MC Matching

Tia Crane

2025 Belle II Summer Workshop
24 June 2025

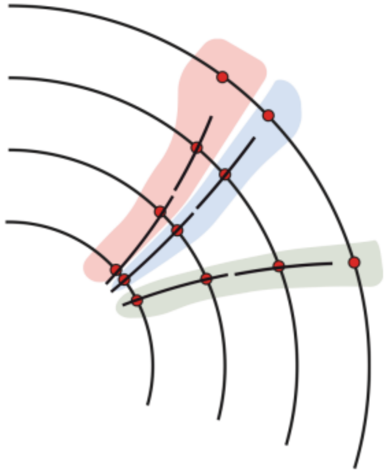
Introduction

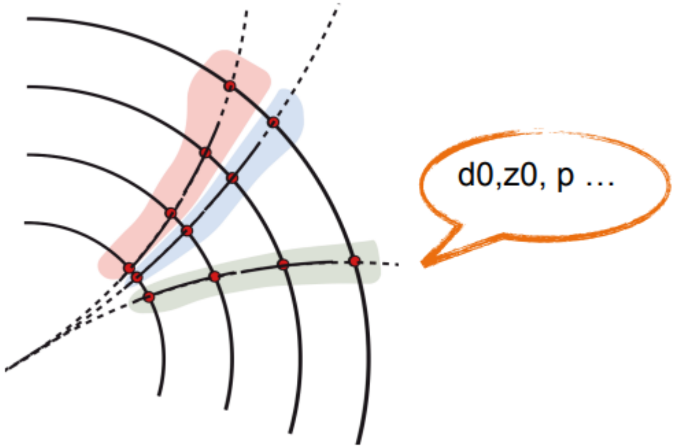
- Why do we need MC matching?
  - Optimization of selection criteria
  - Signal identification
    - Efficiency, purity
  - Sample composition
    - Background studies, selection optimization

- MC matching is NOT an exact science
  - Many valid methods
  - Requires interpretation to decide how we relate tracks and cluster candidates with MC particles

Track Matching

- Reconstructed track candidates (particle = data object) undergo hit pattern recognition (PR)
- Generated MC particles have GEANT simulated interaction with the detector ($hits_{MC}$)

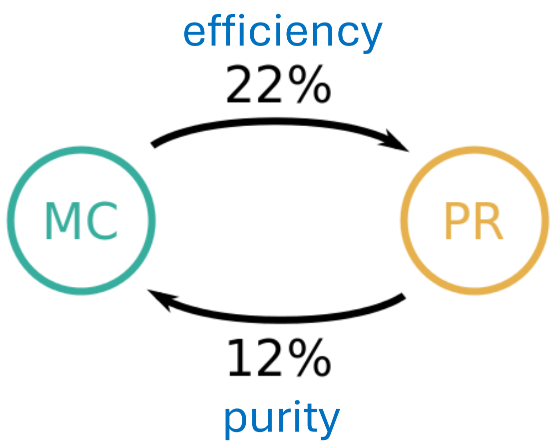




Yubo Han, 2023 Belle II Physics week

Track Matching

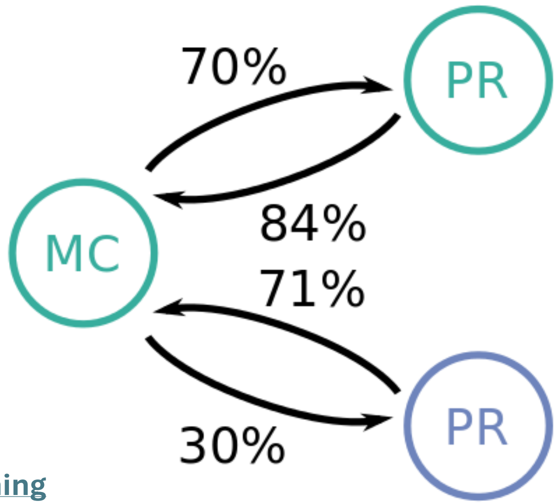
- Compare generated $hits_{MC}$ with overlap of hits used to reconstruct track object



Purity: PR has 12% of its hits coming from MC

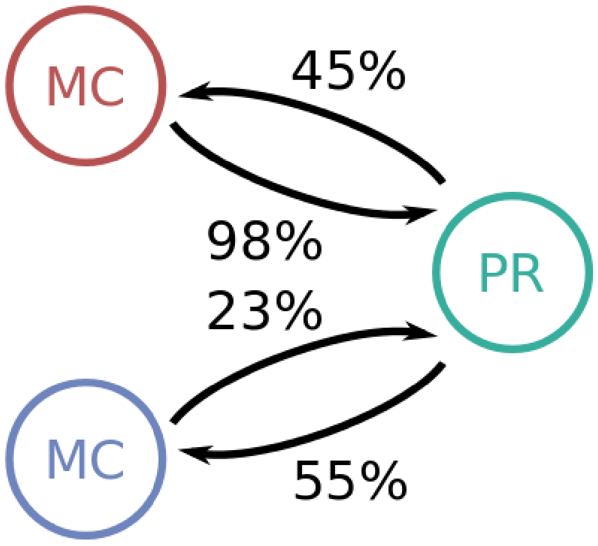
Efficiency: 22% of $hits_{MC}$ from MC particle are used in PR

Mainly background hits in reconstructed track pattern recognition (PR) = BACKGROUND

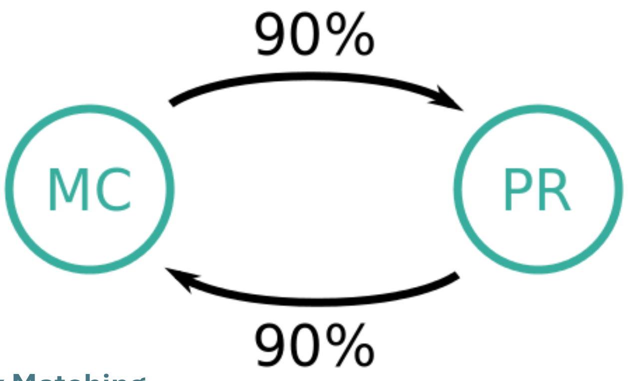


PR of 2 reconstructed tracks meet BOTH the efficiency (5%) and purity (66%) threshold to match to MC particle = CLONE

Track Matching

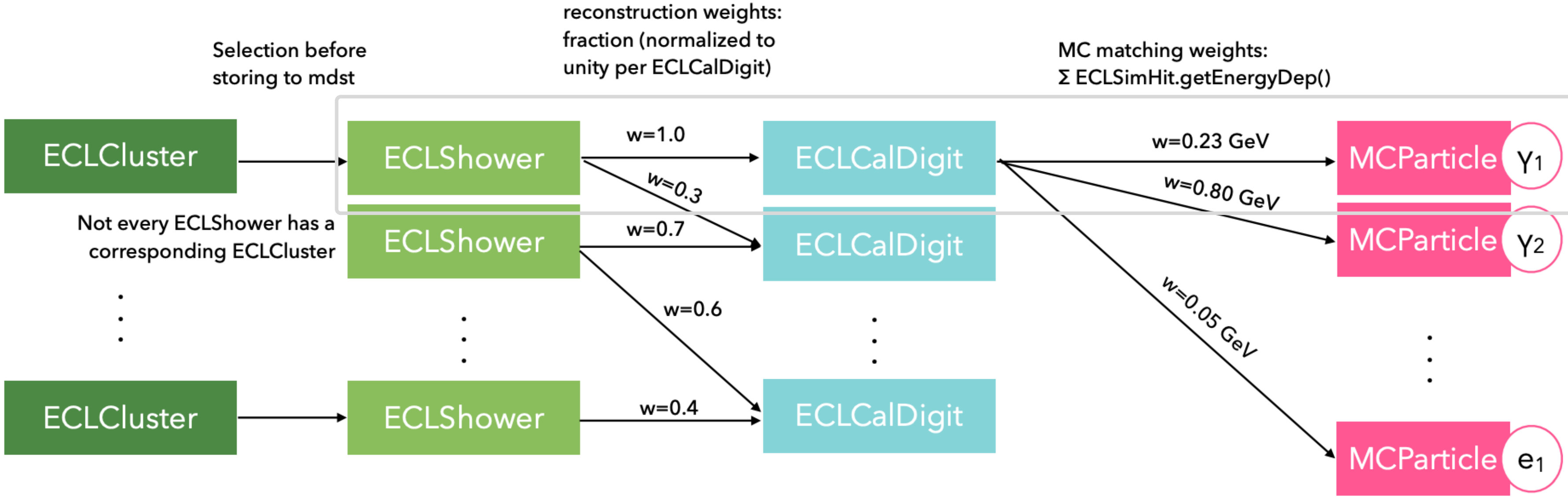


No MC matches track PR with enough purity AND efficiency = GHOST



One-to-one connection = MATCHED

Cluster Matching



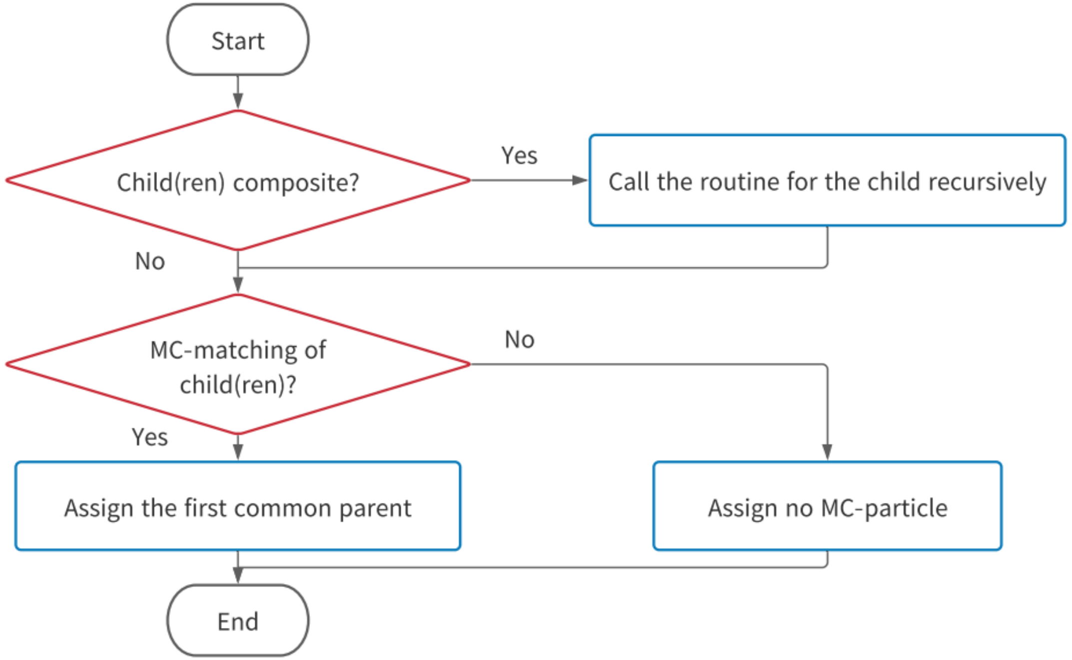
$$\omega = 1.0 * 0.23 \\ = 0.23\ GeV$$

- Track matching between tracks and shower objects
- Unmatched clusters reconstructed as photons if $\omega/E_{rec} > 0.2GeV$ AND $\omega/E_{MC} > 0.3GeV$
  - If clone = ONLY match candidate with highest ω

MC Matching

- How can we complete this in our reconstruction?

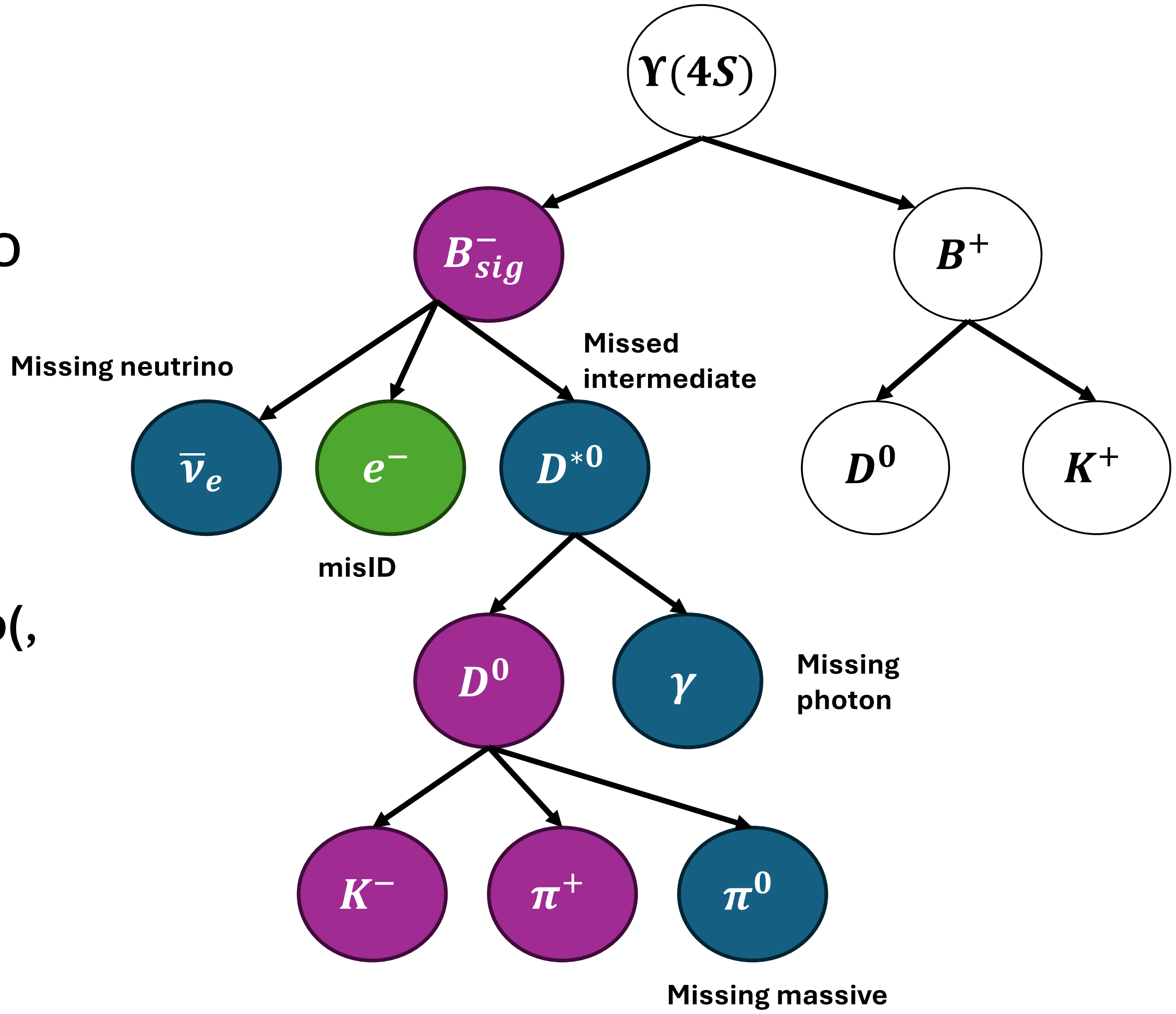
ma.matchMCTruth('B-:rec', path=pathname)



Yoto Sa, Monte Carlo matching in the Belle II software

MC Matching

- What are some ways this can go wrong?
  - Clone tracks
  - Fake tracks
  - Multiple cluster matches
    - nMCMatches, photonHasOverlap(, ‘gamma:all’, ‘e-:all’)
  - Missed intermediate particles
  - Etc.



Interpretation

- MC info

  mcPDG

  mcErrors = information on success/failure in the association of reconstructed particle objects and MC particles

- **isSignal** (**mcErrors==0**)

  isSignalAcceptMissingGamma, **isSignalAcceptMissingNeutrino**, etc.

- MC reconstruction
  - Counting experiment: determine how many reconstructed signal events were generated

  fillParticleListFromMC(decayString, cut="", addDaughters=True, path=pathName)

  reconstructMCDecay(decayString, cut="", path=pathName)

Interpretation

- Decay strings, markers, and keywords

B- -> D0 pi- : decays via intermediate resonances and/or with radiative photon are counted as signal

B- =direct=> D0 pi- : decays via intermediate resonances are NOT signal but decays with radiative photon are counted as signal

B- -> D0 (decay)pi- : decays in flight (e.g. $\pi^- \to \mu\nu_\mu$) are counted as signal

B- -> D0 (misID)pi- ?nu : decays with a misIDed pion and a neutrino (e.g. $B^- \to D^0 e^- \bar{\nu}_e$) are counted as signal

Demo + Discussion

Generated Physics Processes

Tia Crane

2025 Belle II Summer Workshop
24 June 2025

Introduction

- How is the MC information of the reconstructed particle objects different than the generated physics process? Why is this important?

Introduction

- How is the MC information of the reconstructed particle objects different than the generated physics process? Why is this important?

Generated event level interpretation

Remove reconstruction bias + Provide exact MC truth information + Represents true (simulated) physics process + ...

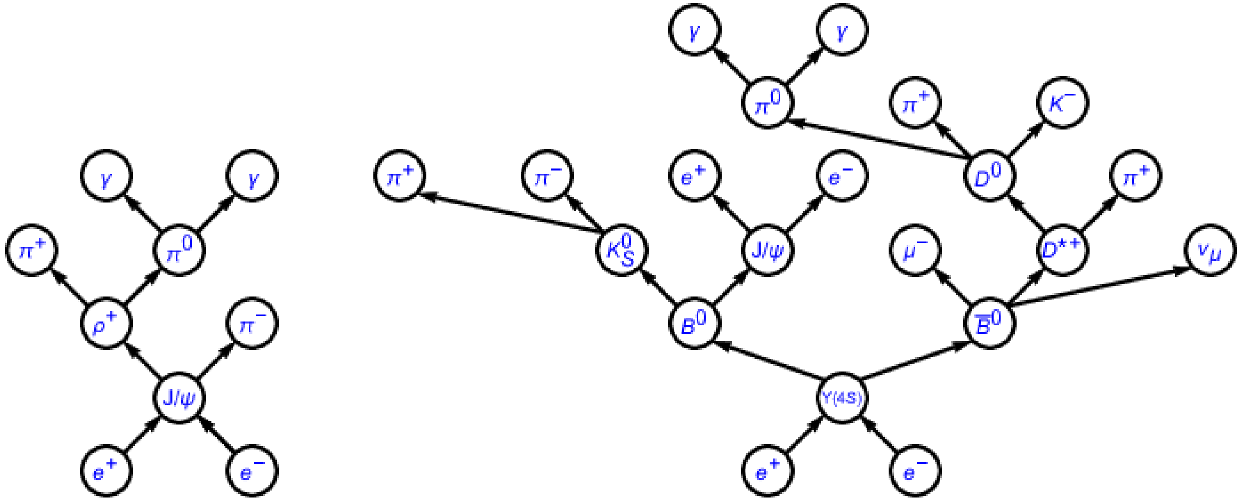
Matched particle lists

Requires interpretation + Reflects reconstruct object (NOT actual decay process simulated + ...

Important to know the true simulated decay (background/efficiency studies) AND truth of particles (misID rate, where things goes wrong)

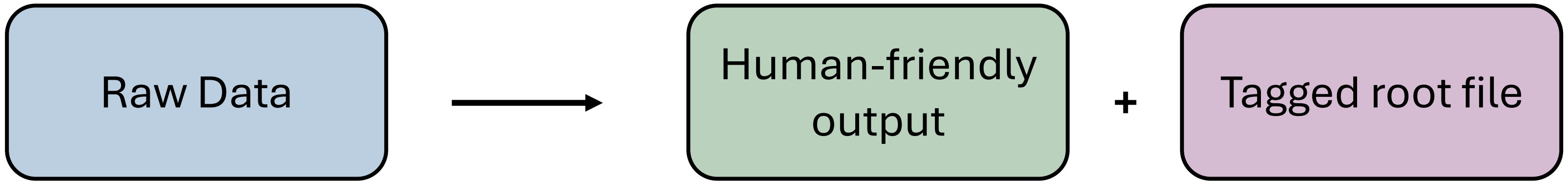
TopoAna

- Offline topology analysis
- Input = raw MC truth information (root TTree entries):
  - Number of MC particles
  - PDG codes of MC particles
  - Mother indices of MC particles
- Based on configured attributes (card file):



(a) (b)

TopoAna, GitLab



e.g. PDF

TopoAna

- [User Manual]: configuration options
- **Component analysis**: indices based on yields in sample (subject to change)

Table 1: Decay trees and their respective initial-final states.

rowNo	decay tree (decay initial-final states)	iDcyTr	nEtr	nCcEtr	nAllEtr	nCEtr
1	$\Upsilon(4S) \to B^+B^-, B^+ \to e^+\nu_e\bar{D}^{*0}, B^- \to \pi^-\omega\bar{K}^*J/\psi, \bar{D}^{*0} \to \bar{D}^0\gamma, \omega \to \pi^0\pi^+\pi^-,$ $\bar{K}^* \to \pi^+K^-, J/\psi \to \pi^+b_1^-, \bar{D}^0 \to \pi^-K^+, b_1^- \to \pi^-\omega, \omega \to \pi^0\pi^+\pi^-$ $(\Upsilon(4S) \dashrightarrow e^+\nu_e\pi^0\pi^0\pi^+\pi^+\pi^+\pi^+\pi^-\pi^-\pi^-\pi^-\pi^-K^+K^-\gamma)$	10	2	0	2	2
2	$\Upsilon(4S) \to B^+B^-, B^+ \to e^+\nu_e\pi^0\bar{D}^0, B^- \to e^-\bar{\nu}_eD^{*0}, \bar{D}^0 \to \pi^-K^+, D^{*0} \to \pi^0D^0,$ $D^0 \to e^+\nu_eK^-$ $(\Upsilon(4S) \dashrightarrow e^+e^+e^-\nu_e\nu_e\bar{\nu}_e\pi^0\pi^0\pi^-K^+K^-)$	109	2	0	2	4
3	$\Upsilon(4S) \to B^+B^-, B^+ \to e^+\nu_e\bar{D}^{*0}, B^- \to \pi^0\pi^-D^+D_s^-, \bar{D}^{*0} \to \pi^0\bar{D}^0, D^+ \to \pi^+\pi^+\pi^-\eta,$ $D_s^- \to \pi^-\eta', \bar{D}^0 \to \pi^0\pi^-K^+, \eta \to \gamma\gamma, \eta' \to \rho^0\gamma, \rho^0 \to \pi^+\pi^-$ $(\Upsilon(4S) \dashrightarrow e^+\nu_e\pi^0\pi^0\pi^0\pi^+\pi^+\pi^+\pi^-\pi^-\pi^-\pi^-\pi^-K^+\gamma\gamma\gamma)$	144	2	0	2	6

Table 2: Cascade decay branches of B^- (only the first four hierarchies are involved).

rowNo	cascade decay branch of B^-	iCascDcyBrP	nCase	nCcCase	nAllCase	nCCase
1	$B^- \to e^-\bar{\nu}_eD^{*0}, D^{*0} \to \pi^0D^0, D^0 \to \pi^+K^-$	15	31	26	57	57
2	$B^- \to e^-\bar{\nu}_eD^{*0}, D^{*0} \to D^0\gamma, D^0 \to \pi^+K^-$	0	13	21	34	91
3	$B^- \to e^-\bar{\nu}_eD^0, D^0 \to \pi^+K^-$	24	11	19	30	121
4	$B^- \to e^-\bar{\nu}_eD^{*0}, D^{*0} \to \pi^0D^0, D^0 \to \pi^0\pi^+K^-$	18	2	4	6	127
5	$B^- \to e^-\bar{\nu}_e\pi^0D^0, D^0 \to \pi^+K^-$	32	2	2	4	131

Table 5: Exclusive components of $B^- \to e^-\bar{\nu}_e + anything$.

rowNo	exclusive component of $B^- \to e^-\bar{\nu}_e + anything$	iDcyBrIncDcyBr	nCase	nCcCase	nAllCase	nCCase
1	$B^- \to e^-\bar{\nu}_eD^{*0}$	0	56	56	112	112
2	$B^- \to e^-\bar{\nu}_eD^0$	2	15	25	40	152
3	$B^- \to e^-\bar{\nu}_eD_1^0$	8	6	4	10	162
4	$B^- \to e^-\bar{\nu}_eD_0^{*0}$	6	3	2	5	167
5	$B^- \to e^-\bar{\nu}_e\eta D^0$	3	1	3	4	171

- **Signal analysis**: indices based on card file (fixed)

Table 15: Signal cascade decay branches.

rowNo	signal cascade decay branch	iSigCascDcyBr	nCase	nCcCase	nAllCase	nCCase
1	$B^- \to \tau^-\bar{\nu}_\tau D^0, \tau^- \to e^-\bar{\nu}_e\nu_\tau$	0	4	2	6	6
2	$B^- \to \tau^-\bar{\nu}_\tau D^{*0}, \tau^- \to e^-\bar{\nu}_e\nu_\tau, D^{*0} \to \pi^0D^0$	1	0	0	0	6
3	$B^- \to \tau^-\bar{\nu}_\tau D^{*0}, \tau^- \to e^-\bar{\nu}_e\nu_\tau, D^{*0} \to D^0\gamma$	2	0	1	1	7

TopoAna: Cheat Sheet

iDcyTr: index of decay tree (used to reference decay in output nTuple)

nEtr: number of entries (charge dependent)

nCcEtr: number of entries of charge conjugate

nAllEtr: nEtr+nCCEtr

nCEtr: cumulative sum of nAllEtr

iCascDcyBr: index of decay (used to reference decay in output nTuple)

nCase: number of entries (charge dependent)

nCcCase: number of entries of charge conjugate

nAllCase: nCase+nCcCase

nCCase: cumulative sum of nAllCase

Table 1: Decay trees and their respective initial-final states.

rowNo	decay tree (decay initial-final states)	iDcyTr	nEtr	nCcEtr	nAllEtr	nCEtr
1	$\Upsilon(4S) \to B^+B^-, B^+ \to e^+\nu_e\bar{D}^{*0}, B^- \to \pi^-\omega\bar{K}^*J/\psi, \bar{D}^{*0} \to \bar{D}^0\gamma, \omega \to \pi^0\pi^+\pi^-,$ $\bar{K}^* \to \pi^+K^-, J/\psi \to \pi^+b_1^-, \bar{D}^0 \to \pi^-K^+, b_1^- \to \pi^-\omega, \omega \to \pi^0\pi^+\pi^-$ $(\Upsilon(4S) \dashrightarrow e^+\nu_e\pi^0\pi^0\pi^+\pi^+\pi^+\pi^+\pi^-\pi^-\pi^-\pi^-\pi^-K^+K^-\gamma)$	10	2	0	2	2
2	$\Upsilon(4S) \to B^+B^-, B^+ \to e^+\nu_e\pi^0\bar{D}^0, B^- \to e^-\bar{\nu}_eD^{*0}, \bar{D}^0 \to \pi^-K^+, D^{*0} \to \pi^0D^0,$ $D^0 \to e^+\nu_eK^-$ $(\Upsilon(4S) \dashrightarrow e^+e^+e^-\nu_e\nu_e\bar{\nu}_e\pi^0\pi^0\pi^-K^+K^-)$	109	2	0	2	4
3	$\Upsilon(4S) \to B^+B^-, B^+ \to e^+\nu_e\bar{D}^{*0}, B^- \to \pi^0\pi^-D^+D_s^-, \bar{D}^{*0} \to \pi^0\bar{D}^0, D^+ \to \pi^+\pi^+\pi^-\eta,$ $D_s^- \to \pi^-\eta', \bar{D}^0 \to \pi^0\pi^-K^+, \eta \to \gamma\gamma, \eta' \to \rho^0\gamma, \rho^0 \to \pi^+\pi^-$ $(\Upsilon(4S) \dashrightarrow e^+\nu_e\pi^0\pi^0\pi^0\pi^+\pi^+\pi^+\pi^-\pi^-\pi^-\pi^-\pi^-K^+\gamma\gamma\gamma)$	144	2	0	2	6

Table 2: Cascade decay branches of B^- (only the first four hierarchies are involved).

rowNo	cascade decay branch of B^-	iCascDcyBrP	nCase	nCcCase	nAllCase	nCCase
1	$B^- \to e^-\bar{\nu}_eD^{*0}, D^{*0} \to \pi^0D^0, D^0 \to \pi^+K^-$	15	31	26	57	57
2	$B^- \to e^-\bar{\nu}_eD^{*0}, D^{*0} \to D^0\gamma, D^0 \to \pi^+K^-$	0	13	21	34	91
3	$B^- \to e^-\bar{\nu}_eD^0, D^0 \to \pi^+K^-$	24	11	19	30	121
4	$B^- \to e^-\bar{\nu}_eD^{*0}, D^{*0} \to \pi^0D^0, D^0 \to \pi^0\pi^+K^-$	18	2	4	6	127
5	$B^- \to e^-\bar{\nu}_e\pi^0D^0, D^0 \to \pi^+K^-$	32	2	2	4	131

Table 5: Exclusive components of $B^- \to e^-\bar{\nu}_e + anything$.

rowNo	exclusive component of $B^- \to e^-\bar{\nu}_e + anything$	iDcyBrIncDcyBr	nCase	nCcCase	nAllCase	nCCase
1	$B^- \to e^-\bar{\nu}_eD^{*0}$	0	56	56	112	112
2	$B^- \to e^-\bar{\nu}_eD^0$	2	15	25	40	152
3	$B^- \to e^-\bar{\nu}_eD_1^0$	8	6	4	10	162
4	$B^- \to e^-\bar{\nu}_eD_0^{*0}$	6	3	2	5	167
5	$B^- \to e^-\bar{\nu}_e\eta D^0$	3	1	3	4	171

Table 15: Signal cascade decay branches.

rowNo	signal cascade decay branch	iSigCascDcyBr	nCase	nCcCase	nAllCase	nCCase
1	$B^- \to \tau^-\bar{\nu}_\tau D^0, \tau^- \to e^-\bar{\nu}_e\nu_\tau$	0	4	2	6	6
2	$B^- \to \tau^-\bar{\nu}_\tau D^{*0}, \tau^- \to e^-\bar{\nu}_e\nu_\tau, D^{*0} \to \pi^0D^0$	1	0	0	0	6
3	$B^- \to \tau^-\bar{\nu}_\tau D^{*0}, \tau^- \to e^-\bar{\nu}_e\nu_\tau, D^{*0} \to D^0\gamma$	2	0	1	1	7

GenMCTagTool

- Generalization of TauDecayMarker
- Analyzes array of MC particles to determine decay mode
- Based on a generic .dec file
  - Index labels fixed (found in user manual)
  - Missing "newly added" modes have VERY small contribution, see here
- Contains:

825 $B^{\pm}$ Modes	84 $D^{\pm}$ Modes
1000 B^{0} Modes	136 D^{0} Modes
3 $D^{*\pm}$ Modes	40 $\tau^{\pm}$ Modes
84 $D_s^{\pm}$ Modes	

- [Warning]: only links to first occurrence of particle

genUpsilon4S

Basf2 variable

`genUpsilon4S(variable)`

[Eventbased] Returns the `variable` evaluated for the generator-level $\Upsilon(4S)$. If no generator level $\Upsilon(4S)$ exists for the event, NaN will be returned.

E.g. `genUpsilon4S(p)` returns the total momentum of the $\Upsilon(4S)$ in a generic decay. `genUpsilon4S(mcDaughter(1, p))` returns the total momentum of the second daughter of the generator-level $\Upsilon(4S)$ (i.e. the momentum of the second B meson in a generic decay).

- Extract exact generated identification, kinematic, etc. information at event-level

genUpsilon4S

MC PDG of first B: **genUpsilon4S(mcDaughter(0, mcPDG))**

- B info:

Generic MC (decay.dec)

$\Upsilon(4S)$ → **0th daughter** 50%-50% split between generic B^+/B^- or $B^0/\bar{B}^0$ decays + **1st daughter** 50%-50% split between generic B^+/B^- or $B^0/\bar{B}^0$ decays

Dedicated MC

$\Upsilon(4S)$ → **0th daughter** 100% decay specified in dedicated decay.dec e.g. 100% B^-, 50% B^0 + 50% $\bar{B}^0$ + **1st daughter** 100% generic decay of CC e.g. 100% B^+, 50% $\bar{B}^0$ + 50% B^0

- B daughter info: indices follow order specified in .dec file

Guided Exploration

Let’s Play a Game…

Our signal mode is $B^0 \to [D^{*-} \to [\overline{D}^0 \to K^+\pi^-, K^+K^-]\,\pi^-]\,\pi^+$, define your signal, investigate the sample composition, and compare methods. To do so, discuss with your team, use your resources, and play with different approaches.

At the end, some of you may be selected to describe your technique, why you choose this technique and what you observe.

Expert Options

1. Comparison of MC Matching and Loose MC Matching

2. Create your own custom python module to extract the generated topologies for the event
    - How would you change this to be based on a reconstructed particle list object?

Summary

	Pro	Con
isSignal	Easy, Clean	Limited Acceptance
mcPDG	Easy, Clean	Related to specific reconstructed particle object
mcErrors	Informative, shows were matching goes wrong	Does not show physics process generated
genUpsilon4S	Custom access to generated processes	Slightly complex extraction / analysis
GenMCTagTool	Easy access to generated processes	Only links to first occurrence of particle of interest, limited to first order decay
TopoAna	Clean customizable access to information on generated processes, clean pdf output for analysis of signal and sample composition	Requires a little understanding of TopoAna, can be problematic with messy events, requires more variables and computation, not linked to basf2
Custom module	Custom access to generated processes	Requires a little understanding of the backend of how things are stored / access, prone to error

Reconstructed truth

Event truth