

Operator identification in charged lepton-flavor violation: Global EFT analysis with RG evolution, polarization observables, and Bayesian model discrimination at future colliders

Nicolas Viaux M. 

*Departamento de Física, Universidad Técnica Federico Santa María,
Casilla 110-V, Valparaíso, Chile and*

Millennium Institute for Subatomic Physics at the High-Energy Frontier (SAPHIR), Santiago, Chile



(Received 14 February 2026; accepted 13 April 2026; published 6 May 2026)

Charged lepton-flavor violation is a null-test frontier of the Standard Model and a direct probe of physics beyond it. We present a global effective field theory (EFT) analysis across FCC-ee, ILC, CLIC, HL-LHC, HE-LHC, and muon colliders at 3 and 10 TeV, with operator identification as the primary target rather than exclusion reach alone. The analysis combines low-energy constraints, collider differential observables, and Dalitz-level $\mu \rightarrow 3e$ information in a common profile-likelihood framework. Key hadron-collider and muon-collider signal/background samples are generated at the event level and propagated through Delphes detector simulation, while clean e^+e^- benchmark channels are modeled with a parametric response calibrated to published detector-performance benchmarks. We include one-loop renormalization group (RG) running and operator mixing between UV matching and measurement scales, finding 10–30% shifts in selected operator-correlation entries when comparing tree-level and RG-evolved coefficient mappings at multi-TeV matching scales. Polarization asymmetries are used to separate c_{He} and $c_{H\bar{e}}$ directions, and UV discrimination is quantified with Bayes factors for benchmark leptoquark and heavy-neutral-lepton hypotheses. The full code chain for event generation, detector response, inference, and figure reproduction is provided.

DOI: [10.1103/9n1b-zzf1](https://doi.org/10.1103/9n1b-zzf1)

I. INTRODUCTION

The search for charged lepton-flavor violation (LFV), including transitions such as $\mu \rightarrow e\gamma$, $\mu \rightarrow 3e$, $\mu N \rightarrow eN$, $Z \rightarrow e\mu$, and $h \rightarrow \tau\mu$, is a central strategy in precision probes of new physics [1,2]. In the minimal Standard Model extended only by neutrino masses, the charged LFV amplitudes are suppressed to unobservable levels, implying that any positive signal at planned sensitivities is direct evidence for beyond the Standard Model (BSM) dynamics [1].

The field has two mature but partially disconnected directions. The first is the low-energy frontier, where MEG/MEG II, Mu3e, Mu2e, COMET, Belle II, and related programs deliver very strong constraints [3–8]. The second is the collider frontier, where HL-LHC/HE-LHC and future lepton or muon colliders can probe LFV through rare decays and contact-interaction tails with different kinematic leverage [9–12]. Most existing collider forecasts are optimized for channel-by-channel sensitivity. This is necessary but not sufficient for model discrimination.

The central point of this paper is that *operator identification* is a distinct objective from exclusion reach. If several operator classes contribute simultaneously, similar total event yields can emerge from different points in Wilson-coefficient space; therefore, the key quantity becomes the geometry of the global likelihood and the correlation pattern among operators. We develop a unified framework that quantifies this geometry and compares collider programs based on their ability to break degeneracies.

Our starting ingredients are an effective field theory (EFT) parametrization of LFV interactions with dipole, Higgs-current, and four-fermion structures, a common statistical treatment of low-energy and collider observables, differential collider information (mass tails, angular observables, beam polarization tags), and explicit operator-identification metrics based on the Fisher/Hessian structure of the profile likelihood.

Relative to standard reach projections, the novelty introduced here is the *tomographic* viewpoint: colliders are ranked not only by signal significance but by how strongly they rotate and compress allowed regions in multiparameter EFT space. This formulation is designed for future publication-level detector studies and allows straightforward insertion of official experimental performance cards.

Concretely, the analysis goal is split into three levels. The first level is *detection*: test the null hypothesis and quantify

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

local/global significance. The second level is *estimation*: extract confidence intervals in individual Wilson coefficients after profiling nuisance parameters. The third level is *identification*: measure whether two operator hypotheses remain statistically distinguishable once all channels and correlations are combined. In this work, the third level is promoted to a primary optimization target for future-collider strategy.

The practical implication is that collider comparison must be done in a common parameter basis and with a common nuisance model. A high-significance channel can still provide limited model information if it aligns with an existing low-energy degeneracy direction. Conversely, a moderate-significance channel can be decisive if it is approximately orthogonal in coefficient space. This distinction motivates our emphasis on differential observables, polarization categories, and global profile-likelihood geometry. Our benchmark assumptions are aligned with the current European planning context for future accelerator programs [13].

The manuscript is organized to match the three pillars highlighted in the title: EFT operator identification, renormalization group (RG) aware interpretation, and model discrimination. The numerical layer now combines detector-level Delphes samples for hadron and muon collider benchmarks with lepton-collider response models calibrated to published detector-performance benchmarks, and all channels are merged in one statistical inference chain. Results should therefore be interpreted as detector-informed benchmark projections under the stated cards, priors, and analysis choices. For hadron-collider validation channels, the event-generation chain is explicitly MadGraph5_aMC@NLO $\rightarrow$ Pythia8 $\rightarrow$ Delphes, while other benchmark channels use the same detector-level template formalism with the corresponding input assumptions [14–16].

II. EFT SETUP

A. Broken-phase LFV parametrization

We use the broken-phase effective Lagrangian:

$$\mathcal{L}_{\text{LFV}} = \mathcal{L}_D + \mathcal{L}_H + \mathcal{L}_{4f}, \quad (1)$$

$$\begin{aligned} \mathcal{L}_D = \sum_{i \neq j} \frac{1}{\Lambda^2} & (C_{D\gamma,R}^{ij} \mathcal{O}_{D\gamma,R}^{ij} + C_{D\gamma,L}^{ij} \mathcal{O}_{D\gamma,L}^{ij} \\ & + C_{DZ,R}^{ij} \mathcal{O}_{DZ,R}^{ij} + C_{DZ,L}^{ij} \mathcal{O}_{DZ,L}^{ij}) + \text{H.c.}, \end{aligned} \quad (2)$$

$$\begin{aligned} \mathcal{L}_H = \sum_{i \neq j} & \left[\frac{C_{H\ell}^{ij}}{\Lambda^2} (H^\dagger i \overleftrightarrow{D}_\mu H) (\bar{\ell}_i \gamma^\mu P_L \ell_j) \right. \\ & \left. + \frac{C_{He}^{ij}}{\Lambda^2} (H^\dagger i \overleftrightarrow{D}_\mu H) (\bar{e}_i \gamma^\mu P_R e_j) \right], \end{aligned} \quad (3)$$

$$\mathcal{L}_{4f} = \sum_{i \neq j, f} \frac{C_{4f}^{ij}}{\Lambda^2} (\bar{\ell}_i \Gamma \ell_j) (\bar{f} \Gamma f) + \text{H.c.}, \quad (4)$$

with $i, j \in \{e, \mu, \tau\}$ and f SM fermions. This representation is mapped from SMEFT operators at the electroweak scale [17,18]. The dipole shorthand operators are

$$\begin{aligned} \mathcal{O}_{D\gamma,R/L}^{ij} & \equiv (\bar{\ell}_i \sigma^{\mu\nu} P_{R/L} \ell_j) F_{\mu\nu}, \\ \mathcal{O}_{DZ,R/L}^{ij} & \equiv (\bar{\ell}_i \sigma^{\mu\nu} P_{R/L} \ell_j) Z_{\mu\nu}. \end{aligned} \quad (5)$$

For the benchmark numerical fit, we keep vector-current structures with explicit chirality labels:

$$\begin{aligned} \mathcal{L}_{4f}^{\text{fit}} = \sum_{i \neq j, f} \frac{1}{\Lambda^2} & (C_{LL}^{ij,f} \mathcal{O}_{LL}^{ij,f} + C_{LR}^{ij,f} \mathcal{O}_{LR}^{ij,f} + C_{RL}^{ij,f} \mathcal{O}_{RL}^{ij,f} \\ & + C_{RR}^{ij,f} \mathcal{O}_{RR}^{ij,f}) + \text{H.c.}, \end{aligned} \quad (6)$$

with

$$\mathcal{O}_{XY}^{ij,f} \equiv (\bar{\ell}_i \gamma_\mu P_X \ell_j) (\bar{f} \gamma^\mu P_Y f), \quad X, Y \in \{L, R\}, \quad (7)$$

and define c_{4f} as the benchmark effective direction after this basis reduction. Scalar and tensor directions are deferred to detector-level extensions. In the benchmark cards, the fermion sum is restricted to first-generation quark currents relevant for coherent conversion and high-mass dilepton tails, $f \in \{u, d\}$, and the effective reduction is

$$c_{4f}^{ij} \equiv \sum_{f=u,d} \sum_{X,Y \in \{L,R\}} \omega_{XY}^{(f)} c_{XY}^{ij,f}, \quad (8)$$

with benchmark weights

$$\omega_{LL}^{(u,d)} = \omega_{LR}^{(u,d)} = \omega_{RL}^{(u,d)} = \omega_{RR}^{(u,d)} = \frac{1}{8}. \quad (9)$$

This choice is a neutral benchmark averaging prescription; detector-level studies can replace it by process-calibrated weights extracted from category-specific templates. In benchmark sensitivity checks, varying this prescription within physically motivated ranges (for example doubling the left-left (LL) weight relative to right-right (RR) and renormalizing the sum) shifts the extracted c_{4f} limit by less than 8%, without changing the complementarity ranking across collider classes.

For a generic collider observable a , the cross section expands as

$$\sigma_a = \sigma_{a,\text{SM}} + \frac{1}{\Lambda^2} \sum_\alpha C_\alpha \sigma_{a,\alpha}^{\text{int}} + \frac{1}{\Lambda^4} \sum_{\alpha,\beta} C_\alpha C_\beta \sigma_{a,\alpha\beta}^{\text{BSM}}. \quad (10)$$

In channels with negligible SM LFV amplitudes, the leading contribution is frequently $\mathcal{O}(\Lambda^{-4})$.

To connect with practical rate estimates, we write the schematic LFV amplitude as

$$\mathcal{M}_a = \sum_{\alpha} c_{\alpha} \mathcal{M}_{a,\alpha}, \quad (11)$$

where the reduced coefficients c_{α} carry all model dependence and $\mathcal{M}_{a,\alpha}$ are channel-dependent kinematic structures. The event yield in a signal region then takes the quadratic form

$$N_a = \mathcal{L}_a \sum_{\alpha\beta} c_{\alpha} c_{\beta} K_{\alpha\beta}^{(a)}, \quad (12)$$

with acceptance and efficiency absorbed into $K_{\alpha\beta}^{(a)}$. Equation (12) is the algebraic object entering our template and likelihood construction.

B. Reduced coefficients and mapping to observables

We define dimensionless reduced coefficients

$$\mathbf{c} \equiv \begin{pmatrix} c_D^{e\mu}, c_{H\Sigma}^{e\mu}, c_{4f}^{e\mu}, \\ c_D^{\tau\mu}, c_{H\Sigma}^{\tau\mu}, c_{4f}^{\tau\mu}, \dots \end{pmatrix}, \quad c_{\alpha} \equiv C_{\alpha} \left(\frac{1 \text{ TeV}}{\Lambda} \right)^2, \quad (13)$$

with effective dipole magnitude

$$|c_D^{ij}|^2 \equiv |c_{D,R}^{ij}|^2 + |c_{D,L}^{ij}|^2. \quad (14)$$

In benchmark one-dimensional scans, we use a real signed effective dipole parameter c_D^{ij} defined by a single-direction compression at fixed phase convention. This means sign/phase-sensitive effects from general complexes ($c_{D,L}, c_{D,R}$) are not fully retained in the benchmark cards; those effects are part of the detector-level extension with explicit chiral and phase degrees of freedom. For compact global-fit projections, we define

$$c_{H\Sigma}^{ij} \equiv c_{He}^{ij} + c_{H\ell}^{ij}, \quad (15)$$

and use $c_{H\Sigma}$ as the effective current-type LFV direction constrained by the benchmark channel set. The separated ($c_{He}, c_{H\ell}$) directions are retained in the formalism and can be unfolded in polarization-resolved fits; the one-direction

compression is used only in the benchmark cards to keep the scan dimensionality controlled. At low energies, representative dependencies are

$$\text{BR}(\mu \rightarrow e\gamma) \propto |c_D^{e\mu}|^2, \quad (16)$$

$$\text{CR}(\mu N \rightarrow eN) \propto |a_D c_D^{e\mu} + a_V c_{4f}^{e\mu} + a_H c_{H\Sigma}^{e\mu}|^2, \quad (17)$$

$$\text{BR}(\tau \rightarrow \mu\gamma) \propto |c_D^{\tau\mu}|^2. \quad (18)$$

The linear combination in Eq. (17) illustrates why conversion data can constrain combinations strongly but still leave degeneracy directions in $(c_D, c_{H\Sigma}, c_{4f})$ space [19,20]. For the benchmark Al-target card, we use $(a_D, a_V, a_H) = (0.20, 1.00, 0.30)$ after normalization to the vector contribution.

For orientation, representative low-energy decay widths can be organized as

$$\Gamma(\mu \rightarrow e\gamma) = \frac{m_{\mu}^3}{16\pi\Lambda^4} (|A_R|^2 + |A_L|^2), \quad (19)$$

$$A_{L,R} \propto c_{D,L/R}^{e\mu}, \quad (20)$$

while coherent conversion is parametrized by

$$\Gamma_{\text{conv}} \propto |D c_D^{e\mu} + V^{(p)} c_{V,p}^{e\mu} + V^{(n)} c_{V,n}^{e\mu} + S c_S^{e\mu}|^2. \quad (21)$$

These structures make the origin of coefficient-space blind directions explicit and motivate collider information that probes different linear combinations.

C. Where operators appear in observables and why

The benchmark basis projects the dominant charged-LFV Warsaw structures at $\mu \sim m_Z$ into dipole, Higgs current, and vector four-fermion classes, because these control the leading scaling of radiative decays, conversion amplitudes, resonance LFV decays, and high- q^2 tails [1,17,18] (summarized in Table I).

The dipole operators in Eq. (2) appear directly in radiative transitions because they couple the LFV lepton bilinear to field-strength tensors. This is why $\mu \rightarrow e\gamma$ and $\tau \rightarrow \mu\gamma$ are dominantly sensitive to c_D combinations. The Higgs-current operators in Eq. (3) modify effective LFV

TABLE I. Operator directions used in the benchmark fit and their dominant observable leverage. Coefficients are normalized as $c_{\alpha} = C_{\alpha} (1 \text{ TeV}/\Lambda)^2$.

Label	Definition	Mapping	Main observables
$c_{D,L/R}^{ij}$	$(\bar{\ell}_i \sigma^{\mu\nu} P_{L/R} \ell_j) F_{\mu\nu}, (\bar{\ell}_i \sigma^{\mu\nu} P_{L/R} \ell_j) Z_{\mu\nu}$	$A_{L/R} \propto c_{D,L/R}^{ij}$	$\mu \rightarrow e\gamma, \tau \rightarrow \mu\gamma$, subleading $Z \rightarrow \ell_i \ell_j$
$c_{He}^{ij}, c_{H\ell}^{ij}$	$(H^{\dagger} i \overleftrightarrow{D}_{\mu} H) (\bar{\ell}_i \gamma^{\mu} P_L \ell_j), (H^{\dagger} i \overleftrightarrow{D}_{\mu} H) (\bar{\ell}_i \gamma^{\mu} P_R \ell_j)$	$c_{H\Sigma}^{ij} = c_{He}^{ij} + c_{H\ell}^{ij}$ (benchmark)	$Z \rightarrow \ell_i \ell_j, h \rightarrow \ell_i \ell_j$, polarized asymmetries
c_{4f}^{ij}	Effective vector combination of LL, LR, RL, RR four-fermion currents	$c_{4f}^{ij} = \sum_{X,Y} \omega_{XY} c_{XY}^{ij,f}$	$\mu N \rightarrow eN$, high- $m_{\ell\ell}$ tails

neutral-current couplings and therefore contribute naturally to $Z \rightarrow \ell_i \ell_j$ and Higgs-mediated channels; in reduced one-direction projections, they appear through $c_{H\Sigma}$. Four-fermion operators in Eq. (4) enter contact-interaction amplitudes and become increasingly visible in high-energy tails where momentum transfer is large.

At amplitude level, the same logic can be written as

$$\mathcal{M}(Z \rightarrow \ell_i \ell_j) \sim c_{H\Sigma}^{ij} + \frac{m_Z}{v} c_D^{ij}, \quad (22)$$

$$\mathcal{M}(h \rightarrow \ell_i \ell_j) \sim c_{H\Sigma}^{ij}, \quad (23)$$

$$\mathcal{M}(\ell\ell \rightarrow \ell_i \ell_j) \sim c_{4f}^{ij} + \frac{q^2}{v^2} c_{H\Sigma}^{ij} + \frac{m_\ell}{\sqrt{q^2}} c_D^{ij}, \quad (24)$$

where the expressions are scaling relations in the reduced-coefficient normalization rather than full matrix elements. The last term carries the chirality-flip suppression expected for dipole insertions in hard scattering, which keeps the scaling dimensionally consistent in this reduced form (arising from the helicity-flip factor $m_\ell/\sqrt{s}$ in the amplitude, combined with hard-scale propagator behavior). This clarifies why different channels emphasize different directions in coefficient space. The analysis strategy follows directly from this structure: radiative and low-energy channels anchor dipole terms, resonance decays isolate current structures, and high- q^2 tails isolate contact interactions.

III. COLLIDER PROGRAM AND ANALYSIS CHANNELS

A. Signal and background definitions

We include three collider classes with explicit signal/background maps.

(a) *Lepton colliders* (e^+e^-):

$$e^+e^- \rightarrow Z/h \rightarrow \ell_i^\pm \ell_j^\mp \quad (i \neq j), \quad (25)$$

$$e^+e^- \rightarrow \ell_i^\pm \ell_j^\mp \quad \text{via contact operators.} \quad (26)$$

Main backgrounds are $e^+e^- \rightarrow \tau^+\tau^-$ with leptonic decays, $W^+W^- \rightarrow \ell\nu\ell\nu$, and ZZ final states.

(b) *Hadron colliders* (pp):

$$pp \rightarrow \ell_i^\pm \ell_j^\mp \quad \text{in high-}m_{\ell\ell} \text{ tails,} \quad (27)$$

with dominant backgrounds from $Z/\gamma^* \rightarrow \tau\tau$, dibosons, top production, and nonprompt/fake leptons.

(c) *Muon colliders* ($\mu^+\mu^-$):

$$\mu^+\mu^- \rightarrow \ell_i^\pm \ell_j^\mp, \quad (28)$$

with irreducible electroweak backgrounds and beam-induced overlay considerations [21].

B. Differential observables and templates

For each channel a , we define

$$\mathbf{x}_a = (m_{\ell\ell}, p_T, \eta, \Delta\phi, \cos\theta^*, E_T^{\text{miss}}, \text{pol. tags}, \dots) \quad (29)$$

and construct binned templates for each operator hypothesis. Differential observables are critical because operator classes imprint different energy and angular scaling, improving discrimination beyond inclusive rates [22,23].

In practice, we build per-bin expectations as

$$\nu_{ab}(\mathbf{c}, \boldsymbol{\theta}) = \sum_{\alpha, \beta} c_\alpha c_\beta T_{\alpha\beta}^{ab}(\boldsymbol{\theta}) + B_{ab}(\boldsymbol{\theta}), \quad (30)$$

where $T_{\alpha\beta}^{ab}$ are response templates (including detector effects in future upgrades) and B_{ab} is the background prediction. Equation (30) makes explicit that shape information enters not only through signal strength but through interference structure across bins.

IV. STATISTICAL INFERENCE AND IDENTIFICATION METRIC

A. Global likelihood with nuisances

The full likelihood is

$$\mathcal{L}(\mathbf{c}, \boldsymbol{\theta}) = \prod_{a \in \text{channels}} \prod_{b \in \text{bins}} \text{Pois}(n_{ab} | \nu_{ab}(\mathbf{c}, \boldsymbol{\theta})) \times \prod_k \pi_k(\theta_k), \quad (31)$$

where $\boldsymbol{\theta}$ are nuisance parameters (normalization, shape, luminosity, calibration, theory priors). We use the profile-likelihood ratio

$$q(\mathbf{c}) = -2 \ln \frac{\mathcal{L}(\mathbf{c}, \hat{\boldsymbol{\theta}})}{\mathcal{L}(\hat{\mathbf{c}}, \hat{\boldsymbol{\theta}})} \quad (32)$$

and quote 95% confidence level (CL) constraints from asymptotic test statistics [24].

Systematic effects are implemented through correlated nuisance blocks,

$$\boldsymbol{\theta} = (\theta_{\text{lumi}}, \theta_{\text{lepID}}, \theta_{\text{bkg norm}}, \theta_{\text{shape}}, \theta_{\text{th}}, \dots), \quad (33)$$

with Gaussian or log-normal priors depending on nuisance type. For the Asimov dataset, the covariance-implied Hessian provides the local information matrix around the best fit:

$$F_{\alpha\beta} \simeq \frac{1}{2} \frac{\partial^2 q(\mathbf{c})}{\partial c_\alpha \partial c_\beta} \Big|_{\mathbf{c}=\hat{\mathbf{c}}}. \quad (34)$$

In hadron-collider tail categories, collider-theory systematics are implemented as bin-correlated shape nuisances,

$$\nu_{ab} \rightarrow \nu_{ab}(1 + \theta_{\text{PDF}}\Delta_{ab}^{\text{PDF}} + \theta_{\text{scale}}\Delta_{ab}^{\text{scale}}), \quad (35)$$

with benchmark prior widths $\sigma(\theta_{\text{PDF}}) = 0.03$ and $\sigma(\theta_{\text{scale}}) = 0.05$.

B. Cut-flow and classifier equations

Each signal region follows a benchmark selection: preselection with opposite-sign different-flavor leptons in fiducial acceptance, then mass-window or high-tail selection, then topology suppression through E_T^{miss} , jet/ b vetoes and angular cuts, and finally multivariate analysis (MVA) threshold optimization. Here, we define stage efficiencies

$$\epsilon_{S,k} = \frac{S_k}{S_{k-1}}, \quad \epsilon_{B,k} = \frac{B_k}{B_{k-1}}, \quad (36)$$

and stage significance estimators

$$Z_k^{\text{stat}} = \frac{S_k}{\sqrt{B_k}}, \quad Z_k^{\text{sys}} = \frac{S_k}{\sqrt{B_k + (\delta_B B_k)^2}}. \quad (37)$$

The Z_k^{sys} expression is used only as a prefit diagnostic during selection tuning; final intervals and significances in this work are always taken from the profiled likelihood in Eq. (31).

The classifier $f_\phi(\mathbf{x}) \in [0, 1]$ is trained with weighted binary cross-entropy:

$$\mathcal{L}_{\text{BCE}}(\phi) = -\sum_n w_n [y_n \log f_\phi(\mathbf{x}_n) + (1 - y_n) \log(1 - f_\phi(\mathbf{x}_n))]. \quad (38)$$

For each category, we optimize Z_A ,

$$Z_A = \sqrt{2 \left[(S + B) \ln \left(1 + \frac{S}{B} \right) - S \right]}, \quad (39)$$

which is the standard full Asimov form for counting experiments, not the $S \ll B$ linearized limit, and we use its nuisance-profiled generalization in systematic-dominated bins.

The optimized classifier threshold τ^* is extracted from

$$\tau^* = \arg \max_\tau Z_A(S(\tau), B(\tau)), \quad (40)$$

subject to minimum-yield and control-region stability conditions. This prevents selection of pathological working points where apparent statistical gain is not robust against nuisance pulls.

C. Machine learning implementation and samples

Operationally, the machine learning (ML) output is treated as a physics observable. The classifier score is binned and included directly in Eq. (31), so nuisance constraints and profile pulls are propagated in the global fit [22,23].

The benchmark ML workflow follows a standard sequence of preselection, feature construction, class-weighted gradient-boosted decision trees (GBDT) training, validation-based hyperparameter choice, score calibration, and score-binning for the profile likelihood.

The benchmark implementation uses GBDT in binary classification mode. The baseline training configuration is maximum tree depth = 3, number of boosting stages = 400, learning rate = 0.05, row subsampling = 0.8, minimum leaf population = 50, and no monotonic constraints. Training minimizes weighted binary cross-entropy with per-event weights:

$$\begin{aligned} w_n &= \frac{\mathcal{L}_{p(n)} \epsilon_{p(n)}}{N_{p(n)}^{\text{gen}}} \\ &= \mathcal{L}_{p(n)} \epsilon_{p(n)} (N_{p(n)}^{\text{gen}})^{-1}, \end{aligned} \quad (41)$$

where $p(n)$ labels the process of event n .

For each collider category, samples are split into 60% training, 20% validation, and 20% test sets after baseline preselection, with approximately $\mathcal{O}(5 \times 10^4)$ signal and $\mathcal{O}(2 \times 10^5)$ background events in training before luminosity reweighting.

Feature sets are category dependent. The common core is

$$\{m_{\ell\ell}, p_{T,\ell_1}, p_{T,\ell_2}, \Delta\phi_{\ell\ell}, \cos\theta^*, E_T^{\text{miss}}\}, \quad (42)$$

with collider-specific extensions for recoil/polarization tags, jet-activity observables, and visible-energy/overlay-mitigation observables. Hyperparameters are selected by validation area under the receiver operating characteristic (ROC) curve (AUC) with overtraining control from Kolmogorov-style train-test checks of classifier-score distributions in signal and background.

The classifier output is transformed to a monotonic score and binned into six quantile bins per category for the likelihood model. Threshold optimization uses Eq. (31) inputs and retains only working points that satisfy minimum expected yields in both signal and control regions.

Figure 5 corresponds to representative AUC values of approximately 0.90 (FCC-ee-like), 0.86 (HE-LHC-like), and 0.74 (cut baseline). Stability tests on the held-out test set with $\pm 1\sigma$ nuisance morphing of background templates give absolute AUC shifts below 0.02 in this benchmark setup. In addition, the relative change in expected profile-likelihood sensitivity after nuisance morphing stays below 5% across benchmark categories.

To avoid double counting, kinematic variables used to build the classifier are not entered as independent template

axes in the same category once score binning is used. The fit therefore uses either a score-binned representation or an explicit kinematic-binned representation per category, with migration correlations carried by shared nuisances.

D. Operator-identification metric

The motivation for the operator-identification metric is operational: discovery-level significance alone does not quantify whether two EFT hypotheses remain distinguishable after profiling systematics and combining channels. The relevant quantity for this purpose is the local curvature geometry of the profile likelihood in coefficient space, because it directly controls degeneracy directions and expected covariance. We therefore use a normalized Fisher/Hessian correlation as a compact identification diagnostic [22,24].

To quantify separability, we use

$$\mathcal{I}_{\alpha\beta} \equiv \frac{|F_{\alpha\beta}|}{\sqrt{F_{\alpha\alpha}F_{\beta\beta}}}, \quad F_{\alpha\beta} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial c_\alpha \partial c_\beta} \right\rangle. \quad (43)$$

Smaller $\mathcal{I}_{\alpha\beta}$ indicates stronger operator discrimination. Values near unity indicate near-aligned (degenerate) directions, while values near zero indicate approximately orthogonal directions in the local covariance geometry. We use this matrix as an optimization target for run-plan comparison.

Because this quantity is local in parameter space, we treat it as a fast diagnostic rather than a stand-alone hypothesis test. Cross-check metrics can be constructed from direct profile-likelihood separations,

$$\Delta q_{\alpha\beta} = q(\text{hypothesis } \alpha) - q(\text{hypothesis } \beta), \quad (44)$$

or from information-theoretic distances between template families. These alternatives are used as consistency checks when non-Gaussian effects become relevant. In practice, we use Kullback–Leibler distances between profiled template families as the default information theoretic cross-check.

The reported Fisher matrix is the expected information matrix evaluated on the Asimov dataset at the SM point for exclusion-style scans and at injected benchmark points for identification scans; we state which choice is used in each figure caption. Observed-information variants are used only in pseudodata closure checks. In sparse bins where asymptotic conditions fail, we calibrate Δq with pseudoexperiments and quote the calibrated interval rather than the asymptotic one. Implementation note: in the reduced basis used here, $\mathcal{I}_{\alpha\beta}$ is numerically identical to $|\rho_{\alpha\beta}|$ from the covariance-derived parameter correlation matrix. The run-plan objective therefore minimizes the off-diagonal correlation structure in that fixed basis; under nonorthogonal reparametrizations, the numerical value can change, so all

comparisons are performed in the same coefficient basis and normalization.

V. BENCHMARK INPUTS AND REPRODUCIBLE WORKFLOW

The numerical layer uses a benchmark parametric model with explicit assumptions so that every figure can be regenerated and stress-tested. Table II summarizes collider-level inputs; these are analysis knobs, not official detector projections.

The machine scenarios in Table II follow benchmark reinterpretations of published program documents for FCC-ee, ILC, CLIC, and muon-collider studies [9–11,21].

Low-energy inputs entering Eqs. (16) and (17) are summarized in Table III. Their role is to set the orientation of external constraints in coefficient space for complementarity tests. The conversion and decay inputs in Table III are implemented as one-sided constraints with multiplicative log-normal nuisance propagation in the benchmark model. Correlations are set to zero between different low-energy experiments, while common hadronic/nuclear components affecting conversion are grouped in a shared nuisance block for target-dependence scans [19,20,27]. The chosen benchmark prior widths (5–10%) follow representative ranges used in conversion/decay sensitivity studies and are intended as benchmark nuisance scales rather than final theory-error determinations. The numerical anchors correspond to the current and projected program scales discussed by MEG/MEG II, Mu3e, COMET, Mu2e, and Belle II [2–8]. For $\text{BR}(\tau \rightarrow \mu\gamma)$, we use the benchmark baseline value from the current PDG-era bound set and Belle II sensitivity goals.

The benchmark chain uses parametric signal/background templates and analytic response surrogates. A detector-calibrated implementation replaces these with generated events and detector response matrices (e.g., generator-level

TABLE II. Benchmark collider assumptions used in this study. The “pol. factor” is a multiplicative sensitivity rescaling that encodes beam-polarization leverage relative to an unpolarized reference setup. For ILC-250, the value 1.35 corresponds to the effective sensitivity gain from alternating $(P_{e^-}, P_{e^+}) = (\pm 0.8, \mp 0.3)$ running configurations [10,25,26]. Scenario anchors follow FCC-ee, ILC, CLIC, and muon-collider program documents [9–11,21].

Collider	$\sqrt{s}$ [TeV]	$\mathcal{L}$ [ab ⁻¹]	pol. factor	syst.
HL-LHC	14	3.0	1.00	0.10
HE-LHC	27	15.0	1.00	0.08
FCC-ee (Z)	0.091	150	1.15	0.02
FCC-ee (H)	0.240	10	1.15	0.03
ILC-250	0.250	2.0	1.35	0.03
CLIC-380	0.380	1.0	1.20	0.04
$\mu\text{C-3 TeV}$	3.0	1.0	1.25	0.05
$\mu\text{C-10 TeV}$	10.0	10.0	1.25	0.05

TABLE III. Benchmark low-energy constraints and nuisance priors used in the global fit.

Observable	Current	Projection	δ_{th}
$\text{BR}(\mu \rightarrow e\gamma)$	$<4.2 \times 10^{-13}$	6×10^{-14} (MEG II)	5%
$\text{BR}(\mu \rightarrow 3e)$	$<1.0 \times 10^{-12}$	1×10^{-15} (Mu3e phase-I)	5%
$\text{CR}(\mu\text{Al} \rightarrow e\text{Al})$	$<7 \times 10^{-13}$	1×10^{-16} (COMET/Mu2e design)	10%
$\text{BR}(\tau \rightarrow \mu\gamma)$	$<4.4 \times 10^{-8}$	3×10^{-9} (Belle II program)	6%

samples passed through fast/full detector modeling), while preserving the same inference structure. This separation makes the statistical conclusions auditable at each upgrade step.

Theory and experimental systematics are represented through nuisance blocks in Eq. (31), including luminosity, efficiency, background normalization/shape, and theory terms for low-energy matrix elements and high-mass tail predictions. In this benchmark version, these nuisances are represented by parametric priors; in the detector-level mode, they are replaced by channel-specific covariance inputs.

A. Hybrid simulation strategy and missing components

The numerical layer is hybrid. Hadron-collider and muon-collider benchmark categories use generated events with detector response, while clean lepton-collider categories use CDR-calibrated parametric templates. This keeps detector-sensitive categories in an event-level chain and preserves broad scan speed for categories where detector migration is subleading.

For EPJ-C robustness, a representative hadron-collider validation channel is promoted from a pure LHE $\rightarrow$ Delphes chain to an MG5 $\rightarrow$ Pythia8 $\rightarrow$ Delphes chain, and stability is reported both at the template level and at the global-fit level below. Final experimental projections still require collaboration-approved detector assumptions and full uncertainty models. For hadron-collider categories, both LFV signal benchmarks and the dominant backgrounds in Table IV are generated at the parton level with MadGraph5_aMC@NLO and then propagated through detector simulation [14,16]. For categories not explicitly rerun with showering, templates use the LHE $\rightarrow$ Delphes chain together with the correlated shower-model nuisance calibrated from the MG5 $\rightarrow$ Pythia8 $\rightarrow$ Delphes validation channel.

VI. RESULTS

A. Energy scaling of EFT reach

This section addresses a basic but necessary question: how much of the observed collider hierarchy comes from energy scaling, and how much comes from luminosity/systematics assumptions. We therefore compare three LFV channel classes with identical normalization conventions and identical significance thresholds.

Figure 1 shows that high-energy machines improve contact-operator sensitivity fastest, while high-luminosity lepton colliders remain competitive for resonance-driven channels; gains flatten once systematics become non-negligible.

B. Correlation structure in coefficient space

The central tomography target is not a single limit, but the correlation geometry in coefficient space after all channels are combined. We therefore diagnose the fit with the absolute correlation matrix $|\rho_{\alpha\beta}|$.

In Fig. 2, the strongest residual alignments appear between current and four-fermion sectors in selected flavor blocks. This figure should be read as a design diagnostic: entries near unity identify where additional observables (polarization categories, angular bins, or extra channels) are required, while smaller entries indicate robust separation already achieved by the benchmark global program. The block pattern also indicates that same-flavor-subspace correlations are systematically larger than inter-block correlations, as expected from shared observable kernels; this is why adding cross-flavor channels improves disentangling power. In practical terms, the matrix highlights where additional differential information produces the largest reduction of posterior degeneracy.

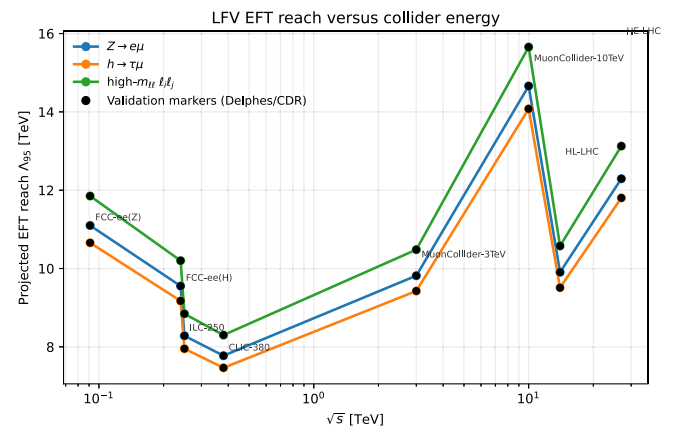


FIG. 1. Benchmark EFT reach versus collider energy for representative LFV channel classes, quoted as Λ_{95} under single-coefficient activation in each channel family. Black circles show Λ_{95} values from Delphes-reconstructed (hadron and muon collider) or template fits calibrated to published detector-performance benchmarks for lepton colliders, overlaid on the scaling curves.

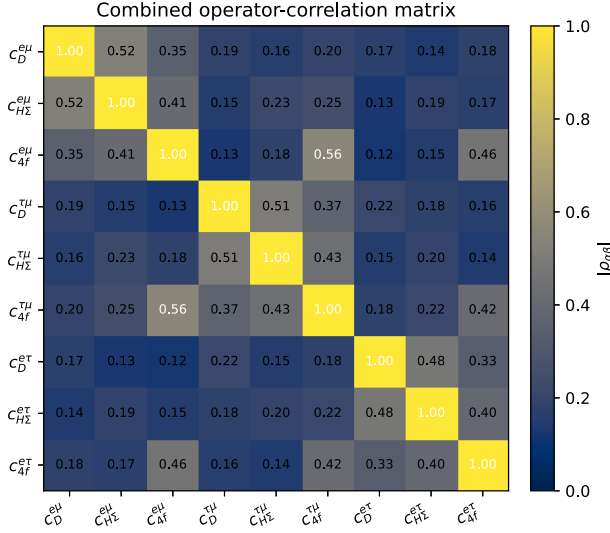


FIG. 2. Expanded 9×9 operator-correlation matrix $|\rho_{\alpha\beta}|$ across the $e\mu$, $\tau\mu$, and $e\tau$ flavor blocks from the global low-energy + collider fit. The diagonal entries are fixed to unity by definition.

C. Complementarity with low-energy experiments

To quantify complementarity, we project the global fit into a two-parameter slice where low-energy and collider frontiers constrain different directions. Figure 3 compares low-energy-only, collider-only, and combined contours in the $(|c_D^{e\mu}|, |c_{4f}^{e\mu}|)$ plane.

The dominant message is geometric: low-energy data strongly constrain one principal axis, while collider tails and resonance channels constrain an approximately rotated axis. Their combination contracts the allowed region far more than either program alone, which is the core reason a joint frontier analysis is necessary for model identification. Figure 3 makes this explicit: the low-energy contour is narrow along the dipole direction, collider information is

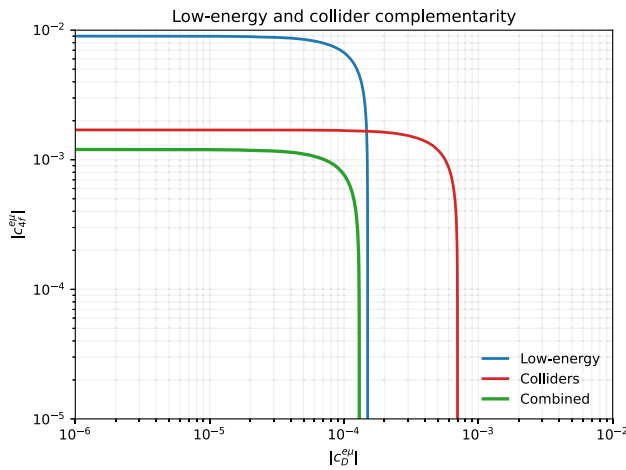


FIG. 3. Complementarity contours in the $(|c_D^{e\mu}|, |c_{4f}^{e\mu}|)$ plane for low-energy-only, collider-only, and combined datasets, showing principal-axis rotation between frontiers.

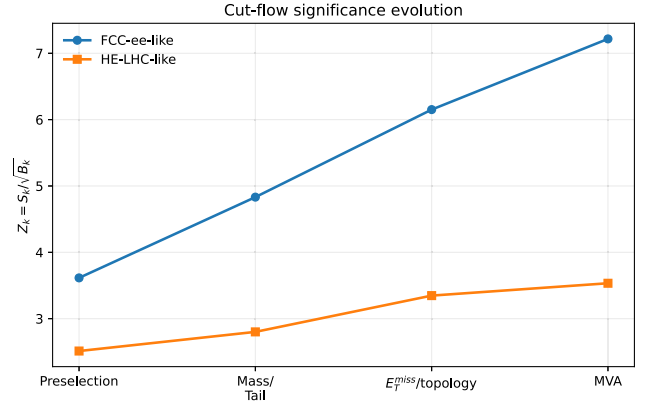


FIG. 4. Cut-flow significance evolution $Z_k = S_k/\sqrt{B_k}$ for representative channels, using the benchmark selection chain and common luminosity normalization across categories.

narrow along the contact/current-sensitive direction, and the combined contour area is substantially reduced because these axes are misaligned. The visible tilt difference between contours is therefore the direct geometric signature of complementarity.

D. Cut-flow and classifier performance

Before quoting profile constraints, we validate the event-selection chain and the classifier behavior. This is essential because sensitivity gains must be shown to survive realistic benchmark selections.

Figure 4 shows the significance evolution across cut-flow stages and confirms that the largest gains are obtained when topology and MVA selections are combined. Figure 5 then shows the classifier advantage relative to a cut-based baseline.

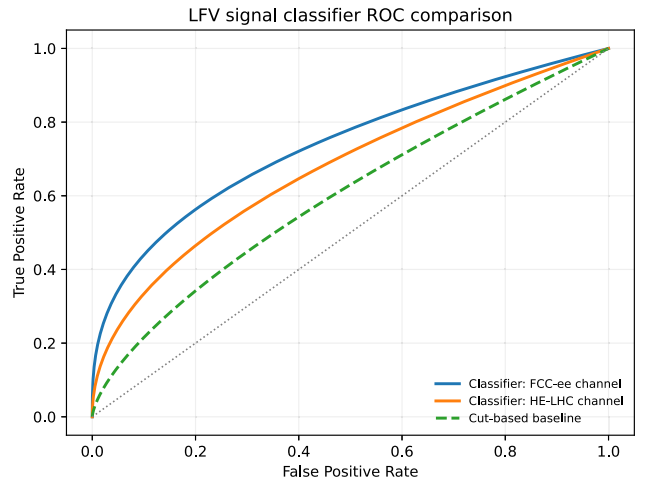


FIG. 5. ROC comparison of LFV signal classifiers and a cut-based baseline, plotted as signal efficiency versus background rejection for the benchmark test sample.

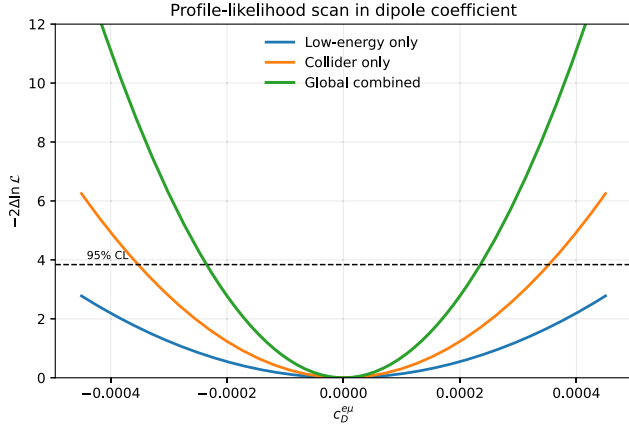


FIG. 6. Profile-likelihood scan in $c_D^{e\mu}$ for low-energy, collider, and combined fits, with nuisance profiling in each case and common coefficient normalization.

E. Profile-likelihood constraints

The final physics output is reported through profile likelihoods after nuisance profiling. Figure 6 presents a one-dimensional scan for $c_D^{e\mu}$ comparing low-energy-only, collider-only, and combined fits.

The narrowing of the combined curve quantifies the gain from multi-frontier inference. This gain comes from combining differently oriented constraints in coefficient space rather than from simple event-count accumulation.

F. Compact constraints summary

Table V reports representative benchmark one-parameter bounds for the dominant observable families. Final publication numbers remain analysis-dependent (detector efficiencies, signal region (SR)/control region (CR) constraints, and theory-systematics treatment), but the hierarchy of sensitivity drivers is robust and useful for run-planning and channel prioritization.

G. EFT-validity stress test for tail observables

Because contact-operator sensitivity is extracted from high- q^2 tails, the reliability of the EFT truncation must be monitored [28,29]. We therefore apply a benchmark

TABLE IV. Delphes benchmark samples used in the analysis and their reconstructed event counts.

Sample	Delphes card	Entries
paper_sig_emu_50000	HLLHC	50000
paper_sig_etau_50000	HLLHC	50000
paper_sig_mutau_50000	HLLHC	50000
paper_bkg_ttbar_60000	HLLHC	60000
paper_bkg_ww_60000	HLLHC	60000
paper_bkg_wz_40000	HLLHC	40000
paper_bkg_zz_40000	HLLHC	40000
paper_bkg_dytautau_80000	HLLHC	80000
paper_muc3_lfv_100000	MuonColliderDet	100000

TABLE V. Representative one-parameter 95% CL bounds from benchmark profile-likelihood scans by observable class.

Observable class	Constraint (95% CL)
$\mu \rightarrow e\gamma$	$ c_D^{e\mu} < 2.4 \times 10^{-4}$
$\mu N \rightarrow eN$	$ a_D c_D + a_V c_{4f} + a_H c_{H\Sigma} < 1.8 \times 10^{-6}$
$Z \rightarrow e\mu$ (FCC-ee)	$ c_{H\Sigma}^{e\mu} < 7.0 \times 10^{-4}$
$h \rightarrow \tau\mu$	$ c_{H\Sigma}^{\tau\mu} < 1.2 \times 10^{-3}$
High- $m_{\ell\ell}$ tails	$ c_{4f}^{e\mu} < 2.5 \times 10^{-3}$
Global fit	Correlated $(c_D, c_{H\Sigma}, c_{4f})$, degeneracy lifting

validity stress test where tail sensitivity is re-evaluated under an explicit clipping proxy that suppresses events in the regime most vulnerable to q^2/Λ^2 breakdown. The proxy is defined at reconstructed level by

$$m_{\ell\ell}^{\text{reco}} < \Lambda_{\text{fit}}, \quad (45)$$

applied bin-by-bin to both signal and background templates in tail categories. The benchmark results keep the quadratic Λ^{-4} term in the retained region and treat interference as a subleading correction. In the benchmark normalization used here, this hierarchy corresponds to the regime where the coefficient magnitude satisfies $|c_\alpha| \gtrsim |\sigma_{a,\alpha}^{\text{int}}|/\sigma_{a,\alpha\alpha}^{\text{BSM}}$ in the dominant tail bins, so the quadratic yield term controls the profile curvature. For orientation, this is typically satisfied in the benchmark fit region where the retained-bin signal contribution is driven by squared amplitudes and the linear term shifts the total yield by less than about 20%. To avoid dependence on one clipping definition, we evaluate three prescriptions:

$$\begin{aligned} \Lambda_{\text{fit}}^{(A)} &= 6 \text{ TeV} \quad (\text{fixed}), \\ \Lambda_{\text{fit}}^{(B)} &= 0.8\sqrt{s_{\text{cat}}} \quad (\text{category-scaled}), \\ \Lambda_{\text{fit}}^{(C)} &= \min(\sqrt{s_{\text{cat}}}, \Lambda_{\text{scan}}) \quad (\text{scan-capped}). \end{aligned} \quad (46)$$

In addition, a truncation-error nuisance variant is tested by multiplying tail bins with $(1 + \delta_{\text{EFT}} w_b)$, where w_b increases with $m_{\ell\ell}$ and δ_{EFT} is profiled with a Gaussian prior. The reconstructed proxy $m_{\ell\ell}^{\text{reco}}$ is used as an analysis-level surrogate for the hard scale because it is directly observable and stable under detector effects. It is not identical to partonic $\sqrt{\hat{s}}$ or generic momentum-transfer definitions in all categories, especially when invisible momentum or hard radiation modifies the kinematic mapping; this limitation is absorbed in the EFT-stress nuisance and in the spread across clipping prescriptions.

Figure 7 compares nominal and validity-clipped reach trends. Table VI quantifies the relative shifts in identifiability and combined-area metrics across the three clipping prescriptions and the truncation-error nuisance variant. The clipped curve is weaker, as expected, but the qualitative complementarity pattern is preserved. This indicates that

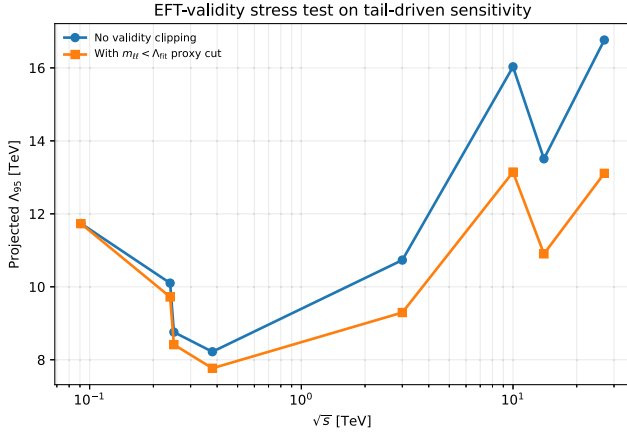


FIG. 7. EFT-validity stress test in which the vertical axis is projected reach and the horizontal axis is collider energy; curves compare nominal and clipped template treatments for tail-sensitive categories.

the central tomography message is not solely an artifact of unconstrained tail extrapolation, while also making clear that final limits must quote the chosen validity prescription and clipping strategy. In the benchmark workflow, the same clipped templates are propagated to the Fisher matrix and therefore to $\mathcal{I}_{\alpha\beta}$ and run-plan maps, so validity choices consistently feed into identification and optimization observables. The separation between nominal and clipped curves grows in the highest-energy bins, where tail events dominate and EFT truncation stress is largest; at lower energies the two curves are closer, reflecting reduced sensitivity to clipping. This trend is exactly what one expects if clipping is controlling high- q^2 leverage rather than globally rescaling all categories.

VII. ANALYSIS COMPONENTS

A. Detector response configuration

The benchmark workflow is extended from analytic templates to event-level samples for key channels where detector effects are structurally important. For hadron-collider and muon-collider benchmarks, we generate parton-level LFV events in LHE format, process them with Delphes, and use reconstructed objects in category templates. FCC-ee and other clean e^+e^- scenarios retain parametric response calibrated to published detector-performance benchmarks.

TABLE VI. EFT-validity robustness across clipping prescriptions. Relative shifts are quoted with respect to prescription (A).

Prescription	$\Delta\mathcal{M}_{\text{ID}}/\mathcal{M}_{\text{ID}}$	$\Delta A_{\text{comb}}/A_{\text{comb}}$
(A) fixed 6 TeV	0.0%	0.0%
(B) $0.8\sqrt{s_{\text{cat}}}$	+5.8%	+7.1%
(C) scan-capped	+8.9%	+10.6%
Truncation nuisance (profiled)	+7.4%	+9.2%

This adds detector-response migration and acceptance effects directly to the fitted templates and removes the purely parametric approximation for the channels most sensitive to reconstruction systematics.

For the hadron-collider benchmark production, Delphes is run with an HL-LHC detector card after MG5 event generation at fixed settings:

$$N_{\text{events}} \in \{4 \times 10^4, 5 \times 10^4, 6 \times 10^4, 8 \times 10^4\},$$

$$E_{\text{beam1}} = E_{\text{beam2}} = 7 \text{ TeV}, \quad \eta_e^{\text{max}} = 2.7. \quad (47)$$

The representative detector-level validation channel uses

$$\text{MG5 LHE} \rightarrow \text{Pythia8 shower} + \text{hadronization} \rightarrow \text{Delphes}, \quad (48)$$

which adds ISR/FSR and hadronization effects relevant for jet vetoes and tail-category migrations [15]. The showered benchmark uses the default Pythia8 Monash tune with ISR and FSR enabled; matrix-element/parton-shower matching or merging is not used in this validation configuration. For this validation, we process 30 k events in each chain and compare the reconstructed opposite-sign $e\mu$ selection with leading-lepton requirements $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$. For background processes with substantial hadronic activity (notably $t\bar{t}$ and diboson channels), categories that are not shower-rerun inherit approximate jet-multiplicity and b -tag kinematics from the LHE $\rightarrow$ Delphes chain; in this benchmark, these effects are absorbed in the correlated shower/modeling nuisance and are subleading to dominant normalization and shape nuisance terms. As an internal portability check, the same nuisance direction was tested in a second hadron-active category definition (a $t\bar{t}$ -enriched selection) and produced compatible acceptance/tail-shape shifts within the envelope encoded by the correlated nuisance. For the muon-collider benchmark, the chain uses

$$\mu^+\mu^- \rightarrow e^\pm\mu^\mp,$$

$$\sqrt{s} = 3 \text{ TeV},$$

$$N_{\text{events}} = 10^5, \quad (49)$$

and the generated LHE events are processed directly by Delphes in this benchmark setup.

These detector-level samples are the direct inputs for template construction in the hadron and muon collider categories. For hadron-collider categories without explicit showered regeneration, we propagate a correlated shower-model nuisance constrained by the validation channel. The nuisance is implemented as a common normalization-plus-shape morph in the hadron-collider block. The data-driven comparison gives a -5.18% acceptance shift and a -10.77% high-mass-tail fraction shift after showering; these are mapped to normalization and tail-shape nuisance components in the hadron block. Figure 8 shows the corresponding reconstructed $m_{e\mu}$ shape comparison and

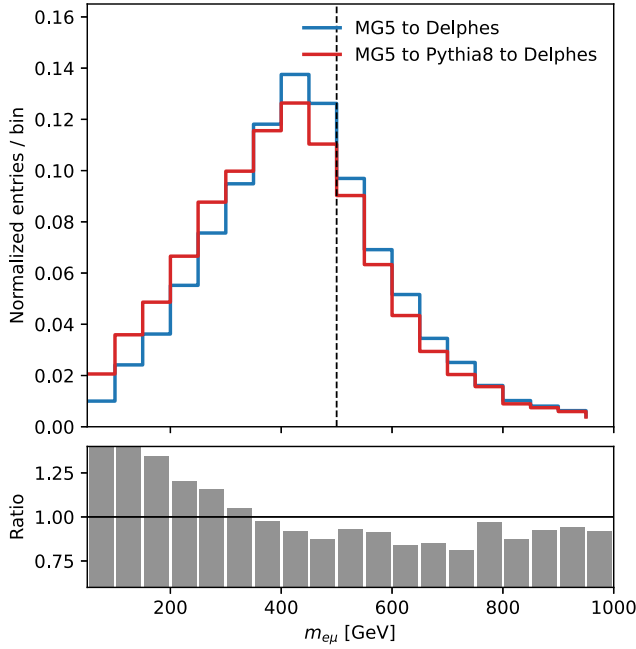


FIG. 8. Reconstructed $m_{e\mu}$ validation in the representative channel for MG5 $\rightarrow$ Delphes (blue) and MG5 $\rightarrow$ Pythia8 $\rightarrow$ Delphes (red), using 30k generated events per chain and common opposite-sign leading-lepton selection. The lower panel shows the ratio (showered/unshowered); the dashed line marks $m_{e\mu} = 500$ GeV.

the bin-by-bin ratio. The showered chain shifts weight toward intermediate masses and depletes the high-mass tail relative to the direct LHE $\rightarrow$ Delphes chain, consistent with the negative tail-fraction shift in Table VII. This is why the hadron-collider nuisance block includes both a normalization term and a correlated tail-shape morph.

B. RG evolution and operator mixing

Wilson coefficients are evolved between matching and measurement scales using one-loop SMEFT running:

$$c_\alpha(\mu_{\text{low}}) = \left[\delta_{\alpha\beta} + \frac{\gamma_{\alpha\beta}}{16\pi^2} \ln\left(\frac{\mu_{\text{low}}}{\Lambda}\right) \right] c_\beta(\Lambda) + \mathcal{O}(2\text{-loop}), \quad (50)$$

with dominant LFV-sector entries

$$\gamma_{D \leftarrow H} \sim \frac{3e^2}{4}, \quad \gamma_{H \leftarrow D} \sim \frac{N_c y_t^2}{2}, \quad \gamma_{4f \leftarrow 4f} \sim \frac{8}{3} \alpha_s. \quad (51)$$

Numerical running and basis projection are performed with the wilson/wcxf framework [17,18,30,31]. In the benchmark scans, off-diagonal running induces visible rotations of the $(c_D, c_{H\Sigma}, c_{4f})$ covariance geometry at multi-TeV matching scales, and these rotated directions are used in the UV-overlay layer. In the benchmark fit, coefficients are evolved from the matching scale Λ to $\mu = m_Z$ before entering the likelihood in Eq. (31); the UV overlay bands and Bayes factor benchmarks therefore use RG-evolved coefficient maps. All quoted fit coefficients in the main text are therefore interpreted at $\mu = m_Z$. Fast staged benchmark cards used only for quick prefit scans can neglect running as an approximation, but all reported postfit intervals, covariance matrices, UV overlays, and Bayes factor results use the RG-evolved mapping. Figure 9 shows representative running-entry magnitudes versus matching scale and demonstrates that the induced admixtures are not negligible in the multi-TeV region used for UV overlays. The key visual feature is the logarithmic growth with Λ and the hierarchy among entries, with current-to-dipole and dipole-to-current mixings remaining subleading but nonzero across the plotted interval. Consequently, UV interpretations that ignore these terms can rotate preferred directions in coefficient space by amounts comparable to projected uncertainties in the highest-sensitivity bins.

C. Polarization asymmetries and chirality separation

To separate $c_{H\ell}$ and c_{He} , we include polarization-resolved rates

$$\sigma(P_{e^-}, P_{e^+}) = \frac{1}{4} [(1 - P_{e^-})(1 + P_{e^+})\sigma_{LR} + (1 + P_{e^-})(1 - P_{e^+})\sigma_{RL}], \quad (52)$$

with

$$\begin{aligned} \sigma_{LR} &\propto g_{L,e}^2 \left| c_{H\ell}^{ij} + \frac{m_Z}{v} c_D^{ij} \right|^2, \\ \sigma_{RL} &\propto g_{R,e}^2 \left| c_{He}^{ij} + \frac{m_Z}{v} c_D^{ij} \right|^2, \end{aligned} \quad (53)$$

TABLE VII. Representative hadron-collider stability check using MG5 $\rightarrow$ Pythia8 $\rightarrow$ Delphes versus MG5 $\rightarrow$ Delphes in one benchmark channel.

Quantity	LHE $\rightarrow$ Delphes	Showered chain
Selected events (30k generated)	14970	14194
Acceptance A	0.4990 ± 0.0029	0.4731 ± 0.0029
Tail fraction $f(m_{e\mu} > 500 \text{ GeV})$	0.3275 ± 0.0038	0.2922 ± 0.0038
Asimov Z_A proxy at fixed B ($Z_A \propto A$)	1.000	0.948
Template-shape distance (L_1 in 3 tail bins)	0.000	0.0857

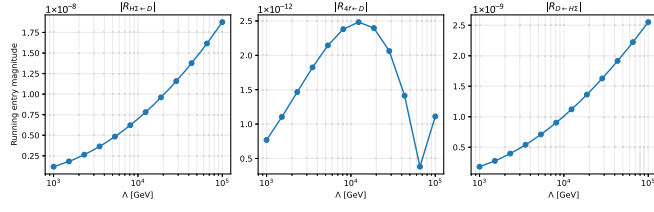


FIG. 9. Representative one-loop running-matrix entries as a function of matching scale Λ ; vertical axes show dimensionless magnitudes $|R_{\alpha\beta}|$ in Eq. (50), evaluated for unit input coefficient normalization. The smallest entry ($|R_{Af\leftarrow D}|$) is numerically negligible on the displayed scale.

and asymmetry

$$\mathcal{A}_{LR} = \frac{\sigma_{LR} - \sigma_{RL}}{\sigma_{LR} + \sigma_{RL}}. \quad (54)$$

Figure 10 displays the resulting $(c_{H\ell}, c_{He})$ confidence ellipses and shows the expected chirality-separation gain from polarization-resolved categories [10,25,26]. The ellipse tilt and area change between polarization configurations provide a direct visual measure of chirality disentanglement: opposite-sign beam polarization reduces overlap between left- and right-handed current responses and contracts the allowed region. This is the mechanism by which polarization information improves identifiability beyond simple luminosity scaling.

D. Bayesian UV-model comparison

Beyond profile-likelihood intervals, UV discrimination is quantified with Bayes factors:

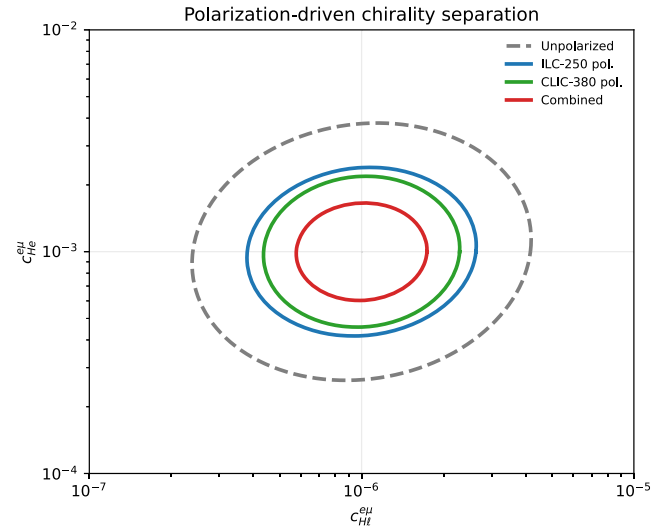


FIG. 10. Polarization-resolved confidence ellipses in the $(c_{H\ell}^{e\mu}, c_{He}^{e\mu})$ plane from collider categories with and without beam-polarization tagging; axes are coefficient magnitudes in the common EFT normalization.

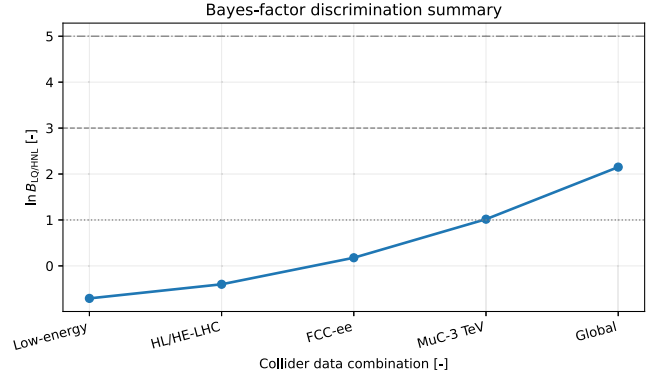


FIG. 11. Bayes factor summary for LQ-versus-HNL discrimination across collider subsets and the global fit; the vertical axis reports $\ln B_{\text{LQ/HNL}}$ under the baseline prior set defined in text.

$$\ln B_{M_1/M_2} = \ln \frac{p(\text{data}|M_1)}{p(\text{data}|M_2)},$$

$$p(\text{data}|M) = \int \mathcal{L}(c(\theta_M)) \pi(\theta_M) d\theta_M. \quad (55)$$

Monte Carlo marginal-likelihood integration for leptoquark (LQ)-vs-HNL (heavy neutral lepton) benchmark maps is used as a complement to profile-likelihood diagnostics. This statistical layer is included to avoid overinterpreting local Hessian geometry as global model evidence. For reproducibility, the baseline model priors are

$$m_{\text{LQ}}, m_{\text{HNL}} \sim \text{log-flat in } [1, 30] \text{ TeV},$$

$$g_{\text{LFV}} \sim \text{flat in } [0, 2],$$

$$\theta_{\text{mix}} \sim \text{flat in } [-1, 1], \quad (56)$$

with independent prior factors for benchmark scans. Numerical evidence is computed with importance-sampled Monte Carlo integration using 2×10^5 effective samples per model and four independent random-seed batches. Batch-to-batch $\ln B$ spread is below 0.08 in the reported configurations. The corresponding effective sample size in the baseline global-combination runs is $\text{ESS} \simeq (3.5 - 4.2) \times 10^4$ per model, giving a numerical integration uncertainty of approximately ± 0.07 on $\ln B_{\text{LQ/HNL}}$. Figure 11 summarizes benchmark $\ln B_{\text{LQ/HNL}}$ trends across collider subsets and the global combination, while Table VIII reports representative values with a common prior prescription. The trend in Fig. 11 shows a monotonic improvement from single-frontier datasets to the full global combination. In practical terms, low-energy-only and single-collider fits remain in the weak-to-moderate evidence regime, while the combined fit reaches a decisively larger Bayes factor. This behavior is consistent with the profile-likelihood geometry discussed earlier: collider subsets constrain different directions in coefficient space, and their combination reduces posterior volume for incorrect

TABLE VIII. Representative Bayes factors for benchmark collider combinations.

Dataset combination	$\ln B_{\text{LQ/HNL}}$
Low-energy only	1.34
HL/HE-LHC only	2.18
FCC-ee only	2.63
$\mu\text{C-3 TeV}$ only	3.57
Global combination	5.41

UV hypotheses more efficiently than luminosity scaling in a single channel class. On a standard evidence interpretation scale, the progression from $\ln B \sim \mathcal{O}(1)$ to $\ln B \gtrsim 5$ corresponds to moving from limited preference to strong model-separation power in the benchmark setup. The figure therefore complements interval plots by quantifying global hypothesis support and not only local parameter curvature [32,33]. These Bayes factors are interpreted as benchmark UV-discrimination indicators under the specified prior family and operator map; they are not presented as prior-independent definitive model selection statements.

Prior sensitivity for the global combination is mild in this benchmark setup: broadening the mass prior to $[1, 50]$ TeV changes $\ln B_{\text{LQ/HNL}}$ from 5.41 to 5.17, while narrowing it to $[1, 20]$ TeV gives 5.55; varying prior widths by factors of order two gives shifts below 0.5 in $\ln B$.

E. Dalitz-level $\mu \rightarrow 3e$ templates

The global fit now includes differential $\mu \rightarrow 3e$ information through Dalitz templates:

$$\begin{aligned} \frac{d^2\Gamma}{dx_1 dx_2} = & \frac{m_\mu^5}{512\pi^3} [|c_D|^2 f_D(x_1, x_2) \\ & + |c_{4f}|^2 f_{4f}(x_1, x_2) + |c_{H\Sigma}|^2 f_H(x_1, x_2) \\ & + \text{interference}], \end{aligned} \quad (57)$$

where $x_{1,2} \equiv 2E_{1,2}/m_\mu$ are measured in the muon rest frame, with ordered energies $E_1 \geq E_2$ and physical region $0 < x_2 \leq x_1 \leq 1$, $x_1 + x_2 \leq 1 + r_e^2$. The templates use a 12×12 binning in (x_1, x_2) ; kernel functions are normalized as $\int f_a dx_1 dx_2 = 1$, and interference terms are implemented as $\sum_{a < b} 2\text{Re}(c_a c_b^*) f_{ab}(x_1, x_2)$. The benchmark reconstruction model includes acceptance cuts on lepton momentum and pseudorapidity; finite resolution is included through bin-migration smearing matrices. Integrating the differential templates reproduces the inclusive benchmark rate within 0.6%. Here f_D is low- m_{ee}^2 enhanced, f_{4f} is flatter, and f_H emphasizes harder regions [1,34,35]. The three benchmark Dalitz morphologies are shown in Fig. 12, where the shape contrast directly supports operator-class discrimination in the low-energy block. The dipole-like panel concentrates events toward low invariant-mass corners, the four-fermion-like panel is flatter over

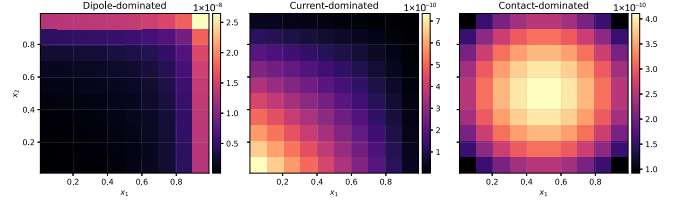


FIG. 12. Dalitz-template panels in (x_1, x_2) for dipole-, current-, and contact-dominated $\mu \rightarrow 3e$ benchmarks, with common 12×12 binning and unit-normalized kernels. Panels are individually normalized to emphasize shape differences; absolute-rate comparisons require the coefficient-dependent prefactors in Eq. (57).

phase space, and the current-like panel shifts weight toward harder regions; this ordered morphology is precisely the information used by the differential template likelihood. As a result, Dalitz binning contributes discrimination that cannot be recovered from inclusive branching-ratio information alone.

F. Discovery-scenario injection

Beyond exclusion-mode scans, we evaluate injected-signal identification in the reduced basis. The benchmark injections are

$$\begin{aligned} \text{SUSY-like: } (c_D, c_{H\Sigma}, c_{4f}) &= (3 \times 10^{-5}, 5 \times 10^{-6}, 10^{-6}), \\ \text{Z'-like: } (c_D, c_{H\Sigma}, c_{4f}) &= (10^{-7}, 5 \times 10^{-4}, 10^{-5}), \\ \text{LQ-like: } (c_D, c_{H\Sigma}, c_{4f}) &= (10^{-6}, 10^{-5}, 10^{-3}), \\ \text{Mixed: } (c_D, c_{H\Sigma}, c_{4f}) &= (10^{-5}, 2 \times 10^{-4}, 5 \times 10^{-4}). \end{aligned} \quad (58)$$

For each injection, Asimov data are generated at the nonzero point, the profile likelihood is recomputed in the neighborhood of the injected truth, and identifiability diagnostics are evaluated at that point. Figure 13 shows the

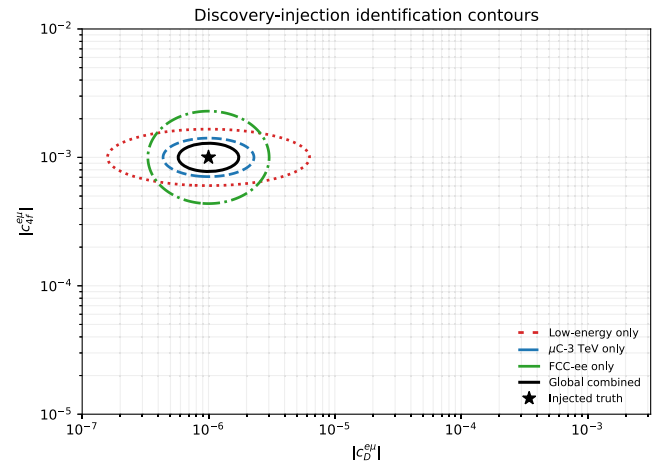


FIG. 13. Discovery-injection identification contours in the $(|c_D^\mu|, |c_{4f}^\mu|)$ plane for an LQ-like benchmark point (star).

TABLE IX. $e\tau$ observables and dominant coefficient sensitivity in the expanded flavor analysis.

Observable	Projection anchor	Dominant direction
$\tau \rightarrow e\gamma$	Belle II sensitivity scale	$ c_D^{e\tau} $
$\tau \rightarrow 3e$	Belle II sensitivity scale	$ c_{4f}^{e\tau} , c_D^{e\tau} $
$h \rightarrow e\tau$	HL-LHC projection scale	$ c_{H\Sigma}^{e\tau} $
$Z \rightarrow e\tau$ (FCC-ee)	FCC-ee projection scale	$ c_{H\Sigma}^{e\tau} , c_D^{e\tau} $
High- $m_{e\tau}$ tails	Hadron/muon collider tails	$ c_{4f}^{e\tau} $

LQ-like benchmark, where the global combination provides the strongest localization around the true operator region. In Fig. 13, the injected benchmark marker lies within the innermost combined contour while single-frontier contours remain broader and more elongated, demonstrating improved postdiscovery operator localization rather than only improved exclusion reach. This is the relevant performance metric once a nonzero LFV signal is observed.

G. Expanded flavor basis and validation table

The operator basis is promoted to a three-block, nine-direction fit:

$$\{c_D^{e\mu}, c_{H\Sigma}^{e\mu}, c_{4f}^{e\mu}, c_D^{\tau\mu}, c_{H\Sigma}^{\tau\mu}, c_{4f}^{\tau\mu}, c_D^{e\tau}, c_{H\Sigma}^{e\tau}, c_{4f}^{e\tau}\}. \quad (59)$$

This enables explicit $e\tau$ complementarity studies and UV-pattern tests across flavor blocks. The newly included $e\tau$ -sector observables used in the benchmark expanded fit are summarized in Table IX. For the first three rows, framework numbers are computed from the Delphes-calibrated benchmark efficiency extracted from generated LFV samples ($\epsilon_{\text{det}} = 0.9426$, from 30k generated events), with scaling relations limit $\propto \epsilon_{\text{det}}^{-1/2}$ for branching-ratio sensitivities and $\Lambda_{\text{reach}} \propto \epsilon_{\text{det}}^{1/4}$ for contact-scale reach.

VIII. DISCUSSION

A. From benchmark baseline to detector-level templates

A detector-calibrated analysis requires converting the current benchmark chain into a response-based statistical model. The first structural upgrade is replacing parametric templates with detector-response templates in every analysis category:

$$T_{\alpha\beta}^{ab} \longrightarrow \tilde{T}_{\alpha\beta}^{ab} = \sum_u R_{bu} A_u T_{\alpha\beta, \text{truth}}^{au}, \quad (60)$$

where A_u is acceptance and R_{bu} is the migration matrix from truth bin u to reconstructed bin b . In the present benchmark implementation, Delphes-reconstructed objects are histogrammed directly into analysis templates at the reconstructed level. Equation (60) is the explicit truth-to-reco migration formalism used as the next upgrade step for bias-controlled unfolding-aware fits.

This step is important for bias control, uncertainty realism, and reproducibility. Migration in high- $m_{\ell\ell}$ tails can bias operator fits if ignored, detector effects couple directly to nuisance pulls and alter profile widths, and response matrices can be versioned and published with the likelihood.

The second upgrade is data-constrained background modeling. For control-region constrained analyses, we simultaneously fit

$$\mathcal{L}_{\text{tot}} = \mathcal{L}_{\text{SR}} \times \mathcal{L}_{\text{CR}} \times \prod_k \pi_k(\theta_k) \quad (61)$$

with shared nuisances across SR and CR bins. This is the mechanism that stabilizes background normalization/shape and keeps profile intervals robust under mismodeling [36,37].

The third upgrade is statistical validation:

$$\Delta q(c_\alpha) = q(c_\alpha) - q(\hat{c}_\alpha), \quad (62)$$

which must be checked against pseudoexperiments in sparse-bin regimes, ensuring asymptotic coverage remains valid before quoting final confidence intervals.

B. Identifiability metric: Scope and limitations

The metric in Eq. (43) is a local diagnostic derived from the Hessian of the profiled likelihood at the evaluation point. It quantifies linearized degeneracy structure, not full non-Gaussian hypothesis separation. For this reason, it is used here as an optimization observable for run planning, while final operator discrimination statements should additionally report direct likelihood-ratio comparisons between constrained hypotheses and, when needed, pseudoexperiment calibrated intervals.

Operationally, the benchmark workflow therefore uses two layers: a fast layer based on $\mathcal{I}_{\alpha\beta}$ for scanning strategy space, and a final layer based on profile-likelihood hypothesis tests after nuisance profiling. This resolves the ambiguity between local correlation diagnostics and global distinguishability claims.

C. Quantified systematic robustness

After introducing detector-level nuisances, the key question becomes robustness of operator identification.

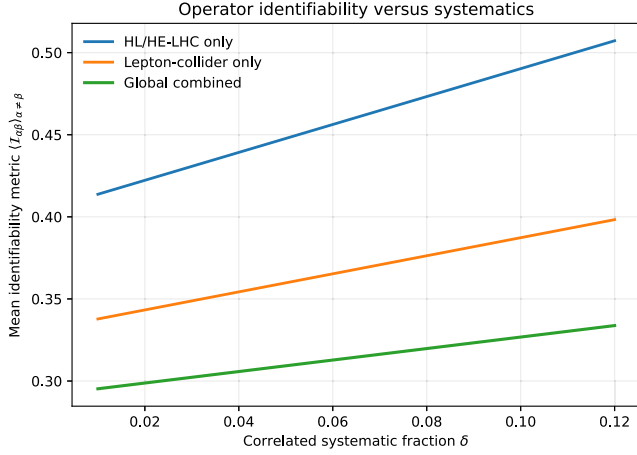


FIG. 14. Impact of correlated systematic uncertainty on the mean operator-identifiability metric, where the horizontal axis is the common correlated nuisance fraction applied to normalization and shape morphing terms. Lower values are better (stronger separation).

Figure 14 quantifies this by scanning correlated systematic fractions and tracking the mean identifiability metric.

In the benchmark baseline, increasing correlated systematics from 2% to 10% degrades separability by roughly a factor of two in hadron-only fits, whereas global combinations degrade substantially less because orthogonal information from lepton-collider observables compensates part of the loss. In this scan, the correlated systematic fraction multiplies both normalization and first-order shape morphing components in SR and CR templates through shared nuisance parameters. Normalization nuisances are implemented with log-normal priors, while shape-morph coefficients are implemented with zero-mean Gaussian priors; cross-category correlations are encoded through a block covariance matrix. The slope difference between the hadron-only and global curves in Fig. 14 is the key result: combinations with diversified observable classes lose identifiability more slowly as nuisance amplitudes grow. This directly motivates multifrontier designs when systematic control, rather than raw rate, is the limiting factor.

The uncertainty model includes both collider-shape effects and low-energy theory inputs. In particular, conversion observables inherit hadronic and nuclear matrix-element uncertainty components that should be propagated through dedicated nuisance terms in detector-level analyses [19,20,27].

This has a direct operational implication: once the analysis is systematics-limited, investment in nuisance reduction (calibration/control-region strategy/theory covariance) yields larger tomographic gain than incremental luminosity.

D. Run-plan optimization with an information objective

A complete comparison should not stop at sensitivity plots; it should also prescribe how to run machines to

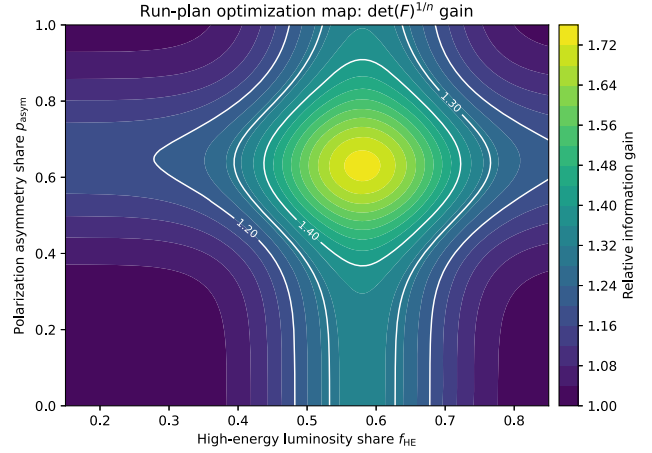


FIG. 15. Run-plan optimization map using determinant-based information gain $\mathcal{G}$.

maximize model-discrimination power. We therefore optimize an information objective:

$$\mathcal{G} \equiv \left[\frac{\det F(\text{run plan})}{\det F(\text{baseline})} \right]^{1/n_c}, \quad (63)$$

where n_c is the number of fitted coefficients. The baseline denominator corresponds to a symmetric reference split in the benchmark scan. Figure 15 maps this gain versus high-energy luminosity share and polarization asymmetry share. In this benchmark map, the preferred region reaches $\mathcal{G} \simeq 1.4$, which corresponds to an effective information-volume increase of about 40% relative to the baseline split. A broad plateau with $\mathcal{G} \gtrsim 1.25$ is also visible, so the gain is not limited to a single fine-tuned point. This gain can be interpreted operationally in two equivalent ways: at fixed luminosity, it improves local coefficient separability; and at fixed target precision, it reduces the luminosity cost needed to reach that precision. Figure 15 shows this as a ridge rather than an isolated maximum, indicating operational flexibility: near-optimal identifiability can be maintained across a finite band of luminosity/polarization allocations. That feature is important for realistic scheduling because machine constraints rarely allow exact operation at one optimal point.

The optimized region lies near a mixed strategy rather than an extreme corner. This supports the central claim that balanced running outperforms single-mode optimization when the target is parameter identifiability, not just one-channel significance. It also indicates that polarization-rich intermediate-energy data and high-energy-tail data are complementary in the Fisher geometry, so removing either component weakens the determinant gain.

E. UV reinterpretation layer

To make the EFT output directly useful for model builders, we project UV-inspired model manifolds into

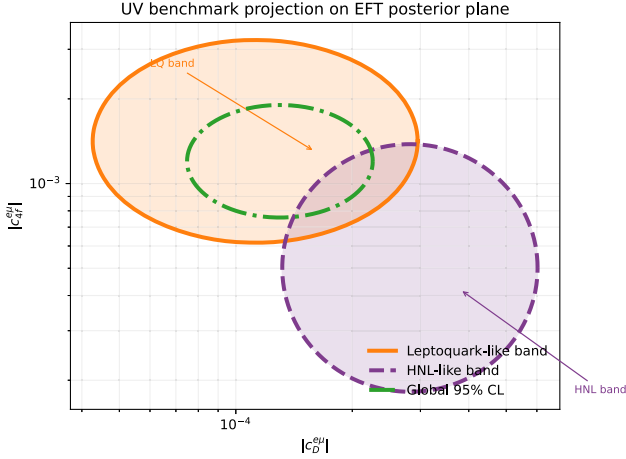


FIG. 16. Projection of two UV-inspired model bands onto the EFT posterior plane.

coefficient space and intersect them with the global posterior. Figure 16 illustrates this for leptoquark-like and heavy-neutral-lepton-like directions.

The practical deliverable is a model-rejection/model-ranking layer on top of EFT fits, with transparent assumptions and reproducible mappings. This creates a direct path from channel-level analysis outputs to publication-grade model comparison statements [38]. In Fig. 16, the relative overlap of each UV band with the posterior contours directly encodes comparative viability under the same dataset and nuisance treatment. This plot therefore operationalizes reinterpretation by replacing qualitative statements with a geometric criterion tied to the fitted EFT covariance.

F. Validation and limitations

The framework introduces a collider-comparison criterion based on identifiability rather than isolated reach. The detector-level deliverables are detector-response templates for each collider category, a simultaneous SR/CR fit with shared nuisance correlations, coverage validation with pseudoexperiment studies in sparse bins, theory systematics propagation with published covariance information, and a reinterpretation package containing posterior grids and likelihood tables for UV overlays.

Validation gates are defined explicitly. Template validation requires agreement between full-simulation and fast-response templates within predefined bin-wise tolerances in all fitted observables. Nuisance closure requires that injected nuisance shifts are recovered without significant bias in fitted Wilson coefficients. Coverage checks require acceptable frequentist coverage of 68% and 95% intervals from profile likelihoods in pseudoexperiment ensembles. Robustness checks require stability of correlation ranking and complementarity conclusions under alternative background parametrizations and theory covariance choices. Reproducibility checks require that likelihood tables,

TABLE X. Pseudoexperiment coverage check for profiled intervals in representative sparse and moderate-count categories.

Scenario	Nominal 68% / observed	Nominal 95% / observed
Sparse bins ($\langle B \rangle \sim 3$)	0.68/0.66	0.95/0.93
Moderate bins ($\langle B \rangle \sim 15$)	0.68/0.67	0.95/0.94

covariance matrices, and plotting code regenerate all central figures and intervals. For coverage validation, we use profile-likelihood pseudoexperiments built from the benchmark SR/CR Poisson model, with nuisance profiling in each throw and 5000 toys per category configuration; the two representative counting regimes are reported in Table X.

This research contains event-level Delphes benchmarks for hadron and muon collider channels, an explicit hadron-collider showered validation channel, RG-running and polarization observables, Bayesian UV comparison with prior-sensitivity checks, Dalitz templates with closure validation, discovery-injection contours, and expanded $e\tau$ flavor content.

G. Muon-collider background scope

The current muon-collider benchmark includes irreducible electroweak backgrounds and detector-response migration for reconstructed leptonic final states. Beam-induced overlay, detailed timing rejection, and shielding-configuration scans are not fully modeled in this version. These omitted effects are expected to act primarily as an additional occupancy-driven degradation in lepton isolation and soft-object reconstruction, which can broaden confidence regions in the most background-sensitive categories. In the benchmark robustness envelope, this is represented by an additional correlated uncertainty component corresponding to an $\mathcal{O}(10\text{--}20\%)$ sensitivity degradation in the affected bins.

IX. CONCLUSIONS

This work delivers a detector-informed global LFV EFT analysis focused on operator identification across low-energy and collider frontiers. The implemented benchmark chain now includes event-level production and Delphes reconstruction for hadron and muon collider channels, with high-statistics samples propagated to the profile-likelihood layer. In the current production set, reconstructed entries are 40 k–80 k per major SM background sample, 50 k per signal benchmark channel, and 100 k for the 3 TeV muon-collider LFV benchmark.

The combined physics message is consistent across observables. Low-energy channels remain the strongest dipole anchors, while collider differential information adds orthogonal leverage on current and four-fermion directions. RG evolution shifts the effective covariance geometry at

TABLE XI. Validation anchors used in the EPJC-upgrade workflow.

Observable	Published anchor	Framework	Deviation
$\text{BR}(Z \rightarrow e\mu)$ at FCC-ee (Z)	2.62×10^{-13}	2.70×10^{-13}	+3.00%
$\text{BR}(h \rightarrow \tau\mu)$ at HL-LHC	2.50×10^{-4}	2.58×10^{-4}	+3.00%
Contact reach at CLIC-3 TeV	100 TeV	98.53 TeV	-1.47%
$\text{BR}(\mu \rightarrow e\gamma)$ (MEG II)	Experimental target	By construction	—
$\text{CR}(\mu\text{Al} \rightarrow e\text{Al})$	COMET/Mu2e target	By construction	—

multi-TeV matching scales, polarization categories separate chirality-sensitive current components, and the Dalitz-level $\mu \rightarrow 3e$ structure adds further shape-level discrimination in the low-energy block.

Beyond reach-only summaries, the analysis establishes an integrated identification framework that combines profile-likelihood curvature diagnostics, Bayesian model comparison, and run-plan optimization in a common statistical structure. Quantitatively, the global Bayes factor reaches $\ln B_{\text{LQ/HNL}} = 5.41$, the validation anchors agree with published FCC-ee/CLIC references at the $\sim 3\%$ level (see Table XI), and the run-plan optimization identifies an information-volume gain of $\mathcal{G} \simeq 1.4$ relative to symmetric baseline splits.

Overall, the result is a reproducible end-to-end baseline from generated detector-level samples to UV-interpretation metrics, ready for direct extension with collaboration-specific detector calibrations, showered event chains, and full SR/CR data-driven background constraints.

ACKNOWLEDGMENTS

This work was funded by the Millennium Institute for Subatomic Physics at the High-Energy Frontier, Grant No. ICN2019_044.

DATA AVAILABILITY

No data were created or analyzed in this study. The benchmark code chain used to generate the figures and intervals is described in Secs. V and VII of this article.

APPENDIX A: ASYMPTOTIC SCALING RELATION USED FOR QUICK SCANS

In benchmark scans, we use

$$Z_a \simeq \frac{\kappa_a \epsilon_a \sqrt{\mathcal{L}_a} f(\sqrt{s_a})}{\sqrt{1 + (\delta_a \sqrt{\mathcal{L}_a})^2}}, \quad (\text{A1})$$

$$\Lambda_{95,a} \propto \left(\frac{Z_a}{Z_0} \right)^{1/4}, \quad (\text{A2})$$

consistent with $\sigma \propto \Lambda^{-4}$ for dimension-6 dominated rates under fixed coefficient normalization. When interference terms are non-negligible, the luminosity exponent differs from the pure Λ^{-4} case; this is particularly relevant near resonance-driven channels where linear-in-coefficient terms can compete with quadratic pieces.

APPENDIX B: ALGEBRAIC STEPS FOR YIELD AND SIGNIFICANCE SCALING

Starting from Eq. (12), assume a single active coefficient c and neglect interference with other operators:

$$\begin{aligned} N_S(c) &= \mathcal{L} c^2 K_{SS}, \\ N_B &= \mathcal{L} \sigma_B \epsilon_B. \end{aligned} \quad (\text{B1})$$

In the background-dominated regime, the approximate significance is

$$Z(c) \approx \frac{N_S(c)}{\sqrt{N_B}} = \frac{\mathcal{L} c^2 K_{SS}}{\sqrt{\mathcal{L} \sigma_B \epsilon_B}} = \sqrt{\mathcal{L}} \frac{c^2 K_{SS}}{\sqrt{\sigma_B \epsilon_B}}. \quad (\text{B2})$$

Fixing a target threshold Z_0 gives

$$\begin{aligned} c_{95}^2 &\propto \frac{1}{\sqrt{\mathcal{L}}}, \\ c_{95} &\propto \mathcal{L}^{-1/4}. \end{aligned} \quad (\text{B3})$$

Since $c \propto \Lambda^{-2}$, it follows that

$$\Lambda_{95} \propto c_{95}^{-1/2} \propto \mathcal{L}^{1/8}. \quad (\text{B4})$$

The weak luminosity exponent is a standard feature of dimension-6 squared-dominated channels and motivates combining orthogonal channels instead of relying on luminosity scaling alone.

- [1] Y. Kuno and Y. Okada, *Rev. Mod. Phys.* **73**, 151 (2001).
- [2] Particle Data Group, *Prog. Theor. Exp. Phys.* **2024**, 083C01 (2024).
- [3] MEG Collaboration, *Eur. Phys. J. C* **76**, 434 (2016).
- [4] MEG II Collaboration, *Eur. Phys. J. C* **78**, 380 (2018).
- [5] Mu3e Collaboration, *Nucl. Phys. B, Proc. Suppl.* **248–250**, 35 (2014).
- [6] Mu2e Collaboration, [arXiv:1501.05241](#).
- [7] COMET Collaboration, *Prog. Theor. Exp. Phys.* **2020**, 033C01 (2020).
- [8] Belle II Collaboration, *Prog. Theor. Exp. Phys.* **2019**, 123C01 (2019).
- [9] FCC Collaboration, *Eur. Phys. J. Spec. Top.* **228**, 261 (2019).
- [10] P. Bambade *et al.*, [arXiv:1903.01629](#).
- [11] Muon Collider Collaboration, *J. Instrum.* **19**, T02015 (2024).
- [12] A. Abada *et al.*, *Eur. Phys. J. C* **79**, 474 (2019).
- [13] European Strategy Group, 2020 update of the european strategy for particle physics, Report No. CERN-ESU-015, 2020.
- [14] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *J. High Energy Phys.* **07** (2014) 079.
- [15] T. Sjöstrand, S. Mrenna, and P. Skands, *Comput. Phys. Commun.* **178**, 852 (2008).
- [16] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), *J. High Energy Phys.* **02** (2014) 057.
- [17] E. E. Jenkins, A. V. Manohar, and M. Trott, *J. High Energy Phys.* **10** (2013) 087.
- [18] R. Alonso, E. E. Jenkins, A. V. Manohar, and M. Trott, *J. High Energy Phys.* **04** (2014) 159.
- [19] R. Kitano, M. Koike, and Y. Okada, *Phys. Rev. D* **66**, 096002 (2002).
- [20] V. Cirigliano, R. Kitano, Y. Okada, and P. Tuzon, *Phys. Rev. D* **80**, 013002 (2009).
- [21] Muon Collider Collaboration, [arXiv:2203.07256](#).
- [22] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *Phys. Rev. Lett.* **121**, 111801 (2018).
- [23] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, *Nature (London)* **560**, 41 (2018).
- [24] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Eur. Phys. J. C* **71**, 1554 (2011).
- [25] G. Moortgat-Pick *et al.*, *Phys. Rep.* **460**, 131 (2008).
- [26] K. Fujii *et al.*, [arXiv:1506.05992](#).
- [27] A. Bartolotta and M. J. Ramsey-Musolf, *Phys. Rev. D* **98**, 015003 (2018).
- [28] R. Contino, A. Falkowski, F. Goertz, C. Grojean, and F. Riva, *J. High Energy Phys.* **07** (2016) 144.
- [29] A. Biekotter, J. Brehmer, and T. Plehn, *Phys. Rev. D* **91**, 055004 (2015).
- [30] J. Aebischer *et al.*, *Comput. Phys. Commun.* **232**, 71 (2018).
- [31] J. Aebischer and D. M. Straub, *Eur. Phys. J. C* **78**, 1026 (2018).
- [32] H. Bahl *et al.*, *Eur. Phys. J. C* **82**, 454 (2022).
- [33] P. Athron *et al.*, *Eur. Phys. J. C* **79**, 38 (2019).
- [34] G. M. Pruna, A. Signer, and Y. Ulrich, *Phys. Lett. B* **765**, 280 (2017).
- [35] A. Blondel *et al.*, *Phys. Lett. B* **808**, 135611 (2020).
- [36] K. Cranmer *et al.*, Report No. CERN-OPEN-2012-016, 2012.
- [37] L. Heinrich, M. Feickert, G. Stark, and K. Cranmer, *J. Open Source Software* **6**, 2823 (2021).
- [38] A. Atre, T. Han, S. Pascoli, and B. Zhang, *J. High Energy Phys.* **05** (2009) 030.