DISS. ETH NO. 25124

ACTIVE MAGNETIC SHIELDING AND AXION-DARK-MATTER SEARCH

a thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZÜRICH

(Dr. sc. ETH Zürich)

presented by

MICHAŁ RAWLIK

magister of Jagiellonian University in Kraków

born on 18.06.1990

citizen of Poland

accepted on the recommendation of

PROF. KLAUS KIRCH

PROF. PHILIP HARRIS

PROF. HANS-ARNO SYNAL

2018

To my family.

SUMMARY

Measurements of the electric dipole moment of the neutron (nEDM) provide insight into unanswered questions of contemporary physics, such as "Why is there more matter than antimatter in the Universe?" and "Why does the strong interaction in the Standard Model appear to be fine-tuned?". The most sensitive nEDM measurement to date was performed in the Paul Scherrer Institute, Villigen, Switzerland. This work is concerned with two aspects of this experiment. One is active magnetic field shielding for its successor—n2EDM. The other is a search for a signature of dark matter in its measurements.

Stability of the magnetic field is crucial in measurement of the nEDM. The experiment at PSI employed an active magnetic shield, which stabilised the field by detecting variations in it and counteracting them with a set of large coils around the apparatus. In the design of an active shield for n2EDM tight spatial constraints excluded the use of known coil geometries, and methods of design thereof. In this work a new method of magnetic field coil design is presented. It achieves an unprecedentedly large ratio of the fiducial volume to the size of the coil. Spatial constraints are easily incorporated, as the coils are designed on a predefined grid, which may be shared between multiple coils. A small-scale active magnetic shield demonstrated the method's designs. Field maps showed the homogeneity to be on a 2 % level. Actively shielding homogeneous fields increased the stability 2–30 times down to 0.3 nT over times from seconds to hours. Based on those developments a design for an active shield for the n2EDM experiment is proposed. The n2EDM shield can perform better when tailored to the particular magnetic environment. To that purpose a device to map magnetic fields in a large volume and in a short time was developed: a mobile tower equipped with magnetic sensors, whose position and orientation was continuously measured with string potentiometers. A small-scale prototype was tested at Laboratoire de Physique Subatomique & Cosmologie (LPSC), Grenoble, France, demonstrating a measurement reproducibility of 138 nT for the field and 3.8 nT/cm for its gradient. A full-scale 8 m high tower was used to map the experimental site of n2EDM. The maps can be used to refine the proposed design of the n2EDM active shield.

Dark matter is an estimated 84 % of the matter content of the Universe, based on astrophysical observations of its gravitational interaction with the visible matter. Yet, the dark-matter constituents remain elusive for particle physics. An axion is a prominent candidate for a dark-matter particle, which is expected to couple to ordinary matter in other ways, besides gravity. In particular, it may couple to the neutrons in the PSI nEDM experiment via a scalar axion-gluon or a derivative axion-nucleon coupling, causing harmonic oscillations in their spin-precession frequency. In this work such oscillations were sought. The search covered periods between 1.5 year and 300 s, which corresponds to axion masses $\sim 10^{-22}$ eV – 2×10^{-17} eV. The null result put the first laboratory constraints on the axion coupling to gluons and improved those on the axion-nucleon coupling by a factor of 40.

ZUSAMMENFASSUNG

Messungen des elektrischen Dipolmoments des Neutrons (nEDM) bringen Einsicht in offene Fragen der heutigen Physik wie zum Beispiel: „Warum gibt es mehr Materie als Anti-Materie im Universum?“ und „Warum scheint die starke Wechselwirkung im Standardmodel eine unnatürliche Feinabstimmung zu haben?“ Die Messung des nEDM mit der bislang grössten Empfindlichkeit wurde am Paul Scherrer Institut (PSI) in Villigen, Schweiz, durchgeführt. Die vorliegende Arbeit behandelt zwei Aspekte dieses Experimentes. Einer ist die Entwicklung der aktiven Magnetfeldabschirmung für das Nachfolgeexperiment n2EDM. Der andere ist die Suche nach Hinweisen auf Dunkle Materie in den Messdaten.

Die Stabilität des magnetischen Feldes ist entscheidend bei der Messung des nEDM. Das Experiment am PSI verwendet eine aktive magnetische Schirmung, die das Feld stabilisiert, indem es Veränderungen detektiert und diesen mithilfe grosser Spulen um den Apparat entgegenwirkt. Bei der Konstruktion der aktiven Abschirmung für n2EDM verhinderten räumliche Einschränkungen die Verwendung bekannter Spulengeometrieen und üblicher Methoden solche zu erzeugen. In dieser Arbeit wird eine neue Methode zur Erzeugung einer Spulengeometrie vorgestellt. Diese erreicht ein beispielloses Grössenverhältnis zwischen den Spulen und dem geschirmten Volumen. Platzeinschränkungen können sehr einfach berücksichtigt werden und die Spulen werden auf einem vordefinierten Gitter angeordnet. Mehrere Spulen können dabei das gleiche Gitter verwenden. Ein verkleinertes Model der Magnetfeldabschirmung wurde für Demonstrationszwecke gebaut. Feldkartografierungen zeigten eine Homogenität auf dem 2 % Niveau. Das aktive Schirmen von homogenen Feldern verbessert deren Stabilität um das 2- bis 30-fache auf 0.3 nT über Zeitskalen von Sekunden zu Stunden. Basierend auf diesen Erkenntnissen wird ein Design für die aktive Magnetfeldabschirmung für das n2EDM-Experiment vorgeschlagen. Die Schirmung für n2EDM kann noch besser werden, wenn sie auf die dortige magnetische Umgebung angepasst wird. Dafür wurde ein Gerät zum Kartographieren des Magnetfeldes in einem grossen Volumen entwickelt, nämlich ein fahrbarer Turm mit Magnetfeldsensoren, dessen Position und Ausrichtung kontinuierlich mit Seilzugsensoren gemessen wird. Ein kleiner Prototyp wurde am LPSC Grenoble in Frankreich getestet. Eine Reproduzierbarkeit der Messergebnisse von 138 nT für die Feldwerte und von 3.8 nT/cm für die Feldgradienten wurde demonstriert. Der Turm in voller Höhe (8 m) wurde verwendet um das Experiment-Areal für n2EDM zu vermessen. Die Feldkarten können verwendet werden, um das vorgeschlagen Design für die aktive Schirmung von n2EDM zu optimieren.

Astrophysikalische Beobachtungen der sichtbaren Materie des Universums führen zu der Schätzung, dass das Universum zu 84 % aus Dunkler Materie besteht, die per Gravitation, aber nicht elektromagnetisch mit gewöhnlicher Mate-

rie interagiert. Allerdings kann Dunkle Materie bislang nicht von der Teilchenphysik erfasst werden. Das Axion ist ein besonderer Kandidat für ein Dunkles Materie-Teilchen, da es neben der Gravitation auch noch anderweitig mit der sichtbaren Materie interagiert. Insbesondere könnte es an Neutronen mit einer skalaren Axion-Gluonen oder vektorartigen Axion-Nukleonen Wechselwirkung koppeln und so im nEDM-Experiment eine harmonische Oszillation derer Larmorpräzissionsfrequenzen erzeugen. In dieser Arbeit wurde nach solchen Oszillationen gesucht. Die Suche deckte Perioden von 1.5 Jahren bis 300 s ab, was einer Masse des Axions von $\sim 10^{-22}$ eV bis 2×10^{-17} eV entspricht. Das Nullergebnis stellt die erste experimentelle Obergrenze der Kopplung von Axionen zu Gluonen dar, und verbessert die Grenze der Kopplung von Axion zu Nukleonen auf das 40-fache.

CONTENTS

Part I

NEUTRON ELECTRIC DIPOLE MOMENT

The existence of an electric dipole moment of the neutron breaks fundamental symmetries of nature and, thereby, gives insight into big problems of contemporary physics. For example the dominance of matter over antimatter in the Universe. Its discovery has been constantly pursued since the 1950s, the newest effort being conducted in the Paul Scherrer Institute in Switzerland. In the first part we explain how the measurement of the neutron electric dipole moment motivated the original work of the author, described in the two further parts of this thesis.

1 AN EXCITING OBSERVABLE

The chief forms of beauty are order and symmetry and definiteness, which the mathematical sciences demonstrate in a special degree.

— Aristotle

1.1 SYMMETRIES

The essence of classical physics is the three symmetries: with respect to spatial translation (homogeneity of space), with respect to rotation (isotropy of space) and with respect to translation in time (homogeneity of time). As Noether showed, they correspond to the conservation laws of momentum, angular momentum and energy, respectively [1]. By saying a symmetry is conserved, we understand that no physical system becomes any different by *only* having been moved in space, rotated, or looked at later.

These are continuous symmetries. No less fundamental is the triad of symmetries with respect to discrete transformations: P, parity transformation, mirroring the spatial dimensions; T, time reversal, flipping the arrow of time; and C, charge conjugation, flipping *all* charges of particles. One would expect a beautiful universe to work exactly the same in a mirror, or with matter swapped with antimatter. This expectation is extended to combined symmetries, like CP, which flips the charges *and* mirrors the space. The combined CPT-symmetry has a special place in physics, as it is obeyed by any Lorentz invariant local quantum field theory with a Hermitian Hamiltonian [2].

Andrei Sakharov, pacifist and human-rights activist, was awarded the 1975 Nobel Peace Price. He was called "a spokesman for the conscience of mankind".

Yet, in a perfectly symmetric universe after the Big Bang an equal amount of matter and antimatter would be created, and they would perfectly annihilate. Luckily, this did not happen, but the question *why* remains one of the big unsolved problems in physics. In 1967 Andrei Sakharov proposed that the mechanism leading to the leftover matter is based on a violation of the CP-symmetry [3].

Already a decade before, physicists had observed for the first time violations of one of the discrete symmetries in their laboratories. In 1956 Chien-Shiung Wu shocked the physics world with her discovery of P violation in beta decay of cobalt-60 [4]. Shortly thereafter, in 1964, the discovery of CP violation in $K_L^0 \to \pi^+\pi^-$ decay left the physics world stunned again [5]. Now, both are explained in the scope of the Standard Model of particle physics. In the model the CP-violation appears in the weak sector as a complex phase in the Cabbibo–Kobayashi–Maskawa matrix, the matrix that mixes the mass and weak interaction eigenstates of quarks [6].

CP violation also appears in the strong sector of the Standard Model's quantum chromodynamics, as an additional term parametrised by θ_{QCD}. Yet, no CP

violation in the strong sector has been observed so far. θ_{QCD}, expected to be $\mathcal{O}(1)$, is limited to be smaller than 10^{-10} [7]. A puzzle in itself, this another unanswered problem in physics is referred to as the strong CP problem.

Altogether, the degree of CP-symmetry violation in the Standard Model is not enough to explain the Universe we observe. The observed matter–antimatter asymmetry is around 10^{-10}, while only 10^{-18} can be attributed to the model [8]. Despite the tremendous success of the Standard Model, we know it is not the full picture. We *know* that the CP-symmetry is violated somewhere still, but do not know where or how [9].

Many theories beyond the Standard Model have been proposed that offer to solve this problem, among others. A very popular idea is to introduce additional particles around the scale of weak interactions (several-hundred GeV), as, for example, supersymmetry does [10]. Those theories typically introduce additional mechanisms of CP-violation, which provide an opportunity to test their prediction [11].

1.2 ELECTRIC DIPOLE MOMENTS

Electric dipole moments (EDMs) are excellent probes for physics beyond the Standard Model [9]. An EDM is a T-violating, and assuming the CPT conservation, also CP-violating observable. In a non-relativistic case of a particle in an electric $\mathbf{E}$ and magnetic field $\mathbf{B}$, with the magnetic and electric moments μ and d, the Hamiltonian is

$$H = -\mu\,\mathbf{B}\cdot\frac{\mathbf{S}}{S} - d\,\mathbf{E}\cdot\frac{\mathbf{S}}{S}\,. \tag{1}$$

For a spin $S = \frac{1}{2}$ particle:

$$H = -2\,(\mu\,\mathbf{B} + d\,\mathbf{E})\cdot\mathbf{S}\,. \tag{2}$$

Vectors, like $\mathbf{E}$, are P-odd, while pseudovectors, $\mathbf{B}$ and $\mathbf{S}$, are P-even. The Hamiltonian as whole is P-violating:

$$H_P = -2\,(\mu\,\mathbf{B} - d\,\mathbf{E})\cdot\mathbf{S} \neq H\,. \tag{3}$$

Magnetic fields are produced either by magnetic moments, which flip under T together with spins, or by a movement of charges, which also is reversed under T.

Under time symmetry the spin is reversed and so is the magnetic field, rendering the Hamiltonian T-violating:

$$H_T = H_{CP} = +2\,(-\mu\,\mathbf{B} + d\,\mathbf{E})\cdot\mathbf{S} = -2\,(\mu\,\mathbf{B} - d\,\mathbf{E})\cdot\mathbf{S} \neq H\,. \tag{4}$$

Note how in both cases setting $d = 0$ restores the symmetry.

EDMs are measured mainly in three kinds of systems. Firstly, beams of cold paramagnetic molecules, like ThO and YbF, provide a great sensitivity to the EDM of the electron [12, 13]. Secondly, measurements of vapours of diamagnetic atoms, notably ^{199}Hg [14], are the most sensitive EDM measurements. Finally, the electric dipole moment of a free neutron, nEDM, is the topic of this chapter. While the Standard Model predicts the nEDM to be below $10^{-30}\,e\,\mathrm{cm}$, its extensions foresee it to be larger [11]. Measurements of the nEDM provide an opportunity to validate, or falsify, those theoretical models. Also, the neutron EDM, provides (together with the EDM of ^{199}Hg [14]) the most sensitive handle on measurement of the θ_{QCD}, parametrising the strong CP problem [7].

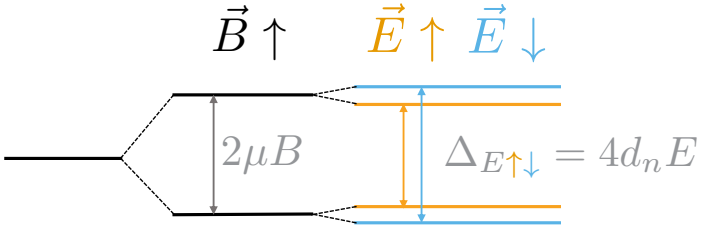

Figure 1: The energy states of a neutron in a combination of a magnetic and an electric field. The Hamiltonian is $H = -2\left(\mu_\mathrm{n}\,\mathbf{B} + d_\mathrm{n}\,\mathbf{E}\right)\cdot\mathbf{S}$. The first term causes the first splitting of $2\mu_\mathrm{n}B$. The second term increases or decreases the splitting by $2d_\mathrm{n}E$, depending on the direction of the electric field relative to the magnetic one and the sign of d_n. In the figure the situation for a positive d_n is depicted.

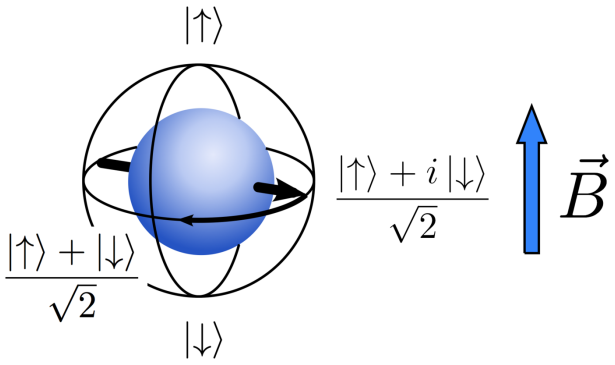

Figure 2: A spin 1/2 particle in a magnetic field on a Bloch sphere. The poles correspond to the pure spin-up and spin-down states. On the equator lie the states with an equal contribution of the two and the longitude corresponding to the quantum phase.

1.3 MEASUREMENTS OF THE NEDM

Already in 1950, before Wu's discovery of P-violation in the weak sector, Smith, Purcell and Ramsey proposed a measurement to test P-violation in the strong sector using the nEDM as the probe [15]. Their result, published in 1957, was consistent with zero $d_\mathrm{n} = (-0.1 \pm 2.4) \times 10^{-20}\,e\,\mathrm{cm}$ [16].

Figure 1 illustrates the energy of a neutron in a magnetic field B, where it has two energy states separated by $2\mu_\mathrm{n}B$. The apparatus of Smith, Purcell and Ramsey could measure this separation, as a frequency, very precisely. In addition to the magnetic field, there was an electric one, either parallel or antiparallel to it. If there had been an nEDM, the energy separation would have increased in one configuration and decreased in the other. The difference between the energy separations measured in the two field configurations is proportional to the nEDM d_n.

To measure the energy separation between the two spin states they used what is now called the Ramsey method of separated oscillatory fields [17]. In order to explain it, let us first consider a neutron in a magnetic field, as pictured in Fig. 2. The neutron's spin is depicted there on the Bloch sphere, where the

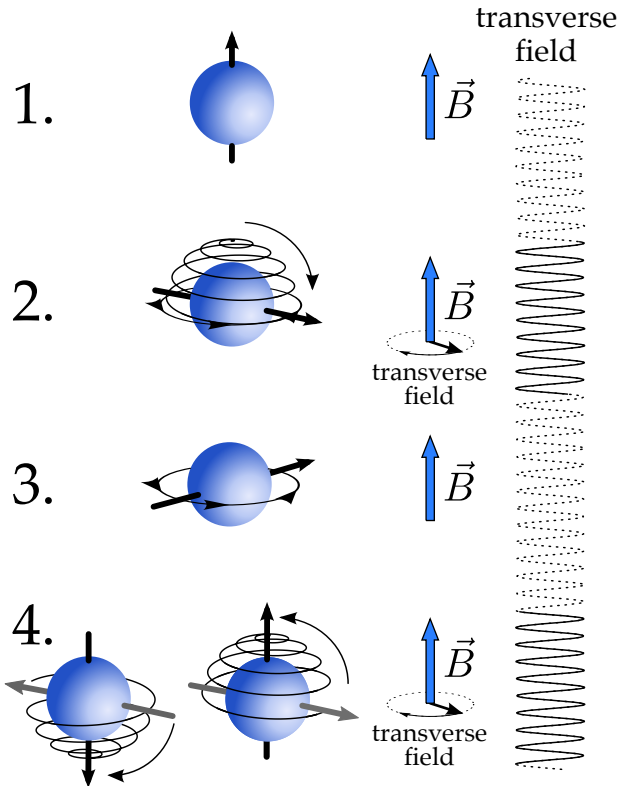

Figure 3: The principle of the Ramsey method, explained with the spin on the Bloch sphere. *1.* A polarised spin ensemble is in a magnetic field. *2.* A pulse of an oscillating transverse field flips the polarisation into the horizontal plane. *3.* The spin is allowed to freely precess in the field. *4.* A second pulse of a transverse field is applied, in phase with the first one. The direction of the polarisation's flip depends on the relative phase between the spin and the transverse field. After a figure by P. Harris [18].

poles correspond to the pure spin-up and spin-down states. On the equator lie the states with equal content of the two, the longitude marking the quantum phase. When the spin state is not vertical, the interaction between the magnetic moment and the magnetic field causes the phase to spin. The frequency of the motion, called the Larmor precession, is proportional to the energy difference between the two states.

In the experiment a nearly monochromatic neutron beam passed through a polariser, a region with an electric field, an analyser, and hit a neutron counter at the end. The whole setup was in a magnetic field. At the beginning and the end of the region with the electric field, there were coils producing an oscillating magnetic field in the direction perpendicular to the main magnetic field. The spin evolution between the two coils is depicted in Fig. 3. When a neutron precessing in a magnetic field feels an additional oscillating field, transverse to the main one and of frequency close to the Larmor one, its spin undergoes a nutation—the precession plane moves along the main field (North or South on the Bloch sphere). The direction is determined by the relative phase between the spin precession and the transverse oscillating field. In Ramsey's experiment the length of the coils was set to flip the neutrons' spins by $\pi/2$, so that after having passed the first coil they were precessing on the Bloch sphere's equator. The direction of the nutation in the second coil, and with it the probability of passing the analyser, depended on the relative phase between the precession and the transverse field. A slight change in the frequency of the generator powering the coils caused a considerable change in the phase difference that built up while the neutrons flew precessing between the coils. Scanning the frequency of the generator and monitoring the counting rate produced a reso-

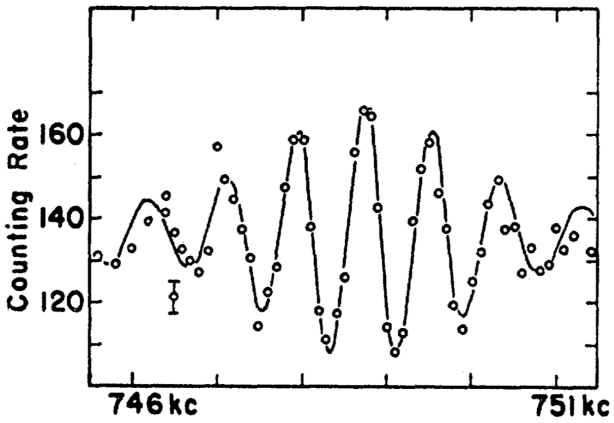

Figure 4: The original resonance curve measured by Ramsey [16], showing the counting rate versus the frequency of the generator powering the spin-flipping coils. The width of the fringes is the inverse duration of the free spin precession. The envelope arises, as at a large detuning the spins were not flipped into the horizontal plane anymore.

nance curve (Fig. 4). The middle of the central fringe is the resonance frequency corresponding to the transition energy between spin-up and spin-down states. Comparing the positions of the resonance with the magnetic and electric fields parallel and antiparallel gave the d_n estimate.

The community uses this technique to measure the nEDM until this day, their ever-more-sensitive efforts summarised in Fig. 5. Until the 1970s the measurements were done using a beam of cold neutrons (marked in green), after which storage experiments greatly improved the time of the free precession and thereby the sensitivity. It was made possible by neutrons with kinetic energies below approximately 300 neV (referred to as ultracold), which are storable in some materials, by undergoing a total internal reflection [34]. In 1980 in the Leningrad Nuclear Physics Institute (LNPI, former USSR) the first measurement was performed with the neutrons being stored in a bottle [25]. The measurements in LNPI continued (orange) and in 1984 were joined by an international effort in the Institute Laue–Langevin in France (black).

The next effort took place in the Paul Scherrer Institute (PSI) in Villigen, Switzerland. The experiment collected data over the years 2015–17 and at the time of writing the result was still being evaluated. In the next chapter the nEDM measurement at PSI is described. It serves as an introduction to the main part of this work, the original work of the author, which focuses on two aspects closely related to all nEDM measurements: stabilisation of magnetic fields and exotic physics that can be explored with these highly sensitive experiments.

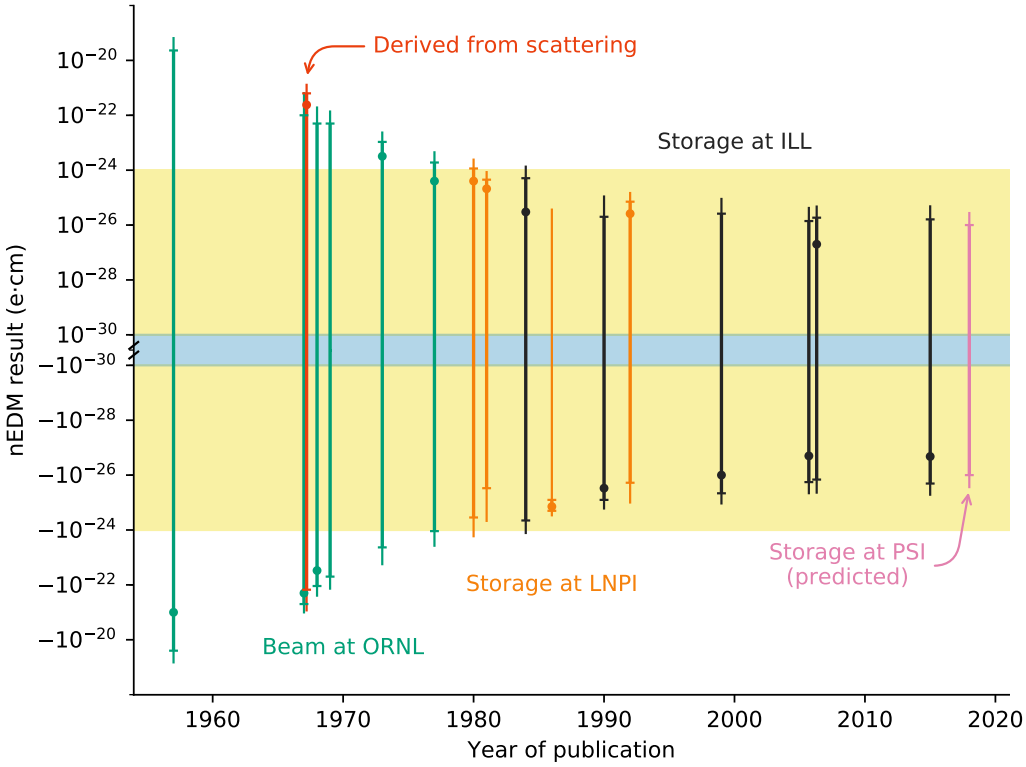

Figure 5: The history of the neutron electric dipole moment measurements. For each published result the vertical line corresponds to the 3σ allowed region, horizontal bars depict the 1σ one. The data markers are located at the central values of the measurements. The vertical scale is a combination of two logarithmic ones—one for each sign of nEDM. The region of the Standard Model's prediction is depicted in blue, the one of the proposed Model's extensions in yellow. The red line indicates a result obtained from a neutron scattering experiment. The sources are, in chronological order: [16, 19–33].

1.4 MAGNETIC STABILITY

The quality of the magnetic field is the main challenge in the measurements of the nEDM. Recall the principle of the measurement, as depicted in Fig. 1. The electric dipole moment d_n is proportional to the difference in the measured energy separations between the field configurations. Let us ask ourselves the following question: How big is a change in the magnetic field that causes a comparable change in the separation of the states?

The current nEDM limit is around $|d_\mathrm{n}| < 10^{-26}\, e\,\mathrm{cm}$ [32]. Taking the electric field to be 132 kV/12 cm, as in the PSI experiment, this corresponds to an energy difference

$$\Delta_{E\uparrow\downarrow} = 4 d_\mathrm{n} E = 4 \times 10^{-26}\, e\,\mathrm{cm}\, \frac{132\,\mathrm{kV}}{12\,\mathrm{cm}} = 4.4 \times 10^{-22}\,\mathrm{eV}\,. \tag{5}$$

With the approximate neutron magnetic dipole moment [35]

$$\mu_\mathrm{n} = -9.7 \times 10^{-27}\,\mathrm{J\,T^{-1}} = -6 \times 10^{-8}\,\mathrm{eV\,T^{-1}}\,, \tag{6}$$

the size of a change in magnetic field corresponding to this energy is

$$\frac{\Delta_{E\uparrow\downarrow}}{2\mu_n} = 3.7 \times 10^{-15}\,\mathrm{T} = 3.7\,\mathrm{fT}\,. \tag{7}$$

This is about the strength of the magnetic field of a car passing several kilometers away. The field needs to be controlled on this level so that the two measurements, with the magnetic and electric fields parallel and antiparallel, can be subtracted from one another without the difference being dominated by the instability of the magnetic field.

In reality the requirement is relaxed by a factor of 10–100, as the final measurement is an average of several thousand measurements.

The nEDM experiments continue to set the world's standard in terms of stabilising and measuring magnetic fields [36–39]. When it comes to stabilisation, a newcomer in the field is an active magnetic shielding. This was first used in the nEDM measurement at PSI, increasing the field stability by a factor of 5–50 [40]. Part II is dedicated to research in this area. A novel method of designing coils is introduced, which makes active stabilisation systems more compact and effective. It is followed by a presentation of a system constructed at ETH Zürich, intended as a small-scale prototype of a next-generation system for the nEDM measurement at PSI—the n2EDM experiment. Finally, a survey of the magnetic field at PSI's experimental area is described, a part of research on the magnetic field compensation there.

1.5 NEW PHYSICS

nEDM experiments are sensitive to, besides the electric dipole moment, other new physics. For example, a search for a short range spin-dependent interaction mediated by an axion, a hypothetical new particle [41]; or testing the Lorentz invariance by looking for variations arising due to the Earth spinning in an non-isotropic Universe [42, 43]; or searching for mirror particles, proposed to restore the global P-symmetry, by detecting neutron to mirror-neutron oscillations with an nEDM apparatus [44].

In this work a dark-matter search with the nEDM experiment at PSI is presented. One of the candidates for dark matter are axions, extremely light particles generalising the idea of promoting the θ_{QCD} parameter (the same as in the strong CP problem) into a field [45]. An axion dark matter could form a coherently, very slowly (as slowly as days) oscillating field. This would induce coherent variations in the measured values of nEDM. A search for such variations is discussed in the Part III.

2

NEDM MEASUREMENT AT PSI

The aim of the nEDM at PSI collaboration is to perform the world's most sensitive nEDM measurement in the Paul Scherrer Institute (PSI) in Villigen, Switzerland, and then continue to improve it. The experiment had been planned in two stages. In the first one, the apparatus used in the ILL experiment was moved to PSI, where it was installed to benefit from a new, highly intense source of ultracold neutrons [46]. The first stage finished in Autumn 2017, after having collected enough statistics to be the world's most sensitive nEDM measurement to date (Fig. 6). As of Spring 2018 the data analysis is still ongoing.

Already in the first stage the apparatus underwent numerous improvements, leaving only few parts of the original. The improvements were part of the research and development plan for the second stage, a newly built apparatus called n2EDM. The new experiment, designed by the personnel experienced with running, modifying and improving the first stage, would have the goal of exploring the $10^{-27}\,e\,\mathrm{cm}$ range.

2.1 THE APPARATUS

This work focuses on the stage-one apparatus. It employed the Ramsey method with ultracold neutrons stored for the time of the spin precession. The neutrons, produced in the PSI source, were guided with a pipe system into a storage chamber, where they underwent the Ramsey procedure. The polarisation was then measured by letting the neutrons fall into a spin-state–sensitive neutron detector. A diagram of the system is shown in Fig. 7.

A single measurement cycle was triggered by a pulse of ultracold neutrons incoming from the PSI source, from which the neutrons were guided in metal-coated glass pipes. First, they were polarised by passing through a five-tesla magnet. Then, a rotary three-way valve, called the switch, directed the neutrons upwards into a vacuum tank, where they filled a 12 cm high cylindrical precession chamber. The chamber was sandwiched between electrodes. The bottom electrode was grounded and a potential of $\pm 132\,\mathrm{kV}$ was applied to the top one, which created an electric field in the chamber. The whole stack was submerged in a $1\,\mu\mathrm{T}$ vertical magnetic field B_0. Once the neutrons filled the cylinder, the entrance was shut and the particles were stored for around 200 s to undergo the Ramsey procedure.

At the nEDM experiment at PSI the neutrons were dropped into a two-armed detector, capable of counting both spin states simultaneously [47].

First, a pulse of oscillating transverse field was applied, with its amplitude and length tuned so that the neutrons' spins flip from the vertical orientation, along the B_0 field, to the horizontal plane (a $\pi/2$ pulse). This set the spins into a Larmor precession. After 180 s a second pulse, in phase with the first one, was applied. Then the chamber's entrance was opened, and the neutrons were allowed to fall through the switch and a spin analyser into the neutron detector.

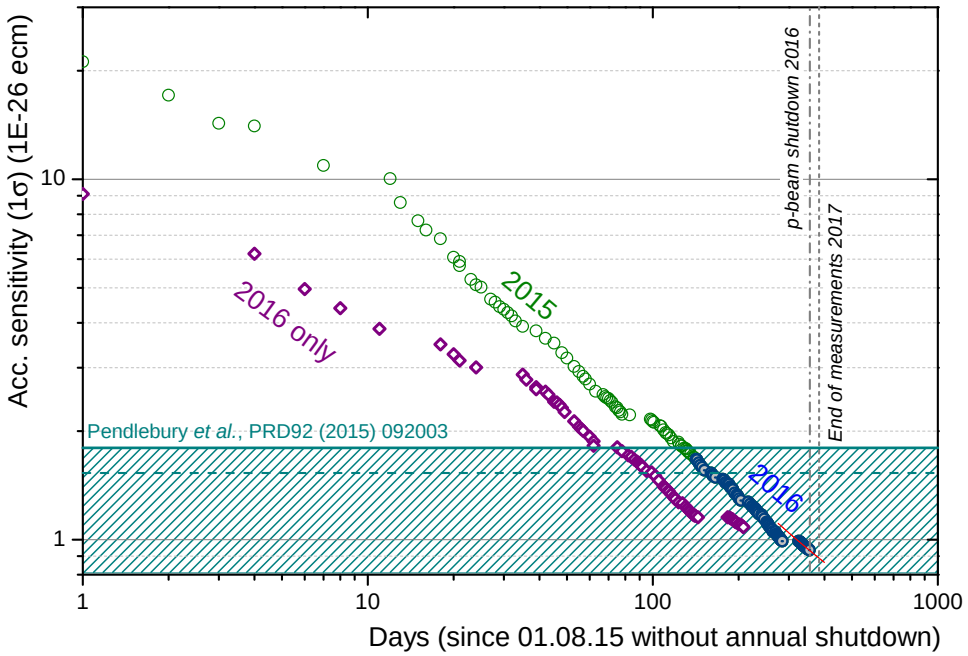

Figure 6: The accumulated statistical sensitivity of the nEDM experiment at PSI. The sensitivity of the data collected in 2015 is shown in green, of 2015 and 2016 in blue, and only 2016 in violet. The dashed region marks the yet unexplored area below the best limit [33]. Courtesy of Dr. Philipp Schmidt-Wellenburg.

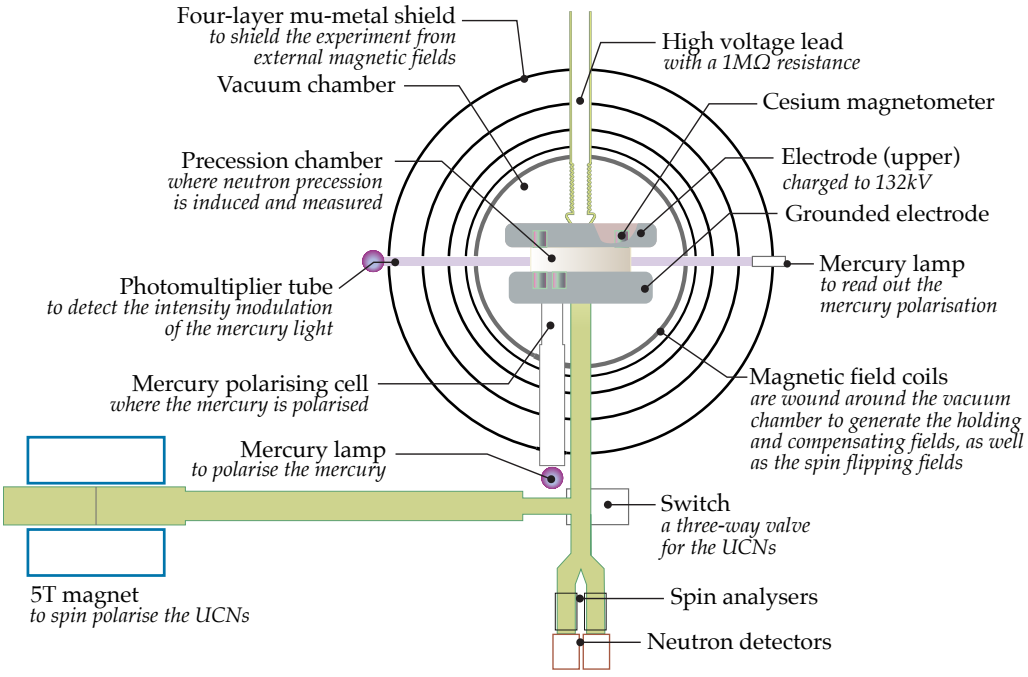

Figure 7: The scheme of the nEDM apparatus at PSI. Courtesy of Dr. Zema Chowdhuri and Dr. Philipp Schmidt-Wellenburg.

The majority of the remaining components either provided a stable magnetic field environment in the precession chamber or measured the field.

2.2 MAGNETIC FIELDS

The most important method of stabilising the magnetic field is passive shielding. The vacuum tank, with the precession chamber in it, was covered by four layers of highly permeable material, called mu-metal. Each layer attenuated the magnetic field changes by a factor of approximately ten. The high magnetic permeability of the mu-metal makes the magnetic field prefer to go inside the metal and around the volume it encloses, than to penetrate into the volume.

The exact attenuation factor of the shield depended on the direction and varied from 1600 to 13300 [48].

Inside the shield, on the vacuum chamber, there were a number of coils wound. Besides the one producing the main magnetic field B_0, there were numerous others, producing fields of various shapes. Those were, on one hand, used to homogenise the field, which reduced the depolarisation rate of the neutrons. On the other hand, they were used to deliberately apply vertical gradients $\partial_z B_0$; operating in different gradients was part of the measurement procedure and is explained later in this chapter.

In an ensemble precessing in an inhomogenous field, members in different places precess at slightly different rates. This leads to depolarisation by the loss of coherence.

The stability of the field inside the shield was in the picotesla range. The remaining variations were measured and corrected for, primarily with a mercury-based magnetometer [48, 49]. Below the precession chamber there was a cell containing a vapour of polarised ^{199}Hg atoms, which was released into the precession chamber once it had been filled with neutrons. There, a dedicated $\pi/2$ pulse of a rotating transverse magnetic field started their coherent spin precession, which was read on-line optically. A mercury discharge lamp shone a circularly polarised light through the precession chamber onto a photomultiplier tube. As the mercury atoms' photon-capture cross-section depended on the phase of the spin precession, the transparency of the chamber to the polarised light oscillated at the mercury's Larmor frequency. The oscillating signal of the photomultiplier was then analysed to estimate its frequency, proportional to the B_0 field, as averaged by the mercury atoms.

In the 1 μT B_0 field the spin-precession frequencies of neutrons and ^{199}Hg atoms are approximately 30 Hz and 8 Hz, respectively. The respective spin-flip pulses affect the other species only very slightly.

The mercury magnetometer may seem an ideal measure to correct the neutron measurement for the drifts of the magnetic field, as the two species filled exactly the same volume (hence it is often called a *comagnetometer*). However, they did not measure the exact same field. The ultracold neutrons were slow enough to be affected by gravity, which shifted their centre of mass by around 2.4 mm relative to the warm, homogeneous mercury vapour [50]. In the presence of a vertical gradient, the two species saw different fields. A detailed discussion of additional systematic effects related to this magnetometer can be found in [50].

The $\pi/2$ pulses for the neutrons were oscillating (linear), the ones for ^{199}Hg atoms were rotating (circular).

A way to measure vertical gradients, as well as other high-order components of the magnetic field, was provided by an array of caesium magnetometers [38]. These sensors were located inside the electrodes—seven in the top one, nine in the bottom one. Each magnetometer was a cell filled with caesium vapour, through which a circularly polarised light, delivered with a light guide, shone at 45° inclination angle. The light simultaneously polarised the atoms and, as

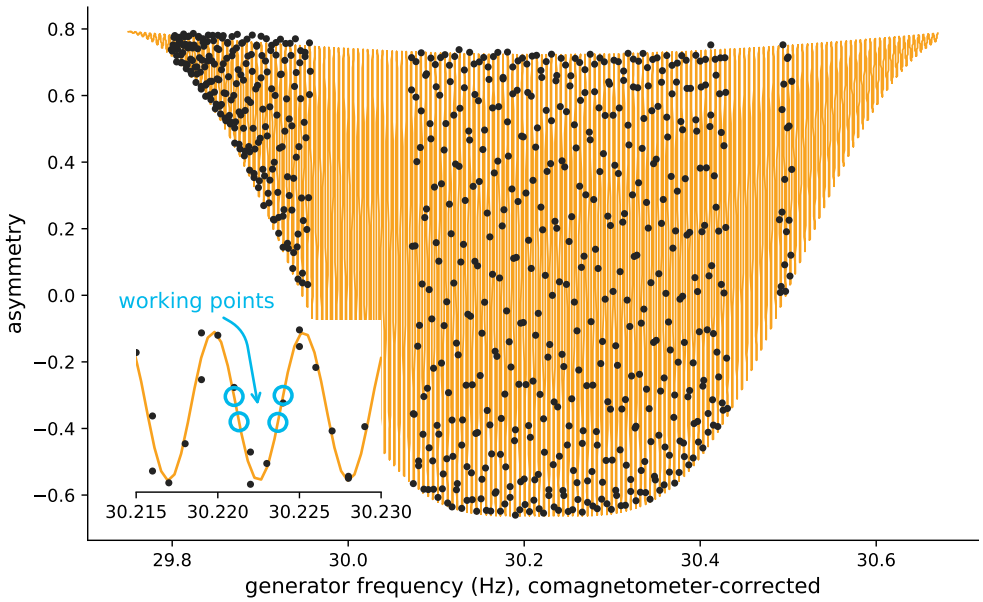

Figure 8: A scan of the Ramsey resonance curve in the nEDM experiment at the PSI. The asymmetry (the difference between spin-up and spin-down counts, normalised to their sum) is measured as the function of the detuning of the spin-flip generator's frequency. The black points depict the measured points, the orange line is the fit of the theoretical model (Eq. 3.125 in Ref. [52]). In the inset, the central fringe is enlarged and the working points are marked, which is where the experiment took data during normal operation. The generator frequency has been corrected with the ^{199}Hg comagnetometer using the formula $\nu_{\text{generator}}/\nu_{\text{Hg}} \times \overline{\nu_{\text{Hg}}}$, where $\overline{\nu_{\text{Hg}}}$ is the average over the scan. The resonance curve was scanned in runs 12666 and 12678.

it met a surface of a photodiode behind the cell, probed their spin precession. A coil wound around the cell, axial with the light beam, was driven in a feedback loop with the diode's signal to resonantly drive the Larmor precession. Even though the magnetometers measured only the magnitude of the magnetic field, they provided information about the higher-order components of the field thanks to their distribution around the chamber. For example, one could naïvely estimate the vertical gradient $\partial_z B_z$ by evaluating the average readings of the sensors in each electrode, taking the difference and dividing by the vertical separation. Instead, the high-order field terms, including the vertical gradient, were obtained by fitting a second-order parametrization of the magnetic field to the readings of the sensors [51].

In the next section we describe how all these components came together to perform a measurement of the electric dipole moment of the neutron.

2.3 MEASUREMENT PROCEDURE

Recall the Ramsey method (Fig. 3). Two coherent pulses of a transverse oscillating field are applied to an ensemble of polarised neutrons, with a period of free

spin precession in between. Measuring the polarisation after the second pulse as the function of the frequency of the transverse field yields a resonance curve. The curve measured in the PSI experiment is reproduced in Fig. 8. Comparing it with Ramsey's original curve (Fig. 4) summarises fifty years of progress in measuring the electric dipole moment of the neutron.

In the normal operation not the whole curve was sampled, but only the region most sensitive to the position of the central fringe—its steepest slope. In Fig. 8 the four *working points* are marked, which the experiment was programmed to aim at. Assuming all the data are taken in this most sensitive region, the sensitivity for the nEDM is (see e.g. [49]):

$$\sigma(d_\mathrm{n}) = \frac{\hbar}{2\alpha E T\sqrt{N}} \quad , \tag{8}$$

where α is the relative height of the curve ($\alpha = 1$ means the curve spanning from -1 to 1 in asymmetry), E is the strength of the electric field, T is the duration of the free spin precession (180 s, its inverse is the width of the fringes), and N is the counting statistics. This is the statistical limit on the sensitivity; the value obtained from the analysis is expected be slightly worse.

Each cycle of the experiment, filling in the neutrons, performing the Ramsey procedure and counting them, sampled one of the working points on the resonance curve. Assuming that the only parameter of the curve that changed between the cycles was the position of the resonance (due to drifts of the field), one could estimate the position of the central fringe, the neutron spin-precession frequency ν_n for each cycle. In other words, the neutrons were a very accurate magnetometer operating on a cycle basis.

Naturally, measuring the magnetic field with the neutrons was not the goal. It was the much, much smaller, if any, effect of the electric field on the spin-precession frequency that the experiment was after. To even out the variations in ν_n due to fluctuations of the magnetic field its ratio to the frequency of the comagnetometer's mercury atoms ν_Hg was used instead:

$$R \equiv \frac{\nu_\mathrm{n}}{\nu_\mathrm{Hg}} = \frac{\mu_\mathrm{n}}{\mu_\mathrm{Hg}} \pm \left(d_\mathrm{n} \mp \frac{\mu_\mathrm{n}}{\mu_\mathrm{Hg}} d_\mathrm{Hg} \right) \frac{2E}{h\nu_\mathrm{Hg}} + \Delta \, , \tag{9}$$

where the signs correspond to the parallel and antiparallel configuration of the magnetic and electric fields. (The derivation of the formula can be found in App. A.) It is immediately visible that R is sensitive to d_n just as ν_n, but the fluctuations due the changes of the magnetic field are suppressed. Δ encapsulates all higher-order terms and systematic effects. In particular, the already mentioned effect due to the gravitational sag of the neutrons' population, relative to the geometrical centre of the chamber, in combination with a vertical gradient of the magnetic field.

This correction was already included in the depiction of the resonance curve in Fig. 8.

R is actually sensitive to $d_n - \frac{\mu_n}{\mu_{Hg}} d_{Hg}$, but it has been measured that $|d_{Hg}| < 7.4 \times 10^{-30}\, e\,\mathrm{cm}$ [14].

In order to obtain the d_n estimate the electric field was modulated. The experiment automatically reversed the polarisation of the electric field every 72 cycles. In between the two polarisations a few cycles were measured without the electric field (a total of 10% of the data, which are not sensitive to the nEDM). Then a linear model of R vs. E yielded an estimate of the electric dipole moment, which we call d_n^* (Eq. 9, Fig. 9).

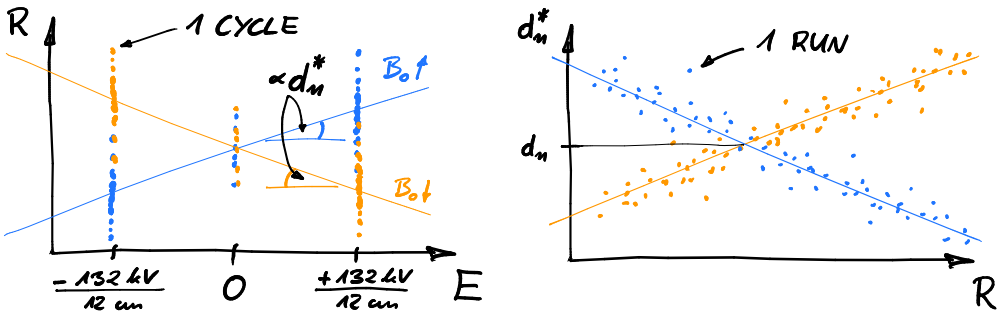

Figure 9: The crossing lines in the nEDM estimation. *Left:* For each sequence, data taken in one magnetic field configuration, d_n^* is estimated as the slope of a linear regression of R vs. E (Eq. 9). The two lines correspond to the B_0 field pointing upwards (blue) and downwards (orange). The d_n^* are tainted by a systematic effect proportional to the gradient. *Right:* a linear regression is done on d_n^* vs. R, R being an estimate of the gradient. The crossing point of the two lines, one for the magnetic field pointing upwards, one for downwards, is the first-order estimate of d_n, being free from the gradient-proportional systematic, although other corrections still apply [33].

Those estimates, however, still contained a large contribution of a systematic effect (hence the star). As a result of a conspiracy between radial magnetic field components and rotating magnetic components arising from the Lorentz transformation of the large electric field into the rest frame of the mercury atoms, a frequency shift proportional to the applied electric field was observed, mimicking the effect of a non-zero electric dipole moment of the neutron, a false EDM. This effect is, however, directly proportional to the gradient of the applied vertical magnetic field $\partial_z B_z$, and reverses with the sign of the applied magnetic field. The false EDM effect has been analysed in more detail in [55–57] and a direct measurement of this effect was made in [58].

Before d_n^* was estimated from the lines' crossing point, R values had also been corrected for the effect of the rotation of the Earth. For details see [33, 53, 54].

The solution was to operate in different vertical gradients and then interpolate to the effect-free zero gradient. Every few hundred cycles a different gradient (up to $\pm 60\,\mathrm{pT\,cm^{-1}}$) was set. Finally, for each direction of the magnetic field, a linear regression of d_n vs R, R itself being linear with the gradient (for small gradients), was performed (Fig. 9). The vertical position of the intersection of the two lines corresponded to the d_n measured at zero gradient, free from this systematic effect.

The blinding mimicked an nEDM signal by changing the logged neutron counts by around one neutron each cycle. The directon of the change depended on the electric field.

On top of that, in order to mitigate the bias due to the human factor, the nEDM experiment implemented data blinding. The data were artificially altered in a way that mimicked a non-zero neutron electric dipole moment, big enough to be visible in the data. The exact offset value is secret and would be revealed only after the analysis is complete.

In this short description we have only scratched the surface of the quite complex experiment. We discussed only those elements necessary for understanding the original work presented in the next chapters. More details can be found in the following references: [48–50, 58–60].

Part II

ACTIVE MAGNETIC SHIELDING

Stability of the magnetic field is a major challenge in measurements of the electric dipole moment of the neutron. An active magnetic shielding system had been succesfully used already in the nEDM experiment. This part begins with a discussion of the principles of active magnetic shielding and a description of the nEDM's system.

Spatial constraints made the design of coils for an n2EDM active shield challenging. This motivated the development of a new method of magnetic field coil design, which is part of this work.

Further, a prototype of an active magnetic shield featuring coils designed with the new method is described. The system had an unprecedentedly large useful volume. It constructively showed, that an active magnetic shield for the n2EDM experiment could be built.

Finally, a magnetic field mapping device is presented. It was used to characterise the magnetic field on the n2EDM site, providing crucial input to the design of an active shield.

The developments described in Part II laid the foundation for an active magnetic shield for the n2EDM experiment, which would reduce the tens-of-microteslas variations down to just a few.

3 THE NEDM ACTIVE MAGNETIC SHIELDING

In this chapter the principle of an active magnetic shielding is explained, followed by a description of the system of the PSI nEDM experiment. The matrix-based feedback algorithm and the properties of the matrix are discussed. Finally, the challenges for the design of a next-generation system for the n2EDM experiment at PSI are stated.

3.1 THE PRINCIPLE OF ACTIVE MAGNETIC SHIELDING

Passive methods of shielding the magnetic field, like mu-metal shields, rely on magnetic properties of materials. In contrast, in active methods the disturbances are first detected, processed and then counteracted. It is not unlike the recently popular active-noise-cancellation headphones. Standard ones provide only passive damping of the ambient noise by covering the ears. Active headphones additionally feature microphones that resolve the profile of incoming sound, which is then inverted and emitted from the speakers. The two waveforms, when overlaid, cancel.

Active magnetic shields follow the same principle. A volume, often called the fiducial volume, is encircled by coils. Within the volume magnetic field sensors are distributed. It is schematically presented in Fig. 10, taking the PSI nEDM experiment's system as an example. The fiducial volume is filled with the violet structure in the middle (the passive shield). The green dots depict the magnetic field sensors. Around there are three orthogonal Helmholtz-like pairs of coils, depicted in shades of orange. The sensors detect the variations of the magnetic field, an appropriate counteraction is calculated and currents are applied to the coils.

Active shields do not substitute passive ones. The shielding factor of passive shields degrades for frequencies slower than 1 Hz–10 Hz [62, 63]. At the same time active systems perform best at DC and reach up to the kilohertz regime. The combination of the two shielding methods provides a stable magnetic field over the whole range of frequencies [62–65].

Since the 1980s numerous active shields have been built [40, 62, 64–68]. The applications span from ion beams, through bio-magnetism, all the way to nEDM measurements, in particular the one conducted at PSI.

Shielding factor, measured in dB is a tenth of the logarithm of the ratio of the power of the magnetic field as if it were not compensated to the compensated one.

3.2 THE NEDM AT PSI SFC SYSTEM

The construction of an active shield for the nEDM measurement at PSI, called the SFC (Surrounding Field Compensation) system, was a part of Beatrice Franke's PhD thesis [61], later also published [40].

There was an inside joke that SFC really stands for SULTAN Field Compensation, SULTAN having been the magnet with by far the strongest influence on the magnetic field.

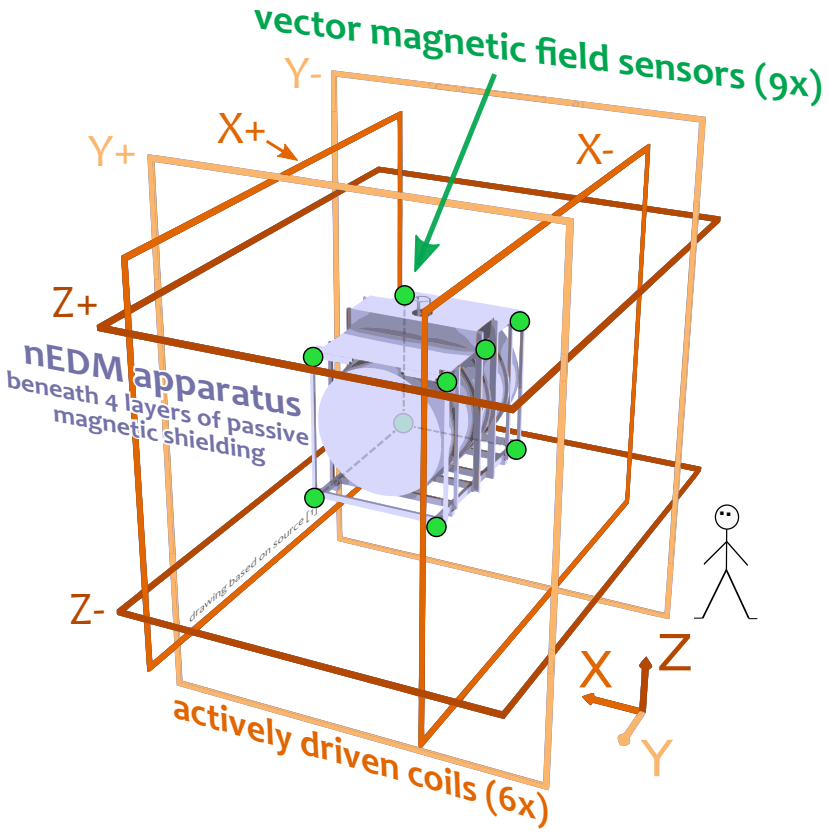

Figure 10: The nEDM at PSI active magnetic shield (SFC). The experiment, in violet, was in the middle, with nine 3-axis magnetic field sensors, depicted as green circles, attached to it. Everything was encircled by six coils (orange). Adapted from [61].

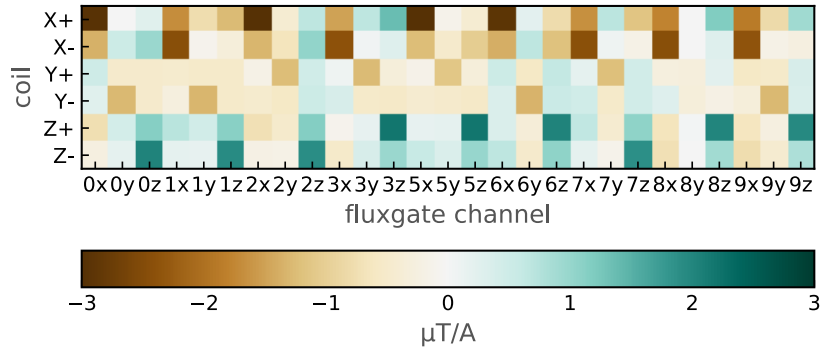

Figure 11: The SFC matrix measured by Franke on 2012–11–07 [61]. For each coil (rows) and each channel of the magnetic field measurement (columns) the proportionality constant between the current in the coil and the field measured by the sensor is depicted. The coils and the sensors are labelled as in Fig. 10.

The distinct feature of the PSI's SFC system was the use of a feedback matrix. Consider a single coil of any shape in air. For every point in space, and for every spatial component, the magnetic field is directly proportional to the current in the coil. The SFC system had six coils and measured the field in nine points, three spatial components for each. In total there were $6 \times 27 = 162$ proportionality constants, which were gathered in the feedback matrix M. The matrix M is a property of the active compensation system, more precisely of the geometry of the coils and sensors. It could have been calculated analytically, if not for the mu-metal shield. Ferromagnetic materials distort the magnetic field around them. Different to similar systems [62, 64–68], which used independent feedback for each coil, the SFC system at PSI used the matrix M that mixes all sensors and all coils.

A matrix-based feedback was first proposed by [69]. As their proposed system did not include µ-metal, they could calculate the matrix analytically.

To measure the matrix the following procedure was used: A current in one coil was scanned in steps over the whole available range, while the field change was measured with the sensors. Then for each sensor, and each spatial component, a linear regression was performed. The slope corresponded to the proportionality constant, an element of the matrix M. The procedure was repeated for all coils. The matrix measured by Franke is depicted in Fig. 11.

3.3 THE FEEDBACK ALGORITHM

The PSI's SFC system followed the established norm [62, 64, 66, 67] and used a PID loop to control the currents. (PI in this particular case, the derivative term was not used.) The jth current in the nth iteration was [61]:

$$I_j^n = I_j^0 + \underbrace{\alpha_j^{\mathcal{P}} \Delta I_j^n}_{\text{proportional term}} + \underbrace{\alpha_j^{\mathcal{I}} \sum_m \Delta I_j^m}_{\text{integral term}} . \tag{10}$$

ΔI_j^n is the error value. It was obtained from the deviation of the measured and target field, $\Delta B_{k'}^n$ for the k'th sensor, by multiplying it by the pseudoinverse of the matrix M, denoted $M^\dagger$ [61]:

$$\Delta I_j^n = \sum_{k'} M_{jk'}^\dagger \, \Delta B_{k'}^n \ . \tag{11}$$

The SFC system, when started, was static. Then the currents could be changed manually to achieve a desired field (or chosen from a predefined set). Then the system was switched to the dynamic mode [61].

Besides the proportional and integral terms, the feedback formula also included the I_j^0 term. It may seem puzzling to give such a big of a role to particular currents that had happened to be there at the zeroth iteration. The reason behind this choice was a particular property of the SFC system: the target field was always the one measured at the moment of switching the system from the dynamic mode (feedback running) to the static one (static current output). At that moment the error value was zero, and so was the integral term, so, according to the Eq. 10, the output currents would immediately have switched to zero. In result the system would violently destabilise. The additional term I_j^0 prevented that from happening.

Here we conclude the brief description of the system and proceed to present original insights of the author.

3.4 INTERPRETATION OF THE FEEDBACK MATRIX

There is a fundamental meaning behind the matrix M. The currents in the six coils can be gathered into a vector $\mathbb{I}$ in a space we call the *current space*. Similarly, the 27 readouts of the magnetic field can be gathered into a vector $\mathbb{B}$ in the *field space*. Then, the matrix M provides the transition from the current into the field space. In particular, we have:

$$\mathbb{B} = M\mathbb{I} + \mathbb{B}_{\text{offset}} \ . \tag{12}$$

Finding the pseudoinverse of a matrix A is equivalent to finding least-squares solution of a system of linear equations described by the matrix A (although computationally more complex).

Which is read: The measured values are a linear combination of the currents in the coils plus the ambient field. The other direction, from the field to the current space, cannot be done exactly. Nevertheless, the optimal, in the least-squares sense, transition can be done with the use of the Moore-Penrose inverse (pseudoinverse) of the matrix M, denoted $M^\dagger$ [70]. In other words, the matrix M tells us what field, as measured by the sensors, a given set of currents will produce. The matrix $M^\dagger$ tells us, what currents to apply to best approximate a given field.

3.5 THE SPECTRUM OF THE FEEDBACK MATRIX

The feedback matrix used during the data taking of the nEDM PSI experiment at PSI (2014–17) was the one measured by Franke (Fig. 11). The inverse of the matrix was regularised [61]. Let us now elaborate on regularisation.

The feedback matrix represents coefficients in a system of linear equations that need to be solved in order to make the optimal transition from the field to the current space. As the system of equations is over-determined, the best

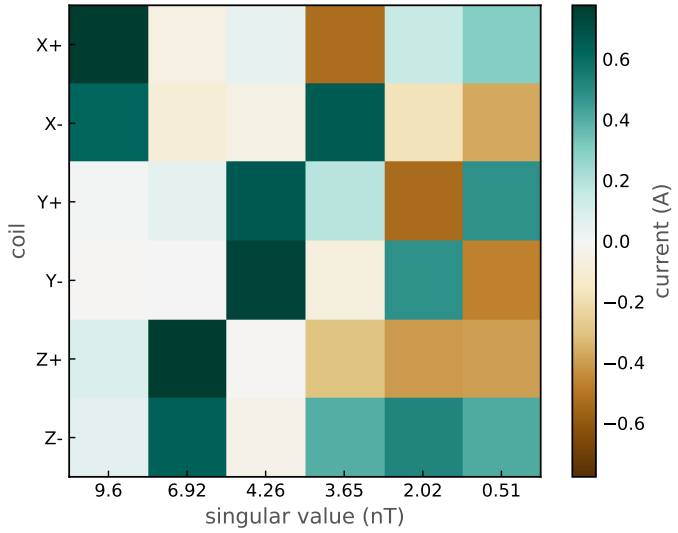

Figure 12: The coil-singular values of the SFC matrix. Columns correspond to singular combinations of the coils (labeled as in Fig. 10). For each column the corresponding singular value is indicated.

solution is found by the least-squares. This is equivalent to calculating the pseudoinverse matrix. A solution can be found reliably if the system is well-defined, i.e. the least-squares well is steep in every direction in the parameter space. If the well stretches, like a valley, in some directions, the solution is not globally well-defined. Still, it may be defined up to a parameter, the one pointing along the direction of the valley. We then speak of an ill-defined set of equations, or an ill-defined matrix. Regularisation helps an ill-defined problem to become better defined, at the cost of the solution having larger sum of the residuals squared.

The *Singular Value Decomposition* (SVD) of a real matrix M is

$$M = USV^T \,, \tag{13}$$

where U and V are unitary, and S is diagonal [71]. The singular values lie on the diagonal of S, which is called the *spectrum*. The *condition number* of a matrix is defined as a ratio of the extreme values of its spectrum [72]. For a matrix with a flat spectrum, all singular values equal, the condition number is one. A set of linear equations represented by such a matrix is well defined. The more extreme values differ, the higher the condition number and the worse defined the problem is. The effect is, that the noise in the original matrix becomes amplified by the condition number in the pseudoinverse. Figure 12 presents the spectrum of the PSI's SFC feedback matrix. The condition number is $9.6/0.51 = 18.2$. A factor of 18 amplification of noise points to why the regularisation was necessary.

It is interesting to ask *why* the system had an ill-defined feedback matrix. A very small singular value means that there exists a coil, or a combination thereof, which has about 18 times smaller influence on the field then other ones. In Fig. 12 the columns are the singular values with their corresponding

coil-vectors. The first three, starting from the left, are easiest to interpret. Each of them is a pair of coils configured as Helmholtz-pair with roughly the same current, producing a homogeneous field in each of the spatial directions. The last column has also a clear interpretation: it corresponds to all pairs configured as anti-Helmholtz—currents flowing in the opposite directions in each pair. The magnetic field that was created by such a configuration was complicated and high-order. The fact that this combination has so little influence on the measured field means that it hardly changes any solution for currents when added to it. It spans a valley in the parameter space in the least-squares problem. If not regularised, a small change in the measured field would cause a large change in the currents along the valley. Rapid oscillations in this direction would render the system unstable.

An important observation is that the feedback matrix is defined solely by the configuration of the coils, sensors and ferromagnetic materials. It follows that already at the design stage one can, and should, take care about the definiteness of the system.

3.6 DESIGN CHALLENGES FOR N2EDM

There was relatively a lot of space available around the passive magnetic shield of the nEDM experiment at PSI (see Fig. 10). The SFC system exploited that space by making the coils much larger than the shield. However, the successor experiment, n2EDM, was going to have a much larger passive magnetic shield. Spatial constraints were a major concern in the design of coils of a new active shield.

Assume, that all there is to compensate are homogeneous fields. That is the zeroth order approximation concerning the space expansion of the magnetic field. To compensate them, a system needs to have coils that can produce a homogeneous field. Yet, the volume where the field of a coil system is homogeneous is limited. For example, in a case of a Helmholtz pair, the (linear) size of the volume where the generated field is homogeneous on a 2 % level is only around 0.4 that of the (linear) size of the coils. Increasing the relative size of the homogeneous volume requires more elaborate designs [73]. When three Helmholtz pairs are used to compensate a homogeneous change, they only do so inside this limited volume. Outside, the field may even be more unstable, the extreme case being right next to the coils, where the field can get arbitrarily large. The situation is similar for any type of field: an active shield can only stabilise the field in a limited volume.

Figure 13 depicts the major spatial constraints in the design of the n2EDM active shield. Firstly, the passive shield was going to be 5 m large in each direction, with the available space being ≈ 7.5 m (a ratio of 0.7). In nEDM the shield was around 2 m and the SFC coils 6–8 m (a ratio of around 0.3). Secondly, other bodies put additional constraints (Fig. 13): support legs of the passive shield, the polarising magnet and an emergency exit from a nearby cyclotron.

For the design of the n2EDM active shield it was necessary to come up with coils that could produce such a field, that they are likely to compensate, and

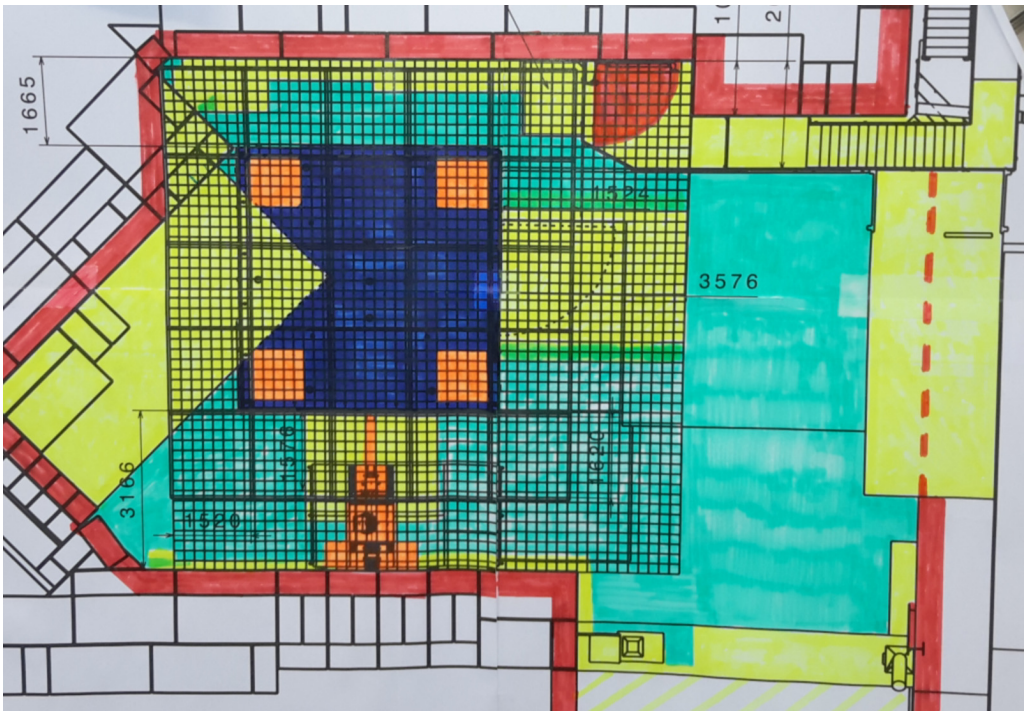

Figure 13: A plan by Michael Meier highlighting the potential conflicts between coils of an n2EDM active shield and other structures. The vicinity of the biological shielding (red) forced the coils to be designed relatively close to the magnetic shield (dark blue). Other bodies, like the support legs of the passive shield (orange), polarising magnet (orange) and an emergency exit from a nearby accelerator (red) gave additional constraints.

do that in the whole volume occupied by the shield. Unfortunately, because of spatial constraints no known design could be used [74–79]. This, together with a concern about ensuring a low condition number of the system, called for an in-depth investigation into the geometry of the future system's coils, which led to the development of a new approach to magnetic field coils design—the topic of the next chapter.

4

COIL DESIGN

In this chapter a new method to design a coil producing an arbitrarily shaped magnetic field is introduced. In this method the path of the coil's wires are restricted to a regular grid, the solution is then found by a simple least squares minimum. Practical applications, in particular in an active magnetic shielding, are discussed.

4.1 INTRODUCTION

This chapter largely reproduces text of the author's publication [80]

How can we design a coil, or more generally, an arrangement of wires, to produce a desired magnetic field? In its simplest form this is a textbook problem (e.g. ex. 6.55 and 6.62 in Ref. [81]). Yet in general it is surprisingly hard, and the solutions (how the wires making up the coils should be laid) complicated. The most widespread application of high-performance coils is in Magnetic Resonance Imaging (MRI), where gradient coils give the possibility to produce spatial images. In the 1980s elaborate methods of MRI coil design had already been developed. These methods range from optimizing positions of discrete windings, where the use is made of symmetries specific to MRI, to analytical methods yielding surface current density, which is then discretised. A general overview can be found in Ref. [76]. Another field known for complex, precise coils is plasma confinement, in particular stellarators [82]. There analytical solutions for the surface current density find their use, too.

Here, a new method is presented that may not be competitive in terms of precision, but is distinct in its simplicity—also when it comes to construction of the designs. The procedure relies on an algebraic representation of the problem, where coil design is simplified to a simple linear least squares problem. In the method the coils are restricted to a user-defined mesh, resulting in easy to deal with spatial constraints.

The discussion is based on textbook linear algebra techniques, notably solving an over-determined system of linear equations, thoroughly discussed e.g. in Ref. [83]. The main physics problem, calculating the magnetic field of coils composed of straight wire segments, is briefly discussed here. More in-depth discussions can be found, for example, in Ref. [84]. Furthermore, the implementation of the discussed problems, including examples, has been published and open-sourced [85].

We begin with a description of our model for a restricted two-dimensional case and generalize it to three dimensions. We then present how the model is used to design a coil, based on an example. Further we discuss possibilities of simplifying the solution. Another section is devoted to practical considerations, significant for the eventual construction. Finally, the application to active magnetic field shielding is considered.

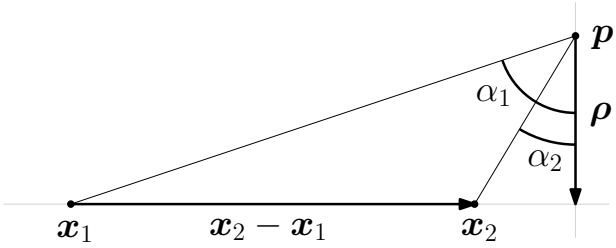

Figure 14: The setting for calculating the magnetic field produced in point $\mathbf{p}$ by a straight wire segment from $\mathbf{x}_1$ to $\mathbf{x}_2$.

4.2 COILS AS A LINEAR SPACE

Consider all the possible coils that can be constructed by laying a wire on a surface of a square. The possibilities are endless. More precisely, as the wires are shifted by arbitrarily small distances, overlapping and crossing other wires, the problem has an infinite number of degrees of freedom. Here, an algebraic representation that reduces the number of degrees of freedom to just a few is proposed.

We start with a straight, finite wire segment spanned between points $\mathbf{x}_1$ and $\mathbf{x}_2$ (represented by vectors in an arbitrary coordinate system) and carrying current I, as depicted in Fig. 14. The magnetic field it produces in the point $\mathbf{p}$ can be calculated with the use of the Biot-Savart law. Consider the vector normal to the wire through the point $\mathbf{p}$:

$$\boldsymbol{\rho} = (\mathbf{x}_1 - \mathbf{p}) - ((\mathbf{x}_1 - \mathbf{p}) \cdot \mathbf{n})\, \mathbf{n} \,, \tag{14}$$

where $\mathbf{n}$ is a unit vector in the direction $\mathbf{x}_2 - \mathbf{x}_1$. The magnitude of the magnetic field at point $\mathbf{p}$ is then [84]:

$$B = \frac{\mu_0 I}{4\pi\rho} \left| \sin\alpha_2 + s \, \sin\alpha_1 \right| \,, \tag{15}$$

where the angles α_i are not directed

$$\sin\alpha_i = \frac{\|(\mathbf{x}_i - \mathbf{p}) \times \boldsymbol{\rho}\|}{(x_i - p)\rho} \,. \tag{16}$$

s is $+1$ if $\boldsymbol{\rho}$ points onto the wire segment (between points $\mathbf{x}_1$ and $\mathbf{x}_2$) and -1 otherwise:

$$s = \operatorname{sgn}\left(\tfrac{1}{2}\,\|\mathbf{x}_2 - \mathbf{x}_1\| - \left\|\mathbf{p} + \boldsymbol{\rho} - \tfrac{1}{2}\,(\mathbf{x}_1 + \mathbf{x}_2)\right\|\right) \,. \tag{17}$$

Another formulation, with better numerical properties close to the axis of the wire, can be found in Ref. [86].

The direction of the field is given by the right-hand principle

$$\mathbf{B} = \frac{B}{\rho}\boldsymbol{\rho} \times \mathbf{n} \,. \tag{18}$$

This formulation is independent of the coordinate system (coordinate-system dependent solutions can be found e.g. in Ref. [87]).

Imagine four wire segments making up a square loop—a coil. It produces a certain magnetic field in the entire space $\mathbf{B}(\mathbf{x})$ following the superposition

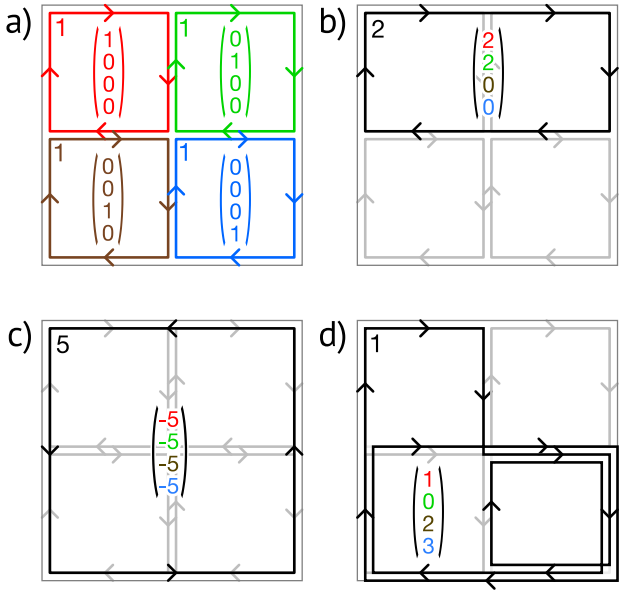

Figure 15: a) A basis of four tile coils on a flat square. Any coil which has its wires restricted to lie on the 2×2 grid can be represented as a linear combination of the four base tile coils. b, c, d) Three coils are presented together with their explicit coordinates in the basis.

principle, by summing the fields produced by each segment of the coil. When the current in the coil is changed only one parameter of the magnetic field is altered—the magnitude, but not its shape. Therefore, it can be said that one coil spans a one-dimensional space of magnetic fields it can produce. Adding a second coil creates a system spanning a two-dimensional space of fields, as the magnetic field is additive. Going a step further, four square coils tiled to form a larger square form a four-dimensional space, as shown in Fig. 15. Any coil restricted to a 2×2 grid can be represented in the base of the four tile-coils.

The range of magnetic fields reachable by coils restricted to a grid is a subset of all possible fields that can be created with coils constructed on the square's surface. The size of the subset is controlled by N, the number of tile-coils forming the grid. In this system a coil is fully described by a vector of N currents, one in each of the tile-coils denoted by $\mathbb{I}$. The problem of coil design is thereby simplified to finding a vector $\mathbb{I}$.

Generalisation onto a cube is simple. A cube is made up of six square faces. Interestingly, for the assembly in the three-dimensional space one degree of freedom is lost. Figure 16 illustrates the simplest case where $N = 6$, a configuration in which finite currents in all six coils cancel and no magnetic field is produced. Such a combination of currents can be added to any solution with no effect on the produced field. Effectively, the space of the fields they can produce has dimension five (i.e. $N - 1$). In other words, the mapping of $\mathbb{I}$ onto fields $\mathbf{B}(\mathbf{x})$ has in this case a one-dimensional kernel. This fact is of importance when it comes to numerically solving the system.

This is the foundation of the method. The restriction to a grid on a cuboid enables all coils in the restricted space to be described by a vector of N numbers.

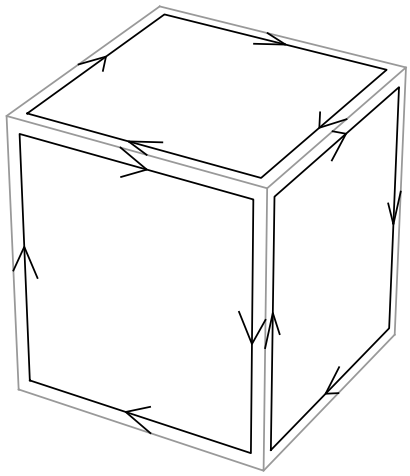

Figure 16: An arrangement of $N = 6$ tile coils on a cube which produces no magnetic field. The currents in the tiles are equal and flow in the directions as indicated. The currents on the invisible faces are analogous to the ones seen in front. For clarity, the coils are depicted slightly smaller. In the model the currents are identical with the edges of the cube.

4.3 COIL DESIGN

In the problem of coil design a coil or an arrangement of coils is sought, which best approximates a given field in a certain volume. The volume we shall call *the volume of interest*. Rather than considering the whole volume, an ensemble of m points of interest on its surface is selected (the surface is sufficient because $\nabla \cdot \mathbf{B} = 0$). Henceforth, only the magnetic field $\mathbf{B}(\mathbf{x})$ at these points is considered. The values $\mathbf{B}(\mathbf{x}_i)$ for $i = 1..m$ are gathered into a vector of dimension $3m$ (B_x, B_y and B_z in each point), which we shall denote $\mathbb{B}$.

As mentioned before, the magnetic field produced by a coil at any given point in space is proportional to the current in this coil. With many coils present it is a linear combination of the currents of all coils in the system. In absence of an external magnetic field the system of N tiles and m points of interest is thus described by a simple linear equation

$$\mathbb{B} = M\mathbb{I}\,, \tag{19}$$

where $M \in \mathbb{R}^{3m} \times \mathbb{R}^N$ is a matrix of proportionality constants. For example, the element $M_{(5,2)}$ is the proportionality constant between the current in the second of N coils and the magnetic field in the y direction in the second of m points of interest, $B_y(\mathbf{x}_2)$. The matrix M can be calculated analytically using the Biot-Savart law.

Equation 19 with $3m > N - 1$ is an over-determined system of linear equations, $\mathbb{I}$ being the vector of unknowns. The optimal least-squares solution $\mathbb{I}_0$ to produce $\mathbf{B}_0(\mathbf{x})$ in the volume of interest is

$$\mathbb{I}_0 = \arg\min_{\mathbb{I}} \left(M\mathbb{I} - \mathbb{B}_0\right)^2 . \tag{20}$$

The optimal solution can be calculated with the normal equation [83]:

$$\mathbb{I}_0 = \left(M^{\mathrm{T}} M\right)^{-1} M^{\mathrm{T}} \mathbb{B}_0 \,, \tag{21}$$

Least-squares is solved with a backslash operator in MATLAB-like languages: `I0 = M \ B0.`

but the problem is typically solved numerically. The majority of numerical software packages use the QR decomposition (a product of an orthogonal and upper-triangular matrix) of the matrix M, which is more numerically stable when compared to the normal equation.

Depending on the properties of M the optimum may be multidimensional. As already mentioned, an arrangement of coils on a cube has a one-dimensional kernel, which will always cause the optimum to be at least one-dimensional. In these cases, we will call $\mathbb{I}_0$ the unique least-norm solution, which minimizes the total current in the system. $\mathbb{I}_0$ is the vector of the optimal currents in the tile arrangement of coils for approximating $\mathbf{B}_0(\mathbf{x})$ in the volume of interest.

Let us consider an example of a coil design on a unit cube with the number of tiles $N = 6 \times (3 \times 3)$ (see Fig. 17). As the volume of interest we pick a centered cube with side length 0.75 (with a regular mesh of 10×10 points on each face, a total of $m = 488$ points of interest). For the sake of simplicity we design a coil for a homogeneous field along an axis of the cube. The solution of Eq. 20 directly gives the currents in each tile. They are graphically depicted in Fig. 17. Note that many currents almost cancel each other out, in particular those along horizontal edges. The magnetic field produced by the solution is shown on the left-hand side in Fig. 18 as a horizontal cut along the central plane. Contours show the relative deviation from the homogeneous field. Inside the volume of interest (dashed line) the design goal of a homogeneous field is reproduced with few per cent accuracy. The solution, and the contours, depend on the choice of the volume of interest. In general, the further away the volume of interest is from the coils, the better the accuracy. If the side length of the volume of interest is decreased to 0.5, the accuracy improves to 1 %, as shown on the right-hand side in Fig. 18. The optimal solution, and thereby the shape of the precision contours change. The accuracy of the field reproduction can also be improved by increasing the number of tiles.

4.4 SIMPLIFICATION OF THE TILE SYSTEM

The tile system may find an interesting practical application. Once independently controllable tiles have been built, it can be used to produce any arbitrary field. However, building many independently driven coils is a high price to pay for producing only one field. Additionally, each edge is shared between two tiles, and the effective current is the sum of these. The currents add either constructively or destructively. If the given solution is dominated by subtraction of large currents, a lot of power is unnecessarily dissipated in the system. We find that both problems can be solved by simplifying the tile solution.

One starts by adding the currents of adjacent tiles and assigning the sum to each common edge. The result is a complicated net of currents (upper left corner of Fig. 19), with each node fulfilling Kirchhoff's laws. The net can then be decomposed into simple current loops by the following algorithm: First find

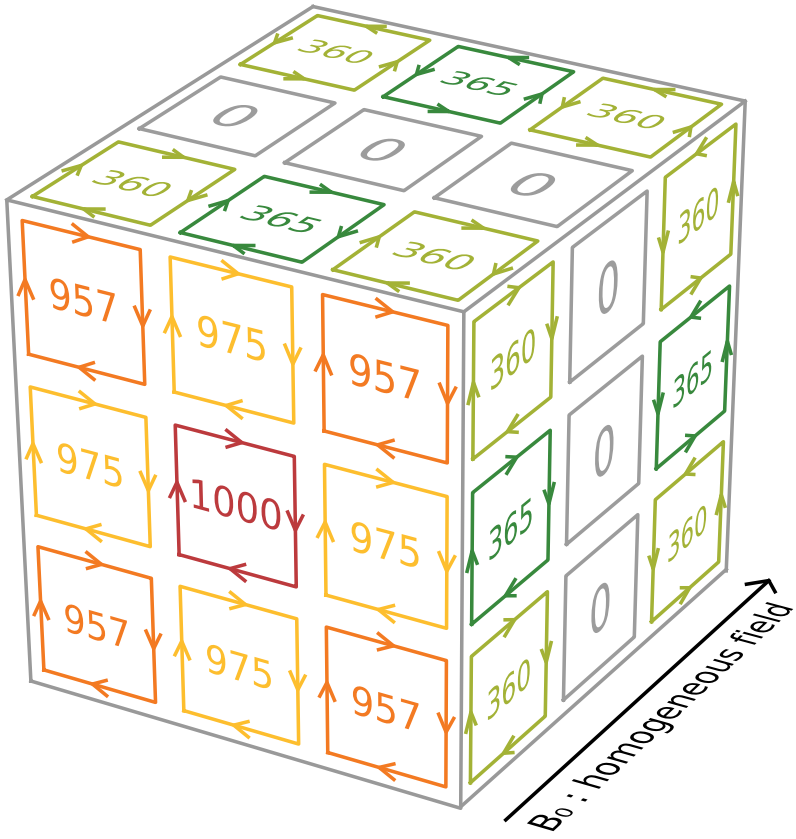

Figure 17: A solution of a tile system with $N = 6 \times (3 \times 3)$ tiles on a unit cube for a homogeneous field. The volume of interest is a cube with side length 0.75, centered inside the unit cube. Numbers indicate currents in the tile coils in arbitrary units. The currents are normalized so that the highest is 1000. For clarity, the coils are depicted slightly smaller; in the model their edges overlap. The currents on the three invisible faces are by symmetry analogous to the visible ones.

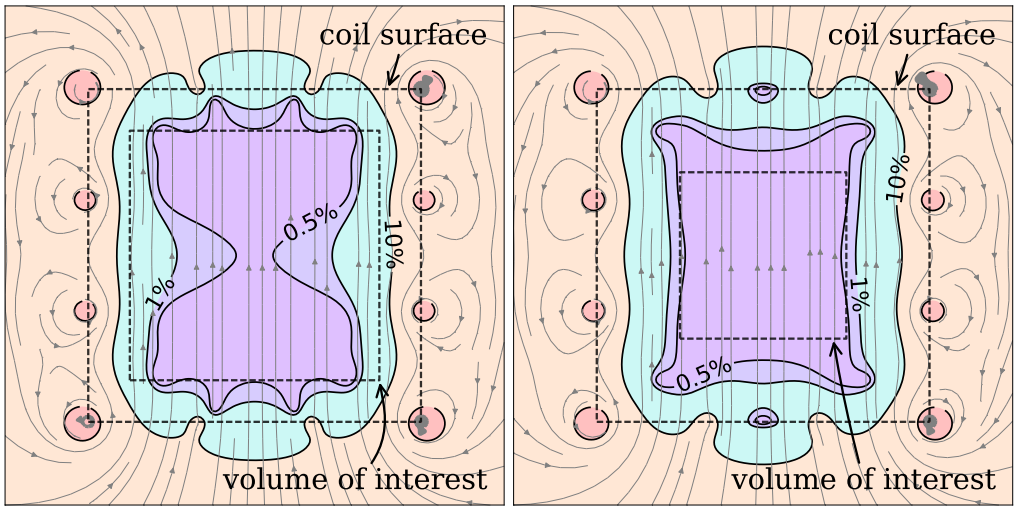

Figure 18: Magnetic field produced by a coil designed for a homogeneous field, with $N = 6 \times (3 \times 3)$ tiles on a unit cube. The field lines are depicted in grey. Contours show boundaries of 0.5, 1 and 10 % magnitude deviation from an ideal homogeneous field. Horizontal cross sections in the middle-height plane are shown. Two designs are presented. Left-hand side: the volume of interest is a cube with side length 0.75 (the individual tile coil currents are depicted in Fig. 17), right-hand side: the size of volume of interest is reduced to 0.5.

in the net the loop with the highest current. In the example it is either of the "597" loops on the front and back faces. This will be the primary loop in the simplified solution (it can be seen in the middle of the right column in Fig. 19, together with the next three loops). Then subtract from the net the current of the primary loop along its edges (the net that remains after subtracting the first four loops is depicted in the middle of the left column in Fig. 19). Finally, continue to find the loop with the highest current in the modified net, which will give the next loop and repeat until the current net is empty. The net remaining after eight loops are found is depicted in the bottom row of Fig. 19, next to the first eight loops. The final simplified solution is shown in Fig. 20. The currents in the simplified coil system are much smaller, the highest being 597 instead of 1000 and they always add constructively. Also, the number of separate loops is decreased from 42 to 10. Still, the total current along each edge of a tile is exactly the same as in the tile configuration.

We conclude here our method of coil design. The simplified arrangement of coils is optimal, given the grid restriction, for approximating the magnetic field in the volume of interest. We continue to consider practical aspects, relevant for constructing the designs of the new method.

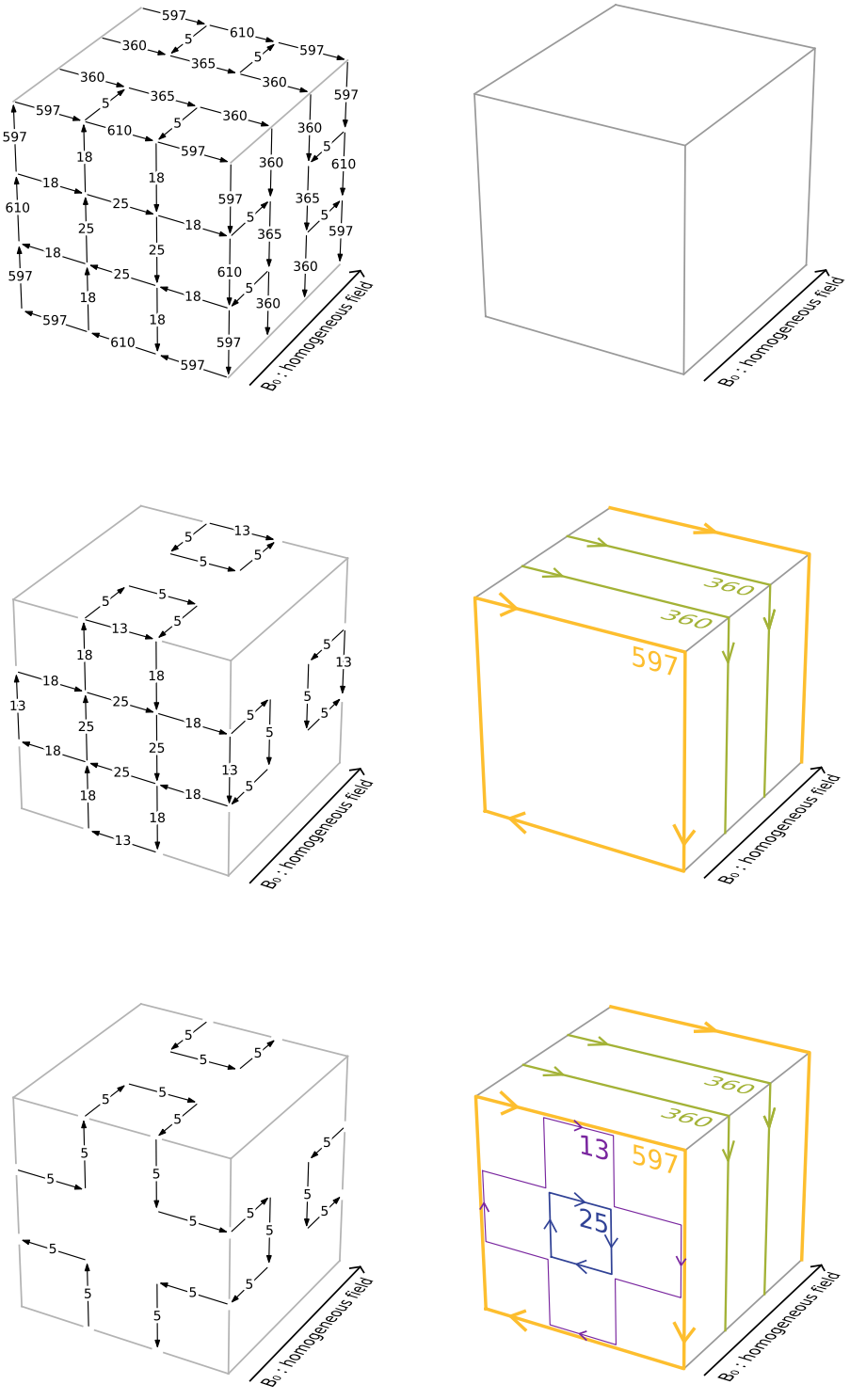

Figure 19: Following the algorithm to simplify a coil. The left column shows the net of a current with the total current along edges of tiles. In each iteration the loop with the highest current is found and transferred onto the simplified solution, shown in the right column. We show iterations, from top: zeroth, fourth and eighth.

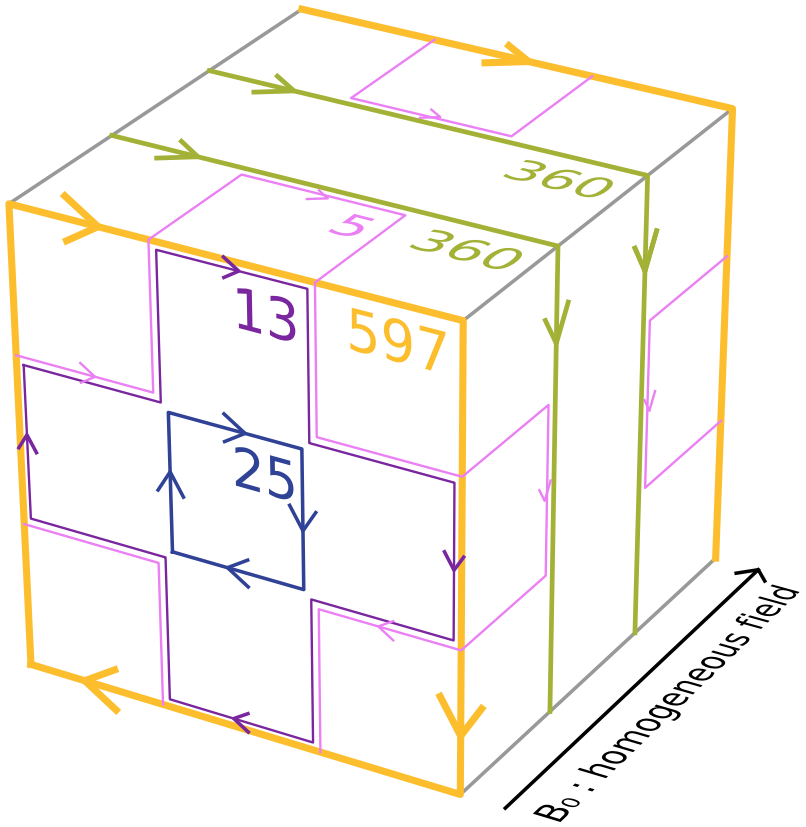

Figure 20: The coil designed for a homogeneous field, with $N = 6 \times (3 \times 3)$ tiles (Fig. 17), simplified by adding the currents along each edge and decomposing into current loops.

4.5 PRACTICAL CONSIDERATIONS

The foremost practical advantage of the new design method is that the coils are constrained to a predefined grid. This is contrary to other methods of coil design, where the position of the wires is the output of the procedure [76, 82]. This may prove useful in applications with spatial constraints. Typically, coils need to be incorporated into a setup in which other components penetrate the surface on which the wires are laid. With the new method it is possible to simply define the grid so that no collisions occur. Although the simple examples presented before used regular grids, we have not used symmetries to solve the problem. When many coils are designed and built, for instance to produce homogeneous magnetic fields in each of the three dimensions, they can all share the same grid. The grid can, for example, be constructed out of cable channels into which the wires are laid.

A limitation associated with the finite size of the channels is the strength of the magnetic field that can be created, which for a given available power is limited by the thickness of the wire. At the same time, the finite size of the cable channels can be neglected in the calculations, only as long as it is small compared to the distance between the coils and the volume of interest. Using an enameled wire, rather than the standard, PVC-insulated cable, we can reduce the overall thickness.

In order to produce the desired field, one still needs a system of several coils, even in the simplified solution. The more complicated the goal field, and the more tiles, the greater the number of currents across the individual loops is required, which quickly becomes impractical. There are several ways to tackle the problem.

The first way is to use only one current and adjust the magnitude by varying the number of windings. In the example when one decides for 60 windings as the maximum, the nominal current that would flow through the wire is $\mathrm{round}(597/60) = 10$. 597, 360, 25, 13 and 5 would be created with 60, 36, 3, 1 and 1 windings, respectively. A discretisation error of $10/597 = 1.7\,\%$ is of the same order as the accuracy of the solution in representing the field (see Fig. 18). For more precise designs the numbers of windings increase, which is more difficult to construct and causes the coils to have larger inductances.

A second way is to use a current divider. This connects the different loops in parallel, each with an appropriately chosen resistance in series. This way the ratios between the currents in each loop can be tuned precisely. However, a practical realization will most likely involve routing all loops out of the system where the current divider is installed. For more complicated coil systems with tens of different currents this may be impractical.

Yet another way is to split the loops into decades of currents. In the coil we use as an example, the currents 597, 360, 13, 7, 5 (in arbitrary units) may be constructed from a set of wires with three relative currents of 100, 10 and 1:

A base different than 10 can be used, too.

n	$\mathbf{P}_n^x(\mathbf{r})$	$P_n^y(\mathbf{r})$	$P_n^y(\mathbf{r})$
1	1	0	0
2	0	1	0
3	0	0	1
4	x	0	$-z$
5	y	x	0
6	0	y	$-z$
7	z	0	x
8	0	z	$-y$

Table 1: Cartesian harmonic polynomials. Further terms can be found for example in Ref. [61].

$$
\begin{aligned}
597 &= 5 \times 100 + 9 \times 10 + 7 \times 1 \\
360 &= 3 \times 100 + 6 \times 10 + 0 \times 1 \\
13 &= 0 \times 100 + 1 \times 10 + 3 \times 1 \\
7 &= 0 \times 100 + 0 \times 10 + 7 \times 1 \\
5 &= 0 \times 100 + 0 \times 10 + 5 \times 1 \, .
\end{aligned}
$$

Using this method, a reproduction accuracy better than 1 % can be reached with only three different currents to control, even for complicated designs. Those can be either separately controlled or split with a current divider.

No claim is made as to the superiority of one of the three above solutions over the other two. The best solution is up to the particular application for which it is required.

4.6 APPLICATION TO ACTIVE MAGNETIC FIELD SHIELDING

Using the method presented here to design coils of an active magnetic shield offers improvements in two areas. Firstly, the size of the coils could be decreased, or the size of the experimental set-up increased, without loss of performance. Better homogeneity in a given volume can always be achieved by choosing a denser grid. Given the tight spatial constraints this was a crucial development for the design of the n2EDM active shield.

Secondly, the method allows one to construct a coil for any field. In particular, one may choose to construct coils that produce fields orthogonal to one another. This makes an active shielding system significantly easier to control [88] and avoids potential problems in the high condition number due to very-high-order fields produced by pathological combinations of coils (recall the discussion in Sec. 3.5). One of the possible orthogonal decompositions of the field is into *cartesian harmonic polynomials* [61]:

$$
\mathbf{B}(\mathbf{r}) = \sum_n H_n \mathbf{P}_n(\mathbf{r}) \, , \tag{22}
$$

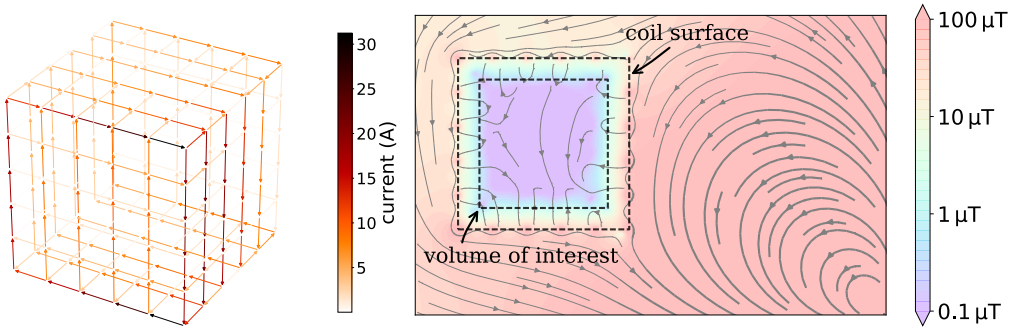

Figure 21: A coil designed on a unit cube with 5×5 tiles per face to shield against a dipole disturbance. The dipole is located, relative to the center of the unit cube, two units to the right and one unit to the front. It is located in the middle height of the cube. The volume of interest has a side length of 0.75. On the left-hand side the total current along each edge of the dipole compensation coil is depicted. On the right-hand side the magnetic field is shown. The magnetic field lines are shown in grey, the volume of interest and the coil surface with dashed lines. The colors depict the magnitude of the magnetic field (capped at 0.1 and 100 µT). A horizontal cross section in the middle height is shown. The dipole source is located in the lower right corner of the plot and points parallel to the plane of the plot. The magnitude of the field in the volume of interest is reduced from tens of microteslas down to below one.

where H_n are the expansion coefficients and $\mathbf{P}_n(\mathbf{r})$ are the cartesian harmonic polynomials, the first eight of which are listed in Table 1. Each term satisfies Maxwell's equations by itself. The first three terms are homogeneous fields, the next five are the five independent linear gradients. Further terms correspond to higher-order gradients.

Additionally, one can consider constructing dedicated coils to counteract a particular known disturbance. In Fig. 21 a showcase design with $N = 6 \times (5 \times 5) = 150$ tiles for compensating a nearby dipole source is presented.

Lastly, the method's unique property of designing coils on a predefined grid makes a large-scale construction particularly simple. It makes the coil array easily incorporated into existing structures by defining the grid in a conflict-free way.

COIL DESIGN – CONCLUSION

Coil design is a complicated and very technical problem, especially when high accuracy is required. This work presents a method that is simple in terms of both underlying maths and computational effort. The design method could find its niche in practical applications, where spatial constraints play a significant role and a percent level in accuracy of the produced field is acceptable. The method has particular advantages when used to design coils of an active magnetic field compensation system. In the next chapter, its application to construction of a compensation system at ETH Zürich is described.

The software implementation of the coil design, including examples, has been published as an open-source and can be accessed in Ref. [85].

5 NEXT-GENERATION ACTIVE MAGNETIC SHIELDING

The coil design method described in the previous chapter opened the door to a next generation of active magnetic shields, where the coil system is not much larger than the fiducial volume and where high-order terms of the magnetic field can be compensated, all while retaining a low number of controlled degrees of freedom.

In a laboratory at ETH Zürich an active shield was constructed. In its first version it featured three coils for the homogeneous components of the magnetic field. The coils were wound on a regular grid. Mapping of the magnetic field created by the coils confirmed that they produce a field at the specified homogeneity in a large volume. The active shield was tested with strong, inhomogeneous disturbances; its long-term stability was also tested.

A next iteration of the active shield was constructed on a grid with one open face. This not only eased access to the inside, but also demonstrated coils of a more complicated design.

Finally, coil design for an active magnetic shield for the n2EDM experiment is proposed. Despite tight spatial constraints, it achieves homogeneity of 1–2 % around the mu-metal shield.

5.1 THE FIRST ITERATION — COIL STRUCTURE

In the discussion of the coil design method a practical way to realise it was indicated—to construct a grid out of cable channels. At ETH Zürich a system pictured in Fig. 22 was built consisting of a $5 \times 9 \times 5$ grid of square tiles. Each tile had side length 262 mm and the total size was $1310 \times 1310 \times 2358$ mm. The vertical axis we refer to as z, the long horizontal one as y, and the remaining axis x. The support frame was made of aluminum construction profiles. To support the cable channels, on each side a large one-piece aluminum sheet was attached, with square cut-outs leaving material only directly below the cable channels. The plastic cable channels were glued onto the aluminum.

For the first version the system featured three coils for the homogeneous components of the magnetic field (the first three cartesian harmonics, Tab. 1). The coils were mostly designed following the method described in Ch. 4, with the exception of the simplification algorithm, which was not ready at the time. Instead, the simplification was carried out manually. The fiducial volume was chosen to be a cuboid centred in the system, with each of its faces 155 mm away from the surface of the coils. The optimal current net for the y-coil (producing a homogeneous field in the y direction) is depicted in Fig. 23 and its manual decomposition into loops in Fig. 24. The design promised a 2 % homogeneity inside the fiducial volume. The current net for x- and z-coils, identical to each other on symmetry grounds, is depicted in Fig. 25. Finally, the individual loops

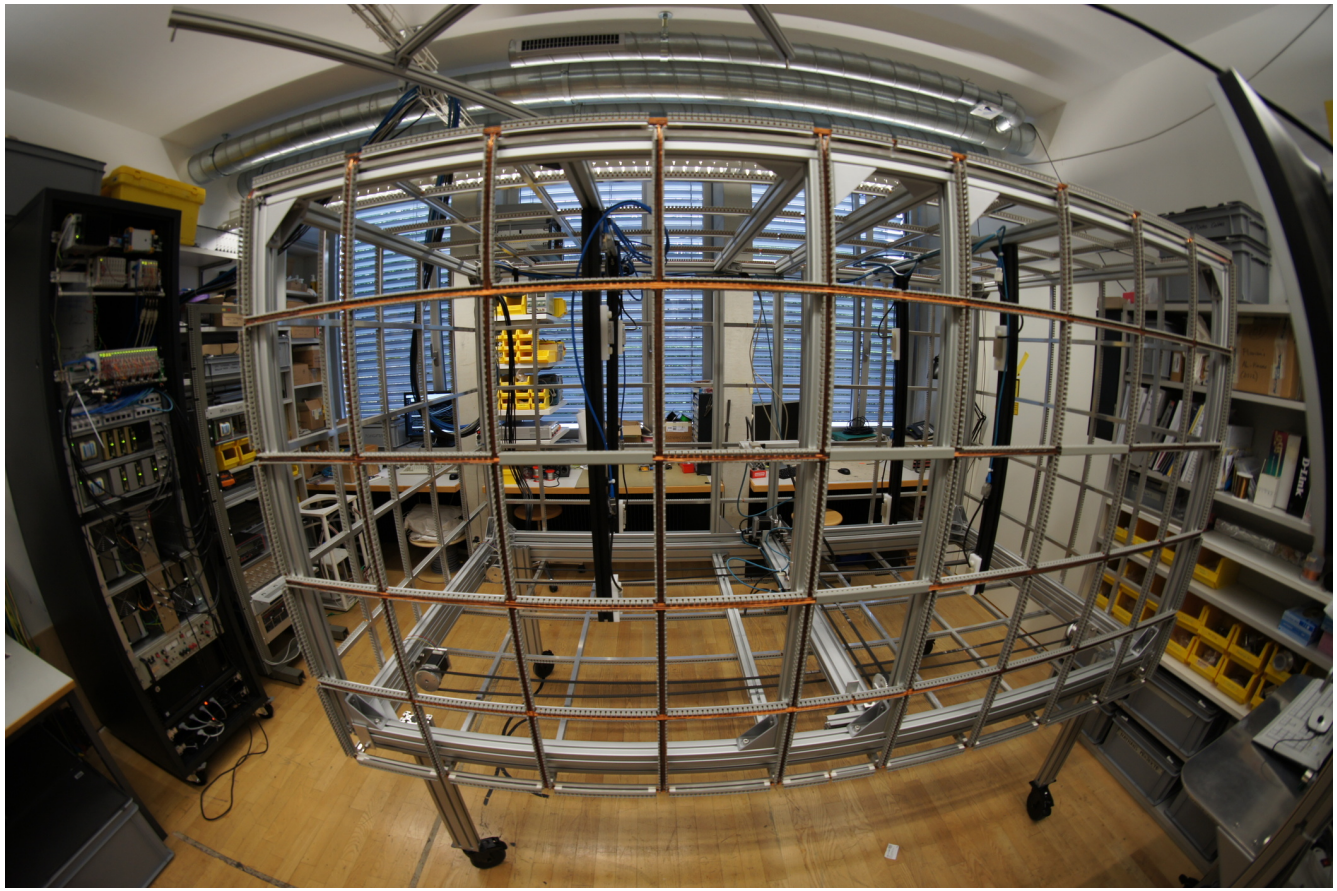

Figure 22: The active magnetic shield in the laboratory at ETH Zürich. In the open cable channels the copper wires making up coils can be seen. The cable channels are held by an aluminum structure. On the left-hand side the control cabinet of the system is visible.

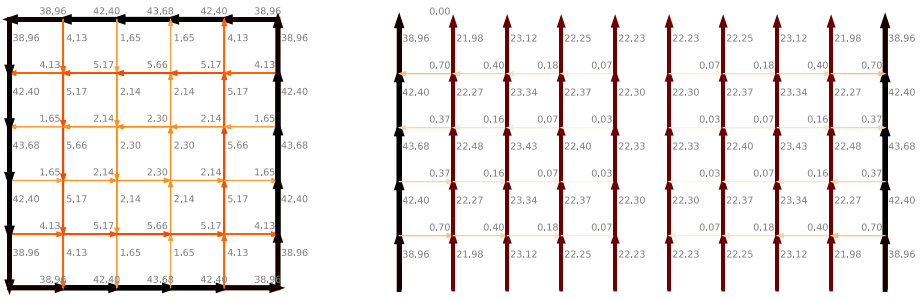

Figure 23: The optimal current net for the y-coil of ETH active magnetic shield. Both 5×5 faces (y = constant planes) are identical and are depicted on the left-hand side. The rectangular faces are all identical, and are depicted on the right-hand side. For each segment the current per 100 µT of generated field is indicated upwards and to the right from its centre.

were discretised into 10 A, 2 A and 0.2 A per nominal 100 µT. For example, the 38.96 A current, indicated in black in Fig. 24, was realised as three windings of the 10 A wire, four of the 2 A wire and five of the 0.2 A wire.

The wire diameter was 1 mm for the 10 and 2 A wires, and 0.8 mm for the 0.2 A one. The wires were prolonged when needed.

Enameled wire was laid in the cable channels according to the discretised design. For one current component of a coil, e.g. 10 A in the x-coil, a single long piece of wire was laid, making up all the windings of all loops. For the three coils, each with three components, nine long wires were laid in total. A close-up of the wires in the cable channels, for all three coils, is shown in Fig. 26.

5.2 MAPPING

The coils were mapped in order to verify that they produced a field of the required homogeneity. To that purpose a robot was built: a fluxgate on an xyz-table, controlled with stepper motors. The robot, called a *mapper*, is pictured in Fig. 27.

A beam seen in the middle-bottom of Fig. 27 could move along the y direction (left-right in the figure), pulled by two timing belts. The belts were wound around pulleys attached to stepper motors, visible to the left. Along the beam (x direction) a cart was moved using the same method. On the cart three vertical rods were attached with a plastic (POM) platform tightly threaded on them. On the platform was a fluxgate magnetic field sensor. Through the centre of the platform was an aluminum threaded rod, passing through a mating threaded hole. The rod was mounted to a stepper motor on the cart. As the motor spun the threaded rod the platform moved vertically along the z direction.

In a fluxgate a ferromagnetic core is periodically driven into saturation. When it is not saturated, it is highly permeable and sucks the external magnetic flux in. When saturated, this does not occur. A pickup coil detects the changes in the external flux as it is alternately sucked in and out of the core.

A simple map consists of moving the fluxgate along one linear direction only. A map of the y-coil along the y direction, in the middle of x and z, is shown at the top of Fig. 28. Only the y-component of the magnetic field is plotted. The coil was set to produce 50 µT. The background field was also mapped and it has been subtracted in the map. The field remained in a ± 0.2 µT range around the average value. In the lower plot of Fig. 28 the maps using just the 10 A component coil, and 10 A and 2 A coils are shown (for 50 µT the currents were 5 A, 1 A and 0.1 A). It is interesting to note how much of the field is produced by

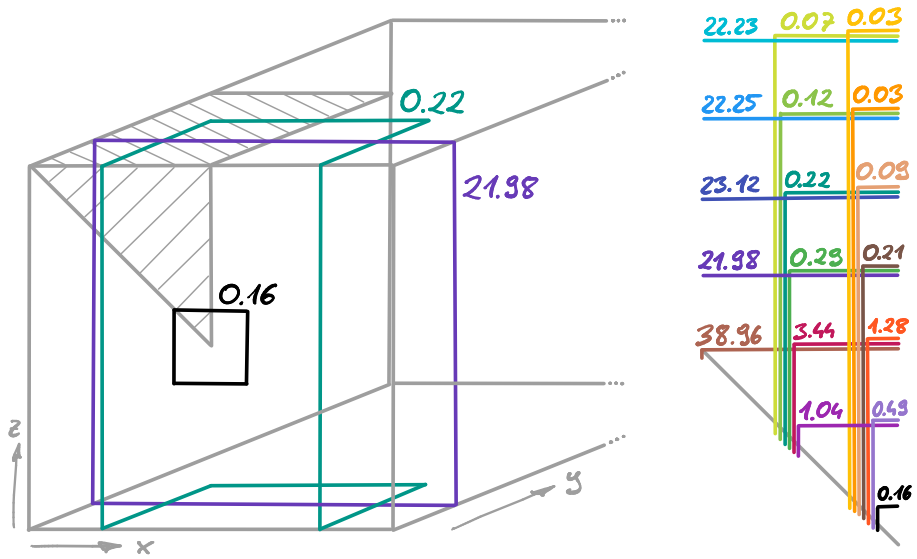

Figure 24: The decomposition of the current net of the y-coil (Fig. 23) into current loops. On the left-hand side only two loops are depicted in an isometric view. On the right-hand side a sixteenth part (dashed on the left) of the system is depicted, all others being identical on the grounds of symmetry. Each loop is indicated with a different colour. The currents are given per 50 µT of generated field.

each of the components. In the middle region the 5 A, 1 A and 0.1 A components produced 86 %, 11 % and 3 % of the 50 µT field, respectively. At the edge the shares change to 70 %, 16 % and 14 %.

Another type of field map was a planar one. A horizontal map in the middle xy-plane, of the y-coil is presented in Fig. 29. The plot shows the maximum deviation among all three components of the magnetic field, i.e. in the area enclosed by the 1 µT isocountour, all components of the field are within 1 µT from the reference field. The reference field would ideally be zero for x and z, and 50 µT for y. In reality the sensor's axis were not perfectly aligned with those of the system, causing the x and z components to be non-zero and the y one to be smaller. The reference field was set as the average field in the area of the highest homogeneity in the middle of the map. Within the fiducial volume the field deviated by no more than 2 % or 1 µT for a 50 µT field, as expected. Maps of the other coils gave similar results.

5.3 CONTROL SYSTEM

In this section the hardware stack used to measure and control the magnetic fields is described. In an active magnetic shield a change in the magnetic field is detected with an array of sensors, and an appropriate response is calculated. The response is then applied by changing the currents in the coils. The feedback loop is closed when the sensors detect the change in the field caused by the coils.

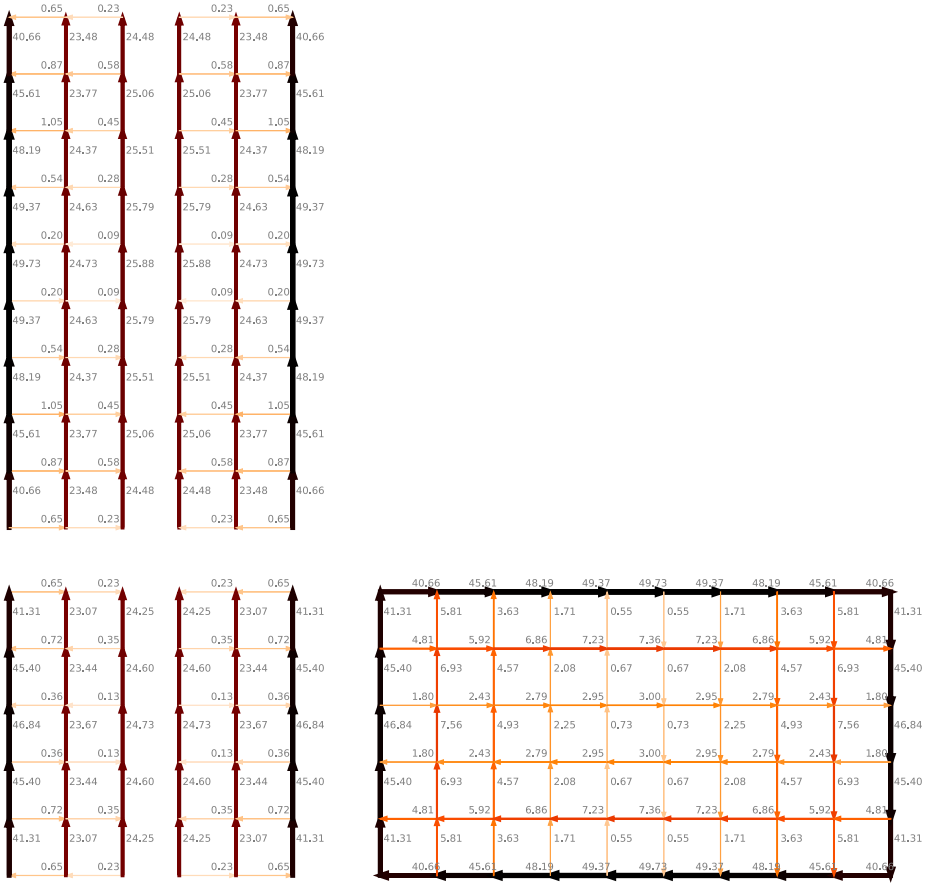

Figure 25: The optimal current net for the x and z-coils of ETH active magnetic shield. Both 5×5 faces (y = constant planes) are identical, and are depicted in the lower-left corner. The rectangular faces perpendicular to the field are depicted to the right, and the ones parallel to it on top. For each segment the current per 100 µT of generated field is indicated upwards and to the right from its centre.

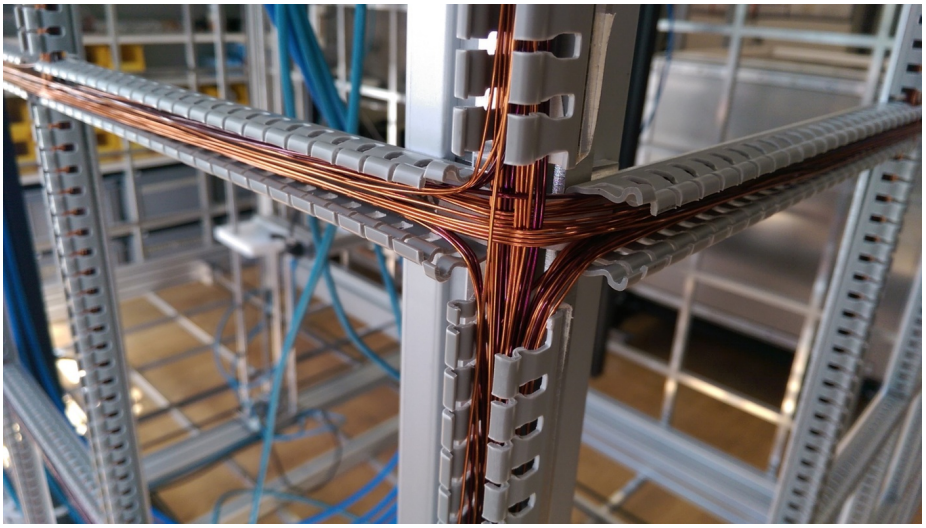

Figure 26: A close-up of the wires in the cable channels. Here all three coils for generating the homogeneous fields (x, y and z) were laid.

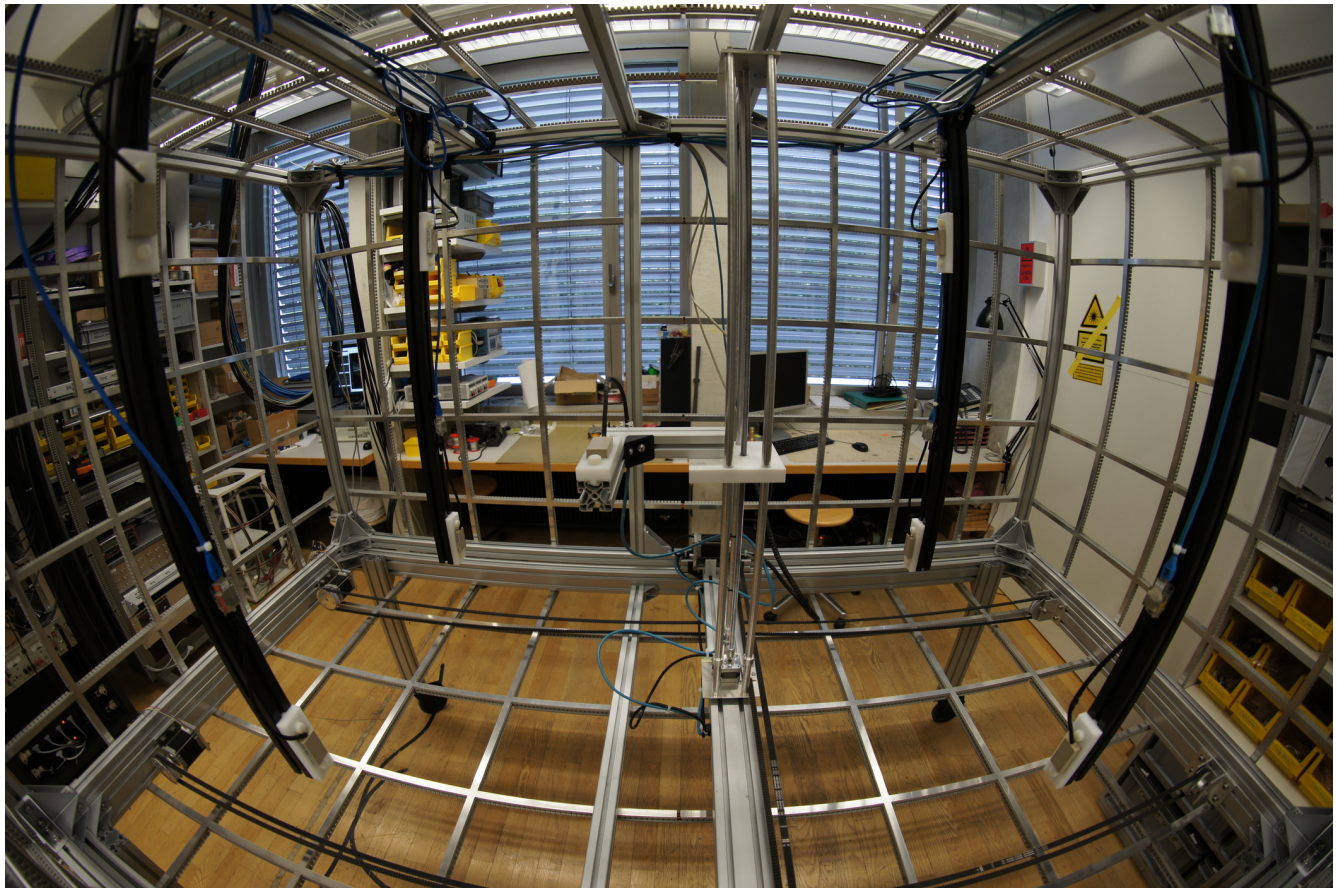

Figure 27: The inside of the active magnetic shield. In the centre the mapping robot (*mapper*) is visible, with a fluxgate magnetic field sensor mounted to a mobile platform. Mounted on black, vertical beams there are eight fluxgates for the active feedback.

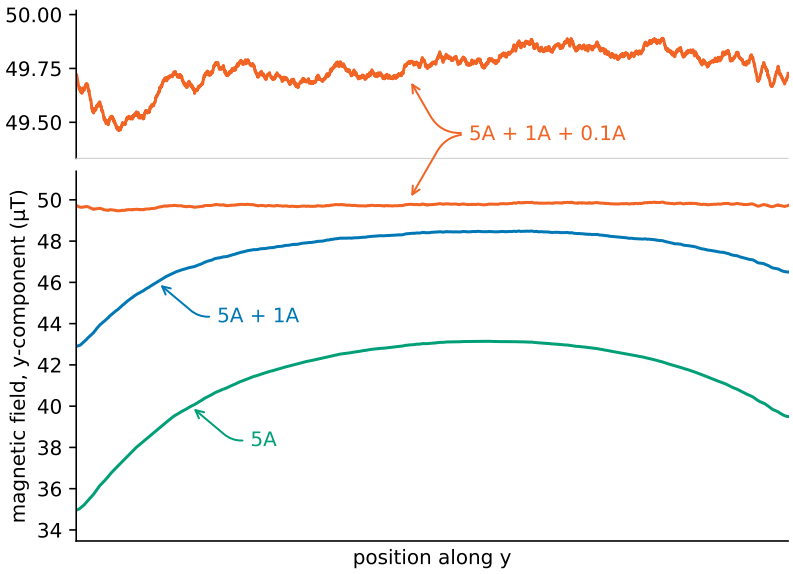

Figure 28: Linear map of the homogeneous field y-coil. The map is along the y-direction, in the middle of x and z. The y-component of the magnetic field is plotted. The green curve depicts the field with only the 5 A/50 µT wire, the blue with the additional 1 A/50 µT wire, and the orange, also zoomed in at the very top, all three wires together.

In the system at ETH the field was measured with eight fluxgates (visible in Fig. 27). They were mounted in a way to make up corners of a cube of approximately 90 cm side length. The sensors were Stefan Mayer Instruments FLC3–70 three-axis fluxgates, ±200 µT range, 1 kHz bandwidth, ±1 % ± 0.5 µT accuracy. They were powered with an in-house built ±5 V power supply. The outputs of the fluxgates were ±5 V signals proportional to the magnetic field. The analogue signals were directly digitised with Beckhoff EL3602 24-bit differential analogue-to-digital converters (ADCs). The digital information was collected in software running on a PC.

The software stack running under OpenSUSE Linux consisted of a low-level Ethercat driver [89], on top of which a custom program written in `Julia` [90] was running. This setup was optimised for high flexibility and close-to-zero turnaround time during the development. In particular, it was possible to develop the software interactively *while* the system was running. The main task of the program was to continuously evaluate the optimal response for the measured magnetic field changes. The response, being the new currents to be applied to the coils, was sent to the digital-to-analogue converters (DACs). The data were recorded and plotted on-line.

The main Julia program published the data on a `ZMQ PUB` socket. The data were stored with a separate `Julia` program, collecting the data on a `ZMQ SUB` socket. Similarly, the plotting program was separate, written in `Python`.

The DACs were 16-bit Beckhoff EL4134. The signals were fed to an array of four-quadrant SERVOWATT amplifiers, configured to regulate the output current with a ±10 V input. For each coil there were three stages: DCP260/30A

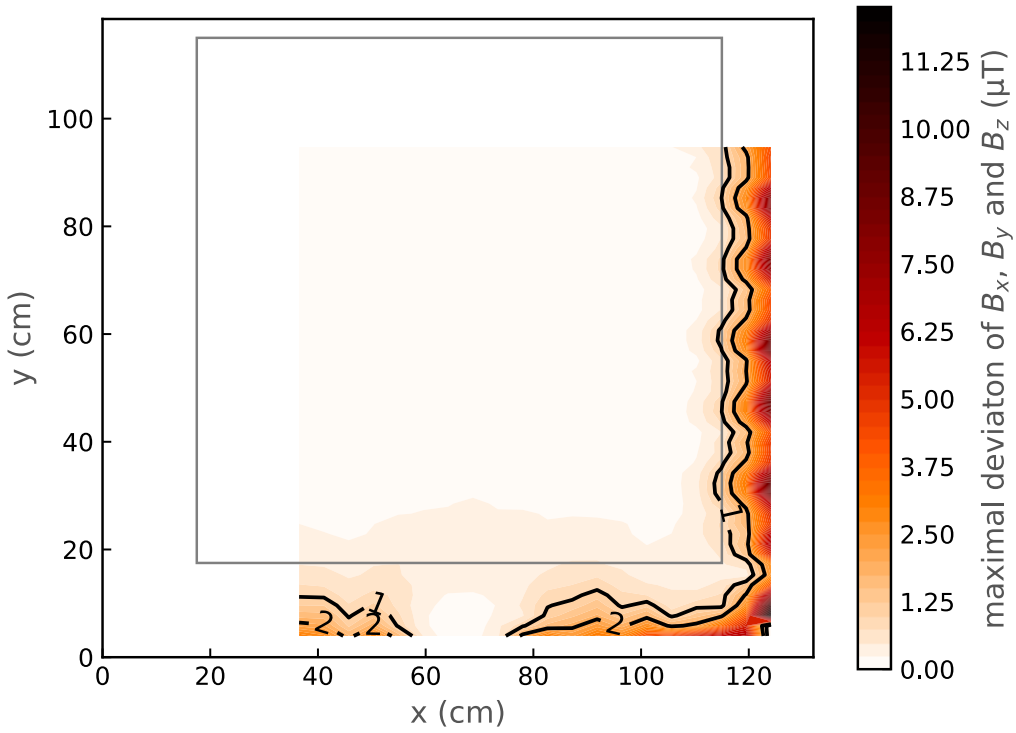

Figure 29: A horizontal map of the y-coil cut along the middle xy-plane. The maximum deviation among all three components of the magnetic field is plotted, i.e. in the area enclosed by the 1 μT isocountour all components of the field are within 1 μT from the nominal field (zero for x and z, 50 μT for y). Only part of the coil was mapped. The border of the plot corresponds to planes where the coil wires are located. The border of the fiducial volume is marked with grey lines.

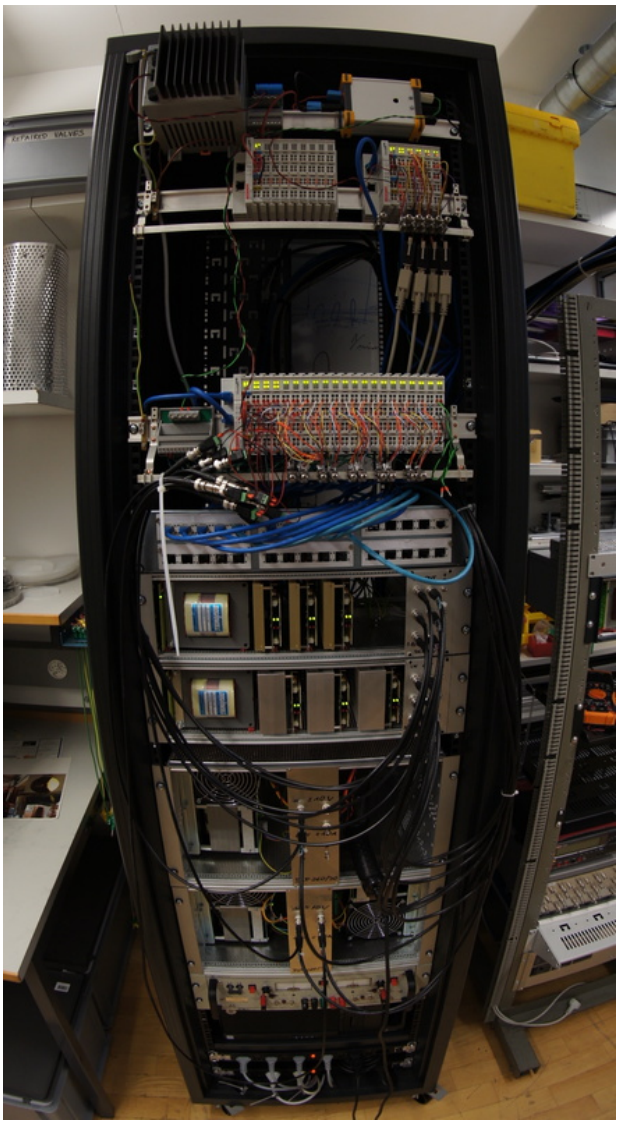

Figure 30: The 19-inch cabinet hosting the electronics for the active magnetic field shield. From top: a 24 V power supply for the Beckhoff EtherCAT clamps and in-house–built ±5 V power supply for the fluxgates; spare EtherCAT clamps and control system for the mapper; a large array of EtherCAT DACs and ADCs for the active stabilisation system; breakout panel for the fluxgate cables (RJ-45 plugs); subrack with ±0.4 A amplifiers; subrack with ±2 A amplifiers; two subracks with ±10 A amplifiers; a general-purpose Kepco ±20 V ±20 A four-quadrant amplifier; the PC.

configured with a ±10 A output, DCE50/30A with ±2 A and DCE10/30A with ±0.4 A. The currents were then directly fed into the coils.

The hardware was hosted in a 19-inch cabinet, pictured in Fig. 30. A detailed list of the components can be found in the figure's caption. All these components were used for the active magnetic field shielding.

5.4 THE FEEDBACK MATRIX

The system was based on a feedback matrix (Ch. 3). The magnetic field measured by each of the 24 sensors is linear with the currents in the three coils. All readouts, gathered in a vector $\mathbb{B}$ (dimension 24) can be written as a linear combination of the currents in the coils $\mathbb{I}$ (dimension 3):

$$\mathbb{B} = M\mathbb{I} + \mathbb{B}_0 \,, \tag{23}$$

where $\mathbb{B}_0$ is the free offset. The matrix of proportionality constants M, dimension 24×3, is the central element of the active shield.

We use the symbol $\mathbb{B}$ for the values measured by the sensors to distinguish it from **B**, *which is used to denote the magnetic field in the whole space.*

Let us recall the discussion of the PSI system's matrix in Sec. 11. There it was pointed out that its high condition number, and the need for regularisation, can be attributed to the geometry of the system. In particular, some linear combinations of the coils created high-order fields which, despite high currents, had little influence on the sensors' readout. PSI SFC matrix was not known a priori. Conversely, the new system already takes the matrix into account at the design phase. The coils were designed to produce a magnetic field corresponding to the terms of the cartesian harmonic expansion of the field; ensuring that they are orthogonal to one another. By design we expect the matrix to have the condition number equal to one. Using the three coils it has been measured to be 1.064 for the homogeneous components, without a mu-metal shield inside.

The active magnetic shield constructed at ETH implemented a new way of measuring the matrix. The procedure changed currents in all the coils simultaneously, shortening the duration of the procedure. It could also measure in close-to-zero field conditions. To measure the matrix, the space spanned by all the coils was considered. In this case it was three-dimensional: the current in the x-coil, in the y-coil and in the z one. A point in this space corresponded to one configuration of the currents in the coils. A set of points on an n-sphere in this space was picked and all currents were changed simultaneously, making their way from one point to another. As the currents changed, the measurements of the sensors were recorded. Then, for each ith readout channel (24 in total) a linear model was fitted to estimate the proportionality constants between the readout and the currents in the coils:

$$B_i = B_i^0 + \sum_{j=x,y,z} M_{i,j}\, I_j \,, \tag{24}$$

where B_i is the readout channel (i going from 1 to 24), I_j are the currents in the coils (j going over x, y and z) and $M_{i,j}$ is the matrix of proportionality constants—the feedback matrix. B_i^0 is the free offset vector—the background field.

It is possible to fit the linear model and estimate the uncertainty of the matrix elements while *measuring; the measurement can be stopped when the uncertainty drops below a threshold.*

If the system is perfectly linear it does not matter where in the current-space the n-sphere is located, nor how large it is (though larger ones give more precise estimates per point). However, we expect non-linearities to appear in the presence of a μ-metal shield inside the system. In that case it is preferable to measure the matrix in the vicinity of the point where the system is later operated. Motivated by this, the following procedure of the matrix measurement was employed: First an n-sphere centred at n-zero and a radius of 50 µT was chosen. Then the system walked over ten random points on the sphere, taking two seconds to pass from one to the next (20 s in total). This gave the first estimate of the matrix. The next sphere was chosen to be centred around the zero field. This is the optimal, in the least-squares sense, solution of Eq. 23 with $\mathbb{B} = 0$:

$$\mathbb{I}_{\text{zero-field}} = -M^\dagger\, \mathbb{B}_0 \,, \tag{25}$$

where $\mathbb{B}_0$ is the free offset from Eq. 24, and $M^\dagger$ is the Moore-Penrose pseudoinverse of the matrix M. This gave the second estimate of the matrix. Finally, the

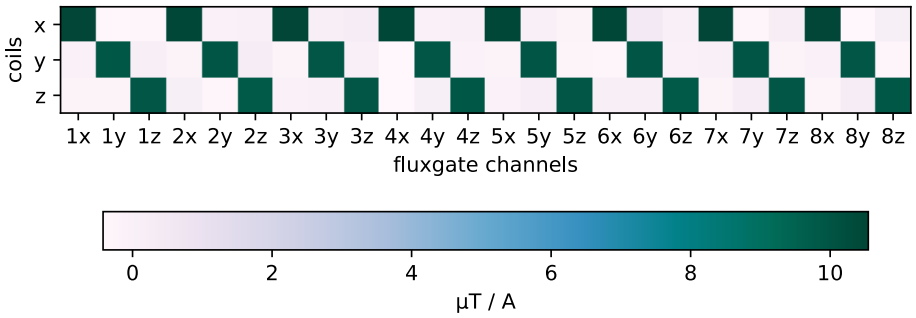

Figure 31: The measured feedback matrix of the active magnetic shielding system. The sensors were mounted $\approx 20\,\mathrm{cm}$ away from the coil surface. The values are microteslas of field per 1 A in the strongest coil (0.2 A and 0.02 A in the other two).

measurement was repeated with a sphere again centred at zero, but with a smaller radius of 1 µT. In the case of an air system, without a mu-metal shield inside, all consecutive estimates of the matrix were almost the same.

The matrix measured with the sensors $\approx 20\,\mathrm{cm}$ away from the coil surface is presented in Fig. 31. The differences between the non-zero elements of different fluxgates are under 1%, as expected from the design and the field map (Fig. 29). The condition number of the matrix is 1.064.

5.5 THE FEEDBACK ALGORITHM

The feedback was based on calculating the optimal solution of Eq. 23 in each step. In the nth iteration (the iteration index marked on top) the equation is

$$\mathbb{B}^n = M\mathbb{I}^n + \mathbb{B}_0^n \,. \tag{26}$$

In the next iteration the field equal to the target field is sought

$$\mathbb{B}^{n+1} \stackrel{!}{=} \mathbb{B}_{\text{target}} \,. \tag{27}$$

We allow $\mathbb{B}_{\text{target}}$ to be any field. This leads to the following requirement for the next currents:

$$\begin{aligned}
\mathbb{I}^{n+1} &= M^\dagger \left(\mathbb{B}_{\text{target}} - \mathbb{B}_0^{n+1}\right) \\
&\approx M^\dagger \left(\mathbb{B}_{\text{target}} - \mathbb{B}_0^{n}\right) \\
&= M^\dagger \left(\mathbb{B}_{\text{target}} - \mathbb{B}^{n} + M\mathbb{I}^n\right) \\
&= M^\dagger \left(\mathbb{B}_{\text{target}} - \mathbb{B}^{n}\right) + M^\dagger M\mathbb{I}^n \\
&= M^\dagger \left(\mathbb{B}_{\text{target}} - \mathbb{B}^{n}\right) + \mathbb{I}^n \,.
\end{aligned} \tag{28}$$

By defining $\Delta\mathbb{I}^n := \mathbb{I}^{n+1} - \mathbb{I}^n$ and $\Delta\mathbb{B}^n = \mathbb{B}^n - \mathbb{B}_{\text{target}}$ we obtain the intuitive rule for the current update

$$\Delta\mathbb{I}^n = -M^\dagger \Delta\mathbb{B}^n \,. \tag{29}$$

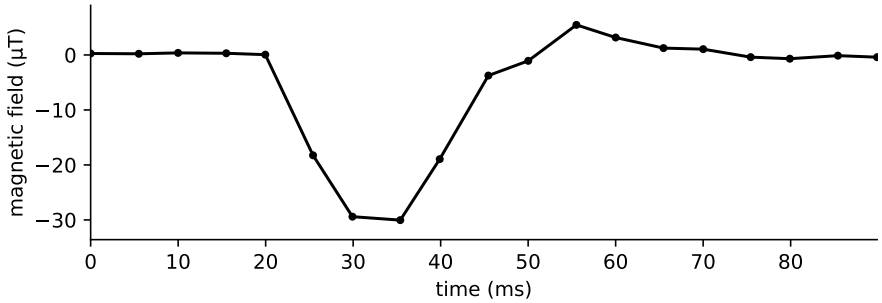

Figure 32: The reaction of the active magnetic field compensation to a step change in the magnetic field ($\approx 30\,\mu T$ in 5 ms). The system was running at a rate of 200 Hz. Each dot marks one iteration. Only one measurement channel is plotted.

A careful reader may be alarmed by the approximation $\mathbb{B}_0^{n+1} \approx \mathbb{B}_0^n$ in Eq. 28. It is an unfortunate necessity, for the calculation needs to be performed just before the iteration $n+1$, when $\mathbb{B}_0^{n+1}$ is not yet known. In other words, the correction needs to be delayed by at least one iteration. This introduces lag, and motivates high feedback frequencies.

In practice, the active shield had a delay of more than one iteration. Although the system was operated at 200 Hz, the quickest turnaround was three cycles (15 ms). This was tested by applying a pulse on a DAC channel and observing the response on a directly connected ADC channel. It appeared only in the third iteration after the pulse had been sent. Knowing that the magnetic field information is delayed, it was crucial to delay the current information too, so Eq. 28 becomes:

$$\mathbb{I}^{n+1} = M^\dagger \left(\mathbb{B}_{\text{target}} - \mathbb{B}^{n-2} \right) + \mathbb{I}^{n-2} \,. \tag{30}$$

Without accounting for the delay the system would spontaneously destabilise. In Fig. 32 a response of the system to a step-like change of the magnetic field is plotted. Note how the system's reaction is delayed by three iterations.

5.6 DYNAMIC STABILISATION

The dynamic stabilisation was tested with a strong permanent magnetic dipole. The dipole was built out of two extremely strong neodymium magnets (200 kg force when attached to iron) connected together with a 1 m long iron rod. The dipole was moved several meters from the system causing a magnetic field disturbance.

The resulting field compensation from the dipole being rotated 7.5 m away from the system's centre are plotted in Fig. 33. The upper part of the plot shows the field as it would be registered without compensation $\mathbb{B}_0$, calculated according to Eq. 23:

$$\mathbb{B}_0 = \mathbb{B} - \mathbb{M}\mathbb{I} \,. \tag{31}$$

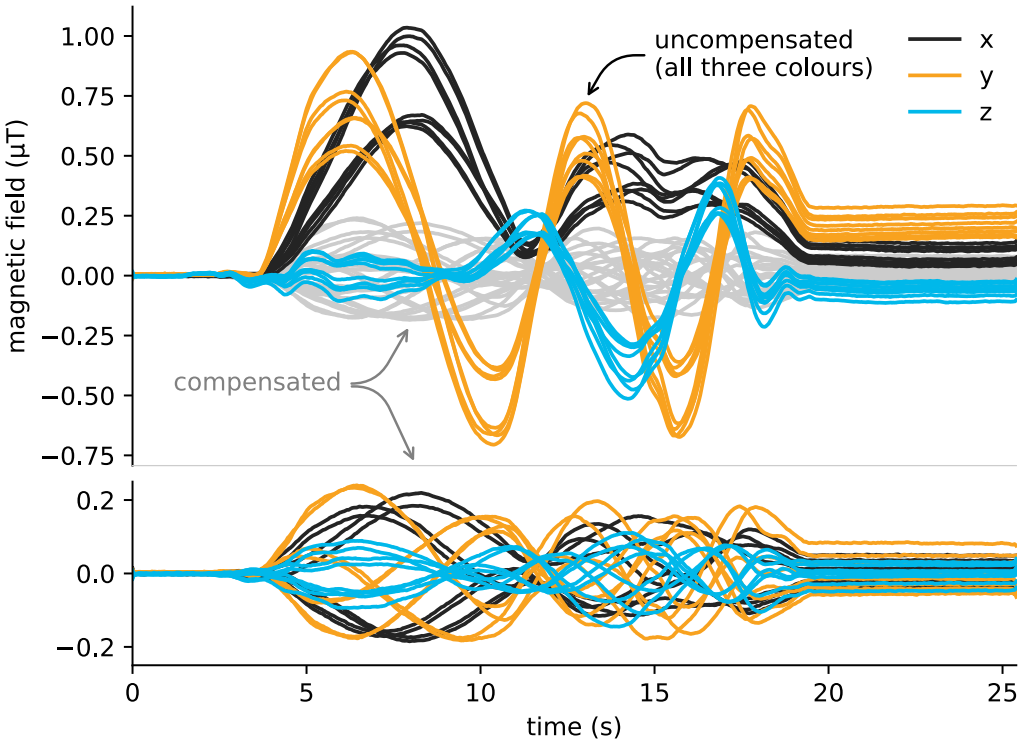

Figure 33: Field variations caused by a rotating dipole 7.5 m away. The upper part shows in colours corresponding to the spatial directions the uncompensated field $\mathbb{B}_0$ for all eight sensors, calculated according to Eq. 31. Each line is one sensor. The compensated field $\mathbb{B}$ is depicted in grey in the upper part and is enlarged in the lower part of the plot in colour. The curves have been smoothed for clarity.

The measured field $\mathbb{B}$ is depicted in the bottom part. The amplitude of the changes was reduced from around 1.75 µT to 0.4 µT (factor four). Yet, the variation is not completely mitigated.

Looking at the orange lines in the upper part of Fig. 33, each representing the readout of a different sensor in the y direction. The lines do not overlap, meaning that the field change was not homogeneous (which, of course, would be the same everywhere). The coils of the system, being able to generate only homogeneous fields can only compensate the homogeneous part. Graphically it may be explained in the following way: each of the coils can keep all lines in one of the colours (one spatial component) steady, but it cannot bring them closer together (homogenise the field). This can be observed in the bottom part of Fig. 33. There lines are centred around zero, but their spread within one colour (spatial component) is not reduced. In order to do that, coils producing higher-order fields would be required.

The amplitude of the variations seen in Fig. 33 can be quantified as the difference between 75th and 25th quantiles of all the measurement points projected on the y-axis. This measure is plotted in the upper part of Fig. 34 as the function of the distance between the dipole disturbance and the centre of the system. The uppermost black curve is the uncompensated field $\mathbb{B}_0$. The orange curve below is the compensated field seen by the feedback sensors $\mathbb{B}$ (the bottom part of Fig. 33). We observe that the compensation improved with the distance to the dipole, as the field became weaker and more homogeneous. In the bottom of the figure the ratio of the two curves, called the *shielding factor*, is plotted in orange.

The next curve (blue) corresponds to the field measured by a sensor placed in the middle of the system. There the system compensates first-order changes (because they are antisymmetric with respect to the centre). It suggests that the difference to the shielding factor for the feedback sensors (factor three) can be attributed to the uncompensated variations of first-order gradients. Furthermore, it suggests that if the system would be extended by a addition of first-order coils, the shielding factor for the feedback sensors (in the whole volume) would be as good as for the centre.

The last curve (green) depicts the variations seen in the average readout of the feedback fluxgates. An ideal system should compensate it perfectly, regardless of the shape of the field changes. Yet, the shielding factor for this measure was around 50. It also did not depend on the distance to the dipole (homogeneity of the field), suggesting that it is a property of the compensation system itself. The dominant factor is probably the inhomogeneity of the field created by the compensation coils (around 1–2 %). Even if the changes of the field were similar in magnitude and pace to the ones discussed, but perfectly homogeneous, this system would not be likely to compensate them better than a factor of 50.

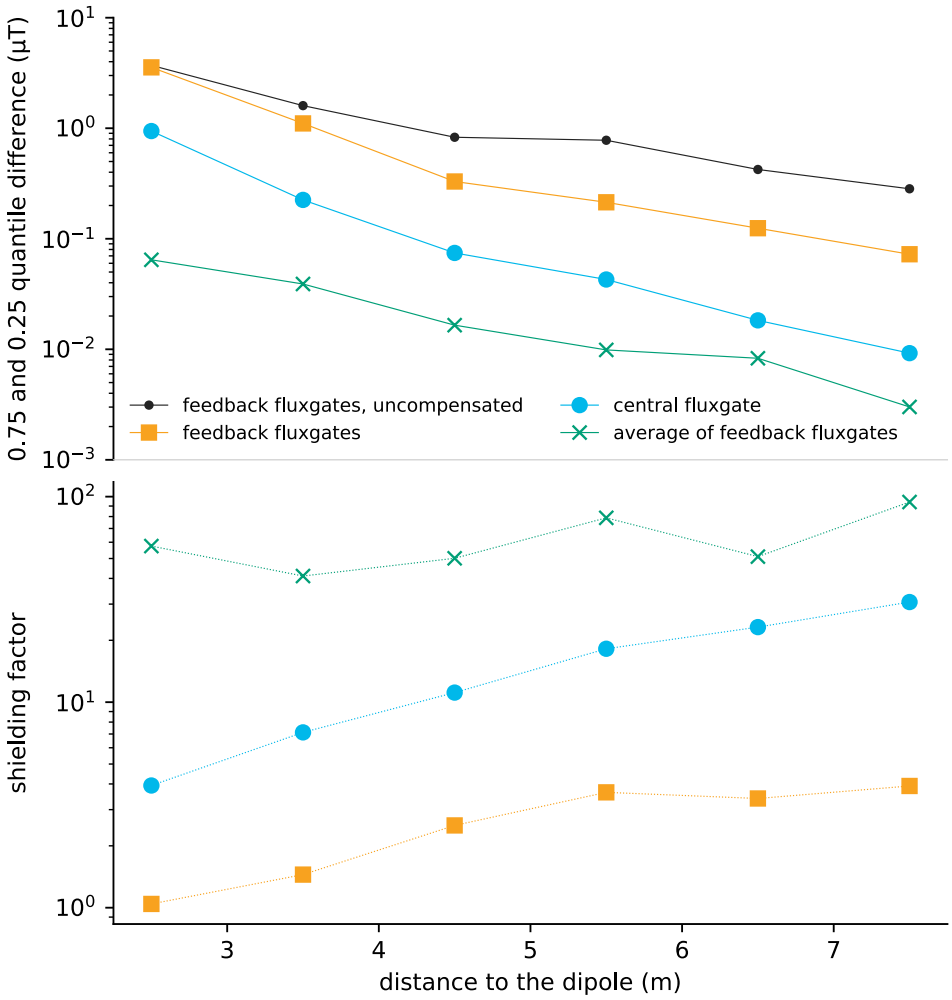

Figure 34: Performance of the active shield compensating a dipole disturbance, as a function of the distance away. *Upper part:* The amplitude of the field variations measured as the difference between the 75th and 25th quantiles of all values. Plotted are the variations of: the uncompensated field (black), the compensated field measured by the feedback sensors (orange), the field measured by a non-feedback sensor in the centre (blue) and the time series of the average readout of the feedback sensors (green). *Lower part:* The shielding factor, defined as the ratio of the curves in the plot above to the uncompensated variation.

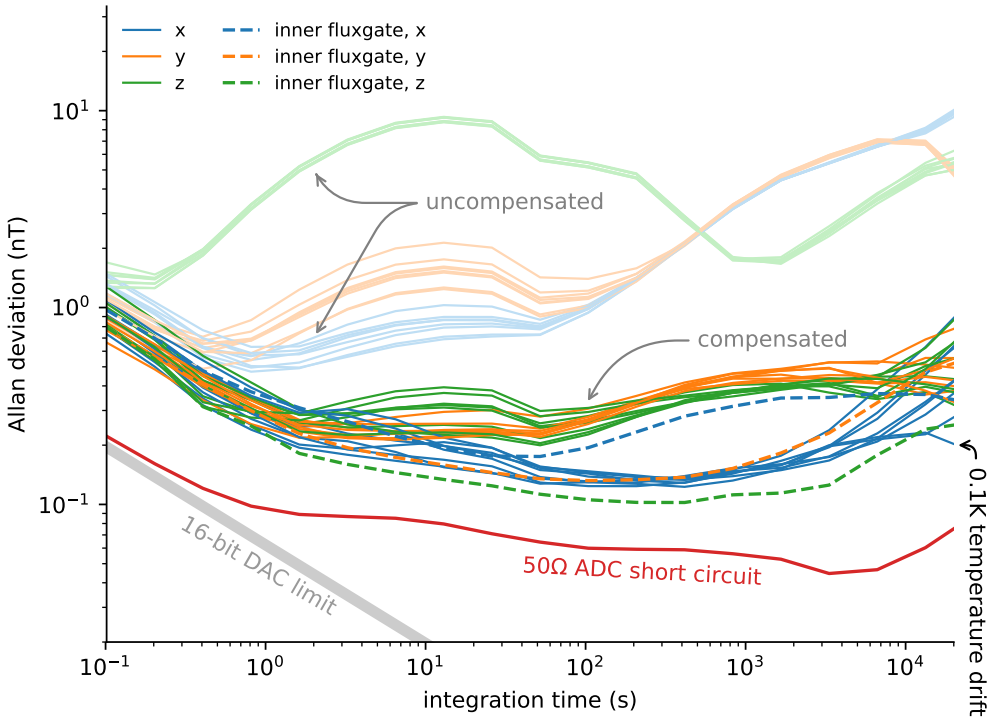

Figure 35: The stability of the active magnetic field compensation. The Allan deviation is plotted as the function of integration time. The three colours depict the three spatial directions. The uppermost three groups of thin, solid lines depict the uncompensated field $\mathbb{B}_0$ measured by the feedback sensors. In the dense group of lines below, around 0.3 nT (corresponding to a temperature stability of 0.1 K, as specified for the fluxgates), are the compensated field measured by the feedback sensors $\mathbb{B}$ (solid) and the compensated field measured by a non-feedback fluxgate in the centre of the system (dashed). Below that is the stability of the readout of a short-cut ADC channel (red) and the limit set by the quantisation noise of the 16-bit DACs (grey).

5.7 LONG-TERM STABILITY

In the PSI environment periods of strong magnetic field changes (tens of microteslas over an hour when nearby magnets ramp) were interleaved with ones of high stability—10 nT at 10 s at night (Fig. 5.3 in Ref. [61]).

We have just discussed how the active shield can stabilise the magnetic field in the case of strong variations. However, the system's internal stability was inevitably finite. In conditions where the environmental magnetic field would be more stable than that, the system would effectively destabilise the field. In this section we will discuss where the limit of the stability lies.

As the measure of stability we use the *Allan deviation*, a special case of the M-sample variance, defined originally by the Eq. 11 in Ref. [91], with $N = 2$ and $T = \tau$. It has a simple interpretation; for a given integration time τ it calculates the RMS variation from one integrated sample to the next.

To assess the stability the system ran overnight, with no known activity in the immediate surrounding of the laboratory. The resultant Allan deviation is plotted in Fig. 35. The three uppermost groups of curves depict the stability of the uncompensated field $\mathbb{B}_0$ (as calculated with Eq. 31), at around 1 nT for x and

y, and 10 nT for z. That the stability differed between the spatial components is not a concern—there is no reason to expect the disturbance to be isotropic. Below, at around 0.3 nT over a wide range of integration time, is a large group of curves. Among them are the thick, solid lines depicting the compensated field $\mathbb{B}$. Even in the most quiet of the conditions in the laboratory the system improved the stability of the field by a factor of two (x) to thirty (z).

In the previous section it was indicated that in a case of high-order variations, the field in the centre of the system was stabilised better then the one measured by the feedback sensors. In the Allan standard deviation plot the stability of the sensor in the centre is depicted with dashed lines; they lie in the large group around 0.3 nT, suggesting that the variations during the measurement were homogeneous.

Why was the improvement not better? The limit came most likely from the temperature drifts. According to the specification of the Stefan Mayer Instruments sensor (FLC3–70), thermal drifts of $2\,\mathrm{nT\,K^{-1}}$ can be expected, meaning the the observed stability corresponds to temperature stability as small as 0.1 K. Even at night the temperature in a non-temperature-stabilised laboratory cannot be expected to be more stable than this. This limit can be pushed a factor twenty lower with higher-quality (and more expensive) sensors. Commercially available Stefan Mayer Instruments FL1–100 and Bartington Mag-03 are specified to drift $0.1\,\mathrm{nT\,K^{-1}}$ with a $\pm 100\,\mu\mathrm{T}$ measuring range.

However, there is an other limitation. Figure 35 features a red line labeled "50 Ω ADC short circuit". This is the stability of the readout of an ADC channel with the terminals connected with a 50 Ω resistor. Even if the magnetic field sensors would have output a perfectly stable voltage, the digitised information would be no more stable than this. A higher-class digitisers could perform better.

Interestingly, the stability was fully defined by the measurement chain: the sensors and the digitisers. Instabilities of the output chain, DACs and amplifiers, indistinguishable from changes in the magnetic field, were corrected by the system itself. One exception is the discrete nature of the currents that can be applied. The lowest curve on the plot is the limit on the stability due to the bit depth of the DACs. A quantisation resolution Δ corresponds to white noise with an RMS amplitude $\Delta/\sqrt{12}$ (derived for example in Sec. IV.A in Ref. [92]). This noise then scales down as $\tau^{-1/2}$ with the integration time. The system's 16-bit DACs had 2^{16} levels mapped onto a $\pm 100\,\mu\mathrm{T}$ range. This defines the quantisation at the feedback time of inverse 200 Hz. In total the limit from the quantisation at the integration time τ is

$$\frac{200\,\mu\mathrm{T}}{2^{16}\,\sqrt{12}\,\sqrt{200\,\mathrm{Hz}\,\tau}}\,. \tag{32}$$

Aside from the obvious—increasing the bit depth of the DACs—this limit can be pushed further by increasing the feedback frequency or decreasing the range of the operation.

5.8 OPEN-DESIGN CAGE

The disadvantage of the first-iteration design was that the cage was fully closed, making installation or removal of large items inside essentially impossible. In particular, the active magnetic shield at ETH could, in future, be used in combination with a mu-metal shield inside it.

The solution was to modify the system in way that one of the faces would hold no wires at all. As neither the coil design method (Ch. 4) nor its implementation are restricted to regular grids, it could be fully realised within the framework. In the scope of this work coils corresponding to the cartesian harmonic polynomials $n = 1, \ldots, 8$ (Tab. 1) were designed and the y-coil ($n = 2$) was wound and mapped. The face that was left open is square (the one to the left in Fig. 22).

In the whole fiducial volume the homogeneity was predicted to be stricly better than 4 %.

In Fig. 36 both simulations and maps of the field of the y coil are shown. In these simulations the homogeneity was predicted to be 2 % (1 µT in the 50 µT field in the figure) in the $97.5 \times 97.5 \times 97.5$ cm volume (depicted in grey the figure). This volume would be occupied by a cubic mu-metal shield intended to be put in the system. The measured field, mapped as described in Sec. 5.2, confirms that the homogeneity could be achieved in practice. Figure 37 details the measured field along the orange lines marked in Fig. 36.

The remaining seven coils were designed. Winding them, however, was beyond the scope of this work. The simulated field produced by the optimal designs for the $n = 1$ (x-coil) and the $n = 5$ coil (Tab. 1) are presented in Fig. 38. Despite lack of one face in the grid, the field is reproduced in a volume large enough to fit the mu-metal cube, even in the case of the linear-gradient coil. In Fig. 39 the first 22 (out of 64) loops of the x-coil are depicted. Note in particular the high density of the cables along the edges of the open face.

The open-design y-coil was a significant step forward from the closed-cage design. It has been demonstrated that the coil design framework is capable of handling irregular grids, and the designs can be successfully realised in practice. While the prototype at ETH could be further extended by winding the remaining seven coils, the positive results of the y-coil already supported the case for designing a system for the n2EDM experiment.

5.9 N2EDM DESIGN

Let us recall the discussion in Sec. 3.6 on the design challenges of an active magnetic field compensation system for the n2EDM experiment. The main point of concern was the spatial constraints. Firstly, there would not be much space available around the mu-metal shield, calling for a coil system with a large fiducial volume. Secondly, only an irregular cage would avoid collisions with other components of the experiment and facilities in the hall.

A cage for the compensation coils was incorporated in an existing CAD model of the experiment. It was composed out of rectangles around 1.5 m large. Care has been taken to avoid conflicts with other parts of the apparatus. The

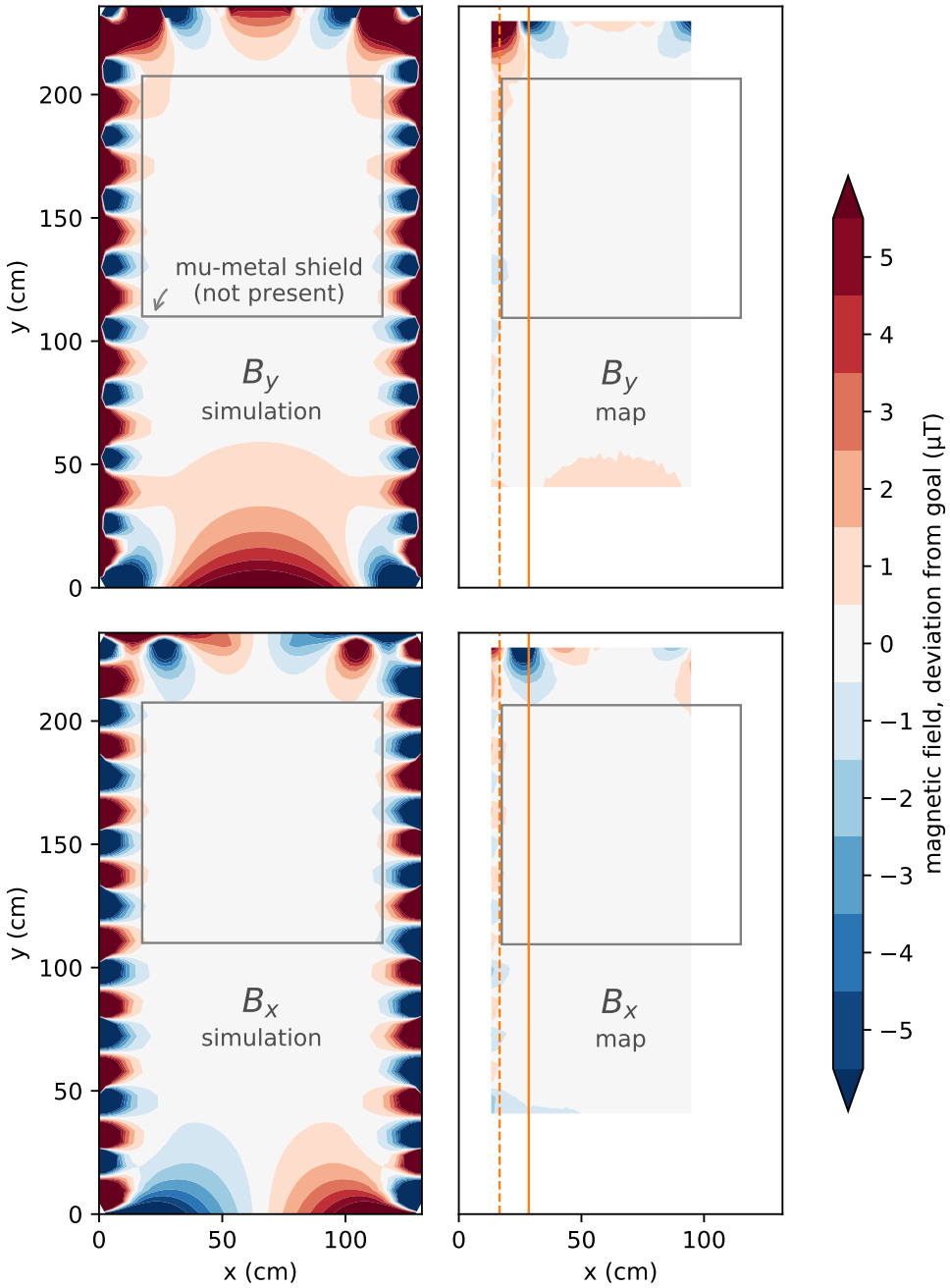

Figure 36: The simulations (left column) and maps (right column) of the field of the y-coil. In the bottom row the x component of the field is shown, in the top row: the y component. The latter relative to the goal value of 50 µT. The contour of a mu-metal shield intended to be put in the system is depicted. To account for misalignments of the sensor, the map has been normalised to the average field measured in the middle region of highest homogeneity. The mapped field along the vertical orange lines is plotted in Fig. 37.

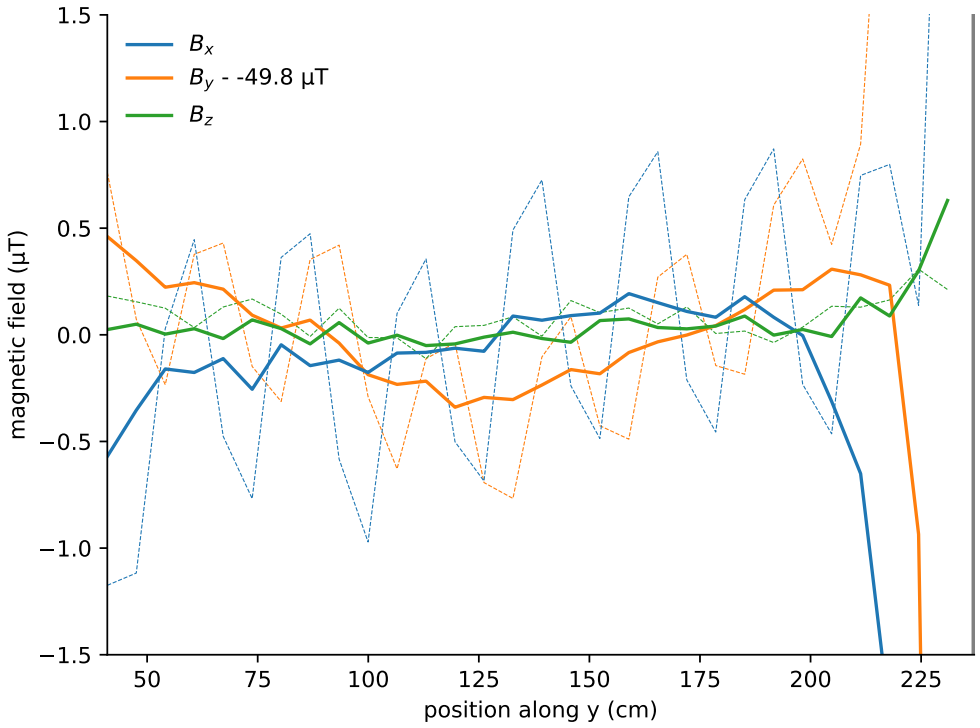

Figure 37: The measured field of the y-coil along the solid and dashed lines depicted in Fig. 36. The thin, dashed lines depict the field 16.6 cm away from the surface of the coils in the x direction (the fiducial volume was 15.5 cm away). The thick, solid lines correspond to a distance of 28.5 cm away. The vertical line on the right depicts the surface of the coils along y.

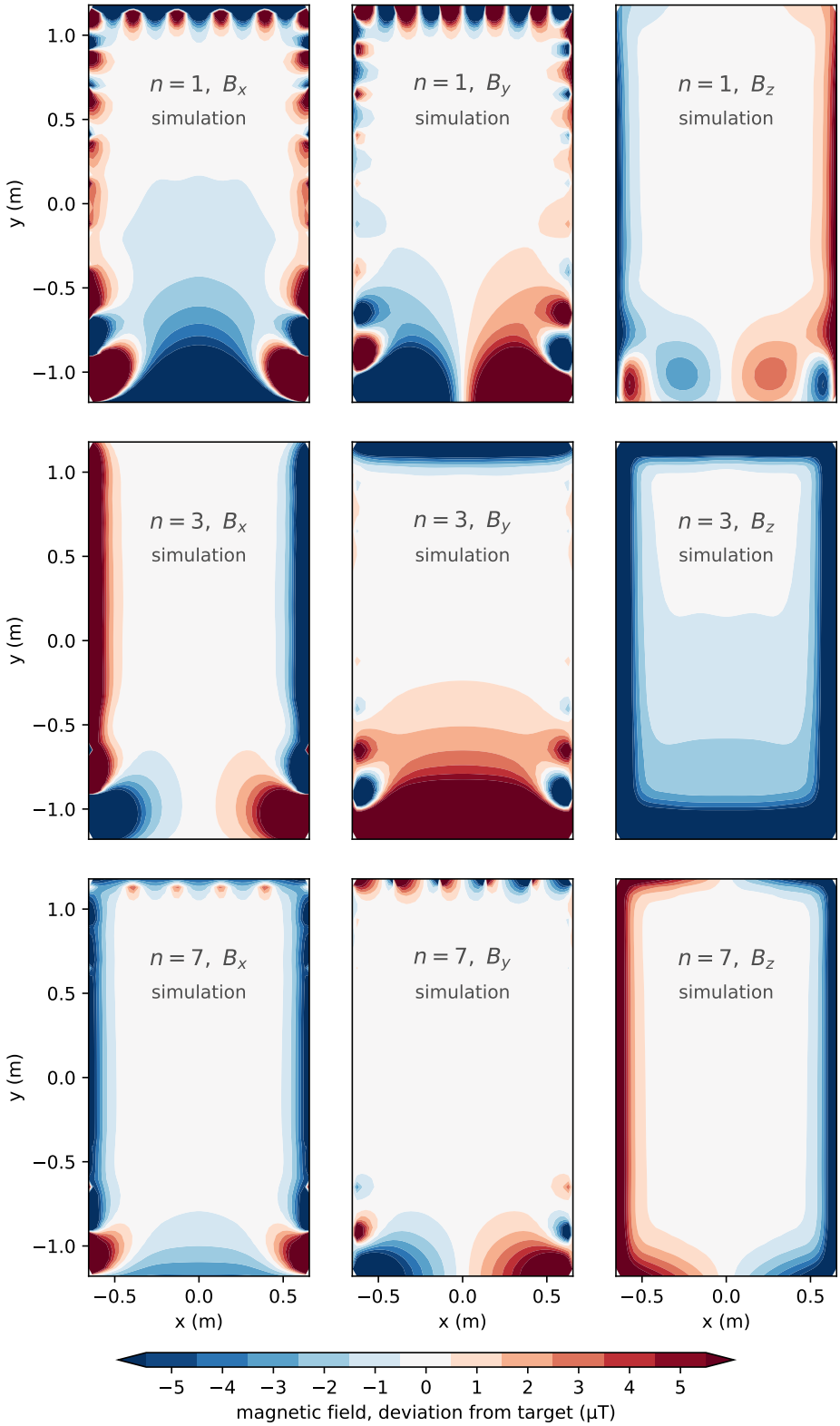

Figure 38: Simulation of the open-design cage for a 50 μT $n = 1$ coil (x-coil, top row, cf. Tab. 1), a 50 μT $n = 3$ coil (z-coil, middle row) and a 20 μT/m $n = 7$ one (linear gradient, bottom row). The colour depicts the deviation of the coil's field x, y and z components from the target field. The section in the XY plane at height $z = 0.3$ m is shown. The open face is at the bottom of the plots. The borders of the plots correspond to the surface on which the coils are wound.

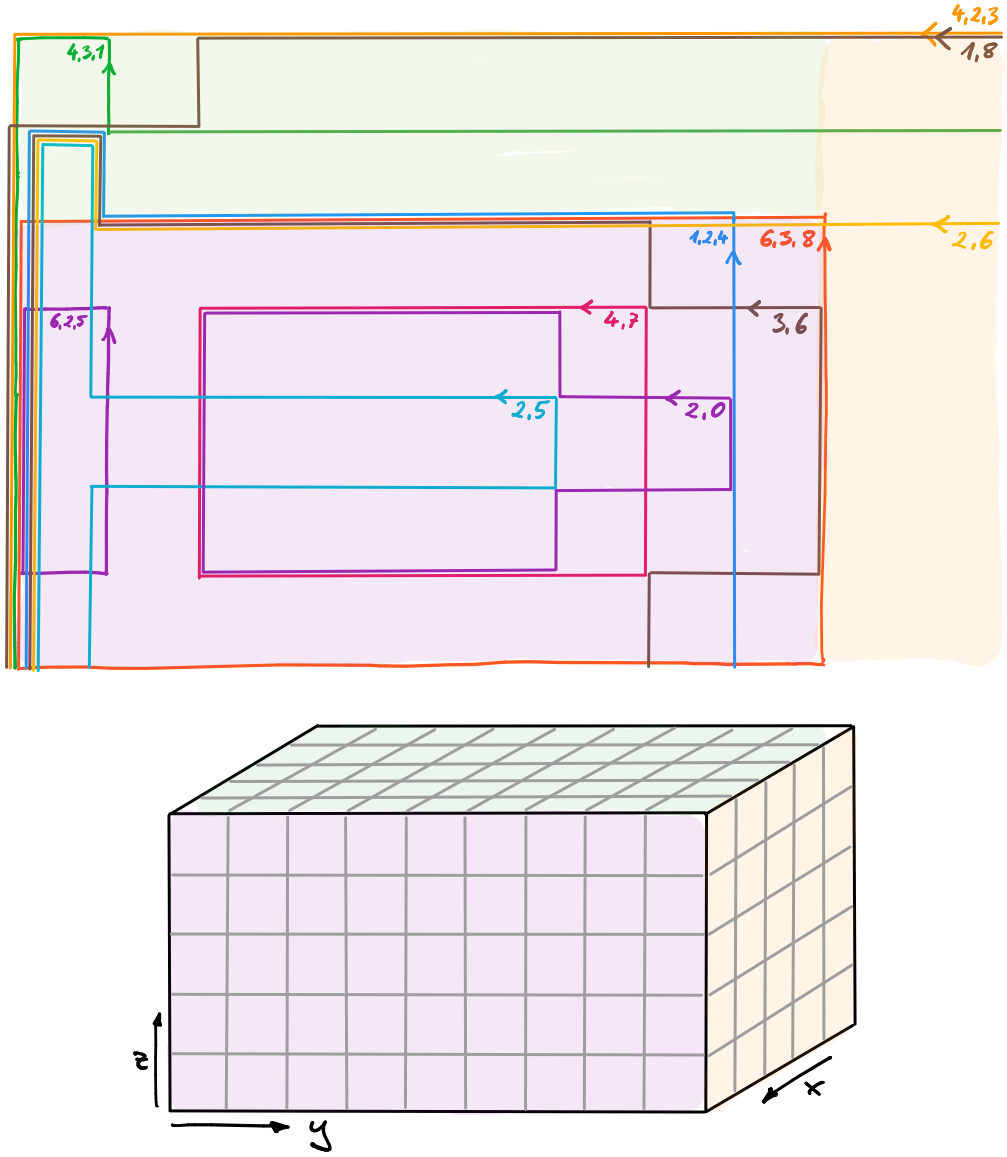

Figure 39: The first 22 (out of 64) loops of the x-coil in the open-design cage. For each loop the digits indicate the number of 5 A, 1 A and 0.1 A windings (per 50 µT field, leading zeros are omitted). Half of the cage is flattened out, the other being identical on symmetry grounds. The open face is located to the left. The faces have been coloured to help with orientation.

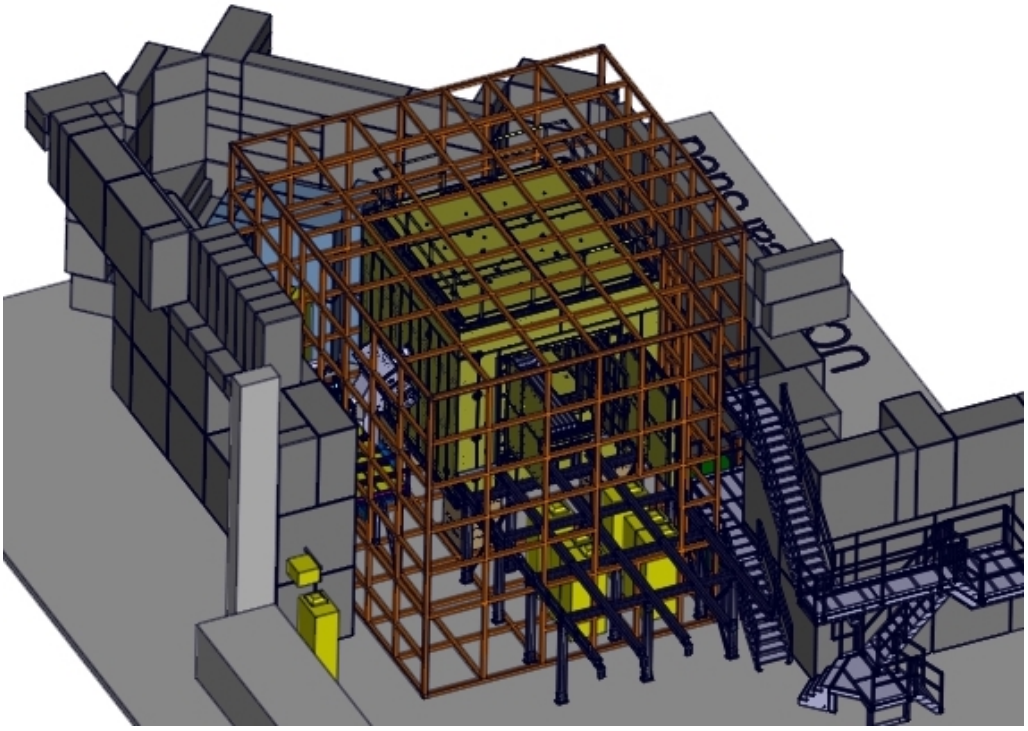

Figure 40: CAD model of the n2EDM apparatus. The passive mu-metal shield is depicted in yellow (with its door open), the proposed cage of an active shield in orange.

cage is rendered in 3D in Fig. 40. In Fig. 41 each face is depicted separately. The model was then implemented in the coil-design framework.

A set of coils has been designed for the first eight cartesian harmonics (Tab. 1). The field simulated for a $n = 1$ homogeneous field coil and a $n = 6$ linear gradient are depicted in Fig. 42. The contour of the n2EDM mu-metal shield is depicted in grey. The quality of the field, as simulated on the surface of the mu-metal shield, is plotted in Fig. 43. The deviation from the pure harmonics is shown. The magnitude of the fields was chosen to be roughly the expected compensation values. When canceling a perfectly homogeneous 50 µT field the remnant is expected to be $< 1\,\mu T$ (which corresponds to 2 %). In the case of a $20\,\mu T\,m^{-1}$ linear gradient the remnant would be $< 6\,\mu T$ (6 % for the 5 m large shield).

The magnitudes of homogeneous fields and gradients have different units. They can only be compared given a characteristic length, in this case the size of the active shield.

Remnant fields this small are unlikely to be a limiting factor in the dynamic stabilisation; recall the discussion at the end of Sec. 5.6, in particular Fig. 34. The ability of the prototype to actively shield was limited by the inhomogeneity of the field changes. The homogeneity in the zeroth order coils for the n2EDM active shield's design was comparable to the ones of the prototype: 1–2 %. Therefore, the limit in the shielding factor was expected to be similar at around 50.

In the design the limiting factor was identified to be the inability of the active shield's coils to reproduce the field changes. It was then crucial to characterise the changes in the experimental site, so that appropriate coils could be built. The characterisation is the topic of the next chapter.

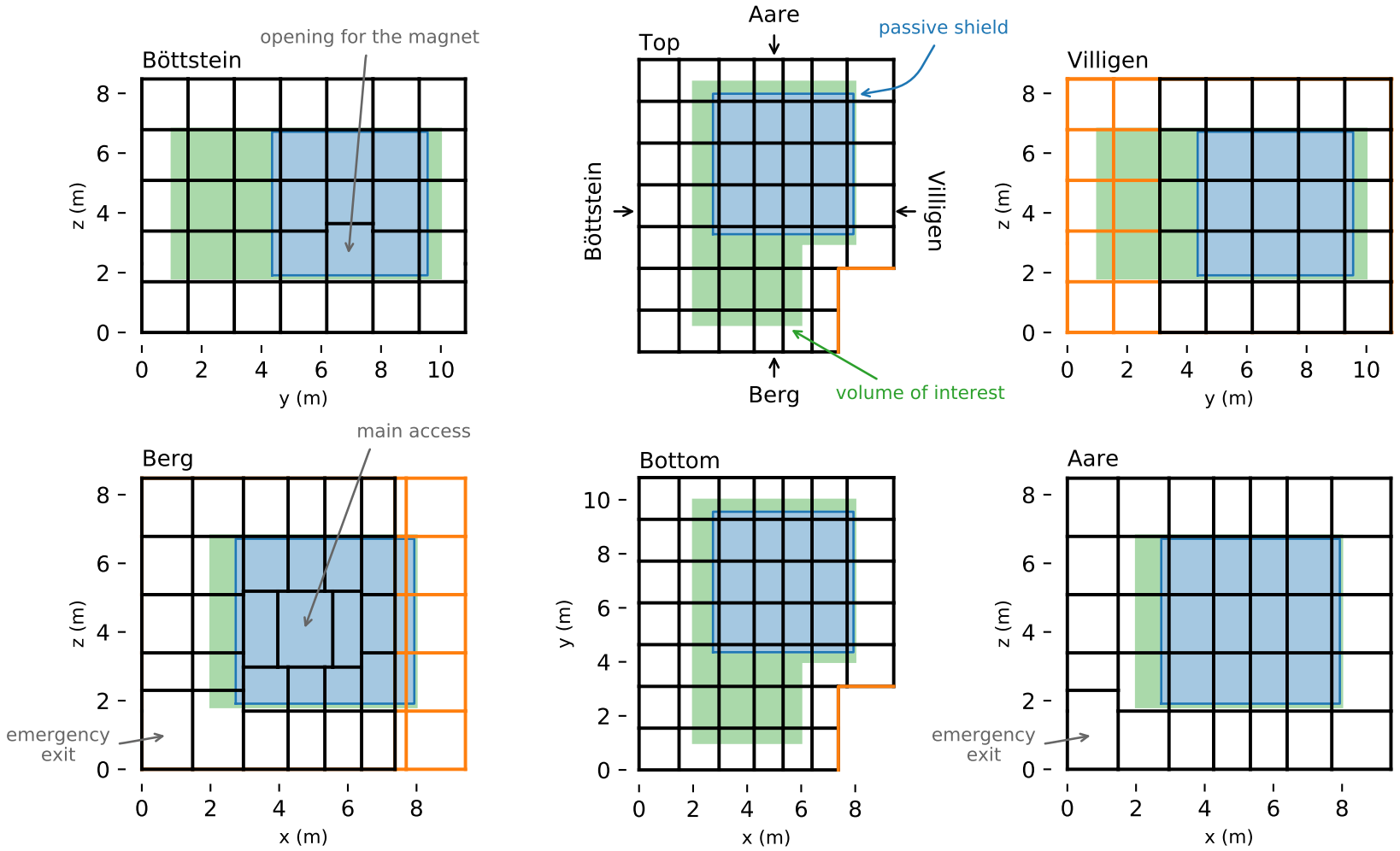

Figure 41: The cage for the proposed n2EDM active shield. On the top view (upper centre) the names of the projections are defined (the names have a geographical origin). A kink to incorporate a staircase to an emergency exit from a nearby facility is marked in orange. The outline of the passive shield is depicted in blue, the volume of interest used in the optimisation in green. Also marked are: an opening for the magnet polarising the neutrons (blue, on the left in Fig. 53), larger openings for emergency exits, and a large opening for the main access to the shield's door.

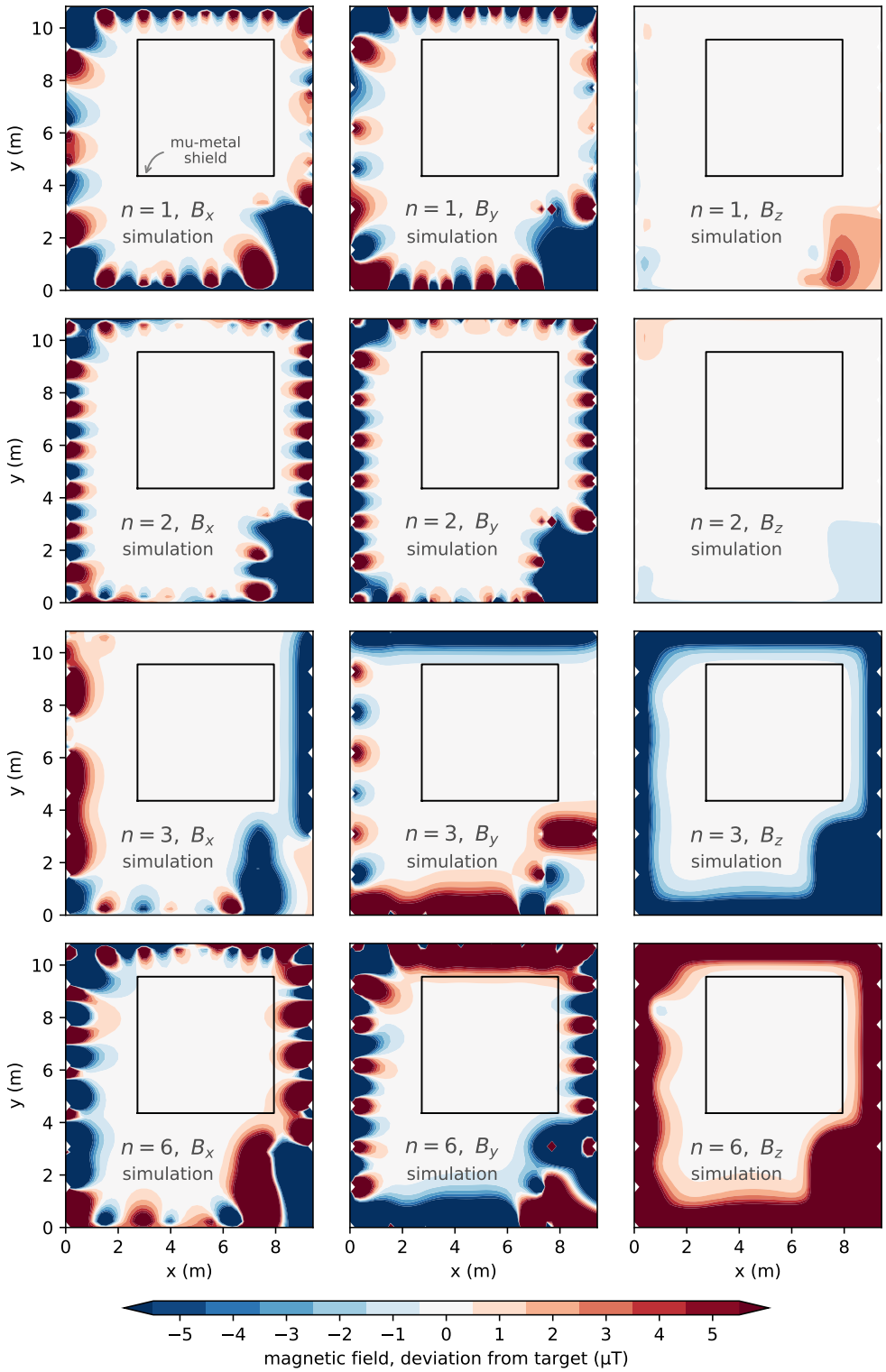

Figure 42: Simulation of the n2EDM design. The colour depicts the deviation from the target field for the x, y and z components of the field. Horizontal sections $z = 4.5\,\mathrm{m}$ above the floor are shown. *First three rows:* Depiction of coils designed to produce 50 µT homogeneous fields along x, y and z, respectively ($n = 1,\ 2$ and 3, cf. Tab. 1). *Bottom row:* The field of coil for a $20\,\mu\mathrm{T/m}$ $n = 6$ linear gradient.

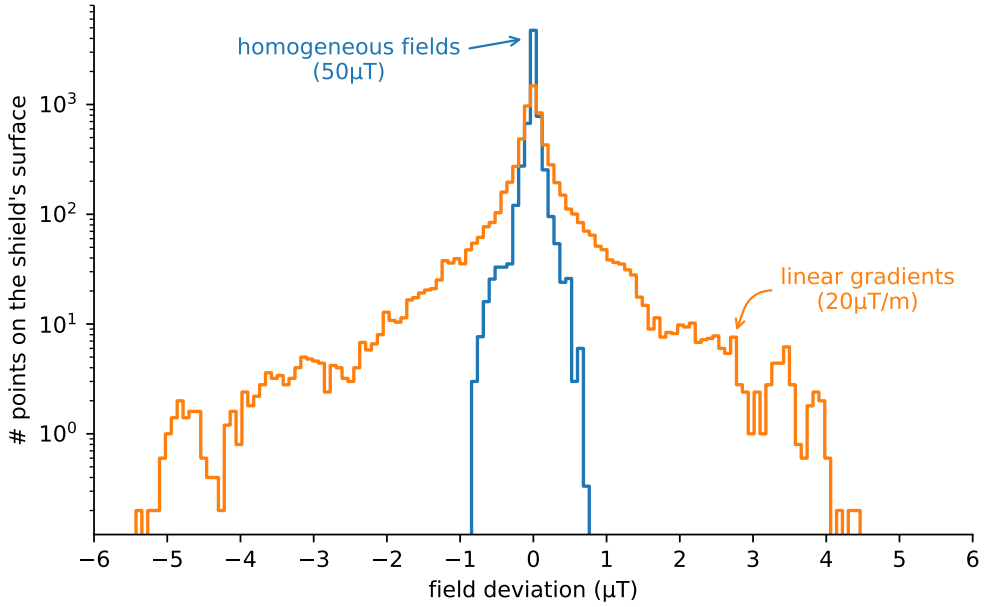

Figure 43: The quality of the field produced by the coils designed for the n2EDM experiment, quantified by a histogram of the deviation of the design's simulated field from the target one, as calculated on the surface of the mu-metal shield. One entry corresponds to a difference in one spatial direction at one point. The designs for the homogeneous coils have been averaged as they are similar. The deviations for the linear gradients were also averaged.

NEXT-GENERATION ACTIVE MAGNETIC SHIELDING – CONCLUSION

Active magnetic shields built on a grid can easily deal with spatial constraints. The grid is shared between different coils, making it easy to overlay a number of them. In particular, it is possible to construct coils for the mutually orthogonal terms of the cartesian harmonic expansion of the field. This simplifies the control of the system.

An active shield built at ETH demonstrated the grid-based approach. Maps of the coils confirmed that the 1310 mm large coils achieve a 2 % homogeneity in a 1000 mm large volume. The prototype showed the practical advantages of using cable channels as a support structure for the coils. The shield performed well when compensating a strong, nearby dipole source. Even in magnetically quiet conditions, the stability of the field was further improved down to 0.3 nT in timescales from seconds to hours.

In the second iteration the coils were designed with one face of the cage left open. This made the designs of the coils significantly more complicated. The maps of an open-design coil demonstrated its adherence to the simulated design.

Finally, a design of an active magnetic shielding cage for the n2EDM experiment was proposed. Thanks to the flexibility of the grid-based approach a design could fully respect the tight spatial constraints. Despite the complicated geometry, it is expected to perform similarly to the small-scale prototype.

6 MAPPING

An active magnetic shield relies on its coils to counteract field variations. A disturbance can only be compensated so well, as it can be approximated with the shield's coils. Therefore, an active shield will perform better if adapted to the magnetic environment, especially if the disturbances are higher-order.

As part of the design process for an active shield for the n2EDM experiment a characterisation of the magnetic fields on the future site was performed. For this, a magnetic field mapper was built in a form of a tower equipped with magnetic field sensors.

First, a small-scale prototype of the mapper was tested during a mapping campaign at LPSC, Grenoble, France. It was then extended to full-scale and used for measurements on the site of the n2EDM experiment.

6.1 THE IDEA

An essential input to the design of the n2EDM active magnetic shield was the scope of fields that would need to be compensated. There are strong magnetic sources in the vicinity of the site, some of them ramping on a daily basis. Taking a number of maps of the magnetic field was planned to characterise the magnetic environment.

Speed was valued more than precision. The less time it takes to map the field in the whole area, the less the map would be influenced by the varying external conditions. Also, the variety of the magnetic environment favoured taking multiple maps under different conditions, rather than fewer precise ones. The implemented solution was a mobile tower equipped with magnetic field sensors at different heights. The position and orientation of the tower could be measured while it was moved.

A small-scale prototype setup is pictured in Fig. 44. The tower houses ten fluxgate magnetometers. In the corner behind the tower a rigid L-piece is visible; it carries three string potentiometers. The potentiometers are attached to the tower via a thin wire. The thin wires are wound onto a spring-loaded spool, connected to a rotary potentiometer, on the inside of the potentiometer unit. The potentiometer gives an analogue signal proportional to the extension of the wire. The combined information from the three sensors was used to determine the position and orientation of the tower.

For maximal linearity, string potentiometers are constructed in a way that the wire is wound flat in one layer only.

Other names for string potentiometers include: cable-extension transducer, draw-wire sensor and string pot.

6.2 THE PRINCIPLE OF A STRING-POTENTIOMETER–BASED MAPPER

For a known length of wire, if one end is fixed, the other end draws a circle. The measurement of the position and orientation of the tower is solved by finding the intersections of those circles. In general, the problem of determining the

Figure 44: The small-scale prototype of the magnetic field mapper. The mobile tower houses ten fluxgate magnetometers. A bundle of readout cables is visible at the bottom of the tower. Behind it a rigid L-piece is visible, which holds three string potentiometers. A wire is extended from each potentiometer to the tower.

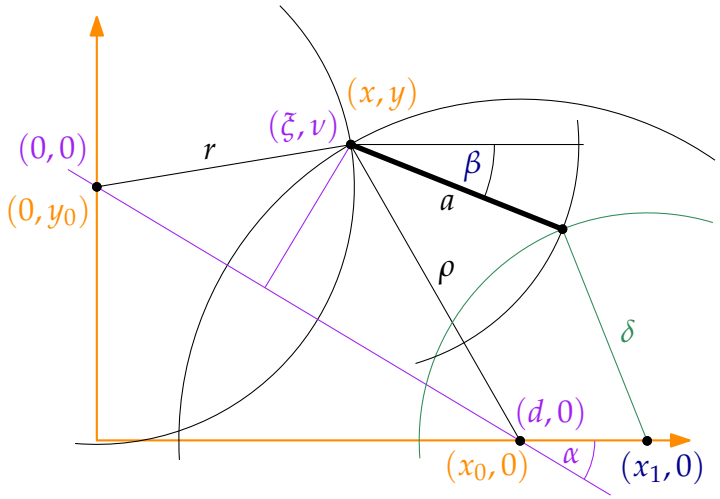

Figure 45: The geometry of the mapper's position and orientation determination. In the orange coordinate system, that of the "L-piece", the position of the centre of the tower is (x, y). The positions of the string potentiometers used to determine the tower's position are $(0, y_0)$ and $(x_0, 0)$, and the extensions of their wires r and ρ. Another coordinate system is depicted (purple), in which the position of one of the string potentiometers is $(0, 0)$ and the other $(d, 0)$. The position of the tower in this coordinate system is (ξ, ν). There is also a third string potentiometer attached to the tower on an arm of a length a; its wire length is denoted by δ.

location based on the measurement of the distance to a set of fixed points is called *trilateration*.

Other position determination methods include triangulation (measurement of angles between lines connecting a set of fixed points) and multilateration (measurement of the difference in distances between a set of fixed points).

The geometry is illustrated in Fig. 45. The two string potentiometers used to determine the position are located at points $(x, y) = (0, y_0)$ and $(x_0, 0)$ (in the L-piece coordinate system, orange in the figure). The length of wire from these locations, r and ρ, are connected to a single point which lies directly on the vertical beam holding the sensors. We will refer to this point as the centre, even though it is not necessarily the geometric centre of the tower. The centre lies on the intersection of the corresponding circles. For the sake of simplicity, we first give the solution for the centre position in the coordinate system depicted in purple, where the first string potentiometer is at $(0, 0)$ and the second $(d, 0)$, where

$$d = \sqrt{x_0^2 + y_0^2} \tag{33}$$

with the tower's centre at (ξ, ν),

$$\xi = \frac{1}{2d}\left(d^2 - \rho^2 + r^2\right) \tag{34}$$

$$\nu = \pm \frac{1}{2d}\sqrt{(-d + \rho - r)(-d - \rho + r)(-d + \rho + r)(d + \rho + r)}\,. \tag{35}$$

There are two solutions, symmetric around the connecting line between the centres of the circles. It is assumed that the tower never crosses this line. The transformation of the solution to the "L-piece" coordinate system (orange) is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \xi \\ \nu \end{pmatrix} + \begin{pmatrix} 0 \\ y_0 \end{pmatrix} , \tag{36}$$

where

$$\alpha = \arctan \frac{y_0}{x_0} . \tag{37}$$

The orientation is determined with a use of the third potentiometer, with its string attached to the tower at a distance a from the centre. Then, the position of the attachment point lies on the intersection of the circle centred at (x, y) with radius a, and one centred at the potentiometer, radius equal to the measured extension of the wire δ. In this case there are also two solutions, but ambiguity can be avoided by keeping the tower's orientation approximately parallel to the x-axis.

In the prototype $x_0 = 0$. This had the consequence that the precision of determining x was poor when both y was large and x was small, meaning the two circles intersected at a very small angle. Due to measurement noise and uncertainty in calibration (the relationship between the analogue signal and the extension of the wires), sometimes the circles did not intersect at all. In these cases, the middle point of the shortest line connecting the circles is taken as the solution.

6.3 LPSC CAMPAIGN

The small-scale prototype setup was used to help Dominique Rebreyend and Guillaume Pignol to characterise the magnetic field environment in two rooms at the Laboratoire de Physique Subatomique & Cosmologie (LPSC) in Grenoble, France. The rooms, referred to as *Bastille* and *Chalet*, were considered to host a magnetic-field-sensitive setup for ^{199}Hg magnetometry. Of particular interest were the gradients, which cause an increase in the depolarisation rate of the mercury atoms [49]. As the condition to be met for the future use 10 nT/cm was specified. The campaign took place during the days 6.-10.03.2017.

The string potentiometers were equipped with custom made attachments that allowed the wire to protrude at an angle out of the device. The point where the wires bent was taken to be the middle of the circle in the geometry solution.

A photograph of the mapping setup is shown in Fig. 44. The 2.5 m high mobile tower was equipped with ten three-axis fluxgate magnetic field sensors (Stefan-Mayer FLC3–70, the same as used in the active magnetic field stabilisation system). The stationary "L-piece" held three string potentiometers (Micro-Eplison WDS-15000-P115-SA-P with a 15 m extension range). The geometry parameters, as defined in Fig. 45, were: $x_0 = 0$, $y_0 = 1871$ mm, $x_1 = 1756$ mm and $a = 623$ mm. The string potentiometer for the orientation measurement was at the point $(1756\,\text{mm}, 0\,\text{mm})$.

The data acquisition system, pictured in Fig. 46, consisted of: a power supply to supply a constant voltage (10 V) to the potentiometers; a custom-built crate for the fluxgates, which supplies power and conditions the incoming signals; and a National Instruments PXI crate, which simultaneously digitises the

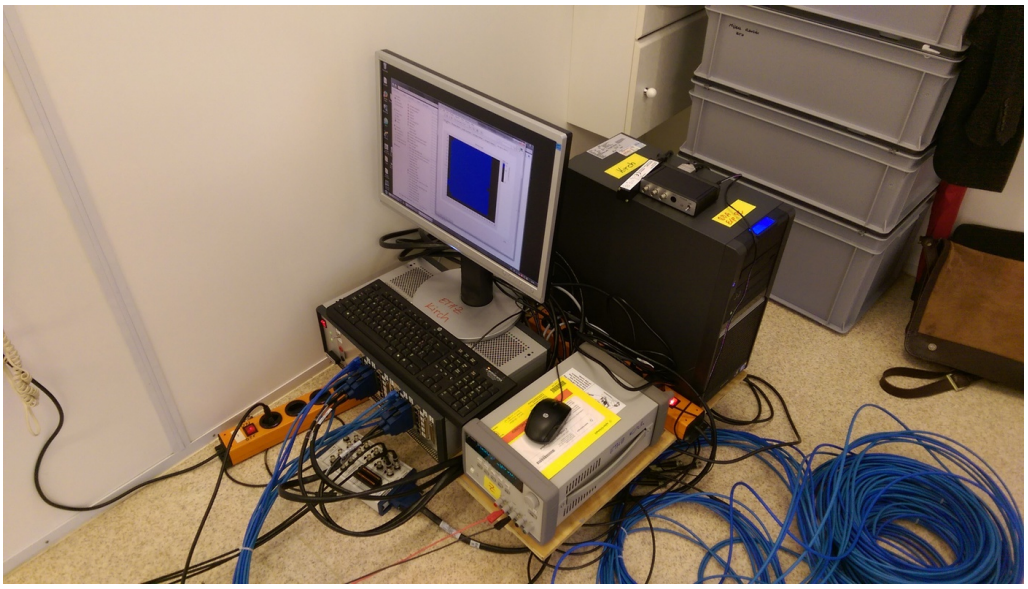

Figure 46: The DAQ system used for the mapping campaign at LPSC: a power supply for constant voltage at the potentiometers (grey, lower corner); a custom-built crate for the fluxgates supplying them with power and conditioning the incoming signals (underneath the keyboard and the screen); a National Instruments PXI crate, which simultaneously digitises the analogue voltage signals from the fluxgates and the string pots (not visible, behind the screen); and a PC computer with peripherals.

analogue voltage signals from the fluxgates and the string pots. The data were collected on a PC computer.

A panoramic shot of the Bastille room is presented in Fig. 47. The L-piece is visible in the upper-left corner. To the right is the entrance door. The room is made out of wood and is located inside a hall made of steel beams and sheets. The hall is equipped a gantry crane, several metres above the roof of the room.

To collect a map the tower was moved around the room by one person with a second taking care that the readout cables did not get in the way of the tower. Care was taken to systematically scan the whole room. A track of one of the collected maps is depicted in Fig. 48. Note that the centre of the tower could only reach a certain distance from the walls, because of the size of the tower's

The position of the mapper was determined on-line. On the screen it was displayed where the tower had already been.

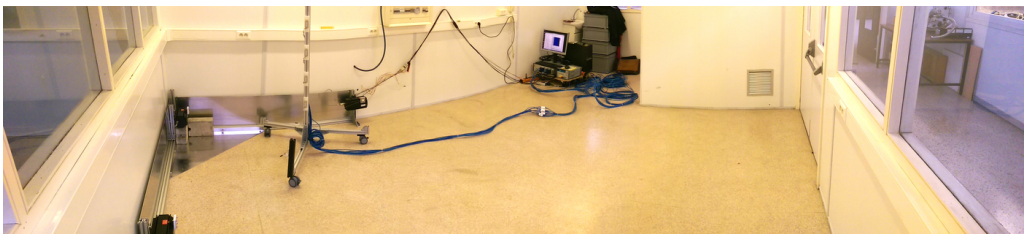

Figure 47: A panoramic view of the *Bastille* room. To the left the "L-piece" and the mapping tower are visible. In the back is the DAQ system and to the right a door with a long handle, which is visible on the magnetic field maps.

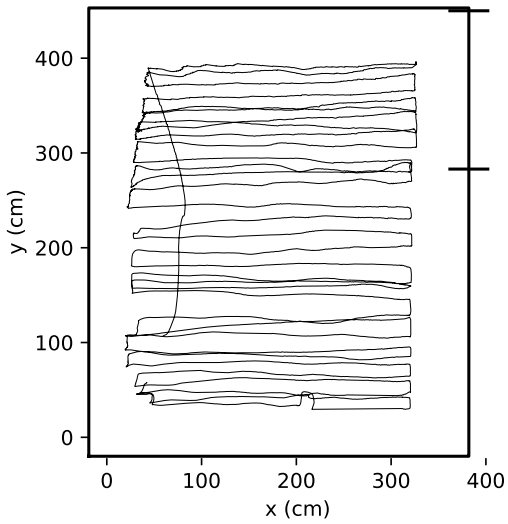

Figure 48: A map of one trace of the tower's centre. The outline of the room borders the plot. In the upper-right corner the location of the entrance door is marked. The "L-piece" was in the lower-left corner.

cart. The data, the position and orientation of the tower, as well as the magnetic field measurements, were down-sampled and registered at a 50 Hz rate.

6.4 DATA TREATMENT

A map of the magnetic field is plotted in Fig. 49. Horizontal slices (each a different sensor) of the magnetic field magnitude are shown, with the height of the slice above the floor indicated. Towards the roof a number of localised sources are visible; they can be attributed to air-conditioning system elements mounted on the roof of the hut.

The measured information was also used to estimate the magnetic field gradients. For this, the mapped area was divided into a 14×14 grid. In each bin the measurements were averaged. Then, the differences between the magnetic field components in the neighbouring pixels were taken, which, when divided by the separation between the bins, give the estimates of the gradient. The map of the gradient 205 cm above the floor is shown in Fig. 50. Only "horizontal" gradients were estimated, ones with respect to x and y. To estimate the vertical ones, with respect to z, would require comparing the measurements of different sensors. The sensors were specified to be only $\pm 1\% \pm 0.5\,\mu\mathrm{T}$ accurate, which in a roughly 50 µT field adds up to a microtelsa. With 22 cm separation between the sensors, the systematic effect on the gradient would be 45 nT/cm. In the map, gradients lower than 10 nT/cm are resolved.

Binned data allowed also for a direct comparison between maps. In order to estimate the reproducibility of the mapping process, two maps were measured directly one after another. The maps were then subtracted, after binning. A histogram of the differences is shown in Fig. 51. The standard deviation of the distribution, a measure of the reproducibility, is 138 nT. The reproducibility of

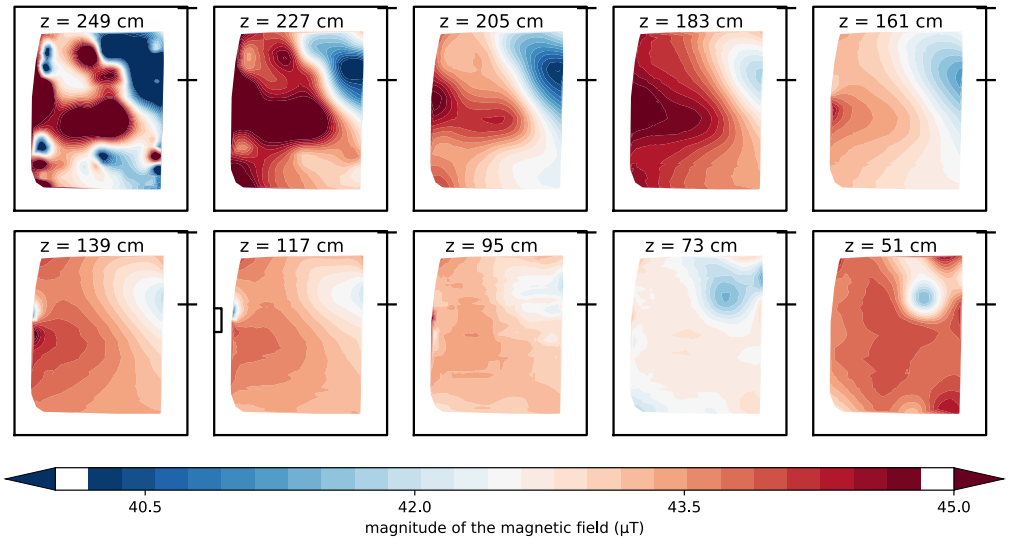

Figure 49: The map of the magnitude of the magnetic field. Each tile depicts a horizontal plane, the field measured by one magnetic sensor. For each point, registered at 50 Hz rate, the magnitude is plotted. The colour in between the measurement points is linearly interpolated. On the ceiling there are small, strong sources of magnetic fields and the colour scale is saturated.

the gradient was estimated in the same way, yielding a standard deviation of 3.8 nT/cm. This is better than a naïve estimate, which assumes that the value of the field component measured in a single, ≈ 25 cm large bin has an uncorrelated error bar of 138 nT, resulting in a gradient of

$$\frac{138\,\mathrm{nT}\,\sqrt{2}}{25\,\mathrm{cm}} = 8\,\mathrm{nT/cm}\,. \tag{38}$$

Reproducibility of the attained field values requires stability over long time periods. For the gradient of the field to be reproducible, the system only needs to be stable from one bin to the next.

Also, two maps taken under different conditions can be compared. Near the roof of the hall where the Bastille hut is located there is a large gantry crane. In order to estimate the change in magnetic field when the crane moved two maps were taken: one with the crane in the far end of the hall and another with the crane directly above the hut. The maps were binned. Their difference is plotted in Fig. 52. The magnetic field produced by the crane in the room is around 1 μT strong half a metre above the room's floor, furthest from the crane, and rises to 3 μT at the roof.

The campaign at LPSC first of all resulted in a detailed magnetic characterisation of the Bastille and Chalet rooms, under consideration for future magnetometry laboratories. (The maps of the Chalet rooms can be found in App. B.) Secondly, the campaign demonstrated the capabilities of the magnetic field mapping with a mobile tower, whose position and orientation is determined with string potentiometers. The prototype was later enlarged in preparation for the mapping of a much larger volume, one where the n2EDM experiment was to be built.

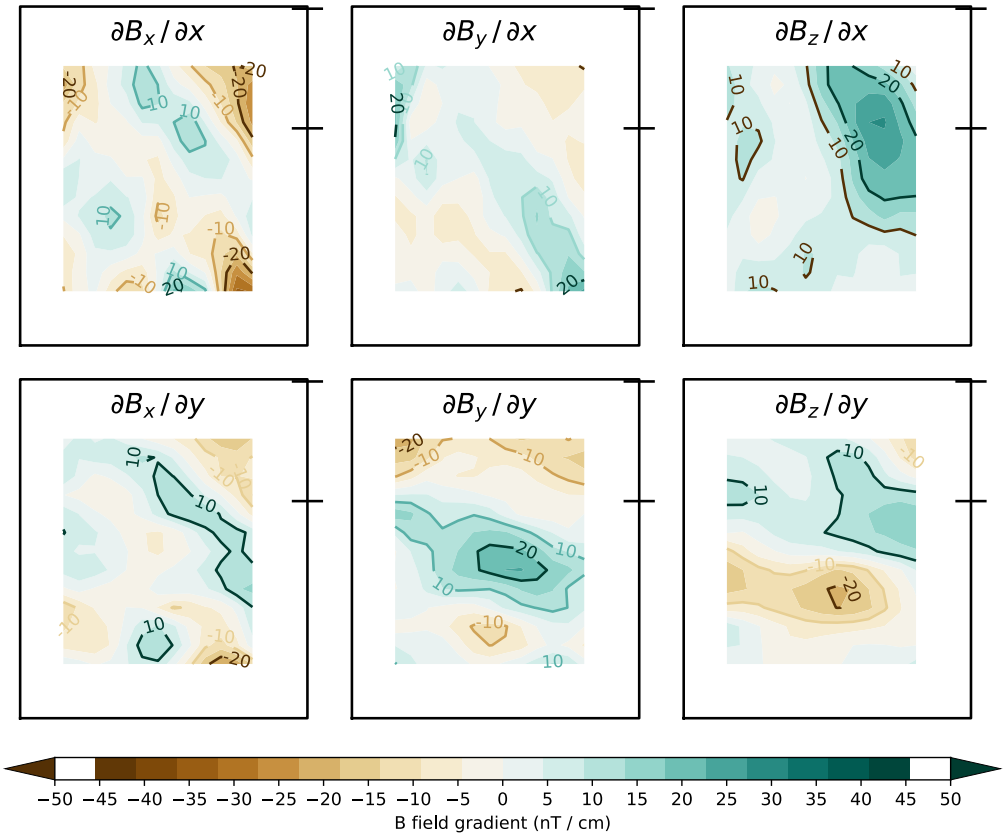

Figure 50: A map of magnetic gradients in the Bastille room 205 cm above the floor. Only the horizontal gradients (with respect to x and y) have been estimated. Estimating the gradients with respect to z would require comparing measurements between different sensors, whose intrinsic offset would dominate the result.

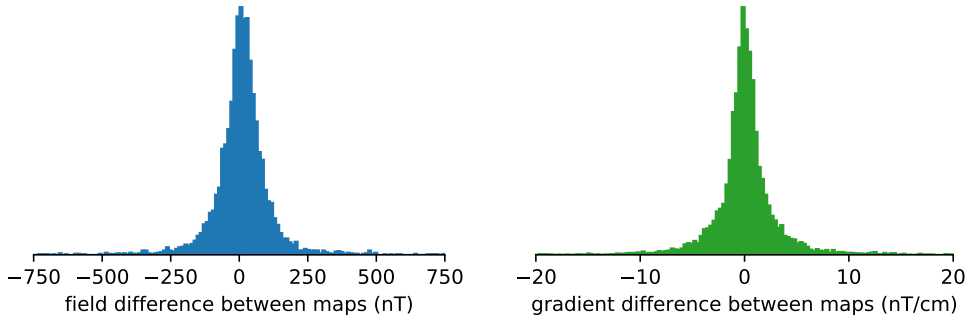

Figure 51: *Left:* Reproducibility of the field, a histogram of the difference in the measured field; one entry is a difference along either x, y or z. The width of the distribution is 138 nT. *Right:* reproducibility of the gradient estimation; one entry is a horizontal gradient estimate, with respect to x or y. The width of the distribution is 3.8 nT/cm.

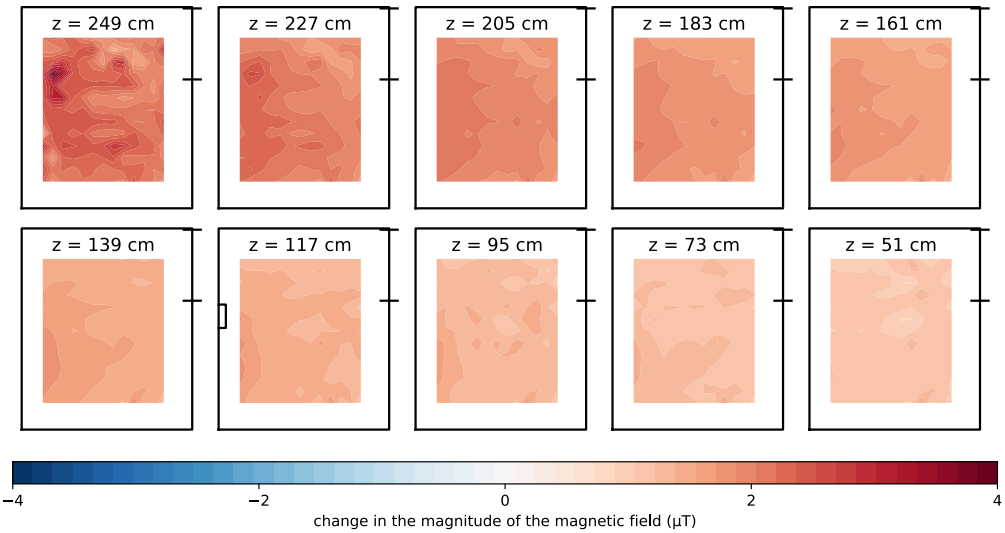

Figure 52: A map of the magnetic field of a gantry crane in the hall where the Bastille room is located. The magnitude of the field difference between two maps is plotted: one with the crane directly above the hut, and one with it moved far away.

6.5 PSI AREA SOUTH CAMPAIGN

The mapping of Area South in the UCN hall at PSI, the location of the nEDM and n2EDM experiments, took place between 19.12.2017–12.01.2018. In that time the area was empty; the nEDM apparatus had already been disassembled and a reinforced concrete foundation for the n2EDM experiment had been laid. The goal of the mapping campaign was to characterise the magnetic field environment, in particular the different fields created by nearby magnets. The maps will provide the information necessary to design an active system that would efficiently compensate for external fields. The mapping campaign was a joint venture with Solange Emmenegger and Jochen Krempel.

The total weight of the n2EDM shield is 47 tonnes.

The prototype mapper used in the LPSC campaign was extended to its full height. Area South, pictured in Fig. 53, is about $10 \times 12\,\mathrm{m}$. The $5.2 \times 5.2 \times 4.8\,\mathrm{m}$ magnetic shield of n2EDM would reach almost 7 m in height. In the full-scale setup the parameters, as defined in Fig. 45, were: $x_0 = 0$, $y_0 = 3814\,\mathrm{mm}$, $x_1 = 3673\,\mathrm{mm}$ and $a = 976\,\mathrm{mm}$. The tower was 8 m high. An additional magnetic field sensor directly above the floor was added. Other than that, no changes from the prototype were made. The string potentiometers, the magnetic field sensors and the data acquisition system remained the same.

The tower was built out of aluminum trusses, typically used for stages.

There are three strong magnets close to the area: COMET, SULTAN and COBRA. COMET is a medical cyclotron located several meters into the concrete biological shielding on the far-side wall in Fig. 53. Its around 10–50 μT field is very inhomogeneous due to the vicinity of the source. COMET is very rarely turned off. Fortunately during the mapping campaign there was a scheduled power-cut for which the cyclotron was ramped down. SULTAN is a superconducting magnet for material research, located $\approx 15\,\mathrm{m}$ away from the site. It

Figure 53: Area South in the UCN hall at PSI—The location of the nEDM and n2EDM experiments at PSI. The brown square outlines the foundation for the n2EDM apparatus, covered with plywood. The mapping tower is being pushed around by one person; another holds the cables. The L-piece with the string potentiometers is visible in the far end of the hall.

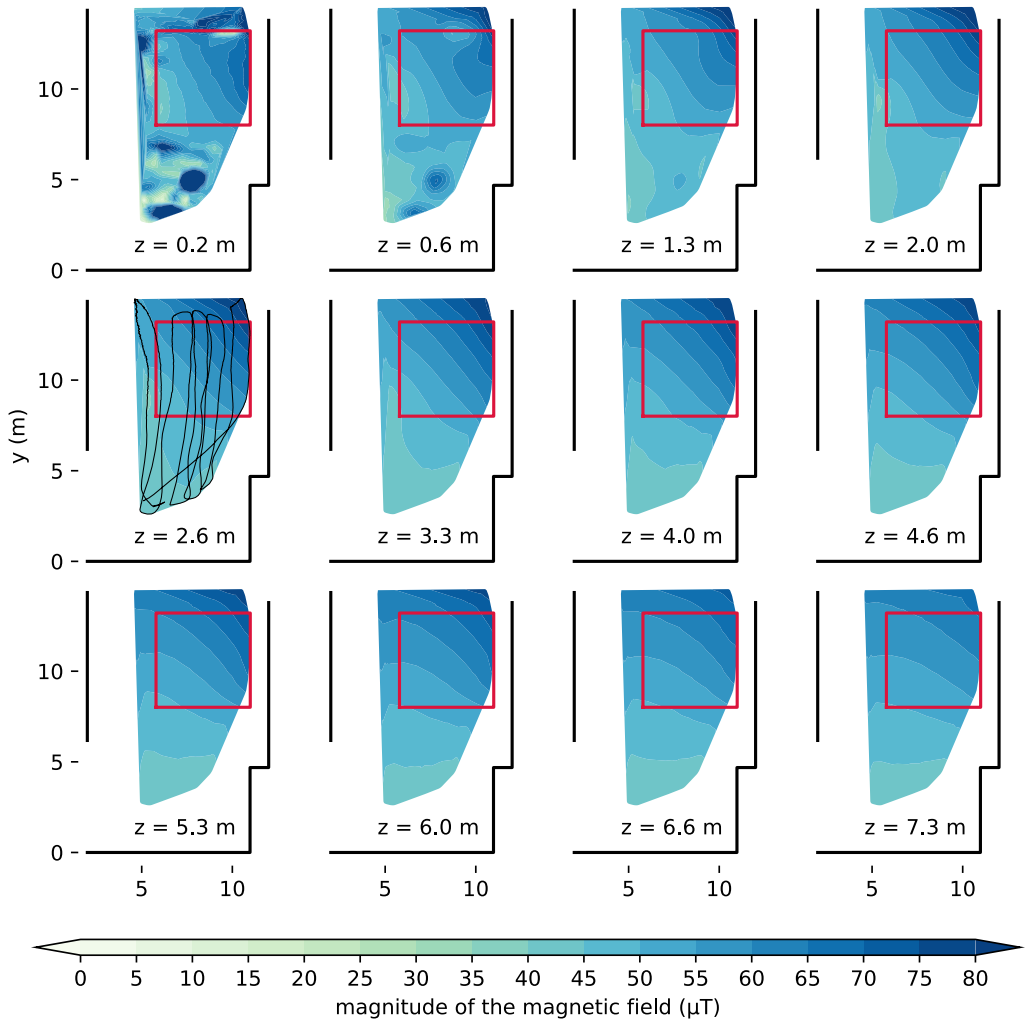

Figure 54: A map of the magnetic field in the Area South at PSI. The field magnitude is plotted in horizontal planes at different heights. The outlines of the n2EDM mu-metal shield (red) and concrete walls (black) are marked. On the map at $z = 2.6\,\mathrm{m}$ the trace of the mapping tower is depicted. The ticks give the distance from an arbitrary origin in metres. Only the COMET magnet was on. Courtesy of Solange Emmenegger [93].

ramps daily, creating a $\approx 50\,\mu T$ field in the area. COBRA is a magnet located in a neighbouring experimental hall. The strength of its field is around $5\,\mu T$.

In total 25 maps were taken with various combinations of the states of the magnets. The maps are summarised in Tab. 2. As of April 2018 the analysis of these maps in ongoing [93]. The procedure of binning and subtracting the maps (Sec. 6.4) will allow the field of each of the three magnets to be estimated. The measurements will be an important input to the design of the n2EDM active magnetic shield. They will help to decide which coils should be built. It is even possible to design coils dedicated for a particular high-order disturbance, such as COMET.

MAPPING – CONCLUSION

Active shielding greatly improved the magnetic field stability for the nEDM experiment at PSI. The design of a shield for its successor, n2EDM, proved to be challenging due to tight spatial constraints.

A new method for magnetic field coil design targeted the problem of spatial constraints. In the design process the coils were constrained to a predefined grid, which was defined to fulfil the constraints. The resultant coils have a very large fiducial volume, which means they can be fitted tightly around the apparatus.

A prototype grid-based active shield demonstrated magnetic field stabilisation in a large fiducial volume, showing that an active magnetic shield for the n2EDM experiment can be built.

The maps of the magnetic field at the site of n2EDM provides a crucial input to the design of the coils of the new active shield. The solution can be implemented within the framework of the new coil design method.

The developments from this work lay the foundation for an active magnetic shield for the n2EDM experiment, which would reduce the tens-of-microteslas variations across the volume down to just a few.

date	time	SULTAN	COBRA	COMET
19.12.2017	17:32	100%	100%	100%
19.12.2017	17:55	0%	100%	100%
21.12.2017	10:04	95%	100%	100%
21.12.2017	10:19	ramping	100%	100%
21.12.2017	11:53	90%	100%	100%
21.12.2017	11:56	90%	100%	100%
21.12.2017	13:38	20%	100%	100%
21.12.2017	14:23	20%	100%	100%
21.12.2017	16:36	20%	100%	50%
21.12.2017	16:47	ramping	100%	50%
21.12.2017	17:35	0%	100%	100%
22.12.2017	10:55	0%	100%	100%
04.01.2018	15:41	0%	0%	100%
04.01.2018	17:01	0%	0%	100%
04.01.2018	17:17	0%	0%	100%
05.01.2018	14:59	0%	0%	0%
05.01.2018	15:22	0%	0%	0%
05.01.2018	16:12	0%	0%	0%
05.01.2018	16:41	0%	0%	0%
05.01.2018	20:26	0%	0%	0%
09.01.2018	10:55	0%	0%	100%
09.01.2018	11:30	0%	0%	100%
12.01.2018	10:31	0%	0%	100%
12.01.2018	10:55	0%	0%	100%
12.01.2018	14:11	0%	0%	100%

Table 2: List of the maps taken. The status of the magnets is given along with an approximate percentage of the maximum field measured during the campaign.

Part III

AXION-DARK-MATTER SEARCH

Most of the universe's matter content, an estimated 84 %, is dark matter. Among the candidates for its constituents is an ultra-low-mass axion. In this part a search for a signature of an axion dark matter in the data of the nEDM experiment at PSI is described. The ratio of the spin-precession frequencies of stored ultracold neutrons and ^{199}Hg atoms was analysed for an axion-induced oscillating electric dipole moment of the neutron and an axion-wind spin-precession effect. No signal consistent with dark matter was observed for the axion mass range $10^{-24}\,\text{eV} \leq m_a \leq 10^{-17}\,\text{eV}$. The null result set the first laboratory constraints on the coupling of axion dark matter to gluons, which improved on astrophysical limits by up to three orders of magnitude. It also improved on previous laboratory constraints on the axion coupling to nucleons by up to a factor of 40.

The first section in this part introduces the subject of axion-like dark matter. In the second the method of choice for the analysis, the least-squares periodogram, is discussed. The methodology developed there is then applied to the PSI nEDM dataset in the third section. The part concludes with a proposal for a resonant search for an oscillating electric dipole moment.

7 INTRODUCTION

This section is largely based on and also reproduces text of Ref. [94].

Based on astrophysical and cosmological observations an estimated 26 % of the total energy density of the Universe and 84 % of its mass content is dark matter (DM) [95]. Observations give hints about the amount and distribution of DM, for example via rotational curves of galaxies or gravitational lensing [96], but the micro-scale properties of DM, in particular its constituents, remain unknown.

Among the candidates for DM is an axion, a new pseudoscalar particle, initially proposed to solve the strong QCD problem (the strong sector in the Standard Model appears to be fine-tuned to be CP-even) [45, 97–103]. It has been later generalised to axion-like particles, or simply axions [104–109]. Light ($m_a \lesssim 0.1\,\mathrm{eV}/c^2$) axions can be produced efficiently via non-thermal production mechanisms, such as vacuum misalignment in the early Universe [110–112]. They fill the universe as almost stationary particles and through gravity participate in the galaxy formation. During the formation they gain speed ($\approx 300\,\mathrm{km/s}$) and, as bosons, condensate into a coherent oscillating field ($\Delta\omega/\omega \sim 10^{-6}$) [109], with the frequency of the oscillation set by the mass of the axion m_a:

$$a = a_0 \cos\left(\frac{m_a c^2}{\hbar}\, t\right) . \tag{39}$$

Due to its effects on structure formation [113], ultra-low-mass axion DM in the mass range $10^{-24}\,\mathrm{eV} \lesssim m_a \lesssim 10^{-20}\,\mathrm{eV}$ has been proposed as a DM candidate that is observationally distinct from, and possibly favourable to, archetypal cold DM [109, 114–117]. The requirement that the axion de Broglie wavelength does not exceed the DM size of the smallest dwarf galaxies and consistency with observed structure formation [118–120] give the lower axion mass bound $m_a \gtrsim 10^{-22}\,\mathrm{eV}$, if axions comprise all of the DM. However, axions with smaller masses can still constitute a sub-dominant fraction of DM [121].

It is reasonable to expect that axions interact non-gravitationally with standard-model particles. Direct searches for axions have thus far focused mainly on their coupling to the photon (see the review [122] and references therein). Recently, however, it has been proposed to search for the interactions of the coherently oscillating axion DM field with gluons and fermions, which can induce oscillating electric dipole moments (EDMs) of nucleons [123] and atoms [124–126], and anomalous spin-precession effects [124, 127, 128]. The frequency of these oscillating effects is dictated by the axion mass, and more importantly, these effects scale linearly in a small interaction constant [123–128], whereas in previous axion searches, the sought effects scaled quadratically or quartically in the interaction constant [122].

Axions are hoped to convert into photons in a strong magnetic field. Helioscopes look for energetic axions produced in the sun. Haloscopes are sensitive to a relict axion dark matter.

In this part two axion couplings are considered: the one to gluons and the one to nucleons:

$$\mathcal{L}_{\rm int} = \frac{C_G}{f_a}\frac{g^2}{32\pi^2} a G^b_{\mu\nu}\tilde{G}^{b\mu\nu} - \frac{C_N}{2f_a}\partial_\mu a\ \bar{N}\gamma^\mu\gamma^5 N\,, \tag{40}$$

where G and $\tilde{G}$ are the gluonic field tensor and its dual, $b = 1, 2, \ldots, 8$ is the color index, $g^2/4\pi$ is the color coupling constant, N and $\bar{N} = N^\dagger\gamma^0$ are the nucleon field and its Dirac adjoint, f_a is the axion decay constant, and C_G and C_N are model-dependent dimensionless parameters. Astrophysical constraints on the axion-gluon coupling come from Big Bang nucleosynthesis [129–131]: $m_a^{1/4} f_a/C_G \gtrsim 10^{10}\ \mathrm{GeV}^{5/4}$ for $m_a \ll 10^{-16}\,\mathrm{eV}$ and $m_a f_a/C_G \gtrsim 10^{-9}\,\mathrm{GeV}^2$ for $m_a \gg 10^{-16}\,\mathrm{eV}$, assuming that axions saturate the present-day DM energy density, and from supernova energy-loss bounds [128, 132]: $f_a/C_G \gtrsim 10^6\,\mathrm{GeV}$ for $m_a \lesssim 3\times 10^7\,\mathrm{eV}$. Astrophysical constraints on the axion-nucleon coupling come from supernova energy-loss bounds [132, 133]: $f_a/C_N \gtrsim 10^9\,\mathrm{GeV}$ for $m_a \lesssim 3\times 10^7\,\mathrm{eV}$, while existing laboratory constraints come from magnetometry searches for new spin-dependent forces mediated by axion exchange [134]: $f_a/C_N \gtrsim 1\times 10^4\,\mathrm{GeV}$ for $m_a \lesssim 10^{-7}\,\mathrm{eV}$.

The axion-gluon coupling in Eq. 40 induces the following oscillating EDM of the neutron via a chirally-enhanced 1-loop process [1] [135–137]:

$$d_{\rm n}(t) \approx +2.4\times 10^{-16}\,\frac{C_G a_0}{f_a}\cos(m_a t)\ e\,\mathrm{cm}\,. \tag{41}$$

The axion-gluon coupling also induces oscillating EDMs of atoms via the 1-loop-level oscillating nucleon EDMs and tree-level oscillating P, T-violating intra-nuclear forces (which give the dominant contribution) [124, 138, 139]. In the case of ^{199}Hg, the oscillating atomic EDM is [124, 130, 140–147]

$$d_{\rm Hg}(t) \approx +1.3\times 10^{-19}\,\frac{C_G a_0}{f_a}\cos(m_a t)\ e\,\mathrm{cm}\,, \tag{42}$$

which is suppressed compared to the value for a free neutron (Eq. 41), as a consequence of the Schiff screening theorem for neutral atoms [148]. The amplitude of the axion DM field, a_0, is fixed by the relation $\rho_a \approx m_a^2 a_0^2/2$. In this work is it assumed that axions saturate the local cold DM energy density $\rho_{\rm DM}^{\rm local} \approx 0.4\,\mathrm{GeV/cm^3}$ [149].

The derivative coupling of an oscillating galactic axion DM field, $a = a_0\cos(m_a t - \boldsymbol{p}_a\cdot\boldsymbol{r})$, with spin-polarized nucleons in (40) induces time-dependent energy shifts according to:

$$H_{\rm int}(t) = \frac{C_N a_0}{2f_a}\sin(m_a t)\ \boldsymbol{\sigma}_N\cdot\boldsymbol{p}_a\,. \tag{43}$$

The term $\boldsymbol{\sigma}_N\cdot\boldsymbol{p}_a$ is conveniently expressed by transforming to a non-rotating celestial coordinate system (see, e.g., [150]):

$$\begin{aligned}\boldsymbol{\sigma}_N\cdot\boldsymbol{p}_a = {}& \hat{m}_F f(\sigma_N) m_a |\boldsymbol{v}_a| \\ & \times\left[\cos(\chi)\sin(\delta) + \sin(\chi)\cos(\delta)\cos(\Omega_{\rm sid}t - \eta)\right]\,,\end{aligned} \tag{44}$$

1 Interaction in Eq. 40 also non-perturbatively induces a mass $m_a \approx 6 C_G\,\mu\mathrm{eV}\cdot(10^{12}\,\mathrm{GeV}/f_a)$. Axions with masses much smaller than this are theoretically fine-tuned.

where χ is the angle between Earth's axis of rotation and the spin quantization axis ($\chi = 42.5°$ at the location of the PSI), $\delta \approx -48°$ and $\eta \approx 138°$ are the declination and right ascension of the galactic axion DM flux relative to the Solar System [151], $\Omega_{\text{sid}} \approx 7.29 \times 10^{-5}\,\text{s}^{-1}$ is the daily sidereal angular frequency, $\hat{m}_F = m_F/F$ is the normalized projection of the total angular momentum onto the quantization axis, and $f(\sigma_N) = +1$ for the free neutron, while $f(\sigma_N) = -1/3$ for the ^{199}Hg atom in the Schmidt (single-particle) model.

The scalar axion-gluon coupling and the vector axion-nucleon coupling would induce harmonic oscillations in the measurements of the nEDM experiment at PSI. In the scope of this analysis the time series of the ratio of spin-precession frequencies of polarised neutrons and ^{199}Hg atoms was tested for statistically significant oscillations. In the next chapter the methodology for this quantitative search is introduced.

8

PERIODOGRAMS

The space of possible axion-induced signals is spanned by their amplitude (the strength of the coupling) and their frequency (the axion mass). The problem is, therefore, naturally set in the frequency domain. The analysis was performed in two steps. First, the measurements were transformed from the time domain into the frequency one by evaluating the periodogram of the time series. In the second step the periodogram was checked for statistically significant signals.

In this chapter the statistical methods used in the analysis are introduced. The statistical properties of the periodogram are discussed, in particular the means of quantifying the significance of detected signals. Then, a Monte-Carlo-based way to set exclusions is presented. The methodology is demonstrated on a toy signal, generated specially for the purpose of a clear illustration of the subject.

8.1 DEFINITION OF THE PERIODOGRAM

A *periodogram* is an estimator of the power spectrum. It was proposed as the preferred way to treat periodic signals as early as 1898 [152]. In its simplest form it is the squared magnitude of the discrete Fourier transform, which, however, is only possible to evaluate for evenly sampled series. Lomb and Scargle have independently described a method to construct a statistically well-behaving periodogram for non-uniformly sampled data with unequal error-bars: the Least Squares Spectral Analysis (LSSA) (also known as the Lomb-Scargle periodogram) [153].

In order to evaluate the LSSA periodogram at a circular frequency ω, a linear least-squares fit (hence the name) is performed to the data with a function

LSSA, in contrast to the fast Fourier transform (FFT), does not require windowing, because it is explicitly phase-aware.

$$A\,\cos(\omega t) + B\,\sin(\omega t) + C\,, \tag{45}$$

where A, B and C are free parameters. The estimator of power $P(\omega)$ is then defined as

$$P(\omega) := \frac{N}{4}\left(A^2 + B^2\right)\,, \tag{46}$$

where N is the number of data points. Different normalisations may be used. Here the one of [153] is used, where the height of $\sqrt{P(\omega)}$ at the noise bed corresponds to the size of the error-bars, if they are all equal. A graphical overview of the method is shown in Fig. 55. Throughout the analysis the figure of merit is either the power $P(\omega)$ or, interchangeably, its square root—amplitude. The latter has conveniently the same unit as the time series.

The noise bed is the flat part of periodogram due to random noise. If there is a signal, it is said to "rise out of the noise bed".

We will now follow an analysis of a toy time series, shown in Fig. 56. The time series had been fabricated as simulated measurements of an oscillating signal

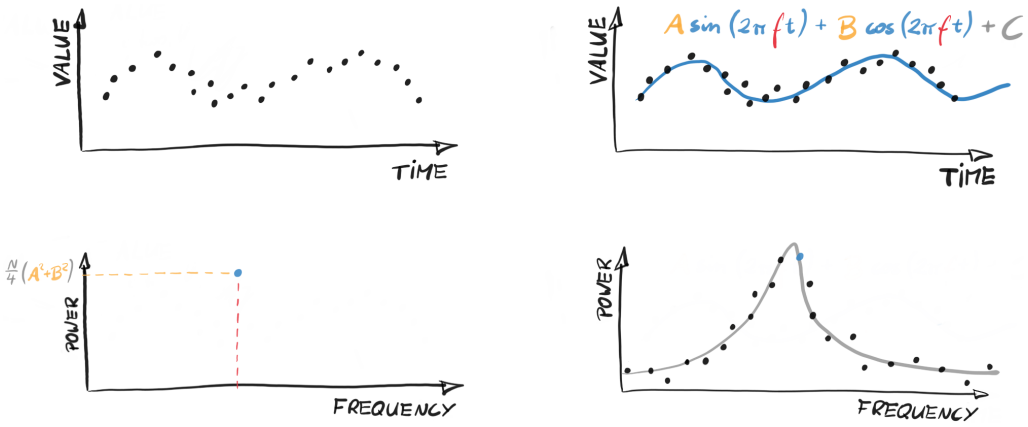

Figure 55: The construction of an LSSA periodogram. The estimate of the LSSA power at the frequency f is the amplitude squared of the least-squares fit of a harmonic oscillation of that frequency to the time-series (with a normalisation factor). The LSSA periodogram, an estimate of the power spectrum, is the LSSA power evaluated for a number of frequencies.

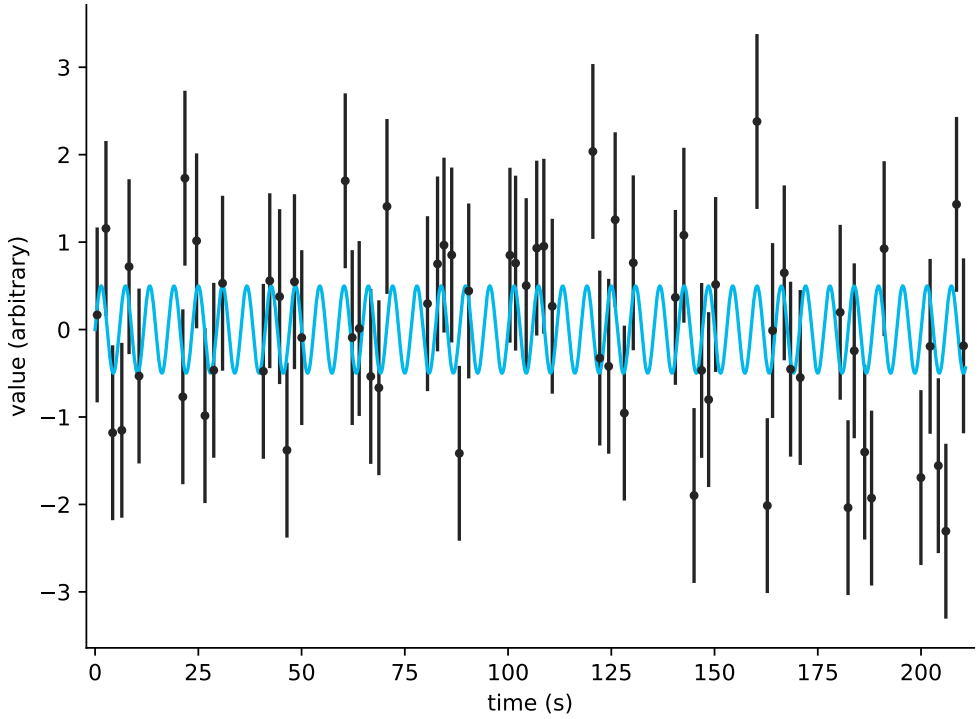

Figure 56: A toy signal generated for the purpose for explaining the general scheme of the periodogram analysis. A harmonic signal (blue) was used to generate unevenly spaced data points (black). Each point was drawn from a normal distribution centred at the blue curve and a unit width.

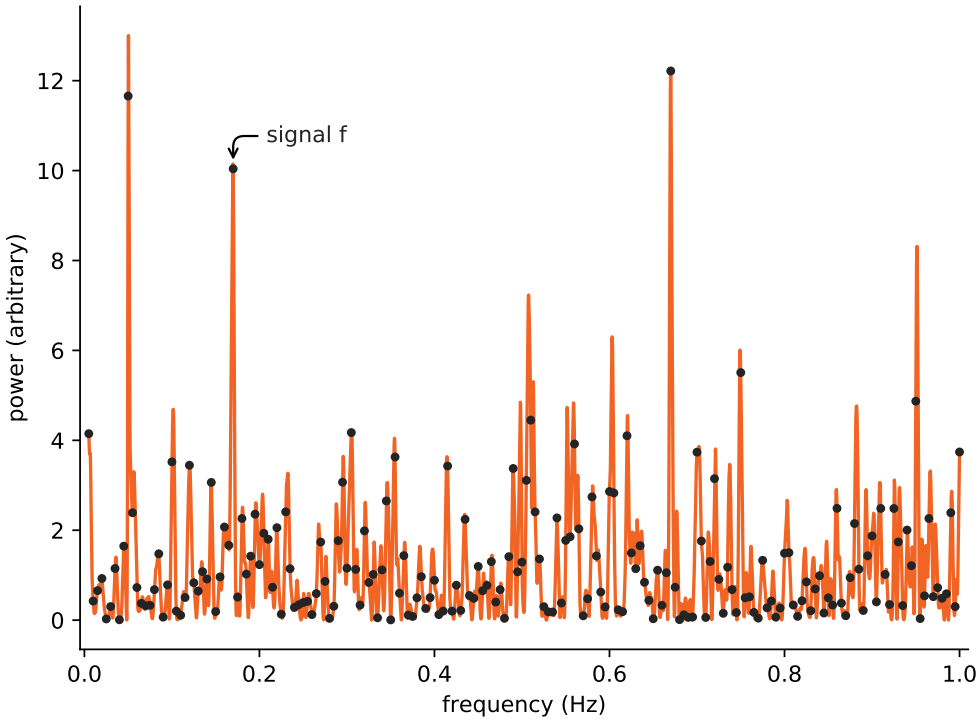

Figure 57: Two periodograms of the time series in Fig. 56. One evaluated at frequencies with a spacing equal to the spectral resolution (black dots), and one evaluated a thousand times more densely (the orange line).

of a frequency 0.17 Hz and an amplitude equal to 0.7 in an arbitrary unit. The series has already some properties of the actual dataset. The measurements are not equally spaced; they are randomly grouped in 10 s long bunches, around 20 s apart. Inside a bunch a "measurement" is taken every 2 s with a 0.3 s jitter. The length of each measurement is 1 s with a 0.1 s jitter. The error-bars are all size one in the arbitrary unit. Each "measurement" averaged the signal over its duration.

The immediate question arising when evaluating the LSSA periodogram is: for which frequencies to evaluate the power? In a case of evenly-spaced series the upper limit is the Nyquist frequency, equal to the half of the sampling rate [154]. This is not the case when the sampling is not uniform. In practice, we can expect little sensitivity to oscillations faster than the period over which each signal is averaged, 1 s in the example. On the low side the limit is zero, which corresponds to the constant offset (a least-squares fit of a horizontal line, which is equivalent to calculating the average of the points). The value of the periodogram at zero is usually not plotted.

When choosing the spacing between the frequencies, the *spectral resolution* is considered, defined as the inverse span of the dataset. It roughly defines the minimal frequency difference between two distinguishable signals. A perfectly coherent oscillation produces a cardinal-sine-shaped peak of a width equal to the spectral resolution [155]. In Fig. 57 there are two periodograms: a "sparse"

A Fourier transform of a rectangular function is a sinc function. Any signal measured for a finite time in the frequency space is necessarily convoluted with the sinc function.

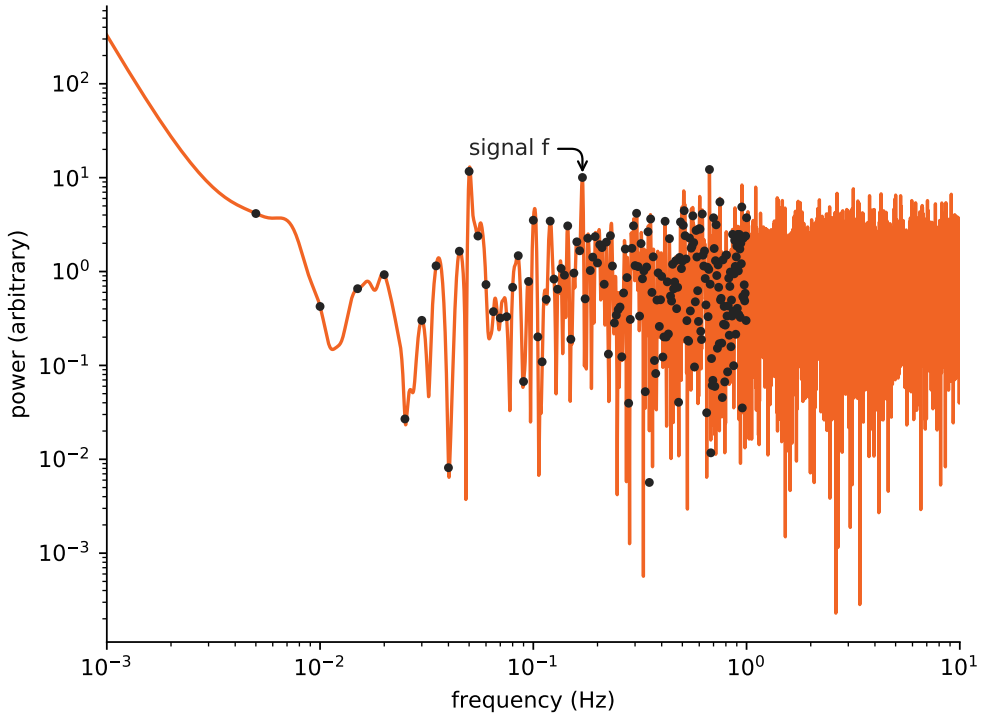

Figure 58: The two periodograms as in Fig. 57 plotted on a log-log scale. The range of frequencies where the periodogram is evaluated is extended. The amplitude of the noise appears to be increasing with frequency, but it is not true. Only the density of the evaluated points increases, making the extreme deviations more pronounce.

one evaluated at frequencies spaced a full spectral resolution apart (black dots), and one a thousand times more dense (the orange line). Each peak visible in the dense line is met with at least one sparse evaluation. However, not for each the point is at the peak's full height, which decreases the sensitivity of the method. As a compromise the frequency spacing is sometimes chosen to be a fraction of the spectral resolution, e.g. a tenth [156]. In this work the spacing of a full spectral resolution was used.

Figure 58 shows the same two periodograms, on a log-log scale. Additionally, the range of frequencies where the periodogram is evaluated has been extended to show its behaviour in the extremes. The logarithmic scale, albeit useful when the potential signals span orders of magnitude in both frequency and amplitude, can lead to misunderstandings. Specifically, the noise appears to increase in amplitude with frequency, but the effect is purely cognitive. The spacing of the points is linear, so on a logarithmic scale the density of points increases for high amplitude. This makes more of extreme deviations likely to appear per unit area of the plot.

An oscillation in the time series produces a peak in the periodogram. The position of the peak is the frequency of the oscillation, the width corresponds to the coherence of the signal. However, in the periodogram there are many peaks besides the one corresponding to the oscillation used when generating

the data. Some are even bigger and there are many smaller ones. In the next section we will consider what, besides an oscillating signal, may give rise to a peak. Most importantly, a way of determining whether a peak is caused by an oscillation is presented.

8.2 A NULL HYPOTHESIS TEST

Once the periodogram of a time series is calculated, one would like to know whether it contains a statistically significant signature of a signal. For a periodic signal this would be a peak. In our case the really interesting statement is the answer to the question:

How likely is it that the highest peak in the periodogram is not only a random fluctuation?

This question has already been stated by Scargle [153]. In this section we will be largely following the reasoning he presented, with few important differences.

In order to describe the question mathematically, let us denote the time series under consideration (Fig. 56) by D. The periodogram is then a set of $P^D(\omega_i)$, depicted with a black line in Fig. 59. In a uniformly sampled case with equal error bars $P^D(\omega_i)$ is exponentially distributed, for those frequencies where no signal is present [153]. In our, more complicated, case the distribution can be generated by a Monte Carlo (MC) simulation in the following way: a new signal is generated, keeping the time, position and the size of the error bars, but with no underlying signal present—the null hypothesis H_0. The value for each simulated measurement is drawn from a gaussian distribution with the width equal to the size of the error bar. Then the periodogram of the generated time series is calculated. This is repeated a number of times, yielding a set of periodograms, which are used to estimate the probability density function (PDF) of $P(\omega_i)$ for each i. The PDF for $\omega = 0.17\,\mathrm{Hz}$ (the frequency of the signal implanted in the time series) is depicted in the right-hand side of Fig. 59. In the left-hand side of the figure bands representing the shape of the PDFs for different frequencies (sigma bands, about to be explained) are depicted is shades of green. For uniformly sampled data with equal error bars all PDFs would be the same and the bands flat [153]. In our case structures appear, despite absolutely no signal being present in the generated time series.

The structures in the $P(\omega_i)$ PDFs are caused solely by the non-uniformity in sampling and in the sizes of the error bars. In particular, we can identify a very large expected rise in power at $0.05\,\mathrm{Hz}$, corresponding exactly to the inverse spacing between the bunches $1/20\,\mathrm{Hz}$, a feature of the toy measurement. A peak is expected in the periodogram to appear at this point, even when there is no significant oscillation of this frequency. This is the reason why the most significant peak should be sought, rather than simply the highest. This conclusion makes the presented reasoning different from the one of Scargle [153].

In the formal treatment we denote the cumulative density function (CDF) of the power estimator at the ith frequency as $F_i(z)$ (z would be the power

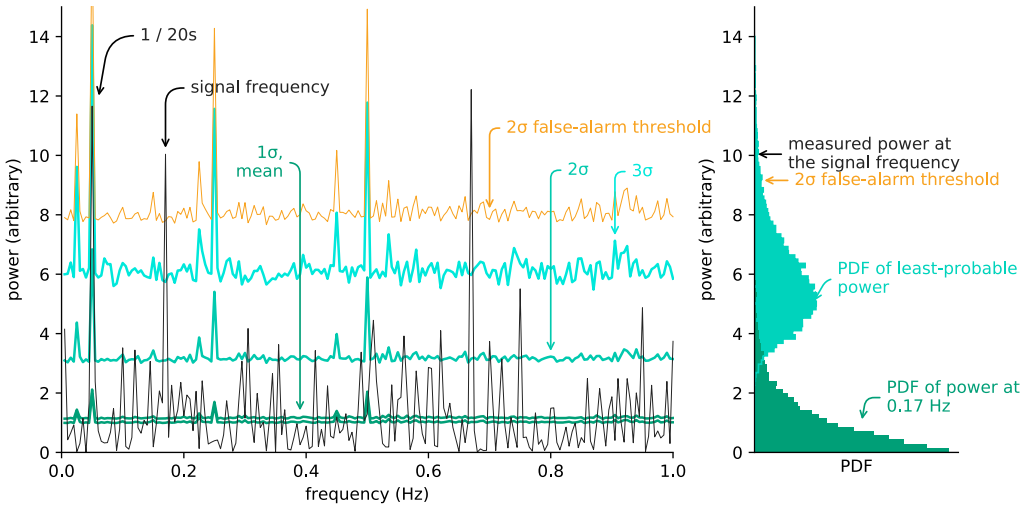

Figure 59: Left-hand side: Juxtaposition of the periodogram of the toy time series (black) and the periodogram PDF under the null hypothesis. For the latter the average and 1-, 2- and 3σ bands are depicted in shades of green. The 2σ global false-alarm threshold is marked in orange. Right-hand side: Aligned with the plot to the left, two PDFs are depicted: the one of the power at 0.17 Hz (the frequency of the signal put into the toy time-series) and the one of the globally least-probable power (across all frequencies).

estimated at frequency ω_i). In an evenly sampled, signal-free case it has a functional form

$$F_i(z) = 1 - e^{-z} \,. \tag{47}$$

P-value is the probability that at least this much power would arise only as a result of a random fluctuation.

In our case it can be estimated from the MC simulations. Then the ith p-value is directly

$$p_i = 1 - F_i\left(P^D(\omega_i)\right) \,. \tag{48}$$

The most significant peak is the one with the lowest p-value.

To construct the discrete CDF estimate (the empirical distribution function [157]) all the MC-generated power estimates were sorted into an array. The points of the CDF were obtained by plotting the power on the x-axis, and the position in the sorted array (normalised to one) on the y-axis. The discrete CDF estimate of power, or actually one-minus-CDF, so that the logarithmic y-scale can be leveraged to resolve the high-power tail, for the frequency 0.17 Hz is depicted in Fig. 60 with the green dots.

The CDF has the advantage over the PDF, that it does not require binning to be estimated [157].

When the power in the time series is very high, as it is the case for 0.17 Hz in the example, obtaining the discrete CDF estimate would require many, in some cases unrealistically many, Monte Carlo samples. Generic solutions of this problem are known (see for example section 39.3.2.2 in Ref. [7]). Here, a more specific approach is taken. Motivated by Eq. 47, the discrete CDF was extrapolated with a linear fit of the form

$$F_i(z) = 1 - A_i\, e^{-B_i z} \,, \tag{49}$$

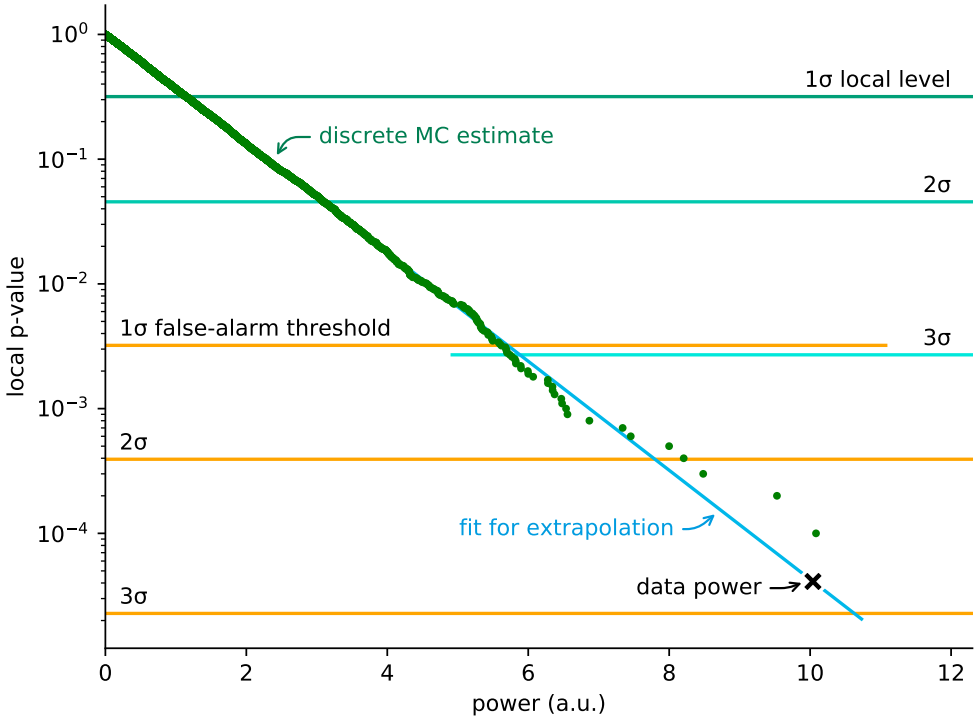

Figure 60: The estimate of the CDF of the power at 0.17 Hz. One-minus-CDF is plotted, so that the use of logarithmic scale details the high-power tail. The discrete estimate from MC samples is depicted with green dots. It is extrapolated with a fit, depicted by the straight blue line. The black cross marks the power in the toy time series at the frequency (x coordinate) and the corresponding local p-value (y coordinate). The local sigma levels are depicted in shades of green and the global ones in orange.

where A_i and B_i are free parameters. The result of the fit is depicted in Fig. 60 as a blue line. This line, different for each frequency, was then used to obtain the correspondence between the power and the local p-value for each frequency.

P-values are often expressed in terms of $n\sigma$, relating them to the ones of the normal distribution at integer multiples n of its width σ:

$$\text{p-value} = \operatorname{erfc}\left(\frac{n}{\sqrt{2}}\right) , \tag{50}$$

where erfc is the complementary error function (intuitively understood as "one minus the integral of the gaussian distribution"). The 1, 2 and 3σ levels of the power distribution are marked in shades of green in Figs. 59 and 60.

In the periodogram of the toy signal there are 15 peaks with p-values on a 2σ level. Naïvely, this may seem to be very improbable—a 2σ event is a five-in-a-hundred one. This is the so-called *look-elsewhere effect* [7], best explained as follows: a one-in-a-thousand event in a system is not a surprise, if it occurs in one of a thousand different systems. By looking at the Fig. 59 and comparing the periodogram of the signal with the sigma bands one essentially performs many, as many as the number of frequencies, largely independent statistical tests, cherry-picking among them the most significant peaks. The p-values p_i are called *local*, because they only measure the local significance at ω_i.

There is an another intuitive way of understanding this phenomenon. The height of the most significant peak is a statistic itself. Its distribution can also be estimated from the MC-generated null hypothesis signals. The distribution is depicted on the right-hand size in Fig. 59. Even when no signal is present, the most significant peak will most of the times have a height placing it between 2- and 3σ local significance bands.

Let us now consider more formally the local p-value of the most significant peak: $p_{\min}|H_0$ and its CDF. If the points of the periodogram were perfectly uncorrelated, this would be a simple case of many hypothesis testing [158]. The CDF of the maximum of a set of uncorrelated variables is a product of their CDFs [159] and the local p-values p_i are, by definition, uniformly distributed. So in this case we have

$$F_p(p) = p \tag{51}$$

$$F_{p_{\max}}(p) = p^N . \tag{52}$$

With $p' := 1 - p$:

$$F_{p_{\min}}(p) = 1 - F_{p'_{\max}}(p') = 1 - (1-p)^N , \tag{53}$$

where N is the number of frequencies tested. The CDF of $p_{\min}|H_0$, which we call F^g, can be estimated from the MC-generated data. Then the *global* p-value is given by

$$p^g = F^g(p^D_{\min}) , \tag{54}$$

where $p^D_{\min}$ is the minimal local p-value in the data set.

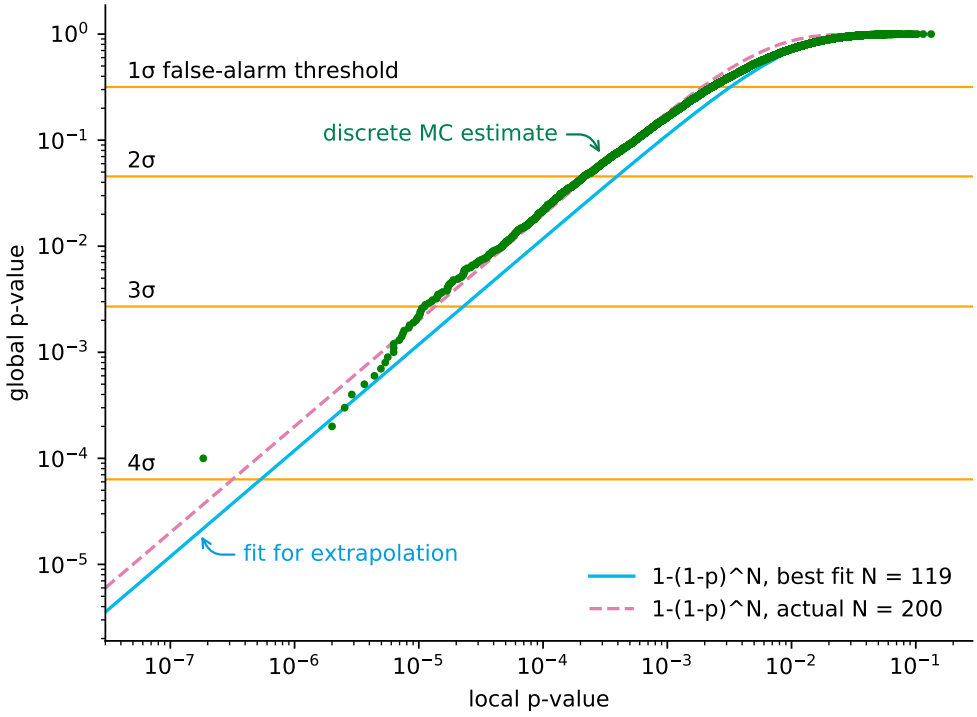

Figure 61: Quantitative illustration of the look-elsewhere effect in the toy data set. The discrete CDF estimate of the lowest local p-value, under the assumption of the null hypothesis, is plotted. For the given lowest local p-value in a periodogram the line gives the corresponding global p-value. The CDF for the case of perfectly uncorrelated power estimates (Eq. 53 with $N = 200$) is depicted with a pink dashed line. The blue line depicts the best-fit CDF with $N = 199$.

The discrete estimate of $p_{\min}\,|H_0$ CDF is depicted with green dots in Fig. 61. It was obtained by generating many periodograms under the assumption of H_0, evaluating the p-value of the most significant peak in each, and then ordering the p-values. The plot has the local p-value on the x-axis and the normalised position in the ordered array on the y-axis. As in the case of the local p-value estimation, in order to access the low-probability tail without straining the computational effort the functional form of the CDF was used. If the power estimates for all the of N tested frequencies are uncorrelated the CDF would be given exactly by Eq. 53. However, it cannot be guaranteed that the power estimates at different frequencies are independent in a case of a periodogram of an unevenly sampled time series. Instead, the parameter N was obtained by fitting a curve described by Eq. 53 to the discrete CDF estimate. CDFs corresponding to both values of N, the actual number of frequencies and the fitted one, are depicted in Fig. 61.

Going further, the global false-alarm thresholds can be determined. The *global* threshold p-value we call $p^g_{\text{f.a.}}$. The corresponding threshold for the local p-value is

$$p_{\text{f.a.}} = (F^g)^{-1}(p^g_{\text{f.a.}})\ . \tag{55}$$

For each frequency the threshold local p-value gives the threshold power:

$$P_i^{\text{f.a.}} = F_{P_i}^{-1}(1 - p_{\text{f.a.}})\ . \tag{56}$$

Note, that the local sigma levels in Fig. 60 and global ones in Fig. 61 have the same position on the y-axis.

The p-values for integer sigma levels are depicted in if Fig. 61 in orange. The 2σ false-alarm threshold for the toy time-series is depicted in Fig. 59, also in orange. It conveys an intuitive message: a single crossing of the 2σ false-alarm threshold anywhere would mean a 2σ confidence, that there is a statistically significant signal in the time-series (which there is). To claim a discovery of a significant oscillating signal, the false-alarm probability has to be at most in the range of 2.87×10^{-7} (5σ) [7].

In summary, the difficulty lies in determining the false-alarm thresholds. Then the detection boils down to comparing the periodogram of the time series to the thresholds on a plot similar to the one in Fig. 59.

8.3 SIGNAL HYPOTHESES TESTS

Should no claim for a discovery be possible, the next question to ask is:

Which oscillations would produce a visible peak, but did not, and can be thus excluded?

In order to answer this question the data need to be tested against being compatible with a number of model signal hypotheses. As an oscillation is characterised by its amplitude and frequency, the space of the hypotheses to test is two-dimensional.

The probability that a hypothetical oscillation of amplitude A and frequency ω would produce less power at frequency ω then observed is

$$\Pr\left(P^{H(\omega,A)}(\omega) < P^D(\omega)\right)\ . \tag{57}$$

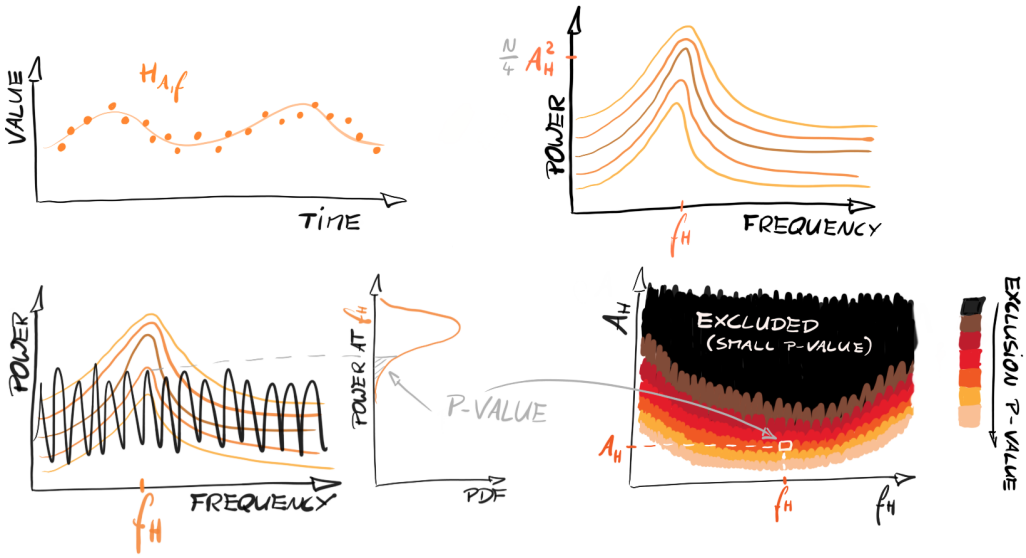

Figure 62: The general scheme for the determination of the exclusion region. First, a hypothesis about a signal, parametrised by its amplitude and frequency f_H, is assumed. Then, the distribution of the LSSA power at the frequency f_H is estimated under this assumption. The p-value of the time series' power, evaluated against the estimated distribution, is the measure of the confidence level on which the signal hypothesis can be rejected. This is repeated to cover the space of possible signals.

This probability is the p-value for the hypothesis $H(\omega, A)$ rejection. The distribution of $P^{H(\omega,A)}(\omega)$ was obtained with the Monte Carlo method in the following way. A signal was generated with a frequency ω and amplitude A. Then, the LSSA power at the frequency ω, $P^{H(\omega,A)}(\omega)$, was evaluated and compared with the one of the time series $P^D(\omega)$. This test was repeated for different ω and A, each time covering a pixel of the space of possible hypotheses. The procedure is schematically shown in Fig. 62. The set of hypotheses excluded at most at certain p-value forms an exclusion region. We took the threshold p-value to be 5 %, which corresponds to the confidence level of 95 %.

The frequency spectrum is covered much less densely than by the null hypothesis test. This procedure estimates the sensitivity of the measurement, solely determined by the timing and precision of the measurement points. No highly resonant structures are expected to appear therein. Also, a broader frequency range was covered, logarithmically spaced 10^{-3} Hz–10 Hz, in comparison to linearly spaced 5×10^{-3} Hz–1 Hz in the test of the null hypothesis. Thereby the behaviour of sensitivity of the method for extreme frequencies is illustrated.

During the Monte Carlo simulations a perfectly coherent signal is assumed. The width of a real axion-induced peak is not resolvable by the nEDM experiment.

The result of this procedure applied to the toy time series is presented on the left-hand side in Fig. 63. In the space of possible signals the colour depicts the confidence level for rejecting the signals. The black region, corresponding to high confidence, is excluded. The exclusion region goes down to small amplitudes only in the region between 10^{-2} Hz (the time series is around 200 s long) and 1 Hz (each of the toy measurements integrated the signal for one second).

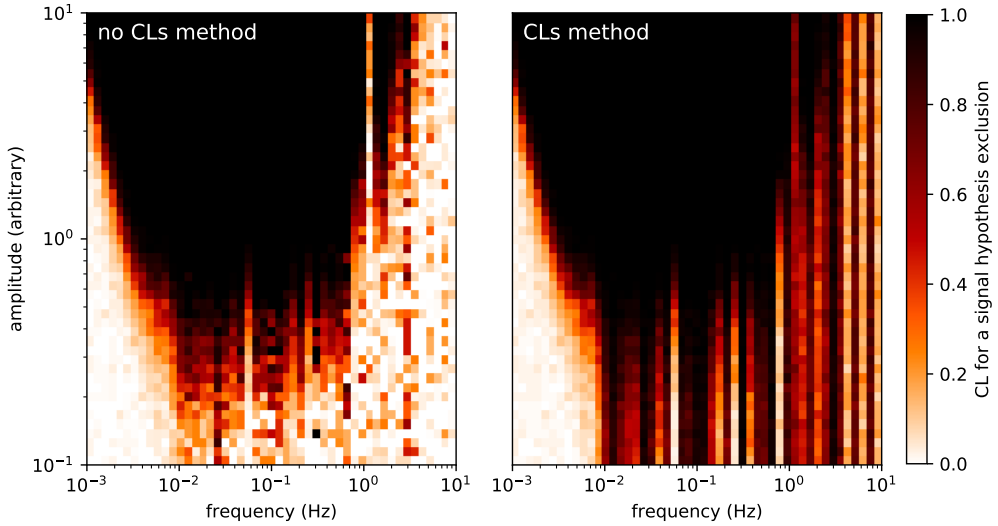

Figure 63: The toy time series tested against hypothetical signals. The signal space is spanned by their frequency and amplitude. The colour depicts the confidence level with which the signal can be rejected. The black region is excluded with a high confidence. *Left:* Without the CLs method. *Right:* The test with the use of the CLs method. Those hypotheses to which the measurement is not sensitive to get a statistical penalty.

Figure 64 illustrates the mechanism behind the loss of sensitivity for high and low frequencies. The average power obtained for various hypotheses is plotted together with the periodogram of the toy time series. Each coloured line depicts how high a peak caused by a generated signal is, as the function of the frequency of the signal. The signal peaks rise distinctly over the null hypothesis' periodogram only in a limited frequency range. Periods significantly longer than the length of the time series (below 10^{-2} Hz) are difficult to exclude, as it is always possible that the time series is located in an antinode of an extremely slow oscillation. This manifests itself as a high amplitude seen even when the null hypothesis is assumed. On the other end the power for frequencies above 1 Hz is suppressed, because the measurements are not point-like, but rather the oscillation is averaged over a period of 1 s. Only little power is observed, even for very large amplitudes of the signal.

The black exclusion region in the left-hand side of Fig. 63 exhibits a number of thin peaks going down to very low amplitudes. Seemingly for some frequencies even tiny signals can be confidently excluded. This is disturbing, and rightfully so. Consider, however, that as the power was evaluated for many frequencies, inevitably at some of them, roughly 5 %, the power is low enough to be rejected at the 95 % confidence level, even when tested against the distribution of power given the null hypothesis itself. It is completely fine from the statistical point of view. Yet, a situation where a hypothesis is rejected based on a measurement which was not sensitive to it is uncomfortable. One possible solution is called the *CLs method*. The method is defined, as well as the problem itself is discussed in detail, in the booklet of the Particle Data Group [7]. Here, a short graphical

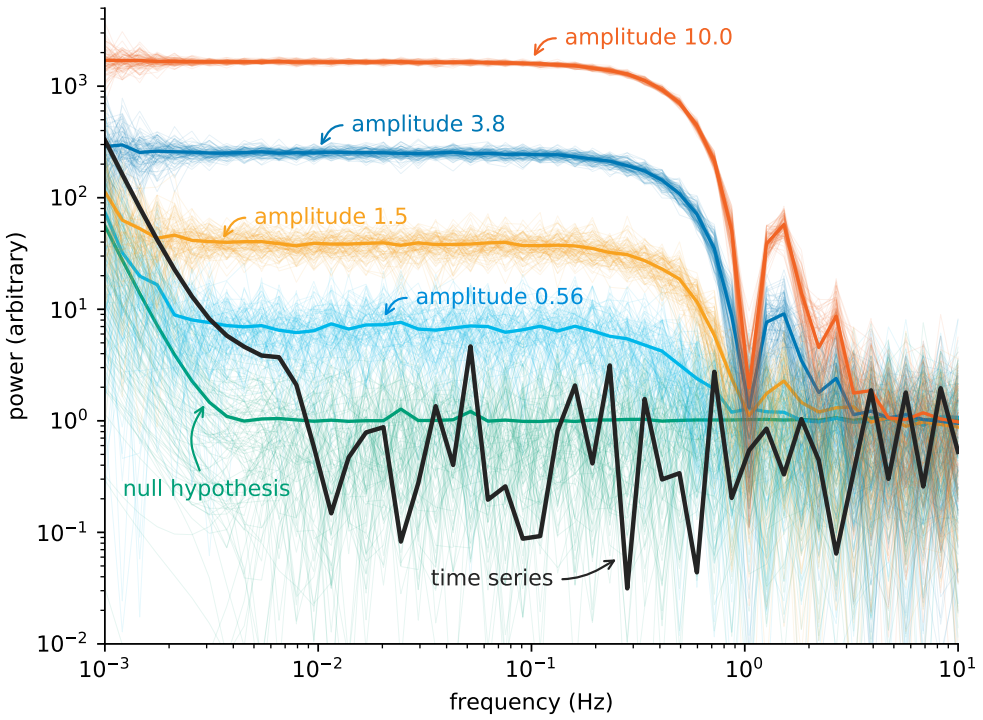

Figure 64: For each frequency, the LSSA power of a simulated signal of that frequency is plotted. Different colours correspond to different amplitudes of the signal. In particular no signal, the null hypothesis, is depicted in green. The lines have the interpretation of the height of a peak that would be observed for different frequencies (the x-axis) and amplitudes (colours). The thin lines represent the different simulation outcomes, the thick ones—their average. The black line is the periodogram of the toy time series itself.

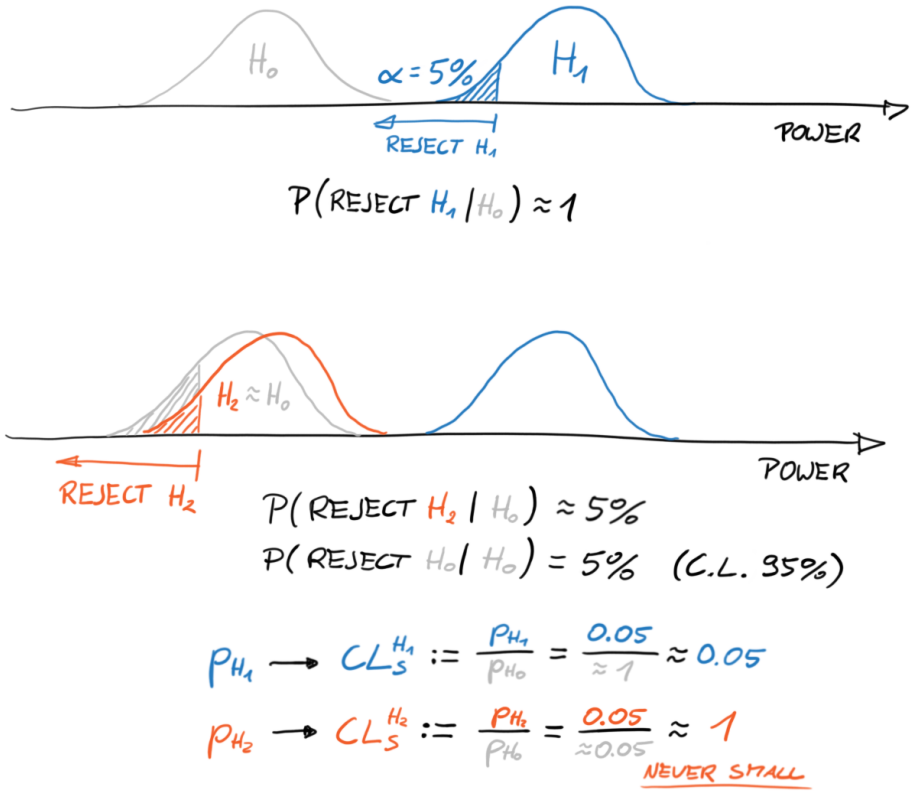

Figure 65: A graphical explanation of the motivation behind the CLs method. From the top: when the distributions of power for the null hypothesis H_0 and an alternative hypothesis H_1 are well separated, the probability of rejecting H_1 given H_0 is close to one. However, when we consider a signal of an arbitrarily small amplitude H_2, it still has roughly 5 % chance to be rejected, on a 95 % confidence level. If many of those are tested, 5 % will be unjustifiably rejected. In the CLs method one considers, rather than the p-value of an alternative hypothesis, the ratio of it to the p-value of the null hypothesis. This imposes a statistical penalty to the hypotheses not well separated from the null one.

explanation can be found in Fig. 65. With use of the CLs method the exclusion is suppressed in the region of low sensitivity, as shown on the right-hand side in Fig. 63.

Rather than calculating each pixel of the alternative hypotheses space, one may resolve only the 95 % C.L. threshold. This can be done, for example with a bisection algorithm run at each frequency.

The phase was treated as a nuisance parameter. In the MC simulations it was taken to be random, uniformly distributed between 0 and 2π. In general, though, the sensitivity may be phase-dependent. In particular, sensitivity for signals with periods much longer than the total span of the data set is linear for sin-like signals and quadratic for cos-like ones. This work does not consider the phase dependence and gives results with the phase marginalised.

In summary, there are three areas where the method loses sensitivity to detect oscillations, each with a different mechanism behind it. Firstly, only signals with amplitude large enough to produce more power than noise can be excluded. Secondly, signals with periods longer than the total span of the data set are difficult to exclude, as only a small part of an oscillation is seen. Thirdly, signals faster than the duration of a single measurement are integrated out, which reduces the sensitivity on the high-frequency side.

PERIODOGRAMS – CONCLUSION

In this chapter the least squares spectral analysis, LSSA, was introduced as a method to look for oscillations in unevenly sampled time series with unequal error bars. The test of the null hypothesis (no signal present) gives an estimate of the level of confidence on which a discovery of an oscillating signal can be claimed. Then a space of possible signals is explored to determine which ones can be excluded on the ground of not having been detected. In the next chapter this methodology is applied to the time series of the neutron electric dipole moment measurements performed at PSI. An oscillation there could be a hint of axion dark matter.

9 AXION ANALYSIS

In the previous chapter a methodology for searching for oscillations in an unevenly sampled time series was introduced. Here it is described how it was applied to look for oscillations in the neutron EDM data taken at PSI in the years 2015–17.

First, the scalar coupling of axions to gluons was considered, acting like an oscillating nEDM signal. The time series of R, the ratio of the spin-precession frequency of stored neutrons and ^{199}Hg atoms, was analysed. No significant signal was found, which allowed the first laboratory limits on the axion coupling to gluons to be put. This analysis was a joint effort with Nicholas Ayres, who analysed the data of the Grenoble-based nEDM measurement in search for the scalar coupling [160]. Rather than the raw R time series, he considered nEDM estimates as obtained on a run basis. The two analyses were complementary, each covering a different range of oscillation frequencies.

An analysis of the derivative coupling of axions to nucleons, acting like on oscillating magnetic field, was also performed. No significant discovery could be claimed, which led to an improvement upon previous laboratory limits.

Finally, a dedicated method to search for an oscillating nEDM is proposed, which could extend the sensitivity to frequencies up to hundreds of hertz.

9.1 THE PSI 2015–16 DATA SET

Let us shortly recapitulate the principle of the measurement (discussed in detail in Ch. 2). The main purpose of the nEDM experiment at PSI was to measure the static neutron electric dipole moment. To that purpose $R = \nu_n/\nu_{Hg}$, the ratio of spin-precession frequencies of neutrons and ^{199}Hg atoms, was measured in a combination of a magnetic and an electric field. A non-zero static nEDM would induce an electric-field-dependent shift in the spin-precession frequency of neutrons, and thereby in R, too. In a zero electric field there would be no shift, while the parallel and antiparallel configurations of the magnetic and electric fields would shift R in opposite directions. The nEDM was estimated based on those shifts. Due to the data blinding, a constant shift corresponding to an nEDM of order $10^{-25}\,e\,\mathrm{cm}$ was expected. Using R, that is normalising the precession frequency of neutrons to the one of ^{199}Hg atoms, suppressed the effect of homogeneous variations in the magnetic field. Even though the neutrons and ^{199}Hg atoms were stored together, the neutrons sagged due to their extremely low speed. Due to the sag vertical magnetic gradients of the magnetic field caused a shift in R. In order to correct for that effect the vertical gradient was modulated on a sequence basis (several hundred cycles).

In the PSI experiment the electric field was automatically changed according to a looped pattern: 48 cycles in one polarity, 8 cycles without the field, 48 cycles in the other, 8 without the field.

The data used in the analysis were collected at PSI between 2015–07–03 and 2016–12–18. No measurement dedicated for an oscillating nEDM search was

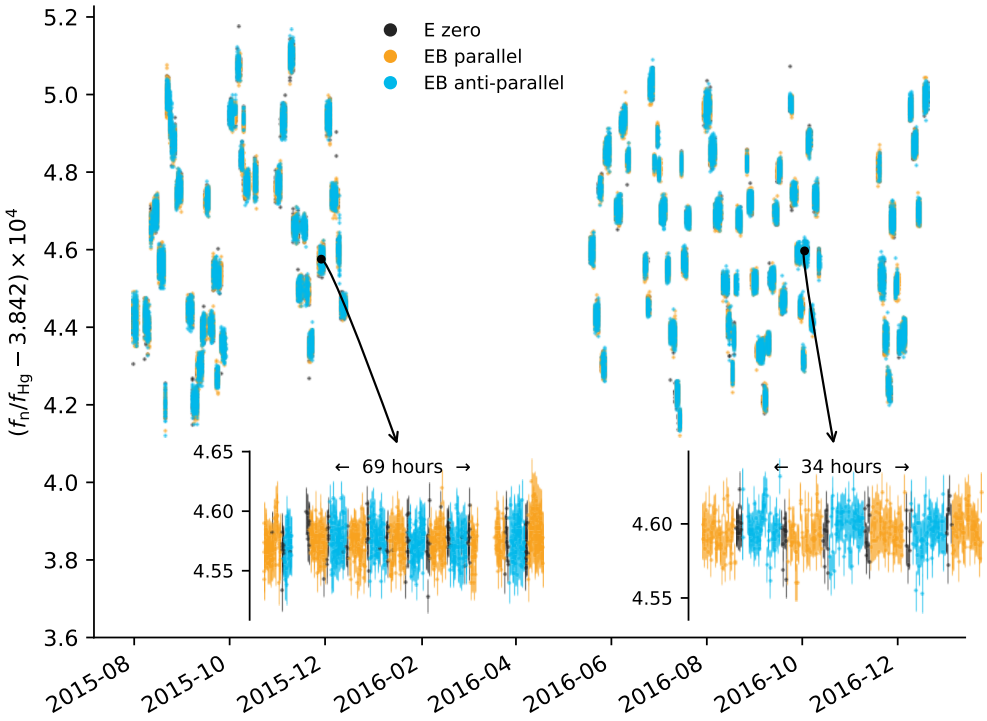

Figure 66: The complete $R = \nu_n/\nu_{Hg}$ time series used in the analysis, measured at PSI between July 2015 and December 2016. Two sequences are enlarged (one 69 hours long, the other 34). Due to the high density of the measurements individual points cannot be resolved. The colours depict the relative orientation of the magnetic and electric fields: $E \uparrow\uparrow B$ (orange), $E \uparrow\downarrow B$ (blue) and $E = 0$ (black). Between sequences the vertical gradient of the magnetic field was changed, so that those systematic effects, which are linear with the gradient could be interpolated to zero. This caused large shifts in R. The R time series has been corrected for gradient drifts (the correction is relative and does not extend from one sequence to another, for details see text). In the insets the points are drawn with their corresponding error bars. See also Fig. 67.

A technical term for uninterupted operation was a run. Sometimes a run was stopped due to technical reasons, and a new started afterwards. A sequence combines those consecutive runs, that could be one run if not for the interruption.

performed. The time series of R, the ratio of the spin-precession frequencies of the neutrons and ^{199}Hg atoms is presented in Fig. 66. Take a look first at the inset in the lower-right corner. It zooms into data collected within one *sequence*, typically 1–3 days long. During a sequence the apparatus completed one cycle after another, one every 300 s, each yielding an estimate of R. The electric field was automatically changed between three states: pointing upwards, being zero and pointing downwards. The different relative orientations of the electric and magnetic fields are depicted in colour in the figure. Sometimes there were technical breaks in the data taking during a sequence, as in the case of the one shown in the lower-left corner of Fig. 66.

A sequence was taken always in one magnetic field configuration. In between the sequences the vertical gradient was changed in 10 pT/cm steps up to ±60 pT/cm. In the static nEDM analysis it allowed for those systematic effects which scale linearly with the gradient to be extrapolated to zero (see Sec. 2.3).

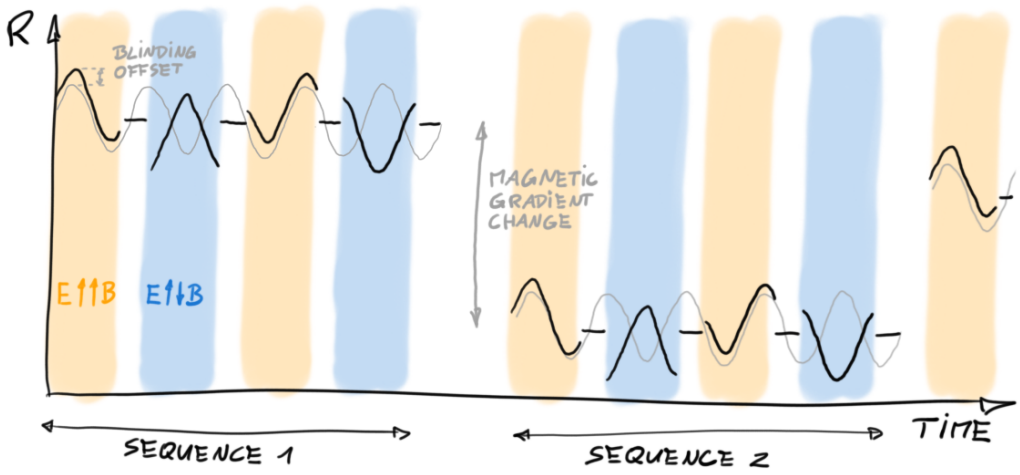

Figure 67: An illustration of an axion-induced oscillation in the $R = \nu_n/\nu_{Hg}$ time series. The colours indicate the configuration of the electric and magnetic fields: parallel to one another ($E \uparrow\uparrow B$, orange) and antiparallel ($E \uparrow\downarrow B$, blue). The oscillation in R has an opposite phase in the two configurations. The effect of a static nEDM (expected to be large due to the data blinding) is a field-configuration-dependent offset. Between sequences the vertical gradient of the magnetic field gradient was changed (so that those systematic effects, which are linear with the gradient could be interpolated to zero). This caused large shifts in R.

These large changes in the vertical gradient caused the large shifts in R from one sequence to another.

9.2 HOW A SIGNAL WOULD LOOK

Now we consider how an oscillating electric dipole moment would have affected the R time series, as measured by the PSI experiment. Should the neutron electric dipole moment oscillate, R would have oscillated as well, even when the electric field was constant. A reversal of the electric field polarity would have reversed the phase of the oscillations. With a zero electric field no oscillations would be visible. In Fig. 67 an R time series is depicted, with the combined effect of a large nEDM oscillation and the blinding offset. Because of the non-zero static nEDM and the phase shift the least-squares spectral analysis could not be applied directly to the R time series. However, the time series is still a part of a simple harmonic oscillation when only one field configuration is considered. (We neglect for the moment the inter-sequence shifts.) In the analysis the R time series was split thus into three: a series of points measured without the electric field (not sensitive to an oscillation of the nEDM), one with the electric and magnetic fields parallel and one with antiparallel. The last two would have the hypothetical oscillation in opposite phases. We will refer to the three data sets as $E = 0$, $E \uparrow\uparrow B$ and $E \uparrow\downarrow B$, respectively. Each of those was treated separately.

A part of the measurement procedure was to deliberately work in a substantial vertical magnetic field gradient. The gradient changed between sequences, causing large shifts in R, as seen in Fig. 66 and illustrated in Fig. 67. The prob-

lem of inter-sequence jumps was solved by allowing the DC offset in the LSSA fit to be different in each sequence:

$$A\sin(2\pi ft) + B\cos(2\pi ft) + \sum_i C_i\,\Pi_i(t)\ , \tag{58}$$

where C_i is the free offset in the ith sequence and $\Pi_i(t)$ is a gate function equal to one in the ith sequence and zero elsewhere. In this way the model could leverage the coherence of the oscillation across sequences despite the unknown offsets. The downside was a reduced sensitivity to oscillations slower than a sequence (2–3 days). These slow oscillations were largely absorbed into the different free offsets. With this modification the LSSA could be applied to the three data sets in a way as it is described in the previous chapter.

9.3 SYSTEMATIC EFFECTS

In the analysis the compatibility of the periodogram of the R time series with the one of pure noise, the null hypothesis, was tested. Variations in R could be considered systematic effects, as they would have resulted in an additional power in the periodogram.

Most prominently, R followed the changes in the vertical gradient of the magnetic field. The gradient changed not only between the sequences, which was accounted for by modifying the LSSA fit, but also drifted within each sequence. The in-sequence drifts were corrected for with the use of the caesium magnetometers. On a cycle basis, a second-order parametrisation of the field was fitted to their readouts, giving an estimate of the gradient [50, 51]. Due to unknown random offsets in the magnetometers' readings the correction was only relative, that is correcting only for the variations of the gradient. In between sequences the caesium magnetometers were calibrated, which altered the offsets. For this reason the correction could not extend across a sequence boundary.

There could have been, potentially, other effects causing the time series of R not to be fully random. An important decision had to be taken on how to treat those. Two possibilities were considered. The first would be a detailed study of time-dependent systematic effects. Here any excess in power, in any dataset ($E \uparrow\uparrow B$, $E \uparrow\downarrow B$ and $E = 0$) would be treated as a signature of some kind of a signal. All effects that could potentially result in that would have to be identified before the analysis was performed and corrected for. This would have required a long and careful systematic study. Moreover, the full-fledged systematic studies for the static nEDM analysis of the PSI data were still ongoing at the time. This approach was considered to be, albeit careful, not necessary for this analysis.

Instead, an easier approach was taken. An axion would produce a very specific signal, in particular:

1. There would be no signal in the $E = 0$ dataset.
2. The signal would appear in both $E \uparrow\uparrow B$ and $E \uparrow\downarrow B$ datasets, with equal amplitude.

3. The signals in $E \uparrow\uparrow B$ and $E \uparrow\downarrow B$ data sets would be shifted in phase by 180°.

4. The signals would have to have a high coherence of $\Delta\omega/\omega = 10^{-6}$ (cf. Ch. 7).

In a case when an excess in the power were observed it would only be called a candidate for an axion signal, if the four above conditions were met. Otherwise, it would be attributed to a, potentially unknown, systematic effect.

This made the systematic study dependent of the fact of having found a signal, opening a line of attack on the analysis. The analysis might be claimed not to have the right to exclude signals, because there might have been a systematic effect that cancelled a real signal out, but was never found nor even looked for. Nevertheless, such an event was highly improbable. In order to cancel an axion signal a systematic effect would have to be fine-tuned in its frequency over five orders of magnitude to a coherence of $\delta\omega/\omega = 10^{-6}$ *and* in its amplitude over ~ 20 orders of magnitude. As the exclusions are anyway probabilistic in nature, in this case a 95 % C.L. threshold is claimed, this approach was considered as justified.

9.4 OSCILLATING NEDM ANALYSIS

The LSSA periodogram of the $E \uparrow\uparrow B$ subset is presented in Fig. 68 in black. The average null-hypothesis periodogram is depicted in green and the false-alarm thresholds in orange. An inset details the region around inverse 300 s, the cycle frequency. There are two regions of expected rise in the oscillation amplitude due to the time structure of the data collection. The one around 28 µHz (the inverse of 10 hours) corresponds to the period of the reversal of the electric field. Each ramp caused a small break in the data taking (one cycle was missed). The other, around 3.3 mHz, the inverse of 300 s, corresponds to the cycle repetition rate.

Recall the discussion in Sec. 8.2 about the peaks in the periodogram solely due to the time structure of the series.

The periodogram of the R time series without the gradient-drift correction is shown in pink in Fig. 68. The correction had an effect only for frequencies slower than the period of the reversal of the electric field and around the narrow window around inverse 300 s. It should be pointed out, that 300 s is the approximate sampling period and as such it is likely to have some of the high power observed in low frequencies folded onto it (see the reasoning in Ref. [154]). The period of the electric field change had been deliberately chosen such, that was possibly infrequent, but still occurring when the magnetic field had not drifted significantly away.

In the periodogram of the gradient-corrected time series there are five trial frequencies for which the 3σ false-alarm threshold is exceeded, two of which, including the largest excess with a 6σ significance, occur in a 100 µHz region around the inverse of 300 s, while the other three are in the low-frequency region, longer than an inverse length of a sequence. The periodograms for the other two datasets, very similar to this one, can be found in App. C. In the other sensitive set there are three excesses above the 3σ threshold (the high-

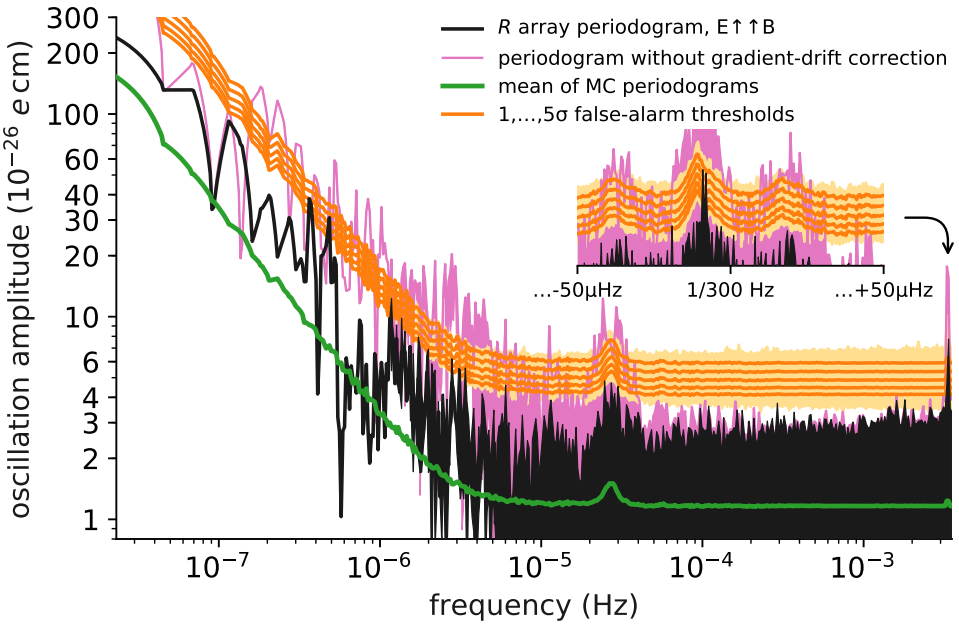

Figure 68: Periodogram of the R time series of the PSI experiment data, sensitive to oscillations in the quantity $d_n - \left(\mu_n/\mu_{Hg}\right) d_{Hg}$, taken with the **E** and **B** fields parallel (black line). The mean of MC-generated periodograms, assuming no signal, is depicted in green. MC is used to calculate 1, 2,...,5σ false-alarm thresholds, depicted in light orange. For clarity, we also plot the smoothed version in orange. There are two regions where a rise in the amplitude is expected, namely around 28 μHz (inverse of 10 hours) and 3.3 mHz (inverse of 300 seconds), due to the time structure of the data taking (see the main text for more details). The periodogram of non-gradient-drift-corrected data is shown in pink.

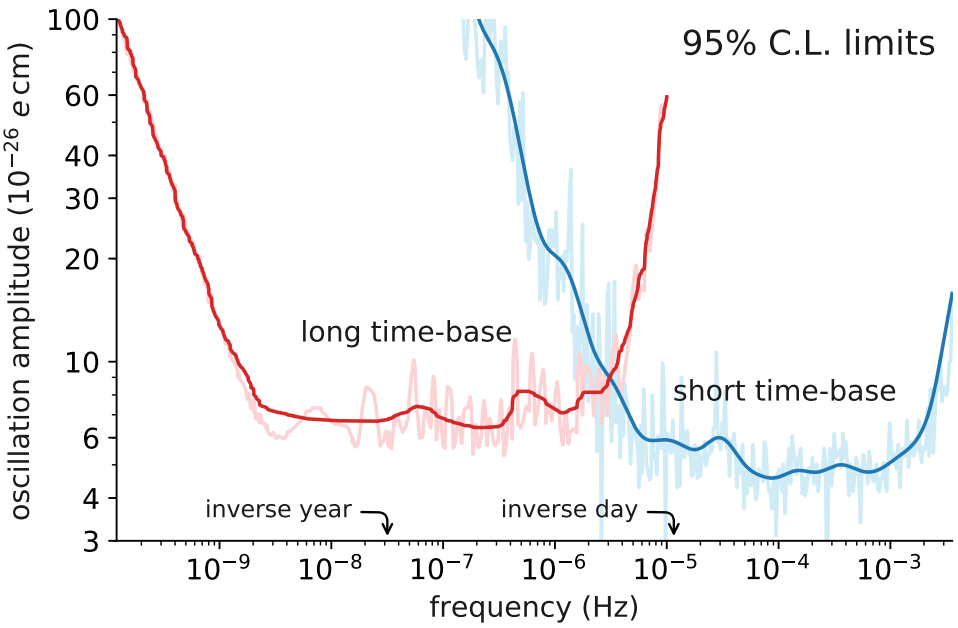

Figure 69: Limits on the amplitude of oscillation in the quantity $d_n - \frac{\mu_n}{\mu_{Hg}} d_{Hg}$, as a function of frequency thereof. The area above the curves is excluded on the 95 % C.L. The limit of this analysis (of the PSI data) are depicted in blue; the red curve depicts the limits of the complementary analysis of the ILL-based experiment's data [94, 160]. The numerically obtained limits are depicted with faint lines; the bold lines are smoothed.

est is 5σ), all constrained to the same two regions. In the control dataset, only the 1σ threshold is exceeded. None of the excesses fulfill the detection criteria, in particular the requirement to be present in both $E \uparrow\uparrow B$ and $E \uparrow\downarrow B$ periodograms with opposite phases.

As no significant signals have been observed, limits could be placed to exclude the observations that would have been detected. The limits, obtained numerically in a way discussed in Sec. 8.3, are depicted in Fig. 69 in blue (labelled "short time-base"). This analysis was most sensitive for periods between the duration of a sequence, around two days, and cycle repetition, 300 s. Amplitudes down to $5 \times 10^{-26}\, e\,\mathrm{cm}$ could be excluded on the 95 % C.L. The limits of the long time-base analysis, the one of the ILL-based experiment's data [94, 160], are depicted in red. Complementarily, they are sensitive to periods just below the duration of a sequence and go down to about a decade.

The oscillating-nEDM limits were interpreted as limits on the axion-gluon coupling (following the Eq. 41). The results are presented in the axion space, spanned by their mass and the strength of the coupling, in Fig. 70. The limits are presented in the landscape of already existing cosmological limits: axions with a mass below 10^{-22} eV have their Compton wavelength larger than the size of the smallest dwarf galaxies and, therefore, could not be the sole constituent of dark matter [109]; the influence of axions in the red-shaded area on the Big Bang nucleosynthesis would result in an underproduction of ^{4}He [129]; the

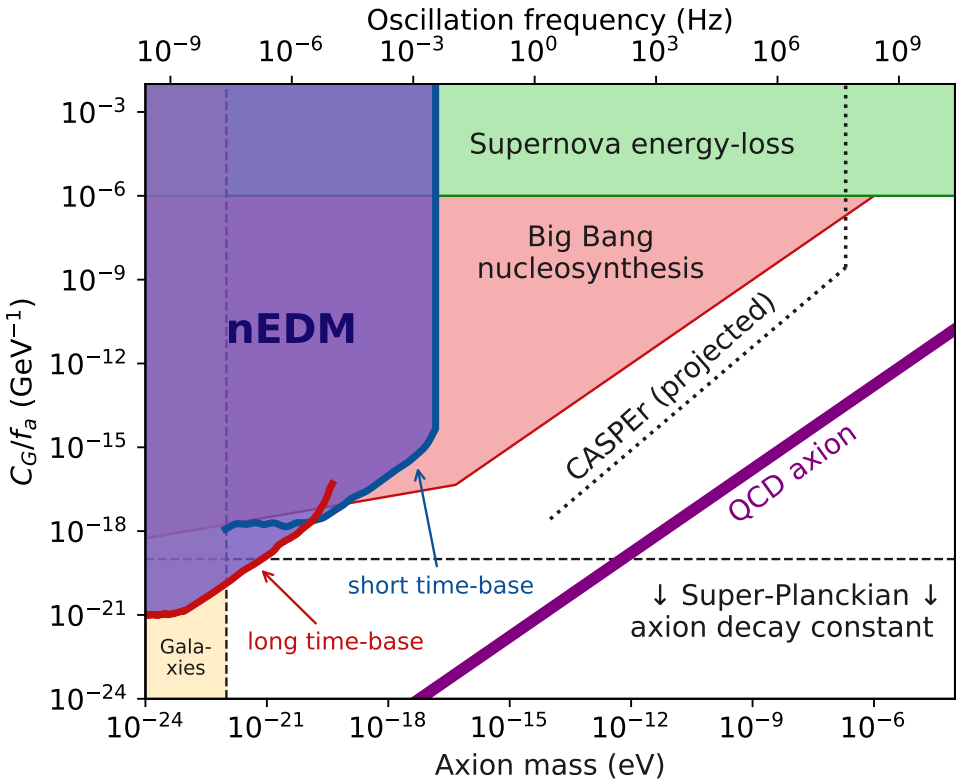

Figure 70: Limits on the interaction of an axion with the gluons (95 % C.L.). The parameter space is spanned by the axion's mass (horizontal) and the strength of the coupling (vertical). It has been assumed that axions saturate the local cold dark-matter density. Other depicted constraints are: Big Bang nucleosynthesis (red, 95 % C.L.) [129–131]; supernova energy-loss bounds (green, order of magnitude) [128, 132, 133] and consistency with observations of galaxies (orange) [109, 118–120]. The projected reach of the proposed CASPEr experiment is depicted with a dotted black line [161], and the parameter space for the canonical QCD axion with a purple band.

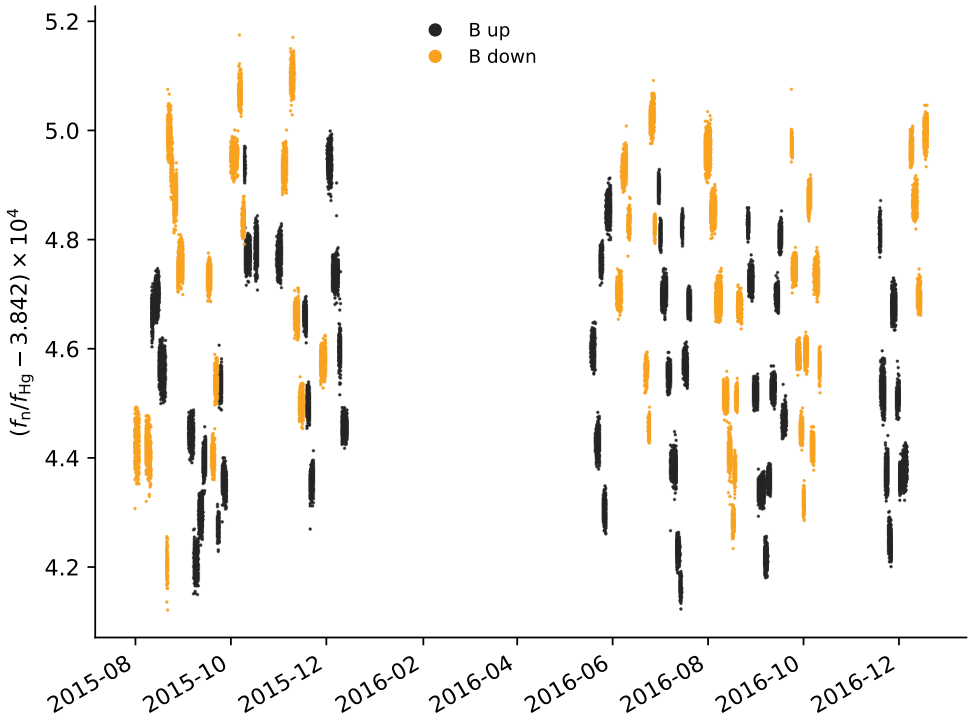

Figure 71: The R time series measured at PSI measured between July 2015 and December 2016 with the orientation of the magnetic field marked. The two orientations form the two datasets for the axion-wind analysis.

green area was excluded based on the observations of the supernova SN1987A, where excess cooling by axion emission would have been observed [128]. The limits are not only the first laboratory constraints on the axion-gluon coupling, but also improve on the existing cosmological exclusions.

9.5 AXION-WIND ANALYSIS

The analysis described so far was concerned with the scalar coupling of the axions to gluons, which looks like an oscillation in the electric dipole moment of the neutron. The same data set, and the same analysis techniques, were also used for a different coupling—a derivative one of axions to nucleons. This coupling acts like an additional dynamic magnetic field. The data were split based on the direction of the holding magnetic field B_0, as indicated in Fig. 71. The axion-wind coupling is insensitive to the electric field in the experiment.

The induced energy level shift, Eq. 43, is proportional to the projection of experiment's quantisation axis on the momentum of the axions. Because the latter arises due to the Earth's traversing the galactic axion field in the Solar System's movement around the Milky Way's centre, the effect is called the axion-wind [124]. The Earth additionally spins, causing the effect to be modulated with the sidereal frequency Ω_{sid}, equal to 23.934 469 9 hours [162]. The modulation would have produced a triplet of lines as a signal in the periodogram. The

The sidereal frequency is the one of the Earth's spinning as seen in the reference of distant stars.

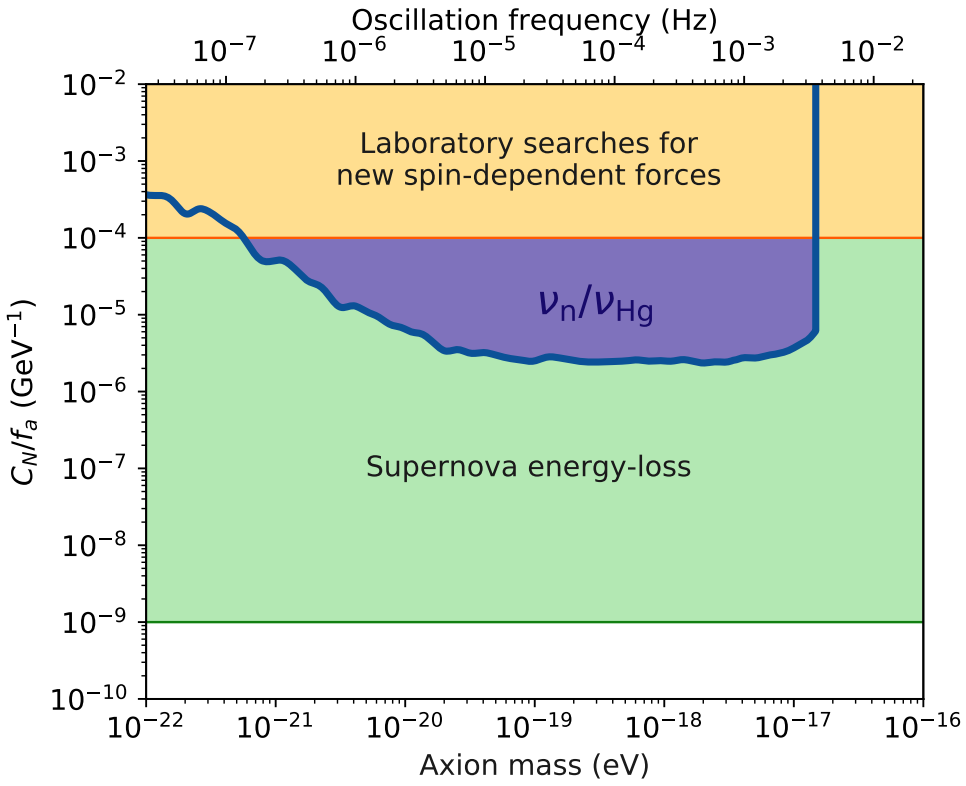

Figure 72: Excluded regions of the space of the coupling of axions to nucleons (Eq. 43). The green region is excluded from observations of the SN1987A supernova [94], and the yellow one from K–^{3}He magnetometry [134]. The blue region, is the exclusion arising from this analysis.

highest peak would be at the frequency of the oscillation of the axion field, and would be accompanied by two additional ones on either side, Ω_{sid} away from it.

There were two datasets, with the magnetic field pointing in either way along gravity, both sensitive to the effect. The periodograms can be found in App. C. There are 44 frequencies with power above the 3σ threshold in $B\uparrow$ and 36 in $B\downarrow$. Ony for two frequencies the threshold is exceeded in both data sets simultaneously: 3.429 69 µHz and 3.325 68 mHz. Neither fulfilled the requirement of the phase being opposite in the two data sets.

The lack of a statistically significant signal compatible with the axion model allowed us to put limits on the axion-nucleon coupling, depicted in Fig. 72. The limits improves upon the existing laboratory constraints by a factor of 40.

9.6 OUTLOOK

There are three directions in which the nEDM-based axion dark-matter search could continue. In order to improve sensitivity vertically, to be sensitive to more weakly coupled or less abundant axions, the overall sensitivity of the nEDM

measurement would need to be improved. Following Eq. 8 it would require more neutrons, a higher electric field or a longer spin-precession time. The global community already spares no effort in this respect.

The second way would be to improve sensitivity for slower oscillations, that is lighter axions. It was limited by the span of the data set—four years in the case of the ILL nEDM data set. Combining the ILL and PSI data into one time series would increase it to 19 years. It could not yet be done at the time for two reasons. Firstly, the PSI data were still blinded. Secondly, the static nEDM analysis, which would produce per-sequence nEDM estimates, was still ongoing. In any case, axions oscillating this slow would be so light, that their Compton wavelength would not fit in small dwarf galaxies [109], which rules them out as the sole dark-matter constituent.

The third direction would be the high-frequency, heavy-axions one. It was limited by the sampling frequency of the system, the cycle repetition rate. The measurements could be conducted with a shorter cycle time. This would, however, worsen the sensitivity, as the loss in Eq. 8 is linear, and the gain from the improved statistics scales with the square root. A repetition period faster than 10 s is hard to imagine. A real improvement in this direction would require changing the principle of the measurement.

9.7 RESONANT OSCILLATING NEDM SEARCH

The periodogram-based searches for dark matter were sensitive to a wide range of frequencies. In contrast resonant searches (for example ADMX [163], or the proposed CASPEr [161]) are sensitive at a any given time only to a relatively narrow band of frequencies. Covering a wide range requires scanning. In this section a resonant search of an oscillating nEDM is proposed, which would give access to faster oscillations than the periodogram-based method.

For polarised neutrons in a magnetic field a transverse oscillating coupling induces a coherent Rabi oscillation between the spin-up and spin-down states. For example, a Ramsey cycle begins with a $\pi/2$ flip induced by an oscillating transverse magnetic field, its frequency tuned to the Larmor one and its length tuned such, that the Rabi oscillation stops when the polarisation is in the transverse plane. Should the nEDM oscillate, an oscillating transverse coupling could be realised with a static electric field *perpendicular* to the holding magnetic field $\mathbf{B}$. Then, when the frequency of the nEDM oscillation is tuned to the Larmor one, the neutrons would undergo a Rabi nutation, which could be detected. Scanning the magnetic field in the range 0.1 µT–10 µT would cover frequencies 3 Hz–300 Hz.

In the proposed scheme, sketched in Fig. 73, the neutrons start polarised along the holding field and are put into the Larmor precession with a $\pi/2$ flip. The oscillating nEDM would then induce a Rabi nutation and a net polarisation along the $\mathbf{B}$ direction would build up. This net polarisation could be measured. The direction of the nutation depends on the phase difference between the Larmor precession (defined by the phase of the $\pi/2$ pulse) and the oscillating nEDM.

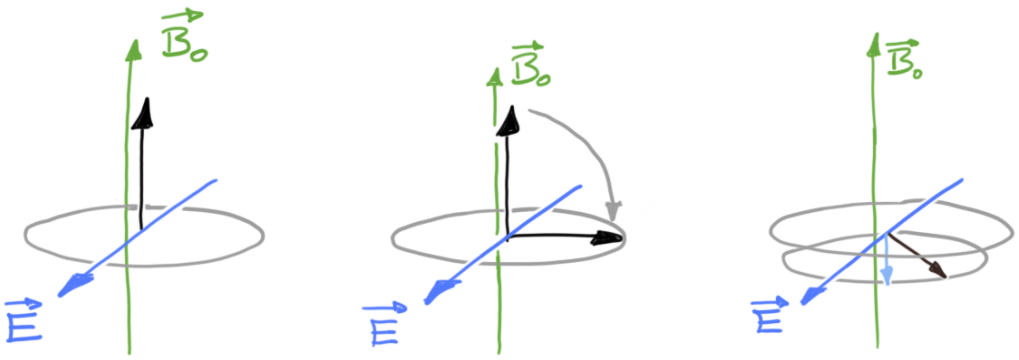

Figure 73: The behaviour of the spin population in a resonant oscillating EDM search. One starts with an ensemble polarised along the holding magnetic field B_0. The electric field is perpendicular. Then, the polarisation is flipped by $\pi/2$ and Larmor precession starts. The combination of the electric field and an oscillating EDM would induce a Rabi nutation. The direction of the nutation is determined by the relative phase of the Larmor precession and the EDM oscillation.

In order to estimate the sensitivity of such a search consider the Hamiltonian for a neutron in a magnetic field

$$H = -\boldsymbol{\mu}_\mathrm{n} \cdot \mathbf{B} = -\gamma_\mathrm{n} \mathbf{S} \cdot \mathbf{B} \tag{59}$$

for which the Larmor frequency is

$$\omega_0 = \frac{\mu_\mathrm{n} B}{S} = \frac{2\mu_\mathrm{n} B}{\hbar} = \gamma_\mathrm{n} B \,. \tag{60}$$

In the presence of an electric field perpendicular to $\mathbf{B}$ a harmonically oscillating nEDM adds an additional term to the Hamiltonian

$$H = -\gamma_\mathrm{n} \mathbf{B} \cdot \mathbf{S} - \mathbf{d}_\mathrm{n}^0 \sin(\omega_0 t) \cdot \mathbf{E} \,, \tag{61}$$

where we assume the Larmor frequency to be resonant with the oscillating nEDM, and a phase difference of 90° between the spin's precession and the nEDM oscillation. The Hamiltonian is the same when nEDM is static and the term $\sin(\omega_0 t)$ comes from an oscillating electric field $\mathbf{E}$. The oscillating field can be decomposed into two fields rotating in the opposite directions, each with the amplitude $E/2$. In a rotating frame spinning around $\mathbf{B}$ with the frequency ω_0 the field $\mathbf{B}$ vanishes. The $\mathbf{E}$ component which rotates with the spin is static [164]:

$$H_\mathrm{rot} = d_\mathrm{n}^0 \frac{\mathbf{E}}{2} \cdot \mathbf{S} \,, \tag{62}$$

whereas the other spin with frequency $2\omega_0$ and can be, in the first approximation, neglected [164]. In the rotating frame the spin precesses with the frequency

$$\omega_\mathrm{nut.} = \frac{d_\mathrm{n}^0 E}{2\hbar} \,. \tag{63}$$

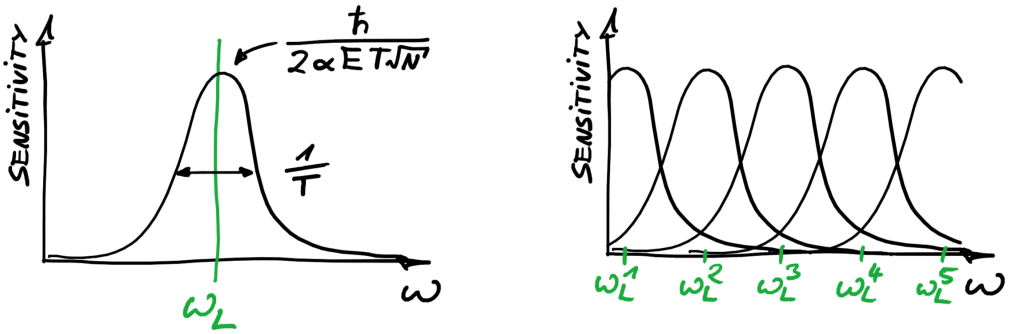

Figure 74: *Left:* the sensitivity of the resonant search has a width equal to the inverse interaction time. *Right:* a broad-band sensitivity needs to be built up with multiple measurements.

After time T the accumulated angle is

$$\theta = \omega_{\text{nut.}} T = \frac{d_{\text{n}}^0 E}{2\hbar} T \ . \tag{64}$$

At the end of the measurement the neutrons pass a spin analyser, which projects their polarisation on the axis of **B**. For N_0 neutrons the resonance curve has a shape

$$N_{\uparrow\downarrow}(\theta) = \frac{N_0}{2} \left(1 \pm \alpha \sin\theta\right) \ , \tag{65}$$

where α is the visibility parameter and the sign depends on which spin state is counted. A change in the neutron counts for a small angle $\delta\theta$ is:

$$\delta N = \frac{N_0}{2} \alpha \, \delta\theta \ . \tag{66}$$

Finally, the sensitivity is obtained from Eqs. 64 and 66, assuming it is dominated by the counting statistics $\sqrt{N_0/2}$

$$\sigma_{d_{\text{n}}^0} = \frac{2\hbar}{\sqrt{N_0/2}\alpha T E} \ . \tag{67}$$

When both spin states are counted the sensitivity increases by a factor $\sqrt{2}$:

$$\sigma_{d_{\text{n}}^0} = \frac{2\hbar}{\sqrt{N_0}\alpha T E} \ . \tag{68}$$

A 90° phase difference between the nEDM oscillation and spin precession (defined by the phase of the $\pi/2$ pulse) is assumed. On average an additional factor of $\sqrt{2}$ is lost due to the non-matching phases. Doing two measurements with the phases of the $\pi/2$ pulses shifted by 90°, one having a sine-like sensitivity other a cosine-like, gives a flat sensitivity for all phases of the nEDM oscillation.

The search is resonant—at any given time the full sensitivity is achieved only for the nEDM oscillation frequency equal to the Larmor one. To cover a wide

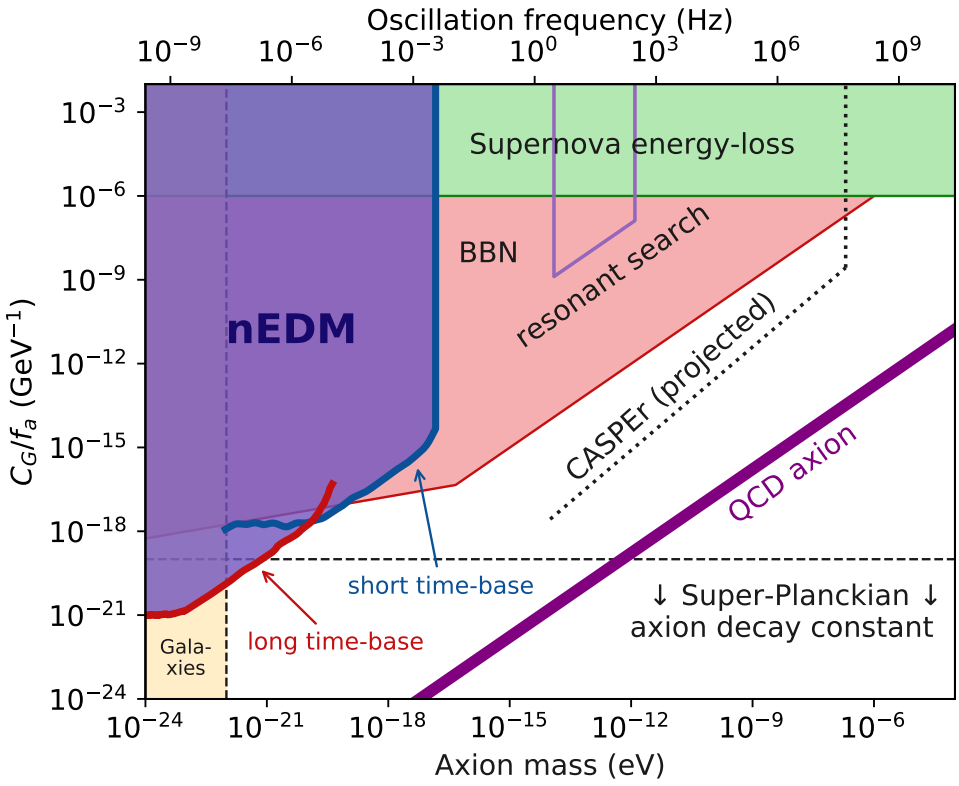

Figure 75: Sensitivity prediction for a 1000 cycles long resonant oscillating nEDM search. A scan of the magnetic field in the range 0.1 µT–10 µT gives the interaction time $T = 3\,\mathrm{s}$. The parameters of the baseline n2EDM design were used: $\alpha = 0.8$, $E = 15\,\mathrm{kV\,cm^{-1}}$, $N_0 = 121\,000$. See Fig. 70 for a detailed of the other limits.

spectrum the strength of the magnetic field **B** would need to be scanned. The width of the sensitivity peak is $1/T$; the broad-band search would need to be built up from many of those, as depicted in Fig. 74.

In order to perform a 3 Hz–300 Hz scan in 1000 cycles (3 days of operation of the n2EDM experiment at PSI) the width of the resonance would need to be

$$T = \left(\frac{1}{1000}\,(300\,\mathrm{Hz} - 3\,\mathrm{Hz})\right)^{-1} \approx 3\,\mathrm{s}\,. \tag{69}$$

With the baseline parameters of the n2EDM experiment at PSI, $\alpha = 0.8$, $E = 15\,\mathrm{kV\,cm^{-1}}$, $N_0 = 121\,000$, the per-cycle sensitivity would be

$$\sigma_{d_\mathrm{n}} = 1 \times 10^{-22}\,e\,\mathrm{cm}\,. \tag{70}$$

The region of an axion coupling to gluons that could be excluded is depicted in Fig. 75.

An important systematic effect would be the quality of the $\pi/2$ pulse. Polarisation remaining after it could be misinterpreted as a signal. However, the effect scales linearly with the electric field E and the interaction time T. Both

could be varied, in particular the electric field reversed, in a series of coherent measurements, which would need to be performed at each frequency. The length of a series is limited by the coherence of the expected signal, 10^6 in the case of an axion-induced nEDM oscillation.

AXION ANALYSIS – CONCLUSION

Ultralight axion-like particles are compelling candidates for dark matter. Besides interacting gravitationally they could couple to gluons and nucleons. This coupling could be detected by the nEDM measurement at PSI. Axions could induce harmonic oscillations in the time series of the ratio of the spin-precession frequencies of stored neutrons and ^{199}Hg atoms. The least squares spectral analysis was used to produce periodograms of the data measured at PSI. The statistical treatment of the periodograms, largely based on Monte Carlo simulations, resulted in no significant signal. The analysis of the electric-field correlated data gave the first laboratory constraints on the scalar coupling of axions to gluons and improved an astrophysical limits by up to three orders of magnitude. The null result of the search in the magnetic-field correlated data improved the existing laboratory constraints by up to a factor of 40.

CONCLUSION

At the time of writing this thesis the analysis of the measurement of the electric dipole moment of the neutron performed at PSI was still ongoing. Once finished, it will give the most precise estimate of the nEDM to date. It might even be the first to hint an non-zero value. When compared with the values predicted by the various extensions of the Standard Model it will support some and contradict others, shedding light onto our understanding of the Universe. This will be an addition to a contribution already made as part of this work—the search for axion dark matter. At PSI the quest for the neutron electric dipole moment will continue with a new, more sensitive apparatus. A part of it is an active magnetic shield, whose design this work was also concerned with.

The new coil design method made it possible to incorporate active magnetic shields in tight spatial constraints. The grid-based design was practically demonstrated in a form of a small-scale active shield. This technology served as a base for a design of a shield for the n2EDM experiment at PSI. The n2EDM shield will perform better if it is tailored for the particular magnetic environment by featuring coils for high-order variations. To that purpose a magnetic field mapper was built—a mobile tower with sensors attached to it, which was used to map the field in the n2EDM areal.

The search for axion dark matter with the data measured in the nEDM experiment at PSI was the first of its kind. The ratio of the spin-precession frequencies of stored neutrons and mercury atoms was checked for statistically significant, axion-induced oscillations. None were found, which resulted in the first laboratory limits for the axion coupling to gluons and an improvement on the ones on the axion-nucleon coupling. In the future a proposed resonant detection scheme may be used to use ultracold neutrons to search for heavier axions.

ACKNOWLEDGEMENTS

I would like to thank Klaus Kirch for giving me an opportunity to pursue a PhD in his group, for his strategic guidance and for sharing his expertise. I thank Florian Piegsa and Jochen Krempel for their day-to-day advice. The search for the axions would never happen without the theoretical support of Victor Flambaum, Malcolm Fairbairn, Yevgeny Stadnik and Doddy (David J.E. Marsh). I greatly enjoyed collaborating with Nick Ayres on the axion analysis. I would never be able to work my way through the data of the nEDM experiment without Elise Wursten's help. I am grateful for insightful discussions about magnetic field coils I had with Gilles Quéméner and Chris Crawford. Towards the end of my work I closely collaborated with Solange Emmenegger. Thank you for all the laughs. Furthermore, I am grateful to the whole nEDM collaboration. The ideas presented in this thesis were discussed over and over again in the friendly and inspiring atmosphere of our meetings. I would like thank, in particular, Philipp Schmidt-Wellenburg, Georg Bison, Bernhard Lauss, Dieter Ries, Sybille Komposch, Chris Abel, Nicolas Hild, Vira Bondar, Małgorzata Kasprzak, Hans-Christian Koch, Guillaume Pignol, Philip Harris and all the other members of the collaboration. I am grateful to the group at the Jagiellonian University in Cracow, Poland, where I did my Master's project. My thanks goes to Kazimierz Bodek, Jacek Zejma, Dagmara Rozpędzik, Grzegorz Wyszyński and Adam Kozela for introducing me not only to the nEDM community, but also to the trade of experimental physics itself. During my PhD I spent most of my working time at ETH, where I enjoyed many discussions during breaks for coffee, tea and juggling. My thanks goes to Andreas Eggenberger, Aldo Antognini, Ivana Belošević, Kim Siang Khaw, Karsten Schuhmann, Gunther Wichmann, Mirosław Marszałek, Zachary Hodge and Manuel Zeyen. I benefited from an excellent technical support of Michi Meier and Fritz Burri at PSI, and Bruno Zehr at ETH. I did a lot of the laboratory work together with apprentices who, despite their young age, amazed me with their skills. Thank you, Mona Hänni, Mirco Dill, Moritz Stöckli and Cyrill Strässle. I also greatly enjoyed working with students. I thank Tizian Bluntschli, Nick Schwegler, Gianluca Janka, Avraam Chatzimichailidis, Daria Cegiełka, Enrico del Re and Hanno Bertle for putting heart in their work. I would be completely lost without administrative help from Anita van Loon, Rosa Bächli, Gaby Amstutz, Bettina Lareida and Caroline Keufer-Platz. For all their support I thank my parents Gabriela and Grzegorz. Most importantly, I owe a great many thanks to my wonderful wife Agata and our darling son Jędrzej.

TOOLS

During my work I used many open-source tools: matplotlib [165], Python, Julia [90], numpy, scipy [166], jupyter [167], pyqtgraph, ipython [168], pandas [169], inkscape, Ipe and countless libraries for Julia and Python. I am grateful to the countless contributors to these remarkable pieces of software. This thesis was typeset using the *classicthesis* template by André Miede.

Part IV

APPENDIX

A

DERIVATION OF THE R-RATIO

Here we derive the expression for the ratio of the neutron and mercury spin-precession frequencies

$$R = \frac{\nu_{\mathrm{n}}}{\nu_{\mathrm{Hg}}} \,. \tag{71}$$

For a spin $S = \frac{1}{2}$ particle in a combination of an electric and magnetic field we have:

$$H = -2\left(\mu\,\mathbf{B} + d\,\mathbf{E}\right)\cdot\mathbf{S} \,. \tag{72}$$

For the parallel (p) and antiparallel (ap) combination of the fields:

$$H_{\mathrm{p,\,ap}} = -\mu B \pm dE \,. \tag{73}$$

The transition between the spin-up and spin-down states is twice the energy:

$$H_{\mathrm{p}\leftrightarrow\mathrm{ap}} = h\nu = 2\left(\mu B \pm dE\right) \,. \tag{74}$$

Substituting into Eq. 71:

$$\begin{aligned} R = \frac{\nu_{\mathrm{n}}}{\nu_{\mathrm{Hg}}} &= \frac{\frac{2}{\hbar}\left(\mu_{\mathrm{n}}B + d_{\mathrm{n}}E\right)}{\frac{2}{\hbar}\left(\mu_{\mathrm{Hg}}B + d_{\mathrm{Hg}}E\right)} \\ &= \frac{\left(\mu_{\mathrm{n}}B \pm d_{\mathrm{n}}E\right)}{\left(\mu_{\mathrm{Hg}}B \pm d_{\mathrm{Hg}}E\right)} \\ &= \frac{\mu_{\mathrm{n}}B}{\left(\mu_{\mathrm{Hg}}B \pm d_{\mathrm{Hg}}E\right)} \pm \frac{d_{\mathrm{n}}E}{\left(\mu_{\mathrm{Hg}}B \pm d_{\mathrm{Hg}}E\right)} \\ &= \frac{\mu_{\mathrm{n}}}{\mu_{\mathrm{Hg}}} \times \frac{1}{1 \pm \frac{d_{\mathrm{Hg}}E}{\mu_{\mathrm{Hg}}B}} \pm \frac{d_{\mathrm{n}}E}{\mu_{\mathrm{Hg}}B} \times \frac{1}{1 \pm \frac{d_{\mathrm{Hg}}E}{\mu_{\mathrm{Hg}}B}} \,. \end{aligned} \tag{75}$$

Since $(1 \pm x)^{-1} \approx 1 \mp x$,

$$R \approx \frac{\mu_{\mathrm{n}}}{\mu_{\mathrm{Hg}}} \mp d_{\mathrm{Hg}}\frac{\mu_{\mathrm{n}}}{\mu_{\mathrm{Hg}}}\frac{E}{\mu_{\mathrm{Hg}}B} \pm d_{\mathrm{n}}\frac{E}{\mu_{\mathrm{Hg}}B} \mp d_{\mathrm{n}}d_{\mathrm{Hg}}\left(\frac{E}{\mu_{\mathrm{Hg}}B}\right)^2 = \tag{76}$$

$$\approx \frac{\mu_{\mathrm{n}}}{\mu_{\mathrm{Hg}}} \pm \left(d_{\mathrm{n}} \mp d_{\mathrm{Hg}}\frac{\mu_{\mathrm{n}}}{\mu_{\mathrm{Hg}}}\right)\frac{E}{\mu_{\mathrm{Hg}}B} \,. \tag{77}$$

We approximate the magnetic field with

$$h\nu_{\mathrm{Hg}} \approx 2\mu_{\mathrm{Hg}}B \;\Rightarrow\; \mu_{\mathrm{Hg}}B \approx \frac{h\nu_{\mathrm{Hg}}}{2} \tag{78}$$

and substitute to get

$$R = \frac{\nu_{\mathrm{n}}}{\nu_{\mathrm{Hg}}} = \frac{\mu_{\mathrm{n}}}{\mu_{\mathrm{Hg}}} \pm \left(d_{\mathrm{n}} \mp \frac{\mu_{\mathrm{n}}}{\mu_{\mathrm{Hg}}}\,d_{\mathrm{Hg}}\right)\frac{2E}{h\nu_{\mathrm{Hg}}} + \Delta \,. \tag{79}$$

B

MAPS OF THE CHALET ROOM AT LPSC

The "chalet" room at LPSC in Grenoble, France, was mapped with the prototype of the mapper (see Ch. 6). A panorama of the room is shown in Fig. 76. Figure 77 shows the map of the magnitude of the field. A large dipole source is visible to the right, which was attributed to a large radiator, visible on the photograph below the window on the far wall.

Figure 76: A photograph of the "chalet" room at LPSC Grenoble. On the picture visible are: the "L-piece" holding string potentiometers (to the right), the mapping tower with magnetic field sensors, a large steel radiator (below the window).

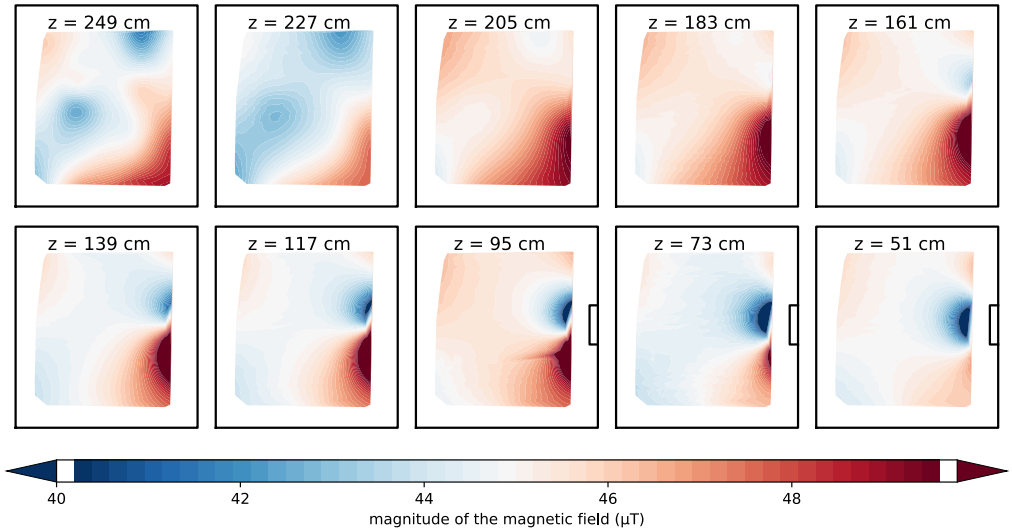

Figure 77: A map of the magnitude of the magnetic field in the "chalet" room at LPSC in Grenoble, France. Each tile depicts measurements in a horizontal plane, as acquired by each sensor.

C

ADDITIONAL PERIODOGRAMS

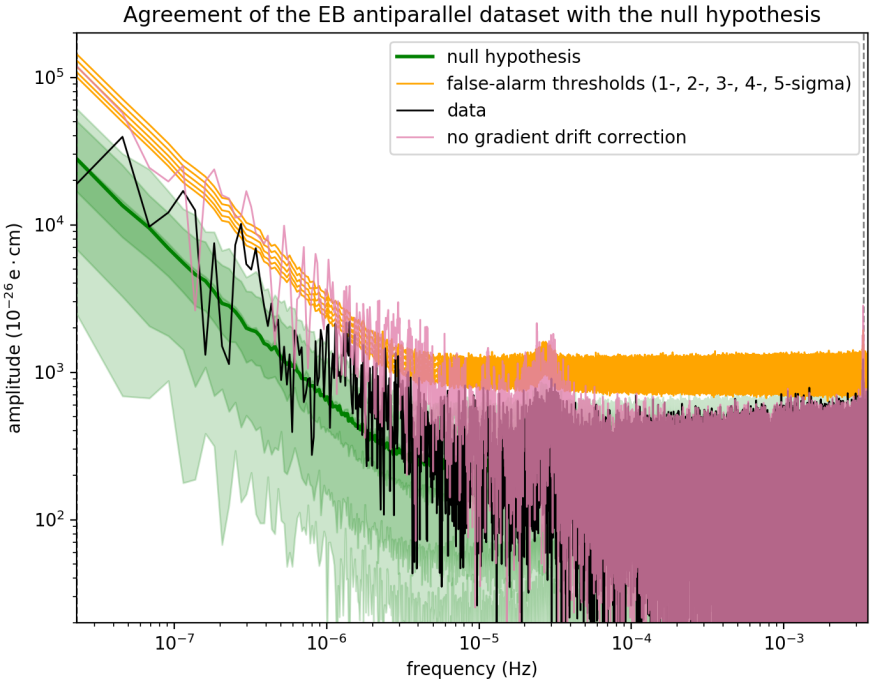

Figure 78: Periodogram of the R time series of the PSI experiment data, sensitive to oscillations in the quantity $d_n - (\mu_n/\mu_{Hg})\, d_{Hg}$, taken with the **E** and **B** fields antiparallel (black line). The distribution of the MC-generated periodograms is depicted in green. The thick line depicts the mean, around which 1, 2 and 3σ bands of the distribution are marked. around it marked. MC is used to calculate 1, 2,...,5σ false-alarm thresholds, depicted in orange. The periodogram of non-gradient-drift-corrected data is shown in pink. For details see Ch. 9.

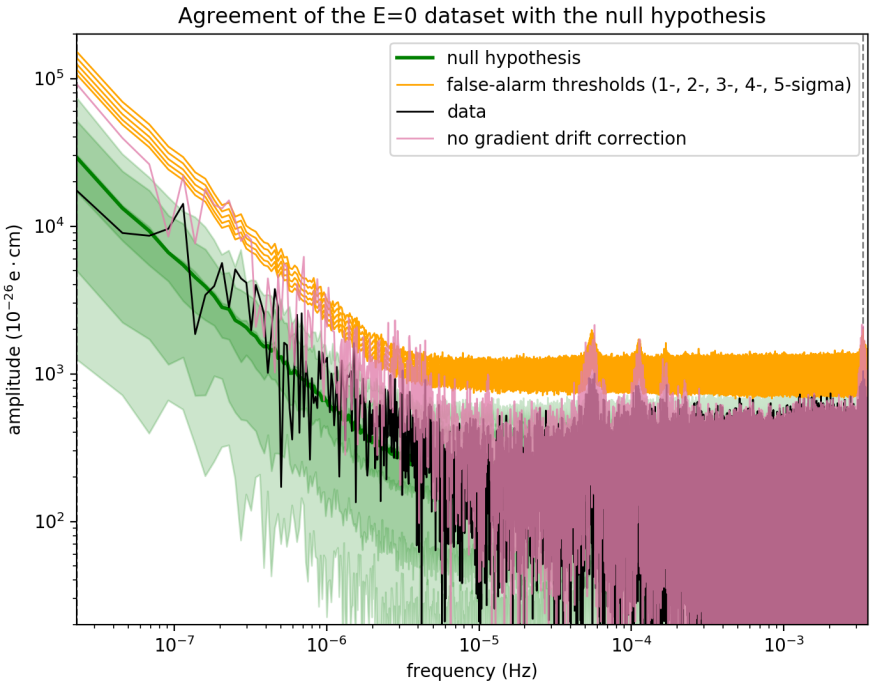

Figure 79: Periodogram of the R time series of the PSI experiment data, not sensitive to nEDM effects, taken with the $\mathbf{E} = 0$ (black line). The distribution of the MC-generated periodograms is depicted in green. The thick line depicts the mean, around which 1, 2 and 3σ bands of the distribution are marked. around it marked. MC is used to calculate 1, 2,...,5σ false-alarm thresholds, depicted in orange. The periodogram of non-gradient-drift-corrected data is shown in pink. For details see Ch. 9.

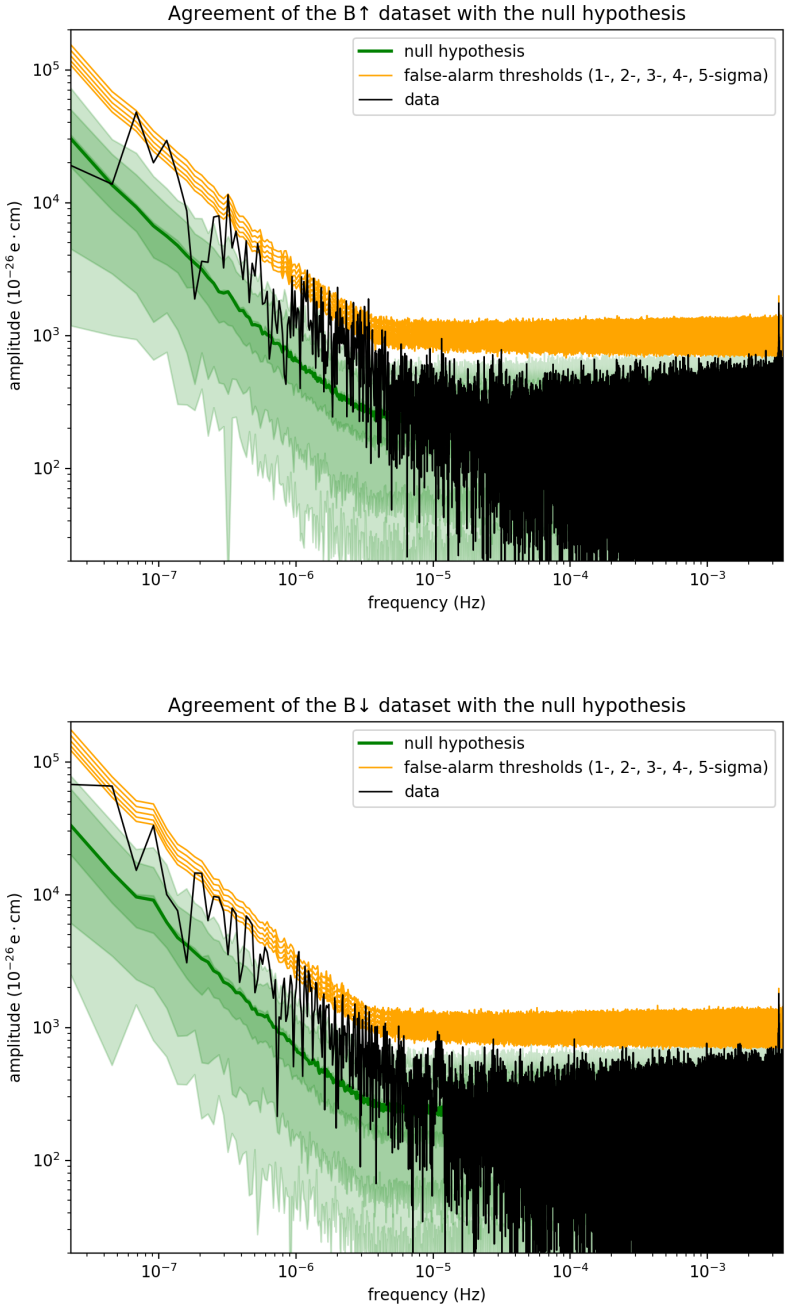

Figure 80: The periodograms of the R time series measured with the main magnetic field pointing upwards (top) and downwards (bottom). The distribution of the MC-generated periodograms is depicted in green. The thick line depicts the mean, around which 1, 2 and 3σ bands of the distribution are marked. around it marked. MC is used to calculate 1, 2,...,5σ false-alarm thresholds, depicted in orange. For details see Ch. 9.

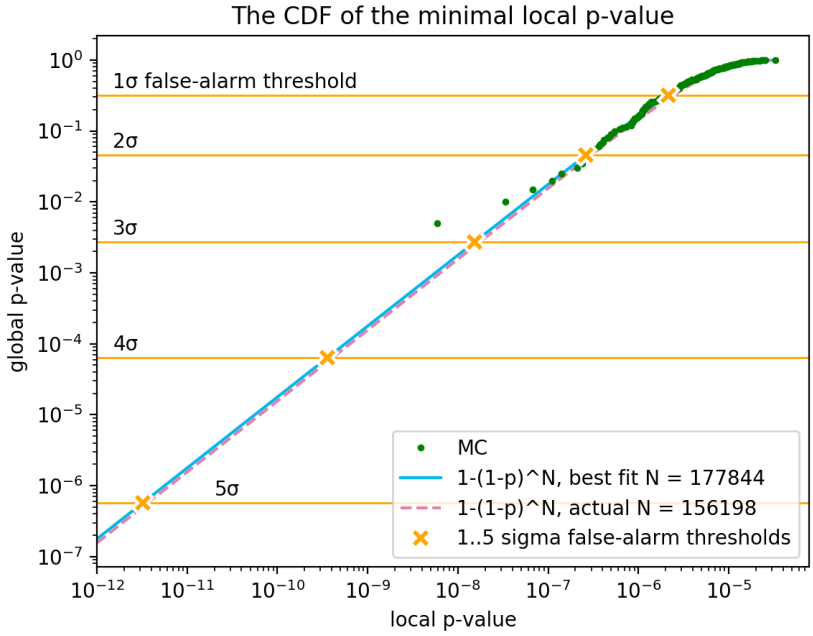

Figure 81: Quantitative illustration of the look-elsewhere effect in the R time series of the PSI experiment data, sensitive to oscillations in the quantity $d_{\mathrm{n}} - \left(\mu_{\mathrm{n}}/\mu_{\mathrm{Hg}}\right) d_{\mathrm{Hg}}$, taken with the **E** and **B** fields parallel. See the caption of Fig. 61 for details.

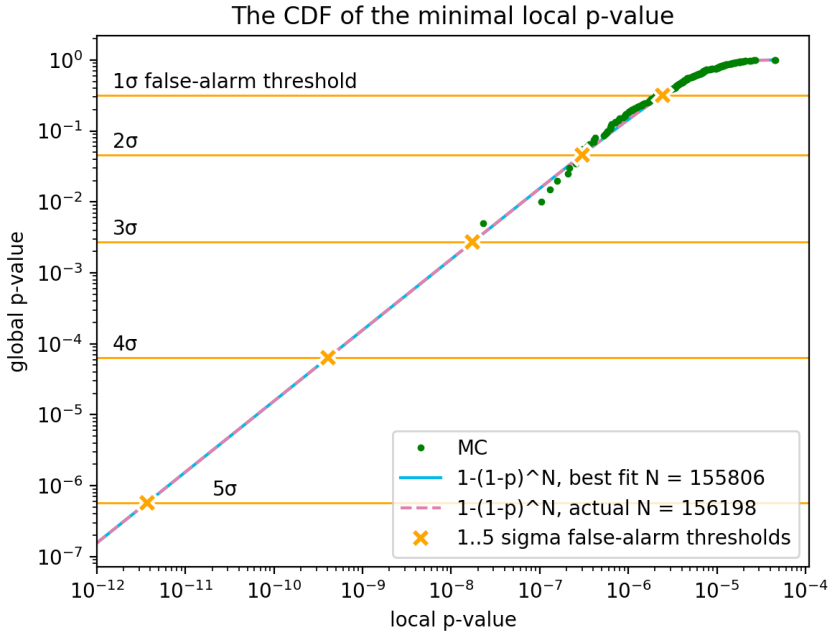

Figure 82: Quantitative illustration of the look-elsewhere effect in the R time series of the PSI experiment data, sensitive to oscillations in the quantity $d_{\mathrm{n}} - \left(\mu_{\mathrm{n}}/\mu_{\mathrm{Hg}}\right) d_{\mathrm{Hg}}$, taken with the **E** and **B** fields antiparallel. See the caption of Fig. 61 for details.

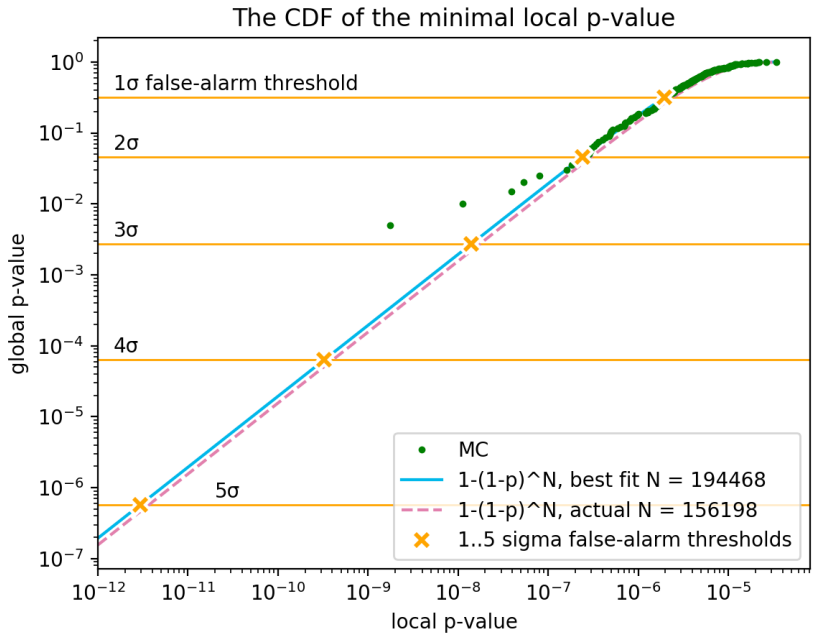

Figure 83: Quantitative illustration of the look-elsewhere effect in the R time series of the PSI experiment data, not sensitive to nEDM effects, taken with the $\mathbf{E} = 0$. See the caption of Fig. 61 for details.

BIBLIOGRAPHY

[1] E. Noether, Nachr. D. König. Gesellsch. D. Wiss. Zu Göttingen , 235 (1918).

[2] R. G. Sachs, *The Physics of Time Reversal*, 1st ed. (University of Chicago Press, Chicago, 1987).

[3] A. D. Sakharov, Soviet Physics Uspekhi **34**, 392 (1991).

[4] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, Phys. Rev. **105**, 1413 (1957).

[5] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, Phys. Rev. Lett. **13**, 138 (1964).

[6] M. Kobayashi and T. Maskawa, Progress of Theoretical Physics **49**, 652 (1973).

[7] C. Patrignani *et al.* (Particle Data Group), Chin. Phys. **C40**, 100001 (2016).

[8] A. Riotto and M. Trodden, Annual Review of Nuclear and Particle Science **49**, 35 (1999).

[9] M. Pospelov and A. Ritz, Annals of Physics **318**, 119 (2005), arXiv:0504231 [hep-ph] .

[10] J. Bagger and J. Wess, *Supersymmetry and Supergravity*, 1st ed. (Princeton University Press, Princeton, 1983).

[11] J. Ellis, Nuclear Inst. and Methods in Physics Research, A **284**, 33 (1989).

[12] The ACME Collaboration, Science **343**, 269 (2014).

[13] J. J. Hudson, D. M. Kara, I. J. Smallman, B. E. Sauer, M. R. Tarbutt, and E. A. Hinds, Nature **473**, 493 (2011).

[14] B. Graner, Y. Chen, E. G. Lindahl, and B. R. Heckel, Phys. Rev. Lett. **116**, 161601 (2016).

[15] E. M. Purcell and N. F. Ramsey, Phys. Rev. **78**, 807 (1950).

[16] J. H. Smith, E. M. Purcell, and N. F. Ramsey, Phys. Rev. **108**, 120 (1957).

[17] N. F. Ramsey, Phys. Rev. **76**, 996 (1949).

[18] P. G. Harris, (2007), arXiv:0709.3100 .

[19] P. D. Miller, W. B. Dress, J. K. Baird, and N. F. Ramsey, Phys. Rev. Lett. **19**, 381 (1967).

[20] C. G. Shull and R. Nathans, Phys. Rev. Lett. **19**, 384 (1967).

[21] W. B. Dress, J. K. Baird, P. D. Miller, and N. F. Ramsey, Phys. Rev. **170**, 1200 (1968).

[22] J. K. Baird, P. D. Miller, W. B. Dress, and N. F. Ramsey, Phys. Rev. **179**, 1285 (1969).

[23] W. B. Dress, P. D. Miller, and N. F. Ramsey, Phys. Rev. D **7**, 3147 (1973).

[24] W. B. Dress, P. D. Miller, J. M. Pendlebury, P. Perrin, and N. F. Ramsey, Phys. Rev. D **15**, 9 (1977).

[25] I. Altarev *et al.*, Nuclear Physics A **341**, 269 (1980).

[26] I. Altarev *et al.*, Physics Letters B **102**, 13 (1981).

[27] I. Altarev, Y. V. Borisov, N. Borovikova, A. Brandin, A. Egorov, S. Ivanov, E. Kolomenskii, M. Lasakov, V. Lobashev, A. Pirozhkov, *et al.*, JETP Lett **44**, 360 (1986).

[28] I. Altarev *et al.*, Physics Letters B **276**, 242 (1992).

[29] J. Pendlebury *et al.*, Physics Letters B **136**, 327 (1984).

[30] K. Smith *et al.*, Physics Letters B **234**, 191 (1990).

[31] P. G. Harris *et al.*, Phys. Rev. Lett. **82**, 904 (1999).

[32] C. A. Baker *et al.*, Phys. Rev. Lett. **97**, 131801 (2006).

[33] J. M. Pendlebury *et al.*, Phys. Rev. D **92**, 092003 (2015).

[34] R. Golub, D. Richardson, and S. Lamoreaux, *Ultra-Cold Neutrons*, 1st ed. (CRC Press, 1991).

[35] G. L. Greene, N. F. Ramsey, W. Mampe, J. M. Pendlebury, K. Smith, W. B. Dress, P. D. Miller, and P. Perrin, Metrologia **18**, 93 (1982).

[36] K. Green *et al.*, NIM A **404**, 381 (1998).

[37] N. Nouri *et al.*, Journal of Instrumentation **10**, P12003 (2015).

[38] S. Groeger, G. Bison, and A. Weis, Journal of Research of the National Institute of Standards and Technology **110**, 179 (2005).

[39] C. Baker *et al.*, NIM A **736**, 184 (2014).

[40] S. Afach *et al.*, Journal of Applied Physics **116**, 084510 (2014).

[41] S. Afach *et al.*, Phys. Lett. B **745**, 58 (2015).

[42] I. Altarev *et al.*, Phys. Rev. Lett. **103**, 081602 (2009).

[43] I. Altarev *et al.*, Physica B: Condensed Matter **406**, 2365 (2011), proceedings of the 8th International Workshop on Polarised Neutrons for Condensed Matter Investigation.

[44] I. Altarev *et al.*, Phys. Rev. D **80**, 032003 (2009).

[45] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).

[46] B. Lauss, Physics Procedia **51**, 98 (2014).

[47] S. Afach *et al.*, European Physical Journal A **51**, 1 (2015).

[48] S. V. Komposch, "Realization of a high-performance laser-based mercury magnetometer for neutron EDM experiments," (2017), PhD Thesis, ETH Zürich.

[49] M. C. Fertl, "A laser based mercury co-magnetometer for the neutron electric dipole moment search," (2013), PhD Thesis, ETH Zürich.

[50] S. Afach *et al.*, Physics Letters B **739**, 128 (2014).

[51] E. Wursten, PhD Thesis, KU Leuven, in preparation.

[52] D. J. R. May, "A High Precision Comparison Of The Gyromagnetic Ratios Of The ^{199}Hg Atom And The Neutron," (1998), PhD Thesis, University of Sussex.

[53] S. K. Lamoreaux and R. Golub, Phys. Rev. Lett. **98**, 149101 (2007).

[54] C. A. Baker, D. D. Doyle, P. Geltenbort, K. Green, M. G. D. van der Grinten, P. G. Harris, P. Iaydjiev, S. N. Ivanov, D. J. R. May, J. M. Pendlebury, J. D. Richardson, D. Shiers, and K. F. Smith, Phys. Rev. Lett. **98**, 149102 (2007).

[55] J. M. Pendlebury *et al.*, Phys. Rev. A **70**, 032102 (2004).

[56] S. K. Lamoreaux and R. Golub, Phys. Rev. A **71**, 032104 (2005).

[57] P. G. Harris and J. M. Pendlebury, Phys. Rev. A **73**, 014101 (2006).

[58] S. Afach *et al.*, European Physical Journal D **69** (2015), 10.1140/epjd/e2015-60207-4, arXiv:1503.08651 [physics.atom-ph] .

[59] J. Zenner, "The search for the neutron electric dipole moment," (2013), PhD Thesis, ETH Zürich.

[60] S. Afach *et al.*, Phys. Rev. Lett. **115**, 162502 (2015).

[61] B. Franke, "Investigations of the Internal and External Magnetic Fields of the Neutron Electric Dipole Moment Experiment at the Paul Scherrer Institute," (2013), PhD Thesis, ETH Zürich.

[62] H. J. M. Brake, W. H. J, and H. Rogalla, Measurement Science and Technology **2**, 596 (1991).

[63] J. Bork, H.-D. Hahlbohm, R. Klein, and A. Schnabel, "The 8-layered magnetically shielded room of the PTB: design and construction," .

[64] V. O. Kelhä, J. M. Pukki, R. S. Peltonen, A. J. Penttinen, R. J. Ilmoniemi, and J. J. Heino, IEEE Transactions on Magnetics **18**, 260 (1982).

[65] J. Voigt, S. Knappe-Grüneberg, A. Schnabel, R. Körber, and M. Burghoff, Metrology and Measurement Systems **20** (2013), 10.2478/mms-2013-0021.

[66] D. Spemann, T. Reinert, J. Vogt, J. Wassermann, and T. Butz, Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms **210**, 79 (2003).

[67] T. Bryś *et al.*, NIM A **554**, 527 (2005).

[68] K. Kobayashi, A. Kon, M. Yoshizawa, and Y. Uchikawa, IEEE Transactions on Magnetics **48**, 4554 (2012).

[69] M. Reta-Hernandez and G. G. Karady, Electric Power Systems Research **45**, 57 (1998).

[70] R. Penrose, Mathematical Proceedings of the Cambridge Philosophical Society **51**, 406–413 (1955).

[71] G. Golub and W. Kahan, Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis **2**, 205 (1965).

[72] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, 1st ed. (John Wiley & Sons, New York, 1980) pp. 100–104.

[73] J. L. Kirschvink, Bioelectromagnetics **13**, 401 (1992).

[74] R. Turner, Journal of Physics D: Applied Physics **19**, L147 (1986).

[75] E. C. Wong, a. Jesmanowicz, and J. S. Hyde, Magnetic Resonance in Medicine **21**, 39 (1991).

[76] R. Turner, Magn Reson Imaging **11**, 903 (1993).

[77] S. Crozier and D. Doddrell, Journal of Magnetic Resonance Series A **103**, 354 (1993).

[78] H. Sanchez, F. Liu, A. Trakic, E. Weber, and S. Crozier, IEEE Transactions on Magnetics **43**, 3558 (2007).

[79] H. Sanchez, M. Poole, F. Liu, and S. Crozier, Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine **Honolulu**, 772 (2009).

[80] M. Rawlik, C. Crawford, A. Eggenberger, K. Kirch, J. Krempel, F. M. Piegsa, and G. Quéméner, (2017), arXiv:1709.04681 .

[81] E. Purcell and D. Morin, *Electricity and Magnetism*, 3rd ed. (Cambridge University Press, 2013).

[82] C. Beidler *et al.*, Fusion Technology **17**, 148 (1990).

[83] H. Anton and C. Rorres, *Elementary Linear Algebra*, 9th ed. (John Wiley and Sons, 2005).

[84] D. J. Griffiths, *Introduction to Electrodynamics*, 3rd ed. (Prentice-Hall, 1999).

[85] M. Rawlik, github.com/rawlik/Coils.jl, or 10.5905/ethz-1007-110.

[86] J. D. Hanson and S. P. Hirshman, Physics of Plasmas **9**, 4410 (2002).

[87] M. I. Grivich and D. P. Jackson, American Journal of Physics **68**, 469 (2000).

[88] F. Roméo and D. I. Hoult, Magnetic Resonance in Medicine **1**, 45 (1984).

[89] I. IgH, `http://www.etherlab.org`.

[90] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, SIAM Review **59**, 65 (2017).

[91] D. W. Allan, Proceedings of the IEEE **54**, 221 (1966).

[92] R. M. Gray and D. L. Neuhoff, IEEE Transactions on Information Theory **44**, 2325 (1998).

[93] S. Emmenegger, PhD Thesis, ETH Zürich, in preparation.

[94] C. Abel *et al.*, Phys. Rev. X **7**, 041034 (2017).

[95] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys **594**, A13 (2016).

[96] J. A. Tyson, F. Valdes, and R. A. Wenk, Astrophysical Journal **349**, L1 (1990).

[97] R. D. Peccei and H. R. Quinn, Phys. Rev. D **16**, 1791 (1977).

[98] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978).

[99] F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).

[100] J. E. Kim, Phys. Rev. Lett. **43**, 103 (1979).

[101] M. Shifman, A. Vainshtein, and V. Zakharov, Nuclear Physics B **166**, 493 (1980).

[102] A. R. Zhitnitsky, Sov. J. Nucl. Phys. **31**, 529 (1980).

[103] M. Dine, W. Fischler, and M. Srednicki, Physics Letters B **104**, 199 (1981).

[104] E. Witten, Physics Letters B **149**, 351 (1984).

[105] J. P. Conlon, Journal of High Energy Physics **2006**, 078 (2006).

[106] P. Svrcek and E. Witten, Journal of High Energy Physics **2006**, 051 (2006).

[107] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, Phys. Rev. D **81**, 123530 (2010).

[108] P. Arias, D. Cadamuro, M. Goodsell, J. Jaeckel, J. Redondo, and A. Ringwald, Journal of Cosmology and Astroparticle Physics **2012**, 013 (2012).

[109] D. J. Marsh, Physics Reports **643**, 1 (2016).

[110] J. Preskill, M. B. Wise, and F. Wilczek, Physics Letters B **120**, 127 (1983).

[111] L. Abbott and P. Sikivie, Physics Letters B **120**, 133 (1983).

[112] M. Dine and W. Fischler, Physics Letters B **120**, 137 (1983).

[113] M. Y. Khlopov, B. A. Malomed, and Y. B. Zeldovich, Monthly Notices of the Royal Astronomical Society **215**, 575 (1985).

[114] W. Hu, R. Barkana, and A. Gruzinov, Phys. Rev. Lett. **85**, 1158 (2000).

[115] D. J. E. Marsh and J. Silk, Monthly Notices of the Royal Astronomical Society **437**, 2652 (2014).

[116] H.-Y. Schive, T. Chiueh, and T. Broadhurst, Nat Phys **10**, 496 (2014).

[117] L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten, Phys. Rev. D **95**, 043541 (2017).

[118] B. Bozek, D. J. E. Marsh, J. Silk, and R. F. G. Wyse, Monthly Notices of the Royal Astronomical Society **450**, 209 (2015).

[119] H.-Y. Schive, T. Chiueh, T. Broadhurst, and K.-W. Huang, The Astrophysical Journal **818**, 89 (2016).

[120] P. S. Corasaniti, S. Agarwal, D. J. E. Marsh, and S. Das, Phys. Rev. D **95**, 083512 (2017).

[121] R. Hlozek, D. Grin, D. J. E. Marsh, and P. G. Ferreira, Phys. Rev. D **91**, 103512 (2015).

[122] P. W. Graham, I. G. Irastorza, S. K. Lamoreaux, A. Lindner, and K. A. van Bibber, Annual Review of Nuclear and Particle Science **65**, 485 (2015).

[123] P. W. Graham and S. Rajendran, Phys. Rev. D **84**, 055013 (2011).

[124] Y. V. Stadnik and V. V. Flambaum, Phys. Rev. D **89**, 043522 (2014).

[125] B. M. Roberts, Y. V. Stadnik, V. A. Dzuba, V. V. Flambaum, N. Leefer, and D. Budker, Phys. Rev. Lett. **113**, 081601 (2014).

[126] B. M. Roberts, Y. V. Stadnik, V. A. Dzuba, V. V. Flambaum, N. Leefer, and D. Budker, Phys. Rev. D **90**, 096005 (2014).

[127] V. V. Flambaum, talk presented at the *9th Patras Workshop on Axions, WIMPs and WISPs*, Schloss Waldthausen, Mainz, Germany, 24th June 2013, `http://axion-wimp2013.desy.de/e201031`.

[128] P. W. Graham and S. Rajendran, Phys. Rev. D **88**, 035023 (2013).

[129] K. Blum, R. T. D'Agnolo, M. Lisanti, and B. R. Safdi, Physics Letters B **737**, 30 (2014).

[130] Y. V. Stadnik, *Manifestations of Dark Matter and Variations of the Fundamental Constants of Nature in Atoms and Astrophysical Phenomena*, (Springer, Cham, 2017).

[131] Y. V. Stadnik and V. V. Flambaum, Phys. Rev. Lett. **115**, 201301 (2015).

[132] G. G. Raffelt, Physics Reports **198**, 1 (1990).

[133] G. G. Raffelt, "Astrophysical axion bounds," in *Axions: Theory, Cosmology, and Experimental Searches*, edited by M. Kuster, G. Raffelt, and B. Beltrán (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008) pp. 51–71.

[134] G. Vasilakis, J. M. Brown, T. W. Kornack, and M. V. Romalis, Phys. Rev. Lett. **103**, 261801 (2009).

[135] R. Crewther, P. D. Vecchia, G. Veneziano, and E. Witten, Physics Letters B **88**, 123 (1979).

[136] Physics Letters B **91**, 487(E) (1980).

[137] M. Pospelov and A. Ritz, Phys. Rev. Lett. **83**, 2526 (1999).

[138] O. P. Sushkov, V. V. Flambaum, and I. B. Khriplovich, Sov. Phys. JETP **60**, 873 (1984).

[139] O. P. Sushkov, V. V. Flambaum, and I. B. Khriplovich, Zh. Eksp. Teor. Fiz. **87**, 1521 (1984).

[140] V. Flambaum, I. Khriplovich, and O. Sushkov, Physics Letters B **162**, 213 (1985).

[141] V. Flambaum, I. Khriplovich, and O. Sushkov, Nuclear Physics A **449**, 750 (1986).

[142] V. A. Dzuba, V. V. Flambaum, J. S. M. Ginges, and M. G. Kozlov, Phys. Rev. A **66**, 012111 (2002).

[143] V. F. Dmitriev and R. A. Sen'kov, Phys. Rev. Lett. **91**, 212303 (2003).

[144] V. F. Dmitriev and R. A. Sen'kov, Physics of Atomic Nuclei **66**, 1940 (2003).

[145] V. F. Dmitriev, R. A. Sen'kov, and N. Auerbach, Phys. Rev. C **71**, 035501 (2005).

[146] J. H. d. Jesus and J. Engel, Phys. Rev. C **72**, 045503 (2005).

[147] S. Ban, J. Dobaczewski, J. Engel, and A. Shukla, Phys. Rev. C **82**, 015501 (2010).

[148] L. I. Schiff, Phys. Rev. **132**, 2194 (1963).

[149] R. Catena and P. Ullio, Journal of Cosmology and Astroparticle Physics **2010**, 004 (2010).

[150] V. A. Kostelecký and C. D. Lane, Phys. Rev. D **60**, 116010 (1999).

[151] NASA, "LAMBDA - Coordinate Conversions," `http://lambda.gsfc.nasa.gov/toolbox/tb_coordconv.cfm`, accessed 25-July-2017.

[152] A. Schuster, Terrestrial Magnetism **3**, 13 (1898).

[153] J. D. Scargle, The Astrophysical Journal **263**, 835 (1982).

[154] C. E. Shannon, Proceedings of the IRE **37**, 10 (1949).

[155] J. T. VanderPlas, (2017), arXiv:1703.09824 .

[156] Debosscher, J., Sarro, L. M., Aerts, C., Cuypers, J., Vandenbussche, B., Garrido, R., and Solano, E., Astronomy and Astrophysics **475**, 1159 (2007).

[157] H. G. Tucker, Ann. Math. Statist. **30**, 828 (1959).

[158] S. Algeri, D. van Dyk, J. Conrad, and B. Anderson, Journal of Instrumentation **11**, P12010 (2016).

[159] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Ḳrocesses* (Tata McGraw-Hill, 2002).

[160] N. Ayres, PhD Thesis, University of Sussex, in preparation.

[161] D. Budker, P. W. Graham, M. Ledbetter, S. Rajendran, and A. O. Sushkov, Phys. Rev. X **4**, 021030 (2014).

[162] The Astronomical Almanac Online for the year 2016, `http://asa.usno.navy.mil/SecK/Constants.html`.

[163] S. J. Asztalos *et al.*, Phys. Rev. Lett. **104**, 041301 (2010).

[164] N. F. Ramsey, *Molecular Beams* (Oxford University Press, Amen House, London E.C.4, 1956).

[165] J. D. Hunter, Computing In Science & Engineering **9**, 90 (2007).

[166] E. Jones, T. Oliphant, P. Peterson, *et al.*, "SciPy: Open source scientific tools for Python," (2001–).

[167] T. Kluyver *et al.*, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by F. Loizides and B. Schmidt (IOS Press, 2016) pp. 87 – 90.

[168] F. Perez and B. E. Granger, Computing in Science Engineering **9**, 21 (2007).

[169] W. McKinney, in *Proceedings of the 9th Python in Science Conference*, edited by S. van der Walt and J. Millman (2010) pp. 51 – 56.