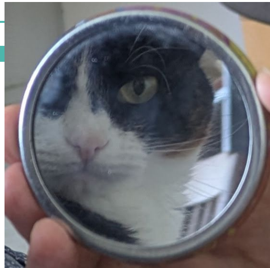
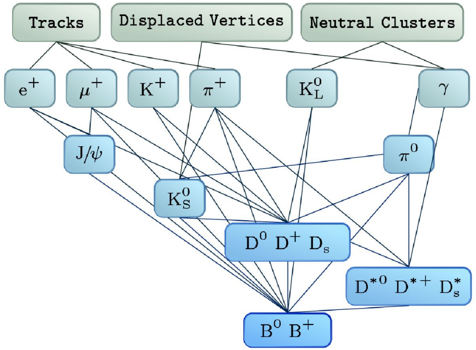
Inclusive tagging

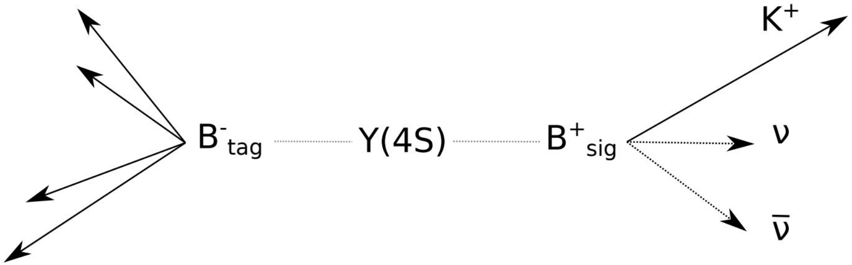
S. Glazov
Belle II Physics Week



Tagging methods at Υ(4S) factory

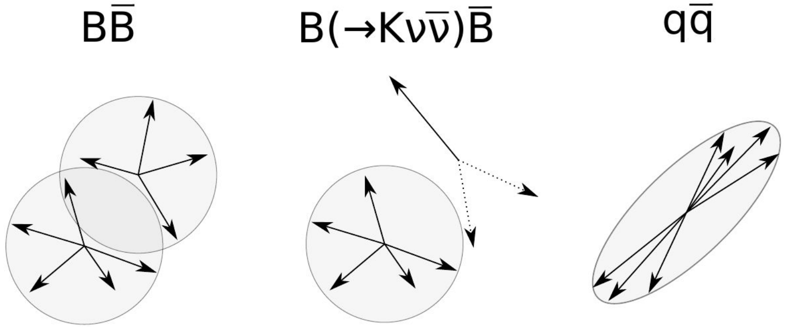
Hadronic tagging	Semileptonic tagging	Inclusive tagging
ε <1%	ε ~ 1%	ε > 1%
highest purity	High purity	Lowest purity
Best Kinematic constraints	Partial constraints	Beam constraints only

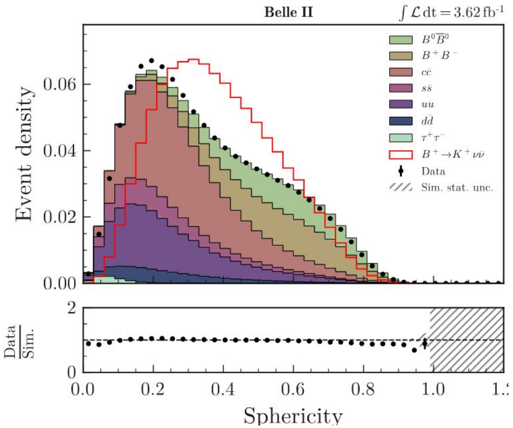


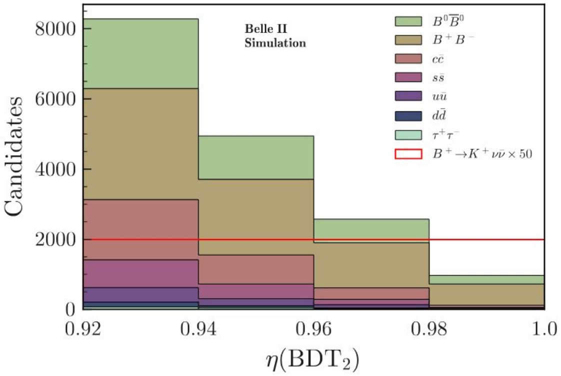


- The partner B-meson ("tag") is an important part of data analysis at Υ(4S) factory
- It can be reconstructed explicitly in one of the hadronic or semileptonic decay modes or its properties can be used inclusively
- The approaches differ in efficiency, purity, and kinematic constraining power

Inclusive tag in a nutshell



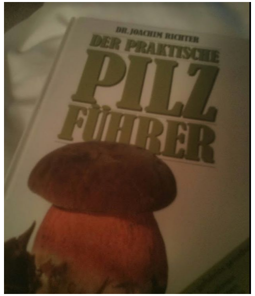




- Use event properties to suppress background with multiple variables combined by a classifier (“**BDT**”). Optimize specifically for channel of interest
- Use classifier output as (one of) the fit variable(s), use **simulation** for signal and background templates
- Use multiple control channels to validate simulation with data

Analysis flow for inclusive tagging

Object definition, candidate(s) selection, variable reconstruction, basic event selection → "BDT_1" basic background rejection with high signal efficiency ("skim") → "BDT_2" optimal training for high purity signal region → PL-Fit



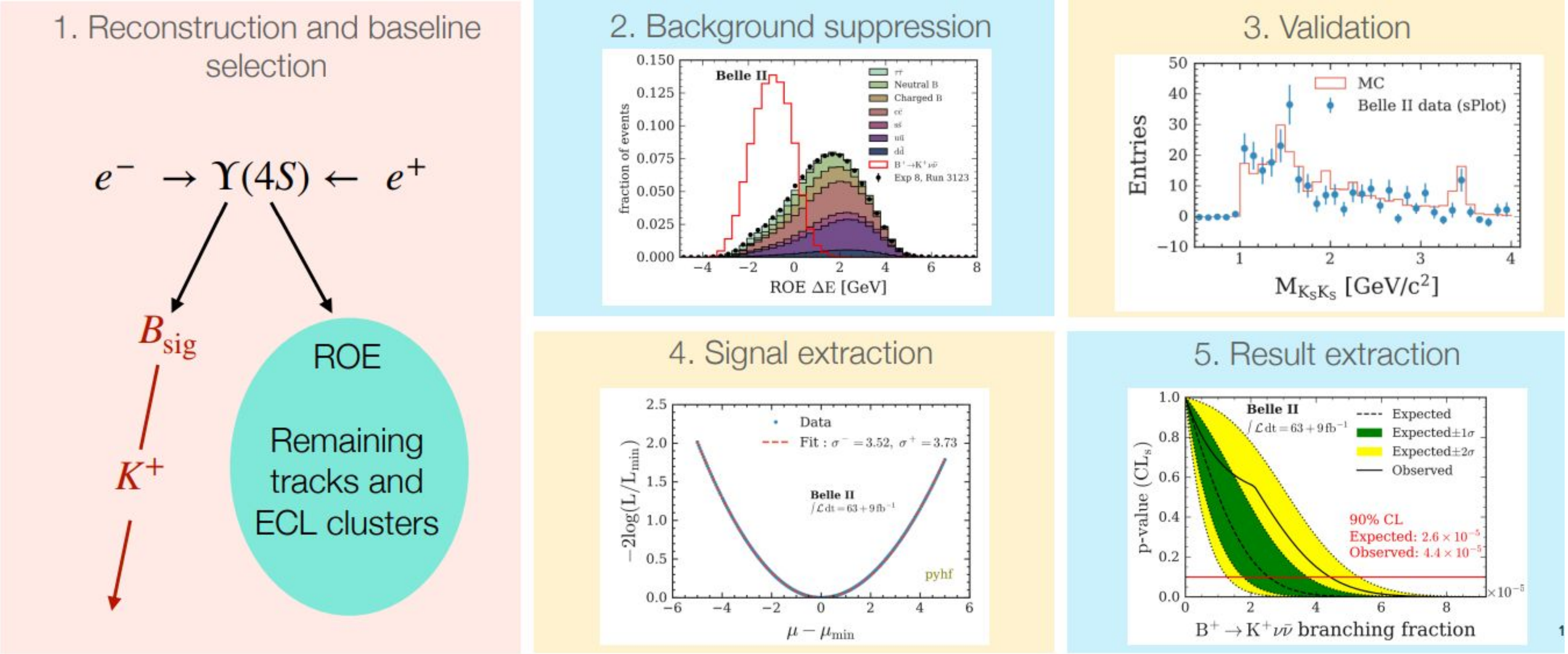
- Start from candidate(s) selection, not from tag: natural flow for a search
- Data volume is large: dedicated skim with **BDT_1**
- **BDT_2** boosts the training for the most interesting signal region
- Sample-composition profile-likelihood fit to extract the signal

Main challenges of inclusive tagging

- Candidate selection for complex final states is a substantial combinatorial problem
- Data volume is significantly increased, efficient analysis framework is essential
- Low purity signal region requires accurate simulation – validation with control samples is essential
- Profile likelihood fit with large amount of nuisance parameters to describe systematic uncertainties can be tricky

Phys.Rev.D 109 (2024) 11, 112006

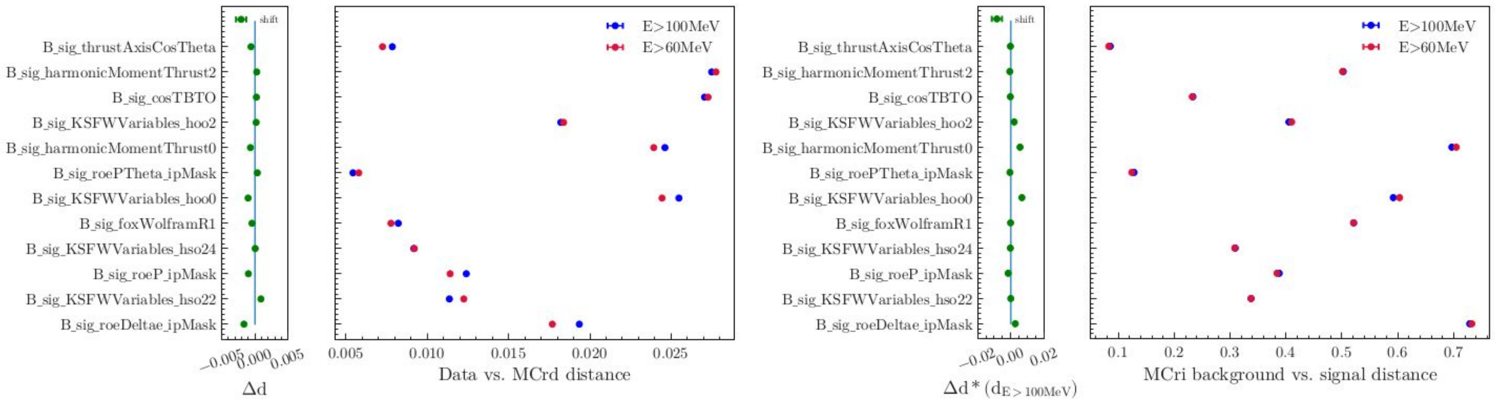
Analysis overview



$B^+ \rightarrow K^+ \nu\bar{\nu}$ analysis is used as the main example to illustrate the method

Candidate(s) Selection

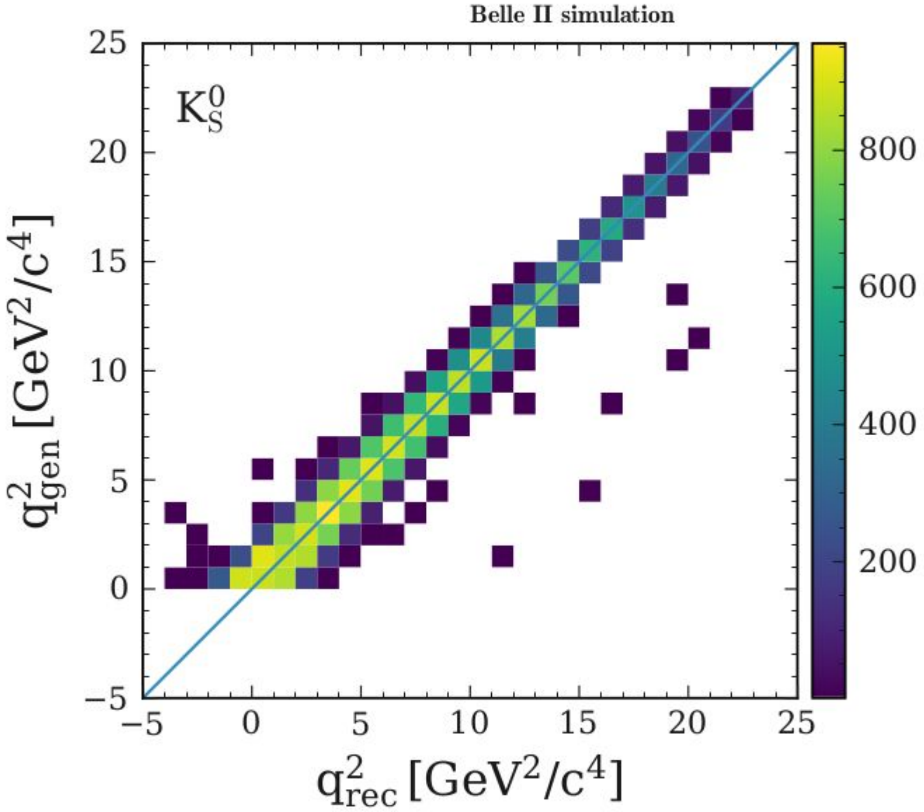
Reconstruction of particles



→MC should look like data while signal should be more separated from background

- Inclusive tagging uses consistent selection for the signal and ROE: perform basic optimization
- Prefer moderately conservative selection: smaller systematic uncertainties
- Build event/ROE from your own lists: simplifies propagation of systematic uncertainty. Remember that e.g. tracking efficiency affects both signal and ROE
- Consider using composite objects for the signal when building event properties: can help to unify e.g. $B^+ \to K^+ \nu \bar{\nu}, B^+ \to K^{*+} \nu \bar{\nu}$ analyses

Reconstruction of q^2_{rec}



$$q^2_{\text{rec}} = s/(4c^4) + M^2_{K^{(*)}} - \sqrt{s}E^*_{K^{(*)}}/c^4$$

- Reconstruct q^2_{rec} as a recoil mass using candidate momentum in CME
- Resolution is much worse compared to hadronic tagging, but sufficient to distinguish main background features

Signal candidate selection

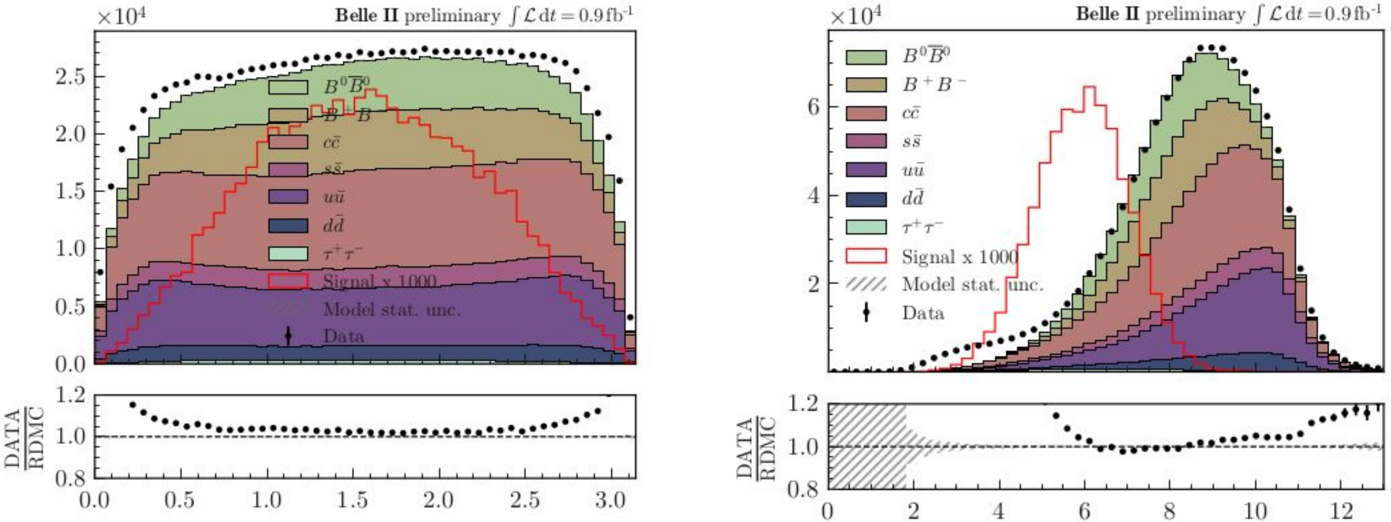
MVA based on $M(K^*), K^*$ momentum in CMS, and vertex - to IP distance

$B^0 \to K^{*0}\nu\bar{\nu}$

cuts	signal efficiency [%]	avg. signal multiplicity	avg. background multiplicity
K^{*0} Mass	39.8	5.2	10.5
K^+ PXD hits and PID	30.2	1.8	2.4
π^- PXD hits and PID	26.2	1.5	1.9
MVA-two-candidate selection	26.1	1.3	1.5

- For wide resonances, there is a large combinatorial background
- Consider applying required cuts early on (e.g. PID)
- Dedicated MVA for signal selection may help
- Keep sufficient amount of candidates to maintain high signal efficiency
- Perform final candidate selection after final selection.

Removal of low multiplicity (γγ) events



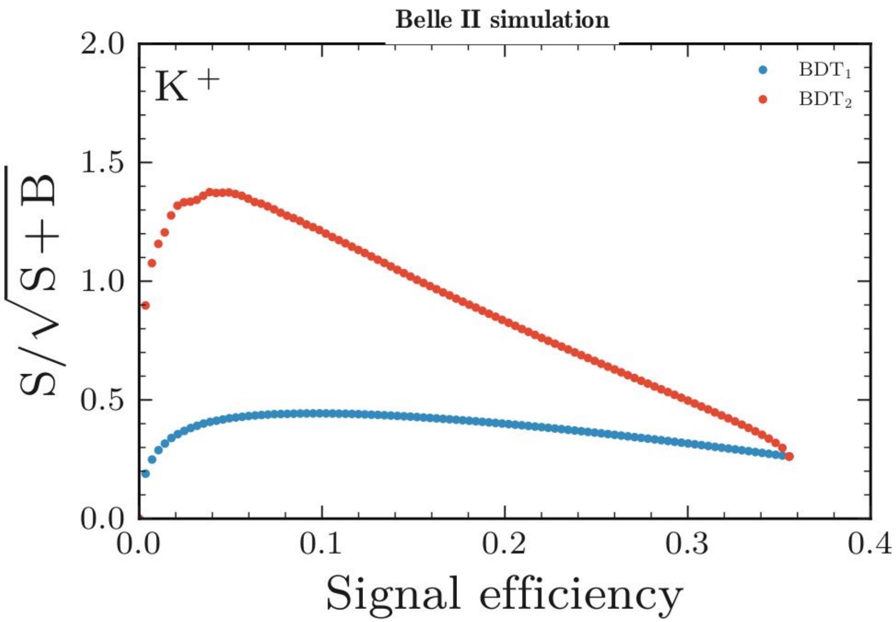
Missing momentum ϑ

Visible energy, GeV

- Belle II simulation does not contain gamma-gamma processes for more complex hadronic final states, e.g. $\gamma\gamma \rightarrow K^* K^*$
- Empirical removal by missing momentum direction and total energy cuts
- Fix residual background by using off-resonance data to tune MC

BDT_1 Skim

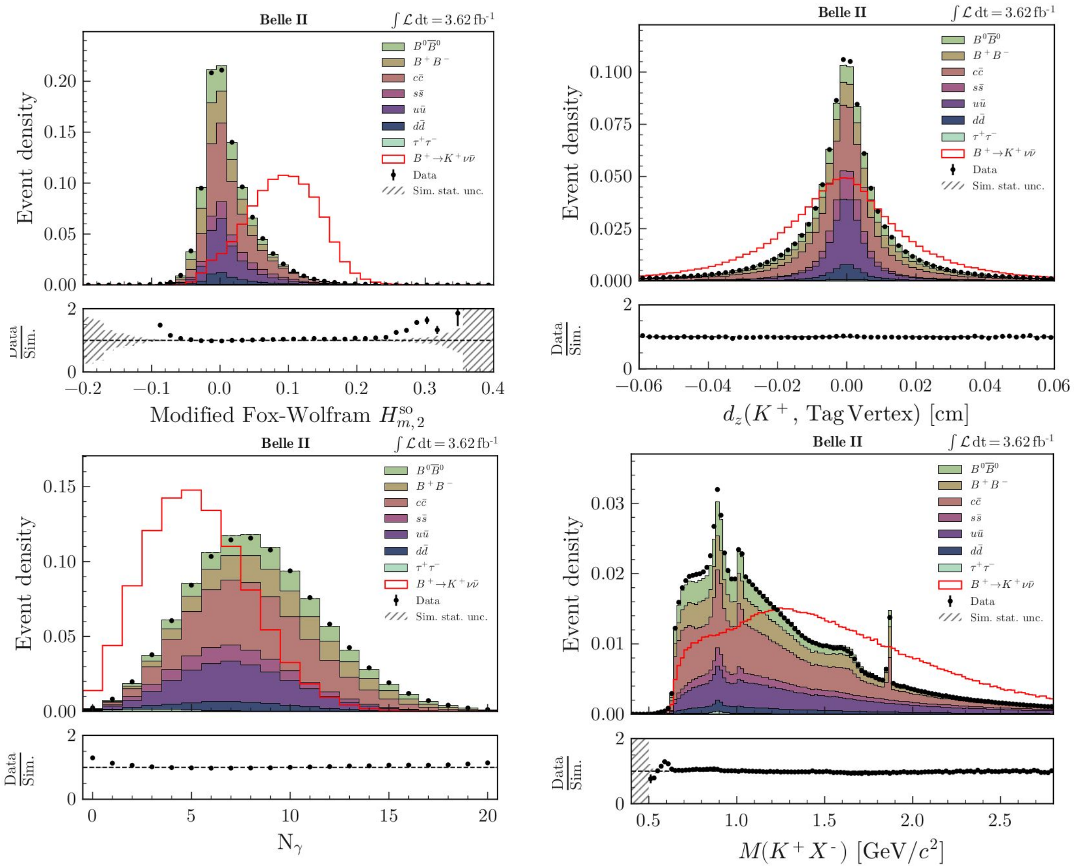
Features of the BDT_1



- Main goal: keep signal efficiency high while removing the bulk of the background
- Use simple event shape/ROE variables: can be the same set for several channels
- Typically, training does not require large MC samples (few fb^{-1})
- Can be made official WG skim

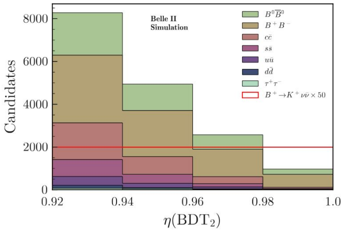
BDT_2 and Signal Region

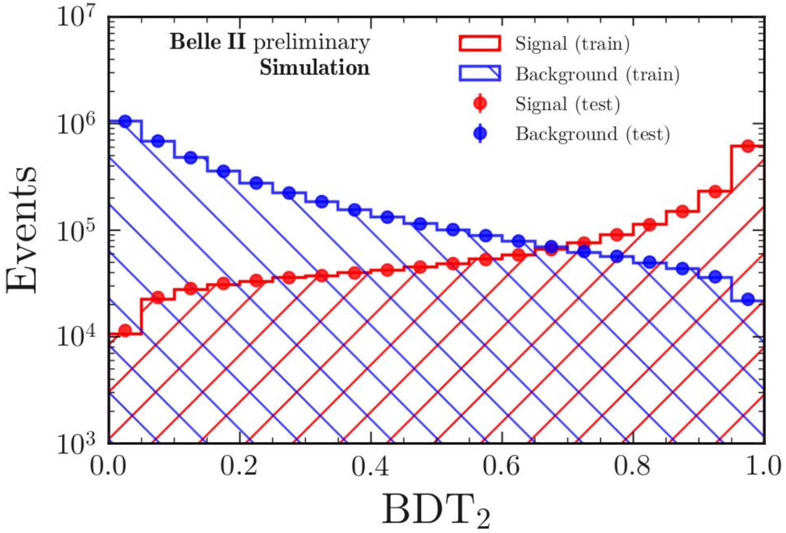
Selection of input variables for BDT$_2$

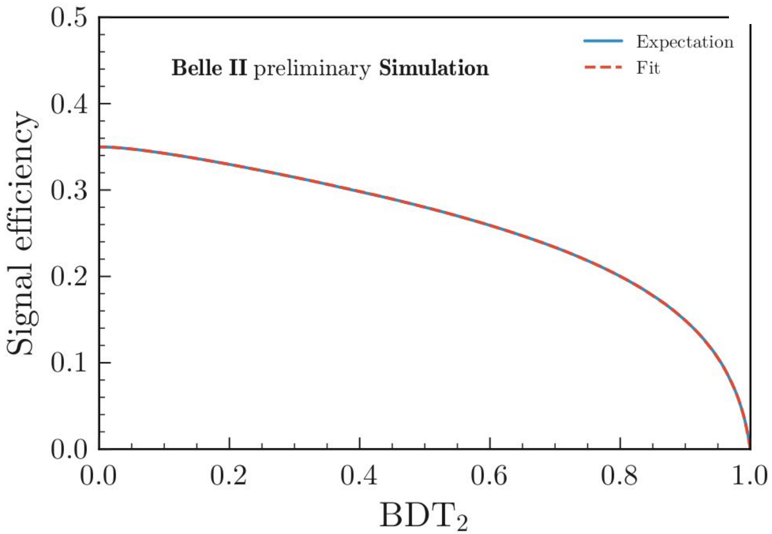


- Use variables with large discriminating power, which are well described by simulation. Residual differences should be covered by systematics
- If the sample composition fit uses additional to BDT$_2$ variable(s), seek for low correlation with them
- Investigate background in the signal region to add dedicated variables

BDT_2 training

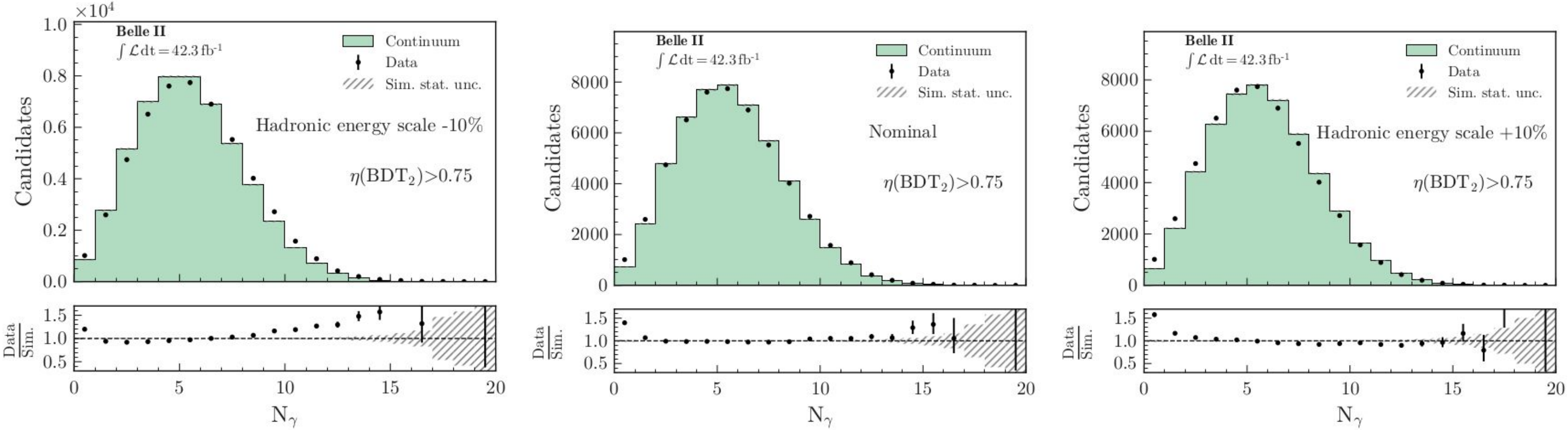






- BDT_2 training is standard: balance signal vs background, optimize FOM (e.g. $S/\sqrt{S+B}$), tune hyperparameters (use e.g. optuna, but check the result)
- Optionally: transform BDT_2 into signal efficiency (better for interpretation, specific sideband studies)

Propagation of detector systematics

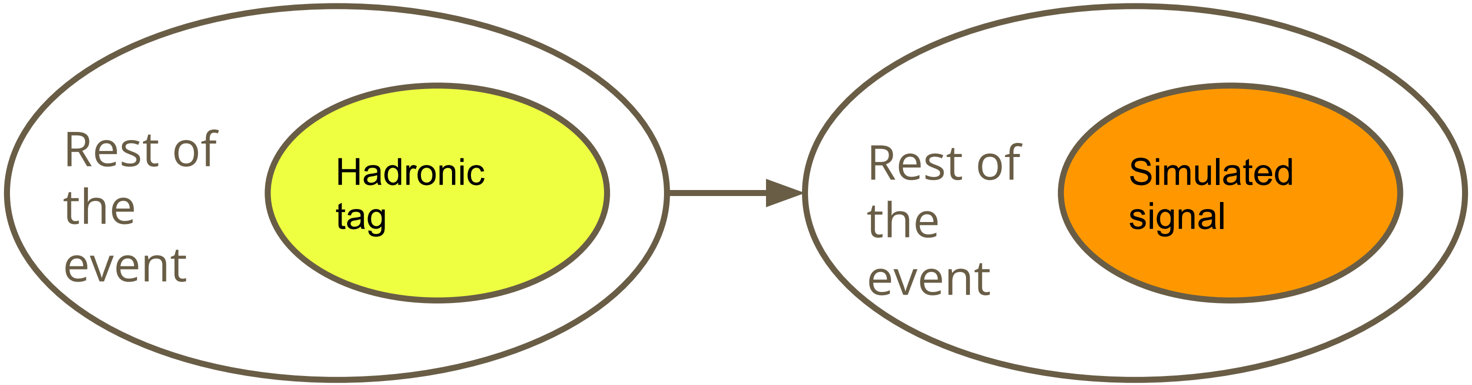


- Detector systematic effects (track efficiency, photon energy, unmatched photon (“hadronic”) energy, K_L efficiency) affect both signal and ROE variables
- Propagate in the analysis by varying them and repeating ntuple production
- Use KDE when comparing signal region templates to avoid double counting of statistical uncertainties

Signal efficiency validation

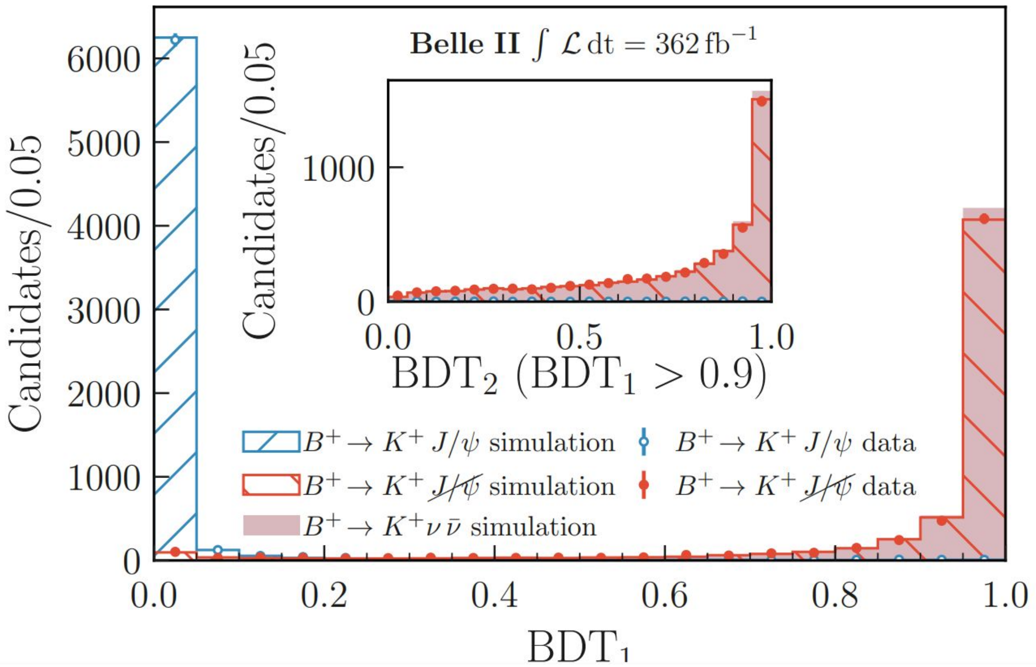
https://software.belle2.org/light-2507-europa/sphinx/analysis/doc/embedding.html#signal-embedding

Signal embedding for signal simulation validation



- Identify B decay by a clean hadronic tag (e.g. $B^+ \to J/\psi K^+$)
- Remove the hadronic tag from the event
- Insert the signal decay instead
- Do the same operation for both data and MC simulation

BDT efficiency validation with embedding



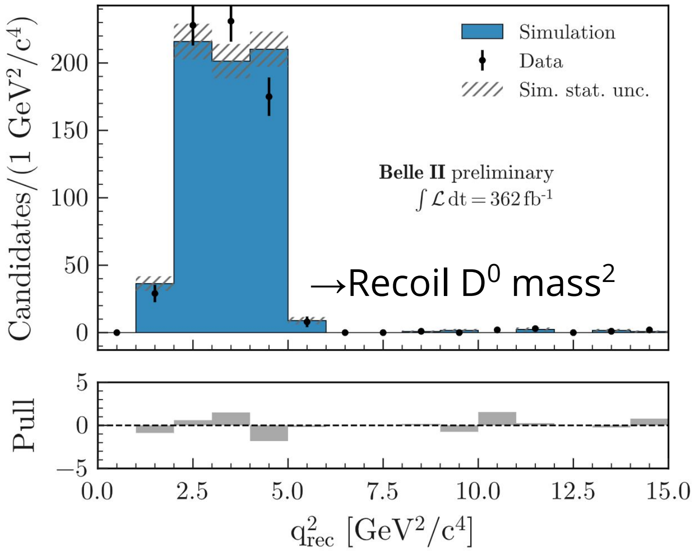
Embedded MC ~ signal: representative

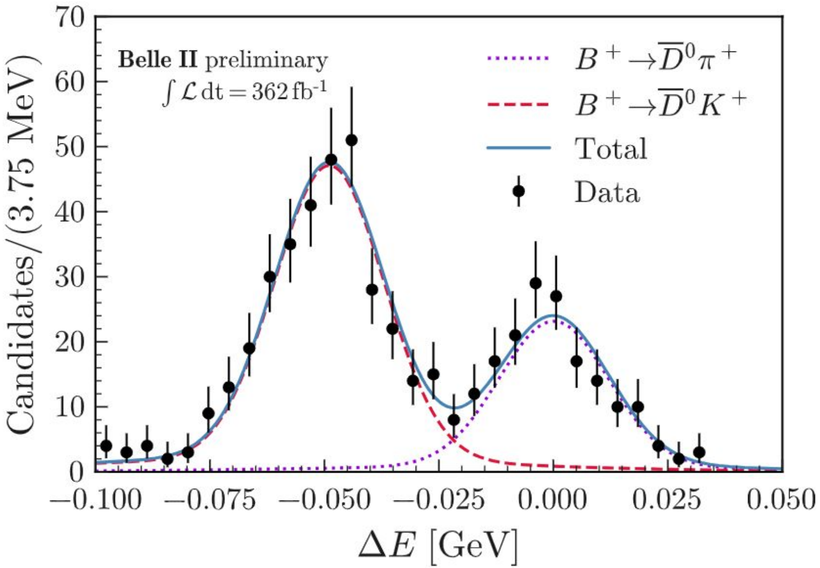
Embedded data vs MC: efficiency check

- Embedding validates ROE modeling for signal topologies
- Variables related to signalB-ROE correlation are validated too
- Signal side is always from simulation: it is not validated. Instead:
  - Use standard performance recommendations
  - dedicated control channels for the signal side, if selection beyond recommendations (e.g. extra checks of $K^{*0} \to K^+\pi^-$ vertex fit)
  - Physics modeling systematics: vary signal form factors

Signal side validation: kaon identification check

$B^{+} \to h^{+} \bar{D}^{0} (\to K^{+}\pi^{-})$ with $K^{+}\pi^{-}$ removed

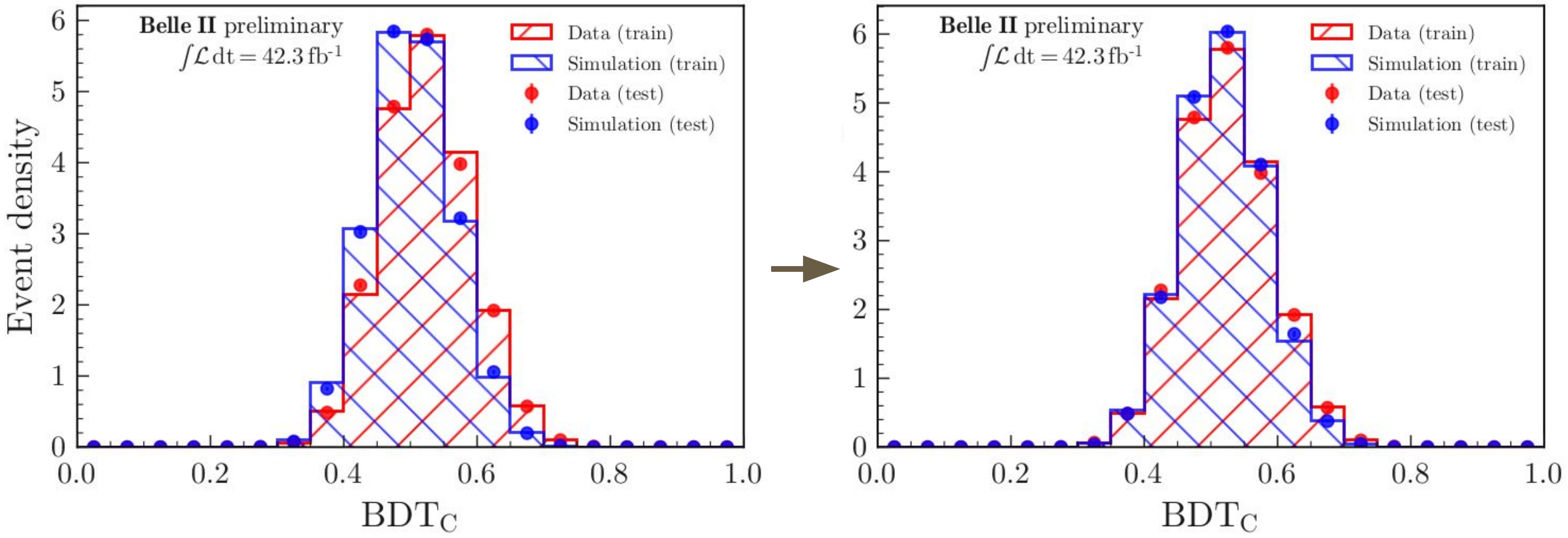




- Kaon identification efficiency/pion fakes: from systematics framework
- Check using $B^{+} \to h^{+} \bar{D}^{0} (\to K^{+}\pi^{-})$ decays. Remove $\bar{D}^{0}$ decay products to mimic the signal, select signal region. Use $\Delta E = E^{*}_{B} - E_{beam}$ distribution to separate $h^{+} = \pi^{+}$ vs K^{+}.

Background Validation

Continuum simulation tuning and validation

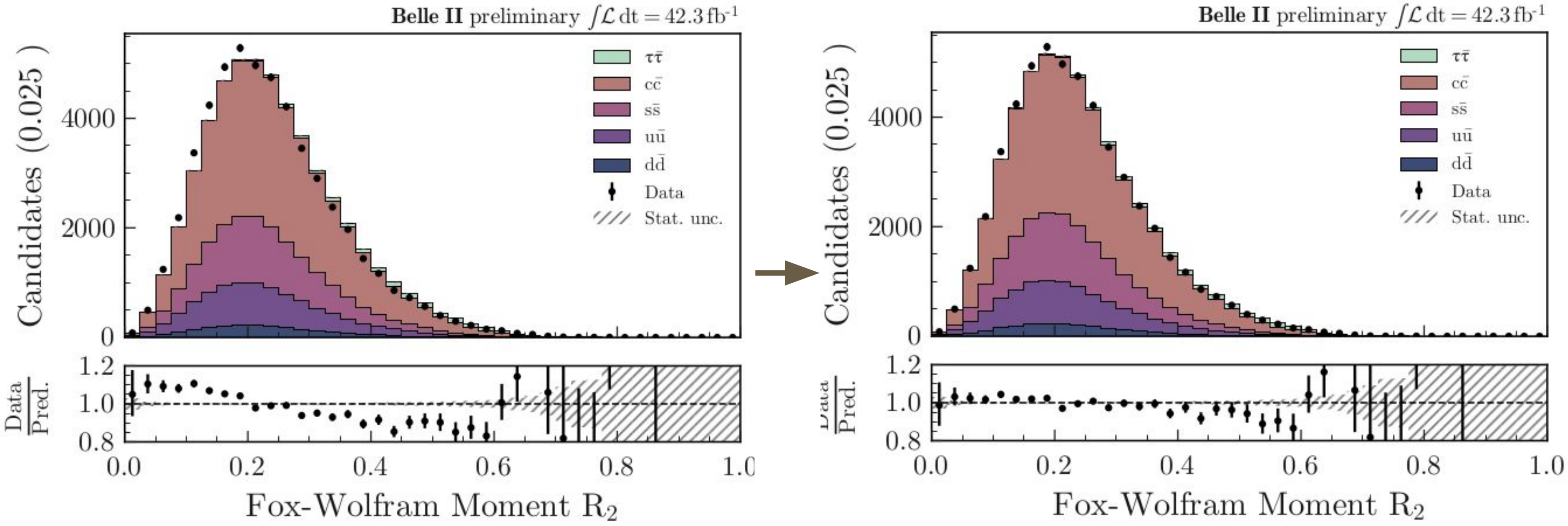


Use off-resonance data sample. Train $\mathrm{BDT_c}$ classifier to distinguish **data from simulation**.
Apply ***w=p/(1-p)*** reweighting where ***p*** is the $\mathrm{BDT_c}$ output.

→Classifier uses the same variables as $\mathrm{BDT_2}$: analysis specific tune

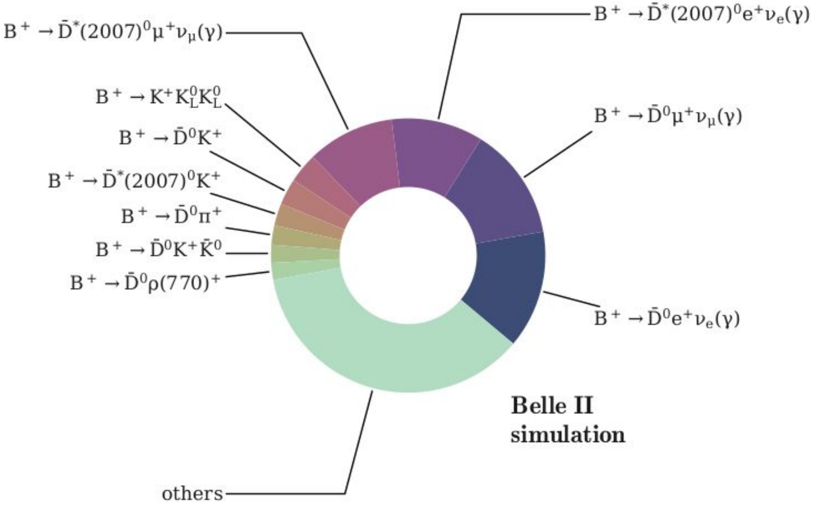
→Given small data sample, make sure that BDTc is not overtrained, use simpler model

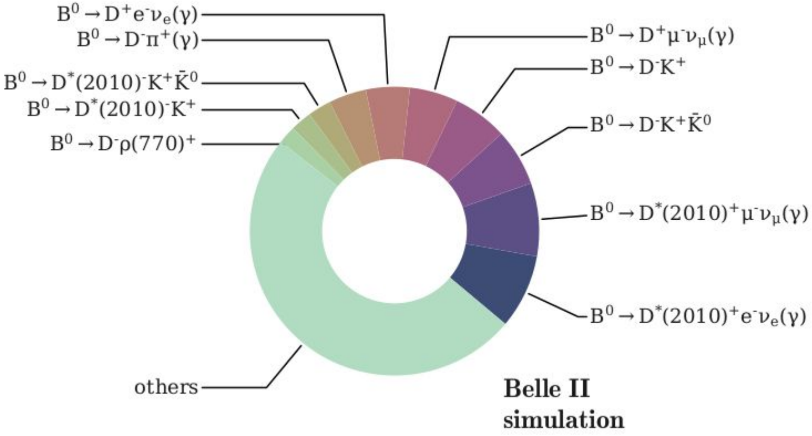
Continuum simulation tuning and validation



→BDT_c weights improve agreement. Include them as **shape uncertainty** with **100%** uncertainty (relative): it is a direction of maximal data-to-simulation disagreement for BDT_2 variables, useful to keep free in the profile likelihood fit.

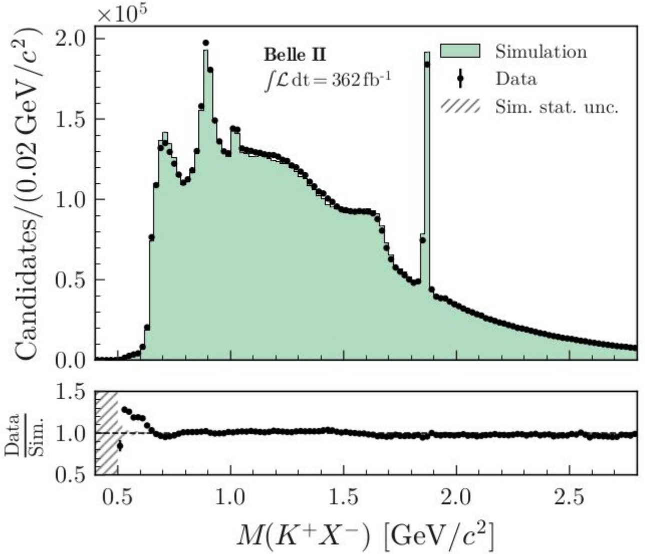
Background composition studies



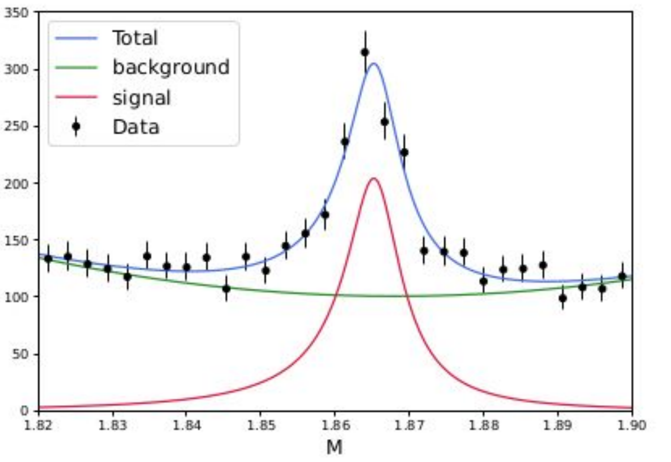


- Use your favorite tool (e.g. topoAna) to study background composition in the signal region; check fractions **and** BDT_2 distribution.
- For analyses with missing energy, semileptonic and decays containing K_L and neutrons are most important
- Check ***D*** decays too (with K_L in particular)

Example background study: B →D →K

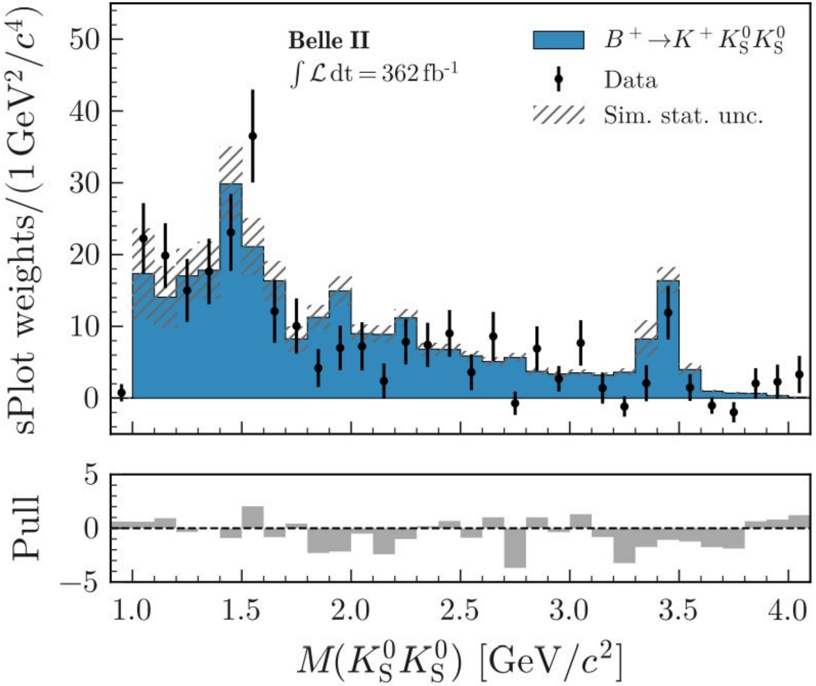


Simulation, $1ab^{-1}$, SR



- Substantial background from K produced in (low multiplicity) ***D*** decays
- Can be suppressed by explicitly reconstructing ***M(K⁺X)*** (BDT_2 variable)
- BDT_2 optimization leaves some events in the SR: can be used to study residual background (note: non-trivial background shape)

Example background study: $B^+ \to K^+ K_S K_S$



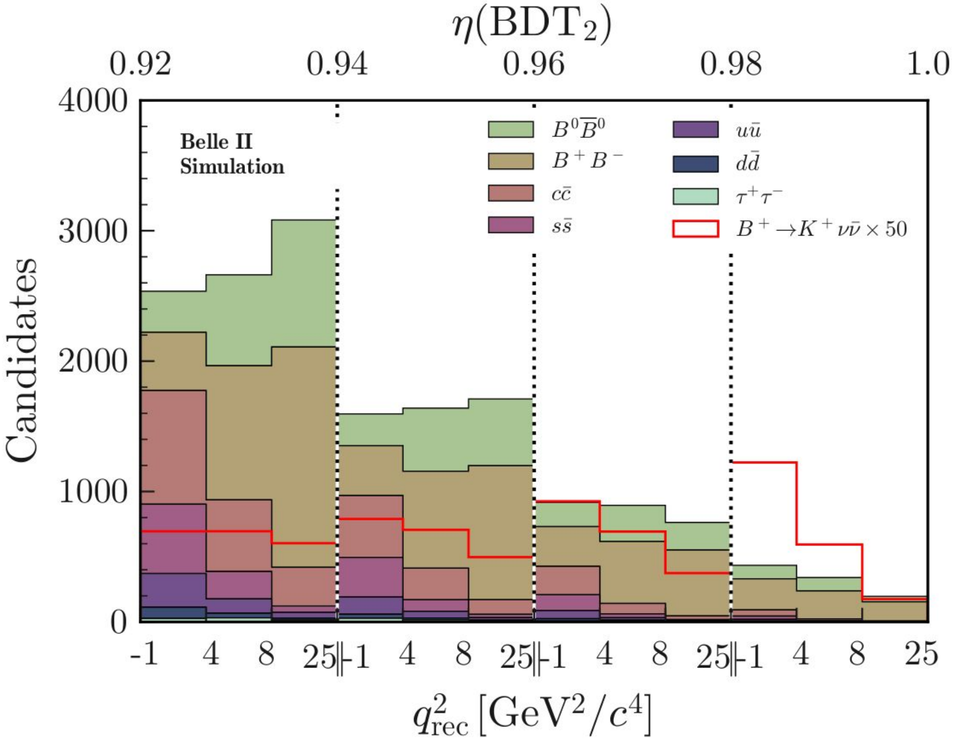
- Backgrounds with K_L in the final state are among most dangerous
- K_L interaction probability in ECL is about 50% and energy deposit is low

  →Dedicated study of $B^+ \to K^+ K_L K_L$ background

- Fully reconstruct $B^+ \to K^+ K_S K_S$ decay. Suppress background while keeping $M(K_S K_S)$ efficiency flat.
- Fit ΔE distribution, use sPlot to extract $M(K_S K_S)$ distribution, compare data to simulation

Signal extraction

Sample composition fit



- Template simulation based fit to binned data counts
- Systematic uncertainties are propagated using nuisance parameters
- Several standard tools: **pyhf** (+ cabinetry), HistFactory, Combine with minimization backends
- Gaussian approximation is often sufficient (χ^2 fit)
- Fast tools: **sghf** (compatible with pyhf input)

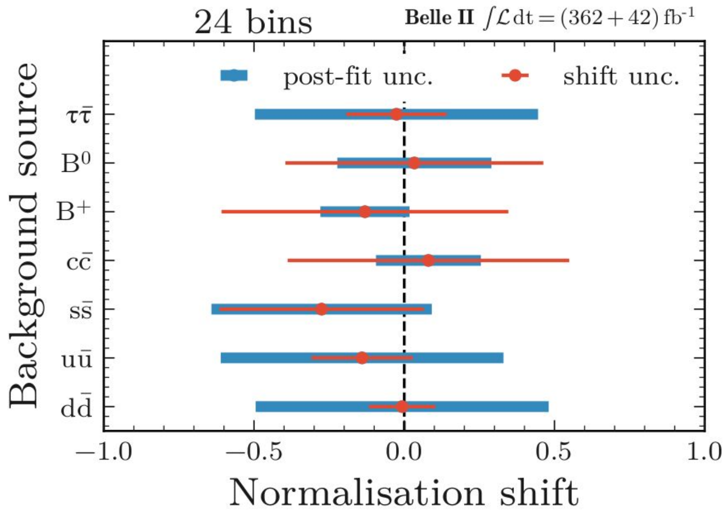
Modeling systematic uncertainties

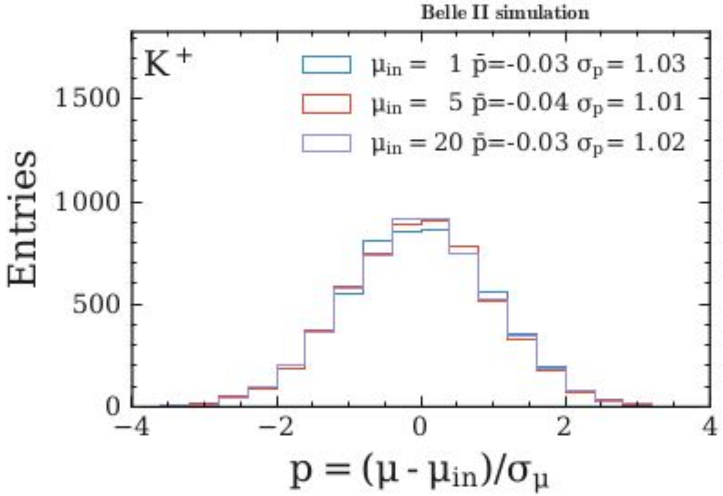
Source	Correction	Uncertainty type, parameters	Uncertainty size	Impact on σ_μ
Normalization of $B\overline{B}$ background		Global, 2	50%	0.90
Normalization of continuum background		Global, 5	50%	0.10
Leading B-decay branching fractions		Shape, 6	$O(1\%)$	0.22
Branching fraction for $B^+ \to K^+ K^0_L K^0_L$	q^2 dependent $O(100\%)$	Shape, 1	20%	0.49
p-wave component for $B^+ \to K^+ K^0_S K^0_L$	q^2 dependent $O(100\%)$	Shape, 1	30%	0.02
Branching fraction for $B \to D^{**}$		Shape, 1	50%	0.42
Branching fraction for $B^+ \to K^+ n\bar{n}$	q^2 dependent $O(100\%)$	Shape, 1	100%	0.20
Branching fraction for $D \to K^0_L X$	+30%	Shape, 1	10%	0.14
Continuum-background modeling, BDT$_c$	Multivariate $O(10\%)$	Shape, 1	100% of correction	0.01
Integrated luminosity		Global, 1	1%	< 0.01
Number of $B\overline{B}$		Global, 1	1.5%	0.02
Off-resonance sample normalization		Global, 1	5%	0.05
Track-finding efficiency		Shape, 1	0.3%	0.20
Signal-kaon PID	p, θ dependent $O(10\text{–}100\%)$	Shape, 7	$O(1\%)$	0.07
Photon energy		Shape, 1	0.5%	0.08
Hadronic energy	−10%	Shape, 1	10%	0.37
K^0_L efficiency in ECL	−17%	Shape, 1	8.5%	0.22
Signal SM form factors	q^2 dependent $O(1\%)$	Shape, 3	$O(1\%)$	0.02
Global signal efficiency		Global, 1	3%	0.03
Simulated-sample size		Shape, 156	$O(1\%)$	0.52

- Each Systematic source is described by one or several nuisance parameters
- Detector, physics modeling systematics is correlated across channels (if several channels are analyzed)
- Analysis specific (e.g. BDT$_c$ - uncorrelated
- Check stability of the result vs correlation assumption

Monitoring fit performance

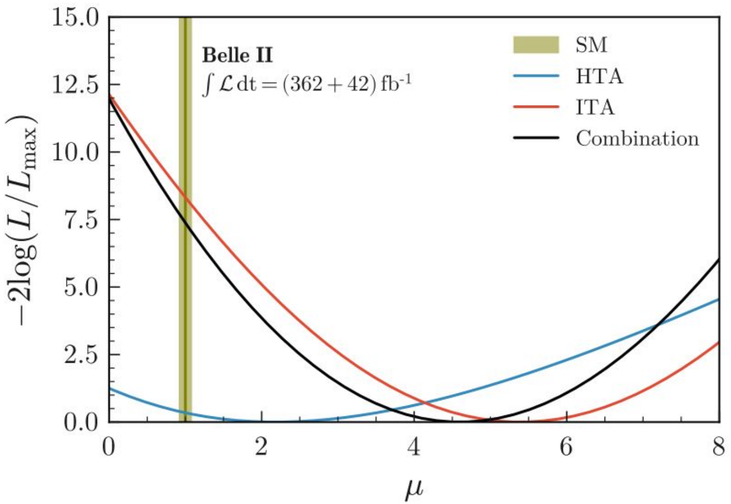
Minuit is not designed to handle > 200 parameters

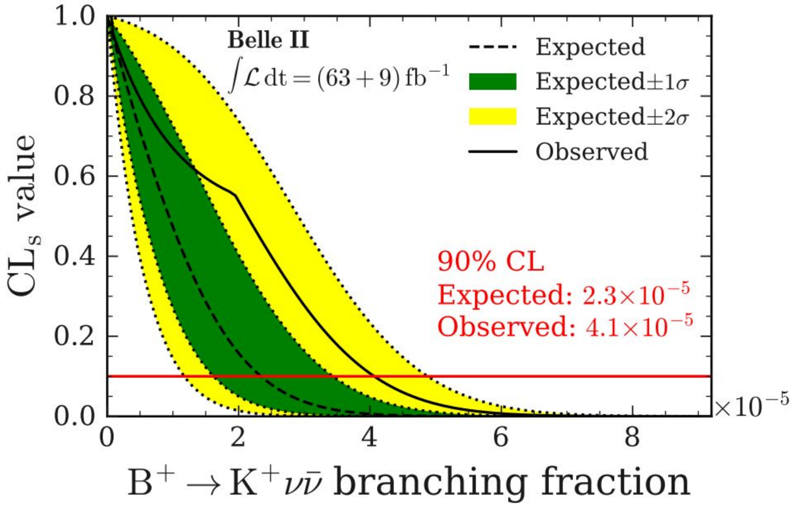




- Study shifts and pulls = shift/√(1-post-fit-error^2) of systematic uncertainties
- Toys to investigate bias/coverage
- Minuit issues: often useful to first find the minimum using scipy minimizer, and use minuit for error analysis (or perform likelihood scan by hand).

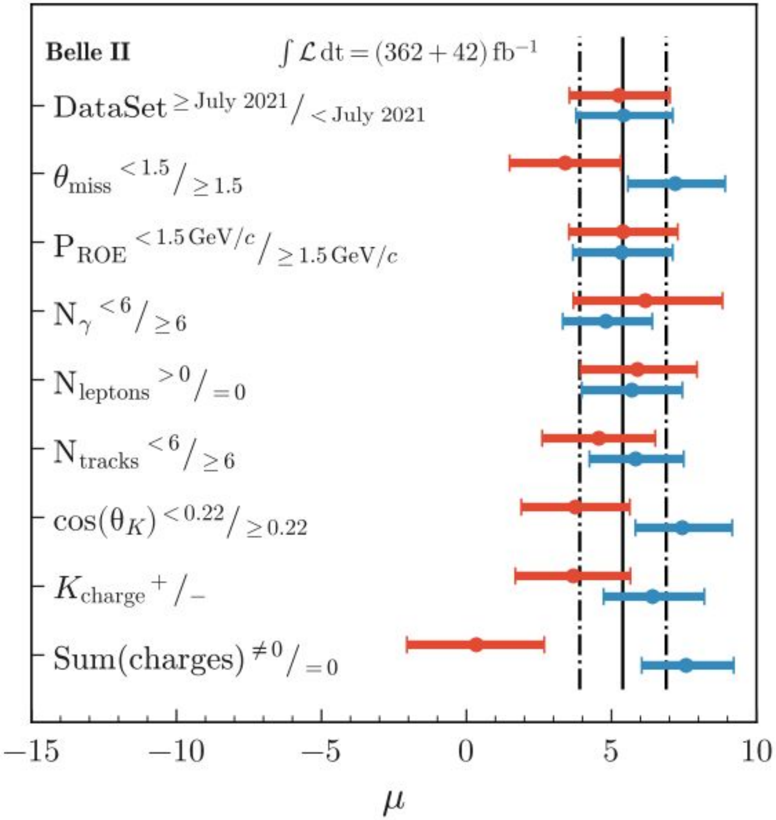
Result representation





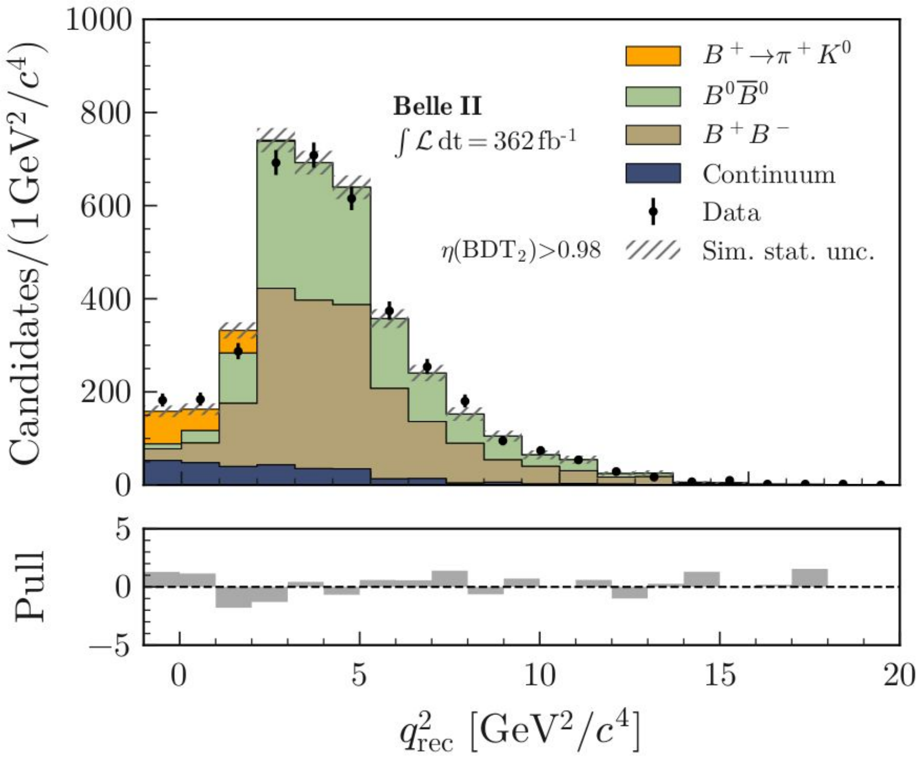
- Signal significance can be determined using log-likelihood scan
- Hadronic tag analysis (HTA) provides a cross check and improved accuracy
- If no significant signal is observed, set limits
- Likelihoods (and example code how to use them) should become public after paper is accepted

Consistency checks



- High statistics of inclusive tagging allows for powerful checks using split samples
- Identify relevant for analysis variables, find median value over the sample, split, and compare results
- Relevant tests for missing energy analysis: missing energy dependence, track/photon multiplicity
- Many tests means > 2sigma fluctuations are possible.

Control channel



- Demonstrate that machinery works using a control channel
- Channels with K_L for missing energy are good candidates
- Analysis of $B^+ \to \pi^+ K_L$:
  - Change PID from K+ to π+
  - Keep the rest of analysis unmodified
  - Tune binning, but keep fit as for the main analysis
  - Extract signal, compare to measured $B^+ \to \pi^+ K_S$
- $B(B+ \to \pi^+ K^0) = (2.5 \pm 0.5)\times 10^{-5}$ vs PDG $(2.39 \pm 0.06)\times 10^{-5}$

Summary

- Inclusive tagging is a powerful method to search for rare decays with missing energy
- All steps in the analysis can be optimized for the specific target
- The method requires accurate modeling by simulation which must be validated by data
- Systematic uncertainties for inclusive and other tagging methods are very different: combination of the approaches is an important cross check and can improve analysis sensitivity.