Rapport de stage

Width measurement of single ion beam for the deterministic implantation of ions in Diamond

Institut für Physik, Mainz

Luc COUTURIER
Master 2 Quantum Devices, Paris Diderot
Maître de stage : Kilian Singer

Du 6 Mars au 27 Juin 2013

Abstract

This report presents the work on a project in the framework the experiment about the deterministic implantation of single ion in diamond, in the AG Schmidt-Kaler, located in the Institut für Physik in Mainz.

In the first part I will give a global overview of the experiment which consists in trapping single ions in a linear Paul trap before accelerating them by the mean of electric fields into a diamond.

In the rest of the report I will present two different optimizations I made for the knife-edge measurement of the single ion beam. They rely respectively on the two interpretations of probability: the frequentist and Bayesian point of view. Then I will explain the working principle of each of them as well as their theoretical background.

Finally I will discuss their performances by using simulations. It turns out that their accuracies are similar but the robustness and the usability make the Bayesian method the best alternative for the single ion beam measurement.

Résumé

Le présent rapport présente le projet effectué dans le cadre de l'expérience sur l'implantation déterministique d'ion dans un cristal de diamant au sein the du groupe de Ferdinand Schmidt-Kaler, situé au "Institut für Physik" à Mayence.

Dans la première partie je donnerai un point de vue d'ensemble de l'éxpérience qui consiste à piéger des ions uniques dans un piège de Paul linéaire avant de les accélérer au moyen de champs magnétiques vers un cristal de diamant.

Dans le reste du rapport je présenterai deux différentes optimisations apportées à la mesure du diamètre du faisceau d'ions unique par la "méthode de la lame de couteau". Elle reposent sur deux interprétation de la probabilité : le point de vue fréquentiste et le point de vue Bayesien. J'expliquerai leur principe de fonctionnement ainsi que leur cadre théorique.

Enfin, je discuterai leur performances en utilisant les résultats de simulations. Il s'avère que leurs exactitudes sont similaires mais la robustesse et la facilité d'utilisation de la méthode Bayesienne en font la meilleure alternative pour la mesure de faisceaux d'ions uniques.

Contents

1 Introduction

1.1 Framework of the internship

This internship was done in the team of Kilian Singer, in the Physikalisches Institut in Mainz among other projects of the Ferdinand Schmidt-Kaler's group. I worked at the experiment for the deterministic implantation of single Nitrogen ions in Diamond. The main goal is to create NV centers[1] whose position and number are well defined. To do so, one uses a segmented Paul trap to catch one single ion, which is first accelerated by an electric field and then focused in the diamond by the mean an electrostatic lens. For the moment, the group was working only with Calcium ions instead of Nitrogen for simplicity. The current goal is to reach better focusing and then to use Nitrogen ions to create NV centers at given places.

1.2 Goals of the internship

This report presents the project I had to carry out. When I arrived in the lab, the team was changing the old electrostatic lens for a new one. We were able to achieve a focus resolution of 8 nm measured with a knife-edge measurement method. However this measurement was relatively basic, not particularly reliable and time-consuming. The main task I had to deal with was to find a way to optimize this measurement to make it more precise and faster. This was mainly simulations using Mathematica and C++[2] in order to write an improved program using a Bayesian approach to run the experiment for measuring the beam diameter.

First, I will present the whole setup in order to give a global view on the experiment in order to to describe with which system we are working. Then, I will present the different approaches we considered to optimize the measurement: first an optimization of the existing measurement, which will be called "Frequentist" method in the present report, by opposition to the second method which is a "Bayesian method" based on the Bayesian experimental design. I will also deepen the explanations about the Bayesian point of view and the consequence on the interpretation of the results. Finally I'll present some results of the simulations, due to unexpected problems with the experiment.

The main point of the report is not about which resolution can be achieved but about how one can perform an efficient and reliable measurement and how to express a result which can be trusted.

[1] Nitrogen Vacancy centers

[2] with the help of MCP, a C++ interface developed by Kilian Singer. See .

2 The implantation experiment

2.1 Introduction

In the past few years the number of transistors on an integrated circuits has increased exponentially, the size of these structures reaching few tens of nanometers. At these scale the statistical Poissonian fluctuations can modify the properties of the device a lot. This is also problematic for any future solid state Quantum computers that are based on the implantation of single atoms acting as Qubit carriers, for example color centers in diamond or phosphorus dopants in silicon, [Dutt et al., 2007, Neumann et al., 2008, Kane, 1998]. Some of these candidates are promising, but the problem of the scalability is still a big issue that needs to be overcome in order to create a quantum computer, [DiVincenzo et al., 2000].

So far very few methods, if none, have achieved to implement particles in a highly deterministic way with a resolution of the order of a nanometer. There are two major problems: the achievable resolution on one hand and the number of implanted ions on the other hand. The challenge is to solve both at the same time. Until recently, implantation relied solely on thermal sources where the fluctuating number of dopants can be only checked a posteriori without any further possible improvement. Moreover these sources uses particles with a high energy which induce defects in the substrate and a fluctuation of dopants number. Some other tricks can be use to implant particles at a given such as [Hausmann et al., 2011] using ordered nanostructure in order to trap the Nitrogen at some specific place or [Meijer et al., 2008] using an AFM tip with a small hole to guide the ion beam to a specific target. But in the end, the number of dopant implanted still fluctuates.

The idea of the current experiment is to trap single ions in a Paul trap, which nowadays is a well-known technology, and accelerate them by applying an electric field before focusing the beam with an electrostatic lens. Trapping the ion initially gives a very small velocity dispersion of $\Delta v/v = 8\text{m.s}^{-1}/110303\text{m.s}^{-1}$ which corresponds to an energy dispersion of $\Delta E/E = 3 \cdot 10^{-4}$ and a very narrow angular spread of the beam about ≈30 µrad, since the initial position and momentum are well defined. Moreover, the energy of the ions is relatively low, about 2 µK. The main limitation up to now was the resolution about 5 µm. With installation of a new lens, we were able to achieve a resolution down to a few nanometers while using using deterministic single ions. Determining this value was the main motivation for this internship.

In this first part we are going to give a global overview of the experiment.

2.2 Experimental setup

The focus of this work is not on the whole experiment, so I am just going to give a global idea of how the experiment works. For more details one can look in the master and PhD thesis from former students as [Wolf, 2012] or to the published papers [Schnitzler et al., 2010], [Schnitzler et al., 2009]. First we are going to give a global view of the setup, before explaining in the next part the implantation process. We won't detail so much all the technical details of the experiment which would be far beyond what has been done during the internship and what could have been done in a such little amount of time.

2.2.1 General scheme

In figure 1 are represented the main elements of the experiment necessary to understand the working principle. We won't detail all the laser setup for the Doppler cooling, neither the technical details of the microscope, the electronics needed to control the electrodes of the trap nor the sequence needed for each step of the experiment.

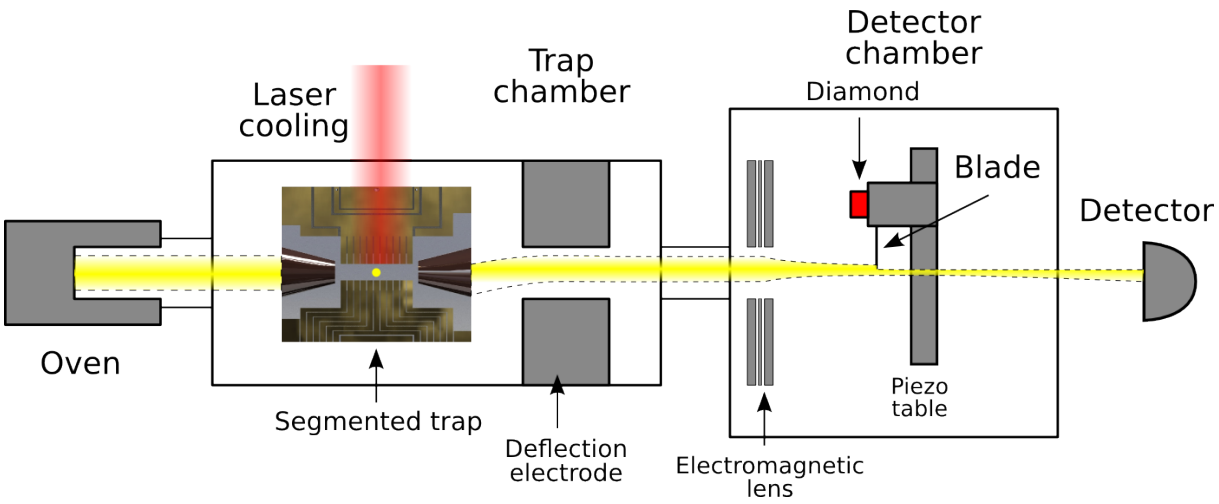


Figure 1: General scheme of the experimental setup

2.2.2 Segmented Paul trap

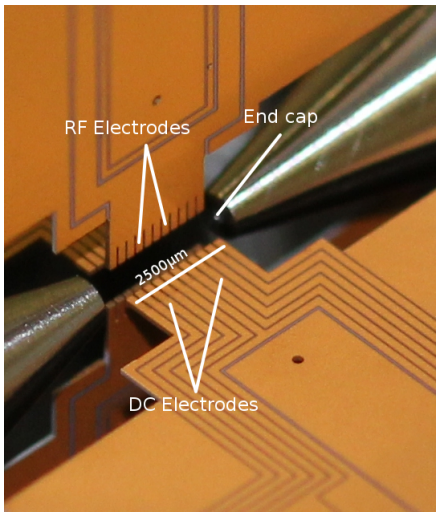


Figure 2: Picture of the trap. Each electrode has a width of 200 µm and the gap between two of them is 30 µm large.

The Paul trap is made up of 4 ceramic blades covered by gold electrodes and assembled in a cross shape as shown in picture 2. Due to the requirements for a Paul trap, two of the blades are supplied with DC voltage for the axial confinement and the two others supplied with RF voltage. The DC electrodes are segmented in 11 independent electrodes[3] which allow to manipulate the potential in axial direction, see [Walther et al., 2012] for the demonstration of a such transport in a similar trap from another setup of the lab. There are two additional end cap electrodes at each side of the trap which are used to extract the ions by applying a high negative voltage. A small hole of 200 µm has been dug by eletroerosion and allows for extraction of the ion. By applying the right voltages and frequencies one can trap the ion, before cooling it down by the mean of Doppler cooling and then extracting it.

If one wants more information about Paul traps and single ion trapping one can refer to [Leibfried et al., 2003] for instance.

2.2.3 Electromagnetic lens

The electrostatic lens is a crucial element to get a small implantation resolution. The design affects the focus size a lot. When I arrived in the lab, we install the new lens with a new design. The previous one led to focus size about 5 µm whereas with the new one resolutions of few nanometers were achieved.

[3] but two of them are linked together due to fabrication defects

Since we are using beam with small divergence and narrow velocity distribution, we don't have to make sophisticated aberration corrections and the design stays simple: it is three consecutive circular electrodes, the first and the third are at ground potential whereas the electrode in the middle is at a positive or negative voltage up to 3 kV. Depending on the voltage sign of the middle electrode, the lens presents different behaviors:

- If the sign is the same as the sign of the particle charge, it is a deceleration-acceleration mode.
- If the sign is the opposite of the sign of the particle charge, it is a acceleration-deceleration mode.

In the accel-decel mode the chromatic and spherical aberrations are smaller. However the decel-accel mode requires lower voltage to achieve the same focal length, which is a good point since in general it is difficult to isolate effectively the electrodes from each other. In addition, it is possible to correct for the spherical aberration by switching the lens to a higher voltage when the particle is passing through the lens. In the experiment we use the decel-accel mode, that is to say we apply a positive voltage on the middle electrode. The details about the lens can be found in [Fickler, 2009] or [Schnitzler, 2010].

For more general information about electrostatic lenses, on can refer to [Rose, 2008].

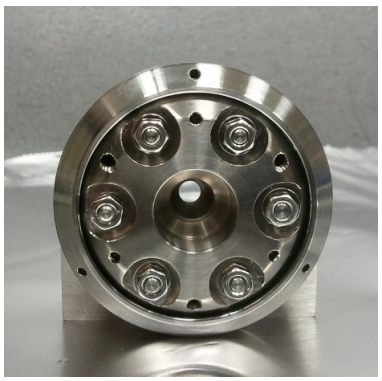


Figure 3: Picture of the new lens, compared to the old lens, a more sophisticated procedure was applied. Also the isolation between the electrodes is better.

2.2.4 MCP

For this report, one important part of the experimental setup is the control of the experiment by a computer. To do so, every experimental procedure is written in C++ using the Master Control Program interface, aka MCP. MCP is a Qt based C++ user interface developed by Kilian Singer which allows easy dynamical handling of C++ functions. It allows also to plot one or two dimensional graphs, histograms or to displays parameters values for example. This can be seen as an equivalent of the front end of LabView but with C++ code in the back end. A screenshot can be seen in figure 14 in the appendix with the interface of the procedure written during the internship. One can use the debugging functionalities of Microsoft visual studio to track errors in our code interpreted by MCP. One of the main advantage is that it can compile the code without restarting the program and losing the front end elements. It naturally benefits from the high computation speed and hardware compatibility of C++.

It also runs every experiment in the lab but can also be used for simulation which requires higher computational power than interpreted language software like Mathematica. Many very complex routines are carried out on the experiment and allows for handling nearly every parameter of the experiment from the computer.

In all the following, most of the simulation were written with Mathematica to start before implementing it in C++ with MCP. However it uses some particular templates to communicates with the front end, which forces to "bad" programmation habits like using global variables or putting class member in the public part.

2.3 Implantation overview

In this part we describe here the different steps of the implantation process.

2.3.1 Loading, cooling and separation of the ions

For the moment we are only working with Calcium ions, because it is easier to produce Calcium than Nitrogen ions, the latter one needing higher energies. First to produce the ions, Calcium powder is heated in an oven to evaporate it, the atoms are first excited by a laser at 423 nm and ionized by a second one at 375 nm. The ions are then Doppler cooled down to 2 mK. Initially more than one ion are trapped in the potential, it is therefore needed to remove the extra ions by lowering the potential barrier. Note that we cannot cool down the Nitrogen directly ions since there is no laser available to do it, therefore the cooling will be done by sympathetic cooling with Calcium ions. This kind of cooling relies the thermalization of two different atom species in a potential: the two species share their energy through the Coulomb interaction in order to reach the same temperature. However since the Calcium ions are cooled by lasers so the whole system is progressively cooled down. The cooling is really important since the temperature will increase the lateral energy dispersion and therefore will increase the size of the focus.

The fluorescence of the Calcium ions is imaged by a camera and a procedure check if there is effectively one single ion in the trap. In the end, after this step there is only one ion trapped in the Paul trap which can be extracted.

2.3.2 Extraction and focusing

To extract the ions from the trap, we apply a negative high voltage on the end cap of the Paul trap to accelerate the ions. They go through a small aperture in the end cap and fly through the main chamber of the setup before being focused by an electrostatic lens.

The working principle of an electrostatic lens is similar to an optical lens and so are the aberrations. Considering that we want a very high accuracy, one needs to take these errors into account. In our experiment, one can neglect the influence of the chromatic aberration. This one changes the focus position depending on the energy distribution which here is the energy distribution of the ions.

However the spherical aberration plays an important role and has to be taken into account. The spherical aberration changes the focus position according to the position of the incident beam on the lens. The aberration is particularly important near the edge of the lens. It's due to the fact that some approximation concerning the lens are done, like for example the thin lens approximation in the case of optical lenses, which are no more valid. To describe the spherical aberration, let's consider a point source at a distance a from the lens, this point is shifted to a distance r_1 from the optical axis. Moreover the beam includes an angle α with the optical axis. In the case of a perfect lens, the distance between the optical axis and the image of the source situated at the position b is given by:

$$r_2 = \left(1 - \frac{b}{f}\right) r_1 + \left(a + b + \frac{ab}{f}\right) \tag{1}$$

and a beam which includes an angle α with the optical axis cross this one after the lens at a distance:

$$f_\alpha = \frac{f\,(r_1 + a\alpha)}{(f + a)\,\alpha - r_1} \tag{2}$$

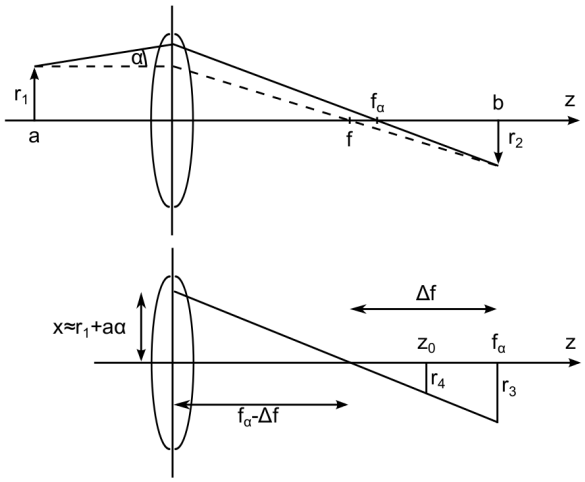


Figure 4: The upper scheme corresponds to a perfect lens, every ray from the object in position a are focused on the image at $z = b$. In the lower scheme is represented the focus length for a lens with spherical aberration.

For a lens with a spherical aberration, these formulae are not valid anymore far from the axis. For a small angle, the distance from the axis can be approximated by $x = \approx r_1 + a\alpha$ the position where the beam crosses the optical axis is shifted by Δf in the direction of the lens. If we consider a rotational symmetry for the lens, the 1st order expansion of the shift can be expressed as:

$$\Delta f = c_2 x^2 + \mathcal{O}(x^4) \tag{3}$$

where c_2 is a coefficient which characterizes the spherical aberration and which we want to determine experimentally. The distance r_3 from the optical axis to the image of the source is:

$$r_3 = \frac{x\Delta f}{f_\alpha - \Delta f} \tag{4}$$

If we want to determine, this distance r_3 but for any arbitrary z_0 position, noting r_4 this gap, one can write it as follows:

$$r_4 = \frac{z_0 - (f_\alpha - \Delta f)}{\Delta f} r_3 = \frac{z_0 + c_2 \left(r_1 + a\alpha\right)^2 - \frac{f(r_1+a\alpha)}{(a+f)\alpha - r_1} + \mathcal{O}(x^4)}{\frac{f}{(a+f)\alpha - r_1} - c_2 \left(r_1 + a\alpha\right) + \mathcal{O}(x^3)} \tag{5}$$

This expression doesn't make any sense if Δf becomes equal to f_α or if $\Delta f > f_\alpha$. It is therefore meaningful to make a Taylor expansion for $\alpha \approx 0$. We also add the assumption that the source is now on the optical axis. Finally r_4 is approximated by:

$$r_4\left(\alpha\right) = \left(z_0 + a\left(\frac{z_0}{f} - 1\right)\right)\alpha + \frac{ac_2(a+f)^2 z_0}{f^2}\alpha^3 + \mathcal{O}(x^4) \tag{6}$$

One can also find the formula in [Rose, 2008]. From a knife edge measurement, one can thus find the focal length and also estimate the spherical aberration. One can also find the z-position where the dispersion of the beam is the lowest, it is at $\Delta f/4$ after the beam focus. At this place, the beam has a width of

$$r_{3,min} \approx \frac{c_2 x^3}{4 f_\alpha} \tag{7}$$

In practice, one wants to have the smallest possible focus, therefore one needs to minimize the spherical aberration which is done by aligning the beam to the center of the lens. For that, considering that the lens has a spherical symmetry, we can scan over the lens by changing the voltage of the deflection electrodes. They are situated at the exit of the trap and allows for adjusting the trajectory of the ion. One can fit the result with the formula 6. Depending on the relative position between the focus plane and the measurement plane, the beam position takes different shapes. For a plane measurement near the focus, the lens shows a large region near the center where the final position doesn't depend on the position on the lens, as shown in 5.

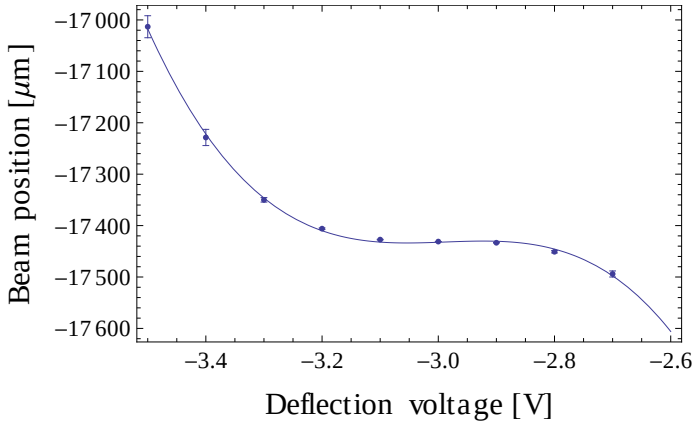


Figure 5: Example of experimental scan showing the position of the beam according to the deflection voltage, fitted with the formula 6. One finds $c_2 = 1.24 \pm 5\text{nm}^{-1}$ for the spherical aberration coefficient.

This symmetry point of the function is where the beam is aligned to the optical axis of the lens and thus the aberration becomes minimal. At this point one might check the size of the beam using a basic knife-edge measurement. This is the part which will interest us in the rest of the the report, so I will discuss that in detail that later.

2.3.3 Implantation and annealing of the diamond

Once the Nitrogen has been accelerated into the diamond at a place selected by the piezo table, one has to anneal the diamond at 800° in order to produce the NV centers. The Nitrogen has to merge with a vacancy to produce color center. This is the only step in the process which is not totally deterministic since the efficiency of creation of NV center is not 100%. The efficiency of NV center creation depends strongly on the ion energy as explained in [Pezzagna et al., 2010].

How to proceed with this step hasn't been decided yet, the problem is that the high temperature can damage the piezo table. There are different possibilities, but for the moment the diamond will be taken out the vacuum chamber.

2.3.4 Checking the NV center creation through spectroscopy

The last step of the process is to check whether the color center creation was successful or not. We use a confocal microscope which should be able to resolve two near NV centers from each other according to the Rayleigh criterion. The resolution for a normal microscope is given by $r = \frac{\lambda}{2n \sin\theta}$ where $2n \sin\theta$ is the numerical aperture. At a distance of few tens of nanometers and at a wavelength of 532 nm, this won't be enough to resolve two close NV centers. There are two possibilities which can work well in the setup to overcome this resolution limit.

- STED microscopy (*Stimulated emission depletion microscopy*: a first laser beam with a non-Gaussian mode is used to saturate the absorption of the color centers in a ring around a central point while a second laser beam will activate the fluorescence of the non saturated centers in the middle with a different wavelength corresponding to the fluorescence wavelength. See [Willig et al., 2007] for more extensive explanations.
- GSD microscopy (*Ground state depletion microscopy*: It's nearly the same excepted that we don't use a second laser, if the NV center is exactly in the middle, it won't emit any light at all. One just has to scan over the sample. Afterwards the image is a convolution of the centers the toroidal shape of the beam and hence has to be deconvoluted.

Since we want to observe the sample in situ, the GSD is used because it needs only one laser beam. According to [Hell, 2007], the resolution achievable for the GSD microscope is given by:

$$\Delta r \approx \frac{\lambda}{2n \sin\alpha} \sqrt{I_{sat}/I_{max} + 1} \tag{8}$$

where $n \sin\alpha$ is the numerical aperture and $\zeta = I_{sat}/I_{max}$ is defined as the saturation factor.

The microscope system using GSD is already in place in the experiment and we are able to detect NV centers from a sample correctly, [Wolf, 2012].

2.3.5 Completion of the pattern

The great advantage of this experiment is that if we observe a missing NV center in the designed pattern, we can always afterwards redo the procedure to implant the missing centers. However since the diamond has already been annealed, few diamond vacancies are left and the Nitrogen ion is still trapped in the diamond. One can create additional vacancies at this place by shooting Carbon ions or Calcium ions and anneal sample again. Thus we can build a fully deterministic pattern of NV centers.

3 Beam measurement: First approach

During the internship I had to deal with a very practical issue, namely the measurement optimization of the distribution of the ion trajectories which can be seen as a "beam". This is an important issue since the beam diameter will determine the achievable precision on the implantation of the Nitrogen ions in diamond and thus the minimum deterministic gap between two NV centers. This will prove the viability of the experiment. With the recent change of the lens, big improvements were done but we were unable to measure this beam diameter with a good confidence. To measure it, we used the well-known Knife Edge measurement method which we will describe in the following. It's commonly used to determine laser beam diameters for example, but in our case this measurement becomes non trivial and has to be improved in order to acquire meaningful data.

3.1 Introduction and problematic

This method is used to determine the diameter of a beam with a given distribution. Very often the shape of the beam follows a Gaussian distribution, e.g. the first Gaussian mode of a good quality laser beam, and we will also assume in first approximation that the ion trajectory follows a Gaussian distribution:

$$P(r,z) = f(z) \times \exp\left(-\frac{\mathbf{r}^2}{2\sigma^2(z)}\right) \tag{9}$$

with:

- The sigma of the beam $\sigma(z)$ such as $w = 2\sigma$ where w is the width of the beam.
- A function f depending on the position z along the propagation axis and which reduces the diameter size at the focal point. For each z position, the PDF should be normalized so $f(z) = \frac{1}{2\pi\sigma(z)}$

This distribution is similar to the shape of the laser beam except that instead of working with macroscopic particle number, we deal with only one particle at a time. The function $\sigma(z)$ in our case has been evaluated as:

$$\sigma(z) = \sqrt{\frac{z^2 f^2 \sigma_\alpha^2 + 2azf(f-z)\sigma_\alpha^2 + (z-f)^2(a^2\sigma_\alpha^2 + \sigma_r^2)}{f^2}} \tag{10}$$

with σ_r the standard deviation of the spatial distribution of the source, i.e. the initial position of the ion, a the distance between the source and the lens and σ_α is the standard deviation of the thermal velocity distribution in the trap.

The principle of the measurement is very simple: we cover a part of the beam for a given z position with a blade and we measure the rest of the beam. By moving the blade until covering the whole beam, we can access the cumulative distribution of the beam, which is, in the case of a Gaussian distribution, an error function:

$$P_{Measured}(x) = \frac{P_0}{2}\left[1 + \mathrm{erf}\left(\frac{(x-x_0)}{\sqrt{2}\sigma(z)}\right)\right] = \frac{P_0}{2}\mathrm{erfc}\left(-\frac{(x-x_0)}{\sqrt{2}\sigma(z)}\right) \tag{11}$$

or

$$P_{Measured}(x) = \frac{P_0}{2}\left[1 - \mathrm{erf}\left(\frac{(x-x_0)}{\sqrt{2}\sigma(z)}\right)\right] = \frac{P_0}{2}\mathrm{erfc}\left(\frac{(x-x_0)}{\sqrt{2}\sigma(z)}\right), \tag{12}$$

depending if it is a rising or falling edge. Just as a reminder, the complementary error function $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$ and the symmetry property of the error function $\mathrm{erf}(-x) = -\mathrm{erf}(x)$. Then by fitting the data set, one can find σ and x_0 for a given z position. By reiterating the measurement for different z position, one can measure the function $\sigma(z)$ and for example find where is the focus of a convergent beam.

This method is usually used to measure the parameters of a laser beam, but in the case of single ion measurement, this is no trivial matter for different reasons.

First, contrary to laser beam, the only outcome values are 0 or 1 since we send only one ion at a time and we observe its transmission through the electrostatic lens. The consequence is that a single measurement won't help to obtain the probability at a given blade position, one needs many ions to estimate the probability.

The direct consequence, considering that a single ion shoot takes roughly a second, is that the measurement can require a huge amount of time. If the system is not perfectly stable and there is for instance

some drift, a long measurement won't let us to determine precisely the parameters. Therefore we can't choose an arbitrary large number of ions for each point without exposing us to some supplementary noise that perturbs the measurement.

Lastly, dealing with a finite number of ions rises the question of the determination of the uncertainty: is the number of ions high enough to estimate it statistically ?

This last question leads to a very interesting point: Should we use a Frequentist approach or a Bayesian approach? We will deal with this problem later, just remind that for the moment x_0, σ and P_0 have a fixed value

Usually, for a given z position, a set of 300 ions distributed over the different blade positions gives reasonable measurement times compared to the drift of the system due to the oven. Thus the goal is to acquire the most useful data with a limited number of ions. We need to evaluate which points will give the most information on the σ. For all the following, we assume that the beams follows a Gaussian distribution.

3.2 Basic approach

Before the start of the internship, by simplicity the most basic approach we can imagine was used for this measurement. As shown on figure 6, we just move the blade position with equidistant steps and we measure an equal number of ion extractions. Then we fit the data with an error function as described before. In this case since all the points are equivalent, we don't have to normalize the result and the amplitude can be simply interpreted as a number of ion detection. The estimation of the error was estimated with the expression of the standard deviation used for a statistical number of points. Thus for the fit every point is equally weighted. Typically we measure the transmission of 20 Ca ions for approximately 15 points distributed equally over the whole blade range position.

This method has many drawbacks and bias which make it inaccurate to evaluate precisely the width of the beam. Intuitively it seems pretty obvious that too many ions are used to measure the probability far away from the center, where the probability is either 1 or 0, with the exception of the noise. Moreover only the points near the mean value can really give information about the position of this one and about the sigma. To compensate that we often searched by fumbling the center and then reduced the position range, but by doing so, we lose some information from the ions used to roughly find the center.

Second, the determination of the uncertainty was totally wrong. We equally weighted every points since we can't use the standard deviation due to the lo number of points. It would have been expressed as follows:

$$\sigma_x = \sqrt{\frac{1}{N_{total}} \sum_{i=1}^{n} (y - \mu)^2} \tag{13}$$

where y is the outcome of one ion transmission and μ the mean value at a given x position. The alternative would have been to use a binomial distribution which is not possible either[4]. From a very practical point of view, it is not easy to measure additional data points after the scan is done. The last disadvantage of this measurement method is the fitting process. We basically use a non-linear least square method provided by any data analysis software with the model presented before. The problem is that the fit doesn't converge well due to the large uncertainty but the equal weight of the data, this give rise to different possible values or absurd ones.

[4]Explanations will follow in 3.3.3.

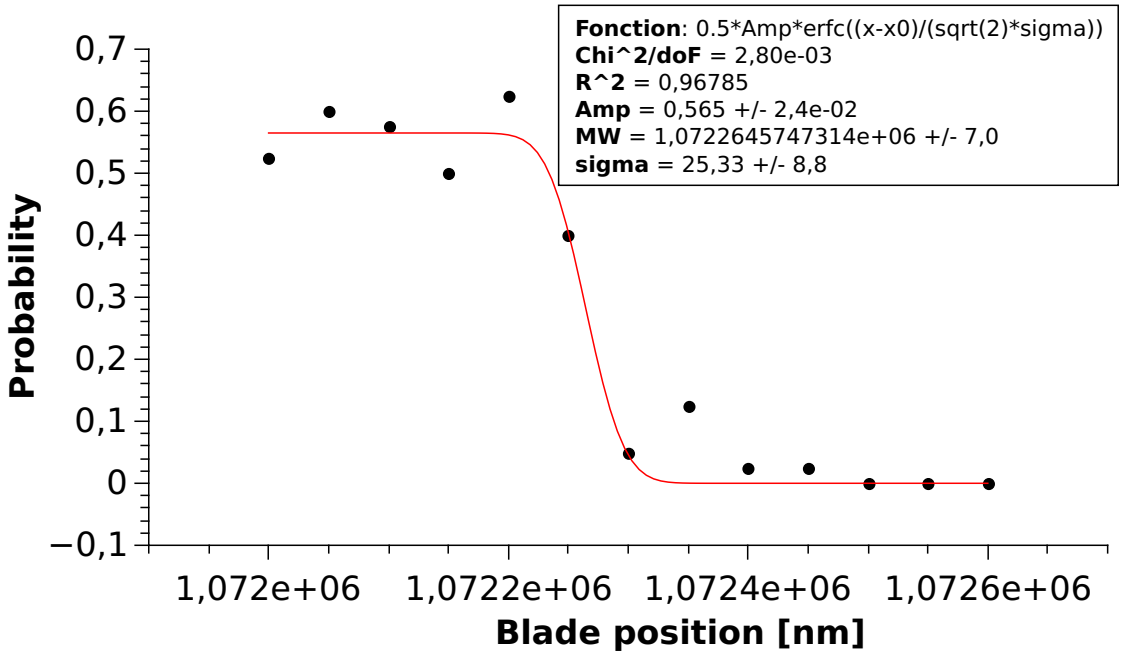


Figure 6: Example of a typical knife-edge measurement before any optimization. Every point has the same weight and don't take the error into account. The number of parameter is high for the number of points. With so few points in the middle, the result could have been different.

Thus we cannot hold as reliable the result of the focus measurement where this one was about "8 nm". To be able to claim a such focusing resolution, one must determine the focus size with a better confidence.

3.3 Frequentist optimization

In this "frequentist optimization" method, by opposition to the Bayesian one that we will describe later, we still want to find the probability for few points at different blade position and then fit the data to find the parameters. This implies taking many points at the same position in order to find an estimation of the probability at this place. The main difference is that an algorithm will seek for the center and measure more extractions near it.

We will explain in 4.1.1 that this method works with the frequentist interpretation of probability.

3.3.1 Global idea

The idea is quite simple:

- Starting far away from where the center x_0 is presumed to be, we scan with as few ions as possible and then reduce the research interval. In the the C++ code this step is referred as "Broadscan".
- Then we reiterate again this step t narrow again the research interval. In the the C++ code this step is referred as "Scan".
- We measure a high number of single ion transmissions at equidistant blade positions in this narrowed interval where the falling edge is. In the the C++ code this step is referred as "Finescan".
- Finally we fit the results as before.

The major improvement is that we use a higher sampling to measure the probability where the intuitive information gain seems to be higher, i.e. around the rising/falling edge. For that, one has to convert the number of ions transmitted into a probability. Moreover, one another big amelioration is that every outcome of ion transmission measurement is saved and used for the final fitting; we can start we any arbitrary large blade position range without losing any information while looking for the center. And finally, I programed it to be able to correct manually any interval between each step or to repeat one of the step described above if we need more points, the constraint being that to be able to calculate a probability, there should be no single ion transmission measurement for a single blade position but always many points for a same blade position.

3.3.2 Detailed explanation

The code of the C++ implementation can be found in the appendix B We will only explain the main idea of the working principle and the critical points, but globally most of the code is not interesting from a physical point of view.

3.3.2.1 Simulation of the transmission process and influence of the noise

To simulate the shooting of one ion through the lens one can use the previous distribution with the error function:

$$P_x = \frac{1}{2}\text{erfc}\left(\frac{(x-x_0)}{\sqrt{2}\sigma(z)}\right) \tag{14}$$

when the blade is uncovering the beam and

$$P_x = \frac{1}{2}\text{erfc}\left(\frac{-(x-x_0)}{\sqrt{2}\sigma(z)}\right) \tag{15}$$

when the blade begins to cover the beam. Then we can just generate a random number between 0 and 1, if the outcome is below the value given by the distribution, the function will return 1, else 0.

This works well for a perfect experiment with a detection efficiency of 100% and no other external perturbation. In practice there is always a small probability that the ion is not detected, even if the outcome should have been one. To take this into consideration we just test afterwards if the simulated ion is "lost" and reduce the result to 0 if it is the case. We could also introduce a factor to translate the fact that the maximum total transmission is lower than 1:

$$P_x = \frac{P_{max}}{2}\text{erfc}\left(\frac{\pm(x-x_0)}{\sqrt{2}\sigma(z)}\right) \tag{16}$$

with P_{max} the maximal transmission efficiency. We chose the first possibility to highlight the fact the noise may be independent of the blade transmission.

Implemented in C++, this gives the following code:

```
double KE::newpointsim(double x){
        int y;
        if(m_reverseDir==true){
                if((double)rand()/RAND_MAX<0.5*erfc(-(x-m_MV)/sqrt(2.)/realS)) y=1;
                else y=0;
                if((double)rand()/RAND_MAX>=noiseloss);
                else y=0;
                return y;
        }
        else{
                if((double)rand()/RAND_MAX<0.5*erfc((x-m_MV)/sqrt(2.)/realS)) y=1;
                else y=0;
                if((double)rand()/RAND_MAX>=noiseloss);
                else y=0;
                return y;
        };
}
```

We can also add a time dependency of the center position, i.e. a dependency of the number of ions in the simulation, $x_0(N)$ to take into account the drift of the system, which might be not negligible for long time measurement.

The influence of the noise has to be considered seriously since it makes complex the algorithm to detect where the center is: observing a "1" over two ions shots doesn't give so much information about the center position if noise is present. It can mean that we are close to the center or very far away but the ion wasn't detected. It can easily perturb the boundaries of the research interval.

3.3.3 Error estimation

One critical point is the error estimation. Previously as described in 3.2, we used the classic definition of the standard deviation to describe the uncertainty. However this can't be right due to the low number of ions.

For example, let's say we measure the probability in the region where the probability is one. We measure 4 ion transmissions and all of them give a detected ion, i.e. "1". That would mean that $\sigma = \sqrt{4(1-1)^2} = 0$ which is totally absurd: we always might observe a "0", no ion transmitted, in the next trial whatever the previous outcomes were. The uncertainty should take that into consideration. This behavior shows up because these statistical analysis assume an important number of events. Usually for the experiments with binary outcome like here, the experiment is repeated so many times that the variance is never 0, there is always an observation of the other event. We need to find another to describe the uncertainty.

The intuitive solution of the problem would be to use the standard deviation of a binomial distribution whose standard deviation is:

$$\sigma = \sqrt{Np\,(1-p)} \tag{17}$$

with N the number of trials and p the probability to observe 1. In our small example, this would give $\sigma = \sqrt{4 \times \frac{4}{4} \times 0} = 0$ which is not realistic. Another model has to be used.

Actually this situation corresponds to the problem of the noisy coin as described in [Ferrie and Blume-Kohout, 2012]. Thus, for a low number of trials, one can describe the average value of the transmission probability through the lens by:

$$\begin{aligned} \langle P_{Hit} \rangle &= \frac{n+1}{N+2} \\ \left\langle P_{Hit}^2 \right\rangle &= \frac{(n+1)\,(n+2)}{(N+3)\,(N+2)} \end{aligned} \tag{18}$$

where P_{Hit} is the probability to measure a hit on the detector, i.e. measuring 1, N the total number of ions shot and n the number of hits on the detector. Then the standard deviation becomes:

$$\sigma_x = \sqrt{\left\langle P_{Hit}^2 \right\rangle - \langle P_{Hit} \rangle^2} \tag{19}$$

That gives in our example $\sigma = \sqrt{\frac{5}{7} - \frac{5}{7}^2} = \frac{\sqrt{10}}{7} \approx 0.45$ which is different from 0 and seems acceptable. This translate well the fact that having only 1 doesn't mean that the next outcome will also be 1 and that we have a knowledge of the probability limited by the number of trials. For high trials number however the standard deviation will converge to the same value as 19 and we find again the behavior of a statistical measurement.

We used this model to describe the error measurement error which seems to describe correctly the results.

3.3.4 Simulation and results

Typically a simulation of this measurement gives the following result plotted on the figure 7 for $x_0 = 1000$, $\sigma = 25$ and $P_{max} = 0.85$.

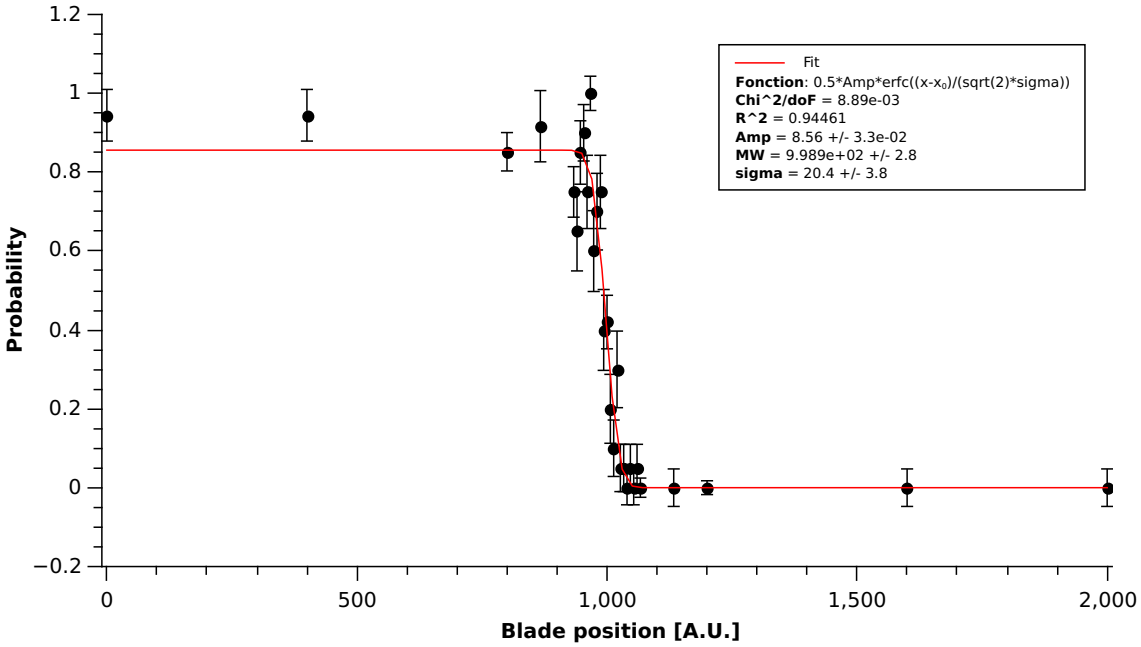


Figure 7: Simulation of a Knife-Edge measurement with the optimized algorithm. Where the error bars are smaller that corresponds to a higher number of measure for this position. The total number of ion is 516. The probabilities near the sides are determined with 6 points whereas the probabilities in the middle are determined from 20 or more measures.

We clearly see that more points are measured near the middle, where more information can be extracted.

However the fit gives $\sigma = 20.4 \pm 3.8$[5] which doesn't cover the the real value. However this is not false due to the interpretation of the error range.

3.3.5 Limitations

Undeniably this measurement methodology is better than the previous one and allows to gain a lot of time in the data acquisition. However, there are still some loopholes.

The algorithm is not very robust either if the initial parameters are wrong or if the noise is important. It's mainly designed to be monitor by someone adjusting the the interval between each step. In its automatic mode, wrong parameters will make the experiment measure points which aren't useful. Moreover, if the drift of the system is fast enough, the center might be brought out of the region of interest since during the last step, where most of the points are measured, there is no real-time update of the boundaries.

Moreover this algorithm is really specific to the experiment and built based on the knowledge of the error function. This wouldn't be adaptable to different measurements with different distribution functions.

[5] The value can be seen as realistic and could be expressed in nanometers

4 Beam measurement: Bayesian Approach

4.1 Introduction to Bayesian approach

4.1.1 Bayesian vs. Frequentist

The Bayes' theorem is well known in statistics and probability but the Bayesian approach is far more general than the Bayes' theorem. This theorem was first formulated by Thomas Bayes (1701–1761), albeit published only later. This opened a debate about the meaning of probability which has no definitive conclusion so far. The Bayesian interpretation has to be opposed to the frequentist interpretation. These are two different ways of extracting and interpret information from observation. They are fundamentally different and there is no "right" one. A good introduction this problematic applied to physics can be found in [Lyons, 2013].

In the frequentist interpretation, the veracity of an assumption is determined by the proportion of the different outcomes; the more observations we have, the more we can predict the real value of the assumption with a good uncertainty. The frequentist interpretation is the most used, particularly in Physics. For an important number of repeted observations, we are able to predict the real value with a good confidence interval. The main idea of this interpretation is that we assume there is a fixed value for the parameter we want to determine, but we don't have enough data to find the true value because we are limited by the finite amount of observations and by the noise on these, therefore we give an statistical estimation of this value based on the frequency of observation. A probability, from a frequentist point of view is objective and is the limit of the relative frequency of an observation. A unique event from the frequentist point of view doesn't mean anything: if I throw a dice twice and two times the result is 6, that doesn't mean the dice is loaded. Another point: the limits of a confidence interval are known but random.

On the other hand, using the Bayesian interpretation, there is no fixed value for the hypothesis and we try to estimate how probable the values are, using the finite amount of data collected. The parameters we want to determine are random and follow a probability distribution, to give the value of a parameter one often uses the mean value of the distribution. Contrary to the frequentist approach, one need an initial degree of belief, i.e. an initial distribution, which is the reason why from a Bayesian point of view, probabilities are subjective, it depends from the initial degree of belief. Each time an observation is made, we can update this degree of belief. Concerning the confidence interval, the limits of this last one are known and fixed, it can be determined precisely. Let's take the previous example with the dice: If the initial belief is that the dice is loaded, observing two 6 will narrow the probability distribution around the parameter value "the dice is loaded" and vice versa. We can also believe initially the possibility of having a loaded dice is equal to not have one, "the dice is loaded" will become more probable. In case of the Bayesian interpretation, it makes sense to attribute a probability to a single event since we can compare it to the initial degree of belief, contrary to frequentist interpretation. Saying that the dice has 75% chance of being loaded after two dice roll has a real meaning since in this interpretation we don't have to choose between one of the statement.

Bayesian approach is often used in other fields of science, for example in Biology to find the cause a disease among a huge amount of genes or in medicine for the image reconstruction [Herman et al., 1979], particularly for brain imaging. It has also applications in Physics, particularly in particle physics where very few events are measured and one needs to extract data from that. To determine the existence of the Higg's boson a Bayesian approach was used, which created a lot of heated discussion about the interpretation of the events, a kind of revival of the quarrel between frequentists and Bayesians. In our case, it is very interesting since we can't achieve a very high number of events: we can't make accurate statistics, so the fit of the data is not really accurate enough and gives sometimes absurd values. By choosing a Bayesian method, we will gain a lot of accuracy and confidence about the result. However, there are also counterparts:

- The Bayesian method needs us to choose an initial distribution for the values of the parameters, an initial "prior". It is an initial assumption which summarizes the knowledge about the parameter. This choice arbitrary but can be improved by determining the prior in the light of some previous experimental data; this approach is called "Experimental Bayes". In our case we can design it, knowing the design of the experiment and the results of previous measurements. Even if in the end

it won't influence the final result due to the big number of updating, this might be an argument against the Bayesian approach. We will discuss this point in 4.2.1.

- Due to the interpretation of the probability, the result of the measurement won't be the value of the sigma with a confidence interval, which means that the real value for any new evaluation will give a result in this interval with a given confidence level. Instead we have to use a "credible interval". It gives the probability of measuring the parameter in this range. In practice, both intervals can be similar but the interpretation is different. We will discuss this point in 4.3.3.
- As said before, the probability is not objective for the Bayesian approach, it depends indirectly on the initial prior. Normally both interpretations should give the same result for a sufficient number of data. In practice, we will see later for example that the data can be interpreted in different ways.

In probability theory and statistics, the most basic expression of the Bayes theorem is:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{20}$$

with A and B two statements. It connects $P(A|B)$, the probability of having A knowing B, also called the posterior, to:

- $P(A)$, the probability of having A independent of the value of B, it is also called the prior, which characterizes the initial degree of belief before acquiring a new evidence. It's an "a priori" probability which can be totally arbitrary.
- $P(B|A)$ the probability of observing B knowing A, also called as the likelihood
- $P(B)$, the marginal probability. It corresponds to the probability to observe A independently of B. It is comparable with a normalizing factor.

One can also write it like:

$$P(A|B) = \frac{P(B|A)}{P(B)} \cdot P(A) \tag{21}$$

where the factor $\frac{P(B|A)}{P(B)}$ translates the impact of the new evidence on the probability $P(A)$.

4.1.2 Bayesian experimental design

How to apply concretely the Bayes theorem to our measurement? Generally speaking, from [Chaloner and Verdinelli, 1995], the application of the Bayes theorem to an experimental setup is called "Bayesian experimental design". It is stated as follow:

$$P(\theta|y,\xi) = \frac{P(y|\theta,\xi)\,P(\theta)}{P(y,\xi)} \tag{22}$$

where:

- θ is a set of parameters we want to determine
- y is the result of the measurement
- ξ is the design of the experiment, which is often the adjustable parameter of the experiment.

To apply it directly to our experiment, if we note σ the standard deviation of the distribution, x_0 the center and x the blade position, the formula becomes:

$$P((x_0,\sigma)|y,x) = \frac{P(y|(x_0,\sigma),x)\,P((x_0,\sigma))}{P(y,x)} \tag{23}$$

where the set of parameters x_0 and σ define a point in the two dimensional parameter space; the x position of he blade is chosen by us, that is the "design" of the experiment. In other words, if we assume an initial distribution of the set of parameters (x_0,σ), we can update our knowledge about the distribution of the parameters by knowing $P(y|x_0,\sigma,x)$ calculated by taking into account the new data point. We can then iterate the process by replacing the initial distribution $P((x_0,\sigma))$ by $P(y|(x_0,\sigma),x)$ evaluated in the previous step. Thus, the knowledge about x_0 and σ is updated constantly and take into account every data acquired.

That is the working principle of the Bayesian experimental design. In theory it can be applied to many different domains; however experimentally it is not so easy to implement. One of the major reasons is that the dimension of the problem depends on the number N of parameters, which can make things really complex, since it would need a N dimensional integration to compute the probabilities. Due to this computation time, most of the experiments which use the Bayesian approach measure a whole set of data before doing a Bayes update before measuring another set. This is how it is done in [Brune et al., 2008], for instance, to count photons in a cavity. It takes advantage of the Bayes' theorem to optimize the measurement. needs to be adjusted dynamically.

However, it was possible to do the Bayesian updating simultaneously with the experiment in our case due to the low number of parameters, the relatively long time between each new data point measurement and the fact that we measure isolated events. This is a dynamical Bayesian experimental design. Concerning our experiment, we actually work in a 3D parameter space: we want to determine x_0 and σ, but we also have to take the experimental imperfection which can be modeled by a single supplementary parameter. We will discuss this issue later in 4.3.1

Nonetheless, even with only 3 parameters, the integration time during the computation can be very long, this is the reason why we had to use a grid, with a size of of $30 \times 30 \times 15$ points approximately, in order to discretize the probability distribution and be able to update this one between each step in a reasonable time. This problem will be discussed in the next part.

4.2 Detailed Explanation of the Bayesian algorithm

We are going to explain in a more detailed fashion the Bayesian algorithm written to optimize the knife edge measurement. First we start with the 2D problem which works well before adding later the supplementary dimension which stands for the experimental noise. The code can be found in the appendix C.

The goal of the algorithm is to determine at which blade position the next ion transmission experiment should be done to maximize the information gained and reiterate the process according to new data acquired.

4.2.1 Initial prior

This Bayesian algorithm assumes that some initial knowledge of the system is available; however even if that is a very rough guess, the program will still converge correctly. The choice of the prior is not without consequence since that a wrong prior can leads to wrong results. This kind of consideration is shortly discussed in [Efron, 2013] as well as a solution which is to use preexisting data and information to choose the prior. In our case, due to the known design of the experiment and the results from previous experiments, one can choose a credible prior with which we can start. A two dimensional Gaussian function can be a reasonable PDF[6]:

$$P_0(x, \sigma; x_0, \sigma_0) = \frac{1}{\sqrt{2\pi}\sigma_{x_0}} \exp\left(\frac{-(x-x_0)^2}{2\sigma_{x_0}^2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma_{\sigma_0}} \exp\left(\frac{-(\sigma-\sigma_0)^2}{2\sigma_{\sigma_0}^2}\right) \tag{24}$$

with σ_{x_0} and σ_{σ_0} the widths of the initial distributions projected along the two axis corresponding to x and σ respectively. This choice is not totally arbitrary, due to the Central Limit Theorem, if we would measure their value a important number of times, in the end the probability distribution would be approximated well by a Gaussian distribution. Another argument is that the Gaussian distribution is a conjugate prior, it means that the result $P(A|B)$ of a Bayesian updating is also a Gaussian if we choose $P(A)$ as being a Normal distribution. The results won't diverge.

As we said before, we are working with a limited grid which is save between each step. The initialization consists in initializing the grid with the presumed distribution; there are also different initializations for the variables, the graphs and so on that we won't detail here. What's follows is already in the loop. In practice, this probability is implemented in C++ by a multi-dimensional table[7] p_theta.

4.2.2 Maximum information finding

This is the first step of the optimization process, therefore we have to decide of the blade position to gain as much information as possible during the measurement. To quantify the information gain, we can

[6] Probability distribution function

[7] A vector would have been preferred but this functionality hasn't been implemented yet.

calculate the utility $U(x)$ which is the gain in Shannon information. In information theory, the Shannon entropy characterizes the unpredictability of a random variable, in other words the information content of this variable.

By computing the gain in Shannon entropy, defined as the utility, at different positions, we can tell in which region the measure will bring the most information. The utility for a given blade position ξ is obtained by summing the utility depending on y and ξ over the outcomes y:

$$U(\xi) = \sum_{y \in \{0,1\}} U(\xi, y) = D_{KL}\left(P(\theta|y,\xi) \,\|P(\theta|\xi)\right) \tag{25}$$

Where $U(\xi)$ is given by the following formula:

$$U(y,\xi) = \int \log\left(P(\theta|y,\xi)\right) P(\theta|y,\xi)\,\mathrm{d}\theta - \int \log\left(P(\theta)\right) P(\theta)\,\mathrm{d}\theta \tag{26}$$

which also corresponds to the definition of the Kullback-Leibler divergence, used in probability theory and information theory. So we can write:

$$U(y,\xi) = D_{KL}\left(P_n(\theta|y,\xi) \,\|P_{n-1}(\theta|\xi)\right) \tag{27}$$

Since we used discrete grid instead of continuous variables, the Kullback-Leibler divergence for the step number n takes the following form:

$$D_{KL}\left(P_n(\theta|y,\xi) \,\|P_{n-1}(\theta|\xi)\right) = \sum_{\theta} \ln\left(\frac{P_n(\theta|y,\xi)}{P_{n-1}(\theta|\xi)}\right) P_n(\theta|y,\xi) \tag{28}$$

where $P(\theta|\xi)$ is the distribution probability for the $n-1$ step. Even if it doesn't appear in the notation, $P(\theta|\xi)$ still contains the information about the outcome of the measurement $n-2$, $n-3$, and so on, but not explicitly. In practice we need a fast numerical computation, this is why we chose a discrete grid and why we have discrete summation.

In our case, with the notations of the previous section, this divergence is written like this:

$$D_{KL}\left(P_n((x_0,\sigma)|y,\xi) \,\|P_{n-1}((x_0,\sigma)|\xi)\right) = \sum_{x_0=x_{0,min}}^{x_{0,max}} \sum_{\sigma=\sigma_{min}}^{\sigma_{max}} \ln\left(\frac{P_n((x_0,\sigma)|y,\xi)}{P_{n-1}((x_0,\sigma)|\xi)}\right) P_n((x_0,\sigma)|y,\xi) \tag{29}$$

where we sum over all the possible values of x_0 and σ. As a reminder, the probabilities used here are the following ones:

- $P_{n-1}((x_0,\sigma)|\xi)$ is the probability distribution for the previous step, which is stored in the array p_theta,
- $P_n((x_0,\sigma)|y,\xi)$ is the new probability distribution after this step n.

The latter is estimated with the Bayes's theorem:

$$P_n((x_0,\sigma)|y,\xi) = \frac{P(y|(x_0,\sigma),\xi)\, P_{n-1}((x_0,\sigma))}{P(y,\xi)} \tag{30}$$

where:

- In this case, $P(y|(x_0,\sigma),\xi)$ plays the role of the likelihood. This the probability distribution when knowing y and the parameters. Here it corresponds to the model used to describe the experiment:

$$P(y|(x_0,\sigma),\xi) = \frac{1}{2}\mathrm{erfc}\left(\frac{\pm(\xi - x_0)}{\sqrt{2}\sigma}\right) \tag{31}$$

  according to the direction of the scan. We called it *designfunction* in the algorithm. Note that we can reverse the distribution only by switching the sign, due to the properties of the error function.

- $P_{n-1}((x_0,\sigma))$ is the prior. In this case, the value is given by the values of the two dimensional array evaluated during the previous steps. Or $P_0(x,\sigma;x_0,\sigma_0)$ for the first iteration.

- $P(y,\xi)$ is the marginal, which is a kind of normalization. It's computed in the following way:

$$P(y,\xi) = \sum_{x_0=x_{0,min}}^{x_{0,max}} \sum_{\sigma=\sigma_{min}}^{\sigma_{max}} P_{n-1}((x_0,\sigma))\, P(y|(x_0,\sigma),\xi) \tag{32}$$

  which is the particular case for a discrete summation over x_0 and σ of:

$$P(y,\xi) = \int_\theta P(\theta)\, P(y|\theta,\xi)\, \mathrm{d}\theta \tag{33}$$

Then, we evaluate the Kullback-Leibler divergence for different ξ positions in order to find the maximum. To do so, we used this simple recursive algorithm:

```
const int recursionN=10;//number of steps of the recursion

double BKE::findmax_Kullback_Leibler(double start,double stop,int cnt){

        double maxval=-DBL_MAX,val;

        double storexi=start;

        for(double xi=start;xi<stop;xi+=(stop-start)/double(recursionN)){

                val=calc_Kullback_Leibler(xi);

                if(val>maxval){

                        maxval=val;

                        storexi=xi;

                }

        }

        if(cnt>0) return findmax_Kullback_Leibler(
        storexi-(stop-start)/double(recursionN),
        storexi+(stop-start)/double(recursionN),--cnt);

        else return storexi;

}
```

Where *calc_Kullback_Leibler* computes the Kullback-Leibler divergence for a certain blade position as described previously. From this part of the algorithm, we obtain the blade position for which the information gain will be the highest. The points are mainly measured where the curve begin to fall down or to rise. Intuitively we understand that the distance between the two positions give the width of the PDF[8]. There are also points measured at the middle which maybe stands for the mean value determination. The dispersion of the points looks like as follows on figure 8.

[8] Probability distribution function

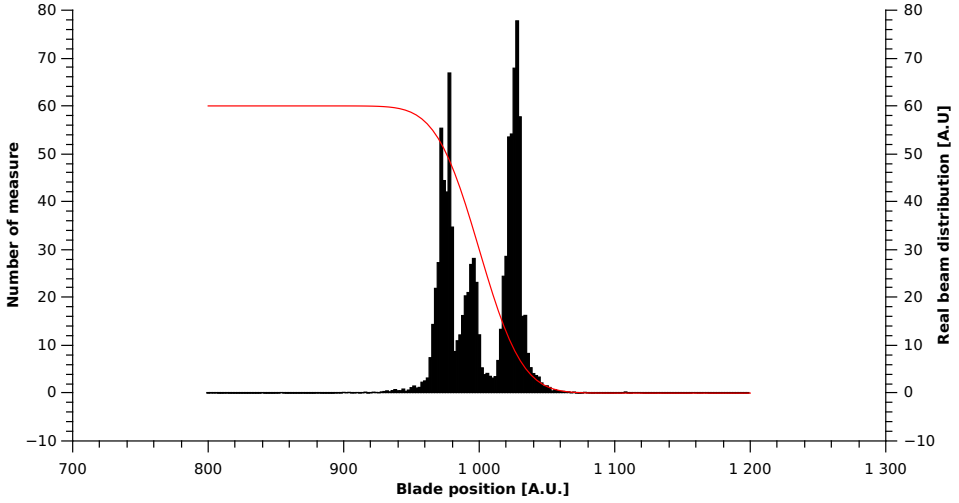


Figure 8: Histogram of the repartition of the points for 100 independent simulations with 1000 points, taking the noise into account as explained later in 4.3.1. The noise introduce an asymmetry since there is no dark count, only "hit" loss.

4.2.3 Measurement and distribution updating

At this point we measure an ion transmission, experimentally or by a simulation, at the blade position guessed before. We obtain a binary outcome: $y = 0$ or $y = 1$. From this result, we now have to update the probability distribution of x_0 and σ, that means updating our knowledge of x_0 and σ.

First, we need to evaluate the current mean value:

$$\begin{aligned} \overline{x_0} &= \sum_{x_0,\sigma} x_0 \cdot P_n\left((x_0,\sigma)\right) \\ \overline{\sigma} &= \sum_{x_0,\sigma} \sigma \cdot P_n\left((x_0,\sigma)\right) \end{aligned} \tag{34}$$

and the associated variances:

$$\begin{aligned} \mathrm{Var}\left(x_0\right) &= \sum_{x_0,\sigma} \left(x_0 - \overline{x_0}\right)^2 \cdot P_n\left((x_0,\sigma)\right) \\ \mathrm{Var}\left(\sigma\right) &= \sum_{x_0,\sigma} \left(\sigma - \overline{\sigma}\right)^2 \cdot P_n\left((x_0,\sigma)\right) \end{aligned} \tag{35}$$

in order to update the parameter PDF afterwards. The probabilities $P_n\left((x_0,\sigma)\right)$ are given by the array p_theta. Theses values give the current estimation of the parameters and the uncertainty[9] on these values. In the end, the final result will be these values given at the last step. The more steps we do, the higher the precision should be.

We can now evaluate the new probability taking the last transmission measurement into account. To do so, we have to use the Bayes theorem again, replacing every value of the array:

$$P_{n+1}\left((x_0,\sigma)\right) = \frac{P\left(y|\left(x_0,\sigma\right),\xi\right) P_n\left((x_0,\sigma)\right)}{P\left(y,\xi\right)} \tag{36}$$

where $P\left(y|\left(x_0,\sigma\right),\xi\right)$ is similar to $P\left(y|\left(x_0,\sigma\right),\xi\right)$ in the previous part , our model for the distribution of the transmission probability, and $P\left(y,\xi\right)$ is similar to the same expression in 4.2.2 , the normalization factor.

At this point of the algorithm, we have updated the values of the parameters by adding the small amount of information we gathered from measuring one more point. The great advantage of this method is that we can repeat these steps as often as we want to obtain, theoretically, a variance as small as we want. We will discuss the accuracy and the relevance with the support of the simulations in the part 5 .

[9] Note for later that this is not totally correct

4.3 Refinement of the algorithm

4.3.1 3rd dimension

For simplicity, so far we have worked only with a two dimensional space containing the mean value of the distribution and its width. However in practice, because of experimental reasons, especially the detector efficiency, we detect less ions which means that the transmission probability is never 100% which mainly tends to increase the width in a 2D space. However, we don't observe any dark count in the experiment, this cause an asymmetry in the PDF. To take this effect into consideration, we add another parameter which we will call P_{max} and change the model as follows:

$$P_x = \frac{P_{max}}{2}\text{erfc}\left(\frac{\pm(\xi - x_0)}{\sqrt{2}\sigma}\right) \tag{37}$$

That also means that we need to work in a three dimensional space and optimize for this new parameter. On the other hand, it increases the computation time a lot. The process is the same except that we add another parameter, so in equation 22, θ stands for the set of parameters $\{x_0, \sigma, P_{max}\}$. The likelihood becomes:

$$P\left(y|\left(x_0, \sigma, P_{max}\right), \xi\right) = \frac{P_{max}}{2}\text{erfc}\left(\frac{\pm(\xi - x_0)}{\sqrt{2}\sigma}\right) \tag{38}$$

The initial prior becomes:

$$\begin{aligned} P_0(x, \sigma, P_{max}; x_0, \sigma_0, P_{max,0}) = &\frac{1}{\sqrt{2\pi}\sigma_{x_0}}\exp\left(\frac{-(x-x_0)^2}{2\sigma_{x_0}^2}\right) \times \frac{1}{\sqrt{2\pi}\sigma_{\sigma_0}}\exp\left(\frac{-(\sigma-\sigma_0)^2}{2\sigma_{\sigma_0}^2}\right) \\ &\times \exp\left(\frac{-(P_{max}-P_{max,0})^2}{2\sigma_{P_{max}}^2}\right) \end{aligned} \tag{39}$$

and the marginal becomes:

$$P\left(y, \xi\right) = \sum_{x_0=x_{0,min}}^{x_{0,max}} \sum_{\sigma=\sigma_{min}}^{\sigma_{max}} \sum_{P_{max}=P_{max,min}}^{P_{max,max}} P_{n-1}\left(\left(x_0, \sigma, P_{max}\right)\right) P\left(y|(x_0, \sigma, P_{max}), \xi\right) \tag{40}$$

Every measurement done in the laboratory will use the three dimensional algorithm since the transmission efficiency is never 100%.

4.3.2 Zoom

Normally, this algorithm is able to work with any initial range for the blade position; however we are rapidly limited by the size of the grid. Due to the pixel size, we lack for resolution which keeps us unable to decrease the uncertainty of the sigma any further and causes also calculation errors.

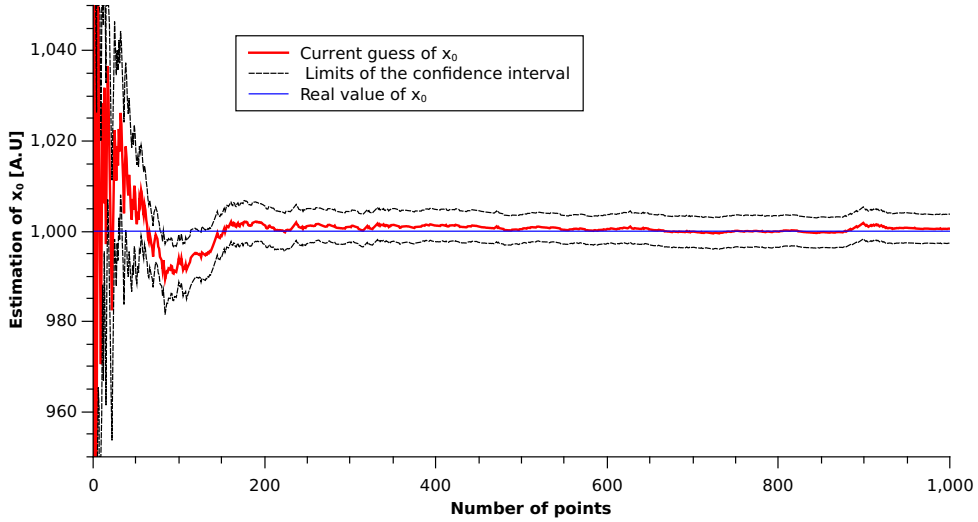
To overcome this problem one had to adapt the size of the grid to the broadness of the parameter distribution. In other words, we have to "zoom" where the distribution of the parameters has its highest value. Basically, we keep the same grid size[10] but we rescale the distribution to fit the "region of interest". This is no trivial operation. Although the distribution can be easily approximated by a Gaussian function, whose value is very small far away the center, every time we resize the picture, we lost a little bit of information contained in the tail of the distribution.

The algorithm works as follows: If the edge of the picture is nearer than a given σ_{x_0}, σ_{σ_0} or $\sigma_{P_{max}}$ from the center for the corresponding parameter, then we delete one row and rescale the image to the array. This is contained in the function *zoomin*. We wrote a similar function to zoom out of the picture. this is useful when the experimental system is drifting and that the center of the distribution tends to go out of the array. As for the inverse operation, we lose some information by compressing and interpolating the data points. Moreover the probability of the new line is set to 0, which is not so far from the real value, but it is still some made-up data.

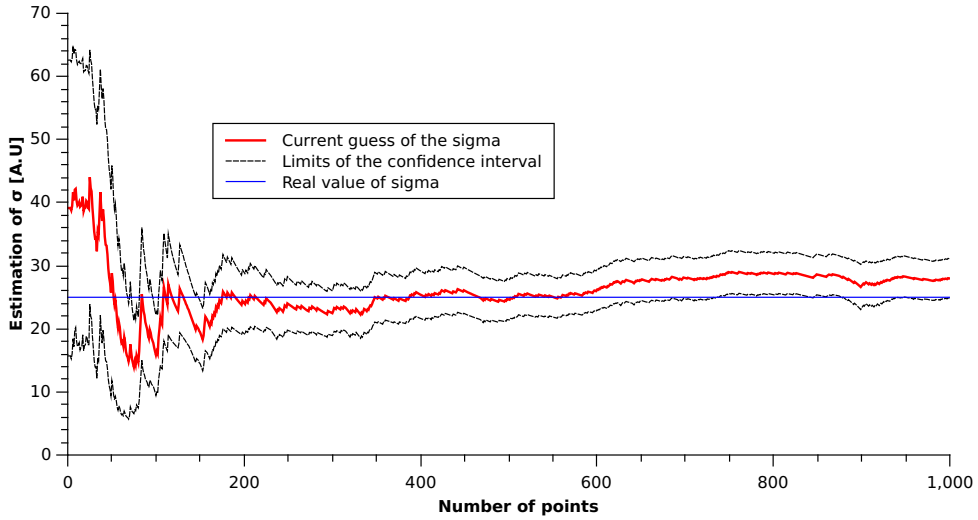
This function doesn't allow zooming fast enough when working with grids with more points.

At this point a simulation looks like figure 9.

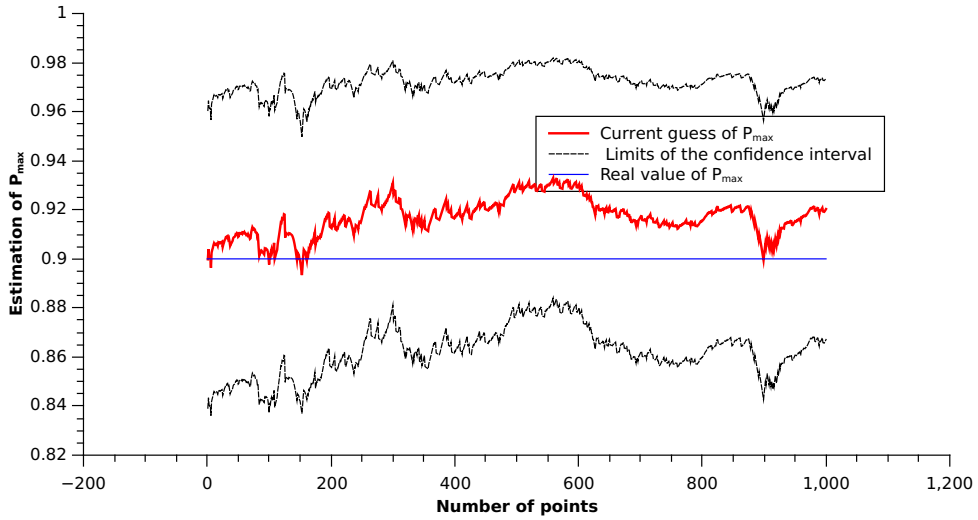
[10] Which is fixed anyway since we work with tables and not vectors



(a) Simulation graph for the convergence of x_0



(b) Simulation graph for the convergence of σ



(c) Simulation graph for the convergence of P_{max}

Figure 9: Simulation of a knife-Edge measurement with the Bayesian algorithm in three dimensions, with a zoom activated, for a measurement of 1000 points. The size of the grid is $30 \times 30 \times 20$. The real values are: $x_0 = 1000$, $\sigma = 25$ and $P_{max} = 0.9$. The final guess for 1000 points are: $x_0 = 1000.56 \pm 3.21233$, $\sigma = 27.9807 \pm 3.1207$ and $P_{max} = 0.920306 \pm 0.0531696$. *The ranges are only a hint* and are not coherent with a Bayesian interpretation even if their value is acceptable.

4.3.3 Confidence and credible interval

Until now we have masked a non-trivial problem: how to express the uncertainty of the result? In a frequentist interpretation, we use an interval which is called "confidence interval", while Bayesians use a "credible interval". Even if in practice their values can be very similar, their interpretations are fundamentally different. We can find a good explanation in [Porter, 2013]

Let's consider an interval $p_l < p < p_u$ with p the mean value and a given probability level, 68% for instance.

In the frequentist interpretation, the ends of the range p_l and p_u changes for each new measurement and are considered like random variables. The meaning of the confidence interval is that if we repeat the experiment many times, 68% of the time the considered range would contain the true value of p. It doesn't say anything about if the mean value of p is currently in the range for the current data set! It just gives information about p_l and p_u. We don't know the probability of having the true value in the range but we do know how often the range will be right. It is often misunderstood with the Bayesian interpretation. To estimate this interval one usually uses the standard deviation. It makes only sense if the experiment can be repeated: If we measured all the samples in the universe[11], the parameter value has to be the mean since it takes a fixed value. Here, the standard deviation just characterizes the value dispersion. In other words, if we repeat the experiment, 68% of the time, the interval $p_l < p < p_u$ will contain the true value.

In the Bayesian interpretation however, the limits of the range p_l and p_u are fixed and determined by the posterior distribution. In this case, the credible interval of 68% tells us that p has 68% of being in this range, but there is still a probability of 32% for p to be outside this range. The Bayesian approach doesn't predict anything about what could be measured next, only gives an estimation about the current estimation! Note that this interval depends on the prior distribution whereas in the frequentist interpretation it only depends on the data.

Since we are working in a fully Bayesian way, one should to work with a credible interval. We estimate it in the easiest way as described in [Yong-Sheng, 2008]. We choose a credible level $CL = 1 - \alpha$ with $\alpha \in [0, 1]$ to determine the credible interval $[\theta_l, \theta_u]$ where $\theta = x_0, \sigma$ or P_{max} is the parameter of which we want to determine the credible interval. The lower bound can be defined by:

$$\int_{-\infty}^{\theta_l} P_n\left((x_0, \sigma, P_{max})\,|y, \xi\right) = \frac{\alpha}{2} \tag{41}$$

And the upper bound defined by:

$$\int_{\theta_u}^{\infty} P_n\left((x_0, \sigma, P_{max})\,|y, \xi\right) = \frac{\alpha}{2} \tag{42}$$

In practice, the probabilities are given by the array p_theta and the boundaries are evaluated numerically in the following way:

```
sum=0;int i=0;double dx=(stopMV-startMV)/(p_theta_size-1);

        while(sum<1-alpha){

                sum+=p_MV[i];i++;

        }

        i=i-1;

        var2[0][0]=abs(mean[0]-(startMV+i*dx));
```

for the evaluation of the lower bound of the x_0 interval.

This is a quite rough approximation:

- We do a small error since we don't interpolate by simplicity. Thus it is works well for large grids.

[11] In the mathematical meaning

- This is correct if the distribution has one central peak and at the edge the probability is nearly 0, which is not always the case for P_{max}.

There are many ways to determine the credible interval depending on the shape, but in our case this simple estimation should be enough.

Note that contrary to the frequentist range, the limit can be asymmetric with respect to the center and the interval won't have any non-physical values: a negative σ or $P_{max} > 1$ for instance. However since the method is computationally consuming and since the grid is too small during the data acquisition, we use it only during the consecutive analysis as described in the next part. Without this interval the results can't be evaluated rigorously.

4.3.4 Analysis of the data

One great functionality is the ability to recalculate the whole experiment afterwards in order to increase the precision or to reinterpret the data in the light of a new prior. In the frequentist approach, all the data is evaluated only afterwards with statistical methods. In the case of the Bayesian approach, the evaluation is quite complex to do in real time due to a permanent updating of the distribution probabilities, but it is possible to do it afterwards with higher computational power: it is often the way it is done in particle physics where few events are used to state about a property of a particle or its existence with a latter computation.

In our case, once we have the data, we can recalculate it with better approximations for the parameter by computing the "story" of the experiment from the start. To do so, we just use the procedure to run another measurement but instead of using simulated points or experimental points, we use the data points from a previous experiment or simulation. A simulation is shown in 5.1. Since we also know the result of the previous measurement, one can reduce the interval parameters and thus using a better initial prior. This solve one of the main critics about the Bayesian approach, which is the arbitrary choice of the initial prior, by using the experimental results we can justify this choice. The computation time is even lower since we don't have to maximize the utility of each point as described by 25.

So we can improve the accuracy of the parameter values only by computing them again with better approximations. This step is necessary to lessen the calculation errors due to the rough discretization. But that highlights the point that even if the data is still the same, their interpretation from the Bayesian point of view can differ only by changing the initial prior. In other words, one has to be careful with the final result which intrinsically depends on the initial prior; it only represents which is the most probable value of the parameter starting from a given prior.

5 Simulation, results and discussion

In this section, we are going to present the results, mainly of simulation, before comparing with the frequentist method.

5.1 Bayesian recalculation

One of the main problems is to produce a result with an uncertainty which is plausible and give a good estimation of the value of x_0,σ and P_{max}. As we said before, one can recalculate[12] the result to counteract the numerical approximation and to give a right estimation for σ.

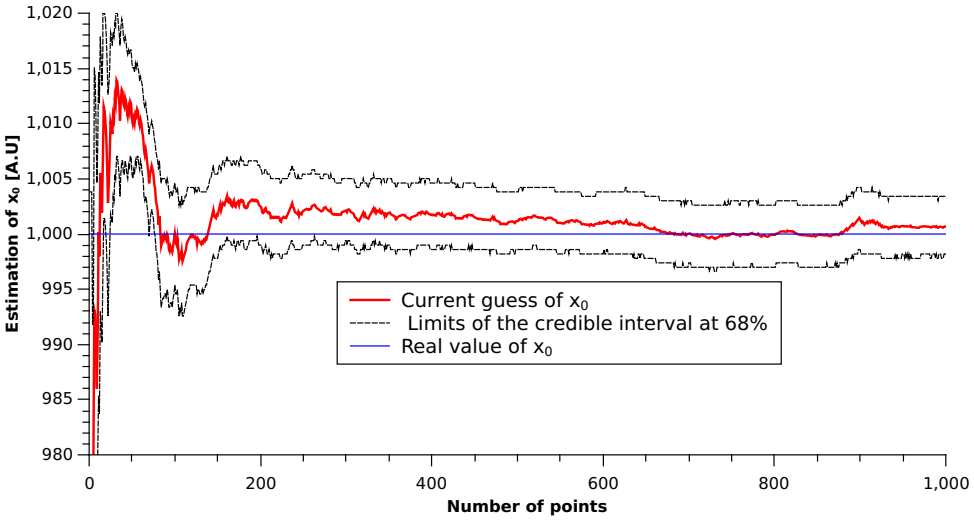
We first did a recalculation plotted in figure 10 of the simulation shown in 9, with a credible interval having a credible level of 68%, i.e. corresponding to the usual confidence level we usually use to present results.

It's obvious that the algorithm doesn't converge really better than before even if we increased the precision. We tried also with even higher grid, that doesn't change the result. The deactivation of the zoom removed an additional error, but since the initial range is narrower, it wouldn't be of any use. We also tried with another simulation 30000 points, the convergence doesn't seem to become better and by the way measurement with more than 1000 points begins to be very long experimentally. At some point we are limited by the set of data we measured. There are different explanations possible for this limitation of accuracy. The first one is that since the Kullback-Leibler optimization works with a smaller grid, at some point the algorithm returns values for the next measurement which doesn't give any new information avoiding to reach the right value. The second hypothesis is that we are limited by the number of data points: the design function is highly non-linear and all the parameters are correlated. This is particularly true for P_{max}.

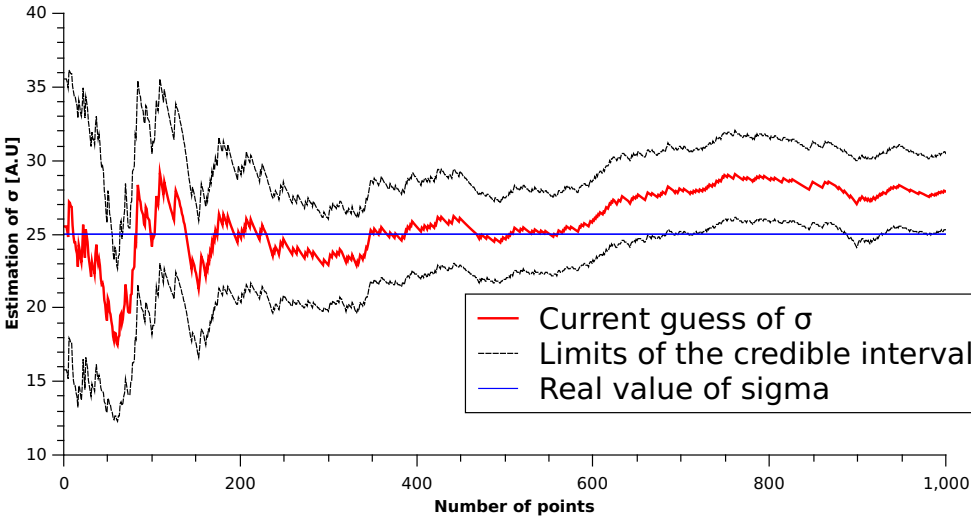
The calculation of the credible interval has to be polished. The discretization limits its precision and the way to calculate it is questionable even if not fundamentally false. For instance, we set the higher boundary of P_{max} credible interval to one since the distribution is quite broad and its value is more than 0 at any time. We can also change the credible level to higher value to give a result which can be believed easily. In the case we choose a 95% credible level, meaning that the value of the parameter has 95% chance of being in this range. Such a graph is shown on figure 11.

With a credible level of 95% the real value is nearly all the time in the error range; we can confirmed that with other simulation. It is not possible however to determine experimentally if the values of the parameters are really close to the "true" one, this is the reason why it is so important to have a result which can be trusted in any situation.

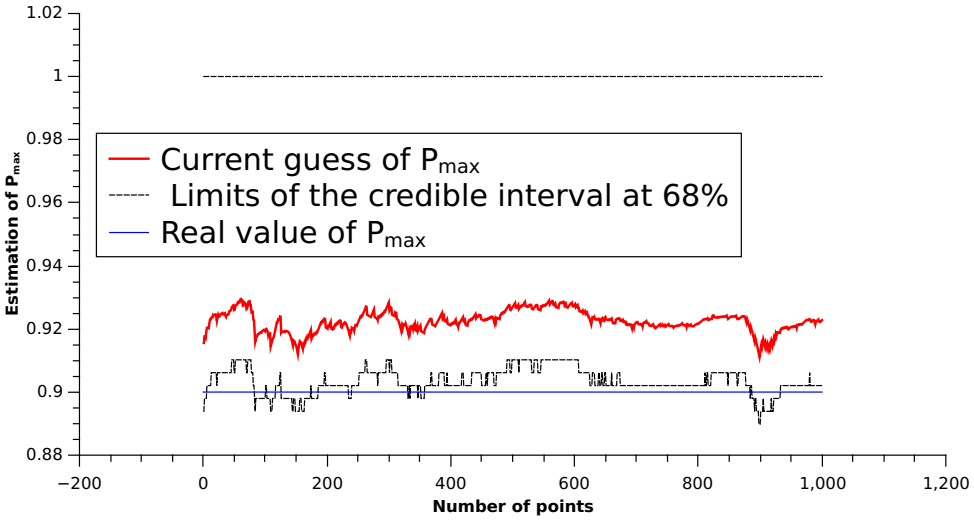
[12] The reason is only the computation time



(a) Simulation graph for the convergence of x_0

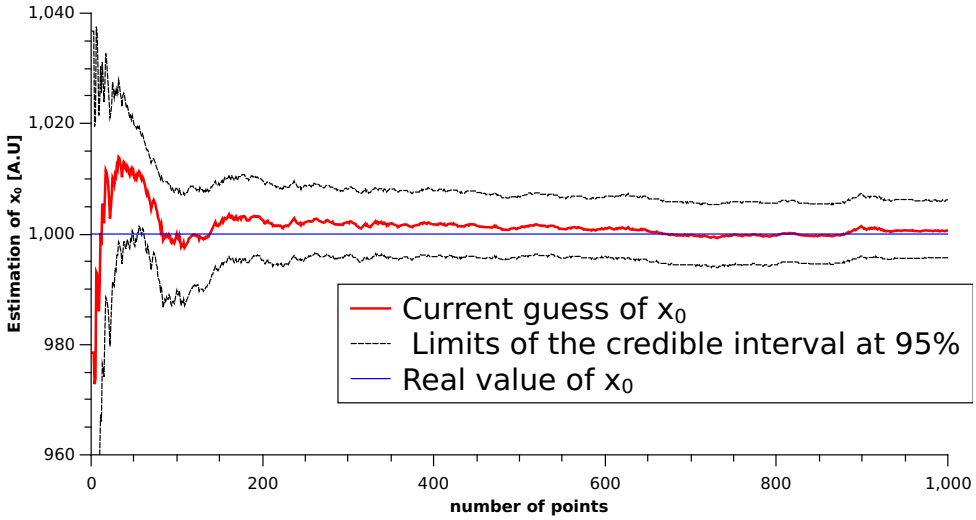


(b) Simulation graph for the convergence of σ

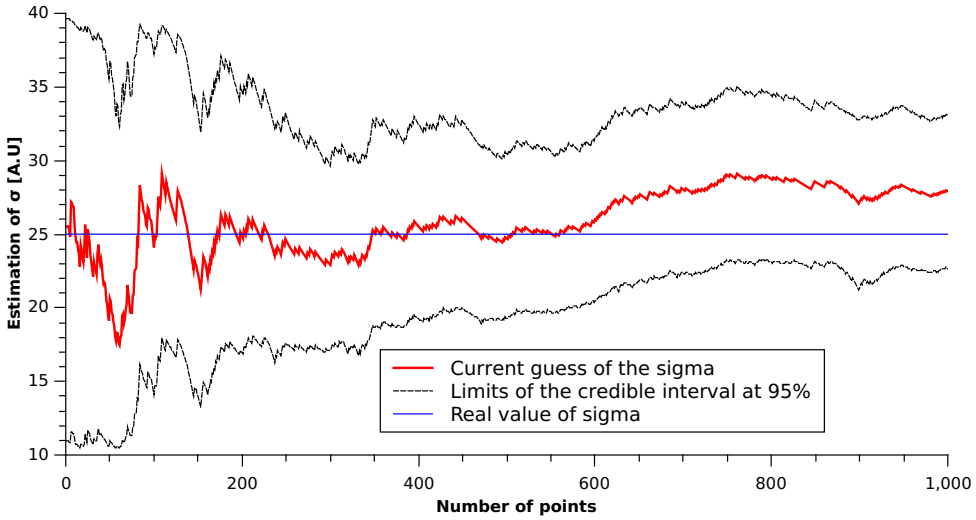


(c) Simulation graph for the convergence of P_{max}

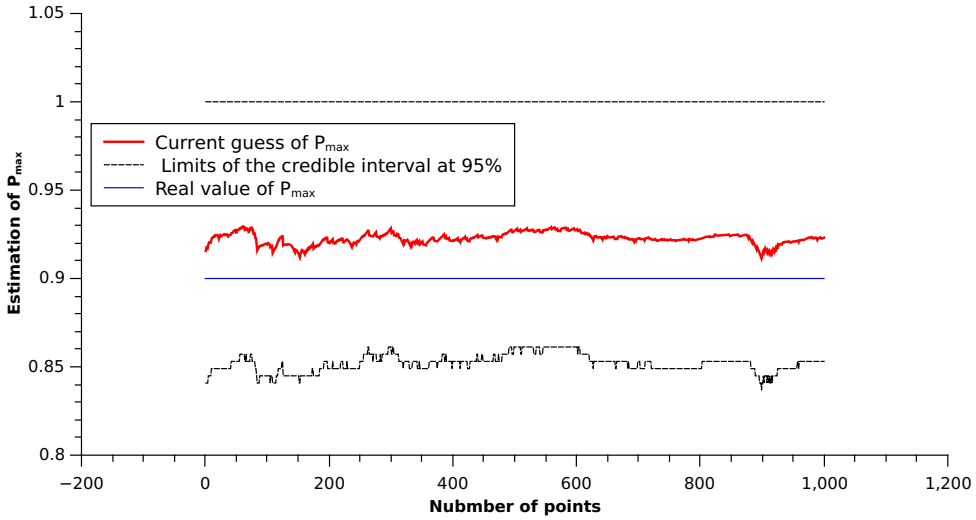
Figure 10: Simulation of a knife-Edge measurement with the Bayesian algorithm in three dimensions, without zooming, for a measurement of 1000 points. The size of the grid is $250 \times 250 \times 50$. The real values were: $x_0 = 1000$, $\sigma = 25$ and $P_{max} = 0.9$. The final guess for 1000 points are: $x_0 = 1000.7 + 2.7 - 2.5$, $\sigma = 27.9 + 2.6 - 2.59735$ and $P_{max} = 0.923 + 0.077 - 0.021$.



(a) Simulation graph for the convergence of x_0



(b) Simulation graph for the convergence of σ



(c) Simulation graph for the convergence of P_{max}

Figure 11: Simulation of a knife-Edge measurement with the Bayesian algorithm in three dimensions, without zooming, for a measurement of 1000 points. The size of the grid is $250 \times 250 \times 50$. The real values were: $x_0 = 1000$, $\sigma = 25$ and $P_{max} = 0.9$. The final guess for 1000 points are: $x_0 = 1000.6 + 5.6 - 5.0$, $\sigma = 27.9 + 5.1 - 5.3$ and $P_{max} = 0.923 + 0.077 - 0.070$.

5.2 Accuracy comparison

We saw that we can't increase indefinitely the accuracy of the parameter values but we at least want to know how wrong we are and compare it to the previous method to see what we have gained. The idea here is to repeat the experiment many times and to see the distribution of values. Since experimentally it is not possible, we can at least simulate it: both for the frequentist and Bayesian approach, we repeated the experiment 100 times independently and plot the results. This gives a hint about how often the different algorithms converge to the true value.

5.2.1 Frequentist Knife-Edge: result dispersion

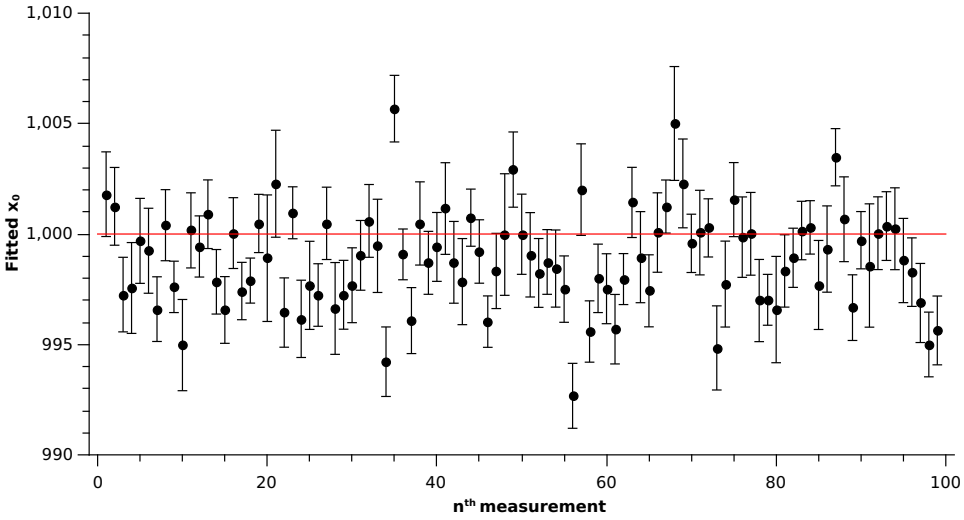
The number of ions for each measurement is around 1000 ions but not always constant depending how fast the algorithm finds the maximum. If we do basics statistics on these values, one finds:

$$\begin{aligned} x_0 &= 998.9 \pm 2.2 \\ \sigma &= 24.0 \pm 2.7 \\ P_{max} &= 0.92 \pm 0.03 \\ \sigma_{x_0,0} &= 2.5 \\ \sigma_{\sigma,0} &= 2.9 \\ \sigma_{P_{max},0} &= 0.03 \end{aligned}$$

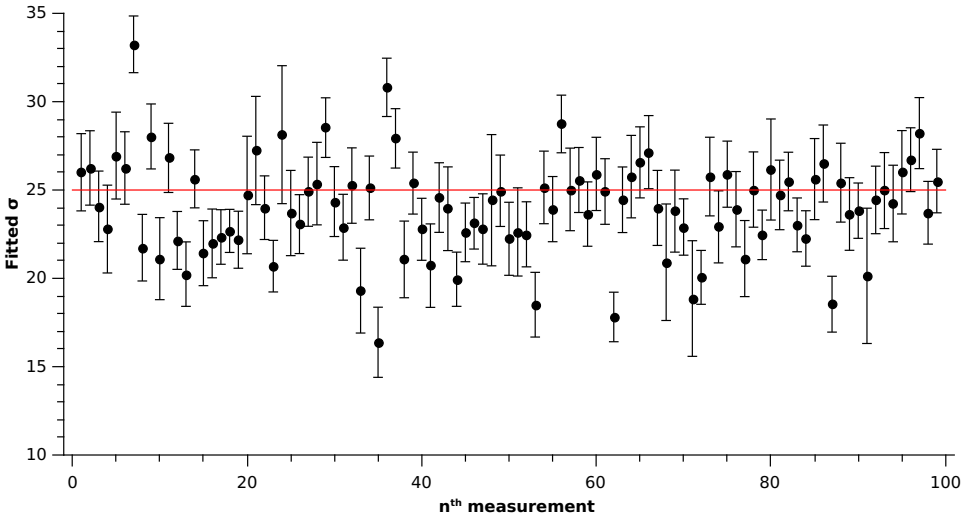
where $\sigma_{x_0,0}$,$\sigma_{\sigma,0}$ and $\sigma_{P_{max},0}$ are the standard deviations from the real values. These values have to be compared with the value of the first method

$$\begin{aligned} x_0 &= 998.3 \pm 6.7 \\ \sigma &= 24.0 \pm 5.9 \\ P_{max} &= 0.93 \pm 0.02 \\ \sigma_{x_0,0} &= 6.9 \\ \sigma_{\sigma,0} &= 6.0 \\ \sigma_{P_{max},0} &= 0.03 \end{aligned}$$

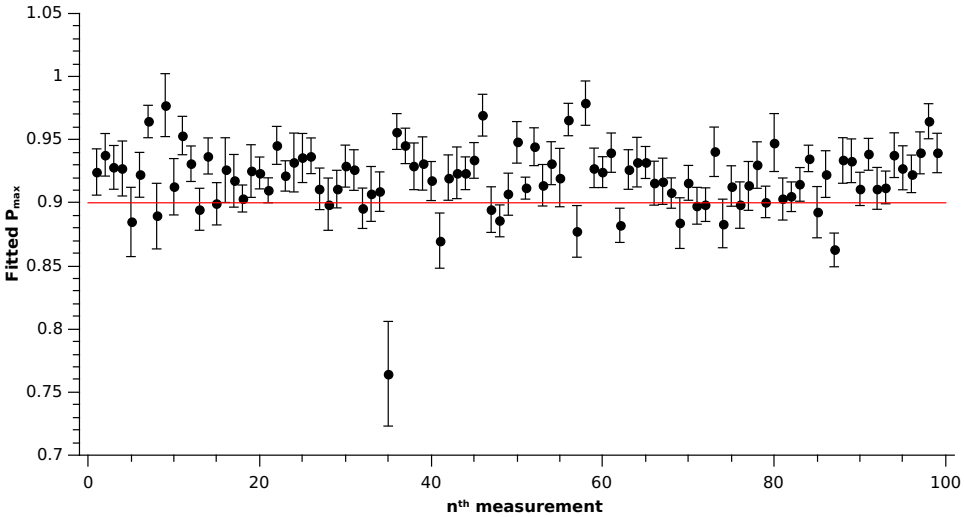
There is clearly an increase in performance, particularly when one observes the standard deviations from their true value. That means that the values are closer to the true value with this method. This point is important because we can't experimentally check how far from the real value we are and one need to trust "blindly" the result.



(a) Final value of x_0 for independent measurements



(b) Final value of σ for independent measurements



(c) Final value of P_{max} for independent measurements

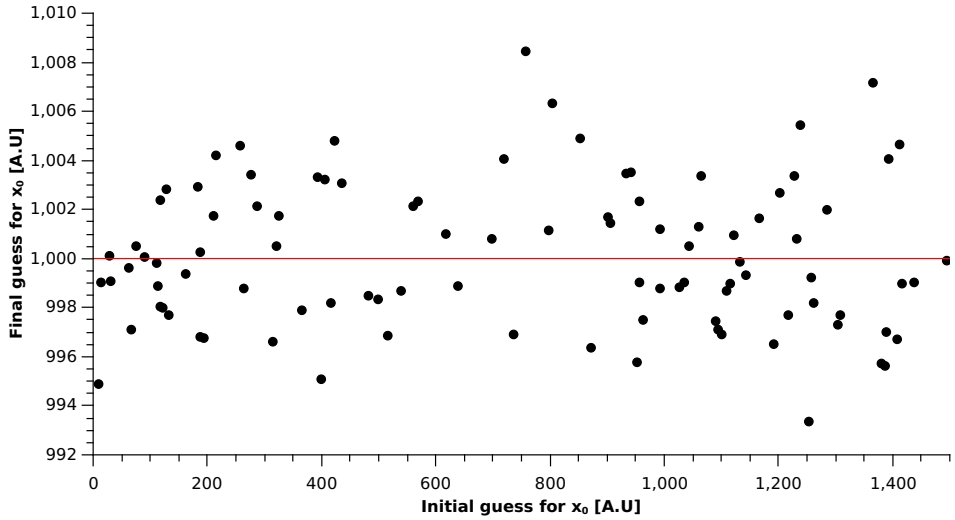
Figure 12: Dispersion of the fitted value for independent Knife-Edge measurements with the frequentist method. The real values were: $x_0 = 1000$, $\sigma = 25$ and $P_{max} = 0.85$.

5.2.2 Bayesian Knife-Edge: result dispersion

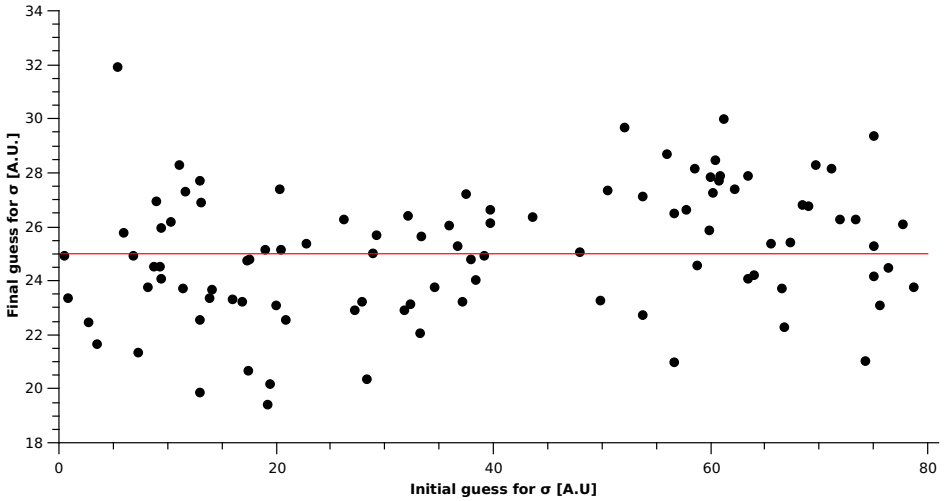
The similar simulation has been done for the Bayesian knife-edge measurement, the results are plotted on figure 13. The first thing we see is that the final guess doesn't seem to depend from the initial guess we chose; the algorithm seems to works relatively well for a broad range without having too much problem of convergence. Experimentally, we always know roughly where the beam and the blade are, so it works fine. We can now compare this with the two previous methods by doing statistics on the measurements:

$$\begin{aligned} x_0 = &\quad 1000.1 \pm 2.9 \\ \sigma = &\quad 25.1 \pm 2.4 \\ P_{max} = &\quad 0.90 \pm 0.05 \\ onewith\sigma_{x_0,0} = &\quad 2.9 \\ \sigma_{\sigma,0} = &\quad 2.4 \\ \sigma_{P_{max},0} = &\quad 0.07 \end{aligned}$$

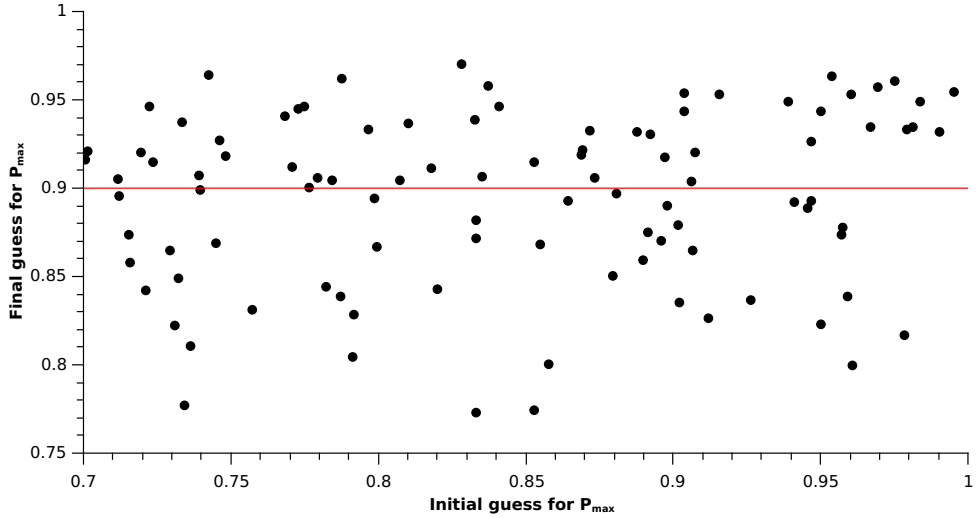
The Bayesian method seems to provide results which are closer than the basic approach but very similar to the results given by the frequentist method. Until now the reasons for that are not so clear. One possible explanation would be that 1000 data points is high enough to provide good statistical estimation of the different probabilities and therefore a good fit. Another possible explanation is that the Bayesian algorithm "forget" some information when going for a very high number of points: all the information is contained in the distribution, since the array use to store the distribution has a finite size, the small errors are accumulating in the distribution, modifying the "memory" of the experiment. One other possible limitation is that, since we use independent measurement, at some point any new measurement will counteract the previous point, creating a kind of “random walk”: the width is limited by some excursions from the real value due to the randomness of the process. However the details are not clear due to the complexity of the Bayesian updating.



(a) Final guess of x_0 according to their initial guess



(b) Final guess of σ according to their initial guess



(c) Final guess of P_{max} according to their initial guess

Figure 13: Simulation of a knife-Edge measurement with the Bayesian algorithm in three dimensions for a measurement of 1000 points with random initial guess in a research range. The size of the grid is $30 \times 30 \times 15$. The real values were: $x_0 = 1000$, $\sigma = 25$ and $P_{max} = 0.85$. The initial research area were: $0 < x_0 < 1500$, $0 < \sigma < 80$ and $0,7 < P_{max} < 1$

5.3 Discussion

So far, we couldn't determine if one method allows to measure the focus more precisely than the other. The dispersion of the final values shows that both optimizations are still better than the basic method but none give really better result according to the final prediction of the real value.

However from a practical point of view the Bayesian algorithm is far better because of its robustness and its usability: one only to enter a large research range and the algorithm will find itself which region to measure. On the other hand, the frequentist methods can become very inefficient if the initial parameters of the program are not carefully chosen, this can lead to the measurement of a huge amount of useless data and therefore a loss of time. Moreover for an efficient measurement the procedure needs a human supervision.

The next step will be to measure the uncertainty according to the number of points, this will give a hint about how fast and how accurately each method converges, for that we evaluate the Allan variance [Allan, 1966], which is a good way to quantify this property. For the Bayesian method it is easy as already implemented; however for the frequentist approach we will have to modify the code to simulate it and implement it in the experiment. It will be rather easy with the basic method, but less with the optimization method we presented. The problem will be the number of ions we can afford before starting this measurement: for the frequentist method, one always needs a small data set to make an initial fitting and determine where is the middle and the sigma as well.

Some improvements could be made for each measurement by using some trick from the other method:

- For the frequentist measurement, instead of measuring equally distributed points in the region of interest, we could use the Kullback-Leibler divergence to determine beforehand where are the interesting positions to measure.
- For the Bayesian method, we could measure few ion transmissions, for example 10, and work with the probability estimated from the 10 measures instead of working with 10 independents probabilities whose values are either 0 or 1. Actually we made some short tests, it turns out that the mean value of each parameter is more stable but the error range is roughly the same. It's actually just a trick to suppress the problem of the "random walk", but in the end the result won't change.

We still also have to determine which method has to be used to present following results of the experiment. Theoretically, for a big amount of points, neither the frequentist nor the Bayesian approach should be better. Since both methods seems to be on an equal foot according to the performance, another point which can tip the scales is how we want to interpret the result: with a Bayesian or Frequentist interpretation? The real value of the beam diameter itself doesn't really matter since we want to find a estimation of the achievable resolution, for this purpose a Bayesian interpretation with a credible interval seems more appropriate because of the fixed limits of the credible interval. A frequentist measurement would be more appropriate to find which minimal beam size we managed to achieve. However in practice the difference won't make any big difference.

5.4 Experimental results

Due to a lack of time and independent experimental problems, we couldn't measure the focus of the single ion beam. However the method itself works fine but no interesting results have been measured yet.

6 Conclusion

6.1 Conclusion and outlook

During this internship we compared mainly two ways of measuring the size of a single ion beam. These two methods turned out to come from two very different theoretical backgrounds: the frequentist and the Bayesian interpretation. Neither one of them is more "correct" than the other one but their ranges of implementation are really different. The frequentist approach works usually well for most of the measurements in Physics[13] and this is the one used most of the time for other knife edge measurement. However it turns out that it doesn't fit so well for the estimation of the size of single particle beam, where few events are observed.

We first tried to optimize this measurement method to obtain better results with the same amount of points by reducing first the range of scan. It appears to work well and could be used.

Then we tried a Bayesian method which was a success for many reasons. First it gave more reasonable results and converges nearly all the time, contrary to the frequentist measurement. Then it is really easier to use far less demanding experimentally. Last but not least it is one of the rare physics experiments which makes a full use of the experimental design: The Bayesian has often been used to analyze data afterwards, but it is probably one the rare experiments which is fully driven by a Bayesian optimization and updated in real-time! However, one of the drawbacks is that we have to be cautious about the result and its interpretation which is slightly different.

In the end, the major difficulty is to determine which methods is the best suited for the single ion beam diameter measurement.

The next step is to use this measurement to get the focus size on the single ion beam in order to estimate the achievable resolution for the implantation. Since the production of Nitrogen ions is nearly operational, we should be able to implant ions soon. The only very limiting step is the baking of the sample in situ.

6.2 Personal contribution and feedback on the internship

The first part of the internship was mainly to observe how the experiment works and learn how to use the specific "tools" of the lab, in particular MCP. The work done during the internship was an essential parallel project done in total autonomy with just some guidance at some points. It wasn't sure if the Bayes method would work or not. What corresponds to 3.3 was mainly my idea, even if not used in the end. All the theoretical explanations about the Bayesian interpretation and theorem come only from what I read in the literature since nearly no one knew a lot about these theoretical problematics in the group.

Most of the work was programing, first in Mathematica to quickly see how to implement the different methods and to have the framework, and then in C++ using MCP. This was the occasion to perfect my programing skills in C++ and Mathematica. All the code one can find in the appendix was written from scratch during the internship and are now able to run perfectly. The Bayesian algorithm is now daily used and creates a huge gain of time, reliability and simplicity for measuring the focus size.

However, due to the complexity of the experiment, the few data acquisition were done with the help of Georg Jakob, a PhD student.

6.3 Aknowledgements

I would like to thank everyone who helped and support me during this internship.

I want to thank Professor Schmidt-Kaler and Kilian Singer for allowing me to work in the group on this project and for giving me a such interesting project despite the short amount of time available. I want to thanks particularly Georg Jakob for his patience, for his explanations all along the internship and for answering my silly questions. I want also to thank Stephan Ulm for his help and the corrections to report with the urgency of the deadline. I want also to thank every other members of the implantation team: Johannes Roßnagel, Karin Root-Berning, Charlotte Degünther and also Sam Dawkins. I thank also every other member of the group.

[13] Except maybe in particle Physics

References

[Allan, 1966] Allan, D. W. (1966). Statistics of atomic frequency standards. *Proceedings of the IEEE*, 54(2):221–230.

[Brune et al., 2008] Brune, M., Bernu, J., Guerlin, C., Deléglise, S., Sayrin, C., Gleyzes, S., Kuhr, S., Dotsenko, I., Raimond, J.-M., and Haroche, S. (2008). Process tomography of field damping and measurement of fock state lifetimes by quantum nondemolition photon counting in a cavity. *Physical review letters*, 101(24):240402.

[Chaloner and Verdinelli, 1995] Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, pages 273–304.

[DiVincenzo et al., 2000] DiVincenzo, D. P. et al. (2000). The physical implementation of quantum computation. *arXiv preprint quant-ph/0002077*.

[Dutt et al., 2007] Dutt, M. G., Childress, L., Jiang, L., Togan, E., Maze, J., Jelezko, F., Zibrov, A., Hemmer, P., and Lukin, M. (2007). Quantum register based on individual electronic and nuclear spin qubits in diamond. *Science*, 316(5829):1312–1316.

[Efron, 2013] Efron, B. (2013). Bayes' theorem in the 21st century. *Science*, 340(6137):1177–1178.

[Ferrie and Blume-Kohout, 2012] Ferrie, C. and Blume-Kohout, R. (2012). Estimating the bias of a noisy coin. *arXiv preprint arXiv:1201.1493*.

[Fickler, 2009] Fickler, R. (2009). *Kalte einzelne Ionen fur die Implantation in Festkorper mit nm Auflösung*. PhD thesis, Universität Ulm.

[Hausmann et al., 2011] Hausmann, B. J., Babinec, T. M., Choy, J. T., Hodges, J. S., Hong, S., Bulu, I., Yacoby, A., Lukin, M. D., and Lončar, M. (2011). Single-color centers implanted in diamond nanostructures. *New Journal of Physics*, 13(4):045004.

[Hell, 2007] Hell, S. W. (2007). Far-field optical nanoscopy. *science*, 316(5828):1153–1158.

[Herman et al., 1979] Herman, G. T., Hurwitz, H., Lent, A., and Lung, H.-P. (1979). On the bayesian approach to image reconstruction. *Information and Control*, 42(1):60–71.

[Kane, 1998] Kane, B. E. (1998). A silicon-based nuclear spin quantum computer. *Nature*, 393(6681):133–137.

[Leibfried et al., 2003] Leibfried, D., Blatt, R., Monroe, C., and Wineland, D. (2003). Quantum dynamics of single trapped ions. *Reviews of Modern Physics*, 75(1):281.

[Lyons, 2013] Lyons, L. (2013). Bayes and frequentism: a particle physicist's perspective. *Contemporary Physics*, (ahead-of-print):1–16.

[Meijer et al., 2008] Meijer, J., Pezzagna, S., Vogel, T., Burchard, B., Bukow, H., Rangelow, I., Sarov, Y., Wiggers, H., Plümel, I., Jelezko, F., et al. (2008). Towards the implanting of ions and positioning of nanoparticles with nm spatial resolution. *Applied Physics A*, 91(4):567–571.

[Neumann et al., 2008] Neumann, P., Mizuochi, N., Rempp, F., Hemmer, P., Watanabe, H., Yamasaki, S., Jacques, V., Gaebel, T., Jelezko, F., and Wrachtrup, J. (2008). Multipartite entanglement among single spins in diamond. *Science*, 320(5881):1326–1329.

[Pezzagna et al., 2010] Pezzagna, S., Naydenov, B., Jelezko, F., Wrachtrup, J., and Meijer, J. (2010). Creation efficiency of nitrogen-vacancy centres in diamond. *New Journal of Physics*, 12(6):065017.

[Porter, 2013] Porter, F. (2013). Interval estimation. http://www.hep.caltech.edu/ fcp/statistics/intervals/intervals/intervals.pdf.

[Rose, 2008] Rose, H. H. (2008). *Geometrical charged-particle optics*, volume 142. Springer.

[Schnitzler, 2010] Schnitzler, W. (2010). *Deterministic ultracold ion source targeting the Heisenberg limit*. PhD thesis, Universität Ulm.

[Schnitzler et al., 2010] Schnitzler, W., Jacob, G., Fickler, R., Schmidt-Kaler, F., and Singer, K. (2010). Focusing a deterministic single-ion beam. *New Journal of Physics*, 12(6):065023.

[Schnitzler et al., 2009] Schnitzler, W., Linke, N., Fickler, R., Meijer, J., Schmidt-Kaler, F., and Singer, K. (2009). Deterministic ultracold ion source targeting the heisenberg limit. *Physical review letters*, 102(7):070501.

[Walther et al., 2012] Walther, A., Ziesel, F., Ruster, T., Dawkins, S. T., Ott, K., Hettrich, M., Singer, K., Schmidt-Kaler, F., and Poschinger, U. (2012). Controlling fast transport of cold trapped ions. *Physical review letters*, 109(8):080501.

[Willig et al., 2007] Willig, K. I., Harke, B., Medda, R., and Hell, S. W. (2007). Sted microscopy with continuous wave beams. *Nature methods*, 4(11):915–918.

[Wolf, 2012] Wolf, S. (2012). *Verfahren zur deterministischen, hochauflösenden Implantation von Farbzentren*. PhD thesis, Johannes Gutenberg-Universität Mainz.

[Yong-Sheng, 2008] Yong-Sheng, Z. (2008). Bayesian credible interval construction for poisson statistics. *Chinese Physics C*, 32(5):363.

A User interface

The Bayesian algorithm display many elements for the user, here are displayed some of them

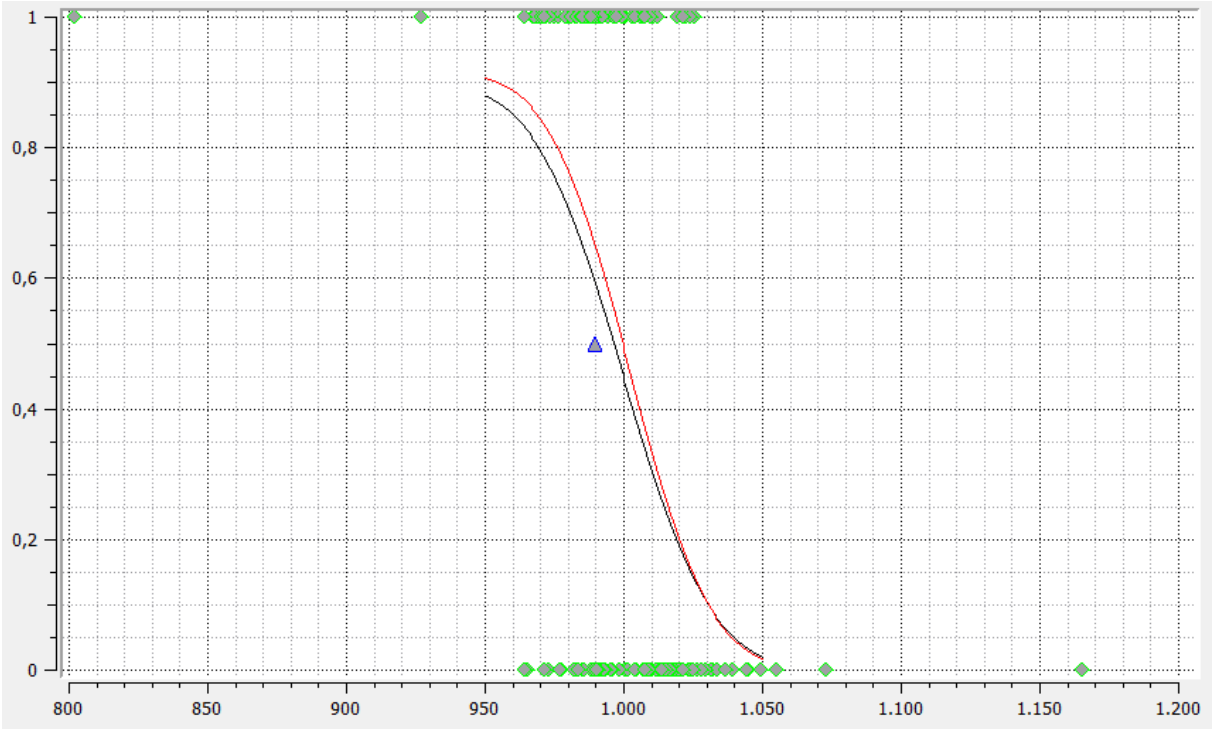
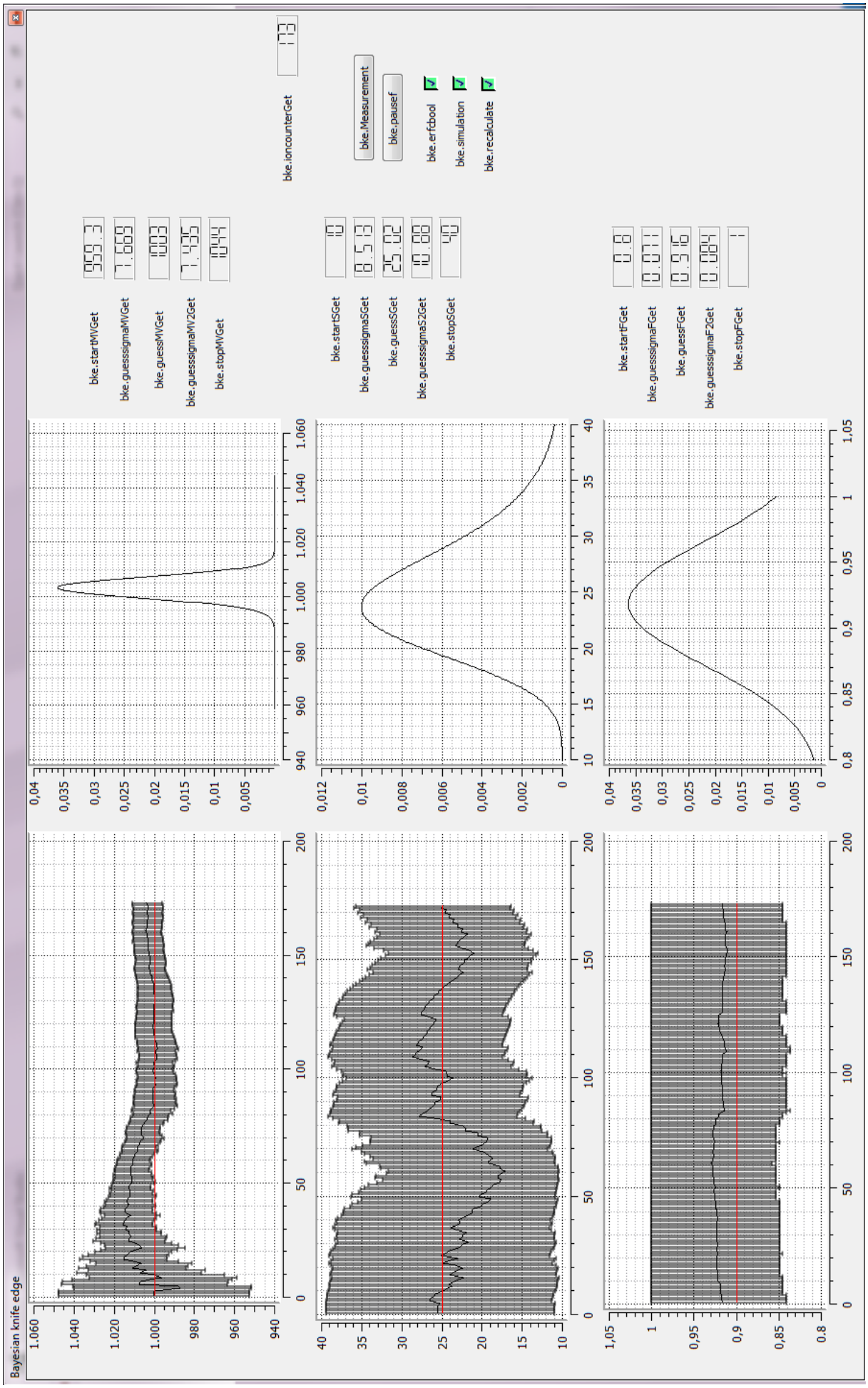


Figure 14: Graph with:
in black, the true beam probability distribution of this simulation,
in red, the current guess for the the beam probability distribution,
in green, the experimental data points. Those which are outside correspond to the start of the experiment, when the algorithm doesn't know at all the position of the beam,
in blue, the current blade position.


Bayesian knife edge
bke.ioncounterGet
bke.Measurement
bke.pausef
bke.erfcbool
bke.simulation
bke.recalculate
bke.startMVGet
bke.guesssigmaMVGet
bke.guessMVGet
bke.guesssigmaMV2Get
bke.stopMVGet
bke.startSGet
bke.guesssigmaSGet
bke.guessSGet
bke.guesssigmaS2Get
bke.stopSGet
bke.startFGet
bke.guesssigmaFGet
bke.guessFGet
bke.guesssigmaF2Get
bke.stopFGet

From left to right:

- The first graph corresponds to the current estimation of the center of the distribution with the credible interval. The second is the the estimation of the sigma and the third the estimation of the maximum probability. The red lines are the real value, in the case of a simulation.
- The first graph is the projection of the three dimensional array along the axis of the center position, the second the projection along the sigma axis and the third along the maximum probability axis
- For the three parameters are presented in this order: the current value for the lower boundary of the parameter space, the negative uncertainty, the current guess for the parameter, the higher part of the credible interval and the current higher boundary for the parameter space
- Some controls among others.

B C++ code of the Frequentist algorithm

The code is too long to be put here, please find the attached file.

C C++ code of the Bayesian algorithm

The code is too long to be put here, please find the attached file.