Year-end Workshop: PH Circle finding

Dorian

Osaka university

December 23, 2019

Table of Content

Table of Contents

Motivation

In a magnetic field, charged particles describe a circular trajectory (2D projection + small momentum). Also, a lot of todays experiments needs high luminosity, depending on the signal you are searching for, you may need to produce a specified particle. Creating this particle can produce a lot of background, but you still need to be able to distinguish signal from backgroud if you want to improve your results sensitivity.

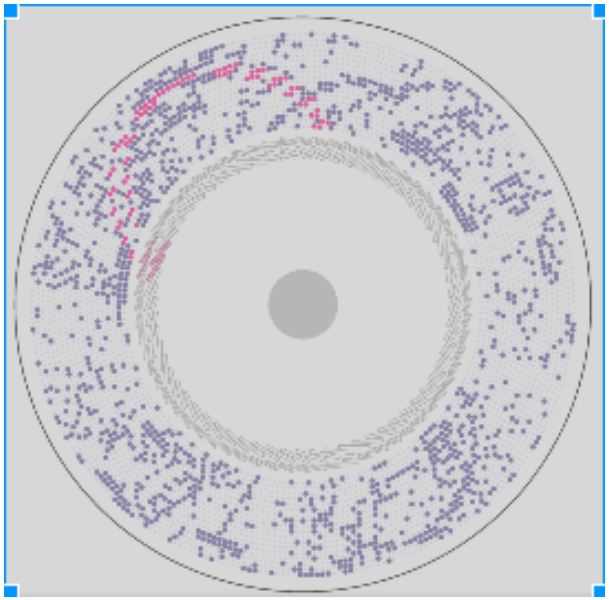


Table of Contents

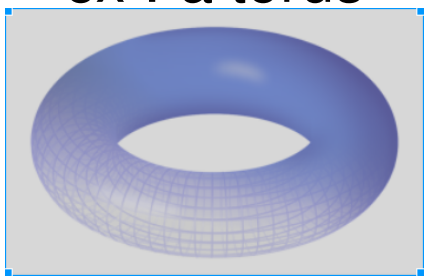
Table of Contents

Persitent Homology

Persistent Homology (PH), is a mathematical tool to express dataset into topological features. Topological features are described by their betti numbers :

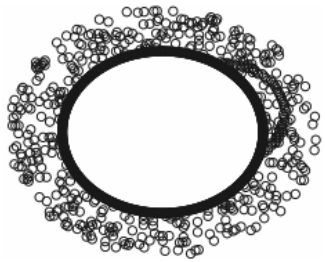
- Number of connected component : b_0
- Number of holes : b_1
- Number of voids : b_2
- etc ...

ex : a torus



$b_0 = 1, b_1 = 2$ *and* $b_2 = 3$

ex : COMET signal (toy model)

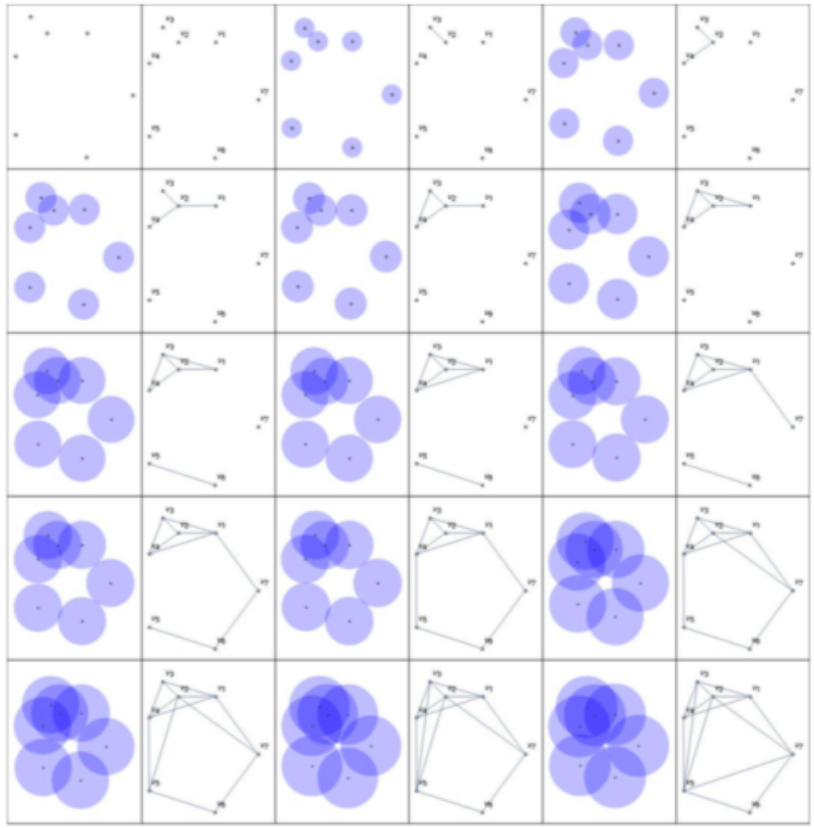


$b_0 = 1, b_1 = 1$

Persistent Homology: a simple example

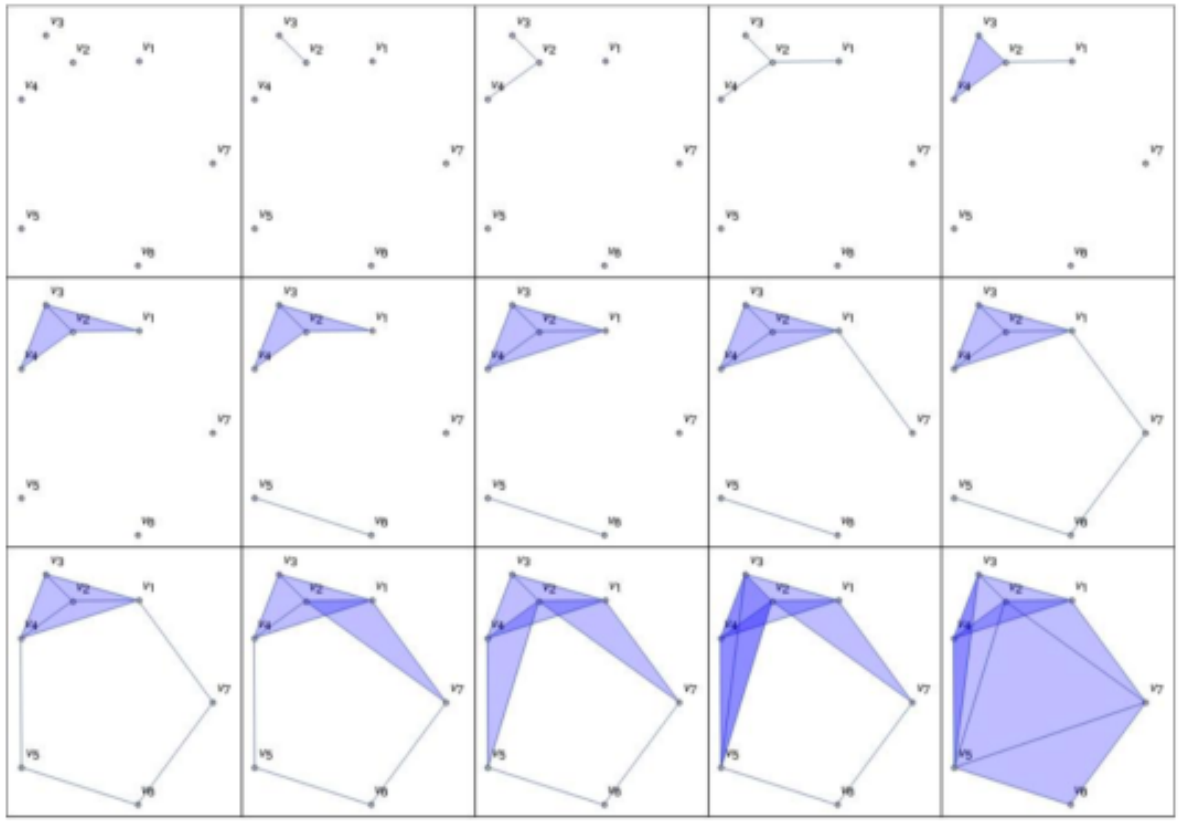
Small exmaple where we would like to observe a loop/hole.
- Replace all data points by vertices
- Draw ball around those vertices with a radius of ϵ
- Link two vertices if their sphere connect
- Increase the radius of the sphere
- Repeat with increasing ϵ

We created a lot of loops/holes at the end of the process, but cannot distinguish the main hole from the others. We only have the created the birth radius of each loop.



Persistent Homology: a simple example

To have more information on the loop created, we need to fill them and thus induce a death radius/time for each loop. We can for example use Vietoris-Rips rules, where we only fill the triangle.



$\rightarrow$ We can see that main hole persist longer than the other one, and pick it up.

Example for COMET (toy model)

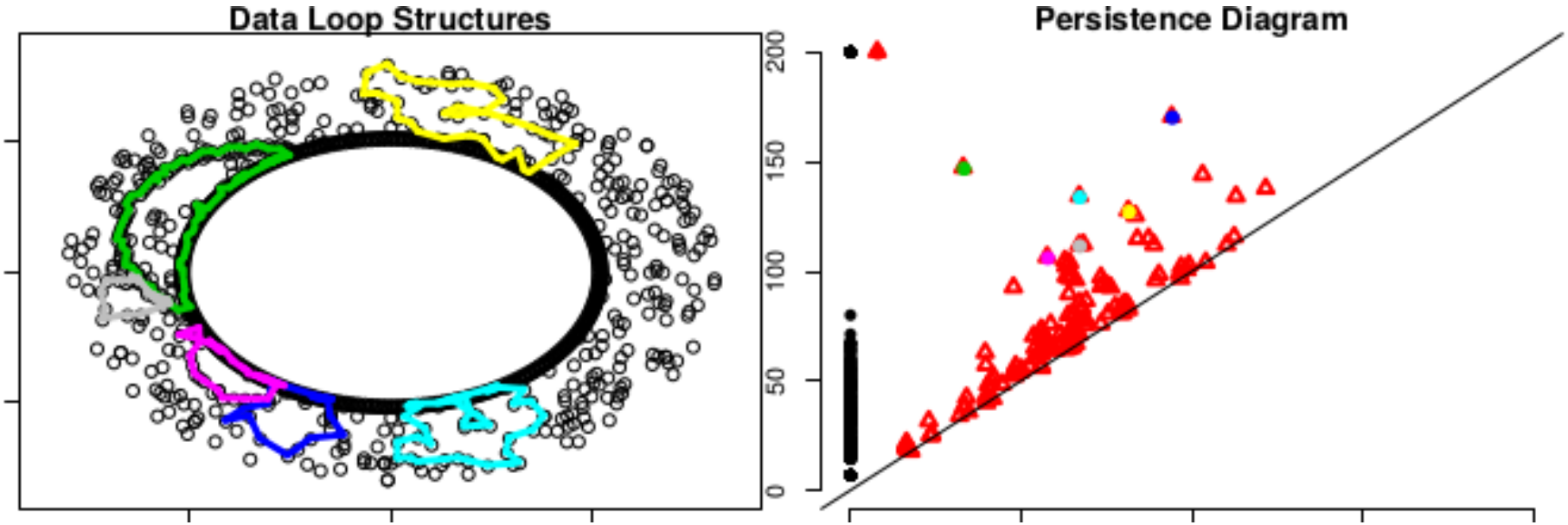


Table of Contents

Weighted PH

The problem is that you don't maximise your chance using PH because you don't use all the information from previous results. The idea is to use those information (energy deposit, drift time etc...) to define a new distance :

$$D(AB) = \frac{1}{P(A)}\frac{1}{P(B)}D_E(AB)$$

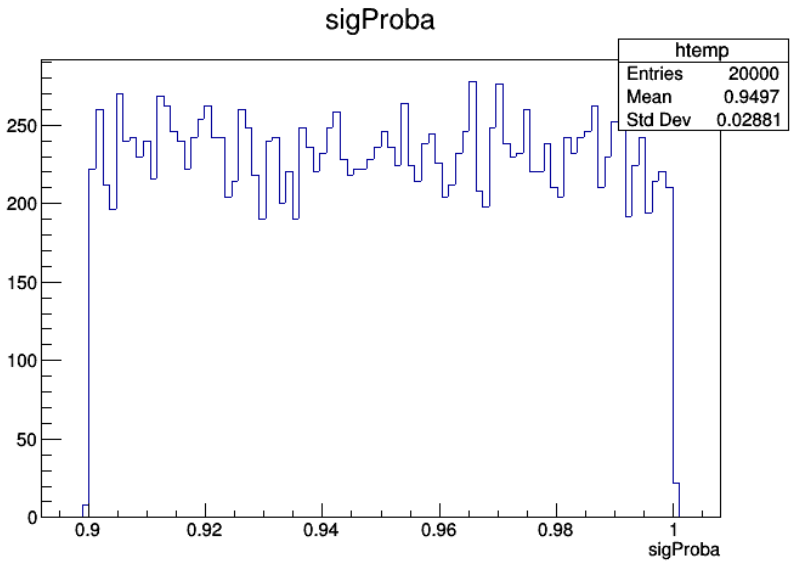
Where $P(X)$ is the probability of X being a signal hit, and $D_E(AB)$ the euclidean distance between A and B.

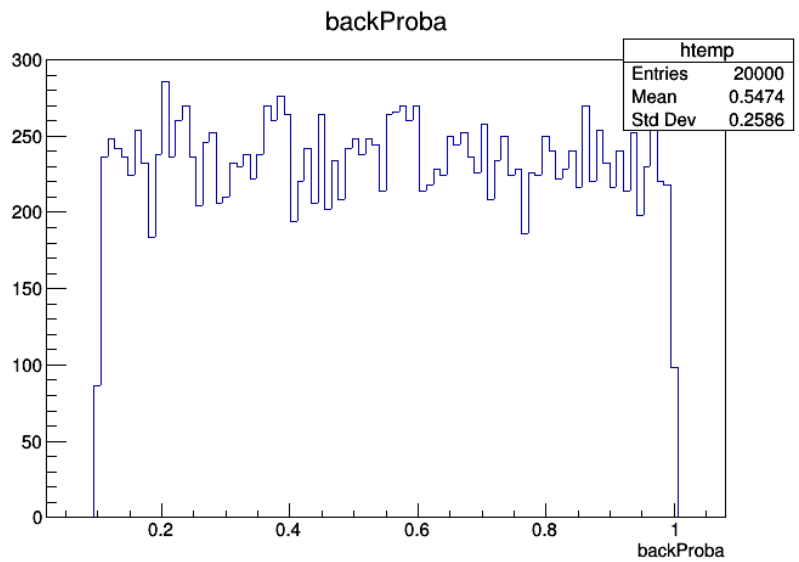
Effects expectation : - Lower the numbers of features (fixed max ϵ)
- Better loop efficiency
- Better loop purity

Table of Contents

Example parameters

To see the difference between both method, I decide to use a simple example, where I sample a circle in 10 points, and add on top of it 10 background hits.
The probability of being a hit for signal hit is defined uniformly distributed between [0.9; 1]. And background probability of being a signal is uniformly distributed between [0.1; 1].

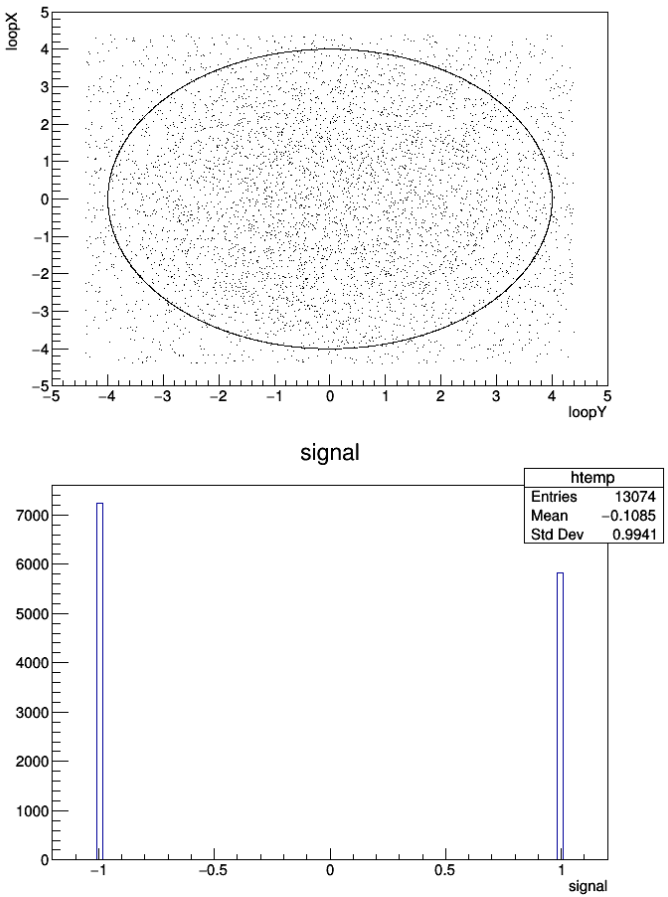




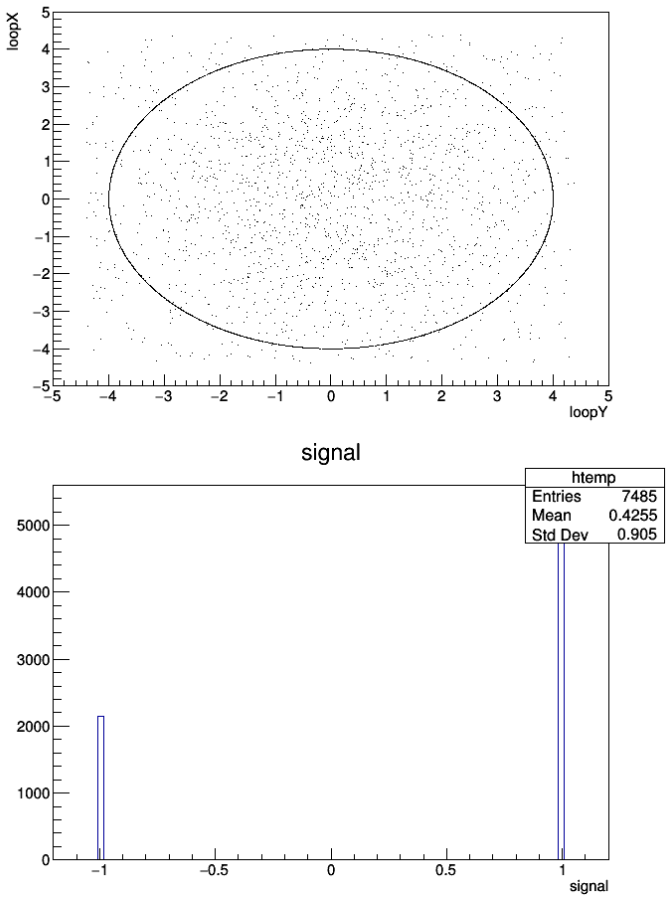
Global purity

First things we can do after that is to compare the ”global purity” of each methods

PH



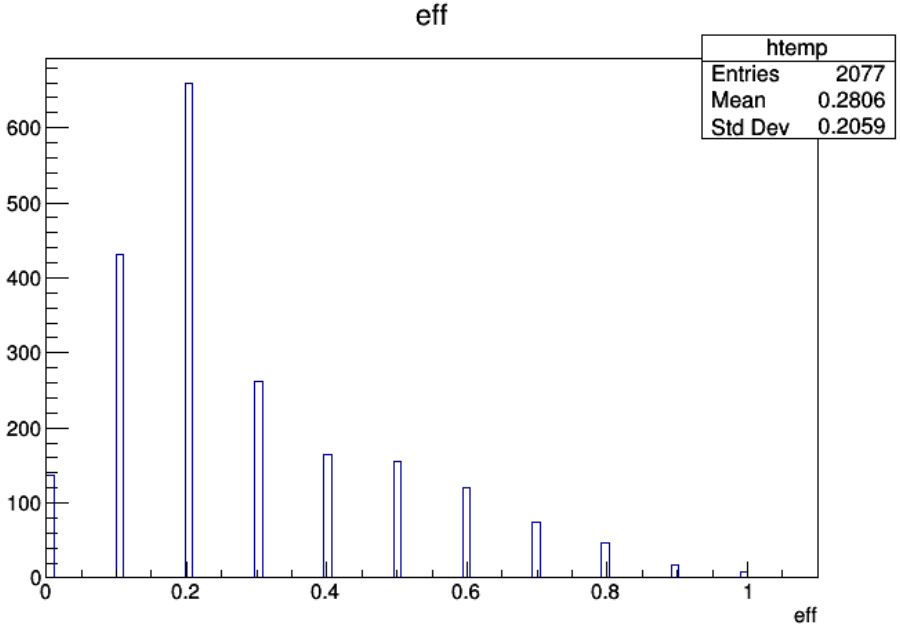
WPH



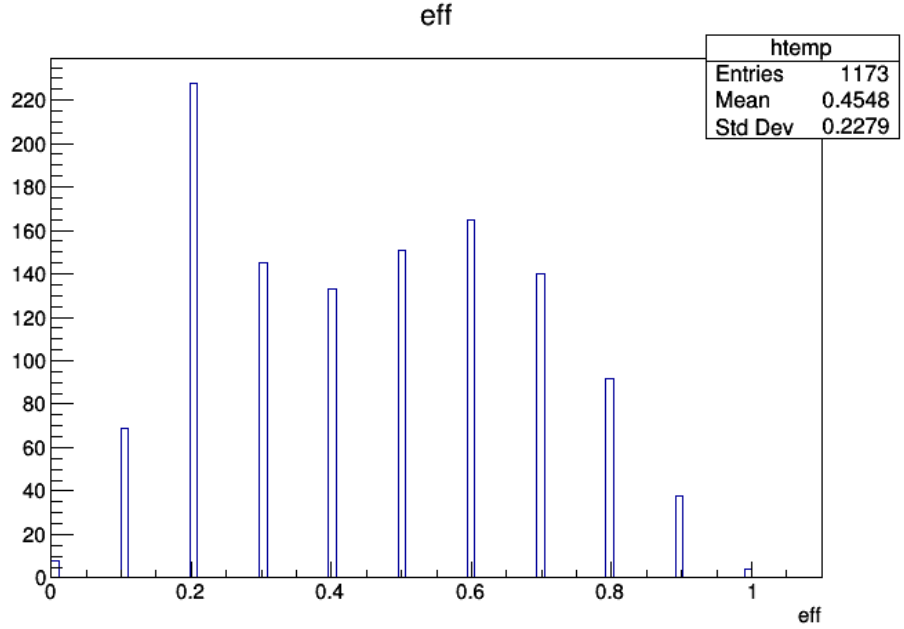
Preliminary results : Efficiency

$$loop\ efficiency = \frac{signal\ hit\ in\ loop}{total\ nb\ of\ signal\ hit}$$

PH



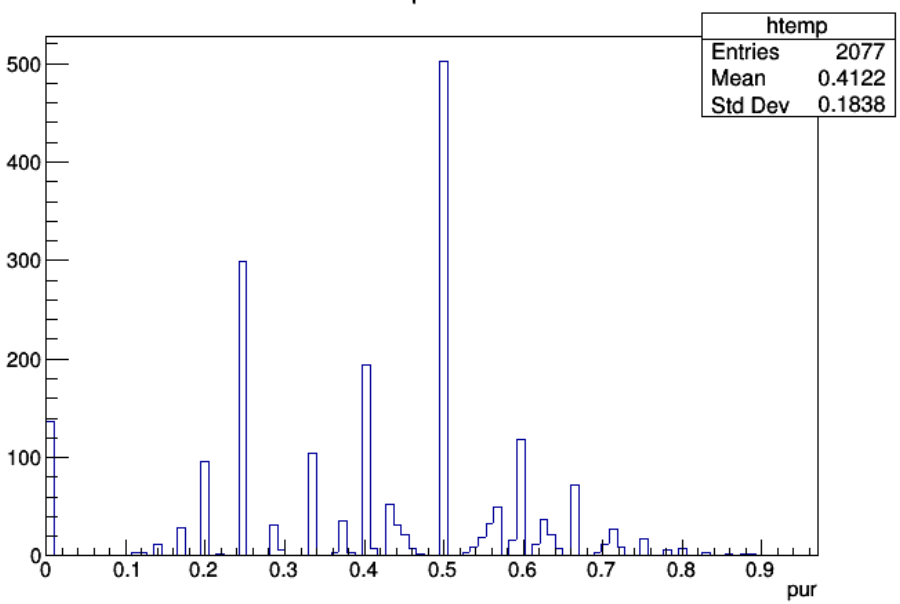
WPH



Preliminary results : Purity

$$loop\ purity = \frac{signal\ hit\ in\ loop}{number\ of\ hit\ in\ loop}$$

PH



WPH

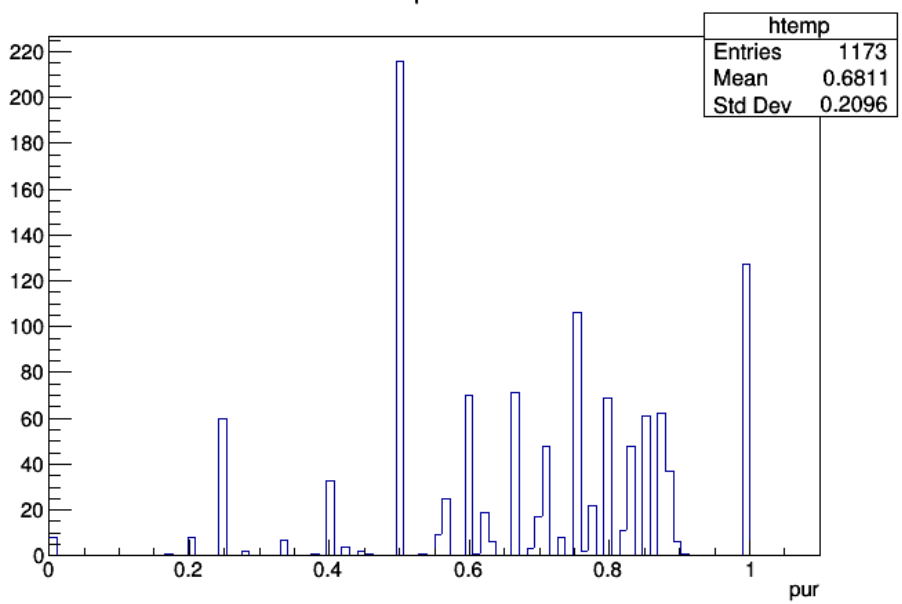
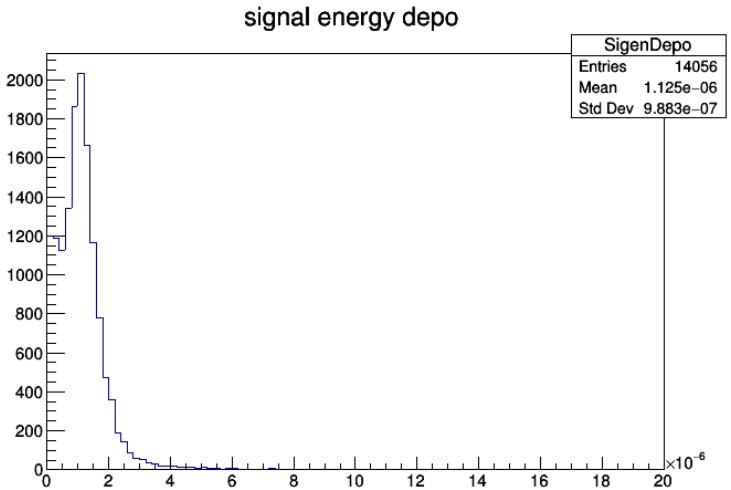
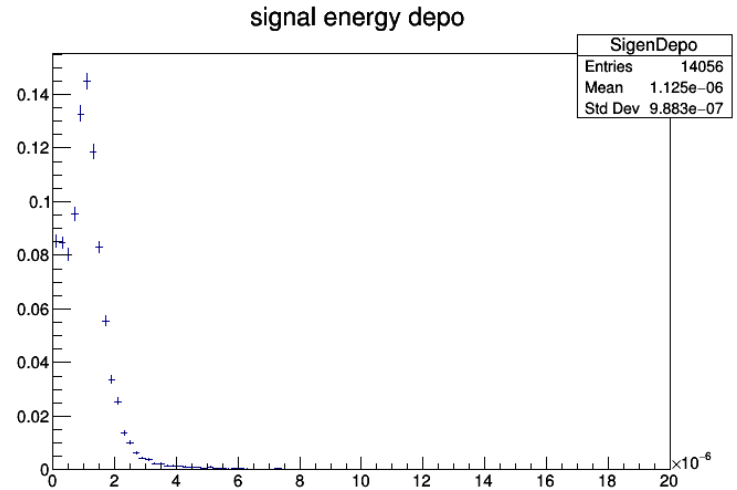


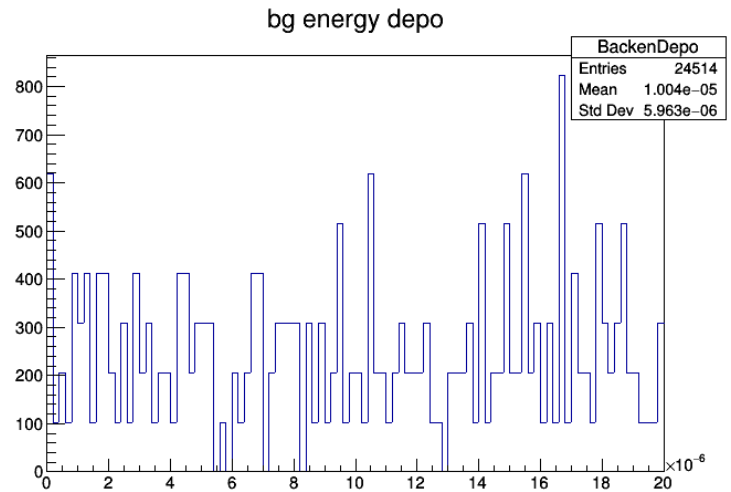
Table of Contents

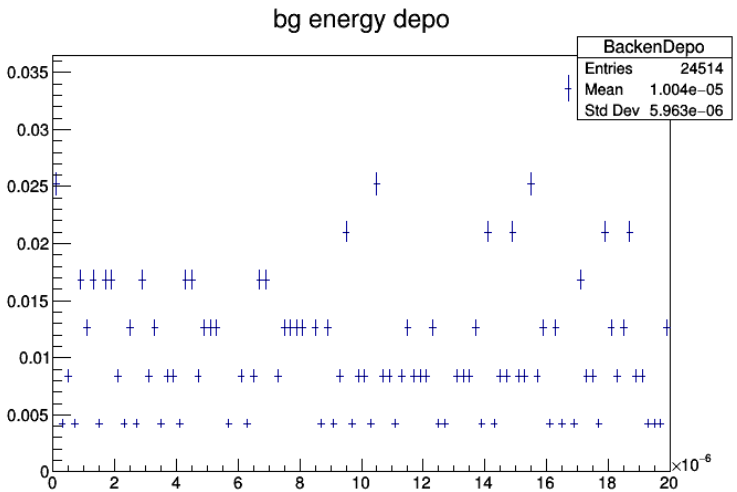
Example of processing probability function

Take the variables distribution (ex: energy deposit), normalize it.
Background has been generated uniformly.









Probability function example

$$P_X(En.Dep) = \frac{Signal \;\; hit \;\; in \; bin(En.Dep)}{Signal \;\; + \;\; Backgroundhit \;\; in \; bin(En.Dep)}$$

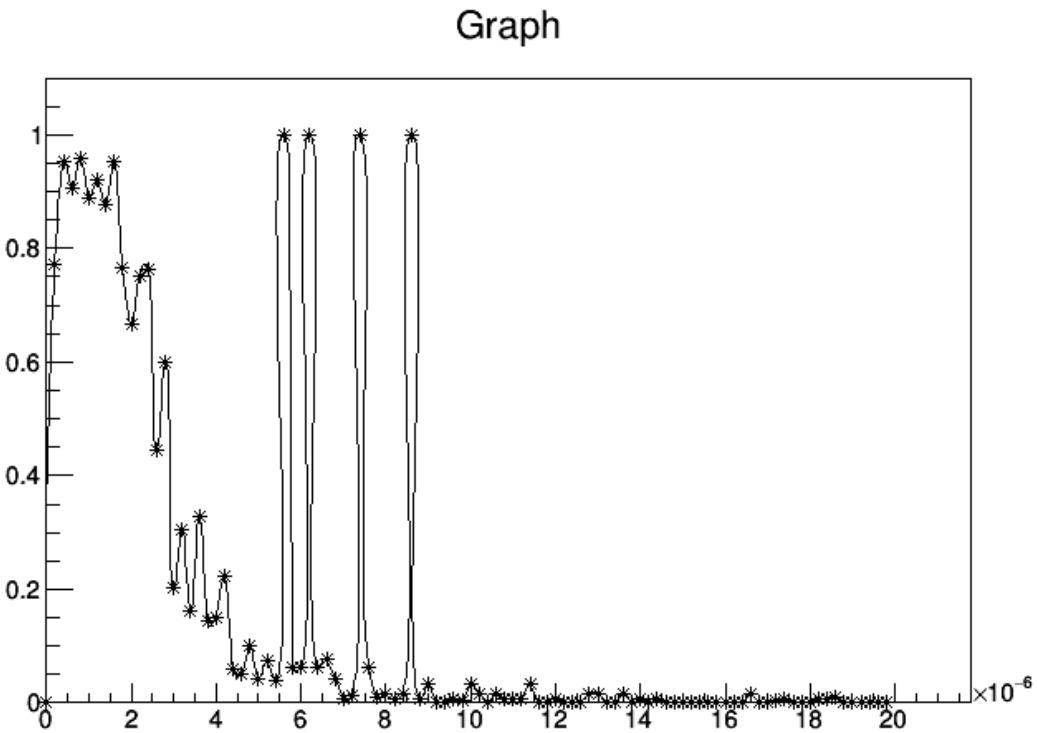


Table of Contents

Summary

- We talk about persistent homology and introduce WPH.
- By comparing both of them, WPH has better results than PH, without sorting the loops and biased statistics.
- Looking closely at the purity results, we see that both methods need circle fit addition to reject background and increase efficiency (Hough Transform, Appolonius Circle, etc..)
- There is a simple way that can be used to make probability function of being a signal (need signal and background distribution) backup plan may be needed (machine learning ?)