Ricerca e studio di stati esotici *charmonium-like* in CMS con il supporto di GPU.

Attività di ricerca del II anno di Dottorato in Fisica XXXI ciclo

Dottorando

Adriano Di Florio

Tutore

Dott. Alexis Pompili

Outline

- **I – Applications of GPU computing in HEP data analyses**
  - **Global statistical significance estimation by MC toys with *GooFit***
  - **Amplitude analysis fit of $B^0 \to J/\psi K^+\pi^-$ with *GooFit***
- **II – Inclusive search in the Jψ/ϕ spectrum of the prompt and non-prompt production of charmonium-like states in pp collisions**
- **III – Convolutional Neural Networks for Track Seed Filtering at the CMS High-Level Trigger**
- **IV – Schools, courses, conferences, workshops and pubblications**

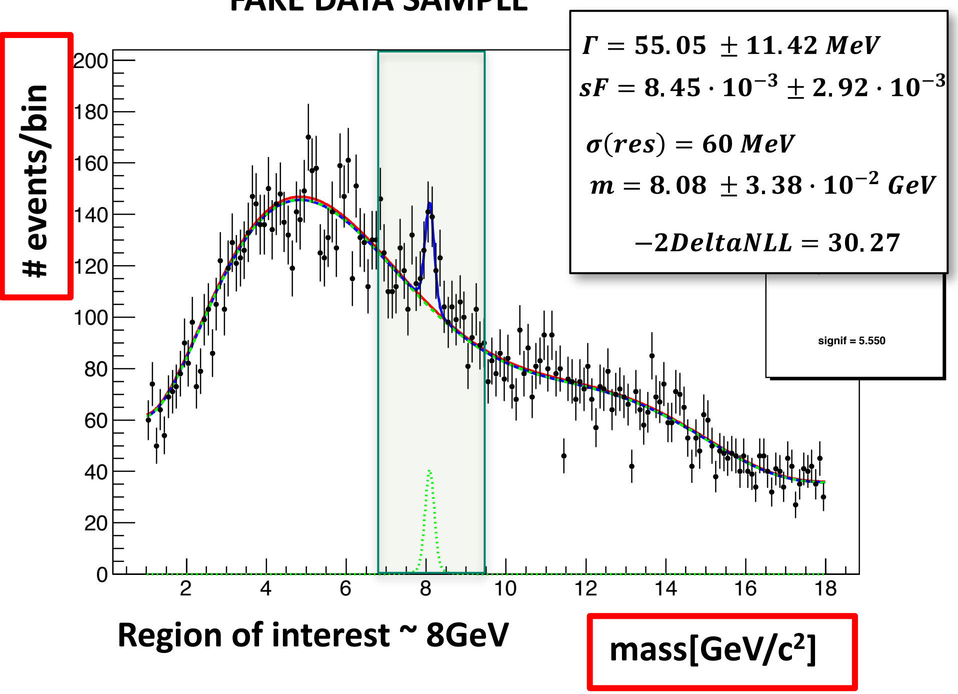
Applications of GPU computing in HEP data analyses

GooFit : LEE & scanning technique

When an unexpected signal is found a ***global significance*** must be estimated. Thus the LEE must be considered and a *scanning technique* must be implemented in order to consider all the relevant peaking fluctuations with respect to the background model everywhere in the overall mass spectrum.

- In order to test the effects of the LEE we generated a *dummy* distribution with a **fake signal** in the 8 MeV region. Mass spectrum goes from 1 GeV to 18 GeV

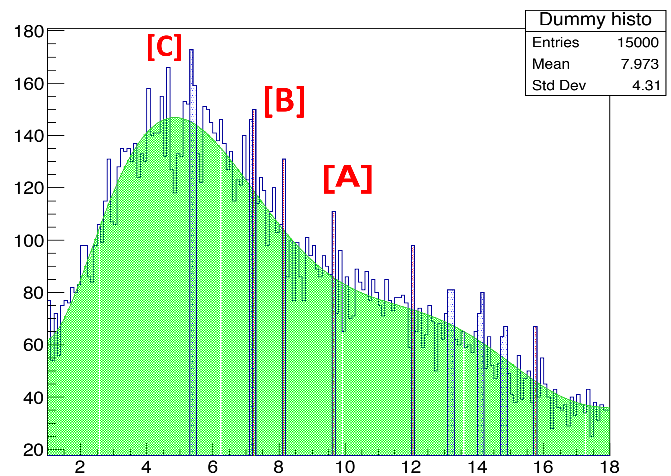
FAKE DATA SAMPLE



- The scanning technique is configured on the basis of a clustering approach with the aim to:

1) Do **not** miss any interesting fluctuation

2) Do **not** select too many marginal fluctuations



Three configuration parameters

- **x (single seed threshold) [A] & [B]**
- **y (side bin threshold) [B] & [C]**
- **z (additional sided seed threshold) [C]**

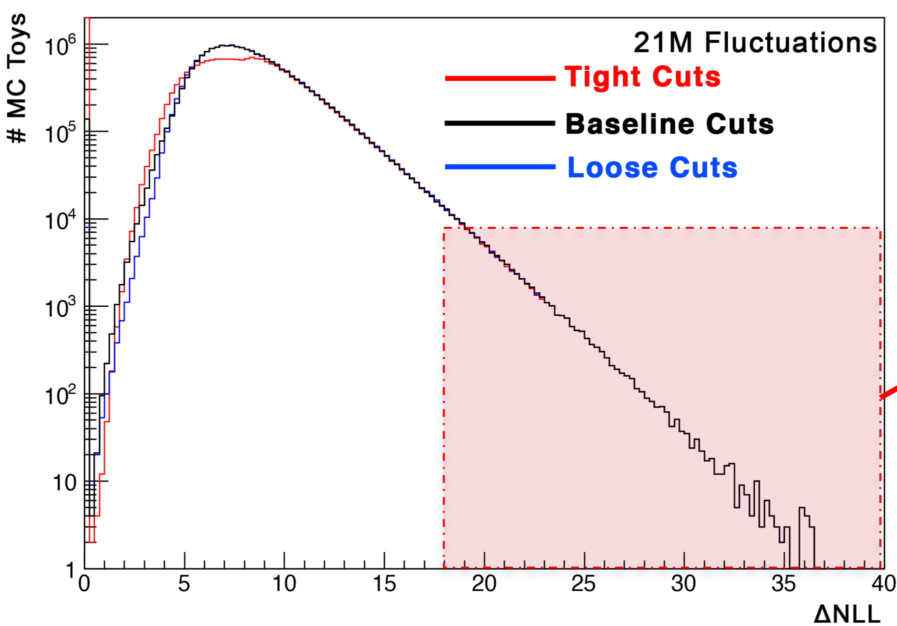
GooFit : LEE & scanning technique

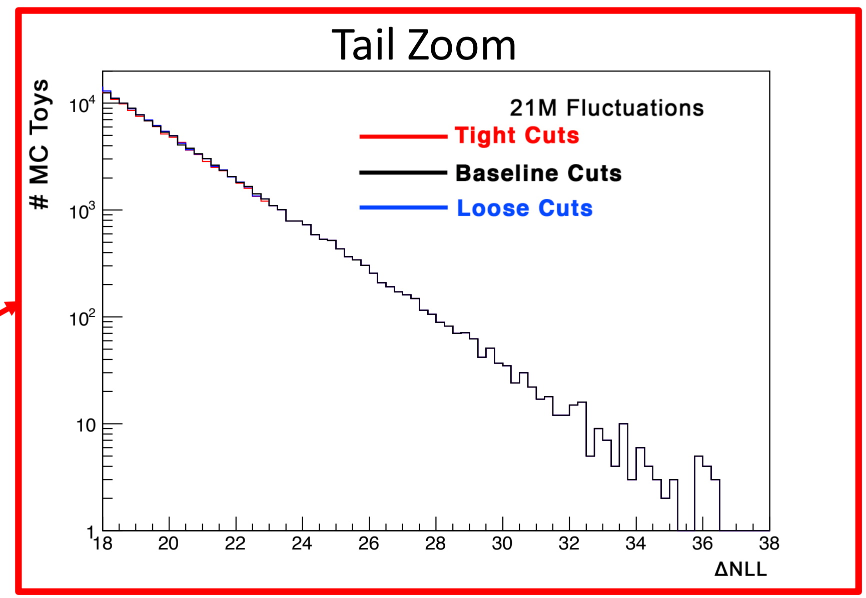
- In order to study the possible bias of this method to the estimation of a global significance we have seleceted three clustering configurations: a ***baseline configuration*** and one ***looser*** and one ***tighter***.

Clustering configs.	$< fit_{H1} >$	f_{nofit}
Tight (3.00, 1.75, 1.00)	2.2	$\sim$10%
Baseline (2.25,1.50, 1.00)	4.5	$\sim$1%
Loose (2.00, 1.25, 1.00)	6.6	0.1%

Local Significance	4.0σ	4.5σ	5.0σ	5.5σ	6.0σ
Tight (3.00, 1.75, 1.00)	2.21	2.91	3.58	4.23	5.19
Baseline (2.25,1.50, 1.00)	2.20	2.91	3.58	4.23	5.19
Loose (2.00, 1.25, 1.00)	2.19	2.92	3.58	4.23	5.19

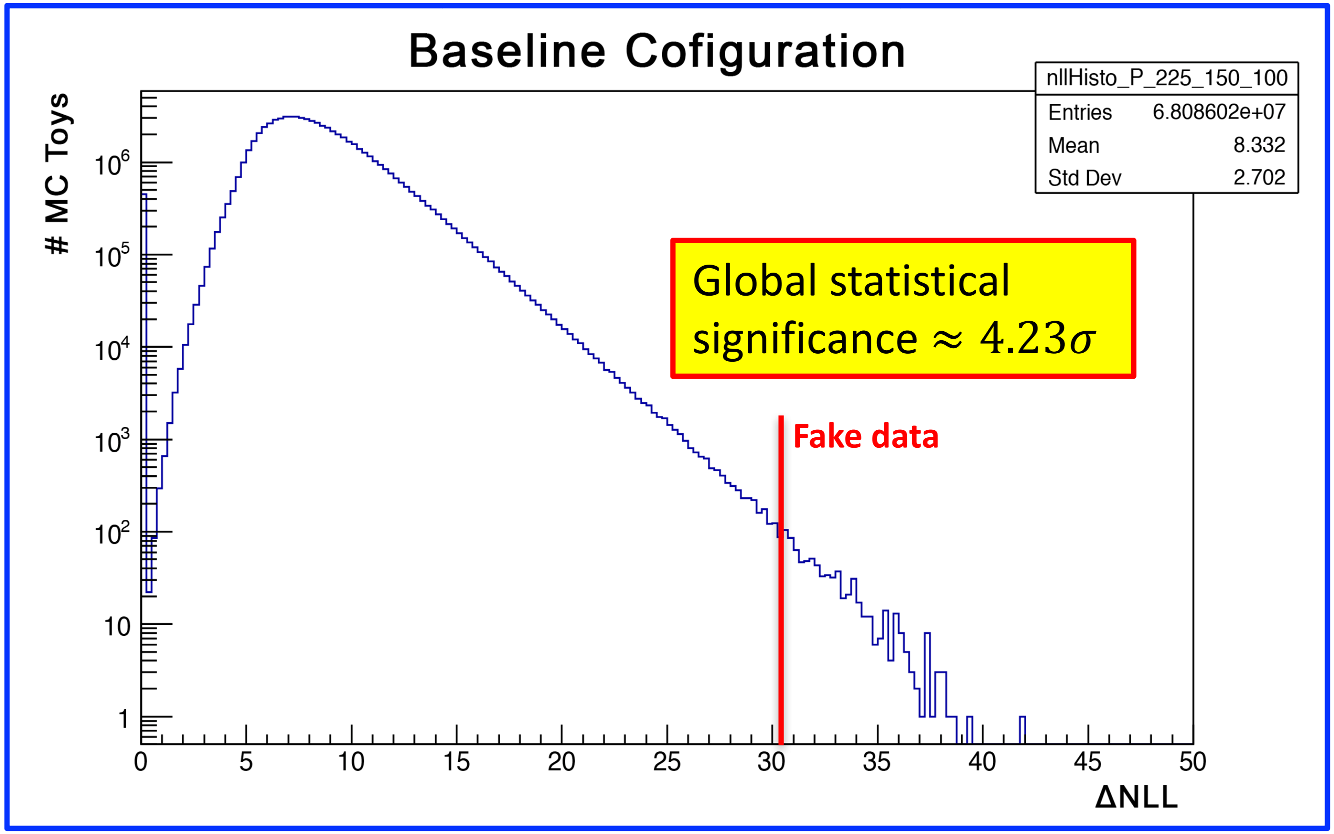
- For any locally significant signal (>4.0) the method **do not introduce** any systematical bias on the global significance.





GooFit : LEE & scanning technique

- Once the method has been validated, the **baseline configuration** has been run on a larger sample of toys (**68M**), obtaining the ΔNLL distribution:



Local significance ≈ $5.5\sigma \longrightarrow$ Global significance ≈ 4.23σ

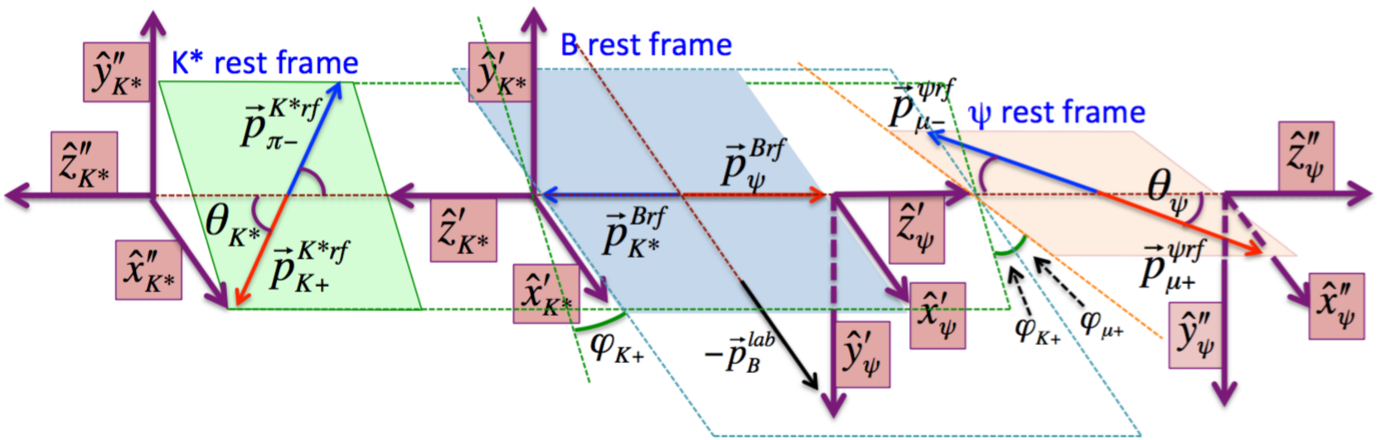
Traditional Dalitz Plot analyses deal with **3-body decays** without any vector state as a daughter. In that case the decay amplitudes are calculated in a *2-D* parameter space, namely the *Dalitz Plot* space itself. In case of with vectors in the final state and with 2-body intermediate states, the decay amplitude is calculated in a 4-D parameter space.

Decay modes needing an amplitude analysis fit being studied in the CMS Bari group

- $B^0 \to J/\psi(nS)K^+\pi^-$ **to search for Z(4430) e Z_c(4240)**

Assuming the only intermediate 2-body states are the K*s : $B_0 \to \psi(\to \mu\mu)K^*(\to K_+\pi_-)$

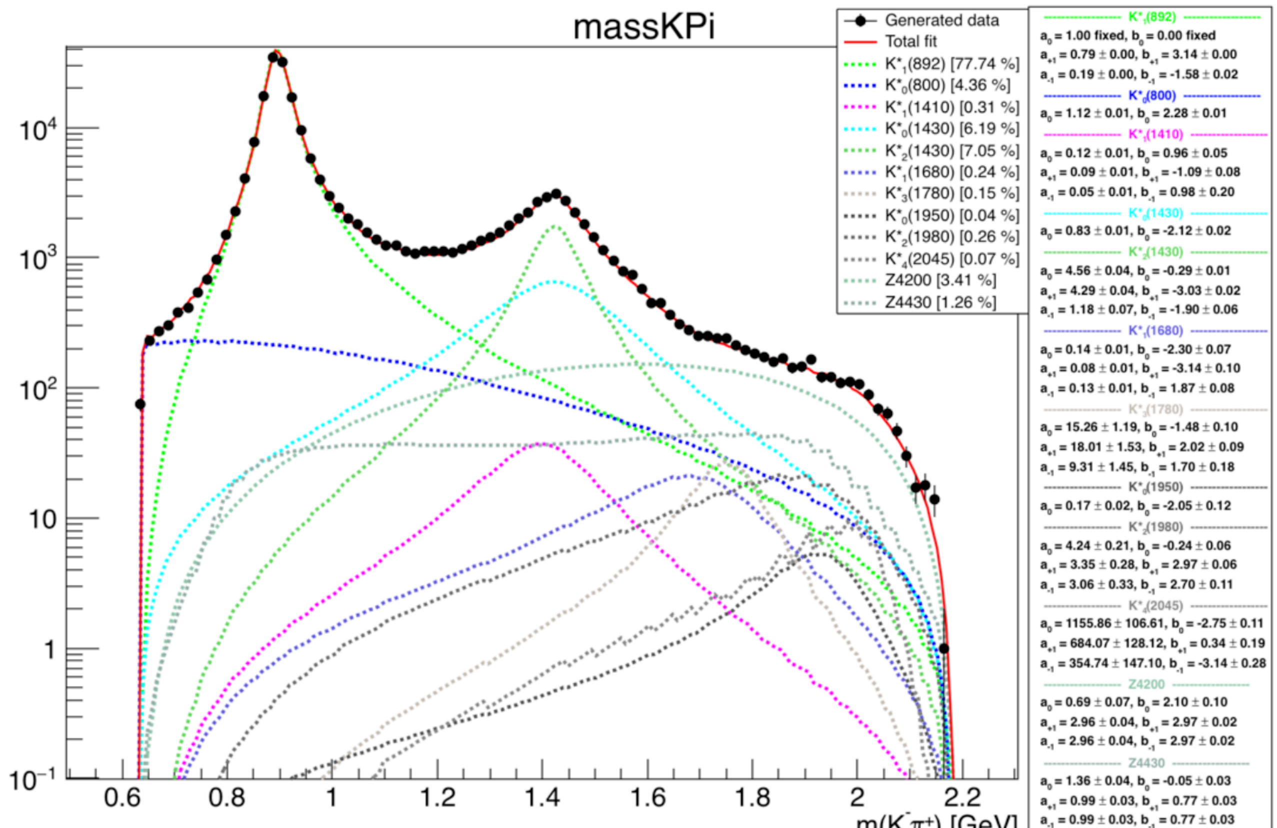
The 4-D parameter space is: $\Phi = (m^2_{K\pi}, m^2_{\psi\pi}, \vartheta_\psi, \varphi_{\psi K^*})$



Taking into account the most predominant K* contribution **28 fit parameters** in a **4 dimensional space** (two factors ($a \times \mathbf{e}^{ib}$) for each helicity amplitude).

GooFit : amplitude analysis

Porting of $B^0 \to J/\psi K^+\pi^-$ A.A. **including:**



- the **efficiency correction** and the combinatorial **background interpolation**
- a model that distinguish B0 and B0bar events.
- an additional model representing the Z(4430) and the Z(4200)

GooFit **fit** takes **25'** performing over 1k MIGRAD calls for **28 fit parameters** in a **4 dimensional space efficiency correction** and **background** included for 100K candidates.

Ongoing efforts aim to develop a ***Goodness of fit*** estimation procedure in order to fullfill the fit procedure validation before starting to **fit the real data.**

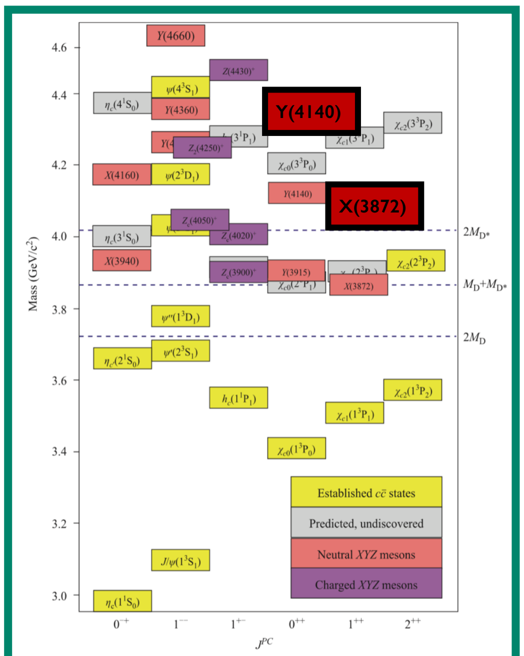
Inclusive search in the $J/\psi/\phi$ spectrum of the prompt and non- prompt production of charmonium-like states in pp collisions

Inclusive search in the Jψ/ϕ spectrum

In the last 13 years about 30 states (**known as *X, Y, Z* states**) observed while decaying to charmonium inspite of being **above the open-charm thresholds**. They show "exotic" characteristics.

Two main hypotheses:

- hadronic molecule
- tetraquark



CMS: Confirmed (2014) two structures seen by CDF (2011) in the $J/\psi\phi$ mass system, with a 1D analysis of $B^+ \to J/\psi\phi K^+$.

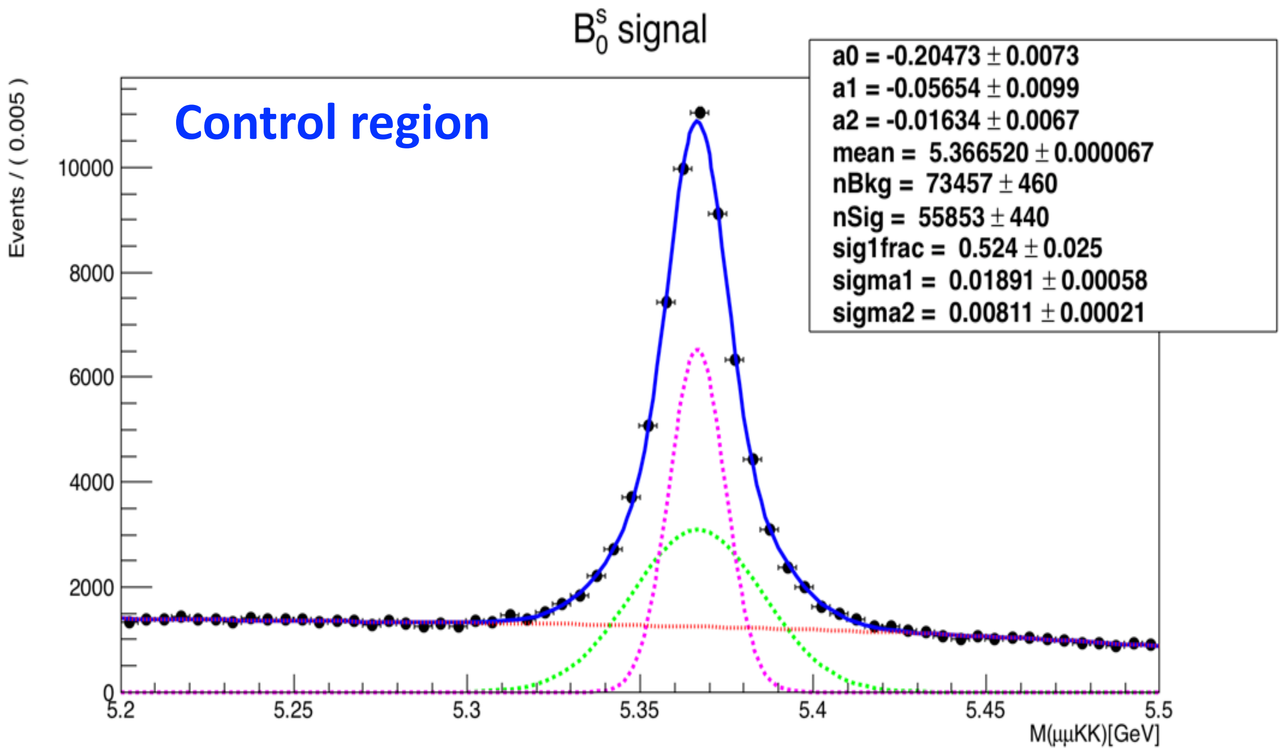
D0: Prompt and non prompt production of Y(4140) state in $\bar{p}p$ collisions [$\bar{p}p \to Y(4140) + X$ with $Y \longrightarrow J/\psi\phi$] by studying inclusively the $J/\psi\phi$ mass spectrum (2015).

LHCb: First amplitude analysis (2016) of the $B^+ \to J/\psi\phi K^+$ observing 4 structures in the $J/\psi\phi$ mass system; among them the Y(4140) structure, the closest to the kinematical threshold, which is slightly *better described as a $D_sD_s^*$ cusp* (resonant interpretation also possible)

It becomes crucial to confirm the D0 result for the inclusive search. In case of a positive confirmation **this would rule out a cusp interpretation of the Y(4140).**

By exploiting Run-I and Run-II data, **CMS** has the capability to look for the presence of the Y(4140) and its partners inclusively in the **J/ψϕ mass spectrum** either **promptly** produced or coming by **decays of beauty mesons**.

- For validation purposes, **two regions** of J/ψϕ invariant mass spectrum need to be explored; a **control region** near the B_s^0 meson mass and a **search region** **[4.1-4.8]** in which these states may appear. In both regions one looks for states decaying to J/ψϕ (J/ψ → μμ and ϕ → KK)
  - By now the whole 8 TeV Run I data have been explored and two results for the control region:

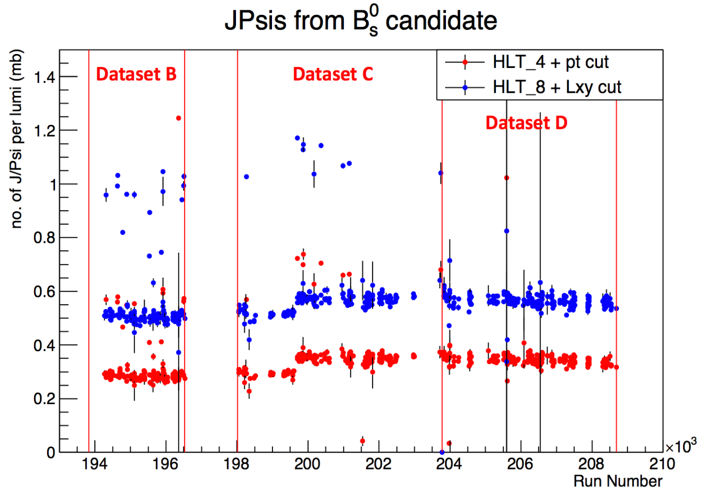


Reconstruction of the B_s^0 **→ J/ψϕ** signal as a control channel. The extracted control signal shows *good agreement* with already published B_s^0 signals for Run I data.

Inclusive search in the Jψ/ϕ spectrum - Outlooks

Study of the **stability of the yield of J/Ψ,** produced from a **B_s^0 → J/ψϕ** decay, along all the Run-I data taking with the aim to

- to check eventual biases in the evts selection
- test the effect of two different possible HLT triggers



The search strategy of a Y resonance in the *μμKK* final state is suitable for both Run-I pp collisions data sample and Run-II data being collected in 2017-2018

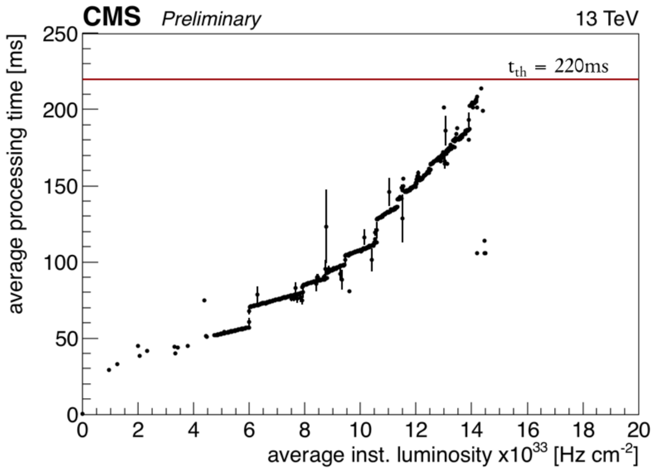
- For Run II data taken in 2017/2018: two new specific high level trigger paths have been implemented, one based on **2muons + 2tracks** and the other on **4muons:**

J/ψϕ background is expected to arise mainly from combinatorial background of **K tracks**, then **an additional strategy can be explored:**

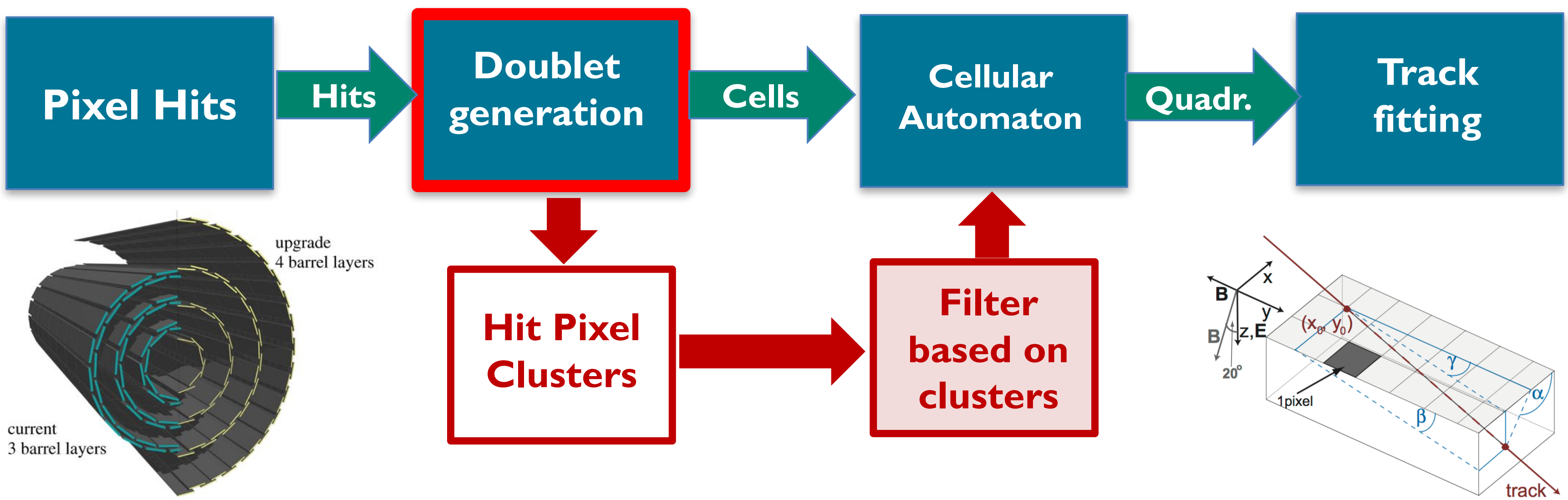
- investigate the J/ψϕ invariant mass spectrum also through the rare but cleaner **ϕ → μμ decay** channel
- J/ψϕ system explored relying **only** on μ reconstruction, bypassing the lack of particle identification for hadron

Convolutional Neural Networks for Track Seed Filtering at the CMS High-Level Trigger

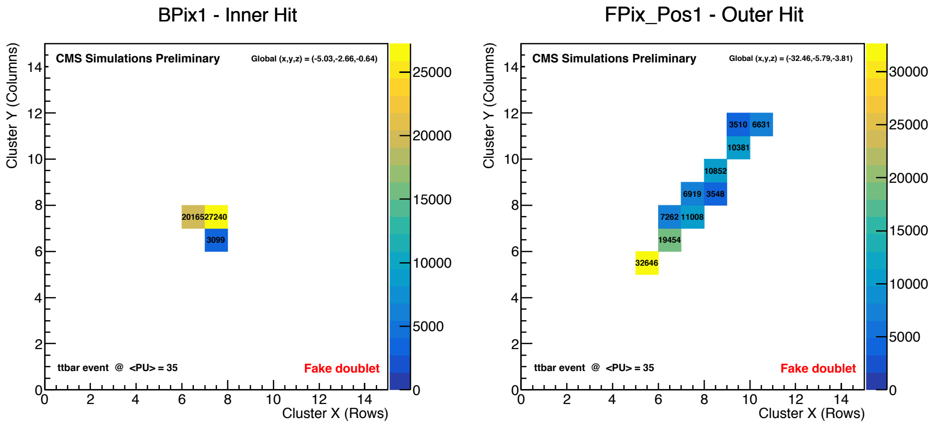
- *Tracking at HLT :*
  - The online farm consists of ~26k CPU Xeon cores
  - A single event per logical core
  - Now tracks are not reconstructed for all the events @ HLT
  - This will be even more difficult at higher pile-up (200 @HL)



- *The **tracking workflow** @ Pixel Tracker in a nutshell*
  - ***Doublets generation:*** *bottleneck due to huge combinatorial background.*
    E.g. ~10^5 - 10^6 doublets produced @ **PU35** *for* ***TTBar13TeV*** *event with* ***a fake ratio of ~300/400***

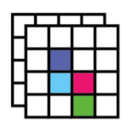


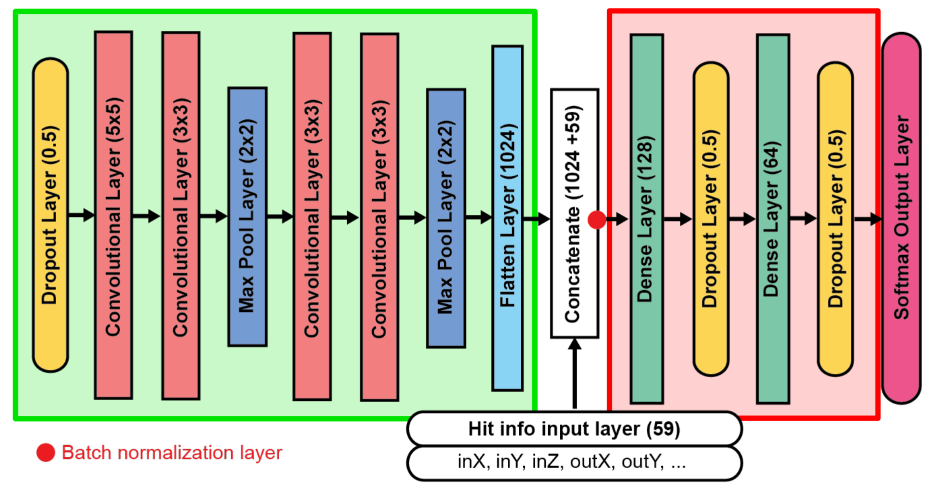
- *Typical binary **classification** problem : **keep true doublets** & **reject fake doublets***



- We considered each hit as a **15x15** pixel **pad/image** centred in the cluster center of charge
- *A single doublet is considered as a couple of 15x15 matrices*
- Pattern recognition problem : suitable for a Convolutional Neural Network approach

- ***"Feature layer map"* model:**
  - *CNN architecture*: stack of convolutional layers (4) and max pooling (2)
  - *"DENSE" architecture*: dense layers (2) fed with the 1-D reduced images + doublets infos

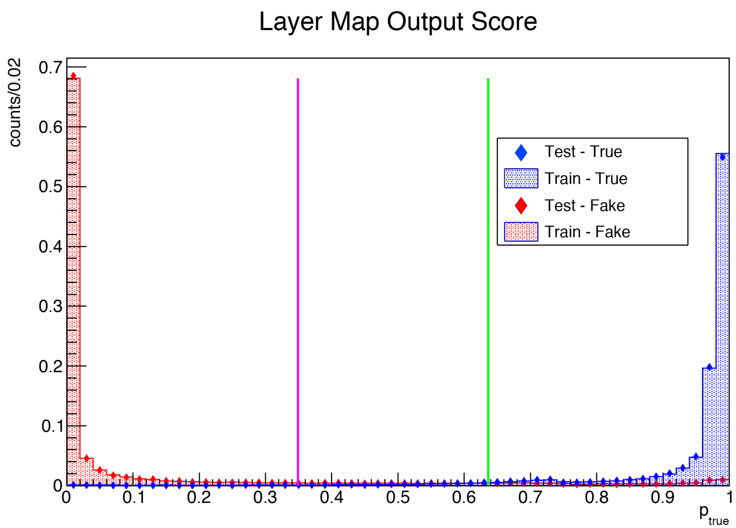




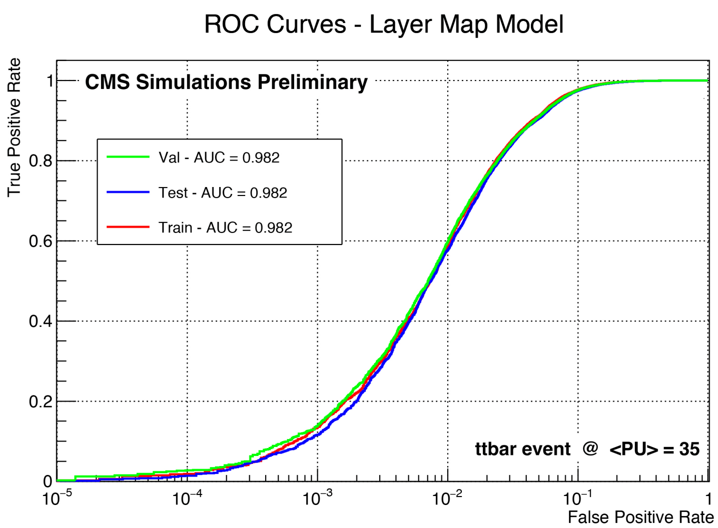
CNN network trained on 2.5M doublets results

			@ Max Accuracy		rej @ eff		eff@rej	
	AUC	*Max accuracy*	*Efficiency*	*Rejection*	*0.99*	*0.999*	*0.99*	*0.5*
Train	0.982	0.940	0.9621	0.9371	0.85426	0.6706	0.5896	0.9997
Test	0.982	0.941	0.9597	0.9296	0.8542	0.6709	0.5899	0.9996
Val	0.982	0.939	0.9611	0.9312	0.8525	0.6707	0.5948	0.9996

Classifier score for signal and background



ROC curve **(efficiency vs rejection)**



CNN techniques for mitigating combinatorial divergence look very promising.
Ongoing work: exploring the integration in the CMS reconstruction Framework

Schools, conferences, workshops and publications

Schools, conferences and workshops

Workshops & Conferences

- 103° Congresso Nazionale della Società Italiana Di Fisica, Trento(Italy), 11-15 Sep 2017 Oral presentation at SIF parallel session with title: “Recent CMS results in the search of new signals in quarkonium physics”

Schools

- Third Machine Learning in High Energy Physics Summer School, Reading (UK), 16-23 Jul 2017
- CMS Physics Object School (CMSPOS), Bari (Italy), 4-8 Sep 2017; Participation to Workshops and Conferences

Publications

- *A.Pompili and A. Di Florio* (on behalf of CMS collaboration), **"GPUs for statistical data analysis in HEP: a performance study of GooFit on GPUs vs. RooFit on CPUs"**, J.Phys.Conf.Ser. 762 (2016), 012044, Proceedings of "17th International workshop on Advanced Computing and Analysis Techniques (ACAT-2016)".

- *A.Di Florio* et al. **"Statistical significance estimation of a signal within the GooFit frame- work on GPUs"**, J.Phys.Conf.Ser. 762 (2016), 012044 Proceedings of "12th Conference on Quark Confinement and the Hadron Spectrum (Confinement XII)".

- *A. Di Florio*, **"Performance studies of GooFit on GPUs versus RooFit on CPUs while estimating the statistical significance of a new physical signal",** Proceedings of "22nd International Conference on Computing in High Energy and Nuclear Physics (CHEP- 2016)". Already reviewed, going to print.

- *A. Di Florio et al., "***Convolutional Neural Network for Track Seed Filtering at the CMS High-Level Trigger"**, Proceedings of "18th International workshop on Advanced Com- puting and Analysis Techniques in physics research (ACAT-2017)", in preparation.

- *A. Di Florio et al.*, **"Performance studies of GooFit on GPUs versus RooFit on CPUs while estimating the global statistical significance of a new physical signal"**, Proceedings of "18th International workshop on Advanced Computing and Analysis Techniques in physics research (ACAT-2017)",in preparation.

- CMS Collaboration Author since 5 October 2017.

Thank you

"I am putting myself to the fullest possible use, which is all I think that any conscious entity can ever hope to do"

HAL9000