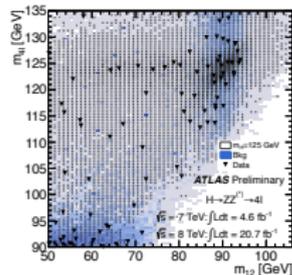
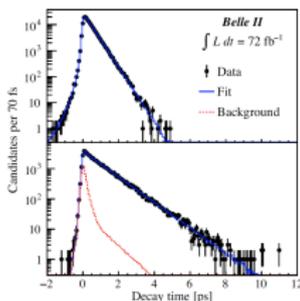


SYSTEMATIC UNCERTAINTIES: An introduction

Bruce Yabsley

School of Physics, University of Sydney

Belle II Physics Week, KEK, 2nd November 2023



Outline

- 1 Introduction: What are systematics? / A taxonomy
- 2 (uncertainties on) the INPUTS to the measurement
- 3 (uncertainties on) any AUXILIARY measurements
- 4 (uncertainties on) the CALIBRATION of the apparatus
- 5 Roundup of methods
- 6 Issues, questions, and tricks of the trade
- 7 Summary

What are systematics? (1) Old-fashioned experimentalist answer

Imagine the answer of Kyūzō-san (Skill Guy from *Seven Samurai*):

Counting uncertainties go as \sqrt{N} , and are reliable.

*Everything else is systematic uncertainty,
and is governed only by judgement and rules of thumb.*

I know my techniques and my mental bank of examples. So I know a systematic uncertainty when I see one, and use an appropriate estimate.

This new-fangled tool MINUIT also returns uncertainty estimates, and empirically it seems reliable when it works (I always check by reading the verbose output carefully). I perk its results in as “statistical uncertainties”, but only on sufferance.

What are systematics? (2) Time-constrained student answer

Imagine the answer of a student six months from submission:

My advisor / Working Group / RC has a standard list of systematics for this sort of analysis, and methods for estimating them.

I have never seen an analysis make it to journal without all of these systematic terms, and only occasionally see an extra term (and then only because a collaborator insisted during CWR).

I have looked at papers from past experiments, and these standard systematics are in most of them.

One seems to be new in Tom Browder's CLEO paper.

I asked him about it once at a Belle II party. He said that this systematic was actually brought to the Hawaiian islands by Maui, and handed down. I am not sure whether he was joking.

What are systematics? (3) New-fashioned experimentalist ...

Imagine the answer of an ATLAS/CMS analysis contact:

Most uncertainties are statistical uncertainties in disguise.

It is straightforward to include them in the fit using Gaussian constraint (or Log-Normal or other constraints if necessary, for extra credit).

Correlations do not scare me, as they can be handled by covariance matrices; complex effects can be modelled by toy Monte Carlo.

As for effects that are only present in the data, these can be accessed via bootstrapping.

My old thesis advisor still includes some systematic terms by hand, and even uses $\pm 1\sigma$ estimates sometimes. We still collaborate (she is a member of the 50-person team on our paper), and I cannot stop her from doing this, but it is seriously embarrassing when I have to show such estimates to the Editorial Board.

What are systematics? A taxonomy

My basic answer is that all three of these answers are correct as far as they go. But they have something to learn from each other.

To the student I would add:

There is a weak presumption in favour of all of the uncertainties on the standard list. Individually they are likely to be appropriate, but maybe not all of them. And some extra terms may be necessary.

But how do you know which uncertainties to drop, and which ones to add?

That is the purpose of this talk

My working taxonomy of systematics:

- 1 (uncertainties on) the INPUTS to the measurement
- 2 (uncertainties on) AUXILIARY measurements
- 3 (uncertainties on) the CALIBRATION of the apparatus

(uncertainties on) the INPUTS to the measurement

- 1 Introduction: What are systematics? / A taxonomy
- 2 (uncertainties on) the INPUTS to the measurement**
 - numbers with uncertainties
 - theoretical uncertainties
- 3 (uncertainties on) any AUXILIARY measurements
- 4 (uncertainties on) the CALIBRATION of the apparatus
- 5 Roundup of methods
- 6 Issues, questions, and tricks of the trade

INPUTS: numbers with uncertainties, e.g. \mathcal{B}

Belle pub632: PRD 107, 072008 (2023); $e^+e^- \rightarrow \Sigma\bar{\Sigma}$ via ISR

- $e^+e^- \rightarrow \Sigma\bar{\Sigma}$ measurement
 - ISR sample, relatively clean \rightarrow
 - (we will discuss the background estimation method later)
 - reconstruct $\Sigma \rightarrow \Lambda\gamma$
 - signal extraction includes event counts, efficiencies, ...
and the known $\mathcal{B}(\Lambda \rightarrow p\pi)$
- \rightarrow 0.8% uncertainty on the result

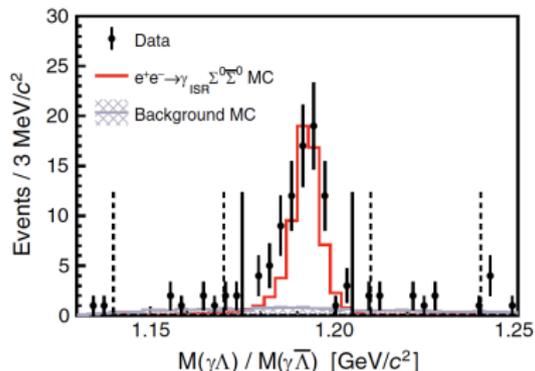
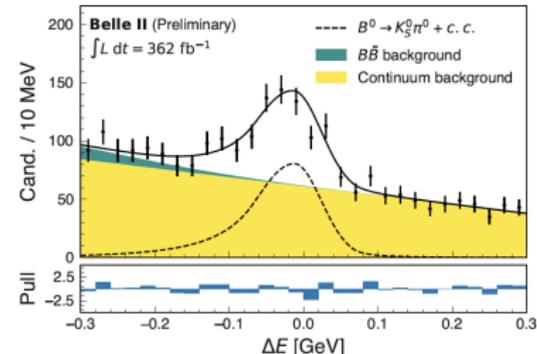
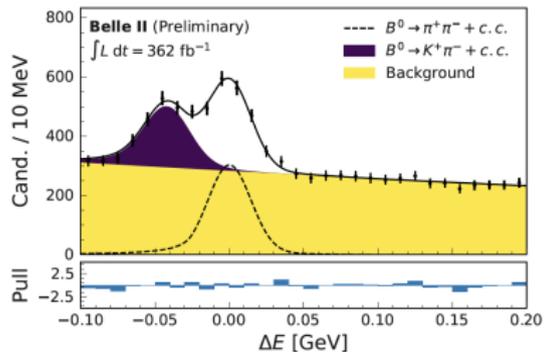
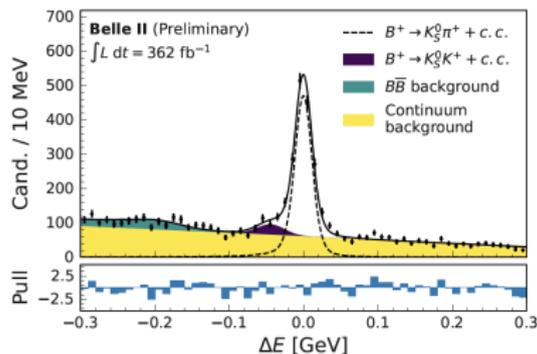
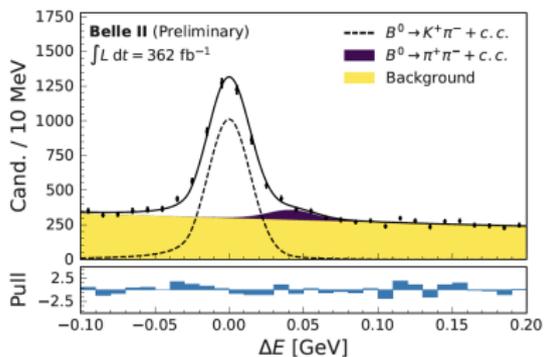


FIG. 2. The invariant mass of the accepted $\gamma\Lambda$ and $\gamma\bar{\Lambda}$ candidates. The points with error bars are experimental data and the red histogram shows the $e^+e^- \rightarrow \gamma_{\text{ISR}}\Sigma^0\bar{\Sigma}^0$ MC events with $\Sigma^0/\bar{\Sigma}^0$ correctly reconstructed. The hatched histogram is a mixture of events in $e^+e^- \rightarrow \gamma_{\text{ISR}}\Lambda\bar{\Lambda}$ and $e^+e^- \rightarrow \gamma_{\text{ISR}}\Sigma^0\bar{\Lambda}$ MC samples and the $e^+e^- \rightarrow \gamma_{\text{ISR}}\Sigma^0\bar{\Sigma}^0$ MC events with $\Sigma^0/\bar{\Sigma}^0$ misreconstructed. The solid and dashed vertical lines denote the $\Sigma^0/\bar{\Sigma}^0$ signal and sideband regions, respectively.

INPUTS: numbers with uncertainties, e.g. $N_{B\bar{B}}$, f^{00}

Belle II pub24: 2310.06381 → PRD; $B \rightarrow K\pi$, $\pi\pi$ BFs and \mathcal{A}_{CP}



INPUTS: numbers with uncertainties, e.g. $N_{B\bar{B}}$, f^{00}

Belle II pub24: 2310.06381 → PRD; $B \rightarrow K\pi$, $\pi\pi$ BFs and \mathcal{A}_{CP}

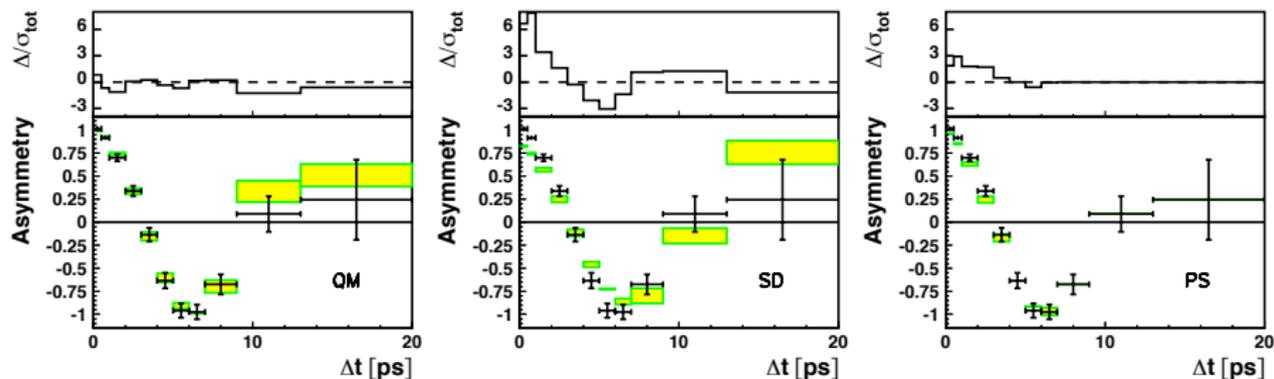
TABLE II. Summary of the relative systematic uncertainties (%) on the branching fractions.

Source	$B^0 \rightarrow K^+\pi^-$	$B^0 \rightarrow \pi^+\pi^-$	$B^+ \rightarrow K^+\pi^0$	$B^+ \rightarrow \pi^+\pi^0$	$B^+ \rightarrow K_S^0\pi^+$	$B^0 \rightarrow K_S^0\pi^0$
Tracking	0.5	0.5	0.2	0.2	0.7	0.5
$N_{B\bar{B}}$	1.5	1.5	1.5	1.5	1.5	1.5
$f^{+-/00}$	2.5	2.5	2.4	2.4	2.4	2.5
π^0 efficiency	–	–	3.8	3.8	–	3.8
K_S^0 efficiency	–	–	–	–	2.0	2.0
CS efficiency	0.2	0.2	0.7	0.7	0.5	1.7
PID correction	0.1	0.1	0.1	0.2	–	–
ΔE shift and scale	0.1	0.2	1.2	2.0	0.3	1.7
$K\pi$ signal model	0.1	0.2	0.1	<0.1	<0.1	0.1
$\pi\pi$ signal model	<0.1	0.1	<0.1	<0.1	–	–
$K\pi$ feed-across model	<0.1	0.1	<0.1	0.1	–	–
$\pi\pi$ feed-across model	0.1	0.2	<0.1	0.1	–	–
$K_S^0 K^+$ model	–	–	–	–	0.1	–
$B\bar{B}$ model	–	–	0.3	0.5	<0.1	0.3
$q\bar{q}$ flavor model	–	–	–	–	–	0.9
Multiple candidates	<0.1	<0.1	1.0	0.3	0.1	0.3
Total	3.0	3.0	5.1	5.2	3.6	5.8

- number of $B\bar{B}$, and B^+B^- vs $B^0\bar{B}^0$ fraction, are uncertain
→ normalisation uncertainty on all branching fractions
- this is the dominant uncertainty for $K\pi$; disappears on the \mathcal{A}_{CP}

INPUTS: numbers with uncertainties, e.g. Δm

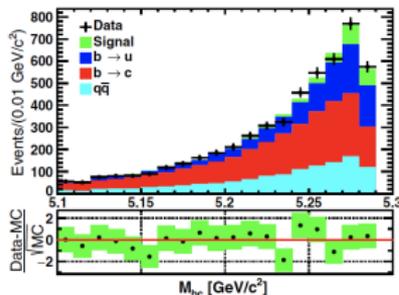
Belle pub197: PRL 99, 131802 (2007); EPR-type flavour entanglement in $B^0\bar{B}^0$



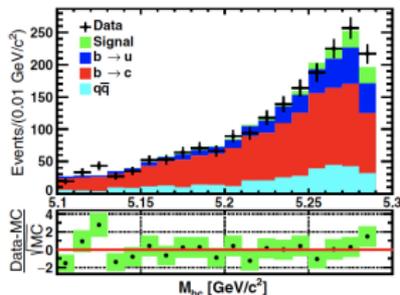
- fits to a Δt distribⁿ in 11 bins; functional form depends on Δm
- Δm is not interesting here, but floating it leads to loss of sensitivity
- world average* measurement: $\langle \Delta m \rangle = (0.496 \pm 0.014) \text{ ps}^{-1}$
- added to the fit by what is now called Gaussian constraint
- large effect on sensitivity: see the yellow boxes

INPUTS: theoretical uncertainties

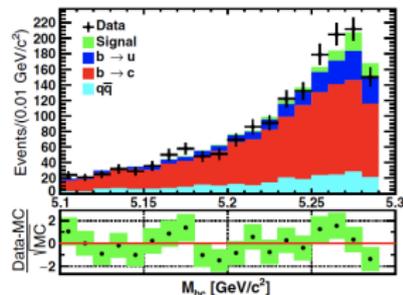
Belle pub582: PRD 106, 032013 (2022); $\mathcal{B}(B^+ \rightarrow \eta^{(\prime)} \ell n u)$



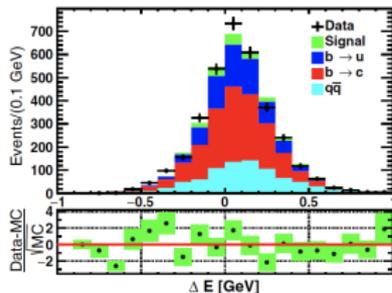
(a) $M_{bc}(\eta \rightarrow \gamma\gamma)$



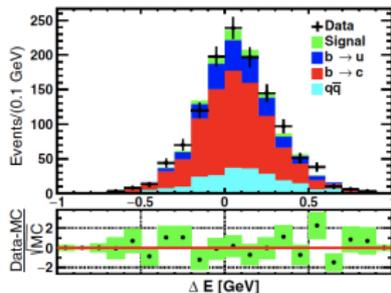
(b) $M_{bc}(\eta \rightarrow \pi^+ \pi^- \pi^0)$



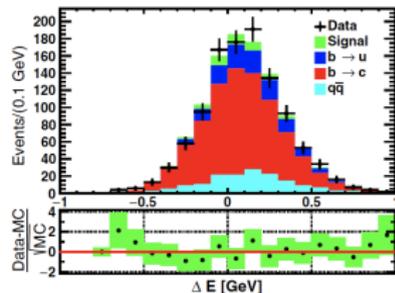
(c) $M_{bc}(\eta' \rightarrow \pi^+ \pi^- \eta(\gamma\gamma))$



(d) $\Delta E(\eta \rightarrow \gamma\gamma)$



(e) $\Delta E(\eta \rightarrow \pi^+ \pi^- \pi^0)$



(f) $\Delta E(\eta' \rightarrow \pi^+ \pi^- \eta(\gamma\gamma))$

INPUTS: theoretical uncertainties

Belle pub582: PRD 106, 032013 (2022); $\mathcal{B}(B^+ \rightarrow \eta^{(\prime)} \ell \nu)$

- form factor assumptions are embedded in signal and background shapes
- uncerTs in FF params propagated as systs
- for some, the underlying *model* is changed, and the shift in result used
- these are *small* uncerTs — see later

The signal decays $B^{\mp} \rightarrow \eta \ell^+ \nu_{\ell}$ and $B^{\pm} \rightarrow \eta' \ell^+ \nu_{\ell}$ are reweighted from the ISGW2 model [30] to the model taken from Ref. [31] with the form factors updated to Ref. [32], using the BZ parametrization and assuming uncorrelated parameters. The decay $B^+ \rightarrow \omega \ell^+ \nu_{\ell}$ is modeled according to Ref. [33] in the MC used and reweighted to Ref. [34] for comparison. The shape of the inclusive component [35] of the $\bar{b} \rightarrow \bar{u} \ell^+ \nu_{\ell}$ transitions is also considered. The form factor uncertainties listed in Table III are based on those reported in the publications they were obtained from. Despite having a slowly varying efficiency the $\eta \rightarrow \gamma \gamma$ mode appears to have the largest such uncertainty.

TABLE III. Breakdown of the systematic uncertainty in %.

Source	$\eta(\gamma\gamma)$	$\eta(\pi^+ \pi^- \pi^0)$	η'
Statistical	22	39	46
Combined Systematic	11	14	11
$\mathcal{B}(B^{\pm} \rightarrow X_{Bkg})$	2.4	1.7	1.3
$\mathcal{B}(\eta^{(\prime)} \rightarrow X)$	0.51	1.2	1.7
$B \rightarrow D^{(*,*)} \ell^+ \nu_{\ell}$ form factor	0.82	1.1	1.3
$B \rightarrow \eta^{(\prime)} \ell^+ \nu_{\ell}$ form factor	3.0	2.9	0.14
$B \rightarrow \omega \ell^+ \nu_{\ell}$ form factor	0.81	2.1	2
$\bar{b} \rightarrow \bar{u} \ell^+ \nu_{\ell}$ shape	0.39	0.15	0.21
Background with K_L^0	3.5	8.6	3.8
Continuum	0.2	0.62	0.63
N_{BB}	1.4	1.4	1.4
$\mathcal{B}(Y(4S) \rightarrow B^+ B^-)$	1.2	1.2	1.2
$\bar{b} \rightarrow \bar{u} \ell^+ \nu_{\ell}$ yield	4.1	5.2	4.4
Monte Carlo statistics	0.86	1.3	2.3
Charged tracks	0.35	1.1	1.1
γ detection	4.0	2.5	4.0
Electron PID	1.6	1.6	1.5
Muon PID	2.1	2.1	2
First π^{\pm} PID	0	0.97	1.1
Second π^{\pm} PID	0	1.3	2.2
Misidentified Leptons	4.3	5.5	2.3
Control Mode	5.0	5.0	5.0

(uncertainties on) any **AUXILIARY** measurements

- 1 Introduction: What are systematics? / A taxonomy
- 2 (uncertainties on) the **INPUTS** to the measurement
- 3 **(uncertainties on) any **AUXILIARY** measurements**
 - ... of efficiencies
 - ... of rates
 - ... of the resolution function
 - ... of the interaction region
- 4 (uncertainties on) the **CALIBRATION** of the apparatus
- 5 Roundup of methods

AUXILIARY measurements . . . of efficiencies (1)

Belle II pub24: 2310.06381 → PRD; $B \rightarrow K\pi, \pi\pi$ BFs and \mathcal{A}_{CP}

TABLE II. Summary of the relative systematic uncertainties (%) on the branching fractions.

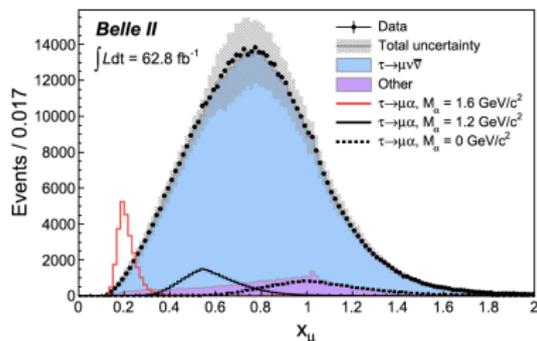
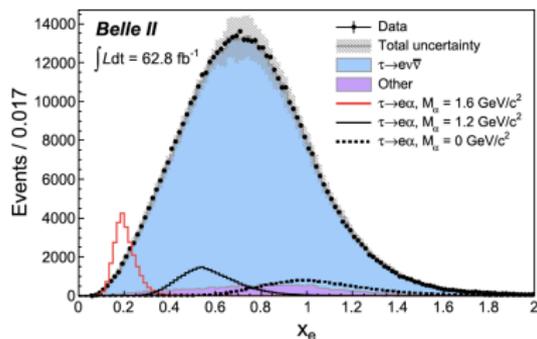
Source	$B^0 \rightarrow K^+\pi^-$	$B^0 \rightarrow \pi^+\pi^-$	$B^+ \rightarrow K^+\pi^0$	$B^+ \rightarrow \pi^+\pi^0$	$B^+ \rightarrow K_S^0\pi^+$	$B^0 \rightarrow K_S^0\pi^0$
Tracking	0.5	0.5	0.2	0.2	0.7	0.5
$N_{B\bar{B}}$	1.5	1.5	1.5	1.5	1.5	1.5
$f^{+-/00}$	2.5	2.5	2.4	2.4	2.4	2.5
π^0 efficiency	–	–	3.8	3.8	–	3.8
K_S^0 efficiency	–	–	–	–	2.0	2.0

- we measure the **track-finding efficiency** in dedicated analyses, but the value has an uncertainty:
this appears according to the number of tracks in each mode
- π^0 -, K_S^0 -finding efficiencies likewise
- again these will “cancel in the ratio” for \mathcal{A}_{CP} , up to possible charge-dependent effects that need to be checked
- in a different sort of analysis, say with a normalisation mode, they will *not* necessarily cancel:
 - they are in general $p, p_T, \cos\theta$ -etc.-dependent, esp. for PID
 - signal and normalisation modes will not have the same distribution . . .

AUXILIARY measurements . . . of efficiencies (2)

Belle II pub10: PRL 130, 181803 (2023); LFV $\tau \rightarrow \ell\alpha$ search

Knowledge of **PID efficiencies** can be limited by calibration on data:



The leading systematic uncertainties originate from the corrections to the lepton-identification efficiency and particle misidentification rate, based on comparison of calibration samples in data and simulated events. These corrections depend on the momentum and polar angle; their typical ranges are summarized in Table II. The resulting uncertainties are asymmetric and strongly depend on x_ℓ ; their ranges and averaged values over the standard-model yields are also reported in the same table. The contribution from lepton-identification efficiency partially cancels in the ratio between signal and normalization channels; while the contribution from particle misidentification rates does not, as it affects only other background sources.

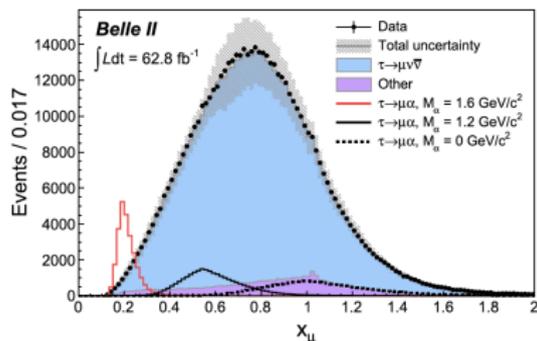
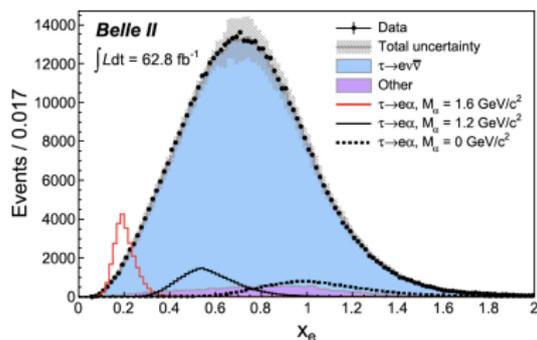
TABLE II. Typical ranges for corrections to the lepton-identification efficiencies and misidentification rates, together with ranges for their respective uncertainties and their average values.

	Correction range	Uncertainty range(%)	Average uncert.(%)
Electron identification	0.84–1.06	0.9–12.6	+5.3, –2.9
Muon identification	0.63–1.02	1.3–32.8	+11.7, –1.6
Electron misidentification	0.6–6.0	4.3–34.6	+17.6, –14.7
Muon misidentification	0.3–1.5	1.4–37.0	+18.0, –18.2

AUXILIARY measurements ... of efficiencies (2)

Belle II pub10: PRL 130, 181803 (2023); LFV $\tau \rightarrow \ell\alpha$ search

Trigger and other efficiencies in this analysis are “hidden” in the \mathcal{L} :



Uncertainties from the trigger and π^0 reconstruction efficiency corrections are also taken into account. Trigger uncertainties range in 0.1%–4% for the electron channel and in 0.2%–1.5% for the muon channel, depending on x_ℓ . Neutral pion reconstruction efficiency is evaluated from studies on independent samples to be 0.914 ± 0.020 . Each of these systematic uncertainties is included in the likelihood as an additional shape-correlated nuisance parameter that is assumed to follow a Gaussian distribution. Other sources of uncertainty from track reconstruction efficiency, beam-energy determination, relative reconstruction efficiency, and momentum-scale correction have negligible impact on the results.

AUXILIARY measurements . . . of efficiencies (3)

Belle II pub13: PRD 107, 112009 (2023); $\pi^0\pi^0$ BF and \mathcal{A}_{CP}

sometimes the reported “uncertainty” of the auxiliary measurement is secondary: the systematic is dominated by our **lack of understanding or confidence** in what is going on . . .

The main sources of systematic uncertainties are listed in Table I and are evaluated as follows. A 3.4% systematic uncertainty associated with the π^0 reconstruction efficiency is determined from data using the decays $D^{*-} \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^-$ and $D^{*-} \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^-$, where the π^0 selection is identical to that of the signal. The π^0 reconstruction efficiency as a function of momentum is also measured using $\tau^- \rightarrow 3\pi\pi^0\nu$ and $\tau^- \rightarrow 3\pi\nu$ decays. A difference of 4.7% in efficiency is observed between the measurement based on D decays and the measurement based on τ leptons. This difference increases the systematic uncertainty for a total of 5.8% per pion. The total systematic uncertainty associated with the π^0 reconstruction efficiency is then 11.6%, as there are two pions and their errors are fully correlated.

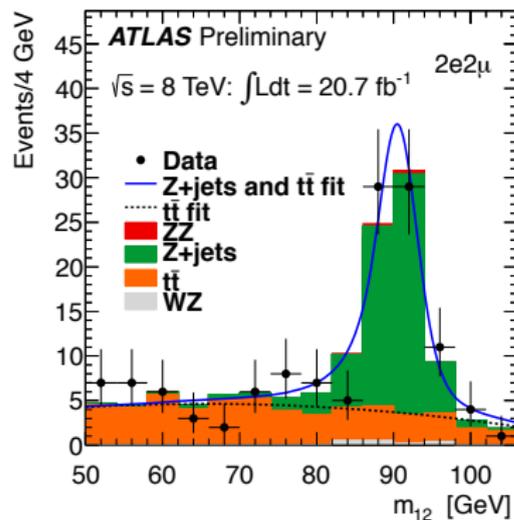
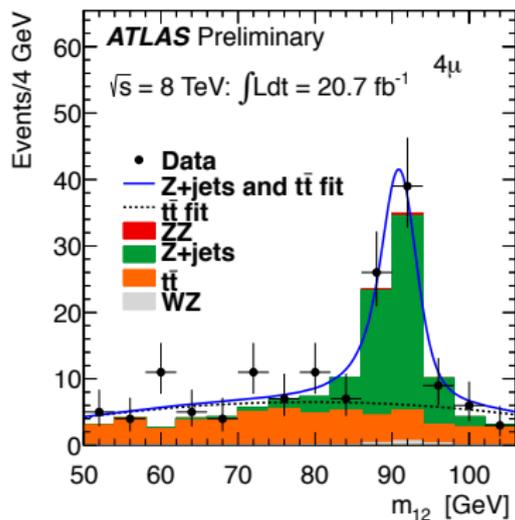
TABLE I. Summary of systematic uncertainties. The total is calculated by adding all systematic uncertainties in quadrature.

Source	$\mathcal{B}(\%)$	\mathcal{A}_{CP}
π^0 reconstruction efficiency	11.6	...
Continuum parametrization	7.4	0.02
Continuum classifier efficiency	6.5	...
$1 + f^{+-}/f^{00}$	2.5	...
Fixed $B\bar{B}$ background yield	2.3	0.01
Fixed signal r bin fractions	2.2	0.01
Knowledge of the photon-energy scale	2.0	...
Assumption of independence of ΔE from r	1.8	<0.01
Number of $B\bar{B}$ meson pairs	1.5	<0.01
Choice of $(M_{bc}, \Delta E)$ signal model	1.3	0.02
Fixed continuum r bin fraction	1.1	<0.01
Branching fraction fit bias	1.0	...
Best candidate selection	0.2	<0.01
Mistagging parameters	...	0.05
Potential nonzero $B\bar{B}$ background \mathcal{A}_{CP}	...	0.03
\mathcal{A}_{CP} fit bias	...	0.02
Continuum $q \cdot r$ asymmetry	...	0.01
Total	16.2	0.07

AUXILIARY measurements ... of rates

ATLAS-CONF-2013-013; Higgs properties in $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$

- $H \rightarrow \ell\ell + \mu\mu$ have “reducible” bkgds due to $t\bar{t}$ & $Z + \text{jets}$ events
- normalisations are set using complementary “control regions”
 - 1 removing $\mu\mu$ isolation cuts, & requiring ≥ 1 isolⁿ failure (not shown)
 - 2 removing $\mu\mu$ isolation cuts, & requiring ≥ 1 IP significance failure:



AUXILIARY measurements . . . of the resolution fn

Belle II pub16: PRD 107, L091102 (2023); B^0 lifetime and Δm measurement

Leading systematic: **function params** that can't all be floated at once . . .

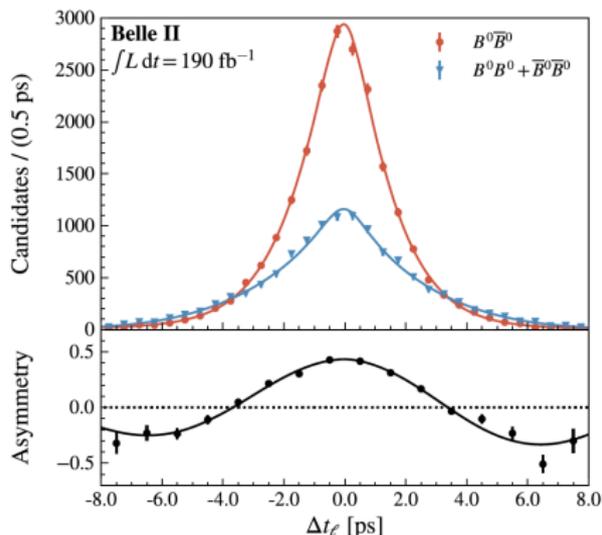


FIG. 3. Distribution of Δt_ℓ in data (points) and the fit model (lines) for opposite-flavor candidate pairs (red) and same-flavor pairs (blue) and their asymmetry (black).

There are several sources of systematic uncertainty; these are listed in Table I and described below. The dominant systematic uncertainty is due to potential discrepancies between the assumed values (fixed in the fit) of the response-function parameters and the true values in the data. For each fixed parameter, we repeat the fit with the parameter allowed to vary. We add all the resulting changes in the result in quadrature and include this value as a systematic uncertainty.

TABLE I. Systematic uncertainties.

Source	τ_{B^0} [ps]	Δm_d [ps^{-1}]
Fixed response-function parameters	0.0063	0.003
Analysis bias	0.004	0.001
Detector alignment	0.003	0.002
Interaction-region precision	0.002	0.001
C-distribution modeling	0.000	0.001
$\sigma_{\Delta t_\ell}$ -distribution modeling	0.001	0.001
Correlations of ΔE or C and Δt_ℓ	0.001	0.000
Total systematic uncertainty	0.008	0.005
Statistical uncertainty	0.013	0.008

AUXILIARY measurements . . . of the IR

Belle II pub16: PRD 107, L091102 (2023); B^0 lifetime and Δm measurement

Smaller systematic: **imprecise knowledge of interaction region**

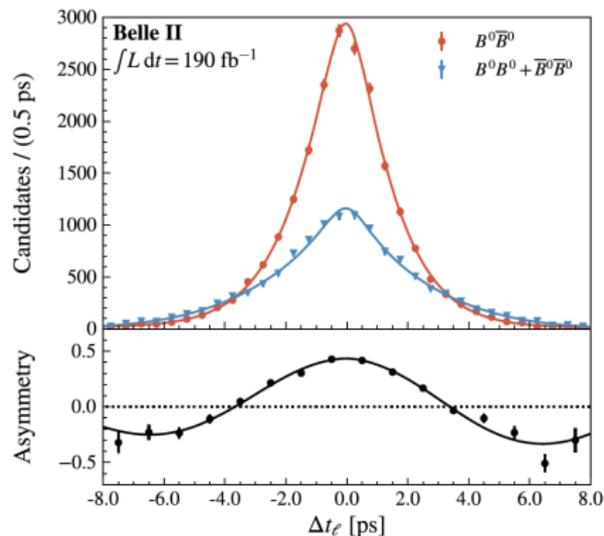


FIG. 3. Distribution of Δt_ℓ in data (points) and the fit model (lines) for opposite-flavor candidate pairs (red) and same-flavor pairs (blue) and their asymmetry (black).

Because we adjust the B_{sig} decay vertex position so that the vector connecting the IR and decay vertex is parallel to the B_{sig} momentum, the precision to which we know the IR affects our determination of ℓ . We repeat our analysis on simulated data in which we shift, rotate, and rescale the IR within its measured uncertainties and assign the changes in the results as systematic uncertainties. We perform an analogous check with changes to \sqrt{s} and the magnitude and direction of the boost vector and find that the results change negligibly.

τ : 0.002 vs 0.008 ps total

Δm : 0.001 vs 0.005 ps^{-1} total

(uncertainties on) the CALIBRATION of ...

- 1 Introduction: What are systematics? / A taxonomy
- 2 (uncertainties on) the INPUTS to the measurement
- 3 (uncertainties on) any AUXILIARY measurements
- 4 **(uncertainties on) the CALIBRATION of the apparatus**
 - ... of measurement of specific quantities
 - ... of “environmental” quantities
 - ... of the experimental technique as a whole
 - ... *of the analyst*: your own choices
 - ... by searching for mistakes
- 5 Roundup of methods

CALIBRATION ... of measurement of specific quantities

Belle II pub15: PRL 127, 211801 (2021); D lifetime

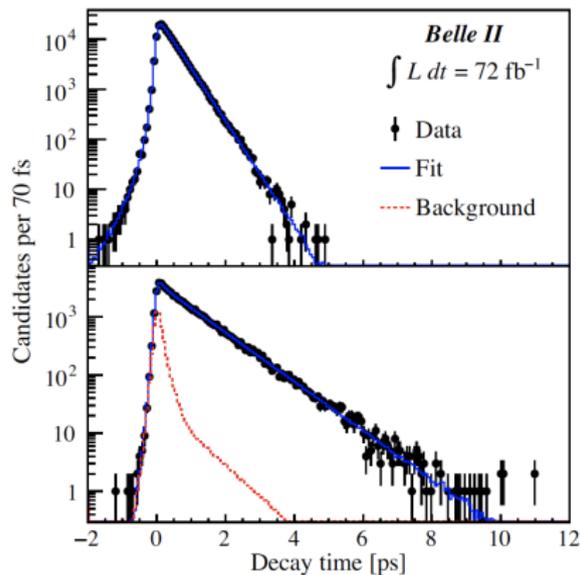


FIG. 2. Decay-time distributions of (top) $D^0 \rightarrow K^- \pi^+$ and (bottom) $D^+ \rightarrow K^- \pi^+ \pi^+$ candidates in their respective signal regions with fit projections overlaid.

The **momentum scale** is important:

- default factor 1.00056
- recommended range [1.00014, 1.00107]
- uncertainty is subleading:

TABLE I. Systematic uncertainties.

Source	$\tau(D^0)$ [fs]	$\tau(D^+)$ [fs]
Resolution model	0.16	0.39
Backgrounds	0.24	2.52
Detector alignment	0.72	1.70
Momentum scale	0.19	0.48
Total	0.80	3.10

Imperfectly known **vertex resolution** is another uncertainty of this type

CALIBRATION ... of “environmental” quantities

Belle II pub15: PRL 127, 211801 (2021); D lifetime

Tracking & vertexing **assumes the alignment** of subdetector elements:

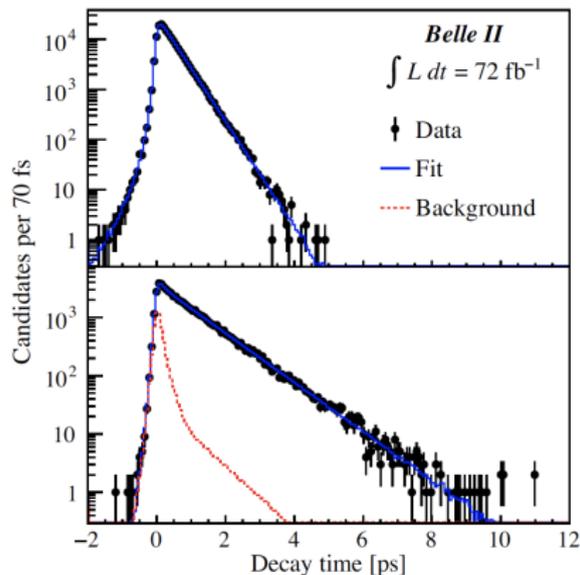


FIG. 2. Decay-time distributions of (top) $D^0 \rightarrow K^- \pi^+$ and (bottom) $D^+ \rightarrow K^- \pi^+ \pi^+$ candidates in their respective signal regions with fit projections overlaid.

background, and cosmic-ray events [26]. Unaccounted-for misalignment can bias the measurement of the charmed decay lengths and hence their decay times. Two sources of uncertainties associated with the alignment procedure are considered: the statistical precision and a possible systematic bias. Their effects are evaluated using simulated signal-only decays reconstructed with a misaligned detector. For the statistical contribution, we consider configurations derived from comparison of alignment parameters determined from data acquired on two consecutive days. These configurations have magnitudes of misalignment comparable to the alignment precision as observed in data averaged over a typical alignment period. For the systematic contribution, we consider configurations derived from simulation studies in which coherent global deformations of the vertex detectors (e.g., radial expansion) are introduced [27]. These deformations have magnitudes, determined by the most misaligned sensors, ranging from about 50 to 700 μm . The alignment procedure determines the magnitude of these deformations within 4 μm accuracy.

CALIBRATION ... of “environmental” quantities

Belle II pub15: PRL 127, 211801 (2021); D lifetime

Tracking & vertexing **assumes the alignment** of subdetector elements:

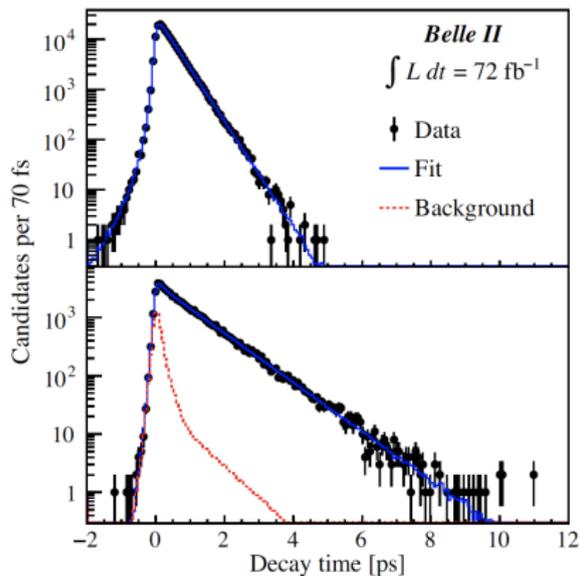


FIG. 2. Decay-time distributions of (top) $D^0 \rightarrow K^- \pi^+$ and (bottom) $D^+ \rightarrow K^- \pi^+ \pi^+$ candidates in their respective signal regions with fit projections overlaid.

We consider configurations in which the CDC is perfectly aligned and configurations in which it is misaligned. Possible effects on the determination of the IR are also introduced by using parameters measured on misaligned samples of simulated $e^+e^- \rightarrow \mu^+\mu^-$ events, to fully mimic the procedure used for real data. For each misalignment configuration, we fit to the reconstructed signal candidates and estimate the lifetime bias. We estimate the systematic uncertainty due to imperfect detector alignment as the sum in quadrature of the largest biases observed in each of the statistical and systematic contributions. The resulting uncertainties are 0.72 and 1.70 fs for $D^0 \rightarrow K^- \pi^+$ and $D^+ \rightarrow K^- \pi^+ \pi^+$ decays, respectively. The absolute length scale of the vertex detector is determined with a precision significantly better than 0.01% and contributes negligibly to the systematic uncertainty.

- Dominant uncertainty for $\tau(D^0)$
- \vec{B} -field anisotropy is another classic example

CALIBRATION ... of the experimental technique as a whole

Here we are putting the whole measurement technique inside a virtual box, as an “instrument” to be calibrated. Typical issues:

- known limitations and omissions in the method
- the equations and parameterizations
- fits: linearity/bias tests
- larger analysis chain: control and validation region studies



Calibration of technique: limitations & omissions

Belle II pub15: PRL 127, 211801 (2021); D lifetime

A small background was left out of the $D^0 \rightarrow K^- \pi^+$ fit:

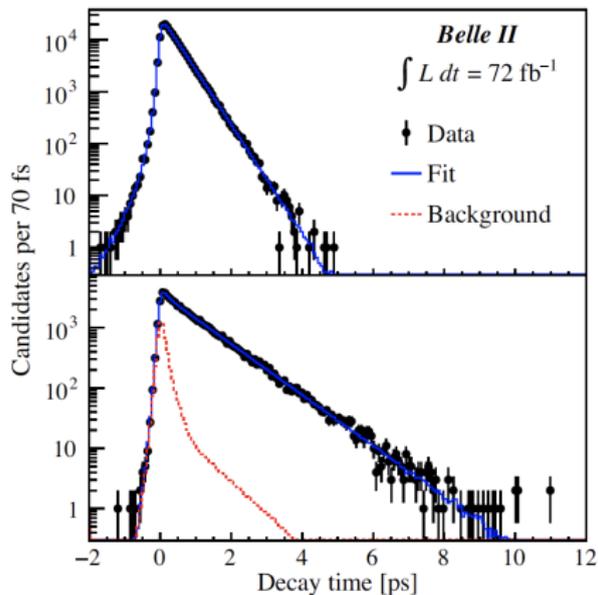


FIG. 2. Decay-time distributions of (top) $D^0 \rightarrow K^- \pi^+$ and (bottom) $D^+ \rightarrow K^- \pi^+ \pi^+$ candidates in their respective signal regions with fit projections overlaid.

The background neglected in the $D^0 \rightarrow K^- \pi^+$ fit could result in a systematic bias on the measured lifetime. To estimate the size of the bias, we fit our model that neglects the background to 500 resampled sets of simulated $e^+ e^-$ collisions, each having the same size and signal-to-background proportion as the data. The measured lifetimes are corrected by subtracting the bias due to the neglected t vs σ_t correlations. The average absolute difference between the resulting value and the simulated lifetime, 0.24 fs, is assigned as a systematic uncertainty due to the neglected background contamination in the $D^0 \rightarrow K^- \pi^+$ fit.

TABLE I. Systematic uncertainties.

Source	$\tau(D^0)$ [fs]	$\tau(D^+)$ [fs]
Resolution model	0.16	0.39
Backgrounds	0.24	2.52
Detector alignment	0.72	1.70
Momentum scale	0.19	0.48
Total	0.80	3.10

Calibration of technique: limitations & omissions

Belle II pub15: PRL 127, 211801 (2021); D lifetime

The sideband method for $D^+ \rightarrow K^- \pi^+ \pi^+$ bkgd may be imperfect:

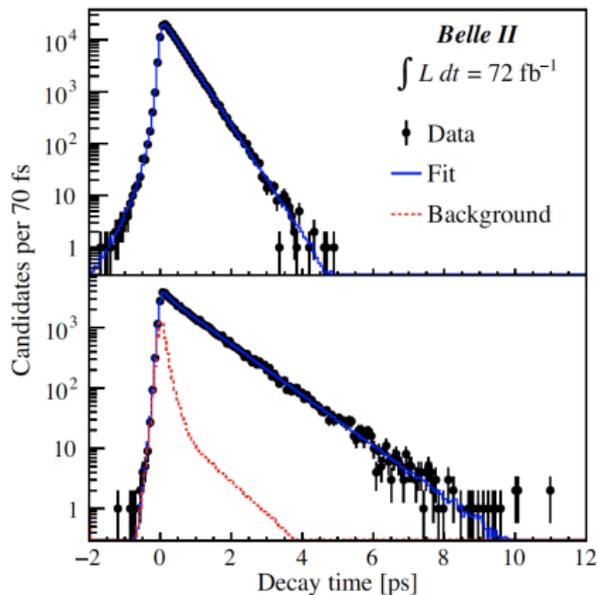


FIG. 2. Decay-time distributions of (top) $D^0 \rightarrow K^- \pi^+$ and (bottom) $D^+ \rightarrow K^- \pi^+ \pi^+$ candidates in their respective signal regions with fit projections overlaid.

The background contamination under the $D^+ \rightarrow K^- \pi^+ \pi^+$ peak is already accounted for in the fit of the D^+ lifetime using sideband data. In simulation, the sideband (t, σ_t) distribution describes the background (t, σ_t) distribution in the signal region well. The same might not hold in data given that some disagreement is observed between data and simulation in the t distribution of the candidates populating the sideband. We fit to one thousand samples of simulated data obtained by sampling the fit PDF for the signal region and by resampling from the simulated e^+e^- collisions for the sideband. The resulting samples feature sideband data that differ from the background in the signal region with the same level of disagreement as observed between data and simulation. The absolute average difference between the measured and simulated lifetimes, 2.52 fs, is assigned as a systematic uncertainty due to the modeling of the background (t, σ_t) distribution.

Source	$\tau(D^0)$ [fs]	$\tau(D^+)$ [fs]
Resolution model	0.16	0.39
Backgrounds	0.24	2.52
Detector alignment	0.72	1.70
Momentum scale	0.19	0.48
Total	0.80	3.10

Calibration of technique: eqns & parameterizations

Belle II pub15: PRL 127, 211801 (2021); D lifetime

Changing how the background is estimated **has a negligible effect:**

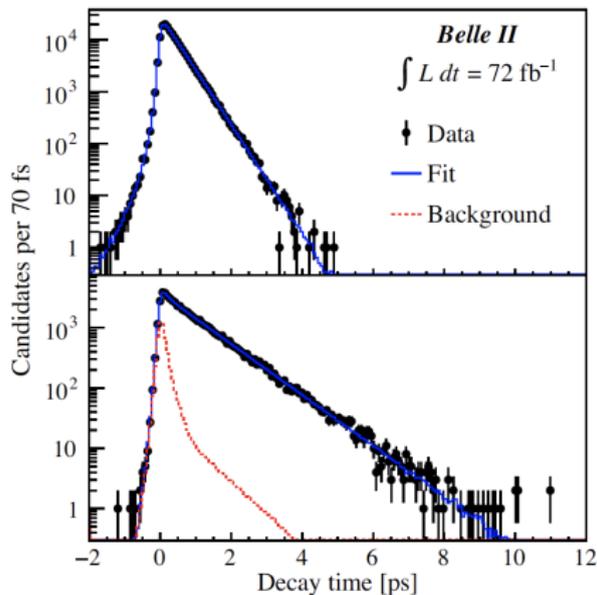


FIG. 2. Decay-time distributions of (top) $D^0 \rightarrow K^- \pi^+ \pi^+$ and (bottom) $D^+ \rightarrow K^- \pi^+ \pi^+$ candidates in their respective signal regions with fit projections overlaid.

In the lifetime fit, the fraction of background candidates in the signal region is constrained from the fit to the $m(K^- \pi^+ \pi^+)$ distribution. When we change this background fraction to values obtained from fitting to the $m(K^- \pi^+ \pi^+)$ distribution with alternative signal and background PDFs, the change in the measured lifetime is negligible.

Calibration of technique: eqns & parameterizations

Belle II pub15: PRL 127, 211801 (2021); D lifetime

Neglect of correlations in fitting has a (small) noticeable effect:

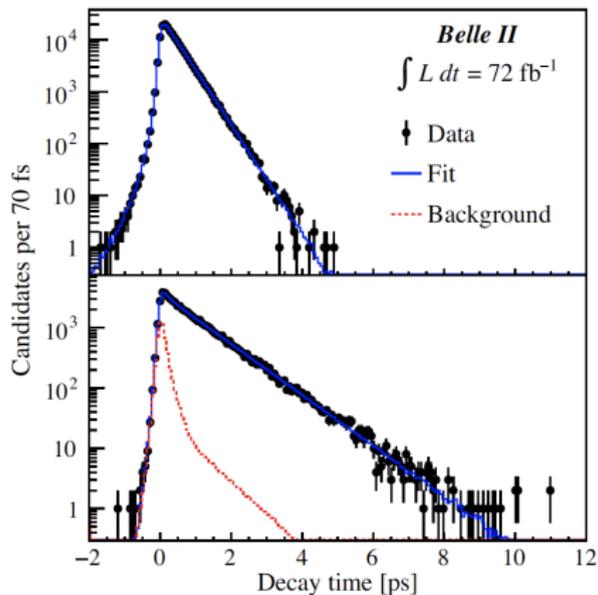


FIG. 2. Decay-time distributions of (top) $D^0 \rightarrow K^- \pi^+$ and (bottom) $D^+ \rightarrow K^- \pi^+ \pi^+$ candidates in their respective signal regions with fit projections overlaid.

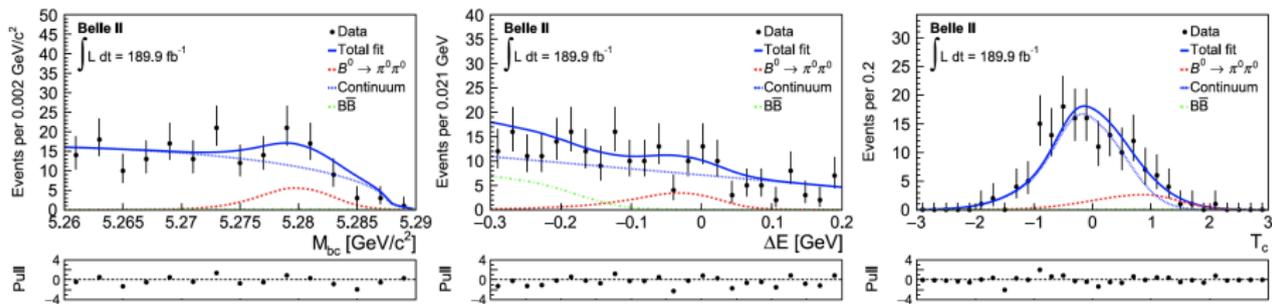
The decay time and decay-time uncertainty are observed to be correlated in data and simulation reproduces these effects well. The dominant effect is that small σ_t values correspond to larger true decay times (and vice versa).

These correlations, when neglected in the fits, result in an imperfect description of the t distribution as a function of σ_t . To quantify the impact on the results, our model that neglects the correlations is fit to 1000 samples of signal-only simulated decays, each the same size as the data. The samples are obtained by resampling, with repetition, a set of simulated e^+e^- collisions corresponding to an integrated luminosity of 500 fb^{-1} . Upper bounds of 0.16 and 0.39 fs on the average absolute deviations of the measured lifetimes from their true values are derived and assigned as the systematic uncertainty due to the imperfect resolution model for the $D^0 \rightarrow K^- \pi^+$ and $D^+ \rightarrow K^- \pi^+ \pi^+$ cases,

Source	$\tau(D^0)$ [fs]	$\tau(D^+)$ [fs]
Resolution model	0.16	0.39
Backgrounds	0.24	2.52
Detector alignment	0.72	1.70
Momentum scale	0.19	0.48
Total	0.80	3.10

Calibration of technique: eqns & parameterizations

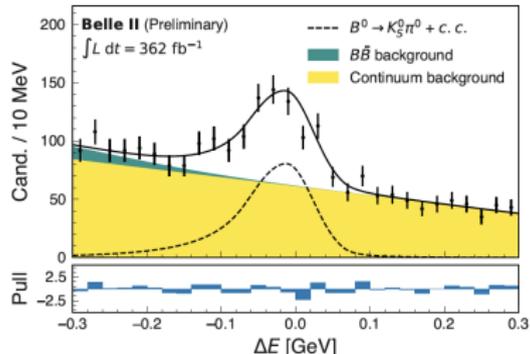
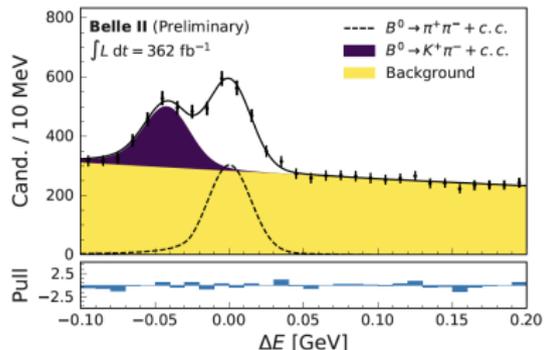
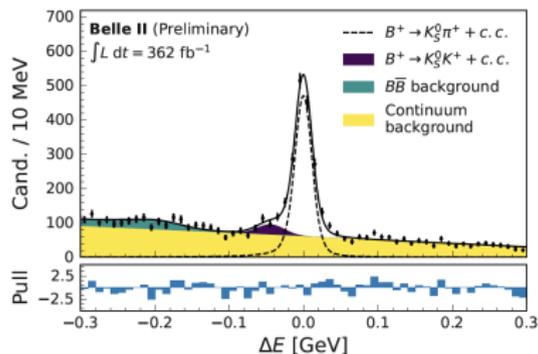
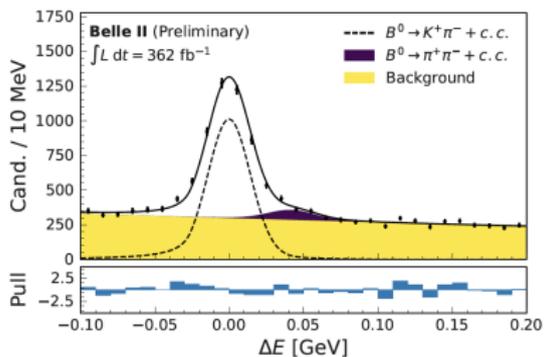
Belle II pub13: PRD 107, 112009 (2023); $\pi^0\pi^0$ BF and \mathcal{A}_{CP}



- some $B\bar{B}$, a *lot* of continuum background under the 3D peak
- continuum shapes taken from the data sideband
 $M_{bc} \in (5.22, 5.27) \text{ GeV}/c^2$, $\Delta E \in (0.1, 0.5) \text{ GeV}$
- uncertainty estimated by shifting shape params by $\pm 1\sigma$ one-by-one (with others shifting per fitted correl^{ns}) and checking yield changes
- after $\epsilon(\pi^0)$, **this is the dominant uncertainty** on $\mathcal{B}(B^0 \rightarrow \pi^0\pi^0)$
- parameterization will come up again, under another heading

Calibration of technique: ML fits (1)

Belle II pub24: 2310.06381 \rightarrow PRD; $B \rightarrow K\pi, \pi\pi$ BFs and \mathcal{A}_{CP}



Calibration of technique: ML fits (1)

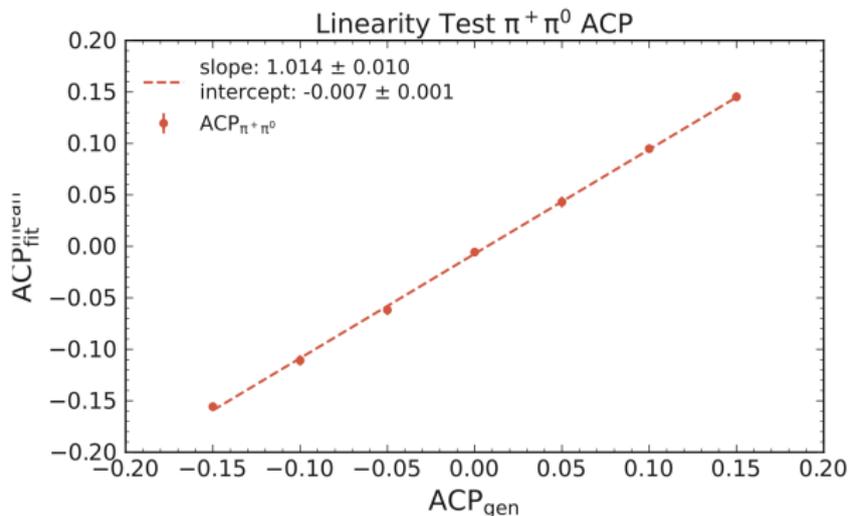
Belle II pub24: 2310.06381 → PRD; $B \rightarrow K\pi, \pi\pi$ BFs and \mathcal{A}_{CP}

Fits may not behave asymptotically/ideally,

even if nothing is “omitted” or wrong: hence “linearity and bias tests”, etc.

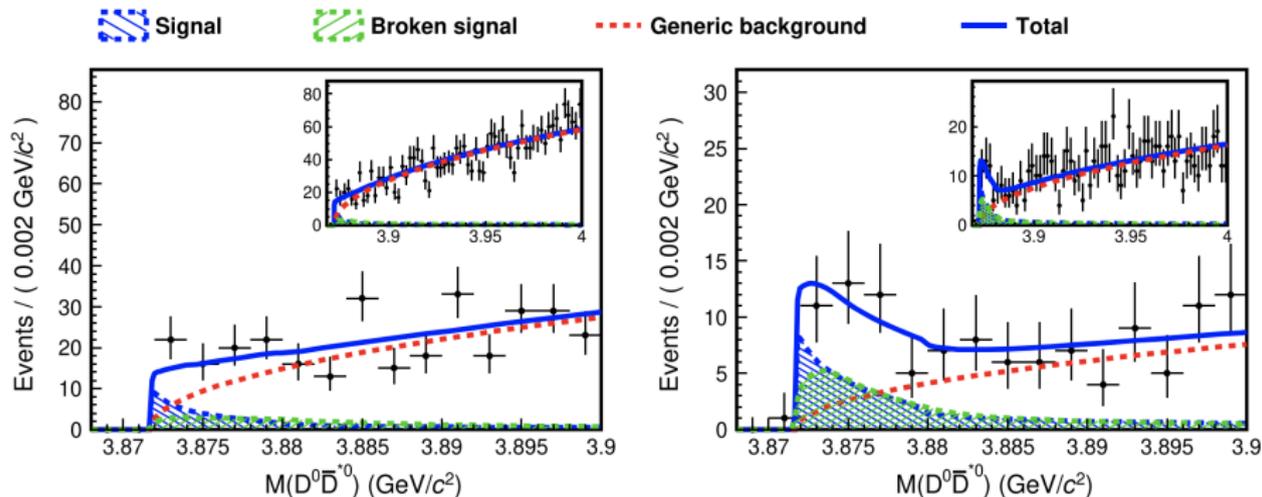
for $\pi^+\pi^0$ and $K_S^0\pi^0$:

- this is the dominant systematic on \mathcal{A}_{CP}
- these modes are statistically limited
- negligible for the branching fraction fits ...



Calibration of technique: ML fits (2)

Belle II pub642: PRD 107, 112011 (2023) X(3872) lineshape in $B \rightarrow D^0 \bar{D}^{*0} K$



Extremely complex analysis, fitting a **signal lineshape** at threshold, over **background**, with **substantial broken signal**, and unstable decay daughters. One of the lineshapes exhibits scaling behaviour in some parameters . . .

Calibration of technique: ML fits (2)

Belle II pub642: PRD 107, 112011 (2023) X(3872) lineshape in $B \rightarrow D^0 \bar{D}^{*0} K$

- BW & Flatté fits
- Flatté fit is very nonlinear:
- significantly changes the reported results: can't deal with this by just "adding a systematic term" ...

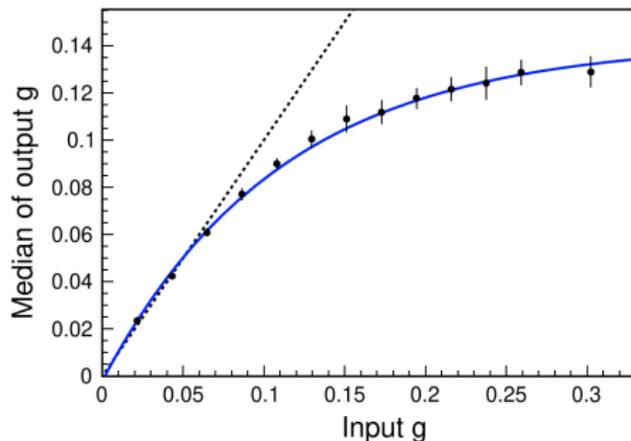


FIG. 5. The median of output values of the coupling constant g , as a function of the input g , evaluated using pseudo-experiments. The dotted black line represents perfect linearity $g_{\text{out}} = g_{\text{in}}$. The solid blue curve represents the threshold function $g_{\text{out}} = 0.14(1 - \exp(-9g_{\text{in}}))$.

Calibration of technique: analysis chain

Belle II pub20: PRD 108, 072012 (2023); CP asymmetries in $B^0 \rightarrow \phi K_S^0$

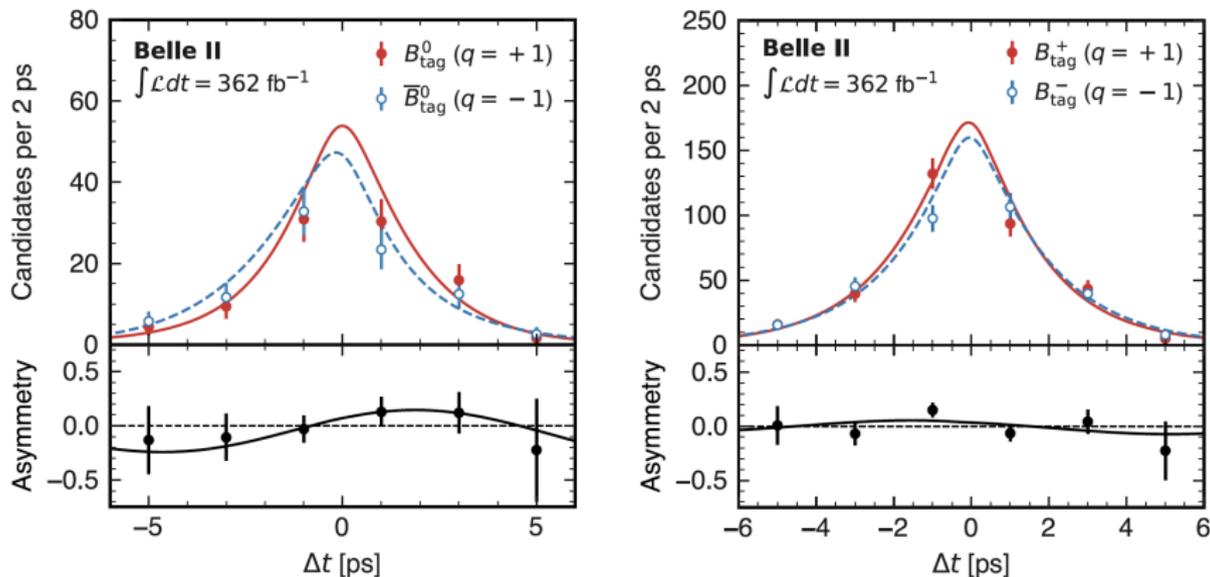


FIG. 3. Distributions, and fit projections, of Δt for flavor-tagged (left) $B^0 \rightarrow \phi K_S^0$ and (right) $B^+ \rightarrow \phi K^+$ candidates subtracted of the continuum background. The fit PDFs corresponding to $q = -1$ and $q = +1$ tagged distributions are shown as dashed and solid curves, respectively. The yield asymmetries, defined as $(N(q = +1) - N(q = -1)) / (N(q = +1) + N(q = -1))$, are displayed in the bottom subpanels.

Calibration of technique: analysis chain

Belle II pub20: PRD 108, 072012 (2023); CP asymmetries in $B^0 \rightarrow \phi K_S^0$

tagging algorithm and Δt resolution calibrated with $B^0 \rightarrow D^{*-}\pi^+$ decays

- this has its own uncertainties

We assess the uncertainty associated with the resolution function and flavor tagging parameters using simplified simulated samples. We generate ensembles assuming for each an alternative value for the above parameters sampled from the statistical covariance matrix determined in the $B^0 \rightarrow D^{(*)-}\pi^+$ control sample. Each ensemble is fitted using the nominal values of the calibration parameters and the standard deviation of the observed biases is used as a systematic uncertainty.

A similar procedure is used to assess a systematic uncertainty due to the systematic uncertainties on the calibration parameters, in which the ensembles are generated by varying each parameter independently within their systematic uncertainty.

TABLE II. Summary of systematic uncertainties.

Source	$\sigma(C)$	$\sigma(S)$
Calibration with $B^0 \rightarrow D^{(*)-}\pi^+$ decays		
Calibration sample size	± 0.010	± 0.009
Calibration sample systematic	± 0.010	± 0.012
Sample dependence	$+0.005$	$+0.021$
Fit model		
Fit bias	$+0.028$ -0.017	$+0.033$ -0.062
$B^0 \rightarrow K^+K^-K_S^0$ backgrounds	$+0.020$	-0.011
Fixed fit shapes	± 0.009	± 0.022
τ_{B^0} and Δm_d	± 0.006	± 0.022
$C_{K^+K^-K_S^0}$ and $S_{K^+K^-K_S^0}$	± 0.014	± 0.013
$B\bar{B}$ background asymmetry	$+0.019$ -0.030	$+0.017$ -0.031
Tag-side interference	<0.001	$+0.012$
Candidate selection	-0.032	-0.002
Δt measurement		
Tracker misalignment	-0.002	-0.002
Momentum scale	± 0.001	± 0.001
Beam spot	± 0.002	± 0.002
Δt approximation	<0.001	-0.018
Total systematic	$+0.046$ -0.052	$+0.058$ -0.082
Statistical	± 0.201	± 0.256

Calibration of technique: analysis chain

Belle II pub20: PRD 108, 072012 (2023); CP asymmetries in $B^0 \rightarrow \phi K_S^0$

tagging algorithm and Δt resolution calibrated with $B^0 \rightarrow D^{*-}\pi^+$ decays

- this has its own uncertainties
- it has to be ported to ϕK_S^0

We estimate the impact of differences in the resolution function and tagging performance between the signal and calibration samples. We apply the resolution function and flavor-tagging calibration obtained from a simulated $B^0 \rightarrow D^{(*)-}\pi^+$ sample and repeat the measurement of C and S over an ensemble of simulated $B^0 \rightarrow \phi K_S^0$ events. The average deviation of the CP asymmetries from their generated values is assigned as a systematic uncertainty.

TABLE II. Summary of systematic uncertainties.

Source	$\sigma(C)$	$\sigma(S)$
Calibration with $B^0 \rightarrow D^{(*)-}\pi^+$ decays		
Calibration sample size	± 0.010	± 0.009
Calibration sample systematic	± 0.010	± 0.012
Sample dependence	$+0.005$	$+0.021$
Fit model		
Fit bias	$+0.028$ -0.017	$+0.033$ -0.062
$B^0 \rightarrow K^+K^-K_S^0$ backgrounds	$+0.020$	-0.011
Fixed fit shapes	± 0.009	± 0.022
τ_{B^0} and Δm_d	± 0.006	± 0.022
$C_{K^+K^-K_S^0}$ and $S_{K^+K^-K_S^0}$	± 0.014	± 0.013
$B\bar{B}$ background asymmetry	$+0.019$ -0.030	$+0.017$ -0.031
Tag-side interference	<0.001	$+0.012$
Candidate selection	-0.032	-0.002
Δt measurement		
Tracker misalignment	-0.002	-0.002
Momentum scale	± 0.001	± 0.001
Beam spot	± 0.002	± 0.002
Δt approximation	<0.001	-0.018
Total systematic	$+0.046$ -0.052	$+0.058$ -0.082
Statistical	± 0.201	± 0.256

Calibration of technique: analysis chain

Belle II pub20: PRD 108, 072012 (2023); CP asymmetries in $B^0 \rightarrow \phi K_S^0$

tagging algorithm and Δt resolution calibrated with $B^0 \rightarrow D^{*-} \pi^+$ decays

- this has its own uncertainties
- it has to be ported to ϕK_S^0

We estimate the impact of differences in the resolution function and tagging performance between the signal and calibration samples. We apply the resolution function and flavor-tagging calibration obtained from a simulated $B^0 \rightarrow D^{(*)-} \pi^+$ sample and repeat the measurement of C and S over an ensemble of simulated $B^0 \rightarrow \phi K_S^0$ events. The average deviation of the CP asymmetries from their generated values is assigned as a systematic uncertainty.

- procedure is a variant of the *control and validation region* studies beloved of the LHC

TABLE II. Summary of systematic uncertainties.

Source	$\sigma(C)$	$\sigma(S)$
Calibration with $B^0 \rightarrow D^{(*)-} \pi^+$ decays		
Calibration sample size	± 0.010	± 0.009
Calibration sample systematic	± 0.010	± 0.012
Sample dependence	$+0.005$	$+0.021$
Fit model		
Fit bias	$+0.028$ -0.017	$+0.033$ -0.062
$B^0 \rightarrow K^+ K^- K_S^0$ backgrounds	$+0.020$	-0.011
Fixed fit shapes	± 0.009	± 0.022
τ_{B^0} and Δm_d	± 0.006	± 0.022
$C_{K^+ K^- K_S^0}$ and $S_{K^+ K^- K_S^0}$	± 0.014	± 0.013
$B\bar{B}$ background asymmetry	$+0.019$ -0.030	$+0.017$ -0.031
Tag-side interference	<0.001	$+0.012$
Candidate selection	-0.032	-0.002
Δt measurement		
Tracker misalignment	-0.002	-0.002
Momentum scale	± 0.001	± 0.001
Beam spot	± 0.002	± 0.002
Δt approximation	<0.001	-0.018
Total systematic	$+0.046$ -0.052	$+0.058$ -0.082
Statistical	± 0.201	± 0.256

CALIBRATION ... *of the analyst: your own choices*



I went forward in time, to view all possible ways we might conduct the analysis.

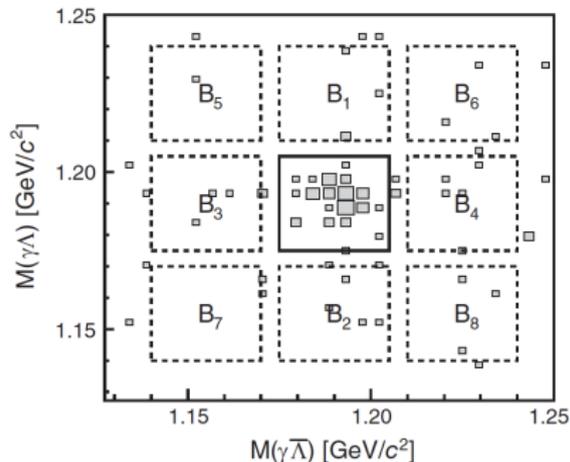
How many did you see? 14,000,605.

How many gave us the right answer? 1.

Calibration of choices: fitting region & other choices

Belle pub632: PRD 107, 072008 (2023); $e^+e^- \rightarrow \Sigma\bar{\Sigma}$ via ISR

The (low) background is estimated using sidebands. But, chosen how?



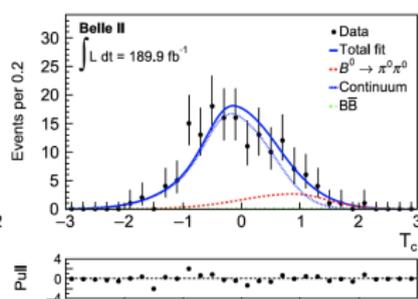
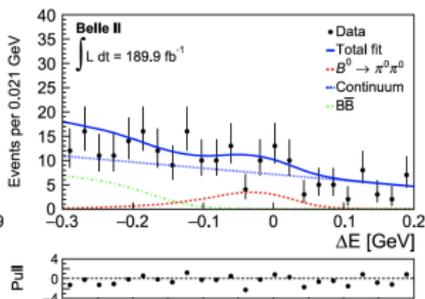
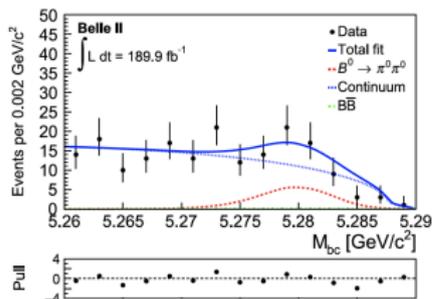
- try larger, smaller, shifted choices of sidebands
- leads to noticeable yield changes: the co-leading systematic at 6%

TABLE IV. Summary of systematic uncertainties for the $e^+e^- \rightarrow \gamma_{\text{ISR}}\Sigma^0\bar{\Sigma}^0$ cross section measurement.

Source	Systematic uncertainty (%)
Tracking	1.4
PID	2.7
Λ reconstruction	5.4
$\Sigma^0/\bar{\Sigma}^0$ mass resolution	0.6
Sideband method	6
$e^+e^- \rightarrow \gamma_{\text{ISR}}\Sigma^0\bar{\Sigma}^0\pi^0$ background	3–9
Other two $\Sigma^0/\bar{\Sigma}^0$ background	2–6
Integrated luminosity	1.4
ISR emission probability	1
PHOKHARA simulation	1
$\Lambda \rightarrow p\pi^-$ branching fraction	0.8
Modeling of angular dependence	3–5
Modeling of energy dependence	1–5
Trigger	3
The fit to efficiency	1
Sum in quadrature	11–16

Calibration of choices: parameterizations

Belle II pub13: PRD 107, 112009 (2023); $\pi^0\pi^0$ BF and \mathcal{A}_{CP}



- reminder: 3D signal peak over $B\bar{B}$ and continuum background
- previously: how well do we know the params of the continuum shape?
- now consider: *choices* were also made in modelling the signal: MC-based KDE in $(M_{bc}, \Delta E)$; what if another shape had been used?
- uncorrelated product of CB functions tried as alternative: 1.3% effect, minor *cf.* 16.2% total systs, but in other cases the effect can be larger
- esp. w limited samples, & limited or absent controls, no assumption-free way to make such choices — so an uncertainty is appropriate

Calibration of choices

- fitting region choices
- parameterizations

In both these cases it is very hard to claim that the default choices are “inevitable” or obviously correct.

Older analyses sometimes include such uncertainties due to

- object selection (how many SVD or CDC hits)
- event selection cuts
- other exact cut values
- dependence on any other choice that has an arbitrary element

This has gone a bit out of style — because where do you stop? — and also for a technical reason (see later). But the basic idea is sound.

CALIBRATION ... by searching for mistakes

- **biases:** signal of $x \rightarrow$ mean measurement of $x + \delta$
- **instabilities:** [butterfly-wingbeat] \rightarrow measured yield changes
- **mis-classifications and omitted categories:**
let's say, the $t\bar{t}$ bkgd has *two components with different behaviour under cuts & differing $m_{4\ell}$ distrib^{ns}*, unresolved by control samples
 - with $2\times$ the sample, if careful, we'll notice problems in fitting
 - with $4\times$ the sample, there will be clear and nasty discrepancies

That was a made-up example, but the phenomenon is very real: Belle $D^0 - \bar{D}^0$ mixing ($D \rightarrow K\pi$) took one year per doubling in sample, to refine method enough to keep systematics under control

- **misunderstanding the relⁿship between auxiliary & principal meas^{ts}**
- **unknown unknowns**

All are **problems that should be fixed:** the corresponding **systematics** are

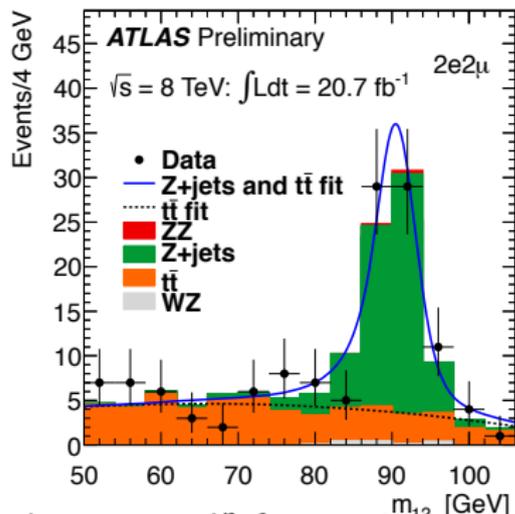
- estimates (guesses?) of possible residual problems
- the tolerances of the tests and cross-checks used

Mistakes? (1) e.g. Relationship with the auxiliaries

ATLAS-CONF-2013-013; Higgs properties in $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$

control region 2 results for $H \rightarrow \ell\ell + \mu\mu$ are extrapolated to the signal region using IP signif. & isolⁿ requirement efficiencies from $Zb\bar{b}$ MC:

- what if this is wrong?
- efficiency validated with *another* control region, requiring $Z + \mu$
- test fails? : stop and try to
 - gain understanding, then
 - fix problem if possible, else
 - back up and change method
- many analyses have dead-ends & side-branches, documented or not
- test succeeded! : a 10% uncert^y on the extrapolⁿ factors is assigned
- note this “data driven” bkgd estimate has embedded dependence on {MC, physical insight, expert judgement on validation, rules of art}



Mistakes? (2) coding and other bugs

```
int getRandomNumber()  
{  
    return 4; // chosen by fair dice roll.  
              // guaranteed to be random.  
}
```

Mistakes? (3) unknown unknowns

Reports that say that something hasn't happened are always interesting to me, because as we know, there are **known knowns**; there are things we know that we know. There are **known unknowns**; that is to say, there are things that we now know we don't know. But there are also **unknown unknowns** — there are things we do not know we don't know.

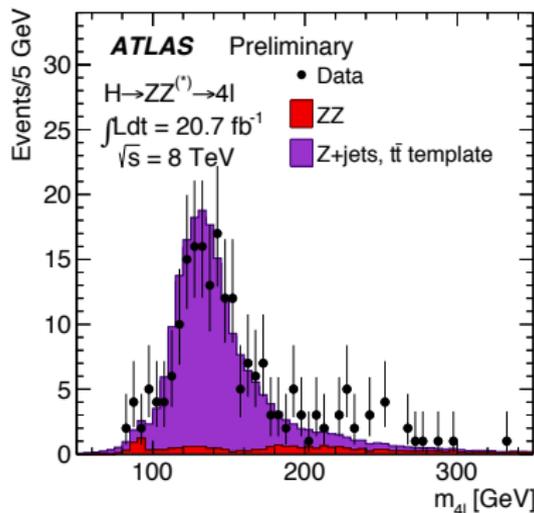
- Donald Rumsfeld got mocked in the media for this, but he had a point
- e.g. “analysis-level” information is vulnerable to subtle problems:
 - what if there is a distinction in response in a drift chamber with stereo layers, never spotted because one must compare the response of $\{U, V\}$ wires to **+ve and -ve** tracks going **forward and backward**?
 - what if **out-of-spill** calorimeter clusters give a decayed but measurable response, not tagged in analysis-level data (& some are **back-to-back**)?
- one *builds confidence* with a new {detector, code, technique} by doing basic & known things first (e.g. so-called “rediscovery” analyses)
- and one relies on . . .

Mistakes? (4) cross-checks

ATLAS-CONF-2013-013; Higgs properties in $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$

additional studies that do not contribute directly to any of the bkgd measurements or systematics directly, but are there to spot problems:

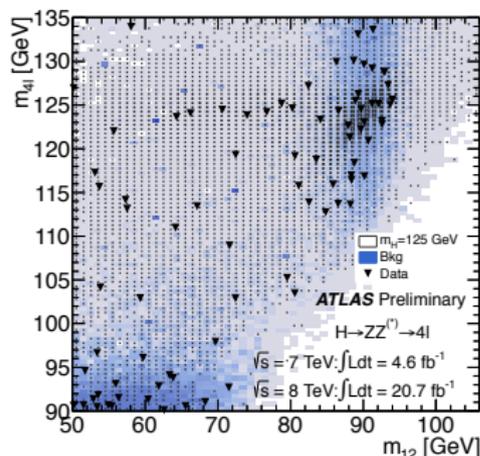
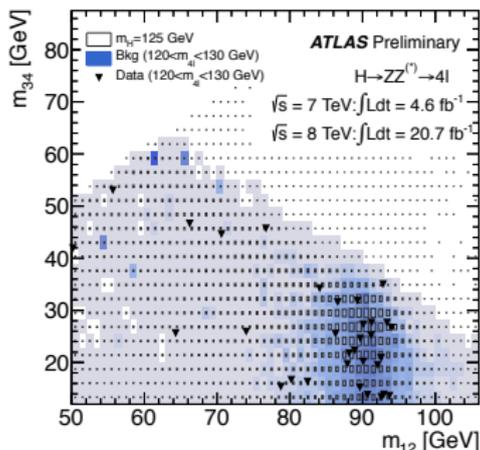
- the $t\bar{t}$ bkgd measurement for $H \rightarrow \ell\ell + \mu\mu$ is cross-checked using a $e^\pm\mu^\mp + \mu^+\mu^-$ sample, with $M(e^\pm\mu^\mp) \in (50, 106)$ GeV
- the $ZZ^{(*)}$ signal and the $Z + \text{jets}$ and $t\bar{t}$ bkgds are checked in another control region:
- agreement is not bad, but is imperfect — how much does this matter?



Mistakes? (4) cross-checks

ATLAS-CONF-2013-013; Higgs properties in $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$

- the final analysis uses a *look-back plot* to check the $(m_{12}, m_{34}, m_{4\ell})$ distribution *in the absence of the Z-mass constraint*:



- your {advisor, RC, journal referee} may ask you for such plots
- “But what are you looking for?”
- “I don’t know, but I may know it when I see it . . .”

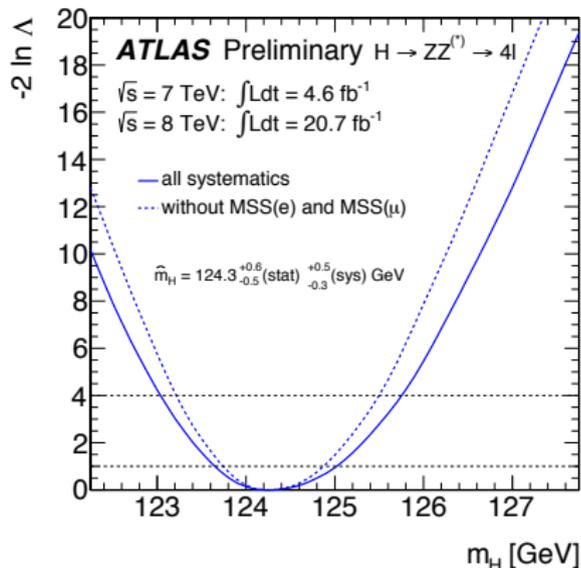
Roundup of methods: (1) profile likelihood

consider

- a quantity of interest μ (signal strength, mass, ...)
- quantities θ , say $\theta = (\theta_s, \theta_b, n_b)$ governing shape & bkgd normⁿ
- form the **profile \mathcal{L} ratio**

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})}$$

- $\hat{\mu}$ and $\hat{\theta}$ are ML estimators
- $\hat{\hat{\theta}}(\mu)$ is the best estimate of θ for the given μ value
- letting “nuisance parameters” θ float at each μ will
 - improve the \mathcal{L} there, and
 - broaden the distribution
- a.k.a. “the MINUIT method”, long used intuitively in HEP



Roundup of methods: (2) Gaussian constraint

ATLAS ATL-COM-PHYS-2012-089; $t\bar{t}$, WW , $Z/\gamma^* \rightarrow \tau\tau$

$$\mathcal{L}_i(N_{sig}, \alpha_j) = \mathcal{P}_i \left(N^{obs} \mid N^{exp}(N_{sig}, \alpha_j) \right) \prod_{j \in \text{syst}} \mathcal{G}(\alpha_j \mid 0, 1)$$

heavily used at ATLAS:

- trigger efficiency officially measured, say $\epsilon = 0.25_{-0.02}^{+0.03}$ at $y = 0$ & $p_T = 20$ GeV, and rising with energy
- ideally, scale for the distⁿ represented by some $f^n f(x)$
- approximate as $\mathcal{G}(x \mid \mu, \sigma)$
- transform to $\alpha = (x - \mu)/\sigma$
- now expressed as $\mathcal{G}(\alpha \mid 0, 1)$

Fit Parameter α	Parameter Value	Symmetric Uncertainty	Asymmetric Uncertainty
Electron trigger	0.00	1.00	-1.00 1.00
Electron reconstruction, ID, isolation	-0.10	1.00	-1.00 1.00
Electron momentum scale	0.00	1.00	-1.00 1.00
Electron momentum resolution	0.00	1.00	-1.00 1.00
Muon reconstruction, ID, isolation	-0.03	1.00	-1.00 1.00
Muon trigger	0.00	1.00	-1.00 1.00
Muon momentum scale	0.00	1.00	-1.00 1.00
Muon momentum resolution	0.00	1.00	-1.00 1.00
Jet vertex fraction	-0.02	1.00	-1.00 1.00
E_T^{miss} Cell-out	0.00	1.00	-1.00 1.00
Pileup	0.00	1.00	-1.00 1.00
WZ/ZZ cross sections	-0.03	1.00	-1.00 1.00
Wt cross section	-0.21	0.99	-1.00 1.00
Luminosity	-0.06	1.00	-1.00 1.00
LHC beam energy	0.00	1.00	-1.00 1.00
ISR/FSR	0.00	1.00	-1.00 1.00
$t\bar{t}$ Parton Shower	0.00	1.00	-1.00 1.00
WW Parton Shower	0.00	1.00	-1.00 1.00
$Z \rightarrow \tau\tau$ Parton Shower	0.00	1.00	-1.00 1.00
$t\bar{t}$ Generator	0.00	1.00	-1.00 1.00
WW Generator	0.00	1.00	-1.00 1.00

Issues, questions, and tricks of the trade

- note that biases must first be corrected!
- do we know what the probability distribution is?
- cancellation *versus* partial cancellation
- rough but conservative estimates (easier/quicker than exact ones, and if the effect still turns out to be small, the exact estimate can be safely skipped)
- possible effects versus proven effects, or, against the 3σ standard
- double-counting of statistical effects;
pitfalls in calculation of differences between overlapping selections
- separating out a large / distinctive systematic term: e.g. f^{00} in $B \rightarrow K\pi, \pi\pi$
- every conceivable effect? every likely/possible effect?
- “two point uncertainties” on complex models are a matter of expert judgement, esp. if large (*cf.* the small FF uncertainties mentioned above)
- combination of symmetric and asymmetric uncertainties produces a bias (*shifts the central value*) and has a nontrivial residual uncertainty, in general (example: Belle II pub31, PRL 131, 171803 (2023); D_s^+ lifetime)

Summary

Not covered: Frequentist & Bayesian Probability

Frequentist: I gotta tell you, he's not really my friend.
Saving his life is more a professional courtesy.



Frequentist: What is your job ... ?

Bayesian: Protecting your reality ... !