PHYS715 Particle Data Analysis in High Energy Physics – Final Exam Questions

The following tasks have been assigned to six groups. The final exam will take place on June 6, 2026, at 13:30. Each group is required to present their assigned tasks orally via slides within 20 minutes. Analysis Tasks are described in the next page. Expected deliverables for the analysis tasks are as follows

- Clear explanation of the physics process
- Description of the dataset and selection strategy
- Plots of invariant mass spectra
- Justification of optimization choices
- Critical comparison of methods if necessary

Finally, each group should prepare a report summarizing the analysis methods and results of all tasks.

Group 1 (6 students)	**Tasks**
Cengizhan Koyutürk Onur Karakaş Emre Yağız Öztürk Nazlıcan Yıldırım Özgür Hancı Burhan Karaduman	P1 (20p). Physics and Upgrades of TimePix Detectors at CMS Experiment P2 (20p). Data Analysis 1 P3 (20p). Data Analysis 2 P4 (40p). Data Analysis 3

Group 2 (6 students)	**Tasks**
Yunus Oynaş Ahmet Sünbül Burak Özdamar Selman Diken Mehmet İnan Yunus Emre Araz	P1 (20p). Physics and Operating Principles of Cherenkov Detectors P2 (20p). Data Analysis 1 P3 (20p). Data Analysis 2 P4 (40p). Data Analysis 4

Group 3 (5 students)	**Tasks**
Şahin Kaya Altar Özmenler Ömer Arslan Özgür Karayel Mehmet Yasin Demir	P1 (20p). KATRIN Experiment and its foundations P2 (20p). Data Analysis 1 P3 (20p). Data Analysis 2 P4 (40p). Data Analysis 5

Group 4 (5 students)	**Tasks**
Kemal Mert Dermirten Elda Koldemir İzel Yeniocak Batır Levashov Yunus Faruk Ağan	P1 (20p). Physics and Upgrades of RPC at ATLAS Experiment P2 (20p). Data Analysis 1 P3 (20p). Data Analysis 2 P4 (40p). Data Analysis 6

Group 5 (5 students)	**Tasks**
Emin Yüksel Neşe Ergin Buğrahan Kılıç Murat Barış Boyacıoğlu Manzoor Ahmed	P1 (20p). Physics and Upgrades of Tracking Detectors at ATLAS Experiment P2 (20p). Data Analysis 1 P3 (20p). Data Analysis 2 P4 (40p). Data Analysis 7

Group 6 (5 students)	**Tasks**
Canay Öz Erkin Ertekin Aysu Keseroğlu Kardelen Kaşif Ece Kondakçı	P1 (20p). DsTau Experiment and its foundations P2 (20p). Data Analysis 1 P3 (20p). Data Analysis 2 P4 (40p). Data Analysis 8

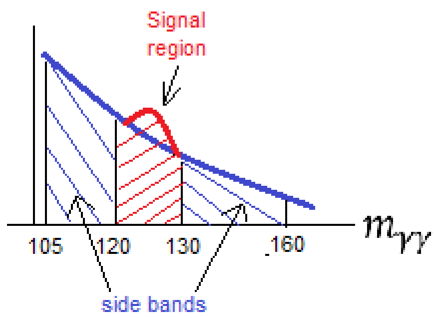
Analysis Tasks

1. Reconstruction of the Higgs Boson using ATLAS Open Data

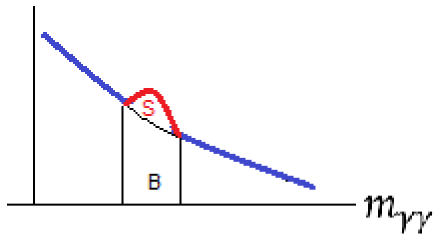
- Use datasets under:

`/hepdata/Atlas/GamGam/MC/*.root`	Monte Carlo simulation for $H \to \gamma\gamma$ signal events
`/hepdata/Atlas/GamGam/Data/*.root`	Small real collision data sample (signal+background)
`/hepdata/Atlas/GamGam/FullData/*.root`	Large real Data sample (signal+background)

- Reconstruct the invariant mass spectrum of two photons ($m_{\gamma\gamma}$) as we did during the lecture to identify the Higgs boson peak and estimate the signal and background contributions in the signal region. We performed cut-based analysis. You don't need to optimize anything up to this stage. However, you should evaluate number of signals (S_0) within mass window $120 \text{ GeV} < m_{\gamma\gamma} < 130 \text{ GeV}$ in data via fitting procedure.
- Generate two sub datasets for machine learning algorithms
  `signal.root` containing $H \to \gamma\gamma$ signal variables (p_T, η, ϕ, isolation, derived variables, etc)
  `background.root` containing $X \to \gamma\gamma$ background variables. (p_T, η, ϕ, isolation, derived variables, etc)
  Use real collision data sidebands as background. To do that select your background candidates around two regions $105 \text{ GeV} < m_{\gamma\gamma} < 120 \text{ GeV}$ and $130 \text{ GeV} < m_{\gamma\gamma} < 160 \text{ GeV}$.



- Develop a machine learning approach to improve signal significance. (Use any BDT classifier, TMVA, TensorFlow or PyTorch). Optimize the Higgs boson selection by maximizing the signal significance $S/\sqrt{B}$ or $S/\sqrt{S+B}$ using the invariant mass spectrum. Here S is the number of signal candidates and B is the number of backgrounds around signal region (peak) of the invariant mass spectrum. Both S and B must be obtained via fitting procedure in root.



- Compare the performance of the cut-based method and machine learning approaches in terms of efficiency (ε), purity (P), and their product (εP) where $\varepsilon = S/S_0$, $P = S/(S+B)$ and $\varepsilon P = \frac{S^2}{S_0(S+B)}$.
- Draw the local p-value vs $m_{\gamma\gamma}$ plot for both methods.
- Best group will get extra points.
- If you share your analysis code, methods, plots, or similar materials with another group, the 20 points assigned to this analysis will be divided equally between the groups involved.

2. Dark Matter Research at CMS

- Consider the study at this link which is related to Search for the Production of Dark Matter Candidates using CMS open data. We'll try to repeat this study at very basics level using this dataset (13 TeV CMS Open Data). You can find a large sample at the path `/hepdata/CMS/DarkMatter/*.root`
- Generate invariant mass spectra of <u>all</u> oppositely charged muon pairs, $m_{\mu\mu}$, for $m_{\mu\mu} > 120 \text{ GeV}$. You can apply suitable selection cuts if required.
- Plot the local p-value vs two-muon invariant mass and comment on the results. See lecture notes.
- If you share your analysis code, methods, plots, or similar materials with another group, the 20 points assigned to this analysis will be divided equally between the groups involved.

3. Reconstruction of $\omega(782)$ meson using ALEPH Data.

- You will use both MC and real data samples.
  `/hepdata/Aleph/AlephMC*.root` and `/hepdata/Aleph/AlephDA*.root`
- Select the dominant decay mode.
- Evaluate the reconstruction performance using a classical cut-based analysis method. Use MC to optimize ω signal selection cuts by maximizing the signal significance $S_{rec}/\sqrt{B_{rec}}$. You can count signals and backgrounds using histogram contents around signal without fitting.
- Use real data to evaluate number of signals S (and B) by fitting procedure in root using optimized cuts in MC analysis.
- Calculate average number of ω mesons generated per hadronic Z decay in data using the following normalization:

$$\langle n_\omega \rangle = \left(\frac{S}{S_{rec}}\right)\left(\frac{S_{gen}}{N_{gen}}\right)\left(\frac{N_{MC}}{N_{DA}}\right)\left(\frac{1}{\mathcal{B}}\right)$$

  where
  S is the number of reconstructed signals in real data (after event selection cuts)
  S_{rec} is the number of reconstructed signals in MC (after event selection cuts)
  S_{gen} is the number of generated signals in MC (before event selection cuts)
  N_{gen} is the number of generated events in MC (before event selection cuts)
  N_{MC} is the number of reconstructed events in MC (after event selection cuts)
  N_{DA} is the number of reconstructed events in real data (after event selection cuts)
  $\mathcal{B}$ is the branching ratio
- Simulate 1 million collision events ($e^+e^- \to Z \to q\bar{q} \to \text{hadrons}$) using Pythia 8 to calculate $\langle n_\omega \rangle$.
- Compare your results with the PDG.

4. Reconstruction of η meson using ALEPH Data.

- You will use both MC and real data samples.
  `/hepdata/Aleph/AlephMC*.root` and `/hepdata/Aleph/AlephDA*.root`
- Select $\eta \to \pi^+\pi^-\gamma$ decay mode.
- Evaluate the reconstruction performance using a classical cut-based analysis method. Use MC to optimize η signal selection cuts by maximizing the signal significance $S_{rec}/\sqrt{B_{rec}}$. You can count number of signals (S_{rec}) and backgrounds (B_{rec}) using histogram contents around signal without fitting.
- Use real data to evaluate number of signals S (and B) by fitting procedure in root using optimized cuts in MC analysis.
- Calculate average number of η mesons generated per hadronic Z decay in data using the following normalization (see problem 3 for details):

$$\langle n_\eta \rangle = \left(\frac{S}{S_{rec}}\right)\left(\frac{S_{gen}}{N_{gen}}\right)\left(\frac{N_{MC}}{N_{DA}}\right)\left(\frac{1}{\mathcal{B}}\right)$$

- Simulate 1 million collision events ($e^+e^- \to Z \to q\bar{q} \to \text{hadrons}$) using Pythia 8 to calculate $\langle n_\eta \rangle$.
- Compare your results with the PDG.

5. Reconstruction of $D^\pm$ meson using ALEPH Data.

- You will use both MC and real data samples.
  `/hepdata/Aleph/AlephMC*.root` and `/hepdata/Aleph/AlephDA*.root`
- Select $D^+ \to 2\pi^+\pi^-$ and $D^- \to 2\pi^-\pi^+$ decay mode.
- Evaluate the reconstruction performance using a classical cut-based analysis method. Use MC to optimize $D^\pm$ signal selection cuts by maximizing the signal significance $S_{rec}/\sqrt{B_{rec}}$. You can count number of signals (S_{rec}) and backgrounds (B_{rec}) using histogram contents around signal without fitting.
- Use real data to evaluate number of signals S (and B) by fitting procedure in root using optimized cuts in MC analysis.
- Calculate average number of $D^\pm$ mesons generated per hadronic Z decay in data using the following normalization (see problem 3 for details):

$$\langle n_{D^\pm} \rangle = \left(\frac{S}{S_{rec}}\right)\left(\frac{S_{gen}}{N_{gen}}\right)\left(\frac{N_{MC}}{N_{DA}}\right)\left(\frac{1}{\mathcal{B}}\right)$$

- Simulate 1 million collision events ($e^+e^- \to Z \to q\bar{q} \to \text{hadrons}$) using Pythia 8 to calculate $\langle n_{D^\pm} \rangle$.
- Compare your results with the PDG.

6. Lifetime measurement of K_S^0 meson using ALEPH Data.

- You will use both MC and real data samples.
  `/hepdata/Aleph/AlephMC*.root` and `/hepdata/Aleph/AlephDA*.root`
- Select the $K_s \rightarrow \pi^+\pi^-$ decay mode.
- Use secondary vertices.
- Optimize the selection of K_s signals using MC.
- Plot the distribution of lifetime of the K_s candidates.
- Evaluate the mean lifetime of K_s via fitting procedure in root.
- Simulate 1 million collision events ($e^+e^- \rightarrow Z \rightarrow q\bar{q} \rightarrow \text{hadrons}$) using Pythia 8 to calculate the mean lifetime.
- Compare your result with PDG.

7 . Lifetime measurement of Λ^0 baryon using ALEPH Data.

- You will use both MC and real data samples.
  `/hepdata/Aleph/AlephMC*.root` and `/hepdata/Aleph/AlephDA*.root`
- Select the dominant mode.
- Use secondary vertices.
- Optimize the selection of Λ^0 signals using MC.
- Plot the distribution of lifetime of the Λ^0 candidates.
- Evaluate the mean lifetime of Λ^0 via fitting procedure in root
- Simulate 1 million collision events ($e^+e^- \rightarrow Z \rightarrow q\bar{q} \rightarrow \text{hadrons}$) using Pythia 8 to calculate the mean lifetime.
- Compare your results with PDG.

8 . Ratio of Branching ratio measurement of D_s^+ meson using ALEPH Data.

- You may use both MC and real data samples.
  `/hepdata/Aleph/AlephMC*.root` and `/hepdata/Aleph/AlephDA*.root`
- Consider decay modes $D_s^+ \rightarrow K^+K_S^0(\pi^+\pi^-)$ and $D_s^+ \rightarrow \pi^+\phi(K^+K^-)$
- Optimize the selection cuts for both decays.
- Evaluate the ratio $R = \mathcal{B}[D_s^+ \rightarrow \pi^+\phi(K^+K^-)]\ /\ \mathcal{B}[D_s^+ \rightarrow K^+K_S^0(\pi^+\pi^-)]$
  This ratio is equivalent to $R = \frac{S_{\pi\phi}/\varepsilon_{\pi\phi}}{S_{KK_S}/\varepsilon_{KK_S}}$ where S is the number of reconstructed particles obtained by fitting procedure in real data and ε is the reconstruction efficiency which can be computed by using MC events. Namely, the efficiency can be computed by $\varepsilon = S_{rec}/S_{gen}$ (see problem 3).
- Simulate 1 million collision events ($e^+e^- \rightarrow Z \rightarrow q\bar{q} \rightarrow \text{hadrons}$) using Pythia 8 to calculate the equivalent ratio.
- Compare your results with PDG.

Good luck.
Ahmet Bingül
May 19, 2026