$\tau \to \mu\mu\mu$ approval presentation

Johannes Albrecht[1], Marta Calvi[2,3], Marcin Chrzaszcz[4,5],
Laura Gavardi[1], Jon Harrison[6], Basem Khanji[3],
George Lafferty[6], Tatiana Likhomanenko[7], Eduardo Rodrigues[6],
Nicola Serra[4], Paul Seyfert[7]

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

[1] Fakultät Physik, Technische Universität Dortmund, Germany
[2] The University of Milano Bicocca, Milano, Italy
[3] INFN Milano Bicocca, Milano, Italy
[4] Physik-Institut der Universität Zürich, Switzerland
[5] Henryk Niewodniczanski Institute of Nuclear Physics Polish Academy of Sciences, Cracow, Poland
[6] The University of Manchester, UK
[7] Yandex, Moscow, Russia
[8] Physikalisches Institut Heidelberg, Germany

University of Zurich[UZH]

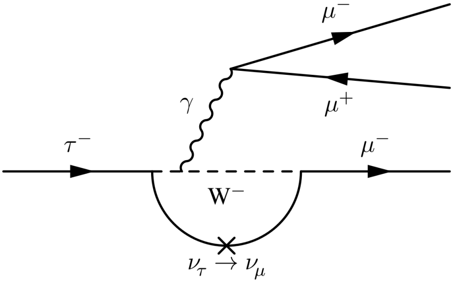
January 30, 2015

Outline

Yellow pages

- 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 $\tau \to \mu\mu\mu$



- 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...

University of Zurich[UZH]

Strategy

- Following same approach as other RD searches.
- Loose stripping selection.
- Multivariate classification in: mass, PID, "geometry/topology".
- Binning optimisation.
- Relative normalisation ($D_s \to \phi(\mu\mu)\pi$).
- 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 MeV $< |m - m_\tau| <$ 30 MeV).
- CLs for limit calculation.

University of Zurich[UZH]

- τ's in LHCb come from five main sources:

Mode	7 TeV	8 TeV
Prompt $D_s \to \tau$	$71.1 \pm 3.0\,\%$	$72.4 \pm 2.7\,\%$
Prompt $D^+ \to \tau$	$4.1 \pm 0.8\,\%$	$4.2 \pm 0.7\,\%$
Non-prompt $D_s \to \tau$	$9.0 \pm 2.0\,\%$	$8.5 \pm 1.7\,\%$
Non-prompt $D^+ \to \tau$	$0.18 \pm 0.04\,\%$	$0.17 \pm 0.04\,\%$
$X_b \to \tau$	$15.5 \pm 2.7\,\%$	$14.7 \pm 2.3\,\%$

$\mathcal{B}(D^+ \to \tau)$

- There is no measurement of $\mathcal{B}(D^+ \to \tau)$.
- One can calculate it from: $\mathcal{B}(D^+ \to \mu\nu_\mu)$ + helicity suppression + phase space.
- `hep-ex:0604043`.
- $\mathcal{B}(D^+ \to \tau\nu_\tau) = (1.0 \pm 0.1) \times 10^{-3}$.

Datasets

- Data from Reco14Stripping20(r1).
- Large MC samples:
  - 24M Inclusive background events ($b\overline{b}$ and $c\overline{c}$).
  - 10M Exclusive background events ($D_s \to \eta(\mu\mu\gamma)\mu\nu_\mu$).
  - 2M Signal events (split over 5 production channels).
  - 12M $D \to K\pi\pi$ (missID studies).
  - 10M $D^* \to D(K\mu\nu_\mu)\pi$ (missID studies).

$\Rightarrow$ Generator level cuts for improved use of computing resources.

  - $\sim$ 14 times more signal statistics after stripping.
  - $\sim$ 2 times more background statistics.

- Mix τ production on ntuple level instead of reweighting.
  $\Rightarrow$ Ease up ntuple usage (no forgotten weighting, no double weighting, ...).

Stripping and selection

	$\tau \to \mu\mu\mu$	$D_s \to \phi\pi$
$\mu^{\pm}$, $\pi^{\pm}$		
p_T	> 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	-
$\tau^{\pm}$ and D_s		
Δm	< 400MeV	< 50MeV
Vertex χ^2	< 15	
IP χ^2	< 225	
$\cos\alpha$	> 0.99	
$c\tau$ (stripping)	$> 100\ \mu m$	
	no PV refitting	
decay time (offline)	> -0.01 ns & < 0.025 ns	
	PV refitting	

University of Zurich[UZH]

Triggers

	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

Triggers in 2012

- Cuts changed through 2012.

$\rightarrow$ emulated two different TCKs for 2012.

$\rightarrow$ Found negligible differences in ratio of signal/normalisation channels.

- Choice of triggers was optimised based on $\frac{s}{\sqrt{b}}$ FOM.

University of Zurich[UZH]

Geometric likelihood

Much work has been put into 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 ($B_s^0 \to \mu\mu$ isolation $\to$ BDT isolation: CERN-THESIS-2013-259).
- Apply Blending technique.

Basic Setup - Step I

- Train 1/3 signal MC against 1/2 background MC.
- Input variables:
    - DOCA
    - Vertex χ^2
    - τ decay time
    - τ IPχ^2
    - min. μ IPχ^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.

University of Zurich[UZH]

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

Blending technique

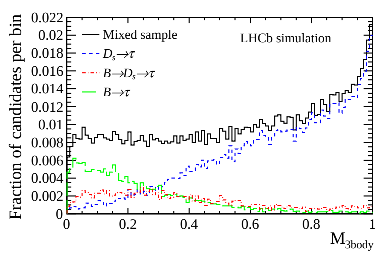
- For each signal channel we train: one BDT, three Fisher classifier, four MLPs, one FDA, one LD classifier and MatrixNet classifier.
- One final MatrixNet classifier using the 13 base variables and the base classifiers as input.

- All evaluation is done on 3*rd* 1/3 signal sample and middle side-bands.
- Splitting into independent samples makes the procedure insensitive to overtraining.

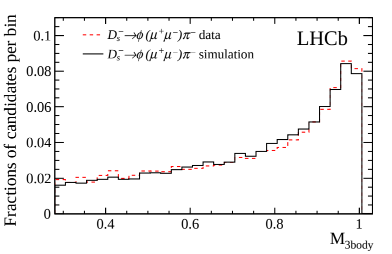
Performance of Blend classifier

- Classifier prefers τ's from prompt $\mathrm{D_s}$, the dominant channel.

MC response for different τ production channels



Response for $D_s \to \phi\pi$ data and MC

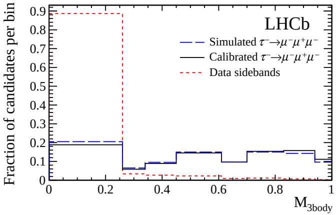


Calibration

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

validation

- done for 2011 analysis, treating smeared MC as data



- $D_s \to \phi\pi$ well modelled in MC.

University of Zurich[UZH]

PID

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

- We tested if PIDCalib samples (J/ψ) are suited for us.
- $D_s \rightarrow \phi\pi$ better representing $\tau \rightarrow 3\mu$ than J/$\psi \rightarrow \mu\mu$.
- Many thanks to Barbara Sciascia for help understanding the details: LINK

PID calibration

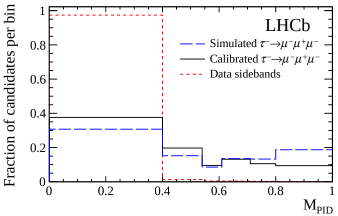
Phenomenological treatment

- correlations are small in $D_s \to \phi\pi$ 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 $\tau \to \mu\mu\mu$

- assign error of 0.02 for c.

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



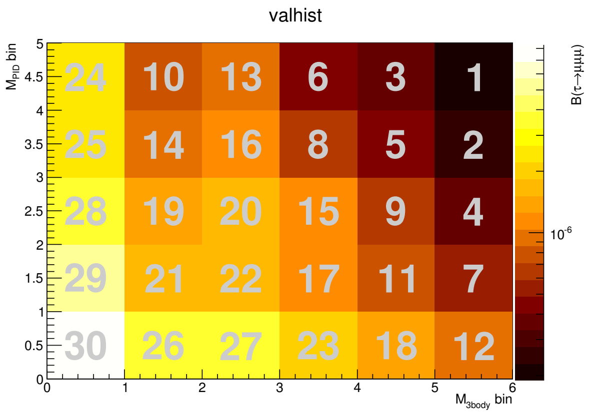
Binning optimisation

- 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 simultaneously).
- 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).

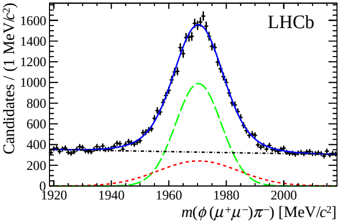
Bin sensitivity (2011 data)

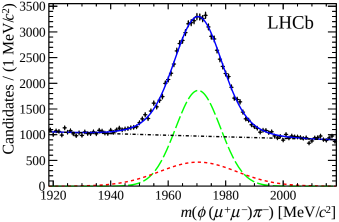


Mass shape

- Double-Gaussian with fixed fraction (70 % inner Gaussian).
- Fix fraction to ease calibration.
- Correct mass by MC:

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





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

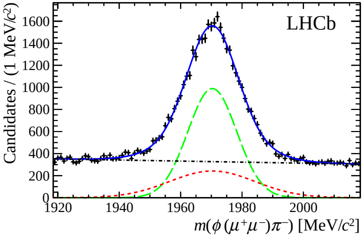
University of Zurich[UZH]

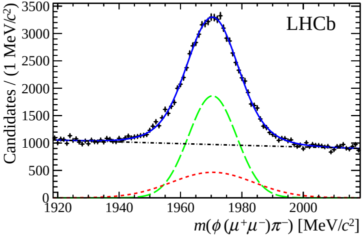
Relative normalisation

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

- where ε stands for trigger, reconstruction, selection,
- $f^{\tau}_{D_s}$ is the fraction of τ coming from D_s,
- norm = normalisation channel $D_s \to \phi\pi$

  i.e. (83 ± 3) % for 2012.





University of Zurich[UZH]

Normalisation in numbers I

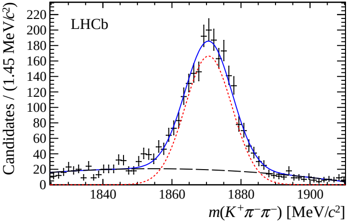
	7 TeV	8 TeV
${\epsilon_{\rm sig}}^{\rm GEN}(\%)$	8.989 ± 0.40	9.21 ± 0.35
${\epsilon_{\rm cal}}^{\rm GEN}(\%)$	11.19 ± 0.34	11.53 ± 0.32
${\epsilon_{\rm sig}}^{\rm REC,isMuon,SEL}(\%)$	9.927 ± 0.028	9.261 ± 0.023
${\epsilon_{\rm cal}}^{\rm REC,isMuon,SEL}(\%)$	7.187 ± 0.022	6.690 ± 0.022
$\frac{{c_{\rm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

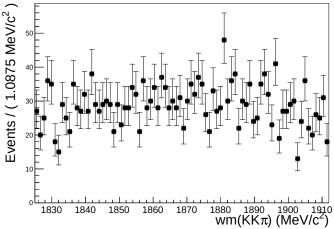
Normalisation in numbers II

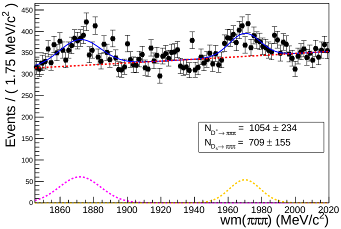
	7 TeV	8 TeV
$\mathcal{B}(D_s \to \phi\pi)$	$(1.317 \pm 0.099) \times 10^{-5}$	
$f_{D_s}^{\tau}$	0.78 ± 0.04	0.80 ± 0.03
$\mathcal{B}(D_s \to \tau\nu_\tau)$	0.0561 ± 0.0024	
${\epsilon_{\rm cal}}^{\rm REC\&SEL}/{\epsilon_{\rm sig}}^{\rm REC\&SEL}$	0.898 ± 0.060	0.912 ± 0.054
${\epsilon_{\rm cal}}^{\rm TRIG}/{\epsilon_{\rm sig}}^{\rm TRIG}$	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 1

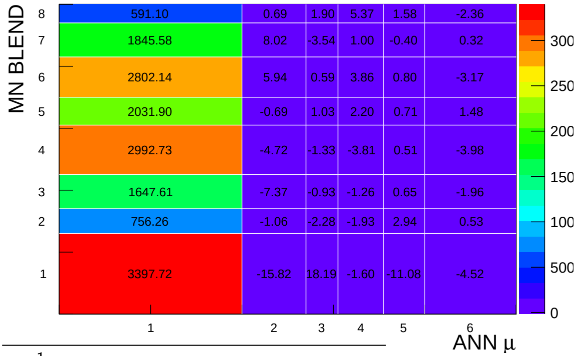
- Most dominant: $D^+ \to K\pi\pi$.
- Also seen $D^+ \to \pi\pi\pi$ and $D_s \to \pi\pi\pi$.
- Looked in all mass hypothesis combinations.







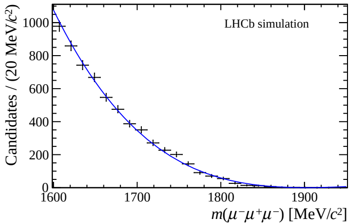
- Many tests were performed to be sure we are safe from $D_x \to 3h$.
- Tested both on MC and data.
- Referees also suggest looking into semileptonic decays.
- Our background is safely contained in "trash"[1] bins.

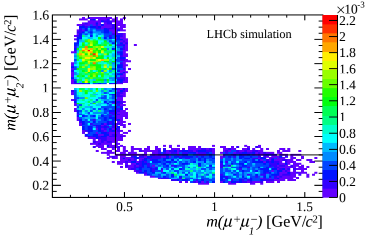


[1] Lowest *ProbNNmu* and M_{blend} bins, not taken for limit calculation.

Dangerous backgrounds

- $\phi \to \mu\mu + X$: narrow veto on dimuon mass.
- $D_s \to \eta(\mu\mu\gamma)\mu\nu_\mu$: not so easy:
  - Modelled in CONF note.
  - Optimised veto in PAPER.
  - 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.



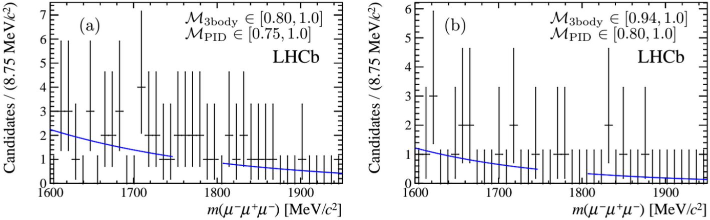


Remaining backgrounds

- Fit exponential to invariant mass spectrum in each likelihood bin.
- Don't use blinded region (± 30 MeV).

$\rightarrow$ Compatible results blinding only ± 20 MeV[2]

Example of most sensitive regions in 2011 and 2012

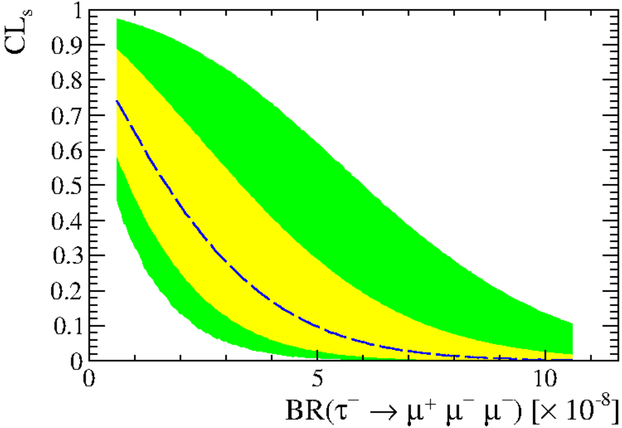


[2] partially used in classifier development

Expected limit

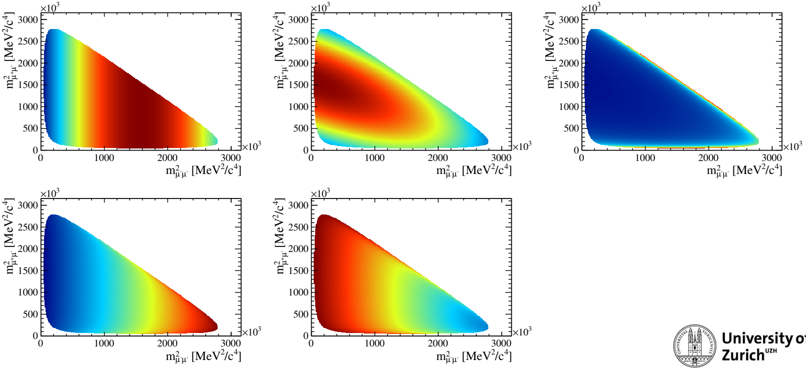
- Consider nuisance parameters from background fit, signal pdf calibration, normalisation.
- Nuisance parameters due to τ production, normalization.
- Limit for combined 2011+2012 analysis.

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



Model dependence

- η veto $\Rightarrow$ our limit not constraining 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 constraining 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 a factor of 1.5 (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]

[a] E.Dickinson

On Monday 4^{th} 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

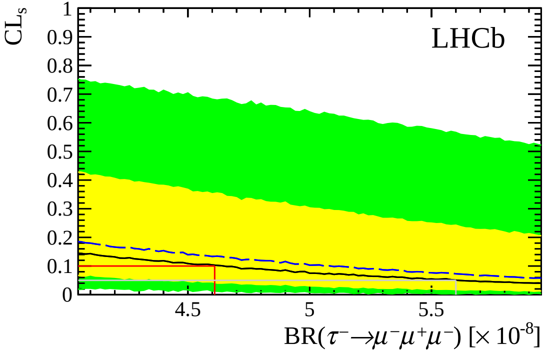
ProbNNmu	M_{blend}	Estimated	Observed
0.54, 0.63	0.28, 0.32	2.327 ± 0.584	6
0.54, 0.63	0.32, 0.46	8.324 ± 1.077	8
0.54, 0.63	0.46, 0.54	2.068 ± 0.534	1
0.54, 0.63	0.54, 0.65	3.291 ± 0.675	1
0.54, 0.63	0.65, 0.80	2.962 ± 0.646	4
0.54, 0.63	0.80, 1.00	3.114 ± 0.687	3
0.63, 0.75	0.28, 0.32	2.688 ± 0.616	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 either 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

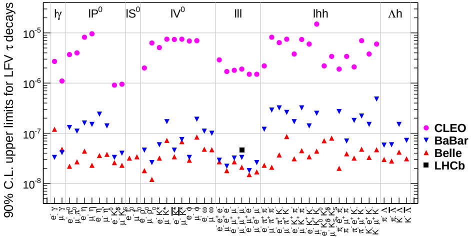


Limits(PHSP):
Observed(Expected)
$4.6\ (5.0) \times 10^{-8}$ at 90% CL
$5.6\ (6.1) \times 10^{-8}$ at 95% CL

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

Conclusions

- We didn't find NP (yet).
- Limits set with full LHCb dataset.
- We wait for the Run 2 dataset!



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

University of Zurich UZH