

$\tau \rightarrow \mu\mu\mu$ approval presentation

Johannes Albrecht¹, Marta Calvi², Marcin Chrzaszcz^{3,4},
Laura Gavardi¹, Jon Harrison⁵, Basem Khanji², George
Lafferty⁵, Tatiana Likhomanenko⁶, Eduardo Rodrigues⁵,
Nicola Serra³, Paul Seyfert⁷

Referees:

Benoit Viaud (Chair) , Matteo Rama, Frederic Machefert (EB)

¹Dortmund, ²Milano Bicocca, ³Zurich, ⁴IFJ Cracow, ⁵Manchester,
⁶Yandex, ⁷Uni. Heidelberg



University of
Zurich ^{UZH}



August 10, 2014

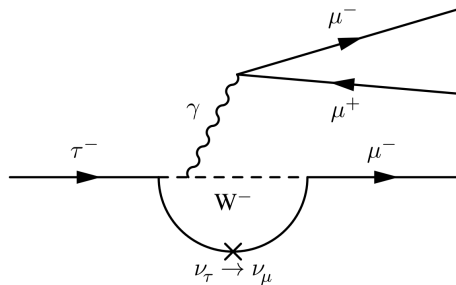
- 1 Introduction
- 2 Selection
- 3 Multivariate technique
- 4 Normalisation
- 5 Backgrounds
- 6 Expected limit
- 7 Model dependence
- 8 Unblinded results



- TWiki: https://twiki.cern.ch/twiki/bin/viewauth/LHCbPhysics/Tau_LFV_3fb
- ANA note: LHCb-ANA-2014-005
- Paper draft: LHCb-PAPER-2014-X
- Target journal: JHEP
- Conference: Tau 2014



Status of $\tau \rightarrow \mu\mu\mu$



- Charged Lepton Flavour Violation process
- Possible as penguin with neutrino oscillation
- SM prediction is beyond experimental reach $O(10^{-40})$

current limits (90% CL)

BaBar 3.3×10^{-8}

Belle 2.1×10^{-8}

LHCb 8.0×10^{-8}

BSM predictions

var. SUSY 10^{-10}

non universal Z' 10^{-8}

mSUGRA+seesaw 10^{-9}

and many more...

- Following same approach as other RD searches.
- Loose stripping selection
- Multivariate classification in: mass, PID, “geometry/topology”
- Binning optimisation.
- Relative normalisation ($D_s \rightarrow \phi(\mu\mu)\pi$)
- Invariant mass fit for expected background in each likelihood bin: fit in $|m - m_\tau| > 30$ MeV
- “middle sidebands” for classifier evaluation and tests. ($20 \text{ MeV} < |m - m_\tau| < 30 \text{ MeV}$).
- CLs for limit calculation



- Consider five production channels (fractions at 8 TeV):
 - Prompt $D_s \rightarrow \tau$ ($72.4 \pm 2.7\%$)
 - Prompt $D^+ \rightarrow \tau$ ($4.2 \pm 0.7\%$)
 - Non-prompt $D_s \rightarrow \tau$ ($8.5 \pm 1.7\%$)
 - Non-prompt $D^+ \rightarrow \tau$ ($0.17 \pm 0.04\%$)
 - $X_b \rightarrow \tau$ (meson or baryon) ($14.7 \pm 2.3\%$)
- Use $\sigma(b\bar{b})$ at 8 TeV from LHCb
- Use Pythia scaling for $\sigma(c\bar{c})$ at 8 TeV

$B(D^+ \rightarrow \tau)$

- There is no measurement of $B(D^+ \rightarrow \tau)$
- One can calculate it from: $B(D^+ \rightarrow \mu\nu_\mu)$ + helicity suppression + phase space
- hep-ex:0604043
- $B(D^+ \rightarrow \tau\nu_\tau) = 1.0 \times 10^{-3}$

- Data from Reco14Stripping20(r1)
- Much MC
 - 24M Inclusive background events ($b\bar{b}$ and $c\bar{c}$)
 - 10M Exclusive background events ($D_s \rightarrow \eta(\mu\mu\gamma)\mu\nu_\mu$)
 - 2M Signal events (split over 5 production channels)
 - 12M $D \rightarrow K\pi\pi$
 - 10M $D^* \rightarrow D(K\mu\nu_\mu)\pi$

⇒ Generator level cuts for improved use of computing resources

- ~ 14 times more signal statistics after stripping
- ~ 2 times more background statistics
- Mix τ production on ntuple level instead of reweighting.
⇒ Ease up ntuple usage (no forgotten weighting, no double weighting, ...)

Stripping and selection

	$\tau \rightarrow \mu\mu\mu$	$D_s \rightarrow \phi\pi$
μ^\pm, π^\pm		
$p\tau$	> 300MeV	
Track χ^2/ndf	< 3	
IP χ^2/ndf	> 9	
track ghost probability	< 0.3	
μ pairs		
$m_{\mu^+\mu^-} - m_\phi$	> 20MeV	< 20MeV
$m_{\mu^+\mu^-}$	> 450MeV	-
$m_{\mu^+\mu^+}$	> 250MeV	-
τ^\pm and D_s		
Δm	< 400MeV	< 50MeV
Vertex χ^2	< 15	
IP χ^2	< 225	
$\cos \alpha$	> 0.99	
$c\tau$ (stripping)	> 100 μm	
	no PV refitting	
decay time (offline)	> -0.01 ns & < 0.025 ns	
	PV refitting	



	signal	normalisation
L0 ¹	L0Muon TOS	
Hlt1 ¹	Hlt1TrackMuon TOS	
Hlt2 2011	Hlt2CharmSemilepD2HMuMu TOS Hlt2TriMuonTau TOS	Hlt2DiMuonDetached ² TOS
Hlt2 2012	Hlt2TriMuonTau ¹ TOS	Hlt2DiMuonDetached ² TOS

¹ triggers in 2012

- Cuts changed through 2012
- emulated two different TCKs for 2012
- Found negligible differences
- Choice of triggers were optimised based on $\frac{s}{\sqrt{b}}$ FOM.



Much work has been put in improving our geometric and kinematic classifier:

- Classify the displaced 3-body decay properties of a signal candidate
- Revisit variable choice
- Revisit classification technique
- More toolkits tried: MatrixNet, NeuroBayes, TMVA
- Retune input variables
($B_s^0 \rightarrow \mu\mu$ isolation \rightarrow BDT isolation:
CERN-THESIS-2013-259)
- Blending technique.



Basic Setup - Step I

- Train 1/3 signal MC against 1/2 background MC
- variables
 - DOCA
 - Vertex χ^2
 - τ decay time
 - τ IP χ^2
 - min. μ IP χ^2
 - τ pointing angle
 - τ p_T
 - max. track χ^2
 - $B_s^0 \rightarrow \mu\mu$ track isolation
 - Cone isolation
 - BDT isolation
- Using these variables, train several classifiers ("Base") for each of the τ source



Step II

- Train using second 1/3 signal MC against second 1/2 background MC.
- Introduce Blending technique

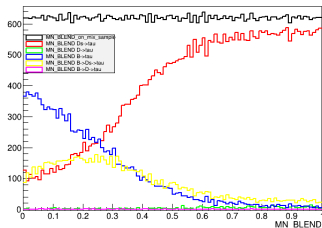
Blending technique

- For each signal channel we train: one BDT, three Fisher classifier, four MLPs, one FDA, one LD classifier and MatrixNet classifier.
 - One final MatrixNet classifier using the 13 base variables and the base classifiers as input
-
- All evaluation is done on 3rd 1/3 signal sample and middle side-bands.
 - Splitting into independent samples makes the procedure insensitive to overtraining.

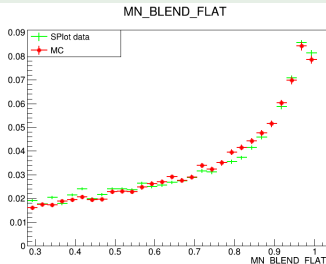
Performance of Blend classifier

- Classifier prefers τ from prompt D_s

MC response for different τ production channels



response for $D_s \rightarrow \phi\pi$ data and MC

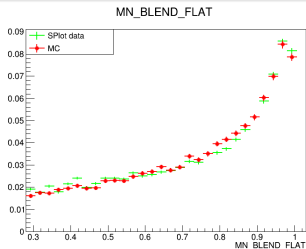


- Assume all differences between $\tau \rightarrow \mu\mu\mu$ and $D_s \rightarrow \phi\pi$ come from kinematics (mass, resonance, decay time)
- Get correction $D_s \rightsquigarrow \tau$ from MC
- Apply corrections to $D_s \rightarrow \phi\pi$ on data

validation

- done for 2011 analysis, treating smeared MC as data

- $D_s \rightarrow \phi\pi$ well modelled in MC



- We used ProbNNmu already in the previous round of the analysis
- Now use MC12TuneV2 (latest)
- Two-fold reason:
 - Expect better performance than CombDLL variables
 - “one variable for everything”:
with CombDLL we needed both $\text{CombDLL}(\mu - \pi)$ and $\text{CombDLL}(\mu - K)$

- Since PIDCalib tool didn't work for us, we choose a phenomenological approach.
- Many thanks to Barbara Sciascia for help understanding this problem.

phenomenologic treatment

- correlations are small in $D_s \rightarrow \phi\pi$ data and MC
 $\varepsilon(\text{cut on one muon})^2 = \varepsilon(\text{cut on two muons})$
- \Rightarrow use $c^3 = (\varepsilon(\text{cut and fit})/\varepsilon(\text{PIDCalib}))^3$ as correction to PIDCalib for $\tau \rightarrow \mu\mu\mu$
- assign error of 0.02 for c

- Many x-checks done.
- Everything works fine.

Binning optimisation

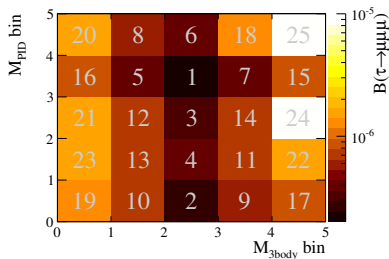
- How to optimise the binning in two classifiers?
- 1 fb^{-1} CONF note: two one-dimensional optimisations as in $B_s^0 \rightarrow \mu\mu$
- 1 fb^{-1} PAPER: iterative loop of one-dimensional optimisations optimising one classifier on the sensitive range of the other classifier
- Now: optimise two-dimensions (optimise bin boundaries in both dimensions at the same time)
- Unchanged: don't use lowest likelihood bins (reflection backgrounds, no sensitivity gain)



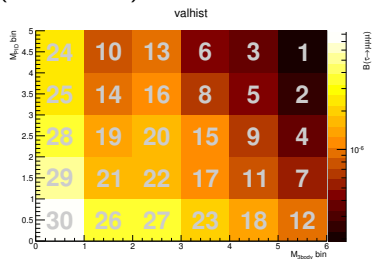
Impact of new binning optimisation

- Removal of tiny bins which contribute negligible sensitivity
- Colour: limit obtained, using only this particular bin
- Number: rank of that bin (1=best sensitivity bin)

old analysis



new analysis
(2011 data)



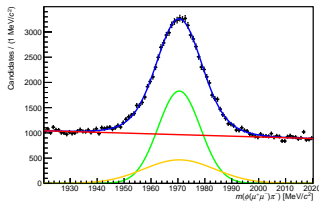
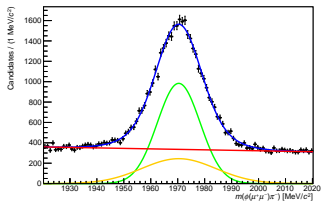
University of
Zurich^{ZH}



Mass shape

- Double-Gaussian with fixed fraction (70 % inner Gaussian)
- Fix fraction to ease calibration:

$$\sigma_{data}^{\tau} = \frac{\sigma_{MC}^{\tau}}{\sigma_{D_s}} \times \sigma_{data}^{D_s}$$

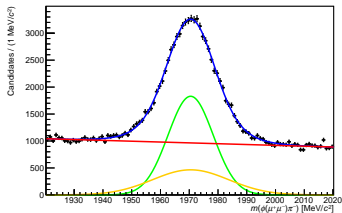
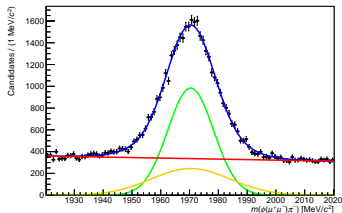


Calibrated τ Mass shape	7 TeV	8 TeV
Mean (MeV)	1779.1 ± 0.1	1779.0 ± 0.1
σ_1 (MeV)	7.7 ± 0.1	7.6 ± 0.1
σ_2 (MeV)	12.0 ± 0.8	11.5 ± 0.5

Relative normalisation

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) = \frac{\mathcal{B}(D_s \rightarrow \phi\pi)}{\mathcal{B}(D_s \rightarrow \tau\nu_\tau)} \times f_{D_s}^\tau \times \frac{\varepsilon_{\text{norm}}}{\varepsilon_{\text{sig}}} \times \frac{N_{\text{sig}}}{N_{\text{norm}}} = \alpha \times N_{\text{sig}}$$

- where ε stands for trigger, reconstruction, selection
- norm = normalisation channel $D_s \rightarrow \phi\pi$
- $f_{D_s}^\tau$ is the fraction of τ coming from D_s
i.e. $(83 \pm 3)\%$ for 2012



Normalisation in numbers I

	7 TeV	8 TeV
$\epsilon_{\text{sig}}^{\text{GEN}} (\%)$	8.989 ± 0.40	9.21 ± 0.35
$\epsilon_{\text{cal}}^{\text{GEN}} (\%)$	11.19 ± 0.34	11.53 ± 0.32
$\epsilon_{\text{sig}}^{\text{REC,isMuon,SEL}} (\%)$	9.927 ± 0.028	9.261 ± 0.023
$\epsilon_{\text{cal}}^{\text{REC,isMuon,SEL}} (\%)$	7.187 ± 0.022	6.690 ± 0.022
$\frac{c_{\text{cal}}^{\text{track}}}{c_{\text{sig}}^{\text{track}}}$	$0.997 \pm 0.009 \pm 0.026$	$0.996 \pm 0.009 \pm 0.026$
$\frac{c_{\text{cal}}^{\mu\text{ID}}}{c_{\text{sig}}^{\mu\text{ID}}}$	$0.9731 \pm 0.0031 \pm 0.0264$	$1.0071 \pm 0.0022 \pm 0.0204$
c^{ϕ}	0.98 ± 0.01	
c^{τ}	1.032 ± 0.006	1.026 ± 0.006
c^{trash}	1.89 ± 0.12	1.96 ± 0.12
$\epsilon_{\text{sig}}^{\text{TRIG}} (\%)$	$35.52 \pm 0.14 \pm 0.14$	$39.3 \pm 1.7 \pm 2.0$
$\epsilon_{\text{cal}}^{\text{TRIG}} (\%)$	$23.42 \pm 0.14 \pm 0.09$	$20.62 \pm 0.76 \pm 1.07$



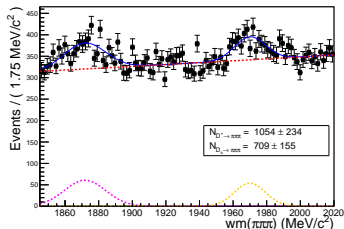
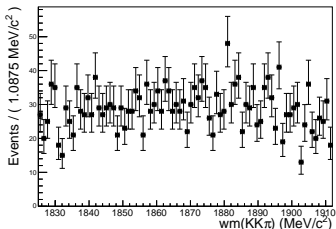
Normalisation in numbers II

	7 TeV	8 TeV
$\mathcal{B}(PD_s \rightarrow \phi\pi)$	$(1.317 \pm 0.099) \times 10^{-5}$	
$f_{D_s}^r$	0.78 ± 0.04	0.80 ± 0.03
$\mathcal{B}(D_s \rightarrow \tau\nu_\tau)$	0.0561 ± 0.0024	
$\epsilon_{\text{cal}}^{\text{REC\&SEL}} / \epsilon_{\text{sig}}^{\text{REC\&SEL}}$	0.898 ± 0.060	0.912 ± 0.054
$\epsilon_{\text{cal}}^{\text{TRIG}} / \epsilon_{\text{sig}}^{\text{TRIG}}$	0.6593 ± 0.0058	0.525 ± 0.040
N_{cal}	$28,207 \pm 440$	$52,131 \pm 695$
α	$(3.81 \pm 0.46) \times 10^{-9}$	$(1.72 \pm 0.23) \times 10^{-9}$
α^{trash}	$(7.20 \pm 0.98) \times 10^{-9}$	$(3.37 \pm 0.50) \times 10^{-9}$



Misidentification

- Most dominant: $D^+ \rightarrow K\pi\pi$
- Experience from last round: cut away low ProbNNmu range
- Check remaining data under $K\pi\pi$ hypothesis for D^+ peak
⇒ misid safely contained in “trash” bin
- $D^+ \rightarrow \pi\pi\pi$ and $D_s \rightarrow \pi\pi\pi$ start to become visible in 2012



Evil backgrounds

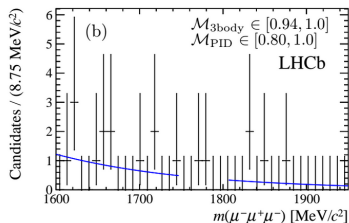
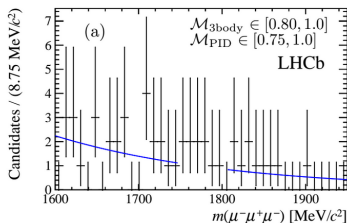
- $\phi \rightarrow \mu\mu + X$: narrow veto on dimuon mass
- $D_s \rightarrow \eta(\mu\mu\gamma)\mu\nu_\mu$: not so easy
 - Modelled in CONF note
 - Optimised veto in PAPER
 - Right now: both versions in the ANA note
- Baseline: veto $m_{\mu^+\mu^-} < 450$ MeV
 - Fits better understood
 - Sensitivity unchanged when removing veto
 - Smaller uncertainty on expected background



Remaining backgrounds

- Fit exponential to invariant mass spectrum in each likelihood bin
 - Don't use ± 30 MeV in the fit
- Compatible results blinding only ± 20 MeV¹

Example of most sensitive regions in 2011 and 2012

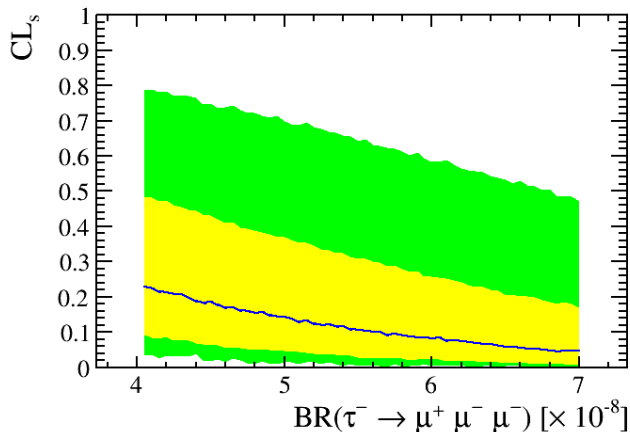


¹partially used in classifier development

- Consider nuisance parameters from background fit, signal pdf calibration, normalisation
- Nuisance parameters due to τ production, normalization.
- Limit for combined 2011+2012 analysis

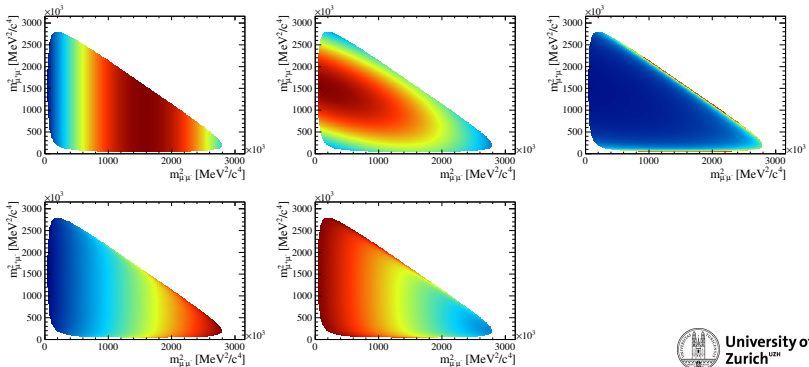


$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < 5.0 \times 10^{-8}$ at 90% CL



Model dependence

- η veto \Rightarrow our limit not applicable to New Physics with small $m_{\mu^+\mu^-}$
- Model description in arXiv:0707.0988
- 5 relevant Dalitz distributions: 2 four-point operators, 1 radiative operator, 2 interference terms



Model dependence

- η veto \Rightarrow our limit not applicable to New Physics with small $m_{\mu^+\mu^-}$
- Model description in arXiv:0707.0988
- 5 relevant Dalitz distributions: 2 four-point operators, 1 radiative operator, 2 interference terms
- With radiative distribution limit gets worse by 51 % (dominantly from the η veto)
- The other four Dalitz distributions behave nicely (within 7 %)



Unblinding 1

" THERE came a day at summer's full
Entirely for us
I thought that such were for the saints,
Where revelations be. " ^a

^aE.Dickinson

On Monday 4th of August we were given the
permission to unblind.

Unblinding 2

- Unfortunately no big "revelations" were there:
- 2011 numbers:

ProbNNmu	M_{blend}	Estimated	Observed
0.4, 0.45	0.28, 0.32	3.172 ± 0.661	4
0.4, 0.45	0.32, 0.46	9.242 ± 1.129	6
0.4, 0.45	0.46, 0.54	2.894 ± 0.632	6
0.4, 0.45	0.54, 0.65	3.173 ± 0.661	4
0.4, 0.45	0.65, 0.80	3.637 ± 0.716	2
0.4, 0.45	0.80, 1.0	3.787 ± 0.802	3
0.45, 0.54	0.28, 0.32	4.223 ± 0.779	6
0.45, 0.54	0.32, 0.46	8.345 ± 1.077	10
0.45, 0.54	0.46, 0.54	2.317 ± 0.568	4
0.45, 0.54	0.54, 0.65	2.828 ± 0.632	8
0.45, 0.54	0.65, 0.80	2.718 ± 0.688	5
0.45, 0.54	0.80, 1.00	4.825 ± 0.900	7

ProbNNmu	M_{blend}	Estimated	Observed
0.54, 0.63	0.28, 0.32	2.327 ± 0.584	6
0.54, 0.63	0.32, 0.46	8.324 ± 1.077	8
0.54, 0.63	0.46, 0.54	2.068 ± 0.534	1
0.54, 0.63	0.54, 0.65	3.291 ± 0.675	1
0.54, 0.63	0.65, 0.80	2.962 ± 0.646	4
0.54, 0.63	0.80, 1.00	3.114 ± 0.687	3
0.63, 0.75	0.28, 0.32	2.688 ± 0.616	1
0.63, 0.75	0.32, 0.46	7.541 ± 1.023	5
0.63, 0.75	0.46, 0.54	2.059 ± 0.534	3
0.63, 0.75	0.54, 0.65	1.996 ± 0.549	5
0.63, 0.75	0.65, 0.80	3.164 ± 0.661	2
0.63, 0.75	0.80, 1.00	4.674 ± 0.836	2
0.75, 1.0	0.28, 0.32	2.192 ± 0.551	2
0.75, 1.0	0.32, 0.46	3.384 ± 0.755	5
0.75, 1.0	0.46, 0.54	1.517 ± 0.457	3
0.75, 1.0	0.54, 0.65	1.280 ± 0.469	1
0.75, 1.0	0.65, 0.80	2.780 ± 0.645	1
0.75, 1.0	0.80, 1.00	4.421 ± 0.833	7



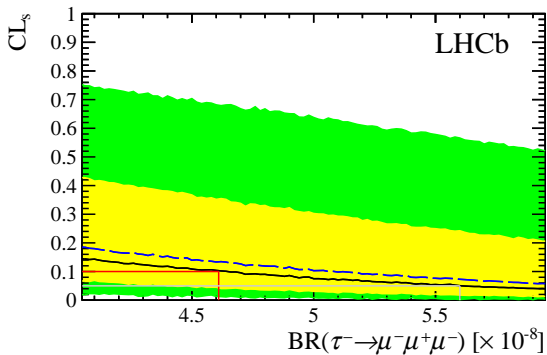
Unblinding 3

- Unfortunately no big "revelations" were also in 2012 data:

ProbNNmu	M_{blend}	Estimated	Observed
0.4, 0.54	0.26, 0.34	39.6 ± 2.3	39
0.4, 0.54	0.34, 0.45	32.2 ± 2.1	34
0.4, 0.54	0.45, 0.61	28.7 ± 2.0	28
0.4, 0.54	0.61, 0.7	9.72 ± 1.22	5
0.4, 0.54	0.7, 0.83	11.38 ± 1.26	7
0.4, 0.54	0.83, 0.94	7.34 ± 1.10	6
0.4, 0.54	0.94, 1.0001	5.98 ± 0.95	0
0.54, 0.61	0.26, 0.34	13.6 ± 1.37	8
0.54, 0.61	0.34, 0.45	12.1 ± 1.29	12
0.54, 0.61	0.45, 0.61	8.32 ± 1.086	13
0.54, 0.61	0.61, 0.7	2.595 ± 0.616	1
0.54, 0.61	0.7, 0.83	1.833 ± 0.601	5
0.54, 0.61	0.83, 0.94	2.929 ± 0.724	6
0.54, 0.61	0.94, 1.0001	2.693 ± 0.632	3

ProbNNmu	M_{blend}	Estimated	Observed
0.61, 0.71	0.26, 0.34	13.457 ± 1.366	7
0.61, 0.71	0.34, 0.45	10.852 ± 1.23	11
0.61, 0.71	0.45, 0.61	9.661 ± 1.18	12
0.61, 0.71	0.61, 0.7	3.346 ± 0.69	2
0.61, 0.71	0.7, 0.83	4.600 ± 0.888	5
0.61, 0.71	0.83, 0.94	4.091 ± 0.809	4
0.61, 0.71	0.94, 1.0001	2.780 ± 0.680	1
0.71, 0.8	0.26, 0.34	7.808 ± 1.067	6
0.71, 0.8	0.34, 0.45	7.001 ± 0.985	8
0.71, 0.8	0.45, 0.61	6.170 ± 0.945	6
0.71, 0.8	0.61, 0.7	1.570 ± 0.556	2
0.71, 0.8	0.7, 0.83	2.987 ± 0.717	0
0.71, 0.8	0.83, 0.94	3.929 ± 0.806	0
0.71, 0.8	0.94, 1.0001	3.222 ± 0.676	1
0.8, 1.0	0.26, 0.34	5.123 ± 0.861	3
0.8, 1.0	0.34, 0.45	4.435 ± 0.792	6
0.8, 1.0	0.45, 0.61	3.802 ± 0.784	5
0.8, 1.0	0.61, 0.7	2.649 ± 0.676	2
0.8, 1.0	0.7, 0.83	3.053 ± 0.674	2
0.8, 1.0	0.83, 0.94	1.740 ± 0.543	2
0.8, 1.0	0.94, 1.0001	3.361 ± 0.702	3

Unblinding 4



Limits(PHSP):

Observed(Expected)

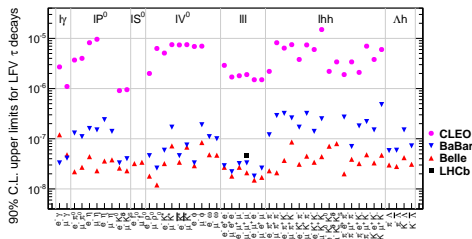
$4.6 (5.0) \times 10^{-8}$ at 90% CL

$5.6 (6.1) \times 10^{-8}$ at 95% CL

Dalitz distribution	$\times 10^{-8}$
$\rho_{V}^{(LL)(LL)}$	4.2 (4.7)
$\rho_{V}^{(LL)(RR)}$	4.1 (4.6)
$\rho_{V}^{(LR)}$	6.8 (7.6)
$\rho_{rad}^{(LL)(LL)}$	4.4 (5.1)
$\rho_{mix}^{(LL)(RR)}$	4.6 (5.0)
ρ_{mix}	

Conclusions

- We didn't find NP (yet).
- Limits set with full LHCb dataset.
- Awaiting for the future data!



- We would like to thank our referees for very friendly, thorough and fruitful review.
- With this presentation we ask collaboration for approval.