

Master Thesis

Statistical Methods for Electric Dipole Moment Measurements at the Cooler Synchrotron COSY

vorgelegt der

Fakultät für Mathematik, Informatik und Naturwissenschaften
RWTH Aachen

angefertigt am

III. Physikalischen Institut B
Prof. Dr. Jörg Pretz

von

Valentin Tempel

Aachen, 01. Oktober 2024

Erstgutachter und Betreuer

Prof. Dr. Jörg Pretz
III. Physikalisches Institut B
RWTH Aachen University

Zweitgutachter

Prof. Dr. Achim Stahl
III. Physikalisches Institut B
RWTH Aachen University

Contents

1. Introduction	1
1.1. CP Violation and Baryon Asymmetry	1
1.2. Electric Dipole Moments	3
1.3. EDM and CP violation	5
1.4. EDMs in the Standard Model	6
1.5. Charged Particle EDM Experiments in Storage Rings	7
1.5.1. Thomas-BMT Equation	7
1.5.2. Magnetic Storage Rings and RF Wien Filter Method	8
2. Cooler Synchrotron	10
2.1. Important COSY Storage Ring Components	10
2.2. Polarimeters	12
2.3. Polarization Determination	14
2.3.1. Turn Number and Time in Cycle	14
2.3.2. Polarimeter Event Rates	14
2.3.3. Left-Right Asymmetry (Vertical Polarization)	15
2.4. Important Quantities of the Data Taking Procedure	17
3. Fundamental Statistic Concepts	19
3.1. Central Limit Theorem	19
3.2. Bayes Theorem	19
3.3. Estimators	20
3.3.1. Variance, Mean Squared Error and Bias	20
3.3.2. Fisher Information and Cramer-Rao Bound	21
3.3.3. Properties of Estimators	22
3.4. Maximum Likelihood Method	23
4. Feldman-Cousins Algorithm	25
4.1. Poisson Distribution with Background	25
4.2. The Downward Fluctuation Discussion	29
5. Parameter Estimation in EDM experiments at COSY	32
5.1. In-Plane Spin Precession and Horizontal Polarization	32
5.2. Up-Down Asymmetry and the Problem of Sine-Fitting	32
5.3. Amplitude Determination and Bias	34
5.4. Statistical Uncertainty of A and B	35
5.5. Three-parameter fit for amplitude estimation	38
5.6. Rice Distribution	39
5.7. Construction of Confidence Intervals	44
5.8. Derivation of the Posterior Probability Density Function	45
6. Turn Fourier Transform	50
6.1. Discrete Fourier Transform	50
6.2. Analytic Fourier Amplitude Model	51
6.3. Fourier Transform Spectrum Simulations	53
6.4. Spin Tune Estimator	57

7. Spin Polarization Decay and Spin Coherence Time	60
7.1. Simple Exponential Decay Model	60
7.2. SCT Results for Precursor Data	64
7.3. Maximum Likelihood Fit	69
8. Beam Oscillation Amplitude	72
9. Conclusion	78
A. Calculations and Proofs	80
A.1. Proof of the MSE decomposition	80
A.2. Fourier Transform of a Sine and Cosine Function	80
A.3. Fourier Transform of the Rectangular Function	81
B. Likelihood Method - Variance of Estimators	82
B.1. Likelihood Method for A and B	82
B.2. Likelihood Method for Amplitude and Phase	84
C. Derivation of the Rice Distribution	85
C.1. Transition from a Rice Distribution to a Gaussian	86
C.2. Bayes-Theorem for Gaussian with Constant Prior	86
D. Mapping Method	87
D.1. Asymmetries from Count Sums and Differences	88
D.2. Precise Spin Tune from Polynomial Fits	90
Bibliography	91

1. Introduction

The Jülich Electric Dipole moment Investigation (JEDI) collaboration at Forschungszentrum Jülich, Germany aims to directly measure the electric dipole moment (EDM) of charged particles – notably deuterons in this case – for the first time using a storage ring, the **Cooler Synchrotron** COSY. According to the Standard Model (SM) of particle physics, these EDMs arise from the violation of charge-conjugation-parity (CP) symmetry.

Since the EDMs predicted by the SM are very small, a measurement of a nonzero electric dipole moment would imply a new source of CP violation, provided the well established CPT-theorem (charge-parity-time symmetry) holds. Therefore, successfully measuring an EDM could provide evidence for physics beyond the Standard Model. The EDMs predicted by the SM are beyond current experimental detection capabilities.

A statistical sensitivity on the order of $\mathcal{O}(10^{-24})\text{e cm}$ is anticipated in the initial phase of the experiment. Subsequently, a proposed dedicated all-electric storage ring could further enhance the sensitivity. In the first stage, the EDM is measured by a signal buildup using a device called Radio Frequency Wien Filter. To achieve the desired sensitivity, it is crucial to minimize systematic effects that could generate false EDM signals. This master thesis written at the chair of the III. Physikalisches Institut B of RWTH Aachen University examines a bias effect introduced by the analysis procedure of experimental data from the polarimeters using statistical methods. Bayesian and Frequentist probability theory is applied to regions of small signal strength, where typical Gaussian uncertainty assumptions are misleading and can be challenged. The improved methods are compared to conventional approaches for quantities relevant for the quality of EDM measurements like signal buildup and Spin Coherence Time.

One goal of the JEDI collaboration measurements is to deepen our understanding of the matter-antimatter asymmetry in the universe, which is closely related to CP violation, through fundamental research in nuclear and particle physics. Resolving the puzzle of this observed asymmetry would be a key to understanding our very existence.

1.1. CP Violation and Baryon Asymmetry

If the charge-parity-time (CPT) theorem holds, the existence of permanent EDMs in elementary particles would imply CP violation, as such moments are generated by processes that violate time reversal symmetry. According to one of the Sakharov conditions, CP violation is essential for generating the matter-antimatter asymmetry in the universe. The search for their violation has been the focus of numerous experiments. The three symmetries can briefly be described as follows [1]:

- **Charge Conjugation:** All particles involved in a process are converted into their antiparticles under charge conjugation transformation.
- **Parity:** The physical process occurs identically under the transformation $\vec{x} \rightarrow -\vec{x}$. However, this is more of a point reflection than a literal mirror image.
- **Time Reversal:** Symmetry of physical laws which are independent of the direction of the time $t \rightarrow -t$.

The first violation of the CP symmetry was observed in the kaon sector in 1964, as indirect¹ CP violation in the Fitch-Cronin Experiment [2], leading to a Nobel prize in 1980. Kaons can transform into their antiparticles and vice versa, but the probability for such a transformation is not exactly the same in both directions [3].

Starting in the 1990s also direct CP violation was observed in again Kaon decay processes, with the more recent additions in the B meson sector (B Factories in the 2000s [4] and LHCb 2013 [5]) and also for D Mesons (LHCb 2019 [6]). All this processes relate to CP violation in weak interactions, since CP violation in the strong sector is extremely small or completely missing (strong CP problem).

For the weak sector the CKM-Matrix (Cabibbo-Kobayashi-Maskawa) has been introduced to explain the observed CP violation. In the Standard Model case with three generations of quarks, there are three mixing angles and one CP-violating complex phase. However, the CP violation obtained from the CKM mechanism alone is insufficient to explain baryogenesis i.e. the unexpectedly large number of baryons in the universe. Therefore other sources of CP violation are extensively studied, e.g. the neutrino sector.

There the PMNS-Matrix (Pontecorvo-Maki-Nakagawa-Sakata) has possibly also a CP-violating phase, but this is not certain yet, since the phase δ_{CP} is still compatible with 180° . Such a phase would lead to a zero Jarlskog invariant, which is typically used as a unique measure of CP violation and can be determined comparatively precisely to $J \approx 0.000\,030\,8(15)$ for the quark sector [7]. Since the Jarlskog invariant can be parameterized as $J_{\text{CP}} = J_{\text{CP}}^{\text{max}} \sin(\delta_{\text{CP}})$, the total amount of CP violation even in the case of some existing complex phase also depends on the value of $J_{\text{CP}}^{\text{max}}$. The CP violation in the leptonic sector is of ongoing interest, but notoriously difficult to measure [8]. Another compelling candidate, where additional CP violation could show up, is the permanent EDM of elementary particles.

The baryon asymmetry is one of the unresolved mysteries in cosmology. Given the lack of evidence for primordial antimatter, the excess of matter is quantified by the baryon-asymmetry-to-photon density ratio

$$\eta = \frac{n_B - \overbrace{n_{\bar{B}}}^{\approx 0}}{n_\gamma} = \frac{n_B}{n_\gamma} = 6.16(15) \times 10^{-10} \quad (1.1)$$

where n_B and $n_{\bar{B}}$ correspond to the baryon and anti-baryon density [9]. $n_{\bar{B}}$ is very close to zero, since measurements like the AMS-02 experiment at the ISS (since 2011) see no leftover or in general no heavy antimatter nucleons; only some anti-protons and possibly very few $\mathcal{O}(10)$ anti-helium events [10]. Also antimatter regions or their annihilation radiation were not found in the universe surrounding us.

In other word Equation 1.1 means that there are approximately 1.6 billion photons for each baryon in the Universe. The Standard Model of Cosmology predicts an asymmetry in the order of $\eta_{\text{SM}} = 10 \times 10^{-18}$, which is roughly 8 orders of magnitude below the observed one [11].

Quantitative measurements of the baryon asymmetry and other cosmological parameters were realized by satellite and space probe experiments such as the Cosmic Background Explorer (COBE, 1989-1993) and its successors Wilkinson Microwave Anisotropy Probe

¹Indirect CP violation involves a difference in the oscillation between a particle and the corresponding antiparticle. Direct CP violation is seen in differences in the decay rates of a particle and its antiparticle.

(WAMP, 2001-2010 [12]) and Planck Surveyor (2009-2013 [13]), which all determined the angular distribution and the power spectrum of temperature fluctuations of the cosmic microwave background (CMB) with increasing precision.

According to current theories, the same amount of matter and antimatter was produced from the Big Bang. In 1967 Sakharov introduced three criteria to allow the generation of a baryon asymmetry [14].

- **Baryon number B violation:** Obviously, baryon-antibaryon asymmetry can only be created after the Big Bang in a process, which violates the baryon number conservation.
- **C and CP violation:** To create an imbalance in the production of baryons and anti-baryons, the violation of charge conjugation symmetry (C) and charge-parity transformation symmetry (CP) is required. Otherwise the processes in question could always be compensated for.
- **Universe not in thermal equilibrium:** Particle reactions increasing B would occur at the same rate as the corresponding inverse processes in thermal equilibrium. Therefore, to generate asymmetry, these processes must occur in conditions deviating from thermal equilibrium, to avoid this last compensation effect, which would be ensured by CPT symmetry.

1.2. Electric Dipole Moments

The classical definition of an EDM describes the charge separation of the centers of gravity of positive and negative charges as a vector property

$$\vec{d}_{\text{EDM}} = \int_V \vec{x} \cdot \rho(\vec{x}) \, dx^3, \quad (1.2)$$

with the charge density $\rho(\vec{x})$. The magnetic dipole moment with the current density $\vec{j}(\vec{x})$ is given as

$$\vec{\mu}_{\text{MDM}} = \frac{1}{2} \int_V \vec{x} \times \vec{j}(\vec{x}) \, dx^3. \quad (1.3)$$

The EDMs and MDMs are fundamental properties of the particles and are both aligned parallel or anti-parallel with respect to the spin $\vec{S} = \hbar\vec{s}$:

$$\vec{d} = \eta_{\text{EDM}} \frac{q\hbar}{2mc} \frac{\vec{S}}{\hbar}, \quad (1.4)$$

$$\vec{\mu} = g \frac{q\hbar}{2m} \frac{\vec{S}}{\hbar}, \quad (1.5)$$

where m indicates the mass of the particle and q is the charge (typically elementary charge $\pm e$ for all particles discussed in this work). The spin g -factor and the parameter η_{EDM} are dimensionless quantities. The structure of the equations for the EDM and MDM only differs by a factor $1/c$. Note that for baryons or nucleons the MDM is often given in another form, which does *not* directly include the particle mass in the definition, but instead uses the proton mass m_p and elementary charge e relating to the properties of the proton. This is done because Equation 1.4 does not work for neutral particles like the

neutron which have no charge, where q and therefore also g could then not be defined. Using \vec{I} , which is the spin angular momentum of composite baryonic particles, one gets

$$\vec{\mu}_{m_p} = g_{m_p} \frac{e\hbar}{2m_p} \frac{\vec{I}}{\hbar} = g_{m_p} \mu_N \frac{\vec{I}}{\hbar} \quad (\text{for baryons or nucleus}). \quad (1.6)$$

The magnetic dipole moment is typically expressed in terms of the Bohr magneton μ_B (for leptons) or the nuclear magneton μ_N (for hadrons), which differ by the ratio $m_e/m_p \approx 5.45 \times 10^{-3}$.

$$\mu_B = \frac{e\hbar}{2m_e} = 5.788\,381\,798\,2(18) \times 10^{-5} \frac{\text{eV}}{\text{T}} \quad \mu_N = \frac{e\hbar}{2m_p} = 3.152\,451\,254\,17(98) \times 10^{-8} \frac{\text{eV}}{\text{T}}$$

particle	$ \vec{I} $ or $ \vec{S} $	rest mass in GeV	μ in μ_N or μ_B	g -factor
proton	$\hbar/2$	0.938 272 089 43(29)	+2.792 847 344 63(82)	+5.585 694 689 3(16)
neutron	$\hbar/2$	0.939 565 421 94(48)	-1.913 042 76(45)	-3.826 085 52(90)
deuteron	\hbar	1.875 612 945 00(58)	+0.857 438 233 5(22)	+0.857 438 233 5(22)
electron	$\hbar/2$	0.000 510 998 950 69(16)	-1.001 159 652 180 46(18)	-2.002 319 304 360 92(36)
muon	$\hbar/2$	0.105 658 375 5(23)	-0.004 841 970 48(11)	-2.002 331 841 23(82)

Table 1: Magnetic properties of the composite particles proton, neutron and deuteron, and for the leptonic Dirac particles (electron and muon) compiled from the constants given in the NIST Reference [15].

Table 1 lists the magnetic properties of typical particle candidates for EDM measurements. The rest mass is given in units of energy and the spin in terms of the Planck constant. Note that the g -factor for composite particles is simply the inverse of the spin quantum number ($s = \frac{1}{2}$ or $s = 1$) times the MDM measured in units of the nuclear magneton, if the typical g_{m_p} convention is used (as done e.g. in the NIST database [15]). However this convention has to be considered misleading, when working again with formulas involving the particle mass, as described later. The MDMs of leptonic particles are predominantly influenced by the mass ratios (as can be seen for the muon and tau), since $g_\mu \approx g_e$. Their spin g -factors are close to the prediction of the Dirac equation $g = 2$ for point like spin- $\frac{1}{2}$ -particles.

The anomalous gyromagnetic G -factor is defined as

$$G \equiv a \equiv \frac{g-2}{2} \quad a_{e,\mu,\tau} \approx \frac{\alpha}{2\pi} \approx 0.001\,16 \quad (1.7)$$

where the variable G is commonly used for hadrons and a for leptons. The magnetic moment g -factors g_{m_p} as defined in the equation Equation 1.6 based on the proton mass, cannot be used directly in the Thomas-BMT equation, which describes spin physics in electromagnetic fields (see subsection 1.5.1 for details). These g_{m_p} obtained from literature have to be corrected by the ratio of the particle mass relative to the proton mass. This comes from the fact, that indeed the particle mass is relevant in the derivation

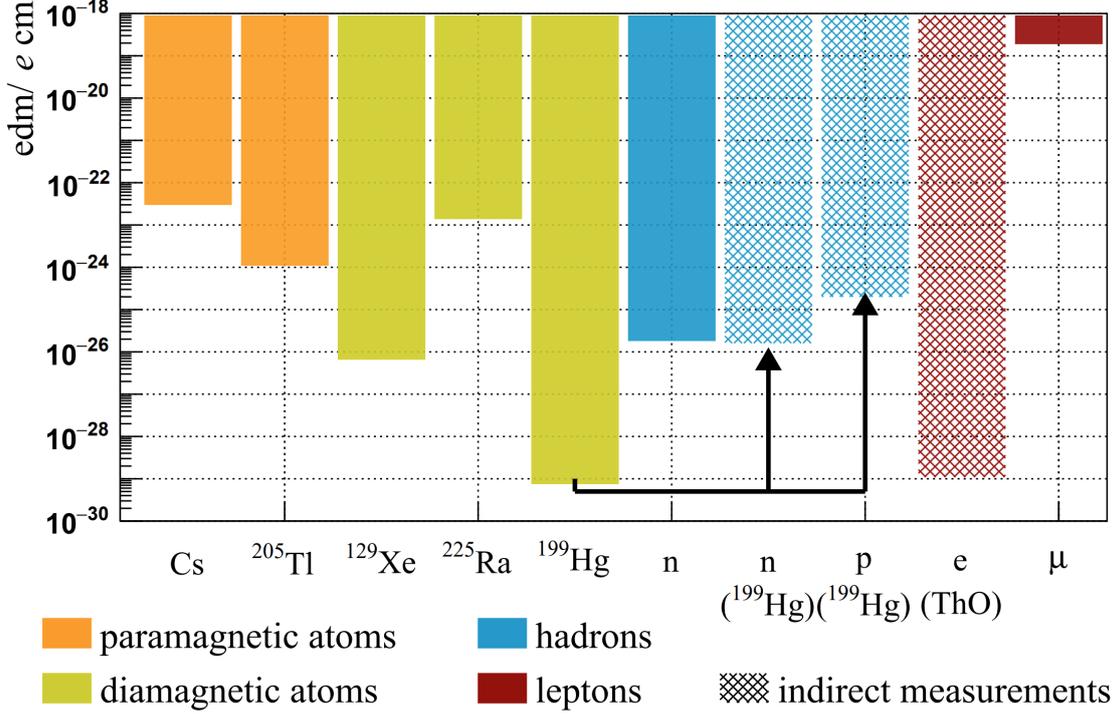


Figure 1.1: Existing limits for the electric dipole moment of various particles (figure taken from [16]).

of the Thomas-BMT equation (and not the reference proton mass).

$$g_{m_d} = \mu_d \frac{2m_d}{q_s} = 1.714\,025\,460(45) \quad g_{m_p} = \mu_d \frac{2m_p}{q_s} = 0.857\,438\,233\,5(22) \quad (1.8)$$

$$G \equiv \frac{g_{m_d} - 2}{2} = \frac{g_{m_p} \cdot \frac{m_d}{m_p} - 2}{2} = \frac{g_{m_p} \cdot 1.99901 - 2}{2} = -0.142\,987\,269\,7(22) \quad (1.9)$$

$$\neq g_{m_p} - 1 \approx -0.142\,56 \quad (1.10)$$

Current EDM limits for various particles are given in Figure 1.1. Note that the measurement for the proton is indirect and that no measurement for deuterons exists – both particles are promising candidates for direct EDM measurements in storage rings. It should be noted however that because the EDMs are so small, $|\eta_{\text{EDM}}| \ll 1$, while the typical g is of order $\mathcal{O}(\pm 1)$ as seen in Table 1.

1.3. EDM and CP violation

The non-relativistic Hamiltonian of a particle with EDM and MDM in external magnetic \vec{B} and electric \vec{E} fields is given by

$$\mathcal{H} = -\vec{d} \cdot \vec{E} - \vec{\mu} \cdot \vec{B} = -d \cdot \vec{s} \cdot \vec{E} - \mu \cdot \vec{s} \cdot \vec{B}. \quad (1.11)$$

Applying the parity operator \mathcal{P} and the time reversal operator \mathcal{T} leads to sign flips

$$\mathcal{P}(\mathcal{H}) = +d \cdot \vec{s} \cdot \vec{E} - \mu \cdot \vec{s} \cdot \vec{B}. \quad (1.12)$$

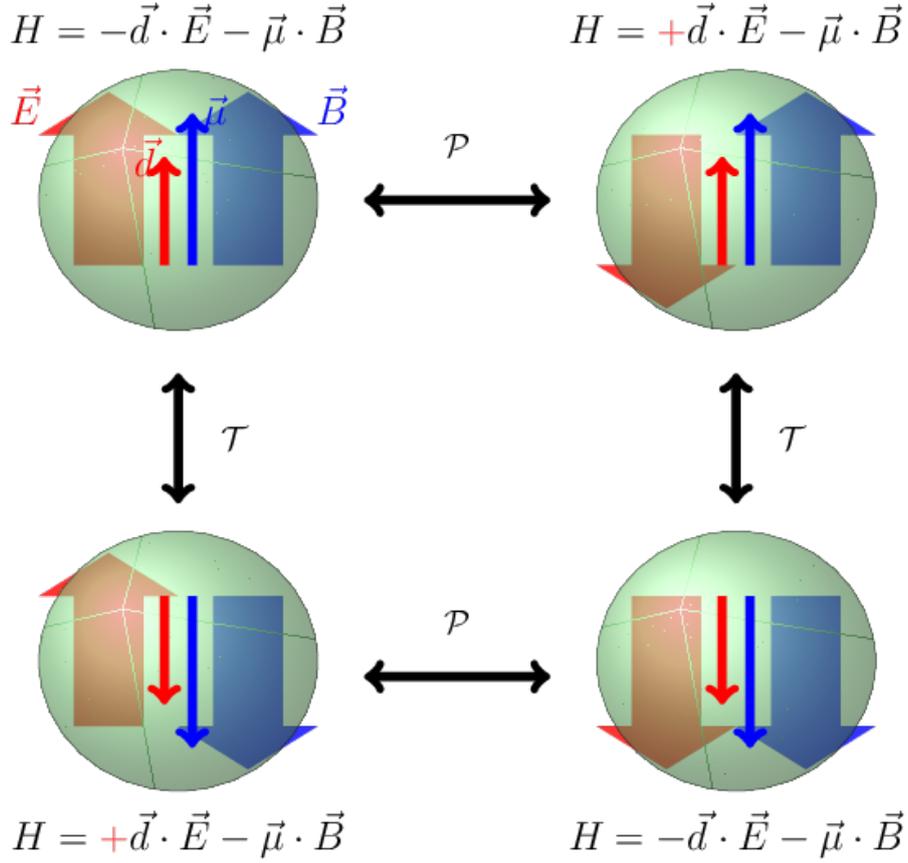


Figure 1.2: Schematic of a particle with a MDM and EDM in an external magnetic and electric field under parity transformation (\mathcal{P}) and time reversal (\mathcal{T}). The graphic is taken from [17].

The Parity transformation in Equation 1.12 changes the sign of the electric field \vec{E} , but does not affect the spin \vec{s} and the orientation of the magnetic field \vec{B} (both axial vectors)

$$\mathcal{T}(\mathcal{H}) = +d \cdot \vec{s} \cdot \vec{E} - \mu \cdot \vec{s} \cdot \vec{B}. \quad (1.13)$$

Time reversal in Equation 1.13 changes the sign of the spin vector \vec{s} and the magnetic field \vec{B} , which finally leads to the same change in the Hamiltonian as the parity transformation. The symmetry violations are visualized in Figure 1.2. Consequently, both symmetries (\mathcal{P} and \mathcal{T}) are violated assuming $d \neq 0$, due to the sign change of the electric dipole moment term. In addition, CP is violated, provided the CPT theorem is valid. Notice how the magnetic dipole moment term remains unchanged by the transformations and does not violate the symmetries.

1.4. EDMs in the Standard Model

The CKM-Matrix in the quark sector is currently the only certain source of CP violation in the Standard Model. The predicted EDMs of the quarks are generated by a three-loop level Feynman diagram and therefore they are very small at [18]

$$d_q^{\text{CKM}} \approx 10 \times 10^{-34} e \text{ cm} - 10 \times 10^{-35} e \text{ cm}. \quad (1.14)$$

Three orders of magnitude smaller values occur for the electron, since here even four-loop diagrams are needed [19]

$$d_e^{\text{CKM}} \approx 10 \times 10^{-38} e \text{ cm} . \quad (1.15)$$

For the neutron enhanced loop effects (from “strong penguin” diagrams [19]) could in principle lead to EDMs of order

$$d_n^{\text{CKM}} \approx 10 \times 10^{-32} e \text{ cm} . \quad (1.16)$$

However, all predictions from theory are orders of magnitude smaller than the current sensitivity of EDM experiments. Still the measurements can put constraints on Super-Symmetry (SUSY) models, which naturally predict much larger EDM values due to additional effects like possible one loops contributions involving hypothetical SUSY particles [20]. If an EDM was measured in current experiments, this would be a clear sign of new physics.

In the strong sector of the Standard Model a possible CP violating $\bar{\theta}$ -term is directly connected to the neutron and proton EDM [21]

$$d_n^{\bar{\theta}} \approx \bar{\theta} \cdot (-2.9 \pm 0.9) \times 10^{-16} e \text{ cm} , \quad (1.17)$$

$$d_p^{\bar{\theta}} \approx \bar{\theta} \cdot (1.1 \pm 1.1) \times 10^{-16} e \text{ cm} . \quad (1.18)$$

The “natural” choice of the $\bar{\theta}$ -term would be expected as $\mathcal{O}(1)$, however the current neutron EDM limits is roughly $d_n \leq 10^{-26} e \text{ cm}$ (compare Figure 1.1), which constrains $\bar{\theta} \leq 10^{-10}$. This is the strong CP “fine-tuning” problem.

1.5. Charged Particle EDM Experiments in Storage Rings

A storage ring experiment designed to precisely determine the EDM necessitates a thorough understanding of the spin motion in electromagnetic fields. The particle motion is determined by the Lorentz force

$$\vec{F}_L = q \left(\vec{E} + \vec{v} \times \vec{B} \right) , \quad (1.19)$$

where q is the charge and \vec{v} is the velocity vector of the particle, while \vec{E} and \vec{B} denote the electric and magnetic field inside the accelerator. To keep a particle on the desired orbit magnetic bending or electrostatic deflection (both dipole fields) can be used; also beam focussing is needed using higher multipole components. Simultaneously, the particle’s spin is tilted, when interacting with a electromagnetic field. According to the Thomas-BMT equation, which describes the spin motion [22], this tilt is proportional to the applied electric field and the particle’s electric dipole moment.

1.5.1. Thomas-BMT Equation

A spin vector \vec{S} in an electric \vec{E} and magnetic \vec{B} field begins to precess with an angular frequency Ω perpendicular to the axis $\vec{\Omega}$. The electromagnetic fields are usually described in the curvilinear laboratory reference frame (here an accelerator). The spin vector \vec{S} is defined in the rest frame of the particle. A calculation of the spin motion leads directly to

the Thomas-BMT (Bargmann-Michel-Telegdi) equation, which also includes the electric dipole moment (for a derivation see e.g. [23])

$$\frac{d\vec{S}}{dt} = \vec{\Omega}_s \times \vec{S} = \vec{\Omega}_{\text{MDM}} \times \vec{S} + \vec{\Omega}_{\text{EDM}} \times \vec{S}, \quad (1.20)$$

$$\vec{\Omega}_{\text{MDM}} = -\frac{e}{m} \left[\left(G + \frac{1}{\gamma} \right) \vec{B} - \frac{G\gamma}{\gamma+1} (\vec{\beta} \cdot \vec{B}) \vec{\beta} + \left(G + \frac{1}{\gamma+1} \right) \left(\frac{\vec{E}}{c} \times \vec{\beta} \right) \right], \quad (1.21)$$

$$\vec{\Omega}_{\text{EDM}} = -\frac{e}{m} \frac{\eta_{\text{EDM}}}{2} \left[\frac{\vec{E}}{c} - \frac{\gamma}{\gamma+1} \left(\vec{\beta} \cdot \frac{\vec{E}}{c} \right) \vec{\beta} - \vec{B} \times \vec{\beta} \right]. \quad (1.22)$$

In Equation 1.20 the spin motion is decomposed into a term $\vec{\Omega}_{\text{MDM}}$ caused by the MDM and a term $\vec{\Omega}_{\text{EDM}}$ caused by the EDM (note again that $\eta \ll G$). The Lorentz factor is $\gamma = 1/\sqrt{1-\beta^2}$ and $\vec{\beta} = \vec{v}/c$.

Assuming that the particle motion $\vec{\beta}$ is always exactly perpendicular to the electromagnetic fields, i.e. $\vec{\beta} \cdot \vec{E} = \vec{\beta} \cdot \vec{B} = 0$, which would be the case in ideal EDM experiments in storage rings, the Thomas-BMT equation simplifies to

$$\vec{\Omega}_s^{\text{RING}} = -\frac{e}{m} \left[\left(G + \frac{1}{\gamma} \right) \vec{B} + \left(-\frac{1}{\gamma+1} - G \right) \left(\vec{\beta} \times \frac{\vec{E}}{c} \right) + \frac{\eta_{\text{EDM}}}{2} \left(\frac{\vec{E}}{c} + \vec{\beta} \times \vec{B} \right) \right]. \quad (1.23)$$

However taking into account longitudinal \vec{B} fields may be important for calculations of systematical errors and longitudinal \vec{E} fields are associated with beam acceleration. Remember that the anomalous magnetic moment $G = \frac{g-2}{2}$ used in the Thomas-BMT equation needs to be calculated from a magnetic moment g corrected to the true mass of the particle and not from the commonly found g_{m_p} (defined using the proton mass). One usually considers the spin motion relative to the beam direction and rewrites Equation 1.23, by introducing the angular rotation velocity of the momentum in beam direction $\vec{\Omega}_p^{\text{RING}}$

$$\vec{\Omega}_p^{\text{RING}} = \frac{e}{m} \left[\frac{1}{\gamma\beta^2} \left(\vec{\beta} \times \frac{\vec{E}}{c} \right) - \frac{1}{\gamma} \vec{B} \right]. \quad (1.24)$$

Finally the angular velocity of the spin rotation relative to the beam direction is given by the most commonly used form

$$\vec{\Omega} = \vec{\Omega}_s^{\text{RING}} - \vec{\Omega}_p^{\text{RING}} = -\frac{e}{m} \left[G \vec{B} + \left(\frac{1}{\gamma^2 - 1} - G \right) \left(\vec{\beta} \times \frac{\vec{E}}{c} \right) + \frac{\eta_{\text{EDM}}}{2} \left(\frac{\vec{E}}{c} + \vec{\beta} \times \vec{B} \right) \right] \quad (1.25)$$

The basic principle of measuring an EDM relies on its interaction with external electromagnetic fields. For an EDM measurement, the experimental parameters like the particle momentum (defines β and γ) and the electromagnetic fields \vec{E} and \vec{B} have to be chosen in such a way, that a macroscopic signal involving changes of the spin axis can be measured.

1.5.2. Magnetic Storage Rings and RF Wien Filter Method

For a magnetic ring like COSY, where in a good approximation $\vec{E} = 0$ can be assumed and \vec{B} is the main magnetic field (oriented vertically to bend the beam) Equation 1.25 is further simplified to

$$\text{magnetic ring: } \vec{\Omega} = \Omega_{\text{MDM}} + \vec{\Omega}_{\text{EDM}} = -\frac{e}{m} \left[G \vec{B} + \frac{\eta_{\text{EDM}}}{2} \left(\vec{\beta} \times \vec{B} \right) \right]. \quad (1.26)$$

In the case of a vanishing EDM component ($\eta = 0$), the spin precession is mainly driven by the magnetic field of the ring. Therefore the spin polarization vector prepared in the horizontal plane precesses around the vertical axis with a constant vertical component. The so-called spin tune is then defined as the ratio of the spin precession frequency f_{spin} and the stored beam revolution frequency $f_{\text{beam}} = 2\pi\omega_{\text{beam}}$

$$\nu_s = \frac{f_{\text{spin}}}{f_{\text{beam}}} = \frac{\Omega}{\omega_{\text{beam}}} = \frac{\frac{e}{m}GB}{\frac{eB}{\gamma m}} = \gamma G, \quad (1.27)$$

where $\omega_{\text{beam}} = \frac{eB}{\gamma m}$ is the revolution frequency of the particles in the machine². The spin tune is a measure of how many additional spin revolutions happen per beam revolution. Typical values of operation suitable for COSY are listed in Table 2.

In the case of a non vanishing EDM component ($\eta \neq 0$) in Equation 1.26, the spin polarization vector will not only precess around the vertical axis. Instead it will be tilted by a small angle and gets an additional oscillating vertical component. However, this amplitude is too small to be detected directly. Using the so-called radio frequency (RF) Wien Filter method, the signal is enhanced. A Wien Filter has perpendicular electric \vec{E}_{WF} and magnetic \vec{B}_{WF} fields. The total force on the beam axis must be zero at the desired momentum, to let the particles pass the filter. The velocity where no deflection in the device happens is given by

$$v = \beta c = \frac{|\vec{E}|}{|\vec{B}|}. \quad (1.28)$$

The RF Wien Filter works at resonant frequency with the spin tune frequency.

$$f_{\text{WF}} = f_{\text{beam}}|\gamma G \pm k| = f_{\text{beam}}|\nu_s \pm k| = |f_{\text{spin}} \pm k f_{\text{beam}}| \text{ with } k \in \mathbb{Z}. \quad (1.29)$$

Every turn, the magnetic field of the RF Wien Filter gives a small kick to the spin vectors, breaking the symmetry between parallel and antiparallel spin and particle motion, which leads to a vertical spin polarization (signal) build up. This buildup of the polarization is linear in time, but note that also various systematic effects like small misalignments in the Wien filter can cause similar fake signals [24].

	p [MeV/c]	β	γ	G	$\nu_s = \gamma G$	f_{beam} [kHz]	f_{spin} [kHz]
d	970	0.459	1.126	-0.143	-0.161	750	120

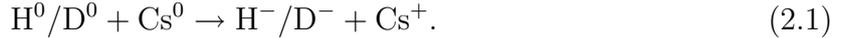
Table 2: Rough overview of the machine parameters for EDM measurements using a magnetic storage ring (circumference 184 m) for deuterons (d) as done with COSY.

²This is a simplification in the case of a circular machine with constant B ; in general one has to calculate field integrals around the storage ring.

2. Cooler Synchrotron

The facility has three key components: Sources that provides polarized (and unpolarized) hydrogen H^- and deuterium D^- ion beams, a pre-accelerator cyclotron called JULIC and the main accelerator COSY. COSY (**C**ooler **S**ynchrotron) is an accelerator with a circumference of 184 m that provides a beam momentum range from $0.3 \frac{\text{GeV}}{c}$ to $3.7 \frac{\text{GeV}}{c}$ for proton and deuteron beams. A sketch of the accelerator facility is shown in Figure 2.1.

Initially, the sources provide unpolarized or nuclear polarized neutral H^0 and D^0 beams. These beams collide with an intense neutralized caesium beam and are ionized in the reaction [25]



The isochronous sector cyclotron JULIC (**JU**elich **L**ight **I**on **C**yclotron) serves as a pre-accelerator for the ions. A small polarimeter called LEP (**L**ow **E**nergy **P**olarimeter) is installed at the injection beamline (IBL). It measures the initial polarization of the beam as provided by the Cyclotron. After transferring the particles via the 94 m long IBL a charge exchanging stripper carbon foil is used for the injection of the particle ensemble into COSY [26]. The accelerator COSY follows a race track design and consists of two arcs (52 m/arc) connected by straight lines (40 m/line). COSY offers four internal (relevant for EDM measurements) and three external (fixed target) experimental sites, where hadron experiments can be carried out. In the arcs, 24 normal conducting water cooled dipole magnets provide the bending force with a maximum magnetic field of 1.58 T. In total 56 quadrupole magnets are used to focus the beam, while sextupole magnets correct chromaticity effects, i.e. they minimize the beam dispersion. The RF-cavity specifies the revolution frequency and momentum of the ions [27].

The main experimental stations for the EDM measurement – including the polarimeters, solenoids and the RF Wien filter – are located in the straight section of COSY as seen in Figure 2.1. The polarization is measured by extracting a small fraction of the circulating particles onto a carbon target. The scattering distribution and asymmetry of the detected events is then analyzed. Solenoid fields allow to manipulate the spins of the particles. COSY has the ability to store polarized ion beams providing long in-plane polarization lifetimes up to $\mathcal{O}(1000 \text{ s})$. Therefore COSY is the natural choice to study systematic effects on the road to a high precision charged particle electric dipole moment experiment [28].

2.1. Important COSY Storage Ring Components

- **Radio-Frequency Solenoid:** The RF solenoid at COSY is a 57.5 cm long air filled, water cooled copper coil. It provides a sinusoidal magnetic field parallel to the beam momentum vector. The RF solenoid at COSY is used to flip the initial vertical polarization of the beam provided from the source into the horizontal plane (also compare Thomas-BMT Equation 1.20). The solenoid is operated on resonance, i.e. on a harmonic of the spin precession frequency $f_{sol} \approx |K + G\gamma|f_{COSY}$ with $K \in \mathbb{Z}$. At COSY $K = 1$ is used, and the exact frequency needs to be determined for the experiment, since imperfections in the ring and fields lead to deviations from the ideal case $\nu_s = G\gamma$ given above. It is crucial to switch off the solenoid at the right moment, when the vertical polarization vanishes [16]. The in plane spin precession

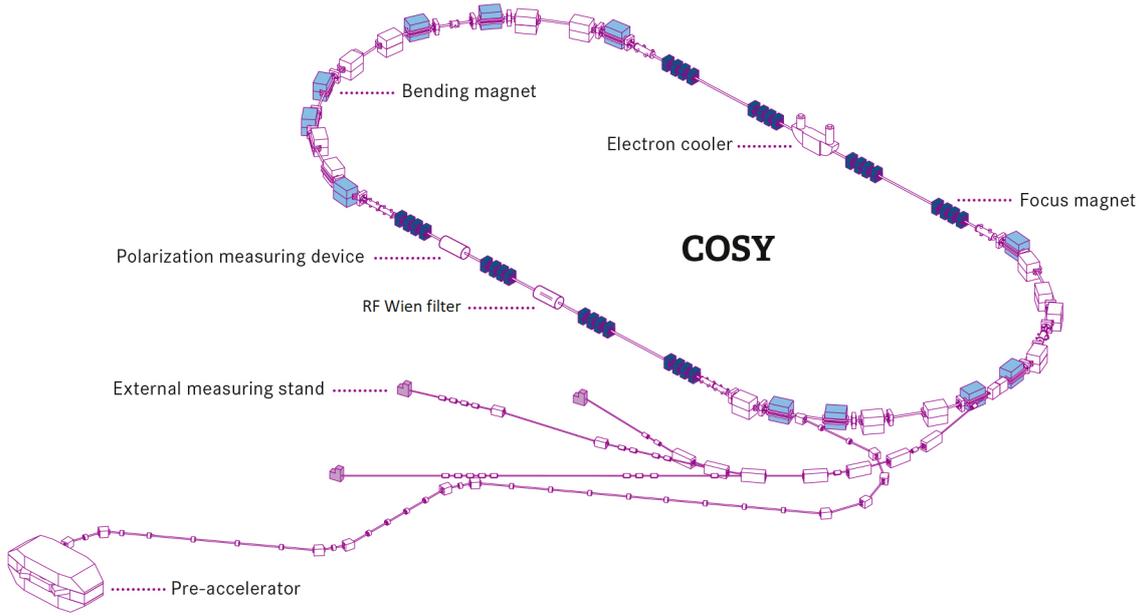


Figure 2.1: Sketch of the COSY facility [29].

frequency f_s is given by the product of the spin tune ν_s and the RF cavity (COSY) frequency:

$$f_s = \nu_s \cdot f_{\text{beam}} \approx 0.1609 \cdot 750 \text{ kHz} \approx 120 \text{ kHz}. \quad (2.2)$$

- Electron Coolers:** Electron coolers reduce emittance and momentum spread of the ion beam by providing a guided coaxial electron beam with the same mean longitudinal velocity, but smaller velocity spread than the ion beam. The improved beam quality is crucial in polarization experiments to increase the so-called Spin Coherence Time [16]. The Coulomb interaction between oppositely charged electrons and ions leads to an energy transfer which reduces the momentum spread of the ion beam. 90° bending dipole magnets guide the electrons in and out of the beam after a short 2 m joint interaction distance, where the particles are also guided and focused by a solenoidal magnetic field, that counteracts the Coulomb repulsion [30]. Two electron coolers are installed in COSY, one in every straight section. Only the 100 keV electron cooler was used for cooling in the experiments discussed in this work. Cooling with the newer 2000 keV electron cooler is not necessary when working with a deuteron beam of up to 970 MeV eV, but its solenoid field was used as an additional spin manipulator for the EDM experiments.
- Reference Frequency System:** For the EDM experiments three RF components are essential: The COSY RF cavity, the RF solenoid and the RF Wien filter. In summer 2020 the fiber-optics-based reference frequency distribution system was installed to synchronize the devices to a GPS-driven reference signal to avoid off-resonance behavior between the beam (RF from cavity) and RF solenoid or RF Wien filter.

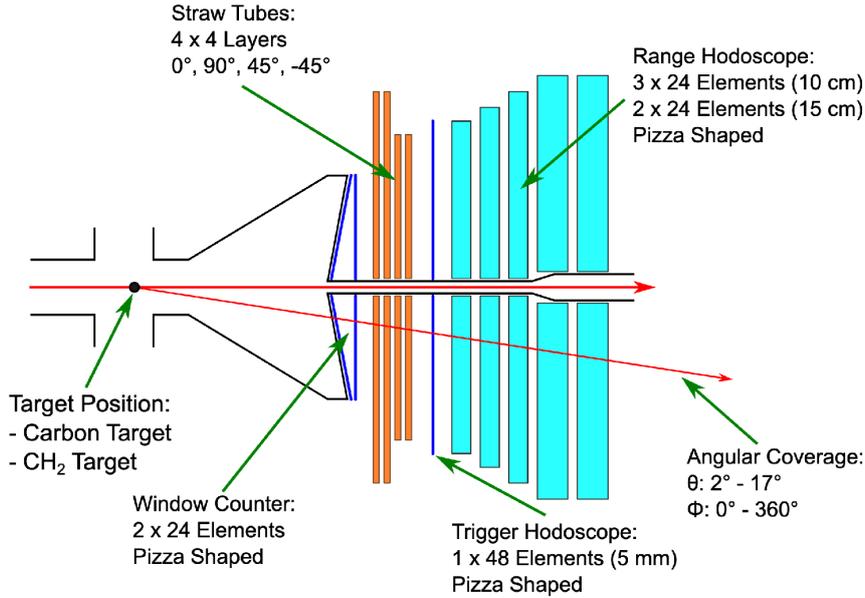


Figure 2.2: Schematic of the WASA forward detector as a cross section [31].

2.2. Polarimeters

Polarimeters are used in COSY to measure the vertical polarization component and the in-plane (horizontal) polarization of the particle bunch circulating the accelerator; both components are reflected in the form of asymmetries of count rates in the detectors. The two polarimeters used for the EDM experiments at COSY were the forward detectors of the “recycled” WASA detector (for so-called Precursor 1 runs, until late 2019) and JEPO, a new polarimeter built by the JEDI collaboration dedicated directly for proton and deuteron EDM experiments (for so-called Precursor 2 runs). The working principle and analysis for data taken by both of the polarimeters is similar: Particles from the halo of the beam scatter with a carbon target and are then detected in the polarimeter.

The detectors are segmented into four regions called “up”, “down”, “left” and “right”, as seen from the beam direction. The up-down asymmetry of count rates in the detectors can then be used as a measure of the in plane (horizontal) polarization, while the left-right asymmetry is linked to the vertical polarization.

- **WASA (forward) detector:** The WASA (**W**ide **A**nge **S**hower **A**pparatus) is a detector built in 1996 at the Department of Radiation Sciences at the University Uppsala originally for the storage ring CELSIUS. The entire detector was moved to COSY in 2006. It consisted of two main parts: a central and a forward part. After taking data for eight years, the physics program of WASA ended in 2015, the detector was removed from the COSY ring and the central part was moved to GSI in Darmstadt. The forward part of the detector was reinstalled and used as a polarimeter for the EDM project [32].

In the first experimental period to measure the orientation of the Invariant Spin Axis in 2018, an unpolarized carbon target was used to scatter the deuterons. The scattered particles can be detected using two layers of forward window counters

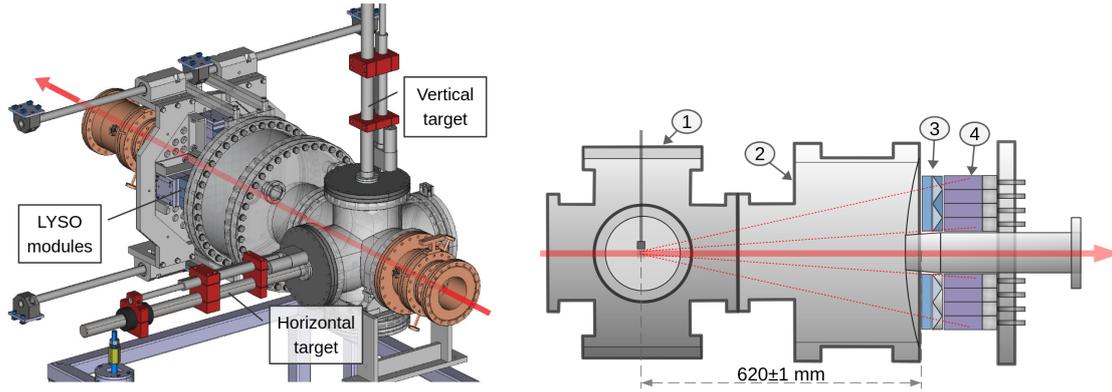


Figure 2.3: Technical drawing of the detector with the target rods in the beamline (left) and sketch of the cross section (right) from the JEDI Polarimeter (JePo) (graphics taken from [34]).

(FWC), followed by four layers of straw tubes, a single layer forward trigger hodoscope (FTH) and finally five layers of the forward range hodoscope (FRH). Each layer of this component is made from plastic scintillators arranged in 24 pizza shaped elements. A sketch of the detector with its layers is shown in Figure 2.2. The angular coverage is given by $\Theta = 2^\circ - 17^\circ$ and $\phi = 0^\circ - 360^\circ$. For the use as a polarimeter, the full detector is subdivided into four sections, each covering an azimuthal angle of $\Delta\phi \approx 90^\circ$.

- **JEDI Polarimeter:** In November 2019, a new dedicated polarimeter called JEPO (**Jedi Polarimeter**) based on LYSO inorganic crystal scintillators was installed in COSY, to increase the sensitivity of the polarimeter for measuring the polarization build-up due to the EDM. Each LYSO (cerium-doped **L**utetium-**Y**ttrium **O**xyorthosilicate, a **S**ilicon **O**xide) crystal measures $3 \times 3 \times 8 \text{ cm}^3$ [33].

The crystals are arranged in blocks of $3 \times 4 + 1$ for the four detector regions. The positioning is done in such a way, that the so called Figure of Merit $\text{FOM} = \sigma(\theta)A_y^2(\theta)$ (with the angular cross section σ and vector analyzing power A_y squared) is maximized, since this minimizes the statistical uncertainty of the measurement. A technical drawing and cross section of the JEPO are shown in Figure 2.3. For the EDM experiments relevant for this work, only the vertical target was used. It can be moved close to the halo of the beam by a rod. The particle extraction rate is hold constant over time by a feedback system. After traversing the target chamber (1) and a vacuum flight chamber (2) the scattered particles are detected in tracking plastic scintillators (3) and finally are stopped in the LYSO crystals (4).

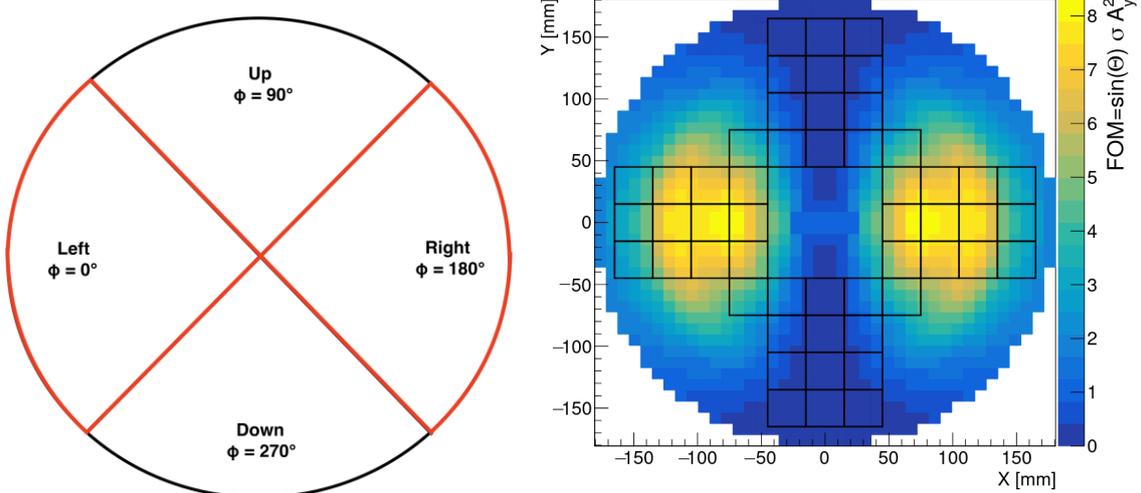


Figure 2.4: Left picture: The convention for Up, Down, Left, Right detector regions. Right picture: LYSO crystals of the JEPO (black boxes) are placed strategically to maximize the information gained from the scattered particles by checking their corresponding figure of merit in the left-right case. Note that the FOM for the up-down case is given by a rotation by 90° [34]).

2.3. Polarization Determination

2.3.1. Turn Number and Time in Cycle

For the determination of the spin precession frequency in the horizontal plane, each event is assigned to a number n of particle turns in the storage ring. The length of a turn number interval is proportional to the time in the accelerator cycle (tic) for operation at fixed energy and therefore fixed beam frequency (up to tiny variations not relevant for this work). Both scales (time and turn number) can and will be used to indicate the duration of the measurement ($\Delta n = 10^6 \equiv 1.3 \text{ s} = \Delta t$). Typically the measurements relevant for this work using the polarimeters start at $t_0 = 90 \text{ s}$ – after performing all previous steps like e.g. beam acceleration, preparation and cooling. The cycle data taking ends at $\approx 264 \text{ s}$. Note that t_0 usually gets turn number $n = 0$. For the EDM measurement campaigns precursor 1 and precursor 2 both time in cycle and turn number is given directly in the ROOT data files of recorded events in the polarimeters.

2.3.2. Polarimeter Event Rates

The event rates $\dot{N}_X = \frac{dN_X}{dn} \approx N/\Delta t$ are measured in each detector quadrant $X = (L, U, R, D)$ in order to determine the horizontal and vertical asymmetry parameter (polarization). In Figure 2.5 the counting rates of the four detector quadrants are shown for a 264 s cycle. There are no events saved from the first 90 s of the cycle, because the beam is prepared by bunching and cooling. As soon as the extraction of the beam onto the target starts the event rates increase. All the rates are (mostly) flat and roughly the same amount of events is detected in each detector segment per time bin in the main data acquisition period after the flip of the spin axis at $t \approx 100 \text{ s}$.

Note that a beam current monitor output would show a decline of stored beam intensity,

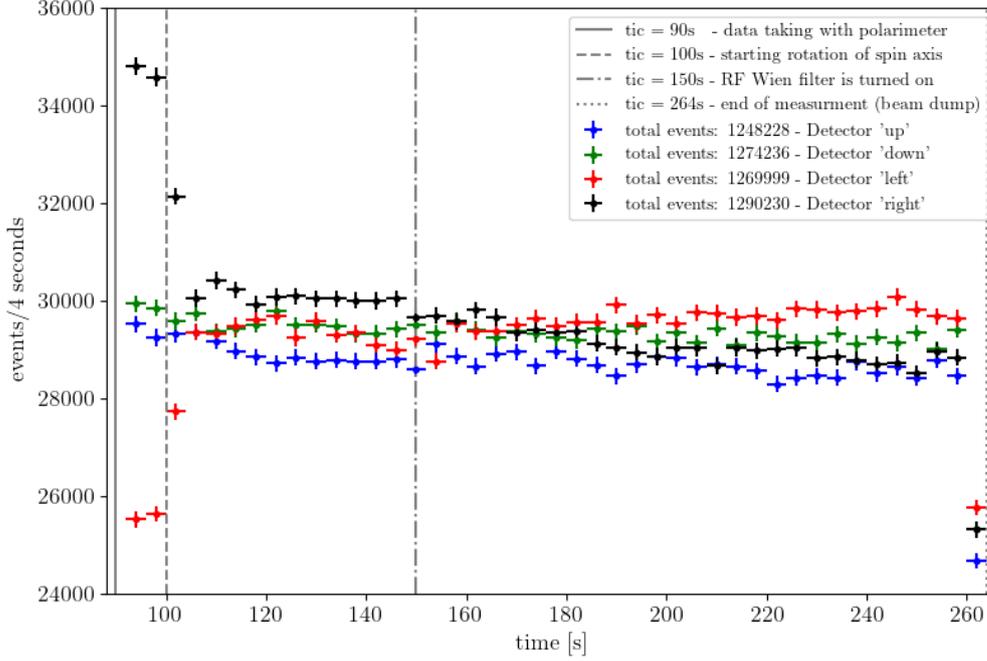


Figure 2.5: Raw counting rates of each detector quadrant for a cycle (Precursor 1 experiment), where the WASA detector was used. Measurement principle (vertical lines indicate the corresponding time): a) Until ≈ 90 s beam preparation, including bunching and cooling, no data taking in the detectors. b) At ≈ 100 s the RF Solenoid turns the polarization into the horizontal plane and feedback preparation takes place. c) At ≈ 150 s the RF Wien Filter is switched on and the frequency is set to the spin precession frequency $f_{WF} = f_s$, while adjusting and maintaining the relative phase between the polarization precession and the RF Wien Filter. c) After ≈ 264 s the measurement for one cycle ends.

similar to Figure 2.6, as particles are lost over time e.g. in interactions with the target of the polarimeter. However, this is corrected for by a feedback system changing the fields around the target, which ensures a constant rate of detected events and by that similar statistical uncertainty during the full measurement. The absolute counting rates for all four detectors are a bit different, since the acceptances of the individual detectors are not identical. Additionally, the beam might not go exactly through the geometric center of the detector and also the target is not symmetric, which also yields to asymmetric counting rates. The efficiency for JEPO is in the order of 1% and therefore roughly 100 times more particles are lost at the polarimeter than detected in the segments over the measurement time. [34].

2.3.3. Left-Right Asymmetry (Vertical Polarization)

To compute an analytical expression for the left right asymmetry two point-like detectors are considered, which are placed at $\varphi_L = 0^\circ$ and $\varphi_R = 180^\circ$ (compare Figure 2.4). The number of detected events N_L , N_R during a (macroscopic) time interval Δt_{mac} for a

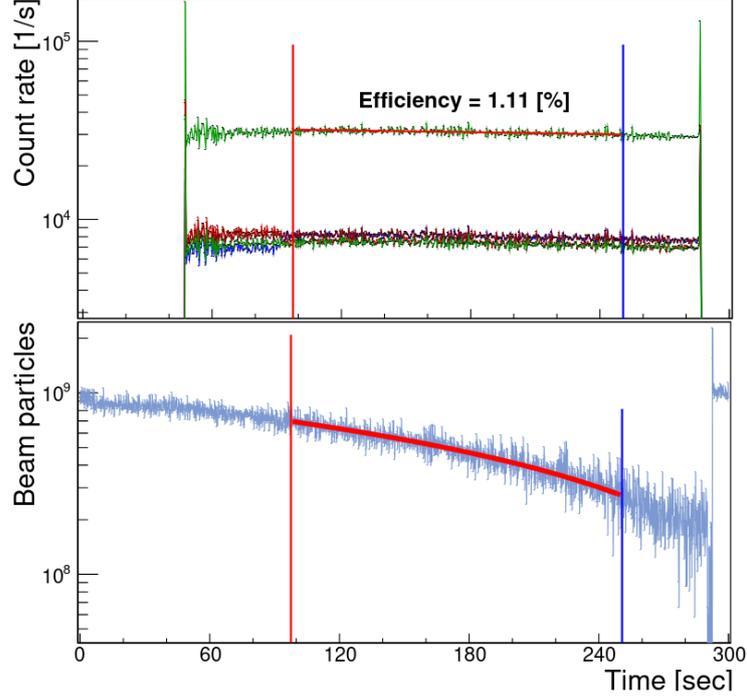


Figure 2.6: In the upper panel the counting rates for the four quadrants of the JEPO polarimeter (used for Precursor 2) appear as colored traces. They are nearly flat for the data acquisition period and similar, but not identical. The green trace shows the summed count rate. The start of beam extraction onto the target (here at ≈ 60 s) generates a spike in rate. In the lower panel the blue line traces the declining stored beam intensity. The efficiency is measured between the two vertical lines during the length of a typical measurement cycle (figure taken from [34]).

vanishing tensor polarization (only vector polarization p_Z considered) is given by

$$N_R = I d_t \bar{\sigma}_{0R} \Delta t_{\text{mac}} \left(1 - \frac{3}{2} p_Z \bar{A}_y \right), \quad (2.3)$$

$$N_L = I d_t \bar{\sigma}_{0L} \Delta t_{\text{mac}} \left(1 + \frac{3}{2} p_Z \bar{A}_y \right). \quad (2.4)$$

Here I denotes the beam current intensity, d_t is the target density, $\bar{\sigma}_{0L}$ and $\bar{\sigma}_{0R}$ are the averaged differential cross sections for both detectors. The averaged analyzing powers \bar{A}_y are assumed to be the same in this simple approach. To determine the vertical polarization, the conventional calculation of the left-right asymmetry ϵ_{LR} is given by

$$\epsilon_{LR} = \frac{N_R - N_L}{N_R + N_L} = \frac{3}{2} p_Z A_Z \quad (2.5)$$

Thus, the asymmetry of the counting rates is ideally proportional to the vertical vector polarization ($\epsilon_{LR} \propto p_Z$), if the cross sections are exactly the same $\bar{\sigma}_{0L} = \bar{\sigma}_{0R}$. In reality this is not necessarily the case and a possible correction procedure for the left-right asymmetry is described in [35].

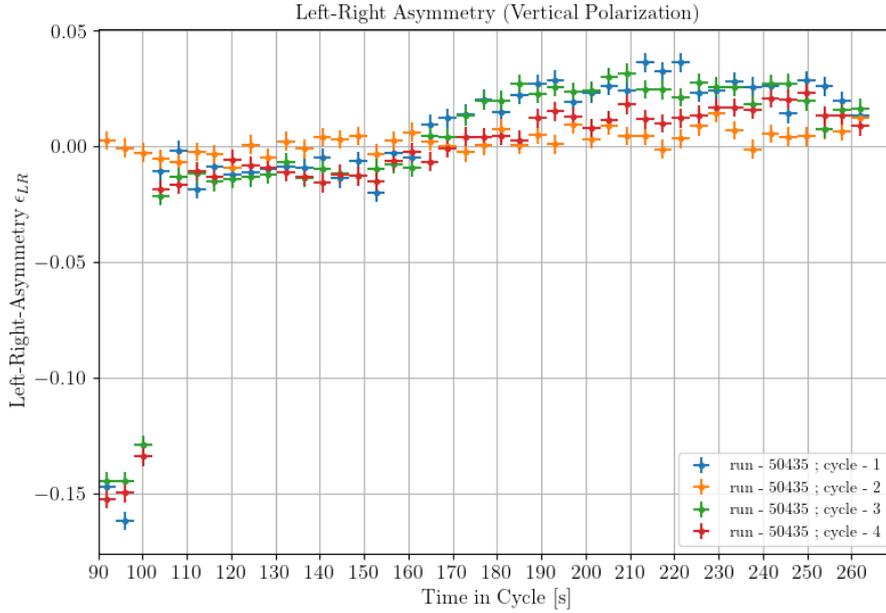


Figure 2.7: Raw left-right asymmetry for multiple cycles in one run, using the counting rates. Note that the unpolarized cycle (here number 2) does not show a notable initial asymmetry (between 90 s and 100 s), while the following cycle are similar in asymmetry, but also not identical.

The statistical uncertainty of Equation 2.5 can be calculated directly using Gaussian error propagation and assuming uncorrelated counting rates, which follow Poisson statistics with $\sigma_N = \sqrt{N}$ for $N \gg 1$

$$\sigma_{\epsilon_{LR}} = \sqrt{\frac{4N_R^2 N_L + 4N_R N_L^2}{(N_R + N_L)^4}} = \sqrt{\frac{4N_R N_L}{(N_R + N_L)^3}} \approx \sqrt{\frac{4N^2}{(2N)^3}} = \sqrt{\frac{1}{2N}}. \quad (2.6)$$

Note that these uncertainties approximately follow the $\sqrt{1/(2N)}$ behavior, when assuming $N_R \approx N_L \equiv N$, which is valid for large parts of the experiment.

A typical examples of the left-right asymmetries for four cycle of a Precursor 1 run is shown in Figure 2.7. Note that a run consists of multiple polarized (here cycles 1,3 and 4) and an unpolarized cycle, which shows no starting asymmetry. This cycle (plotted in orange) can be used for fake-asymmetry corrections as done in [35]. Optimally the left-right asymmetry for every cycle would amount to zero, after the vertical polarization is turned into the horizontal plane (spin precession for $t > 100$ second until the Wien filter is turned on).

2.4. Important Quantities of the Data Taking Procedure

The runs probing for a RF Wien filter signal buildup follows this procedure: A vertically polarized deuteron beam is injected into the COSY ring, accelerated and cooled. After 90 s the target of the polarimeter is moved next to the beam and the extraction starts. At

100 s the RF solenoid is used to rotate the polarization into the accelerator plane, where it starts to precess.

The phase feedback system measures the phase relation between the spin precession frequency and the (turned off) RF Wien filter to adjust frequency and phase of the device. At 155 s the RF Wien filter is turned on, and a buildup of vertical polarization can be measured. This vertical polarization component is linked to the orientation of the Invariant Spin Axis and therefore also to the EDM. To summarize, the most relevant information found in the data for each cycle for the determination of the direction of the Invariant Spin Axis are:

- **Vertical asymmetry (linked to vertical polarization):** The vertical asymmetry is computed directly from an asymmetry of detected particles in the left and right quadrant of the polarimeter, which scales with the vertical polarization. The asymmetry of detected events needs to be corrected using cycles without initial polarization to correct systematic effects and to deduct fake signals for an EDM measurement using a linear model. A nonzero EDM provides a linear signal buildup, after the Wien Filter is turned on.
- **Horizontal asymmetry (linked to horizontal polarization):** The horizontal asymmetry is computed from an asymmetry of detected particles in the upper and lower detectors segments. The in-plane polarization is rotating with the spin precession frequency (spin tune), which is much faster than the detector rate. Therefore this polarization component cannot be determined directly, but many oscillation periods have to be taken into account (Mapping Method or Fourier Transform approach). An oscillation with maximum amplitude is occurring, when the right spin precession frequency is chosen in the analysis. The resulting amplitude is called up-down asymmetry, which scales with the horizontal or in-plane polarization. The horizontal polarization declines over time due to a depolarization effect. Therefore a long spin coherence time is wanted in the experiment. The analysis of the up-down asymmetry and its precession amplitude will be discussed in greater detail in section 5.
- **Total asymmetry (linked to total polarization):** The total polarization is calculated from the sum of the squared left-right asymmetry and up-down asymmetry. It describes the amount of polarization of the particle bunch.
- **Angle between Vertical and Horizontal Polarization:** Determined by calculating the arc tangent of the left-right (vertical) and up-down (horizontal) asymmetry. A linear slope can be interpreted as a EDM signal signal, or as a systematic effect/fake signal.

3. Fundamental Statistic Concepts

This work discusses the estimation of observables obtained from the polarimeters in the COSY storage ring, based on ideas developed by Dennis Eversmann during his PhD Thesis [36] using statistical methods. For this purpose, the relevant statistical concepts for the Bayesian and Feldman Cousins approach, which is purely *frequentist*, are discussed in this and the next chapter. Later these concepts are applied to parameter estimation for the asymmetry (polarization) amplitude and its time dependence (spin coherence time) in the COSY EDM measurements.

3.1. Central Limit Theorem

The central limit theorem (CLT) is a fundamental concept, which states that for a sequence of independent and identically distributed random variables X_1, \dots, X_n with expectation value $E[X_i] = \mu$ and variance $\text{Var}[X_i] = \sigma^2 < \infty$ the resulting summed up distribution converges to a Normal distribution. It is a remarkable property of the CLT that no assumption is to be made about the shape of the distribution of the individual X_i .

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_i^n X_i \rightarrow \mathcal{N}(\mu, \sigma/\sqrt{n}) \quad (3.1)$$

Also one should notice, that for increasing n the width of the distribution (standard deviation σ) will get narrower with $1/\sqrt{n}$. This behavior, that the variance of a parameter falls regularly with $\frac{1}{n}$ for large n is also key to parameter estimation practices [37].

3.2. Bayes Theorem

Bayes' theorem is another important law of probability theory and later relevant for including so called prior knowledge of distributions in the parameter estimation process. It can be written as

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_x P(B|x)P(x)} \quad (3.2)$$

$P(A)$ and $P(B)$ are the probabilities for events A and B happening, without any other requirements. $P(A, B)$ is the probability of A and B happening. The conditional probability $P(A|B)$ represents the probability of observing A given that B is true and vice versa for $P(B|A)$. Notice that the probability $P(B)$ can be expressed as a sum of all its possible conditional events weighted by how likely they are each, as long as those are disjoint with $\sum_x P(x) = 1$ (marginalization).

According to the Bayesian interpretation, the unconditional and conditional probabilities can be seen as a degree of belief before and after considering the evidence. In this sense, $P(A)$ is called the *prior*, which represents the first level of belief in A , while $P(A|B)$ represents the *posterior*, when B is taken into account. The ratio $P(B|A)/P(B)$ is considered as the support B provides for A .

In a certain sense, Bayes' theorem allows conclusions to be reversed: one starts from a known value $P(B|A)$, but one is actually interested in the value $P(A|B)$. In the context of this thesis B is typically some obtained measurement result, and A is its true physical value. Of course one would like to directly conclude the knowledge of the true value and

its uncertainty from the measurement, but it has to be noted, that the *prior* probabilities and correct distributions have to be considered to avoid fallacy.

It can be shown however, that for an unbiased estimator without constraints, and uncertainties that follow a Gaussian distribution, $P(B|A)$ and $P(A|B)$ can indeed be interchanged when assuming a so called **constant prior** (for details see subsection C.2 in the appendix).

This is e.g. implicitly done when performing a standard χ^2 -fit on some data. Here one typically assumes that the measurement is unbiased, so that the best estimate for the fit of the true behavior is indeed the measurement point, and the uncertainties (represented by errorbars) are assumed to be Gaussian. The Gaussian behavior will also be the typical “baseline” for comparisons in this work. The focus in the following chapters will be on cases, where it is not from the start clear, whether e.g. Gaussian simplifications are justified and one needs to discuss the necessity of more elaborate statistical methods and its effects on the analysis.

3.3. Estimators

A statistic estimator is in general a rule to determine an estimate of a quantity based on observed data, typically also providing an uncertainty of the estimate. If some fixed parameter θ (true value) is to be estimated, the estimator is given by a function, which takes the measurement (observed data) as input. In this work estimators (and later their output) are denoted by a hat above the symbol $\hat{\theta}$. For a random variable (X) which describes the distribution of the measured data as input, the estimator becomes a function of a random variable $\hat{\theta}(X)$, but for simplicity this is written still as $\hat{\theta}$. The estimate for some given data set (x) is denoted as $\hat{\theta}(x)$, which is always a fixed value, since only deterministic estimators are discussed in this work.

3.3.1. Variance, Mean Squared Error and Bias

The **variance**

$$\text{Var}(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2] \quad (3.3)$$

is an indicator in how far the estimated parameter is off from the expected value of the estimator on average. A small variance however does not necessarily indicate a good description of the true parameter. The estimator could be always wrong the same way. To deal with this case, the **mean squared error**

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \quad (3.4)$$

can be introduced. It is given by the expectation value of the squared errors, defined as deviation from the true value θ . For some given sample of data x , the **error** of the estimator $\hat{\theta}$ is defined as $e(x) = \hat{\theta}(x) - \theta$. The error $e(x)$ depends on the estimation formula and the sample. The MSE is the probability weighted squared error for all $x \in X$. A small MSE ensure that the results of the estimator are clustered closely around the true value. Finally the **bias**

$$B(\hat{\theta}) = E[\hat{\theta}] - \theta \quad (3.5)$$

is defined, which corresponds to the distance between the average of the estimates and the true value of the parameter θ . If $B(\hat{\theta}) = 0$ the estimator $\hat{\theta}$ is called an unbiased estimator.

The mean squared error can also be written as a combination of variance and bias

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + (B(\hat{\theta}))^2. \quad (3.6)$$

A proof for this relation is given in subsection A.1 in the appendix. Therefore typically the goal is to minimize not only the variance or the bias, but the MSE as a trade off [38].

3.3.2. Fisher Information and Cramer-Rao Bound

The Fisher information indicates the amount of information, that a measurable random variable X carries with respect to an unknown parameter θ . Below the simple case of one scalar parameter will be described including possible bias. For multiple parameters the Cramér-Rao bound corresponds to the covariance matrix.

First the scalar unbiased case: Consider the problem, that the parameter θ is to be estimated from n independent observations of the random variable X , called x_1, \dots, x_n or in short x to denote the full set. Each observation comes from the same distribution according to some probability density function (PDF) $f(x; \theta)$. The variance of any **unbiased** estimator $\hat{\theta}$ with $E[\hat{\theta}] = \theta$ is then bounded by the reciprocal of the Fisher information [39]:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}(\theta)}. \quad (3.7)$$

The Fisher information $\mathcal{I}(\theta)$ is defined by the expected value of the square of the **score function** $\frac{\partial}{\partial \theta} \log(f(x; \theta))$ with respect to the PDF $f(x; \theta)$:

$$\mathcal{I}(\theta) = \text{E} \left[\left(\frac{\partial}{\partial \theta} \log(f(x; \theta)) \right)^2 \right] = \text{E} \left[-\frac{\partial^2}{\partial \theta^2} \log(f(x; \theta)) \right]. \quad (3.8)$$

A more general form of the bound also considers estimator \hat{T} , whose expectation is some (differentiable) function of θ , i.e. $E[\hat{T}] = \psi(\theta)$. In this case, the bound is given by:

$$\text{Var}(\hat{T}) \geq \frac{[\psi'(\theta)]^2}{\mathcal{I}(\theta)} \quad (3.9)$$

where $\psi'(\theta)$ is the derivative $\frac{\partial}{\partial \theta} \psi(\theta)$, and $\mathcal{I}(\theta)$ is the Fisher information from Equation 3.8. Notice how for $\psi_{B=0}(\theta) = \theta$ the unbiased limit reoccurs naturally, since $\frac{\partial}{\partial \theta} \psi_{B=0}(\theta) = 1$.

Finally consider a **biased** estimator $\hat{\theta}$ with a bias $B(\hat{\theta}) = E[\hat{\theta}] - \theta \equiv b(\theta) \neq 0$ given by a differentiable function $b(\theta)$. With the result from Equation 3.9, this corresponds to $E[\hat{\theta}] = \psi(\theta) = \theta + b(\theta)$ and the bound is:

$$\text{Var}(\hat{\theta}) \geq \frac{[1 + b'(\theta)]^2}{\mathcal{I}(\theta)}. \quad (3.10)$$

Also the MSE of a biased estimator directly follows to be

$$\text{MSE}(\hat{\theta}) = \text{E}[(\hat{\theta} - \theta)^2] \geq \frac{[1 + b'(\theta)]^2}{\mathcal{I}(\theta)} + b(\theta)^2 \quad (3.11)$$

using the standard decomposition of the MSE (see Equation 3.6). Note, however, that if $1 + b'(\theta) < 1$ the variance bound will always and the MSE might be smaller than the

unbiased Cramér–Rao bound $1/\mathcal{I}(\theta)$, depending on the bias. Biased estimators, which indeed surpass the unbiased Cramér–Rao Bound are sometimes called “super efficient estimators”. In the case of the parameter estimation for a sine-wave, which will be discussed in this thesis, estimators for the amplitude are biased and indeed can achieve super efficiency [40].

In the case of multiple parameters $\boldsymbol{\theta} = [\theta_1, \dots, \theta_d]^T \in \mathbb{R}^d$ the Cramér-Rao bound corresponds to the covariance matrix of the (biased) estimator $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \dots, \hat{\theta}_d]$ and the Fisher information matrix with entries

$$\mathcal{I}_{m,k} = \mathbb{E} \left[\frac{\partial}{\partial \theta_m} \log f(x; \boldsymbol{\theta}) \frac{\partial}{\partial \theta_k} \log f(x; \boldsymbol{\theta}) \right] = \mathbb{E} \left[-\frac{\partial^2}{\partial \theta_m \partial \theta_k} \log f(x; \boldsymbol{\theta}) \right]. \quad (3.12)$$

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \geq \frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathcal{I}(\boldsymbol{\theta})^{-1} \left(\frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T. \quad (3.13)$$

There $\boldsymbol{\psi}(\boldsymbol{\theta}) = [\psi_1(\theta_1, \dots, \theta_d), \dots, \psi_d(\theta_1, \dots, \theta_d)]^T = \mathbb{E}[\hat{\boldsymbol{\theta}}]$ is the possibly biased expectation vector of the estimator and $\frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ denotes the Jacobian matrix with elements given by $\frac{\partial \psi_i(\boldsymbol{\theta})}{\partial \theta_j}$. In the case of unbiased estimators, where $\frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{1}$, the Cramér-Rao bound reduces to the matrix inverse [41]

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \geq \mathcal{I}(\boldsymbol{\theta})^{-1}. \quad (3.14)$$

The variance bound of the estimator $\hat{\theta}_m$ is then given by

$$\text{Var}(\hat{\theta}_m) = [\text{Cov}(\hat{\boldsymbol{\theta}})]_{mm} \geq [\mathcal{I}(\boldsymbol{\theta})^{-1}]_{mm} \left(\geq ([\mathcal{I}(\boldsymbol{\theta})]_{mm})^{-1} \right). \quad (3.15)$$

If it is inconvenient to compute the inverse of the Fisher information matrix, one can take the reciprocal of the corresponding diagonal element (see last inequality in Equation 3.15) to find a possibly more loose lower bound for the variance of the estimator $\hat{\theta}_m$. This more loose bound is not reached by the estimators but at least dimensions and general behavior are right (up to a constant factor) [41].

3.3.3. Properties of Estimators

The theory of statistics provides four key properties of an estimator as stated below:

- **Consistency:** An estimator is said to be (weakly) consistent if it converges in probability to the true value of the parameter, for $n \rightarrow \infty$. For this we write $\hat{\theta}_n \rightarrow \theta$, where n denotes the number of observations of a given sample $x = (x_1, \dots, x_n)$. If a sequence of estimators is unbiased and converges, then it must converge to the correct value θ and is also consistent [42].
- **Asymptotic Normality:** An asymptotically normal estimator is a consistent estimator whose distribution around the true parameter θ approaches a normal distribution with a standard deviation σ shrinking in proportion to $1/\sqrt{n}$ as the sample size n grows (also see subsection 3.1). Note that $(\sigma/\sqrt{n})^2 = V/n$ is simply an approximation of the *true variance* of the estimator, since in the limit $n \rightarrow \infty$, the asymptotic variance reaches zero and $\hat{\theta}_n$ converges to a Dirac delta function around θ .

- **Efficiency:** An efficient estimator is an estimator that estimates the quantity of interest in some “best possible” manner. The notion of “best possible” is of course not unambiguous, therefore one needs to define the desirable properties of estimators: One might want to get an *unbiased* ($B(\hat{\theta}) = 0$) estimator and also have *minimal mean squared error*. This conditions cannot in general be satisfied simultaneously: A biased estimator may have a MSE smaller than any unbiased estimator (see subsection 3.3.2). For unbiased estimators variance and MSE are of course the same. Among unbiased estimators the one with the lowest variance is called the minimum variance unbiased estimator (MVUE).
- **Robustness:** An estimator is considered robust, if the result is the same or at least not *substantially* changed when there are outliers or (small) deviations from the assumptions (e.g. underlying probability densities) provided to construct the estimator. Therefore, one defines the so called breakdown point, when the results of the estimator become unreliable. Typically this is given as a fraction of “bad” observations or a signal to noise ratio, where the incorrect observations begin to dominate.

3.4. Maximum Likelihood Method

In statistical theory the maximum likelihood method is one of the standard techniques for parameter estimation. The method is based on the idea that the observations are fixed, while θ as the model dependent parameter can vary freely. Without loss of generality θ could also be a vector of parameters $\boldsymbol{\theta}$ in the following considerations. The parameter’s real value θ_0 is a priori unknown.

The joint probability density function for a given collection of n independent and identically distributed observations forming a vector $\mathbf{x} = [x_1, \dots, x_n]^T$ is given by:

$$f(\mathbf{x}|\theta) = f(x_1, x_2, \dots, x_n|\theta) = f(x_1|\theta) \times f(x_2|\theta) \times \dots \times f(x_n|\theta), \quad (3.16)$$

When $f(\mathbf{x}|\theta)$ is viewed as a function of \mathbf{x} with the parameter θ fixed, it is considered a PDF and when viewed as a function of θ with the data \mathbf{x} fixed, it is considered a likelihood function:

$$\mathcal{L}(\theta; \mathbf{x}) = \mathcal{L}(\theta; x_1, x_2, \dots, x_n) \equiv f(x_1, x_2, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta). \quad (3.17)$$

Note, that $\mathcal{L}(\theta; x_1, x_2, \dots, x_n)$ is **not** a *posterior* probability distribution. As described in subsection 3.2 one would need *prior* knowledge about the distribution of the parameter θ , to “invert” the PDF.

In order to simplify subsequent calculations, the product of probabilities is transformed into a sum by taking the natural logarithm to get the so-called objective function

$$\ell = \log \mathcal{L}(\theta; x_1, x_2, \dots, x_n) = \sum_{i=1}^n \log(f(x_i|\theta)). \quad (3.18)$$

The maximum likelihood estimator (MLE) corresponds to the value, which maximizes the likelihood

$$\hat{\theta}_{\text{MLE}} = \arg \left\{ \max_{\theta} (\log \mathcal{L}) \right\} \quad (3.19)$$

A maximum likelihood estimator is therefore an extremum estimator. Under conditions discussed e.g. in [43] maximum likelihood estimation provides appealing asymptotic properties. With increases sample size $n \rightarrow \infty$ (typically n needs to be only of moderate size to get adequate results) maximum likelihood estimators have these properties:

- Consistency: MLEs converge in probability to the value being estimated.
- Efficiency: MLEs achieve the Cramér–Rao lower bound of variance, for a sample size $n \rightarrow \infty$. The variance in this case is given by the reciprocal of the Fisher information as described in subsection 3.3.2. This means that no consistent estimator has lower asymptotic mean squared error than the MLE (or other estimators attaining this bound), which also means that MLEs reach asymptotic normality [44].
- Invariance: If $\hat{\theta}$ denotes the maximum likelihood estimator for θ , and $f(\theta)$ is some transformation of θ , then the maximum likelihood estimator for $\alpha = f(\theta)$ is simply given by $\hat{\alpha} = f(\hat{\theta})$. This property is incompatible with unbiasedness, and indeed MLEs are in general still biased of order $\frac{1}{n}$ [45]. Note that $1/n$ is much smaller than the statistical uncertainty, which falls with $1/\sqrt{n}$. In principle it is possible to correct for the bias to get second-order efficient estimators, but this is beyond this thesis.

4. Feldman-Cousins Algorithm

4.1. Poisson Distribution with Background

The Feldman-Cousins algorithm [46] is used to construct so-called confidence interval for (physical) parameters from corresponding probability distributions. The starting point for the construction of purely frequentist Feldman-Cousins confidence intervals is the so-called Neyman construction combined with likelihood ratios.

This is at best discussed for a concrete and comparably simple example. Consider a counting experiment based on a Poisson distribution of observed events, where one knows a background b and is interested in the signal events s_0 . The corresponding probability distribution is given by:

$$P(k; s_0 + b) = \frac{(s_0 + b)^k}{k!} e^{-(s_0 + b)} \quad (4.1)$$

The Neyman construction works as follows: Given a true value of the parameter θ (here the true signal s_0) one considers a probability (density) function $f(x; \theta)$ for the outcome of the experiment. Often x – here the measurement of k events given in reality as a sum of signal and background events – is an estimator for the parameter θ . One wants to define an interval in the measured outcomes of the experiment, in which a specified fraction of outcomes (say e.g. confidence level $cl = 90\%$) are occurring. This means that with a summed up probability of $cl = 90\%$ the measurement results x (or k in our example) are found in the interval for given θ . The construction happens in direction of x , which is also typically chosen as the x -axis in corresponding plots.

Note that there is no unique way to choose the x values, which should be included first in the interval. An arbitrary choice can lead to problems like the so-called “flip-flop” in confidence bands when approaching parameter limits, where one chooses different rules based on the actual measurement result. The corresponding jumps in the confidence intervals as well as resulting over- and undercoverage should be considered bad practice [46]. Therefore the main idea of Feldman and Cousins was to take the likelihood ratio,

$$R(k; s_0 + b) = \frac{P(k; s_0 + b)}{P(k|\mu_{\text{best}})} \quad (4.2)$$

as an ordering principle. μ_{best} maximizes $P(k|\mu_{\text{best}})$ for a fixed k , or x in the general case, but taking into account physical limits of the parameters θ . Note that the true signal s_0 cannot be negative in this construction. The highest likelihood ratio for a given s_0 is achieved at the considered true signal plus background (rounded to the next integer value) at $k \approx \mu \equiv s_0 + b$. For discrete distributions R will not always reach the value $R = 1$ at the maximum as for continuous distributions.

The procedure is indicated for some arbitrary true signal $s_0 = 2.4$ in Figure 4.1. Starting from the highest ratio according to Equation 4.2 one ranks all the possible outcomes (following the height of the black step plot, which is scaled to the probability plot) and adds up the corresponding probabilities (blue dots) effectively following the red dashed line. Note that “jumps” in the construction from the lower end of the distribution to the upper end and vice versa are irregular. To represent this, a maximum rank R_X (printed in red) is defined, which indicates the corresponding rank of the last following k , which is still on the “other side” of the probability distribution. Summation up to this rank R_X is linked to the red dots in the lower plot of Figure 4.1. One has to note that by construction

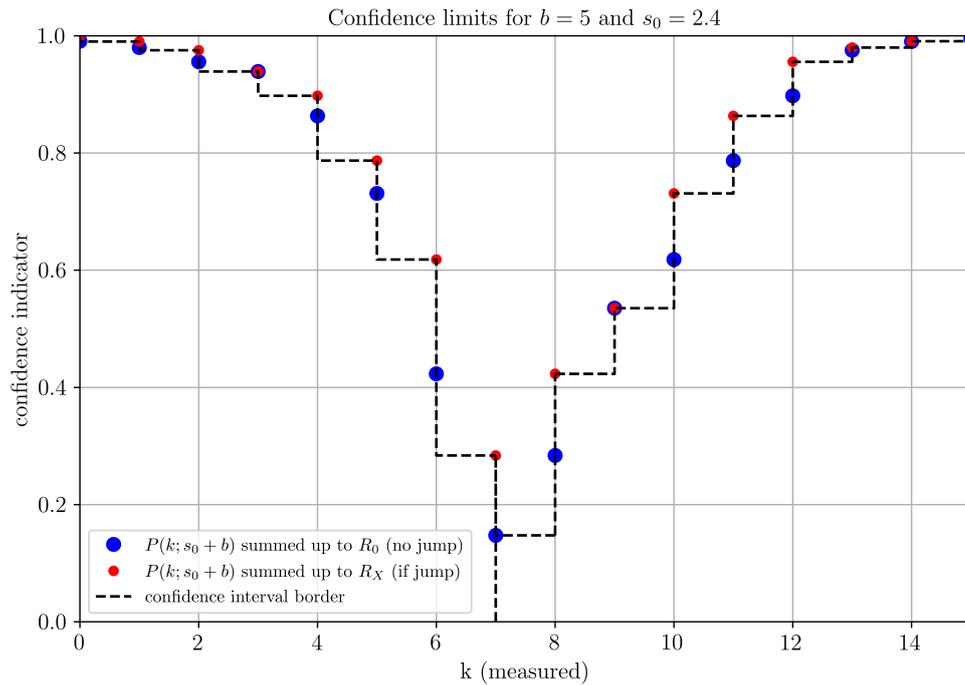
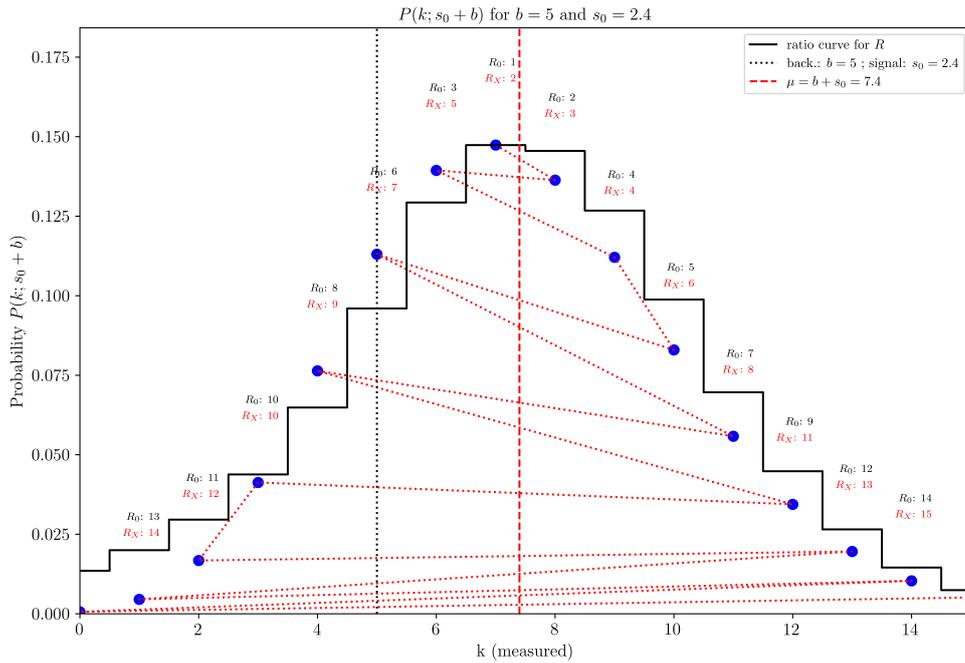


Figure 4.1: Construction procedure. The black step plot indicates the likelihood ratio (normed to match the height of the probability distribution). The summation according to the dotted red line produces the confidence indicators in the lower plot. Note that the actual confidence limits are calculated by comparing the red dots (including jumps) with the confidence limit and the simple summation of probability gives only the blue dots. Note that $X = R_X - R_0$ from the upper plot. Typically X has the value 0, 1 or 2.

of the algorithm, it only stops when the desired confidence level is surpassed (one blue point has to be above the desired confidence). The limits for any desired confidence level at this s_0 are read of simply from the dashed black line in and Figure 4.1.

One does this for all possible true values of s_0 , and builds a confidence belt of these intervals. If one does not specify the desired probability i.e. confidence level, and draws the number of observed events k on the x -axis and the true signal parameter s_0 on the y -axis, it is possible to get a representation of all possible confidence intervals in a 2D plot. Such a plot of calculated confidence limits is given in the upper figure of Figure 4.2. Note that the confidence limits are calculated along the x -axis, in the typical one dimensional case. The obtained confidence intervals for some measurement are read of on the y -axis at the position of the measurement result k . The confidence limits are not always a set of two values, but instead the interval can be shortly interrupted as seen in the corresponding middle and lower plot of Figure 4.2.

Even though the original paper from the late 1990s already mentions such behavior as a so called mild pathology³ common implementations as in CERN ROOT⁴ seem to simply ignore this admittedly rare cases, leading to errors as displayed in Figure 4.2.

Obviously the possible confidence limit values range from 0 to 1, and all values are reached on the x -axis (at least in the continuous case), but *not* for a fixed k along the y -axis. Therefore the sometimes read statement, that the Feldman-Cousins approach never gives an empty confidence interval is not fully correct. In fact there is always a confidence level, for which one can give a confidence interval and typically for the common choice $cl = 90\%$ or even $cl = 68\%$ (compare Gaussian 1σ) the interval is not empty. But for an arbitrary given confidence level (one could in theory e.g. choose 50% or even 10%), the confidence interval can be empty at some given measurement, which happens to be (even only slightly) below the expected background by fluctuation. One can guess that this can only happen for a measurement, which gives a result smaller than the boundary intersection of the maximum likelihood ratio line. In the Poisson with background example this is the linear red line in Figure 4.2, which hits the x -axis at $k = b$. Note that all interval constructions occur along the x -axis and around this line, therefore small intervals cannot be extended over all arbitrarily $k < b$ results. A very similar effect will also be seen again in subsection 5.8 for the amplitude estimation of a sine function. In general this “empty confidence interval case” is however more of an academic remark, which does not lead to major problems in practice.

In general the resulting structure of confidence intervals can be complex. Even in the simple scenario as for a Poisson distribution with background, one has to perform numerical calculations to predict the correct limits. Note that by construction always some over coverage in the confidence intervals is reached, as can be seen in Figure 4.3. Also the irregular “jumps” happen, because the measurement k is a discrete parameter.

³Citation from [46]: “We find that a mild pathology arises as a result of the fact that the observable n is discrete. When the vertical dashed line is drawn at some n [...], it can happen that the set of intersected horizontal line segments is not simply connected. When this occurs we naturally take the confidence interval to have μ_1 corresponding to the bottom-most segment intersected, and to have μ_2 corresponding to the top-most segment intersected.” Note that in their paper the measurement $k \equiv n$, while μ_1 and μ_2 denote the lower and upper limits.

⁴Call `TFeldmanCousins(0.91).CalculateUpperLimit(17,5)` to confirm this bug from the command line in `root`. The obviously wrong output is `(double) 5.4850000`, since it is only slightly larger than the calculated lower limit `(double) 5.4150000`. However there is at least a warning in the documentation that `TFeldmanCousins` is a legacy interface and that there will be no bug fixes or new developments.

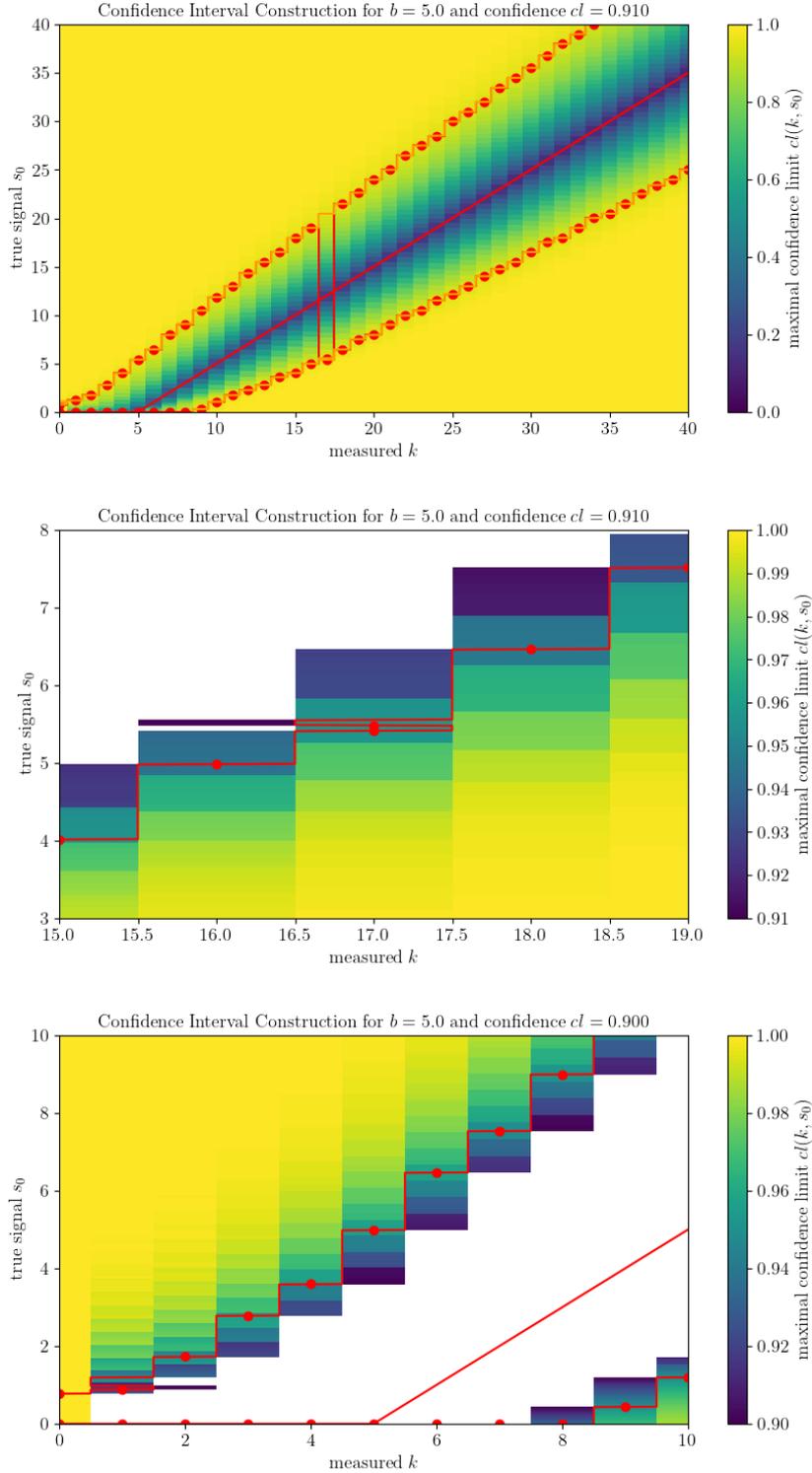


Figure 4.2: Feldman Cousins confidence limits (upper plot) for $b = 5$ and $c_l = 0.910$. The jump in the red confidence limits at $k = 17$ (see missing dot) is due to a bug in the CERN ROOT calculation. A correct calculation by hand gives the orange limits. In a detailed view (middle plot) also the reason for this problem can be observed. Regions with confidence values below the desired limit are white, to pronounce the construction effect, which produces an interrupted interval. The lower plot shows the small signal region with $c_l = 0.900$ (not $c_l = 0.910$). Note that such effects *typically* only produce small deviations in the confidence limits and striking errors as in the first plot are very rare.

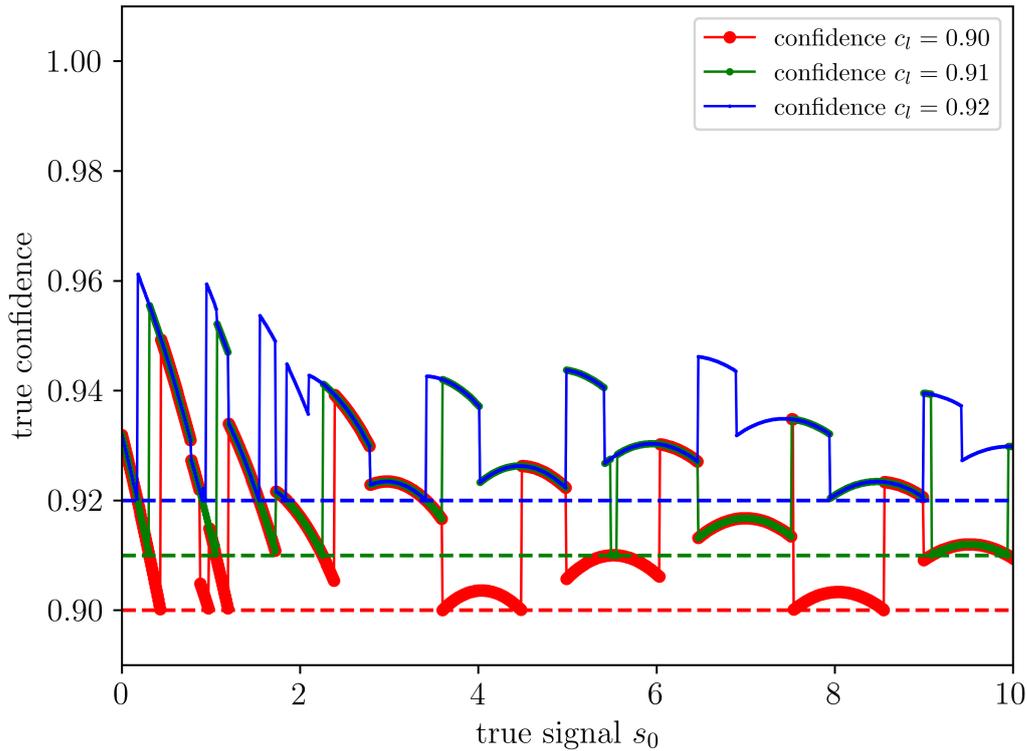


Figure 4.3: Overcoverage in the Feldman Cousins approach for discrete probability distributions (example for $b = 5$). Depending on the true signal the algorithm reaches actual confidence coverage of up to 0.95, for desired confidence limits of 0.90.

The good news is that many such problems only occur, because the underlying Poisson probability density is not continuous. For a continuous distribution the coverage can always exactly reach the desired confidence limit, therefore avoiding any “jumps”. However, still interesting effects, such as empty confidence intervals or discontinuous derivatives of the confidence limits can occur at points, where the construction reaches the boundary of the allowed parameter region. This cases will be discussed later in more detail for the concrete example of confidence limits for the amplitude estimator of a sine wave fit (see subsection 5.8), which follows a so-called Rice distribution.

4.2. The Downward Fluctuation Discussion

An effect which should always be considered when applying the Feldman Cousins approach to a statistical problem is the following: The downward fluctuation. This “paradox” can in principle also occur for continuous distributions and is not limited to the Poisson distribution with background discussed below.

Assume one has a known background of e.g. $b = 3.5$. Now assume one observes for signal plus background $k = 0$ events; so clearly less events than expected by background alone. Therefore the experiment must have observed a downward fluctuation of the background. Note that the 90% upper limit for the signal s_0 in this case is given by 0.75.

A “carefully designed experiment” with no background $b = 0$, also measuring $k = 0$ events, would have to give an upper limit of 2.44, which is counter-intuitive, since this seems to be the “worse” result (factor 3 in the limit). So does a downward fluctuation truly result in “better” upper limits?

One should note that there is a tendency to interpret frequentist confidence intervals as Bayesian objects. That is of course not true. One has to think in terms of repeated experiments. The Neyman construction, where Feldman and Cousins added their ordering principle to, simply calculates for all possible true parameters the regions of experimental results which cover a desired (e.g. 90%) fraction of the possible outcomes. Those confidence intervals then cover scenarios, where the measurement would indeed be in the desired fraction of possible outcomes. And since more drastic downward fluctuations are more unlikely, fewer true signal values cover the observed result in 90% of the experiments, typically only those with really small signal values. This leads to tighter confidence limits, but of course does *not* mean, that the interval obtained from some single experiment contains the true value with the probability given in the definition of the confidence limit. Therefore the discussed limits are not “wrong”, if interpreted correctly. However one typically likes to compare experiments, and the “better” or more sensitive experiment (less background) naturally is expected to give tighter confidence limits.

The original Feldman and Cousins paper addresses this problem as follows: “Our suggestion for [cases] in which the measurement is less than the estimated background, [is to] report both the upper limit and the *sensitivity* of the experiment, where the *sensitivity* is defined as the average upper limit that would be obtained by an ensemble of experiments with the expected background and no true signal [46].” A simpler option is to only state the confidence interval at the expected number of events, so for $s = 0$ and therefore at $k \approx s + b = b$, but since the measurement of k is discrete while the background is not necessarily, both methods are only similar for integer b . By introducing the sensitivity, less background indeed allows to quote tighter (average) limits.

To address the problem of confidence limits in the regime, where the measurements result is smaller or roughly equal to the expected background also some suggestions for corrections exist. A prominent one is the modified approach by Roe and Woodroffe, which introduces conditioning [47]. To quote Roe and Woodroffe one reason for their modifications were that “it seems unwise to regard lower than expected [background] as evidence⁵ against a value of [signal]”. Such modifications typically bring the confidence limits closer to corresponding Bayesian limits. Such modification will not be discussed here, as they typically again introduce conditions in the construction of the limits or mix Bayesian and Frequentist procedures, which can of course be justified given concrete examples, but does not make the interpretation of results easier.

One could of course also criticize the term *confidence limits* as misleading, because it is not true, that those limits represent confidence in the true value which is searched for in the measurement. Note that this point is also not unique to the Feldman Cousins approach. Take for example the so called *CLs method*, originally developed for the LEP Higgs searches and repeatedly use in further Higgs searches (until discovery). In 2011 [48] summarized it as follows: “The method’s name is very descriptive, but also misleading, as the CLs exclusion region is not a confidence interval. The method is neither purely frequentist

⁵Indeed, the smaller FC confidence limits in such $k < b$ cases should *not* be interpreted as direct evidence against a signal value. But the FC construction does not imply this.

nor Bayesian, instead its motivation is practical [...]. Despite its shaky foundations in statistical theory, it has been producing sensible results for over a decade.”

In summary Feldman Cousins deals with physical boundaries on parameters and avoids the so called flip-flop problem. This means that the procedure to quote limits for parameters is automated, and not chosen arbitrarily by the user depending on observed measurement results. Some “unintuitive” confidence limits are the price of the Feldman Cousins approach, but most confusion arises from the fact, that one tries to force Bayesian interpretation on a state-of-the-art purely frequentist approach. If one wants to quote Bayesian limits (also called *credible intervals*), one has to first introduce priors, which could also be criticized, since the justification for the shape of the priors is often not objective. Of course it may be also interesting to compare both the Bayesian and Frequentist approach, which will be done for the sine fitting problem later in this work.

5. Parameter Estimation in EDM experiments at COSY

5.1. In-Plane Spin Precession and Horizontal Polarization

The main problem of unfolding the idle spin precession in the horizontal plane is the fact, that the spin precession frequency $f_s = |\nu_s| \cdot f_{\text{COSY}} \approx 0.16 \cdot 750 \text{ kHz} = 120 \text{ kHz}$ is much faster than the event recording rate. In the experiments only every ≈ 25 th turn of the bunch one event is detected in each of the four detector quadrants. Thus it is not possible to calculate the up-down asymmetries, which are proportional to the horizontal polarization, from the precession in real-time. In the following, procedures to accumulate sufficient statistics during a macroscopic time (and therefore also turn) interval $\Delta T_{\text{mac}} = \Delta n$ are described.

The first analysis method is the so called mapping method. Since it is well described in Dennis Eversmanns PhD thesis *High Precision Spin Tune Determination at the Cooler Synchrotron in Jülich* (see pages 41-46 of [36]), the method itself will not be described here, but a description following his work and adjusting the definitions to match this work is given in the appendix Appendix D. An overview of this technique can also be found in the publication [49]. The occurring distributions are closely related to the concept of the statistical analysis described below.

The second analysis method is based on the discrete turn Fourier transform, which provides Fourier coefficients as estimators and will be discussed in detail in section 6. In both cases the estimator of the asymmetry amplitude parameter is biased. The statistical properties of the different possible estimators will be discussed in this chapter. Also the Cramér-Rao bound is determined in order to specify a lower bound of the statistical error of the estimated parameter, when using multiple possible approaches.

5.2. Up-Down Asymmetry and the Problem of Sine-Fitting

A turn number n is assigned to each recorded event. Then the so-called spin phase advance

$$\phi_s(n) = 2\pi\nu_s^0 n \quad (5.1)$$

of each event can be calculated from the product of the number of spin precessions in the horizontal plane and an assumed (fixed) spin tune ν_s^0 .

With all the detector events from a turn interval with length Δn mapped into one 2π spin phase advance interval by a sophisticated mapping method (see Appendix D for details) the asymmetry can be calculated from the difference of count rates in the upper and lower detector segments of the polarimeter normalized by the total number of counts. Its functional form is finally given by

$$\epsilon_{\text{fit}}(\phi_s) = \epsilon_{\text{UD}} \cos(\phi_s - \phi) + \epsilon_{\text{off}}. \quad (5.2)$$

Note that by construction of the full mapping method the maximal possible amplitude ϵ_{UD} is limited to one, since the count rates in the detectors cannot be negative and ideally $\epsilon_{\text{off}} = 0$. The amplitude is then proportional to the horizontal vector polarization, but this work will instead use ϵ_{UD} (denoted also simply as “amplitudes”), instead of the actual absolute polarization, since mainly the relative change compared to the initial polarization/asymmetry amplitude and the stability over different runs and cycles is of interest.

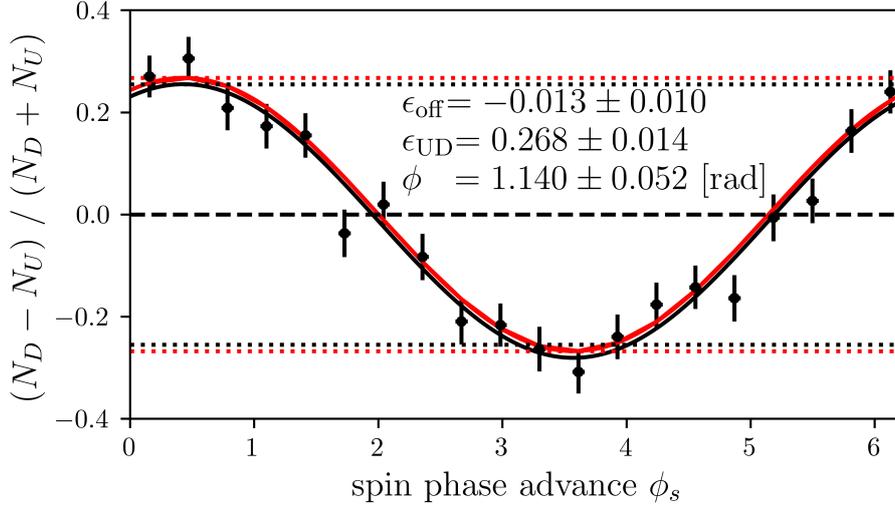


Figure 5.1: Asymmetry mapped from $\Delta n = 10^6$ turns and fitted with Equation 5.2 and Equation 5.3 to extract estimators for the amplitude ϵ_{UD} and phase ϕ .

In Figure 5.1 the asymmetry $\epsilon(\phi_s)$ is shown as an example for some measurement interval of $\Delta n = 10^6$ turns, which corresponds to a length of 1.3 s for the characteristic COSY deuteron revolution frequency of ≈ 750 kHz. Relevant is the size of the errorbars, which correspond to the events per bin. More events from longer macroscopic intervals $\Delta n \equiv \Delta T$ allow for a more precise fit, but the resolution will then be reduced in direction of time.

The oscillation is mapped from count rates to happen around zero. Therefore the data is sufficiently described using only a phase and an amplitude (two parameters), but one can also add an offset as a third parameter (compare the black lines with offset and red lines without in Figure 5.1). The fitted offset is typically compatible with zero and can be ignored.

To better compare different parameter estimation methods (see e.g. also the Fourier transform in section 6) the fitting function can be rewritten using the two orthogonal functions sine and cosine as

$$\epsilon_{\text{fit}}(\phi_s) = A \sin(\phi_s) + B \cos(\phi_s). \quad (5.3)$$

Using the fit results \hat{A} and \hat{B} as estimators⁶, the amplitude and phase

$$\hat{\epsilon}_{\text{UD}} = \sqrt{\hat{A}^2 + \hat{B}^2} \quad (5.4)$$

$$\hat{\phi} = \text{atan2}(\hat{A}, \hat{B}) \quad (5.5)$$

can be reconstructed, where atan2 denotes the arctangent, which provides correct signs for $0^\circ \leq \hat{\phi} < 360^\circ$ typically described in the interval $[-\pi, +\pi)$ (for details see [50] and [51]).

⁶From now on all estimators (this means the fitting results) will be denoted with a hat to differentiate between them from the underlying true parameter values.

The statistical errors without correlations between \hat{A} and \hat{B} calculated by Gaussian error propagation, which is good approximation to get a feeling for the behavior, are given by

$$\hat{\sigma}_\epsilon^2 = \frac{\hat{A}^2 \sigma_A^2 + \hat{B}^2 \sigma_B^2}{\hat{A}^2 + \hat{B}^2} \quad \text{for } \sigma_A = \sigma_B = \sigma \quad \Rightarrow \quad \hat{\sigma}_\epsilon^2 = \sigma^2 \quad (5.6)$$

$$\hat{\sigma}_\phi^2 = \frac{\hat{B}^2 \sigma_A^2 + \hat{A}^2 \sigma_B^2}{(\hat{A}^2 + \hat{B}^2)^2} \quad \text{for } \sigma_A = \sigma_B = \sigma \quad \Rightarrow \quad \hat{\sigma}_\phi^2 = \frac{\sigma^2}{\hat{\epsilon}^2} \quad (5.7)$$

Note that it is not entirely correct to assume zero correlation, as will be described in subsection 5.4 in a detailed analysis discussing the statistical error, but this term will typically be small. It can and will be shown explicitly for a maximum likelihood fit, that the estimators in sine fitting approach the Fisher information limit corresponding to $\sigma_\epsilon^{\text{limit}} = \sqrt{2/N}$ with the total number of events N in the considered turn interval Δn . For a sufficient number of bins this result is also reached by the simpler binned least squares χ^2 -fit, as performed in Figure 5.1.

The following assumptions must hold in order to determine and to unfold the asymmetry properly during a macroscopic time interval $\Delta T = 1/f_{\text{COSY}} \cdot \Delta n$: The (average) spin tune of the beam has to be stable during the interval, otherwise the sinusoidal functional form of the event distribution would get smeared and thus the amplitude would be underestimated. Furthermore, the in-plane vector polarization should not drop much during a macroscopic turn interval Δn . This would cause a false amplitude estimation because the depolarization effect is in general not linear. Generally speaking the macroscopic time interval ΔT has to be small compared to the spin coherence time (see subsection 7.2). A larger SCT therefore also allows for a larger turn interval, which directly results into more precision of the fit and correspondingly the amplitude estimation.

5.3. Amplitude Determination and Bias

From both methods one gets two parameter estimators depending on the chosen reference spin tune: $\hat{A}(\nu_s^0)$ and $\hat{B}(\nu_s^0)$. In the case of the mapping method, the estimators correspond to the parameters of the asymmetry fit (sine and cosine term). For the discrete turn Fourier transform $\hat{A}(\nu_s^0)$ and $\hat{B}(\nu_s^0)$ are simply given by the Fourier coefficients, which are orthogonal, as well.

The problem of fitting bias can be explained as follows. Suppose there is no asymmetry amplitude (no signal) in the data at all. However the two fit estimators \hat{A} and \hat{B} , which describe the sine and cosine component in the data, will be normally distributed random variables and have some (Gaussian) uncertainty. Then the estimated amplitude as in Equation 5.4 is biased in the positive direction, since it is defined as the square root of the sum of the squares of both the estimators. The only way to obtain a zero amplitude, would be to have exactly zero for both \hat{A} and \hat{B} . This means that even if the estimators \hat{A} and \hat{B} are consistent, asymptotic normal, efficient and unbiased (which is the ideal case), the amplitude is systematically overestimated, which becomes more and more significant for small amplitudes and low statistics.

This bias problem of course also occurs when fitting the amplitude P directly, since then the phase ϕ is unknown, and the data still fluctuates randomly. The fitting procedure again finds an amplitude representing this fluctuations.

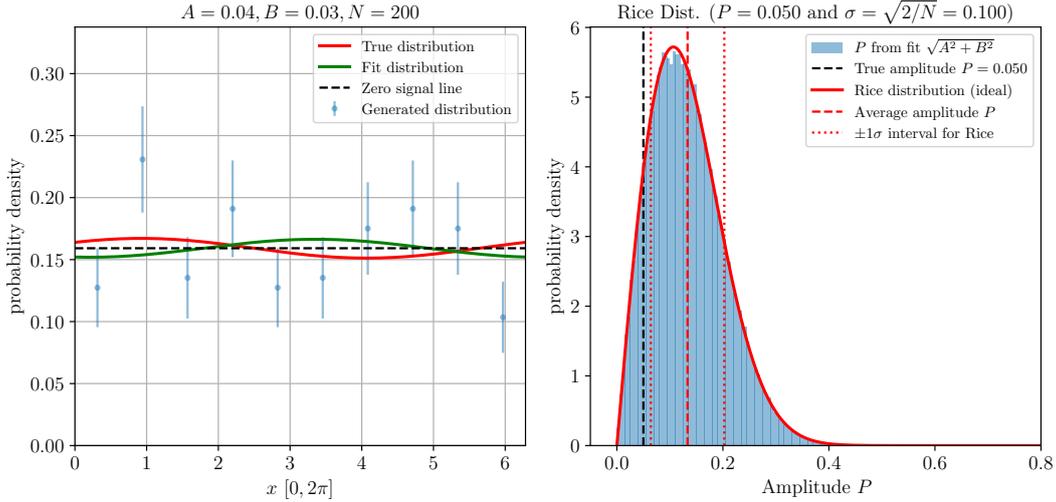


Figure 5.2: Distribution of the amplitude estimator (200000 repetitions), based on random data following a sine with small amplitude and large uncertainties analyzed by a maximum likelihood fit. The binned data with scaled Poisson errorbars in the left plot is for illustration. $N = 200$ points were used in the MLE estimation, which corresponds to $\sigma_P \approx \sqrt{2/N} = 0.1$. The results for the fitted amplitudes (right plot) follow a Rice distribution. The mean of the rice distribution is more than 1σ away from the true P , it is therefore substantially biased.

For simplicity consider the following toy model for the amplitude:

$$f(x; A, B) = \frac{1}{2\pi}(1 + A \sin(x) + B \cos(x)), \quad (5.8)$$

This function is linear in all fitting parameters and represents a probability density for observed events, which are then typically binned in $x \in [0, 2\pi)$. The corresponding amplitude is again given by $P = \sqrt{A^2 + B^2}$, so that P is always positive. The most simple case is to assume no true signal in the data; e.g. also relevant for axion searches, which are possible using COSY data as done in 2023 [52]. Even if the true amplitude $P = 0$ (expected from $A = 0$ and $B = 0$), one finds amplitudes \hat{P} different from zero from the estimators – no matter if maximum likelihood, Fourier transform or χ^2 -fits are used. This is again due to the mentioned bias in fitting. For small true amplitudes $P > 0$ and relatively large uncertainties (fluctuations in binned data), the bias effect is also relevant, as can be seen in Figure 5.2. Actually the estimated amplitude \hat{P} , given some true amplitude P follows a so called rice distribution, which is clearly not Gaussian. The rice distribution will be discussed in subsection 5.6. The model from Equation 5.8 is also relevant for general signal processing [40, 53].

5.4. Statistical Uncertainty of A and B

Using the theory of maximum likelihood estimation, one can calculate the asymptotic variance for the estimators \hat{A} and \hat{B} given N measurements following the PDF in Equation 5.8. The full calculation with intermediate steps can be found in subsection B.1 in

the appendix. Starting from the probability density function, the log-likelihood function reads

$$\ell = \ln\left(\prod_{i=1}^N f(x_i; A, B)\right) = \sum_{i=1}^N \ln\left(\frac{1}{2\pi}\right) + \ln(1 + A \sin(x_i) + B \cos(x_i)). \quad (5.9)$$

By calculating the second derivatives $\frac{\partial^2 \ell}{\partial A^2}$, $\frac{\partial^2 \ell}{\partial B^2}$ and $\frac{\partial^2 \ell}{\partial A \partial B}$ and their expectation values, one finds the inverse of the covariance matrix for $(\theta_1, \theta_2) \equiv (A, B)$ using the abbreviation $C = \sqrt{1 - A^2 - B^2} = \sqrt{1 - P^2} \leq 1$ for $0 \leq P \leq 1$:

$$\text{cov}^{-1}(A, B) = -\left\langle \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right\rangle = \begin{pmatrix} \frac{-(A^2[C-1]+B^2[C^2-C])N}{CP^4} & \frac{(C-1)^2 ABN}{CP^4} \\ \frac{(C-1)^2 ABN}{CP^4} & \frac{-(B^2[C-1]+A^2[C^2-C])N}{CP^4} \end{pmatrix}. \quad (5.10)$$

Note that the results are fully symmetrical $A \Leftrightarrow B$. From the inversion of the inverse covariance matrix one finds the variances. Note the typical behavior of all entries $\propto N^{-1}$:

$$\text{cov}(A, B) = \frac{1}{\det(\text{cov}^{-1}(A, B))} \cdot \text{adj}(\text{cov}^{-1}(A, B)) = \begin{pmatrix} \frac{A^2 C + B^2}{N(1-C)} & -\frac{AB}{N} \\ -\frac{AB}{N} & \frac{A^2 + B^2 C}{N(1-C)} \end{pmatrix}. \quad (5.11)$$

$$\sigma_A = \sqrt{\frac{A^2 C + B^2}{N(1-C)}} \quad \sigma_B = \sqrt{\frac{A^2 + B^2 C}{N(1-C)}} \quad \text{cov}_{A,B} = -\frac{AB}{N} \quad (5.12)$$

The correlation coefficient for A and B is only zero for $A = 0$ or $B = 0$. The results for σ_A and σ_B in the interesting case of small amplitudes $P \ll 1$ are given by

$$\sigma_A^2 = \frac{A^2 \sqrt{1 - P^2} + B^2}{N(1 - \sqrt{1 - P^2})} \stackrel{P^2 \ll 1}{\approx} \frac{2}{N} \left(1 - \frac{3P^2 - 2B^2}{4} - \frac{P^4}{16}\right). \quad (5.13)$$

$$\sigma_B^2 = \frac{A^2 + B^2 \sqrt{1 - P^2}}{N(1 - \sqrt{1 - P^2})} \stackrel{P^2 \ll 1}{\approx} \frac{2}{N} \left(1 - \frac{3P^2 - 2A^2}{4} - \frac{P^4}{16}\right) \quad (5.14)$$

Notice that the only remaining term for $A = B = P = 0$ is $2/N$. For small amplitudes $0 < P \ll 1$ corrections do apply, always reducing σ_A and σ_B because the $\mathcal{O}(P^2)$ term is negative.

The error on $P = \sqrt{A^2 + B^2}$ and on $\varphi = \arctan(A/B)$ can be calculated from simple Gaussian error propagation. Note that a calculation of σ_P and σ_φ from the maximum likelihood fit of the probability density $f(x; P, \varphi) = \frac{1}{2\pi}(1 + P \cos(x - \varphi))$ as done in [54] (see also subsection B.2 in the appendix) gives exactly the same results for the uncertainties, but e.g. does not explain correlations between the sine and cosine term as shown above.

To get the behavior for small amplitudes $P^2 \ll 1$, one can perform a Taylor expansion to $\mathcal{O}(P^4)$.

$$\sigma_P^2 = \frac{-1}{\left\langle \frac{\partial^2 \ell}{\partial P^2} \right\rangle} = \frac{1}{N} \frac{P^2}{\sqrt{1/(1 - P^2)} - 1} \stackrel{P^2 \ll 1}{\approx} \frac{2}{N} \left(1 - \frac{3}{4}P^2 - \frac{1}{16}P^4\right), \quad (5.15)$$

$$\sigma_\varphi^2 = \frac{-1}{\left\langle \frac{\partial^2 \ell}{\partial \varphi^2} \right\rangle} = \frac{1}{N} \frac{\sqrt{1/(1 - P^2)}}{\sqrt{1/(1 - P^2)} - 1} \stackrel{P^2 \ll 1}{\approx} \frac{2}{NP^2} \left(1 - \frac{1}{4}P^2 - \frac{1}{16}P^4\right). \quad (5.16)$$

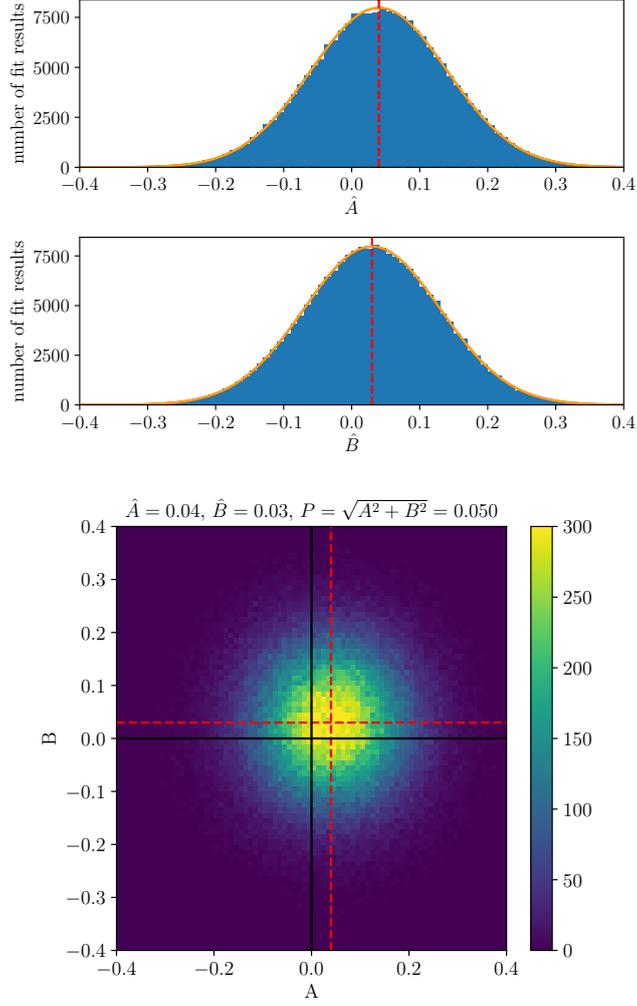


Figure 5.3: Upper two plots: Distributions of the estimators of the parameters \hat{A} and \hat{B} based on random data analyzed by a maximum likelihood fit. $N = 200$ points were used in the MLE estimation, which corresponds to $\sigma_A \approx \sigma_B \approx \sigma_P \approx \sqrt{2/N} = 0.1$. The results for \hat{A} and \hat{B} follow (unbiased) Gaussian distributions, precisely represented by the expected values in mean (here $A = 0.04$ and $B = 0.03$) and uncertainty given from Equation 5.12. Lower plot: 2D representation of fit results for the parameters \hat{A} and \hat{B} .

The approximation $\sigma_P \approx \sqrt{\frac{2}{N}} \equiv \sigma_0$ is the zero order result. The statistical error of the phase scales reciprocally with the amplitude, therefore a larger amplitude makes the phase estimation more precise. If there is no amplitude ($P = 0$), also φ is undefined and a Gaussian error approximation for the phase is not valid.

However one should notice, that the results discussed above are strictly speaking only valid for $N \gg 1$, since the maximum likelihood estimator reaches the Cramér-Rao Bound for $N \rightarrow \infty$. Note that for a sufficiently large number of bins the uncertainties derived here for the unbinned likelihood method coincide with the uncertainties of a binned least squares fit to the data.

5.5. Three-parameter fit for amplitude estimation

Another possible fitting option using matrix calculation is taken from [40] and works as follows. The binned data is given as vectors $\mathbf{y} = (y_1, \dots, y_n)^T$ (with the number of bins $n \gg 3$) and $\mathbf{x} = (x_1, \dots, x_n)^T$, where x is the position of the bins and y the counted events per bin. The fitting model in this case is given by

$$y_n[A, B, C] = A \cos(x_n) + B \sin(x_n) + C. \quad (5.17)$$

A , B , and C are unknown constants, which are sought for in the estimation procedure. Note that one can easily introduce a *known* additional parameter in the procedure e.g. the angular frequency ($x \equiv \omega t_n = 2\pi f t_n$), or the spin tune ($x \equiv 2\pi \nu_s n$), if this parameter can assumed to be fixed. The use of A , B and C is favorable because all the sought for parameters enter the model linearly. In the first step, the three-parameter fit is employed yielding estimates \hat{A} , \hat{B} , and \hat{C} . In the second step, estimated \hat{A} and \hat{B} values are plugged in to $\hat{P} = \sqrt{\hat{A}^2 + \hat{B}^2}$ to obtain an estimate of the amplitude, as done also in the corresponding MLE and Fourier approach. Because of the linearity this problem can be treated with linear algebra. The parameters are gathered in the parameter vector $\hat{\mathbf{x}} = (A, B, C)^T$ and the binned data is written as $\mathbf{y} = (y_1, \dots, y_N)^T$. Then, \mathbf{y} obeys the over determined set of linear equations

$$\mathbf{y} = \mathbf{D}\hat{\mathbf{x}}, \quad (5.18)$$

where \mathbf{D} is the $n \times 3$ matrix

$$\mathbf{D} = \begin{pmatrix} \cos x_1 & \sin x_1 & 1 \\ \cos x_2 & \sin x_2 & 1 \\ \vdots & \vdots & \vdots \\ \cos x_n & \sin x_n & 1 \end{pmatrix} \equiv (\cos(\mathbf{x}), \sin(\mathbf{x}), \mathbf{1}). \quad (5.19)$$

The least squares solution $\hat{\mathbf{x}} = (\hat{A}, \hat{B}, \hat{C})^T$ is given by

$$\hat{\mathbf{x}} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{y}. \quad (5.20)$$

As can be seen in Figure 5.4, for small amplitudes and large statistical uncertainties all biased amplitude estimators \hat{P} follow the expected Rice distribution very well.

If a binned fits is used, there can be noticeable deviations especially for few bins $n_{\text{bins}} = 10$, but those become only noticeable when the statistical uncertainties in the fit are small. This effect can be seen in Figure 5.5. It is possible to reach a good agreement with the expected Rice distribution (red line), by always using the Maximum Likelihood Method or the Fourier Transformation, of when more bins (here $n_{\text{bins}} \approx 100$) are used for the amplitude estimation.

An interesting effect occurs for $P \approx 1$, which is described in Figure 5.6. However, such large amplitudes are not obtained in the experiment. Typically the (asymmetry) amplitude related to the horizontal precession is well below 0.3), therefore even zero order approximation of $\sigma_P \approx \sigma_0$ is considered sufficient.

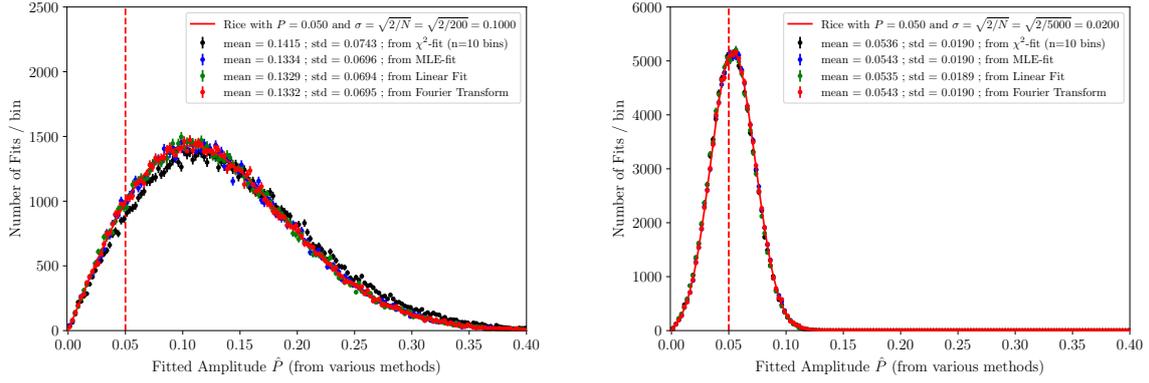


Figure 5.4: Comparison of results for various fitting methods in case of a small true amplitude $P = 0.05$ for $N = 200$ (left, $\sigma_0 = 0.1$) and $N = 5000$ (right, $\sigma_0 = 0.02$) events mapped into $[0, 2\pi)$. All methods roughly follow the expected Rice distribution derived for the MLE approach, when 100000 repetitions of the fitting procedure using random data are performed.

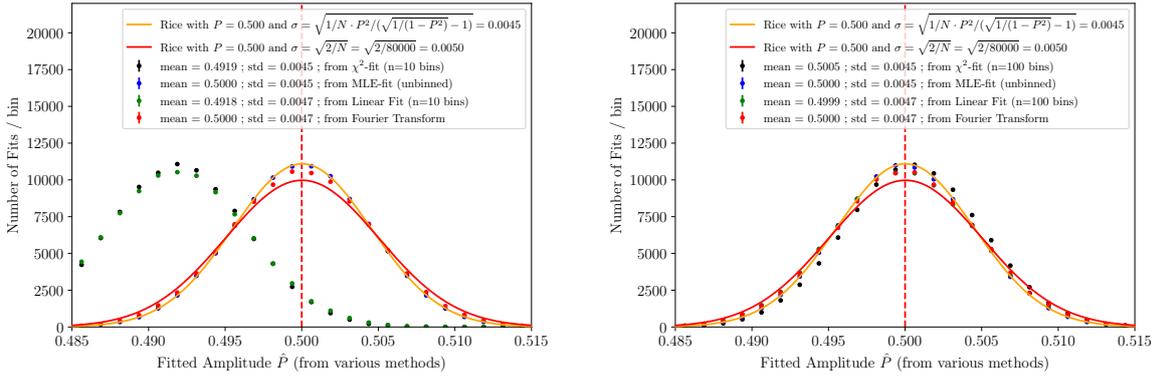


Figure 5.5: Effect of using too few bins on fit results of \hat{P} for the binned χ^2 -fit and Linear Fit using the three parameters $[\hat{A}, \hat{B}, \hat{C}]$ vector. The results for the simple χ^2 -fits and the three parameter estimators given in subsection 5.5 significantly differs from the unbinned methods, but only for a small number of bins ($n = 10$ bins in the left and $n = 100$ in the right plot). Also a very large number of events (here $N = 80000$) has to be used, that ensures small uncertainties to resolve the differences.

5.6. Rice Distribution

It can be shown also analytically that the estimator for the amplitude in fact follows a Rice distribution (see Appendix C taken from [54] and [55]). The probability density of the rice distribution is given by:

$$f(\hat{P}|P, \sigma)d\hat{P} = \frac{1}{\sigma^2} e^{-(\hat{P}^2 + P^2)/(2\sigma^2)} \hat{P} I_0 \left(\frac{\hat{P}P}{\sigma^2} \right) d\hat{P} \quad (5.21)$$

where P is the true amplitude and \hat{P} is the corresponding estimated amplitude. I_0 is the modified Bessel function of the first kind. σ is the uncertainty parameter calculated

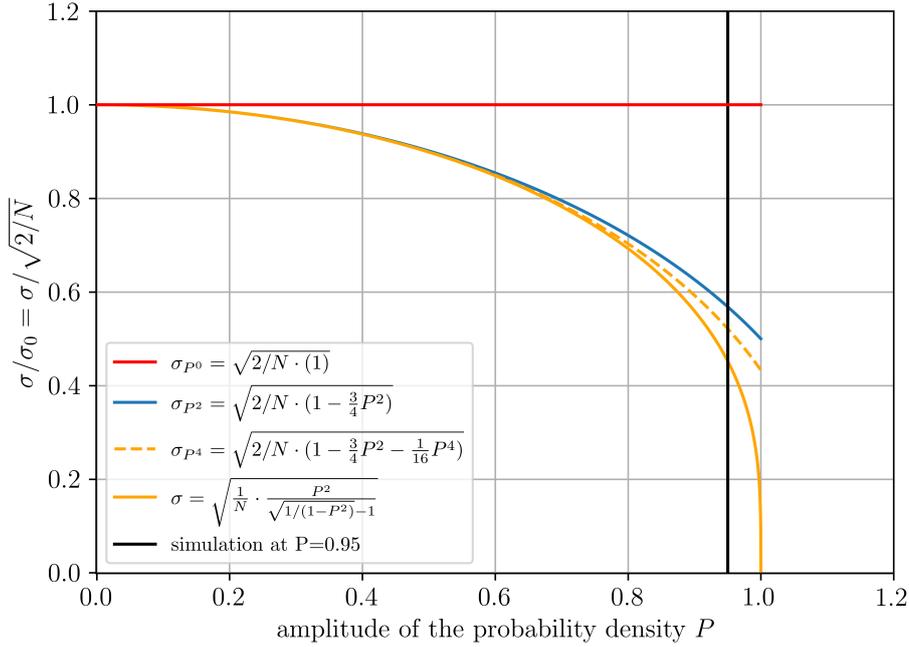
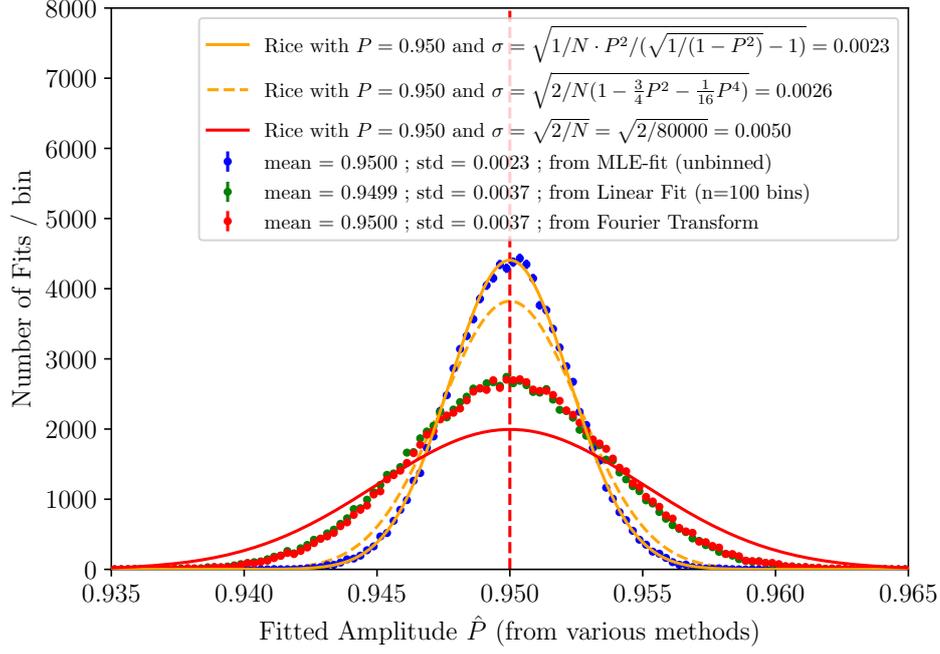


Figure 5.6: Upper plot: In the case of large amplitudes (here $P = 0.95$), the width of the obtained distributions of \hat{P} estimators is noticeably smaller for the MLE (blue dots) when compared to the Fourier transform (red dots), as well as the linear binned three parameter approach (green dots). A very large number of events (here $N = 80000$) is used for the fits. The dots are the number of fitting results with Poisson error per estimated amplitude bin each. In total there were 100000 repetitions of the fit. Lower plot: The MLE sigma parameter is indeed described according to subsection 5.4; compare the solid orange line in both plots. For large P the variance of the estimator goes down steeply, therefore $\sigma_0 = \sqrt{2/N}$ (solid red line) and typical Taylor expansions are not a good representation of the MLE \hat{P} estimator distribution anymore.

from error propagation, which takes into account the correlation between A and B , based on the uncertainties of A and B , as introduced in Equation 5.15.

Also one should *not* misinterpret the σ parameter in Equation 5.21 as the square root of the variance of the distribution. This relation would only be valid in the Gaussian limit, which is given by $\frac{\sigma}{P} \rightarrow 0$, but not for $P \approx \sigma$. Typically even the zero order approximation $\sigma_A \approx \sigma_B \approx \sigma \approx \sqrt{2/N}$ is sufficient to get a decent representation of the distribution especially for small amplitudes. N is still the number of observed events used for the amplitude parameter estimation. In the data analysis of COSY experiments larger N can be obtained by simply choosing longer macroscopic turn intervals $\Delta T = \Delta n$ for each fitting procedure.

If one investigates the mean and variance of the Rice distribution the following results can be derived

$$\mu_{\text{rice}} = \sigma \sqrt{\frac{\pi}{2}} L_{1/2} \left(-\frac{1}{2} \left(\frac{P}{\sigma} \right)^2 \right), \quad (5.22)$$

$$\sigma_{\text{rice}}^2 = 2\sigma^2 + P^2 - \frac{\pi\sigma^2}{2} L_{1/2}^2 \left(-\frac{1}{2} \left(\frac{P}{\sigma} \right)^2 \right). \quad (5.23)$$

where $L_q(\cdot)$ denotes a Laguerre polynomial and $L_q^2(\cdot)$ is its square. For the case $q = 1/2$ the Laguerre polynomial can be expressed using Bessel functions of first kind

$$L_{1/2}(x) = e^{x/2} \left[(1-x) I_0 \left(-\frac{x}{2} \right) - x I_1 \left(-\frac{x}{2} \right) \right]. \quad (5.24)$$

It is seen that as P becomes large compared to σ the mean approaches P and the variance becomes σ^2 . One way to describe the transition to a Gaussian approximation is described in subsection C.1 in the appendix.

For $P = 0$ the Rice distribution is equal to the probability density of the Rayleigh distribution and therefore

$$f(\hat{P}; 0, \sigma) = \frac{\hat{P}}{\sigma^2} e^{-P^2/(2\sigma^2)}, \quad P \geq 0, \quad (5.25)$$

where σ is again the scale parameter of the distribution, but not the standard deviation. In this limiting case the expectation value is given by $\mu_{\text{rice}, P=0} = \sqrt{\frac{\pi}{2}}\sigma$ and the standard deviation is $\sigma_{\text{rice}, P=0} = \sqrt{\frac{4-\pi}{2}}\sigma$. The mode with the maximum of the PDF for $P = 0$ is given by σ .

It is convenient to rewrite the Rice distribution from Equation 5.21 in terms of the relative amplitude \hat{P}/σ , which is a term usually called signal strength:

$$f \left(\frac{\hat{P}}{\sigma} \middle| \frac{P}{\sigma} \right) d\hat{P} = e^{-\frac{1}{2} \left(\left(\frac{\hat{P}}{\sigma} \right)^2 + \left(\frac{P}{\sigma} \right)^2 \right)} \frac{\hat{P}}{\sigma} I_0 \left(\frac{\hat{P} P}{\sigma \sigma} \right) d \left(\frac{\hat{P}}{\sigma} \right). \quad (5.26)$$

The distribution now only depends on ratios of the amplitude and its uncertainty parameter σ . Defining $\hat{\epsilon} = \hat{P}/\sigma$ and $\epsilon = P/\sigma$ one can simply write:

$$f(\hat{\epsilon} | \epsilon) d\hat{P} = e^{-\frac{1}{2}(\hat{\epsilon}^2 + \epsilon^2)} \hat{\epsilon} I_0(\hat{\epsilon} \epsilon) d\hat{\epsilon}. \quad (5.27)$$

The standard deviation σ_{rice} and the corresponding uncertainty parameter σ are shown in the left plot of Figure 5.7. The right plot shows the bias between mean of the distribution

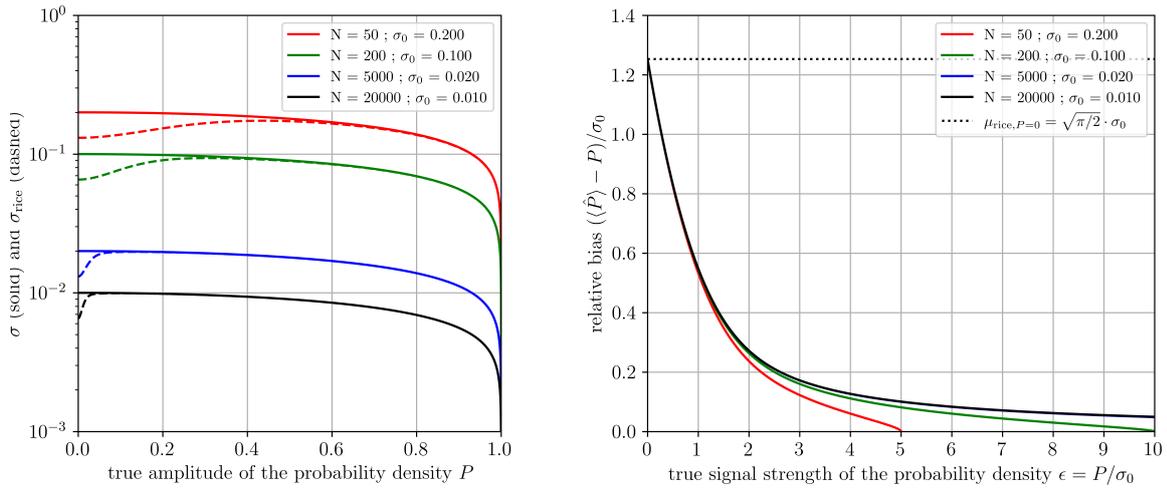


Figure 5.7: Left: Uncertainty parameter σ_P and standard deviation σ_{Rice} of the Rice distribution for various used events N . The limit in case $P = 0$ comes from the Rayleigh distribution. The Rice distribution is narrower for small amplitudes than the σ parameter. For large $P \approx 1$ the behavior of the standard deviation is dominated fully by the behavior of the sigma parameter. Right: Corresponding relative bias in the Rice distribution relevant for amplitude parameter estimation.

and true underlying amplitude P in units of the parameter σ_0 , as a function of the true signal strength $\epsilon = P/\sigma_0$. Only for small signal strengths the bias is similar in size to the uncertainty of the parameter. For larger signal strengths the bias goes down $\propto \epsilon^{-1}$. Note that technically the bias does depend on the uncertainty parameter σ (the number of data points used in the fit), but this effect only becomes relevant if ϵ approaches the *physical* meaningful limit of the signal strength given by $\epsilon_{\text{limit}} = 1/\sigma_0$, which corresponds to $P = 1$ and does not happen in the COSY experiment. Therefore the relative bias can be described with an universal function using $N \rightarrow \infty$. Effectively already $N \approx 2000$ is sufficient for this description (see right plot of Figure 5.7).

In Figure 5.9 some properties relevant for the analysis of the Rice distribution are shown. The solid red line (maximum of $f(\hat{\epsilon}|\epsilon)$ in direction of the true signal strength ϵ , which defines also ϵ_{best}) is equal to the maximum of the likelihood ratio $\max[R(\hat{\epsilon}|\epsilon)] = 1$, with $R(\hat{\epsilon}|\epsilon) = \frac{f(\hat{\epsilon}|\epsilon)}{f(\hat{\epsilon}|\epsilon_{\text{best}})}$.

This implies, that the most likely true ϵ is zero for $\hat{\epsilon} \leq \sqrt{2} \approx 1.41$. Note that the “inverse” distribution $\tilde{f}(\epsilon|\hat{\epsilon})$ constructed from a Bayesian approach using a constant prior will have the same maxima in direction of the true signal strength ϵ .

From Figure 5.9 one can also read off, that the peak (mode) of the rice distribution (dashed red line) is limited to $\hat{\epsilon} \geq 1$ and equality is reached for vanishing true signal amplitude. The mean of the distribution (pointed line) is limited to the region $\hat{\epsilon} \geq \sqrt{\pi/2}$, where equality is again reached for $\epsilon = 0$ representing the Rayleigh distribution.

Both mean and mode approach the true amplitude ϵ for $\gamma \gg 1$ from below, which means that the naive estimate $\hat{\epsilon} = \epsilon$, which assumes that the measured amplitude would be the true amplitude, is always positively biased. However since $\epsilon > 1$ implies also $P > \sigma$, the bias is considerably smaller than the uncertainty σ of the measurements even for

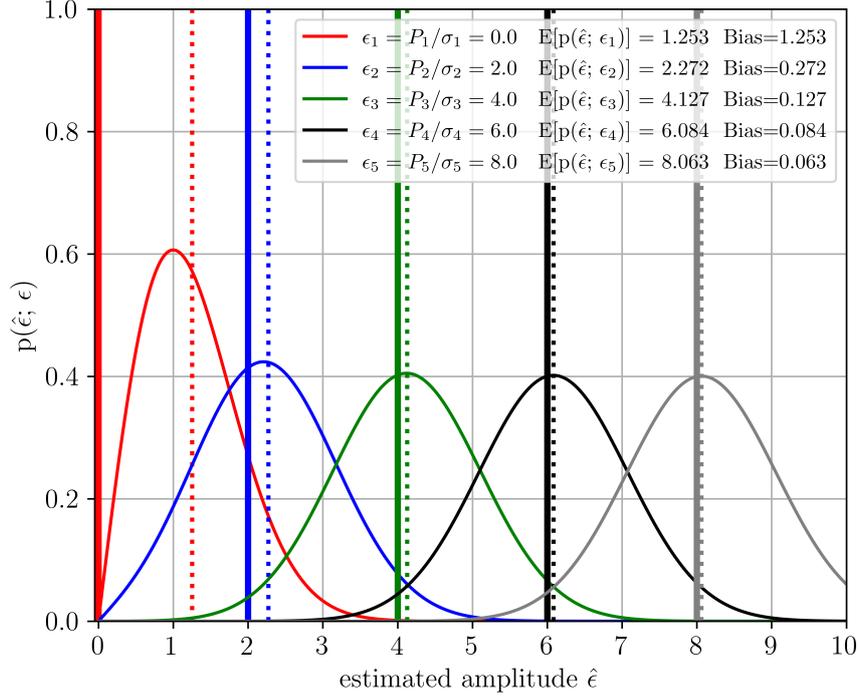


Figure 5.8: Rice probability distribution $f(\hat{\epsilon}|\epsilon)$ for some true signal strengths $\epsilon \in [0, 2, 4, 6, 8]$, with the average estimated signal strength $\hat{\epsilon}$ marked as vertical dashed line. Also note that the rice distributions are well described by a normal distribution if the signal strength $\epsilon \gg 1$ ($\sigma \ll P$).

relatively small signal strengths i.e. for $\epsilon > 3$. Note that at this point, also all relevant properties of the Rice distribution already coincide, but a small bias is still observable as difference to the black diagonal $\hat{\epsilon} = \epsilon$ line.

A simple approximation of the true amplitude signal strength ϵ (given some estimated amplitude signal strength $\hat{\epsilon}$) can be calculated from [56]

$$\epsilon \approx \hat{\epsilon} - \frac{1 - e^{-\hat{\epsilon}^2}}{2\hat{\epsilon}}. \quad (5.28)$$

It is drawn as a blue line in the right plot of Figure 5.9. Note that in the region of small signals with $\epsilon < 1.5$ the formula starts to deviate from the description of the Rice probability mode. Nevertheless, this simple analytic function has the advantage, that it is continuous and the estimated true amplitude is positive for all estimated amplitudes (zero only for $\hat{\epsilon} = 0$). Equation 5.28 also transforms smoothly with the signal strength, while the Feldman Cousins and Bayesian approach put the (most probable) amplitude to zero for multiple estimated amplitudes. For large signal strengths $\epsilon \gg 1$ the bias in the fit behaves as

$$\langle \hat{\epsilon} \rangle - \epsilon \approx \frac{1 - e^{-\hat{\epsilon}^2}}{2\hat{\epsilon}} \approx \frac{1}{2\hat{\epsilon}}, \quad (5.29)$$

$$\langle \hat{P} \rangle - P \approx \sigma^2 \cdot \frac{1 - e^{-\frac{\hat{P}^2}{\sigma^2}}}{2\hat{P}} \approx \sigma^2 \cdot \frac{1}{2\hat{P}} = \sigma \cdot \frac{1}{2\hat{\epsilon}}. \quad (5.30)$$

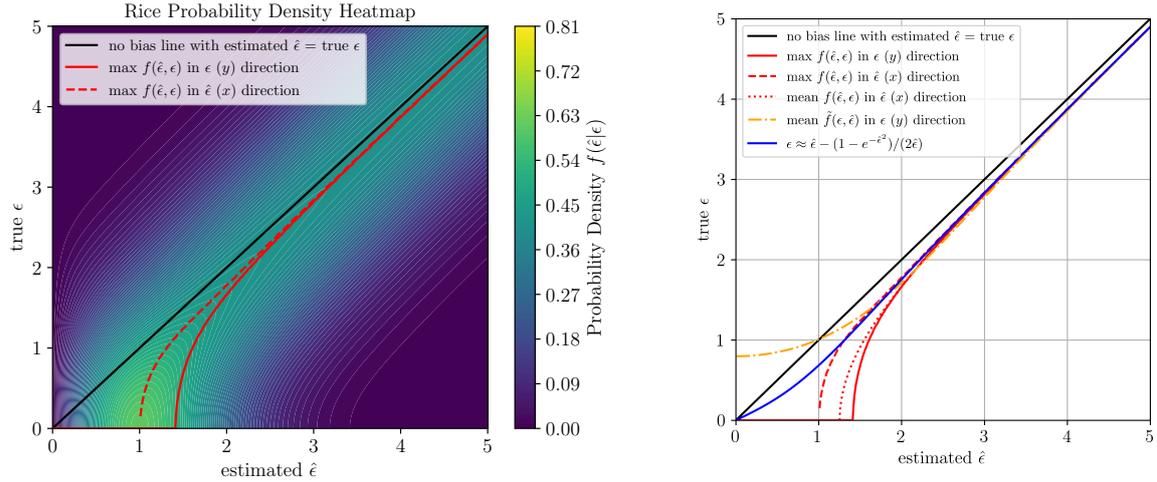


Figure 5.9: Left plot: Two dimensional representation of all possible Rice distributions (for various true signal strengths). Right plot: The corresponding “property lines” such as mean, mode and maximum likelihood ratio line for the Rice distribution for various signal strengths.

One should note that if $\hat{\epsilon} \gg 1$, the bias is therefore much smaller than the uncertainty of the fit, which is of order σ .

5.7. Construction of Confidence Intervals

The Feldman-Cousins algorithm [46] is used to construct the confidence intervals for the Rice distribution. At a given signal strength ϵ the algorithm selects all values of $\hat{\epsilon}$ for which the ratio

$$R(\hat{\epsilon}, \epsilon) = \frac{f(\hat{\epsilon}|\epsilon)}{f(\hat{\epsilon}|\epsilon_{\text{best}})} \quad (5.31)$$

has the largest values until the desired coverage (confidence level cl) of the confidence interval is reached. In the likelihood ratio ϵ_{best} denotes the value of ϵ for which the probability density $f(\hat{\epsilon}|\epsilon)$ has its maximum in the allowed region of ϵ , i.e. $f(\hat{\epsilon}|\epsilon_{\text{best}}) = \max\{f(\hat{\epsilon}|\epsilon)\}$. ϵ_{best} as a function of $\hat{\epsilon}$ is shown in Figure 5.10 as the solid red line. It simply is the maximum of all the possible probability densities in direction of the true signal strength ϵ (where ϵ is limited to the physical region with amplitude $P > 0$). Note that for the following plots $\hat{\epsilon}$ (corresponding to the estimator result) will be plotted in the x -direction, while the true signal strength ϵ is shown in y -direction. Feldman Cousins confidence intervals are constructed in horizontal direction, while they are read off in vertical direction for a given $\hat{\epsilon}$. The construction itself is independent of the measured amplitude value.

Figure 5.10 shows likelihood ratios and the corresponding confidence limits constructed for the signal strength $\epsilon = P/\sigma$. The 68%, 90% and 99% confidence interval limits are shown as dashed black lines. Given the measurement at some signal strength $\hat{\epsilon}$ one reads off the confidence interval in vertical direction (intersection with dashed lines). To get a confidence interval for the amplitude one simply rescales the obtained true signal strength

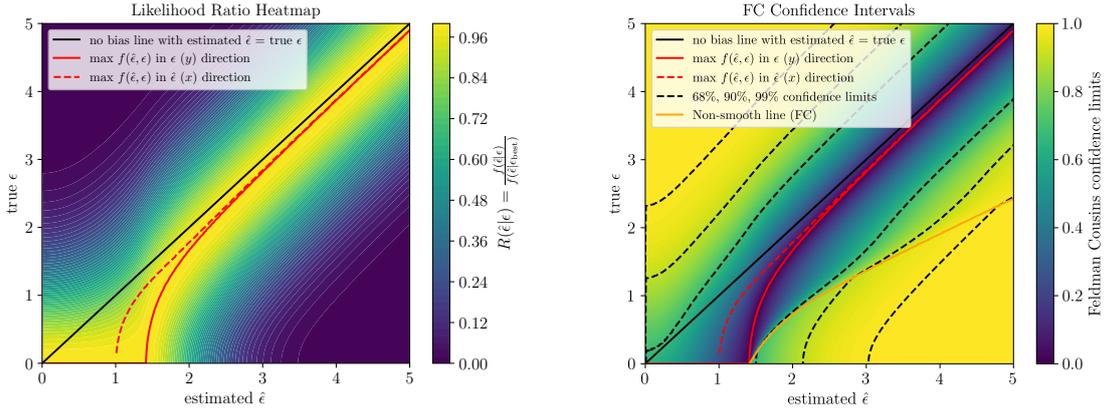


Figure 5.10: Left: The likelihood ratio $R(\hat{\epsilon}, \epsilon)$ as a function of the signal strength $\hat{\epsilon}$ for various true signal strengths ϵ . Right: Feldman-Cousins confidence intervals for the Rice distribution in units of the signal strength.

with the uncertainty parameter σ . Note that the lower limit is always given by zero for values $P < \sqrt{2}\sigma$.

If one only has a single measurement the confidence intervals (e.g. the upper 68% or 90% limit lines) are decreasing for $\epsilon < 1$. By a statistical fluctuation of course one could always find an arbitrarily small estimated amplitude e.g. in the $P = 0$ case. This would also lead to a smaller confidence interval (see Figure 5.10), compared with the expected result, where some amplitude greater than zero is estimated due to the bias of the fitting procedure. This effect is similar in nature to the case discussed in the original work of Cousins and Feldman where one measures signal and background events, but one already knows the expected number of background events (see also the discussion in subsection 4.2). Translated to our case, if one finds a signal strength $\hat{\epsilon}$ far below the expected value $\hat{\epsilon} \approx 1$, it can be noted that a downward fluctuation did happen. If one is interested in the sensitivity in those cases, one should take the limits corresponding to the expectation value for the amplitude.

5.8. Derivation of the Posterior Probability Density Function

The previous section described how a confidence interval can be defined for P given a measured \hat{P} . Note that one typically also wants to quote a point estimate of the true amplitude given some measurement. The natural way to do this with the Feldman Cousins approach, is to quote the P corresponding to the maximum likelihood ratio $R(\hat{P}, P) = 1$, since all confidence intervals start around this “central” estimate.

If the amplitude P is the final result of the experiment, this is sufficient as long as the asymmetrical confidence surrounding the result is considered for small signal strengths. However, if P is used as an input in a subsequent analysis, it is desirable to also have a probability density function for the true amplitude P given some estimator (measurement) \hat{P} . Unfortunately, it is in general not directly possible to construct such a pdf without further assumptions. The typical approach is to use Bayes’ theorem with a constant prior

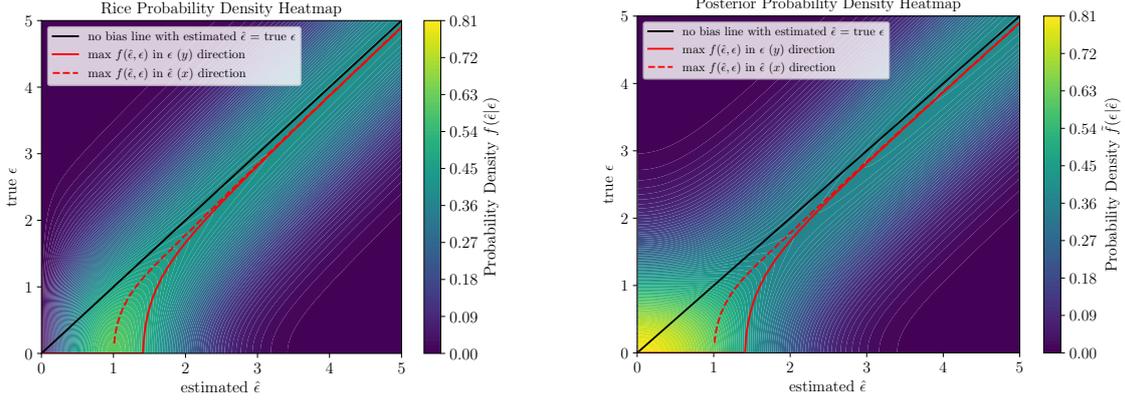


Figure 5.11: The probability distribution $f(\hat{\epsilon}|\epsilon)$ and the posterior Bayesian probability distribution $\tilde{f}(\epsilon|\hat{\epsilon})$. Note that the distributions are only interchangeable if $\sigma \ll P$ (large signal strength). In this case both approach the Gaussian limit.

probability for P , which gives the probability density function

$$\tilde{f}(P|\hat{P}) = \frac{f(\hat{P}|P)}{\int_0^\infty f(\hat{P}|P)dP}. \quad (5.32)$$

Note that $\tilde{f}(P|\hat{P}) \propto R(\hat{P}, P)$ for a fixed \hat{P} . This means, that the maximum of the probability density function for the true amplitude (using a constant prior) is always at the same position as $R = 1$ from the Feldman Cousins approach.

In Figure 5.11 the Rice PDF $f(\hat{\epsilon}|\epsilon)$ and the posterior distribution are compared. Again if the amplitude is large compared to the uncertainties, the distributions approach the Gaussian limit, where $f(\hat{P}|P) \equiv \tilde{f}(P|\hat{P})$ (compare subsection C.2). If $\sigma \approx P$ (small signal strength) the distributions are clearly not Gaussian, and their shape is indeed different. For details on how the rice distribution converges to a normal distribution see [56].

One might ask, why the prior probability is not defined otherwise, e.g. as constant in the range $P \in [0, 1]$, since the maximum amplitude is also naturally limited by one – at least for the typical application related to probability distributions as introduced above. This is not done, since in the real experiment the maximum amplitude $P = 1$ is not reached, and typically there are many σ distance to this limit, therefore the outcome is not significantly changed. Note that various estimators e.g. mapped χ^2 -fits and the Fourier transform can in theory (for general application) produce estimates $P > 1$, because they do not know anything about the underlying probability distribution. Not using a limited prior to $P \leq 1$, but instead allowing for all positive amplitudes, also has the advantage that one can again work with signal strength $\epsilon = P/\sigma$ and simply resale the results. Otherwise the maximum allowed signal strength would depend on the uncertainty (since then $\epsilon_{\max.} = 1/\sigma$) and the distributions would not be universal.

To give Bayesian credible intervals (the Bayesian analogy to Feldman Cousins confidence intervals) for the true amplitude or signal strength ϵ , one simply starts at the most probable signal strength ϵ in the probability distribution $\tilde{f}(\epsilon|\hat{\epsilon})$ and integrates the probability distribution around it with descending probability density. Using this approach the central value for both the Feldman Cousins confidence intervals and the Bayesian credible intervals

is typically the same (as long as $\epsilon \geq \sqrt{2}$). Where the results differ is in how far the intervals extend regarding the true signal strength, given some estimated signal strength.

Figure 5.12 compares the resulting interval limits for all the possible confidence limits cl and all estimates signal strengths $\hat{\epsilon}$. Three “typical” choices 68%, 90%, 99% are marked with dashed lines.

Figure 5.13 compares vertical cross section through this 2D representation at given measured signal strengths $\hat{\epsilon}$. For a given confidence (FC approach) or credible level (Bayes approach) on the x -axis one reads of the corresponding true signal strength on the y -axis. For $\epsilon < \sqrt{2}$ all lower limits are zero (not displayed in the plot). The two horizontal lines correspond to the signal strength where $R = 1$ and the Bayesian posterior PDF has its maximum (red) and to the naive estimator $\epsilon = \hat{\epsilon}$ (blue), which would be used to construct biased Gaussian confidence intervals.

At small signal strength $\epsilon < \sqrt{2}$, the Feldman Cousins approach typically gives narrower confidence limits than the corresponding Bayesian credible intervals. In this case the parameter limits for the Bayesian approach can be calculated from the cumulative distribution function (CDF) of the underlying probability density. Therefore there are never empty credible intervals; for every $\hat{\epsilon}$ and every desired confidence level cl one can give a corresponding interval. This is different for the Feldman Cousins approach, where for small confidence values, the interval can be extremely small and even completely empty for some non vanishing confidence levels $cl > 0$. However empty intervals occur only at $cl \approx 50\%$ and for “typical” $cl = 68.27\% \equiv 1\sigma$ the interval is also never empty, but simply noticeably narrower.

Another interesting effect happens above $\epsilon > \sqrt{2}$. For this consider the second plot in Figure 5.13 with $\hat{\epsilon} = 1.5$. At this point the confidence intervals do not start at zero anymore, but are constructed around some assumed true signal strength (which has the highest likelihood ratio $R = 1$). These “center” values in the interval are indicated by the horizontal red lines. Now there are also lower parameter boundaries at least for small chosen confidence intervals. Therefore both approaches now give non empty intervals no matter what confidence limit is chosen. However the given limits are still very different. The obtained limits for the true signal strengths are then non smooth functions of the confidence limits for both approaches at exactly one point, where small changes in the chosen limit cl can change the obtained interval drastically. Note that this also implies that small changes in the measured signal strength $\hat{\epsilon}$ change the confidence limits non smoothly. The lines where this non smoothness happens are also overlayed in Figure 5.12, but the farther away from the estimated signal strength $\hat{\epsilon} = \sqrt{2}$, the less steep and prominent are those kinks in the limits.

For larger $\epsilon > 3.5$ both the FC confidence and the Bayesian credible levels give more and more similar results. A small bias compared to a naive Gaussian estimator still remains, since the horizontal lines are never identical. In the Gaussian limit FC and Bayesian approach would deliver exactly identical results for the true signal strength limits and start the construction at $\hat{\epsilon} = \epsilon$. In the Gaussian limit also all confidence intervals are “smooth”, since no parameter boundaries exist.

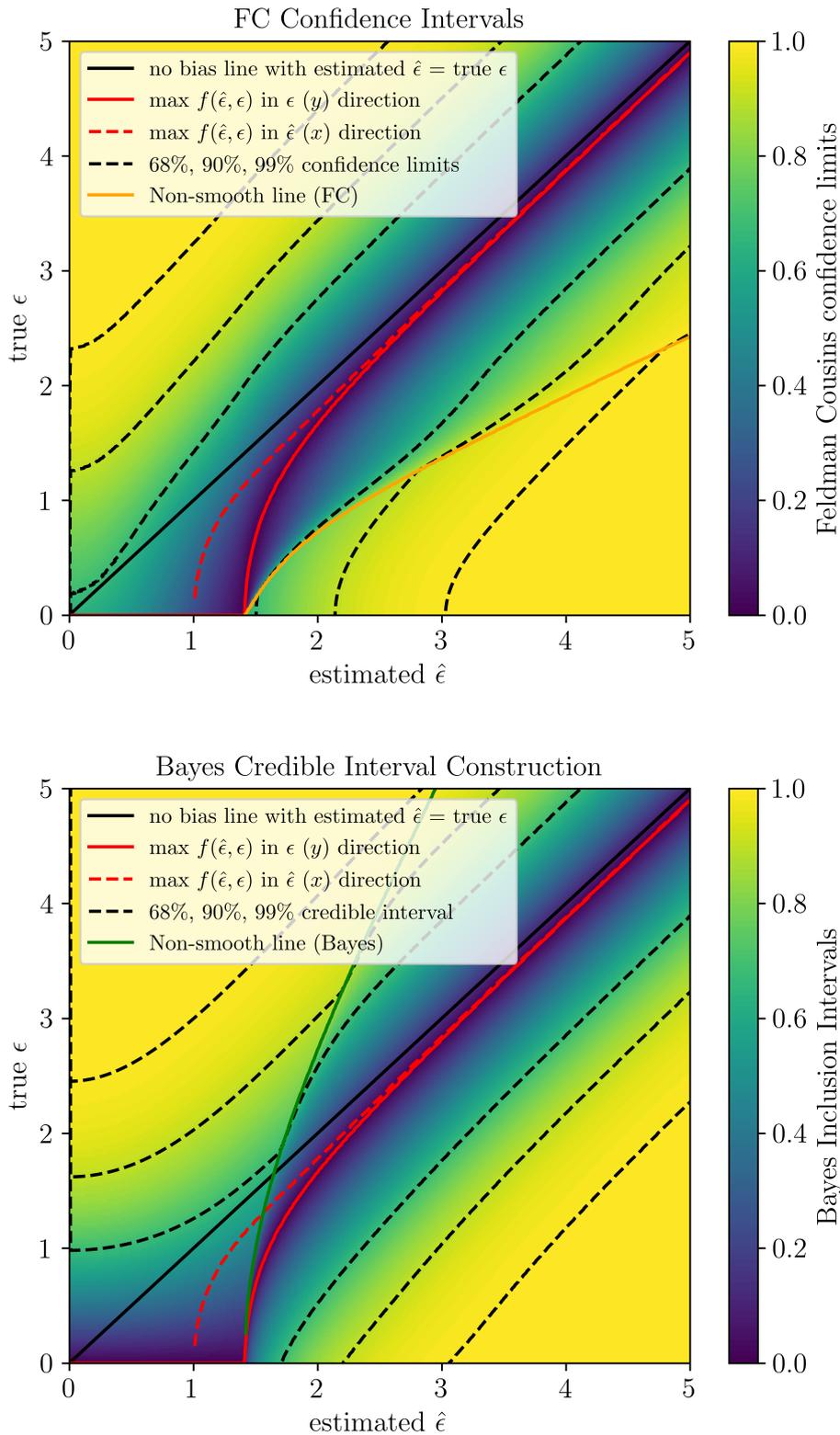


Figure 5.12: Confidence Intervals for the Rice Distribution vs. Bayesian Credible Intervals. Non-smooth points in the confidence intervals are a result of the construction starting at the red central line ($\max R = 1$) and then reaching the limits $\hat{\epsilon} = 0$ in x -direction for FC and $\epsilon = 0$ in y -direction for the Bayesian posterior.

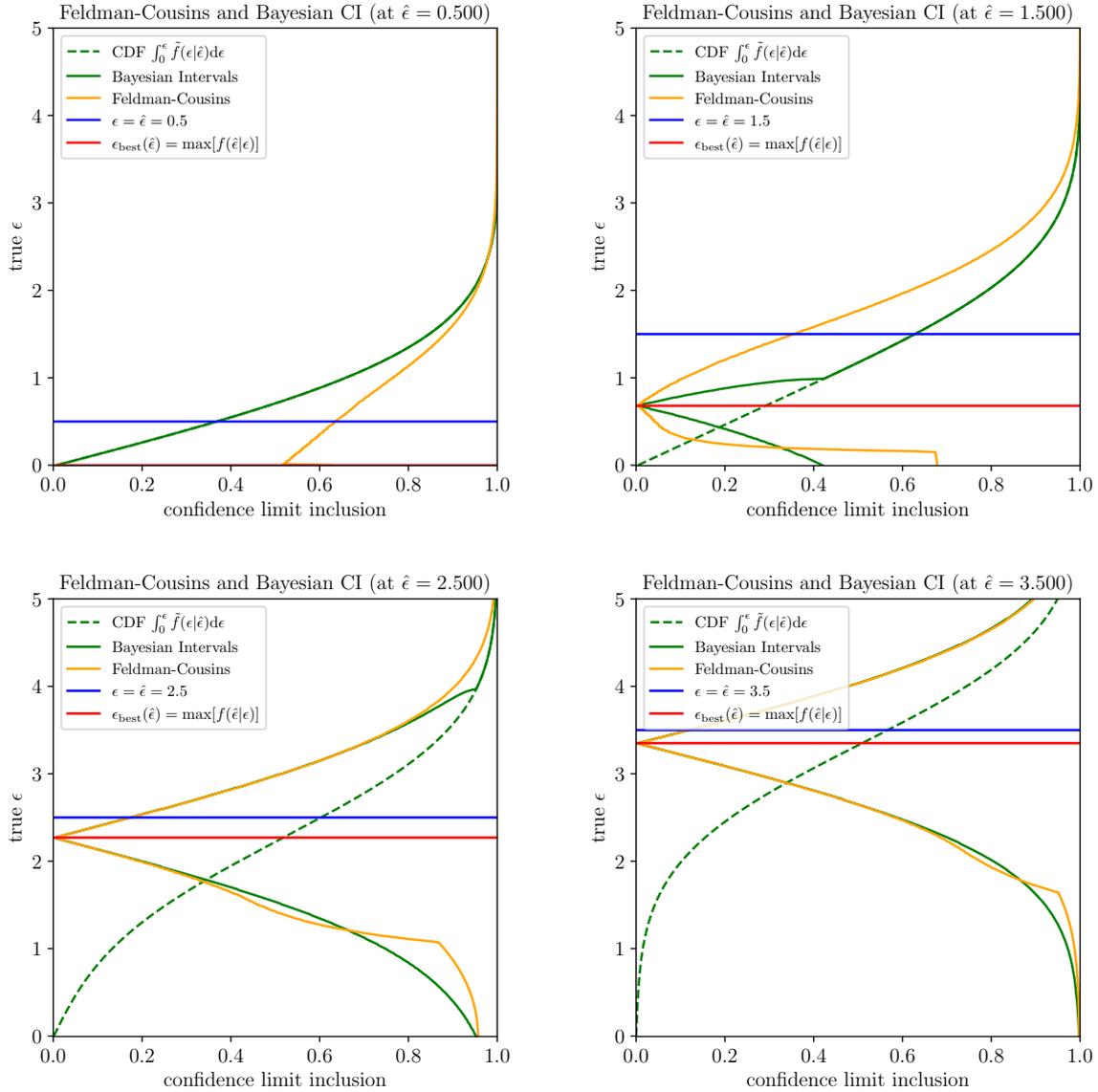


Figure 5.13: Confidence Intervals for the Rice Distribution vs. Bayesian Credible Intervals at given estimated signal strengths $\hat{\epsilon} = 0.5, 1.5, 2.5, 3.5$. This are basically cuts along the vertical direction of Figure 5.12. For small signal strength Bayesian and FC limits are different, for larger signal strength $\hat{\epsilon} > 3$ both approaches become similar.

6. Turn Fourier Transform

Another analysis approach of data is the discrete time Fourier transform or more precise the discrete turn Fourier Transform. In this case the measurement observable is based on the discrete integer turn number n and not on the time stamp (time in cycle).

In a simple model the spin motion function corresponds to a 1-dimensional time periodic function at the polarimeter describing the probability of an event in the detector, where n is a sequence of discrete turn numbers

$$f(t) = \frac{1}{\Delta n} \cdot (1 + P \cos(\omega_s t - \phi_s)) \equiv f(n) = \frac{1}{\Delta n} \cdot (1 + P \cos(2\pi\nu_s n - \phi_s)) \quad (6.1)$$

$$= \frac{1}{\Delta n} \cdot (1 + A \sin(2\pi\nu_s n) + B \cos(2\pi\nu_s n)) , \quad (6.2)$$

with the angular spin precession frequency ω_s , the spin tune ν_s , the phase ϕ_s and an amplitude P . A Fourier transformation can put such a signal from the turn domain into the spin tune domain.

6.1. Discrete Fourier Transform

All detected events are assigned to a turn number $n \in \mathbb{N}$, leading simply to a list of integer numbers. Here the discrete Fourier Transform (DFT) can be used to further process the data. The DFT is a fundamental tool in digital signal processing, and transforms a sequence of numbers given in the time domain into another sequence of (complex) numbers in the frequency domain. Given a sequence of N numbers $x[n]$, where $n = 0, 1, 2, \dots, N-1$, the DFT of the sequence is typically defined as:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-i2\pi n \frac{k}{N}}, \quad k = 0, 1, 2, \dots, N-1 \quad (6.3)$$

Here, $X[k]$ represents the k -th (frequency) component of the transformed signal, i is the imaginary unit, and $e^{-i2\pi n \frac{k}{N}}$ is the complex exponential kernel that can be seen as the basis function for the transformation.

Since one typically chooses $N \approx \mathcal{O}(10^6)$, it would take unnecessary computational power to calculate all sums for $\nu_k = \frac{k}{N}$, especially since a good approximation of ν_k is typically already known e.g. from the Thomas-BMT equation. In practice the discrete Fourier transform for the spin tune is not evaluated for discrete values of $\nu_k = \frac{k}{N}$, but the sampling interval of the spin tune domain is given by $\nu_k = \nu_0 + \Delta\nu k$, with $k \in \mathbb{Z}$ and $|k| \leq N_k$, where $N_k \ll N$. ν_0 is a well educated guess, where the true spin tune might be. Therefore $\nu_k \in [\nu_0 - \Delta\nu N_k, \nu_0 + \Delta\nu N_k]$. Note that the assumed spin tunes are still equidistant (distance $\Delta\nu$).

Not going into detail about the mathematical technicalities, the discrete sequence of the turn signals is transformed into the spin tune domain, using the formula

$$\hat{g}[\nu_k] \equiv g_{\nu_k} = \sum_{n=0}^{N-1} g[n] e^{-i2\pi n \cdot \nu_k} \quad (6.4)$$

$$= \sum_{n=0}^{N-1} g[n] [\cos(2\pi n \nu_k) - i \sin(2\pi n \nu_k)], \quad (6.5)$$

where Euler's formula is used to expand the complex exponential. The Precursor experiments provide signals $g[n]$, which are based on random processes. The probability of detecting an event per one turn of the bunch is small $\ll 1$. Therefore most of the turn entries are zero

$$g[n] = \begin{cases} 1 & \text{for } n = n(n_{\text{event}}) \\ 0 & \text{else} \end{cases} \quad (6.6)$$

Consequently, the summation is given by the sequence of turn numbers $n(n_{\text{event}})$, where indeed events were detected $n_{\text{event}} \in [1, N_{\text{event}}]$. Note that this $g[n]$ with turns chosen at random, is in fact a representation of Equation 6.1. The Fourier coefficients are given by the real and the imaginary part of g_{ν_k} .

$$\hat{A}_{\nu_k} = \frac{2}{N_{\text{event}}} \text{Im}(g_{\nu_k}) = \frac{2}{N_{\text{event}}} \sum_{n_{\text{event}}=1}^{N_{\text{event}}} -\sin(2\pi\nu_k n(n_{\text{event}})). \quad (6.7)$$

$$\hat{B}_{\nu_k} = \frac{2}{N_{\text{event}}} \text{Re}(g_{\nu_k}) = \frac{2}{N_{\text{event}}} \sum_{n_{\text{event}}=1}^{N_{\text{event}}} \cos(2\pi\nu_k n(n_{\text{event}})). \quad (6.8)$$

The factor $\frac{2}{N_{\text{event}}}$ takes into account the normalization based on Parseval's theorem [57] for real input data. A motivation why this factor is needed can also be found in the appendix in subsection A.2 with an application of DFT to a signal coming from pure sine and cosine data, where the factor naturally arises. The uncertainties of both \hat{A}_{ν_k} and \hat{B}_{ν_k} are not derived here, but a calculation can be found in the appendix of [58] based on the principles described in [59]. As simulation shows below, the relevant results for the amplitude follow a rice distribution anyhow and the uncertainty parameter is the expected one from subsection 5.4.

Since \hat{A}_{ν_k} is calculated directly from the sine term and \hat{B}_{ν_k} from the cosine term, the amplitude and phase simply are given by

$$\hat{P}_{\nu_k} = \sqrt{\text{Im}(\hat{g}_{\nu_k})^2 + \text{Re}(\hat{g}_{\nu_k})^2} = \sqrt{\hat{A}_{\nu_k}^2 + \hat{B}_{\nu_k}^2} \quad \sigma_{P_{\nu_k}} = \sigma_{\text{DFT}} \approx \sqrt{\frac{2}{N_{\text{event}}}} \quad (6.9)$$

$$\hat{\phi}_{\nu_k} = \text{atan2}(\text{Im}(\hat{g}_{\nu_k}), \text{Re}(\hat{g}_{\nu_k})) = \text{atan2}(\hat{A}_{\nu_k}, \hat{B}_{\nu_k}) \quad \sigma_{\phi_{\nu_k}} = \frac{\sigma_{\text{DFT}}}{P_{\nu_k}} \approx \frac{\sqrt{\frac{2}{N_{\text{event}}}}}{P_{\nu_k}}. \quad (6.10)$$

Both uncertainty parameters are similar to the theoretical derivation based on the Cramér-Rao bound of a sinusoidal oscillation. However the uncertainties of such a DFT are of course highly correlated between neighboring ν_k , as all the data goes into the calculation of each parameter, therefore a useful interpretation of uncertainty is not trivial, as it tells nothing about the fluctuation in direction of ν_k . A more detailed description therefore requires numerical simulations.

6.2. Analytic Fourier Amplitude Model

Apart from the numerical results, which are usually sufficient for the analysis, one might want to understand the shape and structure also analytically. The following section will argue, how the resulting amplitudes in a Fourier transform are formed, when scanning

through various possible spin tunes. For this one rewrites the Fourier transform $\hat{g}(\nu)$, with the “selection function” $g[n]$ being represented by the probability density function $f(n)$

$$\hat{g}(\nu) = \sum_{n=0}^{N-1} g[n]e^{-in \cdot 2\pi\nu} \approx \sum_{n=0}^{N-1} f(n)e^{-in \cdot 2\pi\nu} \quad (6.11)$$

$$\approx \sum_{n=-\infty}^{\infty} f(n)\text{rect}_N(n)e^{-in \cdot 2\pi\nu} = \sum_{n=-\infty}^{\infty} P \cos(2\pi\nu_s n - \phi_s)\text{rect}_N(n)e^{-in \cdot 2\pi\nu}. \quad (6.12)$$

In this case one can represent the Fourier transform by a product of the sine-like probability density $f(n)$ and a rectangular window function $\text{rect}_N(n)$. This rectangular function analytically selects values of n i.e. $0 \leq n < N$ where $\text{rect}_N(n) = 1$. A multiplication in the turn domain n corresponds to a convolution in the spin tune domain ν , therefore

$$\hat{g}(\nu) = f(\nu) * \text{rect}_N(\nu) = (f * \text{rect}_N)(\nu) = \int_{-\infty}^{\infty} f(\nu - \lambda) \cdot \text{rect}_N(\lambda) d\lambda \quad (6.13)$$

The Fourier transforms of both a sine wave (see subsection A.2) and a rectangular function (see subsection A.2) can be calculated analytically. Without loss of generality and for simplicity one can assume the phase to have some value and choose $\phi_s = 0$: then $A = 0$ and $|B| = P$. In that case it is especially easy to derive the result for the Fourier transform directly even for the amplitude. Otherwise both a sine (with A) and cosine term (with B) have to be considered, but the result for the amplitude remains the same. One gets the Fourier transforms of the parameter B (here equivalent to the amplitude):

$$f_B(\nu) = \frac{NB}{2} (\delta(\nu - \nu_s) + \delta(\nu + \nu_s)) \quad \left[f_A(\nu) = \frac{NA}{2i} (\delta(\nu - \nu_s) + \delta(\nu + \nu_s)) \right]. \quad (6.14)$$

One also finds the Fourier transform of the rectangular function to be

$$\text{rect}_N(\nu) = \sum_{n=0}^{\infty} \text{rect}[n]e^{-i2\pi n \cdot \nu} = \sum_{n=0}^{N_{\max}=\Delta n-1} e^{-i2\pi n \cdot \nu} = \frac{1 - e^{-i2\pi\nu\Delta n}}{1 - e^{-i2\pi\nu}} \quad (6.15)$$

$$= \frac{\sin(\pi\nu\Delta n)}{\pi\nu\Delta n} e^{-i\pi(\Delta n-1)} = \text{sinc}(\pi\nu\Delta n) e^{-i\pi(\Delta n-1)} = \pm \text{sinc}(\pi\nu\Delta n). \quad (6.16)$$

Combining both results one can deduce the Fourier transform of the amplitude

$$\Rightarrow \hat{g}_B(\nu) = \frac{NB}{2} \int_{-\infty}^{\infty} (\delta((\nu - \lambda) - \nu_s) + \delta((\nu - \lambda) + \nu_s)) \frac{1 - e^{-i2\pi\nu\Delta n}}{1 - e^{-i2\pi\nu}} d\lambda \quad (6.17)$$

$$= \frac{NB}{2} \int_{-\infty}^{\infty} (\delta((\nu - \nu_s) - \lambda) + \delta((\nu + \nu_s) - \lambda)) \frac{1 - e^{-i2\pi\nu\Delta n}}{1 - e^{-i2\pi\nu}} d\lambda \quad (6.18)$$

$$= \frac{NB}{2} \left[\frac{1 - e^{i2\pi(\nu_s - \nu)\Delta n}}{1 - e^{i2\pi(\nu_s - \nu)}} + \frac{1 - e^{i2\pi(\nu_s + \nu)\Delta n}}{1 - e^{i2\pi(\nu_s + \nu)}} \right] \quad (6.19)$$

$$= \frac{NB}{2} [\text{sinc}(\pi(\nu_s - \nu)\Delta n) + \text{sinc}(\pi(\nu_s + \nu)\Delta n)]. \quad (6.20)$$

Note that in the equations above starting at Equation 6.15 the length of the turn interval Δn , defines how narrow the sinc function will be, since n is selected as $0 \leq n < \Delta n$. The start at $n = 0$ is arbitrary and does not change the result of the Fourier transform.

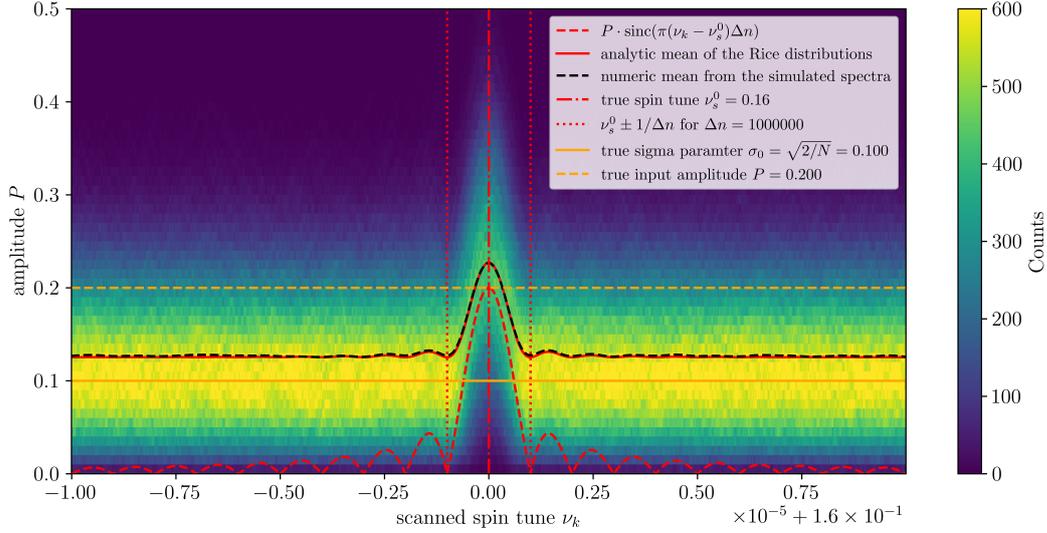


Figure 6.1: Fourier Spectrum for 10000 repetitions of a Fourier scan over a scanned spin tune range. A sinc function structure is formed as the average result for a true amplitude of $P = 0.2$.

From Equation 6.17 one gets the shape of the main peak at $\nu = \nu_s$. The other sinc peak corresponds to $\nu = -\nu_s$, because Equation 6.17 is symmetric. Finally one is mainly interested in the values around the true spin tune, and at $\hat{g}(\nu \approx \nu_s)$ one gets (using $N = \Delta n$)

$$\Rightarrow |\hat{g}_B(\nu \approx \nu_s)| \approx N \frac{B \sin(\pi(\nu_s - \nu)N)}{2 \pi(\nu_s - \nu)N} = N \frac{B}{2} \text{sinc}(\pi(\nu_s - \nu)N) \quad (6.21)$$

$$\Rightarrow |\hat{g}_B(\nu = \nu_s)| \approx \frac{NB}{2} \left[= \frac{NP}{2} \text{ since } A = 0 \right]. \quad (6.22)$$

Indeed $\frac{2}{N} \hat{g}(\nu \approx \nu_s)$ gives the amplitude P in this case. However one has to notice, that this calculation is only fully true for large signal strength ($N \rightarrow \infty$, therefore $\sigma \rightarrow 0$). For few events in the interval and small signal strength one has to take care of possible bias effects (due to fluctuations), as already mentioned in the previews chapters and seen for the fitting results in the toy model case. The Fourier transforms approach described above in general can lead to biased amplitudes. Then the uncertainty parameter $\sigma = \sqrt{2/N}$ has to be included in the description, as will be done below.

6.3. Fourier Transform Spectrum Simulations

Randomly generated data distributed according to Equation 6.1 can be analyzed by the discrete turn Fourier transform. In simulations done below the ideal undisturbed case is studied (fixed amplitude over the full interval), and exactly the same parameters are used to sample thousands of repetitions, which is not possible using real experiential data. For the simulation random turn numbers are generated from a turn interval of length Δn and with a known spin tune of $\nu_s = 0.16$. This is in the region of the true values and the deviations for the simulated spin tune scan are easier to read. The true spin tune

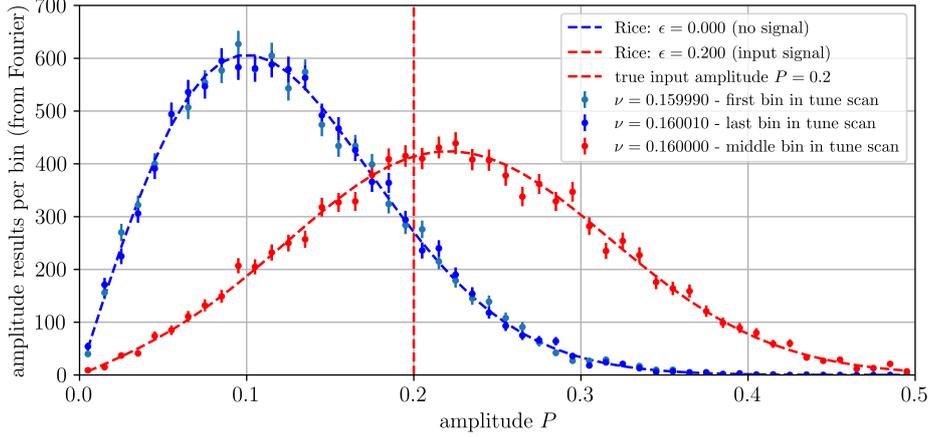


Figure 6.2: The resulting amplitudes at given assumed spin tunes for 10000 repetitions of a Fourier scan using the same initial parameters for a true amplitude of $P = 0.2$. The distributions of Fourier amplitudes P for a fixed assumed spin tune ν_s are well described using a Rice distribution with the expected amplitude from the sinc function of Figure 6.1 and the uncertainty parameter $\sigma_0 = \sqrt{2/N}$.

for simulations is always in the middle of the spin tune axis. As in the experiment the number of events generated N is much smaller than the length (measured in turn numbers Δn) of the interval.

For various assumed spin tunes one calculates the estimator for \hat{A} and \hat{B} following Equation 6.7 and Equation 6.8. For a given set of turn numbers the calculation of amplitudes for neighboring assumed spin tunes is highly correlated, since the same data is used for each calculation. This procedure is repeated multiple time with the same parameters, only the generated turns are random.

One resulting distribution for $P = 0.200$ and $N = 200$ events over $\Delta n = 10^6$ turns can be seen in Figure 6.1. It is noticed that the resulting amplitudes distribution is precisely described by the following model. The underlying amplitude for each spin tune is given by a sinc function

$$P(\nu; \nu_s) = P \cdot \left| \frac{\sin(\pi \cdot (\nu - \nu_s) \cdot \Delta n)}{\pi \cdot (\nu - \nu_s) \cdot \Delta n} \right|, \quad (6.23)$$

which is given as the dashed red line in Figure 6.1. This (unbiased) amplitude is indeed approaching zero for larger deviations from the true spin tune.

However for the estimated amplitude the bias due to the fitting procedure with random possible fluctuations has to be considered. The amplitude from Equation 6.23 therefore has to be used together with the $\sigma_0 = \sqrt{2/N}$ parameter as an input for a rice distribution for all possible spin tunes. As shown in Figure 6.1 the description of simulated data using this model is excellent. The behavior of the mean amplitude for all possible spin tunes (dashed black line in simulation) coincides with the calculation from the procedure give above (solid red line below dashed black line). Note that again the estimated amplitude is on average greater than the underlying true amplitude for every assumed spin tune. The average predicted and simulated amplitude reaches minima for $(\nu - \nu_s) = k \cdot 1/\Delta n$ with $k \in \mathbb{N}$ (without the obvious main maximum for $k = 0$). These (average) minima are given

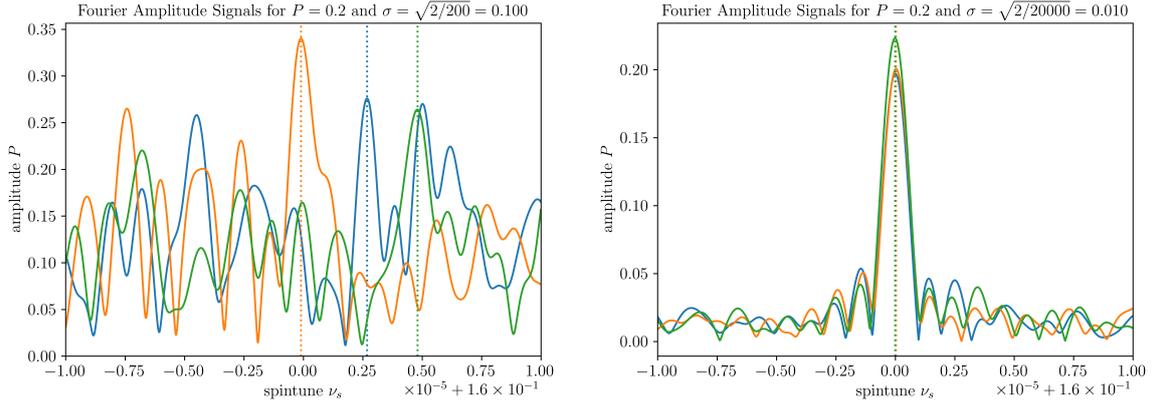


Figure 6.3: Three examples for obtained Fourier amplitudes in the scan through the spin tunes once for a signal strength of 2 (left) and for a signal strength of 20 (right). If the signal strength is small, the main peak can fluctuate out of the (expected) main lobe. The larger the signal strength, the better the sinc-structure is resolved even in individual measurements.

by the expectation value of a Rayleigh distribution with the uncertainty parameter σ .

Due to the uncertainty parameter σ even for large deviation from the true spin tune, the estimated amplitude on average does not reach zero, and the limiting case is a Rayleigh distribution with mean $\sqrt{\pi/2}\sigma$ and the mode (most probable fitting amplitude) is given by σ . The true amplitude P is marked with a horizontal dashed orange line and σ with a solid orange line in Figure 6.1.

One might ask, why a simulation is even needed, when the distribution can be predicted analytically as described above. The reason for this is, that one can easily predict the average behavior of the simulation, but it is much harder to give an accurate model for individual amplitude sets. And since the maximal amplitude is searched for in every individual set of turn numbers, the distribution of its maximum amplitudes (and corresponding spin tunes) is obviously different from the distribution at the peak of the full spectrum.

If one uses surrounding spin tunes (more information) to exclude spin tune result, which do not fall on the expected maximum of the main peak or predicts the spin tune accurately enough for the whole measurement, the distribution of amplitudes at the true spin tune will again follow a Rice distribution, since one then effectively cuts through the distribution at a fixed frequency. This is also the case for any other “quasi-fixed frequency” or “quasi-fixed spin tune” analysis, e.g. when using the so called mapping method.

What is meant by “quasi-fixed spin tune” is explained in more detail in the appendix (see subsection D.2). If one uses the phase information to estimate the spin tune (as briefly described in subsection 6.4) one can use the full data set to determine the spin tune. For more than 133s total measurement time per cycle the total usable turn number interval is given by $100 \cdot 10^6$ and above. Also the phase is more sensitive to changes in the spin tune, and therefore the spin tune can be determined at least two to three orders of magnitude better than from one amplitude alone.

Even really small tune changes of order 10^{-9} and less can be obtained by a phase fit.

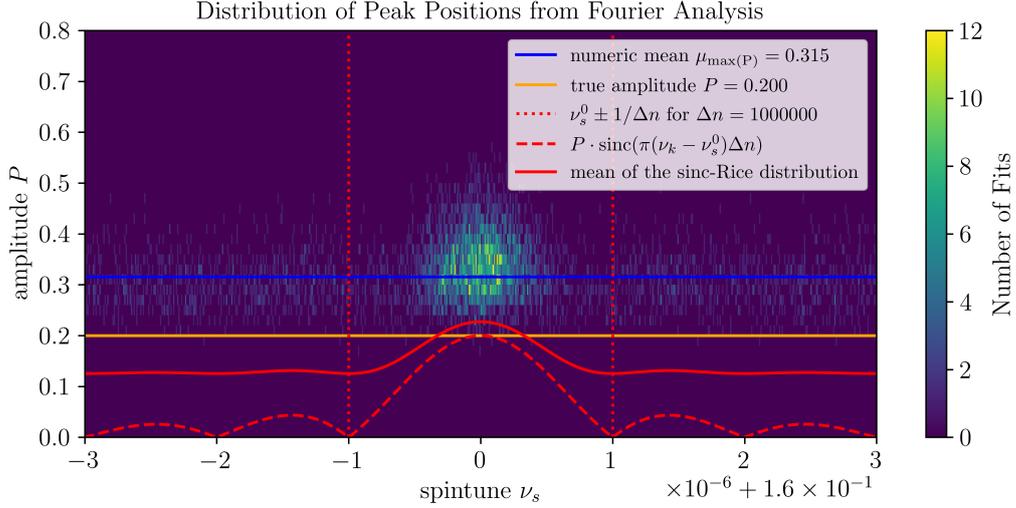


Figure 6.4: Fourier maxima distribution for 10000 repetitions of a Fourier scan using $N = 200$ and $P = 0.2$. If simply the maximum amplitude is chosen in all the measurements the amplitude is overestimated much more, than expected from a simple Rice distribution (following the height of the sinc function drawn in red).

When evaluating the spectra at such spin tunes, one can guess, that the average mean position of the Fourier spectra is correctly estimated. Even if the signal strength drops down, so that large fluctuations of the amplitude peak might happen, the corresponding amplitudes behaves just randomly like a Rice distribution, since they are evaluated at one quasi fixed spin tune positions and do not follow the shifting individual peak positions any longer. Note that the amplitude overestimation coming from the underlying Rice distribution for small signals introduced in the previous chapter subsection 5.3, which is mostly independent of the estimation method chosen, remains even when the correct spin tune is chosen.

For all further analysis, “quasi-fixed spin tune” cases are considered, where the underlying distribution is assumed to be a simple Rice distributions with only two parameters. The estimated amplitude is \hat{P} , the estimated uncertainty is $\hat{\sigma} \equiv \sigma_0 = \sqrt{2/N}$. Since mainly the small amplitude regions are of interest, one can use the universal signal strength approach with $\hat{\epsilon} = \hat{P}/\sigma_0$ and simply scale with the uncertainties afterwards, avoiding unnecessary complications if the uncertainty parameter σ is treated as simple as possible (not regarding it as a function of the amplitude). It was shown in the previous section that this simplification can be justified, especially since it at worst overestimates the uncertainties by a small margin, while still treating the shape of all distributions correctly. Also this approach reproduces the expected results on typical real COSY data for the usual large signal strengths. Finally one should keep in mind, that for $\epsilon \gg 1$ most of the effects discussed in this chapter again become irrelevant as the peak position will fluctuate only in the flat region of the maximum of the main peak and all uncertainties become Gaussian like.

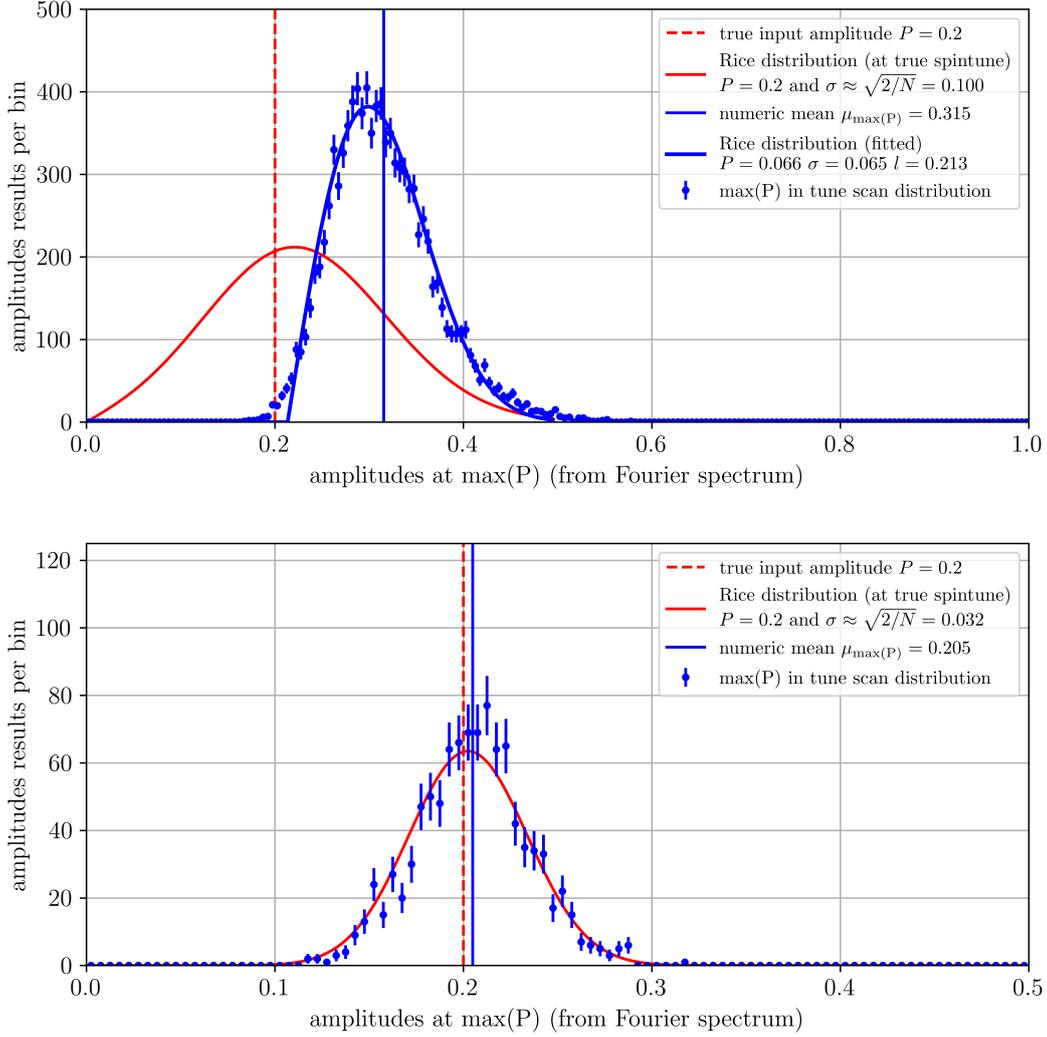


Figure 6.5: Distribution of the maximum amplitude in the fit and the corresponding rice distribution. Upper plot: for $N = 200$ and amplitude $P = 0.2$ (signal strength 2). Lower plot: for $N = 2000$ and amplitude $P = 0.2$ (signal strength 6.3). If the signal strength improves the maxima start to again follow a Rice distribution (approaching the Gaussian limit).

6.4. Spin Tune Estimator

The spin tune can be precisely determined by calculating the turn derivative of the phase information obtained in the analysis procedure described above. The relation of phase and turn depending spin tune is given by

$$\nu_s(n) = \nu_s^0 + \frac{1}{2\pi} \frac{d\phi_s(n)}{dn}. \quad (6.24)$$

where n denotes the turn number and $\phi_s(n)$ is the phase obtained from a spin tune scan. Therefore the spin tune is determined from the interpolation of two consecutive phase estimators. The effect of scanning an off spin tune value, deviating from the correct one,

can be understood by looking at the phase spectrum. Note that the starting phase in this type of plots is fixed to a constant value over the whole spin tune range, to see the effect of a deviating spin tune over the number of turns. Also if the phase goes out of the definition range $\phi_s(n) \in [0, 2\pi]$ it is corrected to be continuous. Assuming that the spin tune matches the true horizontal precession, the phase stays rather constant over the whole range (apart from small actual spin tune drifts or statistical fluctuations). For a data sampling spin tune, which is slightly different than the true spin tune frequency, the phase difference increases (or decreases) more or less linearly with time, since a mismatch in frequency adds the same phase deviation per turn.

Figure 6.6 shows the spin tune scan of an example run (one cycle) from the Precursor 1 experiment for three different binning options in turn number Δn . The peaks at the spin precession frequency are only visible after the first $20 \cdot 10^6$ turn, since the spin tune before the jump (where the solenoid magnetic fields are changed) is outside the scanned range.

The scanned tune range is the same in all plots. Note that the amplitude scan gets narrower $\propto 1/\Delta n$, while the sigma parameter of the amplitudes in the spin tune scan changes with the number of events per bin and therefore $\propto \sqrt{2/\Delta n} \propto \sqrt{2/N}$, assuming constant average count rates in the turn intervals. In the case of $\Delta n = 10^6$, the main peak of the amplitude is limited by spin tunes of $\pm 10^{-6}$ around its maximum, where the first minima are reached.

The peak heights are fluctuating a bit, but the average middle of the peaks stays rather constant. The phase spectra are narrower than the amplitude spectra. Note that the phase spectra start to repeat, if one leaves the region $\pm 1/(2\Delta n)$ around the maximum of the main peak into a secondary spectrum. Therefore one needs to first hit the main lobe using the amplitude spectra, or some good spin tune prediction. The phase estimation uncertainty scales with the inverse of the amplitude; it is also most precise at the main peak.

The spin tune which delivers the most flat phase corresponds to the true reference spin tune. However the spin tune phase is not further analyzed in this thesis, since its detailed description is mostly heuristic, using polynomial fits [58]. Typically a precision of at least 10^{-10} can be reached for the spin tune when looking at the behavior of the phase in a single cycle of 100 s length. Such small deviations are not resolved in the given amplitude spectra, so the spin tune for them can be assumed to be constant (quasi-fixed spin tune). Then the amplitudes of the peaks behave according to the underlying rice distribution (no matter how good the signal strength), as long as one evaluates at fixed (or quasi-fixed) spin tunes, but not at the shifting peak maxima per bin, which can occur at small signal strength.

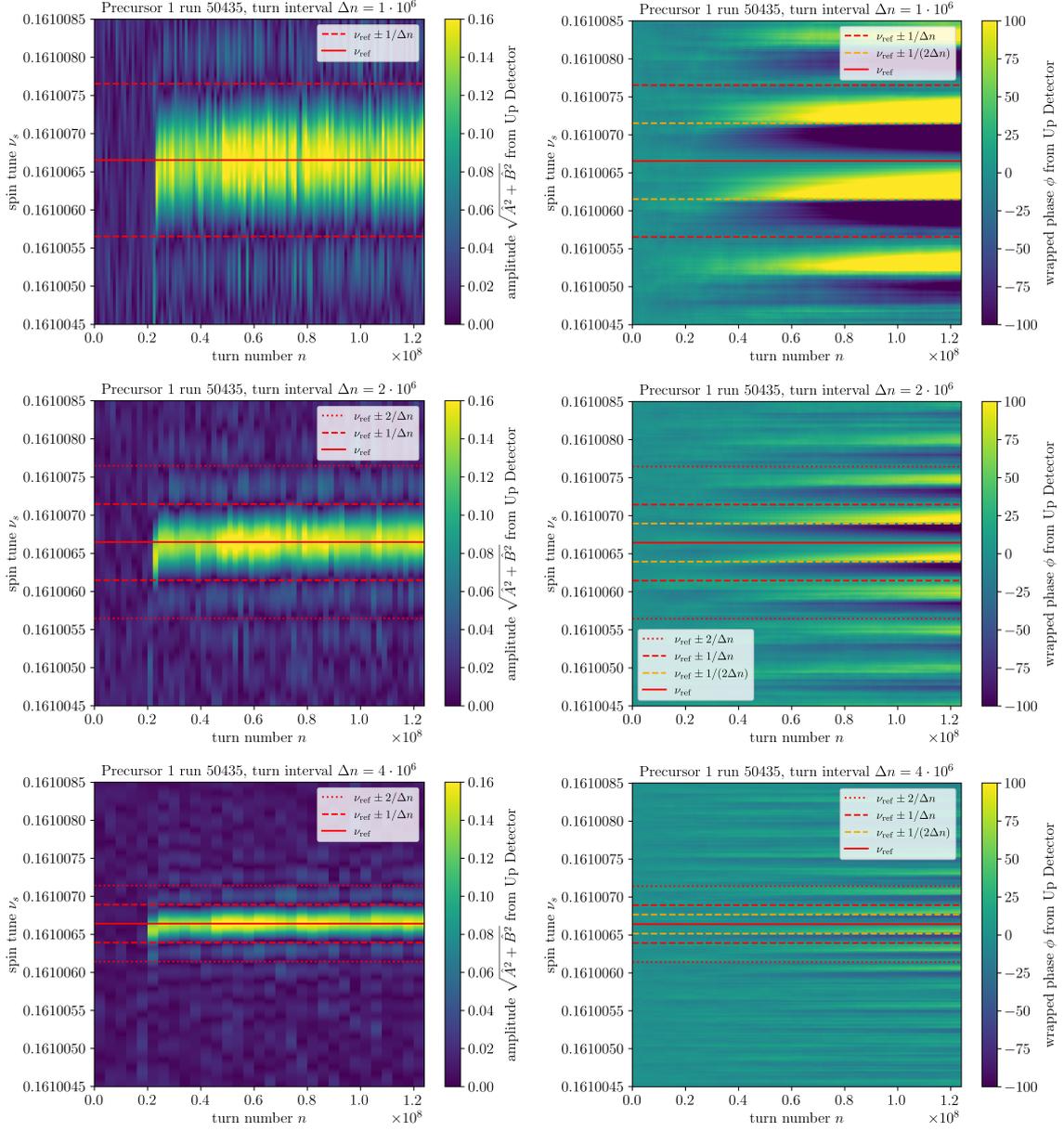


Figure 6.6: Amplitude and phase scans for the upper detector of an example run from the Precursor 1 experiment. The spin tune scan is done for three different turn number intervals $\Delta n \in [1.0 \cdot 10^6, 2.0 \cdot 10^6, 4.0 \cdot 10^6]$ (corresponding to time bins of 1.33s, 2.67s and 5.33s). The spin tune scan is carried out for the same spin tune interval in each plot. The broadness of the main peak around the assumed true reference spin tune ν_{ref} (horizontal red line) is given by $\nu_{\text{ref}} \pm 1/\Delta n$, where on average the first minima of the sinc functions are forming. Second minima can be found at $\nu_{\text{ref}} \pm 2/\Delta n$. Note that the reference spin tune (from the wrapped phase spectrum) coincides with the average center of the peaks in the amplitude spectrum. The pattern of the wrapped phase starts to repeats at $\pm 1/(2\Delta n)$, which is half of the main peak width.

7. Spin Polarization Decay and Spin Coherence Time

7.1. Simple Exponential Decay Model

To study the advantages of the correctly assumed probability distribution of the fitting parameters (in this case the amplitude P) a simple demonstrator model is used. This approach follows the one described in [54], but extends the discussion around the comparison of the results with a simple χ^2 -fit and relating it directly to the spin decoherence in the Precursor EDM measurements.

In figure Figure 7.1 the dashed red line shows an exponential function

$$f_{\text{exp}}(t; P_0, \tau) = P_0 e^{-t/\tau} \quad (7.1)$$

with a decay constant $\tau = 1$ and starting amplitude $P_0 = 0.3$. The data points (also shown in red) were generated at ten different times t_i following the $f(\hat{P}|P(t_i), \sigma)$ rice distribution with a $\sigma = \sqrt{2/N} = 0.045$ corresponding to $N = 1000$ events analyzed per bin. The vertical bands at each time bin indicate the Bayesian probability distribution with constant prior introduced in subsection 5.8 (the darker the higher the probability) $\tilde{f}(P|\hat{P})$ as a function of P , where \hat{P} is given by the generated value (i.e. the data point in red).

The blue curve shows the result of a simple least squares χ^2 -fit to the red data points, which assumes naive Gaussian errors (see the red error bars) for all the data points. Note that these errorbars sometimes reach into the unphysical $P < 0$ region, especially for $t/\tau > 2$. The resulting fitted function is far off from the underlying true one (dashed red line). The orange curve shows the result of a likelihood fit with the likelihood function

$$\mathcal{L} = \prod_{i=1}^{N_{\text{bin}}} \tilde{f}_{\text{Rice}}(P_0 e^{-t_i/\tau} | \hat{P}_i, \sigma) \quad (7.2)$$

varying τ and P_0 to maximize \mathcal{L} . The maximum likelihood estimator (MLE) describes the underlying true behavior much better than the χ^2 -fit results.

In this example the least squares fit yields $\hat{\tau} = 1.56 \pm 0.36$ and $\hat{P}_0 = 0.282(38)$ (blue line) compared to likelihood fit $\hat{\tau} = 1.02 \pm 0.34$ and $\hat{P}_0 = 0.303(44)$ (orange line) for the maximum likelihood estimator; in the latter case the agreement with $\tau = 1$ and $P_0 = 0.4$ is almost perfect.

The more detailed analysis including correlations between the two fitting parameters is shown in the lower plot of Figure 7.1, where the negative log likelihood values are calculated in a grid surrounding the maximum. A decline of the log likelihood to $-\frac{1}{2}$ when compared with the maximum indicates the 1σ interval (dashed red line), and the 2σ interval is reached at ≈ -4 (dashed black line). Note that the 2σ region cannot simply be described by an ellipsis, but is elongated in τ direction. Even for the 1σ region the fit errorbars (symmetric, centered around the fit result) are not fully identical with the extend of the ellipsis region. For comparison the χ^2 fit results with their uncertainties are shown as a black cross. The relative size and shape of the errorbars for both results are typically similar; the χ^2 -fit result is simply shifted away from the true input parameters due to a fitting bias.

Such an individual result are illustrative, but in simulation it is accessible and also more interesting what happens on average. Figure 7.2 compare the fit results for P_0 and τ for

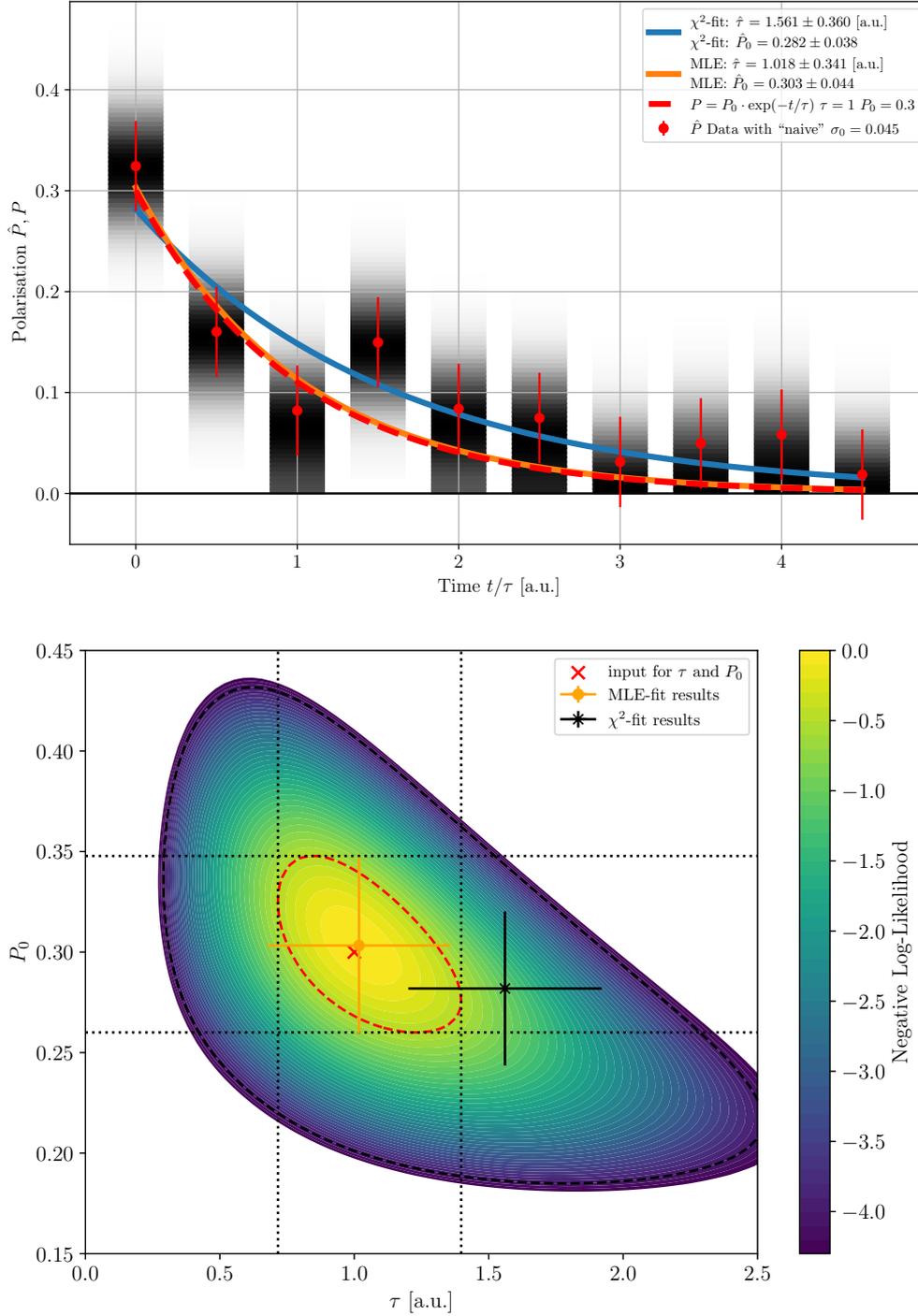


Figure 7.1: **Upper plot:** Maximum Likelihood Fit and χ^2 -fit to a simulated exponential decay, with estimated amplitudes \hat{P} following a rice distribution. Dashed red line: $P(t) = P_0 e^{-t/\tau}$ with $\tau = 1$ and $P_0 = 0.3$. Red data points: random values according to $f(\hat{P}|P(t_i), \sigma)$ shown with Gaussian error $\sigma = \sqrt{2/1000}$. Vertical bands: probability distribution $\tilde{f}(P|\hat{P}, \sigma)$ for true P for the given generated \hat{P} . The lines are the result of a least squares fit (blue) and a maximum likelihood fit (orange) to the data. **Lower plot:** Log Likelihood values around the maximum of the fit from Equation 7.2 and corresponding 1 σ and 2 σ regions in comparison with the χ^2 -fit results.

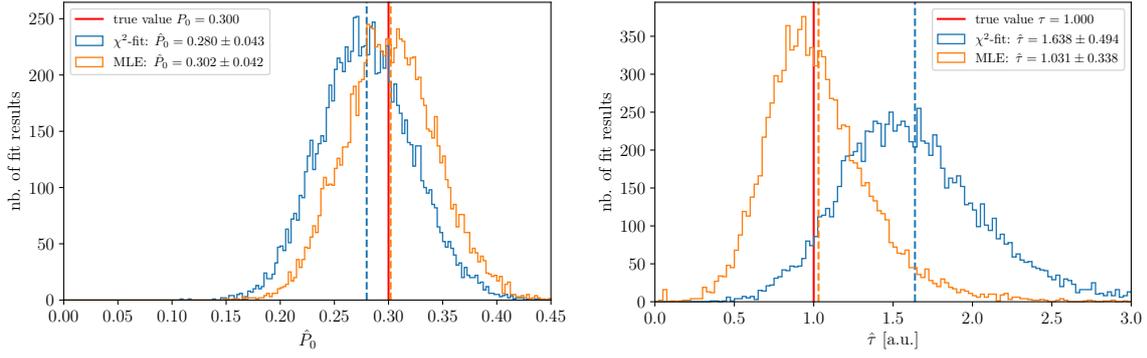


Figure 7.2: Distribution of \hat{P}_0 and $\hat{\tau}$ for 10000 simulations. Results from likelihood fit using the probability distributions and a least squares fit.

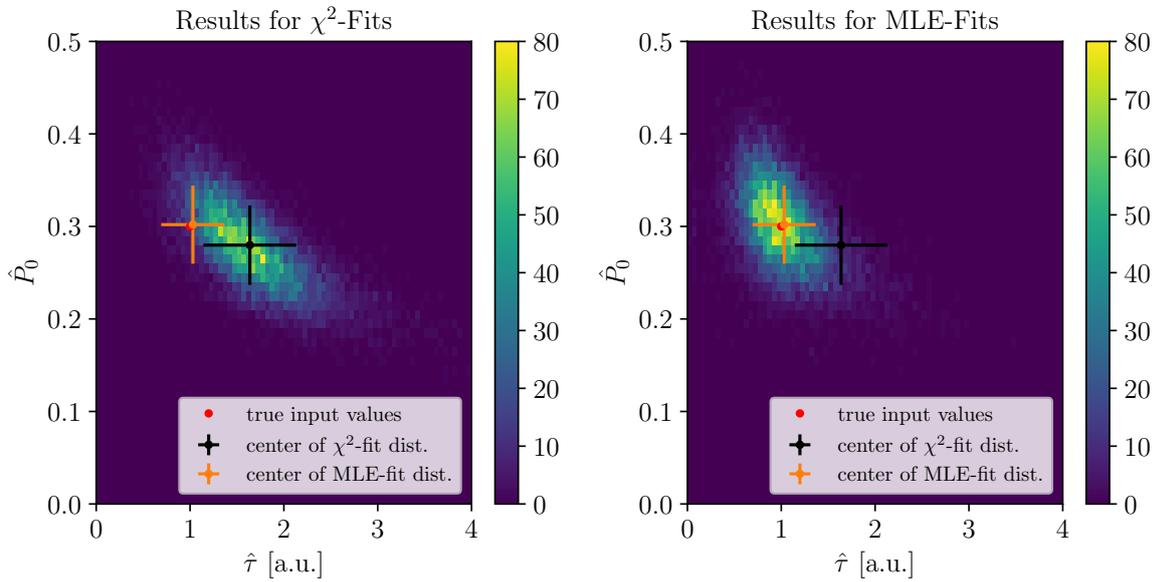


Figure 7.3: Comparison of the 2D histograms for the likelihood method vs. the simple χ^2 fit. For the 1 dimensional comparison along each axis see Figure 7.2.

10000 repetitions of the fit with the same starting parameters. The uncertainty on a fit by fit basis is roughly the same for both the simple χ^2 -fit, were Gaussian distributions for the uncertainty are assumed, and the maximum likelihood estimator using the Rice distribution.

The average of $\hat{\tau}$ is 1.64(49) for the least squares fit and 1.03(34) for the likelihood approach. Note that the uncertainty is not the uncertainty of the mean, but simply the standard deviation of the fit results from the 10000 repetitions.

Therefore on average the likelihood result has a remaining bias of $0.03/0.34 = 0.09$ of its statistical fluctuation, whereas the bias for the least square fit amounts to $0.64/0.49 = 1.29$. This proves that the likelihood fit using $\tilde{f}(\hat{P}|P)$ as a pdf gives a result much closer to the true value $\tau = 1$ in the case of large statistical uncertainties on individual data points.

The average of the amplitude \hat{P}_0 is 0.280(43) for the least squares fit and 0.302(42) for the likelihood approach. The average results for the amplitudes are shifted by a bias of

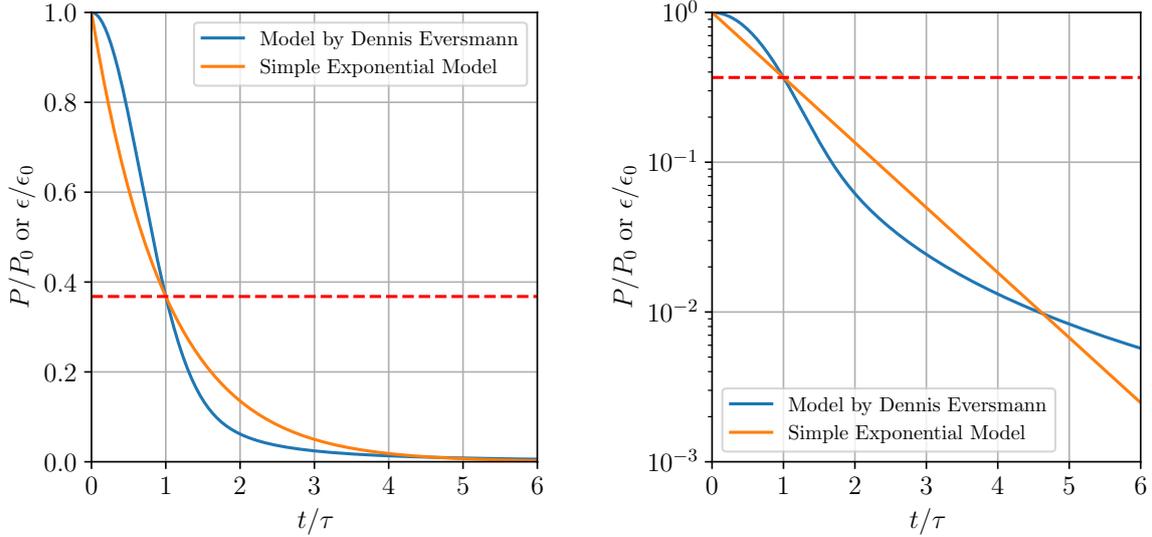


Figure 7.4: Comparison of the simple exponential model and the spin decoherence model by Dennis Eversmann (Equation 7.8) developed in [36] in a lin. and a log. plot.

less than 5% of the statistical uncertainty for the MLE and $\approx 50\%$ for the χ^2 -fit, while both estimators have roughly the same statistical uncertainty of 14%.

To get a better feeling for those results Figure 7.3 shows the 2D distribution of the 10000 fit results in bins of \hat{P}_0 and $\hat{\tau}$. The mean and standard deviations of both estimators are shown in comparison with the input values.

The fit results for the simple χ^2 -fit are both biased, especially the decay time τ , but also the starting amplitude P_0 . If the relative uncertainties get smaller (stronger signal strength per bin), especially relevant in the region where the signal is strong (approx until $t/\tau < 2$), the distributions of the results from the χ^2 -fits approach the MLE results and therefore the true input values. Then one can simply exclude regions with small signal strength from the fit, and obtains basically unbiased results, as it is done typically also in the analysis of the experimental data. Still, in regions with $t/\tau \gg 1$ it is problematic to quote the “naive” estimators with Gaussian errors, as they do not describe the true distributions, cover nonphysical negative amplitudes and imply that on average some small signal would remain. Using Feldman Cousins confidence intervals and or the correct Bayesian probability distributions cures this, being the more elaborate result. Nevertheless, the better the original data (greater signal strength), the less worthwhile the significantly increased effort using methods as described above as the final results the typically changed only withing less than 1σ of one single measurement (compare the bias of the Rice distribution). If one experiment is repeated with (exactly) the same initial conditions $N_{\text{exp.}}$ times, one can quote the uncertainty of the mean, where the statistical fluctuation scales $\propto 1/\sqrt{N_{\text{exp.}}}$. Then the bias becomes more and more interesting, since it stays constant.

7.2. SCT Results for Precursor Data

To study the spin coherence time τ_{SCT} , a polarized $0.97 \frac{\text{GeV}}{c}$ deuteron beam is injected into the storage ring and its polarization is tilted into the horizontal plane by means of the RF solenoid. The setup allows to easily determine the spin tune, since the spin performs an idle precession. Long and short-term measurements are essential to compare the different fitting methods.

In Dennis Eversmann's PhD thesis [36], a mathematical description of the loss of polarization is provided. It is based on the idea that the spin tunes of the individual particles of the ensemble are distributed in relation to the spin tune of the particle on the reference orbit, which is defined as ν_s^{ref} . In this model the probability density distribution of the spin tune ν_s is described using a Rayleigh distribution for $\nu_s > \nu_s^{\text{ref}}$. A scale parameter $\Delta\nu_s$ indicates the width of the distribution, i.e. for larger $\Delta\nu_s$ the spin tune spread of the ensemble is wider, which also leads to a faster depolarization, since the spin vectors are out of phase during less turns. The reference spin tune is assumed to be constant for a cycle.

According to the model the up-down asymmetry ($\epsilon_{UD}(t)$) during the decay of the spin polarization can be described by:

$$\epsilon_{UD}(\gamma_s(t)) = \epsilon_0 \left(\left[1 - \sqrt{\pi} \gamma_s(t) e^{-\gamma_s^2(t)} \text{erfi}(\gamma_s(t)) \right]^2 + \pi \gamma_s^2(t) e^{-2\gamma_s^2(t)} \right)^{\frac{1}{2}}, \quad (7.3)$$

$$\text{with } \gamma_s(t) = \sqrt{2\pi} \cdot \Delta\nu_s \cdot f_{\text{COSY}} \cdot t \quad \text{erfi}(x) = \frac{2}{\sqrt{\pi}} e^{x^2} D(x), \quad (7.4)$$

where $D(x)$ is the Dawson function. The parameter $\gamma_s(t)$ includes the spread of the spin tune distribution $\Delta\nu_s$, and is then the only fit parameter apart from the up-down asymmetry ϵ_0 at the beginning of the analyzed period (when the spin starts precessing in the plane). The Spin Coherence Time in this model can be calculated numerically by solving⁷

$$\epsilon_{UD}(\gamma_s(\tau_{\text{SCT}})) = \frac{1}{e} \quad \Rightarrow \quad \gamma_s(\tau_{\text{SCT}}) = \sqrt{2\pi} \Delta\nu_s f_{\text{COSY}} \tau_{\text{SCT}} \approx 1.571017 \quad (7.5)$$

$$\Rightarrow \quad \tau_{\text{SCT}} \approx \frac{0.353603}{\Delta\nu_s f_{\text{COSY}}} \quad \text{and} \quad \sigma_{\tau_{\text{SCT}}} \approx \frac{0.353603}{\Delta\nu_s^2 f_{\text{COSY}}} \sigma_{\Delta\nu_s}. \quad (7.6)$$

Note that a spin coherence time τ_n measured in turns n is directly linked to a SCT measured in time τ_{SCT} by $\tau_{\text{SCT}} = 1/f_{\text{COSY}} \cdot \tau_n$.

Figure 7.4 compares the shape of the simple exponential model, with the more elaborate spin decoherence model used to describe spin depolarization in the actual experiment. Note that the model starts with a horizontal slope, for $t/\tau \approx 1$ it falls faster than a simple exponential and for $t/\tau \gg 1$ it falls of more slowly in the graph.

Two examples of a measurement of the Spin Coherence Time (from the precursor 2 run) are shown in Figure 7.5 (left short SCT and right long SCT). The RF solenoid is used to rotate the vertical polarization into the horizontal plane and the beginning of the fit interval is therefore at 90s in the cycle. In the left panel, the up-down asymmetry drops

⁷The correct numerical result 0.353 603 that [36] quotes is not identical but close to $1/(2 \cdot \sqrt{2}) \approx 0.353 553$.

If it was exactly the latter value, then $\gamma_s(\tau_{\text{SCT}}) = \pi/2$, as it first appears to be when rounding in Equation 7.5 to less decimal places. But when zooming into Figure 7.4 to 10^{-3} around $t/\tau = 1$, one start to see a small deviation using this approximation.

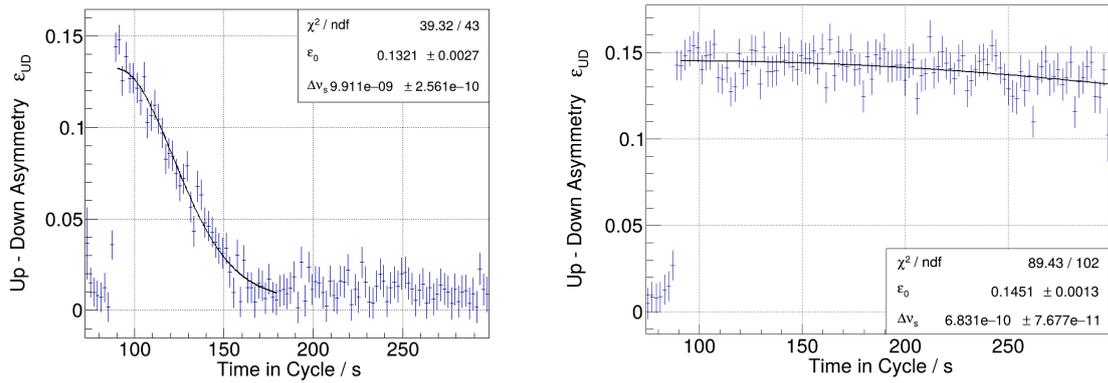


Figure 7.5: Two examples of measuring the Spin Coherence Time while changing the sextupole currents. These two plots (and the corresponding data) were taken from Achim Andres PhD thesis [35]. The left measurement yields a short Spin Coherence Time of (47.53 ± 1.23) s. The right panel shows a measurement of the Spin Coherence Time of (689.6 ± 77.5) s.

fast. The up-down asymmetry is on average rather constant and larger than zero after 150s. The reason behind is that the fitting bias of sine functions with small amplitudes and large uncertainties becomes prominent; also again some errorbars extend below zero. To reduce the influence of a possible bias the fit does not cover the full time range. To optimization the Spin Coherence Time the sextupole magnets in the COSY ring are adjusted by checking a grid of different possible currents. A long Spin Coherence Time along the entire measurement period is crucial for the success of the experiment. The right plot in Figure 7.7 shows a measurement of a better optimized Spin Coherence Time.

As a first step one can use the Feldman Cousins confidence intervals to get rid of any unphysical errorbars at the end of the cycle in Figure 7.5. The results are shown in Figure 7.6. The black data points represent the estimated parameters for the asymmetry amplitude in the time bin with obtained individual (gaussian) errors σ_0 . The average uncertainty per bin is ≈ 0.009 , which corresponds to on average ≈ 25000 evaluated events per bin. The corresponding $1\sigma \approx 68\%$ confidence interval from the Feldman Cousins approach is represented by the red horizontal lines. Note that these intervals can be on rare occasion much shorter than the Gaussian errors, if the measured amplitude is much smaller than the uncertainty (see e.g. the 6th bin before 200s in the short SCT case (upper plot in Figure 7.6) - this what was called a downward fluctuation in subsection 5.7. The FC confidence limits never go into the unphysical negative region. Many bins compatible with zero-asymmetry (as maximum likelihood ratio implies) are seen at the end of the short SCT measurement (see red dots) in Figure 7.6.

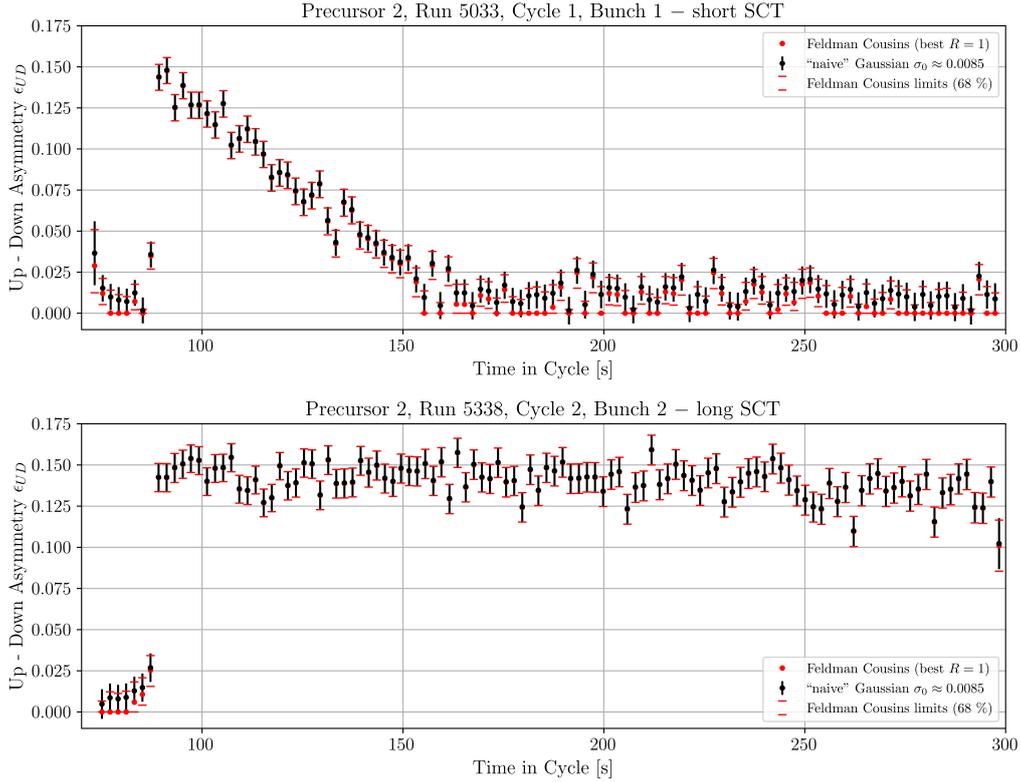


Figure 7.6: Turn depending amplitude for two different runs, where the Gaussian uncertainties and confidence limits from the Feldman-Cousins approach are compared. For the short SCT (upper plot) many measurements at the end of the cycle are linked to a vanishing true amplitude, but not much is changed for the long SCT, where all measurement show a large signal strength.

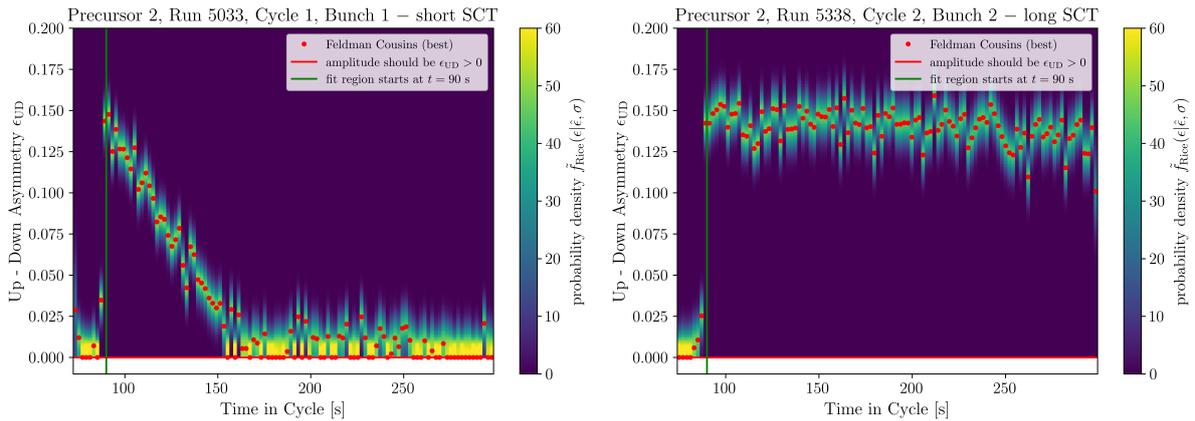


Figure 7.7: Another way to display the data form Figure 7.5. The measurements are first corrected to the corresponding asymmetries with maximum likelihood ratio $R = 1$ of the Feldman-Cousins approach (red points). The Bayesian probability distribution assuming a constant prior is overlaid.

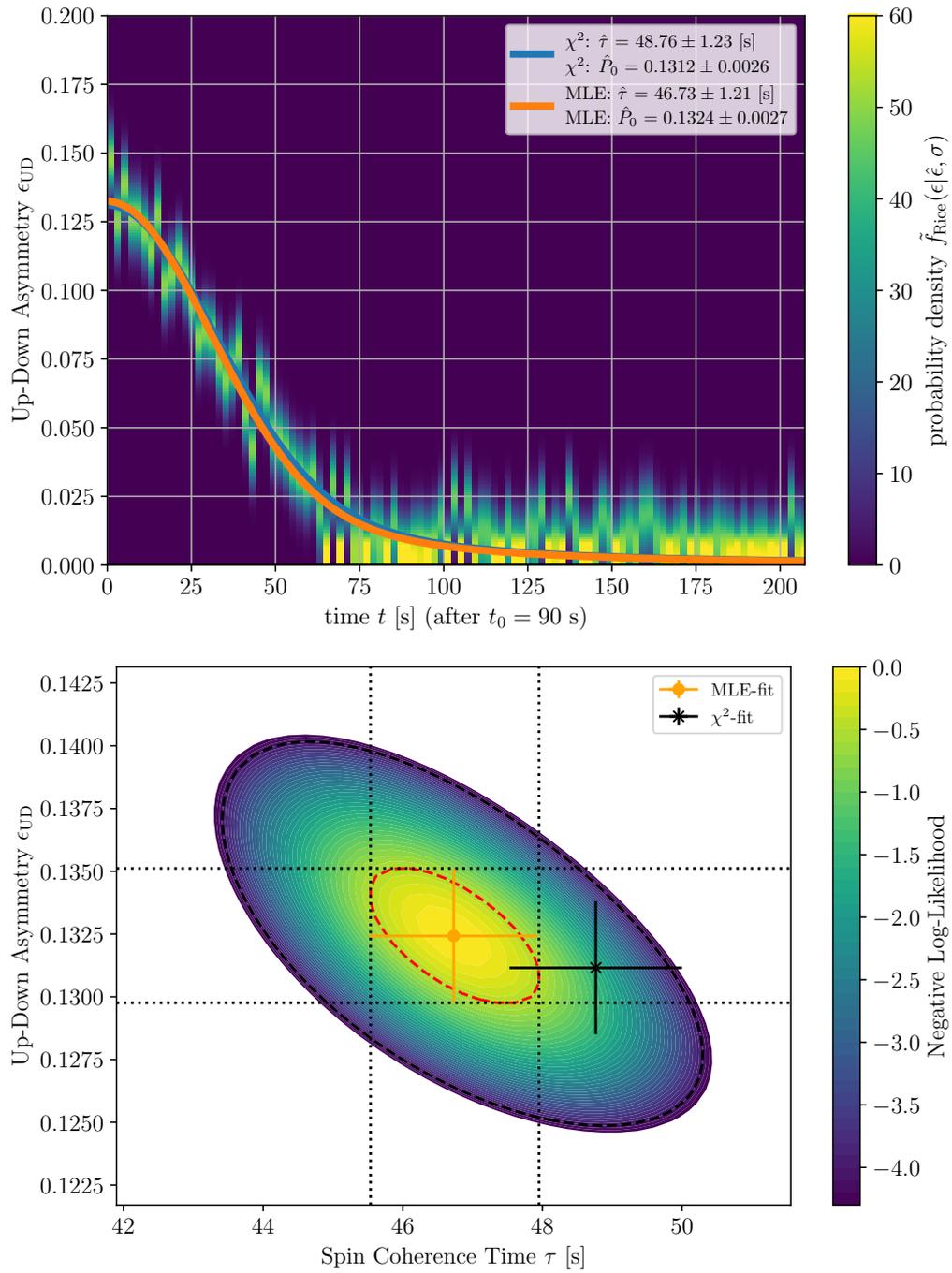


Figure 7.8: **Short SCT Measurement** from Precursor 2 experiment. Upper plot: Turn depending asymmetry amplitude over the measurement. The probability density function of the true amplitude based on the measured amplitude is indicated by the overlaid vertical bands. The orange curve represents a maximum likelihood fit, while the blue curve indicates the corresponding simple χ^2 -fit. Lower plot: The corresponding fit results are compared (including correlations), the dashed red distorted ellipses mark the 1σ region.

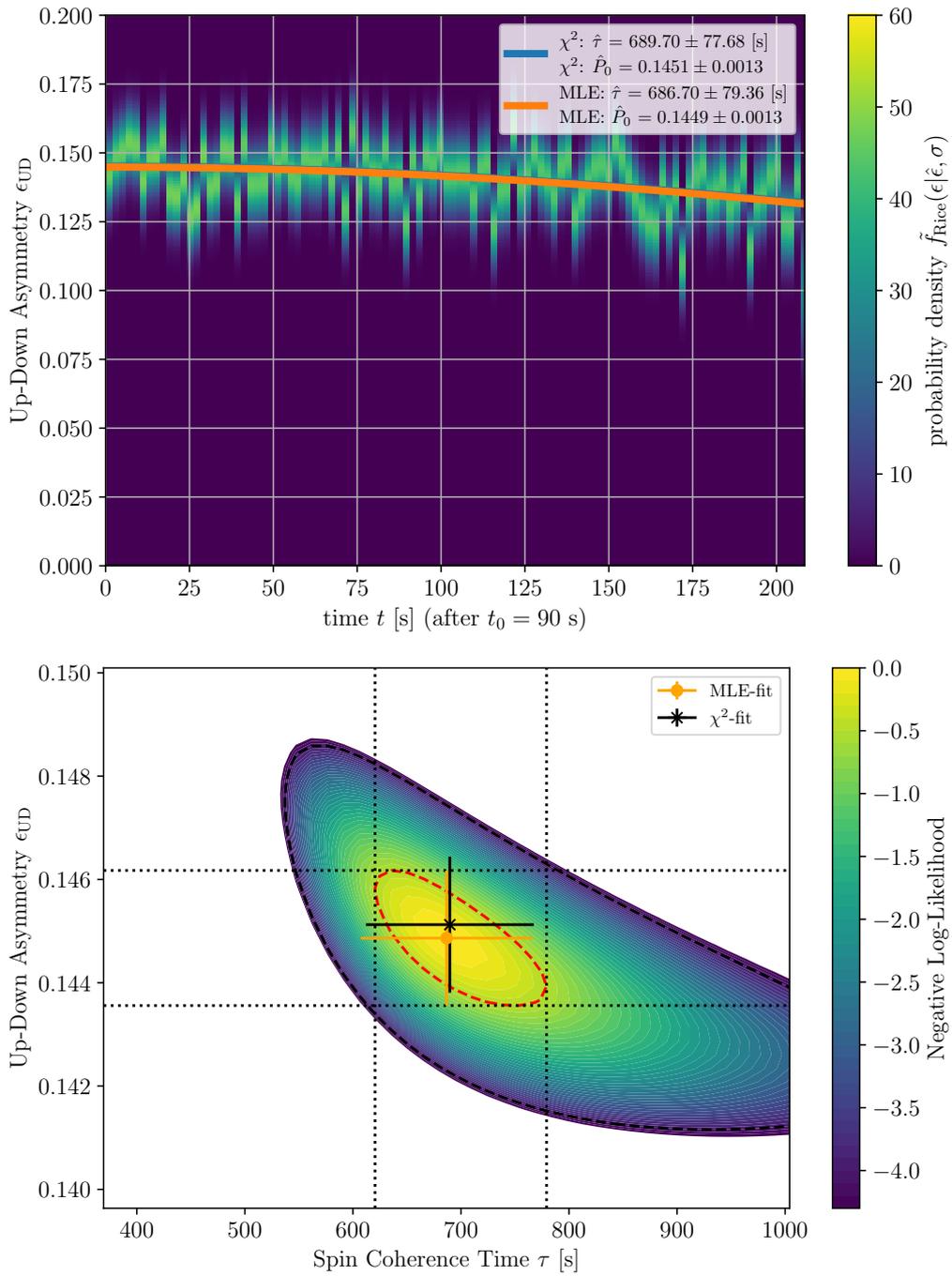


Figure 7.9: **Long SCT Measurement** from Precursor 2 experiment. Upper plot: Turn depending asymmetry amplitude over the measurement. The probability density function of the true amplitude based on the measured amplitude is indicated by the overlaid vertical bands. Again the orange curve represents the fit using the MLE, while the blue curve indicates the corresponding simple χ^2 -fit. Lower plot: The fit results are compared (including correlations), the dashed red distorted ellipses mark the 1σ region.

7.3. Maximum Likelihood Fit

The next step is to see, whether also the fits of the SCT can be substantially improved using the more elaborate statistical estimators. The likelihood function for amplitude fitting is given by

$$\mathcal{L} = \prod_{i=1}^{N_{\text{bin}}} \tilde{f}_{\text{Rice}}(\epsilon_{\text{UD}}^{\text{fit}}(t_i; \epsilon_0, \tau_{\text{SCT}}) | \hat{\epsilon}_i, \sigma_i) \quad \text{with} \quad \sigma_i = \sqrt{\frac{2}{N_{\text{ev},i}}} \quad (7.7)$$

where the uncertainty σ_i of the asymmetry ϵ_i in every bin $i \in [1, N_{\text{bin}}]$ is directly linked to the number of events $N_{\text{ev},i}$ in the bin used to construct the asymmetry.

$$\epsilon_{\text{UD}}^{\text{fit}}(t; \epsilon_0, \tau_{\text{SCT}}) = \epsilon_0 \left(\left[1 - \sqrt{\pi} \gamma_s(t) e^{-\gamma_s(t)^2} \text{erfi}(\gamma_s(t)) \right]^2 + \pi \gamma_s^2(t) e^{-2\gamma_s(t)^2} \right)^{\frac{1}{2}} \quad (7.8)$$

$$\text{with} \quad \gamma_s(t) = 0.353603 \sqrt{2\pi} \cdot t / \tau_{\text{SCT}}. \quad (7.9)$$

Figure 7.9 and Figure 7.8 show the Maximum Likelihood Estimator approach using the Bayesian probability distribution with constant prior for the given asymmetry amplitudes (runs 5033 and 5038, corresponding to the data from Figure 7.5).

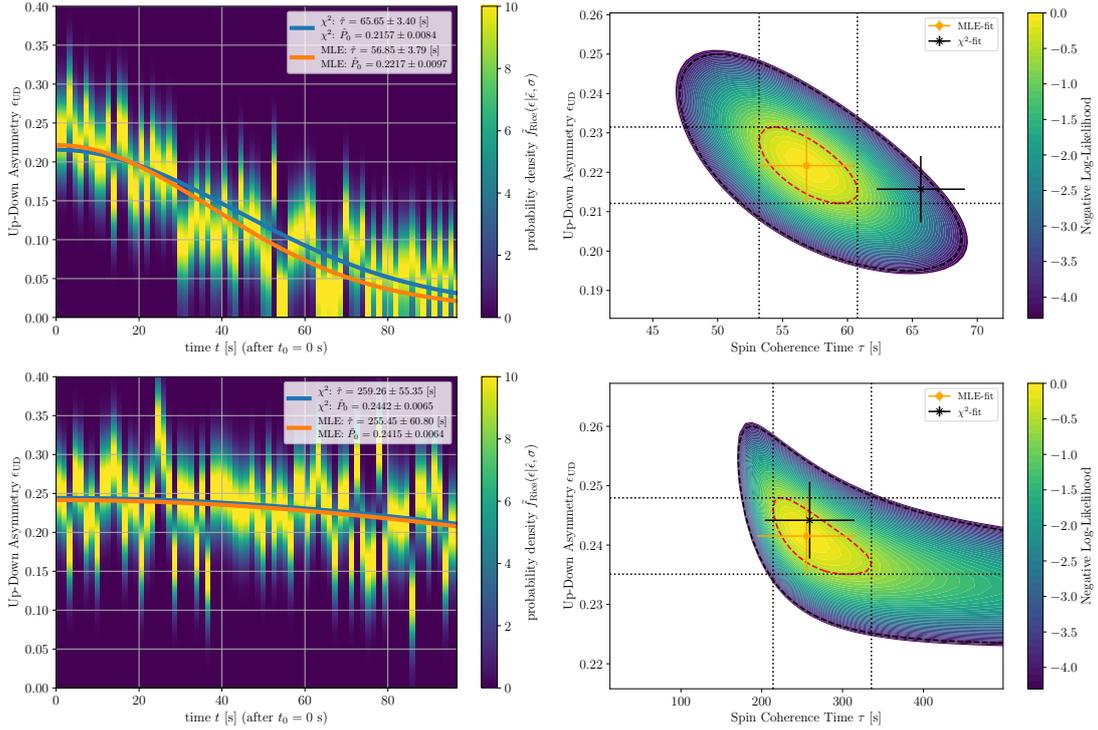


Figure 7.10: Time depending amplitude for a long and short spin coherence time (using old data from 2015 with an average $\sigma_0 \approx 0.036$) corresponding to 1600 events per bin. The color scheme in the left plots again represents the posterior probability density function according to the estimated values of the discrete turn Fourier transform assuming a Rice distribution. The right plots compare the fit results including correlation.

	short SCT τ [s]	short SCT ϵ_0 (P_0)	long SCT τ [s]	long SCT ϵ_0 (P_0)
A. Andres (fit limited)	(47.53 \pm 1.23)	(0.1321 \pm 0.0027)	(689.6 \pm 77.5)	(0.1451 \pm 0.0013)
χ^2 -fit (using full range)	(48.76 \pm 1.23)	(0.1321 \pm 0.0025)	(689.7 \pm 77.7)	(0.1451 \pm 0.0013)
MLE (using full range)	(46.73 \pm 1.21)	(0.1324 \pm 0.0027)	(686.7 \pm 79.4)	(0.1449 \pm 0.0013)
χ^2 -fit (old 2015 data)	(65.65 \pm 3.40)	(0.2157 \pm 0.0084)	(259.3 \pm 55.3)	(0.2442 \pm 0.0065)
MLE (old 2015 data)	(56.85 \pm 3.70)	(0.2217 \pm 0.0097)	(255.5 \pm 60.8)	(0.2415 \pm 0.0064)

Table 3: Parameters obtained by the global maximum likelihood fit based on the data for a long and short spin coherence time shown in Figure 7.9, Figure 7.8 and Figure 7.10.

In the case of the analyzed Precursor 2 runs, both MLE and χ^2 -fit offer a decent description of the data. For the short SCT both measurements are well compatible still withing $\approx 1.5\sigma$, of the statistical uncertainty as estimated from the maximum likelihood fit. For the long SCT measurements no difference in SCT can be resolved. However another interesting effect can be seen from Figure 7.9. For a long SCT measurement the relative statistical uncertainty of the SCT is much larger than for a short SCT measurement. Also the corresponding uncertainties are not ellipses. This can be understood, because the fits of the SCT start off with a flat slope, where it is naturally difficult for the model to predict the correct SCT. Short SCTs are excluded from not seeing a decline, but without extending the measurement, it is hard to exclude longer possible SCTs. The parameter ellipse is therefore elongated into large τ_{SCT} direction in Figure 7.9.

The fit results are tabulated in Table 3. The difference in estimated asymmetries from the fit is shown to be small, which is “good”, because one likes to have at best unbiased asymmetries, as those go into the EDM measurement directly in the form of ratios between horizontal and vertical asymmetry. For typical long SCTs of $\tau_{\text{SCT}} > 700$ s and above (1000s is one often mentioned goal in the experiment, e.g. compare [28] and [35]), the difference between the biased naive χ^2 -fit and the elaborate MLE for the amplitude is considerably less than 20% of the statistical uncertainty and in absolute numbers ≈ 0.0002 .

From the underlying theory introduced in subsection 5.6 this can be confirmed by the following expected bias, using some rough but motivated estimates. For a long SCT the asymmetry is not changing much over the whole measurement. Consider a starting signal amplitude $P_0 \approx 0.2$ and $N \approx 20000$ events per time bin, implying $\sigma_0 = \sqrt{2/N} = \frac{1}{100} = 0.01$. Therefore the average signal strength is $\bar{\epsilon} \approx P_0/\sigma_0 \approx 20$. The remaining bias per bin can be estimated using Equation 5.29 as

$$\langle \hat{\epsilon}_{\text{UD}} \rangle - \epsilon_{\text{UD}} \approx \sigma^2 \cdot \frac{1}{2P_0} = \sigma \cdot \frac{1}{2\bar{\epsilon}} \approx \sigma \cdot \frac{1}{2 \cdot 20} \approx 0.025 \cdot \sigma \approx 0.00025. \quad (7.10)$$

Since every bin has roughly the same bias, also the fit parameter P_0 will be biased the same way, therefore by ≈ 0.0002 . Note that this is compatible with the difference we see between the long SCT fits using Precursor 2 data in Table 3.

Note that also MLE estimators can show some small remaining bias, as was seen for the simple exponential model, but this bias scales with the total used events over the whole cycle (all bins) and will therefore be neglected, as it is much smaller than the χ^2 -fit bias. Also the results are only fully valid (meaning minimal bias is achieved), if the underlying model of the SCT is indeed correct. Within the given uncertainties at least small modification of the underlying fit-function seem possible and would be also

compatible with the data, but such considerations are beyond this work. Nevertheless the Bayesian approach for the observed amplitudes and the corresponding estimators could also be applied to any⁸ other desired SCT model.

Note that the experimental situation was different in the past, when the fitting method using the MLE was originally suggested. The plots in Figure 7.10 show SCT COSY runs from 2015, where the measurement interval was much shorter, and polarimeter i.e. detectors rates were worse too, leading to much greater statistical uncertainties. Therefore the short SCT χ^2 -fits are indeed noticeable different by eye when compared to the MLE approach. For the “long” SCT measurement (only $\tau_{\text{SCT}} \approx 200$ s) in the lower plot, the χ^2 -fit and the MLE results are still similar, since the Rice probability distribution approaches the Gaussian limit for moderate signal strengths already.

⁸The model should not predict negative asymmetry amplitudes in the allowed parameter space, since then the Bayesian Posterior Probability is undefined. If such unphysical regions can be accessed by some parameter combinations, they should be excluded from the fit.

8. Beam Oscillation Amplitude

For the measurement of the in-plane polarization the event rates of the upper and lower detectors are typically combined to cancel out the acceptance and luminosity factors of the beam. When analyzing the count rates in the individual detectors, the following effect was found. Its description here closely follows the work done by Achim Andres for his PhD thesis [35] (see pages 136-139), with then applying the statistical methods introduced above in the end. In the case that the electromagnetic fields inside the RF Wien filter do not exactly match to the momentum of the deuteron beam (a small Lorentz force remains), the beam experiences a kick in each revolution, leading to small coherent oscillations of the beam. An interesting approach to quantify the matching of the RF Wien filter is analyzing event rate changes in the polarimeter. A so called unmatched RF Wien filter induces coherent beam oscillations, where the number of deuterons hitting the carbon target in the polarimeter should periodically change by the frequency the Wien filter is kicking the beam f_{WF} .

This change of interaction rate is then also measured in the four quadrants of the detector after scattering on the carbon block. This can be described by adding an oscillation term to the luminosity of the beam

$$L_{\text{COSY}} \rightarrow L_{\text{COSY}} \cdot (1 + a \cos(2\pi f_{WF}t)), \quad (8.1)$$

where a denotes an amplitude, scaling with the oscillation amplitude of the beam, however the oscillation beam amplitude parameter cannot be trivially translated into an amplitude in “standard units” because a depends on additional beam parameters like beam position and vertical betatron oscillation amplitude. The event counting rates ($dN/dt = \dot{N}$) in the individual detector quadrants are given by

$$\dot{N}_{\text{Up}} \propto (1 + a \cos(\omega_{WF}n + \phi_{\text{rel}})) \cdot (1 - \epsilon_{\text{UD}} \cos(\omega_s n)), \quad (8.2)$$

$$\dot{N}_{\text{Down}} \propto (1 + a \cos(\omega_{WF}n + \phi_{\text{rel}})) \cdot (1 + \epsilon_{\text{UD}} \cos(\omega_s n)), \quad (8.3)$$

$$\dot{N}_{\text{Left}} \propto (1 + a \cos(\omega_{WF}n + \phi_{\text{rel}})) \cdot (1 + \epsilon_{\text{LR}}), \quad (8.4)$$

$$\dot{N}_{\text{Right}} \propto (1 + a \cos(\omega_{WF}n + \phi_{\text{rel}})) \cdot (1 - \epsilon_{\text{LR}}). \quad (8.5)$$

Ideally the Wien Filter and the spin precession matches $\omega_{WF} = \omega_s$. The expected Fourier amplitudes at the frequency of the Wien filter and spin tune in the left and right detector quadrants (needed for vertical polarization determination), where no oscillation of the spin tune does happen, are then directly given by

$$A_{\pm}(\omega = \omega_{WF} = \omega_s) = a. \quad (8.6)$$

One has to notice that those Fourier amplitudes (like other possible fitting methods) are typically biased for small amplitudes, as this is again as cosine estimation procedure, where the amplitude estimators are biased when the signals are small compared to its uncertainties. The measured amplitudes in the upper and lower quadrants $A_{\uparrow\downarrow}$, are more complicated as they are interwoven with the oscillating signal from the spin tune precession. For more details see page 137 of [35].

The findings of such Fourier amplitude considerations from the Precursor 1 data lead to a monitoring tool for the periodic change of luminosity while data taking, which was used during the Precursor 2 runs. This allowed to improve the matching of the RF Wien filter

	(signal) slope m [$10^{-6}\frac{1}{s}$]	(signal) offset b	(pilot) slope m [$10^{-6}\frac{1}{s}$]	(pilot) offset b
χ^2 -fit (Precursor 1)	149.5 ± 25.7	0.0300 ± 0.0016	-	-
MLE (Precursor 1)	151.3 ± 25.7	0.0294 ± 0.0016	-	-
χ^2 -fit (Precursor 2)	24.4 ± 6.5	0.0077 ± 0.0011	1.0 ± 6.5	0.0091 ± 0.0011
MLE (Precursor 2)	38.7 ± 10.1	0.0022 ± 0.0019	5.5 ± 13.3	0.0044 ± 0.0021

Table 4: Parameters obtained by the maximum likelihood and χ^2 -fit with corresponding uncertainties for beam oscillation amplitudes as shown in Figure 8.2.

field, and therefore minimizing the beam oscillation amplitude. For this tool the sum of the count rates in the individual detectors is used, since any polarization effects cancel out (see Equation 8.2) when combining all the data. Of course also more events per time bin reduce the statistical uncertainty of the beam oscillation amplitude parameter a .

$$\dot{N}_{\text{sum}} = \dot{N}_{\text{Up}} + \dot{N}_{\text{Down}} + \dot{N}_{\text{Left}} + \dot{N}_{\text{Right}} \propto 4 \cdot (1 + a \cos(\omega_{\text{WF}}n + \phi_{\text{rel}})). \quad (8.7)$$

By performing a Fourier transformation of Equation 8.7 and scaling by the total number of events per chosen time bin ($\dot{N}_{\text{sum}} = \dot{N}_{\text{Up}} + \dot{N}_{\text{Down}} + \dot{N}_{\text{Left}} + \dot{N}_{\text{Right}}$), the Fourier amplitude does not depend on the polarization or feedback settings and is simply given by

$$A_{\text{sum}}(\omega = \omega_{\text{WF}}) = a. \quad (8.8)$$

In Figure 8.1 the beam oscillation amplitude a for combining all the detectors is shown as a function of time in cycle for a measurement using a run from the Precursor 1 (upper) and the Precursor 2 experiment. For the Precursor 2 experiment two bunches are in the machine, the pilot bunch (middle) and the signal bunch (lower), which show very similar behavior.

In all cases, the RF Wien filter is switched on at 155 s indicated by a vertical black line. An unmatched Wien filter then starts to periodically kick the beam resulting in beam oscillations which are measured in the detectors. For the Precursor 1 data, a clear jump is visible, while this effect was prevented in the precursor 2 experiment. Since both pilot and signal bunch amplitude are almost unaffected by the operation of the RF Wien filter, one can deduce that the filter fields are matched more precisely. For the analysis, this relative change of the parameter when the RF Wien filter is switched off and on is the most relevant result.

However in both picture the naive interpretation would be, that always a small beam oscillation amplitude did in fact remain, and judging from their uncertainties they would be mostly constant and greater than zero. However this is misleading, when using the knowledge of amplitude overestimation, since the uncertainties of the amplitude determination are large compared to the average amplitude, and again this is the regime, where the bias of sine estimation is relevant. Additionally, the error bars (when wrongly interpreted as gaussian errors) sometimes extend to values below 0, which is not a correct description of the data, since the amplitudes can never be smaller than zero with the definitions used here. Correspondingly the error bars should also not be symmetric for small oscillation amplitudes.

By applying the Bayesian method and using a probability density for the Fourier amplitudes – for which the Rice distribution is a good choice as shown in section 6 – one

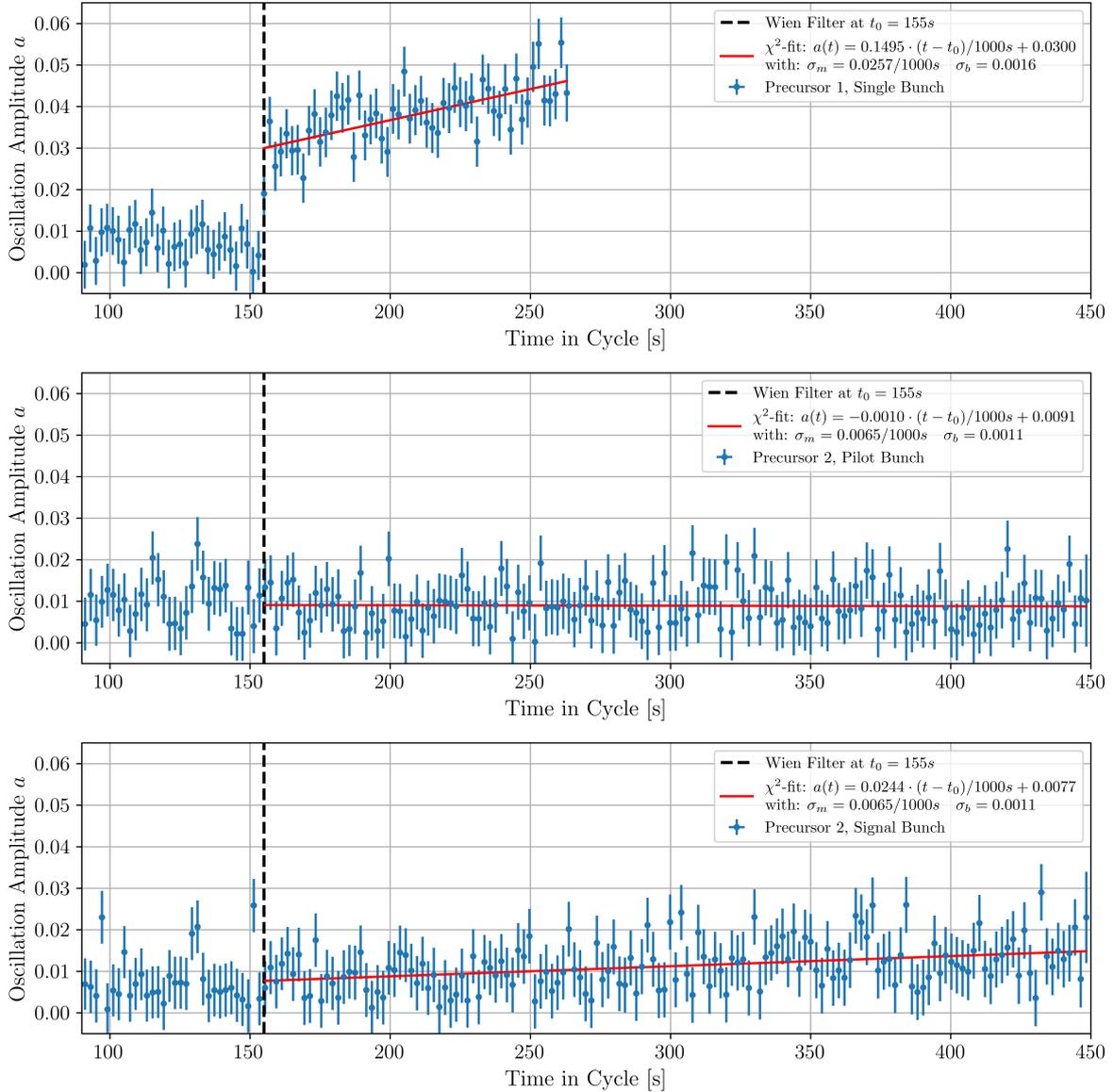


Figure 8.1: Comparison of a beam oscillation amplitude measurement during a Precursor 1 and a Precursor 2 run (where two bunches are in the COSY ring). The raw data included in this plots was directly taken from the PhD Thesis of Achim Andres (see Figure 7.41 of [35]). The style is different (each bunch in its own plot) to better compare it to the analysis using the Maximum Likelihood estimator in Figure 8.2.

can correct or at least vastly reduce the bias effect. The result for such a procedure are shown in Figure 8.2; note that now no estimated amplitude extends below zero. For the Precursor 1 data (only one single bunch in the machine) the beam oscillations parameters (slope m and offset b) are typically both pronounced. Therefore it makes no difference, whether a simple χ^2 -fit or the more elaborate MLE is used. The results for the two fits match within less than the 1σ . Especially the slope is almost identical, for both methods.

For the so called pilot bunch in Precursor 2 data, which is not effected by the RF Wien

filter fields, the Maximum Likelihood fit confirms the result from the χ^2 -fit, that there is no build up of the amplitude over the measurement. The slope (parameter m) in the fit stays well compatible with zero. The starting amplitude (parameter b) is much better compatible with zero in the 2σ region, when compared to the least squares result, where $b = 0$ would be $\approx 8\sigma$ away from the desired zero result.

Interestingly the starting amplitude b shows a very similar behavior for the signal bunch. It is compatible to 0 within $\approx 1\sigma$ for the MLE approach, while a simple χ^2 -fit would result in 7σ . One sees a non-zero slope of the beam oscillation amplitude for the signal bunch already in the simple χ^2 -fit that stays at roughly 4σ in both analysis approaches. Since the simple χ^2 -fit does not start at a $b = 0$ because of the bias, it underestimates the slope of the amplitude of the measurement by a factor of roughly two. The fit results are summarized in Table 4.

Note that the apparent buildup in this analysis, which is seen for the signal bunch, is not directly the buildup of the vertical polarization, which is used to determine the Invariant Spin Axis. The vertical polarization can be calculated directly from the left-right asymmetry of detector events (which is proportional to the vertical polarization) at a given time bin, and does not depend on a frequency related analysis (e.g. Fourier Transform). Therefore it is not suffering from the bias effect discussed in this chapter and naturally starts at zero amplitude. Only when looking at the so called beam oscillation amplitude parameter a , the bias is introduced in the analysis, but this should not disturb the determination of the invariant spin axis with the methods used in Achim Andres PhD Thesis [35].

Still it might be interesting to further investigate also the behavior of the oscillation amplitude, as the behavior of the Wien Filter is of great interest for the precision of the measurement (e.g. precision of Wien filter fields). Anyhow one can confirm the absence of beam oscillations at the Wien filter frequency (assumed equal to the spin tune frequency) for the pilot bunch, and sees some signal buildup in the oscillation amplitude for the signal bunch. This signal would be strongly underestimated using conventional fitting methods in a linear model. To interpret it in terms of a further analysis related to the invariant spin axis however turns out to be difficult, since correction terms would be needed. E.g. for the determination of the invariant spin axis one has to correct the corresponding left-right asymmetry (vertical polarization) using unpolarized cycles by a linear model first (see [35] pages 63-66), to say something about the effects on polarized data. This avoids so called fake buildup, which is seen when the Wien Filter is turned on for the signal data. Similar corrections would be probably needed in this case as well, to get a better feeling for the nature of this beam oscillation amplitude. But if one wants to study a signal buildup using Fourier amplitudes in the region of small signal strength, it is necessary to e.g. use posterior probability distributions, as simple χ^2 -fits are naturally strongly biased in the region.

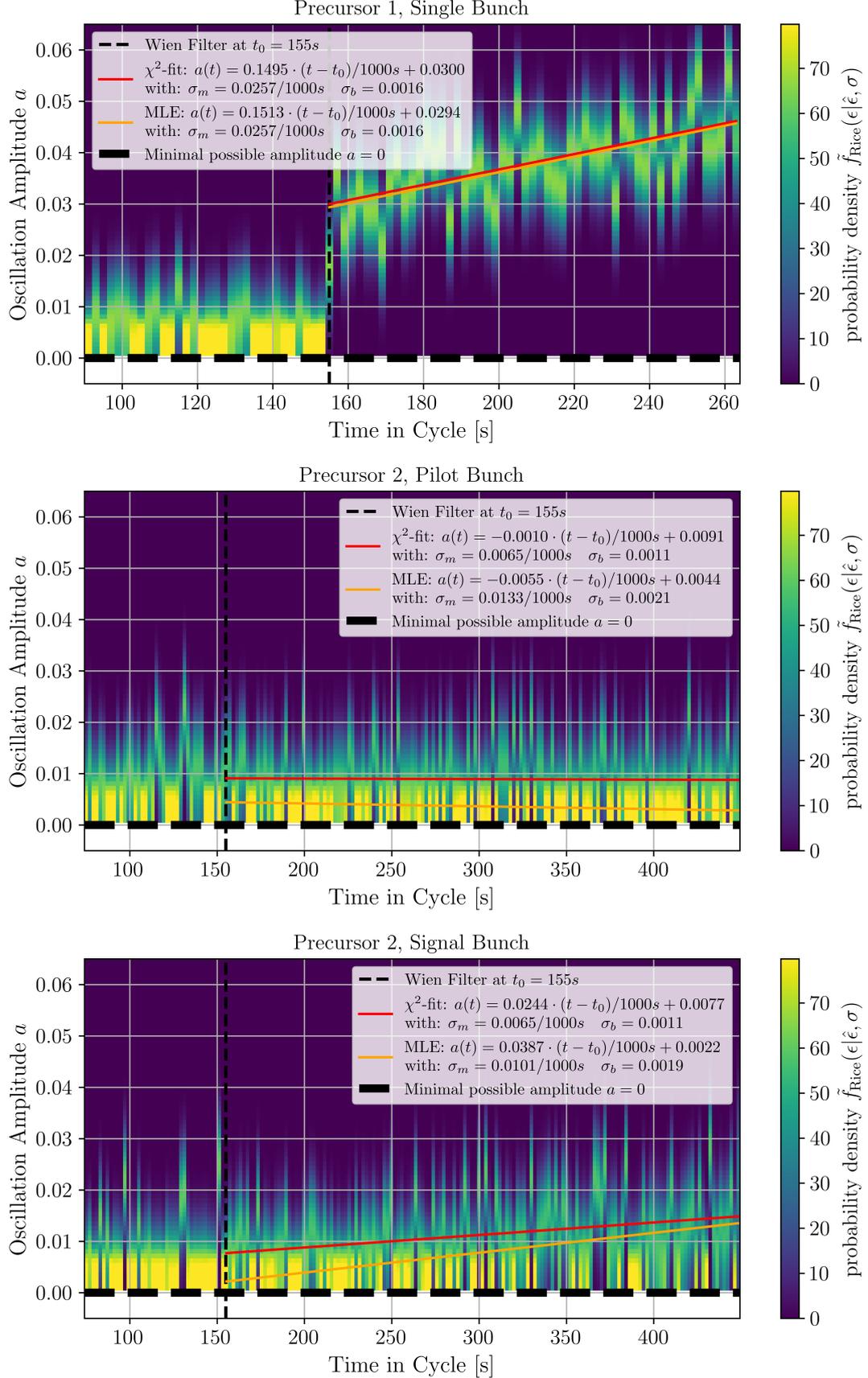


Figure 8.2: Comparison of a beam oscillation amplitude measurement and corresponding fits using Gaussian errors (χ^2 -fit in red) and probability distributions from the Bayesian approach (MLE-fit in orange).

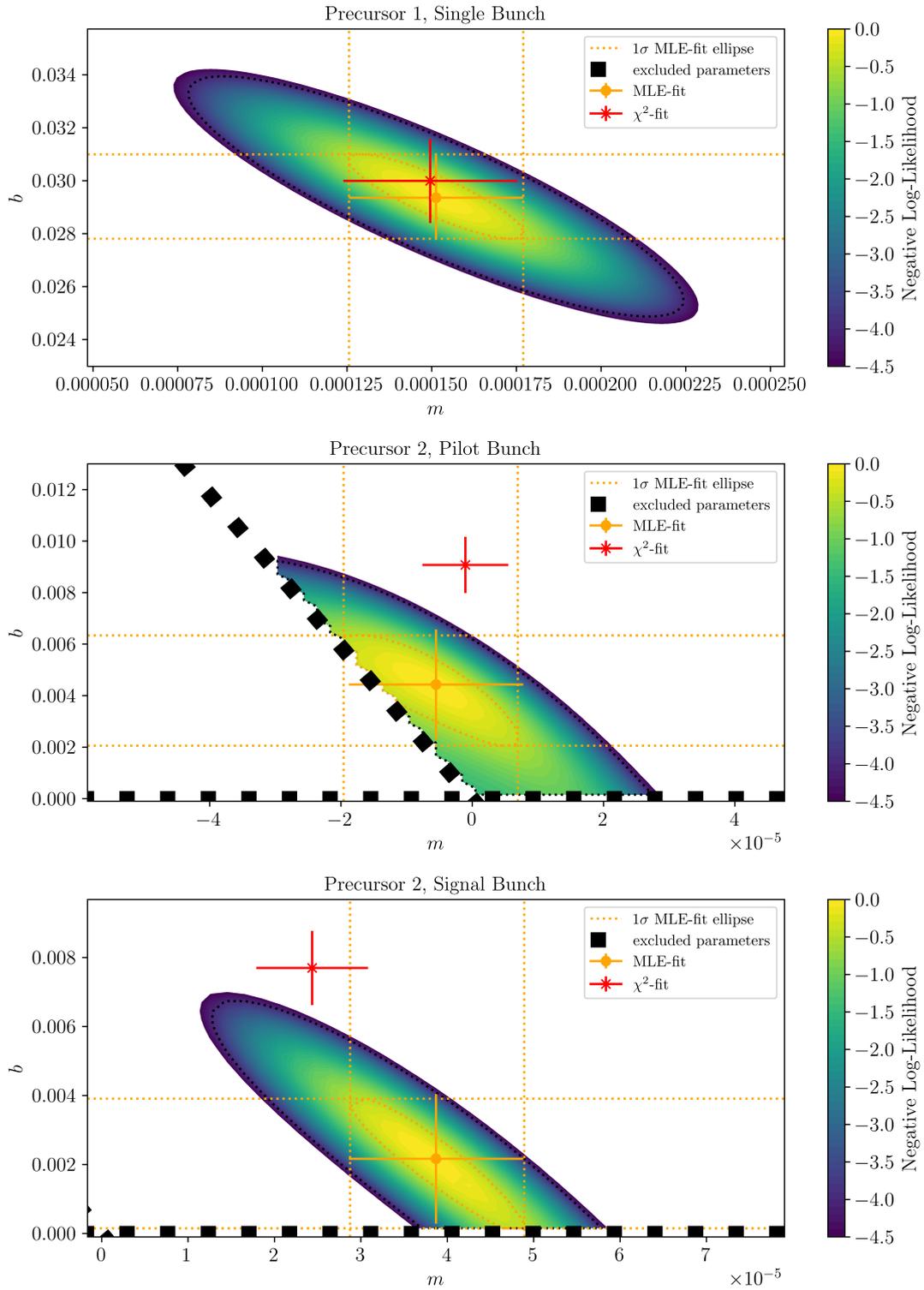


Figure 8.3: Comparison of the obtained fit parameters and their log-likelihood functions showing correlations for the fits shown in Figure 8.2. For the MLE using the RICE distribution only parameter combinations, which fulfill the requirement of non-negative amplitudes for the full measurement cycle, are allowed (black exclusion lines).

9. Conclusion

The JEDI collaboration aims to directly measure the electric dipole moment (EDM) of charged elementary particles, first of all deuterons, with the existing storage ring COSY at Forschungszentrum Jülich during the so-called precursor experiment. In a magnetic ring horizontally polarized particles precess around the vertical axis with the so-called spin tune frequency. The EDM induces an oscillating vertical polarization due to the spin precession, however this oscillation is too small to be detected directly. However measurable signal can be accumulated using a radio frequency Wien Filter which – at the desired momentum – ideally does not affect the orbit of the particles, but acts only as a polarization manipulator. However, also various systematic effects can induce a vertical polarization buildup of the beam, independent from the EDM. Therefore any possible systematic effect has to be reduced to their absolute minimum. One at least possible source of systematic effects is not the machine or experiment itself, but the data analysis procedure. The polarization is measured with a polarimeter: Particles hit a target and scatter, depending on their polarization, into four different segments of a detector called up, down, left and right detector as seen from the beam. From the corresponding counts in the detector, all the signals have to be reconstructed using parameter estimation.

Parameter estimations play an important role in all area of science. However, estimates of parameters obtained from simple least squares fits, which assume Gaussian errors can often be biased, even for simple scenarios like the estimation of the amplitude of a sine-function. This bias is especially important and prominent when the measured quantity is small compared to its uncertainty. If the relative uncertainty gets small, limiting theorems and statistical properties usually ensure Gaussian behavior, where such considerations are not longer essential. Additionally always assuming Gaussian uncertainties may introduce coverage in non-physical regions of the parameters, in the case of large relative uncertainties.

In this work the Feldman-Cousins algorithm was used to construct proper confidence intervals covering only the allowed region $P \geq 0$ for the amplitude P of a sine-function and the results were applied to the corresponding physical quantities such as count rate asymmetries, which are linked to spin polarization. Furthermore those Feldman-Cousins confidence intervals were compared with the corresponding Bayesian credible intervals in detail. The posterior probability density functions for the true amplitude found by applying Bayes' theorem was used to study the reduction of fitting bias effects in spin coherence time measurements (SCT) and applied to a quantity called beam oscillation amplitude.

For the spin coherence time fits it was proven that the Bayesian posterior probability density for a true amplitude, given some measured amplitude, using a maximum likelihood approach leads to better (meaning less biased) fit results compared to simple least squares fits in the case of dominant statistical uncertainties. For the analysis a simple exponential toy model of the SCT and a more elaborated model containing the spin tune spread (developed by Dennis Eversmann) were tested.

In the Precursor experiments it turns out that for the SCT measurements the relative uncertainties are typically already quite small, therefore the Bayesian estimators indeed approach the Gaussian limits. As the SCT should be longer than the measurement time for optimal results of the experiment anyhow (at best $\tau_{\text{SCT}} > 1000$ s), also the problem of fitting in bias dominated regions with small remaining amplitudes can typically be avoided.

The correlations between initial asymmetry (polarization) and SCT are visualized, and by that also an interesting effect is demonstrated: A measurement with a fitted long SCT typically has a much larger relative uncertainty on τ_{SCT} than a short SCT measurement.

The bias from the fitting procedure remains considerably smaller than the statistical uncertainty of individual measurements for short SCTs and is even less relevant for desired long SCT measurements. Only for introductory test measurements with much worse signal strength and very short SCT obvious differences between a simple least squares fit and the Bayesian Maximum Likelihood Estimator approach were found. Since long SCTs are crucial for the quality of EDM experiments, the influence of the described fitting bias would go down further, when the SCT and in general the average signal strength per time bin would be improved even more in potential future experiments.

Important physical quantities – like horizontal and vertical polarization respectively asymmetry – for the established EDM measurement approach involving the determination of the invariant spin axis are mostly unchanged by the bias from the analysis procedure. The left-right asymmetry (linked to vertical polarization buildup) is typically constructed using direct approaches, which do not involve the biased estimators discussed here. The up-down asymmetry (horizontal polarization) is in principle biased from the analysis procedure as discussed in this work, however by ensuring a large SCT and large average signal strengths the bias is small compared to the statistical uncertainty of the measurements. Especially when (rare) regions with small signal strengths are simply excluded from the analysis, the bias problems can be mostly avoided without following the more complex procedures described in this work and sticking to the conventional and established methods.

The situation is different for a quantity called beam oscillation amplitude, which is not directly related to parameters needed for the EDM measurement but related to the behavior of the bunches in the machine and the influence of the Wien filter. This device is interesting, since longitudinal tilts of beam or magnetic field axis have an impact on the buildup of the vertical component of the spin vector and systematic effects in the Wien Filter are not yet fully understood. When studying the Fourier transform of the event rates in all detectors at the Wien filter frequency the behavior of the beam oscillation amplitude with time cannot be correctly described using simple least squares fits, as uncertainties dominate and the Fourier amplitudes of those oscillations are biased. When applying the Bayesian probability theory from this work an expected zero signal for the oscillation amplitude from bunches unaffected by the Wien Filter can be proven. For the signal bunches simple fits underestimate the slope of the signal buildup by a factor of 2, since they start from a biased nonzero amplitude. This amplitude buildup however cannot be easily related to the already established EDM measurement parameters. Therefore, further research would be necessary to determine, whether the behavior of this so-called beam oscillation parameter can actually provide helpful insights for the EDM measurement.

A. Calculations and Proofs

A.1. Proof of the MSE decomposition

The following proof of the MSE decomposition uses only simple and well know properties:

$$\begin{aligned}
B(\hat{\theta}) &= E[\hat{\theta}] - \theta \quad \Leftrightarrow \quad E[\hat{\theta}] = \theta + B(\hat{\theta}) \quad E[E[\hat{\theta}]] = E[\hat{\theta}] \quad E[\theta] = \theta \\
\text{Var}(\hat{\theta}) &= E[(\hat{\theta} - E[\hat{\theta}])^2] = E[\hat{\theta}^2 - 2\hat{\theta}E[\hat{\theta}] + E[\hat{\theta}]^2] = E[\hat{\theta}^2] - E[2\hat{\theta}E[\hat{\theta}]] + E[\hat{\theta}]^2 \\
&= E[\hat{\theta}^2] - 2E[\hat{\theta}]E[\hat{\theta}] + E[\hat{\theta}]^2 = E[\hat{\theta}^2] - E[\hat{\theta}]^2 \\
\text{MSE}(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] = E[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2] = E[\hat{\theta}^2] - E[2\hat{\theta}\theta] + \theta^2 \\
&= E[\hat{\theta}^2] - 2E[\hat{\theta}]\theta + \theta^2 = \underbrace{E[\hat{\theta}^2] - E[\hat{\theta}]^2}_{\text{Var}(\hat{\theta})} + \underbrace{E[\hat{\theta}]^2 - 2E[\hat{\theta}]\theta + \theta^2}_{(B(\hat{\theta}))^2} \\
&= \text{Var}(\hat{\theta}) + (\theta + B(\hat{\theta}))^2 - 2(\theta + B(\hat{\theta}))\theta + \theta^2 \\
&= \text{Var}(\hat{\theta}) + \cancel{\theta^2} + 2B(\hat{\theta})\theta + B(\hat{\theta})^2 - \cancel{2\theta^2} - 2B(\hat{\theta})\theta + \cancel{\theta^2} = \text{Var}(\hat{\theta}) + (B(\hat{\theta}))^2
\end{aligned}$$

When the parameter θ is not a scalar but a vector (multiple parameters), one typically chooses the Euclidean norm to define the MSE [38]. Then an analogous decomposition applies:

$$\text{MSE}(\hat{\theta}) = \text{MSE}(\|\hat{\theta} - \theta\|^2) = \text{trace}(\text{Cov}(\hat{\theta})) + \|B(\hat{\theta})\|^2 = \sum_i (\text{Var}(\hat{\theta}_i) + (B(\hat{\theta}_i))^2) \quad (\text{A.1})$$

where $\text{trace}(\text{Cov}(\hat{\theta}))$ is the trace of the covariance matrix of the estimator and $\|B(\hat{\theta})\|^2$ is the Euclidean square vector norm of the bias vector. Note that this decomposition also still works component wise.

A.2. Fourier Transform of a Sine and Cosine Function

The discrete turn Fourier transform of the function $f[n] = P \cos(2\pi\nu_s n - \phi_s) = A \sin(2\pi\nu_s n) + B \cos(2\pi\nu_s n)$ is defined as:

$$\hat{f}(\nu) = \sum_{n=0}^{\infty} f[n]e^{-in \cdot 2\pi\nu} = \sum_{n=0}^{\infty} [A \sin(2\pi\nu_s n) + B \cos(2\pi\nu_s n)] e^{-in \cdot 2\pi\nu} \quad (\text{A.2})$$

Using Euler's formula, the sine and cosine function can be expressed as:

$$\sin(2\pi\nu_s n) = \frac{1}{2i} (e^{+i(2\pi\nu_s n)} - e^{-i(2\pi\nu_s n)}) \quad (\text{A.3})$$

$$\cos(2\pi\nu_s n) = \frac{1}{2} (e^{+i(2\pi\nu_s n)} + e^{-i(2\pi\nu_s n)}) \quad (\text{A.4})$$

Therefore one gets for the Fourier transform

$$f(\nu) = \sum_{n=0}^{\infty} \left[\frac{A}{2i} (e^{+i(2\pi\nu_s n)} - e^{-i(2\pi\nu_s n)}) + \frac{B}{2} (e^{+i(2\pi\nu_s n)} + e^{-i(2\pi\nu_s n)}) \right] e^{-in \cdot 2\pi\nu} \quad (\text{A.5})$$

$$= \sum_{n=0}^{\infty} \frac{A}{2i} (e^{+i(2\pi(\nu_s - \nu)n)} - e^{-i(2\pi(\nu_s + \nu)n)}) + \frac{B}{2} (e^{+i(2\pi(\nu_s - \nu)n)} + e^{-i(2\pi(\nu_s + \nu)n)}) \quad (\text{A.6})$$

Here only the components at $\nu = \nu_s$ and $\nu = -\nu_s$ are relevant (as the sum goes to infinity). For all other values of ν the exponential terms are cyclic if $2\pi(\nu_s \pm \nu) \cdot n$ approaches some integer multiple of 2π , which happens always for $n \rightarrow \infty$. Therefore those terms are small and zero in the limiting case. Now assume that the sum goes to some index $1 \ll N \rightarrow \infty$ and only evaluate at $\nu = \nu_s$. In the limiting case one gets:

$$\hat{f}(\nu = \nu_s) = \sum_{n=0}^N \frac{A}{2i} \underbrace{\left(e^{+i(2\pi(\nu_s - \nu_s)n)} \right)}_{=e^0=1} + \frac{B}{2} \underbrace{\left(e^{+i(2\pi(\nu_s - \nu_s)n)} \right)}_{=e^0=1}, \quad (\text{A.7})$$

$$= \frac{NA}{2i} + \frac{NB}{2} = \frac{N}{2}(B - iA). \quad (\text{A.8})$$

Therefore one can easily see, that indeed up to some factor $2/N$ the real part of the DFT is given by A and the imaginary part is given by B .

Note that the result from DFT is similar to the (non discrete) Fourier transform of the cosine function (or sine function in complete analogy) $B \cos(\omega_s t) \equiv B \cos(2\pi\nu_s n)$. The prefactor $\frac{1}{2\pi}$ is needed to represent the angular frequency ω instead of ν in the final result. Using again the exponential representation of the cosine the Fourier transform is given by:

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} B \cos(\omega_s t) \cdot e^{-i\omega t} dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{B}{2} \left(e^{+i(\omega_s t)} + e^{-i(\omega_s t)} \right) \cdot e^{-i\omega t} dt. \quad (\text{A.9})$$

This can be split into two integrals:

$$F(\omega) = \frac{1}{2\pi} \frac{B}{2} \left(\int_{-\infty}^{\infty} e^{+i(\omega_s - \omega)t} dt + \int_{-\infty}^{\infty} e^{-i(\omega_s + \omega)t} dt \right). \quad (\text{A.10})$$

Each integral is of the form:

$$\int_{-\infty}^{\infty} e^{i\alpha t} dt = 2\pi\delta(\alpha), \quad (\text{A.11})$$

where $\delta(\alpha)$ is the Dirac delta function. Therefore:

$$F(\omega) = \frac{1}{2\pi} \frac{B}{2} (2\pi\delta(\omega - \omega_s) + 2\pi\delta(\omega + \omega_s)). \quad (\text{A.12})$$

Simplifying this expression one gets

$$F(\omega) = \mathcal{F}\{B \cos(\omega_s t)\} = \frac{B}{2} [\delta(\omega - \omega_s) + \delta(\omega + \omega_s)]. \quad (\text{A.13})$$

In this case again only the true frequency $\omega = \omega_s$ (and the negative equivalent $\omega = -\omega_s$) appear in the Fourier representation.

A.3. Fourier Transform of the Rectangular Function

Consider the rectangular function $\text{rect}_N(t)$, which is defined as:

$$\text{rect}_N(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq N, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.14})$$

Applying the (continuous) Fourier transformation to the rectangular function $\text{rect}_N(t)$:

$$F(\omega) = \int_{-\infty}^{\infty} \text{rect}_N(t) e^{-i\omega t} dt = \int_0^N e^{-i\omega t} dt \quad (\text{A.15})$$

This integral can be evaluated as:

$$F(\omega) = \left. \frac{e^{-i\omega t}}{-i\omega} \right|_0^N = F(\omega) = \frac{1}{-i\omega} (e^{-i\omega N} - e^0) = \frac{1}{i\omega} (1 - e^{-i\omega N}) \quad (\text{A.16})$$

Rewriting the expression into a more useful form gives:

$$F(\omega) = \frac{1}{i\omega} (e^{i\frac{\omega N}{2}} - e^{-i\frac{\omega N}{2}}) e^{-i\frac{\omega N}{2}} = \frac{2}{\omega} \left(\sin\left(\frac{\omega N}{2}\right) \right) e^{-i\frac{\omega N}{2}} \quad (\text{A.17})$$

This expression indicates that the Fourier transform of the rectangular function $\text{rect}_N(t)$ is an unnormalized sinc function, scaled by N and shifted in frequency:

$$F(\omega) = \mathcal{F}\{\text{rect}_N(t)\} = N \frac{\sin\left(\frac{\omega N}{2}\right)}{\frac{\omega N}{2}} e^{-i\frac{\omega N}{2}} = N \cdot \text{sinc}\left(\frac{\omega N}{2}\right) e^{-i\frac{\omega N}{2}} \quad (\text{A.18})$$

where the sinc function is defined as: $\text{sinc}(x) = \frac{\sin(x)}{x}$

Another possible way to define the rectangular function $\text{rect}(t)$ is to be 1 for $|t| \leq \frac{1}{2}$ and 0 for $|t| > \frac{1}{2}$. The unitary Fourier transform of the rectangular function in this case is given by:

$$\int_{-\infty}^{\infty} \text{rect}(t) \cdot e^{-i2\pi ft} dt = \frac{\sin(\pi f)}{\pi f} = \text{sinc}(\pi f). \quad (\text{A.19})$$

B. Likelihood Method - Variance of Estimators

B.1. Likelihood Method for A and B

Using the theory for maximum likelihood estimation, one can calculate the asymptotic variance for the estimators \hat{A} and \hat{B} given N measurements following the PDF in Equation 5.8. Starting from the probability density function, the log-likelihood function reads

$$\ell = \ln\left(\prod_{i=1}^N f(x_i; A, B)\right) = \sum_{i=1}^N \ln\left(\frac{1}{2\pi}\right) + \ln(1 + A \sin(x_i) + B \cos(x_i)). \quad (\text{B.1})$$

Using the second derivatives

$$\frac{\partial^2 \ell}{\partial A^2} = -\sum_i \frac{\sin^2(x)}{(1 + A \sin(x) + B \cos(x))^2}, \quad (\text{B.2})$$

$$\frac{\partial^2 \ell}{\partial B^2} = -\sum_i \frac{\cos^2(x)}{(1 + A \sin(x) + B \cos(x))^2}, \quad (\text{B.3})$$

$$\frac{\partial^2 \ell}{\partial A \partial B} = -\sum_i \frac{\sin(x) \cos(x)}{(1 + A \sin(x) + B \cos(x))^2}, \quad (\text{B.4})$$

and their expectation values, where $C = \sqrt{1 - A^2 - B^2} = \sqrt{1 - P^2} \leq 1$ for $0 \leq P \leq 1$ is used:

$$\begin{aligned}
\left\langle \frac{\partial^2 \ell}{\partial A^2} \right\rangle &= \int_0^{2\pi} - \sum_i \frac{\sin^2(x)}{(1 + A \sin(x) + B \cos(x))^2} \cdot \frac{1}{2\pi} (1 + A \sin(x) + B \cos(x)) dx_i \\
&= -N \int_0^{2\pi} \frac{\frac{1}{2\pi} \sin^2(x)}{1 + A \sin(x) + B \cos(x)} dx = -N \frac{B^2 C - A^2 C - B^2 C^2 + A^2}{C(B^2 + A^2)^2} \\
&= \frac{([A^2 - B^2]C + B^2 C^2 - A^2)N}{CP^4} = \frac{(A^2[C - 1] + B^2[C^2 - C])N}{CP^4}, \\
\left\langle \frac{\partial^2 \ell}{\partial B^2} \right\rangle &= -N \int_0^{2\pi} \frac{\frac{1}{2\pi} \cos^2(x)}{1 + A \sin(x) + B \cos(x)} dx = -N \frac{A^2 C - B^2 C - A^2 C^2 + B^2}{C(B^2 + A^2)^2} \\
&= \frac{([B^2 - A^2]C + A^2 C^2 - B^2)N}{CP^4} = \frac{(B^2[C - 1] + A^2[C^2 - C])N}{CP^4}, \\
\left\langle \frac{\partial^2 \ell}{\partial A \partial B} \right\rangle &= -N \int_0^{2\pi} \frac{\frac{1}{2\pi} \sin(x) \cos(x)}{1 + A \sin(x) + B \cos(x)} dx = -N \frac{-2ABC + ABC^2 + AB}{C(B^2 + A^2)^2} \\
&= \frac{(2ABC - ABC^2 - AB)N}{CP^4} = \frac{-(C^2 - 2C + 1)ABN}{CP^4} = \frac{-(C - 1)^2 ABN}{CP^4}.
\end{aligned}$$

Note that the results are fully symmetrical $A \Leftrightarrow B$. The inverse of the covariance matrix for $(\theta_1, \theta_2) \equiv (A, B)$ is:

$$\text{cov}^{-1}(A, B) = - \left\langle \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right\rangle = \begin{pmatrix} \frac{-(A^2[C-1] + B^2[C^2 - C])N}{CP^4} & \frac{(C-1)^2 ABN}{CP^4} \\ \frac{(C-1)^2 ABN}{CP^4} & \frac{-(B^2[C-1] + A^2[C^2 - C])N}{CP^4} \end{pmatrix}. \quad (\text{B.5})$$

The inverse of a 2×2 matrix is calculated via the determinant and the adjugate matrix. The determinant of the matrix is $\det(\text{cov}^{-1}(A, B)) = \frac{N^2(C-1)^2}{CP^4}$.

$$\det(\text{cov}^{-1}(A, B)) = \frac{N^2(A^4 C^2 - 2A^4 C + A^4 + 2A^2 B^2 C^2 - 4A^2 B^2 C + 2A^2 B^2 + B^4 C^2 - 2B^4 C + B^4)}{CP^8} \quad (\text{B.6})$$

$$= \frac{N^2(C^2 - 2C + 1)}{CP^4} = \frac{N^2(C - 1)^2}{CP^4} \quad (\text{B.7})$$

From the inversion of the inverse covariance matrix one finds the variances. Note the typical behavior of all entries $\propto N^{-1}$:

$$\text{cov}(A, B) = \frac{1}{\det(\text{cov}^{-1}(A, B))} \cdot \text{adj}(\text{cov}^{-1}(A, B)) = \begin{pmatrix} \frac{A^2 C + B^2}{N(1-C)} & -\frac{AB}{N} \\ -\frac{AB}{N} & \frac{A^2 + B^2 C}{N(1-C)} \end{pmatrix}. \quad (\text{B.8})$$

$$\sigma_A = \sqrt{\frac{A^2 C + B^2}{N(1-C)}} \quad \sigma_B = \sqrt{\frac{A^2 + B^2 C}{N(1-C)}} \quad \text{cov}_{A,B} = -\frac{AB}{N} \quad (\text{B.9})$$

The correlation coefficient for A and B is only zero for $A = 0$ or $B = 0$, when the sine

or cosine term vanishes, otherwise it is given by

$$\rho_{A,B} = \frac{\text{cov}_{A,B}}{\sigma_A \sigma_B} = \frac{-\frac{AB}{N}}{\sqrt{\frac{A^2C+B^2}{N(1-C)} \cdot \frac{A^2+B^2C}{N(1-C)}}} = -\frac{AB}{N} \cdot \sqrt{\frac{N(1-C)}{A^2C+B^2} \cdot \frac{N(1-C)}{A^2+B^2C}} \quad (\text{B.10})$$

$$= -\frac{AB(1-C)}{\sqrt{(A^2C+B^2) \cdot (A^2+B^2C)}} = -\frac{AB(1-C)}{\sqrt{(A^4+B^4)C + A^2B^2(C^2+1)}} \quad (\text{B.11})$$

To understand the limiting behavior for σ_A in the case $P \ll 1$ it is convenient to write $A^2 = P^2 - B^2$. The Taylor expansion up to order $\mathcal{O}(P^4)$ is given by

$$\sigma_A^2 = \frac{A^2\sqrt{1-P^2} + B^2}{N(1-\sqrt{1-P^2})} = \frac{1}{N} \frac{(P^2 - B^2)\sqrt{1-P^2} + B^2}{1-\sqrt{1-P^2}} \stackrel{P^2 \ll 1}{\approx} \frac{2}{N} \left(1 - \frac{3P^2 - 2B^2}{4} - \frac{P^4}{16} \right)$$

And of course the same applies to σ_B , with $B^2 = P^2 - A^2$:

$$\sigma_B^2 = \frac{A^2 + B^2\sqrt{1-P^2}}{N(1-\sqrt{1-P^2})} \stackrel{P^2 \ll 1}{\approx} \frac{2}{N} \left(1 - \frac{3P^2 - 2A^2}{4} - \frac{P^4}{16} \right)$$

Notice that the only remaining terms for $A = B = P = 0$ is $2/N$. For small amplitudes $0 < P \ll 1$ corrections do apply, always reducing σ_A and σ_B because the $\mathcal{O}(P^2)$ term is negative. The error on $P = \sqrt{A^2 + B^2}$ calculated from Gaussian error propagation is:

$$\begin{aligned} \sigma_P^2 &= \left(\frac{\partial P}{\partial A} \sigma_A \right)^2 + \left(\frac{\partial P}{\partial B} \sigma_B \right)^2 + 2 \frac{\partial P}{\partial A} \frac{\partial P}{\partial B} \sigma_A \sigma_B \rho_{A,B} = \frac{A^2 \sigma_A^2 + B^2 \sigma_B^2 + 2AB \sigma_A \sigma_B \frac{\text{cov}_{A,B}}{\sigma_A \sigma_B}}{P^2} \\ &= \frac{A^2 \frac{A^2C+B^2}{N(1-C)} + B^2 \frac{A^2+B^2C}{N(1-C)} + 2AB \frac{-AB}{N}}{P^2} = \frac{A^4C + A^2B^2 + A^2B^2 + B^4C - 2A^2B^2(1-C)}{P^2N(1-C)} \\ &= \frac{A^4C + B^4C + 2A^2B^2C}{P^2N(1-C)} = \frac{P^2C}{N(1-C)} = \frac{P^2\sqrt{1-P^2}}{N(1-\sqrt{1-P^2})} = \frac{P^2}{N(\sqrt{1/(1-P^2)} - 1)} \end{aligned}$$

The error on $\varphi = \arctan A/B$ also calculated by Gaussian error propagation is:

$$\begin{aligned} \sigma_\varphi^2 &= \left(\frac{\partial \varphi}{\partial A} \sigma_A \right)^2 + \left(\frac{\partial \varphi}{\partial B} \sigma_B \right)^2 + 2 \frac{\partial \varphi}{\partial A} \frac{\partial \varphi}{\partial B} \sigma_A \sigma_B \rho_{A,B} = \frac{B^2 \sigma_A^2 + A^2 \sigma_B^2 + 2B(-A) \sigma_A \sigma_B \frac{\text{cov}_{A,B}}{\sigma_A \sigma_B}}{P^4} \\ &= \frac{B^2 \frac{A^2C+B^2}{N(1-C)} + A^2 \frac{A^2+B^2C}{N(1-C)} - 2AB \frac{-AB}{N}}{P^4} = \frac{A^2B^2C + B^4 + A^4 + A^2B^2C + 2A^2B^2(1-C)}{P^4N(1-C)} \\ &= \frac{A^4 + B^4 + 2A^2B^2}{P^4N(1-C)} = \frac{1}{N(1-C)} = \frac{1}{N(1-\sqrt{1-P^2})} = \frac{\sqrt{1/(1-P^2)}}{N(\sqrt{1/(1-P^2)} - 1)} \end{aligned}$$

B.2. Likelihood Method for Amplitude and Phase

Starting from the probability density function

$$f(x; P, \varphi) = \frac{1}{2\pi} (1 + P \cos(x - \varphi)), \quad 0 \leq x < 2\pi, \quad (\text{B.12})$$

the log-likelihood function reads

$$\ell = \ln L = \sum_{i=1}^N \ln\left(\frac{1}{2\pi}\right) + \ln(1 + P \cos(x_i - \varphi)). \quad (\text{B.13})$$

Using the second derivatives

$$\frac{\partial^2 \ell}{\partial P^2} = - \sum_i \frac{\cos^2(x - \varphi)}{(1 + P \cos(x - \varphi))^2}, \quad (\text{B.14})$$

$$\frac{\partial^2 \ell}{\partial \varphi^2} = - \sum_i \frac{-P(P + \cos(x - \varphi))}{(1 + P \cos(x - \varphi))^2}, \quad (\text{B.15})$$

$$\frac{\partial^2 \ell}{\partial P \partial \varphi} = - \sum_i \frac{\sin(x - \varphi)}{(1 + P \cos(x - \varphi))^2} \quad (\text{B.16})$$

and their expectation values

$$\left\langle \frac{\partial^2 \ell}{\partial P^2} \right\rangle = N \frac{1 - \sqrt{1/(1 - P^2)}}{P^2}, \quad (\text{B.17})$$

$$\left\langle \frac{\partial^2 \ell}{\partial \varphi^2} \right\rangle = N \frac{1 - \sqrt{1/(1 - P^2)}}{\sqrt{1/(1 - P^2)}}, \quad (\text{B.18})$$

$$\left\langle \frac{\partial^2 \ell}{\partial P \partial \varphi} \right\rangle = 0. \quad (\text{B.19})$$

the variances of P and φ can be calculated directly, since the corresponding matrix is diagonal. The correction terms up to $\mathcal{O}(P^4)$ are also given by

$$\sigma_P^2 = \frac{-1}{\left\langle \frac{\partial^2 \ell}{\partial P^2} \right\rangle} = \frac{1}{N} \frac{P^2}{\sqrt{1/(1 - P^2)} - 1} \stackrel{P^2 \ll 1}{\approx} \frac{2}{N} \left(1 - \frac{3}{4}P^2 - \frac{1}{16}P^4\right), \quad (\text{B.20})$$

$$\sigma_\varphi^2 = \frac{-1}{\left\langle \frac{\partial^2 \ell}{\partial \varphi^2} \right\rangle} = \frac{1}{N} \frac{\sqrt{1/(1 - P^2)}}{\sqrt{1/(1 - P^2)} - 1} \stackrel{P^2 \ll 1}{\approx} \frac{2}{NP^2} \left(1 - \frac{1}{4}P^2 - \frac{1}{16}P^4\right). \quad (\text{B.21})$$

C. Derivation of the Rice Distribution

Starting from the two dimensional probability distribution (where $f(\hat{A}|B)$ and $f(\hat{B}|B)$ are both simple Gaussians) the transformation to \hat{P} and $\hat{\varphi}$ including the ‘‘polar’’ Jacobian determinant \hat{P} yields

$$f(\hat{A}|A)f(\hat{B}|B)d\hat{A}d\hat{B} = \frac{1}{2\pi\sigma^2} e^{-(\hat{A}-A)^2/(2\sigma^2)} e^{-(\hat{B}-B)^2/(2\sigma^2)} d\hat{A}d\hat{B} \quad (\text{C.1})$$

$$f(\hat{P}, \hat{\varphi}|P, \varphi) d\hat{P} d\hat{\varphi} = \frac{1}{2\pi\sigma^2} e^{-(\hat{P} \sin \hat{\varphi} - P \sin \varphi)^2/(2\sigma^2)} e^{-(\hat{P} \cos \hat{\varphi} - P \cos \varphi)^2/(2\sigma^2)} \hat{P} d\hat{P} d\hat{\varphi} \quad (\text{C.2})$$

$$= \frac{1}{2\pi\sigma^2} e^{-(\hat{P}^2 + P^2)/(2\sigma^2)} e^{-(2\hat{P}P(\sin \varphi \sin \hat{\varphi} + \cos \varphi \cos \hat{\varphi}))/2\sigma^2} \hat{P} d\hat{P} d\hat{\varphi} \quad (\text{C.3})$$

Performing the integration over all possible angles $\hat{\varphi}$ the Rice distribution is obtained (with I_0 as the modified Bessel function of first kind).

$$\int_0^{2\pi} e^{(2P\hat{P}(\sin(\varphi)\sin(\hat{\varphi})+\cos(\varphi)\cos(\hat{\varphi}))/2\sigma^2)} d\hat{\varphi} = 2\pi I_0\left(\frac{P\hat{P}}{\sigma^2}\right) \quad (\text{C.4})$$

$$\Rightarrow f(\hat{P}|P) = \frac{\hat{P}}{\sigma^2} e^{-(\hat{P}^2+P^2)/2\sigma^2} I_0\left(\frac{P\hat{P}}{\sigma^2}\right). \quad (\text{C.5})$$

C.1. Transition from a Rice Distribution to a Gaussian

The Bessel function can be approximated as

$$I_\alpha(z) \rightarrow \frac{e^z}{\sqrt{2\pi z}} \left(1 - \frac{4\alpha^2 - 1}{8z} + \dots\right) \text{ as } z \rightarrow \infty \quad (\text{C.6})$$

so, in the large $\hat{P}P/\sigma^2$ region, an asymptotic expansion of the Rice distribution is given by:

$$f(x, P, \sigma) = \frac{\hat{P}}{\sigma^2} \exp\left(\frac{-(\hat{P}^2 + P^2)}{2\sigma^2}\right) I_0\left(\frac{\hat{P}P}{\sigma^2}\right) \quad (\text{C.7})$$

$$\approx \frac{\hat{P}}{\sigma^2} \exp\left(\frac{-(\hat{P}^2 + P^2)}{2\sigma^2}\right) \sqrt{\frac{\sigma^2}{2\pi\hat{P}P}} \exp\left(\frac{2\hat{P}P}{2\sigma^2}\right) \left(1 + \frac{\sigma^2}{8\hat{P}P} + \dots\right) \quad (\text{C.8})$$

$$\text{approx } \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\hat{P} - P)^2}{2\sigma^2}\right) \sqrt{\frac{\hat{P}}{P}}, \text{ as } \frac{\hat{P}P}{\sigma^2} \rightarrow \infty \quad (\text{C.9})$$

When the density is concentrated close around P and $|\hat{P} - P| \ll \sigma$ because of the Gaussian exponent, one can also approximate $\sqrt{\hat{P}/P} \approx 1$ and finally gets the Gaussian approximation:

$$f(\hat{P}, P, \sigma) \approx \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\hat{P} - P)^2}{2\sigma^2}\right), \quad \frac{P}{\sigma} \gg 1 \quad (\text{C.10})$$

This approximation becomes well usable for $\frac{P}{\sigma} \approx 3$ and larger.

C.2. Bayes-Theorem for Gaussian with Constant Prior

Consider a Gaussian probability distribution $\mathcal{N}(x|\mu, \sigma)$ and for simplicity assume the variance $V = \sigma^2$ is known from the start. Therefore one does not have to write σ as a parameter of the distribution, and without loss of generality considers the Gaussian PDF where $\sigma = 1$.⁹

$$\mathcal{N}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \quad (\text{C.11})$$

$$P(B|A) \equiv \mathcal{N}(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - \mu)^2\right). \quad (\text{C.12})$$

Here the constant prior for the true mean is defined as $P(A) \equiv P(\mu) \equiv \frac{1}{C}$ in such a way that $\int_{-\infty}^{\infty} P(A) dA = 1$. Note that no constant $C \in \mathbb{R}$ fulfills this, but obviously

⁹This is equivalent to a PDF considering the Gaussian depending on a signal to noise ratio $\gamma = \mu/\sigma$.

$\int_{-c/2}^{+c/2} \frac{1}{c} dA = 1$, and one simply lets $c \rightarrow \infty$ to reach what is described with C in the limiting case. Also notice how the parameter space is not constrained and therefore $x, \mu \in (-\infty, +\infty)$.

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\int_A P(B|A)P(A) dA} \quad (\text{C.13})$$

$$P(\mu|x) = \frac{\mathcal{N}(x|\mu) \cdot P(\mu)}{\int \mathcal{N}(x|\mu) \cdot P(\mu) d\mu} = \frac{\frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}(x - \mu)^2\right) \cdot \frac{1}{\mathcal{L}}}{\int \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}(x - \mu)^2\right) \cdot \frac{1}{\mathcal{L}} d\mu} \quad (\text{C.14})$$

$$\begin{aligned} &= \frac{\frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}(x - \mu)^2\right)}{\underbrace{\int \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}(x - \mu)^2\right) d\mu}_{=\int \mathcal{N}(x|\mu) d\mu=1}} = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}(x - \mu)^2\right) = \mathcal{N}(x|\mu) = \mathcal{N}(\mu|x) \end{aligned} \quad (\text{C.15})$$

Indeed $P(A|B) = P(B|A)$ for the Gaussian distribution with constant prior, and also both PDF are normal distributions with the same true mean at $x = \mu$.

D. Mapping Method

It is not possible to determine the spin precession frequency f_s directly from the observed event rates by a simple least squares fit with ν_s as a parameter, because the spin precession frequency is given by $f_s = |\nu_s| \cdot f_{\text{RF}} \approx SI0.16 \cdot 750 \text{ kHz} = 120 \text{ kHz}$, and for a detector with an average event rate of a few kilo hertz, multiple spin revolutions happen between each detection.

The algorithm described below is used to map all events from a chosen turn interval Δn into one oscillation period to accumulate enough statistics to extract properly the amplitude and phase of the precession. The algorithm generates asymmetries, which has the advantages, that those are largely independent of variations of acceptance or flux. A turn interval $\Delta n = 10^6$ corresponds to a time interval $\Delta T = 1/f_{\text{RF}} \cdot \Delta n \approx 1.3 \text{ s}$.

For the analysis some (fixed) spin tune ν_s^0 has to be chosen, which is typically close around the expected spin tune. Each of the turn intervals is analyzed independently, and the events are mapped into a 4π interval depending on their spin phase advance

$$\phi_{s,\text{map}}^{\nu_s^0} = 2\pi\nu_s^0 n \bmod 4\pi = \phi_s^{\nu_s^0}(n) \bmod 4\pi. \quad (\text{D.1})$$

This yields the event counts for the up $N_U(\phi_{s,\text{map}}^{\nu_s^0})$ and the down $N_D(\phi_{s,\text{map}}^{\nu_s^0})$ detector quadrant. An example is given in Figure D.1. Here in total $N_{\text{tot}} \approx 10^5$ events were recorded (≈ 6000 for the up and ≈ 4000 for the down detector). The 4π interval is divided into $N_{\text{bins}} = 40$ bins (meaning 20 bins per 2π). The minimum number of events per bin is greater than 50 (up to 200) depending on the detector quadrant and the mapped spin phase advance $\phi_{s,\text{map}}^{\nu_s^0}$. The number of events in each bin is assumed to follow a Poisson distribution, therefore $\sigma_N = \sqrt{N}$ is used for the errorbars. A simple least squares fit is performed by a sine function with three free parameters

$$N_{\text{fit}}(\phi_{s,\text{map}}^{\nu_s^0}) = N_{\text{offset}} + N_{\text{amp}} \sin\left(\phi_{s,\text{map}}^{\nu_s^0} + \phi_{s,\text{fit}}^{\nu_s^0}\right) \quad (\text{D.2})$$

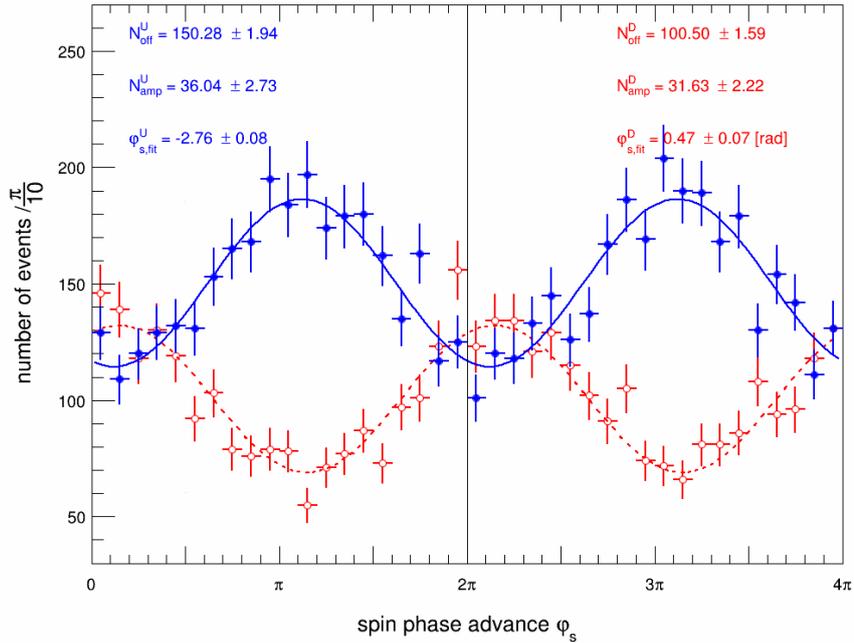


Figure D.1: Counts N_U and N_D after mapping some events recorded during a turn interval of $\Delta n = 10^6$ turns (1.3s) into a spin phase advance interval of 4π . The vertical error bars indicate the statistical uncertainties and horizontal bars represent the corresponding bin width (plot reproduced from [36], also compare [49]).

The quantities N_{amp} and $\phi_{s,\text{fit}}^{\nu_s^0}$ are the amplitude and the phase of the sine. N_{offset} denotes the offset of the function and is equal to the average number of events per bin. Note that this sine function is very similar to the toy model probability density discussed in detail in this work (there for binning in 2π). The total number of events in the analysis is approximately $N_{\text{tot}} \approx N_{\text{offset}} \cdot N_{\text{bins}}$. When norming the average number of events to unity the maximum possible amplitude is again one, since the number of events per bin cannot be negative.

Note that the spin phase advance is mapped properly into the 4π interval only when the assumed spin tune matches the true one $\nu_s^0 \approx \nu_s$. If the assumed spin tune is far off only random fluctuation do occur, but the fit will still find some small amplitude (due to the fitting bias), if the phase is not fixed. The event rates of the two detectors are shifted by $\approx \pi$, since the rate of the up detector quadrants becomes maximal whilst the down detector quadrants reaches the minimum. All the absolute detection rates are depending on the individual detector properties and the beam luminosity. Also the corresponding oscillation amplitude N_{amp} is expected to scale with the total number of detected events and thus depends on the detector geometry and acceptance.

D.1. Asymmetries from Count Sums and Differences

To cancel out the systematic effects, asymmetries are formed by using both detectors. Finally two sinusoidal waveforms oscillating around zero (differences) and two constant

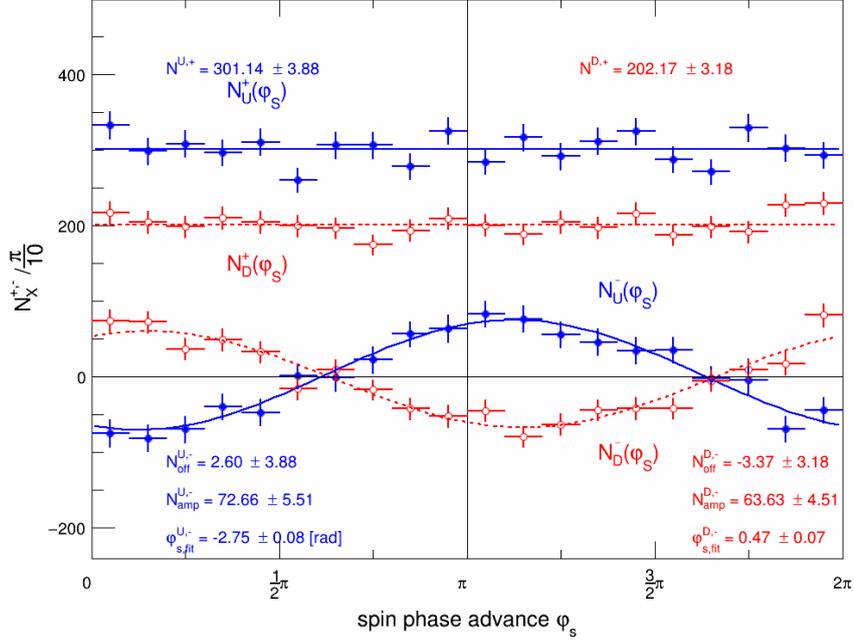


Figure D.2: Count sums $N_{U,D}^+(\phi_s)$ and differences $N_{U,D}^-(\phi_s)$ calculated according to Equation D.3 with $\phi_s \in [0, 2\pi)$ using the counts $N_U(\phi_s)$ and $N_D(\phi_s)$ from both detectors as shown in Figure D.1. Again the vertical error bars indicate the statistical uncertainties, the horizontal bars represent the bin width (plot reproduced from [36] and [49]).

function (sums) are obtained by defining four new event counts for the two quadrants ($X = U$ or D):

$$N_X^\pm(\phi_{s,\text{map}}^0) = \begin{cases} N_X(\phi_{s,\text{map}}^0) \pm N_X(\phi_{s,\text{map}}^0 + 3\pi) & \text{for } 0 \leq \phi_s \leq \pi \\ N_X(\phi_{s,\text{map}}^0) \pm N_X(\phi_{s,\text{map}}^0 + \pi) & \text{for } \pi \leq \phi_s \leq 2\pi. \end{cases} \quad (\text{D.3})$$

The resulting sums ($N_{U,D}^+(\phi_{s,\text{map}}^0)$) and differences ($N_{U,D}^-(\phi_{s,\text{map}}^0)$) of the counts using Equation D.3 are depicted in Figure D.2. The asymmetry

$$\epsilon(\phi_{s,\text{map}}^0) = \frac{N_U^-(\phi_s) - N_D^-(\phi_s)}{N_U^+(\phi_s) + N_D^+(\phi_s)} = \epsilon_{\text{UD}} \cdot \cos(\phi_{s,\text{map}}^0 - \phi) \quad (\text{D.4})$$

now makes use of both these sine-like signals and norms them by the total number of events in the turn interval. The asymmetry is defined in the range $\phi_{s,\text{map}}^0 \in [0, 2\pi)$, where again the maximal possible amplitude ϵ_{UD} is one. The amplitude is proportional to the horizontal vector polarization p_ξ , which can be determined if the cross section $\overline{\sigma}_{0X}$ and the analyzing power are known for both detectors. If one calculates $\frac{1}{2\pi}(1 + \epsilon(\phi_{s,\text{map}}^0))$ i.e. scales values and uncertainties accordingly, a probability distribution fully identical to the one discussed in this master thesis is obtained. Therefore the results from the analysis of the toy model sine function are fully transferable to the spin tune analysis done by the mapping method.

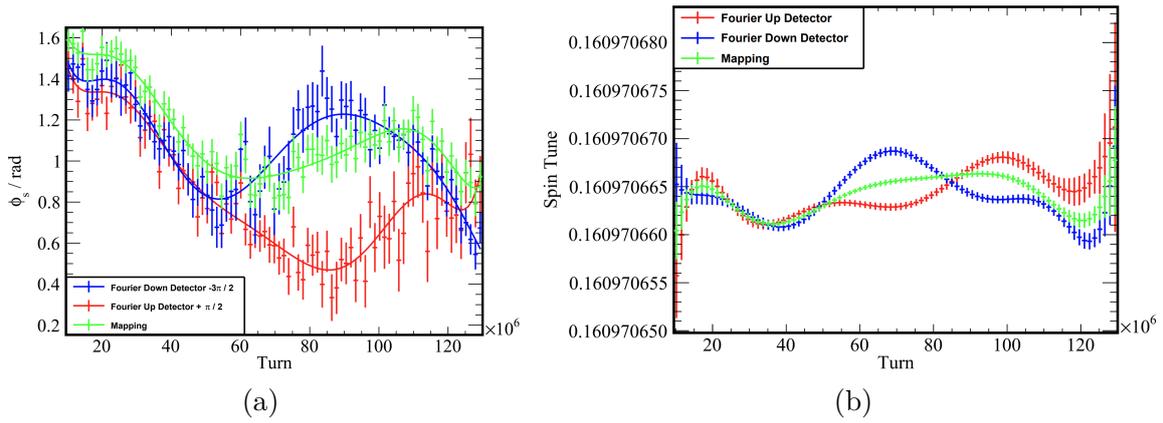


Figure D.3: Spin tune phase and spin tune over the number of turns for COSY Precursor 1 Data. A precision of 10^{-9} is reached, and the typical difference between different analysis methods is shown. The fitting is done using a polynomial of 8th order. This plots are taken from [58].

D.2. Precise Spin Tune from Polynomial Fits

Figure D.3 shows a precise spin tune determination from Precursor data. Relevant for this work is the fact, that the small deviation in the spin tune are not effecting the amplitude estimation procedure. The typical precision obtained when searching for the correct spin tune is more than one order of magnitude better than what is needed for estimating the correct peak amplitude (with corresponding uncertainty). This observation is what was called “quasi fixed spin tune” regarding the amplitude in this work. The observed amplitudes should not be significantly influenced by any statistical effects related to the analysis of the spin tune.

Note that nevertheless the estimated amplitudes between the methods (Up Detector, Down Detector and Mapping) can be and are significantly different, if the beam is e.g. influenced by the Wien Filter Fields, or generally if the symmetry between the detectors is broken. Such effects do cause the differences in phase and spin tune shown in Figure D.3, but even if the spin tune would not change for the analysis, the amplitudes are changed (almost) in the same way as in the detailed analysis. Combining the individual detectors using a (weighted) average can typically reproduce the mapped signal amplitude.

References

- [1] Macro S. Sozzi. *Discrete symmetries and CP violation*. Oxford Graduate Texts. Oxford Univ. Press, 2008. URL: <https://cds.cern.ch/record/1087897>.
- [2] J. H. Christenson et al. “Evidence for the 2π Decay of the K_2^0 Meson”. In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.138>.
- [3] G.C. Branco, L. Lavoura, and J.P. Silva. *CP Violation*. International series of monographs on physics. Clarendon Press, 1999. ISBN: 9780198503996. URL: <https://books.google.de/books?id=5PvpnQEACAAJ>.
- [4] T. Gershon and V. V. Gligorov. “CP violation in the B system”. In: *Reports on Progress in Physics* 80.4 (Feb. 2017), p. 046201. DOI: 10.1088/1361-6633/aa5514. URL: <https://dx.doi.org/10.1088/1361-6633/aa5514>.
- [5] R. Aaij et al. “First Observation of CP Violation in the Decays of B_s^0 Mesons”. In: *Phys. Rev. Lett.* 110 (22 May 2013), p. 221601. DOI: 10.1103/PhysRevLett.110.221601. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.110.221601>.
- [6] R. Aaij et al. “Observation of CP Violation in Charm Decays”. In: *Phys. Rev. Lett.* 122 (21 May 2019), p. 211803. DOI: 10.1103/PhysRevLett.122.211803. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.122.211803>.
- [7] Shu Luo and Zhi-zhong Xing. *First determination of the Jarlskog invariant of CP violation from the moduli of the CKM matrix elements*. 2023. arXiv: 2309.07656 [hep-ph]. URL: <https://arxiv.org/abs/2309.07656>.
- [8] Ivan Esteban et al. “The fate of hints: updated global analysis of three-flavor neutrino oscillations”. In: *Journal of High Energy Physics* 2020.9 (Sept. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep09(2020)178. URL: [http://dx.doi.org/10.1007/JHEP09\(2020\)178](http://dx.doi.org/10.1007/JHEP09(2020)178).
- [9] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. “Matter and antimatter in the universe”. In: *New Journal of Physics* 14.9 (Sept. 2012), p. 095012. ISSN: 1367-2630. DOI: 10.1088/1367-2630/14/9/095012. URL: <http://dx.doi.org/10.1088/1367-2630/14/9/095012>.
- [10] *AMS Days at La Palma, La Palma, Canary Islands, Spain*. 2018. URL: https://indico.mpp.mpg.de/event/5698/contributions/13477/%20attachments/9953/11108/AMS_at_MAGIC_meeting_La_Palma_2018_2.pdf.
- [11] W. Bernreuther. *CP violation and baryogenesis*. 2002. arXiv: hep-ph/0205279 [hep-ph]. URL: <https://arxiv.org/abs/hep-ph/0205279>.
- [12] C. L. Bennett et al. “NINE-YEAR WILKINSON MICROWAVE ANISOTROPY PROBE (WMAP) OBSERVATIONS: FINAL MAPS AND RESULTS”. In: *The Astrophysical Journal Supplement Series* 208.2 (Sept. 2013), p. 20. ISSN: 1538-4365. DOI: 10.1088/0067-0049/208/2/20. URL: <http://dx.doi.org/10.1088/0067-0049/208/2/20>.
- [13] N. Aghanim et al. “Planck2018 results: VI. Cosmological parameters”. In: *Astronomy; Astrophysics* 641 (Sept. 2020), A6. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201833910. URL: <http://dx.doi.org/10.1051/0004-6361/201833910>.

- [14] A. D. Sakharov. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.
- [15] Peter Mohr and Barry Taylor. *Fundamental Physical Constants from NIST*. Nov. 1994. URL: <https://physics.nist.gov/cuu/Constants/index.html>.
- [16] CPEDM Collaboration. *Storage ring to search for electric dipole moments of charged particles: Feasibility study*. en. 2021. DOI: 10.23731/CYRM-2021-003. URL: <https://e-publishing.cern.ch/index.php/CYRM/issue/view/132>.
- [17] JEDI Collaboration. *CP violation from P and T violation*. 2020.
- [18] Andrzej Czarnecki and Bernd Krause. “Neutron Electric Dipole Moment in the Standard Model: Complete Three-Loop Calculation of the Valence Quark Contributions”. In: *Phys. Rev. Lett.* 78 (23 June 1997), pp. 4339–4342. DOI: 10.1103/PhysRevLett.78.4339. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.78.4339>.
- [19] Maxim Pospelov and Adam Ritz. “Electric dipole moments as probes of new physics”. In: *Annals of Physics* 318.1 (2005), pp. 119–169. ISSN: 0003-4916. DOI: 10.1016/j.aop.2005.04.002. URL: <http://dx.doi.org/10.1016/j.aop.2005.04.002>.
- [20] Sebastian A. R. Ellis and Gordon L. Kane. *Theoretical Prediction and Impact of Fundamental Electric Dipole Moments*. 2014. arXiv: 1405.7719 [hep-ph].
- [21] Feng-Kun Guo and Ulf-G. Meißner. “Baryon electric dipole moments from strong CP violation”. In: *Journal of High Energy Physics* 2012.12 (Dec. 2012). ISSN: 1029-8479. DOI: 10.1007/jhep12(2012)097. URL: [http://dx.doi.org/10.1007/JHEP12\(2012\)097](http://dx.doi.org/10.1007/JHEP12(2012)097).
- [22] V. Bargmann, Louis Michel, and V. L. Telegdi. “Precession of the Polarization of Particles Moving in a Homogeneous Electromagnetic Field”. In: *Phys. Rev. Lett.* 2 (10 May 1959), pp. 435–436. DOI: 10.1103/PhysRevLett.2.435. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.2.435>.
- [23] Takeshi Fukuyama and Alexander J. Silenko. “Derivation of Generalized Thomas-Bargmann-Michel-Telegdi Equation for a Particle with Electric Dipole Moment”. In: *International Journal of Modern Physics A* 28.29 (Nov. 2013), p. 1350147. ISSN: 1793-656X. DOI: 10.1142/s0217751x13501479. URL: <http://dx.doi.org/10.1142/S0217751X13501479>.
- [24] William M. Morse, Yuri F. Orlov, and Yannis K. Semertzidis. “rf Wien filter in an electric dipole moment storage ring: The “partially frozen spin” effect”. In: *Phys. Rev. ST Accel. Beams* 16 (11 Nov. 2013), p. 114001. DOI: 10.1103/PhysRevSTAB.16.114001. URL: <https://link.aps.org/doi/10.1103/PhysRevSTAB.16.114001>.
- [25] R. Weidmann et al. “The polarized ion source for COSY”. In: *Review of Scientific Instruments* 67.3 (Mar. 1996), pp. 1357–1358. ISSN: 0034-6748. DOI: 10.1063/1.1146665. eprint: https://pubs.aip.org/aip/rsi/article-pdf/67/3/1357/19299579/1357_1_online.pdf. URL: <https://doi.org/10.1063/1.1146665>.
- [26] D. Chiladze et al. “Determination of deuteron beam polarizations at COSY”. In: *Physical Review Special Topics - Accelerators and Beams* 9.5 (May 2006). ISSN: 1098-4402. DOI: 10.1103/physrevstab.9.050101. URL: <http://dx.doi.org/10.1103/PhysRevSTAB.9.050101>.

- [27] H. Juergen Stein et al. “Current Status of the COSY Electron Cooler (Jülich, Germany)”. In: *Atomic Energy* 94 (2003), pp. 24–26. URL: <https://api.semanticscholar.org/CorpusID:92241850>.
- [28] G. Guidoboni et al. “How to Reach a Thousand-Second in-Plane Polarization Lifetime with 0.97–GeV/c Deuterons in a Storage Ring”. In: *Phys. Rev. Lett.* 117 (5 July 2016), p. 054801. DOI: 10.1103/PhysRevLett.117.054801. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.117.054801>.
- [29] JEDI Collaboration. *JEDI homepage*. visited on 02/09/2024. URL: <https://collaborations.fz-juelich.de/ikp/jedi/>.
- [30] D Prasuhn et al. “Electron and stochastic cooling at COSY”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 441.1 (2000), pp. 167–174. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(99\)01128-6](https://doi.org/10.1016/S0168-9002(99)01128-6). URL: <https://www.sciencedirect.com/science/article/pii/S0168900299011286>.
- [31] Fabian Müller. *Polarimeter Development for Electric Dipole Moment Measurements in Storage Rings*. Nov. 2019. URL: https://collaborations.fz-juelich.de/ikp/jedi/public_files/theses/thesis_fabian_mueller_final_version.pdf.
- [32] Wasa-at-Cosy Collaboration. *WASA Detector*. visited on 02/09/2024. URL: <http://collaborations.fz-juelich.de/ikp/wasa/>.
- [33] O. Javakhishvili et al. “Development of a multi-channel power supply for silicon photomultipliers reading out inorganic scintillators”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 977 (2020), p. 164337. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2020.164337>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900220307348>.
- [34] F. Müller et al. “A new beam polarimeter at COSY to search for electric dipole moments of charged particles”. In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12005. DOI: 10.1088/1748-0221/15/12/P12005. URL: <https://dx.doi.org/10.1088/1748-0221/15/12/P12005>.
- [35] Achim Andres. *The First Direct Measurement of the Deuteron Electric Dipole Moment at the Cooler Synchrotron COSY*. 2024. URL: https://collaborations.fz-juelich.de/ikp/jedi/public_files/theses/.
- [36] Dennis Eversmann. *High Precision Spin Tune Determination at the Cooler Synchrotron in Jülich*. 2018. URL: https://collaborations.fz-juelich.de/ikp/jedi/public_files/theses/ThesisEversmann_v1.pdf.
- [37] Y. Filmus. *Two Proofs of the Central Limit Theorem*. 2010. URL: <http://www.cs.toronto.edu/~yuvalf/CLT.pdf>.
- [38] Marco Taboga. *Mean squared error of an estimator*. Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. 2021. URL: <https://www.statlect.com/glossary/mean-squared-error>.
- [39] B. Roy Frieden. *Science from Fisher Information: A Unification*. Cambridge University Press, 2004.

- [40] Peter Händel. “Amplitude estimation using IEEE-STD-1057 three-parameter sine wave fit: Statistical distribution, bias and variance”. In: *Measurement* 43.6 (2010), pp. 766–770. ISSN: 0263-2241. DOI: <https://doi.org/10.1016/j.measurement.2010.02.007>. URL: <https://www.sciencedirect.com/science/article/pii/S0263224110000382>.
- [41] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai. “Some Classes of Global Cramer-Rao Bounds”. In: *The Annals of Statistics* 15.4 (1987), pp. 1421–1438. DOI: 10.1214/aos/1176350602. URL: <https://doi.org/10.1214/aos/1176350602>.
- [42] Erich L. Lehmann and George Casella. *Theory of Point Estimation*. Second. New York, NY, USA: Springer-Verlag, 1998.
- [43] L. Le Cam. “Maximum Likelihood: An Introduction”. In: *International Statistical Review / Revue Internationale de Statistique* 58.2 (1990), pp. 153–171. ISSN: 03067734, 17515823. URL: <http://www.jstor.org/stable/1403464> (visited on 09/26/2024).
- [44] Johann Pfanzagl. *Parametric Statistical Theory*. Berlin, New York: De Gruyter, 1994. ISBN: 9783110889765. DOI: doi : 10 . 1515 / 9783110889765. URL: <https://doi.org/10.1515/9783110889765>.
- [45] K.V. Mardia, H.R. Southworth, and C.C. Taylor. “On bias in maximum likelihood estimators”. In: *Journal of Statistical Planning and Inference* 76.1 (1999), pp. 31–39. ISSN: 0378-3758. DOI: [https://doi.org/10.1016/S0378-3758\(98\)00176-1](https://doi.org/10.1016/S0378-3758(98)00176-1). URL: <https://www.sciencedirect.com/science/article/pii/S0378375898001761>.
- [46] Gary J. Feldman and Robert D. Cousins. “Unified approach to the classical statistical analysis of small signals”. In: *Physical Review D* 57.7 (1998), pp. 3873–3889. ISSN: 1089-4918. DOI: 10.1103/physrevd.57.3873. URL: <http://dx.doi.org/10.1103/PhysRevD.57.3873>.
- [47] Byron P. Roe and Michael B. Woodroffe. “Improved probability method for estimating signal in the presence of background”. In: *Phys. Rev. D* 60 (5 Aug. 1999), p. 053009. DOI: 10.1103/PhysRevD.60.053009. URL: <https://link.aps.org/doi/10.1103/PhysRevD.60.053009>.
- [48] Amnon Harel. “Statistical methods in CMS searches”. In: *PHYSTAT 2011*. Geneva: CERN, 2011, pp. 88–93. DOI: 10.5170/CERN-2011-006.88.
- [49] D. Eversmann et al. “New Method for a Continuous Determination of the Spin Tune in Storage Rings and Implications for Precision Experiments”. In: *Phys. Rev. Lett.* 115 (9 Aug. 2015), p. 094801. DOI: 10.1103/PhysRevLett.115.094801. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.115.094801>.
- [50] C++ Numerics library. *Element-wise arc tangent choosing the quadrant correctly. Programming language C*. visited on 02/09/2024. URL: <http://en.cppreference.com/mwiki/index.php?title=cpp/numeric/math/atan2&oldid=79630>.
- [51] Numpy Mathematical functions. *Element-wise arc tangent choosing the quadrant correctly. ISO/IEC standard 9899:1999*. visited on 02/09/2024. URL: <https://numpy.org/doc/stable/reference/generated/numpy.arctan2.html>.

- [52] S. Karanth et al. “First Search for Axionlike Particles in a Storage Ring Using a Polarized Deuteron Beam”. In: *Phys. Rev. X* 13 (3 July 2023), p. 031004. DOI: 10.1103/PhysRevX.13.031004. URL: <https://link.aps.org/doi/10.1103/PhysRevX.13.031004>.
- [53] F. Correa Alegria. “Bias of amplitude estimation using three-parameter sine fitting in the presence of additive noise”. In: *Measurement* 42.5 (2009), pp. 748–756. ISSN: 0263-2241. DOI: <https://doi.org/10.1016/j.measurement.2008.12.006>. URL: <https://www.sciencedirect.com/science/article/pii/S0263224109000037>.
- [54] D. Eversmann, J. Pretz, and M. Rosenthal. “Amplitude estimation of a sine function based on confidence intervals and Bayes’ theorem”. In: *Journal of Instrumentation* 11.05 (May 2016), P05003–P05003. ISSN: 1748-0221. DOI: 10.1088/1748-0221/11/05/p05003. URL: <http://dx.doi.org/10.1088/1748-0221/11/05/P05003>.
- [55] Z. Bagdasarian et al. “Measuring the polarization of a rapidly precessing deuteron beam”. In: *Phys. Rev. ST Accel. Beams* 17 (5 May 2014), p. 052803. DOI: 10.1103/PhysRevSTAB.17.052803. URL: <https://link.aps.org/doi/10.1103/PhysRevSTAB.17.052803>.
- [56] S. Plaszczyński et al. “A novel estimator of the polarization amplitude from normally distributed Stokes parameters”. In: *Monthly Notices of the Royal Astronomical Society* 439.4 (Feb. 2014), pp. 4048–4056. ISSN: 0035-8711. DOI: 10.1093/mnras/stu270. URL: <http://dx.doi.org/10.1093/mnras/stu270>.
- [57] Steven W. Smith. *The scientist and engineer’s guide to digital signal processing*. USA: California Technical Publishing, 1997. ISBN: 0966017633.
- [58] Achim Andres. *Polarisation Measurements for Storage Ring Electric Dipole Moment Investigations*. May 2020. URL: https://collaborations.fz-juelich.de/ikp/jedi/public_files/theses/master_AA_JEDI.pdf.
- [59] Jörg Pretz. “Comparison of methods to extract an asymmetry parameter from data”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 659.1 (2011), pp. 456–461. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2011.08.036>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900211016573>.