2019-012

Comparison of Photon Correlation Algorithms for Astronomical Intensity Interferometry

Bachelorarbeit aus der Physik

Vorgelegt von

Frederik Wohlleben

13.08.2019

Erlangen Centre for Astroparticle Physics

Friedrich-Alexander-Universität Erlangen-Nürnberg

Betreuer: Prof. Dr. Stefan Funk

Index

Chapter I

Foreword

While I was preparing for my bachelor thesis an extraordinary incident happened: On April 10^{th} 2019 the Event Horizon Collaboration announced that they were able to take "the first ever image of a black hole". This achievement was of such interest that not only scientists were cheering the collaboration, but all around the world newspapers and TV shows reported about the great success these scientists had.

What most of the public did not take interest in, was the way in which this "image" was taken. When diving deeper into the reports about the "picture taking" one notices that the process of generating this image was not as simple as pressing a button on the best camera ever built, but instead involved vast amounts of data from different radio telescopes, a great deal of logistics and massive resources of calculating power. And the image that came out of this process is not (as assumed by many) the thing one would see when only standing close enough to the black hole, but instead an intensity distribution colored in black and yellow, mapping the measured radio frequency to a visible color.

The concept behind the process of generating the image of the black hole is based on interferometry, in this case very long baseline interferometry (VLBI).

This example shows that new and more precise methods of measurement are not only leading to new scientific discoveries but are also engaging the public in the fascinating worlds that are waiting out there to be found.

Another thing this highlights is that the idea of interferometry is - even today - a state of the art concept that is improved further and further. While the Event Horizon Telescope did its measurements in the radio spectrum, the project this thesis is part of, aims at measuring in the visible spectrum, which allows for the observation of other than radio sources. The presented thesis focuses on the programs behind the correlation of the signals taken by two telescopes, making them more than just two telescopes: an intensity interferometer.

Figure 1.1: First ever image of a black hole generated through VLBI, taken from [1]

In the beginning a short motivation for interferometry as well as a short historic summary will be given. This is followed by an explanation of the physical and mathematical background.

The next chapter is focused on the acquisition and processing of data. It provides the necessary information to understand the process of data taking, as well as the procedures in which the data is analyzed in both the peakfinding and the current correlation method.

In Chapter IV the measurements, conducted to investigate the characteristics of the different algorithms, are explained and evaluated. It is centered around two measurements, one conducted with an LED, the other one with a pulsed laser. These measurements allow for the investigation of the noise level as well as time resolution and signal height. Afterwards the computation runtimes of the different methods are compared in respect to some characteristic parameters.

In the end a conclusion and outlook are given to sum up the findings of this thesis and provide an idea about possible future research.

Chapter II

Intensity-Interferometry

2.1 Motivation

When looking at the night sky one might get the impression that all stars (except for the sun, which is not visible on the night sky anyway) are of equal size and only differ in their luminosity. First of all, this is quite unlikely and furthermore one has to acknowledge, that due to the different distance of the stars only their angular size could be estimated by this method.

When taking a closer look at this method of "measurement" it turns out, that their seen extent is mainly due to effects of the earth's atmosphere and not due to the actual angular size of the star. In fact the angular extent of the stars is so small that even from space optical telescopes are barely able to measure it. This is due to diffraction that limits the angular resolution α_{min} of optical telescopes to

$$\alpha_{\text{min}} = \sin^{-1}\left(1.22 \cdot \frac{\lambda}{D}\right) \tag{2.1}$$

with λ being the wavelength and D the aperture. Taking long wavelength visible light (700 nm) and a telescope with a 10 m diameter the limit of angular resolution would be 17.6 mas. This is still not precise enough to measure most stars. Therefore another method needs to be employed.

In astrophysics interferometry is used to determine the angular size of stars. The easiest setup for such a measurement is a Michelson Interferometer. It uses amplitude interferometry and is already quite good in measuring the angular size of a star. A Michelson Interferometer uses two telescopes in the distance of a certain baseline d pointed at the same star. The light collected by these two telescopes is then guided through beam transportation optics (BTO) and interfered later. It is of utmost importance that the path traveled by the light from both telescopes is of the same length to keep the timing of the amplitudes in sync. Since the path, the light traveled before it gets to the telescopes already, is of different length (at least if the

star is not straight above the center of the baseline), this needs to be corrected for in the BTO by adding a variable extension of the path. The extension needs to be able to change its length constantly as the star "moves" along its path on the sky and therefore the path difference changes all along. This calls for a very high precision in adjusting the variable extension in the range of nanometers. At the same time the required precision limits the length of the baseline as such precision can only be achieved in the range of meters.

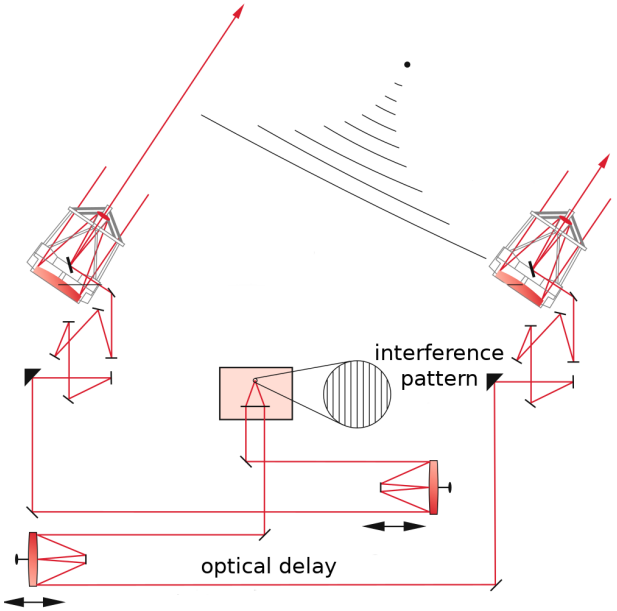


Figure 2.1: Schematics of an amplitude interferometer, adapted from [2]

To break this limit a new technique, called intensity interferometry (II), was developed by Hanbury Twiss & Brown in the 1950s. Instead of interfering the light optically, they measured the intensities at the two telescopes using photomultiplier tubes (PMTs) transforming them into electrical signals. The signals from the PMTs are taken and correlated, either live through wires or just recorded and computationally correlated afterwards [3]. Even though the amplitude and phase information is not contained in this data anymore, the signal still contains enough information to deduce the angular size of the star (compare figure 2.2).

This is due to a property of thermal light sources called photon bunching. Instead of arriving at a steady flux, the photons are likely to be in small bunches (compare figure 2.3). This means that the photons from a thermal source are correlated in time and space. The spatial correlation is used to later reconstruct the light source .

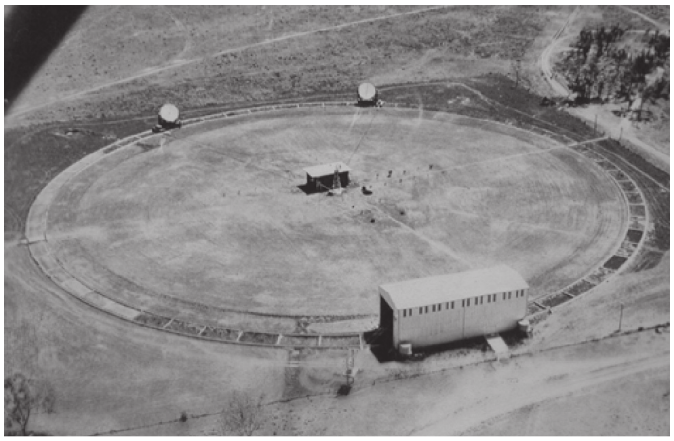


Figure 2.2: Aerial view of the NSII, taken from [4]

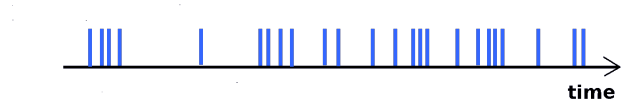


Figure 2.3: Stream of photons in a bunched signal, adapted from [5]

When measuring with an intensity interferometer the temporal correlation is calculated. The recording of the intensity allows for huge baselines only limited by the capability to later reconstruct the timing of the two telescopes and their exact position relative to each other. The use of intensity instead of amplitude also eliminates the consequences of atmospheric effects to the signal, especially the phase.

After the temporal correlation for a certain baseline is recorded with a sufficient quality the baseline is changed. More runs are taken until enough temporal correlations at different baselines are measured to reconstruct the spatial correlation.

It is crucial though to get a high time resolution and measure single photon events. Since the number of photons and the intensity are proportional to each other, actually the single photons are correlated with each other. Hanbury Brown & Twiss built the Narrabri Stellar Intensity Interferometer (NSII) consisting of two telescopes of a 6.7 m diameter mounted on a circular railroad track of 188 m diameter. With their interferometer Hanbury Brown & Twiss were able to measure the angular size of a star with an angular size as small as 0.41 ± 0.03 mas using a maximum baseline of 144.3 m [6]. After having measured the stars in the range their telescope's resolution, the project and with it astronomical intensity interferometry were abandoned. This was mainly due to the long measurement times (up to 88.0 h for one of the baselines) and the limits of the electronics at that time.

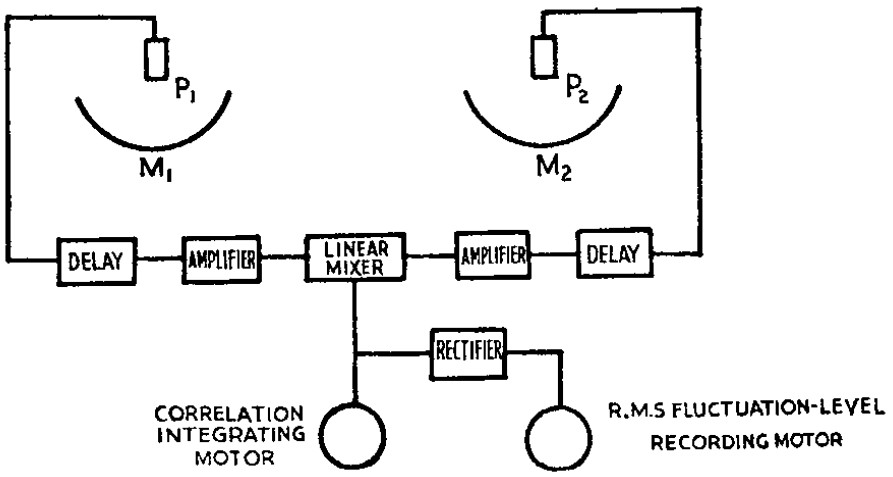


Figure 2.4: Schematics of the experimental setup used by Hanbury Brown & Twiss, taken from [3]

Instead of further working with intensity interferometry astronomers stayed with amplitude interferometry which was improved greatly by adding adaptive optics to correct for atmospheric effects. Amplitude interferometry is therefore still in use today. Furthermore, the concept of very long baseline interferometry (VLBI) was developed. As shown in chapter I this method is quite capable. Instead of optically interfering light, they measure radio-waves. The main advantage of radio waves in this context is that they have a longer period and it is therefore possible to sample them. This allows for a digital measurement containing both the amplitude and the phase information. With very precise timing and spatial information the data from different telescopes can be correlated later on, resulting in information as substantial as the information measured in amplitude interferometry. The problems concerning the effects of the atmosphere or other disturbances on the signal stay the same as with conventional amplitude interferometry.

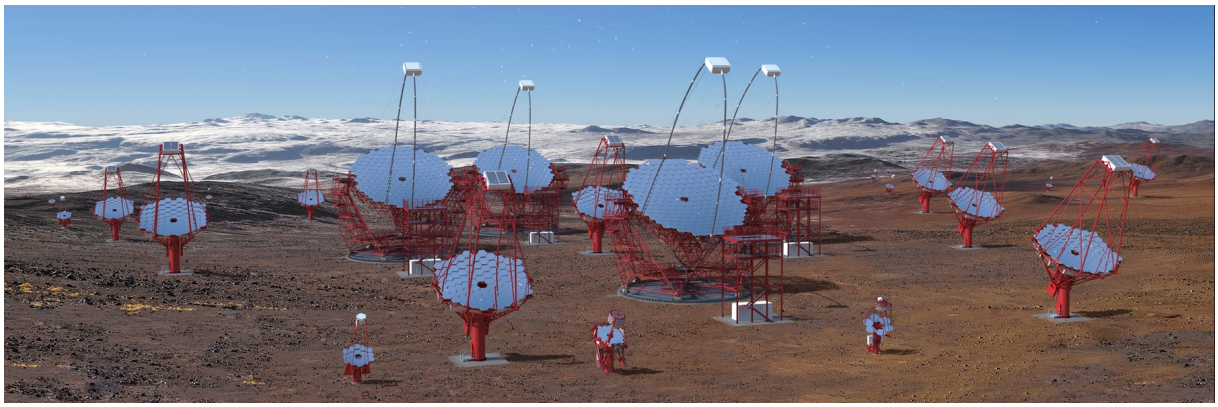


Figure 2.5: Artists impression of the CTA southern array, taken from [7]

The aim of the project currently running at the Erlangen Centre for Astroparticle Physics (ECAP) is to get a better angular resolution than the NSII by using the Cherenkov Telescope Array (CTA) to do intensity interferometry with Cherenkov telescopes. Even though the optical properties of the telescopes are far from perfect, they have the advantage over a dedicated intensity interferometer of already being available and not in use during times where II measurements can be conducted e.g. at times of bright moonlight. In addition the telescopes are already arranged in an array allowing for many different baselines, so that there is no need to move the telescopes around (compare figure 2.6). This could lead to a revival of astronomical intensity interferometry.

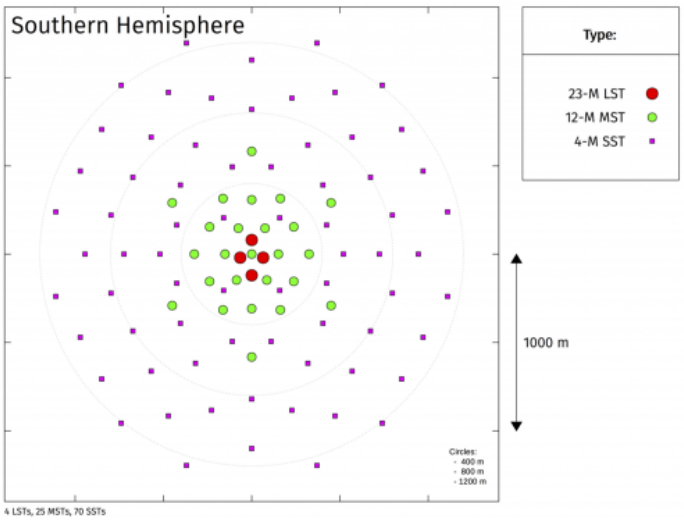


Figure 2.6: Planned layout of the CTA arrays, adapted from [8]

2.2 Physical and mathematical Background

The information about the star's extent or, in case of amplitude interferometry, intensity distribution lies in the coherence of the light. Coherence is a measure for the "stability of light" in space and time. A way to mathematically express the coherence a wave, are the correlation functions.

The $g^{(1)}$-function is defined as [9]

$$g^{(1)}(\vec{r}_1, \vec{r}_2, t, \tau) = \frac{\langle E^*(\vec{r}_1, t) \cdot E(\vec{r}_2, t+\tau)\rangle}{\sqrt{\langle E^*(\vec{r}_1, t) E(\vec{r}_1, t)\rangle \langle E^*(\vec{r}_2, t+\tau) E(\vec{r}_2, t+\tau)\rangle}} \tag{2.2}$$

with E the electric field, $\vec{r}_1$ and $\vec{r}_2$ the points of the telescopes, t the time and τ the time difference between the two signals. The * denotes the complex conjugate and $\langle\rangle$ the timely average. The $g^{(1)}$-function is calculated in VLBI. The van Cittert-Zernicke theorem [10] states that the $g^{(1)}$-function in the far field is proportional to the fourier-transformed intensity distribution $I(x, y)$ of the observed object

$$g^{(1)} \propto \mathcal{F}(I(x, y)) \tag{2.3}$$

This equation enables the VLBI astronomers to reconstruct the intensity distribution via a inverse Fourier transform thereby creating "images" of the observed source.

Intensity interferometry takes the maths one order up by measuring the $g^{(2)}$-function. This gives up the detailed information contained in the fields in turn for easier measurement and less disturbance. As the intensity is $I = |E|^2$ and the square of absolute value destroys all phase information, one can already see that also mathematically information is lost. The spatial $g^{(2)}$-function is defined as

$$g^{(2)}(\tau) = \frac{\langle I_1(t) \cdot I_2(t+\tau)\rangle}{\langle I_1(t)\rangle \langle I_2(t)\rangle} \tag{2.4}$$

with $I_1(t)$ and $I_2(t)$ being the intensities measured at the corresponding telescope, t the time and τ the time difference.

Luckily the Siegert-Relation (only valid for thermal light sources) [11] gives us a relation between the $g^{(1)}$ and the $g^{(2)}$ function.

$$g^{(2)}(\tau, \vec{r_1}, \vec{r_2}) = 1 + |g^{(1)}(\tau, \vec{r_1}, \vec{r_2})|^2 \tag{2.5}$$

As the square of absolute values of the $g^{(1)}$-function in this relation is non-reversable, it is not possible to completely reconstruct the intensity distribution. Normally the shape of the observed object is known. In the case of a star for example, it is reasonable to assume that $I(x, y)$ has the shape of a round disk with a diameter D. Using this knowledge the intensity

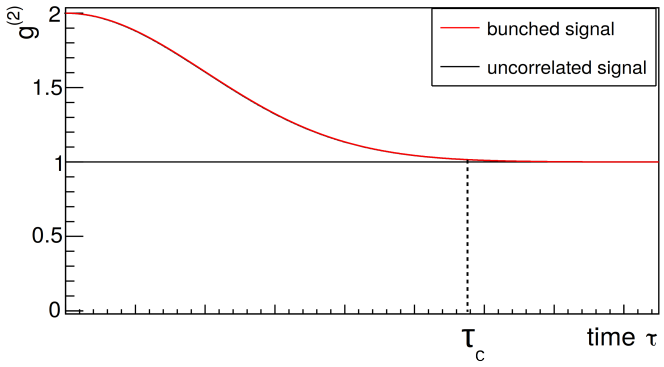


Figure 2.7: $g^{(2)}(\tau)$-function of bunched and uncorrelated signals for $d = 0$

distribution can be written as a two-dimensional Bessel function of the first kind J_1 [12] and equation 2.5 can be rewritten as

$$g^{(2)}(d) = 1 + \left| 2\frac{J_1(\pi D \nu_\mathrm{d})}{\pi D \nu_\mathrm{d}} \right|^2 \tag{2.6}$$

with $\nu_\mathrm{d} = \frac{d}{\lambda \cdot L}$, where λ is the wavelength of the incoming light and L is the distance between the observer and the star. When evaluating this, an additional factor of 1.22 is needed to calculate the first minimum of the $g^{(2)}$-function [13]. This leads to the following equation for the transversal coherence length d_c

$$d_\mathrm{c} = 1.22 \cdot \frac{L\lambda}{D} \tag{2.7}$$

The incoming photons are bunched (compare section 2.1). Therefore their temporal correlation is 2 for $\tau = 0$ and falls to 1 for increasing values of τ. The value τ_c is called coherence time (figure 2.7).

When increasing the baseline, the temporal correlation $g^{(2)}(\tau = 0)$ decreases. Evaluating $g^{(2)}$ for different baselines d at $\tau = 0$ allows to investigate the spatial correlation. These values of $g^{(2)}(\tau = 0, d)$ are plotted and equation 2.6 is fitted to the points to determine $g^{(2)}(d)$. This can be seen in figure 2.8 (note: the curve was normalized by subtraction of 1).

With the $g^{(2)}(d)$ derived from the fit the transversal coherence length d_c can be calculated. Using equation 2.7 and the small angle approximation for angular size of the star this results in an angular size Θ of the star of

$$\Theta = 1.22 \cdot \frac{\lambda}{d_\mathrm{c}} \tag{2.8}$$

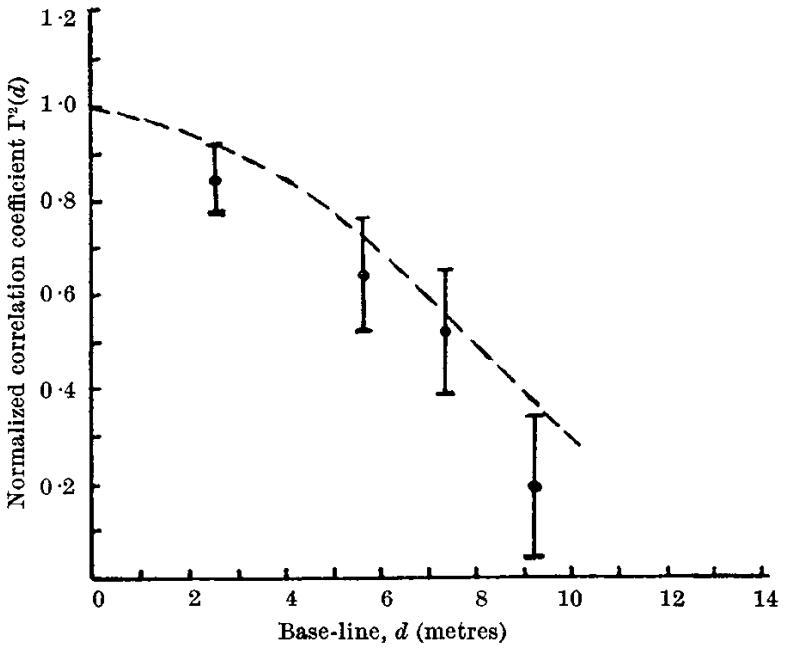


Figure 2.8: Plot by Hanbury Brown & Twiss of $g^{(2)}(d)$ of Sirius, taken from [3].

Knowing the wavelength of the observed light it is now finally possible to reconstruct the angular size Θ of the star.

This process of measuring the angular size of a star seems straight forward but there are quite some limitations. To get a good spatial $g^{(2)}$-function and therefore be able to calculate the transversal coherence length, the error on the points, calculated out of the temporal $g^{(2)}$-functions needs to be small enough for a significant fit. This requires an as good as possible resolution of the temporal $g^{(2)}$, which is limited by the coherence time τ_c (also compare figure 2.7). For interferometers using a filter with a gaussian distribution of transmission it is defined as

$$\tau_\mathrm{c} = \frac{1}{\Delta\nu} \tag{2.9}$$

with $\Delta\nu = 1.06 \cdot \nu_\mathrm{FWHM}$ where ν_FWHM is the full width half maximum of the filter's frequency distribution [11]. It can further be written as

$$\Delta\nu = \frac{c \cdot \Delta\lambda}{\lambda_0^2} \tag{2.10}$$

where c is the speed of light, $\Delta\lambda$ is the sigma of the filter's wavelength distribution and λ_0 is the central wavelength of the filter. The filters used later for the measurements have $\Delta\lambda = 2\,\mathrm{nm}$ and in the case of the experimental setup used a wavelength of $\lambda_0 = 570\,\mathrm{nm}$. This leads to a coherence time of

$$\tau_\mathrm{c} = \frac{(570\,\mathrm{nm})^2}{299792458\,\frac{\mathrm{m}}{\mathrm{s}} \cdot 2\,\mathrm{nm}} = 0.5419\,\mathrm{ps} \tag{2.11}$$

When assuming a sampling rate of 800 ps as with the used sampling card (compare section 3.1), the bin containing the significant information is more than 3 orders of magnitude greater than the coherence time. This means that when measuring the temporal $g^{(2)}$ the signal at $\tau = 0$ will not be at the theoretically expected value of 2 . Instead, assuming no further decrease due to time resolution, the height g of the correlation bin ranging from 0 ps to 800 ps is approximated to be

$$g = 1 + \frac{0.5419\,\mathrm{ps}}{800\,\mathrm{ps}} = 1.00068 \tag{2.12}$$

The small size of the first bin compared to the rest of the temporal correlation is the reason why this measurement is so difficult. All bins have to have an error so small that the correlation found in the first bin is significantly larger than the error. This can only be achieved with the statistics of long time measurements. Additionally, it requires low systematic errors in the signal or at least ones that can be corrected for afterwards.

Chapter III

Procedures of correlation measurements and analysis methods

To measure the incoming light from the light source, a hardware setup is required that allows for a high time resolution. Photomultipliers have such a high resolution. In order to measure two signals that can be correlated later on a beamsplitter is used. It is placed in the light path which divides the stream of light and leads it onto two photomultipliers.

3.1 Data Acquisition

For data acquisition a Spectrum M4i 2212-x8 Card is used. It measures the voltage on each of its 4 channels every 800 ps with a resolution of 8 bit, being decoded to an ADC-value of -128 to 127 [14]. In the used setup only two of the channels are needed, each being connected to a PMT with an amplifier. The measurements are conducted in parts with a defined length (typically 2 GS per channel). Each part is saved to a separate binary file, being numbered by the order in which they were taken. The binary file is a raw chronological stream of data, containing 1 byte of data for each channel and measurement. It is essential to know the number of channels, so each byte can be attributed to its channel.

3.2 Data Analysis

For the processing of the data the root environment is used [15]. Using the C++ programming language the needed functions for the analysis are implemented.

Before starting any analysis the task of processing the input is split into several chunks of work. To parallelize the work and thereby use the multi threading capabilities of the CPU, each file is analyzed separately by a dedicated thread. In the first step of processing the binary file is loaded to memory as a vector of 8-bit integers for each channel. To save memory not

the whole file is loaded at once but instead a sequence of a typical length of 10^5 to 10^6 bytes for each channel is used. Then the peakfinding algorithm is applied to each channel which results in two chronological lists of events. The events are saved as integer numbers, where each number corresponds to the measurement number (timestamp) of the event. The events are then passed on to the correlation function. A width of the correlation needs to be specified as well. For each correlation (auto-correlations channel 0 with channel 0 and channel 1 with channel 1 and cross-correlations channel 0 with 1 and channel 1 with channel 0) the function initializes a vector, with the length of the vector being the width of the correlation and the value of each element being zero. Afterwards all events from both channels are merged together and put in chronological order. To preserve the information in which channel the event occurred this information is stored separately for each event. After putting the events in the right order, the function goes through the events and compares each event to the ones it is followed by until the difference between them exceeds the length specified at call of the function. The so found time differences are then used to increase the integer in the element corresponding to the time difference in the appropriate correlation vector. The algorithm moves on to the next timestamp and repeats this process until it reaches the End Of File. To plot the correlations, the arrays from the different threads, are added up and the result is given to a plotting function, which plots the correlations as histograms. To get an appropriate resolution all time differences within a certain interval are grouped together and put into one bin of the histogram. The histograms are then saved to disk as a ".root"-file.

3.3 Peakfinding

To find the peaks in each channel (compare section 3.2) a peakfinding algorithm needs to be used. The algorithm has to be both efficient as well as precise. Its timely precision is limited by the sampling time of the card (800 ps, compare [14]). As the single-photon peaks fall steeply at their first flank and then depending on various parameters rise again in a neither predictable nor for this measurement important manner, it is favourable to trigger on the falling flank. To differentiate between a photon-peak and noise a threshold needs to be set. The implemented peakfinding algorithm works as follows (compare figure 3.1): It goes through the chronologically ordered ADC values one entry at a time (phase 1). Once it reaches an entry where the ADC value undershoots the threshold value, the entry number (which is equivalent to the timestamp) is added at the end of the event vector. Furthermore, a flag is set. As long as the flag is set no events are saved to the events vector to avoid multiple detections of the same event (phase 2). Once the ADC exceeds the threshold value again the flag is removed and the detection of new peaks is possible (phase 3).

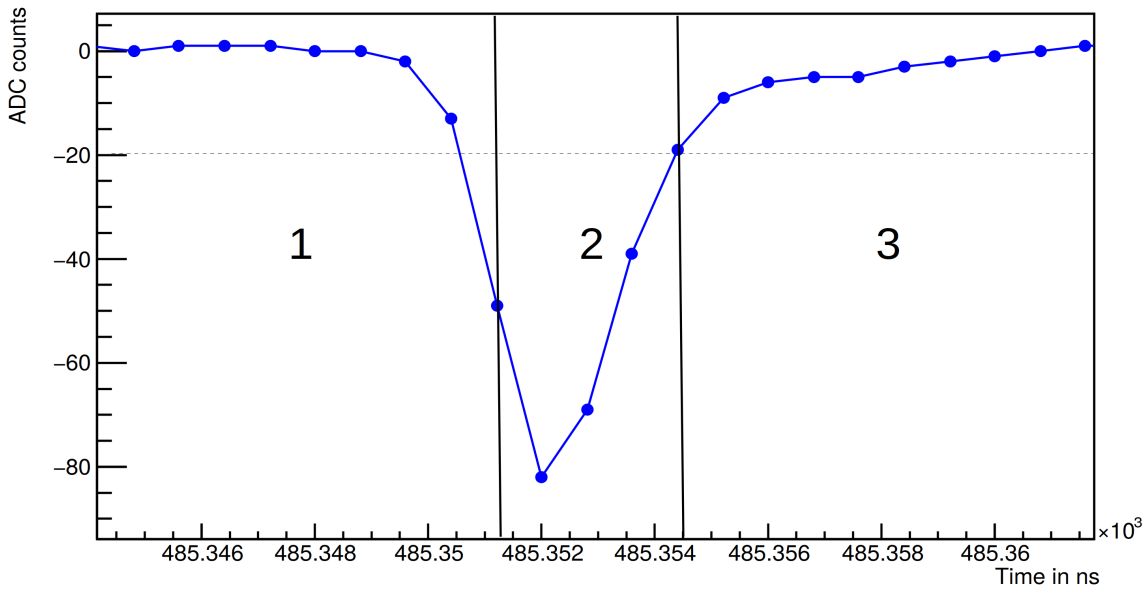


Figure 3.1: Illustration of the peakfinding method.

3.4 Current Correlation

Another method of analyzing the data is to use current correlation. Instead of finding single peaks and correlating those, the whole signal coming from the card is correlated. When starting the current correlation both channels are loaded into memory. Then all values over a certain threshold are set to zero to suppress noise. The remaining vectors now only contain the ADC values of the ranges where peaks are suspected. To calculate the values for the correlation $g^{(2)}(\tau)$ the vectors are shifted by τ for each τ and each value of channel one is multiplied by the corresponding value on channel two. These products are now added up for each τ giving the value of $g^{(2)}$.

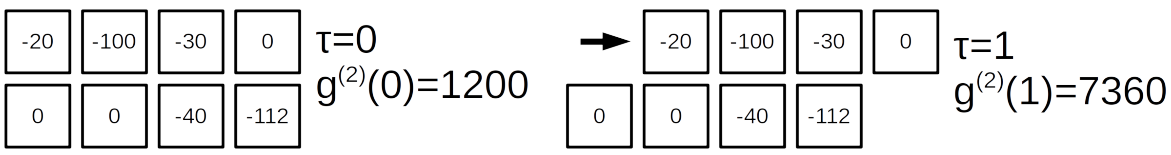


Figure 3.2: Illustration of the Current Correlation Method

Chapter IV

Comparison of Peakfinding and Current Correlation

4.1 Lab setup

To compare peakfinding and current correlation and find the advantages and disadvantages of each method in analyzing waveforms, some measurements were conducted. For data generation a setup that splits an incoming lightbeam and lets it pass into two PMTs was chosen.

In figure 4.1 the schematics of the lab setup are shown. A laser was used to generate the signal. As a high correlation of the signals on both PMTs was desired, a pulsed laser (PILAS PIL040-XSM which is a customized version of the PIL040-FC laser [16]) with a repetition rate of 1 MHz and a wavelength of 404 nm was used. The pulse width of each pulse is below 20 ps so the pulses are quite sharp. The intensities of the laser pulses can be varied as desired. The laser controller does not allow for a direct adjustment of the intensity though. Instead it uses a parameter called tune (d_{tune}) ranging from $100\,\%$ to $0\,\%$, where $d_{\text{tune}} = 0\,\%$ corresponds to maximum intensity and $d_{\text{tune}} = 100\,\%$ to minimum intensity of the pulse.

The laser light is then fed through a fiber and passed into a tube that is connected to a beamsplitter. The beamsplitter lets part of the light pass to PMT1 and deflects the other part to PMT2. PMT1 is a Hamamatsu R12992-100 and PMT2 is a Hamamatsu R11920-100. The quantum efficiency of the PMTs at the lasers wavelength of 404 nm is $\approx 40\,\%$. This means that only about $40\,\%$ of the incoming photons are translated into an electrical signal and therefore "seen" by the used system. The voltage of operation is set at 1398 V for PMT1 and at 1200 V for PMT2.

The signals from the PMTs are now both amplified by a factor of -200 using two ORTEC VT120 Fast Timing Preamplifiers [17]. After leaving the amplifiers the signals run through two long cables, one (PMT1) being 80 m the other one 100 m long (PMT2). The cables are that long

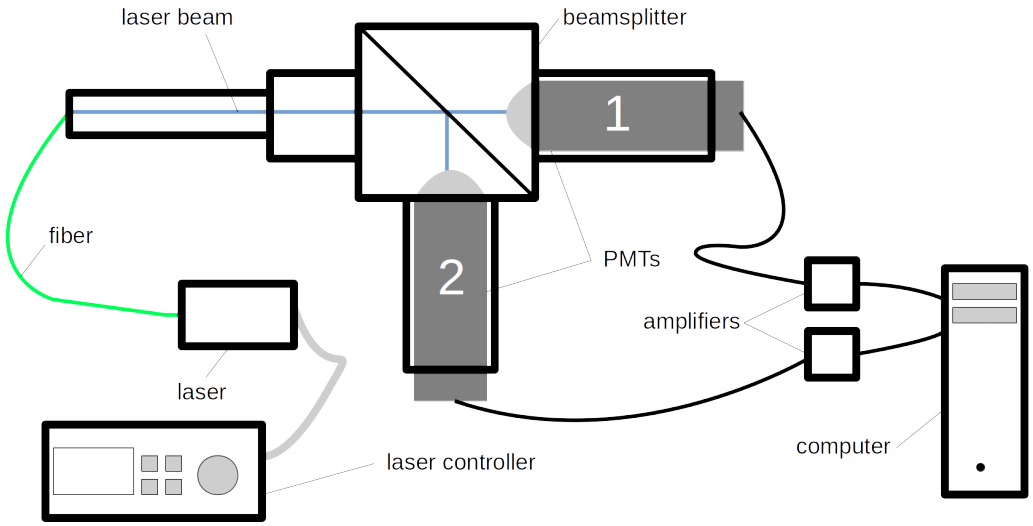


Figure 4.1: Schematics of the lab setup

on purpose to simulate the length of cables which would actually be used in operations on a telescope.

The signal is fed into channel 0 and channel 3 of the card to avoid crosstalk. Later channel 0 will be referred to as channel 1 and channel 3 as channel 2, each corresponding to the number of the PMT they sample. Once the signal is digitized it is stored in files with a file size of 4 GByte containing 2 GSamples in each of the two channels.

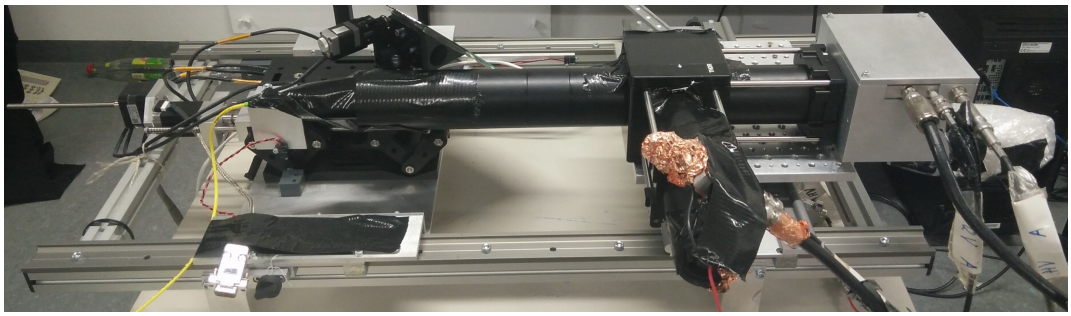


Figure 4.2: Photo of the optical part of the lab setup

4.2 Measurements

To get a broad overview over the behavior of the algorithms under different conditions, multiple runs were conducted.

To test the time resolution as well as the peak height, runs 1 to 5 were conducted. For these runs a laser as described in section 4.1 was used. The laser's tune (d_{tune}) was varied for the different runs which did not only alter the intensity of the peak in the classical view, but also varied the number of photons that are expected for each pulse. This allows for a later analysis of the software's behaviour in relation to single- and multi-photon-peaks.

Additionally, a run was taken in which the laser was replaced by an LED (run 6). This was done to test the algorithms in respect to the noise levels in the correlations. An LED, situated close enough to the beamsplitter/PMTs, allows for such measurements. As its angular size is quite large as seen by the PMTs, no bunching signal is detected, leaving a signal with only noise behavior to analyze. To accomplish this the LED was situated at the same position as the laser and an interference filter with a gaussian wavelength spectrum centered around 570 nm and $\sigma = 2\,\mathrm{nm}$ was added in between the beamsplitter and the LED. As the peak height was of lesser importance than in the laser measurements, but instead a high rate was desired, the voltage of PMT1 was increased to 1500 V.

Run #	tune d_{tune}	length of run (files)	length of run (s)
1	92.8 %	50	85.9
2	82.0 %	50	85.9
3	72.0 %	50	85.9
4	62.0 %	49	84.2
5	52.0 %	50	85.9
6	LED	1354	2326.2

Table 4.1: Conducted runs

4.3 LED measurement

T0 investigate the noise level in the output of the two algorithms, a run with an LED was taken (run 6). The LED was placed in front of the beamsplitter in such proximity that no bunching signal is expected at the PMTs. Therefore no signal but only noise is expected to be measured.

4.3.1 Waveforms and threshold

The data taken by the card, can be plotted as ADC value over time. This shows the waveforms of the PMT signals. To inspect the raw signal of this measurement, some waveforms from run 6 were plotted (figure 4.3).

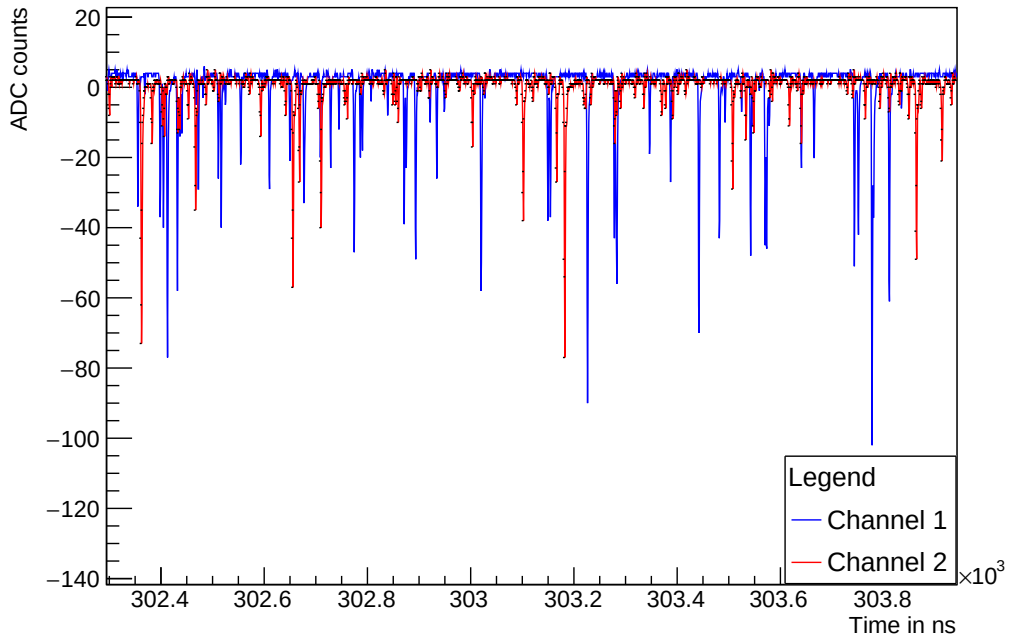


Figure 4.3: Signal heights of the $g^{(2)}$-functions.

One can see that in the waveforms there are no obvious patterns. Instead the position of the photon peaks, as well as there height, are distributed seemingly at random.

To determine an appropriate threshold value for both algorithms, the waveforms are inspected with a higher detail (figure 4.4).

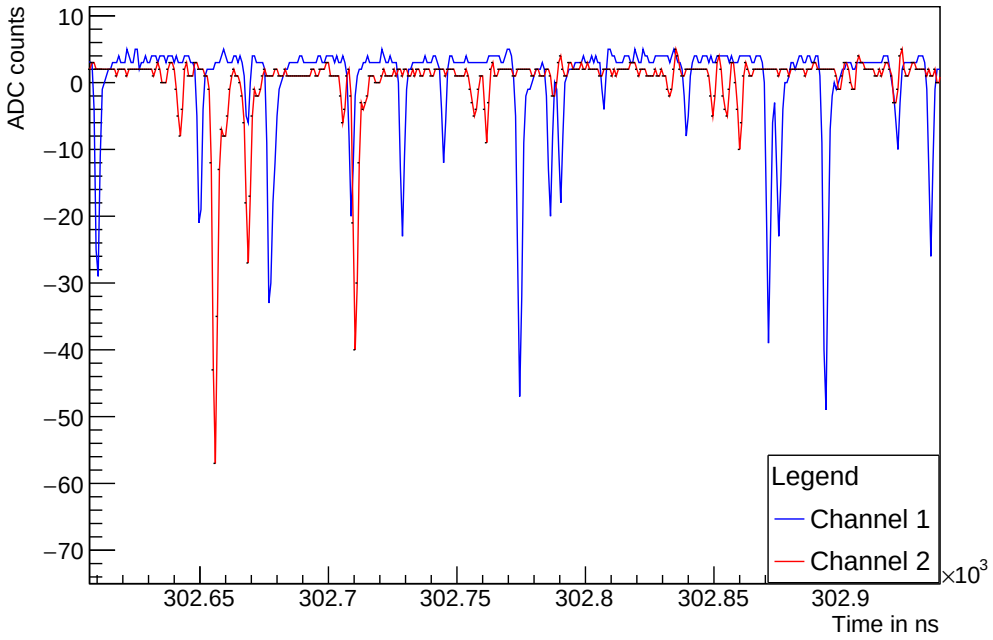


Figure 4.4: Signal heights of the $g^{(2)}$-functions.

The goal of the threshold is to suppress all noise in the original signal so that only actual photon peaks are correlated. To achieve this, a threshold of -5 is set, as the noise level is only around 1 to 3 ADC counts high. While peakfinding uses this value as a trigger for events, current correlation sets all ADC values greater than this value to zero.

One can notice furthermore, that the peaks in channel 2 are sometimes smaller than the ones in channel 1. This is most likely due to a saturation of PMT2, leading to a lower gain and therefore resulting in lower peaks.

4.3.2 Noise Level investigations

Both the peakfinding and the current correlation program are executed for run 6, using the threshold of -5 as determined in subsection 4.3.1. The resulting correlations are then plotted.

As one can see in figure 4.5, the $g^{(2)}$-functions are fluctuating around the value of 1. This is as expected, since the setup is designed in such a way, that no bunching should be visible to the PMTs.

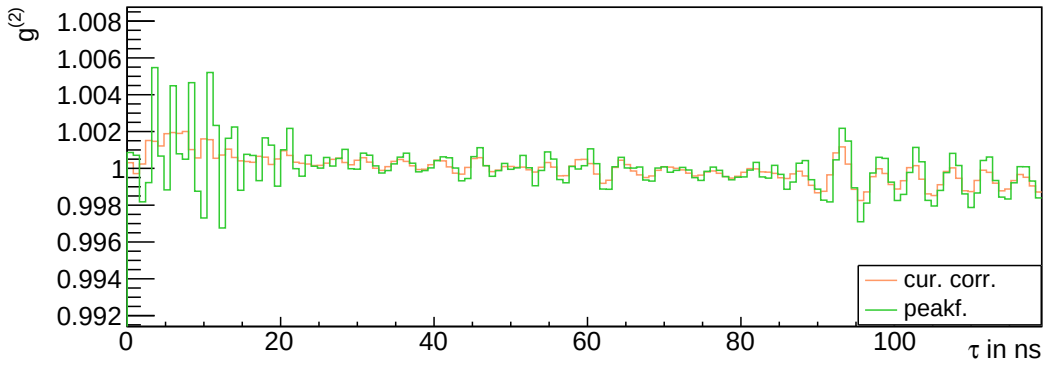


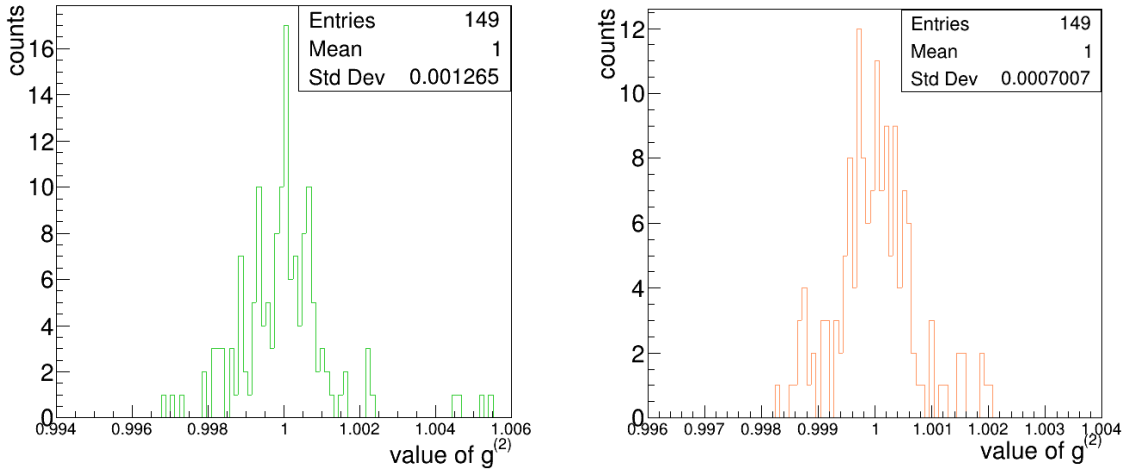
Figure 4.5: $g^{(2)}$-functions of run 6 around the value of 1.

The $g^{(2)}$-functions, as found by the two algorithms, look pretty much alike. However the peakfinding graph shows stronger fluctuations, especially in the range around $\tau = 0$. It is also quite clear that there are periodic oscillations in the signal, which can be assumed to be a systematic error.

For the statistic noise a Poisson distribution is expected. Knowing that each bin of the $g^{(2)}$-function was generated out of an average of $1.461 \cdot 10^9$ photon correlation counts, the expected relative noise σ_{rel} for the $g^{(2)}$ can be calculated

$$\sigma_{\text{rel}} = \frac{\sqrt{1.461 \cdot 10^9}}{1.461 \cdot 10^9} = 2.62 \cdot 10^{-5} \tag{4.1}$$

To compare this calculated σ_{rel} to the real one, the actual distribution of the $g^{(2)}$-function's height is plotted for both peakfinding and current correlation (compare figure 4.6),



(a) Peakfinding

(b) Current correlation

Figure 4.6: Distribution of the $g^{(2)}$-function's values of figure 4.5

The measured standard deviations $\sigma_{\text{meas peak}} = 1.265 \cdot 10^{-3}$ and $\sigma_{\text{meas curr}} = 7.007 \cdot 10^{-4}$ are two orders of magnitude above the expected of $\sigma_{\text{rel}} = 2.62 \cdot 10^{-5}$. This means that most noise in the correlation is likely to be systematic noise.

To separate this systematic noise from the statistic noise a trick is used: Instead of correlating all the files in the measurement, two correlations are calculated. One contains all files with an even file number, the other one the ones with an odd file number. By taking every other file one tries to eliminate the time-stable systematics as they should equally be present in both correlations, whereas the statistics are different and should therefore remain. This was performed for both the peakfinding and the current correlation method. The resulting $\Delta g^{(2)}$ is shown in figure 4.7.

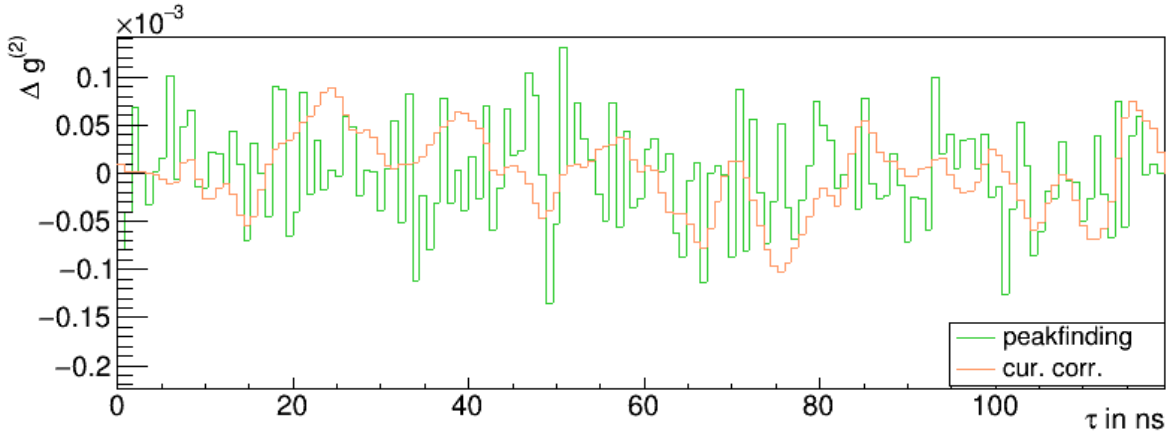
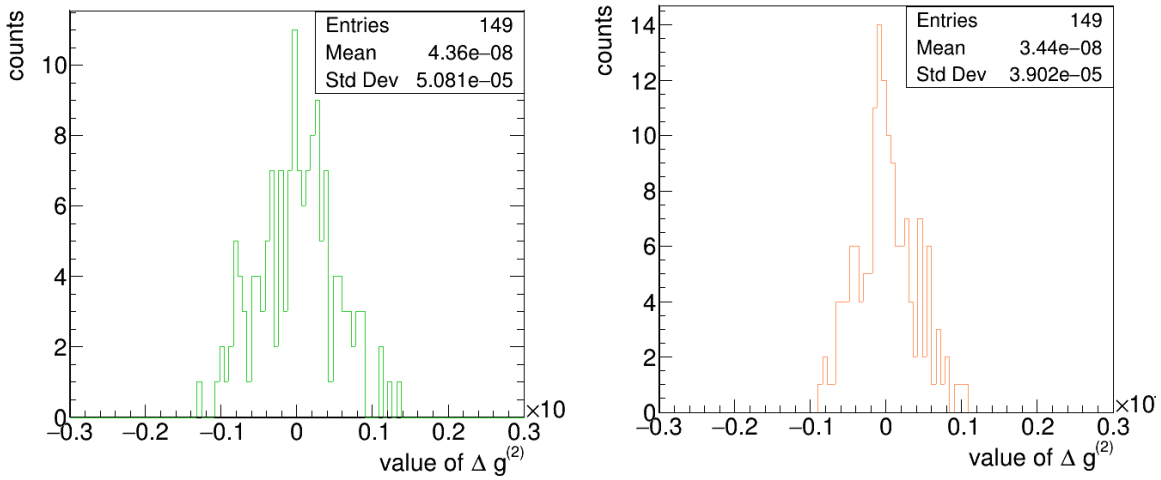


Figure 4.7: $\Delta g^{(2)}$-functions of run 6

It is still possible to see oscillations in both $\Delta g^{(2)}$-functions, even though both the systematic and the overall noise are decreased. Especially for small values of τ the noise in the peakfinding curve is reduced by the subtraction. Hence there have to be systematic characteristic to the peakfinding algorithm which greatly worsen the signal quality in this range.



(a) Peakfinding

(b) Current correlation

Figure 4.8: Distribution of the $g^{(2)}$-function's values of figure 4.7

The distribution of the $\Delta g^{(2)}$-functions' heights are plotted and their standard deviations σ_{red} are determined. As at least part of the systematic noise was eliminated, the standard deviation dropped from $\sigma_{\text{meas peak}} = 1.265 \cdot 10^{-3}$ to $\sigma_{\text{red peak}} = 5.081 \cdot 10^{-5} \pm 2.943 \cdot 10^{-6}$ for the peakfinding

and from $\sigma_{\text{meas curr}} = 7.007 \cdot 10^{-4}$ to $\sigma_{\text{red curr}} = 3.902 \cdot 10^{-5} \pm 2.261 \cdot 10^{-6}$ for the current correlation.

Since the correlations, which $\Delta g^{(2)}$ was calculated from, each contained only half as many counts as the original $g^{(2)}$-functions, a factor of $\sqrt{2}$ has to be multiplied to the expected noise. Additionally the process of subtracting the two correlations from each other results in an error propagation which adds another factor of $\sqrt{2}$. Therefore

$$\sigma_{\text{rel red}} = \sqrt{2} \cdot \sqrt{2} \cdot \sigma_{\text{rel}} = 5.24 \cdot 10^{-5} \tag{4.2}$$

can be calculated. The value of $\sigma_{\text{rel red}}$ is within the error range of $\sigma_{\text{red peak}}$. This means that the estimation of the statistical error fits quite well for the peakfinding algorithm. The current correlation's $\sigma_{\text{red cur}}$ is even smaller. This can be explained by the fact that a number of counts can not be determined for the peakfinding algorithm. Therefore a simple approximation as in the case of peakfinding is not possible. The current correlation program weighs multi photon peaks, which are seen as one peak by the peakfinding, higher. This can be assumed to lead to better statistics as seen in this measurement.

Overall the noise in the obtained correlations is quite high and shows systematics that seem to make a productive measurement pretty hard. Since much of the noise is seen in correlations from both the peakfinding and the current correlation, the systematic noise seems to be at least in big parts due to the experimental setup. The procedure of subtracting the correlations from each other would also remove the signal from the correlation. This means it is not a feasible way to improve the signal to noise ratio either. However, a reference measurement without a signal could be taken. When subtracting this reference run from the run with the signal, only the signal should be left, as the systematic noise is contained in both measurements and is therefore eliminated by the subtraction. When comparing the two programs with each other it is possible to state that the current correlation results in a more steady signal with a lower level of fluctuations

4.4 Laser measurement

4.4.1 Waveforms

To analyze the signal from both channels manually some waveforms from run 1 are plotted.

It is easy to notice on first sight, that the distances between the peaks are quite regular with a distance of about 1 µs. This confirms that the laser works well with its repetition rate of 1 MHz.

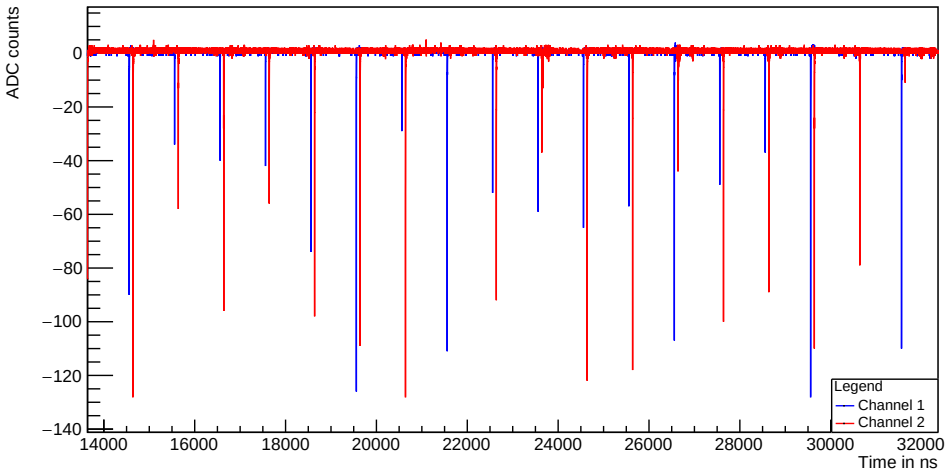


Figure 4.9: Example of equally spaced peaks in waveforms

When zooming in further, one can see the shape of the actual peaks. Once a photon hits a PMT an electron shower is triggered. This results in the steep rise of the current. After reaching its peak the current returns to the baseline with a somewhat flatter flank.

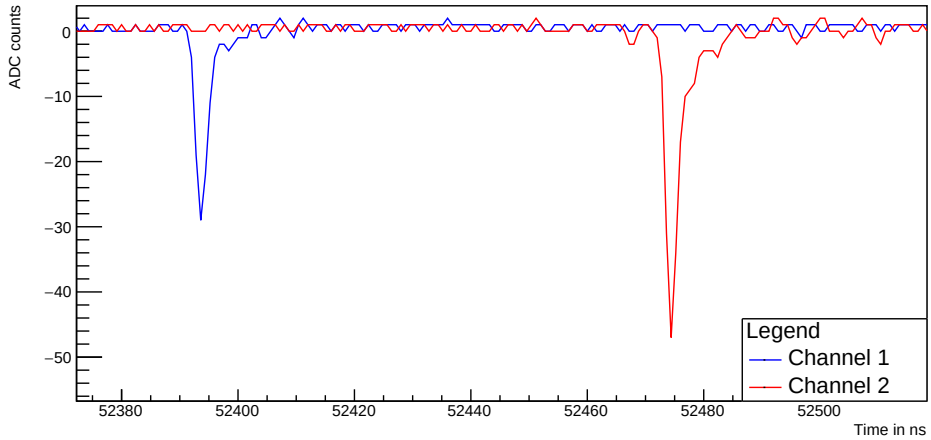


Figure 4.10: Example of peaks in waveforms

In figure 4.10 the time difference between the peak in channel 1 and the one in channel 2 of $\approx 80\,\text{ns}$ can be observed. Even though they originate from the same laser pulse, there is a delay

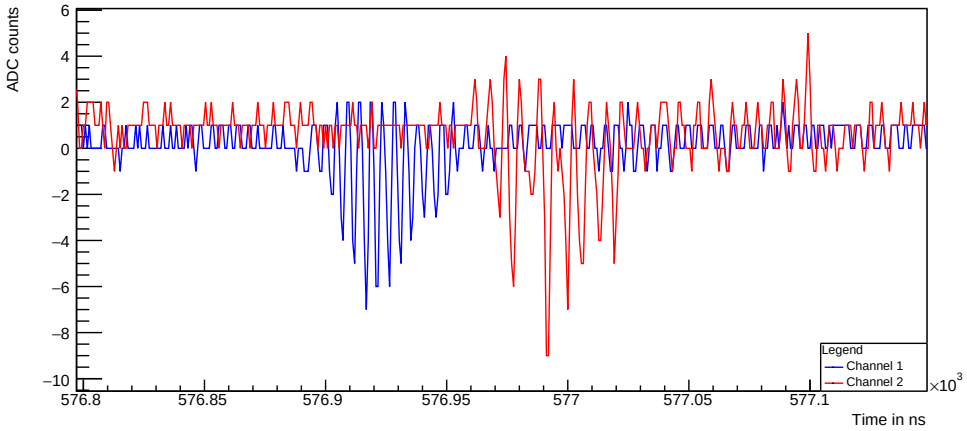


Figure 4.11: Example of noise in waveforms

between these two peaks as measured by the card. This is due to the longer traveling time in the cable between PMT 2 and the card. The expected delay Δt can be calculated to be

$$\Delta t = \frac{\Delta l}{f_{\text{velocity}} \cdot c} = 78.4\,\text{ns} \tag{4.3}$$

with $\Delta l = 100\,\text{m} - 80\,\text{m} = 20\,\text{m}$ the difference in cable length, $f_{\text{velocity}} = 0.85$ the velocity factor of the cable [18] and c the speed of light.

As one can see, the waveforms show a constant noise of 1 bit, which is the usual quantization error and therefore of no interest for the measurement. In addition, oscillations in the signal that appear from time to time can be observed (compare figure 4.11). They are typically only present in one channel at once, hinting that their appearance is not due to any disturbance at the card. As they do not contain any signal that can be used for the analysis, they (as well as the constant noise) should ideally be ignored by the algorithms.

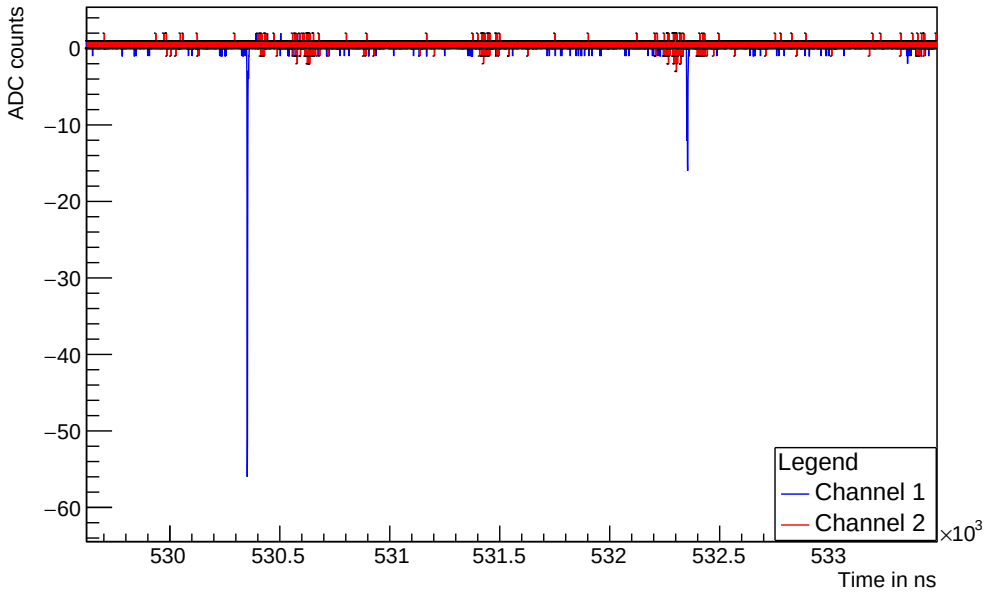


Figure 4.12: Example of different peak heights

Figure 4.12 displays the different height of the peaks. The peak height depends on the gain of the PMT and on the number of photons reaching the PMT as well as the statistic emission of electrons at the first dynode. Even though there is no deterministic way of saying how many photons hit the PMT at a given moment due to the quantum efficiency and the general behaviour of the PMT, higher peaks are more likely to be caused by a number of photons hitting the PMT at the same time. Smaller peaks are therefore considered single-photon-peaks, while larger peaks are considered to be multi-photon-peaks. The width of the peaks is fairly constant, as the duration of the electron shower is limited.

4.4.2 Peak height distribution

For the later analysis it is crucial to understand the peak height distribution of the signal. This knowledge allows not only for a good selection of the ideal threshold parameters for both peak-finding and current correlation but also for a differentiation between single and multi photon runs.

To find the peak height distribution only one file, equivalent to a timespan of 1.72 s, was analyzed for each run, as this already offers reasonable statistics and the system is quite stable in time.

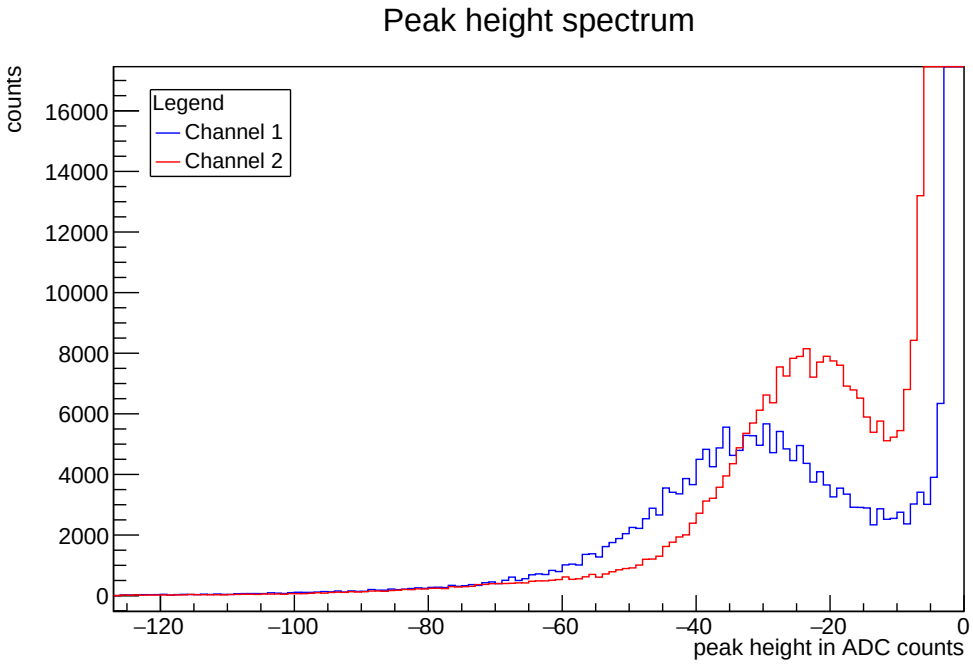
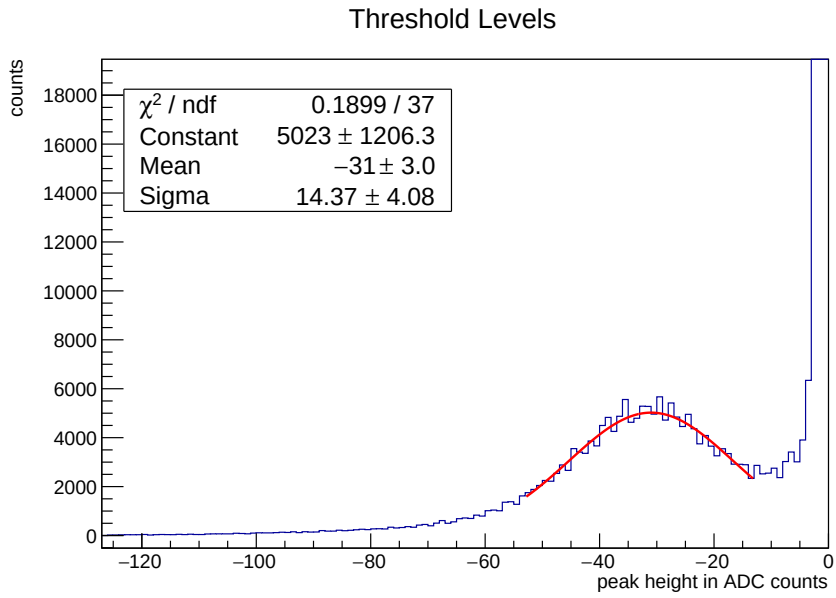
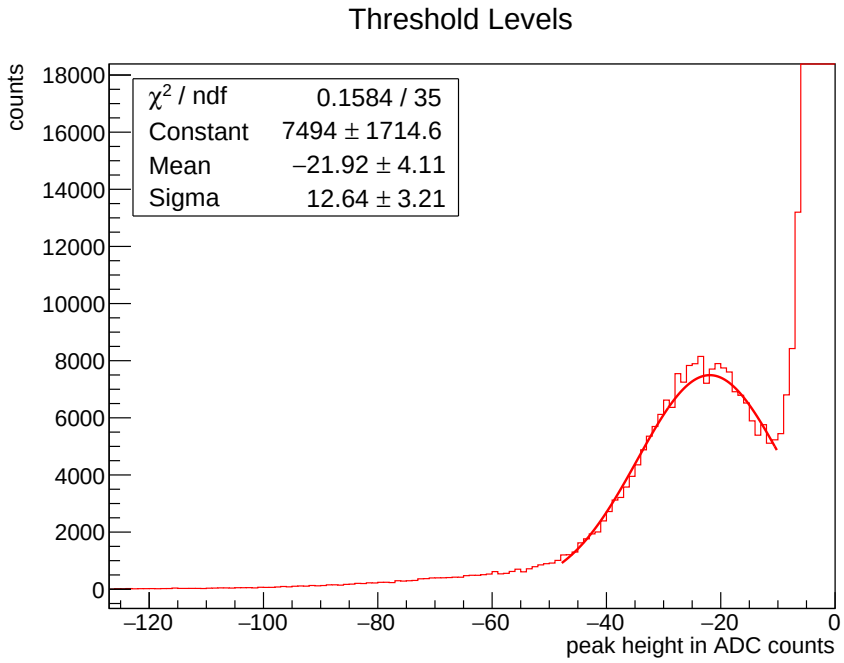


Figure 4.13: Peak height spectra of both channels in run 1

The peak height of the photon peaks for single photon events is assumed to be normally distributed, so it can be approximated by a gaussian curve (compare figure 4.14). The pedestal around 0 ADC counts is noise. As one can see, the histogram for the two channels is quite different. In channel 1 a low pedestal and therefore noise level is present, while the one on channel 2 is larger. The distribution of the photon peak of channel 1 has its maximum at -31.0 ± 3.0 ADC counts whereas channel 2 has its at -21.9 ± 4.1 ADC counts. This is due to the fact, that the two PMTs are different models and different voltages are applied to their dynodes. This leads to different gains, signal to noise ratios and rates. For a productive run on a telescope it is of course not favourable to use different PMTs with unequal properties, as the measurements should be as similar as possible to ensure maximum comparability. For the case of this thesis though, in which only methods to find correlations are tested, this is of lesser importance. The threshold needs to be set in such a way, that as little noise as possible is analyzed while at the same time not loosing to many real events. As the importance of rates and statistics are low for the here conducted analysis of time resolution and peak height, the threshold is set to a level of -20. This ensures that almost all noise is ignored by the algorithms and only actual photon peaks are used for the analysis.



(a) Channel 1



(b) Channel 2

Figure 4.14: Peak height spectra of run 1 with gaussian fits

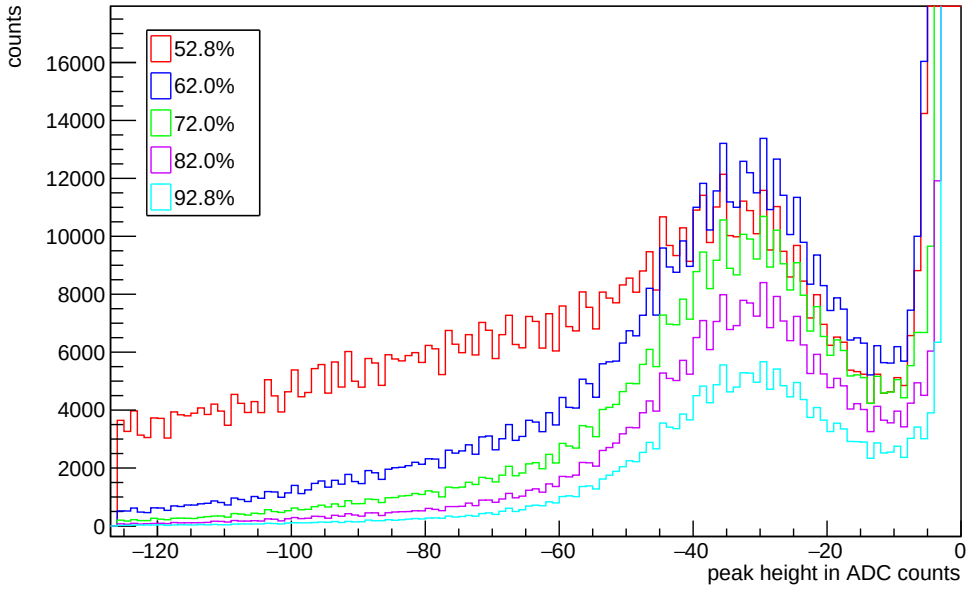


Figure 4.15: Peak height spectrum of channel 1 in different runs

After understanding the behavior of the single-photon-peak as seen in figure 4.13 one can proceed to have a look at signals with multi-photon-peaks.

In figure 4.15 runs 1 to 5 with their different tunes are plotted. One can see, that with increasing intensity from tune $92.8\,\%$ to $62.0\,\%$ both signal and noise increase. Once the tune is decreased to $52.0\,\%$ and the intensity therefore increased, the shape of the peak height distribution changes. One can see, that the peak heights left of the one photon peak increase a lot. Even though it is not possible to identify dedicated peaks in this range, it is obvious that the new shape is made up of the superposition of the multi-photon-peaks. It is therefore possible to conclude that runs 1 to 4 are dominated by single-photon-peaks, whereas run 5 shows single as well as multi-photon-peaks.

Using the threshold parameter of -20 ADC counts as found in section 4.4.2, the process described in section 3.2 is performed. The correlation calculated out of these signals can be seen in figure 4.16. It shows multiple sharp peaks, the first one at $\tau = 80\,\mathrm{ns}$. The others are found each $1\,\mu\mathrm{s}$ further. This is as expected, since the time difference due to cable length between the two channels is about $80\,\mathrm{ns}$ and the repetition rate of the laser ($1\,\mathrm{MHz}$) results in a $1\,\mu\mathrm{s}$ period between pulses. The first peak therefore is the one where photons from the same pulse are correlated, whereas the other peaks are correlations of photons from different pulses. When looking closely one can notice that the peaks slightly decrease in height. This is due to the length of the correlated sequence of the file. When τ is increased the number of datapoints

which are correlated drops linearly. Therefore also the the value of the correlation gets smaller the larger τ becomes. In order to decrease the impact of this effect, the length of the sequence needs to be increased, when calculating the correlation for larger values of τ. In figure 4.16 the peak height only decreases from 1 to 0.96, which is still an acceptable value.

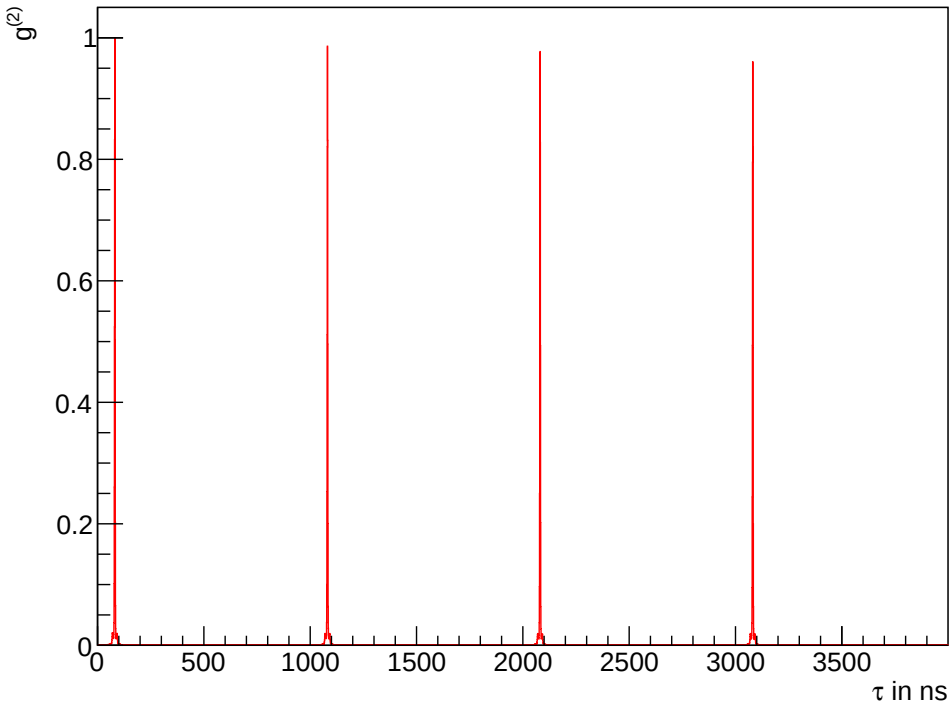


Figure 4.16: Correlation of channel 1 and 2 from run 1 calculated with peakfinding

4.4.3 Time Resolution

To investigate the time resolution of the whole interferometer the width of the peaks is measured. Only the first peak of the correlation is studied, since it is the only peak where photons from the same pulse are correlated. This eliminates possible timing errors of the laser for the time between the pulses and ensures that the limited length of the analyzed sequence is negligible. As the absolute values of the $g^{(2)}$-function determined by both algorithms are quite different, both $g^{(2)}$-functions are normed in such a way that the highest point of the curve represents the value 1.

Single-Photon-Peaks

First the correlation of run 1 is analyzed, because as found out in subsection 4.4.2 it dominantly contains single-photon-peaks.

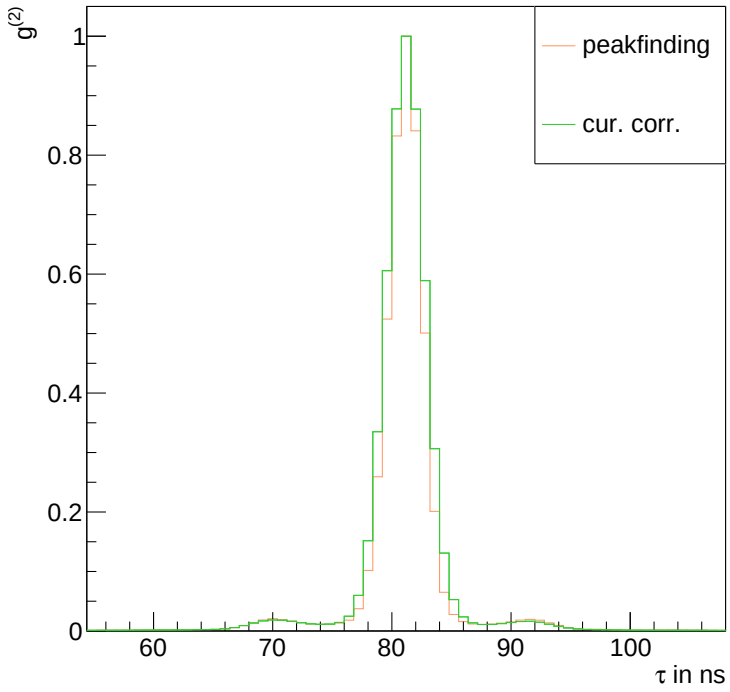


Figure 4.17: Correlation from run 1

It is clearly visible in figure 4.17 that both peaks look quite alike, but the peak found by the current correlation method is a little wider. Left and right of the main peak two little peaks can be found (figure 4.18). Their centers are $\approx 11\,\mathrm{ns}$ away from the center of the main peak, their height about $2\,\%$ the height of the main peak. For the further studies conducted here they are of no special interest, as they are most likely an flaw of either the laser or the PMTs. The time difference to the main peak is too short to be due to after pulsing which is a common effect in PMTs. One possible explanation for the these peaks is the occurrence of delayed pulses, caused by back scattering on the first anode of the PMTs, which would fit quite well with the time difference [19].

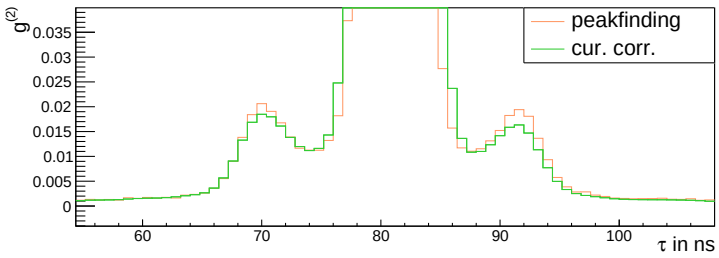
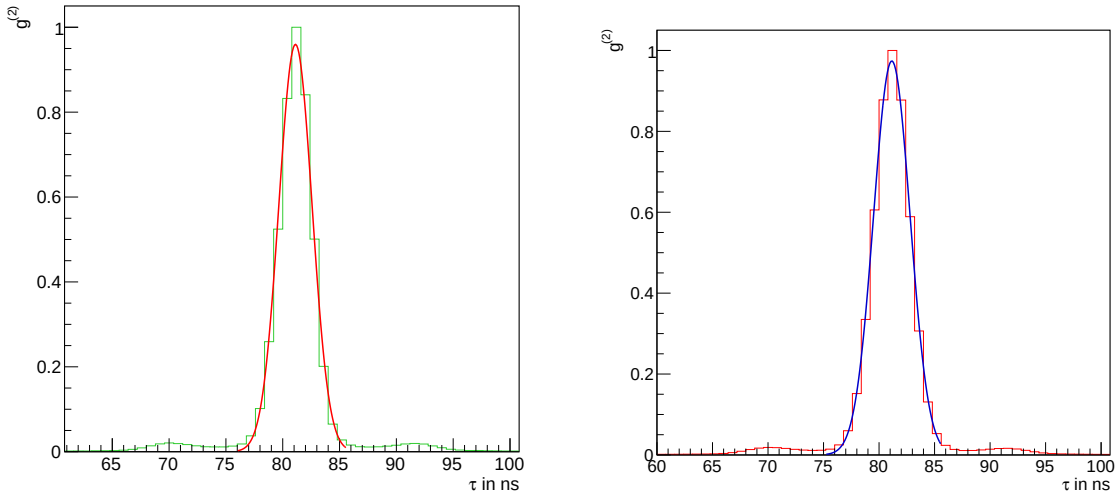


Figure 4.18: Correlation from run 1 (detailed view)

To further investigate the width of the central peak numerically, a gaussian fit (compare figure 4.19) is applied. The fitted curves have the parameters shown in table 4.2.

method	mean a	standard deviation σ	height c
cur. corr.	81.150 20(1) ns	$\sigma_{\mathrm{fc}} = 1.646\,57(1)$ ns	0.973 98(3)
peakf.	81.121(2) ns	$\sigma_{\mathrm{fp}} = 1.460(1)$ ns	0.960(1)

Table 4.2: Fit parameters



(a) peakfinding

(b) current correlation

Figure 4.19: Gaussian fits of the first correlation peak from run 1

The question arises why σ_{fp} is smaller than σ_{fc}. This can be explained by the different shapes of the signals given into the correlation process. While the signals correlated after applying the

peakfinding method can be viewed as δ-functions, the pulses in the current correlation method are of a certain width. A correlation is the equivalent of a mathematical convolution. When convolving δ-peaks the result will be a δ-peak as well:

$$(\delta(t) * \delta(t-\tau)) = \delta(\tau) \tag{4.4}$$

Due to this behavior of the convolution of δ-peaks and the nature of the input to the correlation function in the peakfinding case, the peakfinding program's output actually shows the time resolution of the system. The current correlation on the other hand has a larger input with wider peaks, which means that the input signals are not δ-peaks anymore. Instead the peaks are approximated by a gaussian fit.

Convolving two gaussian peaks $g(t) = \frac{1}{\sigma_\mathrm{g}\cdot\sqrt{2\pi}} \cdot \exp(-\frac{t^2}{2\sigma_\mathrm{g}^2})$ and $f(t) = \frac{1}{\sigma_\mathrm{f}\cdot\sqrt{2\pi}} \cdot \exp(-\frac{(t-\tau)^2}{2\sigma_\mathrm{f}^2})$ with each other results in another gaussian.

$$y(\tau) = (g(t) * f(t)) = \frac{1}{\sigma_\mathrm{y} \cdot \sqrt{2\pi}} \cdot \exp\left(-\frac{\tau^2}{2\sigma_\mathrm{y}^2}\right) \tag{4.5}$$

with $\sigma_\mathrm{y} = \sqrt{\sigma_\mathrm{g}^2 + \sigma_\mathrm{f}^2}$. This means that even if the pulses were absolutely periodical the resulting $g^{(2)}$ would not be a sharp peak but instead a gaussian peak of the calculated shape. Of course the signal is not perfectly periodical but subject to some jitter. The jitter adds another normally distributed error σ_J to the signal, which is therefore convolved again with a gaussian distribution for each channel. One can write $\sigma_\mathrm{g}^2 = \sigma_\mathrm{W1}^2 + \sigma_\mathrm{J1}^2$ respectively $\sigma_\mathrm{f}^2 = \sigma_\mathrm{W2}^2 + \sigma_\mathrm{J2}^2$ assuming that σ_W1 and σ_J1 as well as σ_W2 and σ_J2 are independent of each other. This results in the following expression for σ_fc

$$\sigma_\mathrm{fc} = \sqrt{\sigma_\mathrm{W1}^2 + \sigma_\mathrm{J1}^2 + \sigma_\mathrm{W2}^2 + \sigma_\mathrm{J2}^2} \tag{4.6}$$

where σ_W1 and σ_W2 are the standard deviations of photon peaks in channel 1 and 2, σ_J1 and σ_J2 are the jitter in the respective channel and σ_fc is the actual measured standard deviation of the gaussian peak in the correlation. As the width and jitter of the peaks are similar in both channels, this can be simplified to $\sigma_\mathrm{W1/2} = \sigma_\mathrm{W1} = \sigma_\mathrm{W2}$ and $\sigma_\mathrm{J} = \sigma_\mathrm{J1} = \sigma_\mathrm{J2}$.

Using this equation the actual time resolution for the current correlation analysis, defined as $\sigma_\mathrm{c} = \sqrt{2} \cdot \sigma_\mathrm{J}$, can be calculated

$$\sigma_\mathrm{c} = \sqrt{2}\sigma_\mathrm{J} = \sqrt{\sigma_\mathrm{fc}^2 - 2 \cdot \sigma_\mathrm{W1/2}^2} = 0.86(15)\,\mathrm{ns} \tag{4.7}$$

with $\sigma_{1/2} = 0.995(65)\,\mathrm{ns}$ fitted out of waveforms (figure 4.20). The time resolution for the peakfinding method can be calculated in the same way, leaving out the σ_W's for the peak width

$$\sigma_{\mathrm{p}} = \sqrt{2}\sigma_{\mathrm{J}} = \sqrt{\sigma_{\mathrm{fp}}^2} = \sigma_{\mathrm{fp}} = 1.460(1)\,\mathrm{ns} \tag{4.8}$$

When now comparing $\sigma_{\mathrm{p}} = 1.460(1)\,\mathrm{ns}$ and $\sigma_{\mathrm{c}} = 0.86(15)\,\mathrm{ns}$ which both claim to be equal to $\sqrt{2}\sigma_{\mathrm{J}}$, it is obvious that there has to be an error in one of the two calculations, as the signal and therefore the jitter σ_{J} in the signal is the same. The error is due to the assumption made when deriving equation 4.6 that jitter and the peak width are independent of each other. This actually is not the case and therefore this assumption produces a wrong favourable value for σ_{c}. Since the value of σ_{p} is easier and more precise to calculate, the peakfinding algorithm can be viewed as a more reliable tool to determine the time resolution.

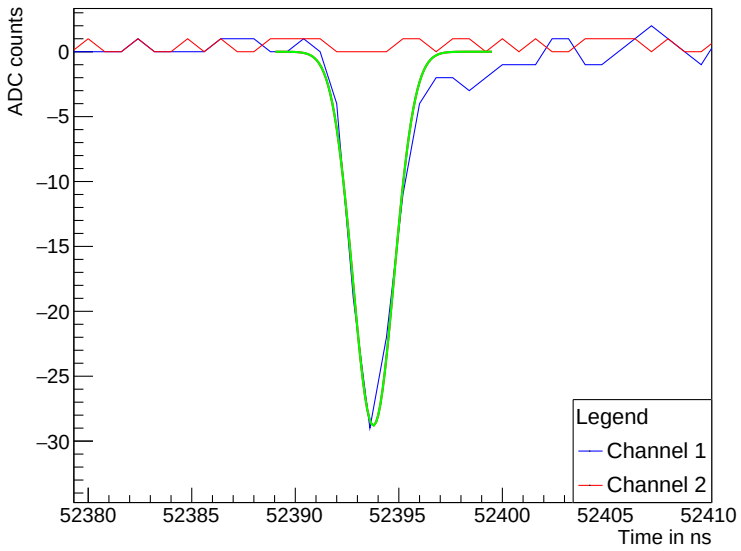


Figure 4.20: Width of photon peaks

Multi-Photon-Peaks

To analyze the behavior of both peakfinding and current correlation in the case of multi-photon-peaks, the correlation for run 5 was plotted and investigated.

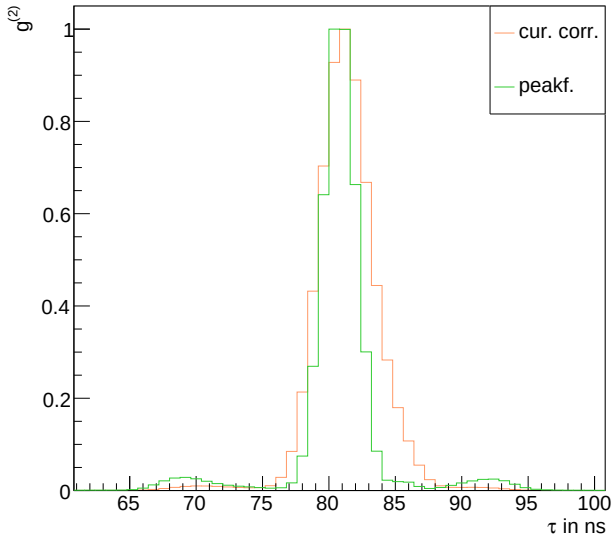
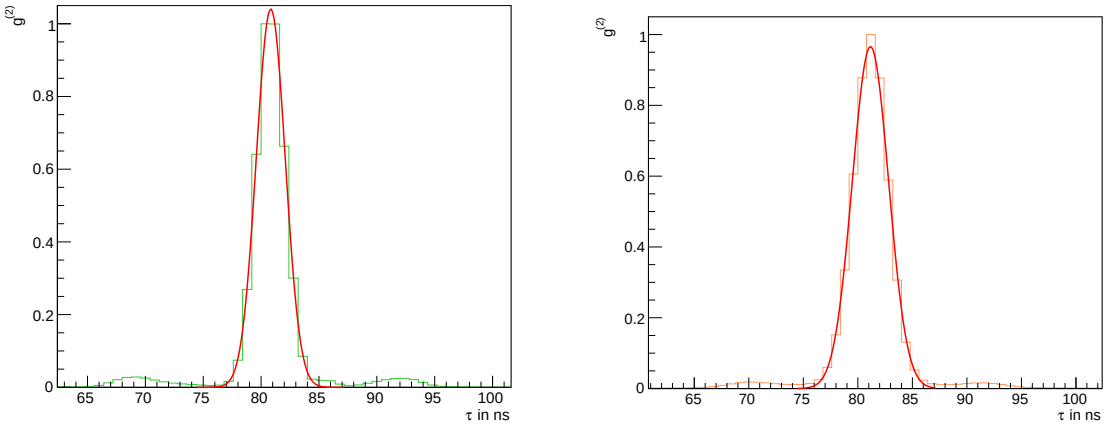


Figure 4.21: Correlation from run 5

Again gaussian functions are fitted to the peaks to determine the time resolution.



(a) peakfinding (b) current correlation

Figure 4.22: Gaussian fits of the first correlation peak from run 5

peak type	method	mean a	standard deviation σ	height c
multi	cur. corr.	$81.157\,40(1)$ ns	$1.663\,55(1)$ ns	$0.965\,88(1)$
single	cur. corr.	$81.150\,20(1)$ ns	$1.646\,57(1)$ ns	$0.973\,98(3)$
multi	peakf.	$80.8349(3)$ ns	$1.2509(2)$ ns	$1.040\,46(30)$
single	peakf.	$81.121(2)$ ns	$1.460(1)$ ns	$0.960(1)$

Table 4.3: Fit parameters of figures 4.19 and 4.22

In table 4.3 the fit parameters for both single and multi photon measurements are shown. Interestingly current correlation only shows negligible changes in its parameters when changing from the single to the multi peak run. The peakfinding on the other hand has quite an improvement in the time resolution and a slight shift in the position of the peak. This can be explained when considering the waveforms.

Multi-photon-peaks rise a lot steeper than single-photon-peaks. This leads to an improved time resolution in the peakfinding algorithm, as the photons are detected closer to their actual arrival at the PMT. Since the peakfinding algorithm only looks at the flank, this effect is quite significant, whereas the current correlation is unaffected by this because its focus is not at the flank but at the whole width of the peak. The shift in the time difference between both channels (the mean a) is likely due to the non perfect behavior of the setup resulting in more measured photons in one channel than the other. Possible reasons for this could be an imperfect beamsplitter, not dividing the light 50:50 or different quantum efficiencies of the PMTs. These circumstances would lead to a steeper flank in the corresponding channel and therefore shorten the measured time difference. As only the peakfinding algorithm is sensitive enough to detect the changed flank, this leads to a change of the results found by the peakfinding but does not effect the current correlation.

4.4.4 Signal height

After investigating the time resolution of the different algorithms, the signal height of the correlations is examined. For this purpose the usual norm of the $g^{(2)}$ is used, where the baseline of the function is defined as 1. This allows for a comparison of the signal height. The signal height is quite important to distinguish the signal from both the background and the noise as described section in 2.2.

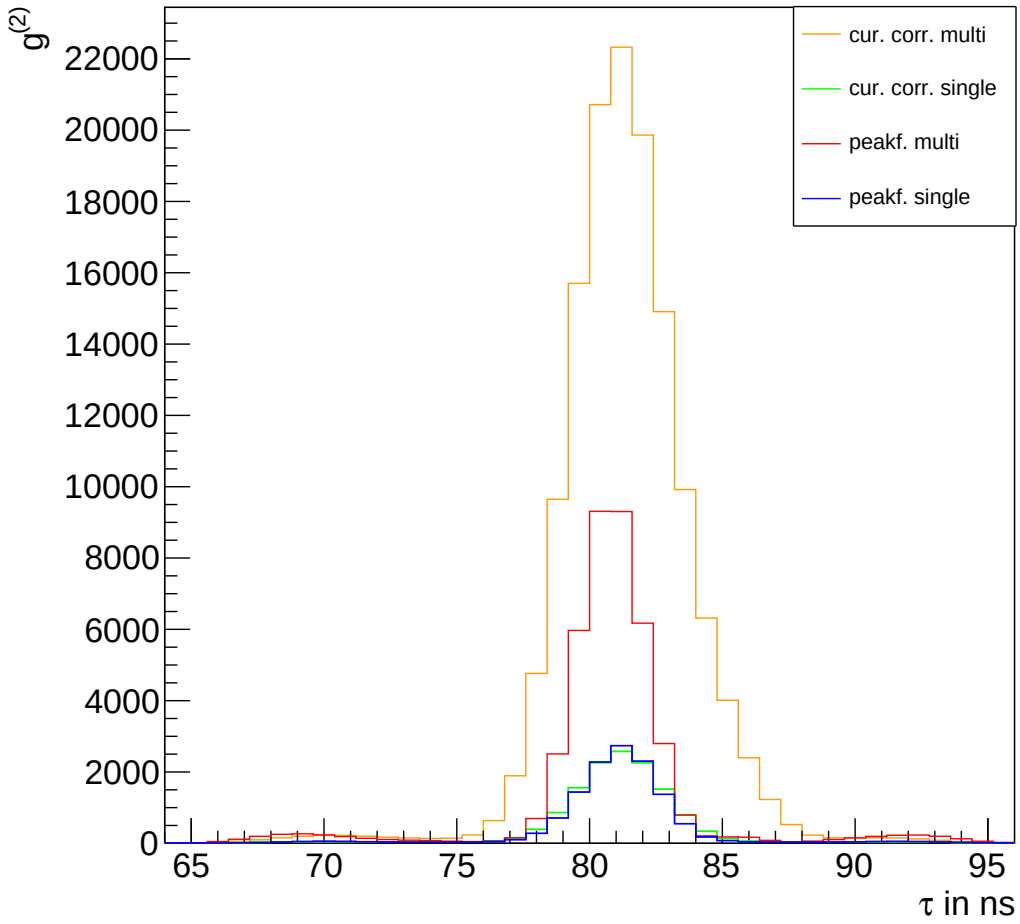


Figure 4.23: Signal heights of the $g^{(2)}$-functions.

Figure 4.23 shows that for the single photon run both algorithms exhibit a similar signal height. When moving on to the multi photon run both of the signals increase, but the signal found by the current correlation increases more than the other one. The similar height for the single photon run can be explained by the fact that all peaks are of a similar height and therefore there is no additional information contained in the height of the peaks. In the multi photon run on the other hand, the peak height does matter. The higher peaks are likely to contain multiple photons and therefore are of a higher significance for the correlation. The current correlation algorithm accounts for this, as it correlates the actual values of the waveforms and therefore weights higher peaks more than lower ones. The peakfinding only looks at the falling flank to

detect the photon but does not care for the height of the pulse following the flank. Therefore the height information is lost. The increase of the peakfinding correlation's peak when going from run 1 to run 5 is therefore only caused by the increased rate of the multi photon run (compare table 4.5).

type	method	peak height h
multi	cur. corr.	22326.3
single	cur. corr.	2576.1
multi	peakf.	9309.7
single	peakf.	2738.9

Table 4.4: Peak heights of the graphs shown in figure 4.23

As the signal height is of high importance for the quality of the temporal correlation and therefore for the measuring time, the current correlation algorithm outshines the peakfinding in this discipline.

type	rate ch1	rate ch2
multi	869 kHz	897 kHz
single	157 kHz	159 kHz

Table 4.5: Rates of the runs shown in figure 4.23

When going out of the lab towards Cherenkov telescopes, much higher rates of about 200 MHz are expected. It is also known that on these telescopes multi-photon-peaks occur quite often. This leads to an higher adequacy of the current correlation technique for the case of astronomic intensity interferometry when considering the height parameter of the correlation.

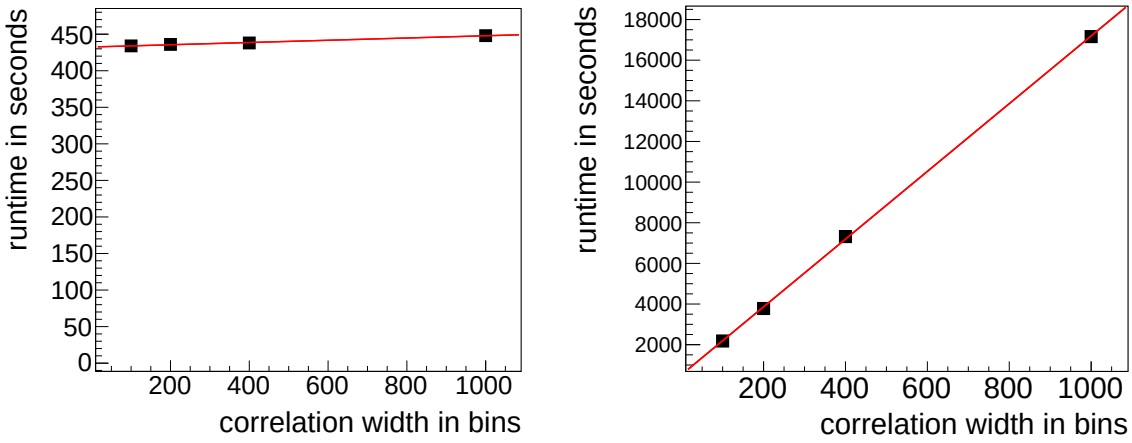
4.5 Computation Runtimes

As resources are limited, runtimes of the programs are quite interesting when actually working with the data. To determine the runtimes' dependency on different parameters, the peakfinding and the current correlation algorithms were executed on a Windows 10 system with an Intel Core i9-7920X @2.9 GHz and 32 GB of RAM. The CPU supports up to 24 threads. As both programs are optimized to work with that many threads, both are able to use the available computing power.

The investigated parameters are correlation rage, amount of processed data and rate of the signal. An overview over the conducted computations can be obtained from table 4.9.

4.5.1 Correlation range

First the effect of the width of the correlation range on the runtime was examined. To do this the number of analyzed files was set to 50 and the same files with an average rate over both channels of 14.8 MHz were analyzed while changing the width of the correlation from computation to computation.



(a) Peakfinding

(b) Current correlation

Figure 4.24: Fits: runtimes over correlation width (computations 27 to 34)

The datapoints in figure 4.24 seem to follow a linear function and are therefore fitted with one. The fit parameters can be found in table 4.6.

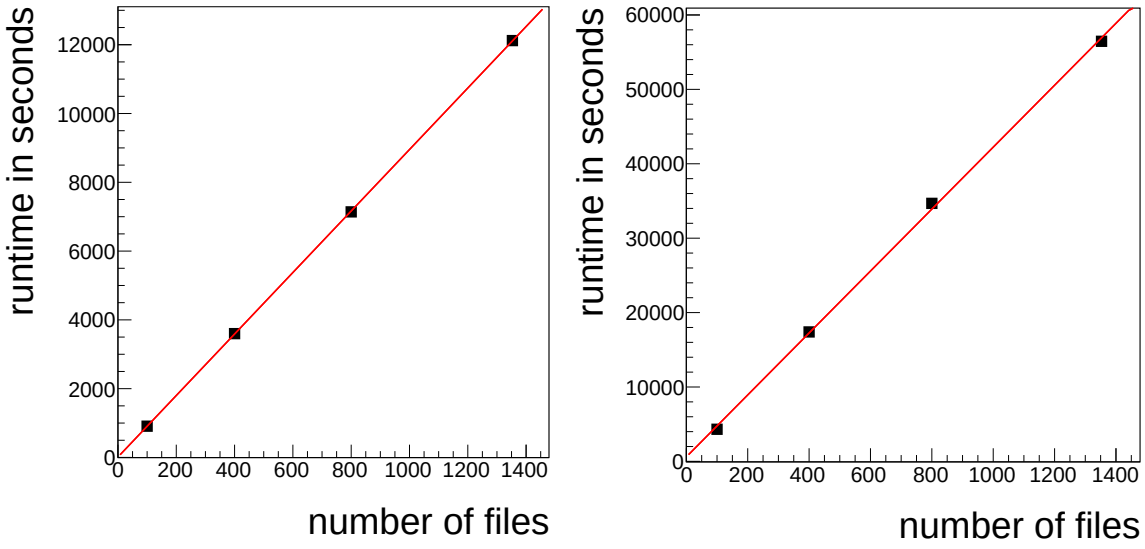
type	slope $\left(\frac{\text{s}}{\text{bin}}\right)$	offset (s)
peakf.	$s_{\text{BP}} = 0.015\,38(80)$	$c_{\text{BP}} = 432.5(4)$
cur. corr	$s_{\text{BC}} = 16.66(16)$	$c_{\text{BC}} = 527(85)$

Table 4.6: Fit parameters of figure 4.24

For both of the algorithms the fit parameters show an offset of several hundred seconds. In case of the peakfinding algorithm this is the larger part of the runtime, whereas for the current correlation this is a minor part of the needed time. For the peakfinding algorithm the offset is most likely the time which is spent on finding the peaks. In case of the current correlation it can be assumed to be the time to load the files, add up the results and write them back to disk. In both algorithms an increase in runtime can be seen for a higher widths of the correlation range. The slope in the current correlation fit is a lot steeper than the one of the peakfinding plot, which translates to a substantially longer runtime for the set parameters.

4.5.2 Number of files

Next the dependency of runtime on the amount of processed data is investigated. Since the number of files is proportional to the amount of processed data, this parameter was changed while keeping the correlation width and the average rate over both channels constant at 100 bins respectively 14.8 MHz. As the number of files can be assumed to be quite high in the experiments and the peakfinding algorithm processes the files 20 files at a time, the number of files was always set to multiples of 20.



(a) Peakfinding

(b) Current correlation

Figure 4.25: Fits: runtime over number of files (computations 19 to 26)

Again, a linear fit was applied to the datapoints to describe their behavior.

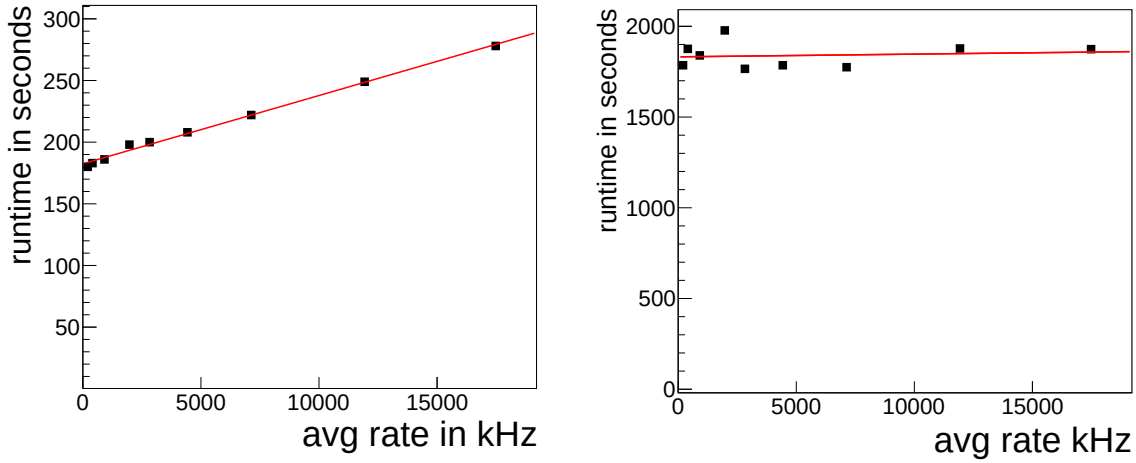
type	slope $\left(\frac{\text{s}}{\text{file}}\right)$	offset (s)
peakf.	$s_{\text{FP}} = 8.94(2)$	$c_{\text{FP}} = 12(18)$
cur. corr	$s_{\text{FC}} = 41.63(77)$	$c_{\text{FC}} = 607(623)$

Table 4.7: Fit parameters of figure 4.25

The high error on the offset values does not allow for any speculations about the nature of the offset. Both plots show that a clear linear behavior is present in the relation between the runtime and the number of files processed. It can again be seen, that the current correlation has a steeper slope than the peakfinding, resulting in higher runtimes.

4.5.3 Rate

To investigate the runtimes' dependency on the rate some more computations were performed. The width of the correlation was set to 5000 and the number of files was 20 for the peakfinding and 1 for the current correlation software, due to the long runtime of the current correlation program.



(a) Peakfinding

(b) Current correlation

Figure 4.26: Fits: runtimes over average rate (computation 1 to 18)

type	slope $\left(\frac{\text{s}}{\text{kHz}}\right)$	offset (s)
peakf.	$s_{\text{RP}} = 0.005\,53(15)$	$c_{\text{RP}} = 182.5(12)$
cur. corr	$s_{\text{RC}} = 0.0015(43)$	$c_{\text{RC}} = 1831(33)$

Table 4.8: Fit parameters of figure 4.26

A linear fit was applied to both plots (compare figure 4.26). The slope for the peakfinding algorithm is quite low but still noticeable, whereas the slope of the current correlation fit is even smaller and its error is massive. The offset in the peakfinding can again be explained by the way this algorithm processes the data. As the peakfinding itself takes a fixed amount of time, no matter how many peaks are found, this can be seen as the offset. The slope of the peakfinding algorithm can be explained by the increased time for the correlation process as more and more peaks are compared to each other when rates rise. For the current correlation the amount of data correlated is independent of the rate, as this algorithm compares the whole signals of both channels to each other no matter how many peaks are in there.

4.5.4 Empirical estimation

Having investigated the dependency of the runtime on the different parameters, it is now possible to emirically estimate the runtime. This is done by taking one representative computation and using the known dependencies to estimate a runtime with varied parameters. The fit of amount of data is approximated as a line through origin, as for larger values this is a valid approximation. The runtime of the peakfinding algorithm t_p can be estimated by

$$t_\mathrm{p} = 12\,120\,\mathrm{s} \cdot \left(\frac{99.6\,\mathrm{bins} + 0.004 \cdot W}{100\,\mathrm{bins}}\right) \cdot \left(\frac{F}{1353\,\mathrm{files}}\right) \cdot \left(\frac{10.212\,\mathrm{MHz} + 0.310 \cdot R}{14.8\,\mathrm{MHz}}\right) \quad (4.9)$$

using computation 25 as a starting point, where W is the width of the correlation in bins, F is the number of files and R is the average rate over both channels. Further, this equation can be rewritten to take the measurement time T as an input, instead of the number of files.

$$t_\mathrm{p} = 12\,120\,\mathrm{s} \cdot \left(\frac{99.6\,\mathrm{bins} + 0.004 \cdot W}{100\,\mathrm{bins}}\right) \cdot \left(\frac{T}{2324\,\mathrm{s}}\right) \cdot \left(\frac{10.2\,\mathrm{MHz} + 0.310 \cdot R}{14.8\,\mathrm{MHz}}\right) \quad (4.10)$$

Of course, a similar equation can be crafted for the estimated runtime of the current correlation algorithm t_c, using computation 26 as a starting point. As no substantial slope can be measured for the rate-dependency, the runtime is assumed not to be dependent on the rate. The fits of amount of data and number of files are approximated as lines through origin, as for larger values this is a valid approximation.

$$t_\mathrm{c} = 56\,460\,\mathrm{s} \cdot \left(\frac{W}{100\,\mathrm{bins}}\right) \cdot \left(\frac{T}{2324\,\mathrm{s}}\right) \quad (4.11)$$

It is easy to spot that the dependency of the runtime on correlation width and measurement time are equal for both algorithms. Its also quite clear that the runtime of the current correlation is more than a factor 4 larger for small rates. As rates rise the advantage in runtime of peakfinding over current correlation vanishes. It is easy to calculate the time where current correlation undershoots the peakfinding algorithm by setting $t_\mathrm{c} = t_\mathrm{c}$ and solving for R. Assuming a reasonable width of the correlation at $W = 100\,\mathrm{bins}$, this calculation results in

$$R_\mathrm{equal} = \frac{14.15\,\mathrm{MHz} - 10.2\,\mathrm{MHz}}{0.310} \cdot \frac{56\,460\,\mathrm{s}}{12\,120\,\mathrm{s}} = 59.4\,\mathrm{MHz} \quad (4.12)$$

Given the fact that the rates expected in the planned intensity interferometry setup are $\approx$ 200 MHz, the current correlation algorithm is expected to surpass the peakfinding algorithm's performance.

computation	method	time	length	files	rate ch1	rate ch2	run
1	peakf.	4 min 38 s	5000 bins	20	33.7 MHz	34.9 MHz	6
2	cur. corr.	31 min 14 s	5000 bins	1	33.7 MHz	34.9 MHz	6
3	peakf.	4 min 9 s	5000 bins	20	32.5 MHz	23.8 MHz	6
4	cur. corr.	31 min 17 s	5000 bins	1	32.5 MHz	23.8 MHz	6
5	peakf.	3 min 42 s	5000 bins	20	30.4 MHz	14.2 MHz	6
6	cur. corr.	29 min 34 s	5000 bins	1	30.4 MHz	14.2 MHz	6
7	peakf.	3 min 28 s	5000 bins	20	27.9 MHz	8.8 MHz	6
8	cur. corr.	29 min 45 s	5000 bins	1	27.9 MHz	8.8 MHz	6
9	peakf.	3 min 20 s	5000 bins	20	24.7 MHz	5.6 MHz	6
10	cur. corr.	29 min 25 s	5000 bins	1	24.7 MHz	5.6 MHz	6
11	peakf.	3 min 18 s	5000 bins	20	21.2 MHz	3.9 MHz	6
12	cur. corr.	32 min 57 s	5000 bins	1	21.2 MHz	3.9 MHz	6
13	peakf.	3 min 6 s	5000 bins	20	12.9 MHz	1.8 MHz	6
14	cur. corr.	30 min 39 s	5000 bins	1	12.9 MHz	1.8 MHz	6
15	peakf.	3 min 3 s	5000 bins	20	6 MHz	0.08 MHz	6
16	cur. corr.	31 min 15 s	5000 bins	1	6 MHz	0.08 MHz	6
17	peakf.	3 min 0 s	5000 bins	20	3.3 MHz	0.04 MHz	6
18	cur. corr.	29 min 45 s	5000 bins	1	3.3 MHz	0.04 MHz	6
19	peakf.	15 min 9 s	100 bins	100	24 MHz	5.6 MHz	6
20	curr. cor.	1 h 12 min	100 bins	100	24 MHz	5.6 MHz	6
21	peakf.	1 h 0 min	100 bins	400	24 MHz	5.6 MHz	6
22	curr. cor.	4 h 50 min	100 bins	400	24 MHz	5.6 MHz	6
23	peakf.	1 h 59 min	100 bins	800	24 MHz	5.6 MHz	6
24	curr. cor.	9 h 38 min	100 bins	800	24 MHz	5.6 MHz	6
25	peakf.	3 h 22 min	100 bins	1353	24 MHz	5.6 MHz	6
26	curr. cor.	15 h 41 min	100 bins	1353	24 MHz	5.6 MHz	6
27	peakf.	7 min 14 s	100 bins	50	24 MHz	5.6 MHz	6
28	curr. cor.	36 min 21 s	100 bins	50	24 MHz	5.6 MHz	6
29	peakf.	7 min 16 s	200 bins	50	24 MHz	5.6 MHz	6
30	curr. cor.	1 h 3 min	200 bins	50	24 MHz	5.6 MHz	6
31	peakf.	7 min 18 s	400 bins	50	24 MHz	5.6 MHz	6
32	curr. cor.	2 h 2 min	400 bins	50	24 MHz	5.6 MHz	6
33	peakf.	7 min 28 s	1000 bins	50	24 MHz	5.6 MHz	6
34	curr. cor.	4 h 46 min	1000 bins	50	24 MHz	5.6 MHz	6

Table 4.9: Runtimes of the programs

Chapter V

Conclusion and Outlook

The goal of this thesis was to examine the differences, advantages and disadvantages of current correlation and peakfinding in the context of intensity interferometry.

After a short introduction to the topic of stellar intensity interferometry, the underlying physics and maths were described. Then the experimental setup as well as the ways of data acquisition and processing were explained.

A short overview over the conducted measurements was followed by the examination of the runs taken. In the analysis of the LED measurement the noise was investigated. The correlations created by the peakfinding program did not only have a higher level of noise, but also strong systematic errors at the range around $\tau = 0$.

Advancing further to the laser measurements runs with multi and single photon dominance were distinguished by an analysis of the peak height distribution. Using this knowledge, the time resolution and peak height were investigated both for single and multi photon runs. This showed that even though the calculated resolution of the current correlation algorithm was better, the peakfinding algorithm shows its capabilities when applied to a run with many multi-photon-peaks. In the time resolution category the peakfinding still seems to be a step ahead. As the time resolution of Cherenkov Telescopes is in the range of nanoseconds, the good time resolution might be helpful in the lab but not important when actually measuring on a telescope.

This picture changes once the signal height is considered. As the peakfinding algorithm, in difference to the current correlation, does not weigh peaks by their size, the signal height does not increase in the same manner when the algorithms analyze multi instead of single photon runs. This leads to better statistics in the current correlation. Accordingly this point goes to the current correlation method.

Finally, the computation runtimes' dependencies on different parameters are investigated, completing the examination of the two algorithms. The runtimes of both algorithms are linearly dependent on the with of the correlation as well as the amount of data to analyze. Only the peakfinding algorithm's runtime is dependent on the rate of events in the file, as it only correlates the found peaks, whereas the current correlation program correlates the whole file anyway.

This gives the peakfinding an advantage in runtime over the current correlation for rates smaller than $\approx 60\,\mathrm{MHz}$, which might be interesting for test cases in the lab. When proceeding to the telescopes rates of $\approx 200\,\mathrm{MHz}$ are expected, so the current correlation will be at an advantage there as well.

The results of the conducted examination leaves the peakfinding algorithm as an interesting tool for lab measurements and for cases where a high time resolution is needed. For all other cases, especially the productive one on the telescope, the current correlation seems to win the race as the method of choice.

With the developed knowledge, the higher performance electronics of these days and the CTA array hopefully soon at the hands of astronomers, stellar intensity interferometry is gaining momentum to play an important role in the future of astronomy. Yet there are still many problems to tackle to get to a well working setup that can be rolled out on an appropriate scale.

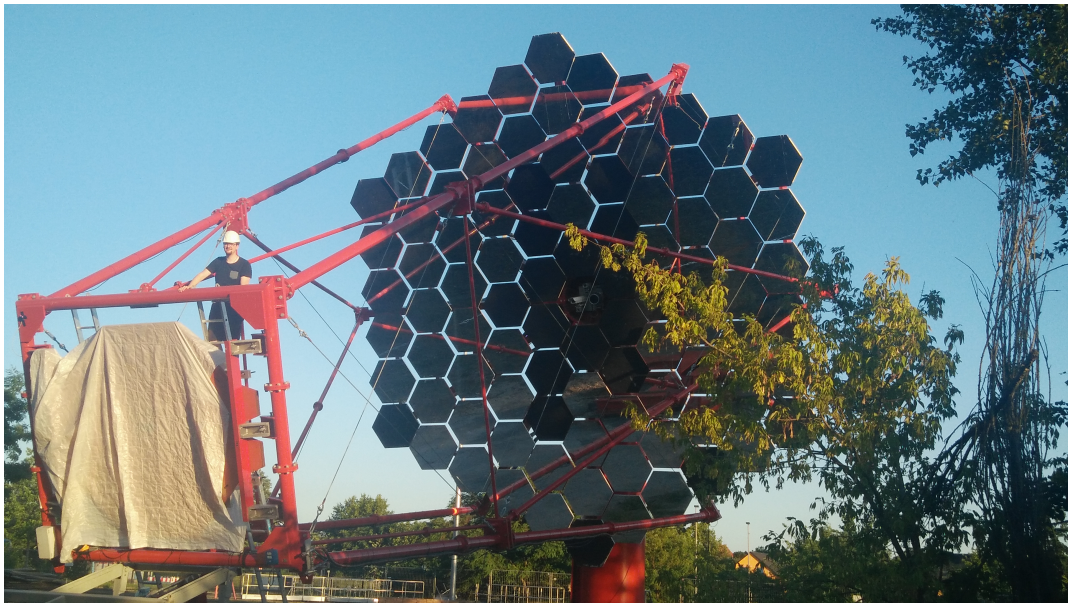


Figure 5.1: The author is sure: intensity interferometry has a bright future. Photo Credit: Andreas Zmija.

Bibliography

[1] Event Horizon Collaboration. Event horizon telescope. Online, accessed 2019-08-06, https://eventhorizontelescope.org/.

[2] Wolfgang Demtröder. *Experimentalphysik 4*. Springer Spektrum, 2014.

[3] R Hanbury Brown and RQ Twiss. A test of a new type of stellar interferometer on sirius. *Nature*, 178(4541):1046–1048, 1956.

[4] John Davis. Forty years of progress in long-baseline optical interferometry: 2005 robert ellery lecture. *Publications of the Astronomical Society of Australia*, 23:94–104, 2006.

[5] Universität Würzburg. G78 klassisches und nichtklassisches licht (1).

[6] R Hanbury Brown, J Davis, and LR Allen. The angular diameters of 32 stars. *Monthly Notices of the Royal Astronomical Society*, 167(1):121–136, 1974.

[7] CTA Collarboration. How cta works. Online, accessed on 2019-07-30 https://www.cta-observatory.org/about/how-cta-works/.

[8] Gabriel Pérez Diaz. How cta works. Online, accessed on 2019-07-30 https://www.cta-observatory.org/about/how-cta-works/.

[9] Katja Gumbert. Laborversuch zur intensitätsinterferometrie mit dem cherenkov-teleskop iceact, 2019.

[10] George W. Swenson Jr A. Richard Thompson, James M. Moran. *Interferometry and Synthesis in Radio Astronomy*. Springer Open, 2017.

[11] Andreas Zmija. Design and characterization of an intensityinterferometer with thermal light sources. Master's thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2018.

[12] Peter Deiml. Investigation of microparticle suspensions using statistical optics. Master's thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2015.

[13] Joseph W. Goodman. *Introduction to Fourier optics 2nd ed.* McGraw-Hill, 1996.

[14] Spectrum Instrumentation GmbH. M4i.22xx-x8 - 8 bit digitizer datasheet. Online, accessed on 2019-07-30 https://spectrum-instrumentation.com/sites/default/files/download/m4i22_datasheet_english.pdf.

[15] Root documentation. Online, accessed 2019-08-08 https://root.cern.ch/documentation.

[16] NKT Photonics A/S. Pilas dxpicosecond pulsed diode lasers, datasheet. Online, accessed on 2019-07-31 https://www.nktphotonics.com/wp-content/uploads/sites/3/2018/06/als-pilas-dx.pdf, 2019.

[17] AMTEC INC. Model vt120fast timing preamplifieroperating and service manual. Online, accessed 2019-08-06 https://www.ortec-online.com/-/media/ametekortec/manuals/vt120-mnl.pdf.

[18] SSB-Electronic GmbH. Datenblatt ecofelx 10. Online, accessed on 2019-08-03 https://ssb.de/download/Kabel/DE/6080%20DE.pdf.

[19] Stefan Eschbach. Pmt characterization for flashcam. Online, accessed 2019-08-07 http://astroteilchenschule.nat.fau.de/schule2015/teilnehmer_vortraege/Eschbach.pdf.

Danksagungen

Last but not least bleibt mir die Ehre mich bei all jenen zu bedanken, die mich bei meiner Bachelorarbeit unterstützt haben. Ich danke insbesondere

Stefan Funk für das in mich gesetzte Vertrauen diese Arbeit hin zu bekommen und seine stets lockere, frische und ermunternde Art.

Gisela Anton für einfache Antworten auf komplizierte Fragen.

Andreas Zmija für die stets hilfreiche Betreuung und Unterstützung bei den kleinen wie großen Problemen, die ich auf meinem Weg dank seiner Hilfe aus dem Weg räumen konnte.

Adrian Zink dafür, dass er immer wenn ich dachte eine Frage sei beantwortet eine neue gestellt hat.

Meinen Bürokollegen für die stets gute Unterhaltung beim gemeinsamen Gummibärchen-Verzehren, zur Mensa laufen und darüber diskutieren was man macht wenn man diese Arbeit endlich fertig hat.

Allen die diese Arbeit gegengelesen haben für die Hilfe beim Finden der vielen Fehler, die man über die Wochen so produziert.

Erklärung

Hiermit erkläre ich, dass ich dir vorliegende Arbeit selbst verfasst habe. Es wurden nur die in der Arbeit angegebenen Quellen und Hilfmittel benutzt.

Erlangen den 13.August 2019

Frederik Wohlleben