chi2 current CMSSW implementation, known limitations & possibilities for 6_X_X

these slides are part of the: “Discussion on rechit flagging and the rejection of anomalous events”

ECAL DPG, July 19th 2012

K. Theofilatos

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

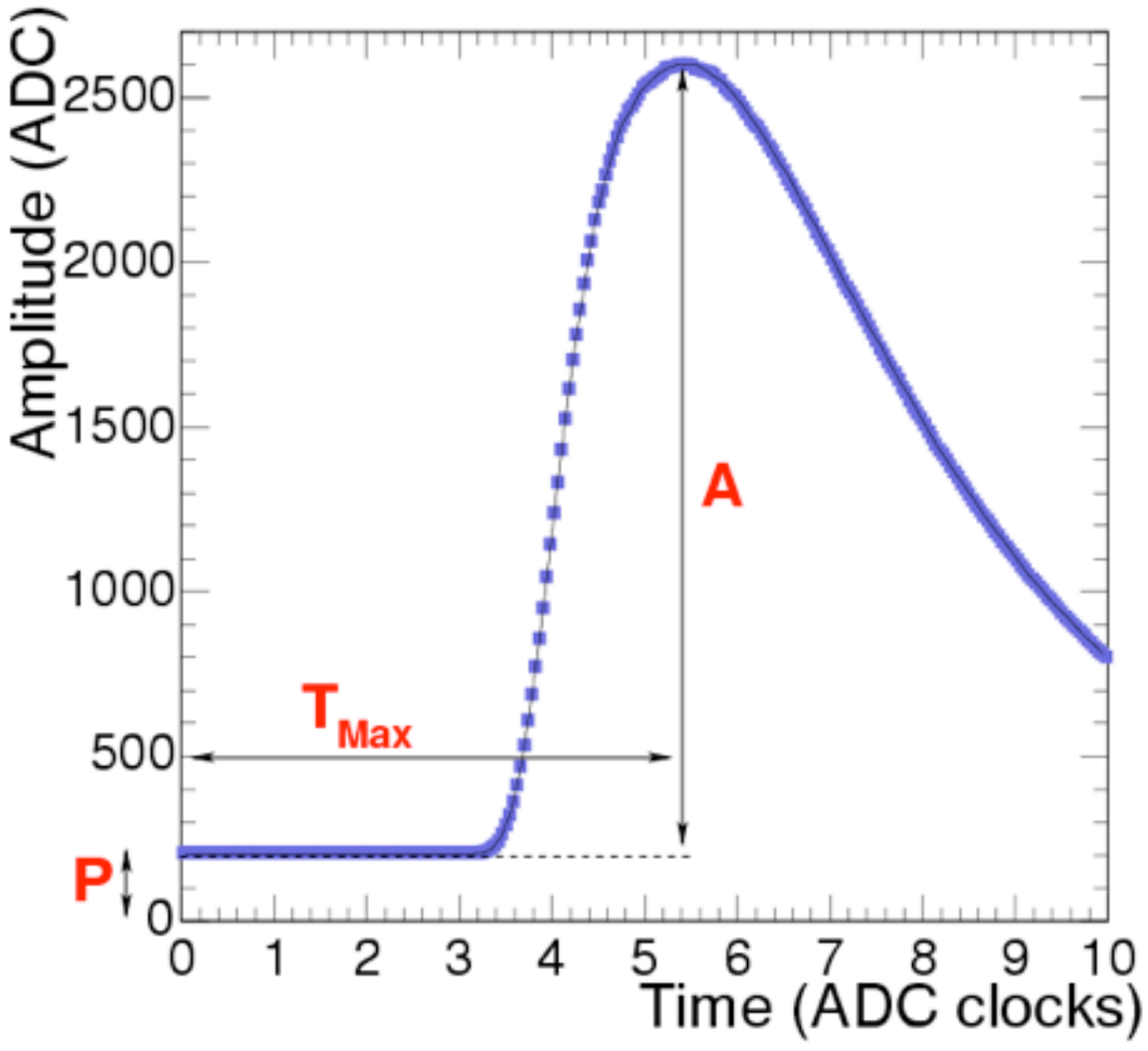
Reminder about chi2

$\chi^2 = \Sigma\ (R_i/\delta R_i)^2$
where $R_i = (S_i - Af_i - P)$

the reference TB shape is used to estimate chi2

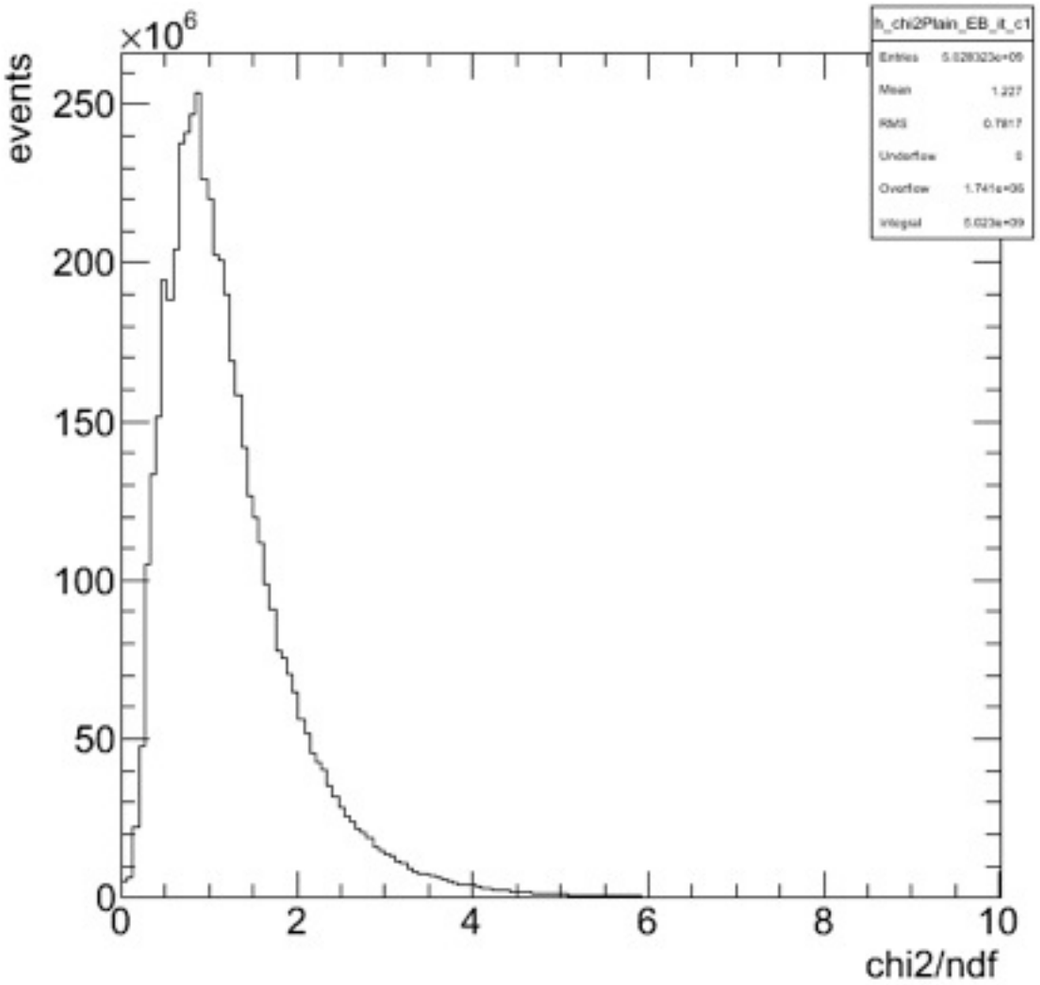
$\delta R_i^2 \sim pedRMS^2 + (A\delta f_i)^2$

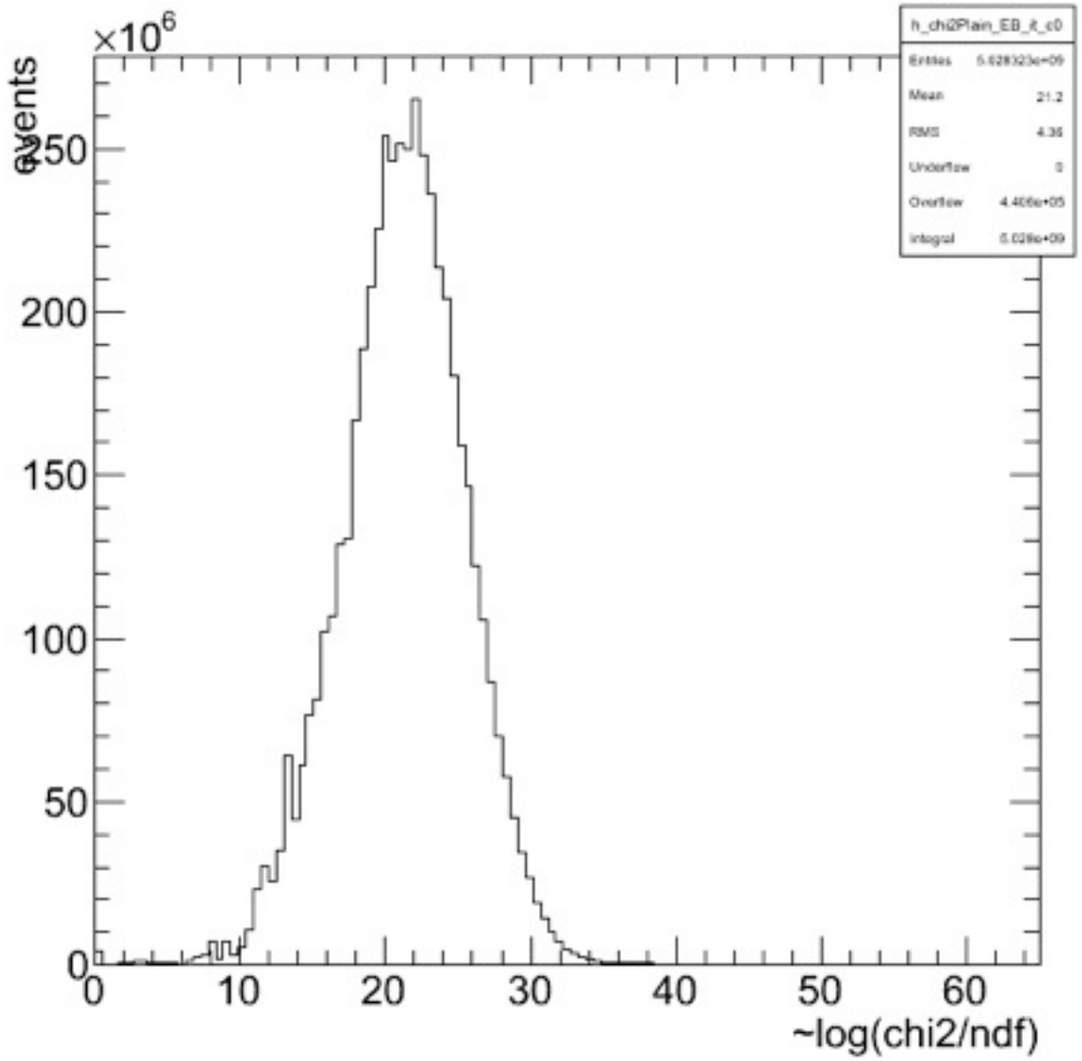
δf_i =sys uncertainty on the shape,
A=amplitude,
P=pedestal



- model the residuals: $<\Sigma\ R_i^2> = N^2 + C^2A^2$
- Noise term related to correlated noise and pedestal fluctuations
- Constant term related to the **sys** uncertainty on the shape

What is actually stored





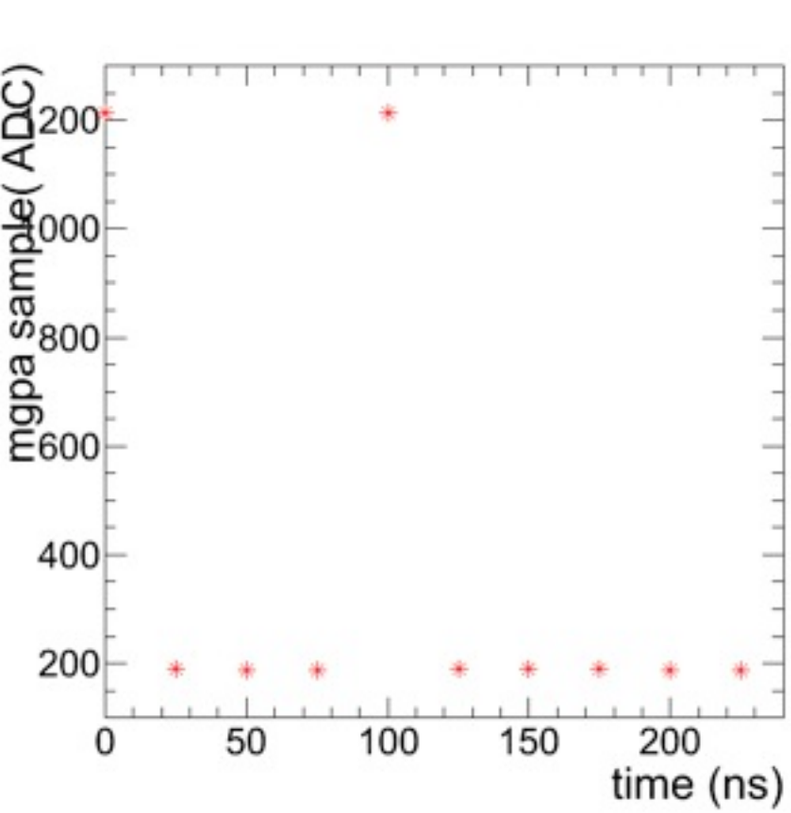
- convenient to use the log(chi2/ndf) instead of the chi2
- calibRechit preserves space in bits for the 0-64 range
- 7[3+log(chi2/ndf)] better exploits available space in the

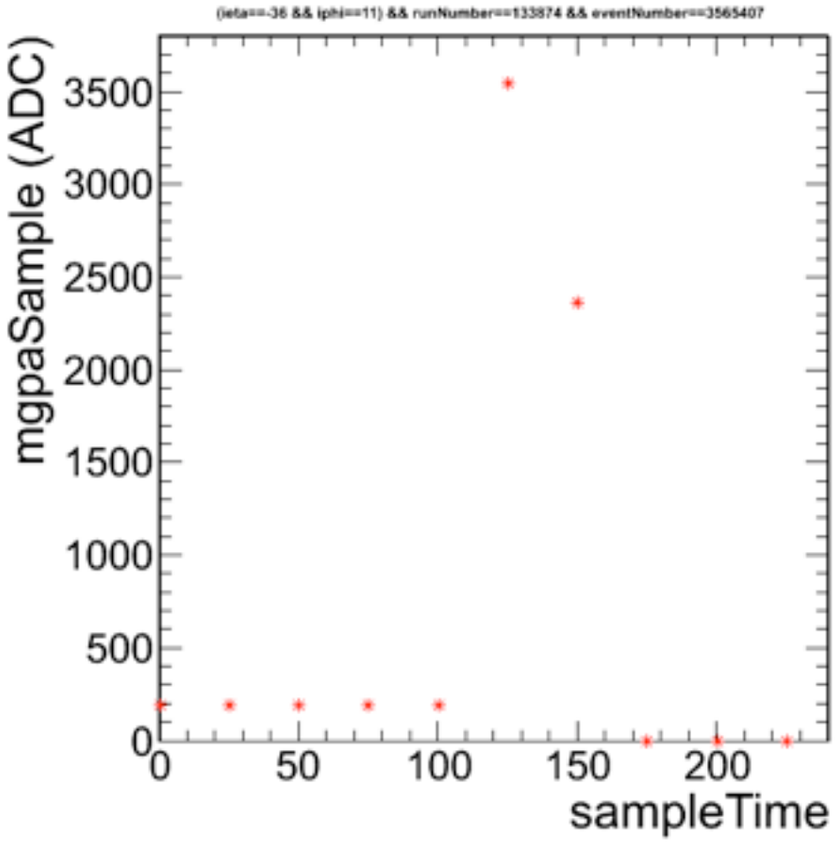
Current implementation

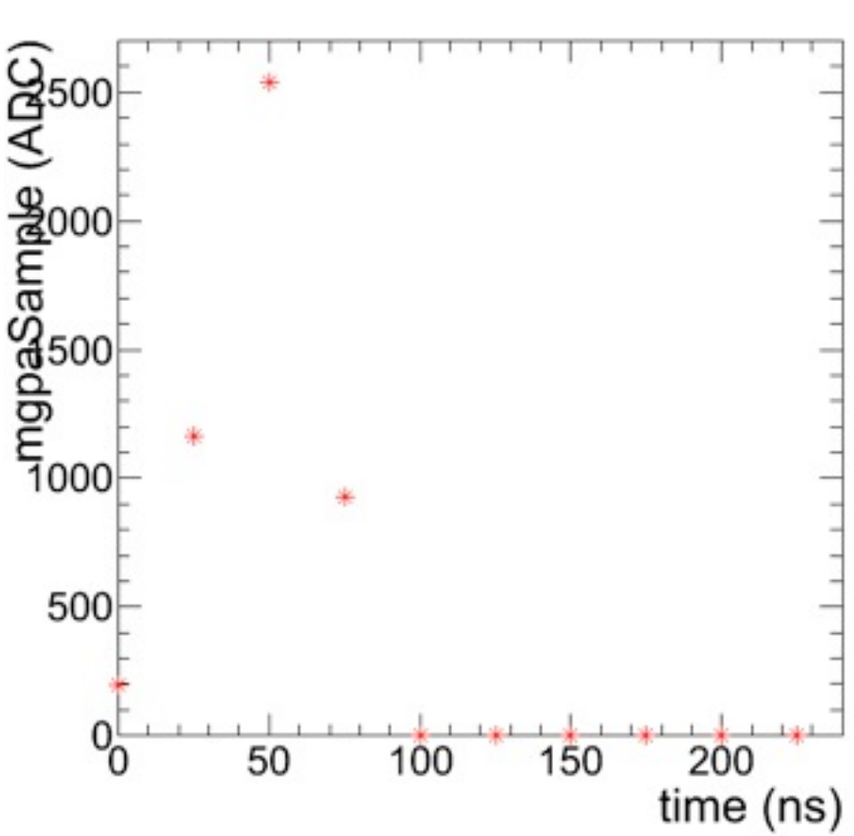
- Two flavors are implemented in the CMSSW:
  - chi2 (for in-time pusles) uses:
    - amplitude from 3+5 weights
    - offline time intercalibrations
  - out-of-time chi2 uses:
    - amplitude from Ratio method
    - signal's measured time
- None of them covers gain switched pulses
- Reference shapes are from testbeam (TB)
- !G12 pulses are characterized as good by default (exception: pulses with readout errors)

Some (rare*) examples

pulses with problematic read-out pattern



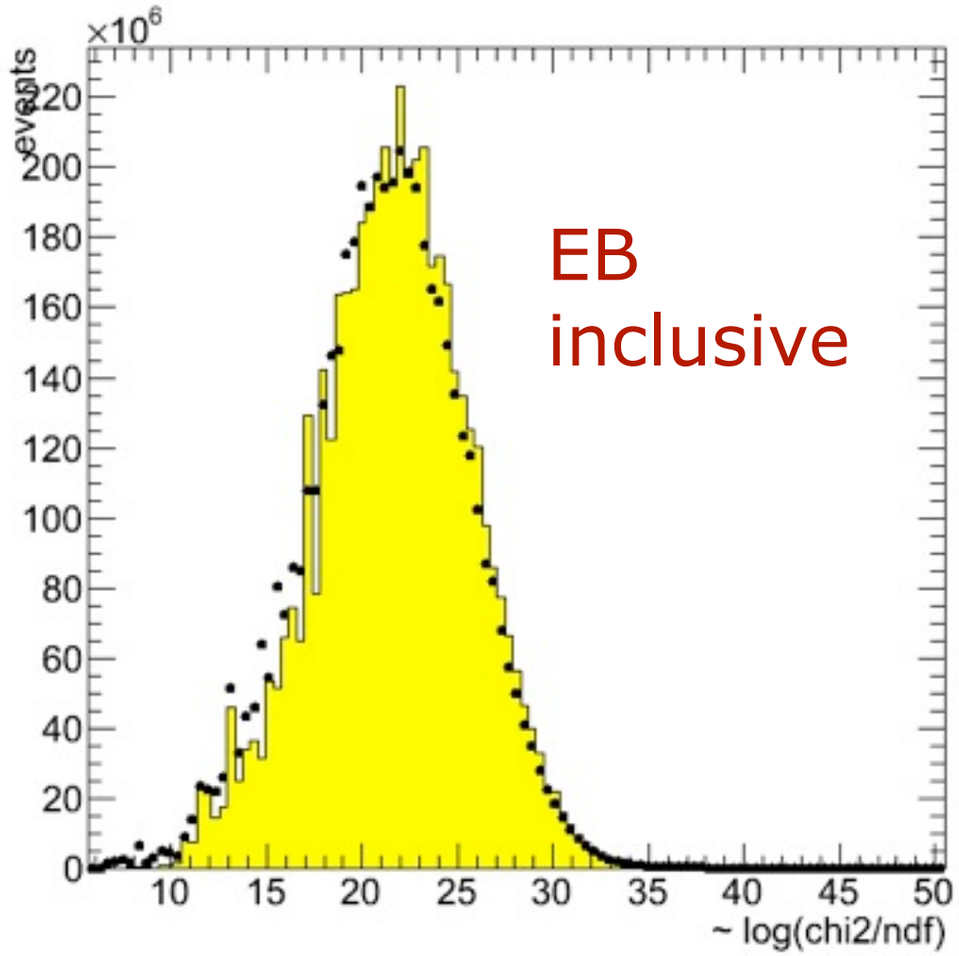


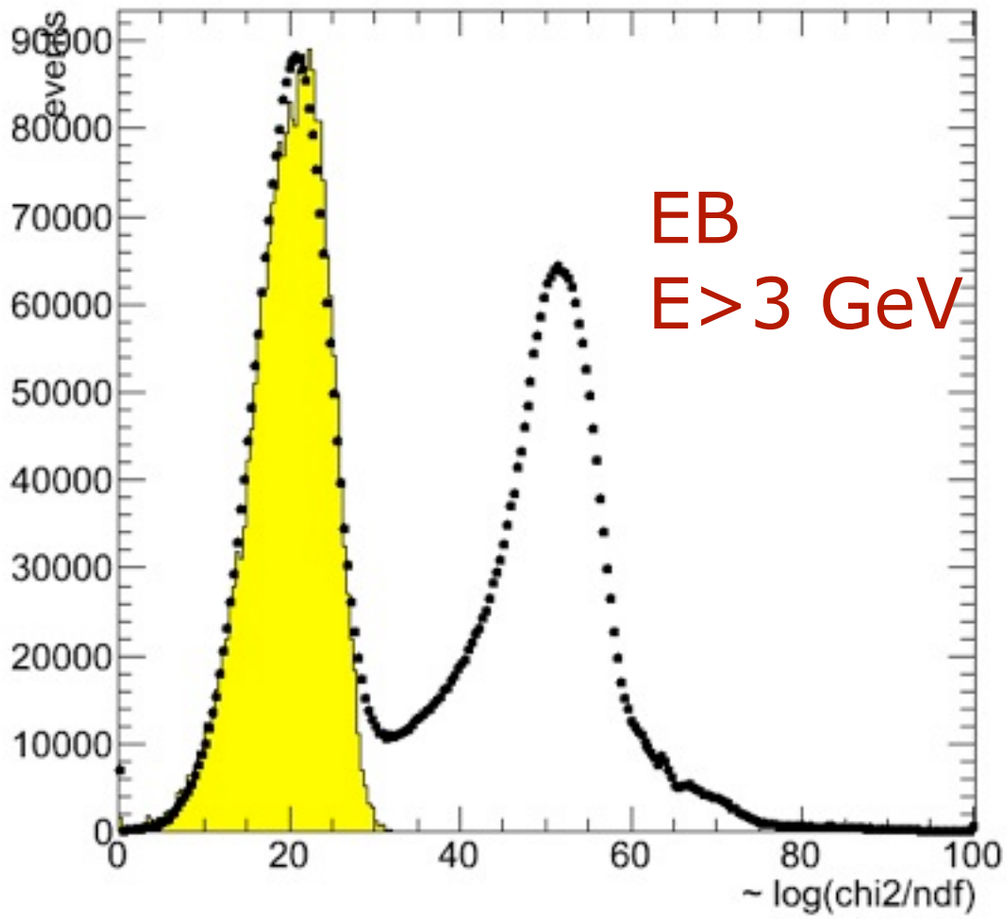


*rare depends on the physics analysis event selection

DATA/MC

- Despite **the rough** shape approximations in EB
  - Some discrimination against spikes is achieved
  - DATA/MC is acceptable* (depending what we aim to)
- However: the size of approximations are not acceptable in EE and produce a huge DATA/MC disagreement (BACKUP)





Limitations (1/2)

- Disagreement between DATA/MC in EE is related to the pulse shape systematics used in the SIM-DIGI step in MC and the real pulse in DATA
  - Either we need just different pulse shape calibration parameters for DATA/MC
    - Or change the reference TB shape in CMSSW (?) *this is really a (backbone)*
- Chi2 CMSSW calibration parameters are hardcoded in a python file which is a show-stopper *if different conditions need to be applied in DATA/MC*

Limitations (2/2)

- The TB pulses which are used as reference pulse shapes are not very precise
- Still good to discriminate spikes in EB, but we can do much better if we want

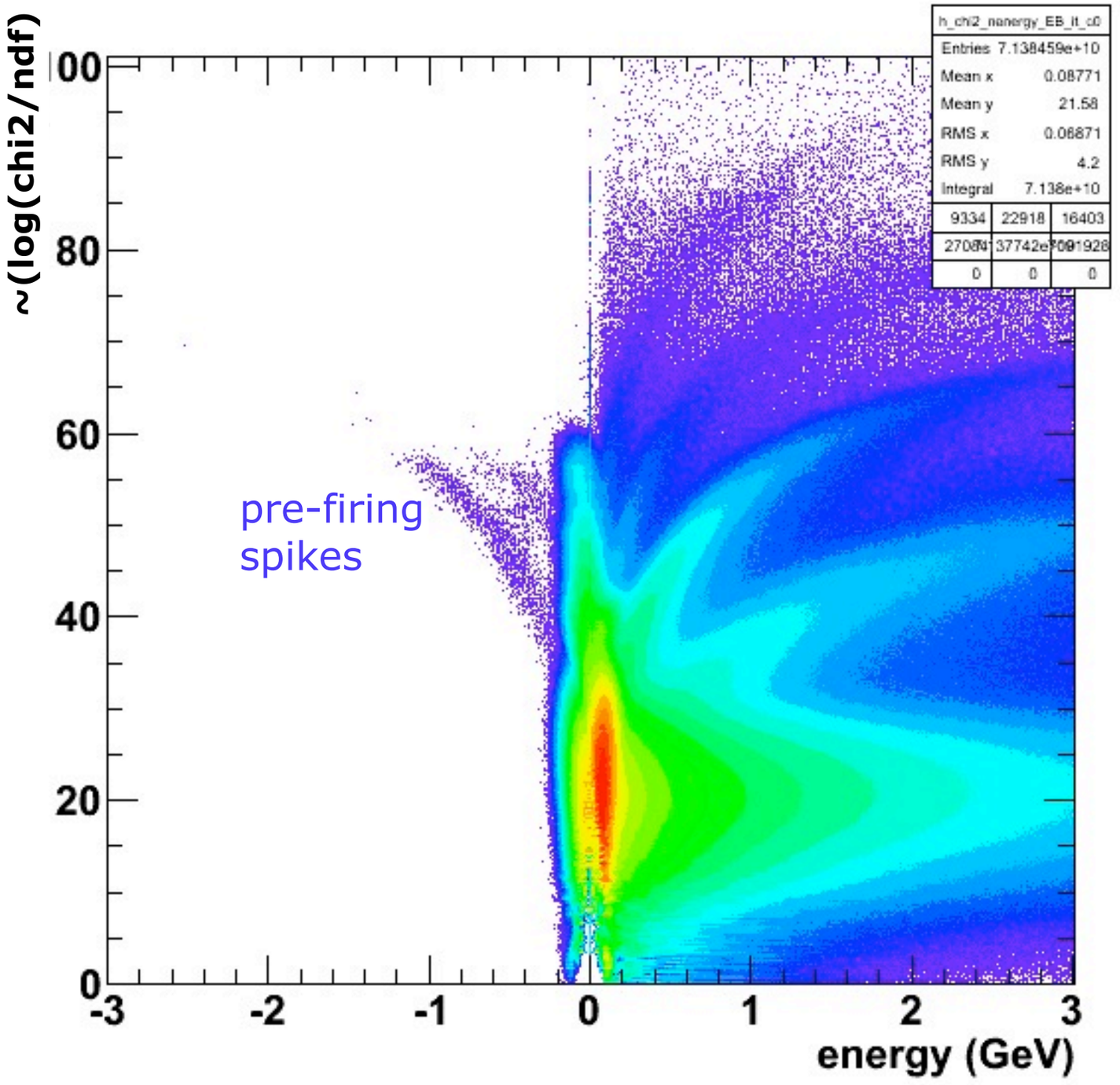
- Don’t forget:
  - Individual channel’s pulse shape may not be constant, current CMSSW implementation cannot handle transient detector conditions
  - Pulse shape has some amplitude dependence

Performance at low p_T

chi2 is stable at low energies no explicit cut off is needed

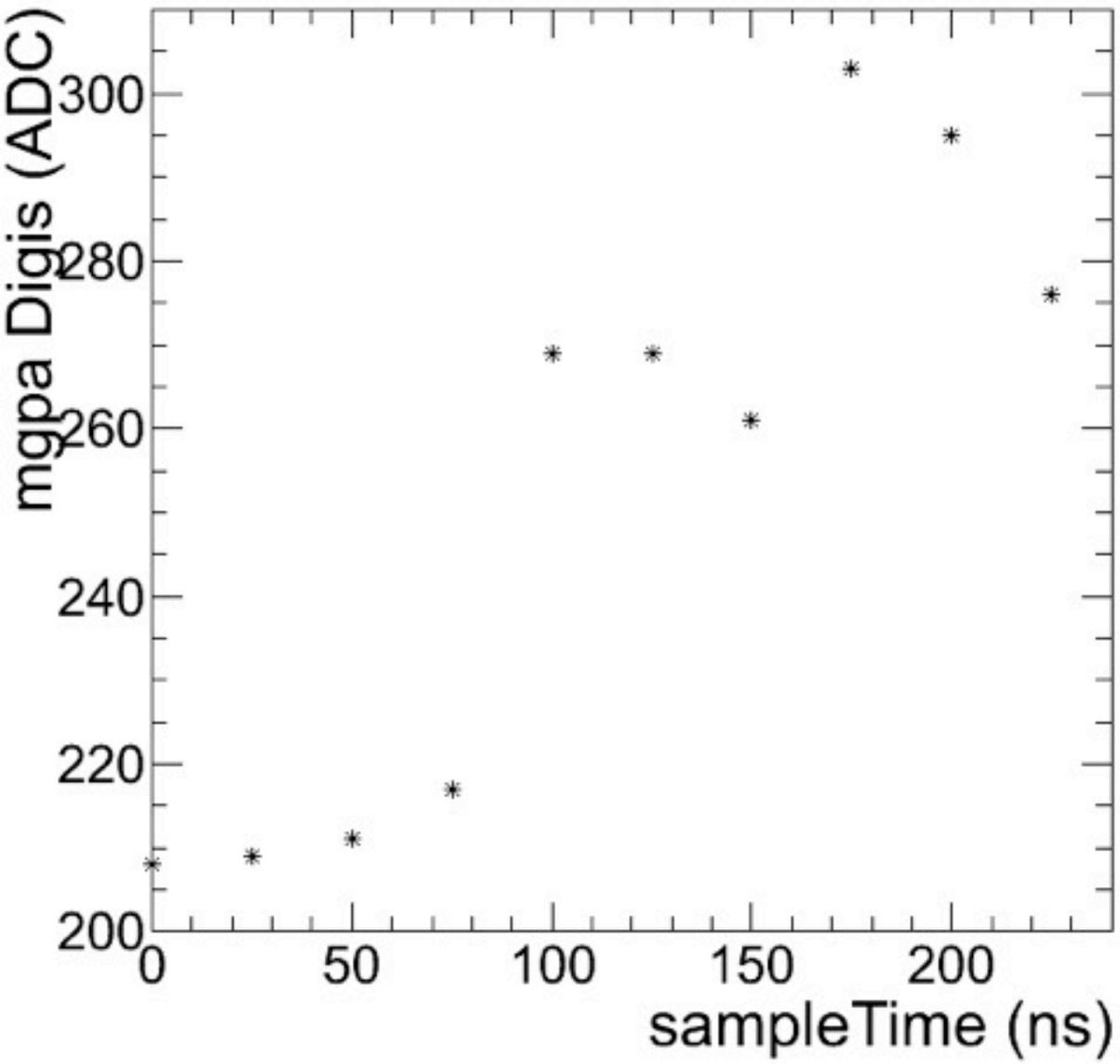
Allowing negative pulses in the chi2 test naturally disentangles the effects:

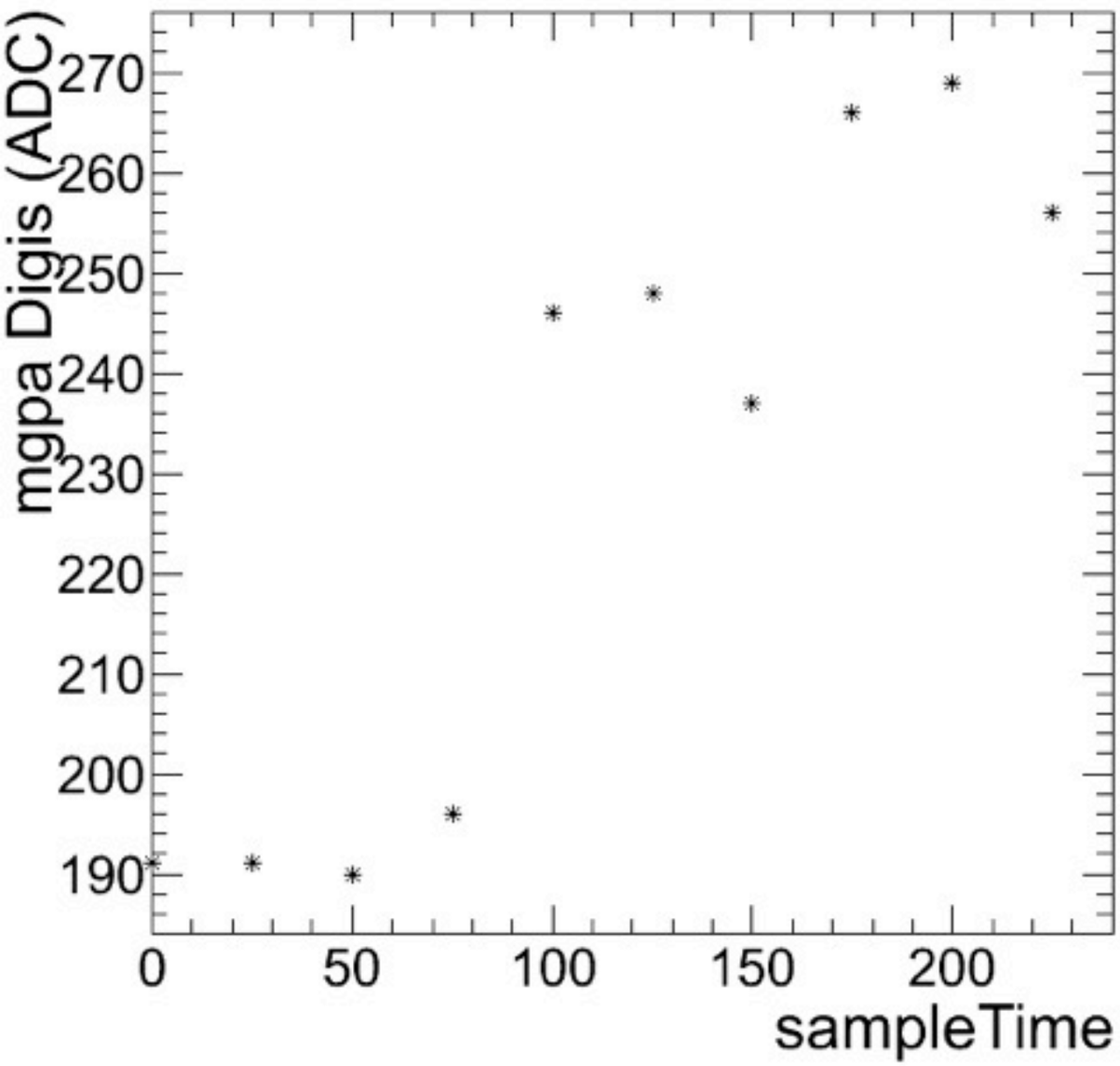
- E<0 due to pedestal breathing
- E<<0 due to problems



BUT not all low p_T anomalous signal are spikes

Some (rare?) low p_T noise





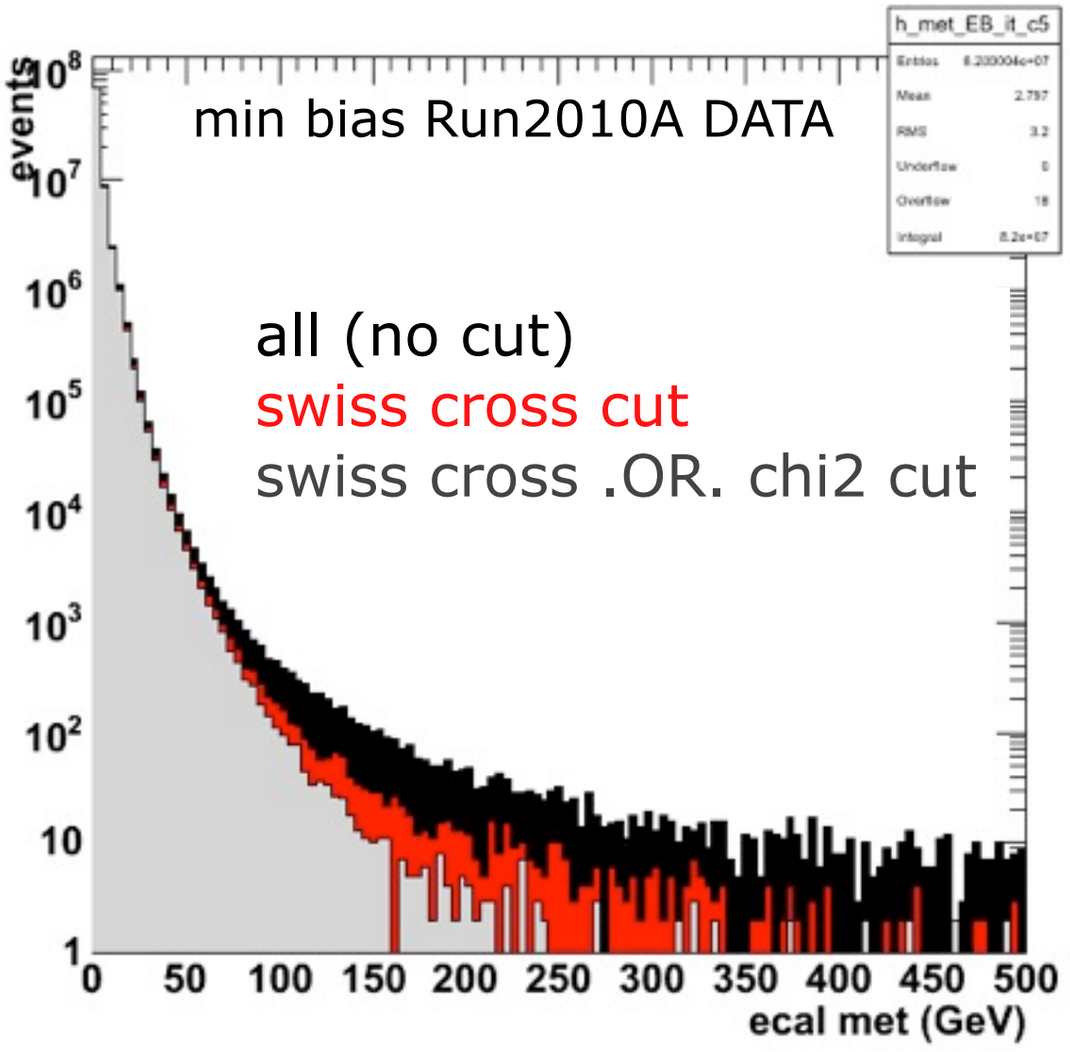
These low E hits degrade ECAL resolution, how much? again this is analysis dependent

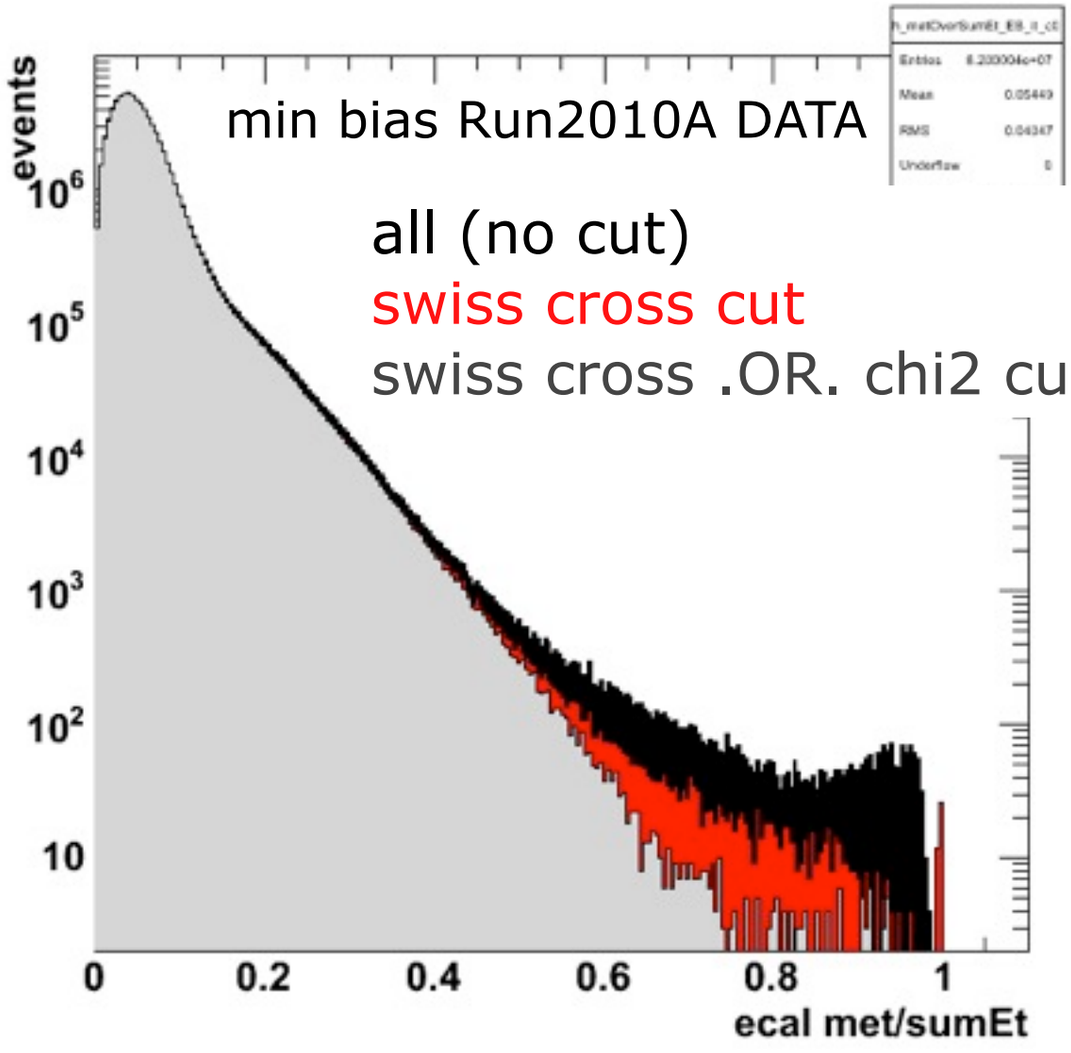
Out-of-time PU

- This is what the chi2 was supposed to control (prior spike epoch)
- Becomes more and more important
- OOT PU gives us low p_T noise that degrades the ECAL resolution
- Possible application: ECAL isolation & clustering for EB can be done with what we have already in CMSSW (since 39X)

Performance at high p_T (1/2)

here cleaning is applied only for G12 digis

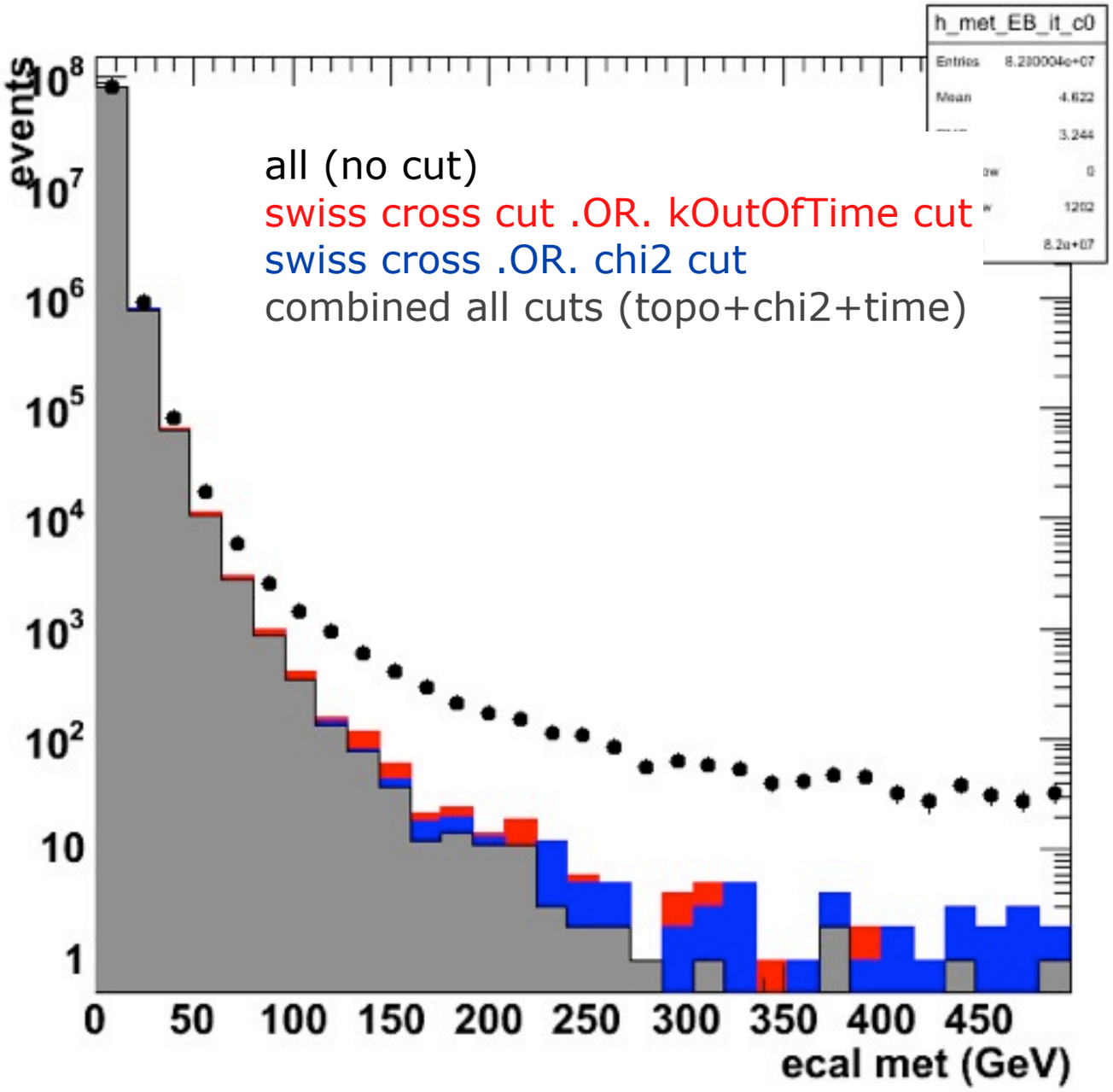




No bad event removal is performed, <u>the number of events is the same in the three histograms</u>

event-by-event improvement of the met & sumEt

Performance at high p_T (2/2)

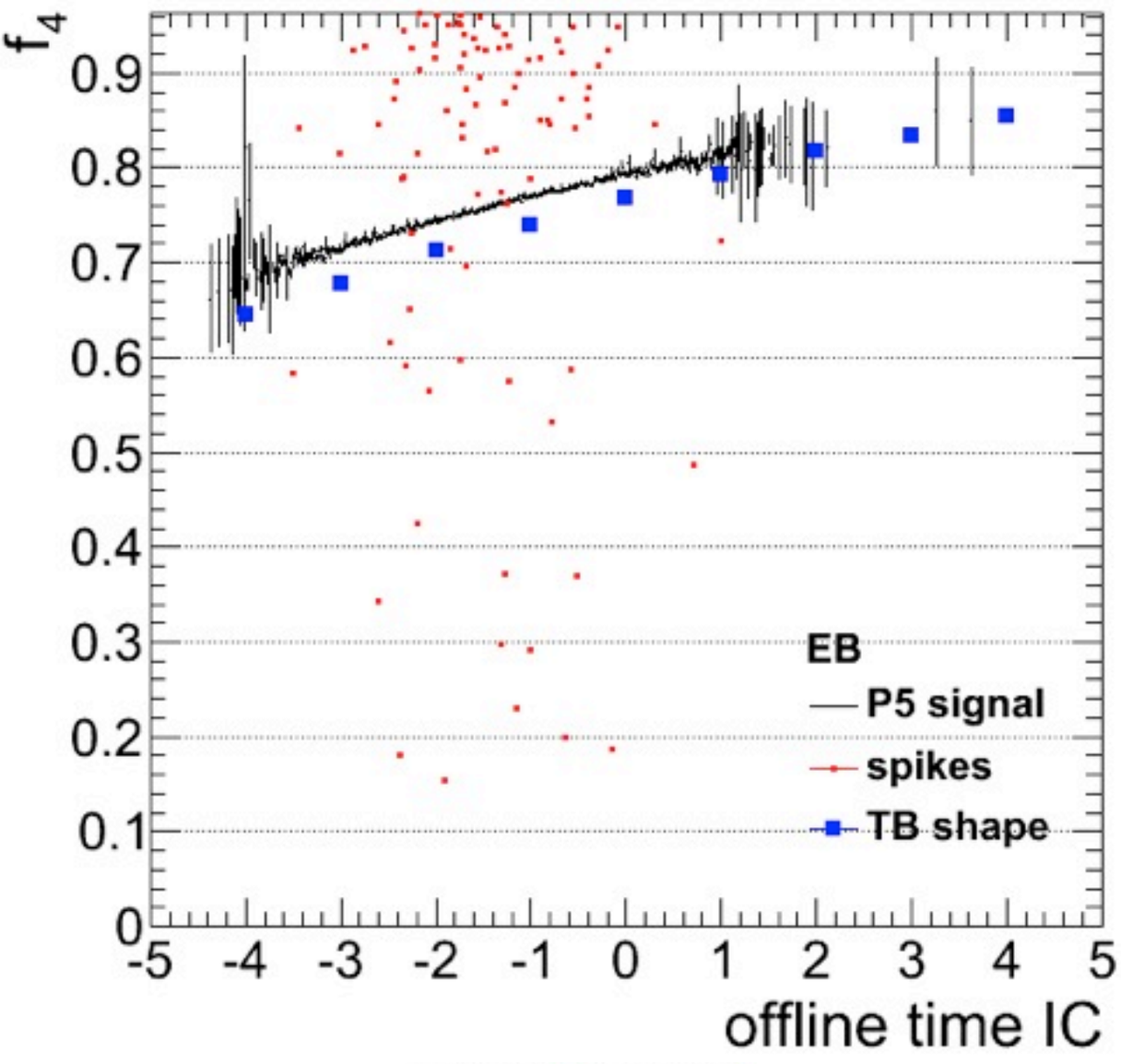


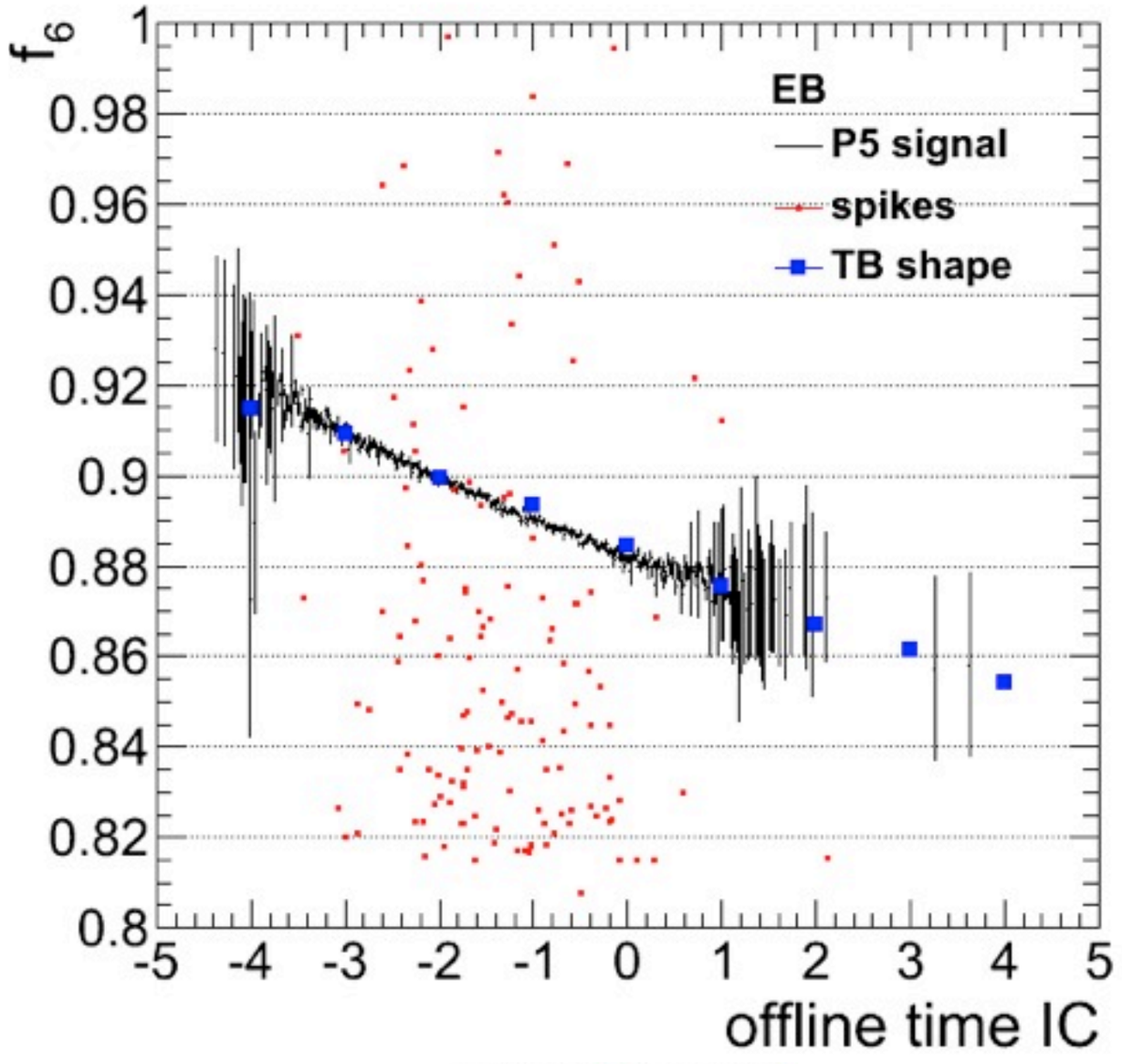
The best cleaning is achieved by combining:
swiss cross + chi2 + time

How can we improve?

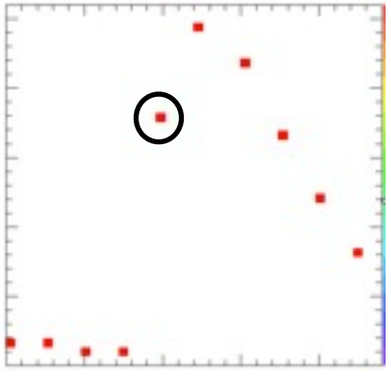
- chi2 (for in-time signals):
  - hardcoded shape calibrations → dynamic in DB
  - average TB shape f[i] → digital from P5 data
  - f[i] for any gain → $f_{G12}[i]$, $f_{G6}[i]$, $f_{G1}[i]$
  - also: binning in energy ?
  - a precise chi2 will need monitoring/calibrations
- chi2 (for OOT signals):
  - study reco-shape systematics for different timing
  - tricky: not many OOT signals in the P5 data
    - ECAL phase scans in 2009
    - TB data re-reco'ed with Ratio method
- both cases are non trivial

EB-P5 data vs theory (TB)

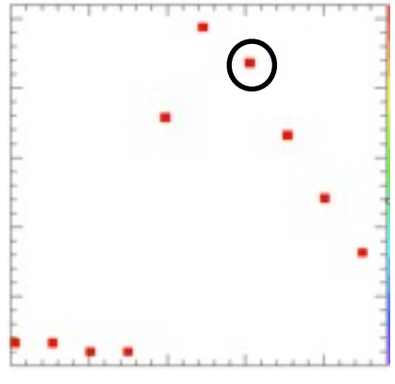




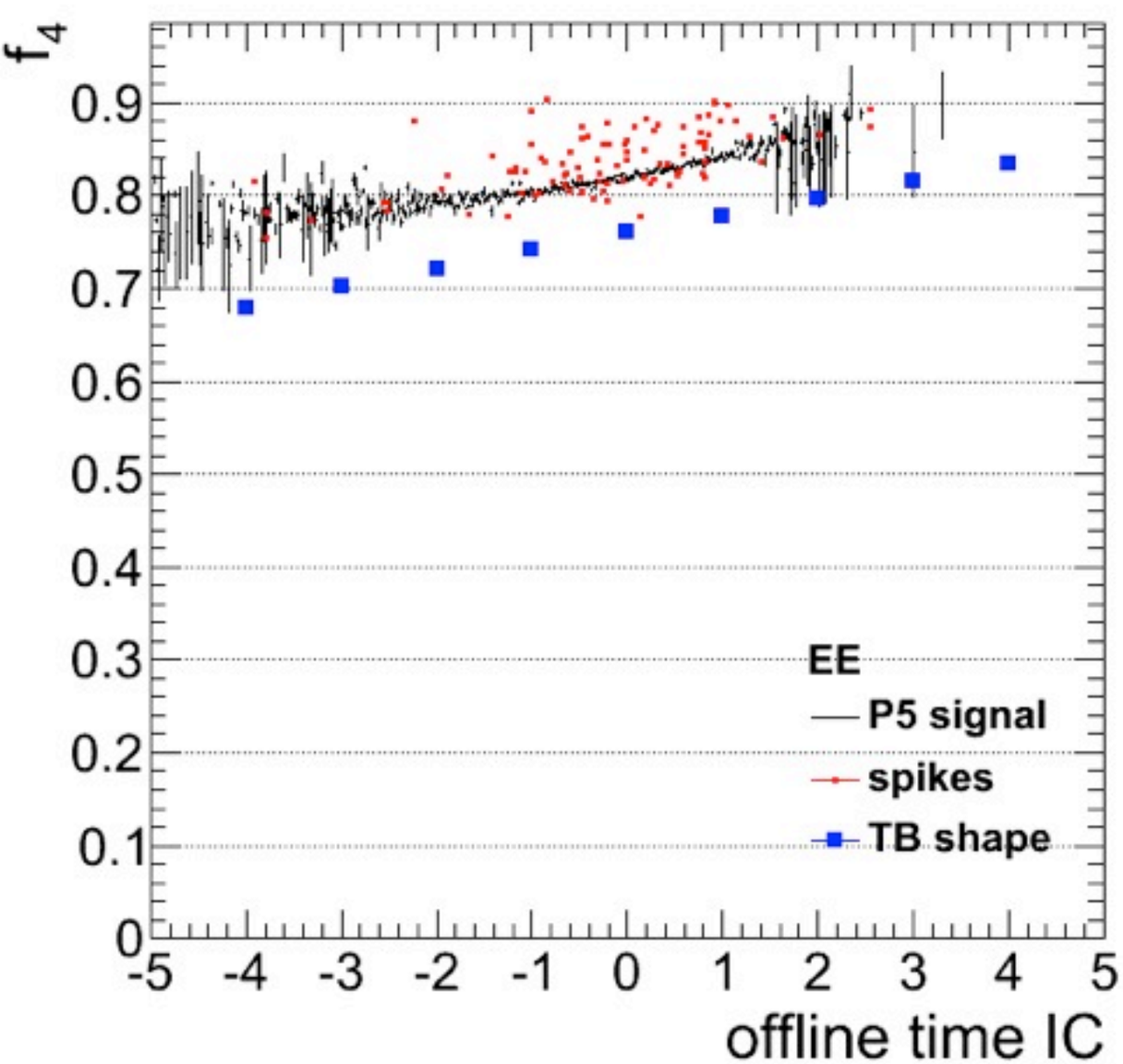
f_4 =

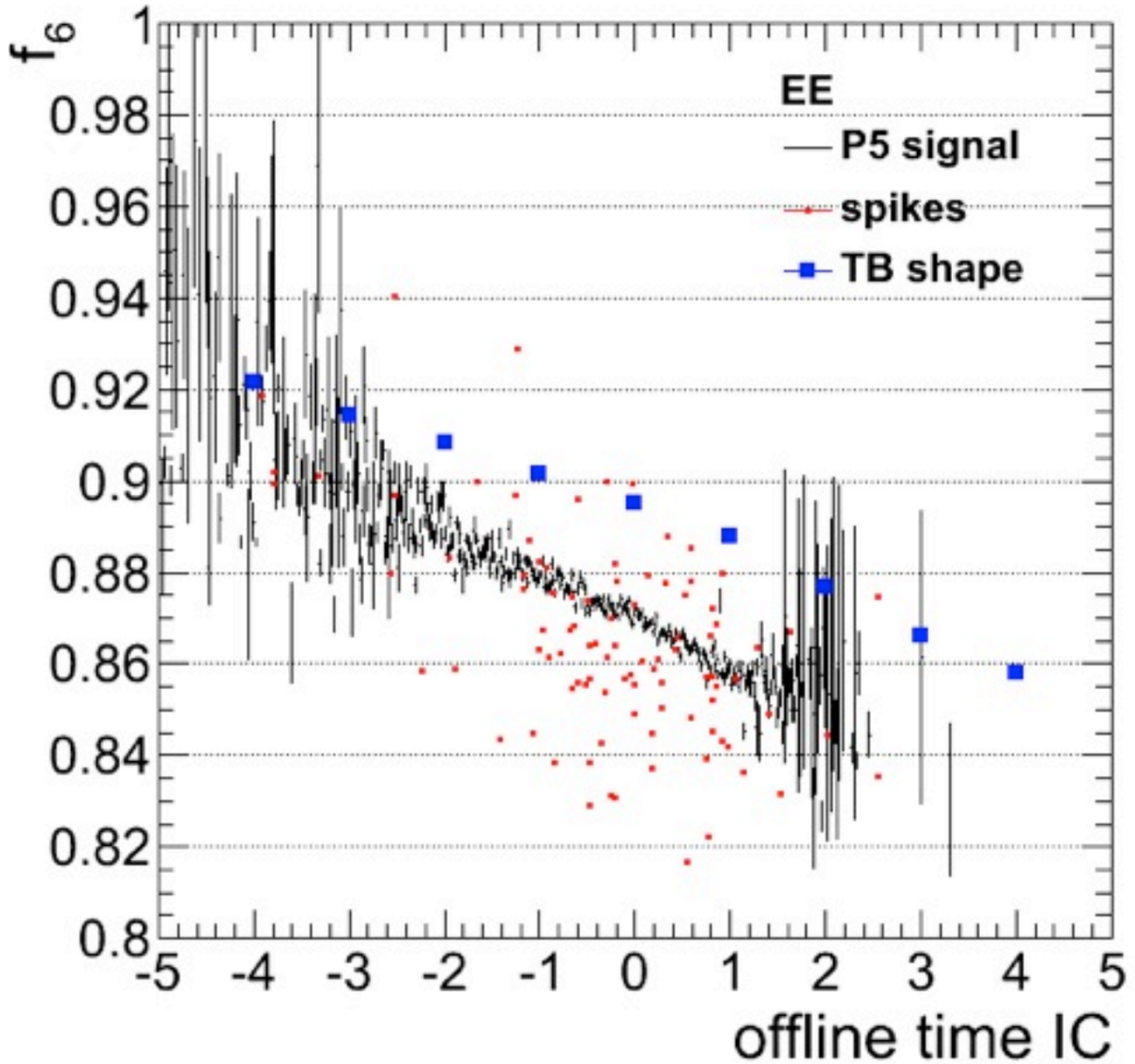


f_6 =

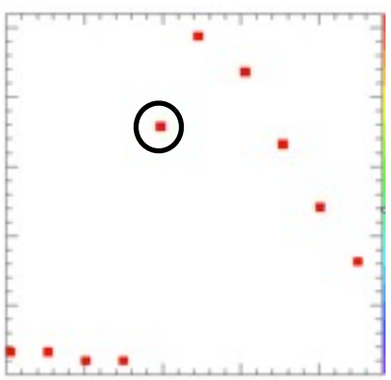


EE-P5 data vs theory (TB)



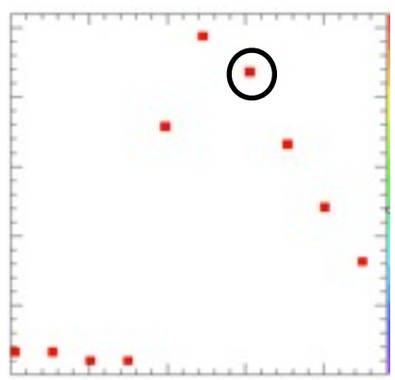


f_4 =



not real spikes in EE, probably signals selected by topological spike definition

f_6 =



Summary

- Signal characterization using ECAL pulse shapes is very primitive
  - Highly anomalous digi patterns not resembling the shape of a scintillating pulse are discriminated BUT we can do much better if we want (?)
- Different prerequisites and amount of effort for
  - low/high P_T
  - DATA/MC
  - G12/G6/G1
  - in-time/OOT chi2
- Different development paths can be followed *depending the precision we aim to and the timescale*