transverse momenta of charged hadrons
plus photons within isolation cone of size $\Delta R = 0.5$

Isolation P_T-sum: $Iso = \Sigma P_T^{h\pm}(dZ < 2mm) + P_T^{\gamma}$

Charged hadrons of $P_T > 0.5$ GeV
within $dZ < 2mm$ of τ_h production vertex [1]

Photons of $P_T > 0.5$ GeV

- Effect of PU on charged hadron isolation minimal due to $dZ < 2mm$ cut
- PU contribution to photon isolation corrected for on statistical basis (see next slide)

[1] τ_h production vertex taken to be the vertex that is closest to the highest P_T ("leading") track of the τ_h

Δβ Correction

Aim: Compensate for pile-up contribution to isolation P_T-sum

- Compute P_T-sum of charged hadrons failing dZ cut within cone R = 0.5 around τ_h direction

- Subtract Δβ correction from photon isolation

$$\Delta\beta = k \bullet \Sigma P_T^{h\pm}(dZ > 2mm)$$

k = ratio of photon/charged hadron energy per unit area in PU events

(k ≈ 0.5 expected from Isospin symmetry, actual value slightly different due to P_T thresholds and reconstruction efficiencies)

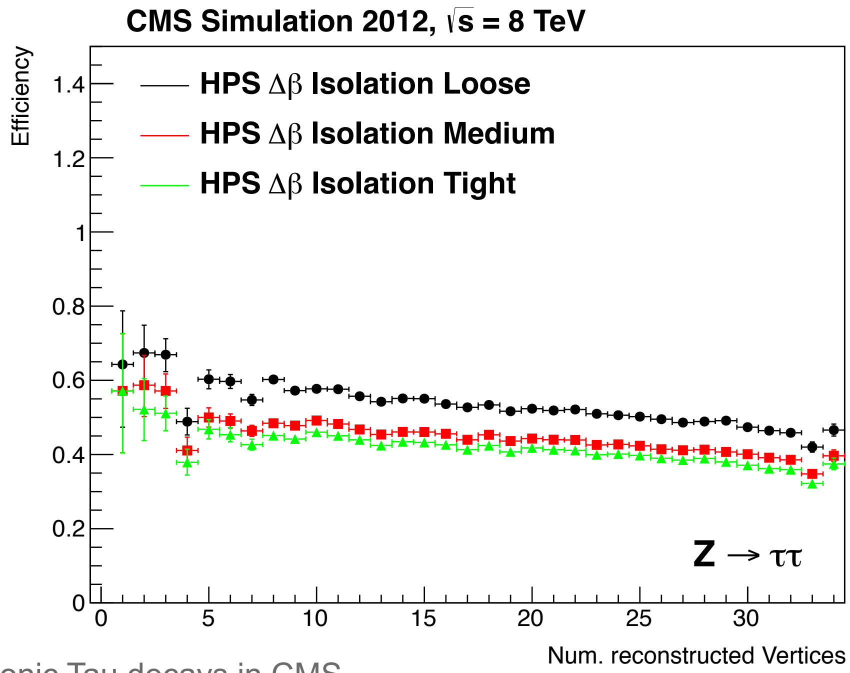
Working-Points

Different cuts on Iso define Loose, Medium and Tight tau ID working-points:

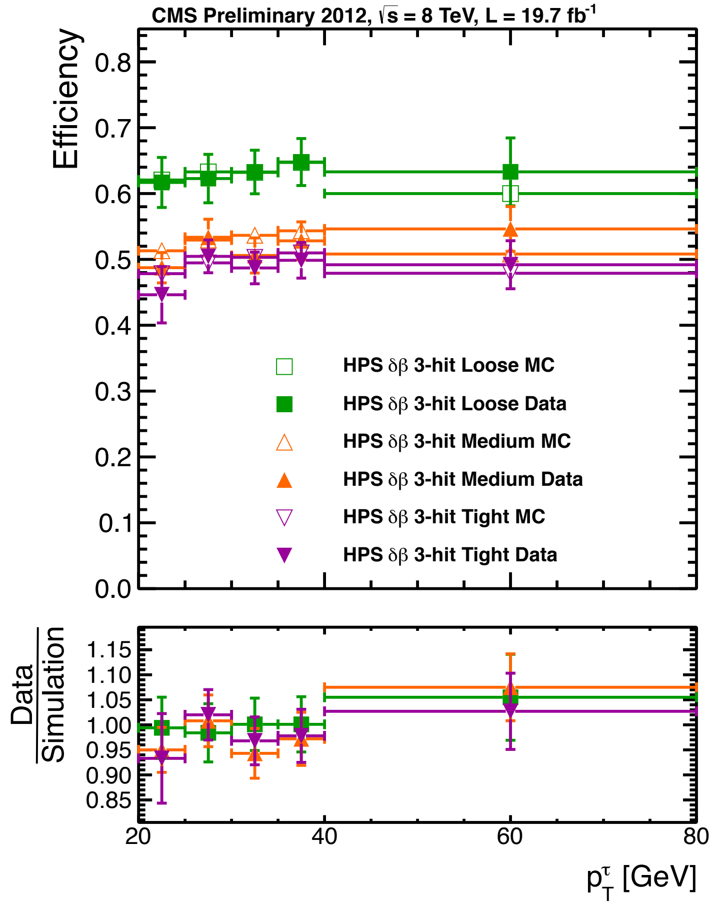
WP	Requirement
Loose	Iso < 2.0 GeV
Medium	Iso < 1.0 GeV
Tight	Iso < 0.8 GeV

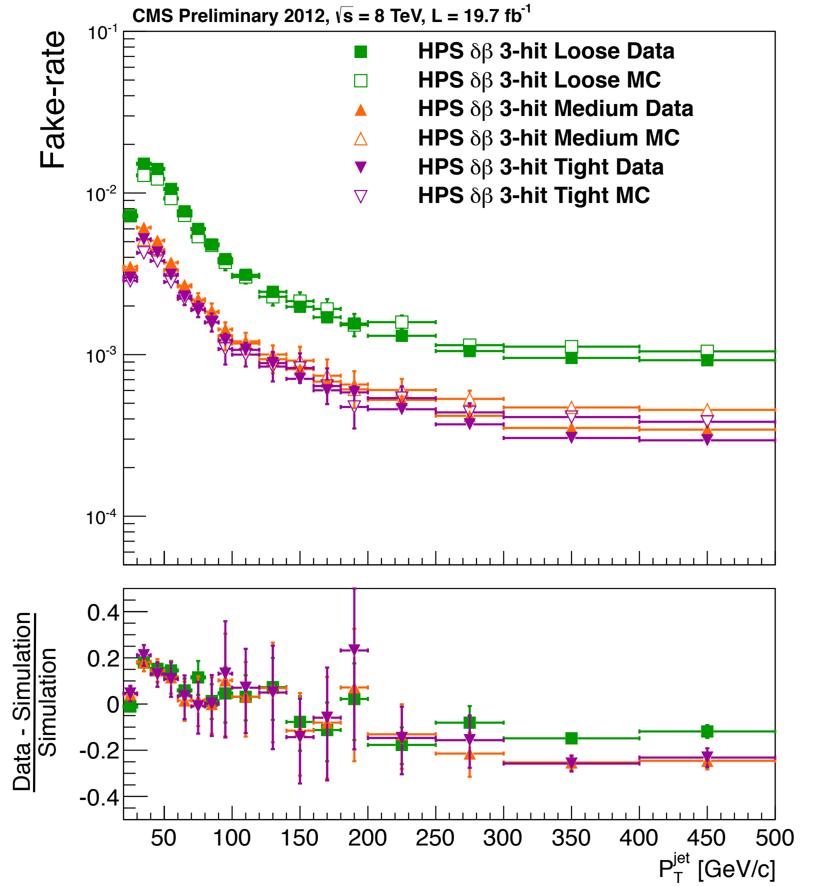
All 3 WPs are actually quite tight: necessary to reduce large QCD background

➔ Tau ID efficiency robust vs. PU if $\Delta\beta$ correction is used



Tau ID Performance (Cut-based)





- ➔ Tau ID efficiency ~60%, flat vs. P_T for $P_T^\tau \geq 40$ GeV
- ➔ Jet → τ_h fake-rate ~1-2% at low P_T, steeply falling with P_T
- ➔ Good agreement between Monte Carlo simulation and data in terms of tau ID efficiency, up to 20% difference in jet → τ_h fake-rate

MVA Tau ID

Idea: Use tau lifetime information to improve τ_h ID performance

Input Variables:

- ΣP_T(charged contribution).
- ΣP_T(neutral contribution).
- Pile up correction.

Cut based isolation variables

- Reconstructed τ decay mode.
- $P_T(\tau)$
- $\eta(\tau)$

*Used to parameterize P_T and η dependence of the MVA input variables

τ reco and kinematic variables

- abs(d_{xy}) of the τ .
- sign(d_{xy}) of the τ .
- d_{xy} significance.

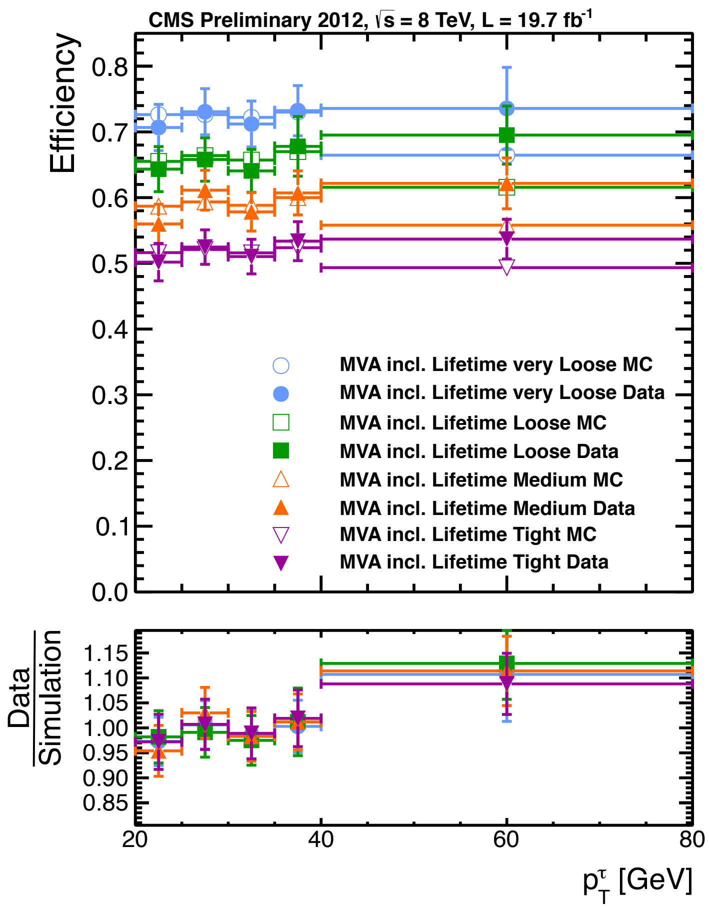
Variables sensitive to the impact parameter

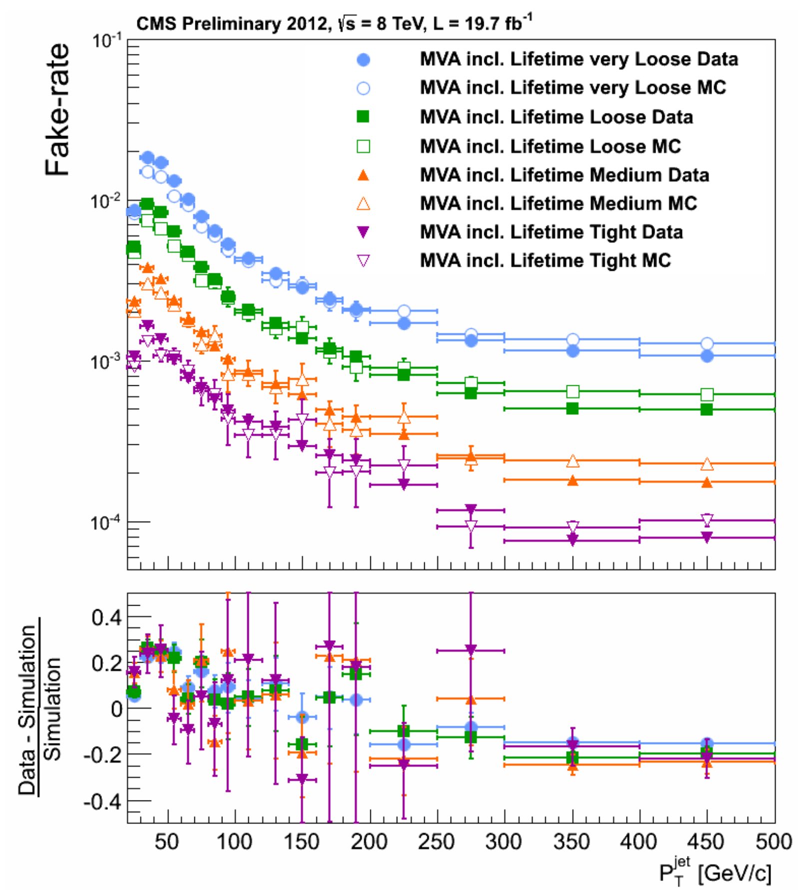
- Presence of a secondary vertex.
- τ flight length.
- τ flight length significance.

SV sensitive variables

➔ Loose, Medium, Tight WPs defined by cut on BDT output

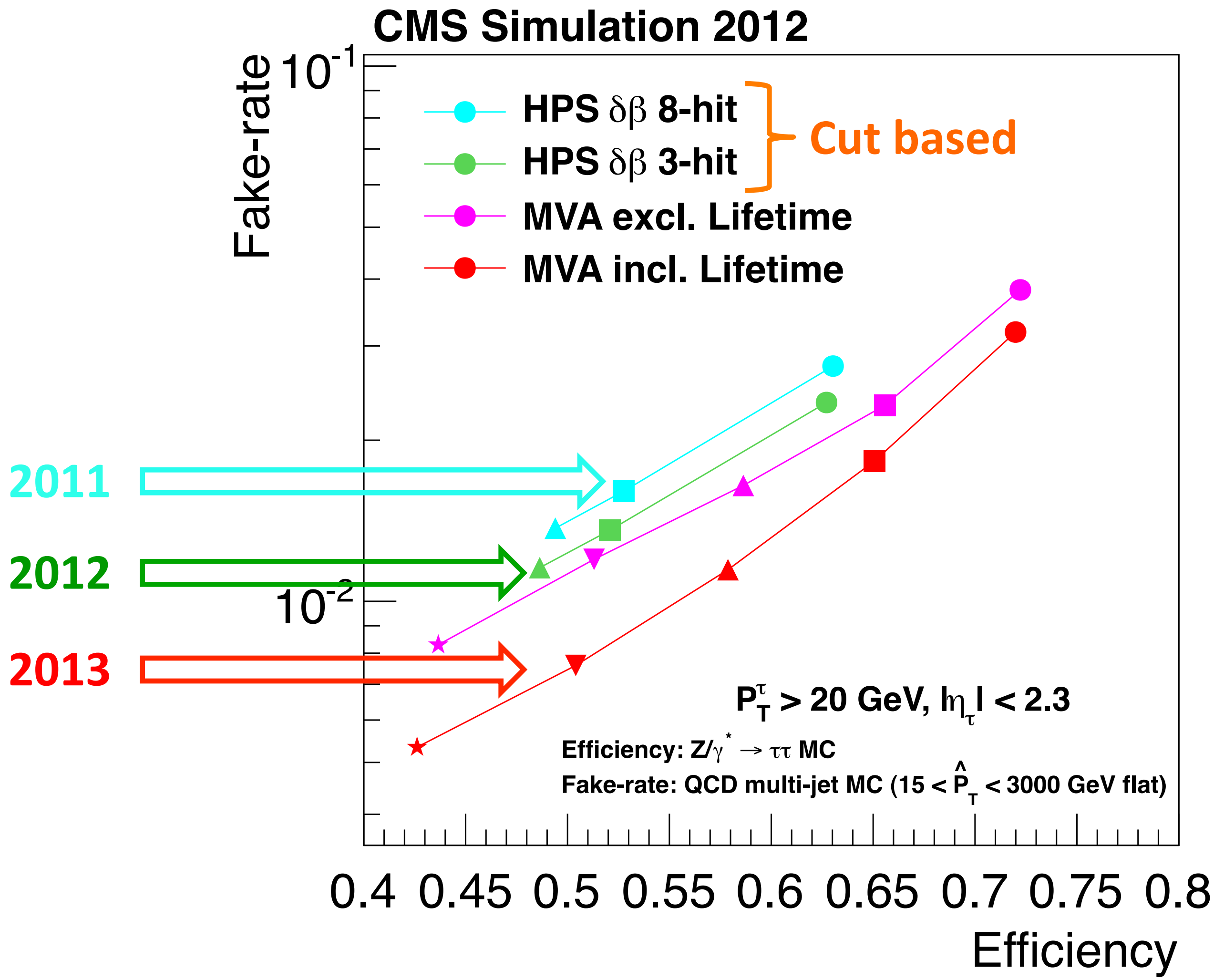
Tau ID Performance (MVA)





- ➔ MVA τ_h ID covers a wide range of tau ID efficiency and fake-rate
- ➔ Very Loose working-point yields tau ID efficiency of ≳ 70%, Tight WP provides fake-rate of 10^{-4} (for high P_T jets)
- ➔ Agreement between simulation and data very similar to cut-based τ_h ID

Tau ID Efficiency vs. Fake-rate



➔ Tau ID algorithm performance improved steadily during LHC run 1, as we gained more experience with the detector and LHC running conditions

e/μ Rejection

Electrons and muons are typically isolated

→ e/μ appear like one-prong hadronic tau decays in tracking detector

→ Information provided by calorimeter/muon system used to discriminate e/u from τ_h

Anti-μ discriminator:

- Cut-based, veto τ_h candidates in case there are nearby track segments in muon system
- Efficiency ≥ 95%, fake-rate ~0.1%

Anti-e discriminator:

- Dedicated MVA trained to separate e from τ_h, using typical electron ID variables as input
- Efficiency ~90% (~70%), Fake-rate ~0.5% (~0.1%) for Loose (Tight) WP

Performance measured in data using Z/γ* → ee (Z/γ* → μμ) events

Summary

- The particle-flow and "Hadron plus Strips" algorithms allow to reconstruct individual hadronic tau decay modes

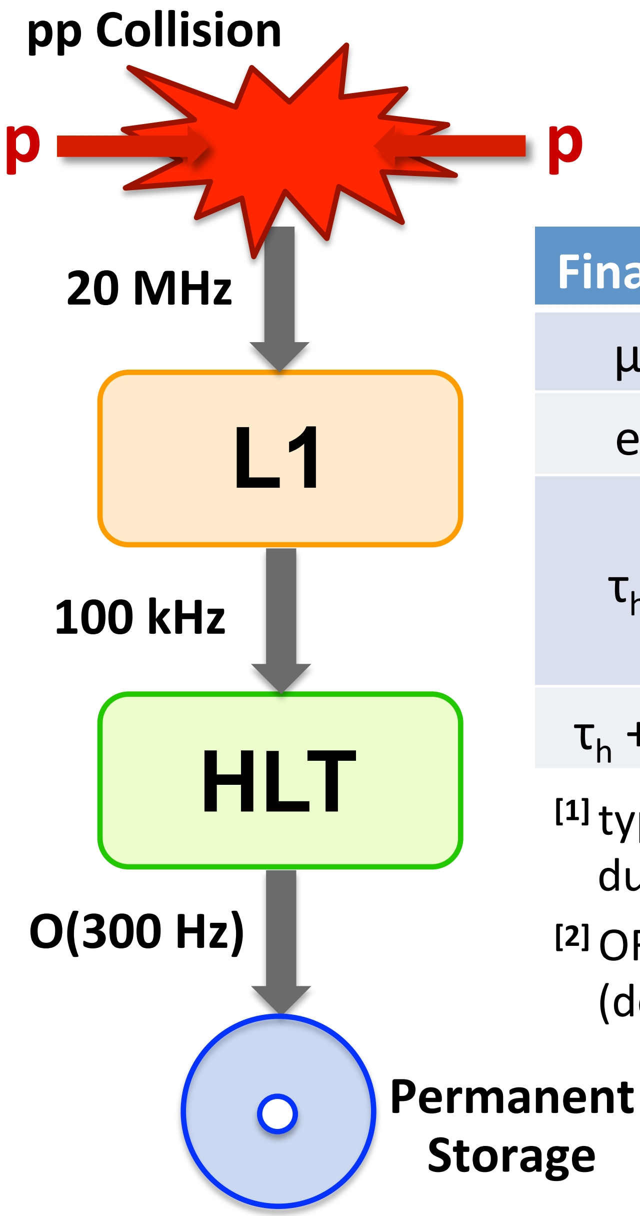
- CMS tau ID performed well during LHC run 1:
  - τ_h identification efficiency typically 50-60%
  - Probability for jets, electrons and muons to be misidentified as hadronic tau decays typically few per mille to O(1%)
  - τ_h reconstruction has been robust against pile-up

- LHC is about to restart taking data at 13 TeV center-of-mass energy, higher pile-up and 25ns collisions next year

➔ CMS is making good progress to prepare its event reconstruction as well as its triggers for these challenging conditions!

Backup

Overview of Tau Triggers



During 2012 data-taking period

Final State	L1	High Level Trigger
$\mu + \tau_h$	SingleMu14 [1]	IsoMu18 + PFTau20
$e + \tau_h$	SingleIsoEle20 [1]	IsoEle22 + PFTau20
$\tau_h + \tau_h$	DoubleTau44 OR DoubleCentralJet64 [2]	DoublePFTau35 DoublePFTau30 + Jet30
$\tau_h + E_T^{miss}$	MET40 [1]	PFMET70 + PFTau35

[1] typical threshold. Small variation of the thresholds occurred during the data-taking period.

[2] OR recovers inefficiency of L1 tau isolation requirements (described on next slide)

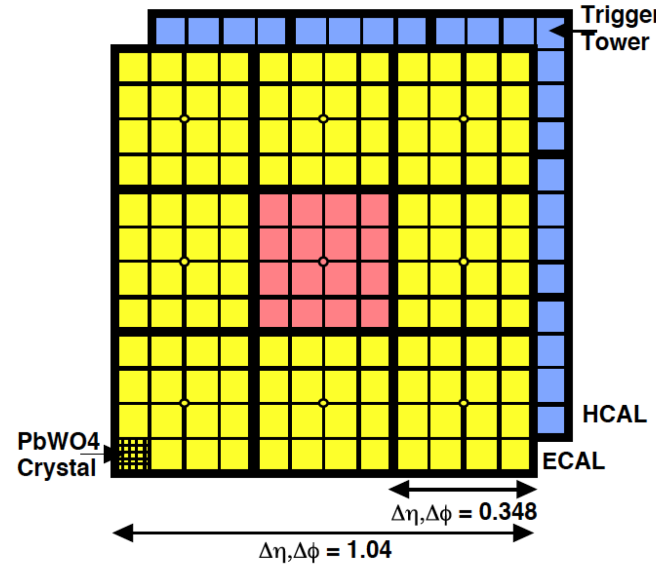
L1 Tau Trigger

Reconstruction of L1 Tau P_T:

Sum of energy deposits in ECAL plus HCAL in a region of 12x12 Trigger Towers

The sum of energy deposits is calibrated using calibration parameters determined for jets.

The calibration overcorrects the true energy of taus by a factor O(1.5)

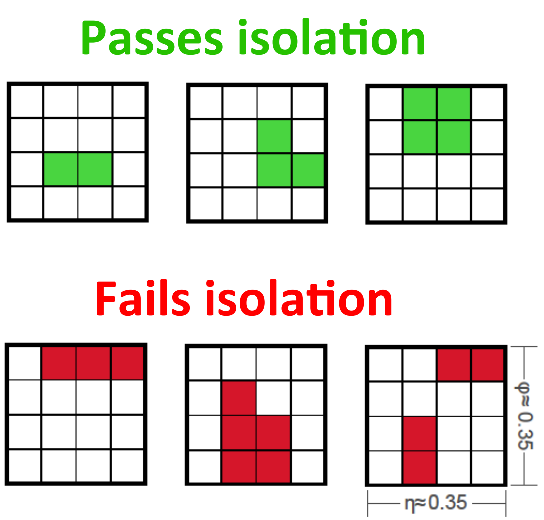


L1 Tau isolation:

- 7 of the 8 non-central 4x4 regions are required to contain $E_T < 2$ GeV
- Energy within the central 4x4 region [1] is required to be concentrated in a region of at most 2x2 Trigger Towers

  Trigger Towers outside of the 2x2 region are required to satisfy $E_T^{ECAL} < 4$ GeV and $E_T^{HCAL} < 4$ GeV

[1] and each other 4x4 region



HLT Tau Trigger ($\tau_h\tau_h$)

L2

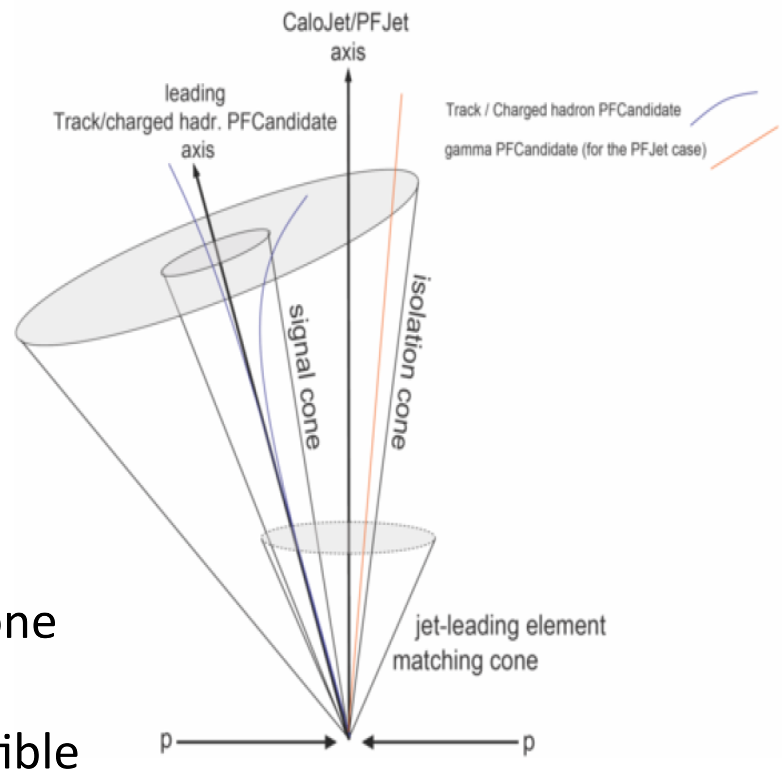
- Reconstruct calorimeter jets in η-ϕ regions matched to L1 Tau OR L1 CentralJet seeds, using iterative cone algorithm with $R = 0.2$
- Require 2 calorimeter jets of $P_T > 30$ GeV and $|\eta| < 2.1$ [1]

[1] 3 calorimeter jets of $P_T > 30$ GeV and $|\eta| < 3.0$ required for DoublePFTau30 + Jet30 trigger. 2 of the jets are required to be within $|\eta| < 2.1$.

L2.5

- Reconstruct tracks from hits in pixel detector, within regions seeded by calorimeter jets
- Reconstruct "hard scatter" vertex
- Require that there be no pixel tracks of $P_T > 1.2$ GeV within annulus of size $0.2 < \Delta R < 0.4$ around jet axis

HLT Tau Trigger ($\tau_h \tau_h$)



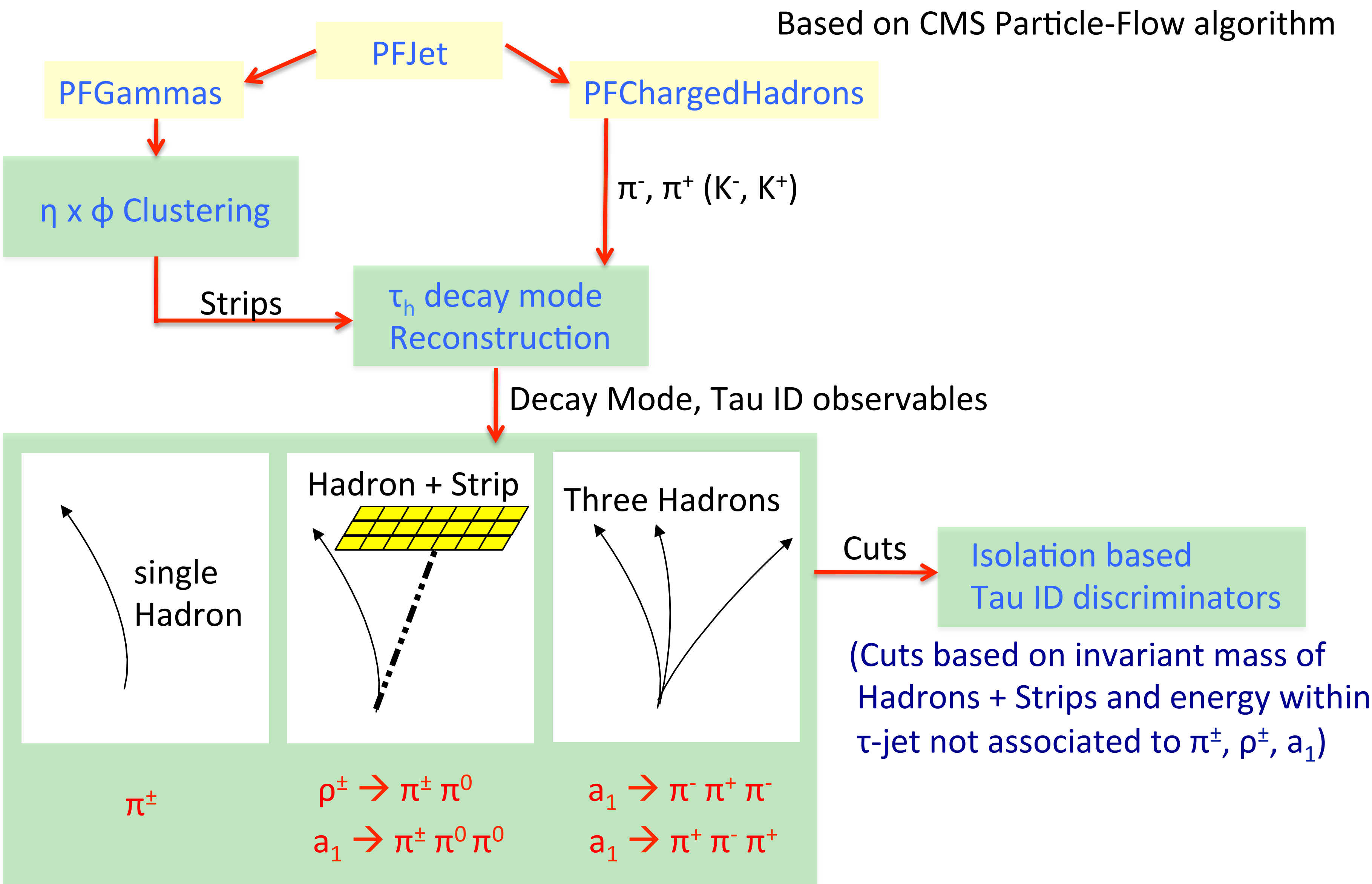
L3

- Reconstruct tracks from hits in pixel plus strip detector
- Run particle-flow algorithm
- Run simple τ_h reconstruction:
  - P_T^τ = sum of particles within "signal cone" of size $\Delta R = 0.15$
  - Highest P_T ("leading") track within signal cone required to pass $P_T > 1$ GeV [1]
  - Require no tracks of $P_T > 1$ GeV and compatible with originating from "hard scatter" vertex within isolation annulus of size $0.15 < \Delta R < 0.45$ around leading track

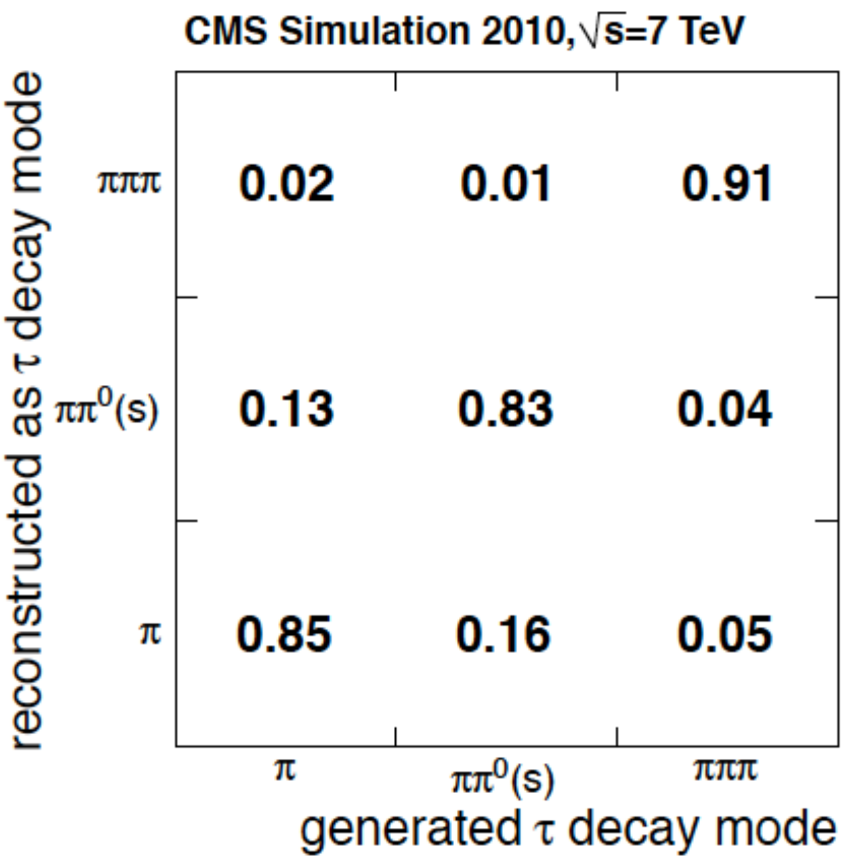
[1] 5 GeV threshold required during part of 2012 data-taking period

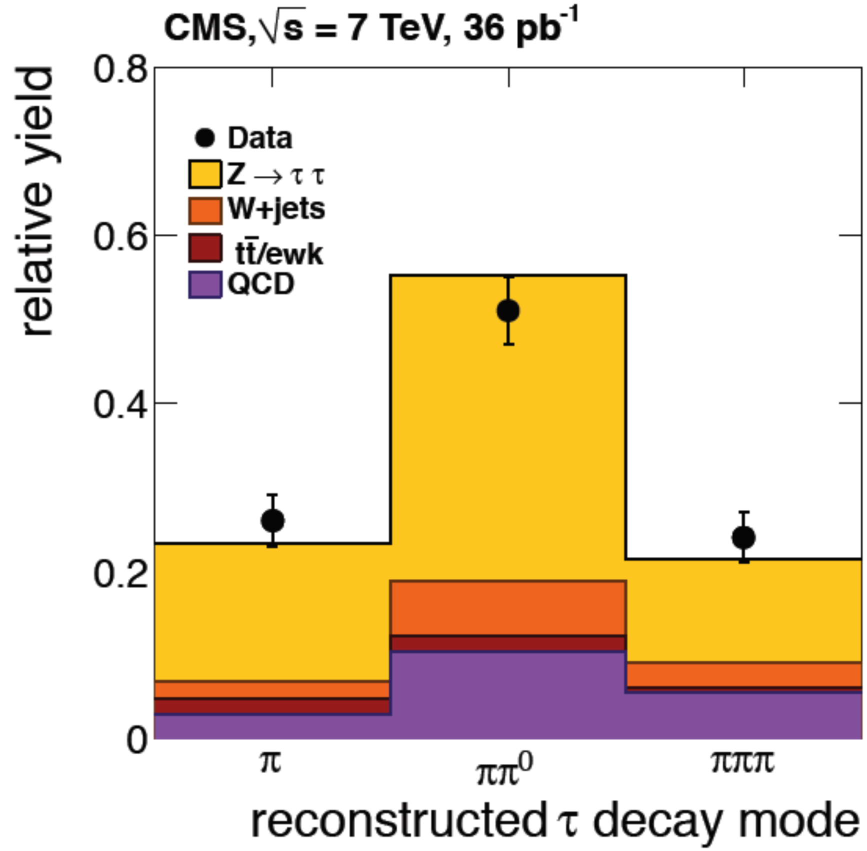
L2, L2.5 and L3 are "virtual" trigger levels, implemented in software
(running simplified fast version of the offline CMS event reconstruction algorithms)

HPS Tau Reconstruction



Decay Mode Reconstruction



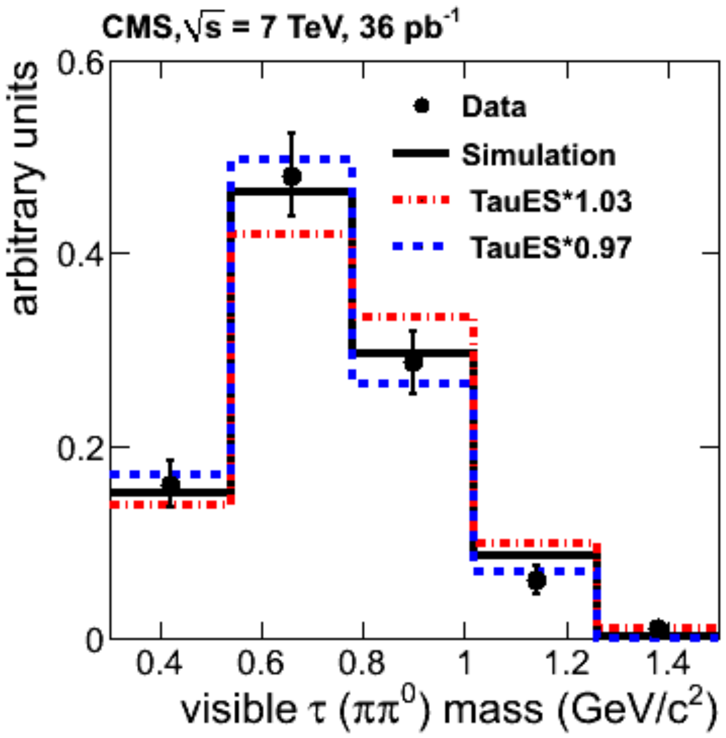


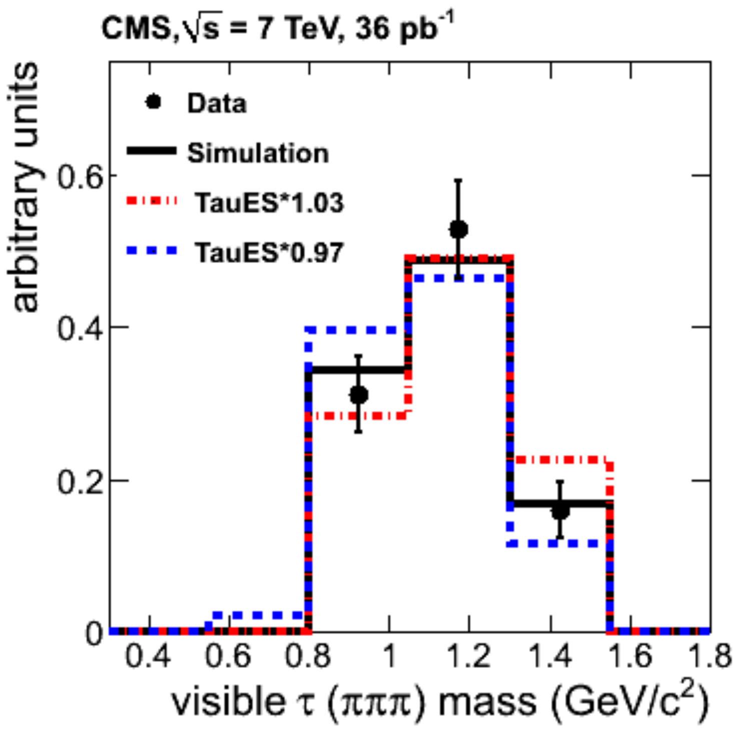
- Purity of tau decay mode reconstruction 80-90%
- Data well modeled by Monte Carlo simulation

Tau Energy Scale

Validated by 2 Methods:

- Fit of visible Mass peak in Z → ττ → $\mu\tau_{had}$ Events
- Data/MC Comparison of τ_{had} Mass





Tau energy scale stable vs. PU

Measurement of e → τ_h and μ → τ_h Fake-rates

e → τ_h and μ → τ_h fake-rates measured via Tag & Probe technique using Drell-Yan events:

- $Z/\gamma^* \to e^+e^-$ for e → τ_h fake-rate
- $Z/\gamma^* \to \mu^+\mu^-$ for μ → τ_h fake-rate

$Z/\gamma^* \to e^+e^-$ ($Z/\gamma^* \to \mu^+\mu^-$) events are:

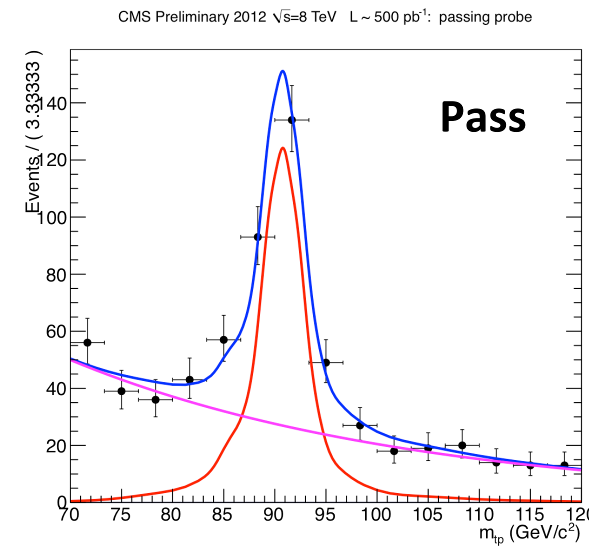
- triggered by single electron (muon) triggers
- selected by requiring 1 well identified and isolated electron (muon) [1] plus 1 τ_h candidate of $P_T > 20$ GeV, $|\eta| < 2.3$ and opposite charge [2]

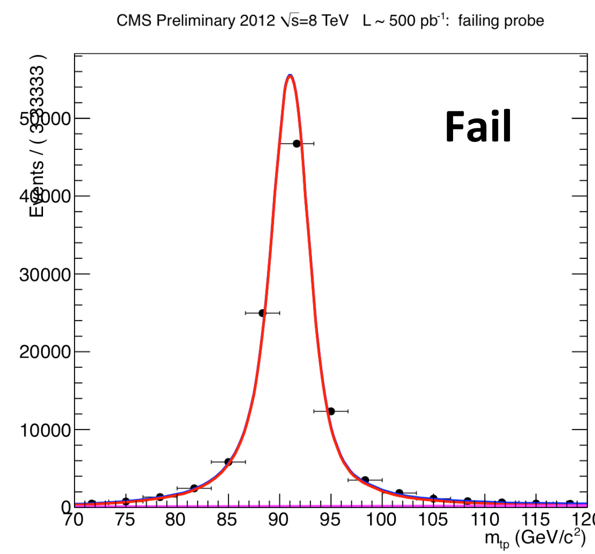
Selected events are split into 2 categories:

- **Pass**: passing tau ID and anti-e (anti-μ) discriminator
- **Fail**: not in **Pass** category

The Drell-Yan contribution to each category is determined by fitting the distribution of m_{tp}, the mass of the Tag + Probe pair

→ $$FR = \frac{N^{DY}_{pass}}{N^{DY}_{pass} + N^{DY}_{fail}}$$





[1] the "Tag"   [2] the "Probe"; no isolation or other tau ID cuts are applied on the probe