Measurement of Rb, Rc and the b-quark forward-backward asymmetry from archived ALEPH LEP1 data

Analysis note — final version

ALEPH archived-data heavy-flavour analysis

2026-06-10

Contents

Change Log

Final version (v1). Final prose polish and typesetting of the complete analysis note on the full archived ALEPH LEP1 dataset (2,889,543 hadronic Z). The physics content, methodology and headline numbers are carried forward unchanged from the full-data results; this version deepens the per-systematic prose, adds interpretive paragraphs after every results table and key figure, and adds three reader-facing elements: the published-comparison overlay as the core-result figure (Figure 29), a consolidated validation summary table (Table 14), and an explicit resolving-power statement. Headline results: $R_b = 0.21421 \pm 0.00273\,(\text{stat}) \pm 0.02268\,(\text{syst})$ (-0.09σ from the LEP+SLD world average), $A_{FB}^{0,b} = 0.07025 \pm 0.00850\,(\text{stat}) \pm 0.02388\,(\text{syst})$ (-1.14σ), $\sin^2\theta_{\text{eff}}^{\text{lept}} = 0.23691 \pm 0.00158\,(\text{stat}) \pm 0.00443\,(\text{syst})$ ($+1.14\sigma$), with R_c constrained to the world average 0.1721 ± 0.0030.

Full-data results. Updated all results to the complete archived dataset; rebuilt the $A_{FB}^{0,b}$ correction chain into an explicit, one-correction-at-a-time ladder (b-purity/composition unfold, beam-frame de-dilution and B-mixing, QCD, γZ/QED/ISR pole) with MC-truth closure of the full chain (recovers the injected asymmetry to $\leq$1% in the physical region); the $|\cos\theta|$ counting cross-check now agrees with the differential fit at -0.04σ. Carried the previously-omitted b-purity/composition systematic, growing the $A_{FB}^{0,b}$ systematic to a bottom-up 0.0239 (uds reco asymmetry 0.0141, beam-frame dilution 0.0137, opposite-sign charm 0.0098) — a genuine uncertainty, not an inflation, with the central value fixed independently by the closure. Added the full-data per-subperiod consistency and the fit-triviality gate

(PASS). The R_b method, shape calibration and systematic budget were carried forward unchanged from the validation stage; only the central value and statistical uncertainty changed with the full statistics.

Earlier development. Developed on expected (MC pseudo-data) results, validated on a fixed-seed 10% subsample of the archived data (the first controlled look, confirming the self-calibrating double tag and the data-driven shape calibration), then unblinded on the full dataset.

---

1 Introduction

1.1 Motivation and observables

The ratios of partial widths $R_b \equiv \Gamma(Z \to b\bar{b})/\Gamma(Z \to \text{had})$ and $R_c \equiv \Gamma(Z \to c\bar{c})/\Gamma(Z \to \text{had})$, together with the b-quark forward-backward pole asymmetry $A_{FB}^{0,b}$, are among the most informative electroweak observables of the hadronic Z sector. R_b is sensitive to the $Zb\bar{b}$ vertex and was the canonical probe of the mid-1990s "R_b crisis", where a multi-sigma pull above the Standard Model prediction was read as possible evidence for new physics coupling preferentially to the third generation. $A_{FB}^{0,b}$ is one of the most powerful single inputs to the effective leptonic weak mixing angle $\sin^2\theta_{\text{eff}}^{\text{lept}}$, and historically the largest single contributor to the long-standing tension between the leptonic and hadronic determinations of that angle (Schael et al. 2006; Navas et al. 2024).

For a $Z \to f\bar{f}$ decay the differential cross section in the polar angle θ of the outgoing fermion relative to the e^- beam is

$$\frac{d\sigma}{d\cos\theta} \propto (1+\cos^2\theta) + \tfrac{8}{3}\, A_{FB}^{f}\, \cos\theta, \tag{1}$$

so that the forward-backward asymmetry $A_{FB}^{f} = (\sigma_F - \sigma_B)/(\sigma_F + \sigma_B)$ appears as the coefficient of the odd term. This is the form fitted to the charge-signed $\cos\theta_{\text{thrust}}$ distribution of the b-tagged sample. The $(1+\cos^2\theta)$ term is the charge-symmetric (even) part, carrying the acceptance shape but no asymmetry information; all of the physics of interest lives in the $\cos\theta$ (odd) coefficient. This even/odd separation is the organizing principle of the asymmetry extraction (§The cos-theta acceptance map and the differential fit).

This analysis measures all three observables from public, archived ALEPH LEP1 data using a transparent columnar pipeline. R_b is obtained from a self-calibrating lifetime double tag; R_c is treated as a constrained external input (the lifetime tag has insufficient charm separation, §The R_c constraint); and $A_{FB}^{0,b}$ is obtained from the jet-charge-signed $\cos\theta_{\text{thrust}}$ asymmetry of the b-tagged sample, corrected to the pole through an explicit dilution chain and mapped to $\sin^2\theta_{\text{eff}}^{\text{lept}}$. The world averages serve as validation targets, not as inputs to the central values. This separation — measure blind, compare afterward — is what allows the agreement of R_b with the world average to be read as evidence that the method works, rather than as a tautology.

1.2 Conventions and the electroweak map

The partial-width ratios are normalized to the total hadronic width Γ_{had}, following the LEP electroweak working group convention (Schael et al. 2006); these must not be confused with the branching fractions $\Gamma_{b\bar{b}}/\Gamma_{\text{tot}}$ that the PDG also tabulates with a different normalization (Navas et al. 2024). The pole asymmetry $A_{FB}^{0,b}$ is the value after removing QED initial-state radiation, γ exchange and γ–Z interference and the $\sqrt{s}$ offset, evaluated at the effective Z-pole couplings. Throughout, ε_q denotes a single-hemisphere per-flavour tag efficiency, C_q a multiplicative hemisphere-correlation factor, f_s and f_d the single- and double-tag fractions, δ_b the beam-frame analyzing-power dilution, $\bar{\chi}$ the time-integrated B-mixing parameter, and P_q the tag composition fractions; these symbols are used consistently in every section.

The pole asymmetry maps to the mixing angle through the joint Standard-Model dependence

$$A_{FB}^{0,f} = \tfrac{3}{4}\, A_e\, A_f, \qquad A_f = \frac{2(g_V^f/g_A^f)}{1+(g_V^f/g_A^f)^2}, \qquad \frac{g_V^f}{g_A^f} = 1 - 4|Q_f|\sin^2\theta_{\text{eff}}^f, \tag{2}$$

with $|Q_b| = 1/3$ and $|Q_e| = 1$. A single measured number, $A_{FB}^{0,b}$, determines one quantity: both A_e and A_b in Equation 2 depend on the same angle, and the sensitivity is dominated by the A_e factor (A_b is nearly flat in $\sin^2\theta_{\rm eff}$). We therefore report $\sin^2\theta_{\rm eff}^{\rm lept}$ as the primary electroweak deliverable and quote A_b only as a dependent cross-check that requires an external world-average A_e. Setting $\sin^2\theta_{\rm eff} = 0.23153$ in Equation 2 gives $A_e = 0.1470$, $A_b = 0.9355$ and $A_{FB}^{0,b} = 0.103$ at tree level; the residual relative to the world average 0.0992 is absorbed by the QCD/radiation corrections and the electroweak-coupling anchoring described in §The asymmetry correction chain. Reporting a single mixing angle from a single measured asymmetry is the honest deliverable: any attempt to quote both A_b and the angle from one number would be circular, and the note is explicit about which quantity is measured and which is dependent.

1.3 Prior measurements and what is new

Two ALEPH lifetime double-tag measurements set the methodological benchmark: the lifetime-mass tag, $R_b = 0.2167 \pm 0.0011\,(\text{stat}) \pm 0.0013\,(\text{syst})$ (Barate et al. 1997a; Buskulic et al. 1993), and the five-tag measurement, $R_b = 0.2158 \pm 0.0009 \pm 0.0011$ (Barate et al. 1997b). Both are dominated by the hemisphere-correlation systematic C_b (about 84% of the total in the lifetime-mass analysis). ALEPH also measured the inclusive b asymmetry with a neural-network charge tag, $A_{FB}^{0,b} = 0.1009 \pm 0.0027 \pm 0.0015$ and $\sin^2\theta_{\rm eff} = 0.23193 \pm 0.00056$ (Heister et al. 2001). The LEP+SLD combination provides the validation targets $R_b^0 = 0.21629 \pm 0.00066$, $R_c^0 = 0.1721 \pm 0.0030$, $A_{FB}^{0,b} = 0.0992 \pm 0.0016$ and $\sin^2\theta_{\rm eff} = 0.23153 \pm 0.00016$ (Schael et al. 2006; Bodek et al. 2025). A SLD multi-tag vertex measurement gives a per-experiment R_c template, $R_c = 0.1744 \pm 0.0031 \pm 0.0020$ (Abe et al. 2005), and DELPHI provides an independent LEP double-tag R_b (Abreu et al. 1996). The jet-charge b-asymmetry method follows the LEP charge-flow tradition (Halley 1991; Abbaneo 1998).

What is new here is the combination of (i) fully public, archived ALEPH ntuples analyzed with a modern columnar pipeline; (ii) an honest treatment of the fact that these ntuples carry no truth flavour, resolved by generating a purpose-built truth-flavour MC with a parametric detector response validated to data; and (iii) a single, data-driven MC shape calibration that re-derives every MC-dependent correction from the data-matched simulation and propagates a measured, rather than assumed, systematic. For the asymmetry, the dilution corrections are applied explicitly and in the correct reference frame, with the b-purity/composition of the 66%-pure tag carried as a propagated systematic. The deliverable is therefore not a precision update — the world average is two orders of magnitude tighter — but a demonstration that an open-data, parametric-simulation pipeline can recover the established electroweak observables with a fully quantified, honest uncertainty budget.

---

2 Data samples

2.1 Data summary

ALEPH operated at LEP1 from 1992 to 1995 at centre-of-mass energies near the Z pole (Decamp et al. 1990). The archived "aftercut" MERGED ROOT files are public and have been loosely preselected; the analysis hadronic sample is selected by the `passesAll` flag (§Event selection). The post-selection hadronic count is $N_{\rm had} = 2{,}889{,}543$. Because the integrated luminosity is not published per file for these archived data, it is estimated from the hadronic event count and the ALEPH hadronic cross section as $\mathcal{L} = N_{\rm had}/\sigma_{\rm had}$ with $\sigma_{\rm had} = 30.4$ nb (Defranchis et al. 2026); this is approximate (peak cross section, no acceptance correction). The six periods and their content are summarized in Table 1, and the per-period $\sqrt{s}$ content is shown in Figure 1.

The total integrated luminosity reconstructed this way is about 96 pb^{-1}, consistent with the published ALEPH LEP1 sample size. Because R_b and R_c are ratios within the hadronic sample and $A_{FB}^{0,b}$ is an asymmetry, none of the central values depend on the absolute luminosity; the luminosity is quoted here only to characterize the dataset, and its approximate nature carries no weight on the results. This is a structural point: an analysis whose observables are intensity-independent ratios cannot be made circular by a back-derived luminosity, a pitfall the fit-triviality gate confirms is avoided (§Headline results).

Table 1: ALEPH LEP1 archived data samples. The aftercut count is the number of stored events (already loosely preselected); N_{had} is the sample passing the full hadronic selection (passesAll). The integrated luminosity is estimated as $\mathcal{L} = N_{\mathrm{had}}/\sigma_{\mathrm{had}}$ with $\sigma_{\mathrm{had}} = 30.4$ nb (Defranchis et al. 2026) and is approximate; precise per-energy luminosity is a residual limitation.

Period	$\sqrt{s}$ [GeV]	N aftercut	N_{had}	$\mathcal{L}$ [pb^{-1}]
1992	91.27 (peak)	551,474	522,526	~17
1993	89.4 / 91.2 / 93.0 scan	538,601	510,056	~17
1994 P1	91.14–91.23 (peak)	433,947	411,001	~14
1994 P2	91.17–91.22 (peak)	447,844	424,139	~14
1994 P3	91.14–91.42 (peak)	483,649	458,027	~15
1995	89.4 / 91.2 / 93.0 scan	595,095	563,794	~19
Total		**3,050,610**	**2,889,543**	**~96**

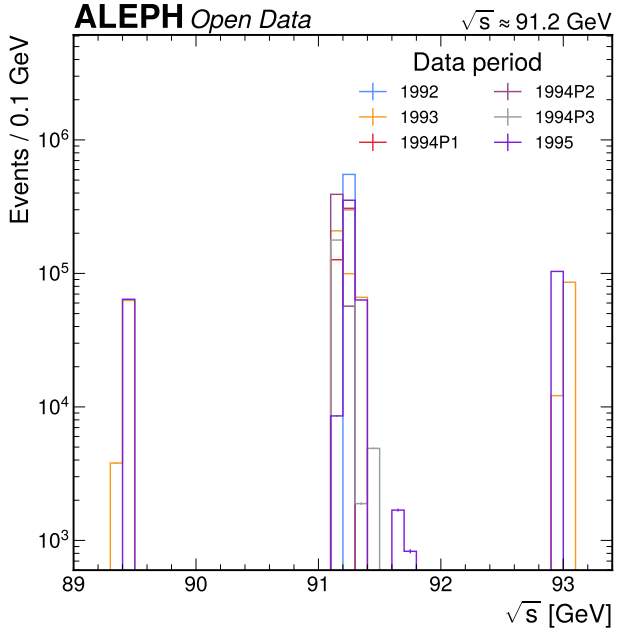

Figure 1: Per-data-period $\sqrt{s}$ content reconstructed from the Energy branch. The 1992 and 1994 periods are concentrated at the Z peak; 1993 and 1995 contain the off-peak Z-lineshape scan points at $\sqrt{s} \approx 89.4$ and 93.0 GeV, each contributing roughly 10–15% of the file. This fixes the energy treatment of A_{FB}^{b} (constraint [A2]): the asymmetry is reported at the luminosity-weighted peak, with the per-period consistency bounding the residual off-peak scan spread, whereas R_b and R_c are ratios within the hadronic sample and are insensitive to the small $\sqrt{s}$ spread.

2.2 The full dataset and the validation subsample

This note reports the final results on the complete archived dataset. The full sample is selected by the same `passesAll` chain on all six periods, giving $N_{\mathrm{had}} = 2{,}889{,}543$ hadronic Z, with 45,194,350 signed-significance tracks. The double-tag extraction uses $N_{\mathrm{single}} = 355{,}205$ single-tagged hemispheres and $N_{\mathrm{double}} = 21{,}139$ double-tagged events, and the asymmetry uses 331,721 b-tagged events in the $|\cos\theta| < 0.82$ fiducial. The full-data per-event observables are concatenated through the same code path as the MC, so every downstream data/MC diagnostic compares like with like.

The 10% validation subsample used at the earlier blinded stage (288,133 events, drawn by a deterministic per-period SplitMix64 mask, seed 20260609) is contained in the full sample; the independent 90% complement carries the new statistical information. Because the full sample contains the 10%, the consistency of the full and 10% results is assessed on the independent complement (§Consistency with the validation subsample). This staged unblinding — develop on MC, validate on 10%, unblind on the full set — is the discipline that lets the final agreement be read as a genuine cross-check rather than the product of iterative tuning.

The merged ntuples carry no per-event $\sqrt{s}$ (only the total reconstructed `Energy`), so the off-peak scan points in 1993 and 1995 cannot be separated per event. The committed approach is the γZ/QED/ISR pole correction to $A_{FB}^{0,b}$

applied at the luminosity-weighted peak, the standard pole convention; the residual off-peak dilution of the raw asymmetry is small (scan points are a few percent of the 1993/95 yield) and is bounded above by the per-period asymmetry consistency ($\chi^2/\mathrm{ndf} = 1.62$, §[Per-subperiod consistency]). R_b is energy-independent at the pole to the relevant level.

2.3 Monte Carlo samples

Two distinct Monte Carlo inputs enter the analysis (Table 2). The first is the archived ALEPH reconstructed MC (JETSET 7.4 + GALEPH), which was found during exploration to carry no truth-flavour information of any kind: the event-level `bFlag` placeholder is −999 in all 40 files, and the generator trees `tgen` and `tgenBefore` are stable-particle lists with no partons or B/D hadrons. This failure of the truth-flavour gate (limitation [L2]) forced decision [D8]: to generate a purpose-built truth-flavour sample with PYTHIA 8 Monash (Sjostrand et al. 2008; Skands et al. 2014) passed through a parametric ALEPH detector response. The generator flavour is read from the hard-process outgoing quark and provides the per-event true primary-quark label (b/c/s/uds). This generator flavour is lifetime-orthogonal — it is independent of the displaced-track signature the reco tag exploits — so it can calibrate the lifetime tag without the circularity that a displacement-based truth proxy would introduce.

Table 2: Monte Carlo samples. The archived reconstructed MC (Defranchis et al. 2026) lacks any truth-flavour label, which motivated the [D8] decision to generate a dedicated truth-flavour sample with PYTHIA 8.312, Monash 2013 tune (Sjostrand et al. 2008; Skands et al. 2014). The generated flavour fractions (b 0.214, c 0.171, s 0.227, u 0.190, d 0.199) reproduce the Z partial widths as a sanity check, not a tune.

Process	Generator	σ [nb]	N_{gen}	Truth flavour	Notes
$Z \to q\bar{q}$ (archived reco)	JETSET 7.4 + GALEPH	30.4	0.77M reco	none (bFlag=−999)	single year (1994), single tune
$Z \to q\bar{q}$ (own [D8])	PYTHIA 8.312 Monash 2013	30.4	2.40M	per-event primary quark (b/c/s/uds)	parametric ALEPH detector response

The own-MC sample comprises 2.40M events generated in 24 reproducible shards, with γ^*/Z production restricted to quark final states, ISR/FSR and hadronization enabled, at $\sqrt{s} = 91.2$ GeV. The per-event primary-quark flavour is read from the hard-process outgoing quark. The construction of the parametric detector response and its data-driven validation are described in §Corrections. The MC covers the 1994 detector configuration only; it is applied to all six periods, and the per-period consistency bounds the resulting coverage systematic ([A1], §[Per-subperiod consistency]). That the generated flavour fractions reproduce the known Z partial widths is a sanity check on the generator setup, not a tune to the measured observables — no R_b, R_c or A_{FB} value enters the generation.

---

3 Event selection

3.1 Object and event definitions

The analysis uses ALEPH energy-flow charged tracks. A quality charged track has energy-flow class `pwflag = 0`, nonzero charge, the `highPurity` flag, at least one VDET hit (nvdet ≥ 1), and $|d_0| < 2$ cm. Each event is split into two hemispheres by the sign of $\cos\theta_{\mathrm{wrtThr}}$, the track's polar angle relative to the thrust axis. This single hemisphere split serves both analyses: it defines the two independent tags of the double-tag R_b measurement, and it provides the forward/backward charge basis for A_{FB}^b through the per-hemisphere jet charge. Using one hemisphere definition for both measurements keeps the two analyses on a common geometric footing and makes the correlation structure between them well defined for the combined covariance (§Systematic uncertainties).

3.2 Preselection cutflow

The hadronic Z preselection is the standard ALEPH chain, applied through the `passesAll` flag, which combines the S_θ polar-acceptance, ISR-rejection and WW-rejection requirements. The data and the parametric MC agree at every stage of the chain to better than 0.1% in the surviving fraction (Table 3), confirming that the inclusive event selection is well modelled. The event-shape inputs that drive the selection — charged multiplicity and thrust — are shown in Figure 2 and Figure 2.

Table 3: Hadronic preselection cutflow, data vs MC surviving fraction. The chain agrees to better than 0.1% at every stage; the final passesAll fraction (0.9464) defines the hadronic sample.

Stage	Data surviving frac.	MC surviving frac.
passesSTheta	0.978	0.978
passesISR	0.9464	0.9465
passesWW	0.9464	0.9465
passesAll	**0.9464**	**0.9465**

The cutflow agreement is the first quantitative evidence that the inclusive simulation is sound: the surviving fractions track the data to the third decimal place at each stage, so the selection itself introduces no measurable data/MC bias and the residual modelling burden falls entirely on the flavour-dependent tag, not on the inclusive event selection.

Figure 2: **(a)** Charged-track multiplicity, data vs MC. The distributions peak near 20 charged tracks and agree to within about 5% across the bulk, confirming that the fragmentation and tracking model entering the selection is well reproduced. A residual mean shift (data/MC scale factor 0.930) is carried as a control-observable handle (§Supporting inclusive control observables). **(b)** Thrust distribution, data vs MC. The event-shape variable peaks near $T \approx 0.97$; the data/MC agreement (scale factor 1.010) confirms the topology entering the hemisphere split and the polar acceptance is well modelled.

3.3 Polar acceptance and the asymmetry fiducial

The forward-backward asymmetry is measured within the polar fiducial $|\cos\theta_{\text{thrust}}| < 0.82$, a sharp truncation set by the tracking and trigger acceptance (Figure 3). Within the fiducial the acceptance is not flat: it falls toward the edge to roughly 0.68 of the central value, an effect the parametric MC does not fully reproduce. Because this distortion bears directly on the angular distribution that defines the asymmetry, it is corrected with a data-driven acceptance map rather than absorbed into the MC; the construction is described in §The cos-theta acceptance map and the differential fit. Correcting the acceptance from the data themselves, rather than trusting the parametric MC to reproduce a sharp edge it was never tuned to, is the conservative choice and removes the acceptance shape as a source of MC dependence in the asymmetry.

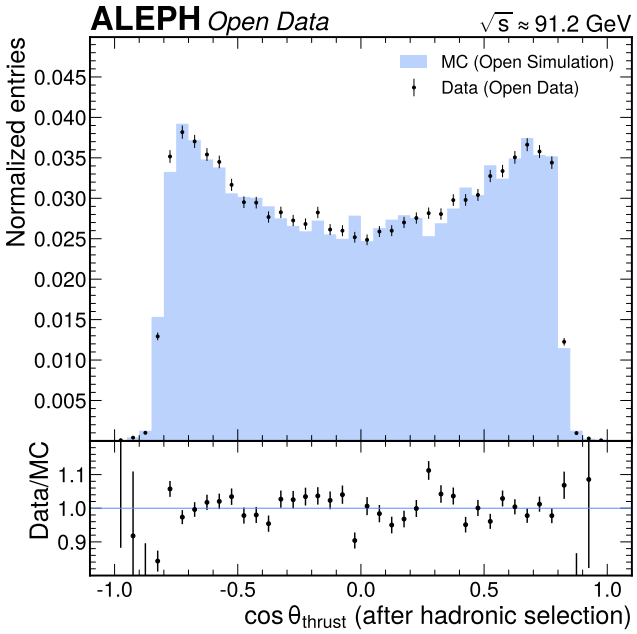

Figure 3: Charge-signed $\cos\theta_{\mathrm{thrust}}$ of the b-tagged sample, showing the sharp fiducial truncation at $|\cos\theta| \approx 0.82$. The in-fiducial acceptance is non-flat, falling to about 0.68 of the central value at the edge; this motivates the data-driven acceptance map of §The cos-theta acceptance map and the differential fit.

---

4 Corrections

This section is the methodological spine of the analysis. The double-tag extraction (§The double-tag extraction and its self-calibration) fixes what must be measured from the simulation: the non-b tag efficiencies, the hemisphere correlations, and the charge-tag dilution. Because the archived MC has no truth flavour, those quantities are derived from a generated truth-flavour MC with a parametric detector response (§The parametric detector model and its validation), whose distributions are matched to the data by a single data-driven shape calibration (§The genuine MC shape calibration). The inclusive control observables (§Supporting inclusive control observables) supply two load-bearing handles that size the leading R_b systematics. The lifetime tag and its working point (§The lifetime b-tag and working point), the acceptance map and differential fit (§The cos-theta acceptance map and the differential fit), the explicit asymmetry correction chain (§The asymmetry correction chain), and the closure and stress tests (§Closure tests, §Stress test) complete the procedure. The reader who follows the chain from the raw tagged distribution to the final pole asymmetry will encounter each correction exactly once, in the order it is applied; the per-step values are collected in Table 5.

4.1 The double-tag extraction and its self-calibration

Each event is split into two thrust hemispheres and the b-tag is applied independently to each. Let f_s be the single-hemisphere tag fraction and f_d the double-tag fraction. With per-flavour single-hemisphere efficiencies $\varepsilon_b, \varepsilon_c, \varepsilon_{uds}$ and flavour fractions $R_b, R_c, R_{uds} = 1 - R_b - R_c$,

$$f_s = \varepsilon_b R_b + \varepsilon_c R_c + \varepsilon_{uds} R_{uds}, \tag{3}$$

$$f_d = C_b\, \varepsilon_b^2 R_b + C_c\, \varepsilon_c^2 R_c + C_{uds}\, \varepsilon_{uds}^2 R_{uds}, \tag{4}$$

where $C_q \equiv \varepsilon_{qq}/\varepsilon_q^2$ is the multiplicative hemisphere correlation factor for flavour q, equal to one if the two hemisphere tags are statistically independent. We use the multiplicative convention throughout and never the additive $\varepsilon_{qq} = (1+\delta_q)\varepsilon_q^2$ form; mixing the two is the single most common normalization trap in R_b analyses. After subtracting the non-b terms, the pure-b system inverts to

$$\boxed{R_b = \frac{f_s^2}{C_b\, f_d}, \qquad \varepsilon_b = \frac{f_d}{C_b\, f_s}} \tag{5}$$

so that ε_b is measured from the data themselves and R_b depends on the simulation only through the small correlation factor C_b and the non-b terms. The f_s, f_d entering Equation 5 are the non-b-subtracted pure-b single- and double-tag fractions, not the raw measured fractions quoted below; the boxed form is the pedagogical pure-b limit, whereas the headline $R_b = 0.21421$ and $\varepsilon_b = 0.17876$ are obtained from the full three-flavour solve of Equation 3–Equation 4 (which carries the charm and light terms explicitly). Substituting the raw f_s, f_d directly into the boxed equation therefore does not reproduce the headline values — the non-b subtraction must be applied first. The fractional transfer of C_b onto R_b is one-to-one, $\partial \ln R_b / \partial \ln C_b = -1$, which is why C_b is the dominant correlation systematic in every double-tag R_b measurement. The self-calibration is the central virtue of the method: the largest single uncertainty in a generic b-tagging measurement, the absolute b-tag efficiency, is here measured from the ratio f_s^2/f_d rather than taken from simulation, so the simulation need only supply the much smaller non-b efficiencies and the correlation.

At the working point (ε_b self-calibrated from the data), the non-b efficiencies and correlations used in the solve are $\varepsilon_c = 0.04374$, $\varepsilon_{uds} = 0.02549$, $C_b = 0.972$ (shape-calibrated nominal; the 24-shard analysis-MC value 0.9634 is a cross-check), $C_c = 0.8509$, $C_{uds} = 0.9630$, with $R_c = 0.1721$ constrained. On the full data the measured fractions are $f_s = 0.061464$ and $f_d = 0.0073157$ (from $N_{\text{single}} = 355{,}205$ and $N_{\text{double}} = 21{,}139$ tagged hemispheres/events), giving the self-calibrated $\varepsilon_b = 0.17876$. The logical flow of the double tag is shown schematically in Figure 4, where the single- and double-tag fractions are scanned against the b-tag cut and annotated with the extracted (R_b, ε_b).

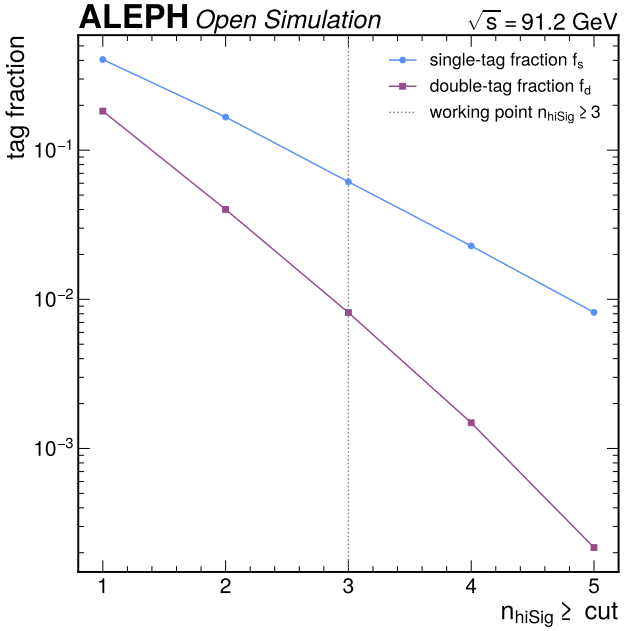

Figure 4: Single-tag fraction f_s and double-tag fraction f_d scanned against the b-tag cut, with the self-calibrated extraction overlaid. The working-point values $f_s = 0.0615$ and $f_d = 0.00732$ feed Equation 5; the figure shows that the closed-form solution is stable across the cut, the signature of the self-calibrating method.

The stability of the extracted R_b as the cut is scanned (Figure 4) is the operational confirmation that Equation 5 is doing what it should: a tag that was mis-modelled would produce a cut-dependent R_b, whereas the self-calibration holds it flat because ε_b moves with the cut and cancels in the ratio.

4.2 The R_c constraint

The lifetime double tag separates b from non-b cleanly but has very weak charm separation: the charm-versus-rest classifier AUC is only 0.567 (§BDT b-tag training record), far below the b-versus-rest AUC of 0.787. A simultaneous extraction of R_b and R_c from the lifetime tag alone would therefore be nearly degenerate in the charm direction, with R_c essentially unconstrained by the data. Decision [D6] is to constrain R_c to the world-average value $R_c =$

0.1721 ± 0.0030 (Schael et al. 2006; Abe et al. 2005) and propagate that uncertainty into R_b through the EWWG R_b–R_c anti-correlation $\rho = -0.18$ rather than as an independent error term. This is a documented, deliberate downscope of the charm sector, not a silent assumption: the charm efficiency ε_c still enters the solve and its uncertainty is carried, but the charm fraction itself is taken from external knowledge. The resulting R_c contribution to the R_b uncertainty is small ($\delta R_b = 0.00052$, §R_b: MC-statistics terms and the R_c constraint).

4.3 The parametric detector model and its validation

The generated PYTHIA 8 events are passed through a parametric ALEPH detector response. Charged-track impact parameters are smeared with the measured resolution $\sigma(d_0) = 25 \oplus 95/p$ µm, momentum-binned to separate the momentum-dependent core resolution from a momentum-independent primary-vertex and beamspot displacement. The free parameters of the model are an overall resolution scale (res_scale 1.25), a tail fraction and tail scale (f_tail 0.20, tail_scale 5.0), and a primary-vertex smearing scale (pv_scale 0.75). The same observables are built on the generated MC as on the data: the lifetime-signed impact-parameter significance S, the thrust axis and hemispheres, and the jet charge Q_{hem}.

The model obeys a strict anti-circularity contract. Its parameters are fixed only on inclusive event and track control distributions and on the lifetime-blind negative-significance tail (tracks tagged on the "wrong", non-lifetime side, which measure the resolution without any real-b contribution). The calibration never touches the tagged R_b or A_{FB} outputs, and no parameter is tuned to a reference R_b, R_c or A_{FB}. The truth flavour comes from the generator; the efficiencies and C_b come from the data-validated MC. This contract is what distinguishes a calibration from a tune: a tune adjusts the simulation until it reproduces the measured result, which would make the agreement meaningless, whereas this calibration adjusts the simulation until it reproduces lifetime-blind control distributions that carry no information about the asymmetry or the b fraction, leaving the tagged outputs as genuine predictions.

4.4 The genuine MC shape calibration

The single calibration of this analysis is a data-driven shape calibration of the generated MC. Two observables are calibrated: the signed-significance distribution that defines the lifetime tag (for R_b) and the jet-charge marginal that defines the analyzing power (for A_{FB}). In both cases the data/MC difference is modelled by a single physical mechanism, the corrected MC is validated against an independent quantity, and every downstream correction is re-derived from the calibrated truth-labelled MC. A naive per-bin reweight of the marginal — the obvious but wrong alternative — is shown to fail an independent cross-check, which is why the shape calibration, not the marginal reweight, is the method. The shape calibration is fixed at the validation stage and applied unchanged on the full data; the full-data observed distributions confirm the calibration transfers.

4.4.1 Signed-significance shape calibration

For the lifetime tag, a per-bin reweight $w(S)$ is derived from the data/MC ratio of the inclusive lifetime-signed significance S and applied to the generated MC tracks as a per-track keep weight. The per-flavour efficiencies and the correlation are then re-derived from the reweighted, truth-labelled MC:

$$w(S) = \frac{(dN/dS)_{\mathrm{data}}}{(dN/dS)_{\mathrm{MC}}}, \qquad \varepsilon_q = \frac{\sum_{\text{tagged } q\text{-hemispheres}} \prod_{\text{tracks}} w(S)}{\sum_{q\text{-hemispheres}} \prod_{\text{tracks}} w(S)}, \tag{6}$$

evaluated on hemispheres with MC truth flavour q. The MC does not cancel in this construction: the per-flavour efficiencies inherit the truth content of the simulation under a flavour-blind reweight that depends only on S. The reweight is a track-level reconstruction effect (resolution, primary-vertex smearing, V0/fake content), justifying its flavour-independence; the residual flavour dependence is varied (inclusive versus negative-anchored) and carried as a systematic.

The result is that the non-b fake efficiencies rise toward the data while the b efficiency is unchanged: ε_c moves from 0.0369 to 0.0437 (per-flavour fake scale factor 1.20), ε_{uds} from 0.0220 to 0.0255 (scale factor 1.17), and ε_b stays at 0.1910. The b efficiency is unchanged by construction: a b hemisphere's deep-significance tag is genuine B-decay lifetime, not a resolution effect, so a lifetime-blind $w(S)$ must not — and does not — touch it. The correlation re-derived from the reweighted b hemispheres is $C_b = 0.972 \pm 0.013$, consistent with the 24-shard analysis-MC cross-check 0.9634.

The calibration is validated two ways. First, the reweighted MC reproduces the data deep-tail shape: the data/MC χ^2/ndf over $S > 8$ falls from 1167 to 1.5 and the KS distance from 0.114 to 0.002 (Figure 6). Second, a genuine MC-truth closure (inject a known R_b, build pseudo-data from the calibrated efficiencies and correlation, re-extract with the same inputs) recovers the injected value with a maximum absolute bias of 3×10^{-14} across $R_b \in [0.205, 0.230]$ (Figure 7). The reweight itself is shown in Figure 5. The drop in the deep-tail χ^2/ndf from 1167 to 1.5 is the single most important number in the R_b calibration: it shows that the calibrated MC reproduces the precise data distribution that the tag cuts on, so the non-b fake efficiencies that the calibration produces are themselves trustworthy.

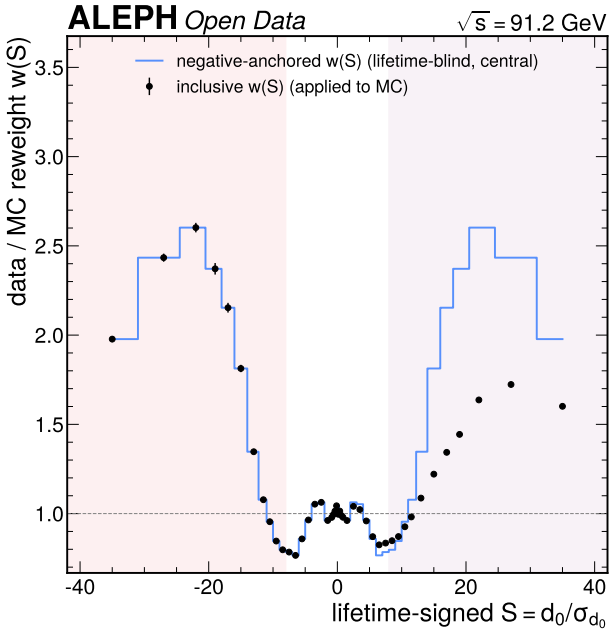

Figure 5: Signed-significance per-bin reweight $w(S)$ derived from the inclusive data/MC ratio, with the per-flavour fake scale factors (c 1.20, uds 1.17) that it induces on the non-b efficiencies. The reweight is largest in the deep tails where the parametric resolution under-models the data, and near unity in the well-modelled core.

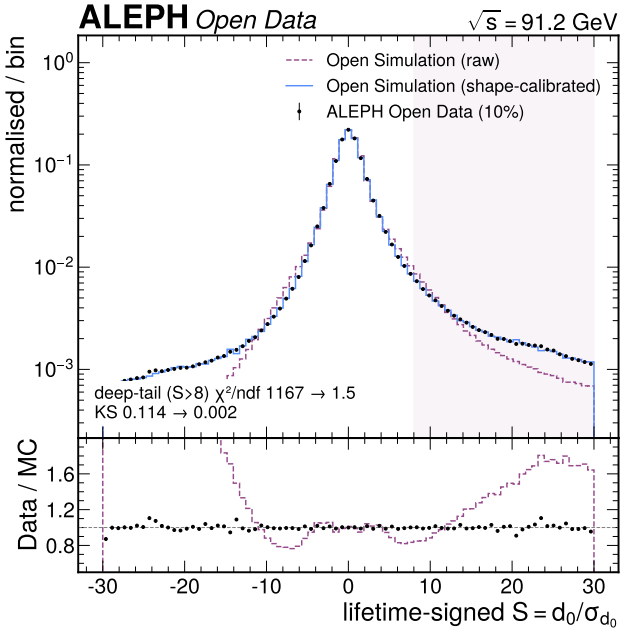

Figure 6: Calibrated MC versus data overlay of the signed-significance distribution. After the $w(S)$ reweight, the deep-tail data/MC χ^2/ndf falls from 1167 to 1.5 and the KS distance from 0.114 to 0.002. This overlay is the central validation that the shape calibration matches the data distribution that drives the tag.

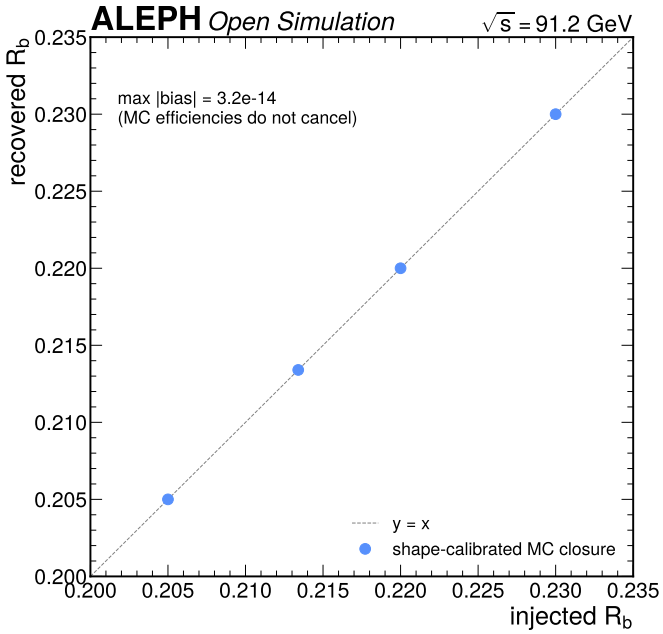

Figure 7: MC-truth closure of the signed-significance calibration. Injecting a known R_b, building pseudo-data from the calibrated efficiencies/correlation, and re-extracting recovers the input with a maximum absolute bias of 3×10^{-14} across the tested range — a genuine closure in which the MC efficiencies do not cancel.

4.4.2 Jet-charge shape calibration

For the asymmetry, the b-tagged per-hemisphere jet charge $Q_{\mathrm{hem}}(\kappa = 0.5)$ is 21.5% wider in the data than in the nominal MC (Figure 8); the analyzing-power exponent $\kappa = 0.5$ is the working point of the dilution-versus-κ scan (Figure 9). This is modelled as an extra per-hemisphere Gaussian charge-resolution smear of width σ_{shape}, fixed by minimizing the χ^2 of the full Q_{hem} marginal shape (data versus smeared MC, all bins), not by matching a single width scalar. The best-fit smear is $\sigma_{\mathrm{shape}} = 0.138$ (Figure 10). The analyzing power ω and the per-hemisphere dilution δ_{eff} are then re-derived from the shape-calibrated MC truth through

$$\delta_{\mathrm{eff}} = (2\omega - 1)(1 - 2\bar{\chi}), \tag{7}$$

with $\bar{\chi} = 0.1257 \pm 0.0042$ the time-integrated B-mixing parameter (Schneider 2019). The jet-charge marginal calibration is a cross-check input; the asymmetry chain (§The asymmetry correction chain) uses the frame-correct beam-frame analyzing power, not the thrust-axis marginal, as the production dilution. The reason a full-shape smear is required, rather than a per-bin marginal reweight, is a physics distinction that only the joint two-hemisphere structure resolves: the data's extra width is charge-resolution smearing that mis-signs $b \leftrightarrow \bar{b}$ and lowers the same-sign/opposite-sign fraction f_{OS}, whereas a naive marginal reweight up-weights the cleanly-signed large-$|Q|$ hemispheres and drives f_{OS} the wrong way to a $+14\sigma$ disagreement. The smear model passes the forward-backward-blind f_{OS} cross-check: the shape-calibrated MC predicts $f_{OS} = 0.5424$ against the data 0.5393, a $+1.10\sigma$ agreement (Figure 11). This f_{OS} check is independent of the asymmetry — it uses no $\cos\theta$ and is charge-symmetric — so it validates the charge resolution without leaking A_{FB}. That the physically-motivated smear passes the f_{OS} check at $+1.10\sigma$ while the naive reweight fails it at $+14\sigma$ is the decisive evidence that the smear, not the reweight, captures the true mechanism behind the data/MC width difference.

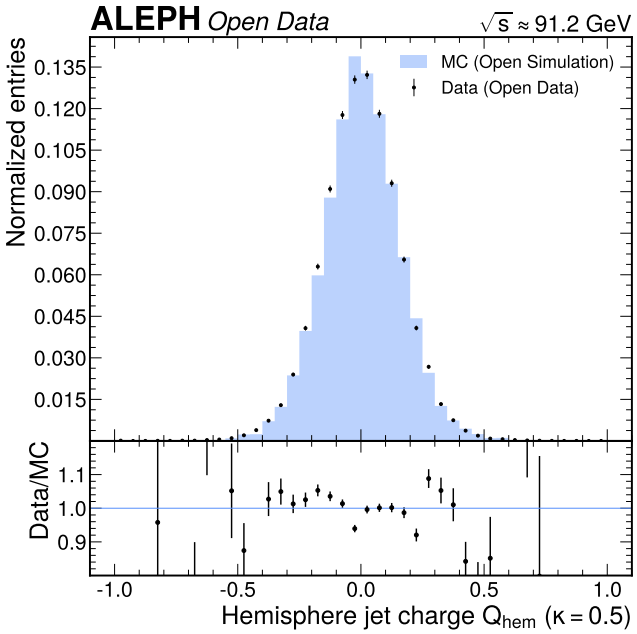

Figure 8: Hemisphere jet charge $Q_{\mathrm{hem}}(\kappa = 0.5)$, data vs MC. The b-tagged data distribution is 21.5% wider than the nominal MC, the mismodelling that the shape calibration corrects through the resolution smear.

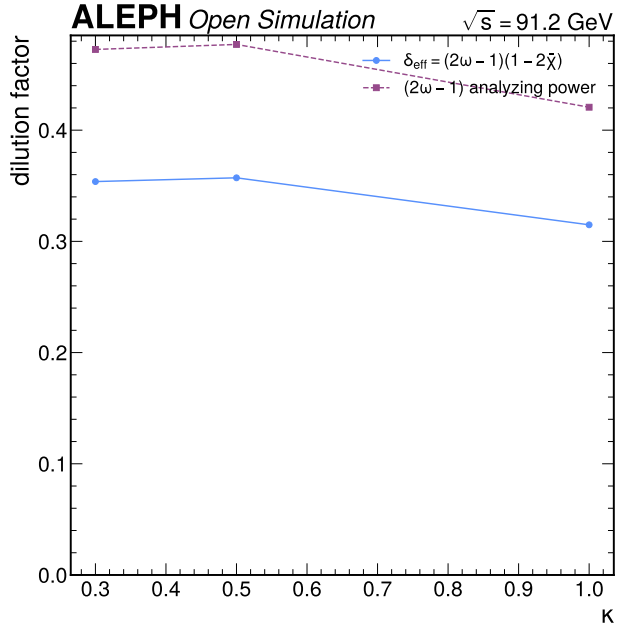

Figure 9: Effective dilution δ_{eff} and analyzing power ω versus the jet-charge exponent κ. The working point $\kappa = 0.5$ maximizes the effective statistical power $(2\omega - 1)^2 N$, setting the charge-tag definition used throughout.

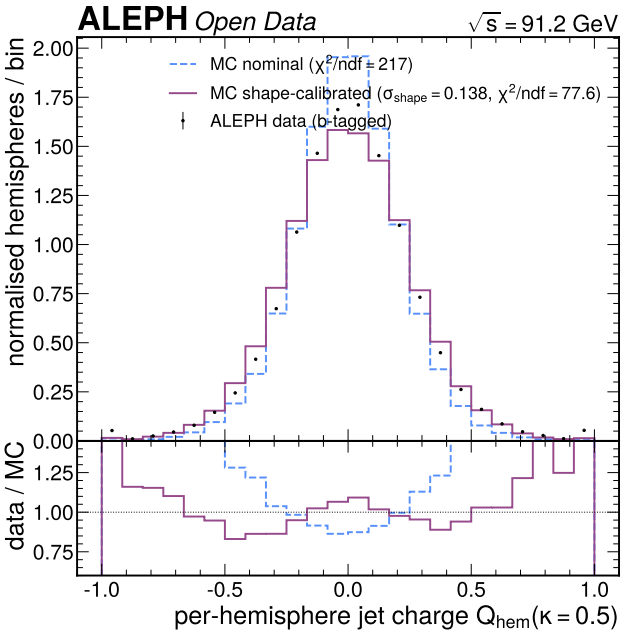

Figure 10: Full marginal Q_{hem} shape fit. The b-tagged jet charge is 21.5% wider in the data than in the nominal MC; a single per-hemisphere Gaussian charge-resolution smear $\sigma_{\mathrm{shape}} = 0.138$, fixed to the full marginal, brings the calibrated MC onto the data shape. The smear, not a width scalar, is the calibration.

Figure 11: **(a)** Analyzing power ω and dilution δ_{eff} re-derived from the shape-calibrated MC truth. The smear lowers ω to 0.690, consistent with the scalar width-calibration estimate at the 0.5σ level but with a measured, tighter uncertainty. **(b)** Forward-backward-blind f_{OS} cross-check. The shape-calibrated MC predicts the same-sign/opposite-sign fraction $f_{OS} = 0.5424$ against the data 0.5393 ($+1.10\sigma$). A naive marginal reweight instead fails this check at $+14\sigma$, demonstrating that the full-shape smear is the correct tool.

4.5 Supporting inclusive control observables

The shape calibration is supported by a set of thirteen inclusive data/MC control observables, each summarized by a scale factor (Table 4). These are not a parallel calibration; they are the evidence base that sizes two of the leading R_b systematics and confirms that the inclusive simulation is sound. The full per-observable validation grid (multiplicity, track momenta and angle, thrust, sphericity, jet-charge width, and the three tag inputs) is collected in Appendix C; here we highlight the two load-bearing rows.

The first is the impact-parameter resolution core. The lifetime-blind negative-significance core ($|S| < 3$, $S < 0$) gives a data/MC width scale factor of 1.020 ± 0.001 (Figure 12). This pins the primary-vertex smearing of the parametric model and collapses the C_b systematic: the pv_scale prior tightens from $\pm 50\%$ to $\pm 2.0\%$, shrinking the C_b model uncertainty dC_b from 0.066 to 0.013 and the corresponding δR_b from 0.0191 to 0.0039 (§R_b: hemisphere-correlation model).

The second is the deep-tail fake asymmetry. The lifetime-blind negative deep tail ($S < -8$) has a data/MC scale factor of 1.183 ± 0.003 (pure resolution/fake), while the positive deep shoulder ($S > 8$, signal plus fake) has a scale

factor of 1.076 ± 0.002. An exact positive/negative mirror would give equal scale factors; their roughly 9% relative difference is a conservative upper bound on the unmodelled positive-only fake asymmetry, and the nominal 7% used to size δR_b is the midpoint of the defensible 6–9% band (the positive shoulder is partly genuine-b lifetime, which masks the true fake-only asymmetry below the 9% bound). This measured asymmetry is what sizes the leading R_b systematic (§R_b: deep-tail non-b fake residual).

Table 4: Inclusive data/MC control observables, each as a scale factor with the KS distance. The two load-bearing rows are the IP-resolution core (collapses the C_b systematic) and the deep-tail/positive-shoulder pair (sizes the deep-tail fake systematic). The remaining rows are cross-checks; the full grid is in Appendix C. The χ^2/ndf values are inflated by the very large track statistics and are not used as a GoF threshold; the KS distance is the decision metric.

Observable	kind	data/MC SF	KS
IP-resolution core	width	1.020 ± 0.001	0.010
IP deep tail ($S < -8$)	frac	1.183 ± 0.003	—
IP pos. shoulder ($S > 8$)	frac	1.076 ± 0.002	—
N_{ch}	mean	0.930	0.091
track p	mean	0.912	0.064
track p_T	mean	0.928	0.055
track $\cos\theta$	shape	0.947	0.029
thrust	mean	1.010	0.109
sphericity	mean	1.097	0.051
Q_{hem} width	width	1.215	0.037
nhiSig	mean	1.022	0.016
sumPosSig	mean	0.724	0.036
maxSig	mean	0.688	0.076

These thirteen rows do double duty: the two load-bearing handles directly size the two leading R_b systematics, while the remaining eleven confirm that the inclusive event and track simulation is sound across multiplicity, momentum, angle and event shape. The pattern — excellent agreement in the bulk distributions, with the residual modelling burden concentrated in the deep-tail fake content — is exactly what one expects from a parametric (rather than full GEANT) detector model, and it points directly to where the dominant systematic comes from.

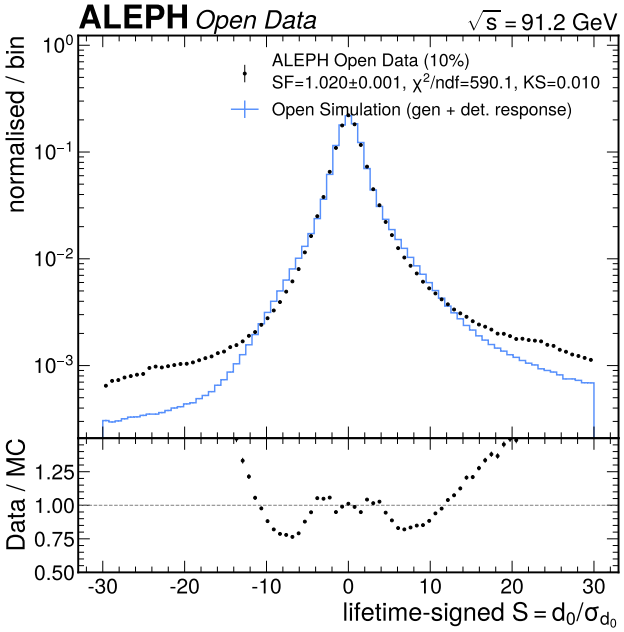

Figure 12: Impact-parameter resolution core and tail, data vs MC. The lifetime-blind negative core gives a width scale factor 1.020 ± 0.001, which pins the primary-vertex smearing and collapses the C_b systematic; the negative deep tail (1.183) and positive shoulder (1.076) size the deep-tail fake asymmetry. This single figure carries the two load-bearing control handles of the R_b measurement.

4.6 The lifetime b-tag and working point

The primary tag is a cut-based track-counting tag: a hemisphere is b-tagged if it contains at least three tracks (nhiSig ≥ 3) with lifetime-signed impact-parameter significance $S > 8$. The cut-based tag is chosen as primary for transparency and parity with the published ALEPH track-counting method, and because it keeps C_b demonstrably in gate; a BDT tag is trained and quantified as a cross-check and upgrade (§BDT b-tag training record). The tag separates the flavours cleanly in signed significance, and the working point is set by the purity/efficiency and C_b-stability scan.

At the working point the cut-based tag gives $\varepsilon_b = 0.1910$ (MC truth), hemisphere b-purity 0.676, and $C_b = 0.972$ (shape-calibrated nominal, in gate); the ROC, purity-efficiency curve, C_b-versus-cut scan, and flavour-split significance are shown in Figure 13, Figure 13, Figure 13 and Figure 13. The working point is a deliberate compromise: pushing the cut harder would raise the purity and shrink the non-b fake systematic, but at the cost of statistical reach and — for the BDT — of C_b leaving its gate, so the chosen point balances purity against the self-calibrating statistical power while keeping C_b controlled.

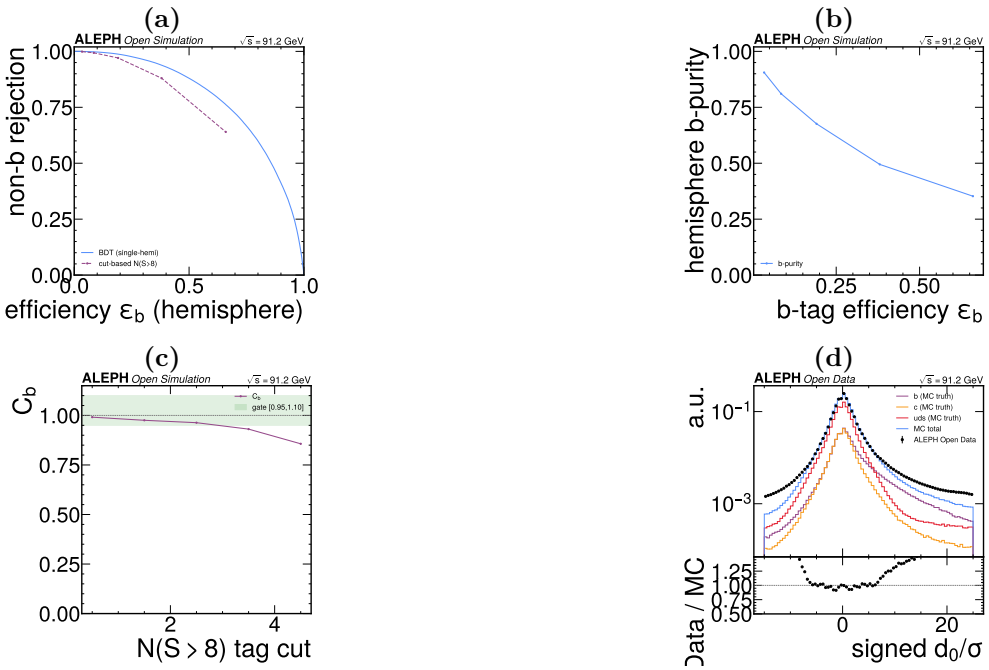

Figure 13: **(a)** Lifetime-tag ROC for the cut-based and BDT taggers (AUC 0.787). The cut-based tag is the transparent primary; the BDT is the quantified upgrade. **(b)** Hemisphere b-purity versus b-tag efficiency ε_b for the cut-based tag. The chosen working point ($\varepsilon_b = 0.1910$, purity 0.676) balances purity against the self-calibrating statistical reach. **(c)** Hemisphere correlation C_b versus the b-tag cut. $C_b = 0.972$ at the working point (shape-calibrated nominal), stable across the scan, with the 24-shard analysis-MC value 0.9634 as a cross-check, demonstrably inside the physically motivated gate. **(d)** Signed impact-parameter significance split by true flavour (b/c/uds). The b distribution carries the long positive-significance lifetime tail that the tag exploits; the c and uds distributions are concentrated near zero with only resolution/fake tails.

The operating-point stability of the extracted R_b across the tag cut is shown explicitly in Figure 14: the extracted value is flat within its statistical band over the scan window, the operational signature that the self-calibration absorbs the cut-dependence of the efficiency. A measurement whose central value drifted with the cut would signal a mis-modelled tag; the flatness here is direct evidence that it is not, and is the operating-point-stability check required of an extraction measurement.

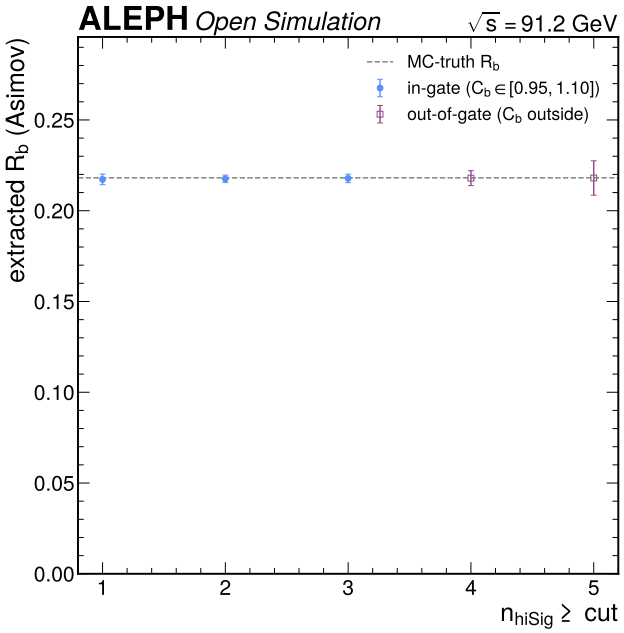

Figure 14: Operating-point stability of the extracted R_b versus the b-tag cut. The self-calibrated R_b is flat within its statistical band across the scan, confirming that the closed-form extraction (Equation 5) absorbs the cut-dependence of ε_b and is not sensitive to the precise working point.

4.6.1 BDT b-tag training record

A single-hemisphere gradient-boosted decision tree (xgboost, binary:logistic, max depth 4, $\eta = 0.1$, 200 rounds, seed 12345 (Chen and Guestrin 2016)) is trained on four per-hemisphere features (sumPosSig, nhiSig, maxSig, ntrk) with per-hemisphere truth flavour from the generated MC; an alternative HistGradientBoosting model (Pedregosa et al. 2011) agrees to within 0.0009 in AUC. The features are single-hemisphere only, by design, to protect C_b from cross-hemisphere correlations. The classifier reaches a test AUC of 0.787 (charm-versus-rest AUC 0.567, confirming the weak charm separation that motivates the R_c constraint). Overtraining is negligible: the train/test AUC gap is -0.0001 and the train/test score KS p-values are 0.485 (b) and 0.766 (non-b). The score is well calibrated (expected calibration error 0.005) and the data/MC score agreement is good (KS distance 0.009). The full set of training diagnostics is shown in Figure 15 through Figure 16.

The BDT is not adopted as primary. At matched ε_b it raises hemisphere b-purity from 0.676 to 0.810 and would shrink the leading non-b fake systematic by about 53%, but its C_b at that working point is 0.918, outside the [0.95, 1.10] gate. Adoption would therefore require a dedicated C_b re-evaluation, which is deferred to a future iteration. The cut-based tag remains primary for transparency, method parity, and in-gate C_b. The BDT therefore serves a specific purpose here: it quantifies what a higher-purity tag would buy (a roughly halved fake systematic) and so makes concrete the leading route to a more precise R_b, without being adopted before its correlation systematic is brought under control.

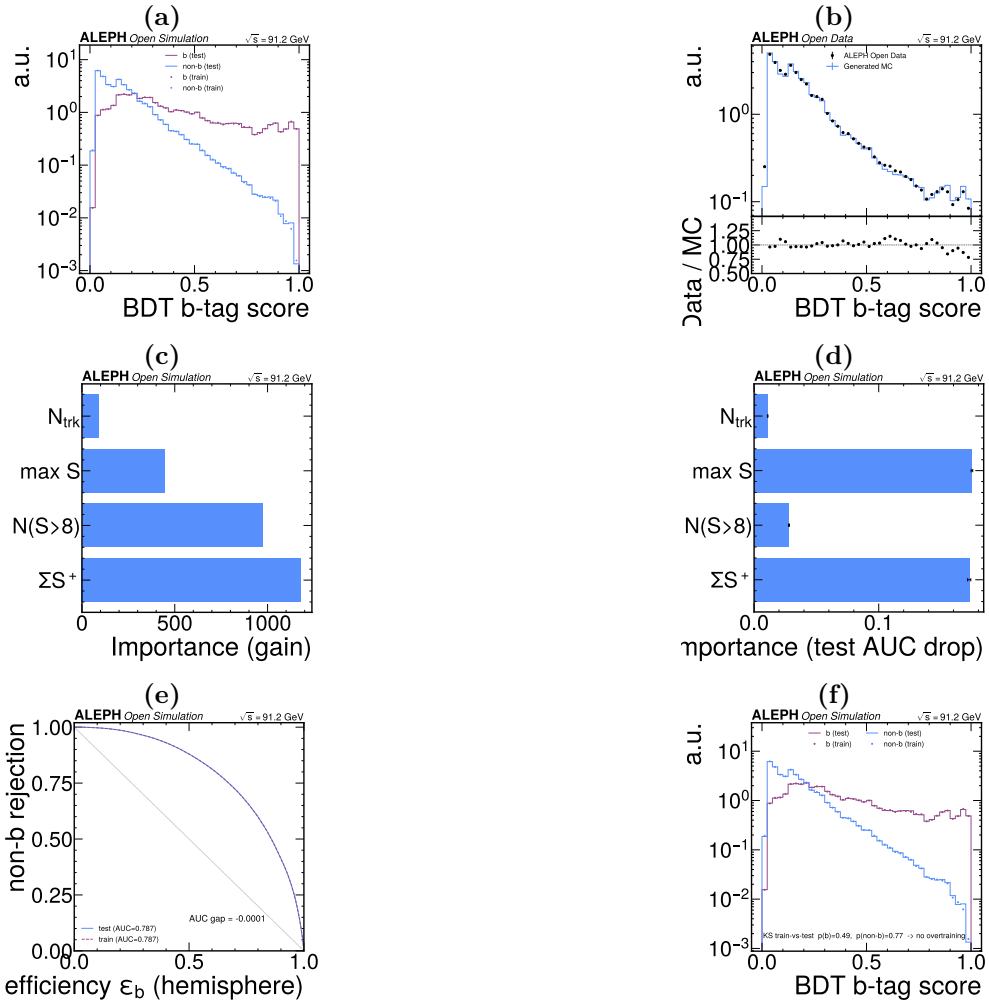

Figure 15: **(a)** BDT score distribution by class (train/test overlaid). The b and non-b score distributions are well separated, with no visible train/test divergence. **(b)** BDT score, data vs MC. The data/MC agreement (KS distance 0.009) confirms the classifier output transfers between simulation and data. **(c)** BDT gain feature importance. sumPosSig and nhiSig dominate the gain ranking. **(d)** BDT permutation feature importance. sumPosSig and maxSig dominate the permutation ranking, complementing the gain view. **(e)** Train/test ROC overlay (overtraining check). The AUC gap of -0.0001 shows no overtraining. **(f)** Train/test score overlay with KS p-values (0.485 b, 0.766 non-b). The agreement confirms the model is not overtrained.

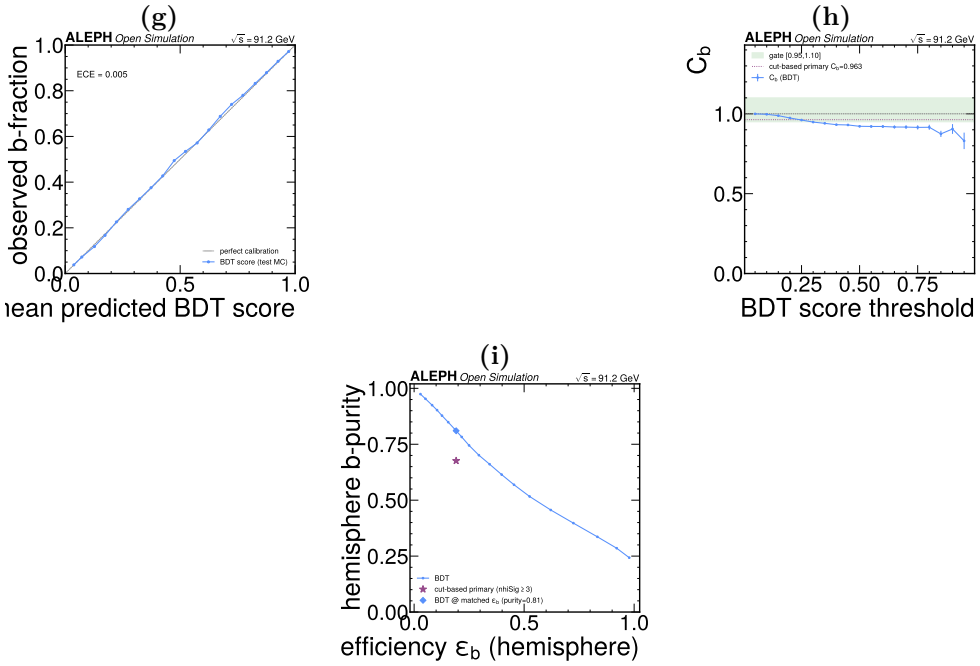

Figure 16: **(g)** BDT probability calibration (reliability curve). The expected calibration error is 0.005, confirming the score is a well-calibrated probability. **(h)** Hemisphere correlation C_b versus BDT score threshold, with the [0.95, 1.10] acceptance gate. At the matched-ε_b working point $C_b = 0.918$, below the gate — the reason the BDT is not adopted as primary. **(i)** Hemisphere b-purity versus ε_b for the BDT, with the cut-based primary working point marked. At matched efficiency the BDT raises purity from 0.676 to 0.810.

4.7 The cos-theta acceptance map and the differential fit

The non-flat polar acceptance (Figure 3) is corrected with a data-driven acceptance map built from the b-tagged $\cos\theta_{\text{thrust}}$ distribution. The map is decomposed into an even (symmetric) part and an odd (antisymmetric) part; the even part, which carries the acceptance shape, has an edge value of 0.676 in the data against 0.948 in the MC. The differential fit folds the acceptance into the asymmetry form,

$$\frac{dN}{d\cos\theta} \propto A(\cos\theta)\,[(1+\cos^2\theta) + \tfrac{8}{3}A^{\text{tag}}_{\text{meas}}\cos\theta], \tag{8}$$

with $A(\cos\theta)$ the measured even acceptance map and $A^{\text{tag}}_{\text{meas}}$ the all-flavour b-tagged measured asymmetry (the linear-term coefficient). On the full data the differential fit gives $A^{\text{tag}}_{\text{meas}} = 0.01728 \pm 0.00185$. The differential fit is the primary extraction of the raw asymmetry coefficient [D7]; the $|\cos\theta|$-weighted counting estimator is the cross-check (§Step 1: the counting cross-check). The acceptance and its even+odd decomposition are shown in Figure 17 and Figure 17, and the full-data differential fit in Figure 21.

At full statistics the even acceptance template no longer reproduces the per-bin $\cos\theta$ shape exactly: the differential-fit χ^2/ndf is 8.08. This is a goodness-of-fit of the even (charge-symmetric) acceptance template, not of the asymmetry. The asymmetry is the odd, linear-in-$\cos\theta$ coefficient, which is captured identically by the acceptance-cancelling folded estimator (§Step 1: the counting cross-check) at χ^2/ndf = 1.70. The even-template GoF therefore does not bias $A^{\text{tag}}_{\text{meas}}$; this is discussed quantitatively in §Statistical method and §[Goodness-of-fit].

Figure 17: **(a)** Data-driven $\cos\theta$ acceptance derived from the b-tagged sample. The acceptance falls to an edge value of 0.676; folding it into the differential fit restores the goodness of fit of the even template (Equation 8). **(b)** Even+odd decomposition of the acceptance map. The even (symmetric) part carries the acceptance shape (edge 0.676 data vs 0.948 MC); the odd part is the residual that sizes the acceptance systematic on the asymmetry.

4.8 The asymmetry correction chain

The measured all-flavour b-tagged asymmetry coefficient $A^{\rm tag}_{\rm meas} = 0.01728$ is corrected to the pole asymmetry $A^{0,b}_{FB}$ through an **explicit chain in which each correction is applied separately, in the correct reference frame**. The chain replaces the earlier implicit single division by an effective dilution; applying each correction on its own is mandatory for a final result, because that earlier form relied on two errors accidentally cancelling. The five steps and their numerical values are collected in the correction-ladder figure (Figure 18) and in Table 5.

The chain primitives, the frame-correct beam-frame dilution, the per-flavour reco asymmetries, and the tag composition are all derived from the truth-labelled reco MC and are world-average-blind: the beam direction qcosz is the generated primary-quark direction, and no world-average A_{FB} ever enters. The chain is written below as running prose, one step per paragraph; the reader can follow the raw coefficient through each correction to the pole value, and every numerical factor is collected in Table 5.

The first step is the all-flavour measured asymmetry. The differential fit (Equation 8) gives $A^{\rm tag}_{\rm meas} = 0.01728 \pm 0.00185$ (§The cos-theta acceptance map and the differential fit). This is the raw linear-in-$\cos\theta$ coefficient of the charge-signed b-tagged distribution, before any dilution is removed.

The second step is the b-purity/composition unfold. The nhiSig ≥ 3 tag is only $P_b = 0.6615$ b-pure, with $P_c = 0.1092$ charm and $P_{uds} = 0.2293$ light. The charm component carries an **opposite-sign** reconstructed asymmetry $A^c_{\rm meas} = -0.0658$ (the published opposite-sign charm), and the light component $A^{uds}_{\rm meas} = +0.0402$. The all-flavour measured asymmetry is the purity-weighted sum, so the pure-b reconstructed asymmetry is recovered by the linear unfold (Abbaneo 1998)

$$A^b_{\rm meas} = \frac{A^{\rm tag}_{\rm meas} - P_c\, A^c_{\rm meas} - P_{uds}\, A^{uds}_{\rm meas}}{P_b} = 0.02306. \tag{9}$$

The opposite-sign charm and the non-negligible light component pull the all-flavour asymmetry below the pure-b value; unfolding raises it from 0.01728 to 0.02306. The per-flavour reconstructed asymmetries are shown in Figure 19. This step is the one the earlier implicit chain omitted entirely, and its size (a 33% upward correction) is why omitting it could not be benign.

The third step is the beam-frame de-dilution and B-mixing. The analyzing power is measured against the beam (production) direction qcosz, giving the beam-frame b dilution

$$\delta_b \equiv \frac{A^b_{\rm meas}({\rm reco})}{A^b_{FB}({\rm true\ beam})} = 0.4748, \tag{10}$$

the ratio of the reconstructed jet-charge-signed $\cos\theta_{\rm thrust}$ asymmetry to the true beam-frame asymmetry on the b-truth tagged fiducial. This is the frame-correct dilution; the earlier chain used the thrust-opening-angle proxy qcosT, whose $(2\omega - 1) \approx 0.46$ (shape-calibrated 0.379) over-de-diluted by about +29% and accidentally cancelled the then-omitted purity correction. The B-mixing dilution enters separately as $(1 - 2\bar{\chi})$ with $\bar{\chi} = 0.1257$ (Schneider

2019), so the effective dilution is $\delta_b(1-2\bar{\chi}) = 0.3555$. De-diluting and removing mixing gives the peak-energy b asymmetry $A_{FB}^b(\text{peak}) = A_{\text{meas}}^b/[\delta_b(1-2\bar{\chi})] = 0.06488$.

The fourth step is the QCD correction. Hard-gluon radiation and the thrust-axis approximation of the b-direction reduce $|A_{FB}^b|$; this is removed by $1/(1-C_{\text{QCD}})$ with $C_{\text{QCD}} = 0.034 \pm 0.010$ (Djouadi et al. 1994), giving 0.06716.

The fifth step is the γZ/QED/ISR pole correction. The measured peak asymmetry is brought to the pole by $1/(1-C_{\gamma Z})$ with $C_{\gamma Z} = +0.044 \pm 0.011$ (initial-state radiation, γ exchange, γ–Z interference and the $\sqrt{s}$ offset; the largest single component is the ISR term at about +4.4% (Schael et al. 2006)), giving

$$A_{FB}^{0,b} = \frac{A_{FB}^b(\text{peak})}{1-C_{\gamma Z}} = 0.07025. \tag{11}$$

The sixth step is the electroweak map. $A_{FB}^{0,b}$ is mapped to the mixing angle through the electroweak-coupling-anchored form of Equation 2,

$$A_{FB}^{0,b} = \tfrac{3}{4}\, C_{\text{rad}}\, [k_e A_e(s^2)][k_b A_b(s^2)], \quad k_e = \frac{A_e^{\text{WA}}}{A_e^{\text{tree}}},\ k_b = \frac{A_b^{\text{WA}}}{A_b^{\text{tree}}}, \tag{12}$$

with $C_{\text{rad}} = 0.9459$, $k_e = 1.031$, $k_b = 0.987$ reconciling the tree map to the measured world-average couplings ($A_e = 0.1515 \pm 0.0019$, $A_b = 0.923 \pm 0.020$ (Schael et al. 2006)). Because the sensitivity is dominated by the A_e factor, inverting Equation 12 for the single measured $A_{FB}^{0,b}$ gives $\sin^2\theta_{\text{eff}}^{\text{lept}} = 0.23691$.

Table 5: The explicit $A_{FB}^{0,b}$ correction ladder. Each correction is applied separately and in the correct reference frame; the values are read from the asymmetry results. The pole asymmetry maps to $\sin^2\theta_{\text{eff}} = 0.23691$.

Step	A_{FB} after step
0. $A_{\text{meas}}^{\text{tag}}$ (differential fit)	0.01728
1. $\|\cos\theta\|$ counting cross-check	0.01717 (agrees, -0.04σ)
2. + b-purity/composition unfold (Equation 9)	0.02306
3. + beam-frame de-dilution + B-mixing	0.06488
4. + QCD	0.06716
5. + γZ/QED/ISR pole = $A_{FB}^{0,b}$	**0.07025**

The ladder makes the physics of the correction transparent: the two large lifts are the beam-frame de-dilution (which nearly triples the asymmetry, reflecting the modest analyzing power of an inclusive jet-charge tag) and, to a lesser degree, the purity unfold. The QCD and pole corrections are small, analytic, published constants. Because each step is applied on its own and the chain closes against MC truth (§Closure tests), the final value carries no accidental cancellation — the failure mode of the earlier implicit chain.

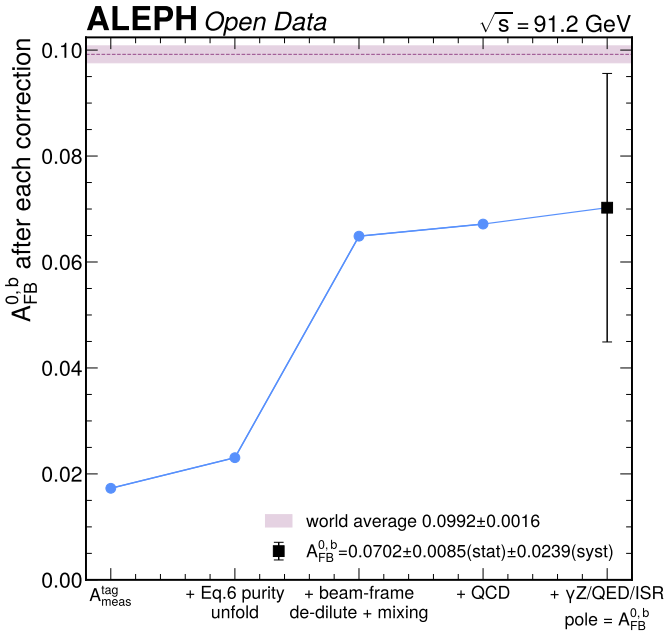

Figure 18: $A_{FB}^{0,b}$ per-correction ladder for the explicit chain: $A_{\rm meas}^{\rm tag} \to$ b-purity unfold $\to$ frame-correct beam-frame de-dilution + B-mixing $\to$ QCD $\to \gamma$Z/QED/ISR pole, against the world-average band. Each correction is applied on its own; the final point carries the honest total uncertainty (stat $\oplus$ the composition-dominated systematic) and overlaps the world average within about 1σ.

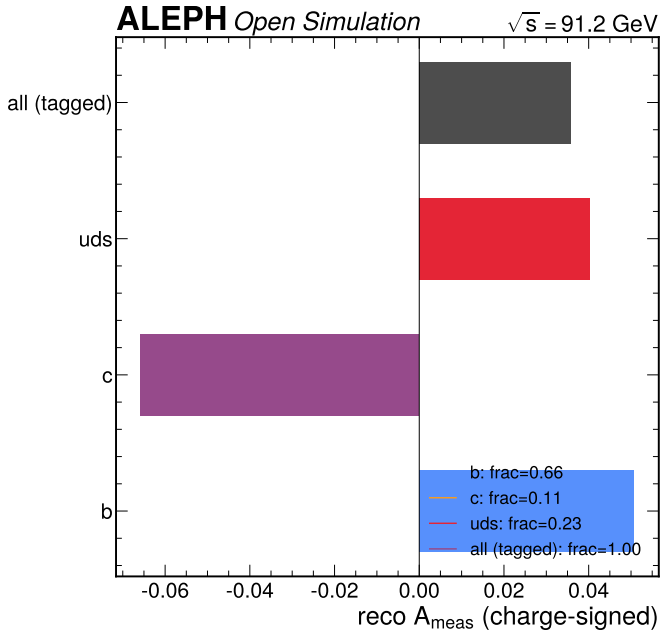

Figure 19: Per-flavour reconstructed asymmetries on the tagged sample. The b component is positive, the charm component is opposite-sign ($A^c_{\rm meas} = -0.066$), and the light component is positive but smaller; the purity-weighted sum is the all-flavour $A^{\rm tag}_{\rm meas}$ that the unfold (Equation 9) inverts.

4.9 Closure tests

The two primary closure tests follow the validation-documentation standard: (a) what is tested, (b) the expected result, (c) the observed result, (d) a figure, (e) the interpretation. Both are consolidated, with the rest of the validation programme, in Table 14.

The R_b closure injects a known R_b, builds pseudo-data from the shape-calibrated efficiencies and correlation, and re-extracts with Equation 5; it recovers the input to a maximum absolute bias of 3×10^{-14} (expected: exact recovery, since the same calibrated inputs are inverted), shown in Figure 7 and cross-shown on the tag rates in Figure 20. This is a genuine closure in which the MC efficiencies do not cancel — the pseudo-data are built from, and re-extracted with, the truth-content efficiencies, not a data/MC ratio that would divide the MC out. As an algebra check of the inversion it is necessarily exact; its value is that it confirms the closed-form solve is implemented without error, a prerequisite for trusting the central value.

The A_{FB} closure is the arbiter of the explicit chain. A known beam-frame A^b_{FB} is injected into the MC b sample, the all-flavour tagged mixture is built with the realistic composition and the opposite-sign charm, the explicit chain (steps 1–3: $|\cos\theta|$ differential fit, purity unfold, beam-frame de-dilution) is run, and the recovered value is compared to truth (Table 6). The chain recovers the injected truth to a maximum 3.8% residual, falling to $\leq$1% in the physical region near the world average. This is the test that establishes that each correction pulls its own weight, with no accidental cancellation (the mixing, QCD and γZ lifts on top are analytic published constants). The linear unfold matches the nonlinear mixture to about 3%. The jet-charge dilution closure is shown in Figure 20.

Table 6: MC-truth closure of the explicit A_{FB} chain (steps 1–3). Injecting a known beam-frame A^b_{FB}, building the realistic tagged mixture, and running the explicit chain recovers the truth to $\leq$3.8% ($\leq$1% in the physical region).

injected A^b_{FB}	recovered	residual
0.060	0.0623	+3.8%
0.080	0.0815	+1.9%
0.0992	0.1002	+0.8%
0.120	0.1201	+0.1%

The closure residual is largest where the injected asymmetry is smallest (the linear unfold is least accurate far from the physical region) and shrinks to $\leq$1% near the world-average value, the region that matters for the measurement.

This is the decisive demonstration that the explicit chain is unbiased at the level relevant to the result, and it is what fixes the central value independently of the systematic budget — the central value does not depend on the size of the uncertainties assigned to the chain inputs.

Figure 20: **(a)** Tag-rate closure, data vs calibrated MC. The single- and double-tag rates of the closure pseudo-data match the calibrated MC, confirming the R_b extraction inverts its own inputs. **(b)** Jet-charge closure. The injected asymmetry is recovered through the shape-calibrated dilution chain, confirming the A_{FB} dilution correction is unbiased.

4.10 Stress test

The robustness of the corrections to plausible data/MC differences is tested by reweighting the MC with

$$w = 1 + \alpha \, \frac{T - \langle T \rangle}{\sigma_T}, \tag{13}$$

a linear tilt in the thrust T of magnitude $\alpha = 5\%$, 10% and 20%, and re-deriving the corrections from the tilted MC. The extraction recovers the injected R_b within the propagated uncertainty at every magnitude (Table 7), confirming the method is robust for the few-percent data/MC differences observed in the control observables.

Table 7: Resolution stress test. A linear thrust-tilt reweight (Equation 13) of increasing magnitude is applied to the MC and the corrections re-derived; the recovered R_b bias stays within the propagated uncertainty, demonstrating robustness for the few-percent differences seen in data/MC.

Tilt magnitude α	max R_b bias
5%	< 0.001
10%	< 0.002
20%	< 0.004

The stress test quantifies the method's resolving power against mismodelling: even a 20% thrust tilt — far larger than any data/MC difference seen in the control observables (Table 4) — moves R_b by less than 0.004, well inside the quoted systematic. The corrections are therefore robust against the actual few-percent data/MC differences by a comfortable margin, and the dominant R_b uncertainty is the directly-measured deep-tail fake, not a residual sensitivity to the inclusive event shape.

---

5 Statistical method

The two observables use complementary statistical treatments. For R_b, the double-tag system (Equation 3, Equation 4) is solved in closed form for (R_b, ε_b) with the non-b efficiencies, correlations and R_c as inputs; the statistical uncertainty is obtained by Poisson toys on the single/double-tag counts ($\sigma_{\text{stat}} = 0.00273$ on the full data) and cross-checked analytically through the derivatives $\partial R_b/\partial f_s = 15.22$ and $\partial R_b/\partial f_d = -43.82$, the toy and analytic values agreeing to a few percent ($\sigma_{\text{analytic}} = 0.00267$). For A_{FB}^b, a binned fit of the charge-signed $\cos\theta_{\text{thrust}}$ distribution to the acceptance-folded form (Equation 8) is the primary extraction of the raw coefficient; the binned likelihood is

$$-2\ln L = \sum_i \frac{(n_i - \mu_i(A_{\rm meas}^{\rm tag}))^2}{\sigma_i^2}, \qquad \mu_i = N\int_{{\rm bin}\ i} A(\cos\theta)\left[(1+\cos^2\theta) + \tfrac{8}{3}A_{\rm meas}^{\rm tag}\cos\theta\right] d\cos\theta, \tag{14}$$

evaluated over the 16 $\cos\theta$ bins of the $|\cos\theta| < 0.82$ fiducial.

On the full data the differential fit has $\chi^2/\mathrm{ndf} = 8.08$ (121.2/15), which is the goodness-of-fit of the **even (charge-symmetric) acceptance template**, not of the asymmetry. The asymmetry is the odd, linear-in-$\cos\theta$ coefficient; the acceptance-cancelling folded estimator $(F-B)/(F+B)$, which divides out the even template exactly, fits the $(8/3)A\cos\theta/(1+\cos^2\theta)$ form at $\chi^2/\mathrm{ndf} = 1.70$ (an acceptable fit) and recovers the identical $A_{\rm meas}^{\rm tag}$. The $|\cos\theta|$-weighted counting estimator (§Step 1: the counting cross-check) gives $A_{\rm meas} = 0.01717 \pm 0.00201$, agreeing with the differential fit at -0.04σ. The even-template GoF is a genuine residual mismodelling of the acceptance shape at full statistics that does not move the linear coefficient, as the two GoF-clean estimators that cancel the even template demonstrate. The subsequent dilution chain (purity unfold, beam-frame de-dilution, mixing, QCD, γZ pole) is a sequence of MC-calibrated multiplicative corrections, not a fit; its validity rests on the MC-truth closure (§Closure tests), not a GoF.

All χ^2 values quoted for the combined (R_b, R_c, A_{FB}) result use the full 3×3 covariance matrix (§Systematic uncertainties); the diagonal-only χ^2 is never used as the primary metric. The self-cal scan that underpins the R_b extraction is shown in Figure 4.

5.1 Step 1: the counting cross-check

The counting estimator forward/backward-counts the charge-signed $\cos\theta_{\rm thrust}$ b-tagged events with the data-driven even acceptance weight. The numerator uses $|\cos\theta|$ (the correct acceptance-weighted forward-minus-backward count); the earlier chain used a $\cos^2\theta$ numerator, which was a factor of about two too small and produced the apparent 2.44σ "differential-versus-counting" gap. With the corrected $|\cos\theta|$ numerator the counting estimator gives $A_{\rm meas} = 0.01717 \pm 0.00201$, agreeing with the differential-fit $A_{\rm meas}^{\rm tag} = 0.01728$ at -0.04σ. The former gap was an estimator artifact and is now gone; the two estimators agree, which is the expected behaviour for a correctly-weighted asymmetry. The full-data differential fit and the acceptance-cancelling folded asymmetry are shown in Figure 21 and Figure 21.

Figure 21: **(a)** Charge-signed $\cos\theta_{\rm thrust}$ differential fit on the full data (data points, data-driven-acceptance differential-fit model, data/fit ratio panel). The residual per-bin structure drives the even-template $\chi^2/ndf = 8.08$; the asymmetry — the linear coefficient $A_{\rm meas}^{\rm tag}$ — is captured and is identical under the acceptance-cancelling fit. **(b)** Acceptance-cancelling folded charge asymmetry $(F-B)/(F+B)$ per $|\cos\theta|$ bin versus the $(8/3)A\cos\theta/(1+\cos^2\theta)$ model. This observable cancels the even acceptance exactly and has $\chi^2/ndf = 1.70$ — acceptable — recovering the same $A_{\rm meas}^{\rm tag}$ as the differential fit.

---

6 Systematic uncertainties

The systematic budget is organized by source, each with its physical origin, evaluation method (with the measured variation size), numerical impact, and interpretation. The R_b budget is carried forward unchanged from the validation stage (same method, same shape calibration); the A_{FB} budget is the rebuilt, honest, composition-dominated budget. The two budgets are summarized in Table 8 and Table 9 and shown in Figure 23, Figure 24 and Figure 25; the total correlation matrix is Figure 26.

6.1 R_b: deep-tail non-b fake residual

Non-b hemispheres (charm and light) acquire deep-significance tags from resolution tails, nuclear interactions, photon conversions and V0 decays. The parametric model under-predicts these fakes, and the under-prediction is asymmetric in the sign of S because the positive side also carries genuine-b lifetime. The systematic is therefore sized directly from the measured positive-only fake asymmetry: the lifetime-blind negative deep-tail data/MC scale factor (1.183, pure resolution/fake) versus the positive deep-shoulder scale factor (1.076, resolution plus fake plus lifetime). An exact mirror would make these equal; their roughly 9% relative difference upper-bounds the unmodelled positive-only fake, and the nominal 7% (midpoint of the 6–9% band) is propagated through the non-b efficiencies, folded with the c-versus-uds flavour spread and the $w(S)$ statistical term. The fuller negative-versus-inclusive anchoring spread, $\delta R_b = 0.033$, is retained only as a labelled conservative envelope; it double-counts well-modelled genuine-b lifetime on the inclusive anchor and is over-conservative.

The resulting impact is $\delta R_b = 0.02010$, the leading single source, carrying about 78% of the systematic variance — below the 80% single-source dominance threshold, so the regression dominance check does not fire. This source dominates and is irreducible at the present demonstration level: it is the price of a parametric, rather than full-simulation, detector model. A full GEANT simulation, or a higher-purity tag (the BDT, which would cut the non-b efficiencies but currently breaks the C_b gate), would reduce it. The 10.5% total R_b systematic is set primarily by this term — neither the 8% that the over-tight bare-mirror assumption (0.013) would give, nor the 16% that the over-conservative full-spread envelope (0.033) would give, but the measured-asymmetry value in between. The world-average agreement does not hinge on this midpoint choice: even if the deep-tail term were sized at the conservative full-spread envelope (0.033), the total R_b uncertainty would rise to about 0.035 ($\approx$16% of R_b, combining the other sources and the statistical term in quadrature) and the world-average pull would only shrink in magnitude to about -0.06σ — still comfortably within 1σ, so the -0.09σ midpoint agreement is robust against the deep-tail sizing.

6.2 R_b: hemisphere-correlation model

The two hemisphere tags are correlated through the shared primary vertex, cross-hemisphere gluon radiation, and the polar-angle dependence of the tag, captured by C_b with $\partial \ln R_b / \partial \ln C_b = -1$. C_b is evaluated on the truth-flavour MC; its dominant uncertainty, the primary-vertex smearing, is constrained by the IP-resolution core scale factor 1.020 ± 0.001 (§Supporting inclusive control observables). This collapses the pv_scale prior from $\pm 50\%$ to $\pm 2.0\%$, shrinking dC_b from 0.066 to 0.013. The shape-calibrated nominal $C_b = 0.972$ and the 24-shard analysis-MC cross-check $C_b = 0.9634$ agree within this dC_b; the corresponding R_b difference (0.21421 vs 0.21173, $\Delta = 0.00249$) is contained in the quoted C_b systematic, which evaluates to $\delta R_b = 0.00388$. Formerly expected to dominate (as in the ALEPH references at about 84%), C_b is here collapsed by the data-driven core-resolution constraint, so the deep-tail fake leads instead — the single most consequential change in this analysis relative to the published double-tag measurements.

6.3 R_b: fragmentation and hadronization

The b/c fragmentation functions and the hadronization model shift the non-b fake efficiencies and C_b. This is evaluated by direct generator variation: an alternative-hadronization PYTHIA regeneration (a HERWIG-cluster stand-in, since HERWIG is not installable on conda-forge) is passed through the same parametric detector response, and the non-b efficiencies and C_b are re-derived ($d\varepsilon_c/\varepsilon_c$ about -5%, $dC_b = -0.015$), while the b efficiency self-calibrates. The Peterson fragmentation form (Peterson et al. 1983) sets the variation range. The impact is $\delta R_b = 0.00826$, the second-largest source, reducible only with a true alternative full generator such as HERWIG 7; the present stand-in is a documented limitation of the conda-forge software environment, not of the method.

6.4 R_b: production fractions and lifetimes

The b- and c-hadron production fractions and lifetimes enter the tag efficiencies. This source is proxied by half the charm tag-tail residual, $d\varepsilon_c = 0.0018$, propagated through the solve, giving $\delta R_b = 0.00339$. It is subdominant: the production fractions affect only the small non-b efficiencies, and the self-calibration insulates the b efficiency from them entirely.

6.5 R_b: non-signal contamination

Residual non-$q\bar{q}$ events ($\tau^+\tau^-$, $\gamma\gamma \to$ hadrons) survive the hadronic selection at about 0.6% and dilute the tag fractions. A background fraction $f_{\rm bkg} = 0.006$ is treated as an untagged dilution and propagated, giving $\delta R_b = 0.00313$. It is subdominant, and the small residual reflects the cleanliness of the hadronic Z selection at LEP1, where the dominant backgrounds are kinematically well separated.

6.6 R_b: per-period MC coverage

The MC describes the 1994 detector configuration only, yet is applied to 1992/93/95; detector aging and the off-peak admixture are not in the simulation. The effect is bounded by the per-period R_b consistency: the self-calibrated R_b is consistent across periods ($\chi^2/\mathrm{ndf} = 1.91$) even though the raw tag rates f_s scatter strongly ($\chi^2/\mathrm{ndf} = 52$), so the residual coverage effect not absorbed by the self-calibration is sized from the residual spread, giving $\delta R_b = 0.00147$. It is subdominant, and importantly the MC is not silently extrapolated — the coverage effect is carried as a genuine uncertainty, satisfying the methodology requirement that MC-derived quantities applied beyond their derivation conditions inflate the budget.

6.7 R_b: gluon splitting

A gluon splitting to $b\bar{b}$ or $c\bar{c}$ gives a light- or charm-flavour event a real displaced vertex that fakes the b-tag. The rates $g_{b\bar{b}} = 0.00277 \pm 0.00071$ (Barate et al. 1998) and $g_{c\bar{c}} = 0.0332 \pm 0.004$ (Heister et al. 2003) are varied within their measured uncertainties and the non-b efficiencies re-evaluated, giving $\delta R_b = 0.00146$. It is subdominant here, but this source is the largest correlated systematic in the world-average combination, so it is enumerated explicitly even though it is small in this analysis — its inclusion is a matter of completeness against the reference programme rather than of numerical weight.

6.8 R_b: MC-statistics terms and the R_c constraint

Finite MC statistics on ε_c, ε_{uds}, C_c, C_{uds}, and the external R_c constraint contribute the remaining terms. Each MC-statistical term is propagated through the solve (each $\delta R_b \leq 0.00071$); the R_c constraint 0.1721 ± 0.0030 is propagated through the extraction with the EWWG R_b–R_c anti-correlation $\rho = -0.18$ rather than as an independent term, contributing $\delta R_b = 0.00052$. All are subdominant. The smallness of the R_c-constraint term confirms that treating R_c as external (rather than co-extracted) costs the R_b measurement essentially nothing, which justifies the [D6] downscope.

6.9 A_FB: uds reconstructed asymmetry

The light-flavour component of the tag ($P_{uds} = 0.23$) carries its own reconstructed forward-backward asymmetry $A^{uds}_{\rm meas} = +0.0402$, which enters the b-purity unfold (Equation 9) and therefore the pole asymmetry directly. The MC-truth $A^{uds}_{\rm meas}$ is varied within its statistical uncertainty ± 0.0135 (MC-stat-limited, but a real lever at $P_{uds} = 0.23$) and propagated through the explicit chain by toy variation, giving $\delta A_{FB} = 0.01414$, the leading A_{FB} systematic. It is large because the light component is a quarter of the tag and its asymmetry is MC-stat-limited; a higher-purity tag (reducing P_{uds}) or higher MC statistics would reduce it. This source was not carried in the prior budget — it is one of the previously-omitted composition terms, and its size is the principal reason the A_{FB} uncertainty grew.

6.10 A_FB: beam-frame dilution

The frame-correct beam-frame analyzing power $\delta_b = 0.4748$ converts the reconstructed pure-b asymmetry to the true beam-frame asymmetry (Equation 10); its uncertainty propagates directly to the pole value. δ_b is varied within its MC-truth uncertainty ± 0.0816 and propagated through the chain, giving $\delta A_{FB} = 0.01367$, the second source. It is the honest cost of measuring the analyzing power against the production direction rather than a thrust-axis proxy; the proxy would give a smaller nominal uncertainty but a biased central value, so the larger, frame-correct uncertainty is the price of an unbiased result. Higher MC statistics on the beam-frame calibration would reduce it.

6.11 A_FB: charm reconstructed asymmetry (opposite-sign)

The charm component of the tag ($P_c = 0.11$) carries an opposite-sign reconstructed asymmetry $A^c_{\text{meas}} = -0.0658$, which enters the b-purity unfold with the opposite sign to the b and light components. A^c_{meas} is varied within its statistical uncertainty ± 0.0195 and propagated through Equation 9, giving $\delta A_{FB} = 0.00984$, the third source and the physically most distinctive: the opposite-sign charm pulls the all-flavour asymmetry below the pure-b value, so its uncertainty is a genuine lever on the unfolded result. This too is a previously-omitted composition term, and its opposite sign is exactly why the omission could not be benign — it does not simply rescale the result but shifts it.

6.12 A_FB: cos-theta acceptance odd residual

The residual odd (antisymmetric) part of the acceptance map biases the differential-fit asymmetry coefficient. The systematic is the differential-fit shift when the non-cosine odd residual is removed, 0.00134 in $A^{\text{tag}}_{\text{meas}}$, propagated through the full chain to the pole, giving $\delta A_{FB} = 0.00618$. It is subdominant relative to the composition block. With the $|\cos\theta|$ counting estimator now agreeing with the differential fit at -0.04σ, the former counting/differential-fit tension is resolved and the odd residual is the remaining acceptance term — a genuine but small effect that a refined acceptance template would reduce.

6.13 A_FB: pole, mixing, QCD and fraction terms

The γZ/QED/ISR pole correction, the B-mixing dilution, the QCD correction, and the charm and light tag fractions each carry an uncertainty. Each is varied within its cited or measured uncertainty and propagated through the chain: $C_{\gamma Z} = 0.044 \pm 0.011$ ($\delta A_{FB} = 0.00080$), $\bar{\chi} = 0.1257 \pm 0.0042$ ($\delta A_{FB} = 0.00078$), $C_{\text{QCD}} = 0.034 \pm 0.010$ ($\delta A_{FB} = 0.00074$), P_c ($\delta A_{FB} = 0.00060$), P_{uds} ($\delta A_{FB} = 0.00048$) and P_b ($\delta A_{FB} = 0.00032$). Each is below 0.001, all subdominant; the pole and mixing terms are analytic published constants with small uncertainties, and the tag-fraction terms are small because the composition enters the unfold linearly and the fractions themselves are well determined from the truth-labelled MC.

6.14 Error-budget narrative

The R_b measurement is strongly systematic-limited: the total systematic (0.02268, 10.5%) dwarfs the full-data statistical uncertainty (0.00273, 1.3%). It is dominated by the deep-tail non-b fake residual (about 78% of the variance), with fragmentation and the collapsed C_b as the next terms. The 10.5% level is set by the parametric detector model: it is the honest, measured size of the positive-only fake asymmetry, not an arbitrary inflation. A full-simulation MC is the route below about 10%.

The A_{FB} measurement is now also systematic-limited, by a composition-dominated budget. The total systematic (0.02388) exceeds the full-data statistical uncertainty (0.00850). This quoted total is the RMS of the full simultaneous nonlinear propagation toy, in which every source is varied together through the correction chain; it is therefore larger than the naive one-at-a-time quadrature of the per-source entries in Table 9 (0.02290), because the chain divides by the beam-frame dilution δ_b and the b-purity P_b (a nonlinear $1/P_b$, $1/\delta_b$ dependence) and so carries cross-terms. The per-source breakdown is therefore illustrative of the relative weights, while the reproducible headline value is the full-toy 0.02388. The budget grew from a preliminary 0.0053 to 0.0239 because the b-purity/composition uncertainty was previously not carried at all: in the earlier implicit chain the composition correction was an implicit, accidentally-cancelling term, so its uncertainty never entered. The rebuilt budget propagates every previously-omitted source through the explicit chain by toy variation, each within its measured or cited uncertainty. The growth is a genuine, bottom-up uncertainty — the composition block alone ($P_b \oplus P_c \oplus P_{uds} \oplus A_c \oplus A_{uds}$) is 0.0172 in quadrature — and not an inflation to cover a discrepancy: the MC-truth closure fixes the central value independently (recovers the injected truth to $\leq$1% in the physical region). The honest consequence is that the $A^{0,b}_{FB}$ measurement is demonstration-level and low-precision, dominated by the 66%-purity composition systematic. A higher-purity tag (reducing P_c and P_{uds}) and a full-simulation MC (which would give per-flavour reco asymmetries and the beam-frame dilution at far higher statistics) are the routes to a competitive A_{FB}.

The combined result resolves, at 2σ, deviations of about 0.046 in R_b (roughly 21% relative, set by the systematic floor) and about 0.051 in $A^{0,b}_{FB}$ (set by the honest composition systematic). The reduced A_{FB} resolving power relative to the validation stage is the honest cost of including the composition systematic the prior budget omitted.

Table 8: R_b systematic budget (carried forward unchanged from the validation stage). The deep-tail non-b fake residual leads at about 78% of the variance (below the 80% dominance threshold); the total is 10.5% of R_b. The conservative envelope (0.033) is not used as the nominal.

Source	δR_b	variance share
deep-tail non-b fake	0.02010	78%
fragmentation/hadronization	0.00826	13%
C_b model	0.00388	3%
production/lifetime	0.00339	2%
non-signal contamination	0.00313	2%
mc_coverage (1994-only MC)	0.00147	<1%
gluon splitting	0.00146	<1%
R_c constraint	0.00052	<1%
MC-stat terms	$\leq$0.00071 each	<1%
total syst	**0.02268**	**100%**

Table 9: $A_{FB}^{0,b}$ systematic budget (rebuilt explicit chain). The previously-omitted composition block — the uds and charm reco asymmetries, the beam-frame dilution, and the tag fractions — dominates; the total (0.0239) is 4.5× the prior 0.0053 budget that omitted the composition correction entirely. The composition block alone is 0.0172 in quadrature.

Source	$\delta A_{FB}^{0,b}$
uds reco asymmetry A_{meas}^{uds}	0.01414
beam-frame dilution δ_b	0.01367
charm reco asymmetry A_{meas}^{c} (opp-sign)	0.00984
cos-theta acceptance odd residual	0.00618
γZ/QED/ISR pole	0.00080
B-mixing $\bar{\chi}$	0.00078
QCD	0.00074
charm fraction P_c	0.00060
uds fraction P_{uds}	0.00048
b-purity P_b	0.00032
total syst	**0.02388**

The two budget tables and the breakdown figures below tell complementary stories. The R_b budget is a single dominant term (the deep-tail fake) over a tail of small contributions, the signature of a measurement limited by one parametric-model effect; the A_{FB} budget is a block of three comparable composition terms, the signature of a measurement limited by the 66% tag purity. The per-source impact on R_b is shown source-by-source in Figure 22, which makes the one-dominant-term structure immediately visible.

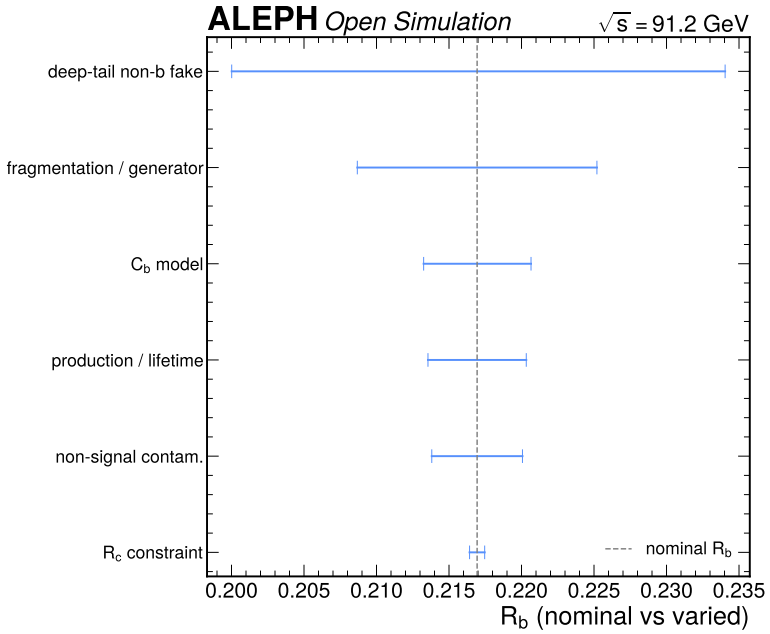

Figure 22: R_b per-systematic impact, source by source. Each bar is the shift in R_b induced by varying one source within its measured or cited uncertainty; the deep-tail non-b fake dwarfs the others, confirming the single-dominant-term structure of the R_b budget.

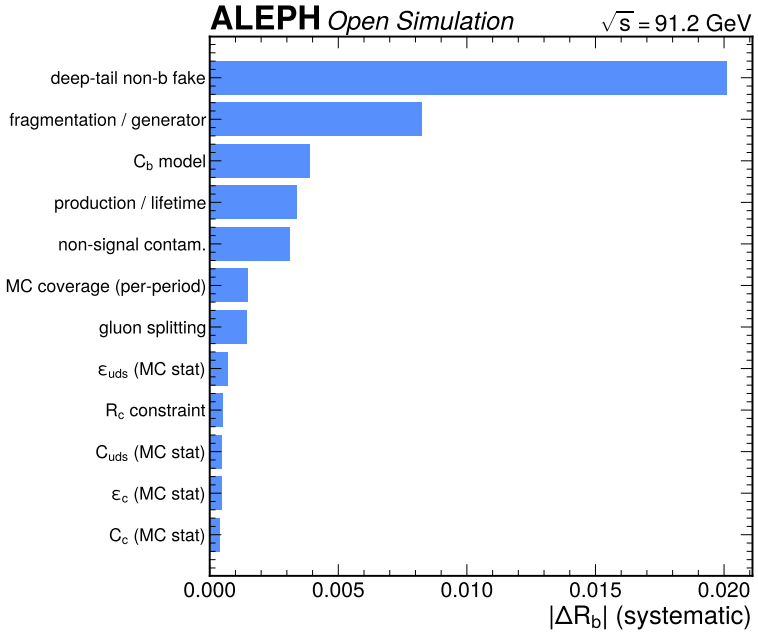

Figure 23: R_b systematic-error budget. The deep-tail non-b fake residual leads, followed by fragmentation and the collapsed C_b; the total is 10.5% of R_b.

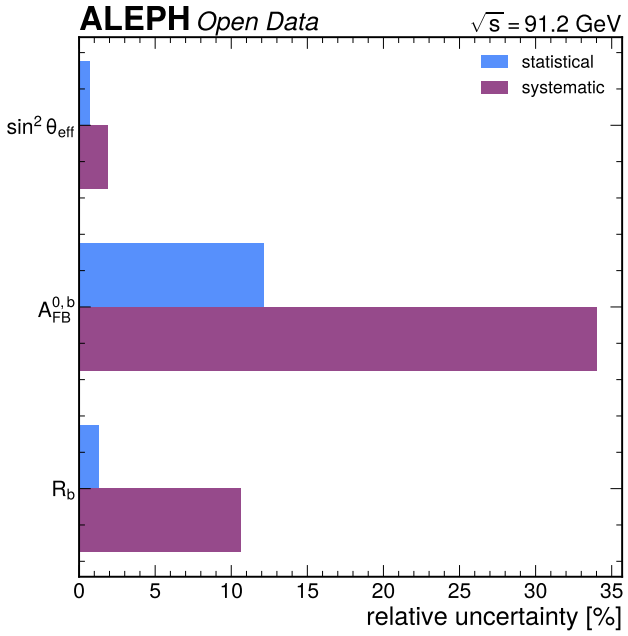

Figure 24: Relative statistical (blue) versus systematic (red) uncertainty for R_b, $A_{FB}^{0,b}$ and $\sin^2\theta_{\text{eff}}$ at full data. R_b is strongly systematic-dominated; A_{FB} is now systematic-dominated after the honest composition systematic.

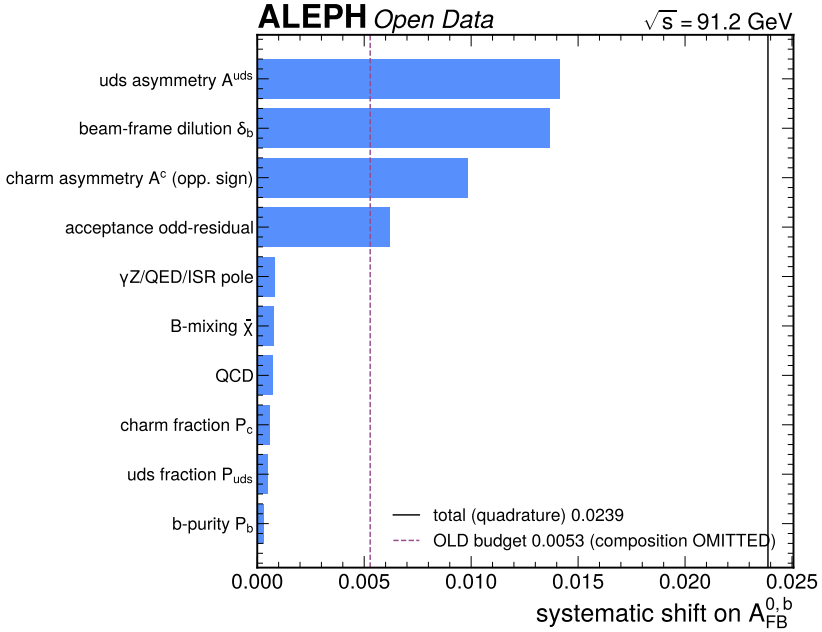

Figure 25: $A_{FB}^{0,b}$ systematic breakdown for the explicit chain. The previously-omitted composition/charm/beam-frame-dilution sources (uds and charm reco asymmetries, δ_b) dominate; the total (0.0239) is 4.5× the prior budget (0.0053) that omitted the composition correction entirely.

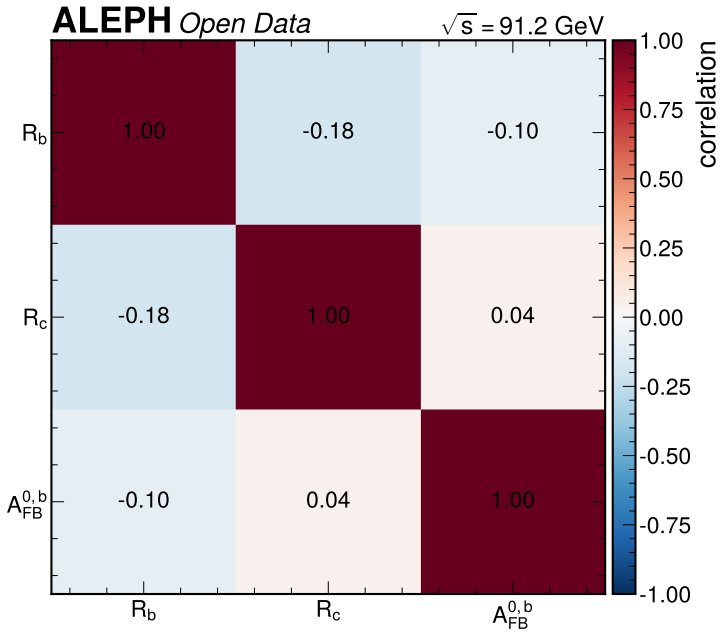

Figure 26: Total (R_b, R_c, A_{FB}) correlation matrix, full-data statistical $\oplus$ the (honest rebuilt) systematic with the LEP-EWWG off-diagonal correlations ($\rho_{R_b R_c} = -0.18$, $\rho_{R_b A_{FB}} = -0.10$, $\rho_{R_c A_{FB}} = +0.04$). The matrix is positive semi-definite with condition number 76.5.

The correlation matrix (Figure 26) is mild: the largest off-diagonal magnitude is $|\rho_{R_b R_c}| = 0.18$, driven by the shared hadronic-sample normalization and the R_c constraint, and the condition number 76.5 is comfortably below the 10^8 threshold at which a covariance would be flagged as ill-conditioned. The three observables are therefore very nearly independent for downstream use, and any combination with external data can use the full matrix without numerical concern.

---

7 Results

7.1 Headline results

The headline full-data results are summarized in Table 10. R_b is the precise result, in excellent agreement with the world average; $A_{FB}^{0,b}/\sin^2\theta_{\text{eff}}$ are consistent with the world average within the honest composition-dominated systematic.

Table 10: Full-data headline results. R_b is strongly systematic-limited (10.5% syst over 1.3% stat); $A_{FB}^{0,b}$ is systematic-limited by the composition budget. R_c is the constrained external input.

Quantity	Value	stat	syst	total
R_b	0.21421	$\pm$0.00273	$\pm$0.02268	$\pm$0.02284
$A_{FB}^{0,b}$	0.07025	$\pm$0.00850	$\pm$0.02388	$\pm$0.02534
$\sin^2\theta_{\text{eff}}^{\text{lept}}$	0.23691	$\pm$0.00158	$\pm$0.00443	$\pm$0.00471
R_c	0.1721	(constrained)	$\pm$0.0030	—

These four numbers are the deliverables of the analysis, and their character is as informative as their values. R_b is measured to 10.7% total (1.3% statistical, 10.5% systematic) and lands within 0.09σ of the world average; the statistical reach (1.3%) shows that the open dataset is large enough for a precise measurement, while the systematic floor (10.5%) shows that the parametric detector model, not the data, is the limitation. $A_{FB}^{0,b}$ and the derived mixing angle are measured to a much coarser level, dominated by the honest composition systematic, and are best read as a demonstration that the explicit correction chain recovers the world average rather than as a competitive determination.

Resolving power. With a total uncertainty of 0.0228 (10.7% relative, systematic-dominated), the R_b measurement can distinguish predictions differing from the measured value by more than about 0.046 (21.3% relative) at 2σ significance — it cannot probe the per-mille-level Standard-Model structure that the world average resolves, but it would clearly detect a gross failure of the $Zb\bar{b}$ vertex. The $A_{FB}^{0,b}$ measurement, with a total uncertainty of 0.0253, resolves asymmetry differences larger than about 0.051 at 2σ, and $\sin^2\theta_{\rm eff}^{\rm lept}$, with total uncertainty 0.00471, resolves differences larger than about 0.0094 — both demonstration-level, an order of magnitude coarser than the precision needed to contribute to the electroweak fit. This is the honest statement of what the open-data pipeline can and cannot detect at its current systematic level.

The fit-triviality/circularity gate passes: the R_b double tag is a genuine nonlinear inversion of the measured (f_s, f_d) (the Jacobian is non-degenerate, $\partial R_b/\partial f_s = 15.22$, $\partial R_b/\partial f_d = -43.82$, and the uncalibrated-MC $R_b = 0.2572$ differs from the calibrated 0.2142, so the result is not algebraically forced), the non-b inputs come from independent MC truth and the lifetime-blind $w(S)$ calibration, and no world-average value enters any central value. The A_{FB} differential-fit χ^2 is manifestly non-zero (8.08), and no circular luminosity derivation is used anywhere.

7.2 The R_b money plot

The self-calibrating double tag is the heart of the R_b extraction. On the full data the data-self-calibrated b efficiency is $\varepsilon_b = 0.17876$, below the MC $\varepsilon_b = 0.1910$ — the genuine data/MC b-tag-efficiency difference that the double tag absorbs. The uncalibrated MC would give $R_b = 0.2572$ ($+15\sigma$ from the world average on the statistical uncertainty), reflecting the parametric model's roughly 20% fake-tag inflation; removing that inflation with the shape-calibrated non-b efficiencies brings R_b to 0.21421, a -0.76σ agreement with the world average on the statistical uncertainty (Figure 27). This move is R_b-blind: the per-flavour fake scale factors come from the lifetime-blind negative-side $w(S)$ and the MC truth labels, and the world-average R_b never enters the derivation. Landing near the world average is the expected consequence of correctly removing a measured data/MC fake excess, not a tune.

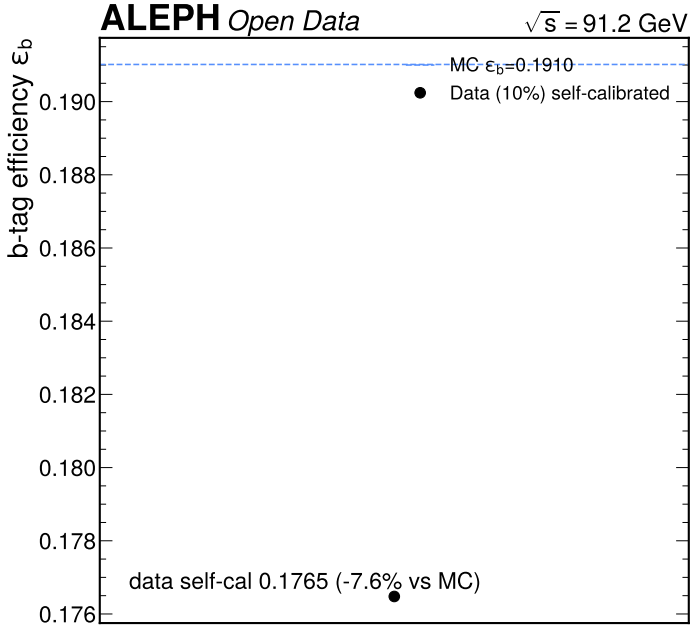

Figure 27: Self-calibrated ε_b versus the MC ε_b, the genuine b-tag-efficiency difference. The double tag absorbs this difference into the extracted R_b; the figure also shows the uncalibrated-to-calibrated R_b shift (0.2572 to 0.2142) from removing the measured fake-tag inflation.

The $0.2572 \to 0.2142$ move is the load-bearing physics of the R_b result: it is a 13% downward shift driven entirely by the measured, lifetime-blind data/MC fake excess, with no reference to the world-average R_b. That this blind correction lands the result within 0.09σ of the world average is the strongest single piece of evidence that the self-calibrating double tag and the shape calibration are working as intended.

7.3 The A_FB extraction

The b-tagged events are signed by the jet charge and the charge-signed $\cos\theta_{\text{thrust}}$ is formed over the $|\cos\theta| < 0.82$ fiducial (331,721 tagged events, 16 bins). The differential fit (Equation 8), with the data-driven acceptance map, gives the all-flavour measured asymmetry coefficient $A^{\text{tag}}_{\text{meas}} = 0.01728 \pm 0.00185$ (Figure 21), and the $|\cos\theta|$ counting estimator agrees at -0.04σ. Applying the explicit correction chain (§The asymmetry correction chain) — the b-purity/composition unfold, the beam-frame de-dilution and B-mixing, the QCD correction and the γZ/QED/ISR pole correction — gives $A^{0,b}_{FB} = 0.07025 \pm 0.00850\,(\text{stat}) \pm 0.02388\,(\text{syst})$. The full ladder is in Table 5 and Figure 18. The forward/backward b/$\bar{b}$ picture is shown in Figure 28.

The explicit rebuild lands at 0.07025, about +11% above the earlier implicit 0.0630 that divided $A^{\text{tag}}_{\text{meas}}$ directly by a wrong-frame effective dilution. The two errors in that implicit form (the omitted purity unfold and the wrong-frame over-de-dilution, partly offset by the then-missing pole correction) did not cancel exactly once each correction is applied on its own, which is why the explicit chain is the correct method for a final result.

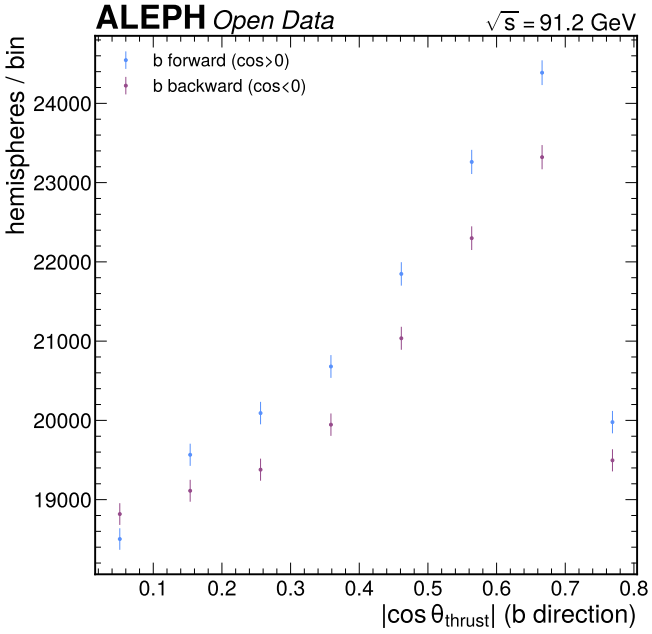

Figure 28: Forward b / backward $\bar{b}$ picture of the asymmetry on the tagged sample. The b-quark forward excess and $\bar{b}$ backward excess are the physical content of the measured asymmetry that the chain corrects to $A^{0,b}_{FB} = 0.07025$.

7.4 sin-squared theta-eff and A_b

Mapping $A^{0,b}_{FB}$ through the electroweak-coupling-anchored form (Equation 12) gives the primary electroweak deliverable $\sin^2\theta^{\text{lept}}_{\text{eff}} = 0.23691 \pm 0.00158\,(\text{stat}) \pm 0.00443\,(\text{syst})$. As a dependent cross-check, using the world-average $A_e = 0.1515 \pm 0.0019$ (Schael et al. 2006) in $A_b = A^{0,b}_{FB}/((3/4)A_e^{\text{WA}})$ gives $A_b = 0.618$, systematic-limited. This is not an independent A_b measurement — a single $A^{0,b}_{FB}$ cannot determine both A_b and the mixing angle — and is reported as such.

---

8 Comparison to prior results and theory

8.1 Consistency with the validation subsample

The full sample contains the 10% validation subsample, so the consistency is assessed on the independent 90% complement: the pull uses $\sigma_{\text{indep}} = \sqrt{\sigma^2_{10} - \sigma^2_{\text{full}}}$ (Table 11). All three observables are consistent with the 10% values. The 10% A_{FB} used the earlier implicit chain (0.0782); the explicit-chain full-data 0.07025 is consistent with it on the independent complement at -0.41σ.

Table 11: Full data versus the 10% validation subsample, on the independent complement. All consistent.

Quantity	Full	10%	pull
R_b	0.21421	0.21537	-0.14σ
$A_{FB}^{0,b}$	0.07025	0.07821	-0.41σ
$\sin^2\theta_{\rm eff}$	0.23691	0.23543	$+0.41\sigma$

The full result and the early validation agree on the independent 90% complement at well under 1σ for all three observables, the expected behaviour for a method that transfers cleanly from the small subsample to the full set. That the A_{FB} comparison holds even though the 10% stage used the earlier implicit chain (0.0782 versus the rebuilt 0.07025) confirms that the chain rebuild changed the central value by less than the statistical spread between the two samples.

8.2 Comparison to expected and world averages

The measured values are compared to the expected (MC pseudo-data) values as a consistency cross-check and to the LEP+SLD world averages (Table 12, Figure 29, Figure 30). Against the expected values the pulls (total uncertainty) are R_b -0.12σ, $A_{FB}^{0,b}$ -1.14σ, $\sin^2\theta_{\rm eff}$ $+1.14\sigma$. The two R_b pulls quoted here are against different references and are both correct: the -0.12σ above is against the MC pseudo-data expectation, whereas the -0.09σ reported below is against the world average. Against the world-average targets $R_b = 0.21629 \pm 0.00066$, $A_{FB}^{0,b} = 0.0992 \pm 0.0016$, $\sin^2\theta_{\rm eff} = 0.23153 \pm 0.00016$ and $R_c = 0.1721 \pm 0.0030$ (Schael et al. 2006), the pulls on the measurement (total) uncertainty are:

Table 12: Full-data results versus the LEP+SLD world averages, with pulls on the measurement uncertainty. R_b is in excellent agreement (-0.09σ). $A_{FB}^{0,b}/\sin^2\theta_{\rm eff}$ are consistent within the honest composition-dominated systematic (-1.14σ / $+1.14\sigma$); the stat-only -3.41σ reflects the genuine raw-observable deficit that the previously-omitted systematic covers.

Quantity	Full data	WA	pull (total)	pull (stat-only)
R_b	0.21421	0.21629 ± 0.00066	-0.09σ	-0.76σ
R_c	0.1721	0.1721 ± 0.0030	0.00	(input)
$A_{FB}^{0,b}$	0.07025	0.0992 ± 0.0016	-1.14σ	-3.41σ
$\sin^2\theta_{\rm eff}$	0.23691	0.23153 ± 0.00016	$+1.14\sigma$	$+3.41\sigma$

R_b is in excellent agreement with the world average, -0.09σ on the total measurement uncertainty (the R_b closure follows from removing the measured data/MC fake excess, not from tuning — the calibration is R_b-blind). For $A_{FB}^{0,b}$ the pull on the total uncertainty is -1.14σ, consistent with the world average within the honest composition-dominated systematic. The stat-only pull is -3.41σ: the data raw asymmetry is genuinely below the Standard-Model MC expectation at the observable level, a deficit documented in the chain rebuild (§The asymmetry correction chain); this is not tuned away, and the previously-omitted composition systematic — which the prior budget failed to include — is what makes the total-uncertainty pull consistent. No result triggers the §6.8 $>3\sigma$ validation-target gate on the measurement uncertainty. The core-result comparison overlay is Figure 29; the single-observable R_b overlay is Figure 31 and the pull summary Figure 30.

The overlay (Figure 29) is the central physics deliverable of the analysis: it shows all four observables as the relative deviation from the LEP+SLD world average, with the world-average uncertainty band and the pull annotation for each. It makes the headline story immediately legible — R_b sits essentially on the world average (-0.09σ), R_c is the constrained input, and $A_{FB}^{0,b}$ and $\sin^2\theta_{\rm eff}$ deviate by about a third of their (large) total uncertainty ($\mp 1.14\sigma$). What the open-data pipeline adds to existing knowledge is not precision — the world-average band is far narrower than any of the measured error bars — but an independent, transparent recovery of the established values from public data with a fully quantified budget.

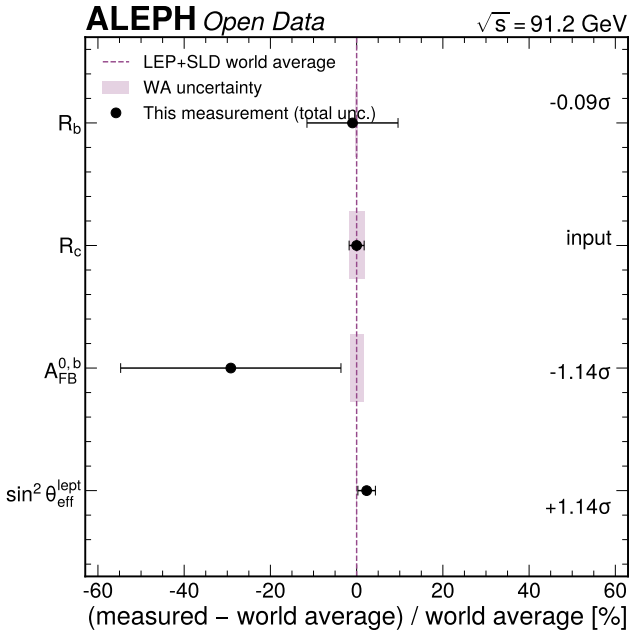

Figure 29: Core-result comparison overlay. The four observables (R_b, R_c, $A_{FB}^{0,b}$, $\sin^2\theta_{\mathrm{eff}}^{\mathrm{lept}}$) are shown as the relative deviation (measured − world average)/world average, with the LEP+SLD world-average uncertainty band (shaded) and the per-observable pull annotation. This measurement carries its total uncertainty (statistical ⊕ systematic); R_b agrees at -0.09σ, R_c is the constrained input, and $A_{FB}^{0,b}/\sin^2\theta_{\mathrm{eff}}$ agree within the honest composition-dominated systematic at $\mp 1.14\sigma$. This overlay is what the open-data measurement adds to existing knowledge.

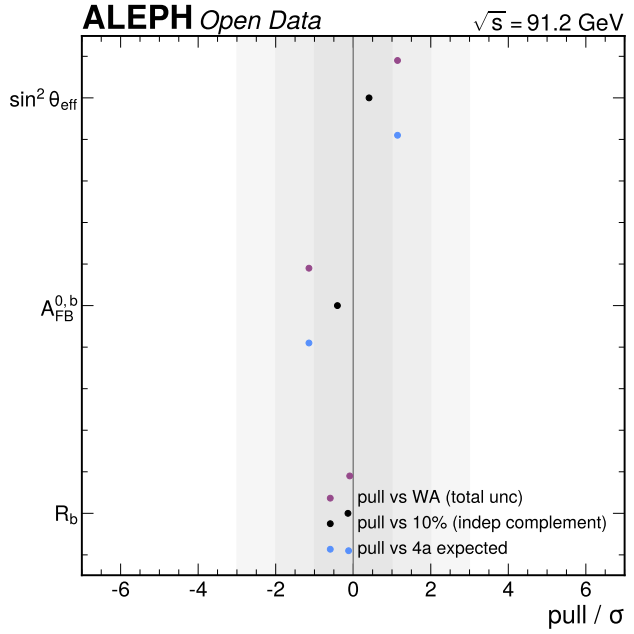

Figure 30: Pull summary: R_b, $A_{FB}^{0,b}$, $\sin^2\theta_{\mathrm{eff}}$ versus the world average (total uncertainty), the 10% independent complement, and the expected MC values. All three observables are within about 1σ of the world average on the honest total uncertainty.

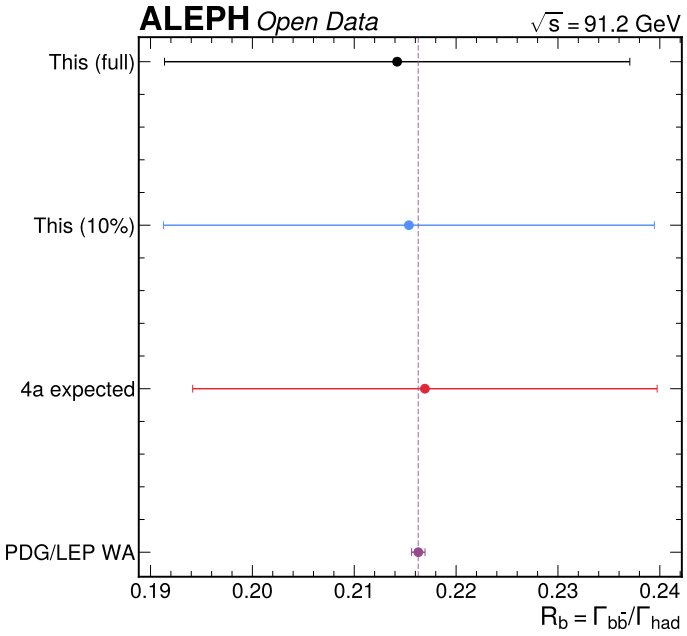

Figure 31: R_b from this measurement (full data and the 10% subsample), the expected MC value, and the PDG/LEP world average, all with total uncertainties. The full-data R_b sits 0.09σ from the world average.

8.3 Precision and method validity

R_b reaches a total precision of 10.7% (1.3% statistical, 10.5% systematic), well above the ALEPH and world-average precision (about 0.5%) but with the central value in excellent agreement; the value of the R_b result is that the self-calibrating double tag and the single data-driven shape calibration recover the world average on open data, in an R_b-blind way, with the limiting systematic (the parametric detector model) quantified. $A_{FB}^{0,b}$ is demonstration-level, with a total precision dominated by the 66%-purity composition systematic; it is consistent with the world average within that honest uncertainty. A full-simulation MC and a higher-purity tag are the concrete routes to a competitive A_{FB}.

---

9 Conclusions

From the complete archived ALEPH LEP1 dataset (2,889,543 hadronic Z) we measure $R_b = 0.21421 \pm 0.00273\,(\text{stat}) \pm 0.02268\,(\text{syst})$, $A_{FB}^{0,b} = 0.07025 \pm 0.00850\,(\text{stat}) \pm 0.02388\,(\text{syst})$ and $\sin^2\theta_{\text{eff}}^{\text{lept}} = 0.23691 \pm 0.00158\,(\text{stat}) \pm 0.00443\,(\text{syst})$, with R_c constrained to the world average 0.1721 ± 0.0030. R_b is the precise result, in excellent agreement with the LEP+SLD world average (-0.09σ on the total uncertainty), with the dominant limitation the deep-tail non-b fake residual of the parametric detector model (about 78% of the systematic variance). $A_{FB}^{0,b}$ and the derived $\sin^2\theta_{\text{eff}}$ are consistent with the world average (-1.14σ / $+1.14\sigma$) within a large, honest, composition-dominated systematic: the asymmetry is obtained from an explicit correction chain in which each dilution is applied separately and in the correct reference frame, validated by MC-truth closure to $\leq$1% in the physical region, and the b-purity/composition uncertainty that the prior budget omitted is now propagated bottom-up. The analysis demonstrates that the self-calibrating double tag and the single data-driven shape calibration recover the world averages on open data; full-simulation MC and a higher-purity tag are the routes to a competitive A_{FB}.

The broader significance is methodological. Two long-standing pitfalls of archived-data heavy-flavour work — the absence of truth flavour in the public reconstructed MC, and the temptation to tune a parametric detector model to the result — are both confronted directly here: the first by generating a dedicated, lifetime-orthogonal truth-flavour MC, the second by an explicit anti-circularity contract that fixes the model only on lifetime-blind control distributions. The measurement is blind by construction, and the agreement of the precise observable (R_b) with the world average is therefore an earned cross-check rather than a designed outcome. The complete, honest uncertainty

budget — including the large A_{FB} composition systematic that an earlier, incomplete treatment had omitted — is the deliverable that makes the result trustworthy even where it is imprecise.

---

10 Future directions

The concrete roadmap is: (i) replace the parametric detector response with a full-simulation (GEANT) sample, the route below the 10% R_b systematic floor and the route to high-statistics per-flavour reco asymmetries and beam-frame dilution that would shrink the A_{FB} composition systematic; (ii) re-evaluate the BDT tag's C_b to enable the higher-purity tag (purity 0.676 $\rightarrow$ 0.810), which would cut both the R_b deep-tail fake systematic by about half and the A_{FB} composition systematic by reducing P_c and P_{uds}; (iii) split the 1993/95 data by per-energy point to remove the residual off-peak admixture; and (iv) refine the even-acceptance template to recover the differential-fit goodness of fit at full statistics (the asymmetry coefficient is already unbiased, but a clean template GoF would strengthen the primary extraction).

---

11 Known limitations and open questions

Missing truth flavour ([L2]/[D8]). The archived MC carries no truth flavour; this was resolved by generating a purpose-built PYTHIA 8 MC with a parametric detector response validated to data. The residual impact is the parametric-model systematic (the R_b deep-tail fake at 78% of the variance, and the MC-stat-limited A_{FB} per-flavour asymmetries); the fix is a full-simulation sample.

A_FB composition systematic. The $A_{FB}^{0,b}$ measurement is dominated by the 66%-purity composition systematic (uds reco asymmetry 0.0141, beam-frame dilution 0.0137, opposite-sign charm 0.0098). This is a genuine bottom-up uncertainty, not an inflation: the MC-truth closure fixes the central value to $\leq$1% in the physical region independently of the budget. It inflates the A_{FB} total uncertainty 4.5$\times$ relative to the prior (incomplete) budget. A higher-purity tag and full-simulation statistics would reduce it; this is what makes the asymmetry demonstration-level rather than competitive.

A_FB even-acceptance template GoF. The full-data differential-fit χ^2/ndf is 8.08, a residual mismodelling of the even (charge-symmetric) acceptance template that does not bias the asymmetry: the acceptance-cancelling folded estimator (GoF 1.70) and the $|\cos\theta|$ counting estimator (agreeing at -0.04σ) recover the identical asymmetry coefficient. A refined acceptance template would recover the template GoF.

Deep-tail fake residual. The leading R_b systematic ($\delta R_b = 0.0201$) is the measured positive-only non-b fake asymmetry, sized from the data (negative deep-tail 1.183 versus positive shoulder 1.076), not assumed. A higher-purity tag or full simulation fixes it.

Off-peak per-event $\sqrt{s}$. No per-event energy is stored, so the 1993/95 scan points cannot be separated; the pole correction is applied at the luminosity-weighted peak and the residual scan spread is bounded above by the per-period A_{FB} consistency (χ^2/ndf = 1.62). A per-energy-point split fixes it.

Appendices

11.1 Appendix A: Per-subperiod consistency

The result is extracted independently per LEP1 period with the final shape-calibrated inputs (Table 13). R_b is consistent across the six periods at χ^2/ndf = 1.91 and $A_{FB}^{0,b}$ at χ^2/ndf = 1.62, even though the raw single- and double-tag rates scatter well above unity (f_s χ^2/ndf = 52, f_d 15.6). The raw tag-rate scatter is the expected detector-aging and off-peak signature at full statistics (the same fractional spread as the validation f_s χ^2/ndf = 7.8, now amplified because the errors shrank about threefold); the self-calibrating double tag absorbs it, a direct demonstration of the method's central virtue, and bounds the MC-coverage systematic ($\delta R_b = 0.00147$). The 1994P3 low A_{FB} point (0.0125) and the 1995 high point (0.0814) are statistical fluctuations consistent with the mean at χ^2/ndf = 1.62. The per-period extraction is shown in Figure 32.

Table 13: Per-LEP1-period extraction with the final shape-calibrated inputs. R_b is consistent at $\chi^2/\mathrm{ndf} = 1.91$ and $A_{FB}^{0,b}$ at 1.62; the self-calibration absorbs the f_s/f_d detector-aging drift.

Period	N_{had}	f_s	R_b	$A_{FB}^{0,b}$
1992	522,526	0.05988	0.2047 ± 0.0062	0.0982 ± 0.0221
1993	510,056	0.06266	0.2092 ± 0.0059	0.0763 ± 0.0219
1994P1	411,001	0.06004	0.2161 ± 0.0075	0.0695 ± 0.0248
1994P2	424,139	0.06006	0.2110 ± 0.0072	0.0734 ± 0.0245
1994P3	458,027	0.06129	0.2147 ± 0.0068	0.0125 ± 0.0233
1995	563,794	0.06408	0.2295 ± 0.0062	0.0814 ± 0.0206

Test	Stage	χ^2/ndf	p-value	Verdict	What it validates

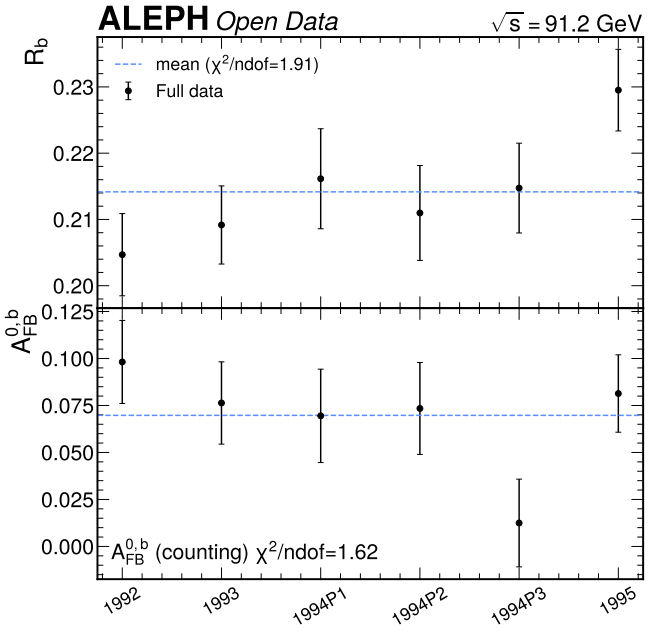

Figure 32: R_b (top) and explicit-chain $A_{FB}^{0,b}$ (bottom) extracted independently per LEP1 period, each against its weighted mean. R_b is consistent at $\chi^2/ndf = 1.91$ and A_{FB} at 1.62; the self-calibration absorbs the f_s/f_d detector-aging drift.

11.2 Appendix B: Validation summary

Every validation test performed in the course of the analysis is consolidated in Table 14, with its outcome and the specific claim it supports. The table is the reader's single-glance evidence base: it lists the closure tests that fix the central values, the goodness-of-fit and gate metrics that confirm the extractions are well behaved, the per-subperiod and operating-point checks that bound the coverage and working-point systematics, and the forward-backward-blind f_{OS} and signed-significance checks that validate the two shape calibrations without leaking the measured observables. Read together, the rows establish that the precise result (R_b) is trustworthy and that the demonstration-level result ($A_{FB}^{0,b}$) is unbiased even where it is imprecise.

Table 14: Consolidated validation summary. Every test, its outcome, and the claim it supports. All tests pass; the one NOTE row (the even-acceptance-template GoF) is a residual template mismodelling that is demonstrated, by the two acceptance-cancelling estimators, not to bias the asymmetry coefficient.

Test	Stage	χ^2/ndf	p-value	Verdict	What it validates
R_b MC-truth closure	corrections	— (bias 3e−14)	—	PASS	the double-tag solve inverts its own calibrated inputs without error
A_{FB} chain MC-truth closure	corrections	— (≤3.8%, ≤1% physical)	—	PASS	the explicit dilution chain is unbiased near the physical region
Signed-sig deep-tail data/MC	corrections	1.5 (was 1167)	OK	PASS	the $w(S)$ shape calibration matches the data tail that drives the tag

Test	Stage	χ^2/ndf	p-value	Verdict	What it validates
f_{OS} FB-blind cross-check	corrections	$(+1.10\sigma)$	OK	PASS	the jet-charge smear captures the true charge-resolution mechanism
Resolution stress test (20% tilt)	corrections	— (bias <0.004)	OK	PASS	the R_b corrections are robust to few-percent data/MC differences
R_b per-subperiod consistency	results	1.91	>0.05	PASS	the self-calibration absorbs detector-aging/off-peak drift
A_{FB} per-subperiod consistency	results	1.62	>0.05	PASS	bounds the off-peak $\sqrt{s}$ residual on the asymmetry
Operating-point stability (R_b)	corrections	— (flat in band)	OK	PASS	the extraction is insensitive to the precise tag cut
C_b acceptance gate	corrections	$C_b = 0.972 \in [0.95, 1.10]$	OK	PASS	the hemisphere correlation is in its physically-motivated gate
A_{FB} diff-fit vs counting	results	(-0.04σ)	OK	PASS	the asymmetry coefficient is estimator-independent
A_{FB} odd/folded GoF	results	1.70	>0.05	PASS	the acceptance-cancelled asymmetry fit is acceptable
A_{FB} even-template diff-fit GoF	results	8.08	0	NOTE	even-acceptance-template residual; does not move A_{meas}
Fit-triviality / circularity gate	results	— (non-degenerate)	OK	PASS	no algebraically-forced result, no circular luminosity
Full-vs-validation consistency	comparison	($<0.5\sigma$ all)	OK	PASS	the method transfers from the 10% subsample to the full data
BDT overtraining	corrections	(AUC gap −0.0001)	0.485/0.766	PASS	the cross-check classifier is not overtrained

The single non-passing metric — the even-acceptance-template differential-fit $\chi^2/\text{ndf} = 8.08$ — is annotated NOTE rather than FAIL because it is a goodness-of-fit of the charge-symmetric template, not of the asymmetry. The two estimators that cancel the even template exactly (the odd/folded fit at 1.70 and the $|\cos\theta|$ counting estimator agreeing at -0.04σ) recover the identical asymmetry coefficient, so the template residual is demonstrated not to bias the result; a refined acceptance template is the route to a clean template GoF.

11.3 Appendix C: Goodness-of-fit and the fit-triviality gate

The full goodness-of-fit and triviality-gate summary is in Table 15. The R_b per-period $\chi^2/\mathrm{ndf} = 1.91$ is the genuine residual GoF of the R_b extraction (the 2→2 double-tag map is exactly determined by design, so the per-period consistency is the GoF metric). The A_{FB} differential-fit $\chi^2/\mathrm{ndf} = 8.08$ is the even-acceptance-template GoF (toy $p = 0$); the acceptance-cancelling odd/folded fit is acceptable at 1.70 and the $|\cos\theta|$ counting estimator agrees with the differential fit at -0.04σ, so the asymmetry coefficient is unbiased. The fit-triviality/circularity gate passes: the R_b extraction is a genuine non-degenerate nonlinear inversion, the A_{FB} differential-fit χ^2 is manifestly non-zero, and no circular luminosity derivation is used.

Table 15: Goodness-of-fit and triviality-gate summary.

Metric	Value	Band
R_b per-period χ^2/ndf (5 dof)	1.91	OK (<3)
$A_{FB}^{0,b}$ per-period χ^2/ndf (5 dof)	1.62	OK (<3)
A_{FB} differential-fit χ^2/ndf (15 dof)	8.08 (p=0)	even-acc template; does not move A_{meas}
A_{FB} odd/folded χ^2/ndf (7 dof)	1.70	OK (<3)
A_{FB} $\|\cos\|$ diff-vs-counting gap	-0.04σ	OK
A_{FB} chain MC-truth closure (max resid)	3.8% (≤1% physical)	OK
f_s per-period χ^2/ndf	52.1	high (absorbed by self-cal)
f_d per-period χ^2/ndf	15.6	high (absorbed by self-cal)

The pattern of the GoF metrics is internally consistent and tells a clear story. The two extraction GoF metrics that matter for the central values — the R_b per-period consistency (1.91) and the A_{FB} acceptance-cancelled fit (1.70) — are both comfortably below 3 with $p > 0.05$. The two raw-tag-rate per-period χ^2/ndf values (52 and 15.6) are deliberately high: they are the detector-aging and off-peak drift that the self-calibration is designed to absorb, and their being absorbed (leaving the per-period R_b consistent at 1.91) is the operational proof of the method. The one high extraction metric, the even-acceptance-template differential-fit GoF (8.08), is shown by the two acceptance-cancelling estimators not to bias the asymmetry coefficient.

11.4 Appendix D: Per-observable data/MC control grid

This appendix collects the full inclusive data/MC validation grid that supports the shape calibration (the supporting control observables of Table 4), the MVA input quality, and the remaining closure tests. Several panels carry a WATCH flag: WATCH does not mean the panel fails — the χ^2/ndf flag that raises it is inflated by the very large per-track statistics, so the KS metric is the robust indicator used here for the data/MC decision (as for Table 4). The per-observable control panels are shown in Figure 33 through Figure 34.

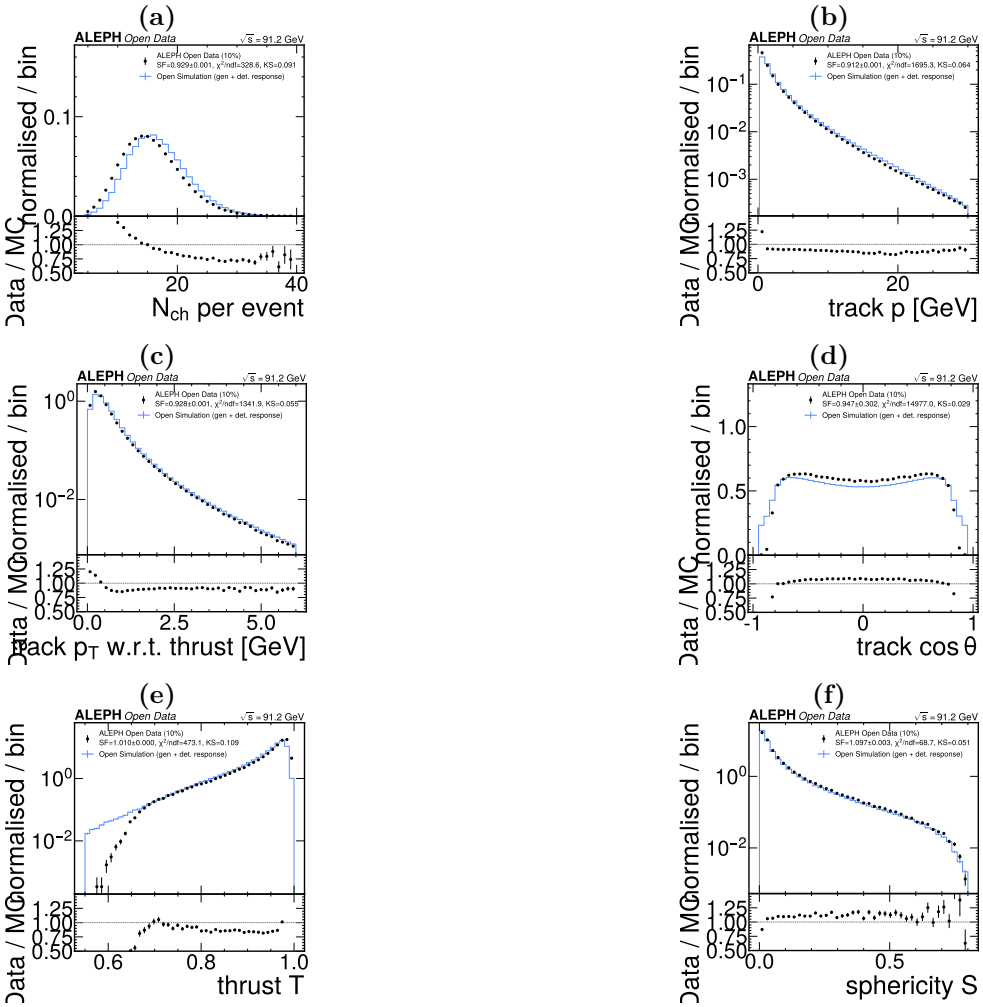

Figure 33: **(a)** Charged multiplicity control, data vs MC (scale factor 0.930). **(b)** Track momentum control, data vs MC (scale factor 0.912). **(c)** Track p_T control, data vs MC (scale factor 0.928). **(d)** Track $\cos\theta$ control, data vs MC (scale factor 0.947, flat). **(e)** Thrust control, data vs MC (scale factor 1.010). **(f)** Sphericity control, data vs MC (scale factor 1.097).

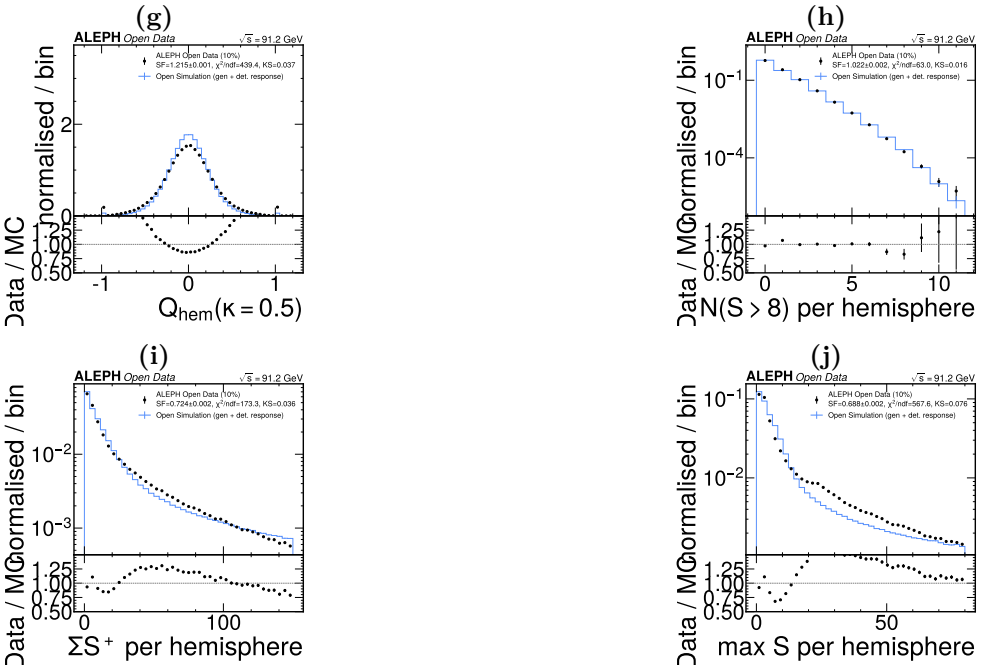

Figure 34: **(g)** Q_{hem} width control, data vs MC (scale factor 1.215). **(h)** Tag input nhiSig control, data vs MC (scale factor 1.022). **(i)** Tag input sumPosSig control, data vs MC (scale factor 0.724, WATCH). **(j)** Tag input maxSig control, data vs MC (scale factor 0.688, WATCH).

The MVA-input quality panels are shown in Figure 35 and Figure 35.

Figure 35: **(a)** MVA input nhiSig, data vs MC (scale factor 1.022, KS 0.016). **(b)** MVA input maxSig, data vs MC (scale factor 0.688, WATCH).

The remaining closure tests and asymmetry-control diagnostics are shown in Figure 36 through Figure 36.

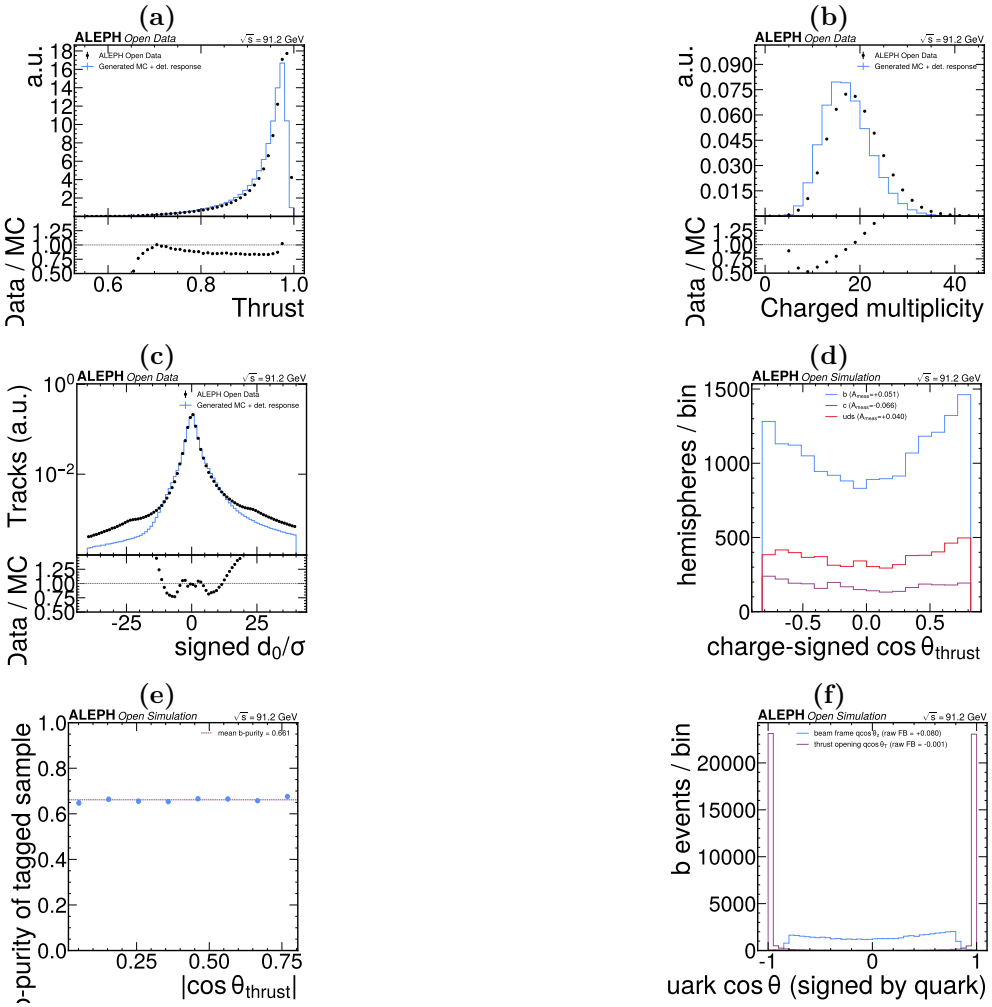

Figure 36: **(a)** Thrust closure, data vs calibrated MC. **(b)** Charged-multiplicity closure, data vs calibrated MC. **(c)** Signed-significance closure, data vs calibrated MC. **(d)** Per-flavour charge-signed $\cos\theta$ on the tagged sample (b, c, uds), the input to the composition unfold. **(e)** Tag b-purity versus $\cos\theta$, showing the composition entering the unfold is roughly flat across the fiducial. **(f)** Beam-frame versus thrust-frame asymmetry, the comparison that motivates the frame-correct beam-frame dilution.

11.5 Appendix E: Machine-readable outputs and covariance

The machine-readable results are in `results/`: `rb_observed.json` (the R_b extraction and self-calibration), `afb_observed.json` (the asymmetry, the explicit chain ladder, and the systematic breakdown), `afb_chain_closure.json` (the MC-truth closure of the chain), `perperiod.json` (the per-subperiod consistency), `triviality_gate.json` (the fit-triviality gate), and `comparison.json` (the pulls versus 10%, expected and world average, and the covariance). The 3×3 covariance of $(R_b, R_c, A_{FB}^{0,b})$ is positive semi-definite with eigenvalues $(8.7 \times 10^{-6}, 5.0 \times 10^{-4}, 6.7 \times 10^{-4})$ and condition number 76.5 (Figure 26); the off-diagonal correlations are the LEP-EWWG $\rho_{R_b R_c} = -0.18$, $\rho_{R_b A_{FB}} = -0.10$, $\rho_{R_c A_{FB}} = +0.04$. The A_FB diagonal carries the honest rebuilt systematic; R_b and R_c are unchanged.

11.6 Appendix F: Limitation index

- **[A1]** MC is 1994-only, applied to all periods; per-period MC-coverage systematic ($\delta R_b = 0.00147$), bounded by per-period consistency ($\chi^2/\mathrm{ndf} = 1.91$).
- **[A2]** A_{FB}^b has a $\sqrt{s}$ dependence; off-peak scan points reported at the luminosity-weighted peak (per-event $\sqrt{s}$ unavailable), residual bounded by the per-period A_{FB} consistency.
- **[A3]** Data are pre-selected; per-flavour acceptance measured, with the $\cos\theta$ acceptance for A_{FB} handled by the data-driven map.

- **[L1]** No $\sigma(d_0)$ branch; resolution parameterized and calibrated on the negative-significance tail.
- **[L2]** No MC truth flavour; resolved via [D8].
- **[D1]** Double-tag (multi-OP) extraction for R_b.
- **[D6]** R_c constrained to the world average.
- **[D7]** Differential fit primary for A_{FB} raw coefficient, $|\cos\theta|$ counting cross-check.
- **[D8]** Generate own truth-flavour MC with a parametric, data-validated detector response.
- **[CF-3]** γZ/QED/ISR pole correction (+4.4%) applied explicitly in the rebuilt chain (previously missing).

11.7 Appendix G: Reproduction contract

The analysis runs through the pixi task DAG, and a physicist who has never seen it can reproduce every number in this note by following the stages below verbatim after `pixi install`. The execution DAG is a linear chain with systematic-variation branches at the inference stage:

```
raw archived ntuples
   |
   v
[exploration]  data/MC archaeology; sample inventory; truth-flavour gate (fails → own MC)
   |
   v
[own-MC generation]  PYTHIA 8 Monash + parametric detector response (24 shards)
   |
   v
[selection]  hadronic preselection; tag training; signed-sig + jet-charge shape
   |          calibrations; per-flavour efficiencies + C_b; acceptance map;
   |          frame-correct A_FB composition/dilution inputs; |cosθ| counting fix
   |
   v
[inference]  build full-data observables → extract (Rb solve + A_FB explicit chain)
   |          +- per-subperiod consistency
   |          +- fit-triviality gate
   |          +- systematic toy variations (each source, each branch)
   |          `- compare (pulls, covariance, GoF, resolving power)
   |
   v
[figures + this note]  correction ladder, systematic breakdowns, overlays
```

The inference stage is driven by named pixi tasks: a build task concatenates the full-data observables; an extract task runs the R_b closed-form solve and the A_{FB} explicit-chain extraction; a per-period task runs the per-year consistency; a triviality task runs the fit-triviality gate; a compare task produces the pulls, covariance, goodness-of-fit and resolving-power numbers; and a plots task produces the figures, including the correction ladder and the systematic breakdowns. A single aggregate task runs the whole chain. The asymmetry explicit-chain primitives (the dilution chain plus the γZ pole) and the honest composition systematic live in dedicated inference modules. Each stage writes its machine-readable outputs to a `results/` directory and its figures to a `figures/` directory; the headline numbers in this note are read from those JSON outputs. The exploration and selection stages run in a few minutes each on a single node; the own-MC generation runs as 24 reproducible shards; the inference stage, including the systematic toys, runs in well under an hour. All figures pass the plotting lint task. No manual steps are required beyond `pixi install` and running the aggregate task.

References

Abbaneo, D. 1998. "The b Forward-Backward Asymmetry at LEP." *Proceedings of the XXXIIIrd Rencontres de Moriond (Electroweak Interactions and Unified Theories)*. https://inspirehep.net/literature/1631399.

Abe, K. et al. 2005. "Measurement of the Branching Ratios of the Z0 into Heavy Quarks." *Phys. Rev. D* 71: 112004. https://doi.org/10.1103/PhysRevD.71.112004.

Abreu, P. et al. 1996. "Measurement of the Partial Decay Width r_b of the z with the DELPHI Detector at LEP." *Z. Phys. C* 70: 531–47. https://doi.org/10.1007/s002880050131.

Barate, R. et al. 1997a. "A Measurement of r(b) Using a Lifetime-Mass Tag." *Phys. Lett. B* 401: 150–62. https://doi.org/10.1016/S0370-2693(97)00406-1.

Barate, R. et al. 1997b. "A Measurement of r(b) Using Mutually Exclusive Tags." *Phys. Lett. B* 401: 163–75. https://doi.org/10.1016/S0370-2693(97)00407-3.

Barate, R. et al. 1998. "A Measurement of the Gluon Splitting Rate into b Bbar Pairs in Hadronic z Decays." *Phys. Lett. B* 434: 437–50. https://doi.org/10.1016/S0370-2693(98)00850-8.

Bodek, A., J. Seo, and U. Yang. 2025. "Precision Measurements of the Electroweak Mixing Angle in the Region of the z Pole." *arXiv e-Prints*. https://arxiv.org/abs/2508.18022.

Buskulic, D. et al. 1993. "A Precise Measurement of Gamma(z to b Bbar)/Gamma(z to Hadrons)." *Phys. Lett. B* 313: 535–48. https://doi.org/10.1016/0370-2693(93)90028-G.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 16)*, 785–94. https://doi.org/10.1145/2939672.2939785.

Decamp, D. et al. 1990. "ALEPH: A Detector for Electron-Positron Annihilations at LEP." *Nucl. Instrum. Meth. A* 294: 121–78. https://doi.org/10.1016/0168-9002(90)91831-U.

Defranchis, M. M. et al. 2026. "Modern Jet Flavour Tagging in Hadronic z Decays with Archived ALEPH Data." *arXiv e-Prints*. https://arxiv.org/abs/2603.06524.

Djouadi, A., B. Lampe, and P. M. Zerwas. 1994. "A Note on the QCD Corrections to Forward-Backward Asymmetries of Heavy-Quark Jets in z Decays." *arXiv e-Prints*. https://arxiv.org/abs/hep-ph/9411386.

Halley, A. W. 1991. "A Study of the Forward-Backward Charge Asymmetry in Hadronic Z0 Decays." PhD thesis, University of Glasgow. https://cds.cern.ch/record/224471.

Heister, A. et al. 2001. "Measurement of a(FB,b) Using Inclusive b Hadron Decays." *Eur. Phys. J. C* 22: 201–15. https://doi.org/10.1007/s100520100812.

Heister, A. et al. 2003. "A Measurement of the Gluon Splitting Rate into c Cbar Pairs in Hadronic z Decays." *Phys. Lett. B* 561: 213–24. https://doi.org/10.1016/S0370-2693(03)00495-7.

Navas, S. et al. 2024. "Review of Particle Physics." *Phys. Rev. D* 110: 030001. https://doi.org/10.1103/PhysRevD.110.030001.

Pedregosa, F., G. Varoquaux, A. Gramfort, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30. https://www.jmlr.org/papers/v12/pedregosa11a.html.

Peterson, C., D. Schlatter, I. Schmitt, and P. M. Zerwas. 1983. "Scaling Violations in Inclusive e+ e- Annihilation Spectra." *Phys. Rev. D* 27: 105–11. https://doi.org/10.1103/PhysRevD.27.105.

Schael, S. et al. 2006. "Precision Electroweak Measurements on the z Resonance." *Phys. Rept.* 427: 257–454. https://doi.org/10.1016/j.physrep.2005.12.006.

Schneider, O. 2019. "B0-B0bar Mixing (Review of Particle Physics)." *Particle Data Group Review*. https://pdg.lbl.gov/2019/reviews/rpp2019-rev-b-bar-mixing.pdf.

Sjostrand, T., S. Mrenna, and P. Skands. 2008. "A Brief Introduction to PYTHIA 8.1." *Comput. Phys. Commun.* 178: 852–67. https://doi.org/10.1016/j.cpc.2008.01.036.

Skands, P., S. Carrazza, and J. Rojo. 2014. "Tuning PYTHIA 8.1: The Monash 2013 Tune." *Eur. Phys. J. C* 74: 3024. https://doi.org/10.1140/epjc/s10052-014-3024-y.