Study of the Rare B-meson Decays with the ATLAS Experiment

Marek Biros
Institute of Particle and Nuclear Physics, Faculty of Mathematics and Physics, Charles University, V Holeovické 2, 180 00 Prague 8, CZ
E-mail: marek.biros@cern.ch

Abstract. The searching for New Physics beyond the Standard Model of particles is one of the main aims of the current particle experiments. Rare b-hadron decays, occurring at the first level through loop Feynman diagrams, are strongly sensitive to the possible existence of new particles. Recent analyses from the LHCb experiment, measuring decay angles distributions in rare B-meson decay into a kaon \(K^0\) and two muons, hints a deviation from the Standard Model prediction. The ATLAS experiment works on a similar analysis of the decay \(B_d^0 \rightarrow K^\mu^\mu\) in order to confirm this deviation. The new particles’ existence could be observed in the deviations from the predicted decay’s angular distribution. This work provides input for the future analysis of this rare decay in the data-taking period of Run 2. To be more specific, observing of the fine deviations requires the precise description of the measured data and the Standard Model’s predictions. The improvement to the Run 1 analysis was achieved by an expansion of the fit function in a series of spherical harmonic functions. Next, the effect of the remaining background events in the signal dataset was repressed by adding masses \(m_B\) and \(m_{K^\mu}\) to the final fit. That required to know functions’ prescriptions of the \(m_B\) and \(m_{K^\mu}\) distributions for the signal decay and the expected background hence looking for their prescriptions was a part of this work too. The invariant \(B\) mass was described by the Gauss distribution with a per-event error. The per-event fit has to take into account the so-called Punzi effect in some cases which was also discussed here.

1. Introduction
The rare decay \(B^0 \rightarrow K^{*0}\mu^\mu^\mu\) is strongly sensitive to physics beyond the Standard Model (SM) of particles [3]. Recent analyses from the LHCb experiment, measuring decay angles distributions in rare B-meson decay into a kaon \(K^0\) and two muons [1], hints a deviation from the Standard Model prediction. The ATLAS experiment works on a similar analysis [2] in order to confirm this deviation. This process is described by loop Feynman diagrams, therefore its probability is very low (branching ratio (BR) is of the order of \(10^{-6}\)). New particles (undescribed by the SM) could flow through the loops and be indirectly measured in the angular distribution of the decay. The most important thing to a measurement of the fine discrepancies between the SM prediction and the real data is the finding of the precise description of the decay’s angular distribution. However this decay channel is very rare, therefore to have the precise measurement, it is extremely important to have the efficient signal from background separation.
1.1. Decay
The decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ is almost immediately followed by the decay of $K^{*0}$ into kaon and pion. To be more specific, a decay into $K^+, \pi^-$ occurs in the $2/3$ of cases and $K^0, \pi^0$ birth in the $1/3$ of cases. Neutral hadrons cannot be detected by the inner tracker of the ATLAS detector [4] which makes impossible the event reconstruction and reduces the total BR $= 1.06 \times 10^{-6}$ to the theoretically detectable BR $= 0.71 \times 10^{-6}$.

![Feynman diagrams](image)

**Figure 1.** Feynman diagrams of the process on the quark level. $b$-quark decays into $s$-quark through loops. The so-called penguin diagram is on the left and the box diagram is on the right.

1.2. Detection problems
There are 4 particles in the final state: two muons and two hadrons with opposite charge. Reconstruction of this type of event is a challenge for the ATLAS detector because the ATLAS does not have sufficient particle identification ability for our range of transversal momenta. The detector is able to distinguish only muons for sure but in the hadrons case, it measures only momentum and sign of charge of two hadrons. The name tagging (if the positive hadron is kaon or pion and the negative vice versa) is carried out after a proper reconstruction of the decay and application of criteria on the reconstructed invariant mass of the $K^*$. The misidentification (exchange of $K^+ \pi^-$ and $K^- \pi^+$) probability is according to the Monte Carlo simulations around $10\%$. This could lead to a reverse angular distribution, because a pair $K^- \pi^+$ indicates an anti-decay $\bar{B}^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-$. Also, protons could be misinterpreted by this type of reconstruction. Therefore decays including 2 hadrons and 2 muons in the final state seem as the signal decay for the detector: $B_s \rightarrow \phi (K^+ K^-) \mu^+ \mu^-$, $\Lambda_b \rightarrow \Lambda(1520) (p K^-) \mu^+ \mu^-$ and $\Lambda_b \rightarrow p \bar{K}^- \mu^+ \mu^-$. The last but not the least type the omnipresence background is the combinatoric background in which case the detector triggers to two muons and two hadrons, that did not birth from one particle decay but from two decays very close to each other. This type could be reached from data out of the range of interest and fitted by polynomials so it is not discussed in the next text.

1.3. The Signal from background separation
The main part of the separation was done using the cut-based selection criteria but there are still some background events remaining which brings inaccuracies into the analysis. The final fit of the angular distribution could be improved by adding masses of $B$ and $K^*$ into a multidimensional maximal likelihood function. If shapes of their distributions were known, fixed values of $m_{K^*}$ and $m_B$ in the fit function help the maximal likelihood estimation method to weigh each event according to the probability of the signal event.

In this work, the used data sets for signal and background were generated by full Monte Carlo (MC) simulations of the decays of interest in the ATLAS experiment. The background events including misidentification of $K^+ \pi^-$ were separated into two subdatasets by the cause: misidentified events when the true reconstructed $m_{K^*}$ lied out of the requested interval and when the fake $m_{K^*}$ was closer to the nominal $K^*$ mass than the true $m_{K^*}$. 

2
2. Multidimensional Maximal Likelihood Function

An unbinned Maximal Likelihood Estimation method is used (works event by event). The unbinned methods carried out better results for small data sets as it is in our case. Data is treated as a sample of independent random variables \( \{X_i\}_{i=1}^n \) with distribution described by an unknown probability density functions (p.d.f.) \(-f\). The function prescription is unknown, but sometimes we could predict it from the theory or from control datasets (e.g. mass sidebands).

Free parameters of the function \( \vec{a} = (\vec{a}_{m_{K^*}}, \vec{a}_{m_B}, \vec{a}_{\theta_K}, \vec{a}_{\theta_L}, \vec{a}_{\phi}) \) are estimated by maximization of the likelihood function:

\[
L(m_B, m_{K^*}, \cos \theta_K, \cos \theta_L, \phi, \vec{a}) = \prod_{i=1}^{n} f(m_{Bi}, m_{K^*i}, \cos \theta_{Ki}, \cos \theta_{Li}, \phi_i, \vec{a}).
\]

If we predicted no correlations between the masses of \( B \) and \( K \) and the angular distribution, then we can look for descriptions of each distribution separately and write \( f \) as:

\[
f(m_B, m_{K^*}, \cos \theta_K, \cos \theta_L, \phi, \vec{a}) = f_{m_{K^*}}(m_{K^*}, \vec{a}_{m_{K^*}}) f_{m_B}(m_B, \vec{a}_{m_B}) f_{\theta_K}(\cos \theta_K, \phi, \vec{a}_{\theta_K}, \vec{a}_{\phi}).
\]

The fit quality could be measured by the value of \( \chi^2/NDF \), where NDF denotes a number of degrees of freedom) and \( \chi^2 \) is defined as (\( O_i \) stands for a number of observables with the same value indexed by \( i \), \( p_i \) is the predicted probability of \( i \)-th observable value population according to the fit function, \( N \) is number of different values of the observable – usually number of bins in a histogram and \( n \) is size of the whole data set):

\[
\chi^2 = \sum_{i=1}^{N} \frac{(O_i - np_i)^2}{np_i},
\]

An ideal description would have value \( \chi^2/NDF = 1 \), a bad fit is connected with \( \chi^2/NDF >> 1 \) and \( \chi^2/NDF < 1 \) means that distribution is so well described by the function, that function started describing random fluctuations too (this is the result of the greater number of free parameters in the fit than is needed). In this work, we defined a good fit as \( \chi^2/NDF \in [1, 2] \).  

The next section is devoted to the determination of the p.d.f. prescriptions for \( m_B \) and \( m_{K^*} \) for all datasets.

3. Mass distributions

Invariant masses of particles are generated according to the Breit-Wigner distribution (BW). This physical shape of the mass generation is combined with the detector resolution through the convolution into the final signal distribution which is measured. The detector resolution is described by a Gauss distribution (Gaussian) for one concrete direction of the particle track, but different directions are described by the Gaussians with different parameters. The total detector resolution could be thus described by the sum of these Gauss distributions for all directions – multi-Gaussian. However, an effective enough description may already be accomplished with few-Gaussians function.

In the case of background, the description is more complicated. The peaking backgrounds originating in the \( B_s \) and \( \Lambda_b \) decays have in addition the effect of sculpting due to wrong track-mass hypothesis.

1 The more standard way is using the stricter criterion \( p(\chi^2, NDF) > 0.05 \), but our freer definition allows us to use simpler fit functions. We need keep on mind the problem of the \( B_d \to K^* \mu \mu \) analysis, the small number of signal events for which the functions with many free parameters would not work.
3.1. Distribution of the invariant mass of $B_d^0$

As was mentioned, the signal distribution of invariant $B$-meson’s mass is theoretically described by the Breit-Wigner convoluted with the multi-gaussian. The half-width of the invariant mass is 0.15 MeV for $B$-meson, which is absolutely negligible in comparison to the detector’s resolution. So the BW practically behaves like a peak of a delta function in 5279.62 MeV. The convolution of the delta function and any distribution is again the same distribution, therefore the convolution of the delta function and the multi-Gaussian is again the multi-Gaussian. A good fit is accomplished already with a double-Gaussian function:

$$G_2(m_B) = \frac{f}{N_1} G(m_B, \sigma_1, \mu_1) + \frac{1-f}{N_2} G(m_B, \sigma_2, \mu_2).$$

There is also one more way to take into consideration different parameters of the Gaussian to the different directions of departing particles. The invariant mass error $\sigma_{m_B}$ could be computed for every event directly from the dataset by using propagation of the uncertainty of the reconstructed tracks and could be added (scaled by free parameter) to the Gaussian description. Formally, we provide the substitution $\sigma \rightarrow s\sigma_{m_B}$. The Gaussian with the different error for each event is called a per-event Gaussian:

$$G(m_B, \sigma_{m_B}) = \frac{1}{s\sigma_{m_B}\sqrt{2\pi}} exp \left( -\frac{1}{2} \frac{(m_B - \mu)^2}{s\sigma_{m_B}^2} \right).$$

Due to simplifications present in the per-event mass error calculations, it is more suitable to use double per-event-Gaussian with two different scale parameters $s_1$ and $s_2$:

$$G_2(m_B, \sigma_{m_B}) = \frac{f}{N_1} G(m_B, \sigma_{m_B}, s_1, \mu_1) + \frac{1-f}{N_2} G(m_B, \sigma_{m_B}, s_2, \mu_2).$$

The double per-event Gaussian was used for descriptions of the signal decay and it turned out that is suitable for backgrounds with misidentified kaon and pion too. In the Fig. 2 there are shown graphs of the signal described by the per-event Gaussian, the normal double Gaussian and the double per-event Gaussian for the comparison.

The other background channels were described effectively: $B_s \rightarrow \phi(K^+K^-)\mu^+\mu^-$ by the sum of gaussian and Crystal-Ball, $\Lambda_b \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b \rightarrow \Lambda(1520)(pK^-)\mu^+\mu^-$ by the sum of the gaussian and reversed Landau (substitution of $m_B \rightarrow -m_B$ and $\mu \rightarrow -\mu$).

3.2. Distribution of the invariant mass of $K^*$

The invariant mass of $K^*$ is also generated according to the Breit-Wigner distribution but the half-width is not negligible compared to the width of the BW. The effects of the BW suppress higher orders of the multi-gaussian expansion and the signal $K^*$ invariant mass would be properly described by the convolution of the Breit-Wigner and the single Gauss distribution.

The shape of the mass of the background events with misidentified kaon and pion is deformed hence the best fit results were provided by effective formulae: $K^*$ masses outside of the requested interval and $K^*$ inside – linear function, $K^*$ masses nearer to mean than $K^*$ – the sum of two gaussians.

The background channels for other processes do not have mass interpreted as $m_{K^*}$ generated by any physical formulae either. After many combinations, we found the following functions: the sum of two exponentials for description of the decay $B_s \rightarrow \phi(K^+K^-)\mu^+\mu^-$, the sum of a linear and an exponential function for $\Lambda_b \rightarrow \Lambda(1520)(pK^-)\mu^+\mu^-$ and the sum of a quadratic function and a gaussian in the case of $\Lambda_b \rightarrow pK^-\mu^+\mu^-$. 
Figure 2. The signal data fit example using different functions: from the left – single per-event gaussian, normal double gaussian and double per-event gaussian. The good fit criterion is not sufficed by any of these functions so it would be possible to improve results by adding more gaussians than two. That would not work in the real analysis. Only a few hundreds of events are expected in the Run 2 analysis, while in the used MC a half million events are present. Running a fit on 10,000 events, $\chi^2$/NDF is almost equal to one, so two gaussians are enough for Run 2 analysis.

4. Punzi Effect

The introduction of the per-event error in fit could cause problems when it is wrongly used, as was pointed out by Giovanni Punzi in his paper[5] from 2003. Since the per-event error is another observable of the fit, the p.d.f. is in fact 2-dimensional. Thus one needs to take into account the error’s distribution ($w(\sigma_{m_B})$). To be more concrete, the p.d.f. in the maximal
likelihood function should be multiplied by \( w(\sigma_{m_B}) \). In this case:

\[
L(m_B, \sigma_{m_B}) = \prod_{i=1}^{n} G_2(m_B, \sigma_{m_B}) w(\sigma_{m_B_i}).
\]

In the case of no background component in the fit (like is in our MC-signal case), the maximal likelihood function could be written as:

\[
L(m_B, \sigma_{m_B}) = \prod_{i=1}^{n} G_2(m_B, \sigma_{m_B}) \prod_{i=1}^{n} w(\sigma_{m_B_i}).
\]

Since the distribution of the error is known, the last term is constant \(-\prod_{i=1}^{n} w(\sigma_{m_B_i}) = C\) and has no influence to the fit results, because of the maximization of the \( L \) is equivalent of the \( CL \) maximization.

In the case when a background component is present:

\[
p.d.f. = G_2(m_B, \sigma_{m_B}) + B_{ckg}(m_B),
\]

the maximal likelihood function can by written as:

\[
L(m_B, \sigma_{m_B}) = G_2(m_B, \sigma_{m_B}) w(\sigma_{m_B}) + B_{ckg}(m_B) w'(\sigma_{m_B}).
\]

The error’s distributions differ for the signal and background in general and therefore cannot be simply factorized out from \( L \). Different prescription of the maximal likelihood function leads to different results of the fit and that is the reason why the Punzi effect have to be consider in analyses of the real measured data that naturally include backgrounds.

5. Angular distribution

The precision of the description of the angular distribution is crucial for the observation of potential discrepancies between the theory and the experiment. The angular distribution of the decay \((\cos(\theta_K), \cos(\theta_L) a \phi)\) was simply described by the product of polynomials for each angle in Run 1. This work proposes the upgrade for the Run 2 – using spherical harmonic functions.

The main disadvantage of using simple 3D polynomial is that it cannot describe any correlations between angles unlike the expansion of the p.d.f. into the three dimensional spherical harmonic functions \((N \text{ is the norm constant)}):\)

\[
f\theta_K,\theta_L,\phi(\cos \theta_K, \cos \theta_L, \phi) = \frac{1}{N} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f_{klm} Y_{l^m}(\cos \theta_L, \phi) P_k(\cos \theta_K).
\]

Here only expansion in the order of \( k_{\text{max}} \) and \( l_{\text{max}} \) is used (spherical harmonics of higher orders are very oscillating which improves the description of the random fluctuations, but they do not contribute to the data shape description):

\[
f\theta_K,\theta_L,\phi(\cos \theta_K, \cos \theta_L, \phi) = \frac{1}{N} \sum_{k=0}^{k_{\text{max}}} \sum_{l=0}^{l_{\text{max}}} \sum_{m=-l}^{l} \sqrt{k + \frac{1}{2}} a_{klm} Y_{l^m}(\cos \theta_L, \phi) P_k(\cos \theta_K).
\]

There is one more advantage of the spherical harmonic expansion, in addition to the better precision. The coefficients \( a_{klm} \) can be simply calculated and do not have to be fitted:

\[
a_{klm} = \sqrt{k + \frac{1}{2}} \int_{\Omega} f\theta_K,\theta_L,\phi(\cos \theta_K, \cos \theta_L, \phi) Y_{l^m}^* (\cos \theta_L, \phi) P^*_k (\cos \theta_K) d\Omega.
\]
Instead of the continuous integral the discrete values are measured, so the integral is substituted by the sum through each event (n - a number of data):

\[ a_{k\ell m} = \frac{\sqrt{k + \frac{1}{2}}}{M} \sum_{i=0}^{n} Y_{\ell m}(\cos \theta_{Li}, \phi_i) P_k(\cos \theta_{Ki}) \]

As a good-fit criterion, the condition \( \chi^2/NDF < 2 \) is used, in this case applied to all 2D projections of the p.d.f. To obtain values of the \( k_{\text{max}} \) and \( l_{\text{max}} \), we set them equal to 10. Then we evaluated coefficients \( a_{k\ell m} \) and sorted them according to their absolute value. After that we used bisection method and set the corresponding number of parameters with the lowest absolute value to zero and again we check the fit goodness until we found the least number of parameters for the good fit. The minimal number of the non-zero parameters (m.n.n.p) and their values of the \( \chi^2/NDF \) for the p.d.f 2D projections are shown in Tab. 1.

| dataset                              | m.n.n.p | \( \cos(\theta_K):\cos(\theta_L) \) | \( \cos(\theta_K):\phi \) | \( \phi:\cos(\theta_L) \) |
|--------------------------------------|---------|--------------------------------------|-----------------------------|----------------------------|
| signal decay                         | 16      | 1.80                                 | 1.55                        | 1.64                       |
| misidentification of K and \( \pi \) | 23      | 1.77                                 | 1.47                        | 1.53                       |
| misidentification of K and \( \pi \) | 120     | 1.93                                 | 1.10                        | 0.93                       |
| \( B_s \to \phi \mu^+ \mu^- \) (4)  | 20      | 1.92                                 | 1.41                        | 1.28                       |
| \( \Lambda_b \to \Lambda \mu^+ \mu^- \) | 25      | 1.43                                 | 1.79                        | 1.11                       |
| \( \Lambda_b \to pK^- \mu^+ \mu^- \) | 5       | 1.82                                 | 1.56                        | 1.54                       |
| combinatorial background              | 15      | 1.80                                 | 1.48                        | 1.29                       |

**Table 1.** The minimal number of the non-zero parameters (m.n.n.p) and their \( \chi^2/NDF \) values for the p.d.f 2D projections.

In the Fig. 3, the comparison of the description by 3D polynomial fit, the data and the fit by expansion into spherical harmonics is shown.

**Conclusion**
This work prepares improved tools for Run 2 analysis of the decay \( B^0_d \to K^* \mu^+ \mu^- \) at the ATLAS experiment.

The p.d.f. prescriptions were found for invariant masses of \( B \)-meson and \( K^* \) for the signal decay, the misreconstructed events from signal decay and background decays \( B_s \to \phi(K^+K^-)\mu^+\mu^- \), \( \Lambda_b \to \Lambda(1520)(pK^-)\mu^+\mu^- \), \( \Lambda_b \to pK^-\mu^+\mu^- \). Adding the precise masses descriptions to the final multidimensional likelihood function (describing the angular distribution) would suppress the effects of background and decrease an uncertainty of the final angular analysis.

The Punzi effect that is connected to the per-event error functions used for the \( m_B \) description was also discussed.

The precision of the description of the detector angular acceptance was improved by the expansion of the p.d.f. into spherical harmonic functions We determined the maximal orders of the expansion to reach sufficiently precise description of the detector angular acceptance. The use of the expansion should again suppress the overall uncertainty of the final analysis.
Figure 3. 2D projections of (from the left) polynomial fit function, data and the fit function expanded into spherical harmonics for the signal decay $B^0_d \rightarrow K^* \mu^+ \mu^-$. The show fit functions are not smooth because the projection of the 3D function into 2D is difficult to compute; an approximation by a histogram is used.

Acknowledgments
Support for this work has been received from the grant LTT 17018 of the Ministry of Education, Youth and Sports of the Czech Republic.

References
[1] Aaij, Roel and others 2016 Angular analysis of the $B^0 \rightarrow K^* \mu^+ \mu^-$ decay using 3 fb$^{-1}$ of integrated luminosity JHEP 1602
[2] M. Aaboud, ATLAS Collaboration Angular analysis of $B^0_d \rightarrow K^* \mu^+ \mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector JHEP 1810
[3] Altmannshofer, Wolfgang and Straub, David M. 2013 New physics in $B \rightarrow K^* \mu \mu$ ? The European Physical Journal C vol 73
[4] G. Aad, ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider JINST 3 S08003.
[5] Giovanni Punzi 2003 Sensitivity of searches for new signals and its optimization eConf C030908