COMPUTING METHODS IN HEP Project work Spring 2025

(To be returned by 24.00 on Friday 23.5.2025)

The CMS experiment was designed to be able to measure muons with a good precision. Your task is to study how fast the Z boson can be detected in the LHC Run 3 with 13.6 TeV center-of-mass energy in the $Z \to \mu\mu$ decay channel.

The main background for this channel is the $t\bar{t}$ production.

1. The events are triggered with a HLT_DoubleIsoMu20_eta2p1 trigger with symmetric thresholds: $|\eta| < 2.1$ and $p_T > 20$ GeV/c for both muons. Ignore the isolation (Iso) at this point.

   Make an educated guess how much events you should simulate. Justify your decision.

   Create generator-level datasets with events passing the trigger for

   a) signal
   b) $t\bar{t}$ background

   Store the events in a ROOT TTree.

   What is the trigger efficiency for the signal?
   (6 points)

2. Analyze your data.

   a) Identify the signal muons. Take as the muon candidates the muons in the event. Apply the following operations and standard selection criteria to each of the muon candidates:
      - Simulate the muon measurement error by applying 1% Gaussian smearing to their momenta and 2 mrad Gaussian smearing to their angles θ and ϕ.
      - Selection of the two signal muons: Require that the muon candidates satisfy $p_T > 30$ GeV/c.
      - Track isolation: Require that the sum of the momenta of charged pions within $\Delta R < 0.3$ of a muon candidate is smaller than 1.5 GeV/c in p_T; after this criterion at least two muon candidates have to exist in the event

      Report the number of events passing the selection. (6 points)

   b) Reconstruct the invariant mass $M_{\mu\mu}$. Produce a ROOT histogram of $M_{\mu\mu}$ with correct normalization (i.e. that each entry in the histogram corresponds to a correct amount of cross-section in fb). Plot sum of the signal and background invariant mass distributions. (6 points)

   c) Fit the signal+background $M_{\mu\mu}$ histogram with appropriate functions over an appropriate mass range. Using the fit, create a function for the background only.

      Choose a reasonable mass window and calculate the number of signal and background events integrating the fit functions over the mass window.

      Estimate the statistical significance of the signal peak with the naive expression $N_S/\sqrt{N_B}$, where N_S and N_B are the number of signal and background events, respectively. For a discovery you need a 5 sigma signal. How much integrated luminosity you need for such a signal? Approximately how long does it take to reach such an integrated luminosity in CMS in 2024 data taking. (6 points)

3. If the data were real data, you'd have to present your work to the world-wide scientific community. Reason why the scientific community should believe in your results. Can you convince yourself? Why/why not? How would you improve your study to make it more realistic? (6 points)

Please push your answers into your public git repository. Include in your answer a written description how you proceed and where you argument any non-trivial choice you make.