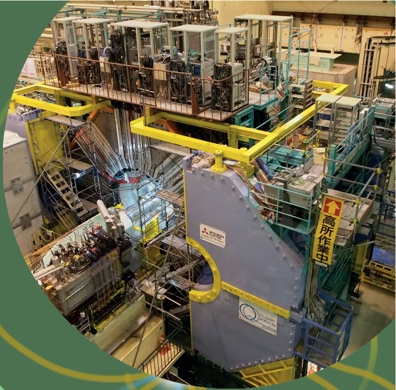
OVERVIEW OF τ LEPTON FLAVOUR VIOLATING DECAYS AT BELLE II

GDR-InF Annual Workshop 2022



Friday 4th November, 2022

Robin Leboucher

Léonard Polat

Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France

CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE CPPM

cnrs

Aix*Marseille université Socialement engagée

erc European Research Council Established by the European Commission

Belle II

Outline

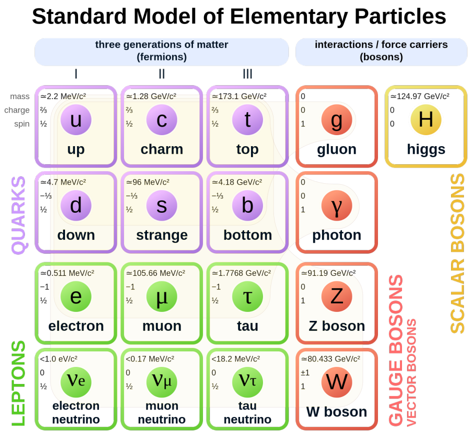
- Lepton flavour violation (LFV) in τ decays
- Overview of the Belle II experiment
- Search for $\tau \to \ell\phi$ (L. Polat)
- Search for $\tau \to \mu\mu\mu$ (R. Leboucher)
- Other τ LFV channels at Belle II

The τ lepton

CPPM

τ ID Card

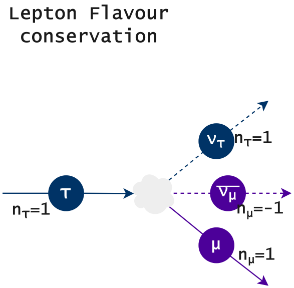
- **Discovery:** 1974-1977 at SLAC-LBL, California
- **Mass: 1.777** $\mathrm{GeV/c^2}$ (heaviest lepton)
- **Lifetime:** $\sim$ **290.3** $\times$ **10**$^{-15}$ s ($10^7 \times$ shorter than muons)
- **Main τ^- decays (95%):**
  - $e^-\overline{\nu}_e\nu_\tau$ or $\mu^-\overline{\nu}_\mu\nu_\tau$
  - $\pi^-\nu_\tau$
  - $\pi^-\pi^0(\pi^0)\nu_\tau$
  - $\pi^-\pi^+\pi^-(\pi^0)\nu_\tau$

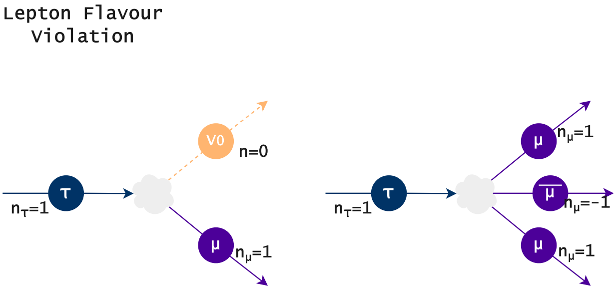


The Lepton Flavour Violation in the Standard Model

Lepton Flavour Violation

$\longrightarrow$ when lepton flavour numbers are not conserved between the initial and final states of a decay.





Flavour-Changing Neutral Currents

Neutrino oscillations

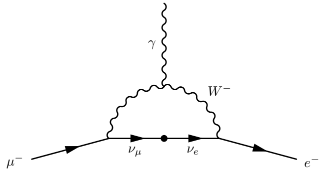
- The only source of lepton flavour violation in the Standard Model.
- First evidence in 1998 by Super-Kamiokande.
- Implies that neutrinos have masses.
- Described by the Pontecorvo–Maki–Nakagawa–Sakata matrix.

$$\underbrace{\begin{pmatrix}\nu_e\\ \nu_\mu\\ \nu_\tau\end{pmatrix}}_{\text{Flavour eigenstates}} = \underbrace{\begin{pmatrix}U_{e1} & U_{e2} & U_{e3}\\ U_{\mu1} & U_{\mu2} & U_{\mu3}\\ U_{\tau1} & U_{\tau2} & U_{\tau3}\end{pmatrix}}_{\text{PMNS matrix}} \underbrace{\begin{pmatrix}\nu_1\\ \nu_2\\ \nu_3\end{pmatrix}}_{\text{Mass eigenstates}}$$

FCNC in the SM

- Neutrino oscillations introduce FCNC via $W^{\pm}$ boson loops.
- Processes heavily suppressed, with rates **below** $\mathcal{O}(10^{-50})$.

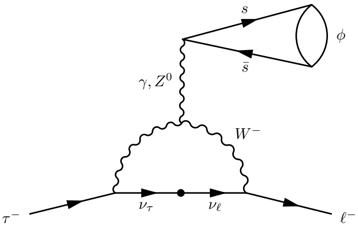
Example: LFV decay $\mu^- \to e^-\gamma$ via neutrino oscillations:



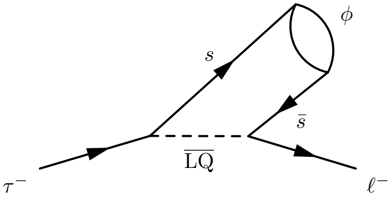
LFV in New physics – $\tau \to \ell V^0$

$\tau \to \ell V^0$ ($\ell = e, \mu$; V^0: neutral vector meson) LFV decays can be enhanced in many new physics models: MSSM, Type-III Seesaw, $SO(10)$ GUT, SM + Heavy Dirac Neutrinos, Littlest Higgs Model with T-parity, Unparticles...

$\tau \to \ell\phi$ ($\phi = s\bar{s}$ meson of mass $\sim$1020 $\mathrm{MeV}/\mathrm{c}^2$) in particular is related to the U_1 vector leptoquark hypothesis $\to$ could explain both $R_{D^{(*)}}$ and $R_{K^{(*)}}$ anomalies.



$\tau^- \to \ell^-\phi$ via neutrino oscillations

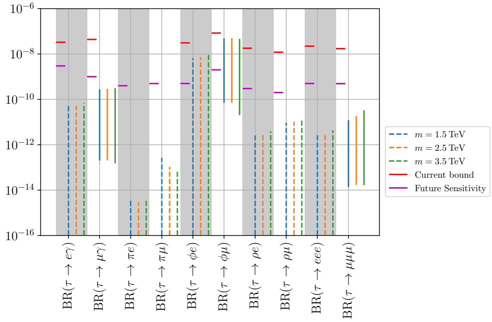


$\tau^- \to \ell^-\phi$ via leptoquark interaction

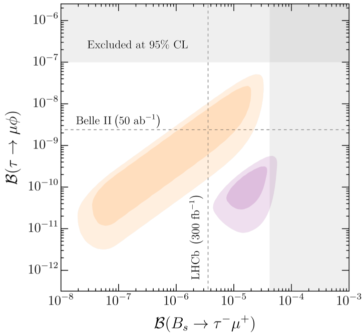
LFV in New physics – $\tau \to \ell V^0$

$\tau \to \ell V^0$ ($\ell = e, \mu$; V^0: neutral vector meson) LFV decays can be enhanced in many new physics models: MSSM, Type-III Seesaw, $SO(10)$ GUT, SM + Heavy Dirac Neutrinos, Littlest Higgs Model with T-parity, Unparticles...

$\tau \to \ell\phi$ ($\phi = s\bar{s}$ meson of mass $\sim$1020 MeV/c^2) in particular is related to the U_1 vector leptoquark hypothesis $\to$ could explain both $R_{D^{(*)}}$ and $R_{K^{(*)}}$ anomalies.



C. Hati et al., Eur.Phys.J.C 81 (2021) 12, 1066



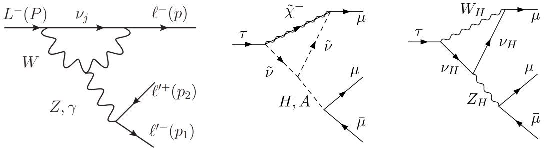
C. Cornella et al., JHEP 08 (2021) 050

LFV in New physics – $\tau \to \ell\ell\ell$

$\tau \to \ell\ell\ell$ ($\ell = e, \mu$) LFV decays involve in:

New Physics models	Limit BF for $\tau^- \to \mu^-\mu^+\mu^-$
Littlest Higgs with T-parity	10^{-8}
R-parity violating SUSY	10^{-8}
Non-universal Z'	10^{-8}
MSSM + seesaw*	10^{-9}
SUSY SO(10)	10^{-10}
SUSY Higgs	10^{-10}
SM + heavy Majorana ν	10^{-10}

Giffels et al. 2008



(a) Neutrino oscillations (b) see–saw MSSM (c) Littlest Higgs with T-parity

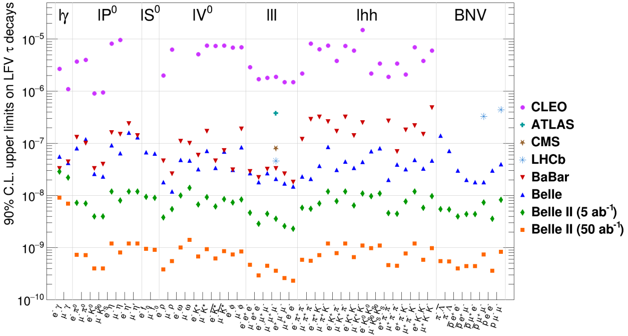
Strength: background free $\Leftarrow$ purely leptonic final state.

τ LFV searches at colliders

Current status

Around 50 τ LFV channels studied in the past two decades; best upper limits on branching fractions set by the Belle experiment (10^{-8} — 10^{-7} range).

Decay channel	$\mathcal{B}_{\text{UL}} \times 10^{-8}$ @90% CL			
	BaBar	Belle	LHCb	CMS
$\tau \to e\phi$	3.1	3.1	-	-
$\tau \to \mu\phi$	19	8.4	-	-
$\tau \to \mu\mu\mu$	3.3	2.1	4.6	8.0

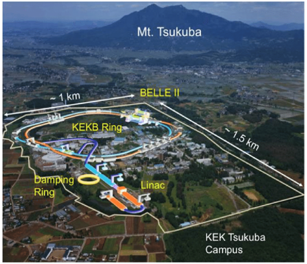


Banerjee et al. 2022; Kou et al. 2019

Prospects

Belle II expected to improve current limits by at least 1 order of magnitude $\to$ sensitive to some NP models.

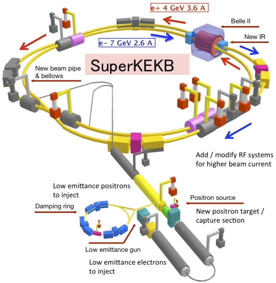
The Belle II experiment



- Located at **KEK, Tsukuba** $\to \sim 60\,km$ from Tokyo Japan
- Upgrade of Belle and KEKB, ran from 1999 to 2010
- **SuperKEKB e^+e^- collider** of $3\,km$ circumference
- **Data taking start in 2019**, with almost complete detector
- 1st Long shutdown since summer 2022 to fully install PXD

- Around 26 countries
- More than 1000 physicists
- 400 students

SuperKEKB and status of Belle II

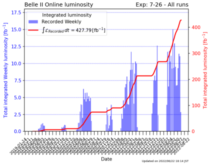


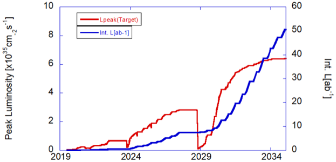
SuperKEKB

- Asymmetric **Electron** (7 GeV) **Positron** (4 GeV) collider
- **Produce:**
  - $e^+e^- \to \Upsilon(4S)[10.58\ \text{GeV}] \to B\bar{B}\ [\sigma = 1.1\ nb]$
  - $e^+e^- \to \tau^+\tau^-\ [\sigma = 0.9\ nb]$
  - and also $u\bar{u}$, $d\bar{d}$, $s\bar{s}$, $c\bar{c}$, e^+e^-, $\mu^+\mu^-$ ...

Targets

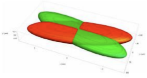
- **Peak luminosity:** today $4.7 \times 10^{34}\ cm^{-2}s^{-1} \to$ target $\sim 6 \times 10^{35}\ cm^{-2}s^{-1}$
- **data collected:** today $427.79\ \text{fb}^{-1} \to$ target $50\ \text{ab}^{-1}$

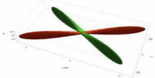




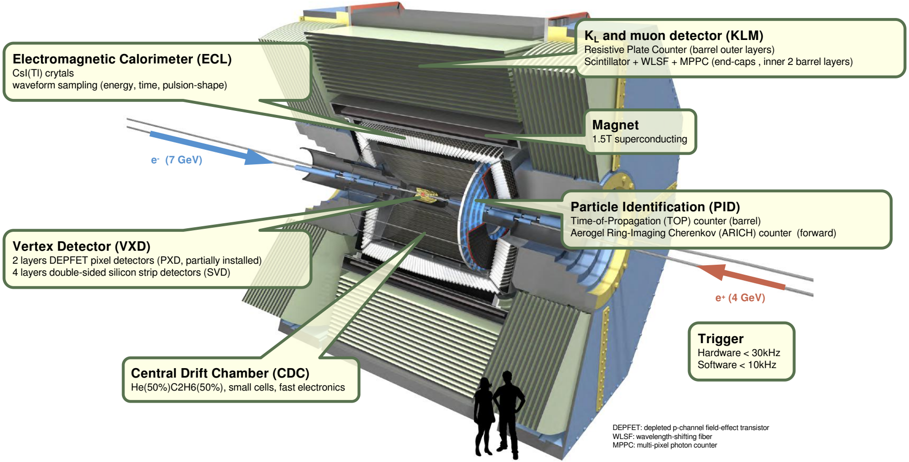
Future improvements

Higher beam currents + Lower beam size $\to$ **30**$\times$ KEKB peak luminosity


KEKB Interaction region beam overlap


SuperKEKB Interaction region beam overlap

Belle II Detector



DEPFET: depleted p-channel field-effect transistor
WLSF: wavelength-shifting fiber
MPPC: multi-pixel photon counter

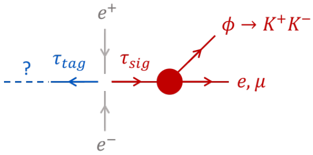
$\tau \to \ell\phi$ - *Analysis strategy*

"Inclusive" reconstruction

Final state of signal τ: 1 charged lepton + 2 charged kaons from ϕ resonance ($\mathcal{B} \sim 50\%$).
Oppositely charged τ (*tag*) not constrained to 1-prong decays $\Rightarrow$ increases signal efficiency by 30%, adding:

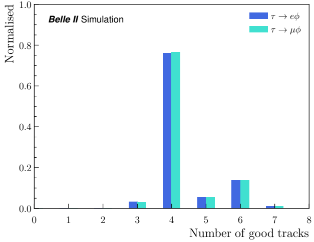
- decays of τ_{tag} into 3-prong ($\mathcal{B} \sim 15\%$),
- events with missing track(s),
- events with additional track(s) (fake/clone/beam-induced)...

... but also other sources of background!



Framework

- 2D space $(M_\tau, \Delta E_\tau)$ where $\Delta E_\tau = E_\tau^{\text{CM}} - \sqrt{s}/2$.
- Define a *signal region* (SR) around expected signal peak ($M_\tau \simeq 1.777$, $\Delta E_\tau \simeq 0$).
- SR blinded in data. Upper limit to be computed in this region.



Background suppression strategy

- Roughly defined **cut-based selection** against background with clear differences from signal in variables distributions.
- **Boosted decision trees** (BDTs) against remaining contributions.

$\tau \to \ell\phi$ - Event reconstruction

Samples

Monte Carlo simulation:

- Signal: $\tau^{\pm} \to e^{\pm}\phi$; $\tau^{\pm} \to \mu^{\pm}\phi$
- Background: $\tau^{+}\tau^{-}$, $q\bar{q}$, $B\bar{B}$, $\ell\ell(XX)$ ($\ell = e, \mu$; $X = e, \mu, \tau, \pi, K, p$).

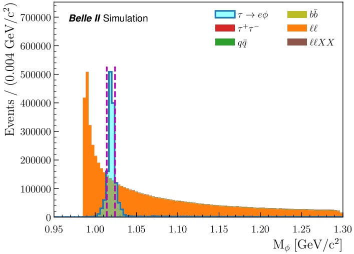
Belle II data: **189.9 fb^{-1}** (2019-2021).

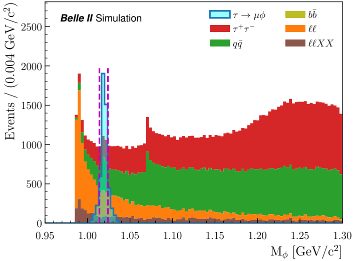
Event requirements

Track selections:

- Originate from the interaction point (IP): "good tracks"; at most 6 in event.
- Lepton tracks selected with particle identification variables (PID).
- Kaon tracks required to be in CDC acceptance $\to$ no PID requirement but selection on ϕ mass instead: $1.014 < M_{\phi} < 1.024$ GeV/c^2.

Events required to fire **low-multiplicity** Level1 (L1) ECL **triggers**.



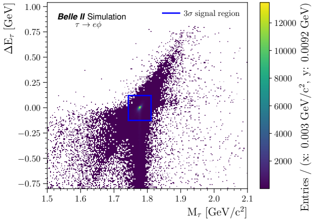


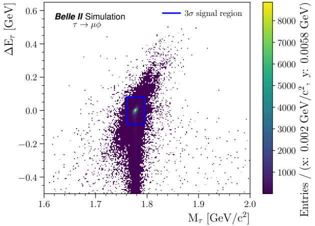
$\tau \to \ell\phi$ - Signal regions

Resolution σ computed from fit to M_τ and ΔE_τ (more details in slide 26).

- Signal region to compute the observed limit: 3σ.
- Region blinded in data: 5σ (as a precaution).
- All events constrained to: 20σ.

	$\tau \to e\phi$		$\tau \to \mu\phi$	
	M_τ [MeV/c²]	ΔE_τ [MeV]	M_τ [MeV/c²]	ΔE_τ [MeV]
3σ region	[1.743, 1.811]	[−0.125, 0.120]	[1.760, 1.794]	[−0.083, 0.080]
5σ region	[1.721, 1.834]	[−0.207, 0.202]	[1.749, 1.805]	[−0.137, 0.134]
20σ region	[1.551, 2.004]	[−0.8, 0.8]	[1.665, 1.890]	[−0.5, 0.5]
Resolution	11.3 ± 0.2	41.0 ± 0.8	5.6 ± 0.2	27.1 ± 0.6





$\tau \to \ell\phi$ - Background suppression: Preselection

Signal efficiency ($\varepsilon_{\ell\phi}$) and background rejection (r_{bkg}) after other preselection criteria are applied:

Mode	Preselection	$\varepsilon_{\ell\phi}$ (%)	r_{bkg} (%)
$e\phi$	$\theta_{\tau-\text{closest}} > 0.02$ rad	98.9	53.5
	CLEO thrust $0 < 8.5$ GeV/c	97.7	63.1
	$\theta_{K_1-K_2} > 0.01$ rad	99.8	5.5
	electronID of lepton > 0.9	91.0	59.7
	KaonID of leading kaon > 0.6	67.8	92.2
	Total (full preselection)	58.7	99.7
$\mu\phi$	Thrust axis magnitude < 0.99	97.8	9.6
	muonID of lepton > 0.99	95.2	27.1
	KaonID of leading kaon > 0.6	69.3	80.8
	Total (full preselection)	64.5	89.8

$\rightarrow$ $e\phi$ preselection mostly against Bhabha scattering background.
$\rightarrow$ kaonID not very efficient, affects signal efficiency.

Punzi figure of merit

Indicator of selection performance:

$$\frac{\varepsilon_{\ell\phi}}{a/2 + \sqrt{B}}$$

$\varepsilon_{\ell\phi}$: signal efficiency; B: background yield; $a = 3$ (significance).

	$\tau \to e\phi$	$\tau \to \mu\phi$
Signal efficiency	10.9% (20σ) 8.5% (3σ)	10.3% (20σ) 8.4% (3σ)
Background yields at 189.9 fb^{-1}	$\sim$1300 (20σ) $\sim$20 (3σ)	$\sim$260 (20σ) $\sim$4 (3σ)
Punzi FOM (3σ)	0.014	0.024

$\rightarrow$ $q\bar{q}$ continuum main background source (KKK, $\pi\pi\pi$, ...) + Bhabha for $\tau \to e\phi$.

$\tau \to \ell\phi$ - Background suppression: BDT optimisation

BDT strategy

MC simulation split into 3 samples:

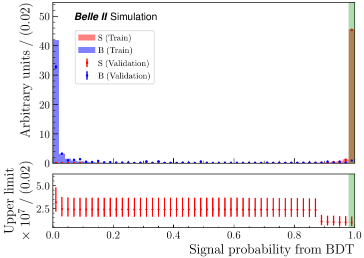
- one for the *training* of the BDT,
- one to *validate* the BDT performance,
- one to extract the expected $\varepsilon_{\ell\phi}$ and B for the limit computation.

Training carried out using the `XGBoost` library.
Evaluation metric: logarithmic loss function.

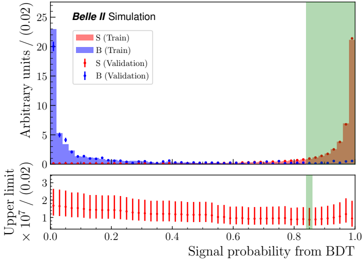
Optimisation of BDT selection

- Expected upper limit (90% C.L.) in validation sample as a function of the signal probability.
- CLs method, asymptotic formula (see slide 24).
- Selection at probability with smallest limit.

$\tau \to e\phi$



$\tau \to \mu\phi$

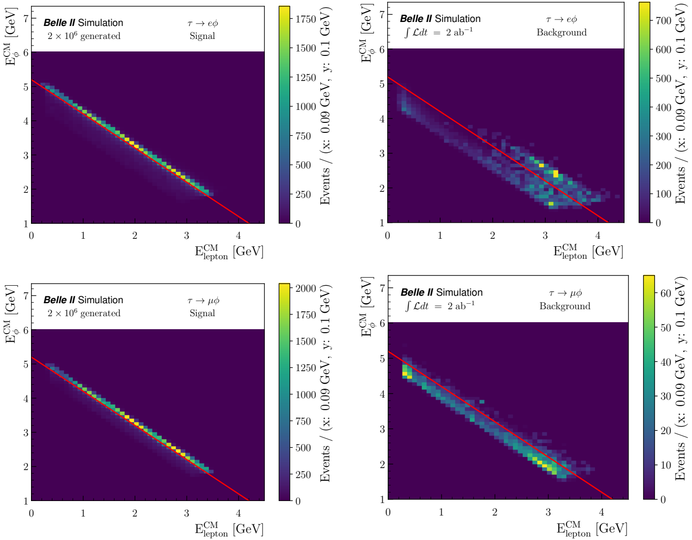


$\tau \to \ell\phi$ – Background suppression: 2- vs 3-body decays

Further background suppression against 3-body decay background:

$$E_{\phi}^{\mathrm{CM}} > -1 \times E_{\mathrm{lepton}}^{\mathrm{CM}} + 5.2.$$

- $\sim$80% relative signal efficiency and $\sim$85% background rejection *if applied before BDT training*.
- Applied after BDT training to increase the statistics.



$\tau \to \ell\phi$ – *Background suppression: Results*

In the SR, after the full selection:

	$\tau \to e\phi$	$\tau \to \mu\phi$
Signal efficiency	7.9%	7.2%
Background yields at 189.9 fb^{-1}	0.92	0.83
Punzi FOM	0.032	0.030

Remaining backgrounds:

- $\boldsymbol{e\phi}$: $KKK, K\pi\pi, e\pi\pi, ee\pi$.
- $\boldsymbol{\mu\phi}$: $KKK, KK\pi, K\pi\pi, \pi\pi\pi, \mu\pi\pi$.

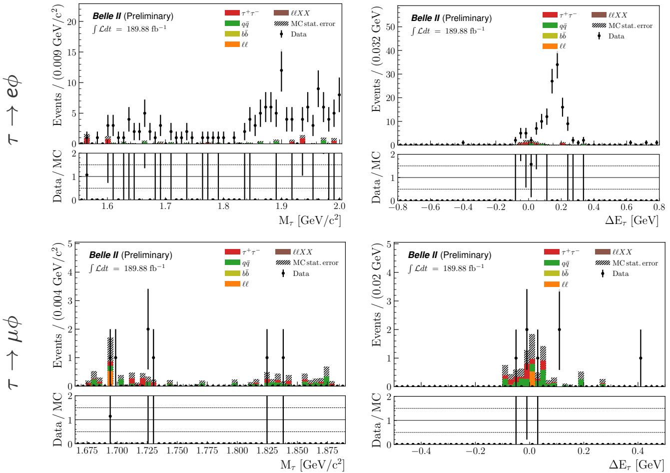
} Misidentification, underperforming kaonID.

$\tau \to \ell\phi$ - Data-MC comparison (1)

Comparison inside the *sidebands*: $5 - 20\sigma$ **region in** $(M_\tau, \Delta E_\tau)$ **space.**

	$\tau \to e\phi$	$\tau \to \mu\phi$
Data yield	138	7
MC yield	6.33	5.06
Data/MC ratio	$22 \pm 4_{\text{stat}}$	$1.4 \pm 0.6_{\text{stat}}$

Clear excess of data events ($\Delta E \sim 0.2$ **GeV)**
$\to$ **background not simulated in MC samples.**



$\tau \to \ell\phi$ – Reduction of data excess (e channel)

Nature of the data excess

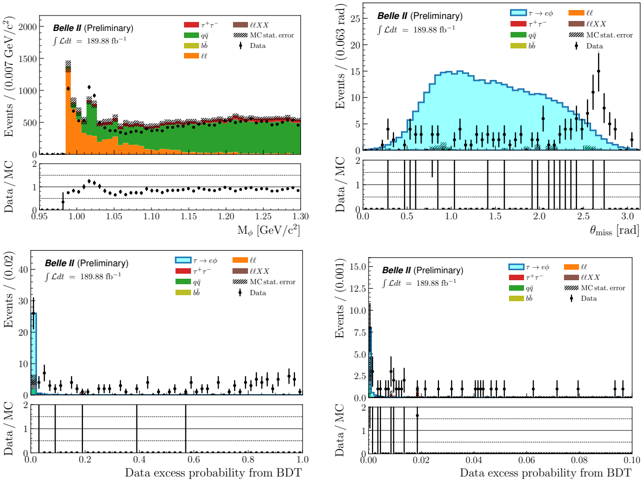
- Peak appears in data at ϕ meson mass after the preselection.
- Missing momentum in backward region (large polar angle).
- Mostly events with 4 tracks.

$\Rightarrow$ probably diphoton events going into hadrons with ϕ resonance, not simulated.

Data excess veto

Second BDT trained on 10% of data, before the selection, *inside the sidebands* (no unblinding).

- Target: data events in $-0.1 < \Delta E < 0.4$ GeV.
- Signal MC added as “background”.
- Veto on probability output at 0.001.



$\tau \to \ell\phi$ - Data-MC comparison (2)

$*$: from validation sample.

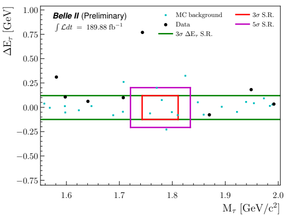
Result		Region	Mode $e\phi$	Mode $\mu\phi$
Signal efficiency $\varepsilon_{\ell\phi}$		SR	6.0%	7.2%
MC background yield		SB	3.34	5.06
		RSB	3.10 / 2.53*	4.10 / 4.06*
		SR	0.38 / 0.36*	0.83 / 0.76*
Background ratio in MC		SR / RSB	0.13	0.19
Data yield	N_{obs}	SB	8	7
		RSB	5	4
	N_{exp}	SR	0.66	0.78
Data / MC ratio		SB	$2.4 \pm 1.0_{stat}$	$1.4 \pm 0.6_{stat}$
		RSB	$1.6 \pm 0.8_{stat}$	$1.0 \pm 0.5_{stat}$

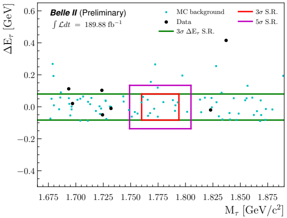
Result	Experiment	Mode $e\phi$	Mode $\mu\phi$
Signal efficiency $\varepsilon_{\ell\phi}$	BaBar	6.43%	5.18%
	Belle	4.18%	3.21%
Data yield N_{exp}	BaBar	0.68	2.76
	Belle	0.47	0.06

- SR: signal region (3σ).
- SB: sidebands ($5 - 20\sigma$).
- RSB: sidebands reduced to 3σ in ΔE_τ.

Data yield expected in SR:

$$N_{exp} = N_{obs}^{RSB} \times \frac{N_{MC}^{SR}}{N_{MC}^{RSB}}$$





$\tau \to \ell\phi$ – Summary of uncertainties

Systematic uncertainties

- **PID**: errors on PID weights, assigned to MC events as corrections according to data. *Provided by the collaboration.*
- **Tracking efficiency**: equal contribution from each of the 3 signal tracks. *From tracking performance studies.*
- **Trigger efficiency**: bias from low-multiplicity trigger requirements. *Measured on signal MC using orthogonal trigger lines.*
- **BDT selection**: bias from data-MC discrepancies for variables in the selection. *Absolute difference with signal efficiency if corrections from control samples ($D_s \to \phi\pi$ & $\tau \to \pi\pi\pi\nu_\tau$) are applied.*
- **Momentum scale**: correction on tracking momentum bias from wrong magnetic field map used in data reconstruction process. *Provided by the collaboration.*

Systematic uncertainties

Affected quantity	Source	Mode $e\phi$	Mode $\mu\phi$
$\varepsilon_{\ell\phi}$	Particle identification	0.11%	0.16%
	Tracking efficiency	0.52%	
	Trigger efficiency	0.4%	0.9%
	BDT selection	0.48%	0.95%
L	Luminosity	0.6%	
$\sigma_{\tau\tau}$	Tau-pair cross section	0.003 nb	
$N_{\rm exp}$	Momentum scale	$^{+0.0}_{-0.0}$	$^{+0.0}_{-0.2}$

Total uncertainties

Affected quantity	Mode $e\phi$	Mode $\mu\phi$
$L \times \sigma_{\tau\tau} \times \varepsilon_{\ell\phi}$	8.25%	13.39%
$N_{\rm exp}$	$^{+0.4}_{-0.4}$	$^{+0.4}_{-0.5}$

$\tau \to \ell\phi$ - Upper limit estimation

CLs method with `pyhf`. With *n* the number of expected events, *s* the signal and *b* the background:

$$E[n] = \mu s + b = \underbrace{\mathcal{B}(\tau \to \ell\phi)}_{\mu} \times \underbrace{2L\sigma_{\tau\tau}\varepsilon_{\ell\phi}}_{s} + b.$$

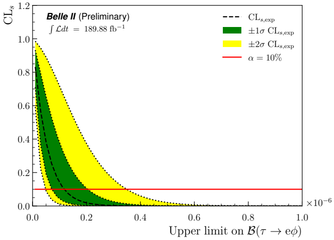
$$\mathrm{CL}_s = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} = 10\%.$$

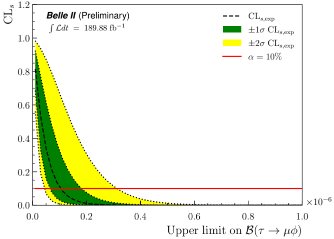
Expected upper limits at 90% C.L.:

$$\mathcal{B}^{90}_{\mathrm{UL,exp}}(\tau \to e\phi) = 1.18 \times 10^{-7},$$
$$\mathcal{B}^{90}_{\mathrm{UL,exp}}(\tau \to \mu\phi) = 1.05 \times 10^{-7}.$$

At equivalent luminosities (BaBar: 451 fb^{-1}; Belle: 854 fb^{-1}):

- $\tau \to e\phi$: expected limit competitive neither with Belle nor BaBar,
- $\tau \to \mu\phi$: same expected limit as Belle, better than BaBar (-17%).





$\tau \to \mu\mu\mu$ - Inclusive reconstruction

Events reconstruction strategy

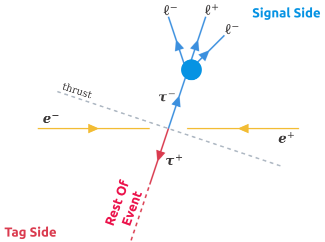
Reconstruct τ pair events with an inclusive strategy:

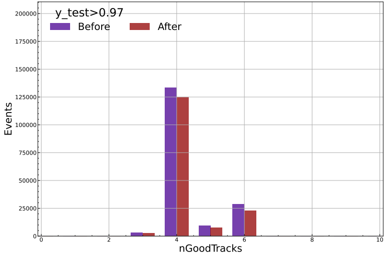
- Signal side: one τ decay into the LFV channel $\to \mu\mu\mu$
- Tag side: opposite one is put in a Rest of Event $\to$ without specific decay

This strategy allow to cover more decays channels for the tag side and improve the reconstruction efficiency.

Analysis strategy

- Define signal region in $M_{\mu\mu\mu}$ and $\Delta E_{\mu\mu\mu}$ plane to remove backgrounds, blind data and estimate expected background
- Reject low multiplicity backgrounds with **cut based preselection**
- Reject main contribution backgrounds with **Boosted Decision Tree**





Reconstruction details equivalent to the $\tau \to \ell\phi$ one in **slide 14**.

$\tau \to \mu\mu\mu$ - Signal Region in $M_{\mu\mu\mu}$; $\Delta E_{\mu\mu\mu}$ 2D plane (1)

Measure resolutions to define a **region** around **peaking signal** at:
$M_{\mu\mu\mu} \simeq 1.777\ \mathrm{GeV}/c^2$ and $\Delta E_{\mu\mu\mu} = E^{\mathrm{CM}}_{\mu\mu\mu} - \sqrt{s}/2 \simeq 0\ \mathrm{GeV}$

Fitting mass and delta energy

Sum 3 distributions:

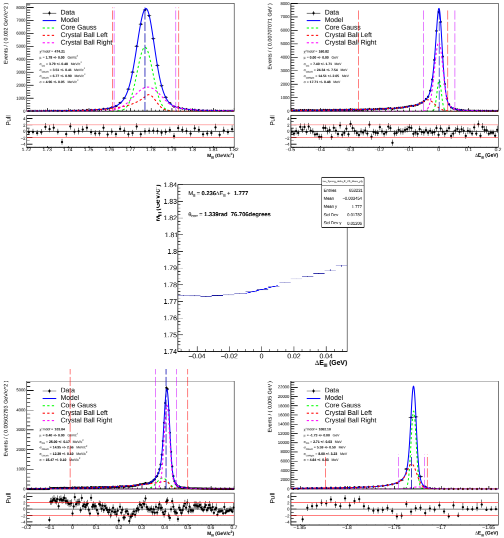
- **Gaussian**: for the peak
- **2 Crystal Balls**: for both side tails

Try decorrelate variables

$$\begin{pmatrix} M'_{\mu\mu\mu} \\ \Delta E'_{\mu\mu\mu} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} M_{\mu\mu\mu} \\ \Delta E_{\mu\mu\mu} \end{pmatrix}$$

with $\theta = 76.71^\circ$

	Mean	Resolution
$M_{\mu\mu\mu}$	$1.7769 \pm 0.0009\ \mathrm{GeV}/c^2$	$4.96 \pm 0.05\ \mathrm{MeV}/c^2$
$\Delta E_{\mu\mu\mu}$	$0.0005 \pm 0.0004\ \mathrm{GeV}$	$17.71 \pm 0.48\ \mathrm{MeV}$
$M'_{\mu\mu\mu}$	0.4044 ± 0.0008	15.47 ± 0.09
$\Delta E'_{\mu\mu\mu}$	-1.7319 ± 0.0003	4.64 ± 0.03



$\tau \to \mu\mu\mu$ - Signal Region in $M_{\mu\mu\mu}$; $\Delta E_{\mu\mu\mu}$ (2)

Use signal region as background suppression

We define several condition with:

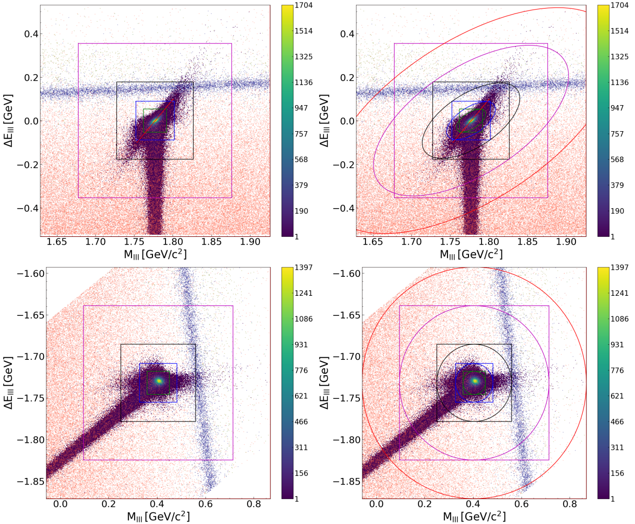
- different variable (rotated or not)
- different shape (boxes or ellipse)
- different sizes (3 or 5 resolution width)

No normalization, no weights:

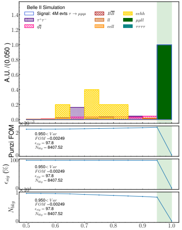
	Var	ϵ_{sig} (%)	N_{bkg}
3	M; ΔE	22.05	372
3	*Ellispe*	21.62	286
3	M'; $\Delta E'$	21.82	295
3	*Ellipse*$'$	21.40	241

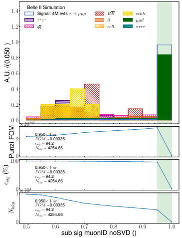
To blind

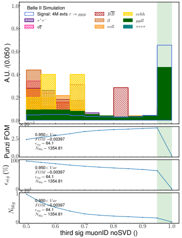
Blind data according to the signal region, only use side-bands, where we also estimate the background.

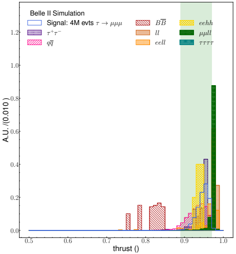


$\tau \to \mu\mu\mu$ – Background suppression: Preselections









Preselectrions

Define **set of preselection** to mainly reject background: $e^+e^- \to e^+e^-$ (Bhabha), $e^+e^- \to \mu^+\mu^-$, $e^+e^- \to \ell\ell XX$

Normalized for 400 **fb**$^{-1}$**:**

MuonID	Preselection	ϵ_{sig} (%)	N_{bkg}
$\mu ID^{lead} > 0.95$	$0.89 < thrust < 0.97$	22.80	1736.71
$\mu ID^{lead} > 0.95$	$0.93 < thrustBm < 0.95$	23.05	2072.30
$\mu ID^{lead} > 0.95$	$5.5 < E^{CMS}_{vis} < 10.5$	22.83	1546.57
$\mu ID^{lead} > 0.95$	$0.5 < E^{CMS}_{mis} < 5.0$	21.40	857.23
$\mu ID^{lead} > 0.95$	$Pt^{CMS}_{mis} > 0.2$	22.48	1108.33
$\mu ID^{lead,sub} > 0.95$	$0.89 < thrust < 0.97$	21.98	735.71
$\mu ID^{lead,sub} > 0.95$	$0.93 < thrustBm < 0.95$	22.25	1128.39
$\mu ID^{lead,sub} > 0.95$	$5.5 < E^{CMS}_{vis} < 10.5$	22.01	630.03
$\mu ID^{lead,sub} > 0.95$	$0.5 < E^{CMS}_{mis} < 5.0$	20.63	328.72
$\mu ID^{lead,sub} > 0.95$	$Pt^{CMS}_{mis} > 0.2$	21.67	458.26

All combination (μID selection + other variable) are tested with a Boosted Decision Tree train to remove other background sources.

$\tau \to \mu\mu\mu$ - Background suppression: BDT (1)

Use BDT for background rejection

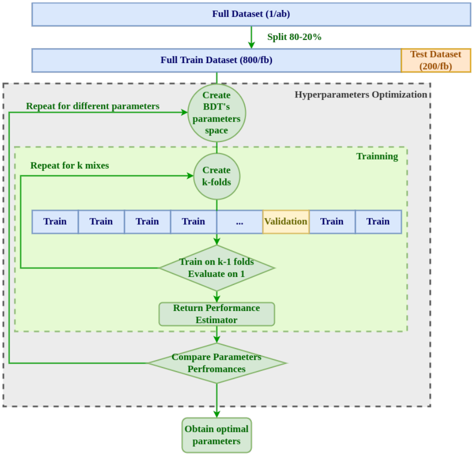
Reject backgrounds coming from continuum $q\bar{q}$ and $\tau^+\tau^-$ with a **Boosted Decision Tree** (BDT) classification **Calafiura, Rousseau, and Terao 2022**.

K-Folding

Machine learning are sensible to arbitrar sample split in test/validation/test subsamples $\Rightarrow$ **k-folds method** minimize this source of overtrainning

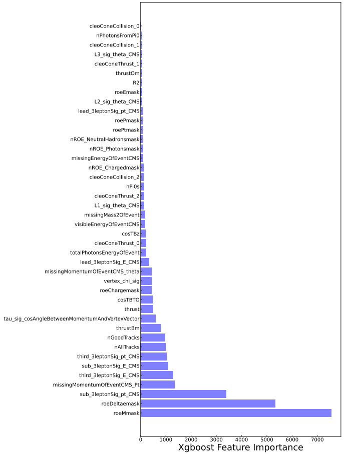
- Split in **Train|Test**
- Split in **k-folds**:
  - Use $k - 1$ folds to train
  - Use last folds to validate
  - Repeat for all $\binom{k}{k-1} = k$ combinations
  - Average k outputs

- Train BDT on Monte-Carlo Simulations
- Using python's **XGBoost** library for BDT
- Using python's **Optuna** library for optimization
- 100 Parameters setups are evaluated according to the **logarithmic loss function**



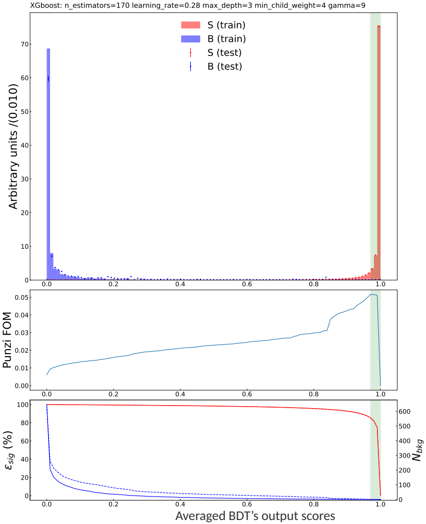
$\tau \to \mu\mu\mu$ - Background suppression: BDT (2)

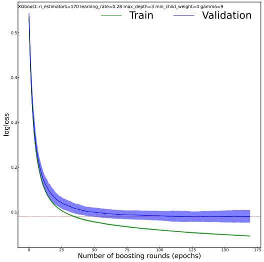
Input BDT variable ranking according to :

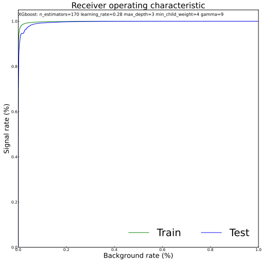


Event	Missing	Leptons	ROE	ROEmask	nROE	ROEthrust	CleoCones
thrust	E^{CMS}_{miss}	Li θ^{CMS}	E_{ROE}	E^{mask}_{ROE}	nROE$prong$	thrustBm	cleoConeThrust_i
E^{CMS}_{vis}	$p^{T,CMS}_{miss}$		M_{ROE}	M^{mask}_{ROE}	nROE$prong^{mask}$	thrustOm	cleoConeCollision_i
E^{tot}_{γ}	θ_{miss}		P_{ROE}	P^{mask}_{ROE}	nROEγ	cosTBTO	
rank $p^{T,CMS}_{sig}$	M^2_{miss}		P^T_{ROE}	$P^{T,mask}_{ROE}$	nROEγ^{mask}	cosTBz	
rank E^{CMS}_{sig}			ΔE_{ROE}	ΔE^{mask}_{ROE}	nROEh^{0}		
χ^2_{vertex}			C_{ROE}	C^{mask}_{ROE}	nROEh^{0,mask}		
$cos(\theta_{p-vertex})$							
nAllTracks							
nGoodTracks							
nPio							
nPhotonsFromPio							

$\tau \to \mu\mu\mu$ - Background suppression: BDT (3)



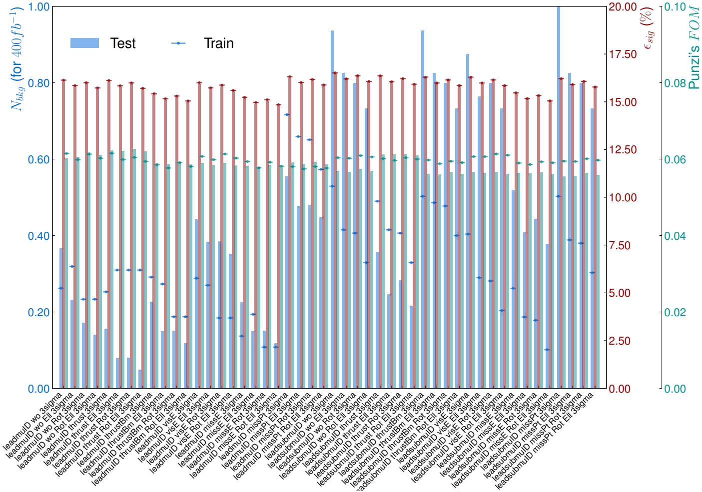




BDT Results

- good separation between signal and background
- evolution of the log loss function are close for train and test samples
- define a cut on BDT output according to the Punzi FOM

$\tau \to \mu\mu\mu$ - Background suppression



Best configuration:

- $\mu ID_{lead} > 0.95$
- $0.89 < thrust < 0.97$
- $M'_{\mu\mu\mu}$; $\Delta E'_{\mu\mu\mu}$ signal region

with:

- $\epsilon_{\mathbf{sig}} = \mathbf{15.93\%}$
- $\mathbf{N_{bkg}} = \mathbf{0.08}$
- **FOM** $= \mathbf{0.062725}$

References in 3×1 topology:

- Belle 2 cut based analysis: $\epsilon_{sig} = 12.7\%$, $N_{bkg} = 0.2$
- **Belle**: $\epsilon_{sig} = 7.6\%$, $N_{bkg} = 0.13$
- **Babar**: $\epsilon_{sig} = 6.6\%$, $N_{bkg} = 0.44$

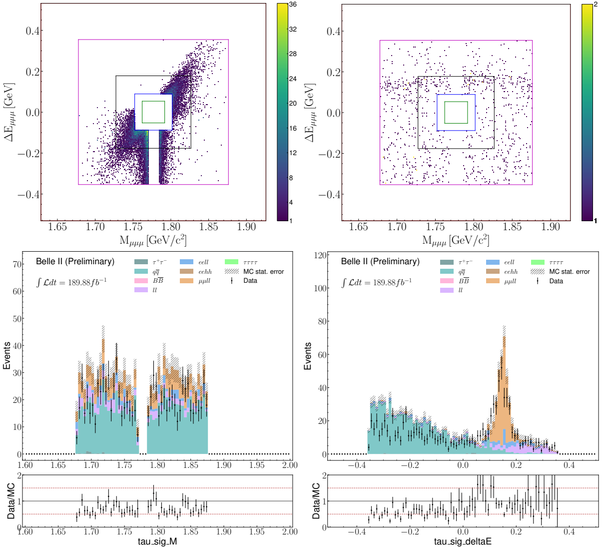
$\tau \to \mu\mu\mu$ - Data-MC comparison

Data/MC Comparison in sidebands

- Reconstruct $\tau \to \mu\mu\mu$ on 189.88 fb^{-1} data
- Blind:
  - inside sidebands the 5 to 20 Signal Region resolution
  - remove window $1.77 < M_\tau < 1.785$

Data/MC agreement

- Monte-Carlo Simulation (scaled): **1295.48 bkg evts** and **1.22%** ϵ_{sig}
- 189.88 fb^{-1} Data: **910 evts**

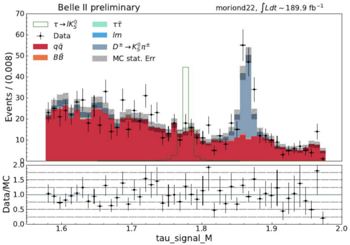


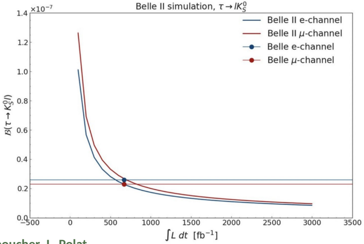
$\tau \to \ell K_S$, $\tau \to \ell\rho$, ...

$\tau \to \ell K_S$

Klemens Lautenbach (CPPM)

- Inclusive reconstruction.
- Overall good data-MC agreement outside of signal region.
- So far, no significant improvement expected w.r.t. Belle's upper limit (at 671 fb^{-1}).

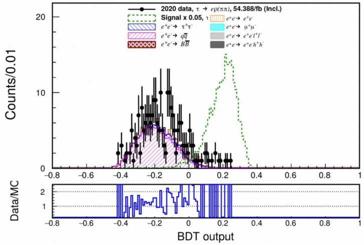




$\tau \to \ell\rho$

Laura Zani (CPPM)

- Reconstruction similar to $\tau \to \ell\phi$.
- Data excess observed as well in electron channel.



Other ongoing studies:

- $\tau \to \ell + \alpha$ (*invisible*) (Cinvestav, HEPHY, DESY, MPP, Pisa)
- $\tau \to \ell\gamma$ (Cinvestav, DESY, Pisa).
- $\tau \to \Lambda\pi$ (Zhengzhou, Fudan).
- $\tau \to \ell K^*$ (CPPM) for the next GdR ;)[2]

Summary

- Large variety of NP models predict lepton flavour violation, notably in tau decays.

- $\tau \to \mu\phi$ (leptoquark hypothesis for *B*-anomalies) and $\tau \to \mu\mu\mu$ (SUSY; background-free decay) are two of the "golden channels" studied at Belle II.

- $\tau \to \ell\phi$: analysis on $\sim$190 fb^{-1} of data; muon channel at the level of Belle and competitive with BaBar for equivalent luminosities; room for improvement in both decay modes (missing background simulation, higher statistics in MC, hopefully better PID, ...).

- $\tau^- \to \mu^-\mu^+\mu^-$: analysis on $\sim$190 fb^{-1} of data; deploy BDT background rejection and inclusive reconstruction $\to$ 2$\times$ better efficiency than Belle, Babar studies and higher than Belle 2 cut based, see data deficit in data/MC comparison.

- Many other channels actively studied; in general, stay tuned for updates next year with full luminosity collected so far ($\sim$430 fb^{-1})!

References I

Banerjee, S. et al. (2022). *Snowmass 2021 White Paper: Charged lepton flavor violation in the tau sector*. DOI: 10.48550/ARXIV.2203.14919. URL: https://arxiv.org/abs/2203.14919.

Calafiura, Paolo, David Rousseau, and Kazuhiro Terao (2022). *Artificial Intelligence for High Energy Physics*. WORLD SCIENTIFIC. DOI: 10.1142/12200. eprint: https://www.worldscientific.com/doi/pdf/10.1142/12200. URL: https://www.worldscientific.com/doi/abs/10.1142/12200.

Giffels, M. et al. (Mar. 2008). "The lepton-flavour violating decay \tau \to \mu \mu \antimu at the LHC". In: DOI: 10.1103/PhysRevD.77.073010. URL: http://arxiv.org/abs/0802.0049.

Kou, E et al. (Dec. 2019). "The Belle II Physics Book". In: *Progress of Theoretical and Experimental Physics* 2019.12. ISSN: 2050-3911. DOI: 10.1093/ptep/ptz106. URL: http://dx.doi.org/10.1093/ptep/ptz106.

Backups

$\tau \to \ell\phi$ - Event reconstruction

Samples

Monte Carlo (MC) simulation:

- Signal: $\tau^\pm \to e^\pm\phi$; $\tau^\pm \to \mu^\pm\phi$
- Background: $\tau^+\tau^-$, $q\bar{q}$, $B\bar{B}$, $\ell\ell(XX)$ ($\ell = e, \mu$; $X = e, \mu, \tau, \pi, K, p$).

Belle II data: **189.9 fb**$^{-1}$ (2019-2021).

Steps of the reconstruction

Track selections:

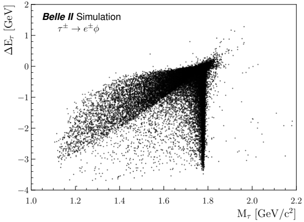
- Originate from the interaction point (IP): "good tracks"; at most 6 in event.
- Lepton tracks selected with particle identification variables (PID).
- Kaon tracks required to be in CDC acceptance $\to$ no PID requirement but selection on ϕ mass instead: $1.014 < M_\phi < 1.024\ \mathrm{GeV/c^2}$.

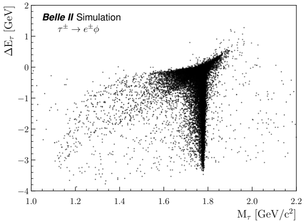
Final state radiations (FSR) = signal events displaced from signal region in $(M_\tau, \Delta E_\tau)$ space $\to$ **Bremsstrahlung corrections** on electron tracks to counteract efficiency loss.

Fit of τ and ϕ **vertices**.

Information on **Rest Of Event** (ROE, unused tracks and clusters) recovered for background suppression.

Events required to fire **low-multiplicity** Level1 (L1) ECL **triggers**.





$\tau \to \ell\phi$ – Background suppression: Variables

Related to	Name	Definition
Lepton	$E^{\mathrm{CM}}_{\mathrm{lepton}}$	Energy in the center of mass frame (GeV).
	$p^{\mathrm{T}}_{\mathrm{lepton}}$	Transverse momentum (GeV/c).
	$\theta^{\mathrm{CM}}_{\mathrm{lepton}}$	Polar angle in the center of mass frame (rad).
Kaons	$\theta_{K_1-K_2}$	Angle between the two kaons (rad).
	KaonID of leading kaon	KaonID of the kaon with the highest transverse momentum.
3-prong	$p^{\mathrm{T,CM}}_{\mathrm{lead}}$, $p^{\mathrm{T,CM}}_{\mathrm{sub\text{-}lead}}$, $p^{\mathrm{T,CM}}_{\mathrm{third}}$	Transverse momentum in the center of mass frame of the leading, sub-leading and third 3-prong tracks (GeV/c).
ϕ meson	M_ϕ	Mass (GeV/c^2).
	E_ϕ	Energy (GeV).
	p^{T}_ϕ	Transverse momentum (GeV/c).
τ	–	τ thrust axis magnitude.
	$\cos\theta^{\mathrm{thrust}}_{\tau-z}$	Cosine of angle between τ thrust axis and z axis.
	$\theta_{\tau-\mathrm{closest}}$	Angle between τ and closest track (rad).
	$\cos\theta_{p-\mathrm{vertex}}$	Cosine of the angle between the τ momentum and the τ decay vertex vector with respect to the interaction point.
Vertexing	–	χ^2 probability of ϕ vertex fit result.
	–	χ^2 probability of τ vertex fit result.
Neutral particles	γ multiplicity	Number of photons in event.
	π^0 multiplicity	Number of neutral pions in event.
	γ from π^0 multiplicity	Number of photons produced by π^0 decays in event.

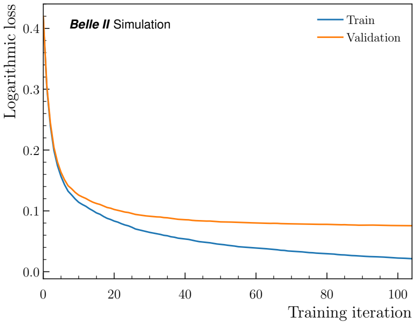
Related to	Name	Definition
Missing momentum	M^2_{miss}	Missing mass squared (GeV^2/c^4).
	$E^{\mathrm{CM}}_{\mathrm{miss}}$	Energy in the center of mass frame (GeV).
	$p^{\mathrm{T}}_{\mathrm{miss}}$	Transverse momentum (GeV/c).
	θ_{miss}	Polar angle (rad).
	$\cos\theta_{\mathrm{lepton-miss}}$	Cosine of angle between lepton and missing momentum.
Event	–	Number of tracks.
	–	Number of good tracks.
	–	Thrust axis magnitude.
	$E^{\mathrm{CM}}_{\mathrm{vis}}$	Visible energy in the center of mass frame (GeV).
	R_2	Reduced Fox-Wolfram moment $R_2 = H_2/H_0$.
	CLEO thrust 0, 1, 2	0^{th}, 1^{st}, 2^{nd} order CLEO cone with respect to thrust axis (GeV/c).
	CLEO collision 0, 1, 2	0^{th}, 1^{st}, 2^{nd} order CLEO cone with respect to collision axis (GeV/c).
Rest Of Event	–	ROE thrust axis magnitude.
	$\cos\theta^{\mathrm{thrust}}_{\tau-\mathrm{ROE}}$	Cosine of angle between τ and ROE thrust axes.
	$M_{\mathrm{ROE(masked)}}$	Mass of the masked ROE (GeV/c^2).
	$E_{\mathrm{ROE(masked)}}$	Energy of the masked ROE (GeV).
	$\Delta E_{\mathrm{ROE(masked)}}$	Difference between the energy of the masked ROE and half the center of mass energy (GeV).
	$p_{\mathrm{ROE(masked)}}$	Momentum of the masked ROE (GeV/c).
	$p^{\mathrm{T}}_{\mathrm{ROE(masked)}}$	Transverse momentum of the masked ROE (GeV/c).
	–	Total charge of masked ROE.
	–	Number of charged particles in masked ROE.
	–	Number of photons in masked ROE.
	–	Number of neutral hadrons in masked ROE.

$\tau \to \ell\phi$ – Background suppression: BDT optimisation

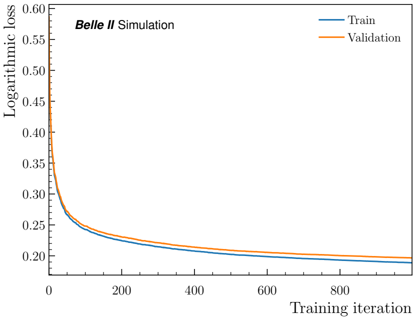
Mitigation of overfitting

- Early stopping implemented: logloss in training sample vs validation sample.
- BDT parameters tested (max. tree depth, min. loss reduction, ...), selected according to area under receiver operating characteristic (ROC) curve in validation sample.

$\tau \to e\phi$

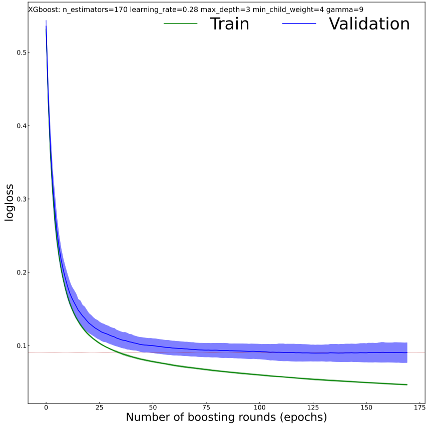


$\tau \to \mu\phi$



$\tau^- \to \mu^-\mu^+\mu^-$ - BDT overtraining health

Train & Validation Logloss in function of the boosting rounds to visualize the training behaviour:



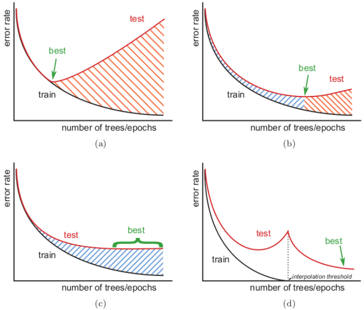


Fig. 3. Overtraining estimation using the error rate as a function of the number of trees (for boosted decision trees) or epochs (for neural networks). Black curves are measured on the training sample and red curves on the validation sample. The optimal classifier corresponds to the "best" label. The hatched areas represent overtraining: beneficial in blue (but underfitting), detrimental in orange (overfitting). (a) Typical curves, with the best model at the minimum of the testing curve, and overfitting beyond with decrease of performance. (b) The best model is overtrained but still improves performance. (c) Typical curves for boosted decision trees with flattening testing error rate: all models in the flat area perform equally well despite increasing overtraining. (d) Interpolation regime: the best classifier is obtained after the training error has reached zero.

P. Calafiura et al., Artificial Intelligence for High Energy Physics