$B^- \to \mu^- \bar{\nu}_\mu$ decay at Belle/BelleII

Alexei Sibidanov

University of Victoria, BC
on behalf of the Belle/BelleII collaboration

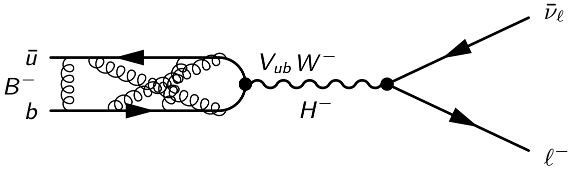
10th International Workshop on the CKM Unitarity Triangle
17-21 September 2018

Introduction

Branching fraction of the purely leptonic decay of the B-meson assuming massless neutrino:

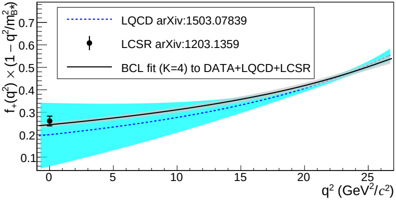
$$\mathcal{B}(B^- \to \ell^- \bar{\nu}_\ell) = \frac{G_F^2 m_B m_\ell^2}{8\pi}\left(1 - \frac{m_\ell^2}{m_B^2}\right)^2 f_B^2 |V_{ub}|^2 \tau_B,$$

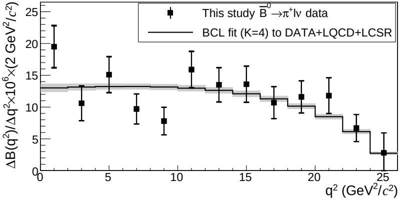
where G_F is the Fermi constant, m_B and m_ℓ are the masses of B-meson and resulting charged lepton correspondingly, f_B is the decay constant obtained from theory (LQCD), τ_B is the lifetime of the B-meson and V_{ub} is proportional to the coupling constant between u and b quarks.



This probes the Standard Model since this branching fraction can be modified by new physics, for example by a charged Higgs boson.

Introduction





Inputs

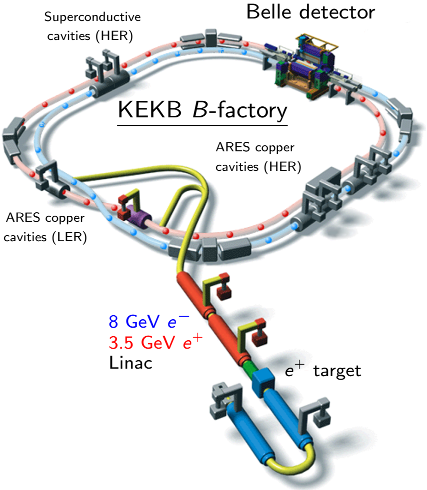
Value of $|V_{ub}| \times 10^3 = 3.736 \pm 0.142$ is from the exclusive $B \to \pi \ell \nu$ fit with the new LQCD input.
Value of $f_B = 185 \pm 3$ MeV is the recent result of HPQCD collaboration [arXiv:1212.0586].

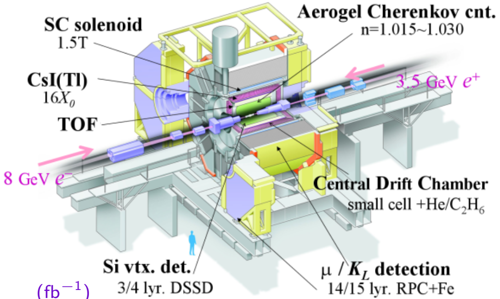
With those input parameters in absence of NP the following branching fractions and number of events in the full Belle/Belle2 data sets are expected:

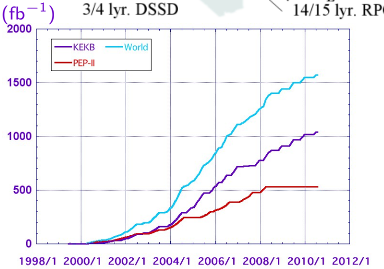
ℓ	$\mathcal{B}_{\mathrm{SM}}$	$N_{\mathrm{SM}}^{\mathrm{Belle}}$(711/fb)	$N_{\mathrm{SM}}^{\mathrm{Belle2}}$(50/ab)
τ	$(8.46 \pm 0.70) \times 10^{-5}$	67419 ± 5570	$(4.74 \pm 0.39) \times 10^6$
μ	$(3.80 \pm 0.31) \times 10^{-7}$	303 ± 25	21300 ± 1760
e	$(8.90 \pm 0.74) \times 10^{-12}$	0.0071 ± 0.0006	0.5 ± 0.04

$B^\pm \to \tau^\pm \nu_\tau$ process has been measured by Belle with hadronic and semileptonic tagging.
$B^\pm \to \mu^\pm \nu_\mu$ process is potentially measurable with the current Belle data set. Whereas not a single decay of $B^\pm \to e^\pm \nu_e$ is expected in the Belle2 data set. Further only results of the search of the $B^\pm \to \mu^\pm \nu_\mu$ decay are considered.

Belle detector and KEKB accelerator (1999-2010)

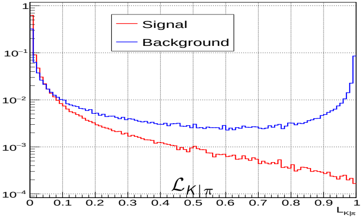






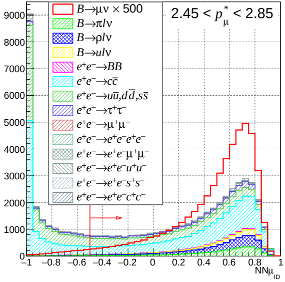
Energy region	$\Upsilon(1S)$	$\Upsilon(2S)$	$\Upsilon(3S)$	$\Upsilon(4S)$	$\Upsilon(5S)$	Off reson./scan	Total
$\mathcal{L}$, fb^{-1}	6	25	3	711	121	100	> 1000

- Signal muon is the highest momentum muon in an event.
- Standard muon selection $\mu_{\text{ID}} > 0.9$.
- Considerable number of kaons are accepted by the standard selection.



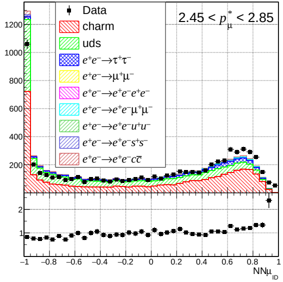
Neural network was built and trained to improve signal muon selection with information from the drift chamber and the calorimeter.

Signal muon and background



$\varepsilon_{\text{sig}} = 97\%$ and $\varepsilon_{\text{bkg}} = 67\%$

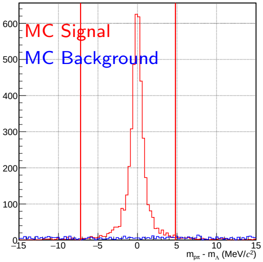
Data/MC off-resonance comparison



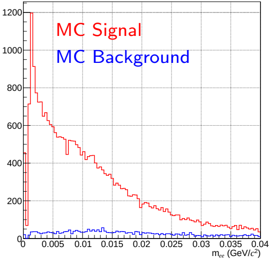
Signal selection – charged particle and photon selection

To build event kinematic variables it is important to properly assign particle species. The following steps were performed to select particles in an event:

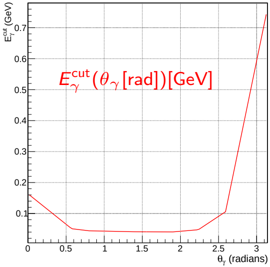
- Filter out low momentum tracks to get only those corresponding to a real particle.
- Select well reconstructed long-lived K_S and Λ particles as well as converted photons.
- The rest of charged particles classified by the following consecutive procedure: $\mu_{\text{ID}} > 0.6$ and $\mathcal{L}_{K|\pi} < 0.25$ is muon, $e_{\text{ID}} > 0.6$ and $\mathcal{L}_{K|\pi} < 0.25$ is electron, $\mathcal{L}_{K|p} < 0.9$ is proton, $\mathcal{L}_{K|\pi} > 0.6$ is kaon and the rest are pions.
- Energy of reconstructed photon has to be above $E_\gamma > E_\gamma^{\text{cut}}(\theta_\gamma)$, where the $E_\gamma^{\text{cut}}(\theta_\gamma)$ function reflects equal probability for photon to be noise or from real B decay (evaluated by MC study).
- K_L and lepton vetos.

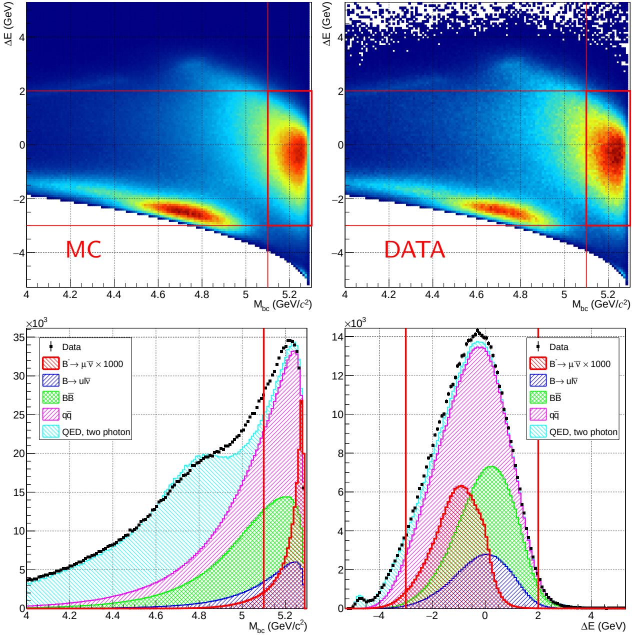


$\varepsilon_\Lambda = 74\%$ and $\varepsilon_{\text{fake}} = 0.09\%$



$\varepsilon_\gamma = 93\%$ and $\varepsilon_{\text{fake}} = 1.3\%$





At least one 2.2 GeV/c muon ($p^*_\mu \notin (2.45; 2.85)$ GeV/c) + selection on angle between signal muon momentum vector and the thrust axis of the rest of event:

$$\frac{\vec{n}_t \cdot \vec{p}_\mu}{|\vec{n}_t||\vec{p}_\mu|} > -0.8.$$

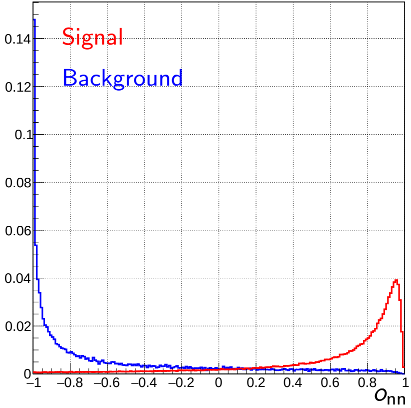
Signal is overlayed on the stack histograms and scaled 1000 times for visibility.

Projection histograms are shown with the other axis selection (shown as the red lines) applied.

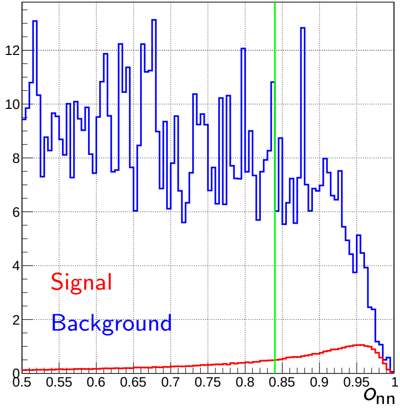
MC qualitatively and quantitatively matches the data. Challenge to see the signal!

Event classification by neural network

The signal-enhanced region $2.644\,\mathrm{GeV}/c < p^*_\mu < 2.812\,\mathrm{GeV}/c$



Both fractions normalized to unity



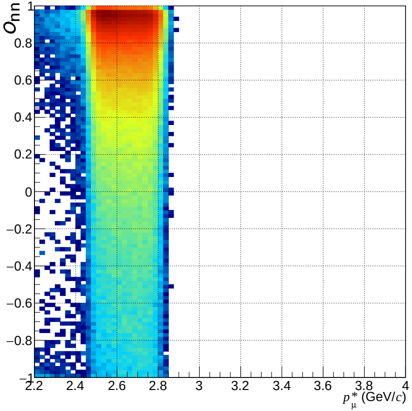
Real scale

Green line – best FoM selection: $\varepsilon_{\mathrm{sig}} \approx 44\%$, $\varepsilon_{\mathrm{bkg}} \approx 1.5\%$, $N_{\mathrm{sig}} \approx 23.4$ $N_{\mathrm{bkg}} \approx 169.5$,

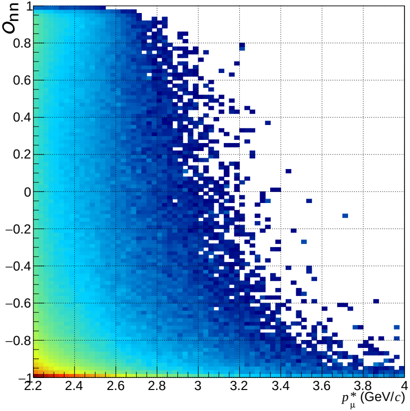
$$\mathrm{FoM} = N_{\mathrm{sig}}/\sqrt{N_{\mathrm{sig}} + N_{\mathrm{bkg}}} \approx 1.7$$

Two dimensional fit in the p^*_μ-o_{nn} plane to improve signal sensitivity.

Neural network output vs p^*_μ



Signal

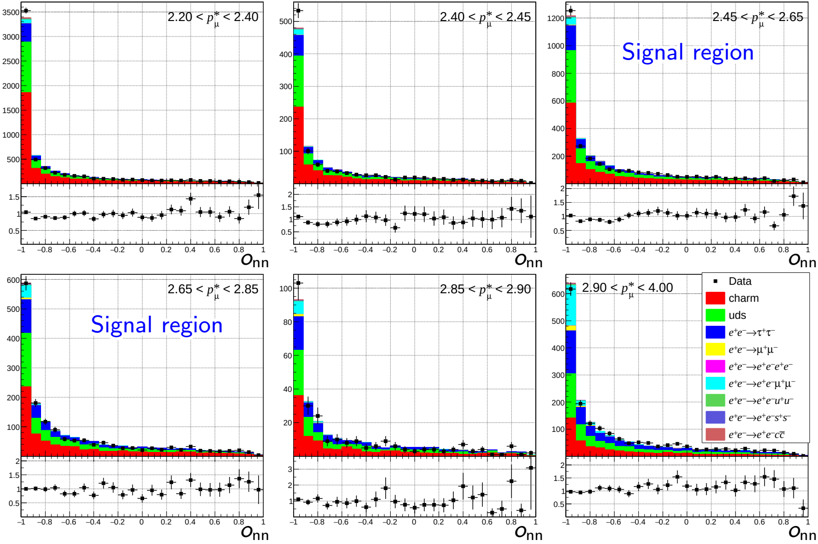


Background

The region $2.45\,\mathrm{GeV}/c < p^*_\mu < 2.85\,\mathrm{GeV}/c$ in data was blind during analysis.

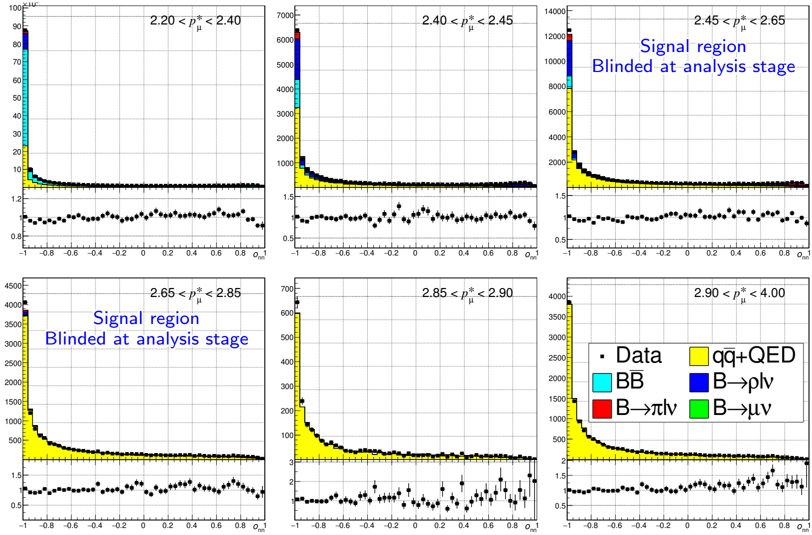
Signal extraction – fit to off-peak(continuum) data

Use MC template histograms for various fit components.
Projections of o_{nn} variable in muon momentum bins:



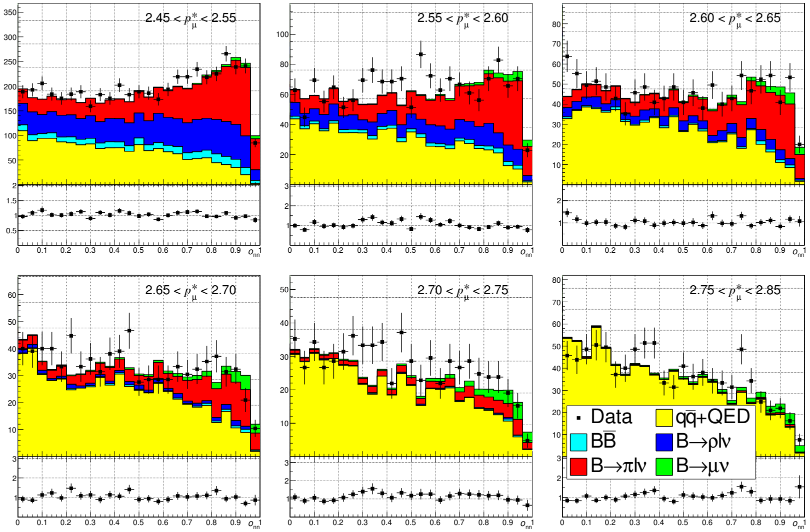
Fit to on-peak data

Projections of o_{nn} variable in muon momentum bins:

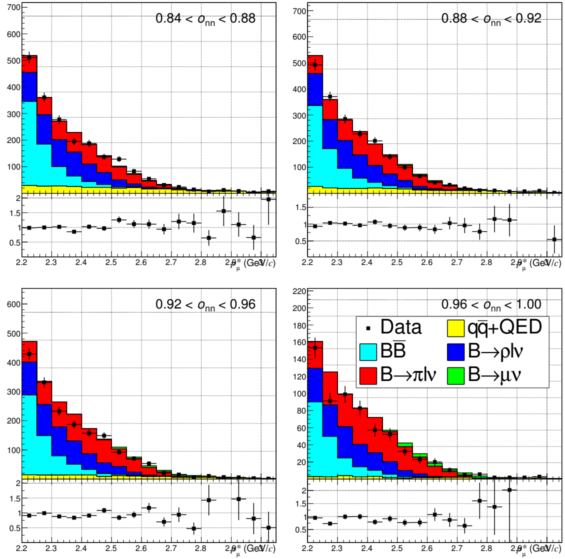


Fit to on-peak data (zoom to the signal region)

Projections of o_{nn} variable in muon momentum bins:



Fit to on-peak data (another projection)



Result is published in PRL**121**, 031801.

- Number of signal events is normalized to the number of $B \to \pi \ell \nu$ events.
- Fit extracts directly ratio $R = N_{B\to\mu\bar{\nu}_\mu} / N_{B\to\pi\ell\nu}$.

Fit result

$$R_{\text{fit}} / R_{\text{MC}} = 1.66 \pm 0.57$$
$$N_{B\to\mu\bar{\nu}_\mu} = 195 \pm 67$$

Branching fraction

$$\mathcal{B}(B \to \mu\bar{\nu}_\mu) = (6.46 \pm 2.22) \times 10^{-7}$$
$$= (6.46 \pm 2.22_{\text{stat}} \pm 1.6_{\text{syst}}) \times 10^{-7}$$

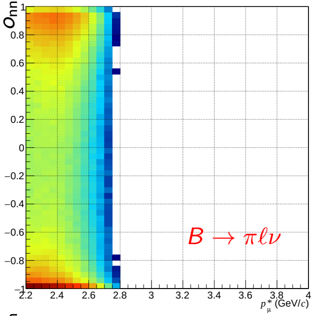
Significance

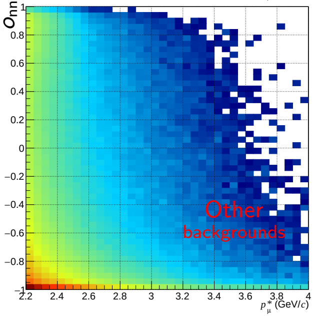
From likelihood ratio

$$\sqrt{2 \log \frac{\mathcal{L}(\mathcal{B}(B \to \mu\bar{\nu}_\mu) = 0)}{\mathcal{L}(\text{free } \mathcal{B}(B \to \mu\bar{\nu}_\mu))}} = 3.4$$

with systematics included drops to 2.4 σ

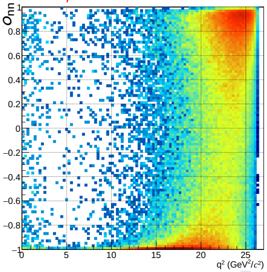
Peaking background from the $B \to \pi \ell \nu$ decay



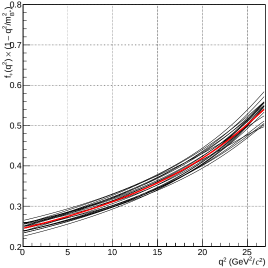


- The $B \to \pi \ell \nu$ decay with soft pion becomes kinematicaly indistinguishable from the signal decay in the untagged search and looks like the signal peak in the neural net output.
- Low momentum pion corresponds to high q^2 value where the form factor previously was poorly known. Now it is tightly constrained by the new LQCD calculations.
- The effect of this peaking background was studied in the sensitivity test with "toy" MC where the $B \to \pi \ell \nu$ template was varied according to form factor uncertainties with the new LQCD data and found to be small $\sim 0.9\%$.

$p^*_\mu > 2.452$ GeV/c



Form factor shapes



Summary of systematic uncertainties

Source	Estimation (%)
$\bar{B} \to \pi \ell^- \bar{\nu}_\ell$ form-factor	0.9
$B \to \rho \ell \nu$ form-factor	12
$B^- \to K_L^0 \pi^-$	5.5
$B^- \to \mu^- \bar{\nu}_\mu \gamma$	6
Continuum shape	15
Signal peak shape	11
Trigger	8
$\mathcal{B}(\bar{B} \to \pi \ell^- \bar{\nu}_\ell)$	3.4
Total (in quadrature)	24.6

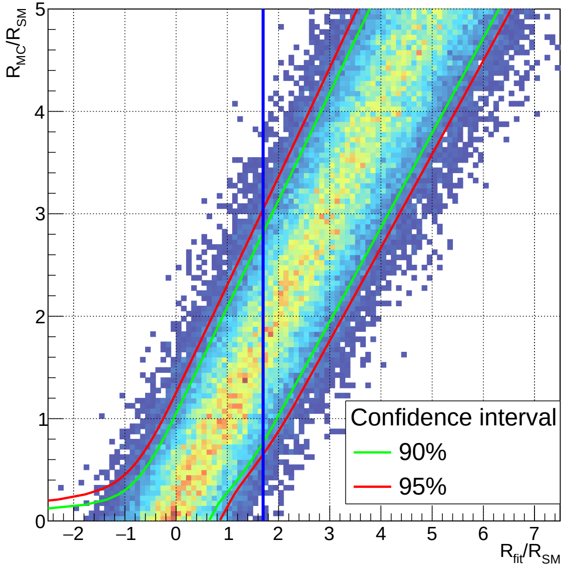
$B \to \rho \ell \nu$ form-factor

Several form factor calculations were employed in the fit the maximal deviation of 12 % as an estimation of the systematic uncertainty.

Trigger

The L4 trigger and HadronBJ skim selection efficiencies are emulated MC. The MC efficiency on the signal events is $\varepsilon = 0.8411 \pm 0.0003$. At the moment we estimate the systematic uncertainty as half of the inefficiency which is 8% since it has to partially cancels in the ratio.

Feldman-Cousins interpretation of the fit result



Using the result of the fit to data (the numbers of events of each type and their covariance matrix) 10^5 "toy" MC samples were generated to interpret the fit result in the Feldman-Cousins approach.

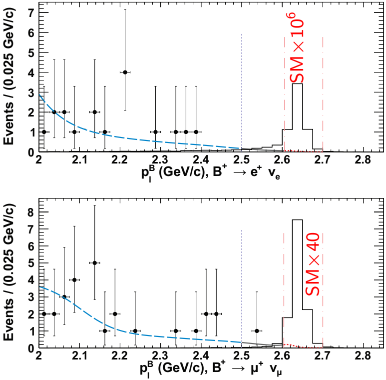
Confidence intervals of branching fraction of the $B^- \to \mu^- \bar{\nu}_\mu$ decay with systematic uncertainty included

C.L. (%)	$\mathcal{B}$ (10^{-7})
90	$[2.9, 10.7]$
95	$[2.5, 11.6]$
99	$[1.6, 13.3]$

Belle II prospects

At the moment it is difficult to give a quantitative estimation of Belle II sensitivity to the $B^{\pm} \to \mu^{\pm}\nu_{\mu}$ decay since actual Belle II detector performance is to be evaluated.

Belle hadronic tag – PRD**91**, 052016 (R) (2015)



Hadronic tag method

Ultimate technique for B decays with missing mass in the final state is full reconstruction of companion B in hadronic mode to infer energy and momentum of the other B and select signal B decays with virtually no background. Problem with the method is extremely low selection efficiency of $\sim 10^{-3}$ with anticipated signal yield of 21 events of the $B^{\pm} \to \mu^{\pm}\nu_{\mu}$ decay.

Untagged selection

- The untagged method still suffers from large irreducible background.
- Hadronic form factors for charmless semileptonic decays have to be measured to tightly constrain their shapes in background templates.
- Naively scaling results of this analysis, 5σ significance can be reached with 6/ab of Belle II data.
- With the full 50/ab Belle II data set about 5% statistical precision is expected.

Conclusion

- Full Belle data sample is analyzed to search for the $B^{\pm} \to \mu^{\pm}\nu_{\mu}$ decay.
- Multivariate classification procedure been developed for signal extraction.
- Measured $2.4\,\sigma$ signal excess corresponds to a branching fraction of $\mathcal{B}(B^{-} \to \mu^{-}\bar{\nu}_{\mu}) = (6.46 \pm 2.22_{\text{stat}} \pm 1.6_{\text{syst}}) \times 10^{-7}$ and consistent with the Standard Model prediction.
- The 90% confidence interval for the obtained branching fraction in the frequentist approach is $\mathcal{B}(B^{-} \to \mu^{-}\bar{\nu}_{\mu}) \in [2.9, 10.7] \times 10^{-7}$.
- The result published in Phys. Rev. Lett. **121**, 031801 (2018), [arXiv:1712.04123].
- Decay discovery is expected with several ab^{-1} from Belle II.

Backup slides

Data set

The following data sets have been used:
Signal MC:

- 2×10^6 of $B^\pm \to \mu^\pm \nu_\mu$.

Generic MC:

- 20 streams of $B \to u\ell\nu$ (main background from B decays).
- 10 streams of charged and mixed B-mesons.
- 6 streams of on-resonance continuum events.
- 6 streams of off-resonance continuum events.

Other backgrounds (not present in the previous Belle untagged analysis):

- $e^+e^- \to \tau^+\tau^-$ with $\mathcal{L} = 3286.120$ fb^{-1}.
- $e^+e^- \to \mu^+\mu^-(\gamma)$ with $\mathcal{L} = 2009.450$ fb^{-1}.
- $e^+e^- \to e^+e^-e^+e^-$ with $\mathcal{L} = 2033.140$ fb^{-1}.
- $e^+e^- \to e^+e^-\mu^+\mu^-$ with $\mathcal{L} = 489.479$ fb^{-1}.
- $e^+e^- \to e^+e^-u\bar{u}$ with $\mathcal{L} = 544.415$ fb^{-1}.
- $e^+e^- \to e^+e^-s\bar{s}$ with $\mathcal{L} = 481.287$ fb^{-1}.
- $e^+e^- \to e^+e^-c\bar{c}$ with $\mathcal{L} = 263.950$ fb^{-1}.

Data:

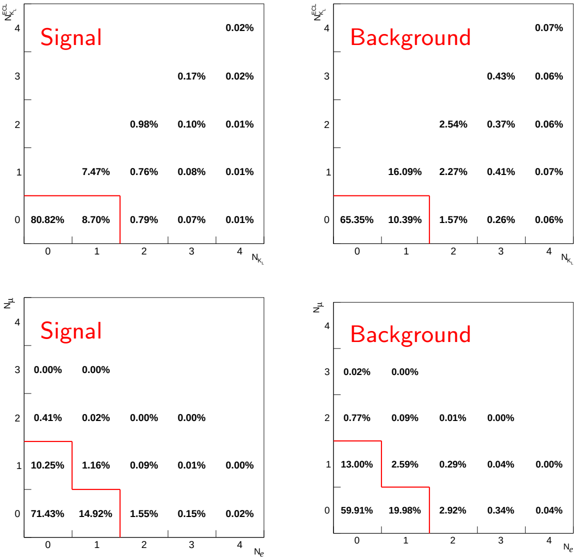
- 702.623 fb^{-1} on-resonance data.
- 79.366 fb^{-1} off-resonance data.

MC and data passed ℓnu skim procedure.

ℓnu skim

- $|\Delta r| < 0.5$ cm and $|\Delta Z| < 2$ cm.
- $e_{\text{ID}} > 0.5$ or $\mu_{\text{ID}} > 0.9$.
- $p^*_\ell > 2.2$ GeV/c.

Signal selection – K_L and lepton veto



K_L veto

It is more probable to reconstruct K_L mesons in a background event than in signal one, since Belle cannot measure the K_L energy and it can therefore mimic missing energy from the signal neutrino. Efficiency correction for K_L is applied.

Lepton veto

Excess of reconstructed charged leptons is a signature of B and D semileptonic decays. Since hadrons can be misidentified as leptons one electron or muon is allowed.

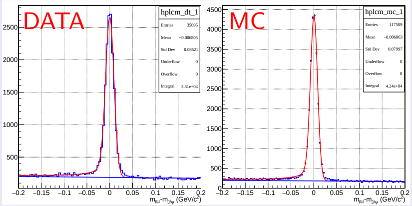
Event classification by neural network (perceptron)

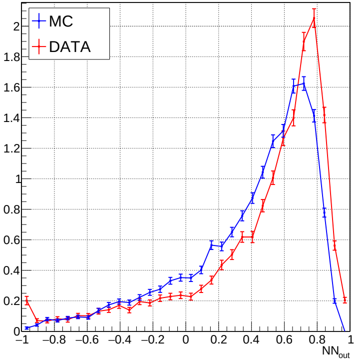
For this analysis a high performance perceptron with back propagation has been implemented from scratch (10^3 times faster than the ROOT implementation).
The best input configuration among 29 tested is:

- $R_1^{\mu o}/R_0^{\mu o}$, $R_1^{\mu o}/R_0^{\mu o}$, $R_1^{\mu o}/R_0^{\mu o}$ – where $R_i^{\mu o} = \sum_j |\vec{p}_\mu||\vec{p}_j|P_i(\cos\theta_{\mu j})$, p_j is in the cm frame, $P_i(x)$ is the i^{th} Legendre polynomial.
- R_1^{oo}/R_0^{oo} – where $R_i^{oo} = \sum_k \sum_j |\vec{p}_k||\vec{p}_j|P_i(\cos\theta_{kj})$, $p_{k,j}$ is in the cm frame
- $R_1^{\text{KFW}} = \sum_k \sum_{j>k} |\vec{p}_k||\vec{p}_j|P_i(\cos\theta_{kj})$, $p_{k,j}$ is in the cm frame
- $\cos(\theta_{\text{miss}})$ – angle of missing momentum in the cm frame
- $\sqrt{\sqrt{\sqrt{\Delta Z^2}}}$ – distance between reconstructed z-coordinates of muon and tag
- $\frac{\vec{n}_\text{t} \cdot \vec{p}_\mu}{|\vec{n}_\text{t}||\vec{p}_\mu|}$ – angle between thrust and muon momenta in the cm frame
- $s = 1 - \vec{n}_\text{t}^2$ – sphericity
- ΔE
- $\frac{\vec{n}_\text{t}^{\text{ECL}} \cdot \vec{p}_\mu}{|\vec{n}_\text{t}^{\text{ECL}}||\vec{p}_\mu|}$ – $\vec{n}_\text{t}^{\text{ECL}}$ is based only on calorimeter information
- $(q_\mu + q_{\text{tag}}) \times q_\mu$ – charge balance
- $\frac{\vec{p}_\mu \cdot \vec{p}_{B_{\text{tag}}}}{|\vec{p}_\mu||\vec{p}_{B_{\text{tag}}}|}$ – angle between muon and tag momenta in the cm frame
- $\cos\theta_\mu$ – muon angle in the cm frame

Validation of $NN_{\mu\text{ID}}$ with $J/\psi \to \mu\mu$ sample

2.2 GeV/c < $|\vec{p}_{\mu}^{\,*}|$ < 2.25 GeV/c





$J/\psi \to \mu\mu$ sample selection

- At least one muon has p_{μ}^{*} >2.2 GeV/c
- Two charged tracks with opposite charges and $\mu_{\text{ID}} > 0.9$, $|\Delta r| < 0.5$ cm and $|\Delta Z| < 2$ cm.
- $|m_{\mu\mu} - m_{J/\psi}| < 0.2$ GeV/c^2
- The highest momentum muon 2.2 GeV/c^2 < p_{μ}^{*} < 4 GeV/c^2.

Efficiency of $NN_{\mu\text{ID}}$

Checked in two momentum regions:

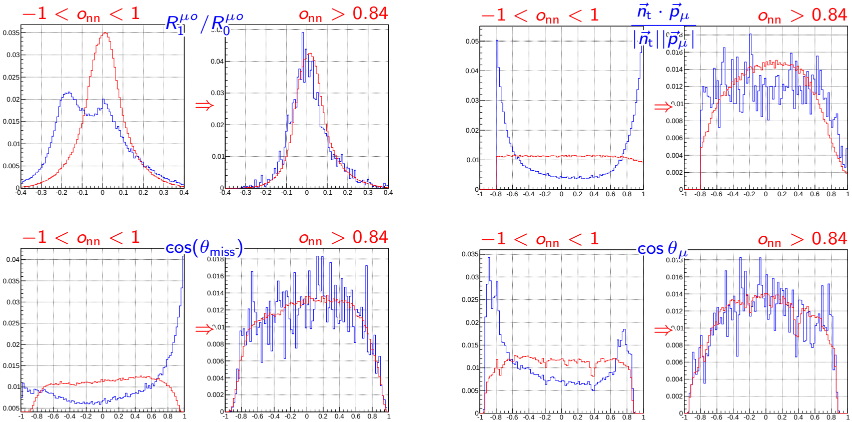
	p_{μ}^{*} (GeV/c)
Fit range	[2.20, 4.0]
Signal range	[2.48, 2.8]

Despite MC/DATA shapes are not fully matched the selection $NN_{\mu\text{ID}} > -0.5$ provides very good agreement in efficiencies:

	Fit range $1-\varepsilon$ (%)	Signal range $1-\varepsilon$ (%)
Data	6.15 ± 0.12	5.23 ± 0.30
MC	4.86 ± 0.09	3.90 ± 0.19
Δ	1.29 ± 0.15	1.33 ± 0.36

Neural network training results

Perceptron configuration is $N_{\text{in}} = 14$, $N_1^{\text{hidden}} = 56$, $N_2^{\text{hidden}} = 28$, $N_{\text{out}} = 1$, the activation function is tanh, in total there are 2456 weights. The training sample contains about 3.9×10^5 signal events and 1.55×10^6 background events. Test sample with the same number of events is used to validate the learning result. Ideally, in the limit $o_{\text{nn}} \to 1$, for a well trained neural network, input variable distributions for signal and background events should be the same.



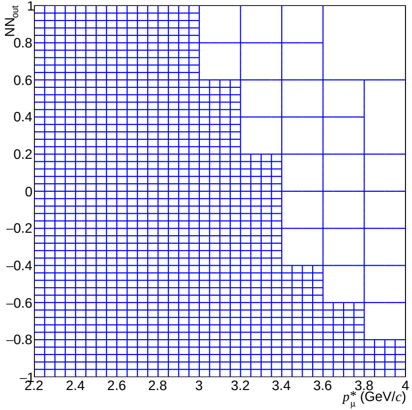
The training procedure shows satisfactory results close to what is expected.

Signal extraction

To extract the signal yield a binned maximum-likelihood fit was performed in the p^*_μ-NN_{out} plane. The p^*_μ-NN_{out} histogram size is 36×50 bins.

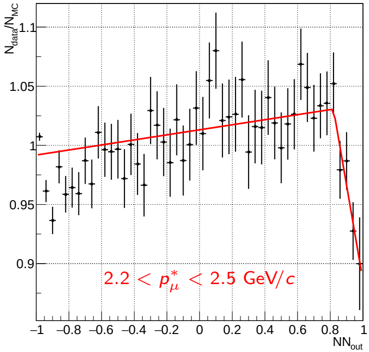
To avoid bins with 0 or a few events in a bin, low-populated bins were merged in the histogram as shown, resulting in a total of 1226 bins.

Background components with a predicted fraction $\leq 1\%$ were fixed in the fit to MC prediction.

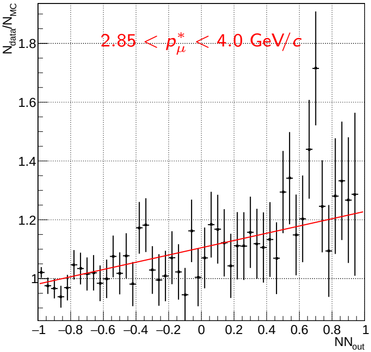


Shape uncertainty estimation

To estimate shape uncertainty the fit sideband residuals were parameterized and applied as shape corrections independenty of muon momentum for corresponding components.



Applied to $B \to \mu\bar{\nu}_\mu$, $B \to \pi\ell\nu$ and $B \to \rho\ell\nu$ MC templates peaking at $o_{nn} \sim 1$.
+11% in signal yield.



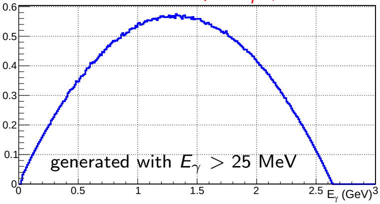
Applied to continuum MC templates.

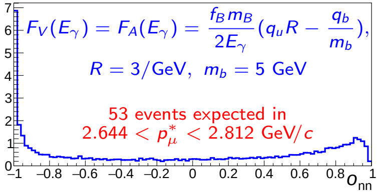
-15% in signal yield.

Both corrections give -2% difference in the signal yield $\Rightarrow$ use conservative approach and estimate the shape uncertainty as a sum of the differences in quadrature.

Other backgrounds

$B^- \to \mu^- \bar{\nu}_\mu \gamma$

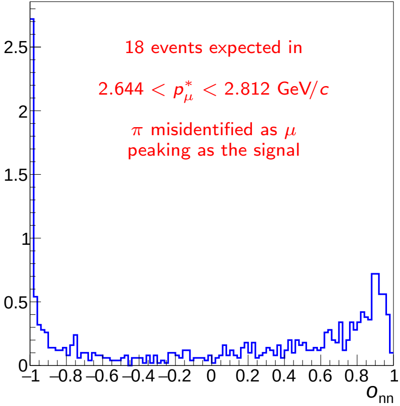




Decay form factor [Phys.Rev.D**61**,114510(2000)]. Fixed to zero in the fit. Uncertainty estimated as a difference in the signal yield between half of the best upper limit $\mathcal{B}(B^- \to \mu^- \bar{\nu}_\mu \gamma) < 3.4 \times 10^{-6}$ at 90% C.L. by Belle [Phys.Rev.D **91**, 112009(2015)] and zero.

6% uncertainty

$B^- \to K_L^0 \pi^-$



Fixed to MC prediction in the fit. Uncertainty estimated as a difference in the signal yield with/out $B^- \to K_L^0 \pi^-$ process since efficiency for K_L is not well modelled.

5.5% uncertainty