JÜLICH
FORSCHUNGSZENTRUM

RWTH AACHEN UNIVERSITY

Analyse der Linienform von Röntgenübergängen nach der Bayesmethode

von
Marian Theisen

Diplomarbeit in Physik
vorgelegt der

Fakultät für Mathematik, Informatik und Naturwissenschaften
der RWTH Aachen

begutachtet von
Univ.-Prof. Dr. rer. nat. Achim Stahl

angefertigt am
Institut für Kernphysik
des
Forschungszentrums Jülich
bei
Prof. Dr. rer. nat. Detlev Gotta

Declaration

I hereby declare that I have written this diploma thesis without any help from others and without the use of documents and aids other than those stated below. I have mentioned all used sources and cited them correctly according to established academic citation rules.

July 17, 2013 Marian Theisen

Acknowledgements

My first thank you goes to my parents, Wilfried Theisen and Dipl.-Ing. (FH) Rosemarie Placzek-Theisen, who taught me reading before i went to school, who always had yet another book to read, who gave everything to enable me and my brother to go to university.

- I want to thank Prof. H. Ströher for giving me the chance to write my diploma thesis at the IKP, Forschungszentrum Jülich.
- Furthermore, i want to thank you Univ.-Prof. Dr. rer. nat. Achim Stahl for taking the time to review this thesis.
- My deepest gratitude goes to Prof. Dr. rer. nat. Detlev Gotta, not only for making this work possible, but for new ways of thinking, new insights into experimental methods and data analysis, for pointing me in the right direction while providing enough freedom to make my own experiences. Thank you for all your input, support, and help in improving this thesis as well as reviewing it (over and over again); for being always available and your endless patience.
- I also want to thank Dr. Leopold Simons, who pioneered the application of the described analysis method to such types of X-ray spectra, whom i had the pleasure to meet for a few days and enjoy passionate discussions about our work. Thank you for helping me solve countless problems by just pointing out a few crucial aspects, which always taught me a lot. And finally thank you for reviewing and supporting this work.
- Dr. Martino Trassinelli for interesting discussions at the team meetup.
- Dr. Vladimir Popov and Dr. Vladimir Pomerantsev for their theoretical work on the ESCM calculations.
- The pionic hydrogen collaboration in total for providing such an interesting topic, and a chance to work on it.
- I would like to thank my grandfather, Dr. rer. nat. Longin Placzek, who passed away in the first year of my studies. I would have liked him to read this thesis.
- A great thank you goes to Dipl.-Phys. Ilhan Engin, who supported me, when i was not sure of finishing my studies at all. With whom i spent an awesome time during our studies, with whom i managed to have a lot of fun, even in solving relativistic quantum-mechanics problems. Thank you for being such a good friend, and for managing the most important thing at the IKP - the coffee list.
- Dr. Hubert Gorke for providing me a computer and support in Linux maintenance.
- Julian Theisen, my brother, who became one of my best friends over the last years.
- Dipl.-Phys. Patrick Greven, my colleague, office mate and now good friend for entertaining times at the IKP and beyond.
- Dr. D.S. Covita, Dr. Thomas Strauch and Dr. Albert Hirtl, for their work in exotic atoms, and effectively providing the data to analyze.

– Dipl.-Phys. Hakan Kayan, my close friend and boss. Thank you for supporting me and this thesis, for giving me as much time as i needed and always trying to keep problems away from me.
– I also want to thank everybody at the IKP for entertaining coffee breaks and discussions, for opportunities to just sit and listen and gain new insights or participate and contribute.
– And last but not least, all my friends for just being there.

Abstract

In this thesis, an alternative method, based on Bayesian statistics, for the analysis of X-ray transition lines has been explored. The focus to test and verify this method laid on existing X-ray spectra of various exotic atoms (pionic hydrogen, pionic deuterium, muonic hydrogen). These atoms show a line shape broadened by numerous effects, including Doppler broadening (due to a complex kinetic energy distribution), spectrometer response function, hyperfine splitting (for muonic hydrogen μH) and hadronic interaction with the nucleus (for the pionic atoms πH and πD).

A software package, named BayXRay, was created to implement Bayesian parameter inference and model-selection methods, especially the nested sampling algorithm and supplemental algorithms: Markov-Chain-Monte-Carlo sampling, ellipsoidal sampling, and clustered ellipsoidal sampling. Using this software, the methodology was verified using the measured spectra. Results were compared to the ones of the previously used analysis method. In general, a good agreement was found with these previous results. In addition, the new method, contrary to the previous χ^2-fitting analyses, allowed a consistent and direct determination of the error margins for the inferred parameters.

Furthermore, the methods of model-selection and model-independent parameter inference opened up new possibilities. Based on these, new techniques like model-averaged spectra (deconvolution) or evaluations with theoretical kinetic energy distributions have been realized and also compared to the previous results.

Zusammenfassung

Das Ziel dieser Diplomarbeit ist es, eine alternative, auf der Bayes-Methode basierende, Methode zur Analyse von Röntgenspektren zu evaluieren. Zur Verifikation dieser Methode wurden vorliegende Röntgenspektren diverser exotischer Atome (pionischer Wasserstoff, pionisches Deuterium und muonischer Wasserstoff) untersucht. Die Linienform der Übergänge dieser Atome zeigt Charakteristika von Dopplerverbreiterung (aufgrund einer kinetischen Energieverteilung), einer Verbreiterung durch die Auflösungsfunktion des verwendeten Spektrometers, Hyperfeinaufspaltung (im Falle des muonischen Wasserstoffs μH) und hadronischer Wechselwirkung mit dem Atomkern (für die pionischen Atome πH und πD).

Um die Bayes Methoden (wiederverwendbar) zu implementieren, wurde ein Software Paket (BayXRay) entwickelt, das numerische Methoden zur Parameter-Inferenz und Modell-Selektion (Hypothesen-Tests) bereitstellt (vor allem den Nested-Sampling Algorithmus, sowie Hilfsalgorithmen wie Markov-Chain-Monte-Carlo-Sampling, Ellipsoidal-Sampling and Clustered-Ellipsoidal-Sampling). Die Analysen, die mit Hilfe dieser Software durchgeführt wurden, konnten die Methodik verifizieren: der Vergleich mit vorherigen Ergebnissen zeigt gute Übereinstimmungen. Zusätzlich erlaubt die neue Methode im Gegensatz zu den früheren χ^2-Anpassungen eine klar definierte Fehlerbehandlung.

Zusätzlich zur Reanalyse wurden neue, bisher nicht mögliche Methoden angewandt. Darunter zählt z.B. “das Model-Averaging” von ganzen kinetischen Energieverteilungen (Entfaltung), oder der Vergleich mit theoretisch berechneten kinetischen Energieverteilungen.

Contents

1. Introduction

1.1. Motivation

Since the 19^{th} century, spectroscopy of the radiation emitted or absorbed from atoms and molecules is an important tool to study their inner structure. A systematics — though of unknown origin in the beginning — became evident from relations between chemical elements and compounds and the combinatorial appearance of spectral line series (*e. g.* from the studies of Kirchhoff, Fraunhofer or Ritz). It became clear that the order and distances of the spectral lines are unique per chemical element, attributed to the effect of an inner structure. An empirical description of the spectral lines was given by the Rydberg formula.

The first successful approach to describe the inner structure of atoms was the Bohr model (1913) by applying the quantization of action. It provided a convincing description of the inner structure of an atom and could reproduce, in simple systems, the experimental results along with their existing description, the Rydberg formula.

After the extension of the Bohr model by Sommerfeld by including angular momentum, the spectroscopy of emission spectra originating from inner shells (characteristic X-radiation) enabled physicists to gain more and more detailed insights. High-precision spectroscopy revealed electron and nuclear spin by fine and hyperfine structure of spectral lines and the determination of the line width yields level life time via the uncertainty relation.

With the discovery of heavier negatively charged particles like mesons or antiprotons, for about six decades so called exotic atoms can be studied, where the heavy particles replace the electron. At modern accelerator facilities, exotic atoms can be formed at high rates. High-resolution spectroscopy of their characteristic X-ray emissions gives access to parameters like the mass of the captured particles and its interaction with the nucleus. Due to the complex nature of these line shapes (see chapter 2), sophisticated and efficient data analysis methods are required to extract particular quantities reliably.

The aim of this work is to apply an alternative scheme in data analysis of spectral lines in general, and specifically of X-ray lines of exotic atoms (see section 1.2). The basis for these new techniques are Bayesian statistics, an alternative approach to parameter inference (*"fitting"*) and hypothesis testing (see section 3.2), which plays an increasingly important role in scientific data analysis and is already established in the astrophysical community.

In order to verify the methods and routines developed, pre-existing data of exotic atoms have been re-analyzed and compared to previous results (pionic hydrogen πH [1, 2], pionic deuterium πD [3, 4, 5], muonic hydrogen μH [6, 7, 8]). In particular, a well defined procedure to compare alternative descriptions and the corresponding error handling (see section 3.2.3) was the driving force behind the exploration of this new techniques.

1.2. Exotic atoms

Atoms are the result of the electromagnetic force between the (positively charged) nucleus and a collection of (negatively charged) electrons forming a bound system having a size determined by the uncertainty relation. Inspecting Bohr's formulae for binding energy and radius (with the fine structure constant $\alpha \approx 1/137$)

$$r_{\mathrm{B}} = \frac{\hbar}{\mu_{ep} c \alpha} \cdot \frac{n^2}{Z} ,$$
$$E_{\mathrm{B}} = -\frac{\mu_{ep} c^2 \alpha^2}{2} \cdot \frac{Z^2}{n^2}$$

suggests immediately that systems formed with heavier negatively charged particles of mass m_x should exist with larger binding energies and smaller radii. Energies and radii scale in leading order with the mass ratio of the heavy particle and the electron m_x/m_e and its inverse, respectively.

These systems are called exotic atoms and have been studied in detail for muons, pions, kaons, Σ-baryons, and antiprotons provided at various accelerator facilities [9]. Pions, kaons, Σ-baryons, and antiprotons are strongly interacting particles as are the nucleons in the nucleus (hadrons). Because of the small size the innermost atomic levels are affected in addition by the nuclear force. These so-called strong-interaction effects are measurable by means of precision X-ray spectroscopy.

1.2.1. Capture and deexcitation cascade in exotic hydrogen

An exotic atom is formed when the particles to be captured in the Coulomb field is slowed down to kinetic energies of about the binding energy of the outermost electrons. After capture the particle is in a highly excited state [10]. For hydrogen the most probable principal quantum number n is about

$$n \approx \sqrt{\frac{\mu_{x^-\mathrm{p}}}{m_{\mathrm{e}}}} ,$$

with the reduced mass of the exotic system $\mu_{x^-\mathrm{p}}$ and the electron mass m_{e}. This yields $n_\mu = 14$ for muons and $n_\pi = 16$ for pions.

From these states, the atom deexcites to the ground state via a series of processes. For such electrically neutral systems, the exotic hydrogen deexcitation is strongly influenced by collisional effects and, hence, strongly density dependent. Collisional processes are Stark mixing, the external Auger effect, Coulomb deexcitation, elastic and inelastic scattering. These are supplemented by internal effects, namely radiative deexcitation, nuclear reactions, and if applicable particle decay. A cascade scheme is shown in Figure 1.1 and a detailed description of the cascade processes can be found in [11, 12, 13].

Stark mixing

As the exotic atom is considerably smaller than a standard hydrogen atom and, electrically neutral, it approaches nuclei to distances of about its own size. Stark mixing is the result of an electromagnetic interaction occuring when the exotic atom penetrates the Coulomb field of neighboring hydrogen atoms. The electric field causes transitions between the l substates, which do not conserve the quantum number l [14]. In dense targets, the result is an almost statistical re-distribution [13, 14] of the l substates at all principal quantum numbers n as long as the widths of the l levels is of the order of the level splitting.

In hadronic atoms, this leads to a depletion of the cascade because of nuclear reactions whenever a high-lying s state is reached. Stark mixing can be described with the following equations

$$\begin{aligned} (\mathrm{x^-p})_{n\,l} + \mathrm{H_2} &\longrightarrow (\mathrm{x^-p})_{n\,l'} + \mathrm{H_2^*}\,, \\ (\mathrm{x^-p})_{n\,l} + \mathrm{H} &\longrightarrow (\mathrm{x^-p})_{n\,l'} + \mathrm{H}\ . \end{aligned}$$

External Auger effect

External Auger effect describes the ionization of another atom in which the required energy is gained by deexcitation of the pionic / muonic system. As the highest probability for Auger emission is found for $\Delta n = 1$, $\Delta l = 1$ transitions [13, 14], it is dominated by states where the $\Delta n = 1$ transitions provide enough energy to ionize the H_2 molecule.

The formal specification is

$$(\mathrm{x^-p})_{n\,l} + \mathrm{H_2} \longrightarrow (\mathrm{x^-p})_{n'\,l'} + \mathrm{H_2} + \mathrm{e^-}\ .$$

Coulomb deexcitation

Coulomb deexcitation is an inelastic electromagnetic scattering process in which the principal quantum number n of the exotic system decreases, and the released energy is transformed into kinetic energy of the exotic atom and its collision partner. Upon collision with a H_2 molecule, possible processes are [15, 16]

$$\begin{aligned} (\mathrm{x^-p})_{n\,l} + \mathrm{H} &\longrightarrow (\mathrm{x^-p})_{n'\,l'} + \mathrm{H}\,, \\ (\mathrm{x^-p})_{n\,l} + \mathrm{H_2} &\longrightarrow \begin{cases} (\mathrm{x^-p})_{n'\,l'} + \mathrm{H} + \mathrm{H} \\ (\mathrm{x^-p})_{n'\,l'} + \mathrm{H_2^*} \end{cases} . \end{aligned}$$

The kinetic energy gained in a two-body collision ($(\mathrm{x^-p}) + \mathrm{H}$) is given by

$$T_{(\mathrm{x^-p})} = \frac{m_{\mathrm{H}}}{m_{(\mathrm{x^-p})} + m_{\mathrm{H}}} \cdot E_{n \longrightarrow n'}\ .$$

Including molecular effects, the energy is distributed among three collision partners: the maximal energy of the exotic system is given by $T_{(\mathrm{x^-p})}$, with the hydogen mass m_{H} is replaced by the molecule mass $2 \cdot m_{\mathrm{H}}$.

It is evident from Table 1.1, that the most significant impact on the kinetic energy distribution stems from Coulomb deexcitation.

transition	kin. energy E [eV]		
$n \longrightarrow n'$	πH	πD	μH
$7 \longrightarrow 6$	12	12	8.8
$6 \longrightarrow 5$	18	20	15
$6 \longrightarrow 4$	52	58	42
$5 \longrightarrow 4$	34	38	27
$5 \longrightarrow 3$	107	119	85
$4 \longrightarrow 3$	73	81	58
$4 \longrightarrow 2$	282	324	225
$3 \longrightarrow 2$	209	240	166

Table 1.1.: Kinetic energy gained in the reaction $(\mathrm{x^-p})_{nl}+H \longrightarrow (\mathrm{x^-p})_{n'l'}+H$ (Coulomb deexcitation) for πH, πD, and μH.

Radiative deexcitation

Radiative deexcitation is the process of a state change to a lower principal quantum number $n \longrightarrow n'$, $n > n'$ by emission of a photon γ. Due to the spin of the photon, it has to obey a change in angular momentum by $\Delta l = \pm 1$.

The reaction can be written as

$$(\mathrm{x^-p})_{n\,l} \longrightarrow (\mathrm{x^-p})_{n'\,l'} + \gamma \ .$$

This effect is of particular importance because the emitted photons make up the measured spectral lines. As the radiative deexcitation is one of the final stages in the atomic cascade, the previous processes affect the energetic distribution of the emitted photons (see chapter 2).

The deexcitation cascade

Which of the above mentioned processes dominates during specific stages of the cascade largely depends on the current energy level. The entire cascade can roughly be subdivided into three parts:

Transitions from $n > 8$ are mainly caused by Coulomb deexcitation and Stark mixing.

Transitions from $n = 7, 8$ are dominated by external Auger effects.

Subsequent deexcitation steps below $n \leq 6$ happen increasingly due to radiative transitions.

As recoils in all processes are of an order of ≈ 1 eV, Coulomb deexcitation dominates the development of the kinetic energy of the exotic system (see Table 1.1).

Theoretical description of the cascade

The theoretical description of the atomic cascade began in the 1960s with the Standard Cascade Model (SCM) [14]. Further developments in the last two decades augmented this model to the so-called Extended Standard Cascade Model by including the level dependent evolution of the kinetic energy [12, 14].

Another improvement resulted from more precise calculations of the cross sections of the various processes. This allows new predicitions of X-ray line yields and more importantly kinetic energy distributions to compare with experimental results [17].

The high-resolution spectroscopy of radiative transitions to the ground state allows the identification of Doppler contributions to the X-ray line width due to acceleration processes like Coulomb deexcitation. Vice versa, a calculated kinetic energy distribution, if verified, along with its induced Doppler broadening, makes a more precise extraction of the hadronic width possible.

Elastic and inelastic scattering, however, slow down the fast exotic atoms. Consequently, the kinetic energy spectrum at the emission of the X-ray is a result of the competition of the accelerating and decelerating processes.

1.2.2. Muonic hydrogen

Muonic hydrogen exhibits another internal effect, namely a hyperfine structure, which is the result of higher order electromagnetic effects. The major proportion is due to coupling between the spins of the proton $\boldsymbol{S}_\mathrm{p}$ and the muon $\boldsymbol{S}_{\boldsymbol{\mu}}$.

This coupling causes shifts in the binding energy by [6]

$$\Delta H_\mathrm{hfs} = A\, \boldsymbol{S}_\mathrm{p} \cdot \boldsymbol{S}_{\boldsymbol{\mu}} \ .$$

In the 1s ground state (with angular momentum $l = 0$), the spins form a singlet state and triplet states of total spin $F_S = 0$ and $F_T = 1$, respectively. These states have an energy shift of

$$\Delta H^F_\mathrm{hfs} = \begin{cases} +\frac{1}{4} A\hbar^2 & F = 1 \\ -\frac{3}{4} A\hbar^2 & F = 0 \end{cases} , \tag{1.1}$$

with A being the constant

$$A = \frac{8}{3} c^2 \alpha^4 \mu_\mathrm{p} \cdot \frac{1}{\hbar^2} \cdot \frac{m_\mu^2 + m_\mathrm{p}^2}{(m_\mu + m_\mathrm{p})^2}$$

which includes the dependence on the masses m_p, m_μ as well as the hyperfine structure constant α and the magnetic moment of the proton μ_p.

The energy shifts eq. 1.1 result in a splitting of the 1s ground state

$$\begin{aligned}\Delta H_{\mathrm{hfs}} &= \Delta H_{\mathrm{hfs}}^{F=1} - \Delta H_{\mathrm{hfs}}^{F=0} \\ &= A\hbar^2 = \frac{8}{3} c^2 \alpha^4 \mu_{\mathrm{p}} \cdot \frac{m_{\mathrm{e}}^2 + m_{\mathrm{p}}^2}{(m_{\mathrm{e}} + m_{\mathrm{p}})^2} \ .\end{aligned}$$

Numerical substitution yields a value of $\Delta H_{\mathrm{hfs}} = 182.443\,\mathrm{meV}$. By considering, in addition, QED corrections a more precise value of $(182.725 \pm 0.062)\,\mathrm{meV}$ is found [18].

1.2.3. Pionic hydrogen and deuterium

Pionic atoms show an additional effect, which is of particular importance. Despite the short range of the nuclear force, a pion bound in the ground state (1s) of *e. g.* a hydrogen atom experiences the strong force of the nucleus. This is caused by the significantly larger mass of the pion (in contrast to an electron)

$$m_\pi \approx 273 \cdot m_e \,,$$

which results in a much smaller orbit of the pion around the nucleus. The consequence of this interaction is a Lorentzian broadening $\Gamma_{1\mathrm{s}}$ and an energetic shift $\varepsilon_{1\mathrm{s}}$ of the 1s ground state.

$\Gamma_{1\mathrm{s}}$ and $\varepsilon_{1\mathrm{s}}$ are directly connected to the QCD scattering lengths $a_{\pi^- p \longrightarrow \pi^- p}$ and $a_{\pi^- p \longrightarrow \pi^0 n}$ via [19, 20]

$$\begin{aligned}\frac{\varepsilon_{1\mathrm{s}}}{B_{1\mathrm{s}}} &= -\frac{4}{r_{\mathrm{B}}}\, a_{\pi^-\mathrm{p} \longrightarrow \pi^-\mathrm{p}}\,(1 + \delta_\varepsilon) \\ \frac{\Gamma_{1\mathrm{s}}}{B_{1\mathrm{s}}} &= \frac{8 q_0}{r_{\mathrm{B}}} \cdot (1 + \frac{1}{P}) \cdot \left[a_{\pi^-\mathrm{p} \longrightarrow \pi^0\mathrm{n}}\,(1 + \delta_{\Gamma_{1\mathrm{s}}})\right]^2\end{aligned}$$

with the Coulomb binding energy of the ground state $B_{1\mathrm{s}} = 3238\,\mathrm{eV}$, the Bohr radius of the $\pi\mathrm{p}$ $r_{\mathrm{B}} = 222.56\,\mathrm{fm}$, the π_0 cms-momentum q_0, and the Panofsky ratio $P = 1.546 \pm 0.009$ between the channels $\pi^-\mathrm{p} \longrightarrow \pi^0\mathrm{n}$ and $\pi^-\mathrm{p} \longrightarrow \gamma\mathrm{n}$ [21]. The corrections δ_ε and δ_Γ can be calculated unambiguously within the framework of chiral perturbation theory (ChPT), which is a modern approach of low-energy QCD [4, 19].

In the case of complex nuclei, like deuterium, the relation between $\Gamma_{1\mathrm{s}}$ and $\varepsilon_{1\mathrm{s}}$ and the complex scattering length $a_{\pi\mathrm{D}}$ is given by the Trueman formula [22]

$$\varepsilon_{1\mathrm{s}} - \mathrm{i}\,\frac{\Gamma_{1\mathrm{s}}}{2} = -\frac{2\alpha^3 \mu^2 c^4}{\hbar c}\, a_{\pi\mathrm{D}} \cdot \left[1 - \frac{2\alpha\mu c^2}{\hbar c}\,(\ln\alpha - 1) \cdot a_{\pi\mathrm{D}} + \delta_{\mathrm{D}}^{\mathrm{vac}}\right]$$

Here, the shift ε_{1s} is the coherent sum of $\pi^-\mathrm{p}$ and $\pi^-\mathrm{n}$ scattering and $\Gamma_{1\mathrm{s}}$ is proportional to the pion absorption strength $\pi^-\mathrm{pn} \longrightarrow \mathrm{nn}$ [4].

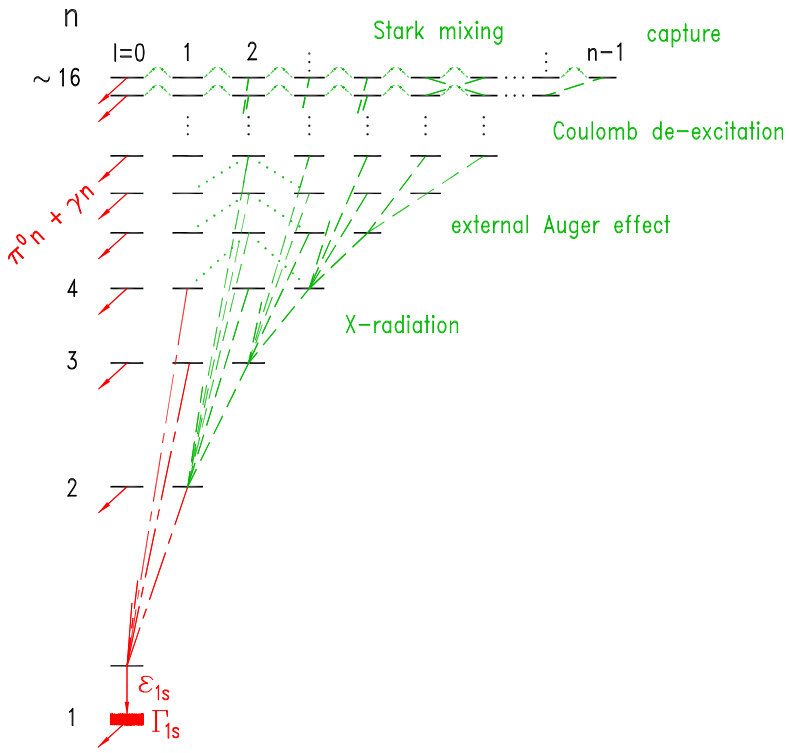


Figure 1.1.: Scheme of the possible state transitions in a pionic hydrogen system contributing to the atomic cascade. The pion is captured in an highly excited state and deexcites via various effects. The Coulomb deexcitation dominates the states of $n > 8$. Below that ($n = 7, 8$) external Auger transitions play the most important role. In the lower states $2 \leq n \leq 6$, radiative deexcitation characterizes the transitions along with Stark mixing between the l substates. In s states, the nuclear reaction $\pi^- \mathrm{p} \longrightarrow \pi^0 \mathrm{n}\,(61\%) + \gamma \mathrm{n}\,(39\%)$ occurs with high probability thus depleting the cascade.

1.3. Experimental setup

The spectra analyzed in this work were collected at the Paul-Scherrer-Institute (PSI), Switzerland at the πE5 beamline [23]. A scheme of the experimental setup is shown in Figure 1.2. The main components are the target cell surrounded by a cyclotron-trap in the pion beamline, the spectrometer (Bragg crystal), and the X-ray detector.

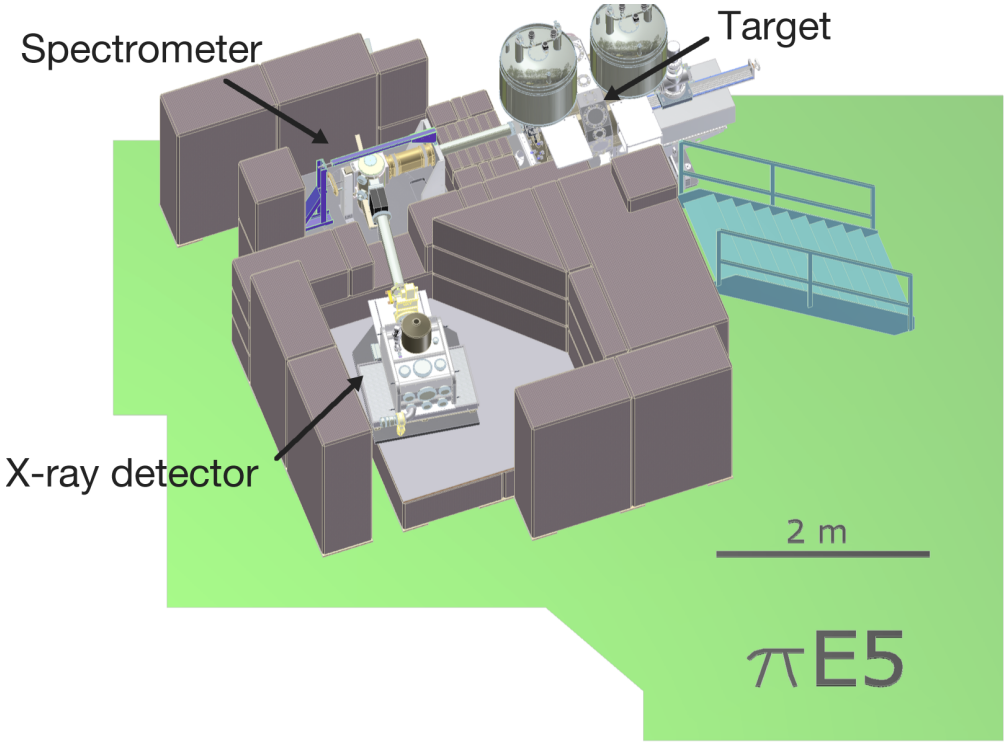


Figure 1.2.: 3D-drawing of the experimental setup at the πE5 area at the PSI.

The πE5 beamline provides a π^- beam with a momentum of about 112 MeV/c [24], which is injected into the cyclotron trap and slowed down using several degraders until they reach the target cell in the center. A muon flux can be provided by adjusting the degraders in a way, that muons from weak π^- decay are stopped in the target gas, where the pion / muon is captured by a hydrogen / deuterium atom. The last step of the following deexcitation cascade is the radiative transition by emission of an X-ray photon.

The X-ray photons emitted are then diffracted at the Bragg crystal and reflected with an angular distribution towards the two-dimensional CCD detector. To illustrate the imaging properties of the setup, the hit pattern in the detector plane is shown in Figure 1.4 as obtained by means of a Monte-Carlo simulation.

To produce a one-dimensional position spectrum, a curvature correction by means of a program called CSDCLUSTER was performed and the corrected image projected along the vertical axis. A detailed description of the setup, particularly of the cyclotron-trap, the target cell as well as the detector and its alignment can be found in [1, 3, 6].

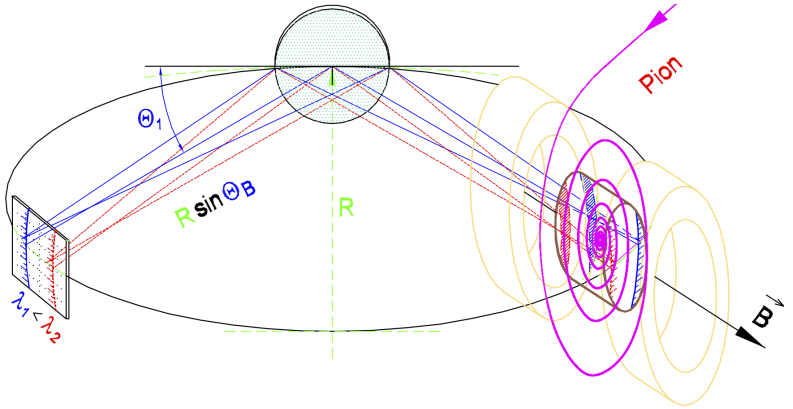


Figure 1.3.: Functional principle of the experimental setup showing the incoming pion beam, trapped in the cyclotron and slowed down by degraders until it hits the central target cell. The pions are then captured in hydrogen or deuterium and experience the deexcitation cascade (see section 1.2). As final step of the cascade, radiative deexcitation and emission of X-rays occur, which are diffracted at the Bragg crystal and subsequently collected in the X-ray detector.

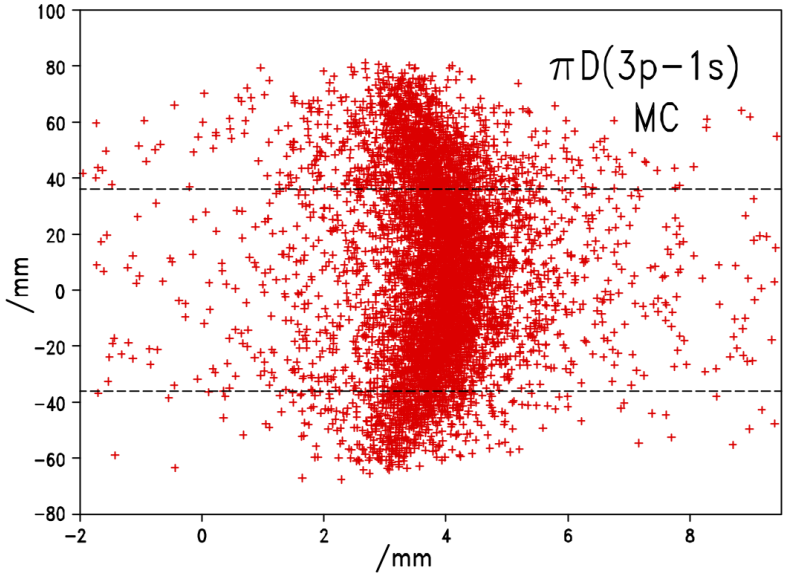


Figure 1.4.: Monte-Carlo simulation of the X-ray hit pattern in the detector plane. Horizontal dashed lines mark the vertical boundaries of the X-ray detector. In order to obtain a position space spectrum, curvature correction has to be applied, followed by a projection along the vertical axis. The horizontal axis is equivalent to an angle and hence also to an energy spectrum.

2. The structure of X-ray lines

2.1. Contributions to the line width

The width, or to be precise, the shape of the line shape is built up from several contributions. These contributions can be individually described by probability density functions P_i, convolution of which describes the total line shape S

$$S = P_1 \otimes P_2 \otimes \ldots \otimes P_3 \ .$$

For the spectra of muonic hydrogen, pionic hydrogen, and pionic deuterium, which are subject of this work, these contributions are:

Doppler broadening $\boldsymbol{D}$ caused by fast motion of the exotic atoms at the moment of X-ray emission.

Hadronic broadening of the ground state $\boldsymbol{L}$ given by a Lorentzian distribution in the case of pionic hydrogen and deuterium.

Hyperfine splitting $\boldsymbol{H}$ of the $\mu^- p^+$ ground state is not a true broadening but a superposition of two close-lying lines.

Spectrometer response $\boldsymbol{R}$ includes several experimental effects: the rocking curve of the Bragg crystal, geometric effects, and an additional Gaussian width.

Thus, the measured line shape reads for pionic atoms

$$S_\pi = L \otimes D \otimes R\,, \tag{2.1}$$

and analog for muonic hydrogen

$$S_\mu = H \otimes D \otimes R \ . \tag{2.2}$$

The individual contributions are described in more detail below.

2.2. Doppler broadening D

The Doppler broadening D of the line shape originates from a complex kinetic energy distribution. As the exact reconstruction of an kinetic energy spectrum from a measured line shape is not possible, the distribution was modeled by a number of narrow kinetic energy intervals. Various combinations of kinetic energy intervals with different positions, widths and relative intensities were tried to achieve the best agreement with the data.

2.2.1. Doppler line shape

In order to reconstruct the kinetic energy distribution from a given X-ray line shape or to verify a predicted ESCM distribution, the relation between velocity and X-ray energy is built up as described in the following.

Given an exotic system $x^- p^+$ with mass $m_{x^- p^+}$ and a kinetic energy of T_0, the magnitude $v_0 = |\underline{v}|$ of its velocity $\underline{v}$ is

$$v_0 = |\underline{v}| = \sqrt{\frac{2\,T_0}{m_{x^- p^+}}} \ . \tag{2.3}$$

Under the assumption of an isotropic velocity distribution, the velocity in the direction of observation v_x is distributed uniformly between $-v_0$ and v_0

$$v_x \sim \mathcal{U}(-v_0, v_0) \ . \tag{2.4}$$

This means a single δ-peak-like kinetic energy distribution

$$prob\big(T \,\big|\, T_0\big) = \delta(T - T_0) \tag{2.5}$$

transforms to a uniform distribution for the velocity in the direction of observation (Figure 2.1a)

$$prob\big(v \,\big|\, T_0\big) = \mathcal{U}\left(-\sqrt{\frac{2\,T_0}{m_{x^- p^+}}}, +\sqrt{\frac{2\,T_0}{m_{x^- p^+}}}\right) \ . \tag{2.6}$$

In order to approximate an arbitrary kinetic energy distribution, non-overlapping, rectangular functions (components) B_i are superposed (similar to a histogram) with relative weights β_i and $\sum_i \beta_i = 1$

$$prob\big(T\big) = \sum_i \beta_i \, B_i(T) \ . \tag{2.7}$$

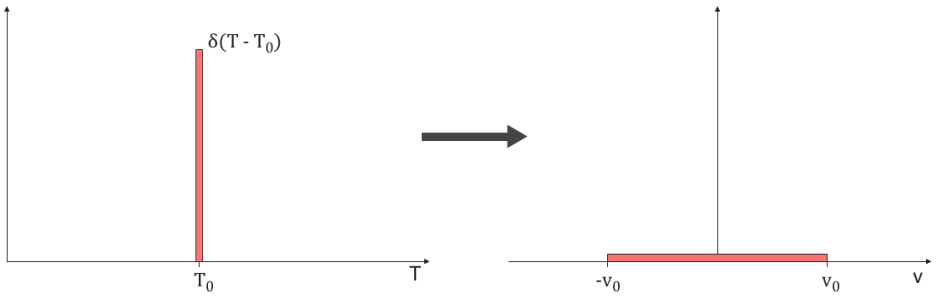
This superposition can be carried over to the velocity distribution

$$prob\big(v_x\big) = \sum_i \beta_i \, B_i(v_x) \tag{2.8}$$

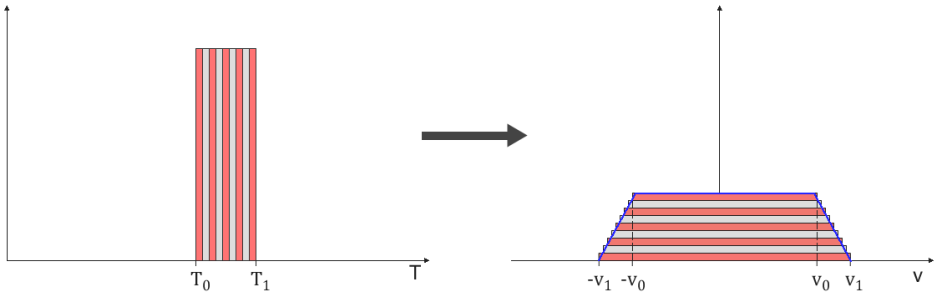
leaving the task of identifying a transformation $B_i(T) \to B_i(v_x)$.

If an rectangular component (from T_0 to T_1) is considered itself a superposition of a large number n of $1/n$ wide rectangular components (converging to δ-functions in the limit $n \to \infty$), the sum of the corresponding velocity distributions take the form of a step-pyramid with a step height of $1/n$. In the limit $n \to \infty$ the superposition approaches a trapezoidal function between $-v_x(T_2)$ and $v_x(T_1)$ (see Figure 2.1b).

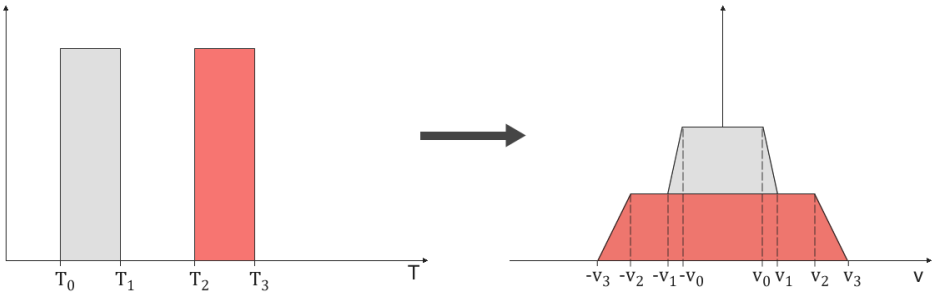
The total line shape of the Doppler broadening is then described by a sum of trapezoidal functions, with the underlying kinetic distribution approximated by rectangular components (Figure 2.1c).



(a) A δ-peak-like kinetic energy distribution transforms to a uniform probability in the velocity domain.



(b) A rectangular component can be transformed by slicing it into superposed δ-peaks, and transforming each slice separately.



(c) A superposition of rectangular components results in a superposition of trapezoids.

Figure 2.1.: Visualization of the transformation from the kinetic energy distribution to a velocity distribution.

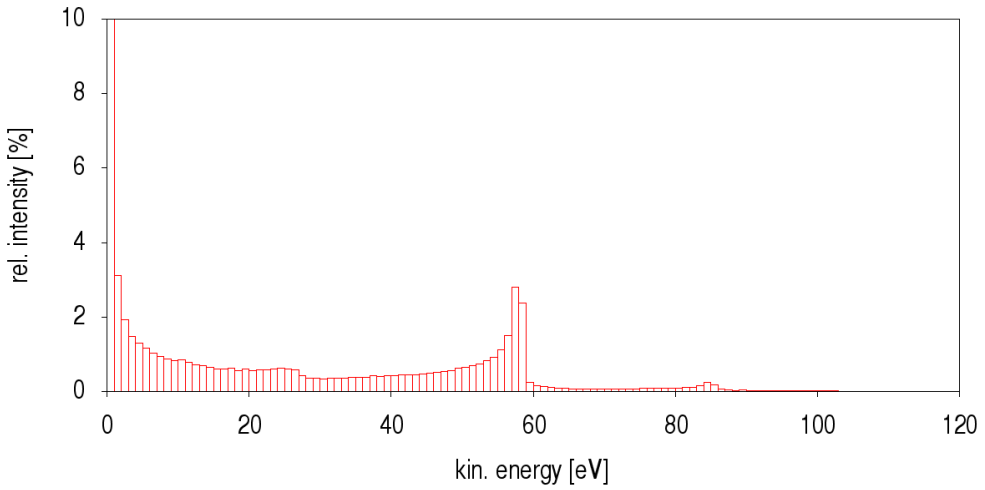


Figure 2.2.: Example of a theoretical kinetic energy distribution for the μH(3p-1s) X-ray lines [25].

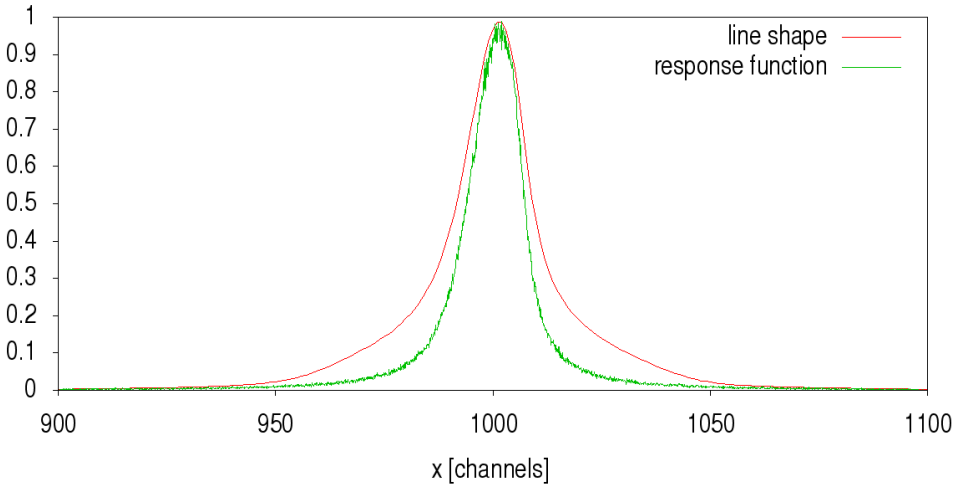


Figure 2.3.: Convolved line shape of the theoretical kinetic energy distribution in Figure 2.2. This line shape includes the response function of the experimental setup (rocking curve + geometrical broadenings + Gaussian width of 75 μrad) as well as the Doppler broadening caused by the kinetic energy distribution (see section 5.4.6).

2.3. Hadronic Lorentz width L

The X-ray line broadening due to the strong interaction with the nucleus is given by a Lorentzian

$$L = \frac{1}{2\pi} \cdot \frac{\Gamma_{1s}}{(E - E_0)^2 + \Gamma_{1s}^2/4} \,,$$

with mean energy E_0 and broadening Γ_{1s} of the 1s ground state. A detailed description and a discussion of the relation of the energy shift ε_{1s} and hadronic width Γ_{1s} to the scattering lengths of the pion-nucleon strong interaction can be found in [9].

The order of magnitude of this broadening is $\Gamma_{1s} \approx 800\,\text{meV}$ and 1200 meV for pionic hydrogen and pionic deuterium, respectively [2, 5].

2.4. Hyperfine structure in μH

Muonic hydrogen exhibits splittings of the energy levels due to spin couplings between the muon and the nucleus (see section 1.2.2). Hence, the ground state transition 2p-1s is effectively a doublet, because the 2p hyperfine structure is comparably small. The energy dependence of the spectrometer response is assumed to be negligible for the small splitting of $\approx$ 180 meV, but to determine the line shape without a bias, the energy difference and the intensity ratio of the two hyperfine components have to be taken into account.

2.5. Spectrometer response R

2.5.1. Bragg rocking curve R_C

The ability to determine X-ray energies relies on Bragg's law

$$n\,\lambda = 2\,d\,\sin\Theta_\text{B} \quad .$$

It relates wavelengths λ (and therefore energies $E = {}^{hc}/\lambda$) of incoming electromagnetic waves to certain angles Θ_B of constructive interferences. These interferences are the result of the superposition of waves diffracted at crystal lattice planes with distance d. In principle, several orders n of diffraction are possible.

Diffraction by principle causes a reflection of an incoming wave into an angular range. This ideal resolution of a plane Bragg crystal is characterized by the so called rocking curve R_C. The shape and width of this rocking curve depend on the energy / wavelength and the material of the crystal.

In order to determine the exact form, a software package called XOP [26] was used, which can simulate a rocking curve for a given material and wavelength / energy. An example of such a rocking curve along with the full response is shown in Figure 2.4.

2.5.2. Geometric constraints R_G

In addition to the rocking curve, broadenings from different parts of the experimental setup arise. These originate mostly from geometrical effects: the curvature and the extension of the bent crystal, apertures on the reflecting crystal surface and the position resolution of the X-ray detector. Given the rocking curve, the geometrical aberrations can be simulated by means of Monte-Carle based X-ray tracking routines (XTRACK) [27]. This yields the ideal resolution function for the given experimental setup.

2.5.3. Gaussian width G

The comparison of the ideal response function with dedicated calibration measurements [28, 29], revealed an additional Gaussian component, which is the cumulative width of all additional broadening effects. This deviation from the ideal case was determined from dedicated calibration measurements and added to the response function. For special cases though, the Gaussian contribution was left as a free parameter in the analysis.

2.5.4. Spectrometer response

The spectrometer response describes the total broadening effect stemming from various experimental effects. The combination of rocking curve R_C of the Bragg crystal, geometrical aberrations R_G and a Gaussian G originating from imperfections of the setup then reads:

$$R = R_C \otimes R_G \otimes G \tag{2.9}$$

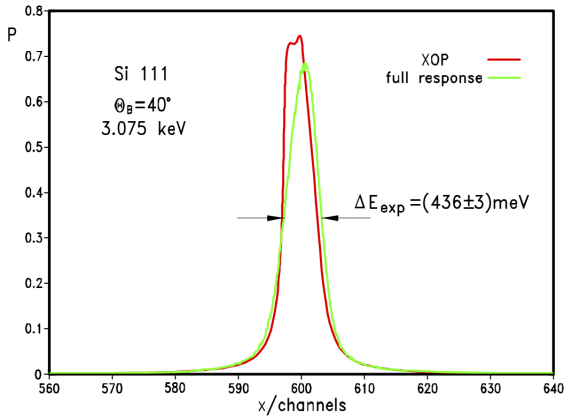


Figure 2.4.: Rocking curve calculated by means of the program package XOP and full response for a spherically bent Si crystal having a radius of curvature of 3 m and an energy of 3.075 keV (πD(3p-1s) transition). The full response is obtained by convolving the rocking curve with the geometry of the setup by means of Monte-Carlo X-ray tracing code (XTRACK) [3, 27].

3. Analysis strategies

3.1. The conventional method

3.1.1. A brief description

The conventional method to analyze the X-ray spectra of exotic atoms consisted of a large number of χ^2 fittings. χ^2 analysis is based on a comparison between a hypothesis $f(x, \underline{p})$ (dependent on a parameter(vector) $\underline{p}$) and a set of data points (x_i, y_i) and their error margins σ_{y_i}

$$\chi^2(\underline{p}) = \sum_{i=1}^{k} \left(\frac{f(x_i, \underline{p}) - y_i}{\sigma_{y_i}} \right)^2 \quad .$$

The best estimated values of the parameters $\underline{p}$ are expected to be found for a minimal $\chi^2(\underline{p})$. This is justified by the presumption that, under certain assumptions, the cumulation of the discrepancies between the data and the model $f(x_i, \underline{p}) - y_i$ weighted by the error margin σ_{y_i} should be minimal [30].

In order to find the optimal parameter vector $\underline{p}$, each parameter dimension p_i has to be varied iteratively and their χ^2 evaluated and compared.

3.1.2. Limitations

Error margins and confidence

One point of the χ^2 based parameter inference is the implicit predefinition of normal distributed parameter values. In reality, the distribution and therefore the error margins / confidence intervals do not have to be symmetric and are not necessarily normal distributed (see chapter 5). This results in an additional layer of complexity to determine asymmetric error margins in the conventional method. In Bayesian methods these problems are solved inherently (see section 3.2). A demonstration of the difference between weighted averages plus sample variance, and the most probable value plus asymmetric error margins can be found in Figure 3.1.

Bias

As the χ^2 function denotes the probability distribution of the sum of the squares of k independent standard normal distributed random variables x_i, using the χ^2-test to infer model parameter values implies the assumption, that the data points are normally distributed. Especially for binned counts in an X-ray spectrum, this assumption can be

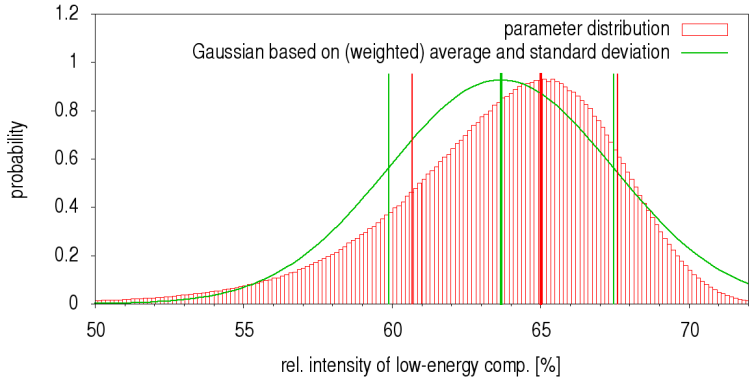


Figure 3.1.: Illustration of the conventional method of parameter inference and error estimation. The calculation of an weighted average and sample variance implies the assumption that the distribution can be approximated by a Gaussian. In this example (taken from the μH analysis in section 5.4) the distribution of the relative intensity is highly asymmetric, and limited to the interval of $0 - 100\,\%$. Thus, the weighted average is lower than the most probable value, and the error margins are not asymmetric.

problematic as integer counts are generally assumed to be Poisson distributed. For large integers, the Poisson distribution converges to a normal distribution and therefore the assumption holds, but for low counts in the tails of an X-ray spectrum, this is too vague.

Effectively, in such analyses a bias occurs, that means a constant but unknown shift of the parameter values [31]. To determine this shift, detailed tests using simulated spectra with known parameter values are necessary. In order to simulate a spectrum, a model has to be chosen and assumed to be true. This leaves little room for hypothesis tests as well as an unambiguous error handling.

3.2. Bayesian approach

3.2.1. Bayesian statistics

Bayesian statistics provide a different view on questions of data analysis and parameter inference. The idea goes back to Thomas Bayes in 1763 [30] and is founded on basic probability theory. Thomas Bayes was an English minister in the 18^{th} century and deduced a first version of the theorem now known as BAYES' THEOREM. Bayes' work was only published after his death by his friend Richard Price, but did not attract much interest at that time. That was until Pierre-Simon Laplace independently rediscovered his ideas and developed them further in the beginning of the 19^{th} century.

In order to understand the Bayesian approach to data analysis, one has to grasp the fundamentally different treatment of models / hypotheses and their parameters. An exhaustive description is provided in [32], but the principles shall be discussed now.

Common frequentist methods usually consider parameters as real, fixed values to infer. Bayesians though, consider each parameter and each hypothesis as a statistical variable and formulate prior knowledge and information as probabilities of these variables. Given these, using probability calculus, one can deduce probability (distributions) for variables (*e. g.* parameters) of interest. This is the very essence of Bayesian statistics and leads directly to the famous BAYES' THEOREM.

The Bayes' theorem

The most common problem in data analysis concerns the inference of model parameters from experimental data. Contrary to frequentistic methods, in Bayesian statistics, one does not pose the (intuitive) question *What is the value of the parameter λ?*, but rather *Which parameter value is the most probable, given the information and data at hand?*, or even *What is the probability for a certain value of a parameter given the information and data at hand?*. This probabilistic interpretation of an unknown but fixed and real parameter value is the source for many discussions and a frequent point of criticism on Bayesian statistics, but shall not be discussed here.

The generic approach, considering the experimental data D as well as the parameter(s) λ as random variables, may be expressed in an (unknown) probability distribution

$$prob\big(D, \lambda \,\big|\, I\big) \ .$$

Using the algebraic rule for conditional probabilities of two random variables A, B, $prob\big(A, B\big) = prob\big(A \,\big|\, B\big) \ \times \ prob\big(B\big)$

$$\begin{aligned} prob\big(D, \lambda \,\big|\, I\big) &= prob\big(D, \lambda \,\big|\, I\big) \\ L(D \,|\, \lambda, I) \times \pi(\lambda \,|\, I) &= P(\lambda \,|\, D, I) \times E(D \,|\, I)\,, \end{aligned}$$

yields the **Bayes' theorem**:

$$\Rightarrow P(\lambda \,|\, D, I) = \frac{L(D \,|\, \lambda, I) \times \pi(\lambda \,|\, I)}{E(D \,|\, I)} \tag{3.1}$$

often abbreviated as (see section 3.2.1):

$$P(\lambda \,|\, D, I) \propto L(D \,|\, \lambda, I) \times \pi(\lambda \,|\, I) \tag{3.2}$$

with a proportionality constant of $1/E(D \,|\, I)$ (constant for a fixed dataset D).

The distributions P, L, E, and π are each of particular importance and shall now be discussed individually. For real-valued, continuous parameters, the distributions of course become probability density functions.

The likelihood function L

The most important component of the posterior probability $P(\lambda \,|\, D, I)$ is the **likelihood function** $L = L(D \,|\, \lambda, I)$, describing the probability that the data point D was measured, given an (unknown) parameter value λ and information I. It is, however, not a probability distribution, but a mere function of λ. These probabilities are usually deducible with the hypothesis, possibly together with assumptions about the statistical nature of the experimental data. Given that, the likelihood function plays a key role, as it contains the connection between data and parameters via a hypothesis.

Example 1. Gaussian distributed events
Let λ_0 be the true value of a parameter λ, and d_k a data point, assumed to be normally-distributed around λ_0, then the probability for a data point to take the value d_k is given by

$$\begin{aligned} prob\big(d_k \,|\, \lambda_0\big) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(d_k - \lambda_0)^2} \qquad \text{(Gaussian distribution with } \sigma = 1\text{)} \\ &\equiv L(d_k \,|\, \lambda_0)\,, \end{aligned}$$

which constitutes a Bayesian likelihood function. The extension to n statistically independent data points consists of combining these distributions of independent variables d_k by multiplication

$$\begin{aligned} L(d_1, d_2, \ldots, d_n \,|\, \lambda_0) &= \prod_{1 \le k \le n} L(d_k \,|\, \lambda_0) \\ &= \prod_{1 \le k \le n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(d_k - \lambda_0)^2} \quad . \end{aligned}$$

The prior probability π

Information and knowledge about the parameter space manifests itself in the **prior probability** $\pi = \pi(\lambda \,|\, I)$. Prior knowledge could be as simple as hard constraints in parameter space or complex results of a previous experiment.

The ambiguity in choosing a prior distribution and the subjectivity associated with this, is often basis for criticism against the whole Bayesian methodology. The criticism, though, can be weakened if either a prior is chosen which has no influence on the posterior, or the likelihood function contains so much information (*e. g.* a combination of large amounts of data points), that the prior becomes negligible. The latter case is demonstrated, in the example in Figure 3.3, where the influence recedes with more informative likelihoods.

Example 2. Confined parameter space
Parameter space constraints $\lambda \in [a, b]$ can be implemented using a confined uniform distribution

$$prob\big(d_k \,|\, \lambda\big) = \begin{cases} \frac{1}{b-a} & \lambda \in [a, b] \\ 0 & \text{otherwise} \end{cases} \equiv \pi(\lambda \,|\, I) \ .$$

The evidence E

In the more usual cases of data analysis, the data D and the information I (containing assumptions and a specific hypothesis) remain constant, making the **evidence** $E(D \,|\, I)$ a constant value and a mere normalization factor of the posterior distribution P. In scenarios of model-selection (comparing different hypotheses) though, the evidence becomes the key component (hence the name) and the computation of $E_i(D \,|\, I_i)$ for different models I_i yields real valued and comparable weightings for each model. As this topic of model selection is of particular importance for this thesis, it is discussed thoroughly in section 3.2.3. For single model problems, the evidence usually can be ignored.

From eq. 3.1 and eq. 3.2 it becomes obvious, that the evidence can be expressed as the normalization of the relative posterior distribution

$$E(D \,|\, I) = \int L(D \,|\, \lambda, I)\, \pi(\lambda \,|\, I)\, d\lambda$$

or as marginalization of $prob\big(D, \lambda \,|\, I\big)$ over λ

$$E(D \,|\, I) = \int prob\big(D, \lambda \,|\, I\big)\, d\lambda \ .$$

Especially the latter underlines the importance of the evidence as it is the combined probability of the data D, given a model, over all possible parameter configurations.

The posterior probability P

The **posterior distribution** $P = P(\lambda \,|\, D, I)$ represents the information available about the parameter λ, given the (experimental) data D and any (circumstantial) information I, namely possible values of λ and the corresponding probability of these values being the 'real' parameter value. This is the expressive result of Bayesian data analysis. In contrast to frequentist methods (*e. g.* χ^2 fitting routines), Bayesian data analysis not only provides an expectation value along with a variance, but an entire probability distribution for the parameter(s), including a most probable value, (asymmetric) uncertainties, confidence intervals etc. More advantages of having a probability distribution of a parameter at hand will become clear later on in the actual analysis.

3.2.2. A simple example

The matter of the Bayesian posterior may best be illustrated by the example, given in the textbook of D. S. Sivia [30], of an experiment to determine whether a simple coin is fair (*i. e.* the probability for *heads* and *tails* are equal) or not. If H denotes the (unknown) rate at which the coin shows *heads*, then the probability for counting R *heads* in N tosses is given by the binomial distribution

$$\begin{aligned} prob\big((R,N) \,\big|\, H, I\big) &= H^R \times (1-H)^{N-R} \\ &\equiv L\big((R,N) \,\big|\, H, I\big) \end{aligned} \qquad \text{(likelihood)}$$

and represents the likelihood in this scenario. Lack of prior information and openness towards the outcome manifests itself in an uniform prior distribution

$$\begin{aligned} prob\big(H \,\big|\, I\big) &= \begin{cases} 1 & 0 \leq H \leq 1 \\ 0 & \text{otherwise} \end{cases} \\ &\equiv \pi(H \,|\, I) \end{aligned} \qquad \text{(prior)}$$

because the rate H cannot be negative and cannot exceed 1 (this would imply more *head* outcomes than tosses in total).

Combining these distributions according to the Bayes' theorem and ignoring the constant evidence yields the posterior distribution

$$P\big(H \,\big|\, (R,N), I\big) = \begin{cases} H^R \times (1-H)^{N-R} & 0 \leq H \leq 1 \\ 0 & \text{otherwise} \end{cases} . \qquad \text{(posterior)}$$

Suppose an experiment was performed by tossing the coin 50 times. The series of outcomes is shown in Figure 3.2 with H for *heads* outcome and T for *tails*.

HHTHTTTTTTHHTTHTHTHTTTHHTTHHTHTHTHTHTHTTTTTHTTTT

Figure 3.2.: Outcomes of a series of coin tosses

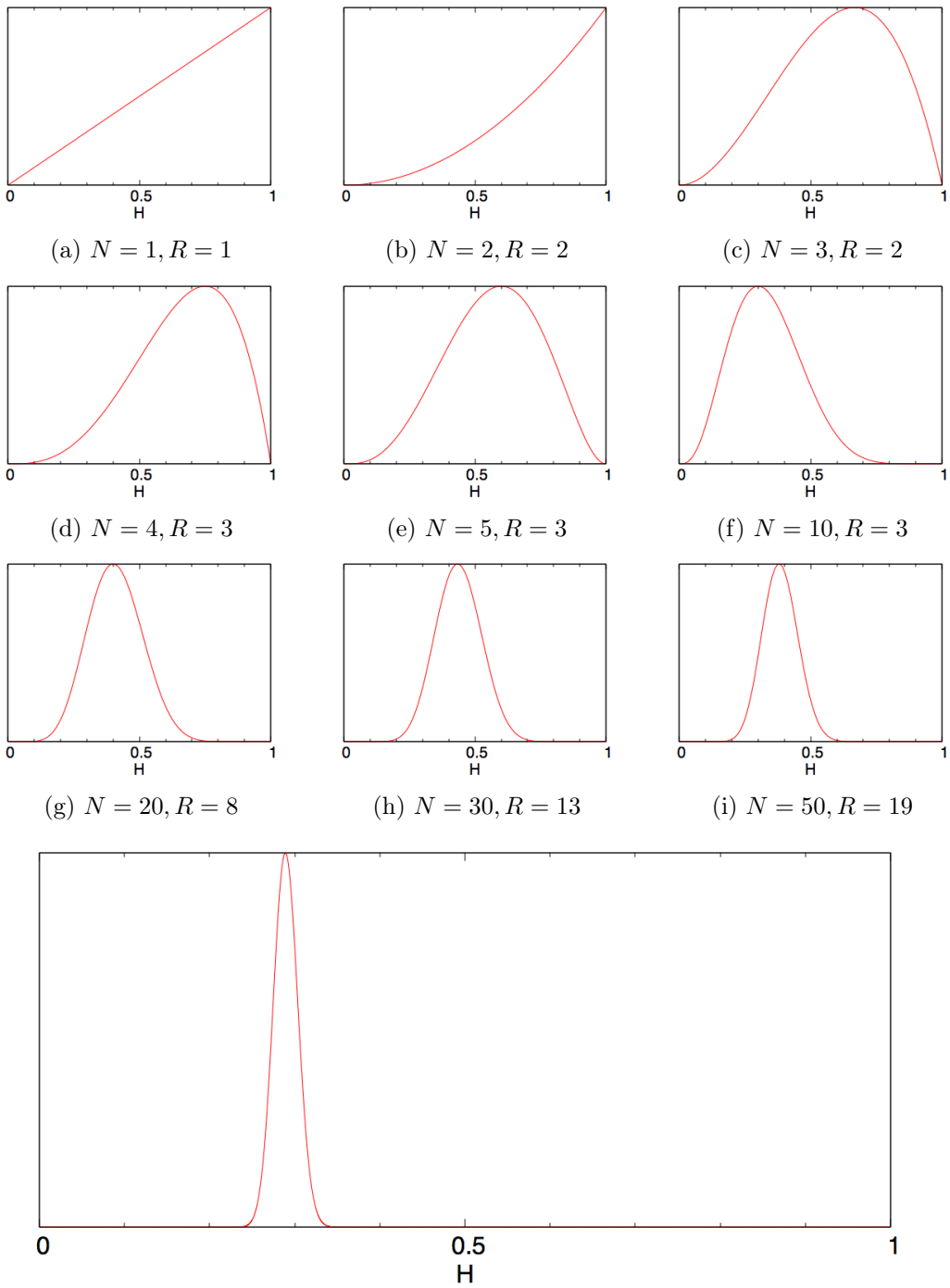
Visualizations of the posterior distribution after each toss illustrate the aspect, that

with every data point the information about the parameter H increases, and the influence of the prior distribution decreases.

Figure 3.3 demonstrates the best knowledge after 1, 2, 3, 4, 5, 10, 20, 30, and 50 tosses, respectively. It is obvious, that with just a few data points the posterior changes vastly with each new piece of information, but that with 20 tosses, it begins to stabilize and to refine around a central value. The important point to make here is, that the central value of this posterior is generally not the real value of H, but the most probable value of H **given** the available data. This is the important difference between conventional data analysis and the Bayesian approach.

Feasability

Despite the obvious advantages, there are, however, some fundamental drawbacks in the practical application. Derivation of an analytical posterior is often not possible and even the likelihood or prior may not exist in a closed analytical form. This means, that for most real world problems, one has to resort to numerical methods. As these are usually computational intensive, Bayesian data analysis has just recently become feasible. The numerical methods and their implementation are treated separately in chapter 4 and appendix A.1 .



(a) $N = 1, R = 1$ (b) $N = 2, R = 2$ (c) $N = 3, R = 2$

(d) $N = 4, R = 3$ (e) $N = 5, R = 3$ (f) $N = 10, R = 3$

(g) $N = 20, R = 8$ (h) $N = 30, R = 13$ (i) $N = 50, R = 19$

(j) Posterior distribution after $N = 1000$ tosses with $R = 289$ *heads*

Figure 3.3.: Posterior distributions for different data sets for N tosses yielding R heads. The approximation improves with every new data point and would contract for $N \to \infty$ to a δ-peak at the real value H.

3.2.3. Model-selection

A powerful aspect of Bayesian statistics is the inherent support for model-selection problems and consequently model-averaging of inferred parameters [33]. Following Bayesian principles, evaluating models M_i against one another becomes a question of finding the probability for this model M **given** the data D

$$prob\big(M \,\big|\, D, I\big) \propto prob\big(D \,\big|\, M, I\big) \times prob\big(M \,\big|\, I\big)$$

Obviously, determining an absolute probability for a particular model is not possible because this would require knowledge about all (uncountable) possible models and their respective probabilities. In order to compare and evaluate a set of models M_i though, relative probabilities are sufficient and provide coherent, relative weights for all tested hypotheses. These relative weights are called **Bayes factors**

$$\begin{aligned} B_{1,2} &= \frac{prob\big(M_1 \,\big|\, D, I\big)}{prob\big(M_2 \,\big|\, D, I\big)} \qquad \text{(Bayes factor)} \\ &= \frac{prob\big(D \,\big|\, M_1, I\big) \times prob\big(M_1 \,\big|\, I\big)}{prob\big(D \,\big|\, M_2, I\big) \times prob\big(M_2 \,\big|\, I\big)} \ . \end{aligned}$$

Without any prior knowledge or assumptions about these models, ${}^{prob(M_1 \,|\, I)}\!/_{prob(M_2 \,|\, I)}$ must be regarded as unity, and therefore the Bayes factor becomes

$$\begin{aligned} B_{1,2} &= \frac{prob\big(D \,\big|\, M_1, I\big)}{prob\big(D \,\big|\, M_2, I\big)} \\ &\equiv \frac{E(D \,|\, M_1, I)}{E(D \,|\, M_2, I)} \ . \end{aligned}$$

This is equal to the *ratio of the evidences*, making the evidences the most important value to be determined in model-selection problems.

Model evaluation

The assignment of (relative) weights, and thus providing an order for a set of models does not complete the task of model-selection. An important part remains, namely the interpretation of these weights to evaluate and eliminate hypotheses. The question is, however, of which magnitude must a Bayes factor be, in order to mark one hypothesis so superior over another, that the other can be eliminated. A good scale is given by Harold Jeffreys [34] to compare Bayes factors B to standard deviations σ and confidence levels (see Table 3.1).

B	$\ln(B)$	σ	p-value	category
2.5	0.9	2.0	0.05	
2.9	1.0	2.1	0.04	*weak*
8.0	2.1	2.6	0.01	
12	2.5	2.7	0.006	*moderate*
21	3.0	3.0	0.003	
53	4.0	3.3	0.001	
150	5.0	3.6	0.0003	*strong*
43000	11	5.0	6×10^{-7}	

Table 3.1.: Comparison of Bayes factors to σ and p-values, given by Harold Jeffreys in [34]. This table provides a lead to quantify the difference between different models and relate it to classical (by σ and p-values) methods of assessment.

Posterior volume

The evidence $E(D \,|\, M, I)$ can be expressed as a marginalization of a broader probability distribution

$$\begin{aligned} E(D \,|\, M, I) &= \int prob\big(D, \lambda \,|\, M, I\big)\, d\lambda \\ &= \int prob\big(D \,|\, M, I, \lambda\big) \times prob\big(\lambda \,|\, M, I\big)\, d\lambda \end{aligned}$$

where the integral is just the volume / normalization factor of the relative posterior distribution. This means, determination of the evidence requires the integration of the relative posterior distribution over the entire parameter space. Integration of non-analytic functions over a high-dimensional parameter space, though, poses a complex challenge and requires efficient numerical algorithms to perform such computations in an acceptable time span.

3.2.4. Model-independent approach

Bayesian methods give rise to the possibility of analysing experimental data without the need to settle on a specific model. This can be extended to a point, where a set of different models is chosen and evaluated, not to verify or falsify one or more of the models, but weaken the dependence on the choice of the model. As every model has a probability weight, given by the Bayes factor, inference of a (common) parameter may be performed across all models by including their weights.

This is an important advantage over the conventional method, especially regarding the bias problem (see section 3.1.2). Having a set of models M_i, along with their evidences E_i or Bayes factors (between two models M_i, M_j) $B_{i,j} = {}^{E_i}/_{E_j}$, two basic options are available:

Selection that means, grading by evidence and subsequent verification / falsification, exclusion of very unlikely models (with $B_{i,j} > 150$, see Table 3.1). Picking one single model over others, though, would require a substantially higher evidence.

Averaging over a set of models, *i. e.* determination of the posterior distributions for common parameters considering each model, if the evidences show no preference for a single model.

In principle, the latter already includes the former because, due to the weighting by evidence, models of significantly larger evidences dominate and unlikely models barely contribute on averaging.

4. Numerical methods

4.1. Posterior sampling

For real world problems, the posterior $P(\lambda \,|\, D)$[1] can rarely be specified in analytical form. An alternative to an analytical form is drawing a substantial amount of random samples of the posterior distribution using Monte-Carlo methods. With the help of these samples, any integral (*i. e.* the expectation value) over a statistical variable λ can be approximated by

$$\begin{aligned} E[f(\lambda)] &= \int f(\lambda) P(\lambda \,|\, D)\, d\lambda \\ &\approx \frac{1}{N} \sum f(\lambda_i) \end{aligned}$$

specifically for the expectation value $E[\lambda]$ of the mean being

$$E[\lambda] \approx \frac{1}{N} \sum \lambda_i$$

with λ_i being N independent samples drawn from the posterior $P(\lambda \,|\, D)$. The real challenge, however, lies in drawing the independent samples λ_i.

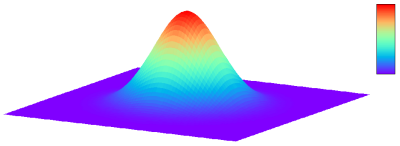
4.1.1. Markov chain Monte-Carlo methods

A common approach to produce such samples are Markov chain Monte-Carlo (MCMC) methods like the Metropolis-Hastings algorithm [35]. These methods usually start with a random point and find new points by creating a Markov chain (*i. e.* via random walk steps) and ensure pairwise independence by taking only each n^{th} point. For these sampling methods relative probabilities are sufficient, making them perfect for problems where the normalization (in the Bayes' theorem: the evidence) is not of interest. But because the Metropolis-Hastings algorithm only produces samples following

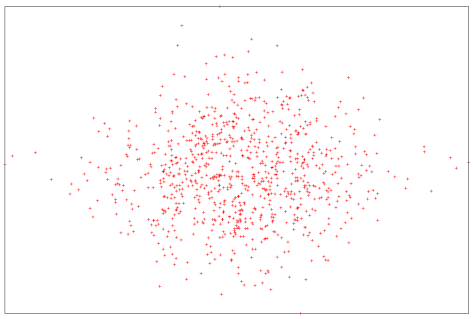
$$P'(\lambda \,|\, D) \propto L(D \,|\, \lambda) \times \pi(\lambda)$$

with an unknown proportionality constant (corresponding to the reciprocal of the evidence), it is not suitable for model-selection problems.

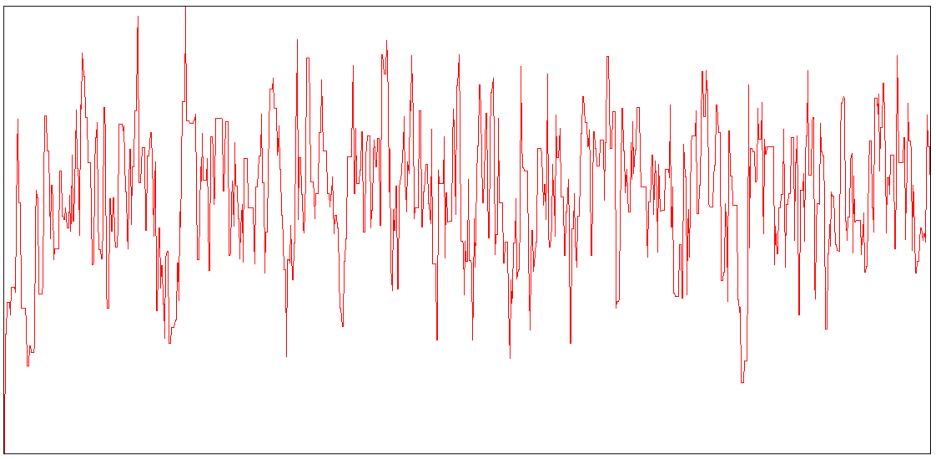
[1] The dependency on the contextual information I is omitted from now on for brevity.



(a) Bivariate normal distribution $P(x, y)$ to draw samples from.



(b) Scatterplot of drawn samples: the density of the samples is proportional to their probability.



(c) First 1000 iterations (x-component) of the Markov chain used to draw samples above.

Figure 4.1.: Visualization of a Metropolis-Hastings sampling run to draw samples from a bivariate Gaussian distribution.

4.1.2. Nested sampling

Nested sampling is a Lebesgue-integration inspired method developed by John Skilling [30, 36] to calculate the evidence of a Bayesian posterior distribution. As a by-product, it can generate posterior samples and therefore provides a complete solution to model-selection problems.

It is based on the idea to map the n-dimensional parameter space to a 1-dimensional, real-valued interval, removing the challenge to integrate over n dimensions. Given an n-dimensional parameter vector $\underline{\lambda}$

$$\underline{\lambda} = (\lambda_1, \ldots, \lambda_n),\ \lambda \in \mathbb{R}^n$$

and a posterior

$$P(\underline{\lambda} \,|\, D) \Rightarrow \tilde{P}(x \,|\, D) = L(D \,|\, \underline{\lambda}(x))\, \pi(\underline{\lambda}(x))\,,$$

then the marginalization (and therefore the evidence) can be expressed with

$$\begin{aligned} & x \in [0,1] \\ \int_{\mathbb{R}^n} P(\underline{\lambda} \,|\, D)\, d\underline{\lambda}^n &= \int_0^1 \tilde{P}(x \,|\, D)\, dx \end{aligned}$$

and thus,

$$\begin{aligned} \int_{\mathbb{R}^n} P(\underline{\lambda} \,|\, D)\, d\underline{\lambda}^n &= \int_{\mathbb{R}^n} L(D \,|\, \underline{\lambda})\, \underbrace{\pi(\underline{\lambda})\, d\underline{\lambda}^n}_{dx} \\ &\approx \sum_i L(D \,|\, \underline{\lambda}(x_i))\, \Delta x_i, \quad \Delta x_i = x_{i+1} - x_i \ . \end{aligned}$$

Calculating the sum now requires knowledge of the transformation $\underline{\lambda}(x)$ with $L(D \,|\, \underline{\lambda}(x))$ being positive and monotonically decreasing and a sequence of x_i so that the $\Delta x_i = \pi(\underline{\lambda})\Delta\underline{\lambda}^n$ are volume elements of the parameter space $\mathbb{R}^n$ weighted by the prior probability density $\pi(\underline{\lambda})$.[2] These Δx_i are therefore called **prior masses**.

To approximate the integral, one starts with the right-most rectangle in Figure 4.2, the area of which is given by

$$\underbrace{L(D \,|\, \underline{\lambda}(x_0))}_{L_0}\, \Delta x_0 \ .$$

If m independent samples $\underline{\lambda}_k$ (called live points) are drawn from the prior distribution $\pi(\underline{\lambda})$, then, because of $\pi(\underline{\lambda}_k)d\underline{\lambda}_k = dx_k$, their corresponding x-values are uniformly distributed within $[0, 1]$.

[2] During the actual analysis a more precise approximation of $\Delta x_i = {}^1\!/_2(x_{i-1} - x_{i+1})$ was used instead of $\Delta x_i = x_{i+1} - x_i$

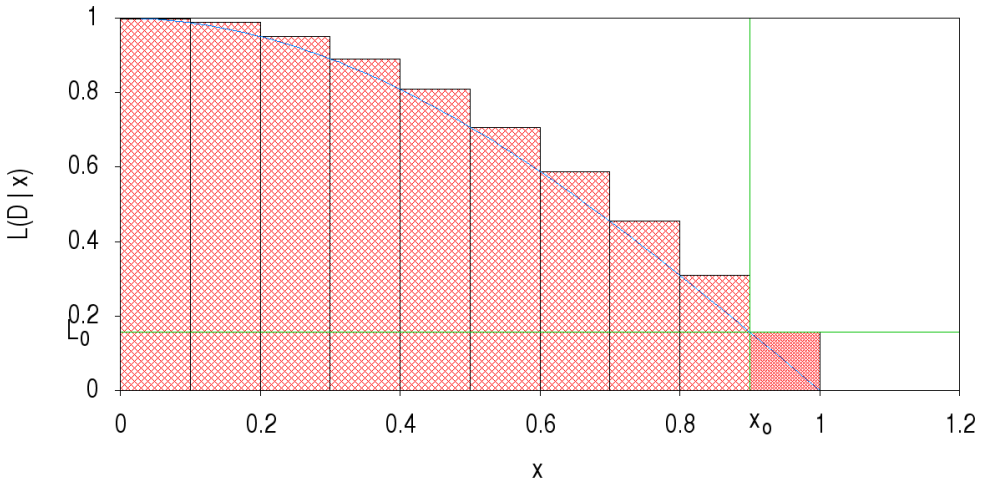


Figure 4.2.: Integral approximation with prior masses on the x-axis.

These uniformly distributed x-values have an average distance (between adjacent values) of $1/m+1$. Sorting these live points by their likelihood $L_k = L(D \mid \underline{\lambda}_k)$ ensures, that the posterior in x is monotonically decreasing.

The rectangular area may now be approximated using the lowest likelihood L_0 of all live points, along with an estimate of $\Delta x_0 \approx 1/m+1$. This approximation of Δx_k must be justified by a large number of iterations and live points.

The rest of the integral can be calculated the same way by iterative nesting. In order to do that, the x-interval has to be reduced by Δx_0 and a new set of m prior live points must be found within this new interval. As there are already $m-1$ samples, which by definition are left of x_0 and therefore in the new, reduced interval, it is sufficient to draw one new point with a likelihood $L > L_0$. After each iteration k, the remaining interval is shrunk by $1/m+1$. This gives a decreasing, but exponential series of intervals Δx_k along with their increasing likelihood values L_k (see Figure 4.3) and thus an approximation to the integral $E(D)$ via

$$E(D) \approx \sum_k L_k \Delta x_k \quad .$$

Although this series is infinite in theory, the exponential character of the series of Δx_k ensures that the summands at some point will tend to zero, adding only marginal amounts to the evidence. This allows the nested sampling iterations to stop as soon as the desired accuracy is achieved.

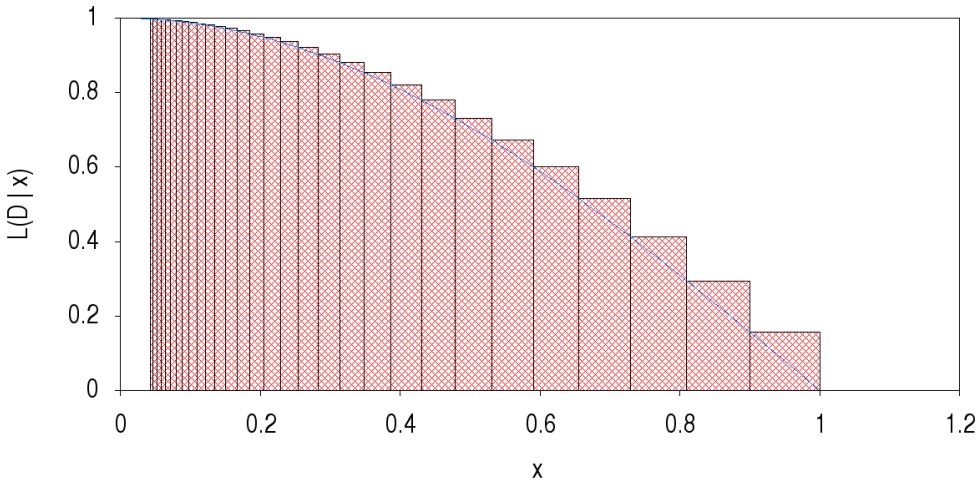


Figure 4.3.: Example of nested sampling with 30 iterations and $m = 10$ samples. The $\Delta x_{i,i+1} = x_{i+1} - x_i$ decrease exponentially, while the likelihoods $L_i = L(x_i)$ are by definition monotonic increasing. The products $L_i \, \Delta x_{i,i+1}$, being the areas of the shown rectangles, eventually tend to zero. The remaining posterior area (left blank from $x = 0$ to the last x_i) has to be approximated using the remaining prior mass and the average of the final set of live points.

The general procedure for a nested sampling evaluation is as follows:

1. Prepare a set of initial live points. These points are sampled from the entire prior distribution.
2. Perform nested sampling iterations and sum up the evidence until the desired accuracy is achieved using the following steps:
   a) Find the live point with the lowest likelihood $L_{\min}$.
   b) Add the weight of the lowest point $\Delta x_i L_{\min}$ to the evidence.
   c) Explore the prior for a new live point with $L_{\text{new}} > L_{\min}$.
   d) Replace the lowest point with the new point.
3. Add the remaining live points to the posterior collection and add their averaged likelihood multiplied by the remaining prior mass as last summand to the (logarithmic) evidence.
4. Post processing using the evidence and the posterior samples (*e. g.* histograming, calculation of means and standard deviations).

Although this constitutes a simple and straightforward algorithm, there are two steps, that usually pose a challenge in implementations of nested sampling.

Likelihood evaluation The evaluation of the likelihood of a specific parameter vector ideally requires an analytical form of the model to apply the parameters and compare it with the data. Even with analytical models, this can be in total rather CPU-intensive, for the models used later in the analysis (see chapter 5), an analytical form does not exist and therefore application of the parameters requires interpolation of discrete, numerical test spectra.

Prior exploration Exploring new samples from the prior distribution while satisfying the requirement of greater likelihood for new points can be the bottleneck in a nested sampling implementation. A simple hit-and-miss sampler for example would not work as the number of rejections increases exponentially. This crucial part requires an efficient solution to make nested sampling feasible.

4.2. Prior space sampling

To resume the explanation in section 4.1.2, exploration of the prior space means sampling a point from the prior distribution under the constraint that the likelihood L_{new} of the new point is higher than a specified likelihood value L_0. The most simple way to explore a new point would be a hit-and-miss loop:

1. Sample a point from the prior distribution.
2. Evaluate the likelihood L_{new} of the new point.
3. If $L_{\text{new}} \leq L_0$, return to step 1.

For this thesis, three alternative methods have been tested:

- MCMC-exploration
- ellipsoidal-sampling
- multimodal-ellipsoidal-sampling

4.2.1. MCMC-exploration

Idea

The concept of this method developed by L. Simons [27] is based on the MCMC technique described in section 4.1.1. The idea is to perform a series of random walk steps within the volume bounded by the equi-likelihood hypersurface with the minimal likelihood L_0 of the current live points collection. This hypersurface must be closed for any likelihood L_0. Figure 4.4 shows a two-dimensional example of a random walk sequence within the likelihood boundary. The shape of this boundary is unknown because there are not enough points to establish a contour. As long as the random walk series stays within that boundary, a new point can be found with n consecutive steps (depending on the dimensionality of the parameter space).

As the sampled points have to be independent of each other, a suitable combination of a number of steps n and step size $\underline{s}$ has to be found to prevent correlation between the starting point and the sampled point. The starting point usually is chosen at random from the collection of live points, the likelihood of which is by definition greater than L_0.

If a step within a sequence crosses the boundary, the point is discarded and a new point is chosen randomly between the previous point and the center of mass (of the live points). Figure 4.5 demonstrates this case. The sequence continues from the new point in this manner until the desired number of steps is reached.

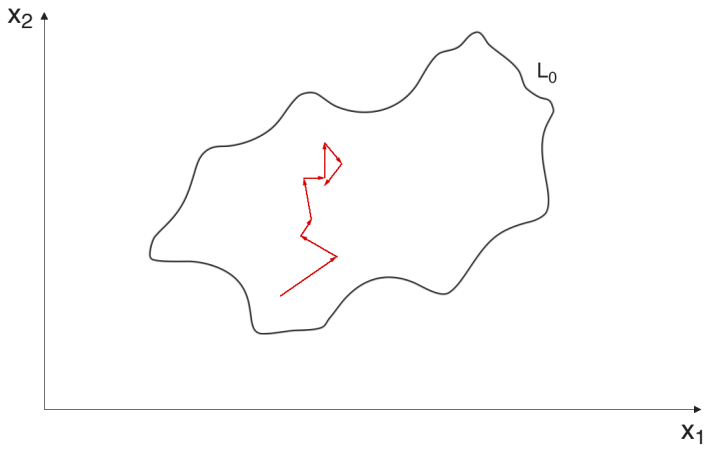


Figure 4.4.: Illustration of a random walk sequence within the likelihood boundary in a two-dimensional parameter space. The contour line L_0 is, of course, unknown and purely hypothetical.

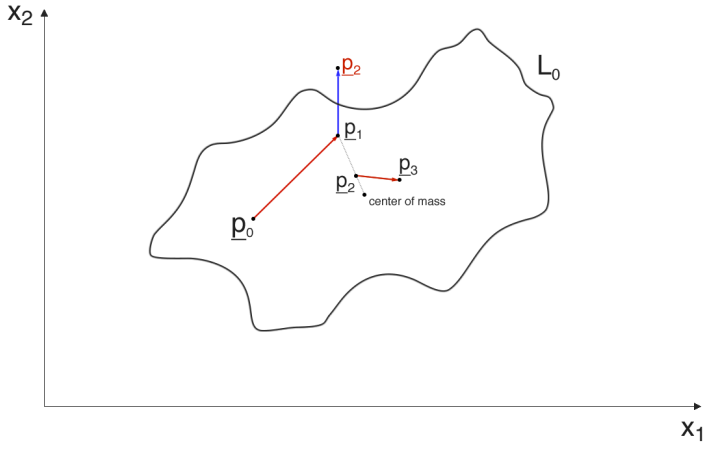


Figure 4.5.: Illustration of a random walk sequence $\underline{p}_{0,1,2,3}$ that hit the likelihood constraint. As soon as a step ($\underline{p}_2$) falls below the constraint L_0, a new point $\underline{p}_2$ is chosen at random between the previous point $\underline{p}_1$ and the center of mass.

The algorithm

In order to find a new point, the following procedure was implemented:

1. Choose a random starting point from the collection of live points.
2. Starting from this point, perform a random walk step and evaluate the likelihood of the new point.
3. If the likelihood is less than L_0, return to the previous point and choose a new point randomly between the previous point and multidimensional mean (center of mass) of the live point collection.
4. Repeat steps 1 to 3 until the desired number of steps is reached.
5. A new point satisfying $L > L_0$ is found.

(Dis-)Advantages

The advantage of this kind of exploration procedure is its simplicity in implementation. Contrary to other techniques, the complexity depends only weakly on the dimensionality n of the parameter space via the desired number of steps (the more dimensions the more steps are needed to ensure uncorrelated samples). These properties make the MCMC-Explorer a computational efficient method to produce new points in nested sampling.

A drawback, however, is the dependence on a number of configuration parameters, which have to be set carefully in order to achieve a good performance and to prevent deadlocks. These $n + 1$ configuration parameters (*i. e.* the desired number of steps m as well as the (n-dimensional) step-size $\underline{s}$) must be chosen in a trial-and-error manner before the main analysis runs.

A problem that occurs occasionally with this exploration method is that a set of configuration parameters works well for one subset of the models to analyze, but causes deadlocks for other (but similar) models. In order to overcome this issue, another method, developed by F. Feroz and M.P. Hobson [37] was implemented and tested, namely **ellipsoid-sampling**, which approximates the contour using an n-dimensional ellipsoid.

4.2.2. Ellipsoidal-sampling

Idea

Ellipsoidal-sampling uses the presumption, that the volume of the parameter space enclosed by the $L > L_0$ likelihood constraint, can be approximated or at least enveloped by an n-dimensional ellipsoid. This is similar to approximating the posterior distribution with a multivariate Gaussian, which has elliptical contour lines [38]. New points are produced by sampling uniformly from within this ellipsoid, followed by a hit-and-miss selection (according to the prior distribution and the likelihood constraint). The challenge in this method is the determination of and sampling within the ellipsoid.
Thus the general strategy is:

1. Calculate an enveloping ellipsoid to the live points.
2. Sample a point from the uniform distribution, bounded by this ellipsoid (see eq. 4.4).
3. In order to ensure that an ensemble of generated points is distributed according to the prior, a hit-and-miss selection step has to be performed, which only accepts points randomly with a frequency proportional to the prior probability.
4. Compare the likelihood with L_0. If $L \leq L_0$, reject the point and start over (step 1).

Determination of an enveloping ellipsoid

As mentioned previously (see section 4.2.1), the specific shape of the L_0 boundary is not known. Available, however, is a (rather large) set of live points within its volume. In order to get at least an approximation to the shape of the k live points distribution, the covariance matrix $cov(\underline{p}_i)$ can be determined by

$$cov(\underline{p}_i) = \frac{1}{k+1} \sum_{i=1}^{k} (\underline{p}_i - \underline{p}_{\text{mean}})(\underline{p}_i - \underline{p}_{\text{mean}})^T$$

with

$$\begin{aligned} \underline{p}_i &\equiv \text{live points} \\ \underline{p}_{\text{mean}} &\equiv \text{mean of live points} \ . \end{aligned}$$

The inverse of the covariance matrix $\mathbf{C} = cov^{-1}(\underline{p}_i)$ (also known as precision matrix) represents the elliptical hypersurface of an (assumed) multivariate Gaussian distribution.

This hypersurface can be parameterized by a distance $f(\underline{p})$, where all $\underline{p}$ with the same $f(\underline{p})$ lie on the same hypersurface. The innermost ellipsoid in Figure 4.6 shows the volume of all $\underline{p}$ with $f(\underline{p}) \leq 1$

$$f(\underline{p}) = (\underline{p} - \underline{p}_{\text{mean}})^T \mathbf{C} (\underline{p} - \underline{p}_{\text{mean}}) \ .$$

By choosing an $f_{\max}$

$$\begin{aligned} f_{\max} &= max[f(\underline{p}_i)] \\ &= max[(\underline{p}_i - \underline{p}_{\text{mean}})^T \mathbf{C} (\underline{p}_i - \underline{p}_{\text{mean}})] \end{aligned}$$

the ellipsoid is re-scaled to include all live points. Although this ellipsoid now contains all k live points, it does not necessarily enclose the L_0 hypersurface entirely.

To minimize the possibility of this case, $f_{\max}$ is usually multiplied by a small factor d^2, $1 < d < 2$, to enlarge the ellipsoid. This configuration parameter d has to be chosen carefully, because the efficiency of this sampling method depends largely on this factor.

Enlargement by a factor d causes an increase in the sampling volume $V_0 \to V$ in an n-dimensional parameter space by

$$V = d^n V_0 \ .$$

If the efficiency of the algorithm is defined by the acceptance rate r of the hit-and-miss selection

$$r = \frac{l_{\text{accepted}}}{l_{\text{total}}}$$

with

$$\begin{aligned} l_{\text{accepted}} &\equiv \text{number of accepted samples} \\ l_{\text{total}} &\equiv \text{total number of samples produced}\,, \end{aligned}$$

then for a uniform prior distribution and a likelihood constraint with the volume $V_{L>L_0}$ the acceptance rate r can be expressed by

$$\begin{aligned} l_{\text{accepted}} &= \frac{V_{L>L_0}}{V}\, l_{\text{total}} \\ \Rightarrow r &= \frac{V_{L>L_0}}{V} = \frac{1}{d^n} \frac{V_{L>L_0}}{V_0} \ . \end{aligned}$$

Under the assumption, that the ellipsoid approximates the $L > L_0$ manifold well enough, r becomes

$$\begin{aligned} r &= \frac{1}{d^n} \underbrace{\frac{V_{L>L_0}}{V_0}}_{\lesssim 1} \\ \Rightarrow r &\lesssim \frac{1}{d^n} \ . \end{aligned}$$

This means that the efficiency can be controlled by choosing d as small as possible, but as large as necessary to ensure the containment of the entire $L > L_0$ manifold within the

ellipsoid.

With a given d and $f_{\max}$, $\underline{p}_{\text{mean}}$, $\mathbf{C}$ calculated from the current collection of live points, a new point $\underline{p}$ must be sampled uniformly within the volume enclosed by

$$(\underline{p} - \underline{p}_{\text{mean}})^T \mathbf{C} (\underline{p} - \underline{p}_{\text{mean}}) \leq d^2 \cdot f_{\max} \tag{4.1}$$

or by consolidating $d, f_{\max}, \mathbf{C}$ to $\mathbf{C}'$

$$\mathbf{C}' = \frac{1}{d^2 \cdot f_{\max}} \cdot \mathbf{C} \tag{4.2}$$

$$(\underline{p} - \underline{p}_{\text{mean}})^T \mathbf{C}' (\underline{p} - \underline{p}_{\text{mean}}) \leq 1 \ . \tag{4.3}$$

This allows the definition of an enveloping probability distribution

$$\mathcal{E}_{L_0} := \begin{cases} \frac{1}{V} & (\underline{p} - \underline{p}_{\text{mean}})^T \mathbf{C}' (\underline{p} - \underline{p}_{\text{mean}}) \leq 1 \\ 0 & \text{otherwise} \end{cases} \ . \tag{4.4}$$

Figure 4.6.: Visualization of the approximating / enveloping ellipsoid. The innermost ellipsoid is the covariance ellipsoid of the live points (1σ), the intermediate one is scaled to include all live points and the outermost is stretched by a constant factor to include the entire $L > L_0$ manifold.

Sampling uniformly within an n-dimensional ellipsoid

Generating random samples from an elliptic volume, specified by a quadratic form (see eq. 4.3) must be done in several steps. Given a Gaussian and a uniform random number generator, the following procedure can be used to generate uniformly distributed points within this ellipsoid.

1. Generate n Gaussian distributed (mean of 0 and $\sigma = 1$), independent values $x_i \sim \mathcal{N}(0,1)$.
2. Normalize the vector $\underline{x} = (x_1, \ldots, x_n)^T$, where $\underline{x}' = \frac{\underline{x}}{|\underline{x}|}$ is now a uniform sample from the surface of the unit hypersphere [39].
3. Sample a value $r \sim \mathcal{U}(0,1)$ uniformly from $[0,1]$, then $\underline{x}'' = \sqrt[n]{r} \cdot \underline{x}'$ is a uniform sample from the volume of the unit hypersphere.
4. Given the n eigenvalues λ_j of $\mathbf{C}'$, using

$$\underline{x}''' = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\lambda_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \ldots & \sqrt{\lambda_n} \end{pmatrix} \cdot \underline{x}''$$

   yields samples from an ellipsoid with the same axes and volume but different orientation and position than the desired ellipsoid.
5. Rotate the sample by multiplication with the eigenvector matrix $\mathbf{E} = (\underline{e}_1, \ldots, \underline{e}_n)$ with $\underline{e}_j$ being the eigenvector to the eigenvalue λ_j

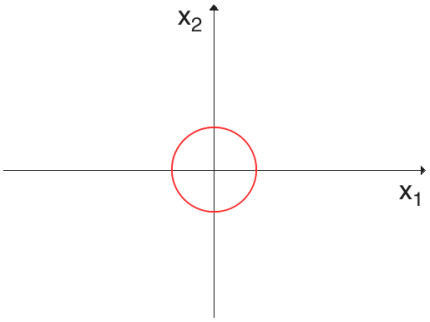
$$\underline{x}'''' = \mathbf{E} \cdot \underline{x}''' \quad .$$

6. And finally translate by the mean $\underline{p}_{\text{mean}}$ to match the center of mass

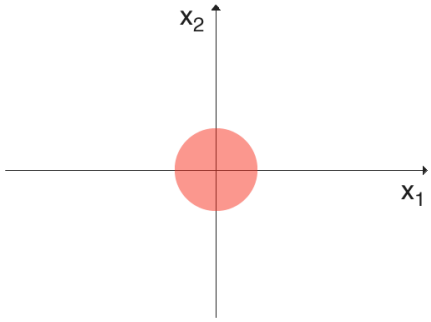
$$\underline{p} = \underline{x}'''' + \underline{p}_{\text{mean}} \quad .$$

This algorithm is visualized in Figure 4.7 for a two-dimensional case. In summary, using $n+1$ samples ($x_i \sim \mathcal{N}(0,1), r \sim \mathcal{U}(0,1)$), a point distributed uniformly within an ellipsoid defined by its quadratic form (see eq. 4.3) can be produced by

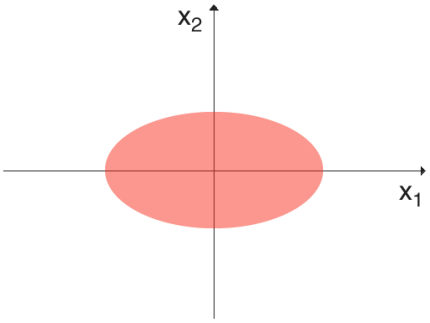
$$\underline{p} = \underline{p}_{\text{mean}} + \mathbf{E} \cdot \begin{pmatrix} \sqrt{\lambda_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\lambda_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \ldots & \sqrt{\lambda_n} \end{pmatrix} \cdot \sqrt[n]{r} \cdot \frac{\underline{x}}{|\underline{x}|} \quad . \tag{4.5}$$



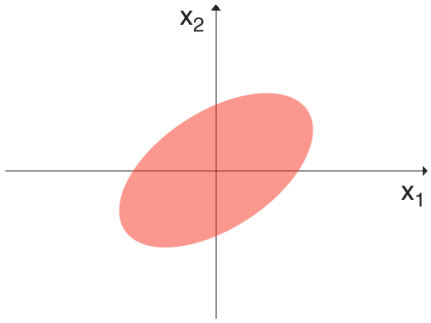
(a) Step 1 and 2: sample from the border of the unit hypersphere (circle), using n Gaussian samples ($x_{1,2} \sim \mathcal{N}(0,1)$) and normalization $x'_{1,2} = \frac{x_{1,2}}{|\underline{x}|}$.



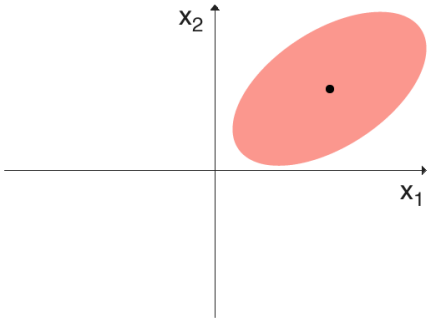
(b) Step 3: using an additional uniform sample $r \sim \mathcal{U}(0,1)$, include the volume within the circle with $x''_{1,2} = \sqrt{r} \cdot x'_{1,2}$.



(c) Step 4: multiply each coordinate with the square root of the corresponding eigenvalue of $\mathbf{C}'$, $x'''_{1,2} = \sqrt{e_{1,2}} \cdot x''_{1,2}$.



(d) Step 5: rotate by the eigenvector matrix $\mathbf{E} = (\underline{e}_1, \ldots, \underline{e}_n)$, $\underline{x}'''' = \mathbf{E} \cdot \underline{x}'''$.



(e) Step 6: translate by center of mass / mean vector $\underline{p}_{\text{mean}}$, $\underline{p} = \underline{x}'''' + \underline{p}_{\text{mean}}$.

Figure 4.7.: Visualizations of the different steps to generate a sample from within an arbitrary ellipsoid (in two dimensions), given its quadratic form (including eigenvalues e_i and -vectors $\underline{e}_i$ and the center of mass $\underline{p}_{\text{mean}}$).

Hit-and-miss selection

To complete the sampling, new points must be selected according to the prior density and the $L > L_0$ likelihood constraint. This so-called rejection sampling is a method to draw samples from a probability distribution $\tilde{\pi}$ by generating points in an enveloping distribution $\mathcal{E}_{L_0}$ and accept them with a probability $\tilde{\pi}/\mathcal{E}_{L_0}$.

The target distribution $\tilde{\pi}$ is of course the prior distribution convolved with the $L > L_0$ likelihood constraint

$$\tilde{\pi} := \pi \otimes \Theta_{L>L_0} \tag{4.6}$$

with

$$\Theta_{L>L_0}(\underline{p}) := \begin{cases} \frac{1}{V_{L>L_0}} & L(\underline{p}) > L_0 \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

and the envelope is the elliptical distribution $\mathcal{E}_{L_0}$ (see eq. 4.4).

After generating a point $\underline{p} \sim \mathcal{E}_{L_0}$, it has to be accepted with a probability of

$$\frac{\tilde{\pi}(\underline{p})}{\mathcal{E}_{L_0}(\underline{p})} = V\tilde{\pi}(\underline{p}) \tag{4.8}$$

which can be achieved by comparing it to a uniform sample $u \sim \mathcal{U}(0,1)$

$$\tilde{\pi}(\underline{p}) > \frac{u}{V} \tag{4.9}$$

$$\Rightarrow \underline{p} \sim \tilde{\pi} \ . \tag{4.10}$$

Special case: uniform prior distribution π In the analyses in this thesis a special case of prior distributions was used to ensure an uninformed approach and simplify the method. This prior consisted of uniform distributions $\mathcal{U}_i$ in each parameter dimension (hyperrectangle)

$$\pi(\underline{p}) = \pi(p_1, p_2, \ldots, p_n) \tag{4.11}$$

$$= \mathcal{U}_1(p_{1,\min}, p_{1,\max}) \otimes \mathcal{U}_2(p_{n,\min}, p_{n,\max}) \otimes \ldots \otimes \mathcal{U}_n(p_{n,\min}, p_{n,\max}) \tag{4.12}$$

with intervals $[p_{i,min}, p_{i,max}]$ chosen for each dimension i depending on the problem at hand.

Thus, to accept a sample $\underline{p} \sim \mathcal{E}_{L_0}$, it was sufficient to check

1. $p_i \in [p_{i,\min}, p_{i,\max}]$ for all n components and
2. $L(\underline{p}) > L_0$.

which reduced the complexity mainly to the evaluation of the likelihood $L(\underline{p})$.

(Dis-)Advantages

The applicability of ellipsoidal sampling depends on the shape of the L_0 hypersurface. Largely elliptical boundaries can be approximated very well and sampled efficiently (see eq. 4.1). Clustered volumes, though, would be enveloped by a single, large ellipsoid. These clusters then occupy only small parts of the ellipsoidal sampling volume. Hence this causes large amounts of rejects, and a lot of computational power is wasted in these scenarios. This was the case with the analysis of μH (see section 5.4), where an ambiguity in the parameterization of the two hyperfine splitting peaks caused a bimodal, but symmetrical posterior distribution.

Additionally, as the collection of live points contract with each iteration, the ellipsoid, *i. e.* the eigenvectors and -values have to be re-determined regularly to keep the efficiency at an acceptable level. This requires a substantial amount of CPU time, so the intervals of re-evaluation t have to be balanced carefully against the increase in time needed to produce a valid sample (as the efficiency decreases with every iteration).

Contrary to the MCMC-exploration, only two configuration parameters are needed, namely the enlargement factor d and the interval of contraction / re-evaluation t. Experience in the analyses showed a mostly stable behavior even for disadvantageous parameter combinations (d, t). The stability was an important fact, because some analyses required evaluation of several thousand models, where determination of a working set of configuration parameters for each model individually was not feasible.

4.2.3. Clustered ellipsoidal sampling

Idea

The great disadvantage of the ellipsoidal sampling method is, that it is unsuitable for distributions with several modes or equi-likelihood hypersurfaces of non-elliptical shapes [37]. The idea of clustered ellipsoidal sampling is to arrange the live points into several clusters, and to approximate each cluster with an ellipsoid. This prevents a lot of rejects, because the ratio of the target volume $V_{L>L_0}$ to a clustered sampling volume $V_1 + V_2$ and in consequence the efficiency is significantly higher than to the volume V_0 of a large enveloping ellipsoid (see Figure 4.8)

$$\begin{aligned} & V_1 + V_2 \ll V_0 \\ \Rightarrow & \frac{V_{L>L_0}}{V_1 + V_2} \gg \frac{V_{L>L_0}}{V_0} \quad . \end{aligned}$$

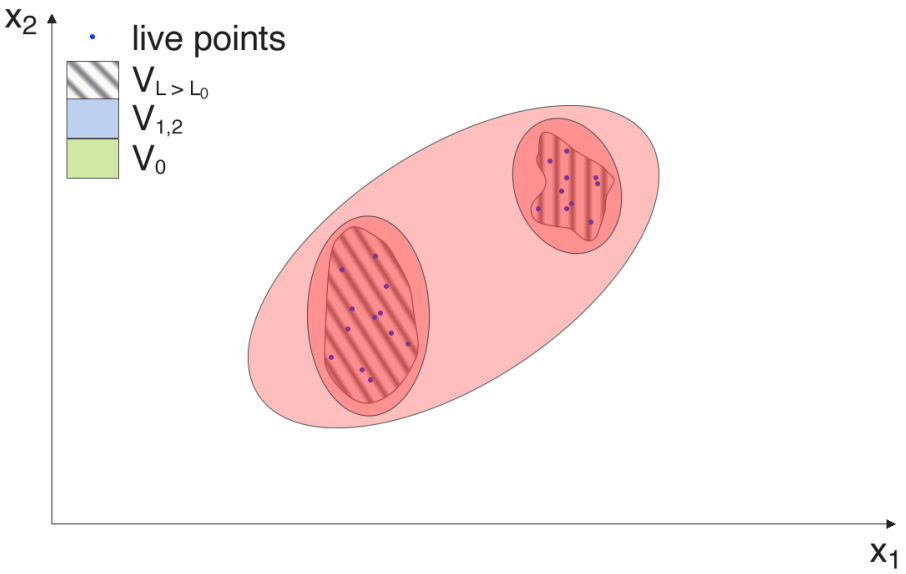


Figure 4.8.: Comparison of sampling volumes with a total enveloping ellipsis or individual ellipses.

An immediate difficulty is to classify the live points into different clusters as well as to decide on how many clusters are necessary. The methods described in section 4.2.2 can then be used to envelop each cluster individually with a separate ellipsoid. When a set of ellipsoids is found, additional challenges arise from sampling from overlapping ellipsoids.

Clustering of live points

The clustering of points in an n dimensional parameter space is a common machine learning problem. To define and quantify a distance relation between two distinct points, a kind of metric is required. For common $\mathbb{R}^n$ vector spaces, an obvious choice would be the Euclidean metric. The Euclidean metric, however, is not suitable to assess the distance between two points in the parameter space, because the scales of the individual dimensions can vary widely. This causes a dominance of large-scale parameter dimensions during the clustering.

Nevertheless, given the precision matrix $\mathbf{C}$ (see section 4.2.2) of all live points, a (very rough) idea of the relative scales and therefore a metric is available, which at least weakens the dominance of large-scale dimensions and provides a more realistic distance between two distinct points.

Based on this metric, well-known algorithms such as k-means-clustering [40] can be used to partition the live points into a set of k clusters. The real difficulty in this step lies defining a number of clusters k.

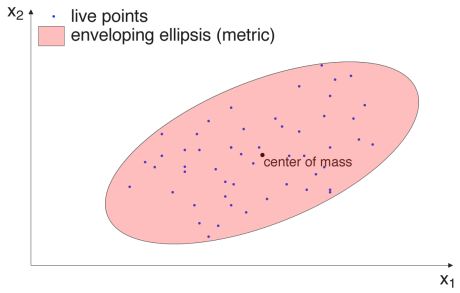
A method developed by J. R. Shaw et al. [41] assumes two clusters recursively until a sufficient partitioning is achieved. Inspired by the discussion of different methods and techniques in a publication by F. Feroz and M. P. Hobson [37], a simplified clustering method was developed.

The basic idea is to start with one cluster and monitor the acceptance rate r of the hit-and-miss selection. If the efficiency drops below a (predefined) fraction f_0 of the expected efficiency r_{exp} (see eq. 4.1), an additional cluster is assumed to be necessary.

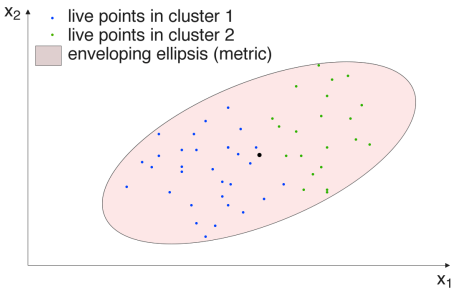
The purpose of this clustering sampler is to approximate the form of the posterior as well as possible and reasonable. With the presumption that the combined ellipsoid volumes envelop the $L > L_0$ posterior volume entirely, and their volume ratio $f = {}^{V_{L>L_0}}/_{V_{\text{tot},k}}$, a quantity f is given to quantify the accuracy of the approximation. The posterior volume $V_{L>L_0}$ is of course not accessible directly, but the ratio ${}^{V_{L>L_0}}/_{V_{\text{tot},k}}$ can be estimated via the acceptance rate (see eq. 4.1)

$$f = \frac{V_{L>L_0}}{V_{\text{tot},k}} = r \cdot d^n \quad .$$

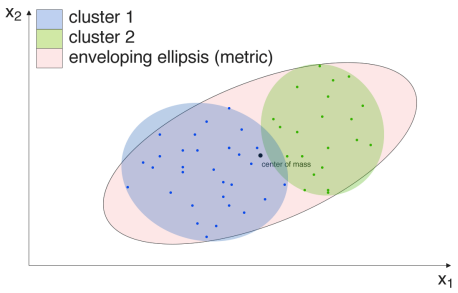
By imposing the condition $f > f_0$ with f_0 being a predefined, minimal fineness of the approximation, a limit is available after which an additional cluster becomes necessary.



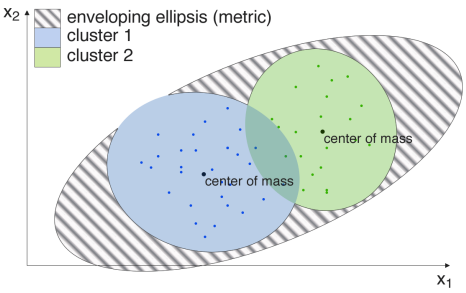
(a) Step 1, the enveloping ellipsis to the collection of points must be determined to provide a metric for the clustering step.



(b) With the given metric, a clustering algorithm (*e. g. k-means*) can be used to partition the points into distinct groups.



(c) Each cluster gets its own elliptic envelope. Their union is a new, better approximation to the total collection of live points then the original ellipsis.



(d) After enlarging the clusters by a constant factor d, the comparison with the total, enlarged ellipsis shows the reduction in sampling volume (hatched area), providing a higher efficiency.

Figure 4.9.: Basic procedure to cluster and approximate the collection of live points using several ellipses.

Sampling uniformly within all clusters

Sampling from an arbitrary number of possibly overlapping ellipsoids with different volumes requires some additional steps to ensure uniformity. Generating a sample from a single ellipsoid was discussed in section 4.2.2, but the question remains, from which of the ellipsoids to sample.

Assuming for now non-overlapping clusters, then for any point $\underline{p}$ within that clusters and a total volume $V_{\text{tot}} = \sum_i V_i$, the density is given by

$$\mathcal{E}_{\text{tot}}(\underline{p}) = \frac{1}{V_{\text{tot}}}$$

The probability to find a point within a certain cluster i is then given by the proportional volume

$$\mathcal{E}_i = \frac{V_i}{V_{\text{tot}}}$$

which of course has to add up to one

$$\Rightarrow \sum_i \mathcal{E}_i = \frac{\sum_i V_i}{V_{\text{tot}}} = 1 \ .$$

By choosing the cluster to sample from randomly with weights $\mathcal{E}_i$, uniformity across all clusters can be guaranteed.

Overlapping clusters destroy the uniformity, because the overlapping volumes are oversampled. Samples from a volume element belonging to n clusters are n times more probable to be sampled than from a volume element of a single cluster. In order to correct this, only every n^{th} sample from that volume element must be accepted. This can be achieved by an additional selection of accepting samples with a probability of $1/n$.

In summary, given a set of clusters, a sample can be generated by the following procedure:

1. Choose a cluster randomly with a probability of V_i/V_{tot} for each cluster i.
2. Sample from this cluster using ellipsoidal sampling (see section 4.2.2).
3. Determine the number of clusters n which contain this sample.
4. Accept the sample with a probability of $1/n$ or return to step 1.
5. Apply prior and likelihood hit-and-miss selection (see section 4.2.2), return to step 1 on rejection.

(Dis-)Advantages

As visualized in Figure 4.9d, a better approximation of the live point distribution by using several ellipses, can reduce the number of rejections significantly. For multimodal

posterior distributions, this effect is even greater. There is, however, a considerable computational overhead in the clustering step (because, instead of only one, k ellipses have to be determined, in addition to the clustering itself), and the sampling (the number n of ellipsoid volumes a sample lies in has to be determined. Depending on the amount of overlaps, a lot of samples have to be rejected).

The additional workload has to be balanced against the advantage of a better approximation. In the case of multimodal posterior distributions, clustered sampling is mandatory, because a single ellipsoid always envelops all modes and thus contains a huge amount of prior space volume between these modes where the likelihood does not fulfill the constraint $L > L_0$. A single ellipsoid could not contract further, and due to the nature of nested sampling, the efficiency would decrease to zero with each iteration.

In order to achieve good performance, both the interval of re-clustering has to be configured along with the efficiency fraction at which a new cluster is assumed, and the maximum allowed number of clusters.

4.3. Likelihood function

A crucial point in Bayesian data analysis is the specification of a likelihood function to relate the experimental data to a hypothesis. The likelihood for a single data point D_k, $L_k(D_k \,|\, \underline{p}, H)$, can usually be partitioned into two functions

$$L_k(D_k \,|\, \underline{p}, H) = prob\big(D_k \,|\, \mu_H(k, \underline{p})\big) \tag{4.13}$$

with $\mu_H(k, \underline{p})$ being the value expected, given the parameter vector $\underline{p}$. It is thus necessary to find an appropriate distribution $prob\big(D \,|\, \mu\big)$, as well as deriving expected values μ_H from a hypothesis H and parameters $\underline{p}$.

4.3.1. Statistical properties of the data

The required distribution $prob\big(D \,|\, \mu\big)$ depends on the nature of the experiment and the collected data. The analyses described later (see chapter 5) are based on X-ray spectra. These spectra were effectively gathered by binning single photon counts in a position-space histogram. Thus, the distribution in each channel was assumed to follow a Poisson distribution.

$$Pois(D_k \,|\, \mu_H(k, \underline{p})) = \frac{\mu_H(k, \underline{p})^{D_k} e^{-\mu_H(k, \underline{p})}}{D_k!} \tag{4.14}$$

where the expected count $\mu_H(k, \underline{p})$ in a channel k must be expressed using the hypothesis H.

Given that, a likelihood function for each channel k can be defined by

$$L_k(D_k \,|\, H, \underline{p}) = Pois(D_k \,|\, \mu_H(k, \underline{p})) \tag{4.15}$$

and, due to the statistical independence of the data points D_k, combined to a total likelihood by multiplication

$$L(D_1, D_2, \ldots, D_m \,|\, H, \underline{p}) = \prod_k L_k(D_k \,|\, H, \underline{p}) \tag{4.16}$$

$$= \prod_k \frac{\mu_H(k, \underline{p})^{D_k} e^{-\mu_H(k, \underline{p})}}{D_k!} \tag{4.17}$$

Working with large products

In floating point arithmetics, real numbers are internally represented similar to the scientific notation $234.5 = 2.345 \cdot 10^2$ (but to a base of 2 instead of 10). On multiplication, the exponents are added and the mantissa are multiplied. For *double precision floating point numbers* the exponent is represented by eleven bits [42], making its minimal value -1023 and its maximum $+1023$. Even with fairly small numbers, *e. g.* $a = 2^{10} = 1024$, after several multiplications $\prod_{k=0}^{1000} a = 2^{10000}$, these boundaries are pushed. Loss of precision

already occurs with much lower values.

Therefore, in these numerical computations, multiplications are usually replaced by summations of the logarithms. Applying this principle on eq. 4.17 yields [30]

$$\tilde{L} := \ln(L) \tag{4.18}$$

$$= \sum_k \ln\left(\frac{\mu_H(k,\underline{p})^{D_k} e^{-\mu_H(k,\underline{p})}}{D_k!}\right) \tag{4.19}$$

$$= \sum_k \left[D_k \cdot \ln(\mu_H(k,\underline{p})) - \mu_H(k,\underline{p})\right] \tag{4.20}$$

4.4. General procedure

The general procedure of the Bayesian analyses was very similar for all of the experiments. Due to the nature of the line shapes (see chapter 2) and their complicated representation, a few preparational steps were necessary before the actual analysis.

The first and foremost step was to decide on a set of models to test. For most of the analyses, the differences in models manifested themselves in varying kinetic energy distributions. Based on that, a reasonable parameterization of these models had to be established.

4.4.1. Parameterization

As discussed in section 2.2, a simplification was made by approximating complex kinetic energy spectra with a few components using a superposition of rectangular-functions. This superposition can be written as

$$\begin{aligned} S &= R \otimes \left(\sum_i D_i \right) \otimes L/G + I_{\mathrm{bg}} \\ &= \sum_i \underbrace{(R \otimes D_i \otimes L/G)}_{S_i} + I_{\mathrm{bg}} \\ &= \sum_i S_i + I_{\mathrm{bg}} \end{aligned}$$

with an additional constant background of I_{bg}. In order to parameterize the line shape S properly, a total intensity I as well as relative intensities I_i have been extracted leaving superposed line shape components $\tilde{S}_i$

$$S = S(I, I_1, \ldots, I_m) = I \left[(1 - \sum_{i=1}^{m} I_i)\tilde{S}_0 + \sum_{i=1}^{m} I_i \tilde{S}_i \right] + I_{\mathrm{bg}} \tag{4.21}$$

Additional parameters arose of course from the Lorentzian component (width Γ), the Gaussian broadening (standard deviation σ), and the position of the peak (defined at the center of mass x_0)

$$\begin{aligned} \tilde{S}_i &= \tilde{S}_i(x_0, \Gamma/\sigma) \\ \Rightarrow S &= S(x_0, \Gamma/\sigma, I, I_1, \ldots, I_m, I_{bg}) \end{aligned}$$

These line shape components $\tilde{S}_i$ are not analytically expressable and must therefore be rasterized and pre-calculated before the analysis. Thus, in order to evaluate parameter values in-between the pre-calculated line shapes, interpolations have to be applied.

The scheme of the pre-calculations is illustrated in Figure 4.10. For each energy component i, a trapezoidal distribution (see section 2.2) had to be generated and convolved with the spectrometer response function (see section 2.5) as well as Lorentzian or Gaussian curves of different widths Γ_j resp. σ_j to obtain a matrix of test spectra $\tilde{S}_{i,j}$.

4.4.2. Modeling the kinetic energy distribution

As discussed in section 2.2, the complex kinetic energy distribution was modeled using a few narrow energy intervals. To get an estimate of the complexity of the kinetic energy distribution, *i. e.* how many different components are necessary to describe an X-ray spectrum, an iterative approach proved to be useful and was generally used throughout the analyses:

Single low-energy component of unkown width was tested.

1^{st} **high-energy contribution** in addition to the low-energy component. The correlation between the width of the low-energy component and the position of the high-energy contribution was studied.

2^{nd} **high-energy contribution** was considered, and also its correlation to the position of the 1^{st} high-energy component.

Model-averaging and parameter inference was done after establishing one of the above model sets as the most probable.

4.5. Interpolation of test spectra

Due to the complexity of the line shapes analyzed in this work, largely stemming from the response function, no analytical description of the line shape is possible. In order to overcome this, measured data spectra were instead compared with large sets of test spectra, which have been pre-calculated at discrete parameter values.

Using these test spectra, some parameter values in-between the discrete points have to be interpolated. As described in section 4.4.1, parameterization has been chosen in a way, that the total line shape S is a superposition of individual line $\tilde{S}_i(x_0, \Gamma/\sigma)$ and a constant background I_{bg} (see eq. 4.21).

Subsequently, these individual line shapes $\tilde{S}_i$ were pre-calculated at discrete x-values for a fixed center of mass $x_0 = 0$ and several Γ_j or σ_j. The evaluation of other discrete center of mass values x_0' was then just a matter of shifting the test spectrum along the x-axis. To apply center of mass values between the discrete samples, the test spectrum was shifted to the nearest discrete value, and an additional cubic spline interpolation was performed. This required the selection of a test spectrum an appropriate Γ/σ, which itself had to be interpolated. In order to achieve this, the following procedure was used [27]:

1. Given the parameter values (x_0, Γ), choose three test spectra $\tilde{S}_{i,j}$ with $\Gamma_j = \lfloor\Gamma\rfloor$, $\Gamma_j = \lfloor\Gamma\rfloor + 1$, and $\Gamma_j = \lfloor\Gamma\rfloor + 2$.

2. Interpolate the center of mass x in each test spectra as described above, yielding three y-values y_0, y_1, y_2.

3. Correct the difference $\Gamma - \lfloor\Gamma\rfloor$ by parabolic interpolation between y_0, y_1, and y_2.

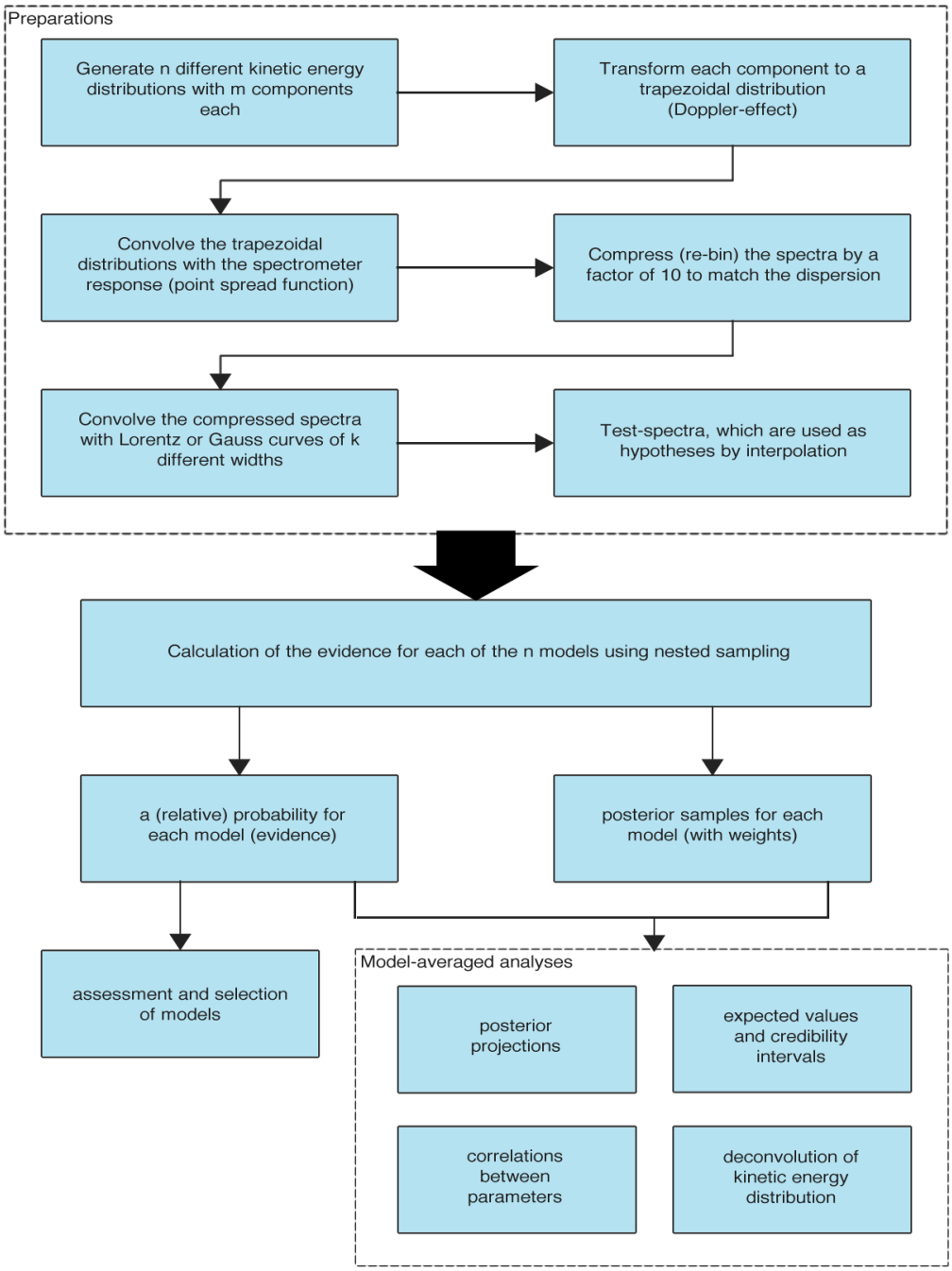


Figure 4.10.: Scheme of the general procedure, common for all analyses. In principle, it consists of three parts: 1. preparation of the test spectra and models, 2. numerical evaluation of the evidences of each model by means of nested sampling, 3. visualization and analyses of the results.

4.6. Post processing

As mentioned above, the method of nested sampling primarily calculates a probability weight for each tested model, but also produces large amounts of posterior samples (with attributed weights), which mimic the posterior distribution of the model parameters.

This gives rise to the opportunity to extract more information about a parameter than an averaged value and standard deviation (conventionally estimated by sample variance). By projecting the posterior samples in one dimension i and subsequent binning, an approximation of the distribution of this parameter λ_i can be achieved. This corresponds to a marginalization over the remaining parameters $\lambda_{j\neq i}$. Using this histogram, an estimate for the most probable value is available via the maximum.

4.6.1. Credibility intervals

The disparity between an actual posterior and a Gaussian distribution renders the quantification of error margins by means of standard deviation ineffective. The interesting point about the standard deviation, or in extend 2σ, 3σ intervals, is the containment of $\approx 68.3\%$, 95.4%, 99.7% of probability volume within $1\sigma,\ 2\sigma,\ 3\sigma$ intervals respectively. By inversion of this relationship, thus calculating the interval which contains *e. g.* 68.3% of the total volume, similar error margins are accessible, called (1σ) credibility interval(s).

4.6.2. Deconvolution

The numerical deconvolution used in this work is a technique developed by L. Simons [27] and is based on bin-wise model-averaging of the kinetic energy distribution. The goal is to determine a kinetic energy spectrum from a line shape with as little assumptions as possible.

The idea is to divide the important part (0 eV to $T_{\max}$) of the kinetic energy scale into n components B_i of width $T_{\max}/n$. Using m pairwise different components out of n, a total of N models are generated

$$N = \frac{n!}{m! \cdot (n-m)!} \ .$$

After the evaluation by means of nested sampling, each of the N models has an attributed probability weight w_k (given by the Bayes factor), and relative intensities for each component (given by the inferred parameters $I_{k,i}$). A model-averaged intensity I_i for the i^{th} channel is then available via[3]

$$I_i = \frac{\sum_k I_{k,i} \cdot w_k}{\sum_k w_k} \ .$$

[3] I_i must not be confused with the circumstantial information I of section 3.2.

5. Analysis

5.1. Simulation of pionic hydrogen (πH) spectra

In order to assess the sensitivity of the analysis methods, simulations of πH transitions were carried out with a few simplified kinetic energy distributions, each consisting of two narrow intervals - one low-energy component at $0-2\,\mathrm{eV}(L)$ and one high-energy contribution at $72-74\,\mathrm{eV}(H)$. All simulations used the same parameter values (specified in Table 5.1), but three different relative intensities, namely $L/H = 50/50, 70/30$ and $85/15$.

The analyses of the simulated spectra helped to understand the limitations of the method itself and the correlation of the precision of the analysis with the statistical quality of the data.

1. How well can the low- and high-energy contributions be distinguished *i. e.* can positions, widths and intensities be reproduced and to which accuracy?

2. To which extend is the hadronic width reproduced and in particular by the correlation of the intensity of the high-energy components and the hadronic width?

parameter	value
peak position	channel 600
hadronic width	900 eV
line intensity	43190 counts
background	8.3 counts / channel
low-energy box	$0-2\,\mathrm{eV}$
high-energy box	$72-74\,\mathrm{eV}$

Table 5.1.: Parameter values used in all of the simulations. The parameters are adapted to a measurement of the πH(2p-1s) transition.

5.1.1. Model-independent approach

To minimize the assumptions, a model-independent approach was chosen, especially concerning the structure of the kinetic energy spectrum. Therefore 8436 different models, consisting of three differently positioned Doppler components, of common width of 4 eV each, were generated and evaluated. This yielded an evidence along with posterior distributions — and along with these, parameter estimates — for each one of the models.

5.1.2. Most probable model

Although the Bayesian methodology provides — in theory — a precise way to assess and weigh different models against each other by comparing each ones evidence, in practice, due to the statistical nature of the data and the nested sampling algorithm itself, the evidence disperses. Thus identification of a single, best model is not possible and model-selection requires a more fine-grained approach. This difficulty can be demonstrated by showing the three most probable models for each simulation in Figures 5.1, 5.2, 5.3.

It is obvious, that the configuration of the high-energy component disperses, although the Bayes factor $\ln(B)$ does not indicate substantial evidence towards the top-most model.

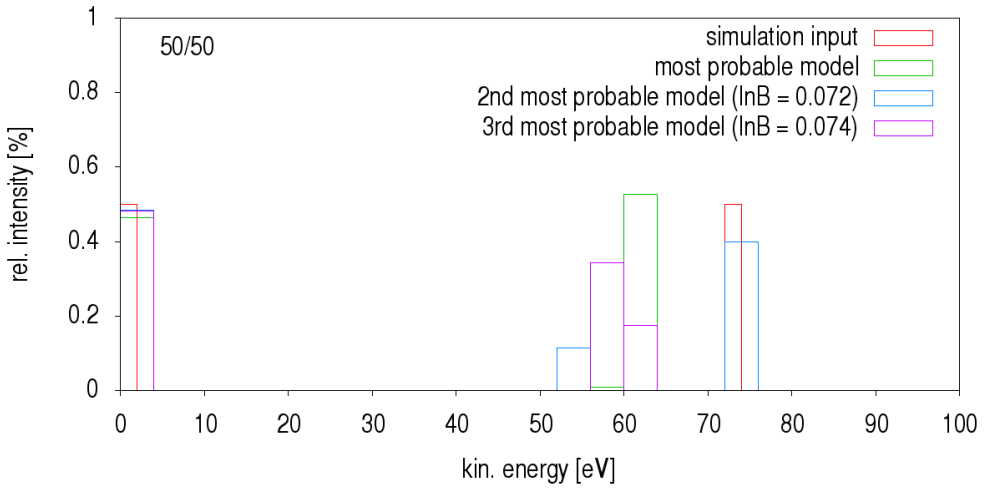


Figure 5.1.: Comparison of the input spectrum of the πH 50/50 simulation with the three most probable spectra / models. These models do not differ much in evidence, but already show a broad range of probable configurations of the high-energy component. The low-energy component is consistent along all models. The simulation input spectrum has a bin width of 2 eV, whereas the model spectra are binned in 4 eV intervals.

Hence a simple model-selection if not feasible. Even for the 50/50 model, a total of 469 (out of 8436) models showed a Bayes factor $\ln(B) < 2$, compared to the highest evidence. In order to work out the structure of the kinetic energy distribution, a procedure to model-average this structure is necessary, such as the deconvolution technique described in section 4.6.2.

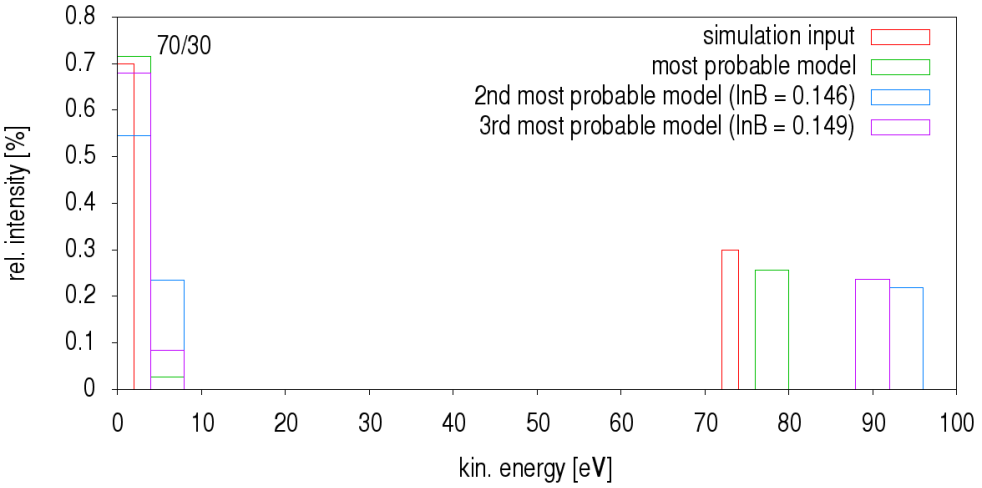


Figure 5.2.: Comparison of the input of the πH 70/30 simulation with the three most probable spectra / models. The larger spread of these models already indicate a decrease in sensitivity.

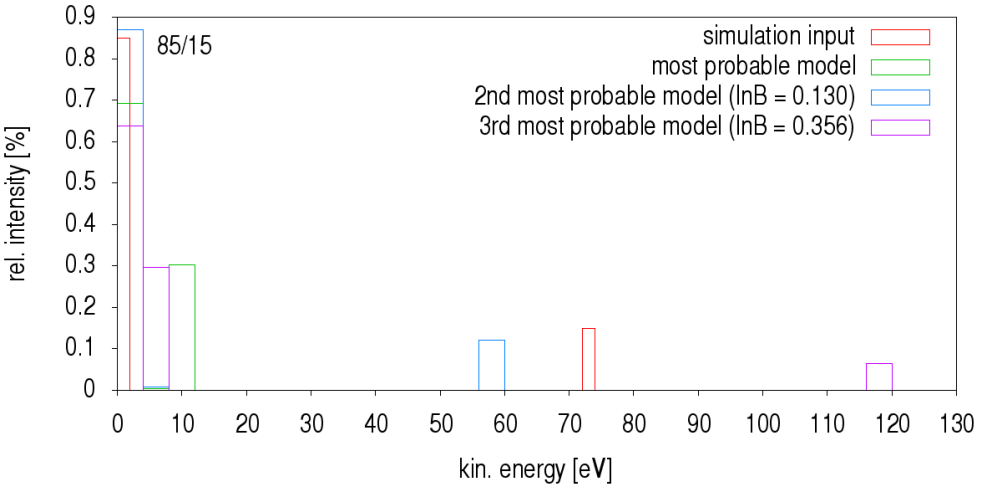


Figure 5.3.: Comparison of the input of the πH 85/15 simulation with the three most probable spectra / models. With only 15% in the high-energy component, the sensitivity is strongly limited by the statistical nature of the data.

5.1.3. Model-averaging

Despite the fact, that even the 50/50 simulation evidences show no sharp peak favoring a single model, taking advantage of all the information contained in all models and posterior distributions can produce valueable results.

Deconvolution of the Doppler spectrum

The model-averaging of the different kinetic energy spectra yields a distribution which approximates the input spectrum, but disperses significantly. This dispersion reflects the uncertainty and limited information contained in the data.

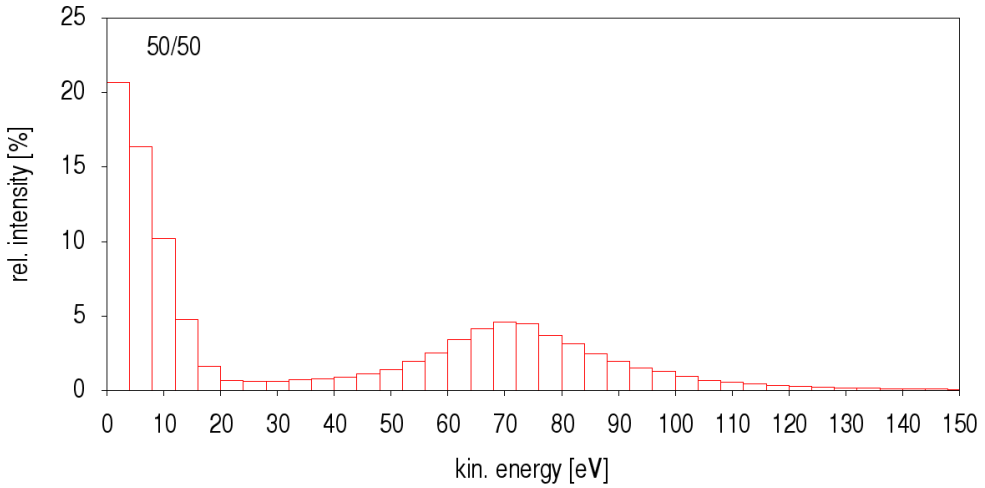


Figure 5.4.: Model-averaged spectrum of the 50/50 πH simulation. The high-energy peak is centered around the expected position of 72 eV, but is significantly broadened.

These examples provide a guidance to the analysis of the πH spectra, concerning the inherent limitations. It is important to state the fact, that these limitations are founded in the limited information content of the data and cannot be substantially improved by more precision (*e. g.* by using more than the 8436 models).

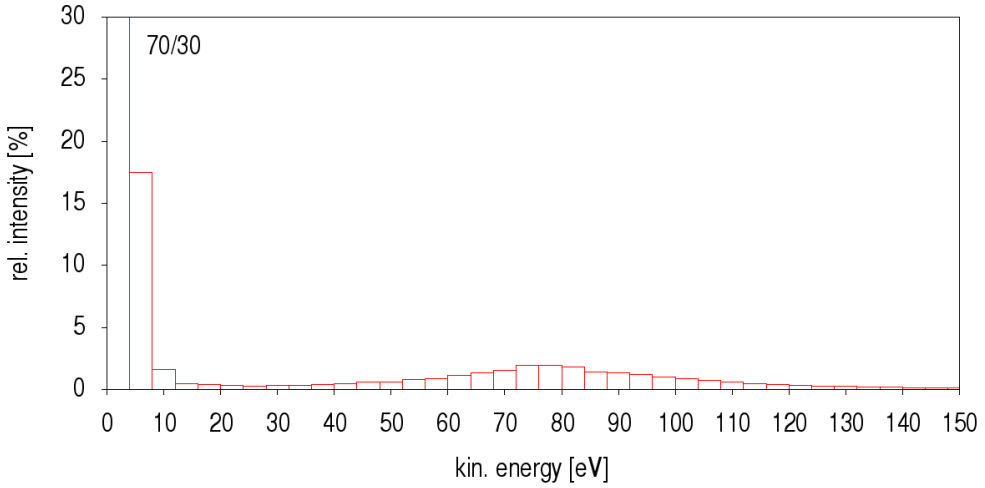


Figure 5.5.: Model-averaged spectrum of the 70/30 πH simulation. Compared to the 50/50 simulation, the dispersion of the high-energy component is even wider with its maximum between 70 eV and 80 eV. The height of the first channel (0 eV to 4 eV) is 55.4%.

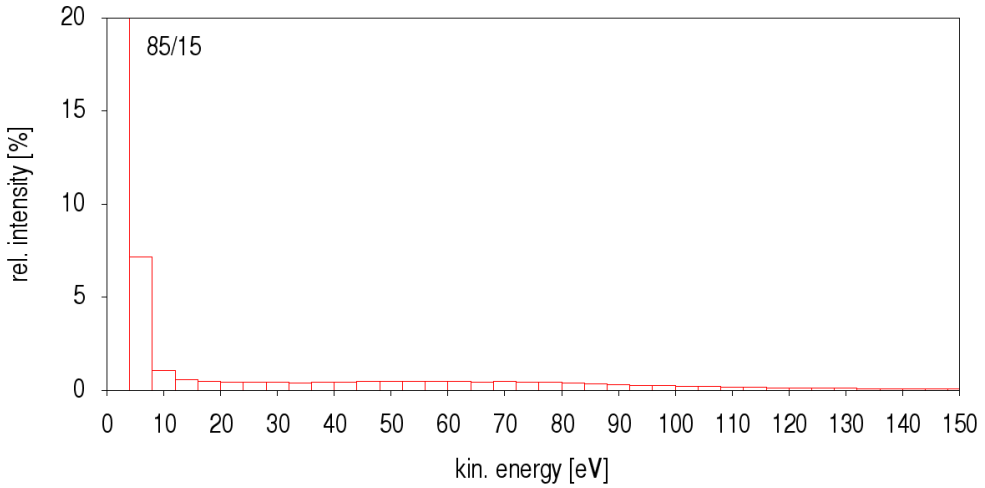


Figure 5.6.: Model-averaged spectrum of the 85/15 πH simulation. For such a small high-energy component no structure or position can be reproduced. The 0 eV to 4 eV channel has a height of 80.5%.

Hadronic width

Another important quantity to be studied in pionic hydrogen is the hadronic width Γ_{1s} of the ground state. Figures 5.7, 5.8 and 5.9 show model-averaged posterior distributions, resp. their projections to the parameter Γ_{1s}. These posterior distributions incorporate all models with a Bayes factor (relative to the maximal evidence) of $\ln(B) > -5$.

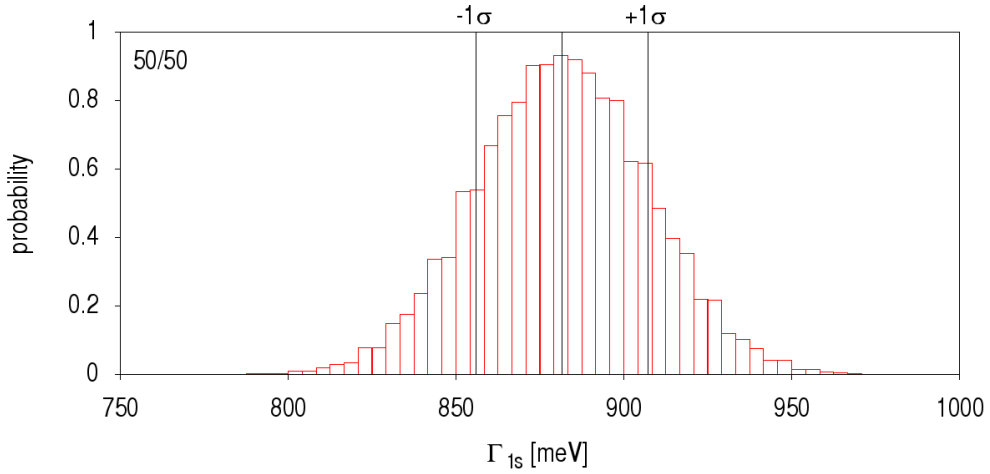


Figure 5.7.: Model-averaged posterior distribution for the hadronic width Γ_{1s} of the simulated 50/50 πH spectrum. This histogram is built from 2016 models with a Bayes factor $\ln(B) > -5$. The most probable value is 882 meV with a credibility (1σ) interval from -26 meV to $+26$ meV.

These inferred Γ_{1s} values are largely consistent with the input parameters of the simulations. It was expected that at least for the 85/15 simulation, the uncertainty increases as the high-energy component could not be reproduced and all width-contributing factors correlate. Contrary to that, apart from a little shift of the most probable value (well within the 1σ interval), the 1σ credibility intervals coincide.

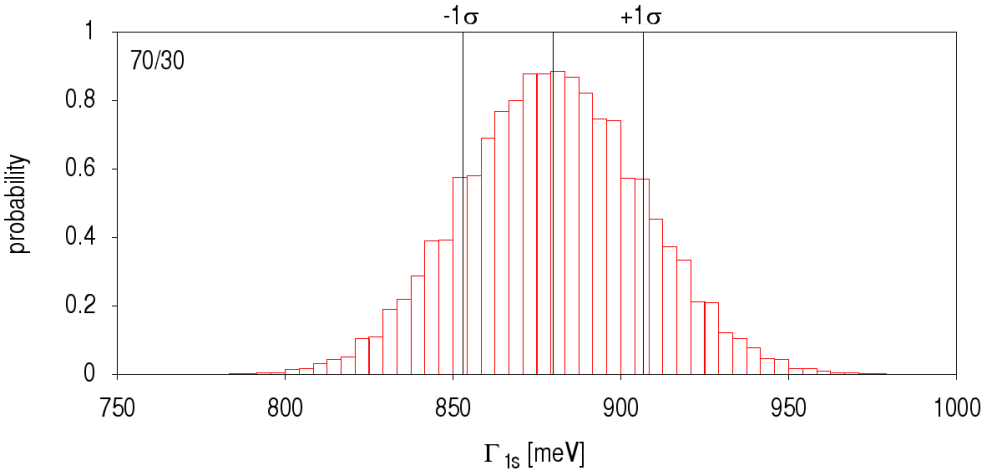


Figure 5.8.: Model-averaged posterior distribution for the hadronic width Γ_{1s} of the simulated 70/30 πH spectrum. 1203 models with a Bayes factor $\ln(B) > -5$ were averaged to create this histogram. The most probable value is 880 meV with a credibility (1σ) interval from -27 meV to $+27$ meV.

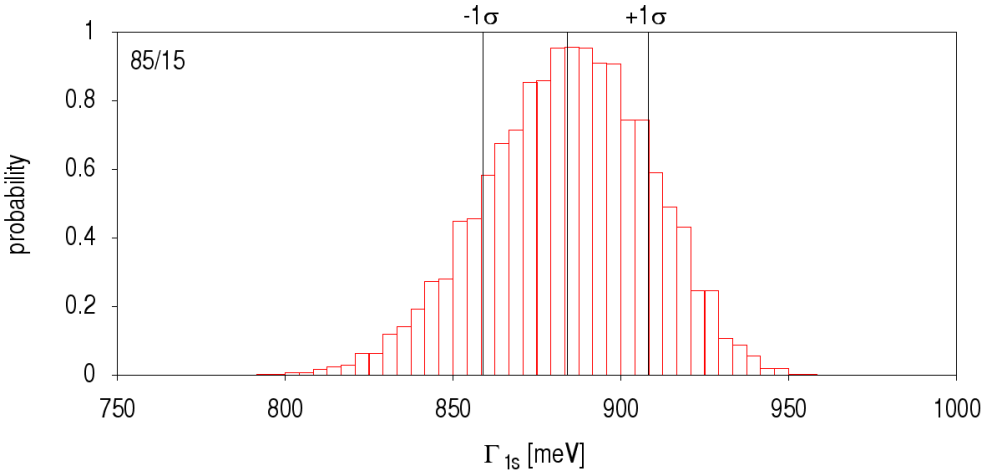


Figure 5.9.: Model-averaged posterior distribution for the hadronic width Γ_{1s} of the simulated 85/15 πH spectrum. It is based on 1095 models with a Bayes factor $\ln(B) > -5$. The most probable value is 884 meV with a credibility (1σ) interval from -25 meV to $+24$ meV.

5.1.4. Position, intensity and background

As for the hadronic width Γ_{1s}, the model-average for the remaining parameters of each simulation have been determined, inluding their 1σ credibility interval, and can be found in Table 5.2. As expected the peak position can be reproduced very accurately. The results are overall consistent with each other and the input parameters, regarding the most probable values and credibility intervals. As difficult as comparing just a few models was (see section 5.1.2), a model-independent approach by evaluating a large mount of different configurations provided a powerful way to extract a maximum of information about the model parameters.

simulation	parameter	value	credibility	
50/50	peak position [channels]	599.669	−0.161	+0.161
	hadronic width [meV]	882	−26	+25
	line intensity [counts]	43663	−278	+278
	bg. [counts / channel]	8.054	−0.280	+0.280
70/30	peak position [channels]	599.658	−0.147	+0.147
	hadronic width [meV]	880	−27	+27
	line intensity [counts]	43647	−281	+281
	bg. [counts / channel]	8.077	−0.284	0.286
85/15	peak position [channels]	599.691	−0.138	+0.138
	hadronic width [meV]	884	−25	+24
	line intensity [counts]	43657	−274	−273
	bg. [counts / channel]	8.030	−0.270	+0.275

Table 5.2.: Model-averaged parameter estimates for the three πH simulations.

5.2. Pionic hydrogen πH(2p-1s)

After the assessment of the sensitivity and limitations of the analysis methods, a first application was to study the 2p-1s transition in pionic hydrogen (πH). The focus was the determination of the kinetic energy distribution in order to achieve a reliable estimate for the hadronic width of the ground state. To accomplish this, a step by step analysis was performed, beginning with a simple, assumed kinetic energy spectrum of one Doppler contribution starting at 0 eV.

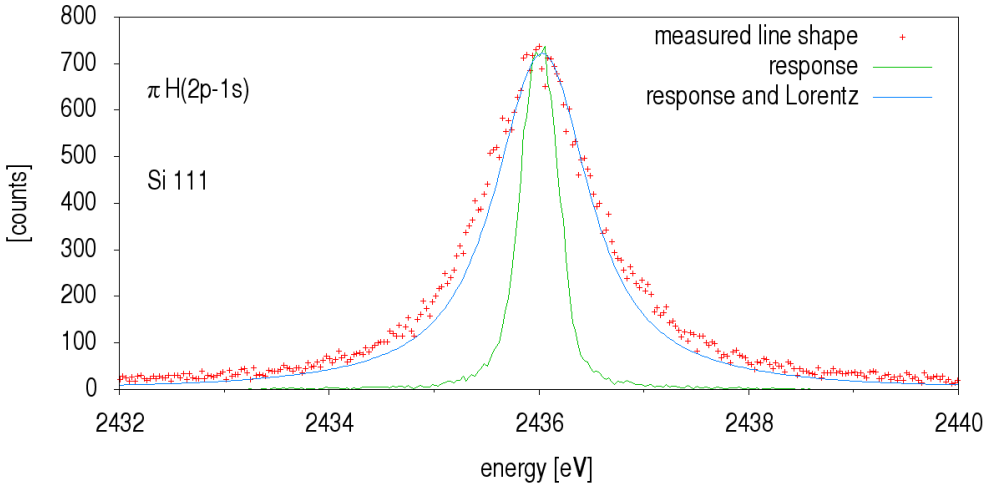


Figure 5.10.: Line shape of the measured πH(2p-1s) transition compared to the spectrometer response and a line shape composed of the response and a Lorentzian width of $\Gamma_{1s} = 823\,\text{meV}$ [2] without any Doppler broadening.

5.2.1. Low-energy component

The simplest possible spectrum consists of a single Doppler contribution at 0 eV with an unknown width. To determine this width, 120 1-component models with widths between 0.25 eV and 30 eV were generated and evaluated (see Figure 5.11).

The results favor a width of about $\approx 12\,\text{eV}$, but with a rather large uncertainty as the evidences between $\gtrsim 6\,\text{eV}$ to $\lesssim 15\,\text{eV}$ provide only weak Bayes factors. The position of the maximum around 12 eV could indicate high-energy components, which would cause an overestimation of the width of a single component model.

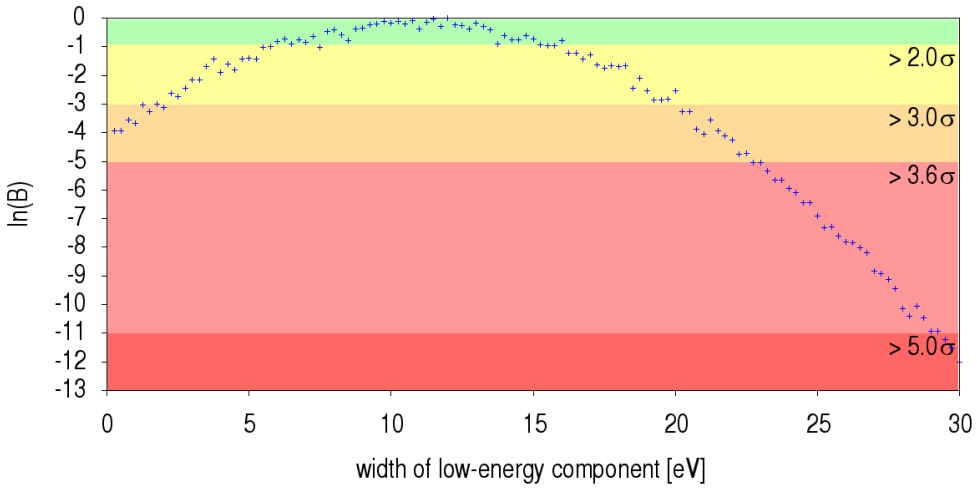


Figure 5.11.: Probabilities (logarithmic) of 120 different models with a single Doppler contribution starting at 0 eV and variable widths. The low evidences for narrow energy intervals ($\lesssim 6\,\mathrm{eV}$) may indicate the presence of high-energy components, which could cause an overestimation of the width. The maximum (reference) value has a (logarithmic) evidence of $e_{\mathrm{max,1\text{-}comp.}} = 211549.07$.

5.2.2. High-energy components

To study the possibility of high-energy contributions and to determine the width of the low-energy component along the way, a set of 17969 models with two components each and different configurations have been generated. These configurations included the same variation of the width of the low-energy component (fixed at 0 eV) along with different positions of a high-energy contribution (of fixed width 2 eV). The evaluated evidences are displayed as scatterplot in Figure 5.12.

To counteract statistical noise and smoothen the distribution, this scatterplot was convolved with a two-dimensional Gaussian distribution with $\sigma_x = 0.375\,\mathrm{eV}, \sigma_y = 1.5\,\mathrm{eV}$. This allowed to determine an approximate maximum at $(1.25\,\mathrm{eV}, 69\,\mathrm{eV})$ as well as contour lines for the Bayes factors (see Figure 5.13). These contour lines reveal a large area of possible 2-component configurations with weak Bayes factors (against the maximum). Nevertheless, the distinctness of the maximum at a position $> 0\,\mathrm{eV}$ suggests, that there is at least one significant high-energy component in the kinetic energy spectrum. This is supported by the Bayes factor $\ln(B) = e_{\mathrm{max,\ 2\text{-}comp.}} - e_{\mathrm{max,\ 1\text{-}comp.}} = 2.86$ between the maximum 2-component and the maximum 1-component model.

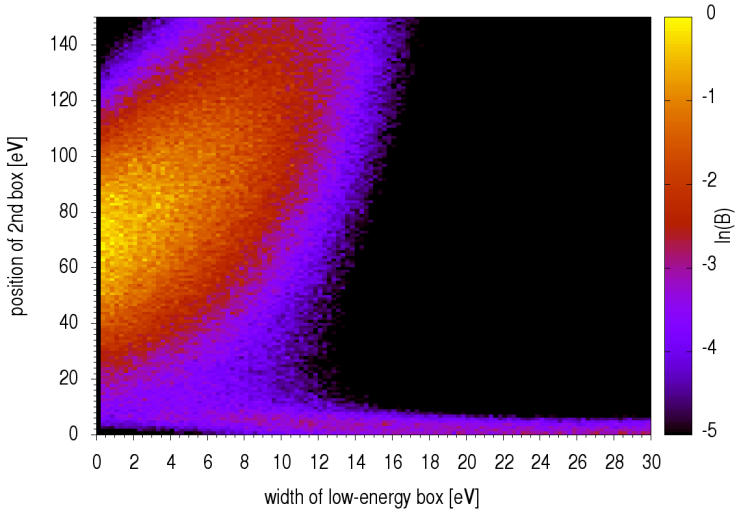


Figure 5.12.: Scatterplot of Bayes factors $\ln(B)$ for different configurations of 2-component models. The position of the low-energy contribution is fixed at 0 eV as well as the width of the high-energy component (2 eV). There is a clear maximum for a narrow ($\lesssim 2$ eV) low-energy contribution with another component around ≈ 70 eV. The maximum (reference) value has a (logarithmic) evidence of $e_{\text{max, 2-comp.}} = 211551.93$.

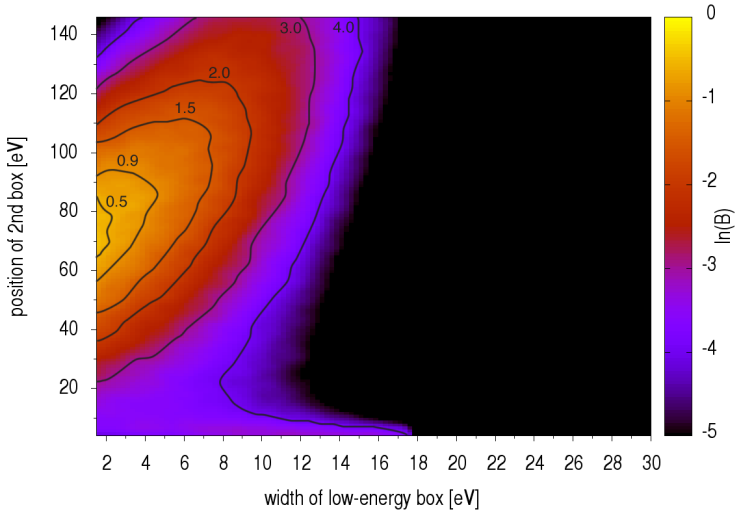


Figure 5.13.: Bayes factors after smoothing with a two-dimensional Gaussian and extraction of contours. A clear maximum ($\ln(B) = 0$) is found at around (1.25 eV, 69 eV) but the area of configurations with a weak (see Table 3.1) Bayes factor to the maximum is rather large.

5.2.3. Search for a 3^{rd} Doppler contribution

The confirmation of high-energy components in the πH spectrum begs the question, if there is a resolvable structure in form of 2 distinguishable high-energy Doppler components. In order to test this hypothesis, a set of 3-component models was generated, each with 2 distinct high-energy contributions at different positions and one fixed low-energy component at 0 eV with a width of 1 eV.

The scatterplot of the evaluated Bayes factors reveals a maximum for a 2^{nd} contribution at $\approx$ 0 eV and a 3^{rd} one at $\approx$ 70 eV, thus an effective 2-component model. This falsification of a 3-component model is supported by the Bayes factor between the best 3-component and best 2-component model $e_{\text{max, 3-comp.}} - e_{\text{max, 2-comp.}} < 300$. This huge difference demonstrates the effect of Occam's razor in Bayesian model-selection: although the favored 3-component model is in principal identical to the 2-component model, the presence of an additional, redundant parameter causes a massive drop in the evidence.

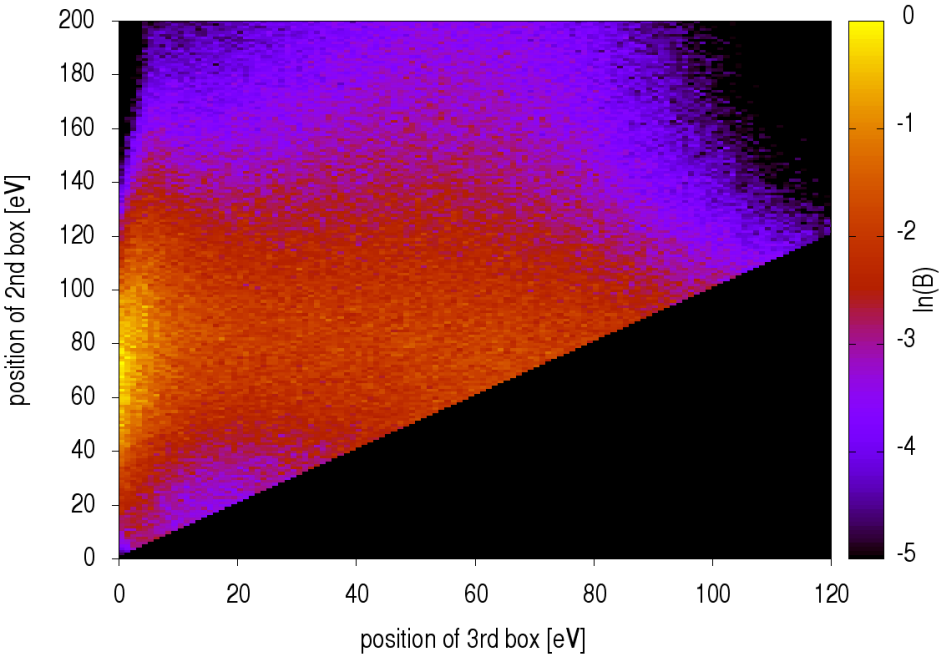


Figure 5.14.: Scatterplot of evidences for 3-component models with the first contribution fixed at 0 eV. There is an obvious maximum for configurations with the 3^{rd} component around 0 eV (effective 2-component models). $e_{\text{max, 3-comp.}} = 211240.86$

model	max. evidence	ln(B)	B
2-comp.	211551.93		
1-comp.	211549.07	2.86	17.5
3-comp.	211240.86	311.07	$> 10^{135}$

Table 5.3.: Comparison of 1, 2 and 3 components modeling the kinetic energy spectrum of the πH(2p-1s) X-ray line. Highest evidence is found for a 2-component model at $0-2\,\mathrm{eV}$ and $\approx 70\,\mathrm{eV}$. The 3-component models are strongly suppressed by their Bayes factors, which demonstrates the effect of Occam's razor in Bayesian model-selection making a 3$^{\mathrm{rd}}$ component superfluous.

5.2.4. Hadronic width

Analogous to the analysis of the simulated πH spectra, a model-averaged posterior distribution for the hadronic width Γ_{1s} is produced by a weighted combination of the posterior distributions of each model, where the weights are the Bayes factors. This allowed the determination of a most probable value for the hadronic width Γ_{1s} as well as credibility intervals.

As there was no evidence found for a 2$^{\mathrm{nd}}$ high-energy component, the posterior was generated using the samples of the 2-component models, more precisely all 9098 models with a Bayes factor $\ln(B) > -5$. The histogram is shown in Figure 5.15 along with the maximum likelihood value and the 1σ credibility interval.

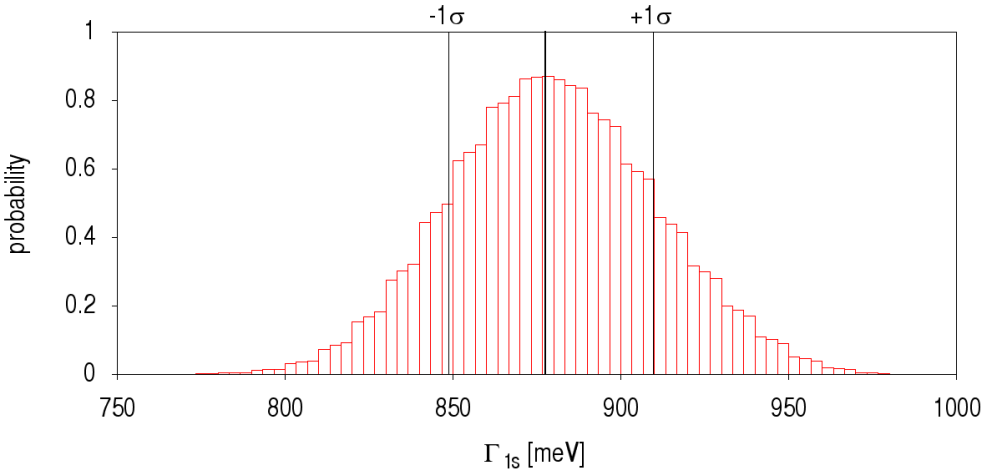


Figure 5.15.: Posterior histogram for the hadronic width Γ_{1s}, based on the samples of 9098 averaged 2-component models with a Bayes factor $\ln(B) > -5$.

parameter	value	credibility					
		1σ		2σ		3σ	
Γ_{1s} [meV]	878	-29	$+32$	-57	$+63$	-86	$+89$
peak pos. [channels]	888.005	-0.141	$+0.141$	-0.281	$+0.282$	-0.422	$+0.423$
peak int. [counts]	43942	-301	$+307$	-599	$+614$	-896	$+918$
bg. [counts / channel]	9.617	-0.349	$+0.345$	-0.694	$+0.686$	-1.026	$+1.027$

Table 5.4.: Estimate for the hadronic width Γ_{1s} with different credibility intervals and the remaining parameters as well.

5.2.5. Relative intensity of the low-energy component

The relative intensity of the low-energy component was inferred similarly to the hadronic width by determining the most probable value via the posterior projection (Figure 5.16).

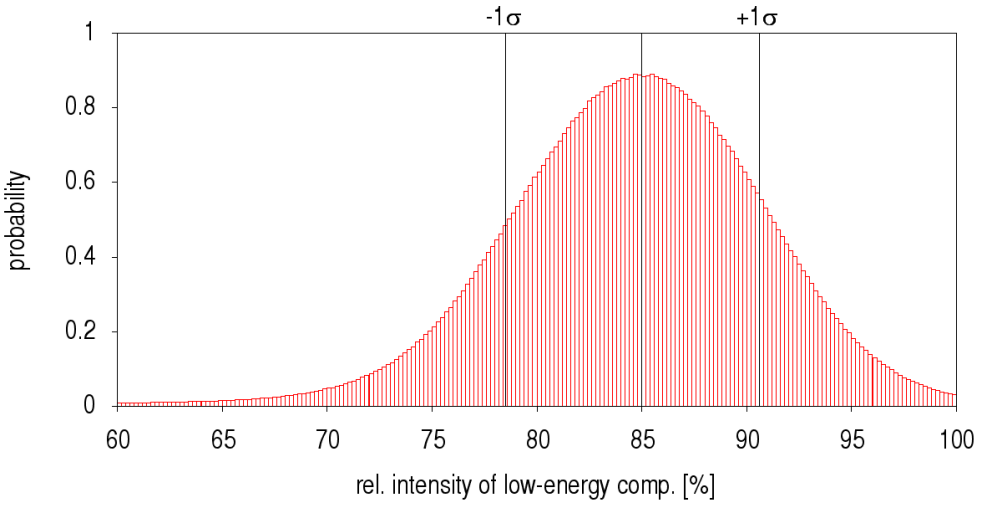


Figure 5.16.: Posterior histogram for the relative intensity of the low-energy component, based on the samples of 9098 averaged 2-component models with a Bayes factor $\ln(B) > -5$.

parameter	value	credibility					
		1σ		2σ		3σ	
rel. int. [%]	85.0	−6.5	+5.6	−21.8	+11.9	−77.9	+14.5

Table 5.5.: Estimate of the relative intensity of the low-energy component with different credibility intervals.

5.2.6. Conclusions

The comparison with the previous results, obtained by frequentistic methods of data analysis (see section 3.1), shows a general consistency [1, 2].
The value of the hadronic width of the ground state of $\Gamma_{1s} = 878\,^{+32}_{-29}$ eV includes the previous result of $\Gamma_{1s} = 902\,^{+25}_{-25}$ meV [43] within its 1σ interval.

Furthermore, the difficulty to distinguish a high-energy structure in the kinetic energy distribution could be resolved, and the assessment of a major low-energy component $85.0\,^{+5.6}_{-6.5}$ % holds (compared to an estimation of $85\,^{+5}_{-5}$ % in [43]). Based on the Bayes factors (Table 5.3, Figure 5.14), it is conclusive, that for the present data, only one high-energy component is measurable.

5.3. Pionic deuterium πD(3p-1s)

The study of pionic deuterium aims at the determination of the hadronic width of the ground state (see section 2.3).

The measured spectrum of the 3p-1s transition is shown in Figure 5.17, along with the spectrometer response function, and a line shape constructed from the response function and a Lorentzian width of $\Gamma_{1s} = 1171\,\text{meV}$ [4].

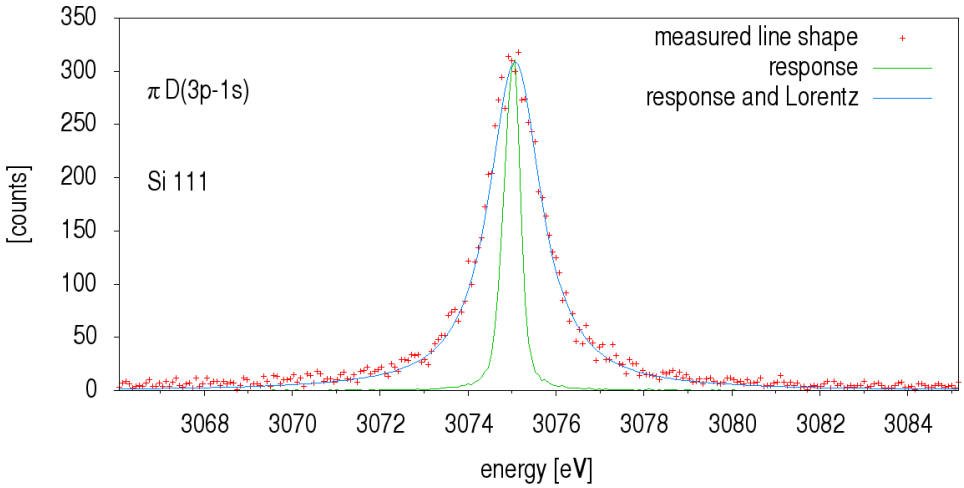


Figure 5.17.: Line shape of the measured 3p-1s transition of pionic deuterium. Additional to the data, the spectrometer response and a Lorentzian convolution ($\Gamma_{1s} = 1171\,\text{meV}$ [4]) are shown. In contrast to the case of the πH(2p-1s) (Figure 5.10), no clear indication for a Doppler broadening is seen here.

5.3.1. Determining the width of low-energy component

Again, the simplest line shape possible to assess contains only a single, low-energy Doppler contribution at 0 eV. To pin down the width of such a low-energy component, 100 models with widths between 0.5 eV and 50 eV have been evaluated and compared (see Figure 5.18). This comparison did not allow for an exact determination of the width, as the statistics of the available data limits the sensitivity of this analysis, but the width could be limited to be $T \lesssim 10\,\text{eV}$. The Bayes factor between the models with the highest and the lowest evidence is just $\ln(B_{\text{min,max}}) = 0.44$.

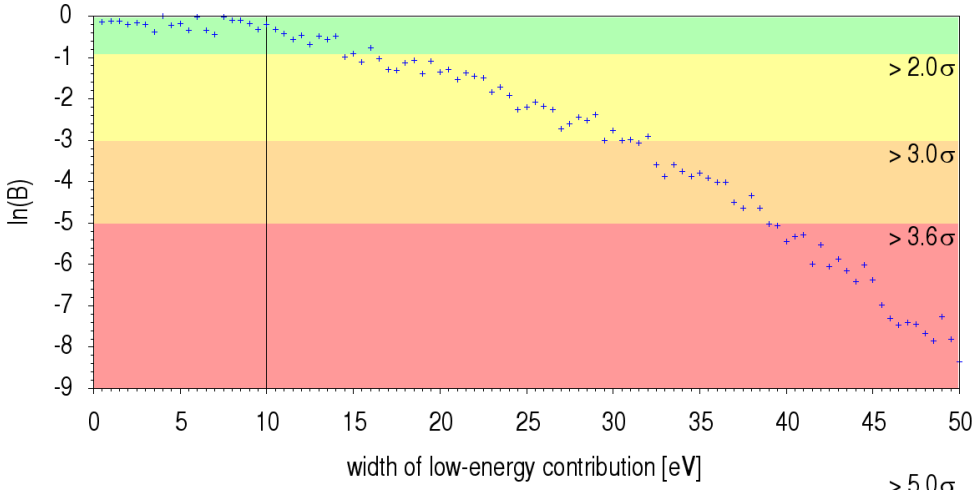


Figure 5.18.: Bayes factors for the πD(3p-1s) line using 100 different single Doppler component models with different widths. The maximal evidence is $e_{\text{1-comp.}} = 33646.38$.

5.3.2. Search for high-energy contributions

To study the possibility of high-energy Doppler components ($T > 0\,\text{eV}$), 1000 models with differently positioned high-energy components (between $0 - 50\,\text{eV}$, with width $2\,\text{eV}$) and varying low-energy contributions (of widths $0.5 - 10.0\,\text{eV}$) have been evaluated and compared. The two-dimensional evidence distribution is shown in Figure 5.19. The purpose of the variation of the width of the low-energy component within the priorily established limits was to eliminate a possible shadowing of a high-energy component by a broad low-energy component. The comparison of the different 2-component models show no indication for a high-energy contribution. In fact, the maximal evidences were found for the effective 1-component models (where the position of the second contribution is at $0\,\text{eV}$ - along with the fixed position of the low-energy component).

kin. energy dist.	max. evidence	ln(B)	B
2-comp.	33646.75		
1-comp.	33646.38	0.37	1.5

Table 5.6.: Comparison of the different models using the evaluated Bayes factors. The difference in evidence between the 1-component and 2-component models is negligible.

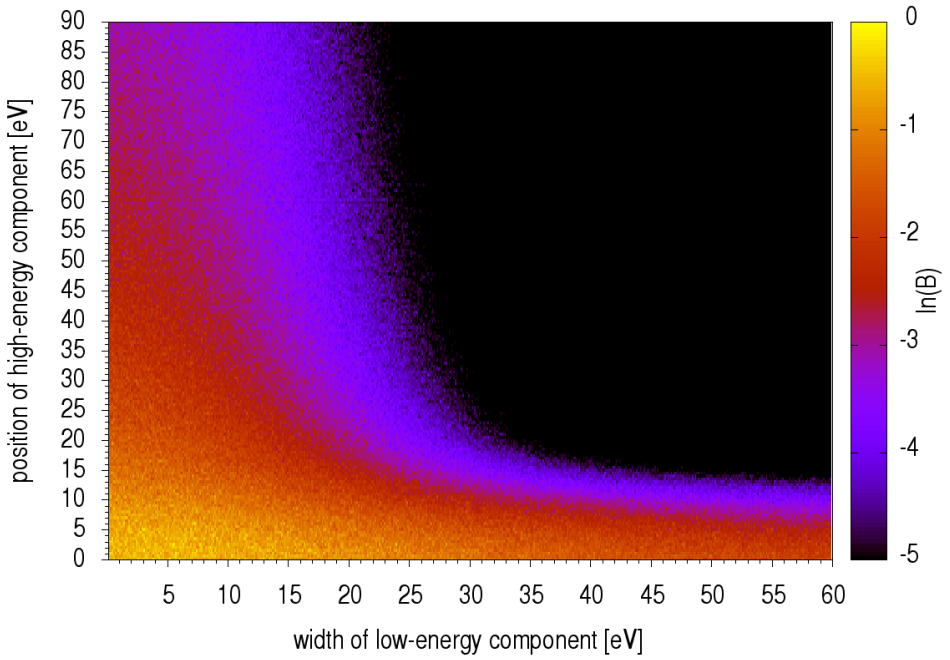


Figure 5.19.: The comparison of 57595 different 2-component models shows no significant indication for a high-energy Doppler contribution. Effective single component models are preferred of a width of about $T \lesssim 10\,\mathrm{eV}$. This is consistent with the result of section 5.3.1.

5.3.3. Hadronic width

The second component contributing significantly to the line shape is the Lorentzian width of the ground state. As both the kinetic energy as well as the hadronic component cause a broadening of the line their parameters strongly correlate. This dependence becomes obvious by plotting each maximum likelihood hadronic width Γ_{1s} of the 100 models in section 5.3.1 against the width of the low-energy component (Figure 5.20). Thus, the determination of the hadronic width Γ_{1s} cannot be accomplished without considering the kinetic energy distribution, even if it contains only a low-energy component.

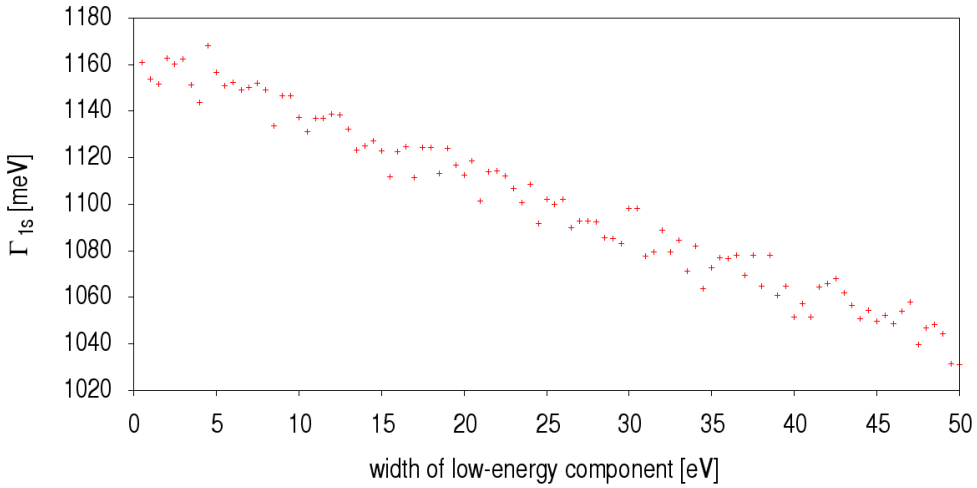


Figure 5.20.: The direct correlation of the width of the low-energy component with the hadronic width Γ_{1s} is shown in this plot. Each point corresponds to a different kinetic energy model with one component and width given at the horizontal axis. The value at the vertical axis shows the most probable Γ_{1s} value for that model.

To produce a model-independent posterior distribution for the hadronic width Γ_{1s}, all posterior samples of all 2-component models in section 5.3.2 have been combined, weighted by their evidences. This posterior projection, shown in Figure 5.21 is based on a total of 6822881 samples

5.3.4. Conclusions

The analysis of the πD(3p-1) data confirmed the previous results, that no high-energy Doppler contributions are detectable in the measured spectrum [4]. The upper limit of the low-energy component was reestablished with the same magnitude $\approx 10\,\mathrm{eV}$.

The comparison with the hadronic width $\Gamma_{1s} = 1171\,^{+23}_{-49}\,\mathrm{meV}$ shows a good coincidence with the result of this work ($1142\,^{+28}_{-32}\,\mathrm{meV}$), regarding both the most probable value, as well as the error margins.

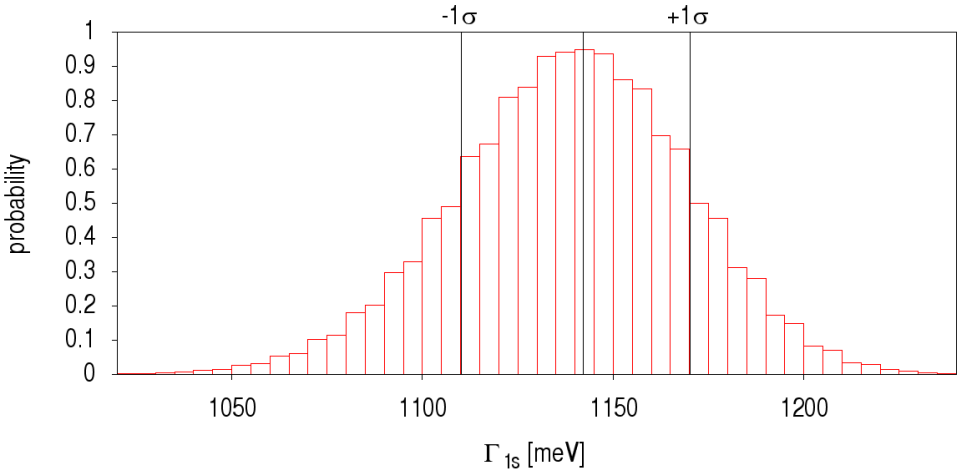


Figure 5.21.: Combined posterior distribution of the hadronic width Γ_{1s} with an average of 1142 meV and a credibility (1σ) interval of -32 meV, $+28$ meV.

parameter	value	credibility					
		1σ		2σ		3σ	
Γ_{1s} [meV]	1142	-32	$+28$	-66	$+56$	-102	$+84$
peak pos. [channels]	817.520	-0.140	$+0.141$	-0.280	$+0.281$	-0.420	$+0.421$
peak int. [counts]	8823	-104	$+105$	-207	$+210$	-310	$+317$
bg. [counts / channel]	3.223	-0.105	$+0.108$	-0.209	$+0.217$	-0.311	$+0.329$

Table 5.7.: Estimate of hadronic width Γ_{1s} and the remaining parameters in πD and common credibility intervals.

5.4. Muonic hydrogen μH(3p-1s)

The last spectrum analyzed was the 3p-1s transition of muonic hydrogen. As discussed in chapter 2, the width of this line shape has no Lorentzian component and is, therefore, ideal to study the Doppler broadening stemming from the kinetic energy distribution.

Figure 5.22 shows the data, collected at the Paul-Scherrer-Institute (PSI), Switzerland, along with the response functions of the individial peaks (see section 2.4) and the sum of these (with the theoretical hyperfine splitting value of 182.725 meV, and the assumed triplet / singlet population ratio of 3/1). All additional width of the line shape is attributed to the Doppler broadening.

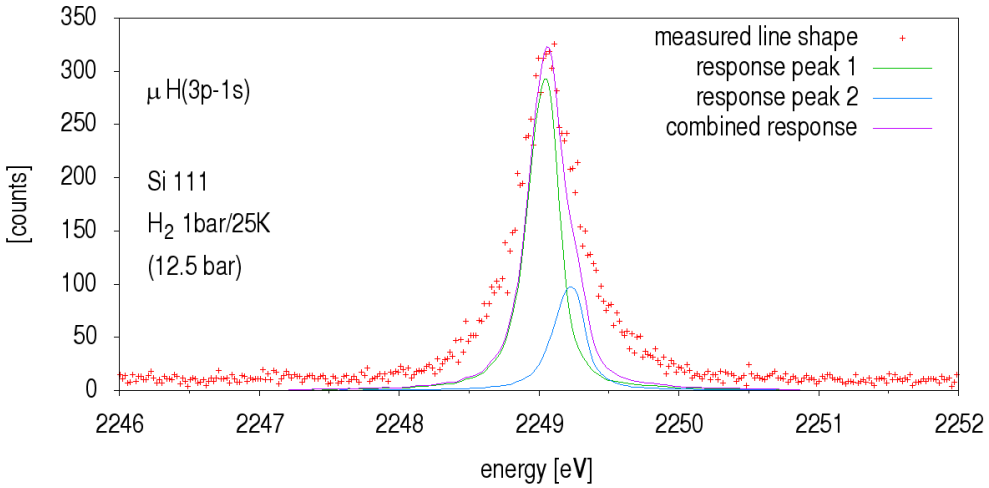


Figure 5.22.: Line shape of the measured 3p-1s transition of muonic hydrogen. Additional to the data, the individual spectrometer responses for each hyperfine peak and the cumulative response function are shown. The measured line shape is considerably wider than the response, which is the effect of the Doppler broadening.

5.4.1. Low-energy component

Although for μH at least one high-energy contribution is expected, for the sake of completeness a brief analysis with a single low-energy component was performed. The Bayes factors for each test width (0.25 eV to 30 eV) are depicted in Figure 5.23. The pronounced maximum at about $T \approx 18$ eV suggests the presence of significant high-energy contributions. Therefore, as in the previous analysis with the frequentist approach [6, 7, 8] the analysis was continued with 2-component models.

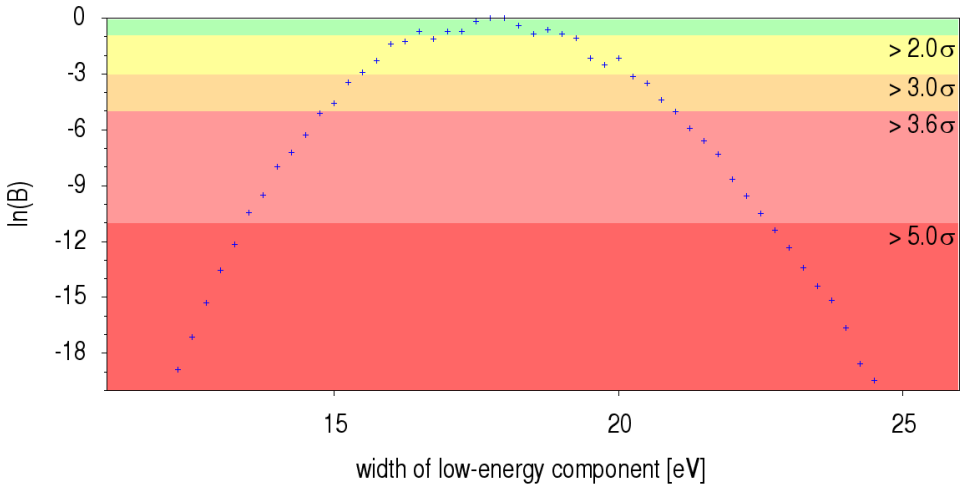


Figure 5.23.: Probabilities (logarithmic) of 120 different models with a single Doppler contribution starting at 0 eV and variable widths. The vanishingly low evidences for narrow energy intervals ($\lesssim 12$ eV) indicate the presence of high-energy components, causing an overestimation of the width. The maximum (reference) value has a (logarithmic) evidence of $e_{\text{max,1-comp.}} = 50171.36$.

5.4.2. Search for high-energy contributions

Analogous to the procedure in the πH and πD analysis, a high-energy component was added, along with variable widths of the low-energy component and finally evaluated. The resulting Bayes factors can be found as scatterplot in Figure 5.24. Apart from a relatively flat global maximum at (≈ 4.25 eV, 41 eV), the Bayes factors show a second local maximum. The contour lines themselves suggest that the shape is made from two ellipsoids. The conclusion that this is a hint for a second high-energy contribution seems natural. Compared to the single low-energy models in section 5.4.1, the Bayes factor $\ln(B_{\text{1-comp.,2-comp.}}) = 57.04$ clearly dictates the presence of at least one high-energy component.

To conclude the considerations of the 2-component models, the relative intensity of the low-energy contribution was studied. Again using model-averaging for all models with a Bayes factor $\ln(B) > -5$, the distribution of the low-energy intensity yields for a most probable value and 1σ credibility interval $65.5 \pm 3.4\%$ (Figure 5.25).

5.4.3. 2nd high-energy contribution

As the evaluation of 2-component models already suggested to search for a 2nd high-energy contribution, the analysis was continued with 3-component models. Basing on the results of section 5.4.2 a low-energy component was fixed at 0 eV to 4 eV and two high-energy contributions were added at different positions. The calculated Bayes factors

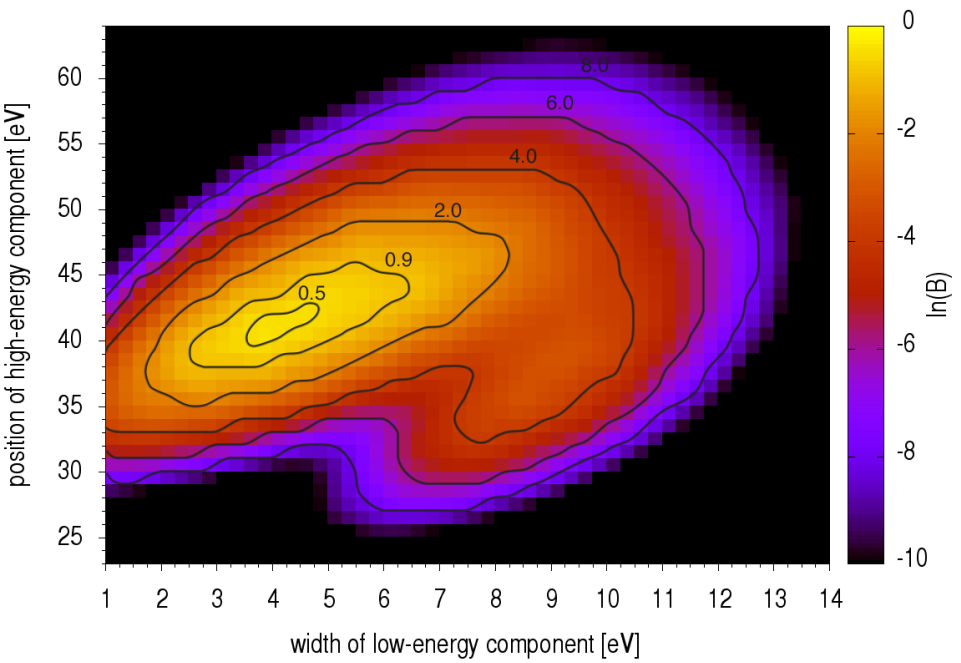


Figure 5.24.: Comparison of different 2-component models, with a low-energy contribution at 0 eV and variable width, and a high-energy component at positions between $23 - 64$ eV (of fixed width 2 eV), max. (logarithmic) evidence $e_{\text{max, 2-comp.}} = 50228.40$

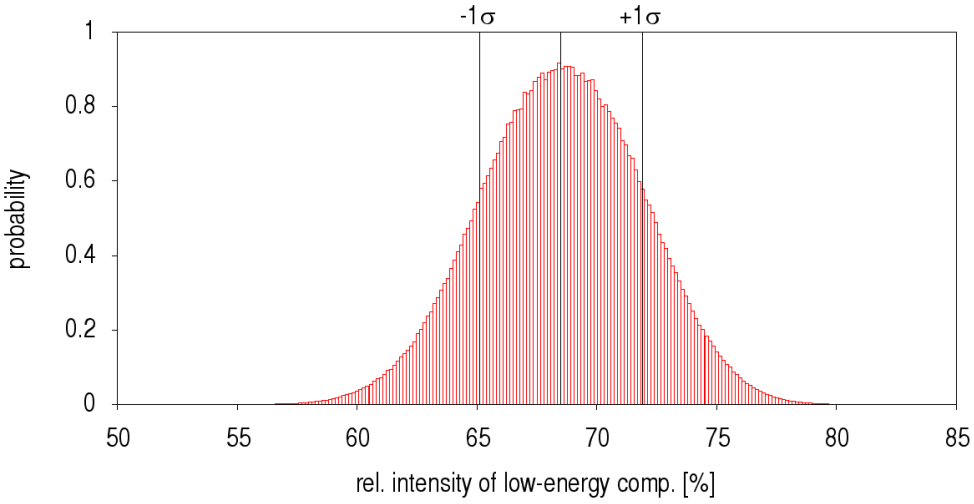


Figure 5.25.: Distribution of the model-averaged relative intensity of the low-energy component in the μH analysis with two Doppler contributions. Mean is 65.5% with a credibility interval (1σ) of -3.4% to $+3.4\%$.

according to the positions of the high-energy components can be found in Figure 5.26. Although there is a flat maximum at about (23 eV, 57 eV), the contours are significantly larger and less peaked compared to the 2-component results. The Bayes factor $\ln(B_{\text{2-comp., 3-comp.}}) = 0.78$ weakly favors a 3-component model. Furthermore, if a 2-component model would be more probable, the maximum should be found near one of the axes, similar to the πH analysis, where no indication for a 3$^{\text{rd}}$ component could be found (see section 5.2.3).

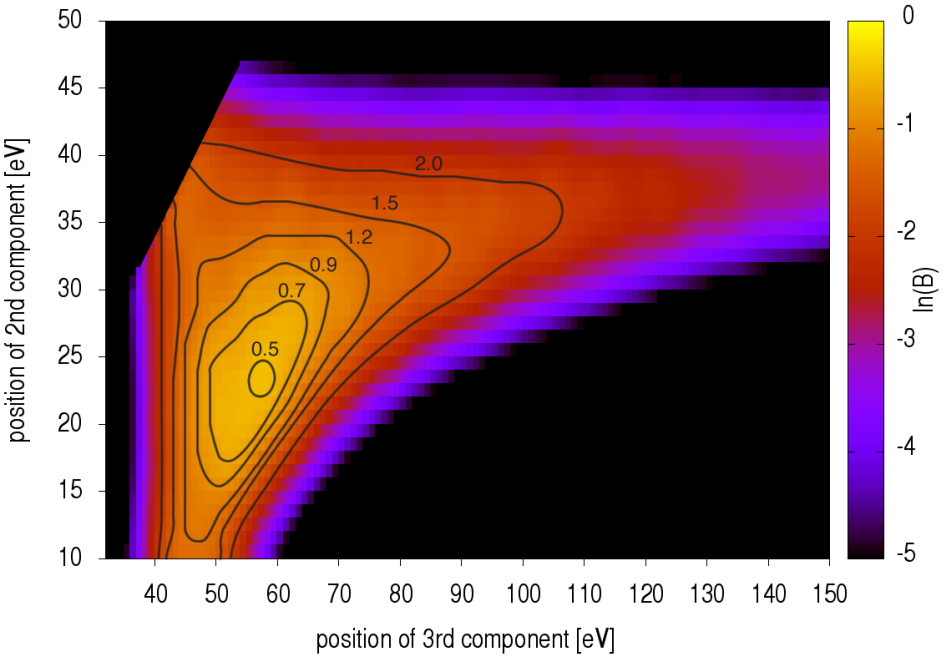


Figure 5.26.: Comparison of different 3-component models with a fixed low-energy contribution at 0 eV to 4 eV and two variable high-energy components. There is a flat maximum at (23 eV, 57 eV). The maximal (logarithmic) evidence is $e_{\text{max,3-comp.}} = 50229.18$.

Low-energy component

The low-energy component of the kinetic energy distribution marks an important quantity to compare the experimental results to ESCM calculations. Using the posterior samples of the 3-component model-selection along with their weights (Bayes factors), a model-averaged posterior projection for the relative intensity of the low-energy contribution is produced (see Figure 5.27).

Compared to the 2-component models (Figure 5.25), the distribution is significantly more asymmetric with a maximum at 65.02 % and a credibility interval (1σ) of -4.33 % to $+2.58$ %.

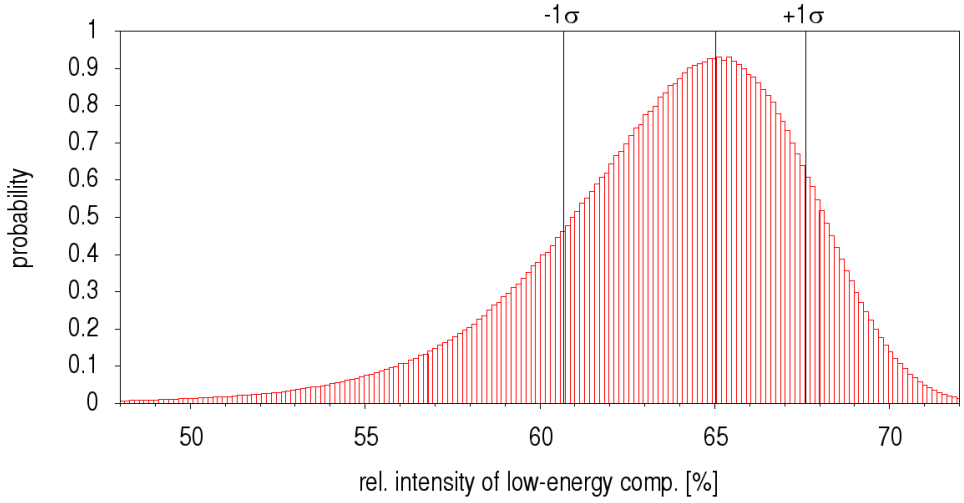


Figure 5.27.: Histogram of the model-averaged relative intensity of the low-energy component in the μH analysis with three Doppler contributions. The most probable value is at 65.02 % with a confidence interval (1σ) of -4.33 % to $+2.58$ %.

Hyperfine structure

The hyperfine structure of the hydrogen atom (see section 2.4) manifests itself in the presence of a second peak. The distance between these peaks corresponds to the energy splitting (see eq. 2.1) and the intensity ratio to the population of the singlet and triplet states.

With the help of the posterior samples and their model weights, a two-dimensional projection (scatter plot) of the posterior for these quantities could be produced (see Figure 5.28). There is a strong correlation present between the hyperfine splitting and the occupation of the states.

The calculated value for the hyperfine splitting of the ground state is, as given in section 2.4, 182.725 meV, which lies well within the 2σ credibility interval of the experimental value (Table 5.8).

parameter	value	credibility					
		1σ		2σ		3σ	
splitting [meV]	212.08	−21.81	+23.18	−47.48	+57.69	−88.01	+298.98
N_T/N_S	3.19	−0.70	+1.59	−1.33	+4.98	−1.94	+21.09
peak 1 pos. [channels]	975.083	−0.526	+0.555	−1.082	+1.158	−1.718	+1.834
peak 2 pos. [channels]	986.379	−1.365	+1.602	−2.84	+3.998	−4.804	+14.342
peak 1 int. [counts]	7921	−573	+563	−1234	+1185	−2119	+1885
peak 2 int. [counts]	2198	−555	+570	−1173	+1231	−1843	+2115
bg. [counts / channel]	10.154	−0.144	+0.147	−0.291	+0.298	−0.441	+0.455

Table 5.8.: Hyperfine splitting and distribution ratio N_T/N_S of the triplet / singlet states for the muonic hydrogen 3-component models in the ground state (1s), along with the remaining parameters.

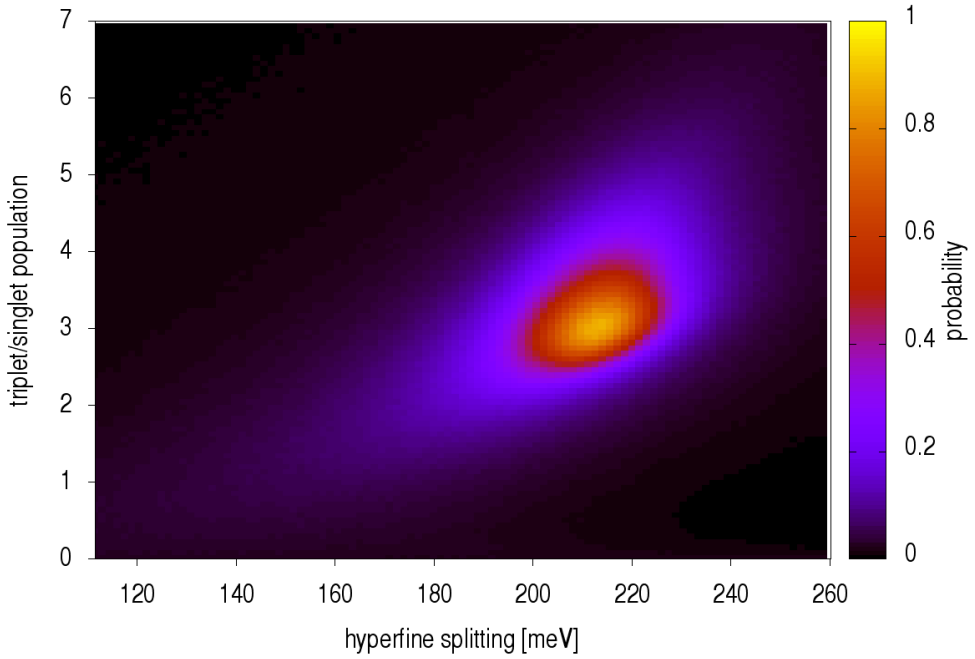


Figure 5.28.: Model-averaged scatterplot of the hyperfine splitting to the intensity ratio (of the two peaks). In total 21734069 posterior samples for 3-component models have been combined to produce this plot. The correlation between the hyperfine splitting and the intensity ratio is evident. A maximum can be found at about (212.08 meV, 3.19).

5.4.4. Deconvolution

Using the method described in section 4.6.2, a deconvolution of the underlying kinetic energy distribution was performed. The basis for this deconvolution were 20825 different 3-component models, with each component 3 eV wide. The resulting spectrum shows the expected dominance of a low-energy contribution with a relative intensity of 60.60 %. Noticable is the discrepancy to the previous result (Figure 5.27) with 65.02 %. This is due to the fact, that this deconvolution technique is based on weighted averages of each component's relative intensity. The previous result of 65.02 % though, is the most probable value of an extremely asymmetrical distribution and therefore not coherent with the weighted average.

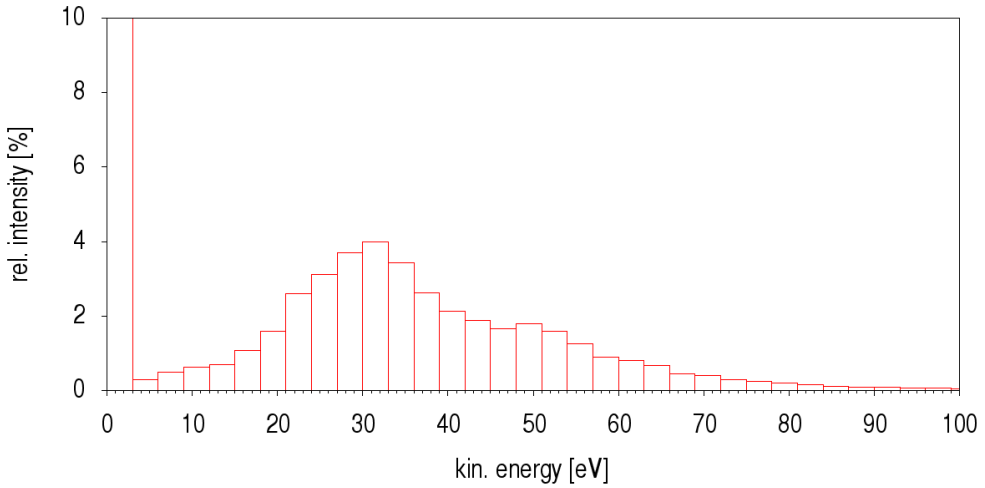


Figure 5.29.: Deconvoluted kinetic energy spectrum of the muonic hydrogen. The low-energy component has a proportion of 60.60 % of the total intensity. There are clear indications for high-energy components, but a reconstruction of a more complex structure was not possible.

5.4.5. Fixed hyperfine splitting

Additionally to the 3-component models with hyperfine splitting h as a free parameter, a set of 3-component models with $h = 182.725\,\text{meV}$ set to the theoretical value [18], was analyzed. The overall shape of the probability (evidence) distribution (Figure 5.30) is similar, but the maximum is slightly sharper (compared to Figure 5.26).

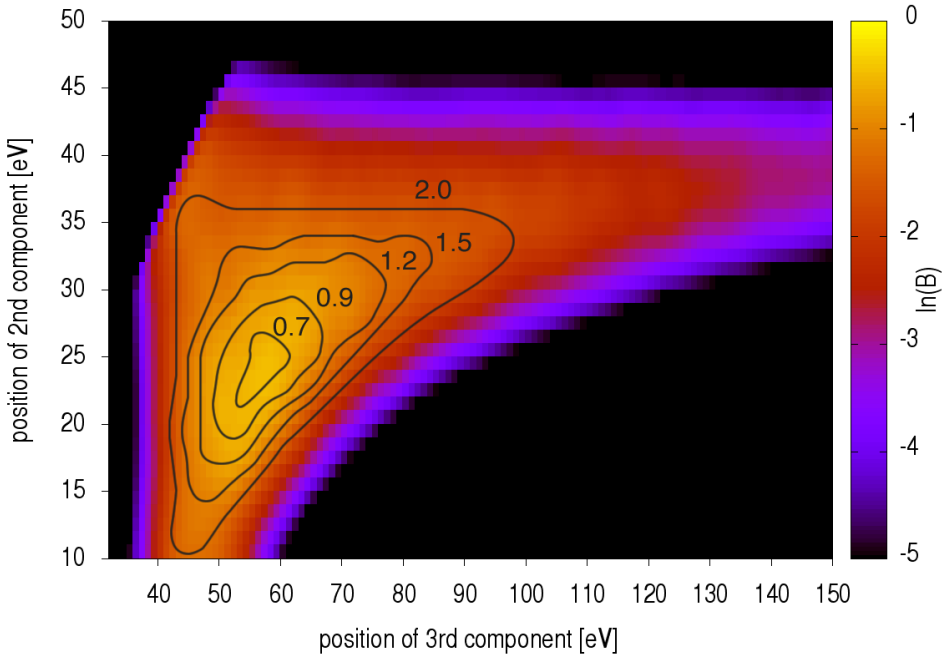


Figure 5.30.: Comparison of different 3-component models with a fixed low-energy contribution at 0 eV to 4 eV and two variable high-energy components. There is a flat maximum at about (25 eV, 55 eV). The maximal evidence is $e_{\text{max,3-comp.}} = 50231.41$, yielding a Bayes factor (in comparison with the analysis in section 5.4.3) of $\ln(B) = -2.23$.

Relative population of the triplet / singlet states

The triplet / singlet population ratio correlates strongly with the hyperfine splitting (see section 5.4.3). It was, therefore of particular interest how the posterior for this parameter evolves, if the hyperfine splitting is fixed to the theoretical value of 182.725 meV.

Based on the 3-component evaluations, the histogram for the N_T/N_S posterior projection was created and its maximal value and credibility intervals were determined. The correlation causes a lower most probable value for the triplet / singlet population as expected, along with smaller credibility intervals.

The expected value of 3, originating from a statistical population, lies well within the 1σ-interval.

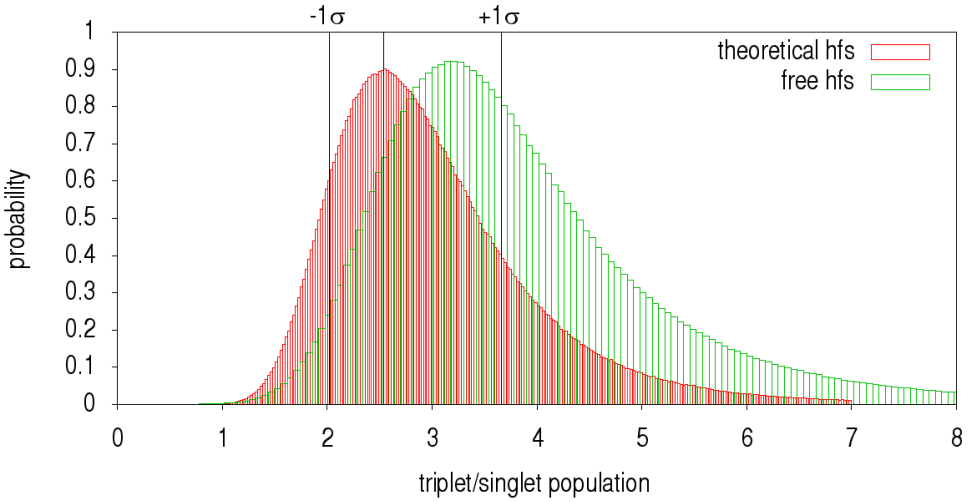


Figure 5.31.: Comparison of the model-averaged intensity ratio of the μH 3-component analysis with fixed, theoretical hyperfine splitting of 182.725 meV and splitting being a free parameter.

parameter	value	credibility					
		1σ		2σ		3σ	
N_T/N_S	2.54	-0.51	$+1.13$	-0.95	$+3.00$	-1.31	$+7.38$

Table 5.9.: Relative population N_T/N_S of the triplet / singlet states for the muonic hydrogen 3-component models in the ground state (1s) with hyperfine splitting fixed to the theoretical value.

Relative intensity of the low-energy component

The most probable value for the low-energy intensity is slightly lower if the hyperfine splitting is fixed to its theoretical value. The values though, are still within their 1σ intervals and therefore do not contradict each other.

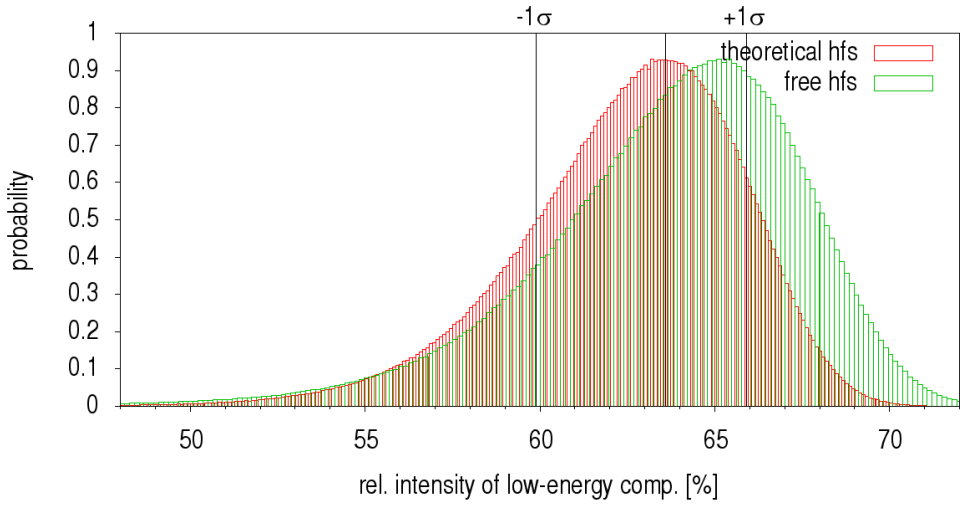


Figure 5.32.: Comparison of the model-averaged relative intensity of the low-energy component in the μH 3-component analysis with fixed and free hyperfine splitting.

parameter	value	credibility					
		1σ		2σ		3σ	
low-energy int. [%]	63.59	−3.70	+2.31	−8.16	+4.36	−13.81	+6.18

Table 5.10.: Relative intensity of the low-energy component with fixed hyperfine splitting.

Deconvolution

The deconvolution of the kinetic energy spectrum with a fixed hyperfine splitting value shows significantly more structure (Figure 5.33). It reveals the presence of two high-energy components. The first maximum is at around 25 eV with a spread of $\approx \pm 15$ eV, and another at 50 eV with a much larger width. This is in agreement with the 3-component evaluation in section 5.4.3.

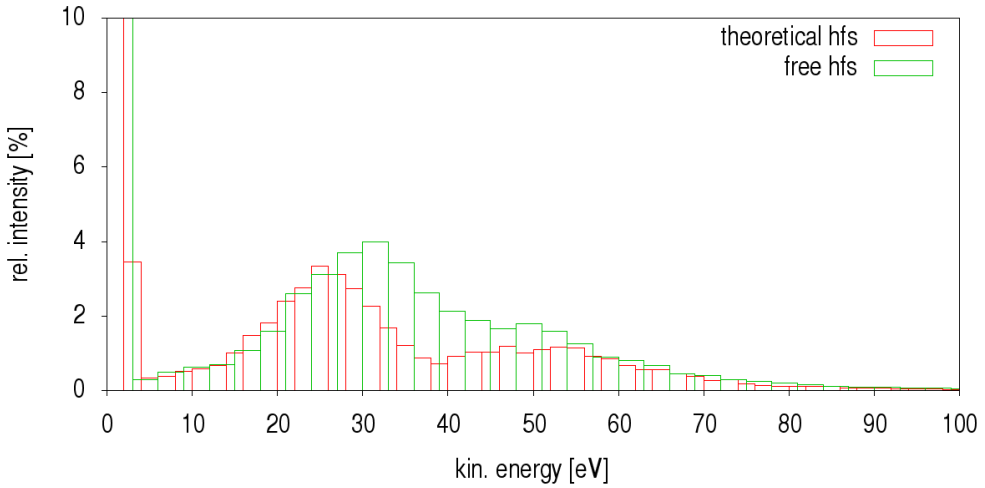


Figure 5.33.: Deconvolution / model-averaged kinetic energy distribution for the μH analysis with fixed hyperfine splitting in comparison to the previous result. The relative intensity of the first bin is 53.7%. This deconvolution shows more structure: two more or less distinctive high-energy components can be recognized, around 25 eV and 50 eV.

5.4.6. Theoretical kinetic energy distribution for the μH 3p state

After the evaluation of different simplified kinetic energy distributions, modeled by $1, 2$ or 3 components, two complex distributions for the 3p-1s transition in muonic hydrogen, calculated with the ESCM framework by V. Popov and V. Pomerantsev [25] from 2008 and 2013 have been applied to the measured data.

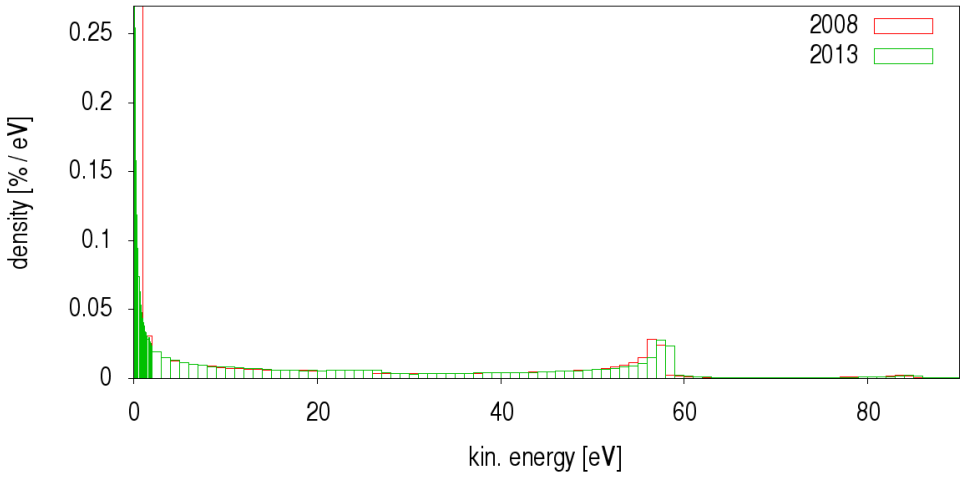


Figure 5.34.: The 2008 and the 2013 kinetic energy distribution of the 3p-1s transition in muonic hydrogen, calculated by means of the ESCM framework by V. Popov and V. Pomerantsev [25]. The predicted spectra differ mainly in the shape of the low-energy part (due to different binning), and slightly around the component between 50 % and 60 %.

The 2008 ESCM kinetic energy distribution was first transformed into a line shape (see section 4.4), by generating a line shape for each bin individually, and superposing these using the bin heights. The thus obtained test spectrum was evaluated against the data 24 times to counteract the dispersion of the evidence and to produce more samples. This yielded an average (logarithmic) evidence of $e_{\mathrm{th},2008} = 50227.77 \pm 0.28$, which is slightly below the maximal evidence for the 2-component analysis of 50228.40.

The assessment of the 2013 kinetic energy distribution, though, resulted in an even smaller average (logarithmic) evidence of $e_{\mathrm{th},2013} = 50225.73 \pm 0.22$. This relatively large difference to the simplified 3-component model was unexpected and the result for the most recent, and more precise 2013 spectrum was puzzling. As the difference between the two kinetic energy distributions is the shape of the low-energy component (Figure 5.35), it was decided to test the 2013 spectrum with a re-binned (to 1 eV) low-energy part, which yielded an improved average (logarithmic) evidence of 50227.66 ± 0.28, very similar to the 2008 result. In addition, the re-binned 2013 spectrum was analyzed 24 times with a fixed hyperfine splitting parameter (182.725 eV), which gave an average (logarithmic) evidence of 50226.96 ± 0.28

This sensitivity to the shape of the low-energy component was studied in an additional analysis (see section 5.4.7).

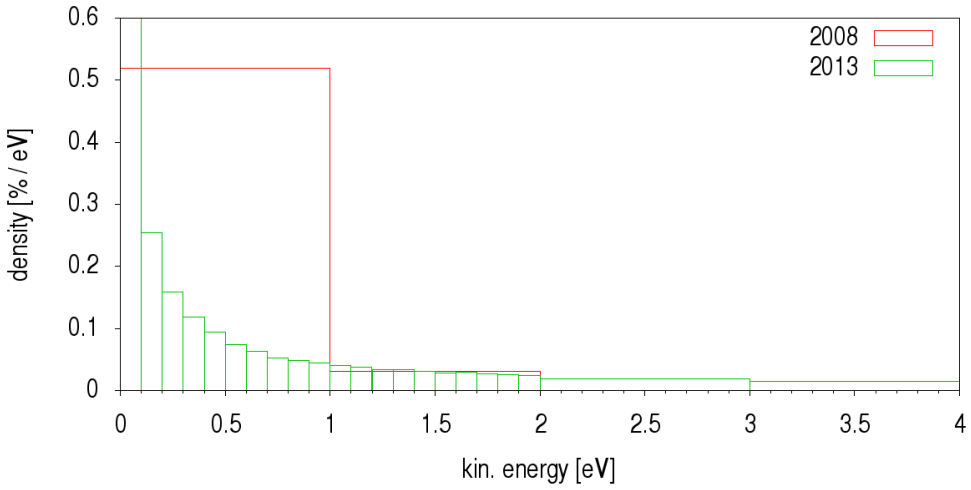


Figure 5.35.: Detailed view of the low-energy part of the two theoretical kinetic energy distributions (see Figure 5.34). The 2013 spectrum is more detailed below 2 eV with a binning of 0.1 eV (contrary to 1 eV for the rest of the 2013 and the entire 2008 distribution).

Hyperfine structure and triplet / singlet population ratio

The results for the hyperfine splitting and the triplet / singlet population ratio (Table 5.11) show smaller confidence intervals than before (see Table 5.8). The most probable values of triplet / singlet population ratio, though, differ greatly from the previous results, and the evidences (mentioned above) suggest that these distributions do not match the data as well as the simplified 3-component model. The same behavior is found in the evaluation of the 2013 re-binned spectrum with fixed hyperfine splitting, where the expected population ratio 3 is far from the result $1.70\,^{+0.33}_{-0.24}$ and even beyond the 3σ interval.

5.4.7. Shape of the low-energy component

Driven by the results of the theoretical kinetic energy spectra and the assumed dependence on the shape of the low-energy contribution, the analysis was extended by two steps:

1. Verifying the width of the low-energy component by varying it in the presence of two high-energy components at the previously established positions of 23 eV and 57 eV.

2. Subdividing the low-energy component into six components, additional to the two fixed high-energy contributions.

kin. en. dist.	parameter	value	credibility					
			1σ		2σ		3σ	
2008	split. [meV]	213.40	−11.09	+12.81	−21.74	+25.53	−32.17	+38.90
	N_T/N_S	2.32	−0.38	+0.47	−0.70	+0.99	−0.98	+1.60
2013	split. [meV]	209.71	−10.03	+11.01	−19.41	+22.06	−28.76	+33.48
	N_T/N_S	2.10	−0.31	+0.38	−0.59	+0.78	−0.82	+1.25
2013 1 eV bins	split. [meV]	213.69	−11.00	+12.75	−21.74	+25.21	−32.21	+38.75
	N_T/N_S	2.31	−0.38	+0.46	−0.70	+0.97	−0.98	+1.56
2013 1 eV bins fixed hfs	N_T/N_S	1.70	−0.24	+0.33	−0.47	+0.69	−0.65	+1.08

Table 5.11.: Hyperfine splitting and distribution ratio N_T/N_S of the triplet / singlet states for the muonic hydrogen theoretical kinetic energy distribution.

Width of the low-energy component

In order to verify the width of the low-energy component, 100 models with widths varying from 0.1 eV to 10 eV and fixed high-energy contributions at 23−25 eV and 57−59 eV have been generated and each one evaluated 20 times. The maximal evidence thus achieved was 50230.18.

The resulting distribution of the width of the low-energy component is shown in Figure 5.36, weighted averaging gave a result of (4.66 ± 1.96) eV, which is consistent with the result of section 5.4.2.

Precise shape of the low-energy component

The next step was to test an 8-component model, with two high-energy contributions at $23.5 - 24.5$ eV and $57.5 - 58.5$ eV (similar to those before, but of smaller width), and a subdivided (by 1 eV components) low-energy part between $0 - 6$ eV. This model was evaluated 120 times with a fixed and free hyperfine splitting parameter to generate enough samples in order to reliably determine the intensities of the low-energy slices.

The average (logarithmic) evidence thus found was 50333.88 ± 0.33 (for free hyperfine splitting) and 50233.84 ± 0.32 (for fixed hyperfine splitting) and, therefore, considerably better than the 3-component models with a rectangular low-energy component.

In order to reconstruct the individual relative intensities of the low-energy components, the most probable value for each of these and the high-energy components was determined individually using all samples and attributed weights of all 120 evaluations. The result is shown in Figure 5.37.

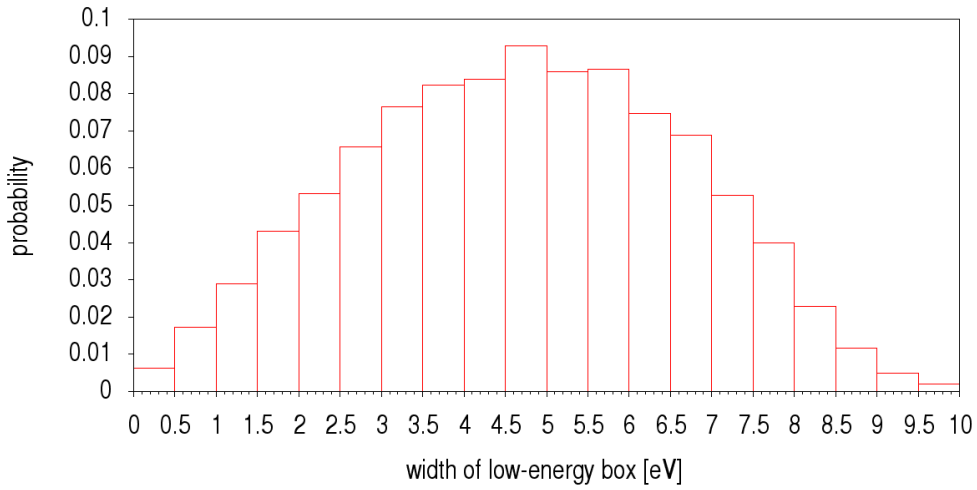


Figure 5.36.: Distribution of the low-energy component width, resulting in a total 20 evaluations of 100 different widths between 0.1 eV and 10 eV. The weighted average gave a width of 4.66 ± 1.96 eV.

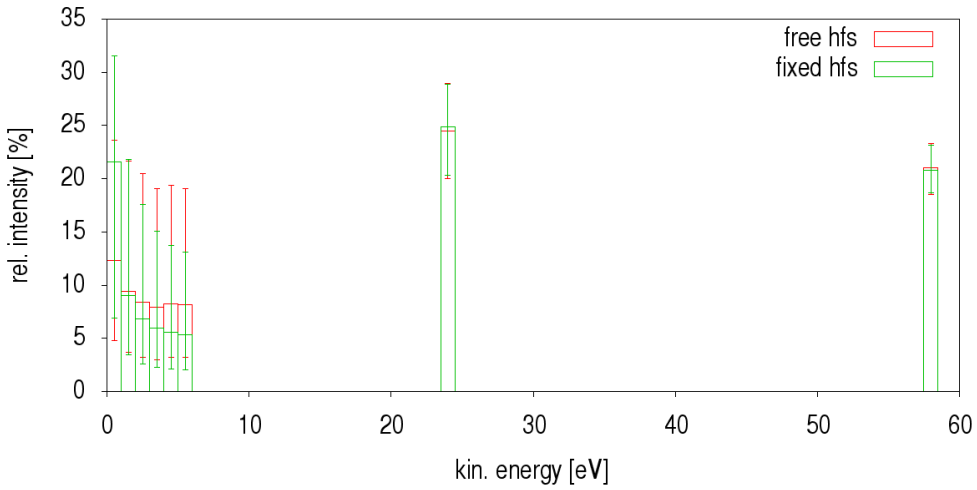


Figure 5.37.: Reconstructed low-energy component of the μH(3p-1s) X-ray line. The shown intensities are the respective most probable values along with the 1σ credibility interval. The model of six low-energy components provides a better agreement with the data than the 3-component model with a single, rectangular low-energy contribution ($\ln(B) = -4.7$). This supports the conclusion in section 5.4.6, that the analysis is sensitive to the shape of the low-energy component.

Although the error margins are relatively high compared to the most probable value, a tendency can be seen for a maximum at 0 eV, followed by smaller contributions. In the case of the theoretical hyperfine splitting, the 0 eV energy component is significantly more distinctive.

hfs	parameter	value	credibility					
			1σ		2σ		3σ	
free	$0-1$ eV	12.35	−7.54	+11.28	−11.47	+24.30	−12.29	+35.76
fixed		21.56	−14.63	+10.02	−20.26	+21.44	−21.47	+31.09
free	$1-2$ eV	9.44	−5.72	+12.22	−8.79	+25.93	−9.40	+38.05
fixed		9.01	−5.54	+12.79	−8.42	+28.16	−8.97	+40.37
free	$2-3$ eV	8.38	−5.17	+12.10	−7.83	+27.71	−8.34	+40.94
fixed		6.86	−4.27	+10.76	−6.42	+25.44	−6.83	+39.13
free	$3-4$ eV	7.92	−4.90	+11.18	−7.41	+25.88	−7.89	+39.67
fixed		5.94	−3.69	+9.11	−5.56	+21.63	−5.93	+34.56
free	$4-5$ eV	8.27	−5.05	+11.10	−7.72	+24.91	−8.23	+37.47
fixed		5.58	−3.46	+8.16	−5.22	+19.41	−5.56	+31.30
free	$5-6$ eV	8.16	−4.91	+10.94	−7.60	+23.87	−8.12	+35.63
fixed		5.35	−3.31	+7.78	−5.01	+18.31	−5.33	+28.80
free	$23.5-24.5$ eV	24.47	−4.42	+4.49	−9.27	+9.72	−14.62	+16.23
fixed		24.87	−4.54	+4.01	−9.35	+7.99	−14.53	+12.13
free	$57.5-58.5$ eV	21.01	−2.50	+2.30	−5.73	+4.77	−10.01	+7.47
fixed		20.84	−2.14	+2.33	−4.41	+4.81	−6.82	+7.42

Table 5.12.: Most probable values and credibility intervals for the relative intensities shown in Figure 5.37.

Relative intensity of the non-rectangular low-energy component

The relative intensity of the total low-energy component does not depend on its shape. The results for these are in agreement with the 3-component result (Figure 5.38, Table 5.13).

hfs	parameter	value	credibility					
			1σ		2σ		3σ	
free	low-energy int. [%]	64.11	-2.39	$+2.52$	-4.95	$+5.22$	-7.65	$+8.15$
fixed		62.41	-2.08	$+2.44$	-4.13	$+5.04$	-6.21	$+7.84$

Table 5.13.: Relative intensity of the combined low-energy components (see Table 5.12).

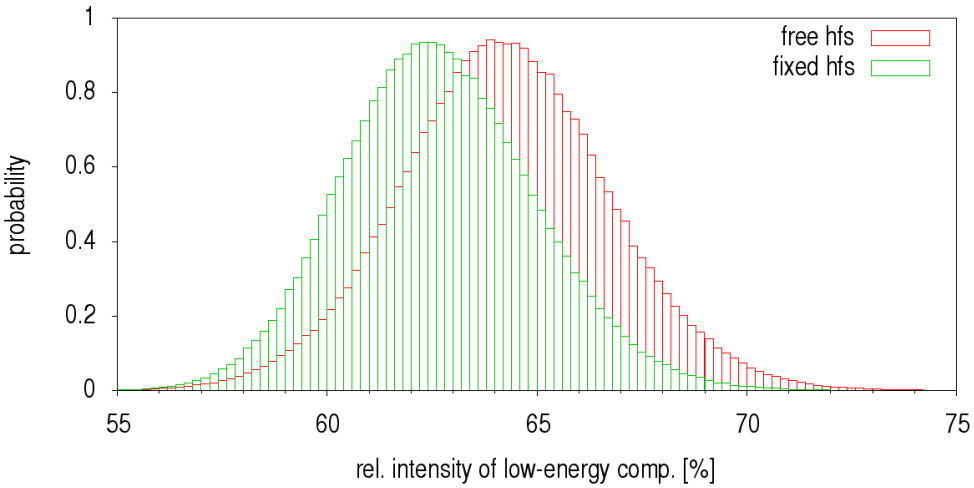


Figure 5.38.: Comparison of the distribution of the cumulative low-energy intensity for the free and fixed hyperfine splitting analyses of the μH(3p-1s) line.

Hyperfine splitting and triplet / singlet population ratio

Analogous to the previous cases, the hyperfine splitting and the triplet / singlet population ratio were determined. The analysis with hyperfine splitting as a free parameter yields values with large credibility intervals far from the theoretical value for the expected hyperfine splitting of 182.725 eV and triplet / singlet population ratio of 3. This broad distribution illustrates the fact, that the width and the shape of the low-energy component correlate with the hyperfine splitting and the triplet / singlet population ratio. In order to pin down both, more information *i. e.* more data is necessary.

In the case of the fixed hyperfine splitting, however, a good agreement of the triplet / singlet population ratio with the expected value could be found.

hfs	parameter	value	credibility					
			1σ		2σ		3σ	
free	splitting [meV]	245.53	−28.94	+68.00	−61.23	+213.59	−113.00	+452.63
	N_T/N_S	6.72	−1.69	+6.23	−3.01	+21.93	−4.01	+63.59
fixed	N_T/N_S	3.12	−0.61	+1.54	−1.14	+4.52	−1.60	+13.13

Table 5.14.: Hyperfine splitting and distribution ratio N_T/N_S of the triplet / singlet states for the muonic hydrogen analysis of the low-energy shape.

5.4.8. Conclusions

Due to the presence of two peaks, the muonic hydrogen 3p-1s spectrum renders substantial difficulties for the analysis. It could be established, that the kinetic energy distribution is made of at least two high-energy components in addition to the low-energy contribution, which is consistent with the results in [7, 8]. The evaluation of the theoretical spectra, though, revealed a correlation between the width of the low-energy component (expressed in the bin width) and the triplet / singlet population ratio. This became obvious upon modeling the low-energy contribution with six 1 eV bins. The additional parameters made it difficult to pinpoint the hyperfine splitting and therefore the triplet / singlet population, and resulted in very broad distributions for these parameters (see Table 5.14). With a fixed hyperfine splitting (at the theoretical value of 182.725 eV), though, a population ratio 3.12 close to the expected value of 3 could be found.

model	hfs	max. evidence	ln(B)	B
8-comp.	free	50333.88		
8-comp.	182.725 meV	50233.84	0.04	1.0
3-comp.	182.725 meV	50231.41	2.43	11.4
3-comp.	free	50229.18	4.70	110
2-comp.	free	50228.40	5.48	239
theory 2008	free	50227.77	6.11	450
theory 2013 (re-binned)	free	50227.66	6.22	502
theory 2013 (re-binned)	182.725 meV	50226.96	6.92	> 1000
theory 2013	free	50225.73	8.15	> 3400
1-comp.	free	50171.36	62.52	$> 10^{27}$

Table 5.15.: Comparison of the evidences / Bayes factors for all tested models for the muonic hydrogen 3p-1s X-ray line. The 8-comp. models are the last tested models with two high-energy and six low-energy contributions.

The comparison of the triplet / singlet population ratio with the previous results [6, 7, 8] showed a principal agreement for the 3-component model, with slightly lower values. For the theoretical spectra, however, they differ significantly, especially in relation to their credibility intervals.

Wheter the broadening of the low-energy component is physical, *e. g.* deceleration of high-energy components cannot be answered decisively given the quality of the data.

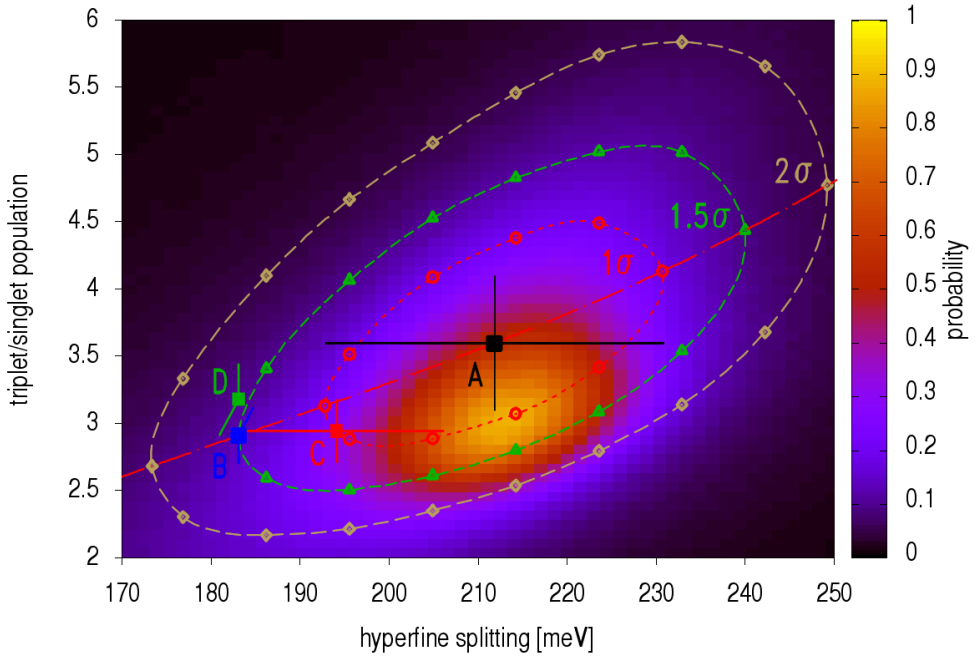


Figure 5.39.: Comparison to the previous results for the hyperfine splitting and state population. A, B, C, and D mark the previous results from [6, 7, 8] given in Table 5.16.

	hfs [meV]	kin. energy	N_T/N_S		β_0 [%]		ln(B)
			this work	[7, 8, 6]	this work	[7, 8, 6]	
B	182.725	3-comp.	$2.54^{+1.33}_{-0.51}$	$2.90^{+0.20}_{-0.20}$	64^{+2}_{-4}	61^{+2}_{-2}	
A	*free*		$3.19^{+1.59}_{-0.70}$	$3.59^{+0.51}_{-0.51}$	65^{+3}_{-4}		2.23
C	*free*	theory 2013 re-binned	$2.31^{+0.46}_{-0.38}$	$2.94^{+0.24}_{-0.24}$	(55)		4.33
D	182.725		$1.70^{+0.33}_{-0.24}$	$3.17^{+0.27}_{-0.27}$			4.97

Table 5.16.: Comparison of the results with previous measurements [6, 7, 8]: triplet / singlet population ratio N_T/N_S and relative intensity of the low-energy component β_0. It is striking that the error margins from the new analyses are quite larger compared to the previous results. This is a consequence of the Bayesian methods, that provide a more consistent access to error margins using confidence intervals from the posterior projection (see Figure 5.31).

6. Conclusion and outlook

6.1. Summary of results

The purpose of this thesis was to explore Bayesian methods regarding its applicability for the analysis of X-ray spectra of exotic atoms at hand. The previous analysis of some of such spectra (namely πH(2p-1s) [1], πD(3p-1s) [3], μH(3p-1s) [6]) by means of conventional frequentist methods, provided results to be reproduced with the new method.

Starting with pionic hydrogen, it soon became obvious that the Doppler broadening in the presence of a hadronic broadening is difficult to pin down. Therefore, the analysis methods were tested at first on simulated spectra with known input parameters. It could be shown, that a structural resolution of the high-energy Doppler components is not possible for a relative intensity $\lesssim 15\,\%$, but the relative intensity itself of the low-energy contribution was reproducible. The following analysis of real pionic hydrogen data then showed a diffuse high-energy contribution, centered at around 70 eV without hints for additional high-energy components. The result for the hadronic width $\Gamma_{1s} = 878\,^{+32}_{-29}$ meV was then in good agreement with the previous result of $\Gamma_{1s} = 902\,^{+25}_{-25}$ eV.

Similar conclusions could be drawn from the pionic deuterium analysis. The absence of any high-energy contributions was expected from the previous work [3, 4], and the hadronic width of the ground state $\Gamma_{1s} = 1142\,^{+28}_{-32}$ eV differed from the value given in [4] by only 1σ.

For the analysis of the muonic hydrogen 3p-1s line, a lot of different models for the kinetic energy distribution, hyperfine splitting, and level population have been tested to fully explore the possibilities as well as limitations of model-selection using the Bayesian approach. As the muons do not interact strongly with the nucleus, the absence of this broadening uncovers more details in the kinetic energy distribution. In agreement with previous works [6, 7, 8], the kinetic energy distribution could be modeled by 3 rectangular components at 0 eV, 23 eV, and 57 eV, which yielded good results for the hyperfine splitting and the triplet / singlet population ratio. The evaluation of the theoretical spectra, however, gave rise to some difficulties to establish the triplet / singlet population ratio. This was attributed to a correlation between the shape of the low-energy component, the hyperfine splitting and the triplet / singlet population ratio. The supplementary study with a more complex low-energy component showed a better agreement with the data in terms of evidence, but due to the large dimensionality of the parameter space, the distributions of the hyperfine splitting and the triplet / singlet population parameters became very diffuse. The fixing of the hyperfine splitting to the theoretical value, though, provided a good result of the population ratio of $3.11\,^{+1.54}_{-0.61}$, close to the expected statistical value of 3.

In general, the analyses have shown that the results of the conventional methods are reproduced, which is the minimal requirement for the adoption of a new data analysis method. Furthermore, Bayesian methods provide plenty of new possibilities. At first, there is a coherent approach to model-selection, which was successfully applied in all three analyses and helped to distinguish between kinetic energy distributions with a single, two or three components.

And despite the fact, that (*e. g.* in the case of the muonic hydrogen), a specific 3-component configuration could not be singled out by simply comparing the evidences of the studied configurations, the distribution of the evidences in a scatter plot made it possible to identify a most probable combination of high-energy components. Furthermore, model-independent parameter estimates could be produced. Another new possibility is the deconvolution, developed by L. Simons [27], based on model-averaged kinetic energy spectra, which could provide an alternative to modeling this distribution using several rectangular components.

Regarding (model-averaged) parameter inference, the new methods provided a more consistent way of obtaining parameter values along with error margins from the corresponding probability distribution for each parameter p_i (by marginalization over the remaining parameters $p_{j\neq i}$).

During the analysis of the muonic hydrogen with widely different models, and parameter space dimensionalities between $4 \leq n \leq 12$, a correlation of the size of the credibility intervals with the parameter space dimensionality n became apparent. As the credibility intervals are determined using projections of the posterior distributions, the marginalization over the remaining $n - 1$ parameters causes a broader (projected) distribution. This can be understood in the case of the triplet / singlet population ratio of muonic hydrogen, when comparing the models of free with the ones with fixed hyperfine splitting. Where the projection of the triplet-to-singlet intensity ratio for the former case is a marginalization over all possible hyperfine splitting values, the projection of the latter corresponds more to a slice of the posterior distribution of the free hyperfine splitting case.

6.2. Software package

The software package developed over the course of this work provides the methods and routines used in the analyses. An important aspect in designing this package was sensibility in order to adapt new experimental conditions and parameterizations, to implement new methods of analysis and post processing, and to improve the numerical algorithms.

An improvisation to the nested sampling algorithm or, in particular, to the likelihood evaluation was considered by using the vector computation unit of the GPU (graphics processing unit). As the combined likelihood is a product of individual, independent likelihoods, each stemming from a separate data point, vector calculation of these seems feasible. Additionally, the interpolation of two- or three-dimensional images (or in this case, matrix of test spectra) is a core competence of graphic processors and could provide another speedup.

6.3. Outlook

Having established and tested a set of methods and procedures to analyze X-ray spectra lines, these can now be applied to various measured spectra. Other transitions (3p-1s, 4p-1s for pionic hydrogen, 2p-1s, 4p-1s for pionic deuterium and muonic hydrogen) are to be re-analyzed and compared with this as well as with previous results.

Furthermore, an analysis of the kinetic energy distribution of pionic nitrogen (πN) promises new insights about the Coulomb explosion of the nitrogen molecule N_2, where also a Doppler broadening occurs. In addition, the analysis of complex X-ray spectra from electronic atoms may profit [44, 45].

Last but not least, simulations, like the ones in section 5.1, could help in designing new experiments by assessing the required precision before hand.

A. Appendix

A.1. Implementation

A.1.1. Overview

As part of this thesis, a software package named BayXRay was created to implement the numerical algorithms as well as support repetitive tasks with common routines. This software package mainly consists of four layers. Each layer adds another level of abstraction to the numerical algorithms, providing different possibilities for extension and integration. The purpose of this architecture is to provide reasonable flexibility and extendibility without the sacrifice of computational performance.

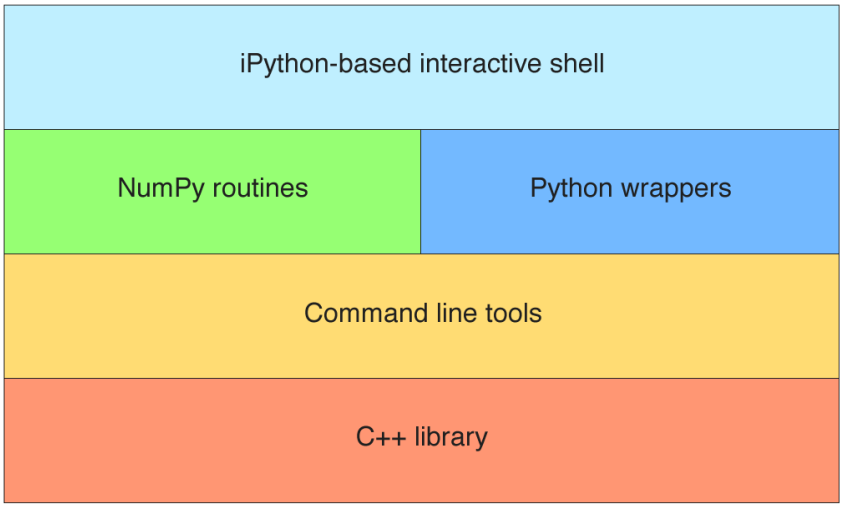


Figure A.1.: The architectural layers of the software package BayXRay.

Interactive shell provides an iPython based interface, similar to MATLAB to perform analyses (generate and evaluate model sets, perform post-processing tasks *e. g.* histograming).

Python wrappers are an API to use the command line tools provided by the software package to prepare model sets and test spectra, convolve input spectra or start parallelized nested sampling runs.

NumPy routines are a set of small routines to post-process the results of nested sampling evaluations. There are methods to generate histograms, model-averages, scatter-plots as well as tools to create fully-contained gnuplot scripts to visualize the data and results.

Command line tools provide small tools to perform file-based task like convolutions, generation of spectra and nested sampling evaluations. These tools could, in general be used directly or in Bash scripts.

C++ library is a set of classes and methods to implement the numerical algorithms along with a set of commonly used tools.

A.1.2. Python and NumPy routines

The most important tools, which were used mostly exclusively through the analysis, are the Python and NumPy routines. These provide the most common functions used in the analysis of X-ray spectra, and encapsulate repetitive tasks.

Some examples of analysis scripts, which make use of these routines, are provided in section A.2. As discussed in section 4.4, several preparation steps are required before the actual evaluation, which can be simplified using the Python tools. The following is a brief overview of the available routines.

Model generators

As it was a common task to create different models for the kinetic energy distribution, with either variable widths, positions, number of components or the like, a few functions to generate these models were written.

`VariedBoxWidth(stop, start=1, step=1, position=0, add=None)` Generates models with a single component at `position` of different widths. The widths generated range from `start` to `stop` in steps of `step`. Additionally one or more fixed components can be added to each model by using the `add` parameter.

`SequentialBoxes(stop, step=None, start=0, width=1, n=1, add=None)` This generates models with `n` (pairwise different) components (of width `width`) in sequential order. The `start`, `stop` (max. value) and step-size (`step`) can be defined. Like `VariedBoxWidth`, it accepts an `add` parameter to include more (static) boxes.

`GridBoxes(rangex, rangey, widthx=1, widthy=1, add=None)` `GridBoxes` can be used to generate a set of models with two components each, covering the entire area specified by `rangex` and `rangey`.

Experiments

These are classes to represent the different experiments, along with additional parameters necessary for the analysis (*e. g.* dispersion, transition energy etc.). They also include the information, which preparational steps and which parameterization is used.

Each experiment can be initialized using the `load("params.dat")` function, which expects the filename of a parameter file. This parameter file should contain the following information (one value per line)

1. Transition energy in [keV]
2. Curvature of the Bragg crystal in [m]
3. Distance of the lattice planes in the Bragg crystal in [nm]
4. Mass of the particle (pion / muon) [eV]
5. Atomicity of the nucleus
6. Order of the Bragg reflex

`MuonicHydrogen.load(params_file)` Load an experimental setup for the muonic hydrogen, with the Gaussian width of the spectrometer response as a free parameter. The preparational steps to create the test spectra, which are performed for this experiment are:

1. Calculates the trapezoidal distributions for the n given kinetic energy components, with a 10-fold resolution (compared to the data).

2. Convolves this distributions with a given response function (10-fold resolution as well).

3. Rebins the outputs of the previous step to match the resolution of the data (combines ten bins into one bin).

4. Convolves the re-binned spectra with m Gaussians of different widths, yielding $m \cdot n$ different test spectra.

`MuonicHydrogenFixedGauss.load(params_file)` Similar to `MuonicHydrogen`, but uses a parameterization with a fixed Gaussian width. In the preparation, only $m = 1$ Gaussian is produced, yielding n different test spectra.

`MuonicHydrogenFixedHFS.load(params_file)` Similar to `MuonicHydrogenFixedGauss`, but with one less parameter. The distance between peaks (hyperfine splitting) is fixed to a value (which must be provided later). Thus, the preparational steps are exactly the same as in `MuonicHydrogenFixedGauss`.

`PionicHydrogen.load(params_file)` **and** `PionicDeuterium.load(params_file)` Represent the pionic experiments. The generation of the test spectra is similar to the case of the muonic hydrogen:

1. Calculates the trapezoidal distributions for the n given kinetic energy components, with a 10-fold resolution (compared to the data).

2. Convolves this distributions with a given response function (10-fold resolution as well).

3. Rebins the outputs of the previous step to match the resolution of the data (combines ten bins into one bin).

4. Convolves the re-binned spectra with m Lorentzians of different widths, yielding $m \cdot n$ different test spectra.

`TestSpectra(name, models, experiment, response, gamma_sigma)`

Returns an object which can generate the test spectra needed for the specified `models`. A unique `name` should be specified to identify this set of models. The `experiment` parameter defines how the test spectra are prepared, `response` identifies the filename with the response function and `gamma_sigma` is a 3-tuple `start, stop, step` to specify which Γ (Lorentzian) or σ (Gaussian) curves to generate.

`generate()` Starts the generation of the test spectra by convolving the n Doppler broadenings with the response and m different Lorentzians / Gaussians. This is concludes the preparation (see Figure 4.10).

`start_nested_sampling(data, multi_modal=False, stretch=2.0)` Launches the nested sampling evaluations in several parallel processes, and logs the progress to the terminal. Depending on the experiment/ parameterization and the number of models, this can take from a few minutes to a few days. This function expects a `Data` object (see section A.1.2) as first parameter.

`load_results(data_filename)` Loads the results from a nested sampling run into memory to perform the post-analysis. Returns a `Results` object (see section A.1.2).

`Data(data_filename, lower_cut=0, upper_cut=-1, col=0)`

Reads in the data spectrum given in column nr. `col` in `data_filename`. Optional cuts can be specified via `lower_cut` and `upper_cut`.

`plot()` Plot the data spectrum (*e. g.* to verify the cuts).

`Results`

A `Results` object, as returned by `TestSpectra#load_results`(see A.1.2) , represents the collection of nested sampling `Result` objects. The results are sorted in descending order by their evidences, with the result that the first element of `Results` returns the best result.

`posterior_map()` This creates a huge array containing all posterior samples from all evaluated models, including their weights (model-weight and likelihood-weight), which can be used to create model-averaged histograms and scatterplots. As the `PosteriorMap` can take up up to several gigabytes of memory, it is cached on the disk.

`with_box(a, b)` Returns all results, whose models contain the energy component given by `(a, b)`.

`with_evidence_gt(b)` Returns all results, whose Bayes factors (compared to the best result) are $B \geq b$. Useful to select only the important results.

`PosteriorMap`

See section A.1.2.

`_column(index_or_callable)` Returns a column (parameter projection) of the posterior given by `index_or_callable`, or alternatively if `index_or_callable` is a function, every set of samples is passed to this function and the return values returned as array. This allows the transformation of the samples,*i. e.* to calculate the peak distance and intensity ratio in muonic hydrogen.

`histogram(index_or_array, bins=10, range=None)` Calculates a histogram for the specified column (`index_or_array`), between the limits given by `range` with the number of `bins` specified. As this function uses the NumPy histogram routine, it returns a tuple consisting of `(densities, bin_edges)`.

`sigma1/2/3(index_or_array, avg)` Determine σ credibility intervals relative to the value given by `avg`.

`Plot()`

Creates an object which can be used to plot data series or histograms and save them as self-contained GNUPLOT scripts.

`add_data(values, using=None, _with=None, title=None)` Add a series of y, x-y or x-y-z values to this plot. Can be called several times to create multiplots.

`add_histo(histogram_values)` Add histogram bins and values to this plot.

A.1.3. Shell

The shell is an iPython based console, in which the analysis can be performed interactively. It can be lanuched using the `console.sh` command, located in the `bin`-directory of the software package. This shall provides access to the methods and programs discussed in section A.1.2, as well as the entire Python standard library and the NumPy / SciPy libraries.

Analysis scripts (similar to those presented in section A.2) can then be typed line by line or run within the current context by calling `run_script("myscript.py")`. After execution of the script, the interactive shell returns, and the local variables defined in the script are available for further steps and inspection.

A.1.4. Commmand line tools

The routines described in section A.1.2 are based on a set of command line tools, which perform the CPU-intensive tasks. The input / output is file based, via simple text data files, values separated by any kind of whitespace and rows separated by linefeeds.

All these commands can also be called from within a Python script or the iPython shell with the options passed as keyword arguments.

```
Convolve --input-1 a.dat --input-2 b.dat --output c.dat
```

Convolves the spectra given in `a.dat` and `b.dat` and saves the result to `c.dat`.

```
CompressSpectrum --input a.dat --compress n --output b.dat
```

Compresses (re-bins) the spectrum in `a.dat` by a factor of n and writes the result to `b.dat`. n must be an integer value.

```
ConvolveCurve --input a.dat --output b.dat --type gauss_or_lorentz
--number-of-channels n
```

Convolves the input spectrum `a.dat` with a Gaussian or Lorentzian curve. The number of channels to be evaluated on the curve has to be specified with n. Additional available options are `--position, --volume, --scale, --deviation`, with the two latter only applicable for Lorentzian or Gaussian type respectively.

```
ConvolveTrapezoid --input a.dat --output b.dat --number-of-channels n
--top t --bottom b --height h
```

Convolves the spectrum given by `a.dat` with a trapezoidal distribution (defined by the half lengths of its `top` and `bottom` edges `t, b`, and the `height h`) and saves the result to `b.dat`.

`Superpose --in a.dat b.dat ...--factors n1 n2 ...--out c.dat`

Superposes the spectra given in `a.dat, b.dat ...`, weighted by the factors `n1, n2` and stores the result to `b.dat`.

`GenerateSpectrum --type type --volume I --number-of-channels n`

Generates a spectrum of type `type` (constant, gauss, lorentz or voigt) with the intensity I and the number of channels n. An additional switch `--random` can be set to generate sampled counts instead of evaluated y-values. It takes the following options, depending on the type of curve chosen: `--position, --scale, --deviation`.

`GenerateBoxes --min-value a --max-value b --output c.dat`

Produces a trapezoidal spectrum for the lower and upper bounds given with `a, b` and saves it to `c.dat`.

`NestedSampling`

This command performs one nested sampling evaluation and takes a lot of mandatory arguments:

`--data` Data file to run against.

`--model` File containing the configuration of kinetic energy components for this evaluation. Each line in the file marks a box-like component, marked by an lower and upper limit *e. g.*

```
0   2
24  26
50  52
```

`--box-pattern` A filename pattern to load the test spectra for each kinetic energy component and Γ, σ-value, *e. g.*

```
model_box_width/test_spectra/box_%s_%s/%s.dat
```

would yield a path of

```
model_box_width/test_spectra/box_0_2/600.dat
```

for the test spectra of a component between $0 - 2\,\mathrm{eV}$ with a convolved Lorentzian of $\Gamma = 600\,\mathrm{meV}$ width.

`--gamma-start` Lowest Γ, σ-value.

`--gamma-stop` Highest Γ, σ-value.

`--gamma-step` Distance between different Γ, σ-values.

`--number-of-channels` Number of channels in the test spectra.

`--lower-cut` Lower channel limit of the data.

`--upper-cut` Upper channel limit.

`--out` Output directory for the result files.

`--type` Type of parameterization to use (*e. g.* muH, piH, piD, muH_hfs, muH_fixed).

Additional to these, the behavior of the nested sampling algorithm can be modified using the following options:

`--live-points 500` Number of live points to use.

`--max-iterations 1000000` Maximal number of allowed iterations before cancel.

`--contraction 30` Interval after that the elliptic sampler is re-determined.

`--stretch 2` Additional scale for the sampling ellipsoid.

`--precision 2` Desired precision of the evidence in $\log_{10}$ scale, so the default of two means precise down to 0.01.

`--seed 0` Seed for the random number generator. Must be specified, if more than one evaluation is performed, to prevent correlations.

`--multi-modal 0` Enable or disable the clustering sampler. Disabled by default for performance reasons.

`--max-clusters 20` Specify a maximal number of allowed clusters (only applicable if `--multi-modal`) is set.

`--min-acceptance 0.5` Minimal fraction of the expected acceptance ratio, before a new cluster is assumed. Also only applicable if the clustering sampler is enabled.

`EvalFit`

In order to apply a set of parameters to a line shape and plot this line shape long with the data, `EvalFit` can be used. It takes mostly the same mandatory arguments as `NestedSampling` plus:

`--params p1 p2 p3 ...` Parameter values to apply, separated by spaces

`--out line_shape.dat` File to save the calculated line shape to.

A.2. Analysis scripts

A.2.1. Introduction

In this appendix, an example script for each of the analysis steps performed is presented. The scripts are a series of Python commands, which can be used interactively via the iPython terminal, or run from a file. These commands use the routines and programs described in (see section A.1).

Most of these scripts are examples from the πH analysis, with the exception of the hyperfine splitting scripts from the μH studies. The steps to create sets of models and run the nested sampling evaluations are of course independent of the experiment.

The first step is always to load an experiment and its configurational parameters. These are transition energy, curvature (crystal), lattice constant (crystal), particle mass, atomicity of the nucleus, order of the Bragg diffraction and should be present in a simple textfile *params.dat* with one value per line.

A.2.2. Width of low-energy component

```
#  Load the experimental parameters (pion-mass, lattice constant, etc.)
m = PionicHydrogen.load("params.dat")

#  Generate 1-component models of different widths
#  (starting at 0.25 eV up to 50 eV in
b = VariedWidthBoxes(30, start=0.25, step=0.25, position=0)

#  Prepare the test spectra for the models b and experiment m
#  collected under the name "box_width". "response.dat" contains
#  the response function, (200, 600, 5) denotes the Γ values
#  (from 200 eV to 600 eV in steps of 5 eV) for the
#  convolution with the Lorentzian.
t = TestSpectra("box_width", b, m, "response.dat", (200, 600, 5))

#  Generates the trapezoidal functions and convolves them with the
#  response and the Lorentzians. These test spectra can be found
#  in ‘model_box_width/‘
t.generate()

#  Load a datafile, but use only the bins 620 – 1180 and the values
#  in column 1
d = Data("piH.dat", 620, 1180, col=1)

#  Start the actual nested sampling processes. This produces a lot
#  of output and takes some time. The results are saved
#  in ‘results_box_width/‘
t.start_nested_sampling(d)
```

```
#  After finishing the evaluation of all models, load the results
r = t.load_results("piH.dat")

#  The results are ordered by evidence in descending order, so r[0]
#  is the best result. Accessing r[0] gives a quick overview.
r[0]
# => NestedSampling result: results_box_width/piH2-1.dat/0000/0048
#    evidence: 211549.067207
#    weight: 0.031273311867
#    model: [(0.0, 12.0)]
#    params: [887.9870603279975, 469.1270536486451, 44251.54923095699,
#             9.217557222163887]

#  Get an x, y list with x being the width of the first component,
#  and y being the Bayes factor ln(B)
ev = map(lambda x: (x.model[0][1], x.evidence() - r[0].evidence()), r)

#  Create a GNUPLOT with this data
p = Plot()
p.add_data(ev)

#  Save this plot to a self-contained file (including the data)
p.save("piH_box_width_evidences.gp")
```

A.2.3. Looking for a high-energy component

```
m = PionicHydrogen.load("params.dat")

#  Generate 2-component models of different widths (0.5 eV
#  to 30 eV) of the low-energy component, and different positions
#  (0 eV to 200 eV) of the high-energy comp.
b = []
for i in range(200):
  b.extend(VariedWidthBoxes(30, start=0.5, step=0.5, position=0, \
      add=[(i, i + 2)])

t = TestSpectra("2box_width", b, m, "response.dat", (200, 600, 5))
t.generate()
d = Data("piH.dat", 620, 1180, col=1)
t.start_nested_sampling(d)
r = t.load_results("piH.dat")

#  Get an x, y, z list with x being the width of the first component,
```

```
#  and y the position of the high-energy contribution, and z the
#  Bayes factor
ev = map(lambda x: (x.model[0][1], x.model[1][0], x.evidence() - \
    r[0].evidence()), r)

#  The list has to be sorted for GNUPLOT
ev = sorted(ev)

#  Create scatter plot (with image)
p = Plot()
p.add_data(ev, _with='image')

p.save("piH_2box_width_evidences.gp")
```

A.2.4. Smoothing / re-binning

```
#  Given an x, y, z list ev (e.g. from previous script),
#  create an ImagePlot
im = ImagePlot(ev, (0.5, 30, 0.5), (0, 200, 1))

#  Get the z values as pixel-matrix (bitmap)
m = im.matrix()
#  Apply a bivariate Gaussian filter, σ is in pixels.
#  As this is a convolution with a finite filter, 4·σ pixels
#  at the edges are lost and should be discarded.
im.from_matrix(gaussian_filter(m, sigma=1.5, mode='nearest'))

p = im.plot()
p.save("piH_2box_width_evidences_smoothened.gp")
```

A.2.5. Searching for additional high-energy contributions

```
m = PionicHydrogen.load("params.dat")

#  Generate a set of 3-component models with a low-energy
#  comp. fixed at 0 − 1 eV and two variable high-energy
#  components between 0 eV and 200 eV.
b = SequentialBoxes(200, step=1, width=2, n=2, add=[(0, 1)])

t = TestSpectra("3box", b, m, "response.dat", (200, 600, 5))
t.generate()
d = Data("piH.dat", 620, 1180, col=1)
t.start_nested_sampling(d)
r = t.load_results("piH.dat")
```

```
#  Get an x, y, z list with x being the position of the third
#  component, and y the position of the second component, and z
#  the Bayes factor
ev = map(lambda x: (x.model[1][0], x.model[0][0], x.evidence() - \
    r[0].evidence()), r)

#  The list has to be sorted for GNUPLOT
ev = sorted(ev)

#  Create scatterplot (with image)
p = Plot()
p.add_data(ev, _with='image')

p.save("piH_3box_evidences.gp")
```

A.2.6. Γ_{1s} posterior projection

```
#  Given a set of results r
r = t.load_results("piH.dat")

#  Select all results with Bayes factor >= -5
r5 = r.with_evidence_gt(5)

#  Load all samples with their likelihood and evidence weights
pm = r5.posterior()

#  The Γ value has to be doubled to meet the conventions
gammas = pm._column(lambda p: p[2] * 2.0)

#  Create a histogram of the samples
h = pm.histogram(gammas, bins=80, range=(750, 1000))

#  Plot the histogram
p = Plot()
p.add_histo(h)

p.save("piH_gamma_histo.gp")

#  Determine maximal value gamma_m (e.g.\ parabola fit with
#  GNUPLOT on the 10 highest bins)
gamma_m = 878

#  Calculate credibility intervals
```

```
pm.sigma1(gammas, avg=gamma_m)
pm.sigma2(gammas, avg=gamma_m)
pm.sigma3(gammas, avg=gamma_m)
```

A.2.7. Posterior projection of the low-energy intensity

```
#  Given a set of results r and loaded samples pm
r = t.load_results("piH.dat")
r5 = r.with_evidence_gt(5)
pm = r5.posterior()

#  The parameter index of the rel. intensity of the first box is
#  6 in the case of the πH 2-comp. analysis.
int_first = pm._column(lambda p: p[6] * 100)

h = pm.histogram(int_first, bins=200, range=(60, 100))
p = Plot()
p.add_histo(h)
p.save("piH_first_box_histo.gp")

int_first_m = 85
pm.sigma1(int_first, avg=int_first_m)
pm.sigma2(int_first, avg=int_first_m)
pm.sigma3(int_first, avg=int_first_m)
```

A.2.8. Hyperfine splitting and triplet / singlet population

```
m = MuonicHydrogenFixedGauss.load("params.dat")
b = SequentialBoxes(150, step=1, width=2, n=2, add=[(0, 4)])

#  Prepare test spectra with a Gaussian component with
#  σ = 27 meV, which corresponds to a FWHM of
#  76 μrad
t = TestSpectra("3box", b, m, "response.dat", (27, 27, 1))
t.generate()
d = Data("muH.dat", 620, 1180)

#  Start nested sampling, use multi-modal (clustering) sampler,
#  and a smaller ellipsoid scale of 1.6 (default is 2.0)
t.start_nested_sampling(d, multi_modal=True, stretch=1.6)

#  Given a set of results r
r = t.load_results("muH.dat")
```

```
#  Select all results with Bayes factor >= -5
r5 = r.with_evidence_gt(5)

#  Load all samples with their likelihood and evidence weights
pm = r5.posterior()

#  Transform the samples to get samples for the triplet / singlet
#  population intensity ratio and the hyperfine splitting distance
hfs_int = pm._column(m.hfs_int)
hfs_diff = pm._column(m.hfs_diff)

#  Create a two-dimensional histogram, this is a numpy routine
#  which returns the x-edges (h[1]), y-edges (h[2]) and the data
#  as two-dimensional matrix (h[0])
h = numpy.histogram2d(hfs_diff, hfs_int, bins=100,
    range=((170, 250), (1, 7)))

#  Convert h to a x, y, z list
xyz = []
for i in range(100):
  for j in range(100):
    xyz.append(h[1][i], h[2][j], h[0][i, j])

p = Plot()
p.add_data(xyz, _with='image')

p.save("muH_hfs.gp")
```

A.2.9. Theoretical kinetic energy distribution

```
#  Read the theoretical distribution in from a file.
#  Column 2 contains the densities for the 3p-1s line.
th = TheoSpectrum("e_dist_mup_12.5bar_02.2012-1.dat", 2)

#  Load an experiment
m = MuonicHydrogenFixedGauss.load("params.dat")

#  Generate the test spectrum
t = th.to_test_spectra("theo_spec", m, "response.dat", (27, 27, 1))

#  Load data file and start nested sampling
d = Data("muH.dat", 620, 1180)
t.start_nested_sampling(d, multi_modal=true, stretch=1.6)
```

```
#  Load results
r = t.load_results("muH.dat")
```

A.2.10. Deconvolution

```
m = MuonicHydrogenFixedGauss.load("params.dat")

#  Generate 20825 different models, based on 3 components of
#  width 3 eV each.
b = SequentialBoxes(150, step=3, width=3, n=3)

t = TestSpectra("deconvolution", b, m, "response.dat", (27, 27, 1))
t.generate()
d = Data("muH.dat", 620, 1180, col=1)
t.start_nested_sampling(d, multi_modal=True, stretch=1.6)
r = t.load_results("muH.dat")

#  Calculate the deconvolution by model-averaging. 5 denotes the
#  index of the first intensity parameter.
dc = Deconvolution(r, 5)

#  Access the spectrum, the first two values indicate the bin limits,
#  the third the rel. intensity and the forth the integrated intensity
# => [[0.0, 3.0, 0.605936448365571, 0.605936448365571],
#     [3.0, 6.0, 0.002815332253710746, 0.6087517806192818],
#     [6.0, 9.0, 0.004918451401902636, 0.6136702320211844],
#     [9.0, 12.0, 0.006262049397827175, 0.6199322814190116],
#      ...
#    ]

#  Create a plot
p = dc.plot()
p.save("muH_deconvolution.gp")
```

Bibliography

[1] A. Hirtl. *Determination of the Strong Interaction Ground State Width in Pionic Hydrogen*. PhD thesis, Technische Universität Wien, Vienna, 2008.

[2] D. Gotta, F. Amaro, D. F. Anagnostopoulos, S. Biri, D. S. Covita, H. Gorke, A. Gruber, M. Hennebach, A. Hirtl, T. Ishiwatari, P. Indelicato, Th. Jensen, E.-O.Le Bigot, J. Marton, M. Nekipelov, J. M. F. dos Santos, S. Schlesser, Ph. Schmid, L. M. Simons, Th. Strauch, M. Trassinelli, J. F. C. A. Veloso, and J. Zmeskal. Pionic hydrogen. 745:165–186, 2008.

[3] T. Strauch. *High-Precision Measurement of Strong-Interaction Effects in Pionic Deuterium*. PhD thesis, Universität zu Köln, 2009.

[4] Th. Strauch, F.D. Amaro, D.F. Anagnostopoulos, P. Bühler, D.S. Covita, H. Gorke, D. Gotta, A. Gruber, A. Hirtl, P. Indelicato, E.-O. Bigot, M. Nekipelov, J.M.F. Santos, Ph. Schmid, S. Schlesser, L.M. Simons, M. Trassinelli, J.F.C.A. Veloso, and J. Zmeskal. Pionic deuterium. *The European Physical Journal A*, 47(7):1–19, 2011.

[5] Th. Strauch, F. D. Amaro, D. F. Anagnostopoulos, P. Bühler, D. S. Covita, H. Gorke, D. Gotta, A. Gruber, A. Hirtl, P. Indelicato, E.-O. Le Bigot, M. Nekipelov, J. M. F. dos Santos, S. Schlesser, Ph. Schmid, L. M. Simons, M. Trassinelli, J. F. C. A. Veloso, and J. Zmeskal. Precision determination of the $d\pi \leftrightarrow nn$ transition strength at threshold. *Phys. Rev. Lett.*, 104:142503, Apr 2010.

[6] D.S. Covita. *High Precision Spectroscopy of the 3p – 1s X-ray Transition in Muonic Hydrogen*. PhD thesis, University of Coimbra, 2008.

[7] D.S. Covita, D.F. Anagnostopoulos, H. Gorke, D. Gotta, A. Gruber, A. Hirtl, T. Ishiwatari, P. Indelicato, E.-O. Bigot, M. Nekipelov, J.M.F. Santos, Ph. Schmid, L.M. Simons, M. Trassinelli, J.F.C.A. Veloso, and J. Zmeskal. Line shape of the μh(3p - 1s) transition. *Hyperfine Interactions*, 193(1-3):61–67, 2009.

[8] D.S. Covita, D.F. Anagnostopoulos, H. Gorke, D. Gotta, A. Gruber, A. Hirtl, T. Ishiwatari, P. Indelicato, E.-O. Bigot, M. Nekipelov, J.M.F. Santos, Ph. Schmid, L.M. Simons, M. Trassinelli, J.F.C.A. Veloso, and J. Zmeskal. Line shape of the μh(3p - 1s) transition. *Phys. Rev. Lett.*, 102, 2009.

[9] D. Gotta. Precision spectroscopy of light exotic atoms. *Progress in Particle and Nuclear Physics*, 52(1):133 – 195, 2004.

[10] James S. Cohen. Isotope effects on antiproton and muon capture by hydrogen and deuterium atoms and molecules. *Phys. Rev. A*, 59:1160–1169, Feb 1999.

[11] T.S. Jensen and V.E. Markushin. Scattering of light exotic atoms in excited states. *Hyperfine Interactions*, 138(1-4):113–116, 2001.

[12] T.S. Jensen and V.E. Markushin. Collisional deexcitation of exotic hydrogen atoms in highly excited states. *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics*, 21(3):271–283, 2002.

[13] V.E. Markushin. Cascade processes in exotic atoms with z = 1. *Proceedings of Physics of Exotic Atoms on Electromagnetic Cascade and Chemistry, Erice, Italy, 1989*, pages 73–96, 1990.

[14] M. Leon and H. A. Bethe. Negative meson absorption in liquid hydrogen. *Phys. Rev.*, 127:636–647, Jul 1962.

[15] T.S. Jensen and V.E. Markushin. Collisional deexcitation of exotic hydrogen atoms in highly excited states. *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics*, 21(3):271–283, 2002.

[16] L. Bracci and G. Fiorentini. Coulomb de-excitation of mesic hydrogen. *Il Nuovo Cimento A*, 43(1):9–30, 1978.

[17] T. S. Jensen, V. P. Popov, and V. N. Pomerantsev. Atomic cascade in muonic and hadronic hydrogen atoms, 2007.

[18] R. N. Faustov and A. P. Martynenko. Muonic hydrogen ground state hyperfine splitting. *J.Exp.Theor.Phys. 98 (2004) 39-52*, 2003.

[19] S. Deser, M. L. Goldberger, K. Baumann, and W. Thirring. Energy level displacements in pi-mesonic atoms. *Phys. Rev.*, 96:774–776, Nov 1954.

[20] G. Rasche and W.S. Woolcock. Connection between low-energy scattering parameters and energy shifts for pionic hydrogen. *Nuclear Physics A*, 381(3):405 – 418, 1982.

[21] J. Spuller, D. Berghofer, M.D. Hasinoff, R. Macdonald, D.F. Measday, M. Salomon, T. Suzuki, J.-M. Poutissou, R. Poutissou, and J.K.P. Lee. A remeasurement of the panofsky ratio. *Physics Letters B*, 67(4):479 – 482, 1977.

[22] T.L. Trueman. Energy level shifts in atomic states of strongly-interacting particles. *Nuclear Physics*, 26(1):57 – 67, 1961.

[23] Psi proposal r-98-01. URL: http://http://collaborations.fz-juelich.de/ikp/exotic-atoms/.

[24] pie5 beam line. URL: http://aea.web.psi.ch/beam2lines/beam_pie5.html.

[25] V. Popov and V. Pomerantsev. private communication.

[26] M. Sanchez del Rio and R.J. Dejus. Xop 2.1: A new version of the x-ray optics software toolkit. *AIP Conf. Proc*, (705):784–787.

[27] L. Simons. private communication.

[28] D.F. Anagnostopoulos, S. Biri, D. Gotta, A. Gruber, P. Indelicato, B. Leoni, H. Fuhrmann, L.M. Simons, L. Stingelin, A. Wasser, and J. Zmeskal. On the characterisation of a bragg spectrometer with x-rays from an ecr source. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 545(1–2):217 – 224, 2005.

[29] M Trassinelli, S Boucard, D S Covita, D Gotta, A Hirtl, P Indelicato, É-O Le Bigot, J M F dos Santos, L M Simons, L Stingelin, J F C A Veloso, A Wasser, and J Zmeskal. He-like argon, chlorine and sulfur spectra measurement from an electron cyclotron resonance ion trap. *Journal of Physics: Conference Series*, 58(1):129, 2007.

[30] D.S. Sivia and J. Skilling. *Data Analysis: A Bayesian Tutorial*. Oxford University Press, USA, 2006.

[31] U.C. Bergmann and K. Riisager. A systematic error in maximum likelihood fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 489(1–3):444 – 447, 2002.

[32] R.D. Cousins. Why isn't every physicist a bayesian? *American Journal of Physics*, 63(5):398–410, 1995.

[33] Roberto Trotta. Applications of bayesian model selection to cosmological parameters. *Monthly Notices of the Royal Astronomical Society*, 378(1):72–82, 2007.

[34] J. Harold. *The Theory of Probability*. Oxford University Press, USA, 3rd edition, 1961.

[35] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[36] J. Skilling. Nested sampling for general bayesian computation. *Bayesian Analysis*, 1(4):833–859, 2006.

[37] F. Feroz and M.P. Hobson. Multimodal nested sampling: an efficient and robust alternative to markov chain monte carlo methods for astronomical data analyses. *Monthly Notices of the Royal Astronomical Society*, 384(2):449–463, 2008.

[38] M. P. Hobson and C. McLachlan. A bayesian approach to discrete object detection in astronomical data sets. *Monthly Notices of the Royal Astronomical Society*, 338(3):765–784, 2003.

[39] G. Marsaglia. Choosing a point from the surface of a sphere. *The Annals of Mathematical Statistics*, 43(2):645–646, 1972.

[40] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. 1:281–297, 1967.

[41] J. R. Shaw, M. Bridges, and M. P. Hobson. Efficient Bayesian inference for multimodal problems in cosmology. August 2007.

[42] IEEE Task P754. *ANSI/IEEE 754-1985, Standard for Binary Floating-Point Arithmetic*. August 1985. Revised 1990. A preliminary draft was published in the January 1980 issue of IEEE Computer. Also standardized as *IEC 60559 (1989-01) Binary floating-point arithmetic for microprocessor systems*.

[43] D. Gotta. private communication.

[44] D.F. Anagnostopoulos, R. Sharon, D. Gotta, and M. Deutsch. Kα and kβ x-ray emission spectra of metallic scandium. *Phys. Rev. A*, 60:2018, 1999.

[45] D. Gotta, K. Rashid, B. Fricke, P. Indelicato, and L. Simons. X-ray transitions from antiprotonic noble gases. *Eur. Phys. J. D*, 47:11, 2008.