**Università degli Studi di Padova**

Dipartimento di Fisica e Astronomia "G. Galilei"
Corso di laurea in Fisica


Tesi di Laurea

# Study of monitoring system of a calibration laser for the iTOP detector at Belle II

**Relatore:**
Dott. Stefano Lacaprara
**Correlatore:**
Dott. Alessandro Mordà

**Laureando:**
Gabriele Labanca
Matricola: 1069556

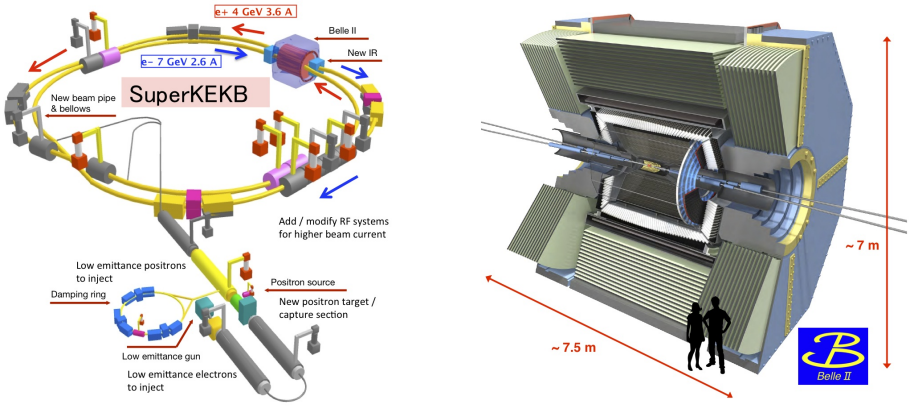
Anno Accademico 2015/2016

# Contents

# 1 Introduction

The work presented in this thesis concerns a study on a calibration system to be used by the iTOP ("Time Of Propagation") detector in the *Belle II* experiment ([1]) at *superKEKB* (upgrades of *Belle* and *KEKB*, respectively), managed by *KEK*, the Japanese High-Energy Accelerator Research Organization. The experiment is currently under installation at Tsukuba (Japan), and is scheduled for first data taking in late 2018.

Figure 1: (Left) The SuperKEKB accelerator, showing the position of Belle II; (Right) The Belle II detector

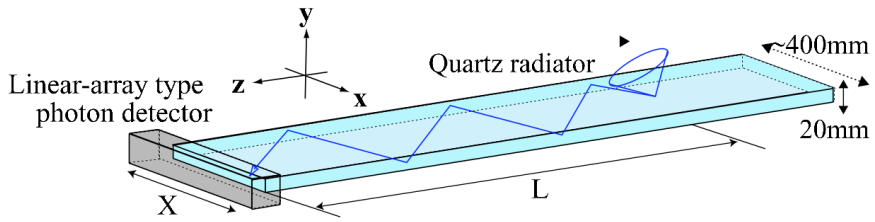



## 1.1 iTOP

The iTOP detector (Figure 2; *Technical Design Report*, [1]) will be used to identify charged particles, $\pi$ and K, and consists of 16 quartz crystal bars surrounding the interaction region arranged in a cylindrical shape. At one end of the bars there is an array of $16 \times 2$ MCP-PMT (Micro Channel Plate - Photo-Multiplier Tubes), at the other end there is a mirror. When a charged particle crosses a bar, it emits Cherenkov light, which is transmitted by internal reflection to the ends of the bar, where photons hit the photodetector. The angle at which the light is emitted depends on the $\beta$ of the particle; by measuring the curvature of the track of the particle in the magnetic field it is possible to compute its momentum: knowing the momentum and the $\beta$ is enough to separate particles with different masses such as $\pi$ and K. The photodetector records a pattern as a function of position and time which depends on the $\beta$ of crossing particles.

**Calibration system** A calibration system for the iTOP uses a laser, whose characteristic emission is to be known and whose light is splitted and reaches

Figure 2: iTOP schematics: the photodetector is highlighted. Image from the *Technical Design Report* [1]




through optical fibers all the photodetectors in the array of each bar, to check if they keep reacting the same way over time: this requires a monitoring system to measure the emission of the laser in different conditions: the resulting parameter is its **yield**, whose **stability** over time and exposure to temperature is to be studied..

# 2 Apparatus and data acquisition

The apparatus and the features of the signal are here exposed. In particular, an extimate of the background is computed.

## 2.1 Device schematics

The laser schematics is reported in Figure 3: for the monitoring of the laser a photodiode is used instead of a MCP-PMT.

The photodiode can be modeled as a capacitor in series with a resistor: when a photon goes through the photodiode, it creates electron-hole pairs which reach to the terminals of the capacitor, making its charge drop by an amount proportional to the energy deposited. As a consequence, the voltage decrement is also proportional to the amount of energy deposit.

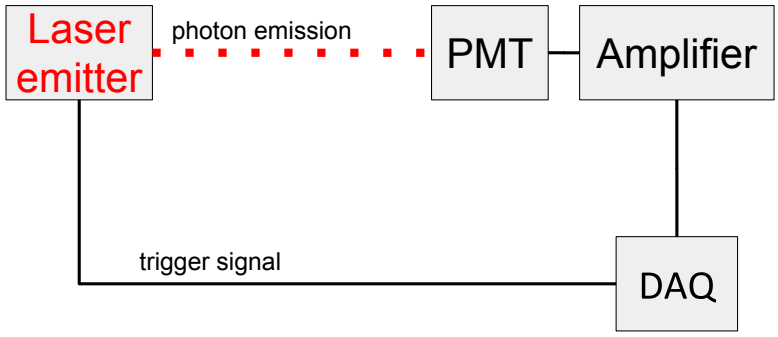
An electronic device (DAQ in Figure 3) reads the voltage value and writes it in a file, which is analyzed in the present work, starting from a preliminar analysis performed by the research team, displayed in this section.

The trigger signal is produced by the laser at a fixed $\Delta t$ prior of the actual light emission and allows a temporal comparison between the signal shapes.

## 2.2 Signal features

The signals (trigger and photon) consist in 1024 bins representing the time: each bin corresponds to 200ps and stores the value of the measured voltage amplitude.

Figure 3: Laser schematics: the laser emitter works at 1kHz frequency, shooting short pulses and its attenuation is tunable; the photon emission is trasmitted through optical fibers to the photodetector, whose signal is then amplified before reaching the data acquisition devices; meanwhile trigger signal is sent directly to the DAQ (data acquisition, namely the electronics that acquires the two signals and save them in a file) through a parallel channel



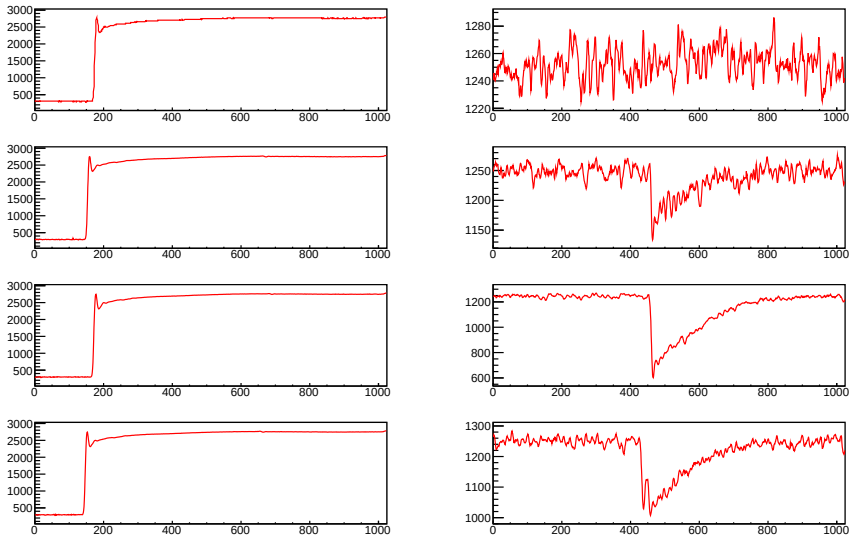

**Shape** Figure 4 shows some typical events (from top to bottom):

- a signal without evident peaks, corresponding to zero photons;
- a negative peak exponentially decaying, corresponding to a one-photon detection;
- a peak with a larger amplitude, due to the detection of several photons at the same time;
- a not very clear event, featuring two consecutive peaks. A possible explanation is the assumption of internal reflections, causing some delay for the arrival of the two photons, or the effect is probably simply due to noise effects.

From now on the generic acquisitions will be called "signals", while the acquisitions of photons will be called "events".

**Background** In the following the voltage will be considered as the variation from the voltage baseline, namely the average value of the voltage before the arrival of a photon.
For each signal this pedestal value has been evaluated by averaging the amplitude over the bins in the interval [10,110] which are consecutive and in a range near the signal region, but clearly separated from it. The histogram in (Figure 5) shows the distribution of the background estimate: as expected, it has a

Figure 4: Typical events: on the left, the trigger signal; on the right, the photon signal. From top to bottom: triggered background, one peak, one higher peak (background scaled down), one double peak [x-axis: time(bins); y-axis: signal amplitude] [bin=200ps]



central peak. **This estimate has been used as a baseline for signal**: the peak for zero-photon signal is centered in zero.

The triggered signals are considered "photons" when their maximum amplitude is bigger than 80 *and* their integral over 400 bins is bigger than 4000, "background" otherwise; this choice has been done considering Figure 12: being the the width of the noise distribution $\sim 10$, the limit is well over $3\sigma$. It is important to notice that the background signals are *not just noise*, since they represent the of zero-photon component of the collected events.

**Trigger and time of arrival** The trigger signal plays an important role: as shown in Figure 6, subtracting the trigger time the "time of arrival" (defined as the bin of the minimum) of signals is distributed around peaks: the first and higher (from now on named "principal") is followed by some others ("secondary"), progressively decreasing; there is another peak around the 860th bin ("tail"). Their nature will be discussed in 3.2.3

Figure 5: Estimate of the background

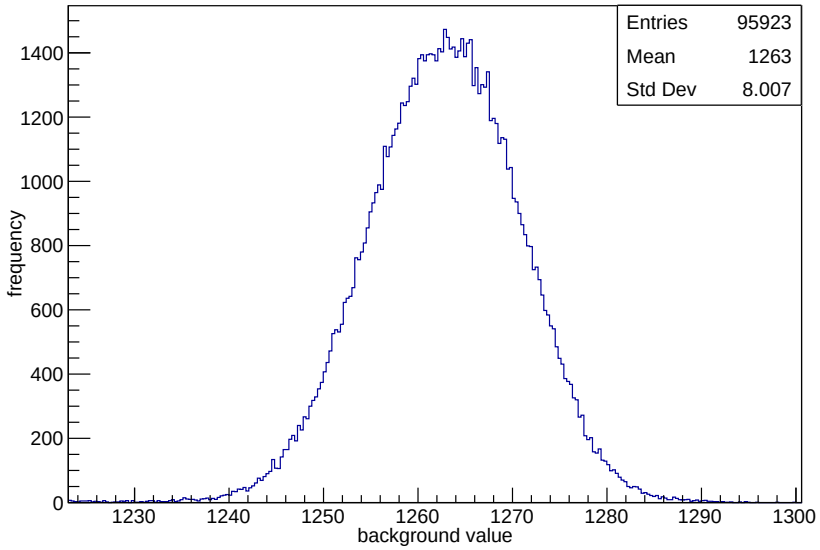
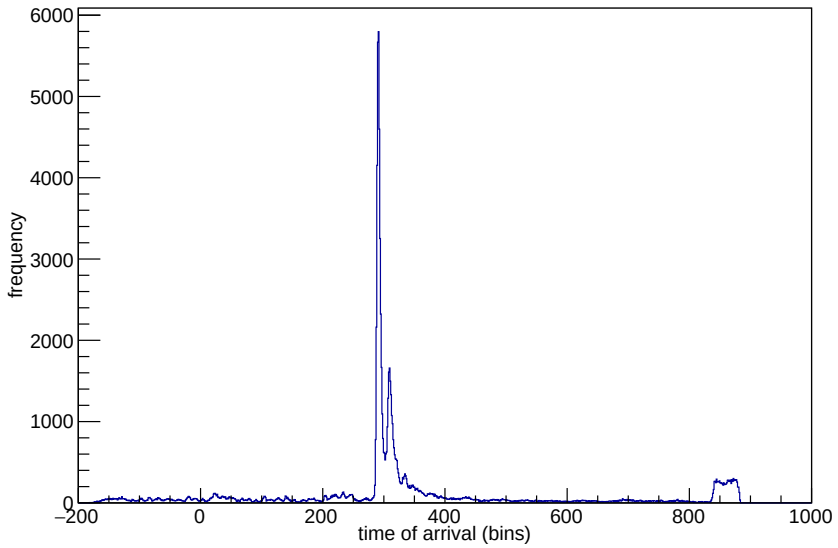



Figure 6: Delay of the time of arrival with respect to the trigger time: there are some steep peaks around 300 ("principal", "secondary") and a block around 850 ("tail") [x-axis: time (bins); y-axis: frequency]

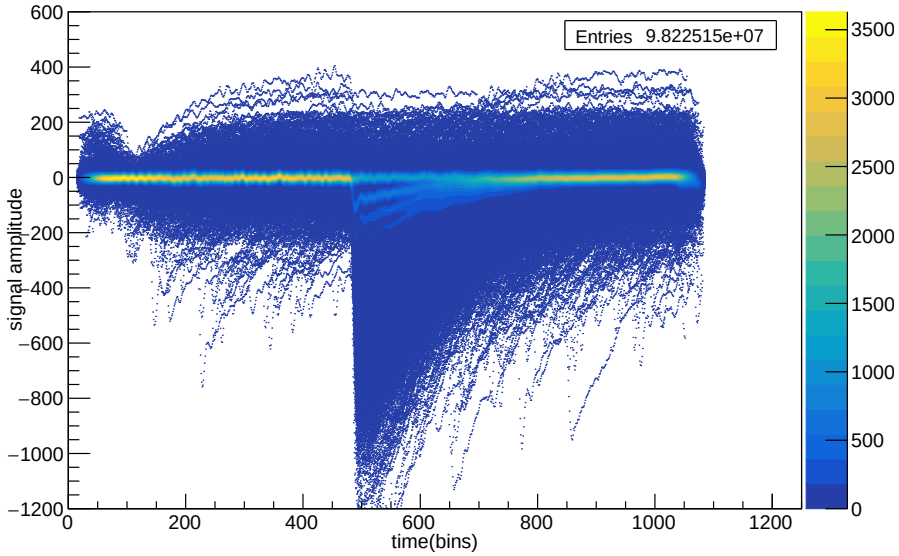


# 3 Signal analysis

**Datasets** The thesis focuses on building an algorithm to extract a variable descriptive of the emission. To do so, two datasets have been used:

- **test**, with high attenuation, suppressing the photon emission (Figure 7);
- **real**, with lower attenuation, close to the working conditions (Figure 8).

Figure 7: 2D plot of all events superimposed (high attenuation): the background it is clearly visible and two or three peaks can be identified, their intensity decreasing when the amplitude is higher [bin=200ps]

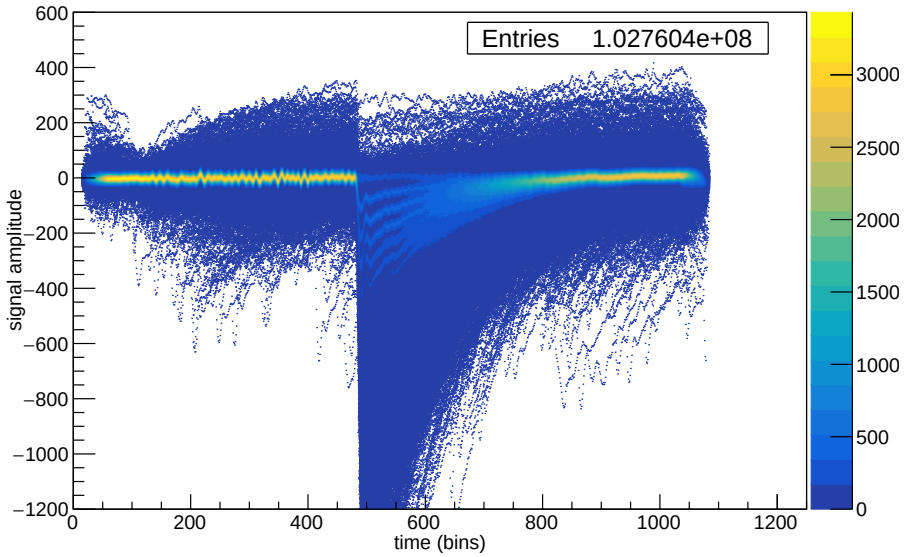



**Variables** Three possible choices are discussed:

- amplitude in a bin;
- integral of the signal;
- integral of a function fitting the signal.

Finally, the chosen one, the integral of the signal, is used to describe the light yield of the laser in the two datasets, by counting the frequency of arrival of n-photons (n from 0 -noise- to the maximum number of photons seen). This distribution, in principle Poissonian (possibly plus instrumental effects), is sensitive to the light yield.

Figure 8: 2D graph of all events superimposed (lower attenuation): the background is less substantial and more stripes are visible [bin = 200ps]




**Analysis** In order to estimate the fraction of events $w_n$ corresponding to the detection of $n$ photons, three variables have been considered: the amplitude in a fixed bin ( 3.1.1), the integral of the signal ( 3.1.2) and the integral of the function which fits the signal ( 3.2). The integral of the signal has then been chosen and fitted with a sum of Gaussian functions ( 3.5): ideally the fractional weights of these functions are expected to follow a Poissonian distribution, but some instrumental effects may be present.

## 3.1 Amplitude and integral distributions

In order to describe quantitatively the signal, two kinds of variables are considered:

- **local**: evaluated considering the amplitude in only one selected bin; they are simpler to evaluate, but more affected by noise effects;
- **global**: evaluated over a range of bins; they are less affected by noise effect.

The maximum amplitude and the amplitude at a fixed bin are in the first category ( 3.1.1), while the integral is in the second ( 3.1.2).

### 3.1.1 Amplitude distribution

A first variable considered for the characterization of the signal is the maximum absolute value of amplitude: its distribution is shown in Figure 9. This variable, however, is biased by construction (being a maximum): this is clearly seen for the first peak, which should be centered in zero.

A less biased variable is the amplitude at a fixed bin: the 292nd one has been chosen, since is the central value of the principal peak (Figure 6). There are issues expected using this variable: first, using a fixed bin does not detect the maximum absolute value of the amplitude if the arrival time is different (which is the case for many events); furthermore it is very sensible to noise. Its distribution is presented in Figure 10: the shape is more regular than the one of maximum amplitude.

### 3.1.2 Integral distribution

A method to refine the analisys is to compute a simple integral of the signal shape, namely a sum of the amplitudes, over 400 bins after the arrival of the signal. The length of integration has been chosen noticing that the decay length of almost all the signals is 400 bins; varying this length does not affect the results substantially. It can be seen, from Figure 11, that the distribution is more similar to the one of the amplitude, and features the advantage that more peaks are visible. The improvements concern primarily the number of peaks clearly identifiable, which increased up to 6-7. Looking at the peak of the background, a shift towards lower number is evident: this can be due to a *drift of the baseline* along the bins. This however does not affect the analysis, since the interesting quantity is the distance between the peaks.

Figure 9: Maximum amplitude histogram: a shift of the peak in zero is visible

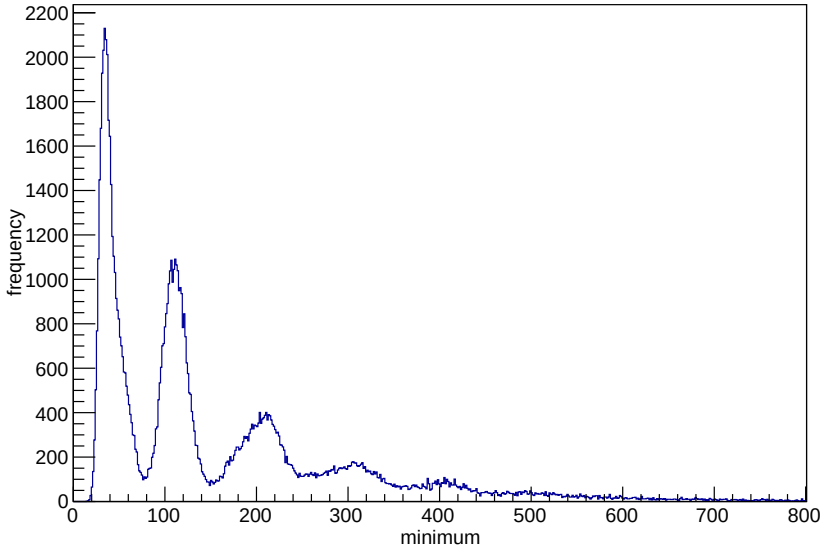



Figure 10: Amplitude at a fixed bin: the highest peak is at zero, with the other progressively decreasing

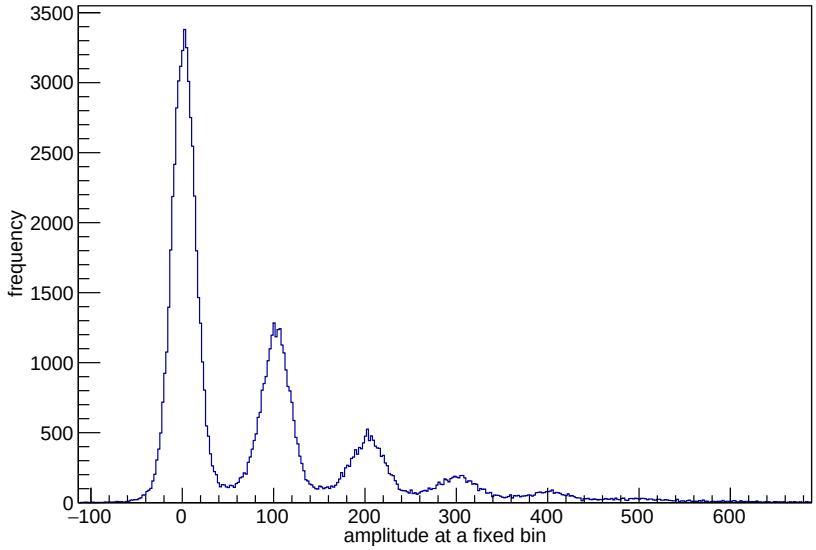
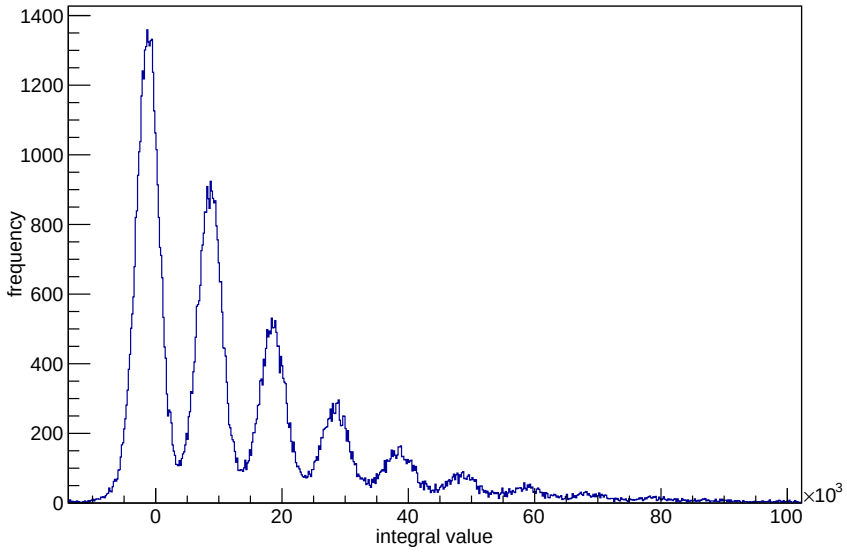


Figure 11: Histogram of the integral over 400 bins: the peaks are shifted, this can be due to a drift in the baseline




### 3.1.3 Discussion on the methods

**Uncertainties** To compare the efficacy of these differents methods, the the peak corresponding to the detection of three photon is fitted with a Gaussian function and its uncertainty is computed and compared in Table 1: the values have been normalized on the distance between this peak and the previous one. It is evident that the distributions of the amplitude at a fixed bin and the integral describe more separated peaks.

Table 1: The uncertainties on the three-photons peak, normalized on the distance between this peak and the previous, compared for the distributions of maximum amplitude, amplitude at a fixed bin, and integral

| | $\frac{\sigma}{\Delta\mu}$ |
|---|---|
| MAXIMUM AMPLITUDE | $41\% \pm 2\%$ |
| AMPLITUDE AT FIXED BIN | $27.0\% \pm 0.7\%$ |
| INTEGRAL | $27.2\% \pm 0.4\%$ |

**Correlations** Evaluating the correlation between these variables helps understanding how well they describe the emission. The correlation between the

amplitudes ( 3.1.1) and the integral ( 3.1.2) is plotted in Figure 12 and Figure 13: they describe the same phenomenon, but it is evident that the amplitude at a fixed bin fails to classify a big part of the events.

In particular, from Figure 12 a discrimination criterion between background and signal can be defined, considering background the events in the left-bottom peak.

The only variable being neatly distributed and well correlated to the number of photons detected is the integral as a sum over the bins. This one is thus the best variable according to the analysis done in this section.

Figure 12: Minimum as a function of direct integral: the correlation is strong; the events with maximum smaller than 80 and integral smaller than 4000 are considered background

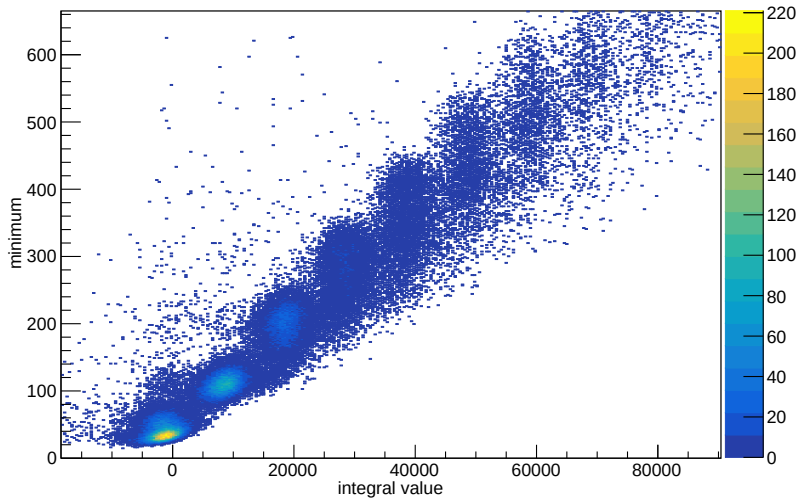
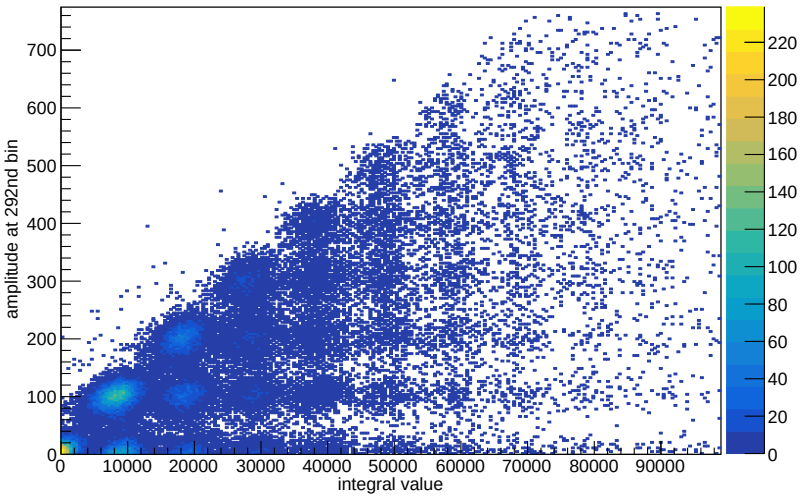



Figure 13: Amplitude at bin 290 as a function of direct integral: the correlation is not good



## 3.2 Fit the signal

The signals have been fitted one by one, then the integral of the fitting function has been evaluated and used as another descriptive variable. This is similar to the direct integral, but possibly even less sensible to background oscillations.

### 3.2.1 Strategies

The signals have been fitted one by one, with different methods, with the purpose of integrating the fitting function, having another descriptive variable, analogous to the simple integral, but possibly less dependent on the noise.

The function chosen to fit the signal is the following (dimension proportional to a voltage):

$$f(x) = e^{A-kx}$$

where $x$ is the bin index, in units of 200ps.

The methods differ in the way the points to fit are chosen:

**simple points** 50 points with coordinates:

$$(x_0 + n\Delta, \mathcal{A}(x_0 + n\Delta))$$

being "$x_0$" the position of the aboslute minimum of the signal amplitude "$\mathcal{A}$", "$\Delta$" a fixed distance between two subsequent points, equal to 10 bins, and "$\mathcal{A}(x_n)$" the amplitude value at the bin "$x_n$".

**points with uncertainties** 50 points with the same values as the previous case for the x and y coordinates with in addition an associated error to both of them.

The error on the y value is given as the uncertainty on the background mean.

**averaged points** 50 points with the central value of y given by the average over the 5 surrounding bins around the central one (*i.e.* $n_{\text{central}} \pm 2$) and associated error on y given by the uncertainty on the background mean. The choice of the average is done in order to reduce the amplitude fluctuations due to the electric noise affecting the y coordinate central value; the estimated error on the y value may represent, with this choice, a conservative estimation. Nevertheless, it will be shown that the obtained $\chi^2$ distribution of the fit will assume reasonable values.

### 3.2.2 Results

**Compare the strategies** The efficiency $\epsilon$ is defined as the number of converged fit over the number of attempts; analogous efficiencies can be defined separately for photon ($\epsilon_\phi$) and background ($\epsilon_B$) triggers. The efficiencies as well as the standard deviation of a Gaussian fit of the first peak of the distribution ($\sigma$) are reported in Table 2.

Table 2: The efficiencies, as defined in 3.2.2, as well as the Gaussian error on the first peak, are reported for all the strategies of interpolation

| | SIMPLE POINTS | UNCERTAINTIES | AVERAGED POINTS |
|---|---|---|---|
| $\epsilon$ | $0.921 \pm 0.003$ | $0.967 \pm 0.003$ | $0.959 \pm 0.003$ |
| $\epsilon_\phi$ | $0.9995 \pm 0.0040$ | $0.996 \pm 0.004$ | $0.994 \pm 0.004$ |
| $\epsilon_B$ | $0.780 \pm 0.005$ | $0.935 \pm 0.005$ | $0.925 \pm 0.005$ |
| $\sigma$ | $1670 \pm 18$ | $1684 \pm 14$ | $1691 \pm 12$ |

Since $\epsilon_B$ is much lower for the first strategy and the number of zero photon events is of interest, the first strategy has been excluded. The performances of the others are similar, but the last one should be less dependent on noise oscillations, since the points are averaged, and is thus the chosen one. The distribution of the integrals of the fitting functions for this method is shown in Figure 14.
The events for which the fit attempt did not converge have been then removed from the analysis.

**$\chi^2$ distribution** The $\chi^2$ distribution of the fit for the chosen method is shown in Figure 15: there are 50 points to fit and two parameters in the fitting function, so there are 48 degrees of freedom. The mean is high, but the higher peak corresponds to a reasonable value (near 50). Since the results are good, using a less conservative estimate of the uncertainty wouldn't be accettable. .

Figure 14: Distribution of the integral of the fitting function: the first peak (cut due to the evaluation process) is background, as defined in 2.2, while the others are photon events

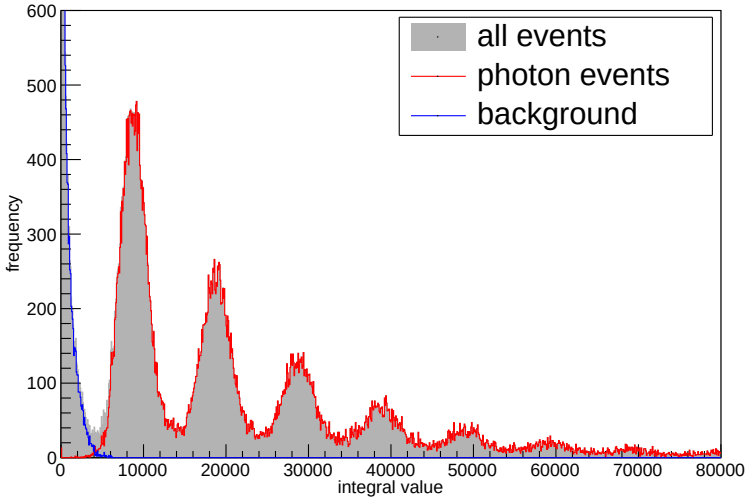
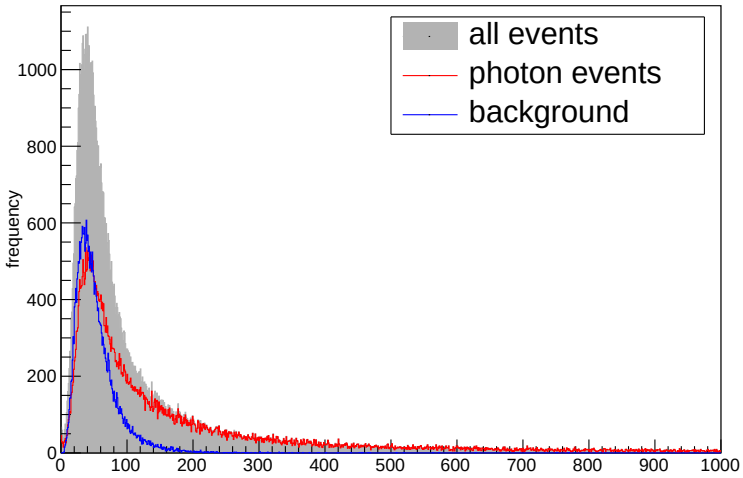



Figure 15: The $\chi^2$ function: points with error, converged fit. For the distinction between photon events and background see 2.2. Since there are 48 degrees of freedom, the peak around 50 is a mildly good result



**Parameters distribution: signal events** The parameter $k$ is expected to be the same for all the events, since it depends on the RC circuit of the photodiode, and in Figure 16 the results agree with this hypothesis. The presence of a shoulder in the left part of the distribution could be due to the drift in the baseline discussed before.

Since $k$ is fixed, $A$ must vary depending of the type of event and looking at Figure 17 a distribution analogous to the one of amplitude can be recognised. Since $A$ is the constant term in the exponent, it is proportional to $\ln n$ (being $n$ the maximum amplitude of the signal in units of the one-photon amplitude); for that reason the distances between two subsequent peaks are given by $\ln \frac{n+1}{n}$.

The same pattern is visible in Figure 18: looking only at the photon events some stripes are present, which correspond to the distributions of the amplitude at 292nd bin and of the integral.

Figure 16: The parameter k and filters: photon events ( 2.2). Only converged fits considered. As expected, there is only one peak

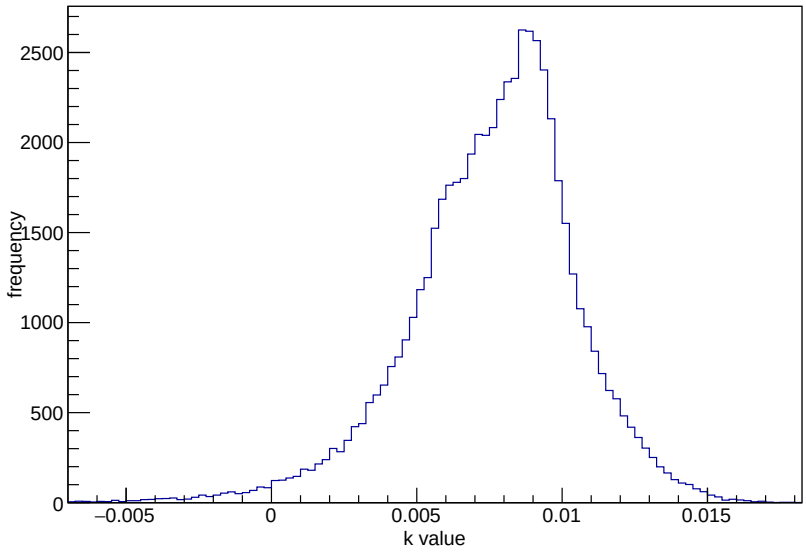


Figure 17: The parameter A and filters: photon events ( 2.2). Only converged fits considered. A pattern similar to the amplitude peaks is evident

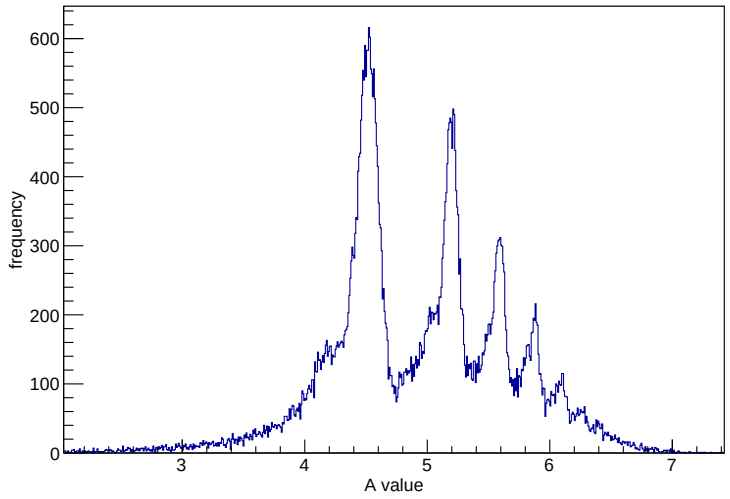
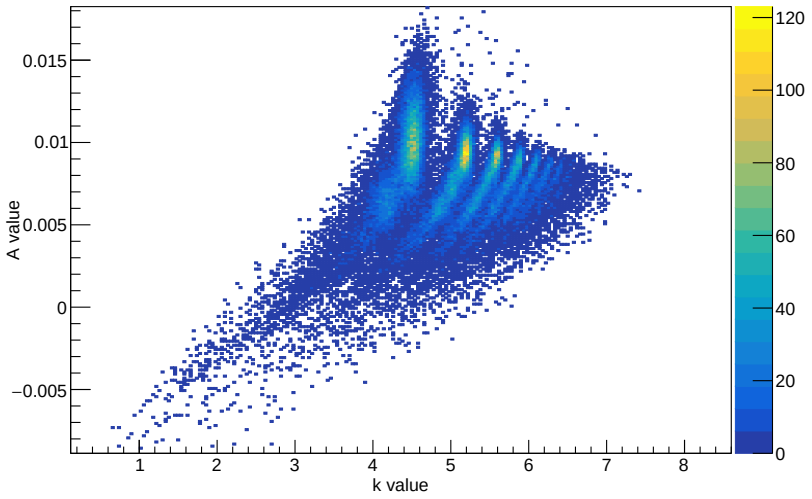


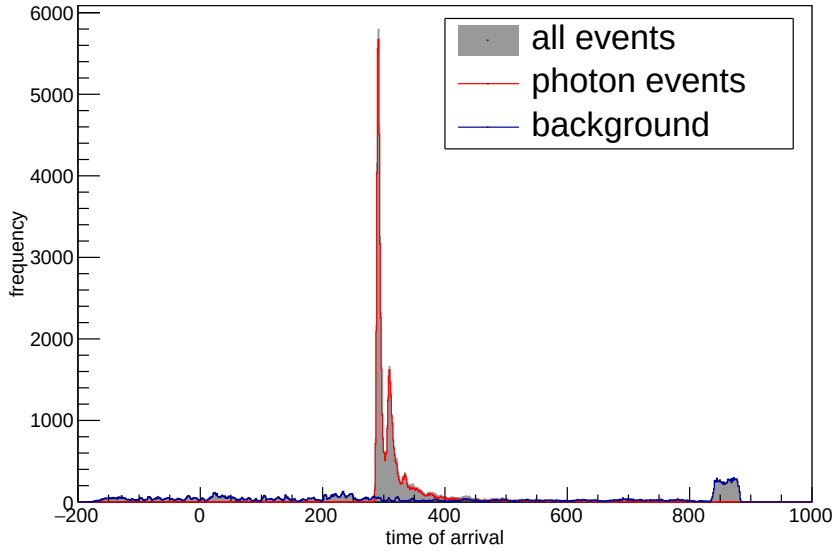

Figure 18: The parameter k as a function of A filtered for photon events ( 2.2). Only converged fits considered. A pattern similar to the amplitude peaks is evident



### 3.2.3 Discussion on time of arrival

Filtering the time of arrival with maximum amplitude it can be seen that the tail peak has only background (Figure 19), while the other two contain photon events.

Figure 19: Time of minimum filtered for photon events and background events ( 2.2). It is clear that the tail peak consist almost only of background events




Concerning the primary and secondary peaks, Figure 20 and Figure 21 show the distribution of k as a function of A, presenting a shape analogous to to the photon events seen in Figure 18.
This is sufficient to explain the nature of these peaks, at least considering the goals of the thesis. Probably a better choice to define the time of arrival would have been the position of the step falling front, rather than the position of the minimum. However, this does not affect the results presented in this work.

Figure 20: The fit parameter k as a function of A, for events with arrival time in the primary peak: the shape resembles the typical signals

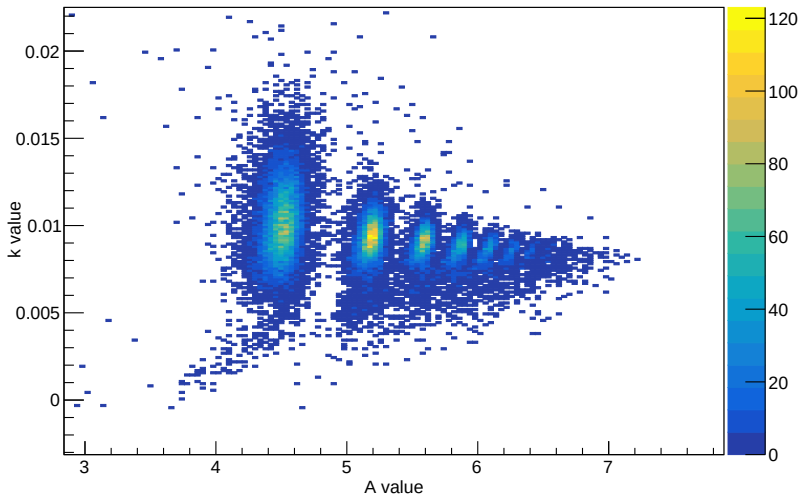
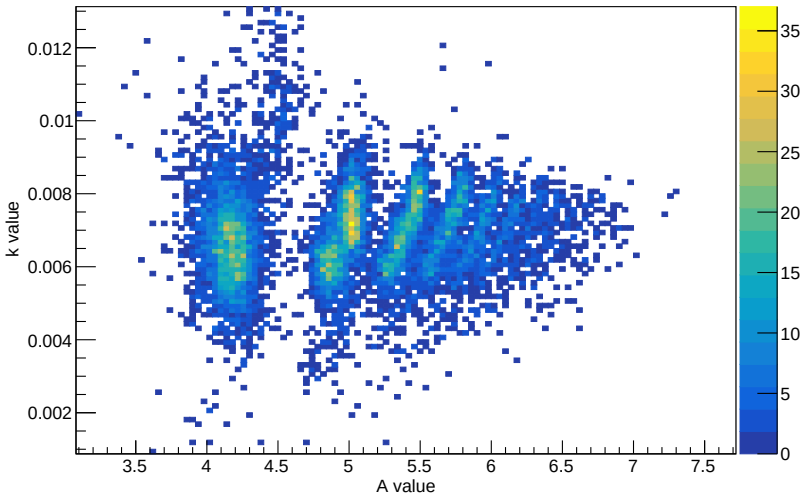


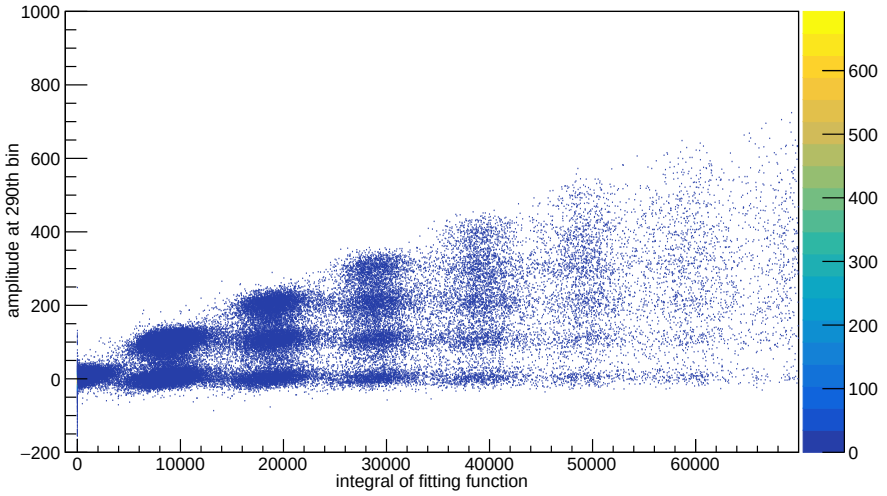

Figure 21: The fit parameter k as a function of A, for events with arrival time in the secondary peak: the shape resembles the typical signals



## 3.3 Discussion on variables

At this point, a choice has to be done, between the simple amplitude, the integral as a sum, or the integral of the fitting function (the more complete method, with averaged points and errors, is used). The first thing to notice is the disagreement between the simple amplitude and the other two: from Figure 13, as seen in 3.1.3 and Figure 22 is evident that the variables behave differently. Instead, the two strategies of integration agree very well (Figure 23). The integral as a sum over bins is chosen, because it is easier to compute and indipendent on the method of fit used.

Figure 22: Amplitude at bin 290 cross-plotted with integral of fitting function: a strong disagreement is noticeable




## 3.4 Real case

After the definition of the method, another acquisition has been analyzed, this time with attenuation set at a value similar to that expected in the *Belle II* experiment, to compare and check the results.

In Figure 8, where all the events in the set are plotted, it can be noticed an increased fraction of events at higher amplitudes, as expected. The results of the different methods of integration are summarized in Figure 24 (the simple sum over bins) and Figure 25 (the integral of the fitting function, with errors and averaged points).

Figure 23: Integral as sum over bins cross-plotted with integral of fitting function: the agreement is good

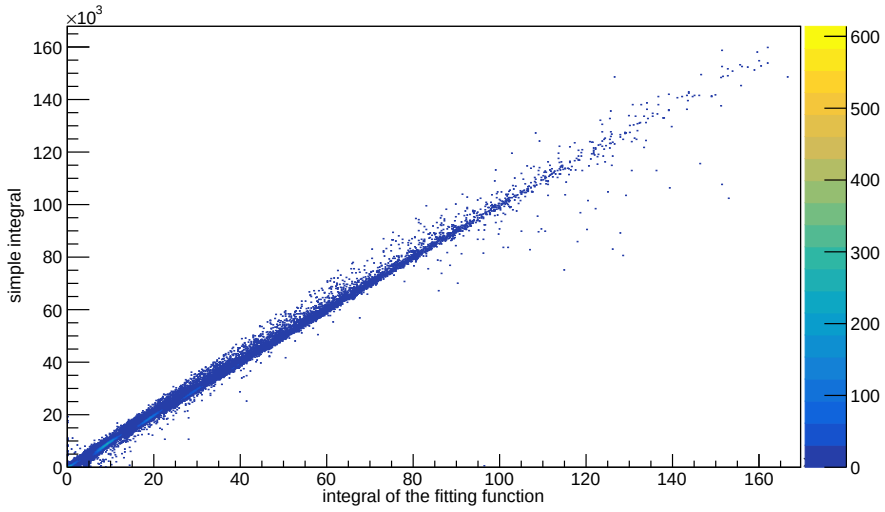



Figure 24: Direct integral histogram, real dataset

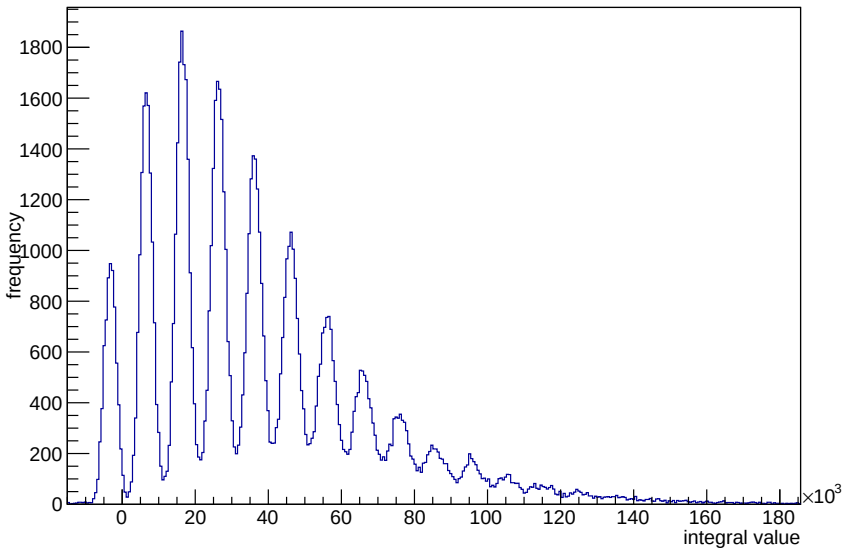
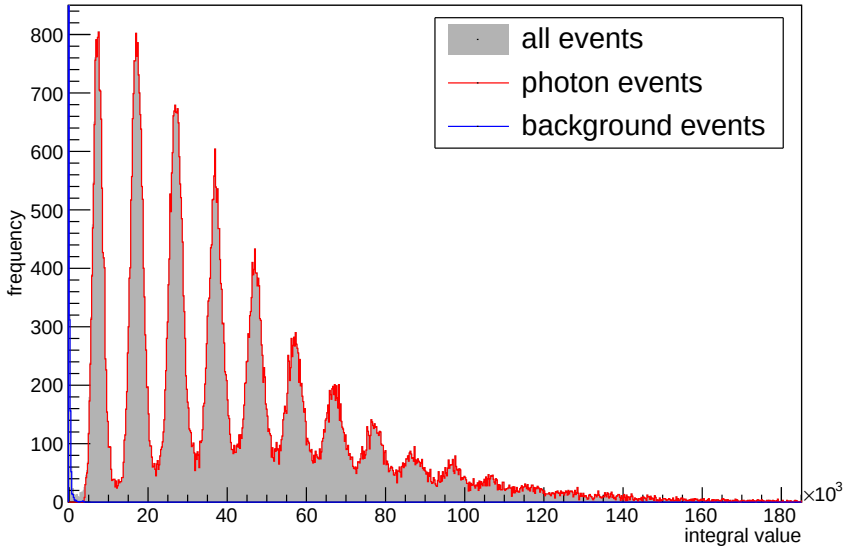


Figure 25: Integral of fit histogram, real dataset: the first peak (cut because of the evaluation process) is background, while the others are photon events




## 3.5 Light yield estimation

**Fit the integral** The direct integral has been chosen as the optimal method to analyze each trigger, for the reason that leads to the same results of the integral of the fitting function, however being much easier to compute and independent on errors made during fit. Another fact leading to this conclusion is that the integral of the fitting function doesn't allow to use the peak in zero, despite it being the bigger.

Once the optimal method to analyze each trigger has been defined, the next step is to look at the distribution of the number of photons detected by the photodiode, which depends on the laser light yield. In order to describe the distribution, a fit is performed.

The function chosen to fit the integral distribution is a sum of Gaussian distributions, each centered at points whose distance from one another is fixed to be the same. The peak at zero needed an exception, though: having the tails bigger than those of a normal Gaussian, it has been fitted with two Gaussian functions. The resulting function is

$$F(x) = w_0 \Big( G_a(x) \cdot C + G_b(x) \cdot (1 - C) \Big) + \sum_{i=1} G_i(x) \cdot w_i$$

where the index 0 refers to the peak at zero photons, and the others follow; the Gaussian functions are written as $G_i$ while their weigths are written as $w_i$; C is

a constant used to preserve the normalization; $G_a$ and $G_b$ are the two Gaussian functions used for the peak at zero. The range has been limited at the seventh photon peak for the case with test attenuation (Figure 26) and at the tenth for the case with real attenuation (Figure 27). Each figure shows the true function and the same in logarithmic scale, to better see the deviations from the data; the pull distribution follows. Most of the pulls are less than $3\sigma$, showing that the result of the fit is good.

Figure 26: Fit of the direct integral distribution, test attenuation: from the top, the interpolation, in normal and logaritihmic scale, pull distribution

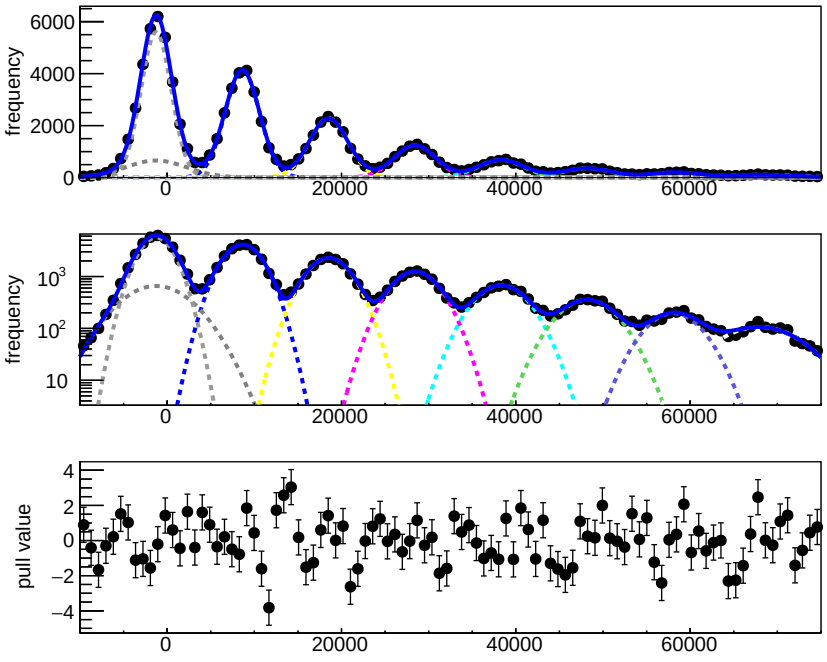
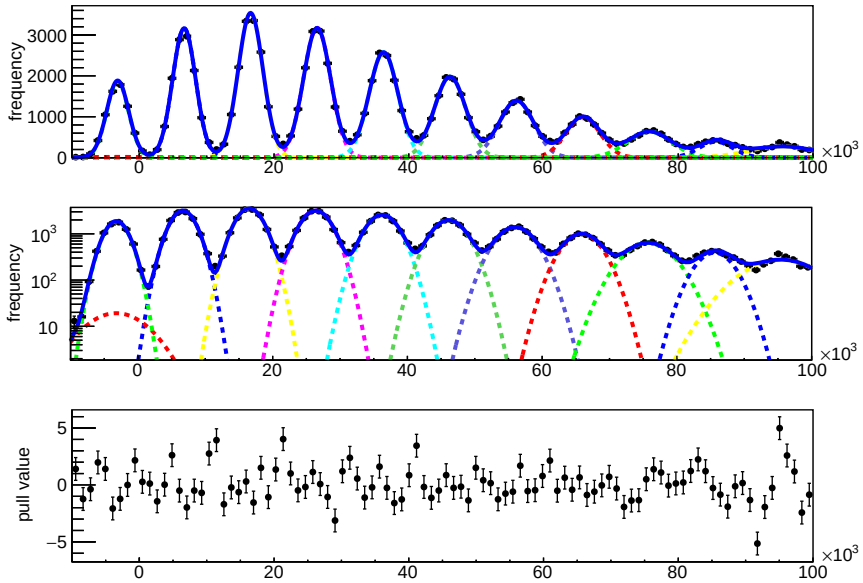


Figure 27: Fit of the direct integral distribution, real attenuation: from the top, the interpolation, in normal and logarithmic scale, pull distribution




**Fit the distribution** Finally, an attempt has been done to fit the distribution of the set (number of photons,weight): in Figure 28 and Figure 29 the blue line is a Poissonian function, while the red line is respectively an exponential and a Gaussian function. It is evident that the distribution isn't purely Poissonian.

Figure 28: Fit of the direct integral distribution, attenuation = 100%; red line: exponential function (parameters in the panel); blue line: Poissonian function (mean = 0.1)

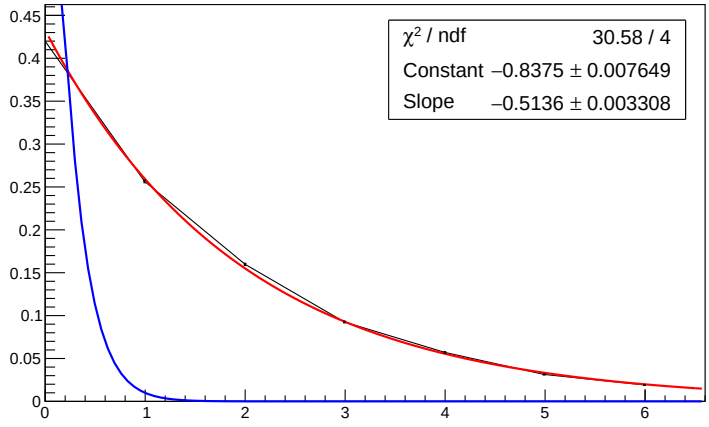
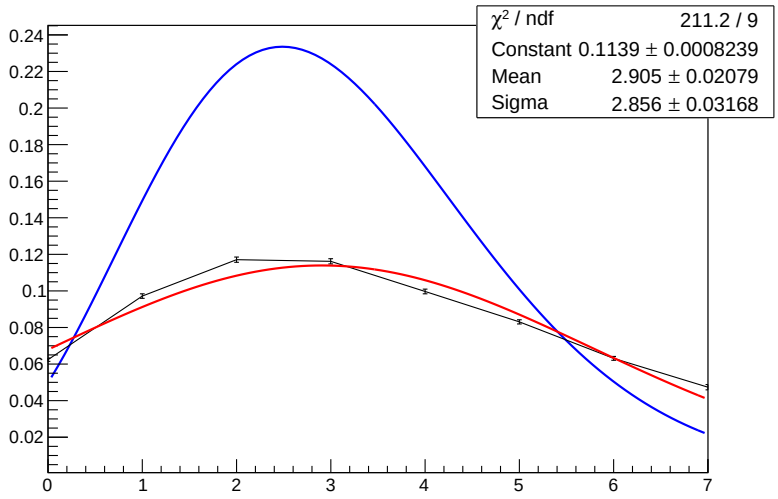



Figure 29: Fit of the direct integral distribution, attenuation = 80%; red line: Gaussian function (parameters in the panel); blue line: Poissonian function (mean = 3)



# 4 Conclusions

The analysis has found a parameter whose distribution can represent the signal emitted by the laser: the simple integral over 400 bins starting from the 290th bin of the array of data (where most of the events have minimum) allows a simple and fast computation. The fit with Gaussian functions works, as expected, but the distribution followed by the photon emission is not simply Poissonian.

**Further developements** An attempt to fit the photon distribution can be done with a convolution of a Poissonian with a Gaussian, so that the efficiency of the detector is taken into account too.

Further steps will be analyzing the behaviour of the parameters of this function under different conditions: not only varying the attenuation, but also the time of acquisition, the amplification and ambient parameters such as temperature.

# Bibliography

# References


[1] Z. Doležal, S. Uno
*Belle II Technical Design Report*
KEK Report 2010-1, October 2010

[2] *Belle II* project site
`https://www.belle2.org/`

[3] *SuperKEKB* project site
`http://www-superkekb.kek.jp/`

[4] K. Inami, et al.
*Nuclear Instruments and Methods*
in *Physics Research*, Section A, 766 (2014), p. 5

[5] F. Bianchi et al.
*The laser calibration system of the TOP detector*
BelleII note 0041, Mar. 15, 2015

[6] Advanced Laser Diode Systems
Picosecond Diode Laser - PiLas Manual