Report NPI ASCR Řež, EXP-01/2005
FZKA Scientific Report 7090
MS-KP-0501

KATRIN Design Report 2004

J. Angrik[i], T. Armbrust[h], A. Beglarian[e], U. Besserer[g], J. Blümer[d,h], J. Bonn[i], R. Carr[i], B. Bornschein[g], L. Bornschein[h], T. Burritt[m], M. Charlton[n], A.J. Davies[n], C. Day[f], P.J. Doe[m], L. Dörr[g], O. Dragoun[l], G. Drexlin[d,h], F. Eichelhardt[h], K. Eitel[d], K. Essig[a], A. Felden[d], B. Flatt[i], J. Formaggio[m], B. Freudiger[d], R. Gehring[f], H. Gemmeke[e], E. Geraskin[j], F. Glück[d], M. Glugla[g], S. Grohmann[d], R. Gumbsheimer[d], F. Habermehl[h], G. Harper[m], G. Hellriegel[g], J. Herbert[p], T. Höhn[d], M. Howe[m], H. Hucker[d], O.V. Ivanov[j], K.-P. Jüngst[f], J. Kašpar[l], O. Kazachenko[g,j], N. Kernert[d], P. Komarek[f], A. Kovalík[l], C. Kraus[i*], H. Krause[d], M. Leber[m], R. Lewis[n], V. Lobashev[j], X. Luo[d], K. Maier[a], O. Malyshev[p], M. Mark[d], I. Meshkov[o], K. Müller[d], S. Mutterer[d], G.R. Myneni[r], M. Neuberger[c], H. Neumann[f], M. Noe[f], H.W. Ortjohann[k], S. Osipov[j], A. Osipowicz[b], B. Ostrick[a,k], E.W. Otten[i], P. Plischke[d], A. Povtschinik[k], M. Prall[a,k], R. Reid[p], K. Rielage[m], R.G.H. Robertson[m], M. Ryšavý[l], P. Schäfer[g], K. Schlösser[d], U. Schmitt[d], F. Schwamm[d], P. Schwinzer[i], D. Sevilla Sanchez[i], F. Sharipov[q], A. Skasyrskaya[j], A. Špalek[l], M. Steidl[d], M. Süßer[d], E. Syresin[o], H.H. Telle[n], T. Thümmler[a,k], N. Titov[j,k], V. Usanov[j], K. Valerius[a,k], T. VanWechel[m], D. Vénos[l], R. Vianden[a], C. Weinheimer[k], A. Weiss[c], J. Wendel[g], J.F. Wilkerson[m], J. Wolf[h], S. Wüstling[e], S. Zadoroghny[j], M. Zbořil[l]

The KATRIN Collaboration:

[a] *Rheinische Friedrich-Wilhelms-Universität Bonn, Helmholtz - Institut für Strahlen- und Kernphysik, Nussallee 14-16, 53115 Bonn, Germany*

[b] *Fachhochschule Fulda, Fachbereich Elektrotechnik und Informationstechnik, Marquardtstr. 35, 36039 Fulda, Germany*

[c] *Forschungszentrum Karlsruhe, Commercial Project Execution/Contracting, Postfach 3640, 76021 Karlsruhe, Germany*

[d] *Forschungszentrum Karlsruhe, Institut für Kernphysik, Postfach 3640, 76021 Karlsruhe, Germany*

[e] *Forschungszentrum Karlsruhe, Institut für Prozessdatenverarbeitung und Elektronik, Postfach 3640, 76021 Karlsruhe, Germany*

[f] *Forschungszentrum Karlsruhe, Institut für Technische Physik, Postfach 3640, 76021 Karlsruhe, Germany*

[g] *Forschungszentrum Karlsruhe, Tritium-Labor Karlsruhe, Postfach 3640, 76021 Karlsruhe, Germany*

[h] *Universität Karlsruhe (TH), Institut für Experimentelle Kernphysik, Gaedestr. 1, 76128 Karlsruhe, Germany*

[i] *Johannes Gutenberg-Universität Mainz, Institut für Physik, 55099 Mainz, Germany*

[j] *Academy of Sciences of Russia, Institute for Nuclear Research, 60th October Anniversary Prospect 7a, 117312 Moscow, Russia*

[k] *Westfälische Wilhelms-Universität Münster, Institut für Kernphysik, Wilhelm-Klemm-Str. 9, 48149 Münster, Germany*

[l] *Academy of Sciences of the Czech Republic, Nuclear Physics Institute, CZ-250 68 Řež near Prague, Czech Republic*

[m] *Center for Experimental Nuclear Physics and Astrophysics, and Department of Physics, University of Washington, Seattle, WA 98195, USA*

[n] *Department of Physics, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, United Kingdom*

[*] *present address: Department of Physics, Queen's University Kingston, Ontario K7L 3N6 Canada*

Associated members:

[o] *Joint Institute for Nuclear Research Dubna, Joliot-Curie 6, 141980 Dubna, Russia*

Expert groups:

[p] *ASTeC, CCLRC Daresbury Laboratory, Daresbury Warrington, Cheshire, WA4 4AD, United Kingdom*

[q] *Departamento de Fı́sica, Universidade Federal do Paraná, Caixa Postal 19044, Curitiba, 81531-990, Brazil*

[r] *Jefferson Lab, 12000 Jefferson Avenue, Newport News, VA 23606, USA*

ABSTRACT

After the KATRIN Letter of Intent in 2001 and the Addendum to the LoI in 2002, this Design Report 2004 gives a detailed overview of the KATRIN experiment in its phase of building and testing of the first components. We critically assess the actual status in neutrino physics with respect to direct and indirect searches on the neutrino mass in particle physics and cosmology and the impact by KATRIN. The physical parameters as well as the schemes for the technical realization of the central experimental components and their status are reported. Having optimized the experimental configuration compared to the LoI, we anticipate a KATRIN sensitivity on the neutrino mass of $m(\nu_e) = 0.2\,\mathrm{eV}$ (90% C.L.) corresponding to a $5\,\sigma$ discovery potential for $m(\nu_e) = 0.35\,\mathrm{eV}$, based on a detailed assessment of systematic and statistical uncertainties.

Contents

1 Introduction and Motivation

In this *Design Report* we give an overview of the present reference design of the Karlsruhe Tritium Neutrino (KATRIN) experiment. An initial outline of KATRIN has been reported in 2001 in a Letter of Intent (LoI) [1] and, in May 2002, in an Addendum to the LoI [2]. The 2004 reference design of KATRIN presented below is based on a two year design optimization phase from mid–2002 to mid–2004. This work has enhanced the neutrino mass sensitivity of the experiment. For the envisaged KATRIN measuring time of 3 full years, the statistical and systematic uncertainties contribute equally. If no neutrino mass signal is observed, the upper limit will be $m_\nu < 0.2\,\mathrm{eV/c^2}$ (90 % C.L.), as compared to the initial 2001 estimate of $m_\nu < 0.35\,\mathrm{eV/c^2}$ [1]. In case of a positive signal, the new reference design has a discovery potential of $5\,\sigma$ ($3\,\sigma$) for a neutrino mass $m_\nu = 0.35\,\mathrm{eV/c^2}$ ($m_\nu = 0.3\,\mathrm{eV/c^2}$). This significant improvement of KATRIN's sensitivity to $0.2\,\mathrm{eV/c^2}$ is a full order improvement with respect to previous direct neutrino mass experiments. It will allow KATRIN to probe all quasi-degenerate neutrino mass scenarios and the full cosmologically relevant neutrino mass range.

Design Modifications

In the following we give a brief executive summary of the major design works of the past 2 years, which have improved the neutrino mass sensitivity of KATRIN to the present reference value:

- the luminosity of the windowless gaseous tritium source has been improved by a factor of 2 by enlarging the source diameter
- the potential problem of source charging has been investigated in detail and hardware remedies have been worked out as well as methods to determine the exact electric potential in the source
- the diameter of the main spectrometer has been enlarged from 7 m to 10 m to retain the standard energy resolution of $\Delta\mathrm{E} = 0.93\,\mathrm{eV}$ for analysis of the enlarged source
- the technical feasibility of the main spectrometer UHV vessel has been demonstrated by two industrial studies
- a novel method to suppress spectrometer related background by an inner repelling grid has been worked out and been successfully tested at the Mainz spectrometer
- the vacuum concept of KATRIN has been successfully tested with a UHV test recipient down to the 10^{-12} mbar regime
- a new method for monitoring the retarding HV of the main spectrometer by setting up a monitor beam line has been worked out

- an optimized measuring point distribution with enhanced scanning time in the area of a few eV below the endpoint has been implemented
- a new procedure to reduce the main systematic uncertainty of the measurements associated with the inelastic scattering of β-electrons in the source has been developed
- a detailed planning of the infrastructure, and in particular of the experimental halls, has been compiled

In addition, the commissioning of hardware components on-site has started. The first major component of KATRIN, the pre-spectrometer was delivered to the FZK site in autumn 2003. Since then the pre-spectrometer UHV vessel, the superconducting magnets, the pre-spectrometer detector, the HV system as well as the DAQ and SCS system have been commissioned. The WGTS and the main spectrometer vessel have been fully specified, the tender actions completed and the orders given to companies. First UHV test measurements have started which successfully ended autumn 2004. In 2005, the pre-spectrometer will be equipped with an inner electrode system, of which the construction is nearly finished. The electromagnetic properties of this system will be tested in detail in 2005.

1.1 Overview

KATRIN is a next-generation tritium β-decay experiment which will improve the ν–mass sensitivity compared to the present direct neutrino mass experiments at Mainz and Troitsk by one order of magnitude. With an estimated ν-mass sensitivity of 0.2 eV (90 % CL.) KATRIN will allow the investigation of the sub-eV neutrino mass scale, which is of particular interest for particle physics, astrophysics and cosmology. In contrast to other methods such as the search for neutrinoless double beta decay ($0\nu\beta\beta$) or cosmological ν-mass studies using large scale structure (LSS) and cosmic microwave background radiation (CMBR) data, KATRIN will provide a completely *model-independent* measurement of the ν-mass. The KATRIN result will be based only on kinematic relations and energy-momentum conservation. For these reasons, it will complement the other laboratory and cosmological methods to investigate neutrino masses. The combination and comparison of results from tritium β-decay, $0\nu\beta\beta$ and cosmological studies will be essential for our understanding of the rôle of neutrinos in our physical world, both at the Micro - and Macro - scale.

This design report is organized as follows: in the introduction (sect. 1) we discuss the implications of ν-masses in astroparticle physics together with recent experimental results for ν-mixing and ν-masses ($0\nu\beta\beta$ and cosmology). In sect. 2 we describe the present tritium β decay experiments at Mainz and Troitsk. Then we outline the requirements for a next-generation tritium β-decay experiment with sub-eV ν-mass sensitivity followed by an overview of the reference design of the KATRIN experiment in sect. 3. The following hardware components are discussed in more detail: the tritium sources WGTS and QCTS

with the tritium pumping and electron transport section (sect. 4.2), the electrostatic spectrometers (sect. 5), the electron detector (sect. 6) as well as the magnet and vacuum system (sect. 7). The next three sections give an overview of background sources and their suppression (sect. 8), of methods for energy calibration and monitoring (sect. 9) and the control and data acquisition system (sect. 10). Section 11 describes various sources of systematic uncertainties of the measurement. The KATRIN sensitivity and discovery potential are discussed in sect. 12 together with their implications. In sect. 13 we discuss the KATRIN experimental implementation and schedule.

1.2 Evidence for massive neutrinos

The Standard Model (SM) of particle physics has long assumed neutrinos to be massless particles, however, the results of ν-oscillation experiments using solar (SNO, Super-Kamiokande, Kamiokande, GNO, Gallex, Sage, Homestake) as well as atmospheric neutrinos (Super-Kamiokande, Soudan2, MACRO) have provided compelling evidence for flavor transformations of neutrinos (ν-oscillations) and hence for non-zero neutrino masses [3, 4, 5, 25]. These important results have been further confirmed by the observation of neutrino disappearance in experiments using reactor (KamLAND) and accelerator (K2K) neutrinos at long baseline [6, 7]. The parameters of neutrino oscillations, i.e. the mass splitting Δm^2_{ij} and mixing amplitudes $\sin^2\theta_{ij}$ of neutrinos are now known to leading order[1]. Especially the perfect agreement between the parameters obtained by solar ν_e experiments with the ones yielded from the KamLAND reactor $\bar{\nu}_e$ experiment rules out all other possible explanations except neutrino oscillations to describe the observed disappearance or flavor change.

Neutrino oscillations imply that a neutrino from one specific weak interaction flavor, *e.g.* a muon neutrino ν_μ, transforms into another weak flavor eigenstate, *i.e.* an electron neutrino ν_e or a tau neutrino ν_τ, while travelling from the source to the detector. The existence of neutrino oscillations requires a non-trivial mixing between the weak interaction eigenstates (ν_e, ν_μ, ν_τ) and the corresponding neutrino mass states (ν_1, ν_2, ν_3) and, moreover, that the mass eigenvalues (m_1, m_2, m_3) differ from each other. Consequently, the experimental evidence for neutrino oscillation proves that neutrinos have non-zero masses. Unfortunately neutrino oscillation experiments are not sensitive to neutrino masses directly.

1.2.1 Atmospheric neutrinos

The observation of the oscillation of atmospheric neutrinos in 1998 by the Super-Kamiokande experiment [3] was the first direct proof for non-zero neutrino masses. Atmospheric neutrinos are produced by the interactions of cosmic rays in the upper atmosphere, which generate secondary pions and kaons of both charges. The subsequent weak meson decays give rise to ν_μ and ν_e with a flavour ratio of about 2:1 with energies E_ν in the few

[1] The $\bar{\nu}_e$ - appearance signal at short baseline claimed by LSND has not been confirmed by KARMEN and will not be discussed further[8, 9].

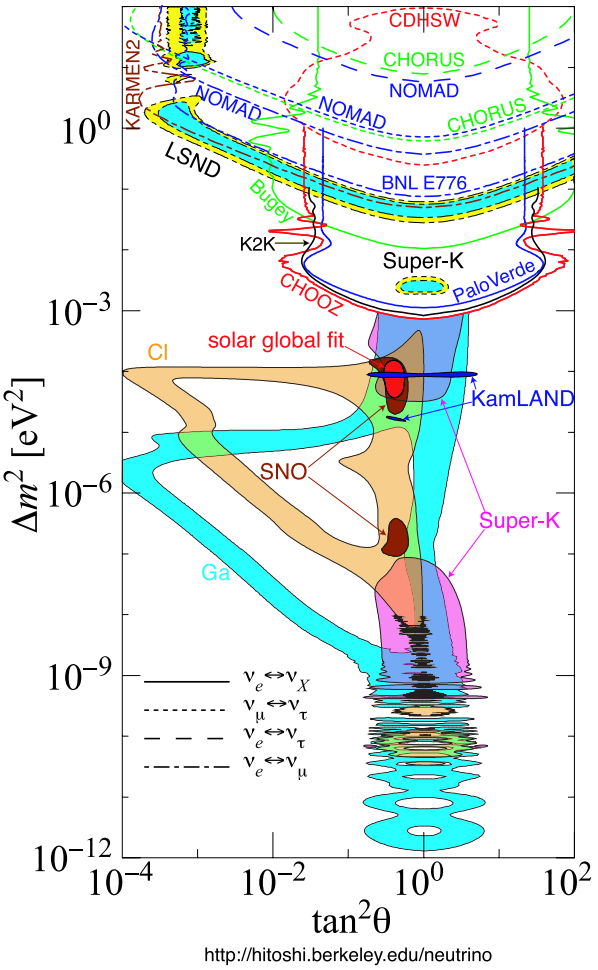

Figure 1: Results from oscillation experiments using solar and atmospheric neutrinos as well as ν's from accelerators and reactors at long and short baseline [10]. The solar-ν as well as the reactor-ν results only allow the LMA solution of the MSW effect, whereas atmospheric ν's as well as long baseline accelerator ν's point to maximum mixing at $\Delta m^2_\nu = 2 \times 10^{-3}\,\mathrm{eV}^2$.

GeV range. In the 50 kt imaging water Cherenkov detector Super-Kamiokande the flavor (ν_μ, $\bar{\nu}_\mu$ with respect to ν_e, $\bar{\nu}_e$) as well as the energy E_ν and their zenith angle θ_ν can be reconstructed. In [3] the observation of a zenith angle dependent deficit of ν_μ was reported, which is a clear indication of ν_μ - flavor transformation. The analysis of the angular spectrum θ_ν allows the determination of the mass splitting parameter $\Delta\mathrm{m}^2_{atm}$ with great precision, as θ_ν selects different neutrino path lengths L_ν. The reduction factor for upgoing ν_μ is ~ 2, implying maximum or near-maximum mixing of ν_μ. The non-observation of the oscillations of ν_e (as well as the results of the intermediate baseline reactor experiments CHOOZ and Palo Verde) points to $\nu_\mu \rightarrow \nu_\tau$ mixing as the dominant oscillation mode (there is also a preference for ν_τ in the Super-Kamiokande data).

At the end of 2003 the Super-Kamiokande Collaboration reported on a re-analysis of their entire data set [11]. Selecting high quality atmospheric ν_μ events with good $\mathrm{L}_\nu/\mathrm{E}_\nu$

resolution, the experimental precision was enhanced so that for the first time a 'dip' in the L_ν/E_ν distribution of atmospheric events was observed. This identification of the first oscillation minimum in the L_ν/E_ν parameter space is the first detection of the oscillatory behavior of neutrino oscillations. The latest atmospheric neutrino oscillation results, based on 1489 days of data taking, have been reported in [12]. A full 3-flavor neutrino oscillation analysis yield the following 'best fit' atmospheric oscillation parameters (3 σ -range) [13]:

$$0.5 \quad < \quad \tan^2\theta_{atm} \quad < \; 2.1 \tag{1}$$

$$1.6 \cdot 10^{-3}\ \mathrm{eV}^2 \quad < \quad \Delta m^2_{atm} \quad < \; 3.6 \cdot 10^{-3}\ \mathrm{eV}^2 \tag{2}$$

The results of atmospheric neutrinos have been further supported by results of the first long baseline accelerator oscillation experiment K2K. This experiment detects ν_μ with mean energy of 1.5 GeV energy from the 12 GeV KEK-PS at a distance of 235 km to look for ν_μ - disappearance with the Super-Kamiokande detector. Initial results from June 1999 - July 2001 based on 4.8×10^{19} protons on target showed a ν_μ - deficit in agreement with oscillation expectations based on atmospheric ν's corresponding to a no oscillation probability of 0.5 %.

1.2.2 Solar neutrinos

The solar-ν problem deficit, originating from the pioneering solar ν's observations of the Cl-37 experiment by R. Davis Jr. in the late 1960's [14] has been the first hint for non-SM properties of neutrinos. In a series of classic solar-ν experiments based on radiochemical (Homestake, GALLEX/GNO, SAGE) as well as on real-time (Kamiokande, Super-Kamiokande) detection methods, the deficit of solar ν_e was confirmed with high precision. A global analysis of this data ensemble left the following ν-oscillation solutions (see fig. 1): enhanced MSW (Mikheyev-Smirnov-Wolfenstein) oscillations at small (SMA) and large (LMA) mixing, vacuum oscillations (VAC) at long baseline, or flavour transitions at an intermediate scale (LOW).

The long-standing problem of the missing solar ν_e was finally solved by the SNO experiment. Following initial results [5] for charged current interactions $\mathrm{d}\,(\nu_e\,,\mathrm{e}^-\,\mathrm{p})\,\mathrm{p}$ of solar ν_e, the first *neutral current* measurement of ^{8}B solar-ν's based on the observation of the deuteron breakup reaction $\mathrm{d}\,(\nu\,,\nu'\,\mathrm{n})\,\mathrm{p}$ was reported in [15]. Assuming a standard ^{8}B-flux, the comparison of the NC and CC as well as the elastic scattering rate allowed SNO to deduce a non-ν_e component $\phi_{\mu\tau} = 3.41^{+0.45}_{-0.45}(\mathrm{stat.})^{+0.48}_{-0.45}\ (\mathrm{syst.}) \times 10^6\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ in the solar neutrino flux, corresponding to a $5.3\,\sigma$ signal for ν_e flavour transformations. These results from the initial SNO phase-I with pure D_2O were confirmed and refined by the SNO phase-II (salt running), making use of a different and more efficient technique for neutron detection [17].

A global solar neutrino fit [18], which incorporates additional SNO data (day-night asymmetries of the NC, CC and elastic scattering events) gives strong preference for the LMA solution of MSW transitions.

1.2.3 Reactor neutrinos

The LMA region of solar MSW transitions can be probed by a terrestrial long baseline reactor oscillation experiment looking for $\bar{\nu}_e \rightarrow \bar{\nu}_x$ disappearance. The KamLAND experiment is a 1 kt liquid scintillator observing $\bar{\nu}_e$ from Japanese nuclear power plants. With a mean distance of 160 km to the power plants, the experiment is located at the appropriate L_ν/E_ν parameter space. In 2002, initial data were released, which show a $\bar{\nu}_e$ - deficit of R=0.611 $\pm$ 0.085 $\pm$ 0.041. New data with 5 times better statistics were presented at the conference Neutrino 2004 yielding not only a similar rate deficit but also a distinguished spectral shape of $\bar{\nu}_e$. Using this spectral shape the L/E-distribution clearly shows a dip at the right position confirming and restricting the LMA parameter region of the MSW effect. Especially the perfect agreement between the parameters obtained by solar ν_e experiments with the ones yielded from the KamLAND reactor $\bar{\nu}_e$ experiment rules out all other possible explanations except neutrino oscillations to describe the observed disappearance or flavor change.

The existence of neutrino oscillations and therefore of neutrino mixing and masses has far-reaching implications for numerous fields of particle physics, astrophysics and cosmology, which are discussed in the following in more detail.

1.3 Massive neutrinos in Particle Physics

The SM of particle physics describes present experimental data up to the electroweak scale. Within the SM the charged fermions acquire mass by Yukawa-couplings to the Higgs boson. Unfortunately these couplings are arbitrary in the theory, the observed pattern of the masses or the mixing of the charged fermions has no natural explanation within the SM. The neutrinos appear purely left-handed and massless within the SM. In principle the SM could be enhanced by introducing right-handed neutrinos and Yukawa-couplings for the neutrino as well, but then the pattern of observed masses and mixings is even more obscure as the masses of the neutrinos are at least 6 orders of magnitude smaller then the masses of the charged fermions. Therefore, the experimental evidence for neutrino masses and mixing is seen as the first clear indication for physics beyond the Standard Model.

There are many theories beyond the Standard Model, which explore the origins of neutrino masses and mixing. In these theories, which often work within the framework of Supersymmetry, neutrinos naturally acquire mass. A large group of models makes use of the so-called see-saw effect to generate masses for Majorana neutrinos [19]. There are two types: The Seesaw I mechanism creates the small neutrino masses by introducing heavy right-handed Majorana neutrinos. The heavier the right-handed neutrino is, the lighter the left-handed neutrino becomes ("Seesaw"). These models prefer an hierarchical pattern of the three light neutrino masses. The Seesaw II mechanism requires a new Higgs triplet which couples directly to the light neutrinos. Many new models predict such a Higgs triplet. These Seesaw II models predict more degenerate or quasi-degenerate neutrino mass scenarios. Therefore, the distinction of hierarchical and quasi-degenerate

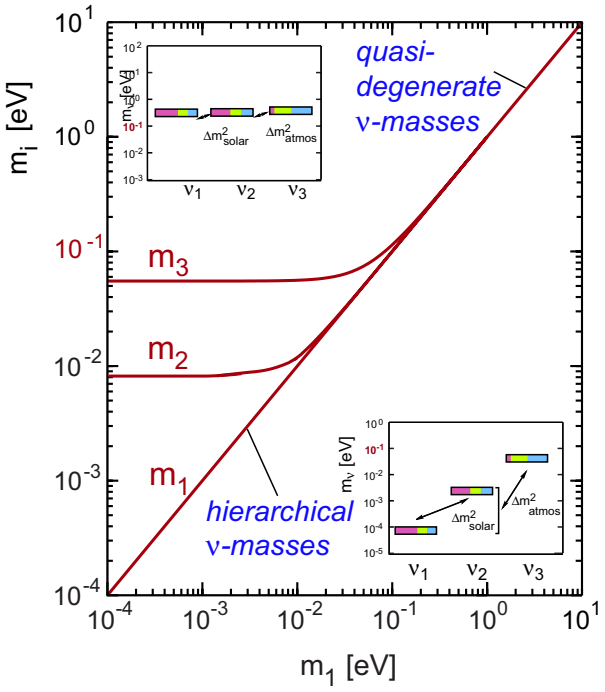

Figure 2: Values of the neutrino mass eigenstates m_1, m_2 and m_3 as a function of the lightest mass eigenstate m_1 for a normal mass ordering $m_1 < m_2 < m_3$, forming either a quasi-degenerate or a hierarchical ensemble of eigenstates.

neutrino mass scenarios would be very important to find the right theory beyond the SM.

Other classes of theories are based on completely different possible origins of neutrino masses, such as radiative corrections arising from an extended Higgs sector [20]. In models with extra dimensions very light Dirac neutrinos can also be created [21].

As neutrino masses are much smaller than the masses of the other fermions, the knowledge of the absolute values of neutrino masses is crucial for our understanding of fermion masses in general. Recently it has been pointed out that the *absolute mass scale of neutrinos* may be even more significant and straightforward for the fundamental theory of fermion masses than the determination of the neutrino mixing angles and CP-violating phases [22]. It will most probably be the absolute mass scale of neutrinos which will determine the scale of new physics.

As described above, all the theories that extend beyond the SM can be grouped into two different classes, leading either to a hierarchical pattern for the neutrino mass eigenvalues m_i (following the pattern of the quark and charged lepton masses)

$$m_1 \ll m_2 \ll m_3 \,, \tag{3}$$

or resulting in a nearly degenerate pattern of neutrino masses

$$m_1 \approx m_2 \approx m_3 \,. \tag{4}$$

As neutrino oscillation experiments are only sensitive to the differences of the squared masses Δm^2 (*e.g.* $\Delta m^2_{12} = |m^2_2 - m^2_1|$ or $\Delta m^2_{12} = m^2_2 - m^2_1$ if matter effect are involved), they cannot measure absolute values of ν masses. While they do not distinguish between the two classes of models, oscillation experiments allow to set a *lower bound* on the absolute ν-mass, as at least one of the neutrino mass eigenvalues should satisfy the inequality :

$$m_i \text{ or } m_j \geq \sqrt{|\Delta m^2_{ij}|} \tag{5}$$

Analysis of the actual results of Super-Kamiokande [3] in terms of oscillations of atmospheric neutrinos thus gives a lower bound on m_3:

$$m_2 \text{ or } m_3 \geq \sqrt{\Delta m^2_{atm}} \sim (0.04 - 0.07) \text{ eV} \tag{6}$$

However, the fundamental mass scale of neutrinos can be located well above this lower bound (*e.g.* at around a few tenths of an eV), as suggested by mass models with partial degeneracy [23]. Discrimination between hierarchical and quasi-degenerate mass scenarios thus requires a sensitivity on the absolute ν-mass scale in the sub-eV range.

Finally, theoretical models come to different conclusions of whether neutrino masses are of the Dirac- or of the Majorana type. A massive neutrino which is identical to its own antiparticle is called a Majorana particle, while for Dirac-type neutrinos the lepton number distinguishes neutrinos from antineutrinos. This requires the development of experimental techniques for ν-masses in the sub-eV range, which do not depend on assumptions about the Dirac or Majorana character of the neutrino mass.

1.4 Massive neutrinos in Cosmology

Neutrinos and their properties also play an important rôle in astrophysics and cosmology. As there are about 1 billion times more neutrinos than baryons predicted to be left over from the Big Bang these relic neutrinos could play an important rôle as *neutrino hot dark matter* (νHDM) in the evolution of large scale structures (LSS) [24]. The imprint of νHDM on LSS evolution is quite distinct from other dark matter candidates such as supersymmetric particles, which act as non-relativistic or Cold Dark Matter (CDM). Cosmological models of structure formation strongly depend on the relative amounts of CDM and νHDM in the universe, hence a determination of the neutrino contribution Ω_ν to the total dark matter content Ω_{DM} of the universe is important for our understanding of structure formation [24].

1.4.1 Neutrino Hot Dark Matter

In the early universe massive relic neutrinos act as hot dark matter in LSS evolution, because they are *relativistic* when decoupling from matter which took place at a temperature of $k_B T_{decouple} \approx 1 - 2\,\text{MeV}$ shortly before $e^- e^+$ annihilation into the CMBR photons. Of course at the present epoch relic neutrinos with masses $m_\nu \gg 0.2$ meV are non-relativistic. With $k_B T_{decouple} = 1 - 2\,\text{MeV}$, the ratio of number densities is thus given by $n_\nu = 9/11\ n_\gamma$,

making neutrinos the second most abundant particles in the universe (112 ν's/cm^3 per flavor, 336/cm^3 in total).

Relativistic neutrinos freely stream out of the density perturbations formed by cold dark matter and baryons. Thereby the neutrino perturbations are completely erased and the growth of density fluctuations is slowed. On the free streaming scale or Jeans length, which is roughly 1 Gpc for $m_\nu = 1$ eV, the matter power spectrum P of cold dark matter and baryons is strongly suppressed by

$$\Delta P \,/\, P \sim -\,8\;\Omega_\nu \;/\; \Omega_m \tag{7}$$

Heavier neutrinos with mass m_ν contribute more to the total cosmological matter density according to

$$\Omega_\nu \; h^2 = \Sigma m_\nu \;/\; 93 \; eV \tag{8}$$

and hence suppress power on smaller scales more strongly.

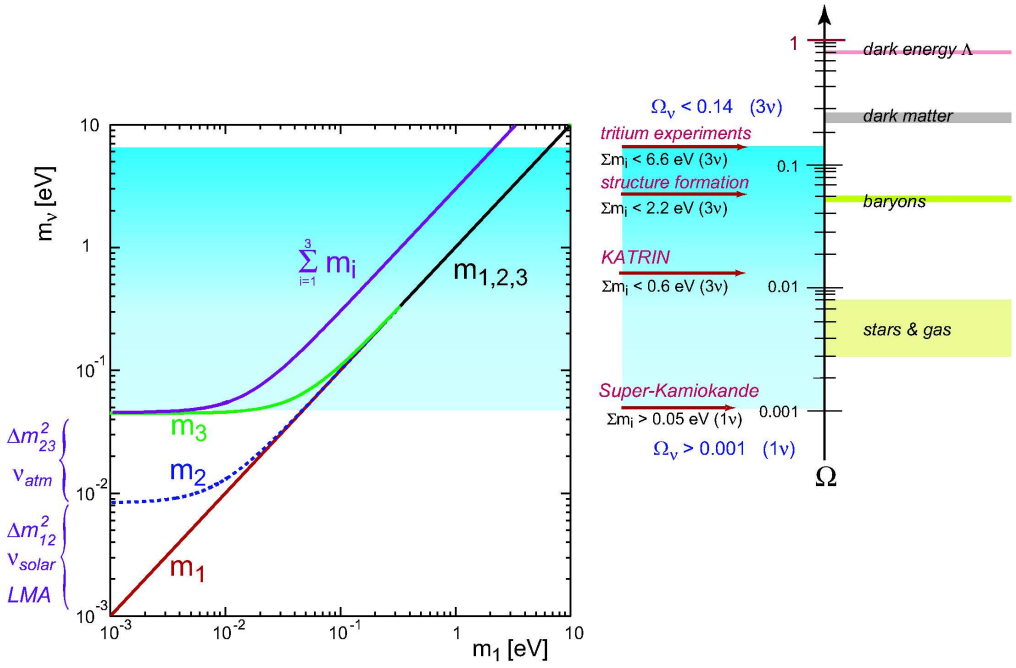

Figure 3: Neutrino masses (left, compare to figure 2) and their contribution Ω_ν (right) to the energy density of the universe in comparison to the contribution from dark energy Ω_Λ, dark matter Ω_{DM} and baryons Ω_b. The luminous matter (stars and gas clouds) is part of the Ω_b, neutrinos are part of Ω_{DM}. The experimentally allowed contribution Ω_ν from neutrino HDM to the total matter energy density Ω of the universe spans two orders of magnitude. The lower bound on Ω_ν results from the analysis of oscillations of atmospheric ν's. The upper bound stems from current tritium β decay experiments and studies of structure formation. One of the main motivations of KATRIN will be to investigate the parameter space $\Omega_\nu > 0.01$, where relic neutrinos from the Big Bang would play a significant rôle as νHDM in the evolution of large scale structures.

The experimentally allowed contributions of the neutrino density Ω_ν to the total matter-energy density Ω of the universe are shown in fig. 3 together with the contributions arising from luminous matter, baryons (Ω_b), dark matter (Ω_m), and dark energy Ω_Λ). These contributions represent the canonical standard cosmological model ('concordance model'). Recent measurements of the CMBR by the WMAP satellite and LSS data from the SDSS and 2dFGRS galaxy redshifts surveys indicate a flat ΛCDM model dominated by dark energy $\Omega_\Lambda = 0.73 \pm 0.04$ and dark matter $\Omega_m = 0.23 \pm 0.04$. The latter is composed of non-baryonic cold dark matter $\Omega_{cdm} \simeq 0.2$, baryons $\Omega_b = 0.044 \pm 0.004$ and neutrino hot dark matter $\Omega_\nu > 0.001$.

The contribution Ω_ν of νHDM can vary in the interval $0.001 < \Omega_\nu < 0.15$. The lower bound on Ω_ν arises from the atmospheric Δm^2-scale determined by Super-Kamiokande [3], assuming hierarchical neutrino masses so that only one neutrino mass eigenstate contributes to Ω_ν. The upper bound is derived from current tritium β decay experiments [25] using eq. (5) and assuming three quasi-degenerate mass eigenstates. Similar upper bounds on Ω_ν have been published from recent studies of the evolution of large scale structures in the universe (see below). The experimentally constrained parameter range of Ω_ν thus spans about two orders of magnitude. A determination of Ω_ν or a significant constraint on the allowed parameter range of Ω_ν would lead to a much better understanding of the role of νHDM in the evolution of large scale structures.

The investigation of the still open role of νHDM in the evolution of large scale structures is one of the main motivations for the next-generation direct neutrino mass experiment KATRIN. With a neutrino mass sensitivity of 0.2 eV (90 %CL.) KATRIN will be sensitive to a neutrino contribution $\Omega_\nu = 0.01$. The KATRIN measurement would either significantly constrain or fix the role of νHDM in structure formation by a model independent method.

1.5 Cosmological studies

The contribution Ω_ν of relic neutrinos to the total density Ω can be probed by a combination of precise measurements of the temperature fluctuations of the CMBR (power spectrum at very large scales) with measurements of the matter fluctuation at large scales by high statistics galaxy redshift surveys. The results for the mass density of neutrinos can then be transformed into neutrino mass results based on the calculated neutrino density of 112 ν's/cm^3 in the framework of the canonical standard cosmological model.

1.5.1 CMBR experiments

Measurements of the fluctuations of the cosmic microwave background radiation (CMBR) provide an early picture of structure formation shortly after the Big Bang [26]. These investigations have made dramatic progress in recent years, in particular due the results from the WMAP satellite[27]. The imprint of νHDM on the analysis of the angular power spectrum of the CMBR temperature fluctuations is however rather small. For example,

the signature of an 0.1 eV neutrino in the CMBR angular power spectra is only 0.1 % [2]. Even the high angular resolution data of WMAP allow the entire dark matter Ω_m to be composed of neutrinos Ω_ν (see fig. 3). The main significance of CMBR data in cosmological studies of the ν-mass lies in their ability to pin down other cosmological parameters such as Ω_m and Ω_b. These constraints can then be used in analyses of complementary[3] LSS data from galaxy surveys, which are much more sensitive to Ω_ν (a neutrino with 0.1 eV mass will reduce the matter power spectrum by $\approx 10\,\%$).

1.5.2 LSS experiments

The matter distribution at the present epoch can be traced by observations of the Lyman-α forest [28], galaxy cluster abundances [29] and, more recently, by high statistics galaxy redshift surveys. Major advances in the field of large scale structures have been the first data release from SDSS in May 2003 with $2 \cdot 10^5$ galaxy redshifts [30], and earlier results by 2dFGRS (2dF Galaxy Redshift Survey) with more than $2.2 \cdot 10^5$ measured galaxies [31]. The combination of these results with WMAP and other high resolution data from the CMBR, as well as other information (Lyman-α forest, X-ray luminosities of clusters) have allowed detailed investigations of νHDM contributions Ω_ν to the dark matter in the universe.

1.5.3 Results from cosmological studies

Early cosmological bounds on Σm_ν, which have been based on pre-WMAP/SDSS/ 2dFGRS data, give upper limits in the range of 3 – 6 eV, comparable to the present laboratory limit from tritium β-decay. In [26], for example, a limit $\Sigma m_\nu < 4.2$ eV was quoted from a combined analysis of CMBR and LSS data.

Recent cosmological studies are based on different combinations of the high quality data from WMAP, small scale high resolution CMBR experiments, and the two large galaxy redshift surveys 2dFGRS and SDSS. In some cases additional structure information from Lyman-α data or X-ray luminosities from clusters have been added.

Table 1 gives an overview of recent cosmological studies, which have provided either upper limits for neutrino masses or tentative evidence for a νHDM component in the universe.

The table shows the considerable spread in the results published recently. This is due to several generic problems associated with cosmological studies. First, these studies suffer from the problem of parameter degeneracy [35]. Different combinations of cosmological parameter can describe the LSS and CMBR data equally well, so additional information is required to break the degeneracy. Serious degeneracies occur for example between Ω_ν and Ω_m and the Hubble expansion parameter h. Further degeneracies have been reported between Ω_ν and the scalar spectral index n of the matter power spectrum and the galaxy

[2] heavy neutrinos with masses in the eV range are non-relativistic at the time of the last scattering of the CMBR and are thus indistinguishable from CDM for CMBR experiments

[3] it is crucial that there is an overlap of scales between both data sets

author	WMAP	CMB_{hi-l}	SDSS	2dF	other data	$\Sigma\, m_\nu$ [eV]
Bar'03 [32]	×	×	×	×	h (HST)	< 0.75
Teg'03 [30]	×	×	×		SNIa	< 1.7
ASB'03 [33]	×	×		×	XLF	$= 0.36$-1.03
WMAP [27]	×	×		×	Lyα , h (HST)	< 0.7
Bla'03 [34]	×			×	Ω_m=1	$= 2.4$
Han'03 [35]	×	×		×	h (HST), SNIa	< 1.01
Han'03 [35]	×	×		×		< 1.2
Han'03 [35]	×			×		< 2.12

Table 1: Survey of neutrino mass results obtained by recent cosmological studies. For each study the specific set of input data is listed: WMAP- CMBR angular / polarisation data from WMAP [27], CMB_{hi-l}-multipole l CMBR data [36, 37], SDSS- LSS data from the SDSS first data release [30, 38], 2dF-LSS data from the 2dFGRS galaxy survey [39].

cluster factor σ_8. Thus, one has to rely on further assumptions (flat universe with Ω_{tot}=1) as well as on input from other experiments for Ω_Λ, Ω_m and Ω_b and the Hubble expansion parameter h. The uncertainties associated with these input parameters imply that the ν-mass results crucially depend on the priors for these parameters. Different priors for cosmological parameters result in limits for the sum of neutrino masses Σm_ν which differ by factors of 2 or more. For these reasons, some of the results listed in table 1, most notably the stringent WMAP upper limit, have been criticized for using too strong priors.

The strong dependence of cosmological neutrino mass results can be illustrated by comparing the strongest upper limit on m_ν reported in the literature, the WMAP upper limit [27] of $\sum_\nu = \sum_{i=1}^{3} m_i < 0.7\,\mathrm{eV}$ (95 % CL.), with a tentative evidence for non-zero neutrino masses reported in [33]. Both analyses use almost identical sets of input parameters, most notably the WMAP and 2dFGRS data. However, while the authors of [27] use additional Lyman-α data for their analysis, the authors of [33] use X-ray luminosity functions (XLF) of galaxy clusters obtained with the orbiting Chandra X-ray telescope. This small change of input data transforms an upper limit into a tentative evidence for non-zero masses with Σm_ν=0.56 eV as best fit value. It is interesting to note that both the use of Lyman-α data as well as the use of the measured XLF have been criticized in the literature. This clearly underlines that cosmological studies, as impressive as they are with regard to the determination of cosmological parameters, still yield *model-dependent* results for neutrino masses.

Further concerns with cosmological neutrino mass studies are associated with systematic errors. LSS data from galaxy surveys suffer from the problem of biasing (i.e. to what extent does the galaxy distributing track that of the distribution of cold dark matter and dark baryons), possible redshift-space distortions and selection effects. A model-independent input of the value of the neutrino mass to cosmological studies would thus be especially important for future high precision cosmological studies. A laboratory measurement of m_ν could break the existing degeneracies between Ω_ν and other cosmological parameters, and thus help to provide a better picture of large scale structure evolution.

1.5.4 Beyond the 'concordance' model

The results of the cosmological studies presented in table 1 have been obtained within the 'canonical standard cosmological model'. In the following we briefly list studies which go beyond the present 'concordance' flat ΛCDM models.

Broken power law

In [34] the authors argue that cosmological data can be fitted equally well in the framework of an Einstein-de Sitter universe with *zero* cosmological constant Λ, albeit at the expense of requiring a very low value for the Hubble constant of $H_0 \simeq 46$ km/s/Mpc. In addition, the authors relax the usual requirement of a single power law for the primordial density fluctuations during inflation. Analysis of CMBR and LSS data with this broken-power law plus the introduction of three degenerate neutrinos of mass 0.8 eV gives acceptable fits to the standard set of data (WMAP+LSS+others). The authors thus conclude that CMB and LSS data seem to imply the existence of a non-cold dark matter component.

Mass varying neutrinos

Recently, several authors [40] have speculated on the interesting coincidence between the smallness of neutrino masses ($m_\nu < 1\,\mathrm{eV}$) and the deduced vacuum density $\rho_V \approx (\,10^{-3}$ eV$\,)^4$ (in units of $\hbar c = 1$) responsible for the cosmological dark energy Λ. Starting with the introduction of sterile neutrinos, which acquire masses of 10^{-3} eV through interactions with a scalar field, the models transmit the sterile neutrino mass to the active sector via the seesaw mechanism. In this framework, the masses of the active ν-species vary like the inverse of the ν-density. Accordingly, the neutrino mass could be of order of 1 eV today. As the so-called cosmo-MSW effect for relic neutrinos [40] could lead to resonant conversion of sterile to active neutrinos (and vice-versa), the results of laboratory neutrino mass measurements would differ from cosmological studies.

Stochastic neutrino masses

Another 'coincidence' problem ($\Omega_\Lambda = \Omega_m$) has been used [41] to investigate the possibility that m_ν is a stochastic variable, which is randomized during inflation. Combining this scenario with *a posteriori* arguments based on observer-related selection effects (large neutrino masses effectively prohibit the formation of galaxy sized objects in the early universe), the authors calculate a probability distribution for the neutrino mass peaking at $\Sigma m_\nu \sim 1\,\mathrm{eV}$.

Neutrinoless universe

Another possible modification to the neutrino sector of the 'concordance' model has been brought up by the authors of [42]. They assume that neutrinos have small extra scalar or pseudoscalar interactions. Should neutrinos possess even only tiny couplings of the order of 10^{-5} to hypothetical scalar ϕ bosons, the neutrino density Ω_ν could be affected strongly by annihilation processes of the type $\nu\,\nu \leftrightarrow \phi\,\phi$. In this scenario neutrinos would not decouple from matter at $T = 1\,\mathrm{MeV}$, but would stay in thermal equilibrium until much

later times (T = 1eV), which would inhibit free streaming.

For couplings of the order of 10^{-5} all relic neutrinos would have annihilated in the present epoch, and we would live in a 'neutrinoless universe'. The important consequence of this model is the possibility, that neutrinos -after annihilation- would have no rôle in the subsequent cosmological evolution. Calculations of the large scale power spectrum in this model show that cosmological neutrino mass limits could be evaded completely and that even very large neutrino masses of 1 eV would be allowed by cosmology.

Relic neutrinos have not been detected at present, and will likely not be detected in the near future due to their very low (μeV) energy. For this reason, the 'neutrinoless universe' scenario can only be tested by comparing the model-independent neutrino mass result from KATRIN (supplemented by results from double beta decay) with cosmological data. Thus, KATRIN has the chance to indirectly check the existence of relic neutrinos.

1.5.5 Future perspectives

At present, the cosmological studies of m_ν are still model-dependent due to systematic effects such as biasing, parameter degeneracy, possible selection effects, possible contributions from non-linear effects and the strong influence of priors on the neutrino mass results. Future high precision studies aim to strongly reduce these effects.

In the field of CMBR experiments, the new promising technique of studying distortions of the CMBR temperature and polarization maps induced by gravitational lensing has been proposed [43]. This technique is sensitive to changes of structure evolution at late times due to massive neutrinos and thus could break the degeneracy between neutrino mass, equation-of-state of the dark energy and the primordial power spectrum amplitude. In [43] the sensitivity of this method is estimated to be as low as Σm_ν=0.3 eV. This is of the same order as that expected in [44] for the ν-mass sensitivity (m_ν=0.14 eV) from the Planck satellite scheduled to start operation in 2007.

These investigations will be complemented by future deep galaxy surveys extending out to redshift parameters z=2 (DEEP2, VLT-Virmos) as well as dedicated studies of lensing effects on galaxy clusters (LSST, Large Synoptic Survey Telescope) [45].

In the case that these future data reveal non-zero neutrino masses a laboratory experiment of similar sensitivity, like KATRIN, is of utmost importance. A non-zero neutrino mass or an upper limit from KATRIN is in any case extremely helpful to interpret the fluctuations in the universe at small scales and to constrain the many correlated cosmological parameters going into those interpretations.

1.5.6 Leptogenesis

Leptogenesis is one of the most attractive theories to explain the observed baryon asymmetry of the universe $\eta = n_B/n_\gamma \approx 6.5 \times 10^{-10}$. There are two distinct classes of leptogenesis models: in type I models the asymmetry is caused by the out-of-equilibrium decay of the right-handed heavy neutrinos which are responsible for generating the light neutrino masses via the (type I) see-saw mechanism. To achieve successful leptogenesis

this type of models demand very small neutrino masses which do not allow a partially degenerate pattern. This leads to an upper bound for the absolute mass scale of neutrinos of $m_\nu < 0.12\,\mathrm{eV}$ [46].

In type II leptogenesis models, such as of [47], the out-of-equilibrium decay of the lightest right-handed (s)neutrino in a type II class see-saw model is discussed in the framework of the Minimal Supersymmetric Standard Model. In this framework the stringent upper limits of type I leptogenesis models can be evaded. Even more, partially degenerate light neutrino mass patterns with scales of about 0.35 eV are rather helpful. Also, and in sharp contrast to type I models, the total decay asymmetry of heavy neutrinos *increases* with the neutrino mass scale. This in turn allows a lower reheating temperature, making thermal type II leptogenesis models also more consistent with gravitino constraints from SUSY.

1.6 Massive neutrinos in Astrophysics

There are two recent developments in neutrino astrophysics, which are of relevance for direct ν-mass measurements. First, there are new approaches for ν-mass measurements with supernova neutrino Time-of-Flight (ToF) studies (see sect. 1.8.1). Second, a new model has been proposed which aims to explain the origin of ultra high energy (UHE) cosmic rays. This so-called Z-burst model would require relic neutrino masses within the sensitivity range of KATRIN for neutrino masses.

1.6.1 Ultra High Energy Cosmic Rays

A possible new interrelation between neutrino and cosmic ray physics has received a lot of interest recently [48, 49, 50, 51] with a Z-burst scenario, where the observed flux of UHE cosmic rays $\geq 10^{20}$ eV is explained by the resonant annihilation of UHE neutrinos with *massive* relic neutrinos into Z-bosons.

The origin and nature of the observed UHE cosmic ray events above the so-called Greisen-Zatsepin-Kuzmin (GZK) cutoff is one of the most important questions in cosmic ray physics today. Conventional scenarios for UHE cosmic rays assume that they are protons from plausible UHE sources (such as quasars) at cosmological distances. However, the GZK cutoff (interaction of ultra-high energetic protons with cosmic microwave background photons forming the $\Delta(1232)$ resonance) limits the maximum energy for UHE cosmic ray protons from distances ≥ 50 Mpc to 4×10^{19} eV. The Z-burst scenario would evade the GZK cutoff, as UHE neutrinos with energies far above the GZK cutoff could travel over cosmological distances of several hundred Mpc without being attenuated (see fig. 4 a).

The Z-burst model requires the annihilation between the UHE-ν's and relic neutrinos to be on the Z-mass shell M_Z, so that the resonance energy $\mathrm{E}_{\nu_i}^{\mathrm{res}}$ is given by

$$E_{\nu_i}^{\mathrm{res}} = \frac{M_Z^2}{2\,m_{\nu_i}} = 4.2 \cdot 10^{21}\ \mathrm{eV} \left(\frac{1\ \mathrm{eV}}{m_{\nu_i}} \right) , \tag{9}$$

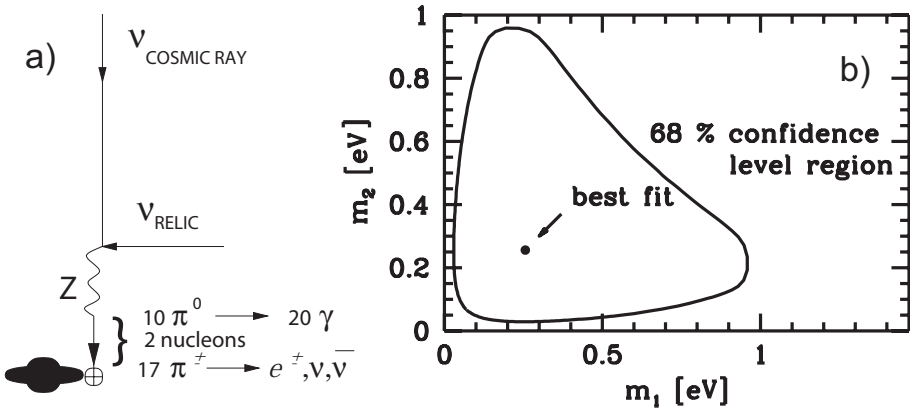

Figure 4: The Z-burst scenario for UHE cosmic rays: a) the resonant annihilation of an UHE-ν with a massive relic ν gives rise to UHE cosmic ray protons from the hadronic decay modes of the Z-boson at local distances of a few Mpc, b) neutrino mass parameters from a fit to a global data set for UHE cosmic ray energies [50].

For relic neutrino masses m_{ν_i} of $\mathcal{O}\,(1\,\mathrm{eV})$, $E^{\mathrm{res}}_{\nu_i}$ would be close to the highest energy cosmic rays. Several detailed calculations in the framework of this model have been performed [50, 51], taking into account known parameters like Z-production and -decay as well as propagation effects, while at the same time relying on assumptions for the unknown UHE-ν fluxes and the relic neutrino number density. Quantitatively, a small relic ν-mass m_{ν_i} requires a large $E^{\mathrm{res}}_{\nu_i}$ to produce a Z-boson. This relationship implies that a measurement of the shape of the UHE cosmic ray energy spectrum above a few $\times\, 10^{19}$ eV would permit statements on m_{ν_i}.

Fits to the available UHE cosmic ray data suffer from the rather limited statistics of the current data from AGASA, Fly's Eye and other experiments. Moreover, systematic errors in the fit procedures arise, as one has to take into account other possible contributions to the UHE cosmic ray spectrum. Both factors limit the precision of the ν-mass estimates in the Z-burst model. Most fits using reasonable input parameters yield however relic ν-masses of the order of $\mathcal{O}\,(1$ eV) [49, 50] (see fig. 4).

There are several ways to test the Z-burst model. In the nearer future the Pierre Auger Observatory [52] will be able to detect, or set limits upon, the flux of UHE-neutrinos, thus constraining one of the input parameters of the Z-burst model. Auger will also measure the shape of the UHE cosmic ray energy spectrum with much better statistics, which should facilitate an improved quantitative analysis of ν-masses in the framework of the Z-burst model. These improved ν-mass estimates could then be compared with future laboratory results from the direct measurement of the electron neutrino mass $m(\nu_e)$. KATRIN will thus be able to test the second input parameter of the Z-burst model, i.e. the rest mass m_{ν_i} of relic neutrinos. The independent results by Auger and KATRIN should test the Z-burst model.

1.7 Neutrinoless double beta decay

The search for neutrinoless double β-decay ($0\nu\beta\beta$) is a very sensitive means of searching for (Majorana-type) neutrino masses. The double β decay with $2\,\nu$-emission ($\beta\beta 2\nu$) is a SM weak process of second order. The $0\nu\beta\beta$ process on the other hand is physics beyond the SM, as the $\bar{\nu}_e$ emitted at the first vertex has to be absorbed as ν_e at the second vertex. This requires neutrinos to be massive and to be identical to their own antiparticles (Majorana particles). Although the exchange of a massive Majorana neutrino is not the only possible mechanism for $0\nu\beta\beta$ (the exchange could be mediated by SUSY particles and/or right-handed currents), the detection of a $0\nu\beta\beta$ - process would unambiguously prove that neutrinos, in contrast to all other SM particles, are Majorana particles. It would also imply that lepton number is not conserved, as the emission of two electrons violates lepton number by two units (ΔL=2). More details on double beta decay can be found in recent reviews (see [53, 54]).

The experimental search for $0\nu\beta\beta$ has a long history. At present, there are several experimental searches for $0\nu\beta\beta$ signals based on $\beta\beta$-unstable nuclei such as ^{76}Ge, ^{100}Mo, ^{116}Cd, ^{130}Te, ^{136}Xe and others. The most sensitive experimental detection techniques today are based on calorimeters (^{76}Ge-experiments Heidelberg-Moscow [55], IGEX [56]) and on bolometers (^{130}Te-experiments MiBeta, Cuoricino [57]). These experiments have yielded upper limits on the effective Majorana mass m_{ee} in the $\mathcal{O}$ (1 eV) range [53]. The most stringent upper limit $m_{ee} < 0.35$ eV (90 % CL.) has been published by the Heidelberg-Moscow experiment, based on data equivalent to 47.7 kg y. The Heidelberg-Moscow experiment operated five enriched high-purity ^{76}Ge detectors with a total mass of 11.5 kg at the Gran Sasso Underground Laboratory from 1990–2003. A detection of a $0\nu\beta\beta$ signal $^{76}\text{Ge} \rightarrow {}^{76}\text{Se} + 2\ e^-$ requires the observation of a statistically significant peak in the Ge-sum energy spectrum at $Q_{\beta\beta} = 2039$ keV.

Reported evidence for $0\nu\beta\beta$ and future $0\nu\beta\beta$ experiments

In 2001, a sub-group of the Heidelberg-Moscow experiment claimed evidence for a $0\nu\beta\beta$ - signal [58] based on a statistics of 54.98 kg y. The conclusions (evidence for a $0\nu\beta\beta$signal) as well as the analysis methods of [59, 58] have been subject to criticisms [62, 61].

After the experiment was terminated in mid-2003, a final analysis of the full data based on 71.7 kg y was presented in [59]. The results are based on a re-analysis of the entire data set (re-calibration, cut of corrupted data). Several statistical techniques based on Bayesian and Frequentist methods were applied to investigate the presence of a $0\nu\beta\beta$ peak at $Q_{\beta\beta}$. Restricting the analysis to a narrow energy interval around $Q_{\beta\beta}$[4], the peak finding algorithms applied by the authors of [59] indicate the presence of a signal with a statistical significance of $\simeq 4\sigma$. The signal extracted would correspond to a $0\nu\beta\beta$ half-life of ^{76}Ge of $T^{0\nu}_{1/2} = (0.35 - 2.03) \times 10^{25}$ y (95% c.l.) with a best value of 1.19×10^{25} y. For a specific set of nuclear matrix elements (h=1), this result translates into an effective Majorana

[4] this procedure relies on additional and crucial assumptions on the nature of the background close to $Q_{\beta\beta}$

neutrino mass $m_{ee}\,/\,h = (0.29 - 0.60)\,eV$ ($3\,\sigma$ range), with a best value of $m_{ee}\,/\,h = 0.44\,eV$. Here we have followed ref. [61] with its notation and introduced a normalization parameter h, which reflects the uncertainties of the nuclear matrix elements (h can vary from $0.6 \div 2.8$ for different calculations).

The open issue of a $0\nu\beta\beta$ signal in [58, 59] has to be clarified by further experimental results, with its implications on the generic Majorana nature of neutrinos and the non-conservation of the lepton number. For the neutrino sector this would imply quasi-degenerate mass pattern, taking into account the small neutrino mass squared differences $\Delta m^2 \leq 10^{-2}$ eV^2 from ν-oscillation experiments.

The current $0\nu\beta\beta$ experiments NEMO3 [60] and Cuoricino [134] have a certain chance to confirm the claim [58, 59], but – due to the uncertainties of the nuclear matrix elements – no real chance to disprove it. New double beta decay experiments with a much improved sensitivity have been proposed. Before the publication of ref. [59] several ambitious experiments with large masses (≈ 500 kg) of enriched (^{76}Ge,^{136}Xe) or non-enriched (^{130}Te) material were proposed to investigate the parameter region corresponding to the inverted hierarchy mass pattern at about 30-50 meV. In response to the claimed evidence for $0\nu\beta\beta$ several groups [63, 64] now aim for experiments of intermediate size (several tens of kg) to verify or refute the claims of [59]. The use of different $\beta\beta$-nuclei and comparison of experimental results should allow to manifest a genuine $0\nu\beta\beta$ signal in the quasi-degenerate mass region over the next decade. The uncertainties of nuclear matrix elements for $0\nu\beta\beta$ - decays will however complicate the precise determination of the absolute mass scale.

The KATRIN experiment will provide a model-independent sensitivity to neutrino masses down to 0.2 eV (90 % CL.), which is well below the central value reported for the effective mass m_{ee} in [59]. A neutrino mass signal corresponding to the central value m_{ee}=0.4 eV could be identified by KATRIN after a very short running time of several months with high statistical confidence (see sect. 12.1 for details). In case that a positive signal is seen in future tritium β-decay and double beta decay searches, the possibility to assess the unknown Majorana CP-phases could open up. This is discussed in more detail below.

1.8 Direct investigations of neutrino masses

In contrast to double β decay experiments, direct investigations of the neutrino mass do not rely on further assumptions on the neutrino mass type (Majorana or Dirac). Direct or kinematic experiments can be classified into two categories both making use of the relativistic energy momentum relation $E^2 = p^2c^2 + m^2c^4$ as well as of energy and momentum conservation.

1.8.1 Neutrino time-of-flight studies

The narrow time signal of a supernova (SN) neutrino burst of less than 10 s in combination with the very long-baseline between source and detector of several kpc allows the

investigation of small ToF effects resulting from non-zero ν-masses [65]. Supernova ToF studies are based on the observation of the energy-dependent time delay of massive neutrinos relative to massless neutrinos. This method provides an experimental sensitivity for the rest masses of ν_e , ν_μ and ν_τ of a few tens of eV. This sensitivity can be pushed into the few-eV range, if additional assumptions concerning the time evolution of the ν-burst are made. However, in this case the ν-mass sensitivity becomes model-dependent.

Recently, new methods have been proposed which do not rely on details of the ν-burst timing and which allow a sensitivity in the few-eV range for ν_μ and ν_τ to be achieved. The two most promising techniques are: a) the measurement of the abrupt termination of the SN-ν signal due to the early formation of a black hole [66, 67] and b) the correlation of SN ν-signals with independent signals from gravitational wave experiments [68].

The first method relies on a scenario where the gravitational core collapse of a very massive star does not stop at the hot protoneutron star stage but proceeds further, resulting in the formation of a black hole. For short time scales of black hole formation of the order of a few seconds, the SN ν-pulse is sharply terminated, when the event horizon crosses the ν-emission sphere. This cutoff can be used as a rather sharp 'time reference' to search for ν-mass effects. Detailed model calculations of this scenario have been performed in [66] for ν-masses in the few-eV range. Fig. 5 shows that the mass sensitivity of this method improves if the distance source-detector decreases. If compared to the sensitivity of present and future tritium experiments, only a nearby supernova would allow to deduce a competitive ν-mass limit. However, core collapse supernovae at distances $< 1\,$kpc are exceedingly rare.

Instead, should future measurements reveal a SN ν-pulse, the information on the ν-mass provided by KATRIN could be used as a reliable input in the interpretation of the time structure of the ν-pulse.

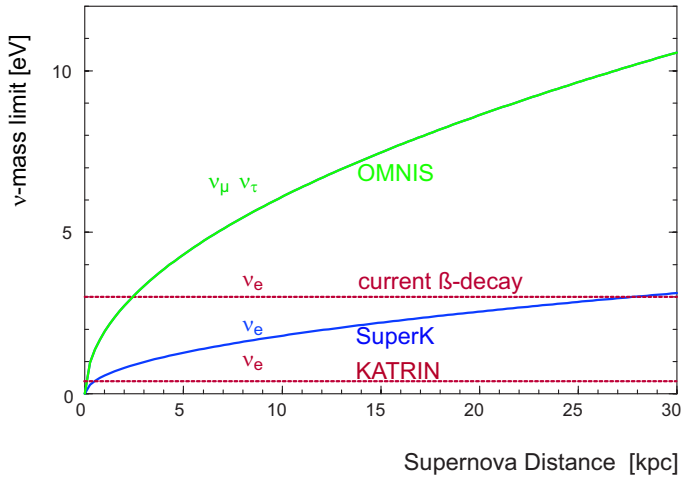

Figure 5: Comparison of the ν-mass sensitivities of astrophysical and laboratory experiments. For the model of a core collapse with early black hole formation the mass sensitivity of Super-Kamiokande for ν_e and of the proposed OMNIS experiment for ν_μ and ν_τ are shown as a function of SN distance (from [66, 67]) and are compared to the ν_e mass sensitivity of present and future ^{3}H β-decay experiments.

A second method with the potential to yield improved ν-mass sensitivities with SN neutrinos is the time correlation of the ν-pulse with data from gravitational wave experiments. This technique has been investigated in detail in [66]. The matching of the signals from the neutronization electron neutrino flash, which could be observed by existing underground detectors like Super-Kamiokande or SNO, with the gravitational waves recorded by the gravity wave antennae soon to be operational (GEO600, LIGO, Virgo), would allow a sensitivity to an electron neutrino mass of 0.75 eV (95% C.L.), if no significant mass effect is found. However, the results will depend to some extent on the ν_e survival probability and the details of ν-flavor oscillations or transformations occurring in the environment of the hot protoneutron star. In the case where all ν_e from the neutronization burst would be converted to ν_μ and ν_τ the sensitivity limit would be 1.1 eV (95% C.L.). Given these estimates, the sensitivity limit of KATRIN for the electron neutrino mass would still be more stringent than these interesting and fascinating possibilities. Again, taking the reverse argument, the neutrino mass measurement from KATRIN could contribute to a better understanding of a future SN signal, in particular with regard to the subtle effects associated with the core neutronization and the emission of gravitational waves, thus adding to a better understanding of these processes.

1.8.2 Kinematics of weak decays

The investigation of the kinematics of weak decays is based on the measurement of the charged decay products. For the masses of ν_μ and ν_τ the measurement of pion decays into muons and ν_μ at PSI and the investigation of τ-decays into 5 pions and ν_τ at LEP have yielded the upper limits:

$$\begin{array}{lcl} m(\nu_\mu) & < & 190 \text{ keV} \quad \text{at } 90\ \% \text{ confidence } [69] \\ m(\nu_\tau) & < & 18.2 \text{ MeV} \quad \text{at } 95\ \% \text{ confidence } [70] \end{array}$$

Both limits are much larger than the interesting range for cosmology and ν HDM (see fig. 3). However, experiments investigating the mass of the electron neutrino ν_e by analyzing β decays with emission of electrons are providing a sensitivity in the interesting eV-range (see section 2).

Other approaches to β decay

A different approach to directly measure the electron neutrino mass is the use of cryogenic bolometers. In this case, the β source can be identical to the β-electron spectrometer. This new technique has been applied to the isotope ^{187}Re, which has a 7 times lower endpoint energy than tritium [71]. The experiments are still in an early stage of development. Current rhenium microcalorimeters reach an energy resolution of $\Delta E = 20 - 30$ eV and yield an upper limit of $m(\nu_e) < 15$ eV [72]. In addition to a significant improvement of the energy resolution to further improve the statistical accuracy, the principle of integration of active source and detector requires the operation of large arrays of microcalorimeters. The expected sensitivity on the neutrino mass in the near future will be below ~ 10 eV [73].

1.9 Tritium β decay

The most sensitive direct searches for the electron neutrino mass up to now are based on the investigation of the electron spectrum of tritium β decay

$$^3\mathrm{H} \rightarrow {}^3\mathrm{He}^+ + \mathrm{e}^- + \bar{\nu}_e \quad .$$

This energy spectrum for a neutrino with mass m_ν is given by

$$\frac{dN}{dE} = C \times F(Z,E)p(E+m_ec^2)(E_0-E)[(E_0-E)^2 - m_\nu^2]^{\frac{1}{2}}\Theta(E_0-E-m_\nu), \qquad (10)$$

where E denotes the kinetic energy of the electron, m_e the mass of the electron, p is the electron momentum, E_0 corresponds to the maximal electron energy for m_ν=0 (endpoint energy), $F(Z,E)$ is the Fermi function, taking into account the Coulomb interaction of the outgoing electron in the final state, the step function $\Theta(E_0-E-m_\nu)$ ensures energy conservation, and C is given by

$$C = \frac{G_F^2}{2\pi^3}\cos^2\theta_C|M|^2 \; . \qquad (11)$$

Here G_F is the Fermi constant, θ_C is the Cabibbo angle and M is the nuclear matrix element. As both M and $F(Z,E)$ are independent of m_ν, the dependence of the spectral shape on m_ν is given by the phase space factor only. In addition, the bound on the neutrino mass from tritium β decay is independent of whether the electron neutrino is a Majorana or a Dirac particle.

The signature of an electron neutrino with a mass of m(ν_e)=1 eV is shown in fig. 6 in comparison with the undistorted β spectrum of a massless ν_e. The spectral distortion is statistically significant only in a region close to the β endpoint. This is due to the rapidly rising count rate below the endpoint dN/dE $\propto (E_0-E)^2$. Therefore, only a very narrow region close to the endpoint E_0 is analyzed. As the fraction of β decays in this region is proportional to a factor $(1/E_0)^3$, the very low tritium endpoint energy of 18.6 keV maximizes the fraction of β decays in this region (in fact, tritium has the second lowest endpoint of all β unstable isotopes). Nevertheless, the requirements of tritium β decay experiments with regard to source strength are demanding. As an example, the fraction of β decays falling into the last 1 eV below the endpoint E_0 is only 2×10^{-13} (see fig. 6), hence tritium β decay experiments with high neutrino mass sensitivity require a huge luminosity combined with very high energy resolution.

Tritium has the following clear advantages as β emitter in ν mass investigations:

1. Tritium has the second lowest endpoint energy of $E_0 = 18.6$ keV.
2. Tritium has a rather short half life $t_{1/2} = 12.3$ a.
3. The hydrogen isotope tritium and its daughter, the $^3\mathrm{He}^+$ ion, have a simple electron shell configuration. Atomic corrections for the β decaying atom -or molecule- and corrections due to the interaction of the outgoing β-electron with the tritium source can be calculated in a simple and straightforward manner.

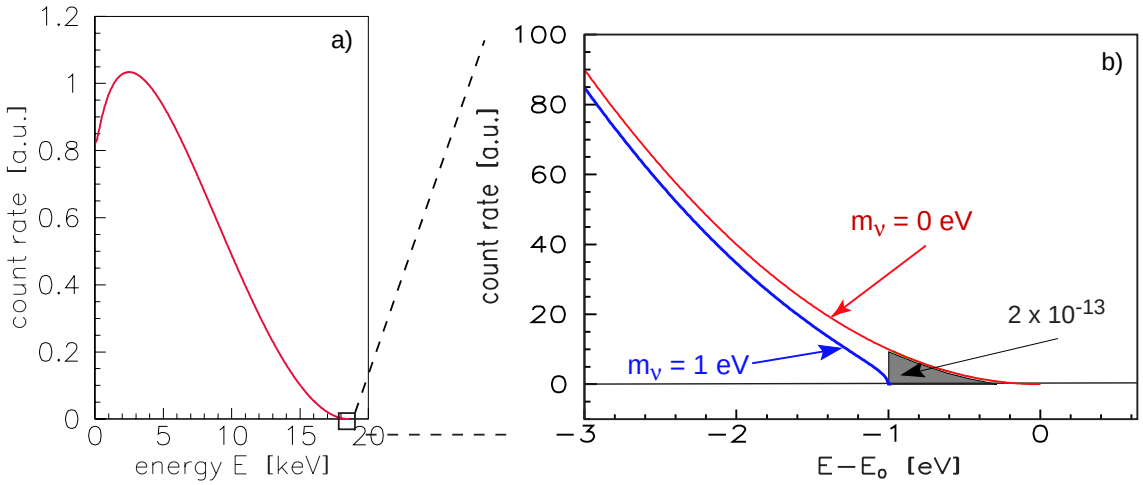

Figure 6: The electron energy spectrum of tritium β decay: (a) complete and (b) narrow region around endpoint E_0. The β spectrum is shown for neutrino masses of 0 and 1 eV.

4. For the same reason (low nuclear charge Z) the inelastic scattering of out-going β-electrons within the β source are small.

5. The tritium β decay is a super-allowed nuclear transition. Therefore, no corrections from the nuclear transition matrix elements M have to be taken into account.

The combination of all these features makes tritium an almost ideal β emitter for neutrino mass investigations.

1.10 Cosmological motivation of KATRIN

In contrast to the cosmological methods described above, the KATRIN experiment will yield a *direct* measurement of the electron neutrino mass m(ν_e). As outlined earlier [1], this value can then be used as a precise input parameter for the interpretation of the CMBR and LSS data. Most importantly, a new input for Ω_ν would further break the degeneracies of the different matter-energy components of the observed CMBR and LSS data, thereby adding to the overall precision in the field of cosmology.

The next generation of LSS or CMBR data will certainly improve the sensitivity of the cosmological results for the neutrino mass. Thus, if a non-zero neutrino mass is indicated, the KATRIN experiment is very important to check this result. If no neutrino mass is found in the analysis of the cosmological data according to the standard cosmological model, then the KATRIN experiment serves as an important test of this concordance model. Such a test is urgently required by cosmologists [74].

When discussing the cosmological relevance of the future KATRIN measurements, it is worth mentioning two recent results. As the corresponding mass splittings are $\Delta m^2 \leq 10^{-2}$ eV^2, the ν_e-mass measurement from KATRIN will allow the determination of the fundamental mass scale of neutrinos. The SNO results have thus added to the cosmological relevance of future tritium β-decay measurements.

A second cosmological implication of the LMA mixing solution is with regard to the number density of relic neutrinos. The exact number density of relic neutrinos from the standard hot Big Bang model depends on the unknown chemical potentials for the three ν-species. Usually, it is assumed that there is a negligible asymmetry between neutrinos and anti-neutrinos of the order of the baryon asymmetry of a few $\times 10^{-10}$. However, there are several cosmological models which predict rather large neutrino asymmetries. According to these models, the relic neutrino number density would not be well defined.

In a recent paper [75] this open issue has been addressed. Taking into account the effects of ν-flavor oscillations shortly after the Big Bang before the nucleosynthesis epoch, the authors conclude that effective flavor equilibrium between all active neutrino species is established, if the ν-oscillation parameters are in the range indicated by the atmospheric ν-results and the LMA solution of solar ν's. This strongly constrains possible neutrino chemical potentials, so that the number density of relic neutrinos is now known to about 1 % [75]. Consequently, existing ν-mass limits from ^{3}H β-decay as well as future measurements of m(ν_e) with KATRIN will provide unambiguous information on the neutrino mass density, essentially free of the uncertainty of neutrino chemical potentials.

1.11 Particle Physics motivation of KATRIN

The effects of neutrino mixing on tritium β decay and $0\nu\beta\beta$ experiments can be significant. Below we discuss the implications of ν-mixing in the determination of the absolute mass scale of neutrinos for both experimental approaches, following largely the discussions in [22, 76].

1.11.1 Neutrino mixing and mass scale

Considering neutrino mixing for $0\nu\beta\beta$ decay, the *effective* Majorana mass m_{ee} is a coherent sum of all neutrino mass states ν_i contributing to the electron neutrino ν_e. The parameter m_{ee} is a combination of mass eigenvalues m_i, Majorana phases α_i and mixing parameters U_{ei} given by

$$m_{ee} = \left| \sum_{i=1}^{3} |U_{ei}^2| \cdot e^{i\alpha_i} \cdot m_i \right| \tag{12}$$

Since the Majorana CP-phases are unknown, strong cancellations in the sum over all neutrino states ν_i can occur.

In the case of tritium β decay, the presence of mixing modifies eq. (10) to:

$$\frac{dN}{dE} = C \times F(Z,E) p (E + m_e c^2)(E_0 - E) \sum_i |U_{ei}|^2 [(E_0 - E)^2 - {m_i}^2]^{\frac{1}{2}} \Theta(E_0 - E - m_i), \tag{13}$$

The step function, $\Theta(E_0 - E - m_i)$, ensures that a neutrino state ν_i is only produced if the energy available is larger than its mass. In general, the effects of mixing will lead to the following spectral modifications :

1. the β spectrum ends at $E_{0'} = E_0 - m_1$, where m_1 is the lightest mass in the neutrino mass spectrum (i.e. the electron spectrum bends at $E \lesssim E_{0'}$)

2. the appearance of 'kinks' at the electron energy $E_e^i \approx E_0 - m_i$, with the size of the kinks being determined by the mixing elements $|U_{ei}|^2$.

For general mixing schemes with 3 neutrinos (3ν) or 3 active neutrinos and 1 'sterile' (4ν), the spectral shape of tritium β decay can be rather complex, requiring the introduction of at least five independent parameters (two mixing parameters and three masses for 3ν-mixing). In [22, 76] all possible 3ν or 4ν mixing schemes (with normal or inverted mass hierarchy) have been discussed extensively. In the following, we restrict our discussion to the schemes explaining the solar and atmospheric neutrino oscillation data. This omits the 4ν schemes incorporating the LSND oscillation results [8], which, however, have not been confirmed by other experiments such as KARMEN [9].

If the pattern of neutrino masses is hierarchical ($m_1 \ll m_2 \ll m_3$), the largest mass, $m_3 \simeq \sqrt{\Delta m^2_{atm}} = (4-7) \times 10^{-2}$ eV, would be too small to be observed in tritium β decay. For models of quasi-degenerate neutrino masses with an absolute mass scale in the range of sensitivity of future tritium β experiments, the effects of non-zero ν-masses and mixing reduce to a single parameter, $m^2(\nu_e)$. With $m_1 \approx m_2 \approx m_3$, the only distortion of the spectrum to be seen is a bending at the electron energy $E_0 - m(\nu_e)$, equivalent to the case of a ν_e with definite mass and no mixing. Therefore the analysis of the β spectrum can be parameterized by

$$m^2(\nu_e) = \sum_{i=1}^{3} |U_{ei}|^2 \cdot m_i^2 \quad . \tag{14}$$

In contrast to eq. (12), eq. (14) describes a non-coherent sum over all (squared) neutrino mass eigenstates to the electron neutrino. It is a real average with weighting factors $|U_{ei}^2| \geq 0$, such that no cancellations can occur. Hence, the neutrino mass $m(\nu_e)$ extracted from the experiment fixes the absolute mass scale ($m_1 \approx m_2 \approx m_3$), taking into account the small values of Δm^2 from oscillation experiments.

1.11.2 Absolute neutrino mass scale

While tritium β decay and $0\nu\beta\beta$ are largely complementary to each other, it is nevertheless interesting to compare the two parameters $m(\nu_e)$ and m_{ee} with each other (assuming neutrinos are Majorana particles) and to investigate their relation to the fundamental neutrino mass scale $\sum_\nu = \sum_{i=1}^{3} m_i$ in the presence of ν-mixing. For a 3ν mixing the following bounds on the β decay mass $m(\nu_e)$ can be derived [22]:

$$m_{ee} < m(\nu_e) < \frac{m_{ee}}{|| \cos 2\theta_\odot | (1 - |U_{e3}|^2) - |U_{e3}|^2 |}, \tag{15}$$

The ν_e mixing parameters in (15) can be deduced from the results of oscillation experiments using solar neutrinos (solar mixing angle $\theta_\odot$) and reactor antineutrinos ($|U_{e3}|^2$).

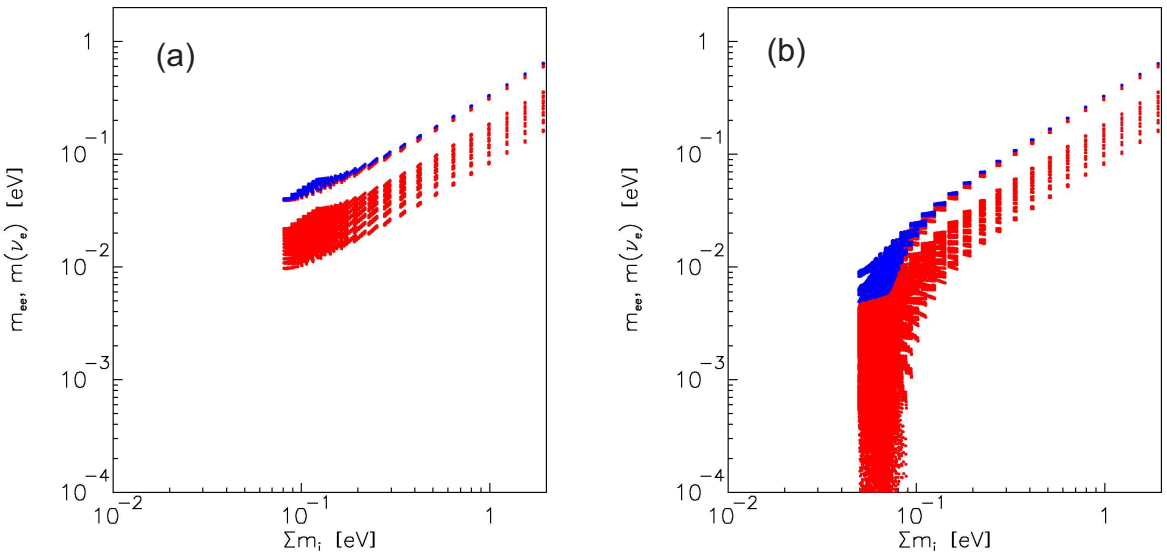

Figure 7: "Electron neutrino mass" $m(\nu_e)$ (blue) from a direct neutrino mass determination and effective neutrino mass m_{ee} (red) from $0\nu\beta\beta$ as function of the cosmological neutrino mass scale $\sum_\nu = \sum_{i=1}^{3} m_i$ for normal neutrino mass scenario ($m_3 > m_2 > m_1$, (a)) and for inverted neutrino mass scenario ($m_2 > m_1 > m_3$, (b)). The parameters of the neutrino mass differences and the neutrino mixing matrix are varied according to a very recent 3-flavor analysis [13] within their 3 σ uncertainties ($7.3 \cdot 10^{-5}$ eV^2 $\leq \Delta m^2_{12} \leq 9.3 \cdot 10^{-5}$ eV^2, $1.6 \cdot 10^{-3}$ eV^2 $\leq \Delta m^2_{23} \leq 3.6 \cdot 10^{-3}$ eV^2, $0.79 \leq |U_{e1}| \leq 0.88$, $0.47 \leq |U_{e2}| \leq 0.61$, $|U_{e3}| \leq 0.20$). The main uncertainty of m_{ee} originates from the unknown Majorana phases α_i, which were taken with the extrema $\alpha_i = 0$ and $\alpha_i = \pi$ (In principle all values of α_i are allowed, which would connect the lower and the upper bands of m_{ee}).

Fig. 7 shows the observables of a direct neutrino mass determination $m(\nu_e)$ and from double β decay experiments m_{ee} as a function of the cosmological neutrino mass scale $\sum_\nu$. It is clearly visible that the uncertainty of the neutrino mass matrix according to eq. (12) results in an uncertainty of half an order of magnitude. The majority of the uncertainty originates from the missing knowledge of the Majorana phases α_i. On the contrary, fig. 7 demonstrates that a direct neutrino mass measurement, or a limit, in the 0.1-1 eV range directly transforms into a limit determination on the fundamental or cosmological neutrino mass scale $\sum_\nu$.

In addition to the uncertainties from the neutrino mixing matrix 3 other uncertainties disfavor $0\nu\beta\beta$ for a precise determination of the fundamental neutrino mass scale:

1. Against the prejudices based on theoretical models, it is not experimentally proven that neutrinos are Majorana particles. In the case of Dirac neutrinos $0\nu\beta\beta$ is forbidden.

2. The uncertainties of the nuclear matrix element calculation of $0\nu\beta\beta$ has not improved significantly over the last decade. It is agreed that it corresponds to an uncertainty of factor 2 on the determination of the effective neutrino mass m_{ee}. One

way to reduce this uncertainty is to search for $0\nu\beta\beta$ with more than one isotope.

3. Neutrinoless double β decay does not need a neutrino to be exchanged between the two nucleon vertices. Also other new particles, *e.g.* from Supersymmetry, could be exchanged or other processes like right-handed currents could allow $0\nu\beta\beta$.

Majorana CP-phases

The argument that the unknown Majorana masses do not allow double β decay to be used for a precise determination of the neutrino mass scale could be reversed: The comparison of the "electron neutrino mass" $m(\nu_e)$ from a direct neutrino mass determination or the cosmological neutrino mass scale $\sum_\nu$ from astrophysics with the effective neutrino mass from $0\nu\beta\beta$ m_{ee} is the only realistic chance to get access to the unknown Majorana phases α_i [22, 77, 78, 79]. For a quasi-degenerate neutrino mass spectrum with generic Majorana masses, a measurement of the electron neutrino mass $\mathrm{m}(\nu_e)$ from ^{3}H β-decay and of the effective Majorana neutrino mass m_{ee} from $0\nu\beta\beta$ could, in principle, give evidence for Majorana CP-violating phases, even though no CP-violation would be directly observed.

Hence, the existence of CP-violation could be established by a precise experimental determination of the effective Majorana neutrino mass m_{ee} and the mass $\mathrm{m}(\,\nu_e\,)$ measured in tritium β-decay, taking into account independent precision measurements of the solar mixing angle $\theta_\odot$ and the mixing matrix element $|U_{\mathrm{e}3}|^2$. These measurements would also imply a non-trivial constraint on the two CP-violating phases α_1 and α_2. For more detailed discussions of the reconstruction of the neutrino mass spectrum (quasi-degenerate, inverted, normal or partial hierarchy) and of measuring Majorana CP-phases we refer to [22, 77, 78]. In [80] an update is given on these issues, taking into account the latest SNO results.

2 Tritium β decay experiments

The almost ideal features of tritium as a β emitter for neutrino mass measurements (low endpoint energy $E_0 = 18.6\,\mathrm{keV}$ at reasonably short half life $t_{1/2} = 12.3\,\mathrm{a}$, pure phase space spectrum because of super-allowed decay, simple final states due to lowest Z isotope) have been the reason for a long series of tritium β decay experiments. Figure 8 shows the

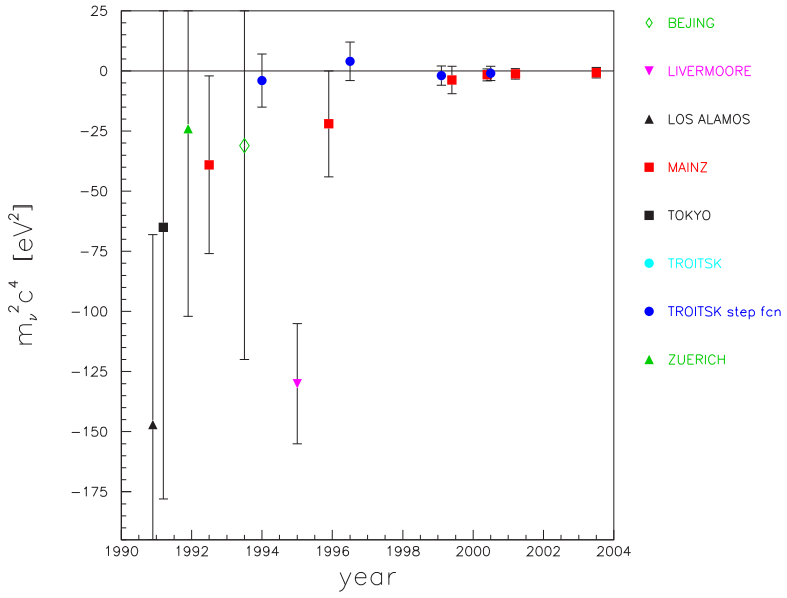

Figure 8: Results of tritium β decay experiments on the observable m_ν^2 over the last decade or so. The experiments at Los Alamos, Zürich, Tokyo, Beijing and Livermore [81, 82, 83, 84, 85] used magnetic spectrometers, the experiments at Troitsk and Mainz [86, 87] are using electrostatic spectrometers of the MAC-E-Filter type (see text).

evolution of the observable m_ν^2 of the various tritium β decay experiments over the last decade or so. It is remarkable that the uncertainties on m_ν^2 have decreased by nearly two orders of magnitude. Equally important is the fact that the problem of negative values for m_ν^2 of the early nineties has disappeared due to better understanding of systematics and improvements in the experimental setups.

2.1 MAC-E-Filter

The high sensitivity of the Troitsk and the Mainz neutrino mass experiments is due to a new type of spectrometer, the so-called MAC-E-Filter (Magnetic Adiabatic Collimation combined with an Electrostatic Filter). This new type of spectrometer was first proposed in [88]. Later, this method was reinvented specifically for the search of the electron neutrino mass at Troitsk and Mainz [89, 90], independently. It combines high luminosity

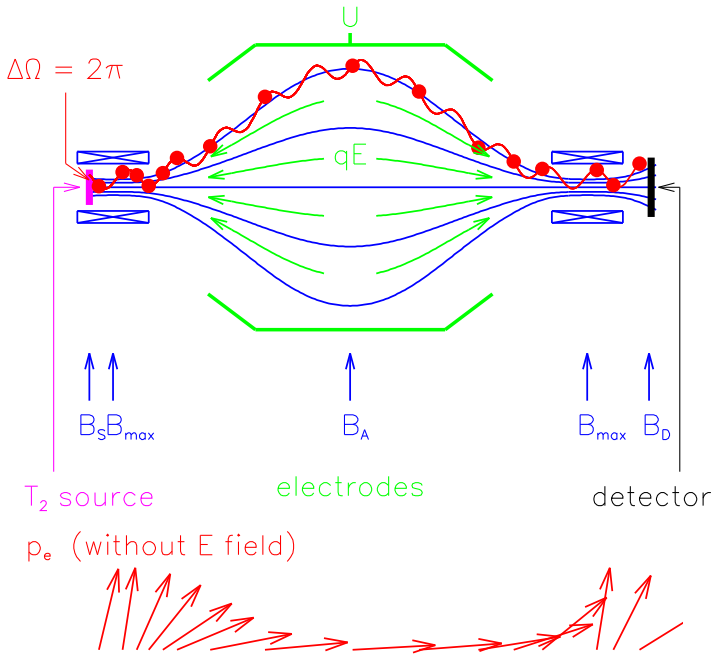

Figure 9: Principle of the MAC-E-Filter. (a) Experimental setup, (b) Momentum transformation due to the adiabatic invariance of the magnetic orbit momentum μ in the inhomogeneous magnetic field.

with a high energy resolution, both essential prerequisites to measure the neutrino mass from the endpoint region of a β decay spectrum.

The main features of the MAC-E-Filter are illustrated in fig. 9(a). Two superconducting solenoids produce a magnetic guiding field. The β-electrons, which start from the tritium source in the left solenoid into the forward hemisphere, are guided magnetically on a cyclotron motion around the magnetic field lines into the spectrometer, thus resulting in an accepted solid angle of up to 2π. On their way into the center of the spectrometer the magnetic field B drops by many orders of magnitude. Therefore, the magnetic gradient force transforms most of the cyclotron energy $E_\perp$ into longitudinal motion. This is illustrated in fig. 9(b) by a momentum vector. Due to the slowly varying magnetic field the momentum transforms adiabatically, therefore the magnetic moment μ keeps constant[5]:

$$\mu = \frac{E_\perp}{B} = const. \tag{16}$$

This transformation can be summarized as follows: The β-electrons, isotropically emitted at the source, are transformed into a broad beam of electrons flying almost parallel to the magnetic field lines.

This parallel beam of electrons is running against an electrostatic potential formed by one or more cylindrical electrodes. All electrons with enough energy to pass the electrostatic barrier are reaccelerated and collimated onto a detector, all others are reflected.

[5]Equation (16) is only correct in non-relativistic approximation. For relativistic particles $(\gamma+1)\cdot E_\perp/B$ is an adiabatic invariant of the motion.

Therefore the spectrometer acts as an integrating high-energy pass filter. From eq. (16) follows directly, that the relative sharpness $\Delta E/E$ of this filter is given only by the ratio of the minimum magnetic field B_A in the center plane and the maximum magnetic field $B_{\max}$ between the β-electron source and the spectrometer:

$$\frac{\Delta E}{E} = \frac{B_{\mathrm{A}}}{B_{\max}} \quad . \tag{17}$$

Varying the electrostatic retarding potential allows the measurement of the β spectrum in an integrating mode.

In order to suppress electrons which have a very long path within the tritium source and therefore exhibit a high scattering probability, the electron source is placed in a magnetic field B_{S} (see fig. 9), which is lower than the maximum magnetic field $B_{\max}$. This restricts the maximum accepted starting angle of the electrons $\theta_{\max}$ by the magnetic mirror effect to:

$$\sin\theta_{\max} = \sqrt{\frac{B_{\mathrm{S}}}{B_{\max}}}\,. \tag{18}$$

Following eq. (16), (17) and (18) the normalized transmission function of the MAC-E-Filter with retarding potential U is analytically given for an isotropic electron source of energy E by

$$T(E,qU) = \begin{cases} 0 & E - qU < 0 \\ \dfrac{1-\sqrt{1-\frac{E-qU}{E}\cdot\frac{B_{\mathrm{S}}}{B_{\mathrm{A}}}}}{1-\sqrt{1-\frac{\Delta E}{E}\cdot\frac{B_{\mathrm{S}}}{B_{\mathrm{A}}}}} & 0 \le E - qU \le \Delta E \\ 1 & E - qU > \Delta E \end{cases} \tag{19}$$

where q denotes the electron charge.

2.1.1 Time-of-flight mode (MAC-E-TOF)

The MAC-E-Filter is a very superior spectrometer concerning energy resolution and luminosity and it is specially suited for the investigation of the very upper end of a spectrum. However, for some applications the integrating feature might be a disadvantage, since any low energy feature under investigation is combined with the background of all the electrons of higher energies from the same spectrum.

There is a very elegant solution [91]: The electrons under investigation have energies close to the retarding potential and, therefore, are very slow when passing the analyzing plane. Electrons of higher energies have already a significant fraction of the speed of light. By measuring the time-of-flight of the electrons those of interest can be distinguished from the high energy electrons and the MAC-E-Filter transforms from an energy high-pass filter into an energy band-pass filter of the same energy resolution and nearly the same luminosity as the MAC-E-Filter without time-of-flight measurement [91].

The only problem is to determine the time-of-flight. For applications with a pulsed electron source (*e.g.* photoelectrons from a pulsed photon source) this task involves simply

measuring the time-of-arrival at the detector. When using a radioactive source like the tritium β decay, which does not provide the possibility to detect the time-of-decay by a coincidence, the only possibility is to chop the source frequently and very fast by applying a retarding high voltage (*e.g.* a sharp rectangular HF signal with a frequency of 100 kHz). The width of the "on-time" window reduces somewhat the energy resolution, the duty factor of the on-off cycle reduces the luminosity accordingly. Simulations and experiments at the Mainz setup have shown [91] that in the MAC-E-TOF chopped mode a triangle-like energy resolution function of the width (FWHM) as the full width ΔE of the transmission function of the integrating MAC-E-Filter can be achieved at a loss of about a factor of 4 in the count rate.

For tritium β decay experiments an ideal application for this non-integrating mode is the precise investigation of systematic corrections like the energy loss function.

2.2 The Mainz and the Troitsk experiments

The experiments at Troitsk [86] and Mainz [87] use similar spectrometers of the MAC-E-Filter type but they differ somewhat in size: the diameter and length of the Mainz (Troitsk) spectrometers are 1 m (1.5 m) and 4 m (7 m). However, the major differences between the two setups are related to the tritium sources.

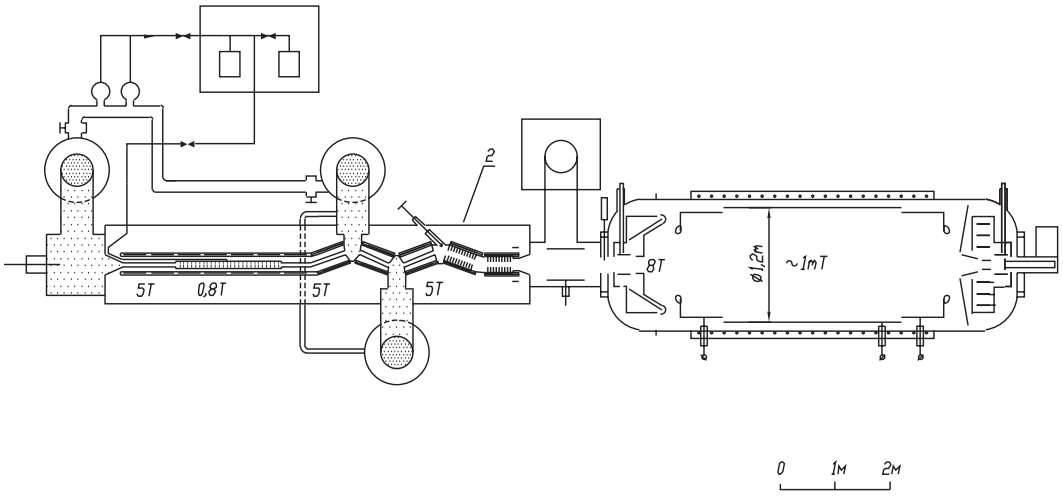

Figure 10: Schematic view of the Troitsk experimental setup [92].

The Troitsk experiment uses a windowless gaseous tritium source (WGTS) which is based on the adiabatic transport of electrons in a strong longitudinal magnetic field and circulation of tritium gas at low pressure by means of a differential pumping system [89] (see fig. 10). This approach was first pioneered in an experiment at Los Alamos [81]. An essential refinement made at Troitsk was the use of a strong magnetic field for electron transport. This technique permits the use of multiple bends in the transport channel, thus providing better differential pumping and smooth coupling to the MAC-E-Filter spectrometer. The Troitsk WGTS (a 3 m long tube of 50 mm diameter filled with

0.01 mbar of T_2) provides a number of beneficial features for the study of the tritium β spectrum, such as guaranteed homogeneity over the cross section of the source and reliable on-line control of inelastic energy losses of electrons in the source by use of a monoenergetic electron source at the rear end. Furthermore, it allows use of theoretical calculations of free molecular final state corrections and almost totally suppresses backscattering.

Mainz used a film of molecular tritium quench-condensed onto a substrate of highly oriented pyrolytic graphite. The film had a diameter of 17 mm and a typical thickness

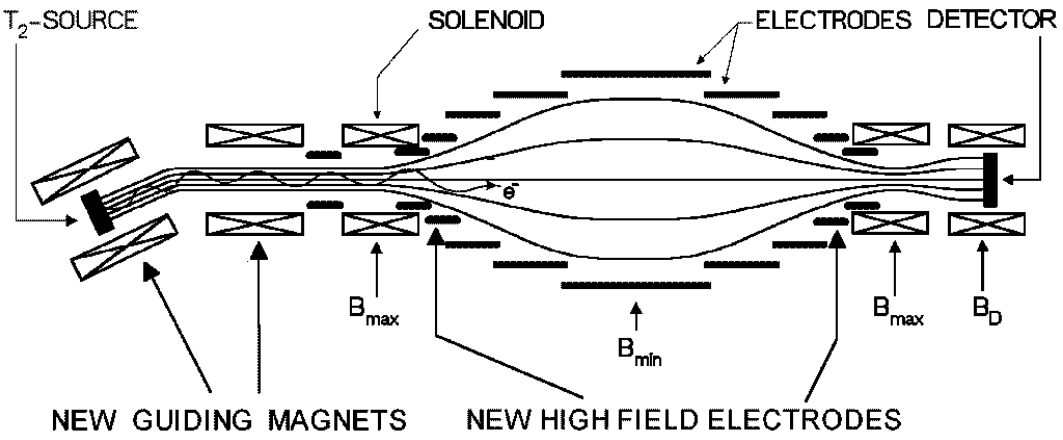

Figure 11: The upgraded and improved Mainz setup (schematically, not in realistic scale). The outer diameter amounts to 1 m, the distance from the source to the detector is 6 m [87].

of 40 nm, which was measured by laser ellipsometry. In the years 1995–1997 the Mainz setup has been upgraded to enhance the signal count rate, and to decrease the background, by placing a magnetic bend with cryotrapping between the tritium source and spectrometer. This measure reduced the tritium-related background at even higher source strengths. A second substantial improvement was the installation of a new cryostat which provided temperatures of the tritium film below 2 K to avoid a roughening transition of the film, which had been a source of systematic errors in earlier Mainz measurements. The roughening process is a temperature-activated surface diffusion process, therefore low temperatures are necessary to get time constants much longer than the duration of the measurement [105, 106]. The full automation of the apparatus and remote control allowed long term measurements over several months per year to be performed. Figure 11 gives a sketch of the Mainz setup. Since this upgrade, the count rate, background and energy resolution of the Mainz setup were similar to those of the Troitsk experiment.

2.2.1 Results of the Troitsk experiment

The Troitsk experiment has taken tritium data for 200 days since 1994. From the start, the Troitsk experiment has observed a small anomaly in the energy spectrum, located a few eV below the β endpoint E_0. The distortion resembles a sharp step in the count rate [92]. Since a MAC-E-Filter is integrating, a sharp step corresponds to a narrow line in the

primary spectrum. The data indicate a relative intensity of about 10^{-10} of the total decay rate. From 1998 onwards, the Troitsk group reported that the position of this line seems to oscillate with a frequency of 0.5 years between 5 eV and 15 eV below E_0 [86]. The origin of such an anomaly is not known. Indications for such a distortion were found at Mainz only once before stabilizing and reducing the background by applying high frequency pulses in measurement pauses. In addition, parallel measurements at Troitsk and Mainz clearly indicate that the anomaly at Troitsk is an experimental artefact. Experimental upgrades which lowered the background rate at Troitsk also reduced the size of the anomaly.

Fitting a standard β spectrum to the Troitsk data results in significantly negative values of $m(\nu_e)^2 \approx -10$ to -20 eV^2. However, describing the anomaly phenomenologically by adding a monoenergetic line with free amplitude and position to a standard β spectrum results in values of m_ν^2 compatible with zero. The average over all Troitsk runs from 1994–1999 and 2001 amounts to

$$m^2(\nu_\mathrm{e}) = -2.3 \pm 2.5 \pm 2.0 \text{ eV}^2 \tag{20}$$

which corresponds to an upper limit on $m(\nu_e)$ –under the assumption that the phenomenological anomaly description is correct– of

$$m(\nu_e) < 2.05 \text{ eV} \quad (95\ \%\ \text{C.L.}) \tag{21}$$

2.2.2 Results of the Mainz experiment

Figure 12 shows the endpoint region of the Mainz 1998 and 1999 data in comparison with the former Mainz 1994 data. An improvement of the signal-to-background ratio by a factor 10 following the upgrade of the Mainz experiment in 1995–1997, as well as a significant enhancement of the statistical quality of the data by long term measurements, are clearly visible. The main systematic uncertainties of the Mainz experiment are connected to the physics and the properties of the quench-condensed tritium film and originate from the inelastic scattering of β-electrons within the tritium film, the excitation of neighbor molecules due to the β decay, and the self-charging of the tritium film by its radioactivity. As a result of detailed investigations at Mainz [94, 96, 97, 98] –mostly by dedicated experiments– the systematic corrections became much better understood and their uncertainties were reduced significantly.

From 1998 onwards the background was stabilized and reduced by disturbing the storage condition for charged particles in the Mainz spectrometer by applying high frequency pulses at one electrode within measurement pauses. The comparison of the data taking in Mainz in 2000 under bad background conditions and the data of 2001, obtained after very carefully preparing the experiment –with respect to vacuum, high-voltage and the quench-condensed tritium source– proved how important clean experimental conditions are. The data of 1998, 1999 and 2001 (see fig. 12), which profited from the background stabilization and the clean experimental conditions neither show a Troitsk-like anomaly nor any other residual problem. These Mainz data proved, that the Troitsk anomaly is an experimental artefact appearing at Troitsk.

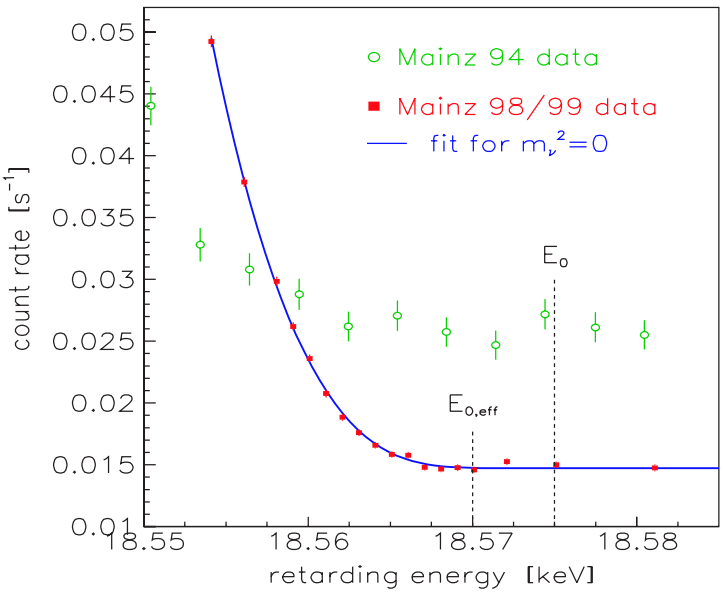

Figure 12: Averaged count rate of the Mainz 1998 and 1999 data (points) with fit (line) in comparison with previous Mainz data from 1994 [93] as function of the retarding energy near the endpoint E_0, and effective endpoint $E_{0,\mathrm{eff}}$. The position of the latter takes into account the width of response function of the setup and the mean rotation-vibration excitation energy of the electronic ground state of the $^3HeT^+$ daughter molecule.

The high statistics Mainz data from 1998-2001 allowed for the first time to determine the amplitude of the so-called neighbor excitation[6].

The most sensitive analysis on the neutrino mass, in which only the last 70 eV of the β spectrum below the endpoint are used, results for the Mainz 1998, 1999 and 2001 data including the experimentally obtained value for the neighbor excitation in (see figure 13)

$$m^2(\nu_\mathrm{e}) = -0.6 \pm 2.2 \pm 2.1 \ \mathrm{eV}^2 \tag{22}$$

which corresponds to an upper limit [96] of

$$m(\nu_e) < 2.3 \ \mathrm{eV} \quad (95\ \%\ \mathrm{C.L.}) \tag{23}$$

This analysis of the Mainz data up to 2001 improved not only the published former upper limit of $m(\nu_e) < 2.8\,\mathrm{eV}$ [87], which was based on the Mainz 1998 data alone, it is also the first analysis for which all critical systematic corrections were determined by dedicated experiments and analyses. Together with the Troitsk results, they represent the world's best sensitivity on a neutrino mass in a direct experiment.

[6]The sudden change of the nuclear charge during β decay can result in the excitation of neighboring molecules. The only theoretical calculation predicts this for 5.9 % of all tritium β decays for an ideal solid tritium crystal [95]. The Mainz group has corrected this value down to 4.6 % by considering the pores within the tritium film and the change of excited levels in the solid phase [87]. By the self-consistent analysis from the Mainz data a value of $5 \pm 1.6 \pm 2.2$ % was obtained [96], which is in very good agreement with the previously used correction.

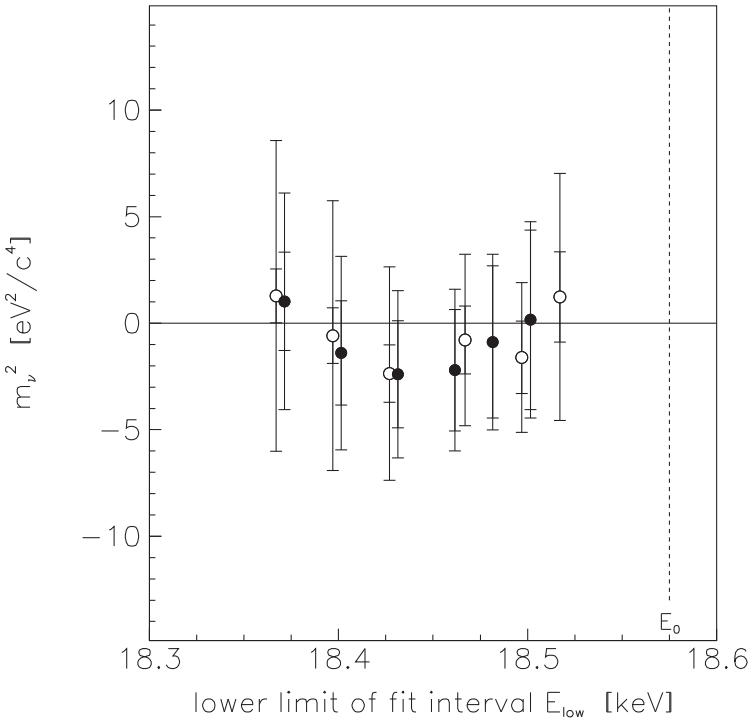

Figure 13: Mainz fit results on $m^2(\nu_e)$ as a function of the the lower boundary of the fit interval (the upper bound is fixed at 18.66 keV, well above E_0) for data from 1998 and 1999 [99] (open circles) and from the last runs of 2001 (filled circles) [96]. The error bars show the statistical uncertainties (inner bar) and the total uncertainty (outer bar). The systematic uncertainties for large fit intervals with a lower boundary $E_{low} < 18.5$ keV are correlated.

The Mainz tritium measurements were completed by the end of 2001. In two stages in 2002 and 2004 the Mainz group modified the spectrometer to study new ideas for background suppression in the KATRIN experiment:

- The modification of the shape of the electrodes to avoid Penning traps for electrons in corners of the electrode system resulting in a much lower pressure dependent background rate.
- The application of a "massless" screening electrode prohibiting electrons created by cosmic rays or environmental background at the electrodes from entering the sensitive magnetic flux tube and causing background (see sect. 8),
- The application of an electric dipolar field to eject trapped particles by a $\vec{E} \times \vec{B}$ drift (see chapter 8.3.1).

All these new ideas developed by computer simulations for the KATRIN experiment were successfully proven by the experiments at Mainz (see section 8).

2.3 Design criteria for a next generation tritium β decay experiment

New data from both the Mainz and Troitsk experiments will not improve the sensitivity on $m(\nu_e)$ significantly. This fact clearly underlines the importance of a next generation tritium β-decay experiment with a sub-eV sensitivity.

Fig. 6 already illustrated that the main requirements of a tritium β decay experiment to determine the neutrino mass are:

1. high signal rate S in the endpoint region
   The energy interval of interest below the endpoint $(E_0 - E)$ is proportional to the value of the neutrino mass aimed for, thus becoming smaller for a higher sensitivity on $m(\nu_e)$. However, according to eq. (10) the count rate within this interval depends to the cube power on $(E_0 - E)^3$ showing the need for a strong increase of the signal rate S when aiming for significantly smaller neutrino masses.

2. high energy resolution $E/\Delta E$
   The absolute energy resolution ΔE required scales with the value of the neutrino mass aimed for. Since the energy resolution of the Mainz and Troitsk experiments was not a critical or limiting factor an improved energy resolution down to $\Delta E < 1$ eV is sufficient for a factor 10 higher sensitivity on the neutrino mass of $\Delta m_\nu \approx 0.2$ eV.

3. low background rate B in the endpoint region

We will show that the signal rate S and the energy resolution $E/\Delta E$ are mainly defined by the size of the analyzing plane A_A. The former is a consequence of the conservation of the magnetic flux Φ and the fact that the signal strength cannot be enhanced by increasing the source column density above a limit due to inelastic processes [7], the latter is due to eq. (17) in combination with the conservation of the magnetic flux Φ.

2.3.1 General layout

Fig. 14 illustrates the general principles underlying a tritium β experiment: A tritium source of area A_S is located in a magnetic field B_S. B_S is usually lower than the maximum magnetic field B_max (magnetic pinch) to reject electrons with very large starting angles and therefore large path lengths within the tritium source. According to eq. (18) the ratio of B_S and B_max defines the maximum accepted starting angle $\theta_\mathrm{max} = \arcsin\sqrt{B_\mathrm{S}/B_\mathrm{max}}$. Assuming homogeneous magnetic fields over a cross section A, the conservation of the magnetic flux within which the electrons are transported results in

$$\Phi \;=\; \int B \; \mathrm{d}A \;=\; B_S \cdot A_\mathrm{S} \;=\; B_\mathrm{max} \cdot A_\mathrm{S,eff} \;=\; B_\mathrm{A} \cdot A_\mathrm{A} \tag{24}$$

[7] Any inelastic scattering process requires a minimum amount of energy loss of about 10 eV (see fig. 120), therefore only the zero loss fraction P_0 contributes to the count rate very close to the endpoint.

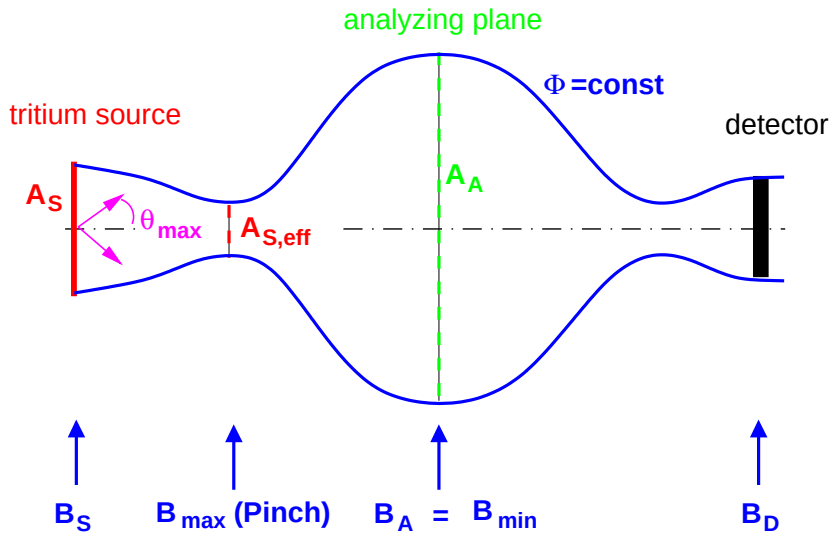

Figure 14: Schematic view of magnetic fields and cross sections of a tritium β decay experiment consisting of a source, a spectrometer of MAC-E-Filter type and a detector (omitting an optional pre-spectrometer).

The maximum cross section A_{A} of the electron flux tube is reached at the minimum magnetic field B_{A} in the analyzing plane. The ratio $B_{\mathrm{max}}/B_{\mathrm{A}}$ determines the relative energy resolution according to eq. (17). The electrons, which are able to pass the electrostatic barrier in the analyzing plane, are guided further to the detector located at the magnetic field B_{D}.

2.3.2 Signal rate and energy resolution

The signal rate S very close to the endpoint is proportional to

- the number of tritium molecules $N(T_2)$:

$$N(T_2) = A_{\mathrm{S}} \cdot \varepsilon_{\mathrm{T}} \cdot \rho d \tag{25}$$

  with the tritium purity ε_{T} and the column density ρd,

- the relative accepted forward solid angle $\Delta\Omega/2\pi$:

$$\Delta\Omega/2\pi = 1 - \cos\theta_{\mathrm{max}} \quad , \tag{26}$$

- and the probability $P_0(\rho d, \theta_{\mathrm{max}})$
  for a β electron not to undergo any inelastic scattering processes. $P_0(\rho d, \theta_{\mathrm{max}})$ is an average over all possible electron paths, *i.e.* an average over all starting angles up to θ_{max} and all starting positions within the source column density ρd.

With the proportionality constant a, S then reads

$$S = a \cdot N(T_2) \cdot \frac{\Delta\Omega}{2\pi} \cdot P_0(\rho d, \theta_{\mathrm{max}}) \tag{27}$$

$$= a \cdot A_{\mathrm{S}} \cdot \varepsilon_{\mathrm{T}} \cdot \rho d \cdot (1 - \cos\theta_{\mathrm{max}}) \cdot P_0(\rho d, \theta_{\mathrm{max}}) \tag{28}$$

By using the conservation of the magnetic flux eq. (24), the energy resolution eq. (17) and the maximum accepted starting angle eq. (18) the source area A_S can be expressed as

$$A_\mathrm{S} = A_\mathrm{A} \cdot \frac{\Delta E}{E} \cdot \frac{1}{\sin^2 \theta_\mathrm{max}} \tag{29}$$

Thus the signal rate S becomes

$$S = a \cdot \frac{A_\mathrm{A} \cdot \Delta E}{E} \cdot \frac{\varepsilon_\mathrm{T}}{1 + \cos \theta_\mathrm{max}} \cdot \rho d \cdot P_0(\rho d, \theta_\mathrm{max}) \ , \tag{30}$$

which transforms by eq. (24) into

$$S = a \cdot A_\mathrm{S,eff} \cdot \varepsilon_\mathrm{T} \cdot \underbrace{\frac{1}{1 + \cos \theta_\mathrm{max}} \cdot \rho d \cdot P_0(\rho d, \theta_\mathrm{max})}_{:= \ (\rho d)_\mathrm{eff}} \ . \tag{31}$$

The last three terms of eq. (31) can be understood as an effective column density $(\rho d)_\mathrm{eff}$ of a virtual source of unscattered electrons. This virtual source

- of effective column density $(\rho d)_\mathrm{eff}$ is
- placed in the maximum magnetic field B_max and
- emits with a relative strength of ε_T w.r.t. to pure tritium
- into the full forward solid angle of 2π
- from an effective source area $A_\mathrm{S,eff} = A_\mathrm{A} \Delta E / E$.

Due to the factor ρd one would expect that the effective column density $(\rho d)_\mathrm{eff}$ should increase for all θ_max with larger values for ρd. But since $P_0(\rho d, \theta_\mathrm{max})$ decreases at the same time the effective column density $(\rho d)_\mathrm{eff}$ approaches a maximum asymptotic value (see fig. 15):

$$(\rho d)_\mathrm{eff} \ = \ \frac{1}{1 + \cos \theta_\mathrm{max}} \cdot \rho d \cdot P_0(\rho d, \theta_\mathrm{max}) \quad \overset{d \to \infty}{\to} \quad \frac{(\rho d)_\mathrm{free}}{2} \ = \ \frac{1}{2 \cdot \sigma} \tag{32}$$

Equation (32) means that averaging over all emitted starting angles and weighting with the accepted solid angle the maximum effective source thickness is restricted to half of the mean free column density $(\rho d)_{free} = 1/\sigma = (2.94 \pm 0.06) \cdot 10^{17}$ cm^{-2}, where σ is the total inelastic cross section at an electron energy of 18.6 keV [94]. Thus,

$$S = a \cdot A_\mathrm{S,eff} \cdot \varepsilon_\mathrm{T} \cdot (\rho d)_\mathrm{eff} \ \leq \ a \cdot A_\mathrm{S,eff} \cdot \varepsilon_\mathrm{T} \cdot \frac{(\rho d)_\mathrm{free}}{2} \tag{33}$$

$$= \ a \cdot A_\mathrm{A} \cdot \frac{\Delta E}{E} \cdot \varepsilon_\mathrm{T} \cdot \frac{(\rho d)_\mathrm{free}}{2} \, . \tag{34}$$

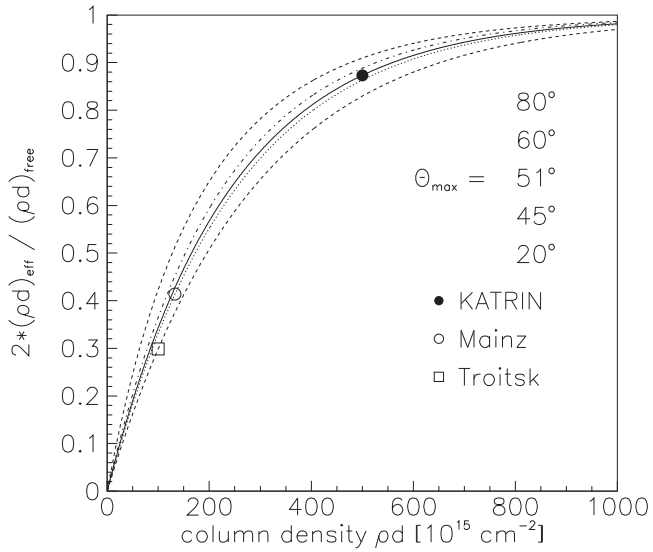

Figure 15: Ratio of effective to free column density $(\rho d)_{\text{eff}}/\left((\rho d)_{\text{free}}/2\right)$ (being proportional to the signal rate S) as function of the source column density ρd for different maximum accepted starting angles $\theta_{\max}$. The filled circle indicates the proposed WGTS parameters of $\rho d = 5 \cdot 10^{17}\,\text{molecules/cm}^2$, $\theta_{\max} = 51°$. The open circle (open square) show the parameters of the Mainz (Troitsk) experiments [87, 86].

Equation (34) can be transformed into

$$S \cdot \frac{E}{\Delta E} \leq a \cdot A_{\text{A}} \cdot \varepsilon_{\text{T}} \cdot \frac{(\rho d)_{\text{free}}}{2} \tag{35}$$

which shows that the maximum signal rate S times the relative energy resolution $E/\Delta E$ are governed only by the size of the analyzing plane A_{A} of the electrostatic spectrometer, the tritium purity ε_{T} and how close $(\rho d)_{\text{eff}}$ comes to $(\rho d)_{\text{free}}/2$.

Therefore, the design requirements of a new tritium β decay experiment with utmost sensitivity on the neutrino mass can be summarized as follows:

1. The analyzing plane A_{A} of the spectrometer should be as large as possible to achieve the necessary improvements in energy resolution $\Delta E/E$ and signal rate S in the endpoint region.

2. The tritium purity ε_{T} should be as close to 100 % as possible.

3. The column density ρd should be large enough that the effective column density $(\rho d)_{\text{eff}}$ is close to $(\rho d)_{\text{free}}/2$. If ρd is too large the signal rate S will not increase anymore, rather the systematic uncertainties will be enhanced. Fig. 15 shows, that a good compromise is a column density of $5 \cdot 10^{17}\,\text{molecules/cm}^2$, which provides a ratio of the effective column density to the maximum possible of 90%.

2.3.3 Design parameters of the KATRIN experiment

To reach a sensitivity on the neutrino mass of 0.2 eV a relative energy resolution of $E/\Delta E = 20000$ is proposed, which –according to eq. (17)– corresponds to the ratio of the maximum magnetic field $B_{\max}$ to the one in the analyzing plane $B_{\rm A}$. The magnetic fields have to be sufficiently high to guarantee the adiabatic transformation of eq. (16) and to minimize the influence of the Earths magnetic field or other stray fields. On the other hand, the magnetic field strengths should not be higher than 7-10 T to avoid unnecessary technical complications. We therefore propose to use the following magnetic fields:

$$B_{\rm A} = 3 \text{ G}\,, \quad B_{\max} = 6 \text{ T } (10 \text{ T})\,. \tag{36}$$

The $B_{\max}$ value in brackets can be used for special measurements with enhanced energy resolution.

Since the analyzing plane of the spectrometer should be as large as possible, we propose to use the maximum value for the diameter of the spectrometer vessel of about 10 m, which seems to be technically feasible. Considering the inner screening electrode system to suppress background (see section 8), and some safety margin for clearance of the transported magnetic flux to the screening electrode, we assume that at the analyzing plane the usable diameter is given by:

$$R_{\rm a} = 9 \text{ m} \quad \Rightarrow \quad A_{\rm A} = 63.6 \text{ m}^2\,. \tag{37}$$

From eq. (36), (37) and (24) the magnetic flux, which has to be transported from the tritium source through the whole system to the detector is thus

$$\Phi = \int B \mathrm{d}A = B_{\rm A} \cdot A_{\rm A} = 191 \text{ Tcm}^2\,. \tag{38}$$

The source magnetic field should be lower than the maximum magnetic field to reject electrons with large starting angles. On the other hand it should not be too low to avoid to have a very large source column density in order to reach a high ratio of $2{\cdot}(\rho d)_{\rm eff}/(\rho d)_{\rm free}$ (see fig. 15). We propose to use the following values:

$$\rho d = 5 \cdot 10^{17} \text{ cm}^{-2} \qquad B_{\rm S} = 3.6 \text{ T} \quad \Rightarrow \quad \theta_{\max} = 51^\circ\,. \tag{39}$$

The tritium purity $\varepsilon_{\rm T}$ has to be as close to 100 % as possible. Due to the tritium supply and purification by the Tritium Laboratory Karlsruhe (TLK) it is expected that we will achieve a value of

$$\varepsilon_{\rm T} = 95 \text{ \%}\,. \tag{40}$$

These figures given above correspond to an improvement of a factor 5 (4) in relative energy resolution, of a factor of 40 (20) in effective source area, of a factor 1.4 in tritium purity and a of a factor 2 (3) in effective column density yielding in total a factor of about 100 (100) in signal rate compared to the previous Mainz (Troitsk) experiments, respectively.

3 The KATRIN experiment

The requirements for a next-generation tritium β-decay experiment with sub-eV ν-mass

Figure 16: Aerial view of the site of Forschungszentrum Karlsruhe. A sketch of the location of the planned KATRIN experimental halls adjacent to Tritium Laboratory Karlsruhe (TLK) is superimposed.

sensitivity, which have been outlined in the previous chapter, constitute the basis for the design of the KATRIN experiment. An initial outline has been reported earlier in a Letter of Intent (LoI) [1] and an Addendum to the LoI [2]. The 2004 reference design of KATRIN presented below differs from the earlier outline due to the following major design changes:

- the luminosity of the tritium source has been enlarged by a factor of two by doubling the area of the transported flux tube at the source
- the diameter of the electrostatic spectrometer has been enlarged correspondingly from 7 m to the new reference value of 10 m
- the electromagnetic layout of the entire setup has been readjusted to and optimized for a reference magnetic flux of $\phi = 191\ \mathrm{T} \times \mathrm{cm}^2$
- the reference value of the tritium purity has been updated from an initial estimate of 75% to the new design criterion of 95%

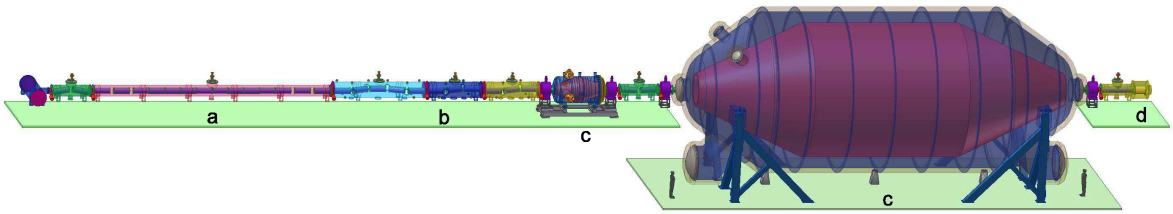

Figure 17: The 70 m long KATRIN reference setup with its major components: a) the windowless gaseous tritium source WGTS, b) the transport elements, consisting of an active pumping part and a passive cryotrapping section, c) the two electrostatic spectrometers and d) the detector for β-counting (not shown is the monitor spectrometer).

- the reference experimental configuration is a *linear* setup which minimizes the overall magnetic inhomogeneities due to stray fields
- the monitor spectrometer with its separate beam line

These modifications as well as corresponding design work (better scanning procedure, reduction of systematic effects) during a two year long optimization phase have improved the neutrino mass sensitivity of KATRIN from the initial estimate of m_ν=0.35 eV/c^2 [1] to the new reference value of m_ν=0.2 eV/c^2 (for details see section 12). In the following we give an overview of the new reference setup of KATRIN, which supersedes the earlier outlines reported in [1, 2].

3.1 Experimental overview

The Karlsruhe Tritium Neutrino (KATRIN) experiment will be performed on the site of Forschungszentrum Karlsruhe (FZK). Locating the KATRIN experiment at FZK allows to make use of the unique expertise of the on-site Tritium Laboratory Karlsruhe (TLK), which is the only scientific laboratory equipped with a closed tritium cycle and licensed to handle the required amount of tritium (licence: 40 g tritium $\approx 1.5 \cdot 10^{16}$ Bq). A further unique advantage of choosing TLK as host laboratory is the possibility to operate the tritium related parts (and in particular the tritium source) of KATRIN within the existing TLK building close to the tritium handling facilities. The non-tritium related parts of KATRIN, in particular the electrostatic spectrometers, will be housed in new buildings at the 'green field' site north of TLK.

The reference setup of KATRIN shown in fig. 17 corresponds to a $\sim$ 70 m long linear configuration with about 40 superconducting solenoids, which adiabatically guide β-decay electrons from source to detector. The experimental configuration of KATRIN can be grouped into four major functional units:

- a high luminosity Windowless Gaseous Tritium Source (WGTS) delivering 10^{11} β-decay electrons during the standard operation mode of the experiment

- an electron transport and tritium pumping section, comprising an *active* Differential Pumping Section (DPS) and a *passive* Cryogenic Pumping Section (CPS)
- a system of two electrostatic retarding filters with a smaller pre-spectrometer for pre-filtering and a larger main spectrometer for energy analysis of β-electrons
- a semi-conductor based high-resolution low background detector to count the β-electrons transmitted through the electrostatic filters.

In the following we briefly summarize the main features of the tritium-related part of KATRIN as well as the electrostatic spectrometers and the detector. More detailed descriptions for the WGTS and the transport section (sect. 4), the electrostatic spectrometers (sect. 5) and the detector (sect. 6) are presented later in this design report.

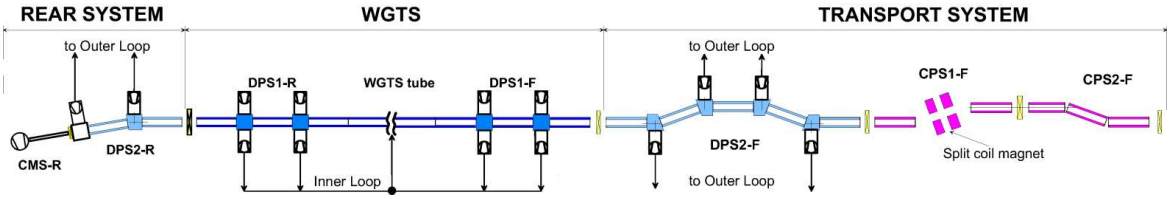

Figure 18: The tritium-related parts of KATRIN: the windowless gaseous tritium source WGTS and the front transport section, comprising the differential pumping unit DPS1-F as well as the two cryotrapping sections CPS1-F and CPS2-F.

3.2 Tritium-related section

The tritium-related parts of KATRIN are shown in more detail in fig. 18. They comprise the windowless gaseous tritium source WGTS as well as the differential (DPS) and cryogenic (CPS) pumping sections. The rear system comprises an additional differential pumping section and the calibration and monitoring system (CMS) which houses electron guns and an electron counting detector for calibration and monitoring purposes. An optional quench-condensed tritium source (QCTS) could be installed at a split-coil magnet within the cryogenic pumping sections CPS.

3.2.1 Tritium Source WGTS

The windowless gaseous tritium source WGTS will be the standard β-electron source of KATRIN for the long-term tritium measurements since it offers the highest luminosity and small systematic uncertainties. Ultra-cold molecular tritium gas (T =27 K) of high isotopic purity ($> 95\%$) will be injected through a set of capillaries at the middle of the 10 m long WGTS tube. The gas injection pressure p_{in} allows adjustment of the column density ρd of the WGTS. For a value of $p_{in} = 3.4 \times 10^{-3}$ mbar and a source tube temperature T = 27 K, the column density of the WGTS is fixed to the reference value $\rho d = 5 \times 10^{17}$ molecules/cm^2. After injection the T_2 molecules are transported over a length

of 5 m to both tube ends by diffusion. This process leads to a non-linear decrease of the tritium number density. At the first pumping port this parameter is at least a factor of 20 smaller than at the injection point [100]. The transport time of tritium molecules through the WGTS is of the order of 1 s. The decay probability of a single tritium molecule in the WGTS is thus about 10^{-9}. Electrons from the β-decay processes are adiabatically guided by the WGTS magnetic field of $B_S = 3.6\,T$ to both ends of the tube.

The total β-luminosity is proportional to the area A_S of the WGTS tube. To double the β decay luminosity of the WGTS (as compared to the initial outline in [1, 2]) the source diameter was increased to the new reference value $d_S = 90$ mm. For the present reference design, the WGTS will deliver a total (front and rear hemispheres) of 9.5×10^{10} β-decays per second within the guided magnetic flux of 191 $T \times cm^2$.

The main systematic uncertainty of the WGTS is associated with the column density ρd=5×10^{17} molecules/cm^2 which has to be known to a precision of 0.1 % (see section 11). This requires an appropriate stability of base parameters such as tube temperature and gas inlet pressure etc. Technical solutions to meet these demands are the proposed WGTS tube cooling system by means of two-phase neon (with a temperature stability of 30 mK) and a pressure controlled tritium buffer vessel for stabilized tritium injection.

3.2.2 Calibration and Monitoring System

At the most upstream end of the beam line a rear Control and Monitor Section (CMS) will be installed, which consists of multi-purpose electron guns to measure the overall KATRIN response function (transmission and inelastic scattering) and to investigate systematic effects, as well as of a rear detector to monitor the source activity and a rear WGTS plate to control the plasma effects within the source tube.

3.2.3 Transport System

The background generated by tritium decay within the spectrometers must be less than 10^{-3} counts/s which limits the amount of tritium permissible in the main spectrometer, equivalent to a partial pressure of tritium of about 10^{-20} mbar. This leads to a maximal allowed tritium flow rate into the pre-spectrometer (depending on the detailed design of the vacuum system of the pre-spectrometer and the connecting tube between it and the main spectrometer) of the order of 10^{-14} mbar l/s, the tritium inlet rate is about 2 mbar l/s (STP) (see sections 4 and 7). This design criterion requires that the tritium flow is suppressed by about a factor of 10^{11} between the outlet of the WGTS tube and the entrance of the pre-spectrometer. This very large suppression factor will be achieved by an effective tritium pumping system, based on a combination of differential (DPS) and cryogenic (CPS) pumping sections.

The first part of the tritium flow suppression is based on differential pumping. A series of pumping ports (see fig. 18), instrumented with turbomolecular pumps (TMP's) with high pumping capacity, reduces the tritium flow both at the rear (DPS1-R) and front sides (DPS1-F, DPS2-F). The DPS2-F element with four pumping ports, for example, will

reduce the tritium flow by more than a factor of 10^5. The DPS also has to guarantee a stable operation mode of the WGTS. As the DPS elements adjacent to the WGTS (DPS1-F, DPS1-R) contribute in a non-negligible way to the total column density ρd, they have to be operated under (almost) identical parameter values (temperature, magnetic field) as the WGTS. The requirements for stabilization and monitoring of the DPS-1 parameters are however less stringent than for the WGTS.

The differential pumping system will reduce the tritium flow to a such a low level that the subsequent passive cryotrapping system CPS will receive a tritium load of less than 1 Ci over a period of 60 days, which is equivalent to a standard tritium run cycle. The beam tube of the CPS is kept at a temperature of 4.5 K. At this temperature tritium molecules are passively adsorbed onto the lHe cold inner tube surfaces. To enhance the trapping probability for tritium molecules, the cold surfaces of the CPS transport tubes are anticipated to be covered by a thin layer of argon frost.

3.3 Electrostatic spectrometers

The energies of electrons from the WGTS, the quench condensed sources and the electron guns will be analysed by a set of electrostatic retarding spectrometers of the MAC-E filter type. In a MAC-E filter, the isotropic direction of emission of β-electrons at the source is transformed into a broad beam of electrons flying with a momentum almost parallel to the magnetic field lines. This parallel beam of electrons is energetically analysed by an electrostatic retarding potential. All electrons which have enough energy to pass the electrostatic barrier are re-accelerated and collimated by the upstream spectrometer magnet. Electrons with less energy than the retarding potential are reflected. Thus the spectrometer acts as an integrating high-energy pass filter.

The design of the KATRIN spectrometers is based on a novel electromagnetic concept: the retarding *high voltage* (HV) is directly connected to the hull of the spectrometer vessel itself. A nearly massless inner wire electrode at slightly more negative potential than the vessel itself suppresses low-energy electrons emanating from the inner surfaces of the spectrometer walls, representing a potential source of background. In addition this inner electrode fine-tunes the electrostatic field to avoid the occurrence of Penning traps in corners and to optimize the adiabatic transmission properties of the MAC-E filter.

The spectrometer section of KATRIN (see fig. 19) consists of two electrostatic spectrometers in a 'tandem setup'. Electrons have to pass the smaller pre-spectrometer first, which works at a fixed retarding potential, acting as a pre-filter. In normal tritium mode of the experiment it will reject low-energy electrons below 18.3 keV, which do not carry information on the neutrino mass. The remaining electrons enter the second, much larger spectrometer, where the energy spectrum close to the β decay endpoint is scanned with an energy resolution of 0.93 eV. The advantage of this tandem setup is the reduction of the total flux of electrons from the tritium source into the pre-spectrometer by a factor of 10^{-6}. This will minimize the background from ionization of residual gas molecules in the main spectrometer. In addition the pre-spectrometer serves as a test facility to validate the *extreme high vacuum* (XHV) and electro-magnetic design concepts for the larger main

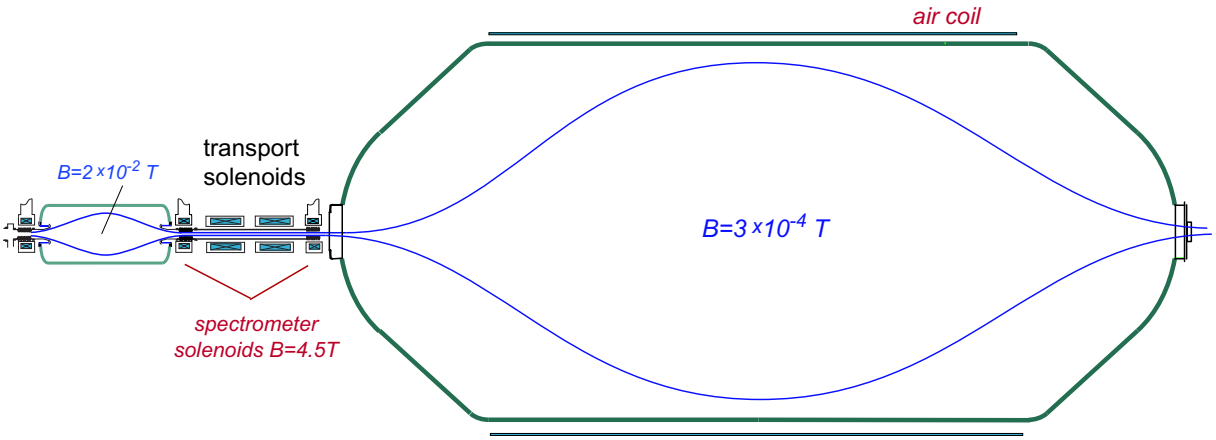

Figure 19: The two electrostatic spectrometers of KATRIN

spectrometer. In particular the requirement to reach and to maintain XHV conditions with a pressure of $< 10^{-11}$ mbar in both spectrometers constitute a major technological challenge for the KATRIN project. Both spectrometers are separated by a 2 m long superconducting 2-solenoid transport element.

3.4 Detector

Downstream of the main spectrometer and separated by a 2-solenoid transport element the electron detector is placed within a detector solenoid, which leaves enough room for an active and passive detector shielding to lower the detector background rate. The detector is a multi-pixel silicon semiconductor detector with ultra-high energy resolution and very thin entrance window.

3.5 Monitor beam line

The 70 m long KATRIN beam line is complemented by a second independent beam line for on-line monitoring of the actual retarding voltage at the main spectrometer. This second 5 m long monitor beam line runs parallel to the main beam line (see section 9.3). The monitor beam line has the following functional units:

- a monitor source (AmCo, CKrS), which emits monoenergetic electrons in the energy range from 17.8 – 32 keV
- a high resolution MAC-E filter (the modified Mainz spectrometer), which is fed by the same retarding HV as the KATRIN main spectrometer and thus provides an on-line monitoring of the retarding HV of the main spectrometer
- a segmented silicon based PIN-diode array for β-counting (the modified detector system used for the pre-spectrometer measurements).

To minimize the stray fields of the monitor beam line at the KATRIN beam line (and vice versa), the two setups are separated into two different buildings.

3.6 Technical challenges

Summarizing the different design requirements, it is obvious that the KATRIN experiment needs to meet the following technical challenges:

1. long term recirculation and purification of tritium on the kCi scale,
2. 10^{-3} temperature stability at 27 K,
3. extreme high vacuum ($< 10^{-11}$ mbar) at very large volumes (≈ 1400 m^3),
4. large number of superconducting magnets (≈ 30),
5. ppm stability for voltages in the 20 kV range, aim to reach ppm absolute precision as well,
6. simulations and Monte Carlo studies.

These items represent very interesting challenges for the KATRIN collaboration to be solved by diploma and PhD students as well as by engineers and physicists. Results from ongoing R & D efforts (*e.g.*: The outgassing rate reached with the KATRIN pre-spectrometer is more than a factor of 10 better than known outgassing rates in the literature.) will have applications in many different fields.

4 Tritium sources and other tritium related parts of KATRIN

The KATRIN experiment will make use of a molecular windowless gaseous tritium source (WGTS) as the standard source for the long-term tritium measurements as it offers the highest luminosity and small systematic uncertainties. Such a source, which is essentially a tube filled with molecular tritium gas, was first used at the LANL experiment [81] and developed further and adapted to the MAC-E-Filter by the Troitsk group [92].

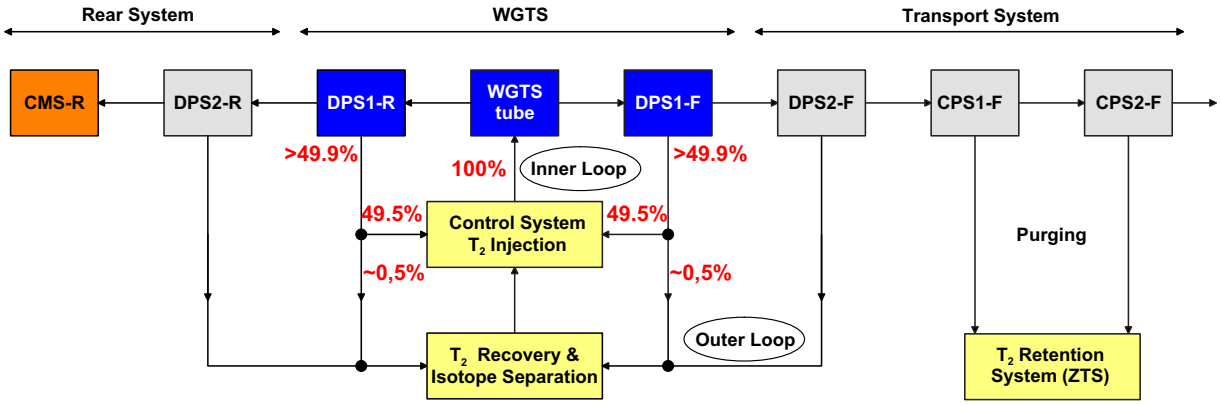

Figure 20: Block diagram of the tritium related parts of KATRIN (Rear System, WGTS, Transport System, Inner Loop) and their interfaces to the infrastructure of the Tritium Laboratory Karlsruhe (TLK).

Fig. 20 shows schematically the arrangement of the various components of the windowless gaseous tritium source and the other tritium related parts of KATRIN which are necessary

a) to achieve the required tritium source column density of $5 \cdot 10^{17}$ cm^{-2} with a stability of 0.1%

b) to adopt the WGTS to the XHV of the spectrometer and

c) to monitor source parameters.

Especially the required stable tritium column density is only achievable with the help of closed tritium loops and a controlled injection system.

As noted earlier the flow of tritium gas from the WGTS-tube into the spectrometer has to be very low for various reasons, the most important one being to avoid background from tritium decaying in the spectrometer volume. Therefore, modular differential pumping systems will be installed in the front side of the WGTS-tube (DPS1-F, DPS2-F). These comprise tubes of about 1 m length, with turbomolecular pumping stations installed between each pair of tubes. Similar differential pumping systems (DPS1-R, DPS2-R)

connect the rear side of the WGTS-tube to the calibration and monitoring system CMS in order to protect the CMS (electron guns, detector) and the rear ground plate from a too high tritium flux. These differential pumping sections also symmetrize the tritium density profile inside the WGTS-tube. Between the downstream differential pumping system and the pre-spectrometer two cryogenic pumping sections (CPS1-F, CPS2-F) will further reduce the tritium flow.

The whole WGTS and the tubes of the differential pumping and cryo-pumping sections will be enclosed by superconducting solenoids to guide the electrons adiabatically from the tritium source into the spectrometers.

The KATRIN experiment will be the only direct neutrino mass measurement, which will reach a sensitivity of 0.2 eV within the next decade. Therefore, it is very important for the KATRIN experiment to check as much as possible its own results and systematic corrections by different and independent methods. This is the main reason, why we still consider as an option to use in addition to the WGTS a quench-condensed tritium source, since most of its systematics are different to those of the WGTS. Such a quench-condensed source was developed and used at the Mainz experiment [87]. At present this kind of source suffers from the self-charging effect which creates in the equilibrium phase an electrical potential gradient of about 20 mV/monolayer (one monolayer of T_2 corresponds to about 10^{15} molecules per cm^2) and thus smears the β spectrum to a non-acceptable level. Currently, it is investigated whether the self-charging effect can be avoided or reduced significantly [101]. In this case, the KATRIN experiment could make use of a QCTS tritium source to check the results obtained with the WGTS. The optional QCTS would be placed within the split-coil magnet within the cryotrapping system CPS2-F.

The tritium related parts of KATRIN as well as their interfaces to the Tritium Laboratory Karlsruhe will be described in more detail in the following subsections.

4.1 WGTS and Inner Loop

The windowless gaseous tritium source WGTS will be the standard β-electron source of KATRIN for the long-term tritium measurements as it offers the highest luminosity and small systematic uncertainties. The main systematic uncertainty of the WGTS is associated with the column density $\rho d = 5 \cdot 10^{17}$ molecules/cm^2 which has to be known to a precision of 0.1 % (see sect. 11). This requires an appropriate stability of base parameters such as tube temperature and tritium gas injection rate. Technical solutions to meet these demands are the proposed WGTS tube cooling system by means of two-phase neon to stability of ± 30 mK and the setup of a so-called Inner Loop to achieve a tritium injection rate of $4.7\,Ci/s = 1.7 \cdot 10^{11}\,Bq \approx 2$ sccs (40 g tritium throughput per day) with a stability level of 0.1 % .

4.1.1 Properties of the WGTS

Figure 21 shows a schematic of the WGTS. A 10 m long cylindrical tube of 90 mm diameter at a nominal temperature of 27 K is filled with molecular tritium gas of high isotopic purity

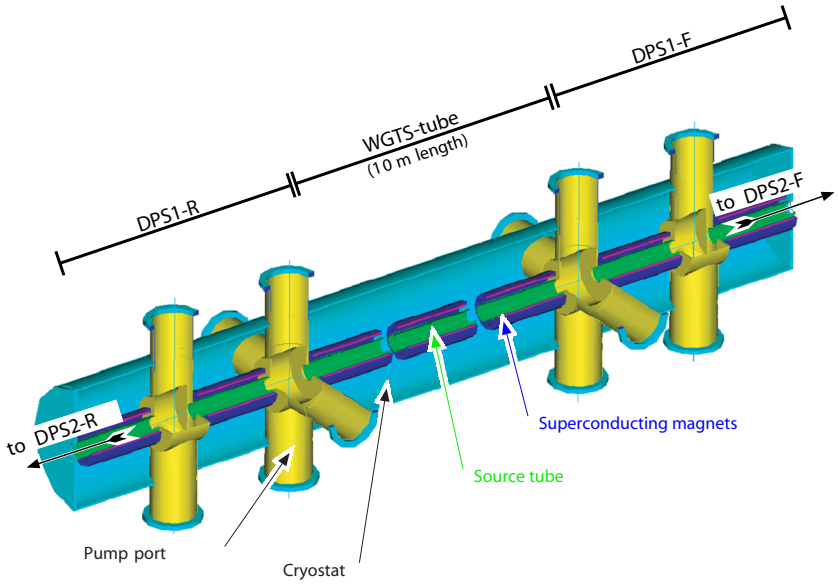

Figure 21: Schematic view of the windowless gaseous tritium source WGTS

($\varepsilon_T > 95\,\%$). The tritium gas will be continuously fed through a capillary to an injection chamber in the middle where the tritium is injected through more than 250 holes of an diameter of 2 mm thus avoiding gas jets. The tritium molecules freely stream from the middle to both ends and are pumped away by a series of differential turbomolecular pump stations (DPS1-F, DPS1-R).

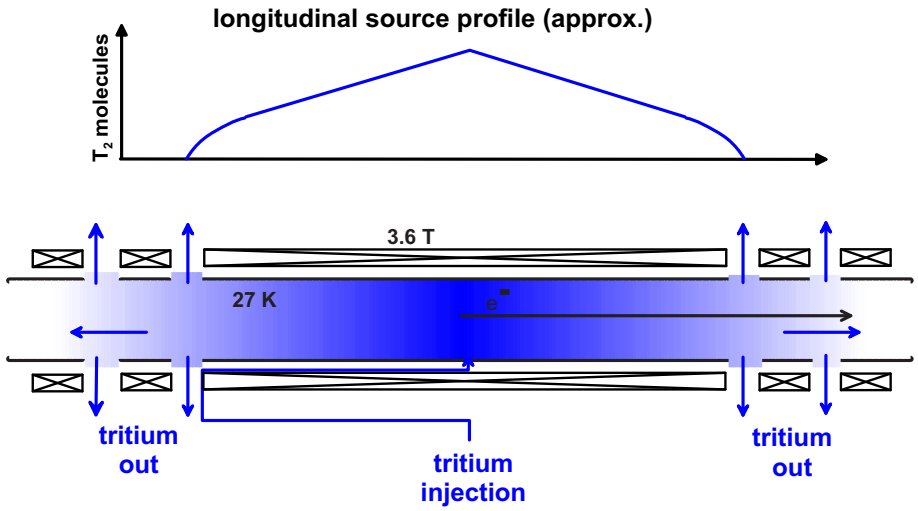

Figure 22: Bottom: tritium gas injection into the WGTS tube and first differential pumping stages. Top: density profile over source length (schematic).

Fig. 22 shows a schematic diagram and fig. 23 the density profile obtained by a detailed calculation considering the different pressure regimes from laminar flow to molecular flow.

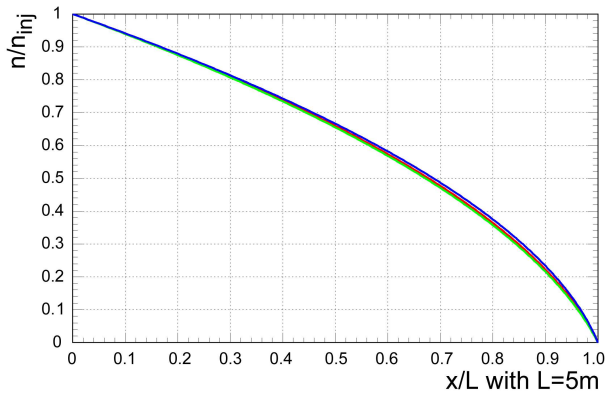

Figure 23: Analytical gas density profile within the WGTS from the injection point ($x = 0$) to the first pump port at $x = L$ for the input pressure $p_{\mathrm{in}} = 3/3.35/4$ μbar (green/red/blue line, respectively) at constant temperature $T = 27\,\mathrm{K}$. In this calculation zero pressure in the first pump port was assumed.

At the operation temperature of 27 K a central T_2 pressure p_{in} at the injection point of $3.35 \cdot 10^{-3}$ mbar results in a total T_2 column density of $5 \cdot 10^{17}\,\mathrm{cm}^{-2}$ and corresponds to a throughput of 1.853 mbar l/s (STP). This calculation assumes that the pressure at the end (i.e. at $x/L = 1$) is zero, which is not quite correct. The real pressure at the ends will be $0.036 \cdot p_{\mathrm{in}} < p(x/L = 1) < 0.05 \cdot p_{\mathrm{in}}$. Calculations have been carried out to study the changes in column density with this correction, as well as with variations of temperature and the input pressure [100]. Their influence on KATRIN systematics is discussed in section 11.4.

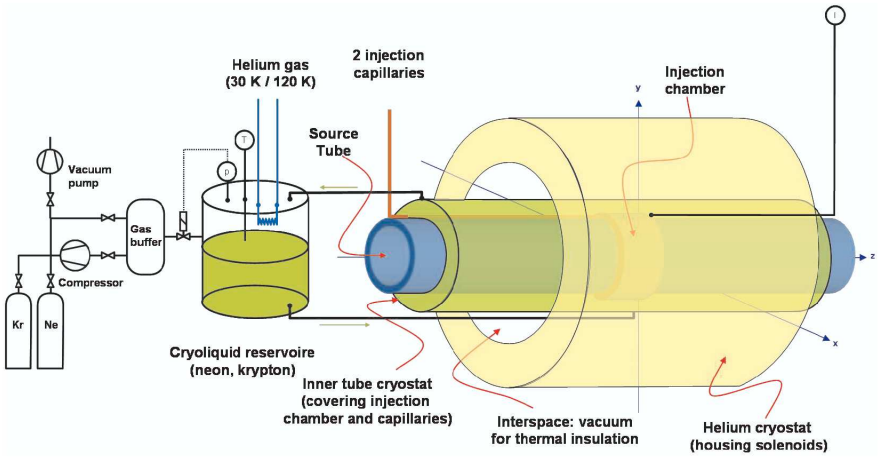

Figure 24: Cooling of the WGTS beam pipe by a cooling liquid at regulated pressure. The liquid is cooled by the helium liquifier.

The low source temperature of 27 K is chosen to obtain this high tritium density at a reasonably low tritium pressure and flow rate and to reduce the Doppler broadening and space charge potential[8]. Very essential for the stability of the column density ρd is the stability of the temperature to 0.1 % which is achieved by a boiling liquid neon cryostat surrounding the tritium tube (see fig. 24). The pressure control of the LNe in the range of 1.0-2.2 bar allows stabilization of the temperature in the range of 27-30 K very precisely. This LNe cryostat is surrounded by the LHe cryostat of the superconducting solenoid (see fig. 25). A second regime of operation at $T = 100 - 150$ K is obtained by exchanging the liquid neon for another liquid, *e.g.* liquid krypton.

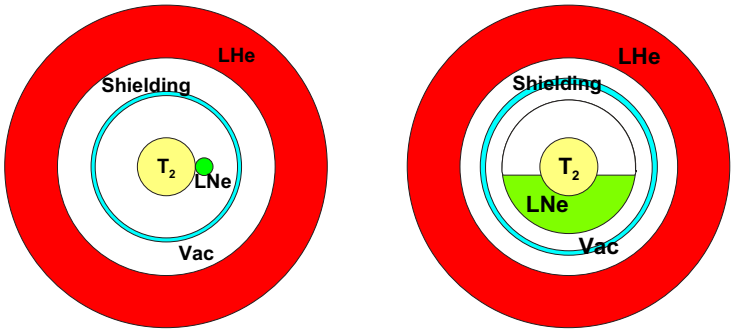

Figure 25: Schematic view of the cross section through the windowless gaseous tritium source WGTS. Shown are the two possibilities to stabilize the WGTS-tube temperature which are currently considered. Not displayed are the superconducting coils and the injection capillaries.

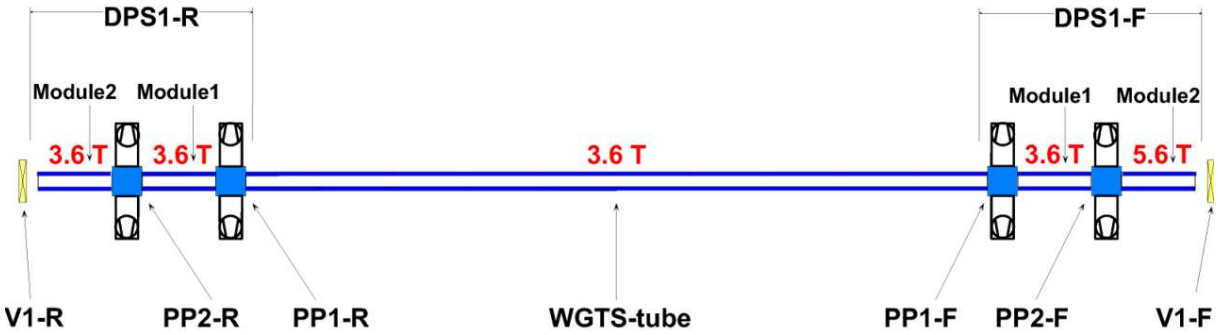

Figure 26: WGTS and magnetic field strengths along the source tube.

The magnetic field over the full volume of the tritium source of $B_S = 3.6$ T guides the β decay electrons adiabatically to both ends (see fig. 26). The ratio of magnetic field strengths at the source and at the pinch magnet B_S/B_{max} is chosen to be 0.6, so that the maximum accepted starting angle in the WGTS is $\theta_{max} = 51°$ and the accepted source

[8]To use much lower temperatures than 30 K is not possible, because tritium clusters may be created out of the T_2 molecules. As the former have different electronic final states this would give rise to uncontrollable systematic corrections.

area is $A_S = 53\,\mathrm{cm}^2$ (compare equations (18), (24) and (37)) resulting in a magnetic flux of $191\,\mathrm{Tcm}^2$. At the connections of the individual coils of the WGTS the magnetic field should not drop by more than 1 % in order to keep the fraction of electrons which are trapped by the magnetic mirror effect in the low magnetic field region very low. A simulation showed that for this small value of the inhomogeneity the influence of trapped electrons, which escape by large angle scattering into the solid angle accepted by the main spectrometer can be neglected (see sec. 11.4.12).

The column density of $\rho d = 5 \cdot 10^{17}/\mathrm{cm}^2$ and the magnetic field ratio are the optimum values following our design discussions in section 2.3 (see fig. 15). With the magnetic field at the source of $B_S = 3.6\,\mathrm{T}$ the necessary source diameter is $d_s = 8.2\,\mathrm{cm}$ according to eq. (24) and (38). The proposed diameter of $d_s = 9.0\,\mathrm{cm}$ leaves a reserve of 4 mm in radial direction to compensate misalignments of the whole system.

We are considering to plate the inner tube of the WGTS with gold to ensure constant and homogeneous electrical surface potentials. For scanning the tritium β spectrum it is technically easier to keep the main spectrometer retarding potential constant but to vary the voltage at the source. Therefore the whole WGTS system is insulated against ground potential. Electrical potentials of up to $\pm 1\,\mathrm{kV}$ are allowed.

4.1.2 Differential pumping sections DPS1-F/R

Each of the elements DPS1-F and DPS1-R has two pumping ports with 4 (PP1) respectively 2 (PP2) ducts. Each of these ducts (channels) has a pumping speed of 2000 l/s and will be equipped with turbomolecular pumps with pumping speed of about 2000 l/s for H_2. A first Monte Carlo study was performed to assess the flow rate reduction factor of DPS1. The inlet conditions of DPS1 are just at the edge of molecular flow, which leads to some overestimation of the reduction factor, and the final geometry still has to be elaborated together with industry. Fig. 27 shows the results of gas flow and density reduction for DPS1 for two pump configurations, namely 2 pumps at the first pumping port and 4 pumps, respectively; the latter being the KATRIN reference solution. It is clearly indicated that, under these assumptions, reduction factors of the order of 10^3 are achievable, which has to be confirmed upon the detailed design of DPS1. A discussion about the combined tritium flow reduction factor of DPS1-F and DPS2-F can be found in sec. 4.2.1

The differential pumping sections DPS1-F and DPS1-R will be equipped with two transversal scanning magnets each, which will allow the beam from the electron gun in the rear system (see sect. 4.3) to be moved along the cross section of the WGTS. To achieve a sufficiently large shift of the electron beam a magnetic dipole of a strength of 0.2 Tm is needed. These scanning magnets could be realized by two pairs of so-called "cosine coils" wound onto the solenoidal coil (see fig. 28).

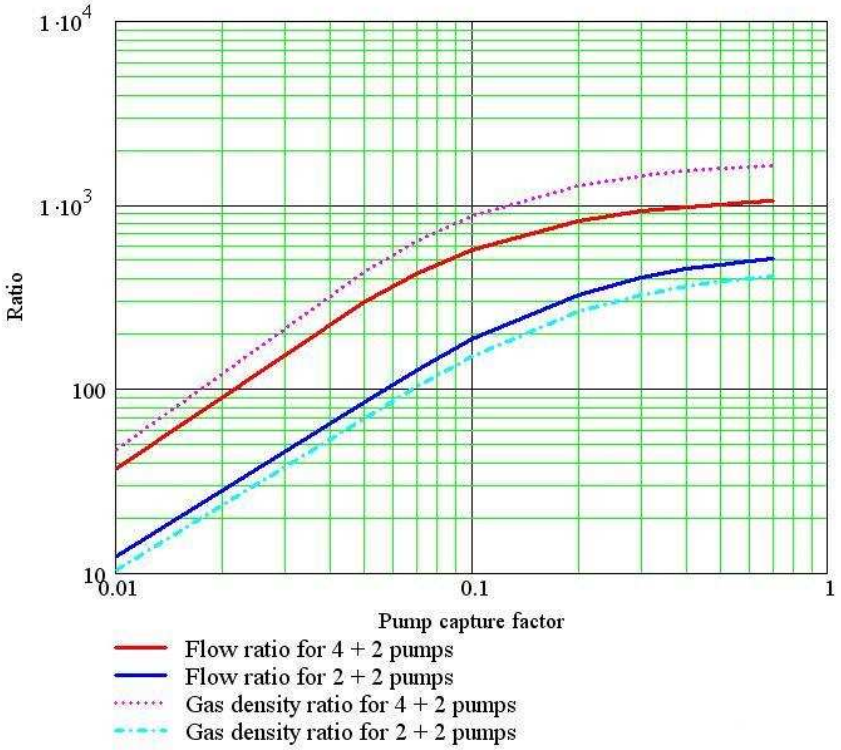

Figure 27: Tritium reduction factor for the differential pumping section DPS1-F as a function of the turbomolecular pump capture factor α, assuming molecular flow conditions.

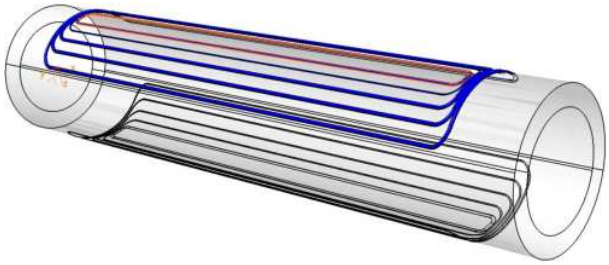

Figure 28: Scheme of a scanning magnet in z-direction. To achieve a rather homogeneous magnetic dipole field with a cylindrical volume the wires are mounted on this cylinder at constant intervals in y.

4.1.3 Inner Loop

The tritium gas flow is formed as a nearly closed loop (see fig. 29). The turbomolecular pumping stations pump the tritium gas through a palladium/silver membrane filter (permeator) back into the injection line. A pressure controlled WGTS buffer vessel keeps the tritium injection rate of $q = 1.853\,\mathrm{mbar\,l/s}$ (STP)[9] constant to obtain the column density of $\rho d = 5 \cdot 10^{17}\,/\mathrm{cm}^2$. The cascaded arrangement of the input and output gas flows of

[9]This injection rate corresponds to a tritium purity of $\varepsilon_\mathrm{T} = 0.95$ to an injected activity of $1.7{\cdot}10^{11}$ Bq/s.

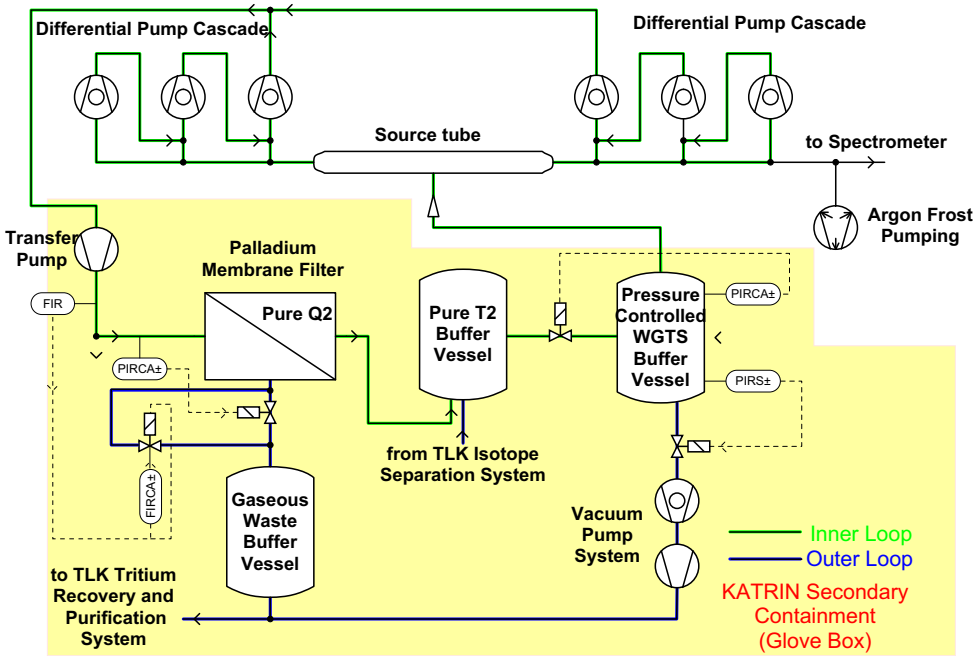

Figure 29: Schematic view of the gas flow within the WGTS, the tritium loop and the differential pumping stations. Not shown is the gaseous krypton supply system, which is needed for the measurements with gaseous ^{83m}Kr added to the gaseous tritium in the WGTS in order to determine, by the spectroscopy of ^{83m}Kr conversion electrons, the electrical potential within the WGTS.

the turbomolecular pumps results in a chain of differential pumping stations with a much higher reduction of the T_2 pressure than a single turbomolecular pump would achieve. Since the permeator can only separate non-hydrogen isotopes from the circulating gas a small part, about 1 %, of the gas flow rate needs to be fed into a tritium recovery and purification system and to be replaced with pure tritium ($\varepsilon_T \approx 98\%$, for TLK infrastructure see section 4.6). This guarantees a permanent high tritium concentration of $\varepsilon_T > 95$ % within the WGTS. The turbomolecular pumps will have a nominal pumping speed of 2000 l/s each and will be tritium compatible. The first pumping port will be equipped with 4 pumps instead of one in order to come close to the ideal molecular flow regime in the differential pumping section. Additionally this measure limits the β-electrons emitting source volume to the WGTS which avoids systematics uncertainties from regions at different electrical potentials and magnetic fields.

Of special importance is the control of the column density on the 1 per mill level by regulating the pressure in the WGTS supply buffer vessel. Fig. 29 shows that two levels of gas flow rate stabilization are foreseen in KATRIN:

- High Level of Stabilization system (HLS)
  The 'Pressure Controlled WGTS Supply Buffer Vessel' (PCBV) with pressure control devices (pressure sensor, valve controller and regulating valve) and injection channel (with known conductance) represent the higher level of stabilization.

- Low Level of Stabilization system (LLS)
  The turbomolecular pumps of DPS1-F/R, transfer pump, palladium/silver membrane filter (permeator) and the 'Pure T_2 Buffer Vessel' with the pressure control system (not shown in fig. 29) represent the lower level of tritium flow rate stabilization.

4.1.4 Test experiment TILO

The test experiment TILO (Test of Inner LOop) currently performed at TLK will check the principle design of the proposed setup of the Inner Loop. TILO will use protium (^{1}H) and deuterium gas instead of tritium. The aim of the experiment is to demonstrate a stabilized injection rate at the level required for KATRIN. Additional important objectives for this test experiment are tests of different tritium compatible pumps for the loop, specification of the pressure distribution along the circulation loop and tests of different modes of loop operation to get information with regard to safety discussion. The flow diagram of TILO

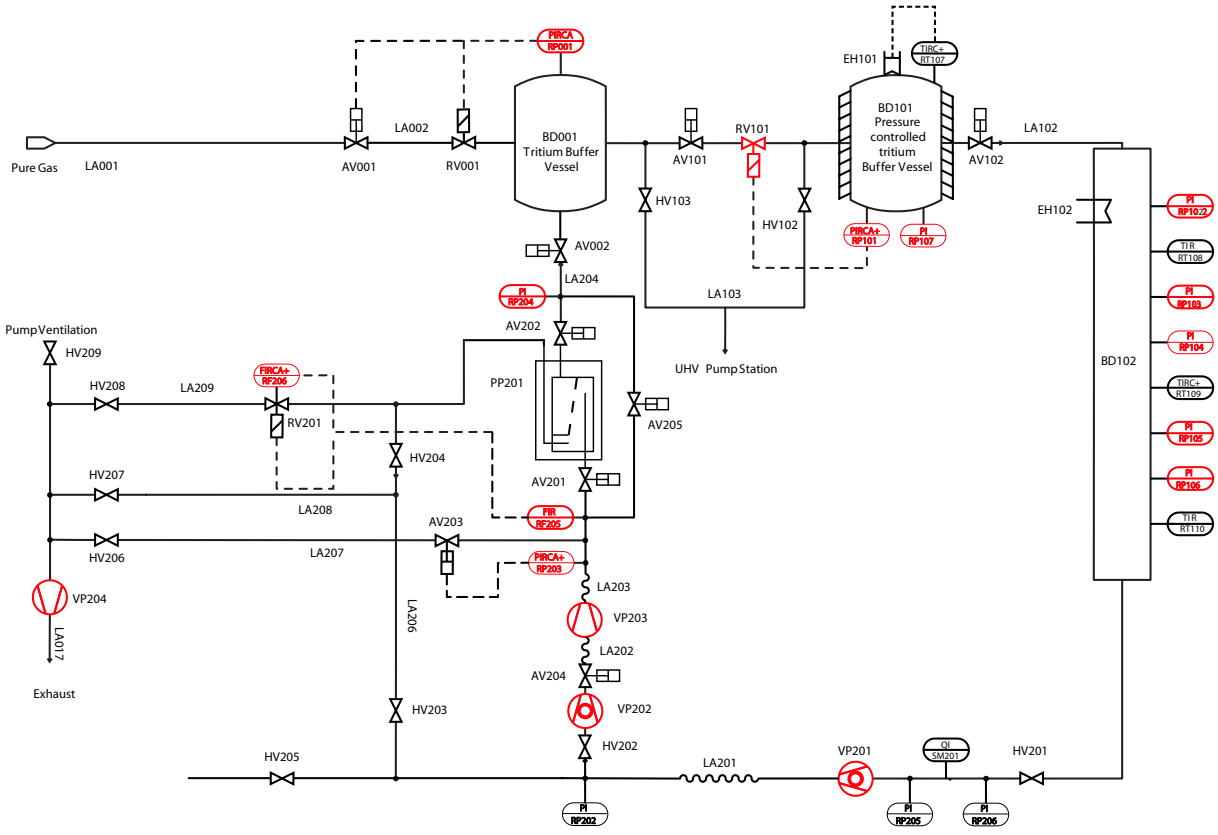

Figure 30: P&I diagram of the TILO experiment.

is shown in fig. 30. A tube of about two meters length (BD102) will be used as a mock-up of the windowless gaseous tritium source. The tube diameter of 65 mm provides the conductance for the Knudsen flow of D_2 with about the same average molecule density

as it is foreseen in the WGTS. The tube temperature will be stabilized in the range of 0.1 % to a few percent allowing to check the temperature dependence of the WGTS column density. The arrangement of high accuracy capacitance pressure sensors at the tube will give information about the pressure distribution along the tube. The capacitance manometers have been chosen because their sensitivity is independent from the type of gases processed in the system. Thus the change from protium to deuterium will not require any additional calibration of the manometers. This fact increases the reliability of the measurements and allows for a later application with tritium. Gas emitting the tube will be pumped by the turbomolecular pump VP201 (with magnetic bearing), compressed to a pressure appropriate for the permeator (PP201) by the transfer pumps VP202 (magnetic bearing) and VP203 (5 stage compressor) and finally passed into the tritium buffer vessel (BD001). Purified protium (deuterium) from the LLS (BD001, RP001, RV001,PP201, VP202, VP203, VP201) will pass over the HLS (BD101, RP101, LA102) through the automatic control valve RV101, which is controlled by a valve controller combined with a BARATRON 690 type pressure sensor (RP101) and a signal conditioner of MKS-670 type.

In principle such a system can provide the pressure stabilization in BD101 with 0.1% accuracy. Nevertheless, an experimental check is necessary. The pressure in the LLS cannot be kept constant on that high level, since elements like the permeator or the transfer pumps deteriorate the pressure stability. But, as mentioned already above, this high stability fortunately is not necessary, since these elements only slightly affect the pressure at the tube outlet. The pressure in BD001 will be maintained at a level of 150 to 200 mbar, however, fast fluctuations of pressure, which can affect the regulating properties of HLS, must be suppressed. To fulfill this requirement the volume of BD001 was chosen to be 2 l (in that case the pV value of BD001 and BD101 is about the same).

The combination of the flow meter RF205 and flow controller RF206 adjusts the exhaust flow rate from the Inner Loop to 1% of the processed value. Pressure controller RP001 together with control valve RV001 are used to compensate this gas losses by feeding pure protium (deuterium) from gas supply into the loop. Valve HV205 will allow to feed different impurity gases into the loop.

A first successful demonstration of the pressure stabilization was done with a simplified gas circulation system, as it is shown in fig. 31. The pressure in the 15 l buffer vessel was controlled by a system consisting of a high accuracy A690, 100 torr range MKS pressure sensor with a 670 type signal conditioner, which was connected to a 250 type MKS valve controller (RP001) operated by a J148, 200 sccm range MKS control valve (RV001). An ASP28 dry roots pump from Alcatel was used as a circulating pump. The circulated gas flow rate was limited by a tube with 4 mm inner diameter and 3 m length. To eliminate possible pulsation of the gas flow rate at the exhaust side of the pump, a second buffer vessel (2 l) was installed between the pump and the regulating valve. The pressure in the 15l buffer vessel was measured with a second high accuracy A690, 100 torr range MKS pressure sensor with a 670 type signal conditioner. The test was performed with helium gas.

Fig. 32 presents one of the first results with the simplified setup. It is clearly visible

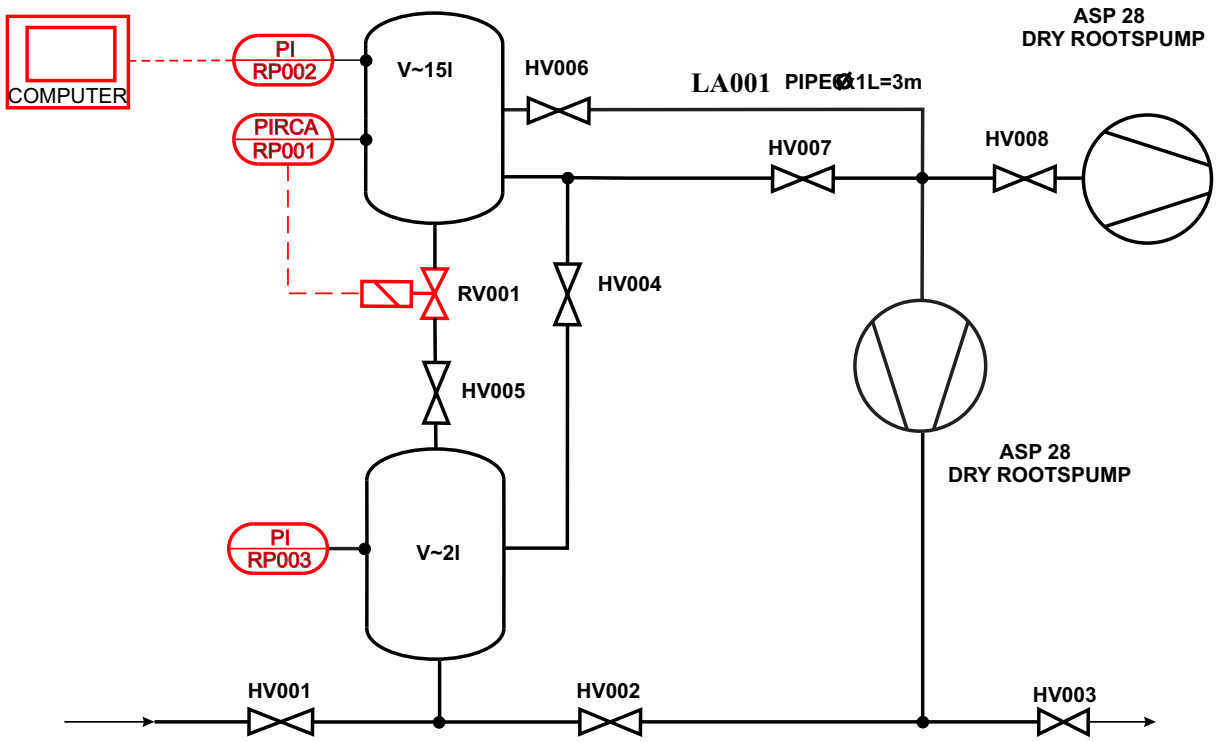

Figure 31: Schematic view of the TILINO test experiment.

that the peak to peak fluctuation is less than 0.1%. We have also examined the influence of different possible perturbations of the system conditions such as temperature changes of different loop elements and pressure changes in the 2 l vessel. These measurements demonstrated that the pressure stabilization on a level of 0.1 % is possible if the perturbations are not on a short time scale. For instance, increasing or decreasing the pressure of the 2 l vessel has almost no influence on the pressure of the 15 l vessel, if the change of the pressure is less than 2 mbar/s.

4.1.5 Potential distribution within the WGTS

The activity of the tritium gas inside the WGTS is about 10^{11} Bq. Most of the β-electrons are leaving the WGTS directly to both ends following the magnetic field lines whereas the positive daughter ions are thermalized immediately. Their diffusion to the wall of the tube is substantially reduced due to the transverse magnetic confinement of the plasma, such that their storage within the WGTS would be determined by the longitudinal pump out time towards the DPS[10]. Therefore a large positive space charge and potential could build up within the WGTS. Fortunately, this positive space charge can be compensated by thermalized electrons in the WGTS, which stem from the 15 or so ionization events caused by each β-electron. These secondary electrons will be cooled by inelastic and elastic

[10] At this location, they are driven out by additional dipolar electric fields, see sect. 11.3 and 11.4.8).

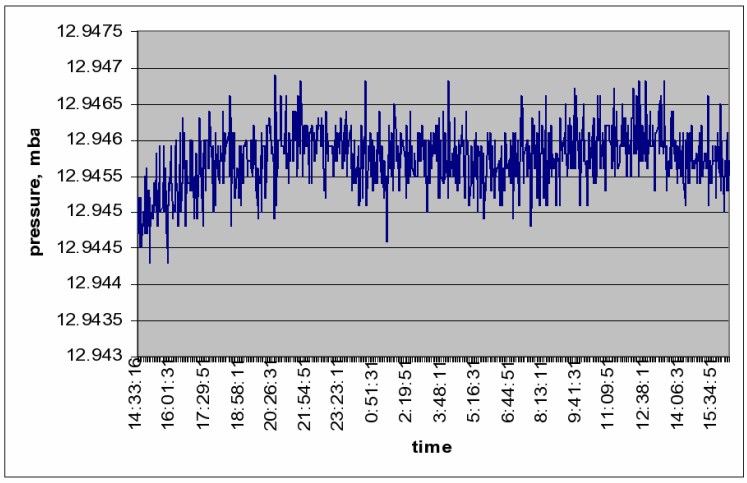

Figure 32: First results of pressure stability obtained with the TILINO test experiment.

scattering processes with the tritium molecules down to thermal energies. Computer simulations show that the total effect for the source potential from WGTS space charges will not exceed 10 mV. Further discussions on effects from WGTS source charging can be found in section 11.4.8.

The technical consequence of the possibility of space charges within the WGTS is a system of monitoring and manipulating devices: This includes the installation of a conducting plate into the magnetic flux tube at the rear side of the WGTS (Rear System), which then defines the electrical potential of the WGTS as well as additional equipment within the 5 beam tubes of the differential pumping section DPS2-F (see section 4.2.1):

- 2 electric dipole electrodes to drift out charged particles by the $\vec{E} \times \vec{B}$ drift (see also sections 8.3.1 and 11.3). At the low plasma density within the DPS2-F differential pumping system, a transverse electric field will not be shielded off substantially by plasma polarization so that a reasonable electrical field strength of about 10 V/cm will be strong enough to drift out particles within a few passages. A semiconductor material will be used as insulating material between the dipole electrodes in order to achieve well-defined potential conditions at the boundaries.

- an axial Penning trap to monitor the charged particle density.

- 2 ring electrodes at positive potential to restrict positive ions from going further downstream and reflect them back towards the drift dipoles and the WGTS. We want to install two ring electrodes, to allow an alternating operation, because any electrode within the solenoidal magnetic field acts as Penning trap for opposite charges. To avoid filling up these traps with charge, thereby shielding off the field of the electrodes, it might be necessary to empty them at regular intervals.

Of course whether the WGTS behaves as the performed computer simulations suggest and whether additional effects like instabilities of the plasma will appear needs to be checked

experimentally, *e.g.* by measurements with gaseous conversion electron emitting ^{83m}Kr added to the tritium gas in the WGTS (see section 9.2.1).

4.1.6 Temperature range of operation

Apart from the standard tritium measurements, specific modes of operation of the WGTS will be required from time to time which need different temperatures and pressures as compared to the standard parameters. Especially the mode with an admixture of gaseous ^{83m}Kr for measuring the electric potential over the volume of the WGTS (see section 9.2.1) requires a WGTS temperature of 100-150 K to avoid the freeze-out of the krypton and a bypass of the palladium filter.

The initial cleaning and out-gassing of all relevant inner surfaces of the tritium source and the electron transport section will rely on the bake out at high temperatures (vacuum pumping at high temperature). For the WGTS a bake out temperature of 500 K will be required, whereas the differential pumping section will be heated up to 400 K and the cryogenic sections to 500 K. This bake out procedure will have to be repeated after some time to remove tritium and argon from the inner surfaces.

4.1.7 WGTS summary

The main advantages of the WGTS can be summed up as follows:

- investigation of the tritium β-spectrum with the highest possible energy resolution, limited only by the spectrum of final state vibrational and rotational excitations of the daughter molecule $(^3\mathrm{HeT})^+$
- use of a maximum specific activity (high signal rate)
- no perturbing solid state effects (the most serious being self-charging of tritium films [97])
- perfect homogeneity over the source cross section

Reasonable solutions have been proposed for the following critical points:

- stability of source strength and column density
- electric potential distribution over the source volume

The WGTS has been ordered end of 2004 and will be delivered in 2007.

4.2 Transport System

The electron transport system adiabatically guides β decay electrons from the tritium sources to the spectrometer, while at the same time eliminating any tritium flow towards the spectrometer, which has to be kept practically free of tritium for background and

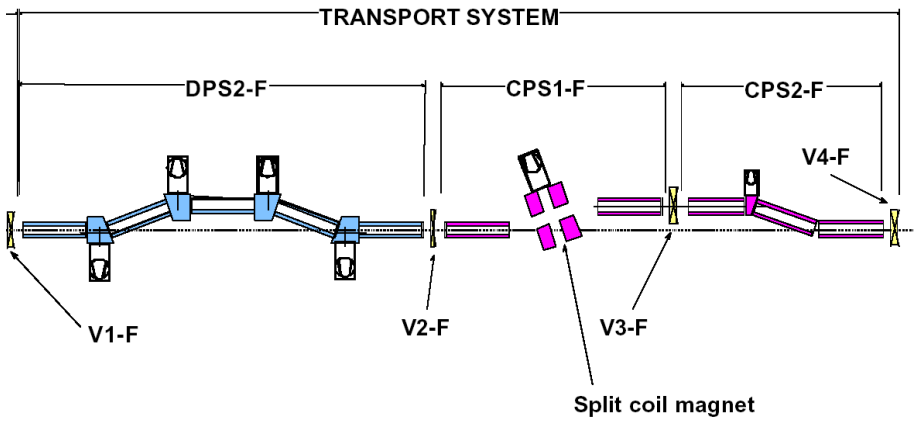

Figure 33: Schematic drawing of the Transport System.

safety reasons. The tritium flow (T_2 and HT molecules) into the pre-spectrometer should be smaller than 10^{-14} mbar l/s to limit the increase of background caused by the decay of tritium molecules in the pre-spectrometer as well as in the main spectrometer to 10^{-3} counts/s (see discussion in section 5 and in 11). Fig. 33 shows the electron transport system schematically. DPS2-F is followed by two cryo-pumping sections (CPS1-F and CPS2-F).

The WGTS is separated by the RT valve V1-F from the second 6.5 m long differential pumping section DPS2-F, which is operated at the higher (*i.e.* LN2) temperature of 77 K. This element has four pump ports. The individual solenoids of this element are tilted by 20 ° to prevent straight trajectories for tritium molecules to travel from one end of the DPS2-F to the other. Each of the 5 solenoids of the DPS2-F delivers a magnetic field of 5.6 T, confining the transported magnetic flux tube to a diameter of 66 mm.

A second RT valve V2-F separates the element DPS2-F from the cryopumping section CPS. The solenoids of the CPS also provide a magnetic field of 5.6 T. The CPS is subdivided into the two cryostats CPS1-F and CPS2-F. Both units are separated from each other by the RT valve V3-F. This design allows the separate cleaning of each element by a flow of warm He gas after a run cycle. The element CPS2-F is thus protected from the back flow of tritium molecules during the cleaning procedure of element CPS1-F, which carries a much higher tritium load than element CPS2-F.

The CPS1-F cryostat also houses a split coil magnet for insertion of quench condensed sources, such as the Quench Condensed (Tritium/Krypton-83) Source (QCTS, QCKrS), which can be inserted into the beam line for study of systematic effects.

4.2.1 DPS2-F

The differential pumping section consists of five 1 m long beam tubes of 75 mm diameter within a superconducting solenoid of 5.6 T field. The magnetic flux of 191 Tcm^2, which needs to be transported from the WGTS to the pre and main spectrometer (see eq. (38)), requires a diameter of 66 mm thus leaving more than 4 mm of radial clearance. The

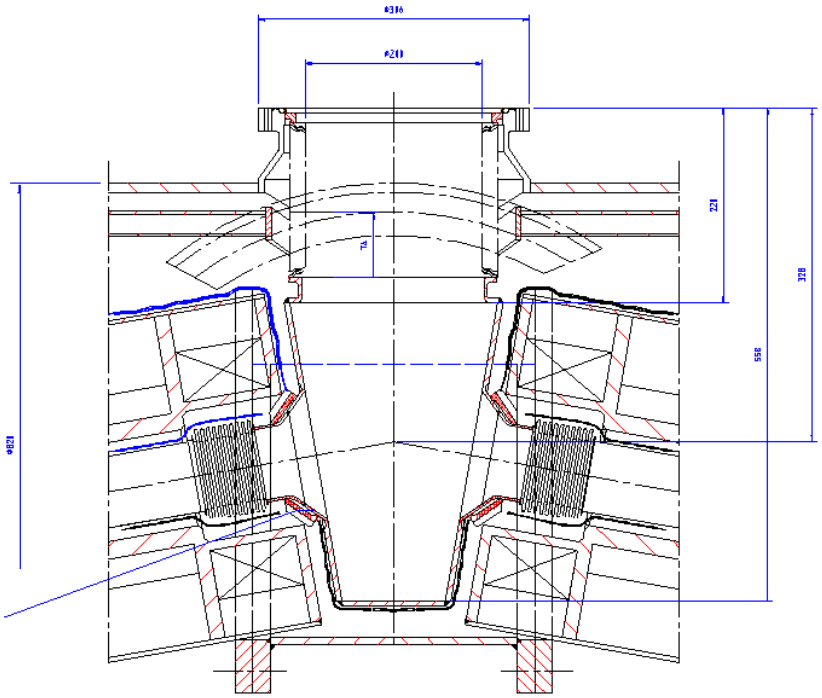

Figure 34: Principle design of a pump port of the differential pumping system.

temperature of the tubes is about 80 K. Within the differential pumping section DPS2-F two beam tubes are intersected by a pump port with a conductance of 2000 l/s. Figure 34 shows the principle design of such a pump port. The differential pumping is achieved by the very different conductance of the beam tube of nominal conductance about 80 l/s with respect to the 15 times higher effective pumping speed at the pumping port. A tilt of 20 degrees of one tube with respect to the other reduces the molecular beaming effect and ensures the differential pumping efficiency. The ports will be equipped with turbomolecular pumps with a pumping speed of $>$ 2000 l/s for H_2 (as specified by one manufacturer). It is estimated that such a pump will have a capture factor for tritium in the range between 0.25 and 0.34.

For a detailed determination of the reduction factor one has to consider the temperature of the gas of 30 K (80 K) and the mass composition of the residual gas (mainly T_2). Fig. 35 shows a detailed Monte Carlo calculation of the T_2 flow rate reduction factor of the differential pumping section DPS2-F as a function of the capture factor of the turbomolecular pump employed. The reduction factor for T_2 is estimated to be of the order of 10^5. With an estimated reduction factor of more than 10^2 for the differential pumping section DPS1-F (see fig.27) the overall reduction factor of DPS1-F and DPS2-F amounts to $> 10^7$.

The differential pumping section DPS2-F is already under manufacture [116]. Fig. 88 shows the layout of the differential pumping section DPS2-F. Inside the beam tube of the section DPS2-F the electrodes with an inner diameter of 75 mm to monitor and to

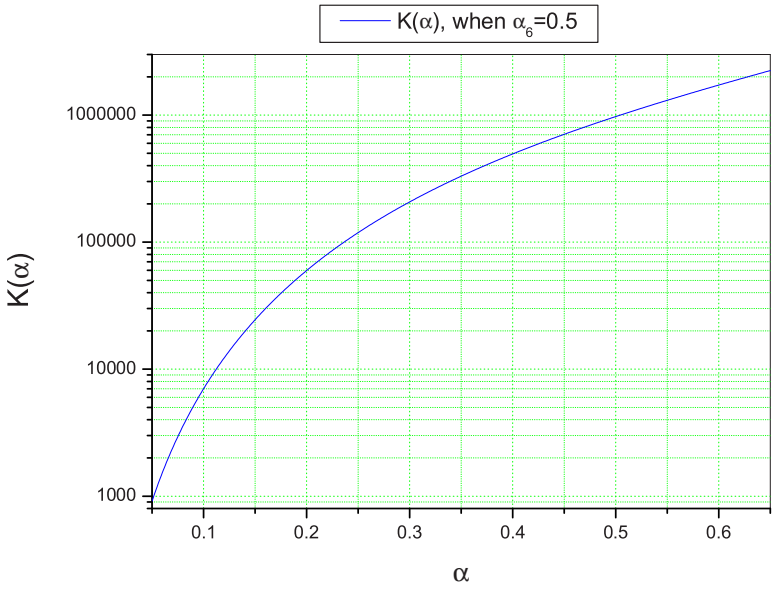

Figure 35: T_2 flow rate reduction factor $K(\alpha)$ for the differential pumping section DPS2-F as a function of the turbomolecular pump capture factor α, assuming a capture coefficient for CPS of $\alpha_6 = 0.5$.

manipulate the charged particles escaping the WGTS downstream will be installed (see sections 4.1.5 and 11.3).

The possibility of much improvement of the reduction factor of the DPS2-F is rather limited because the magnetic field in which the TMPs can operate is limited whereas the magnetic field in the pumping port should be sufficiently high such that adiabatic guiding according to eq. (16) is still guaranteed. For the case of the differential pumping section a good estimator for the adiabaticity violation is the parameter δ^*:

$$\delta^* = \int \frac{|\vec{\nabla}_\perp B(l)|}{B^2(l)}\, dl \tag{41}$$

which is obtained by an integral over the electron trajectory. Fig. 36 shows the expected violation of cyclotron energy in the analysing plane of the main spectrometer $\Delta E_{\perp,\mathrm{A}}$ obtained from microscopic tracking simulations of the differential pumping system for different minimum magnetic fields in the pumping ports B_{min} which led to different parameters δ^*. The simulations show, that for the various designs under consideration the quantity $1/B_{\mathrm{min}}$ can also be used as a good estimator to describe the adiabaticity violation in the differential pumping section instead of δ^*. These simulations show, that the energy violation $\Delta E_{\perp,\mathrm{A}}$ is smaller than 0.01 eV, if the minimum magnetic field B_{min} is larger than 0.3 T [11]. To still have some safety margin the magnetic fields in the pumping ports of the sections DPS1-F and DPS2-F will be kept above 0.4 T by adding to the end

[11] A value for the energy violation of 0.01 eV is 100 times smaller than the width of the transmission function of the KATRIN main spectrometer. If this energy violation would be neglected completely in

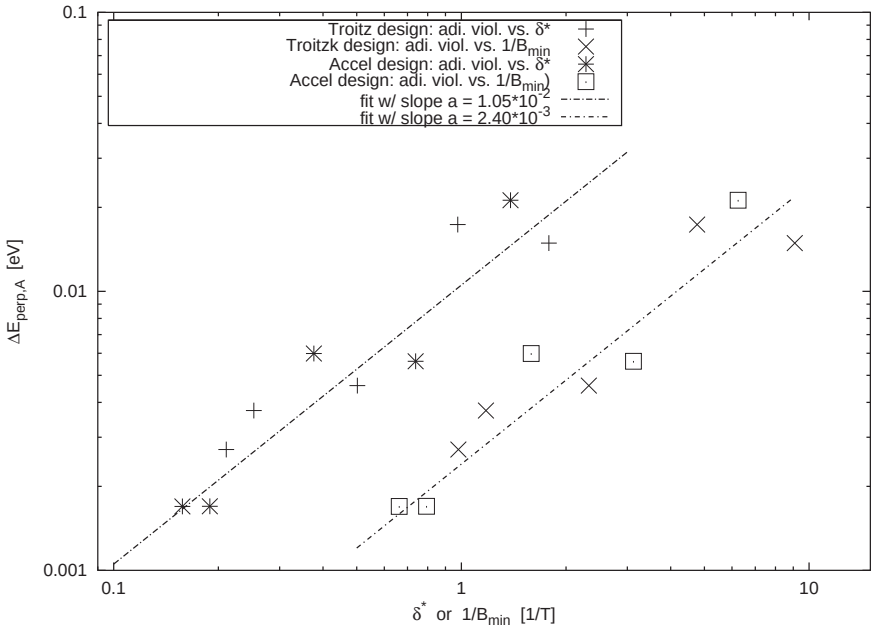

Figure 36: Adiabaticity violation due to low magnetic fields in the pumping ports. Plotted is the violation of the cyclotron energy in the analyzing plane of the main spectrometer $\Delta E_{\perp,\mathrm{A}}$ as function of the adiabaticity violation estimator δ^* (pluses and stars) and as function of the inverse of the minimum magnetic field in the pumping port $1/B_{\mathrm{min}}$ (squares and crosses) for two different designs (labelled as Troitsk and ACCEL design).

of each beam tube short auxiliary coils, which have a larger diameter than the solenoid of the beam tube.

Figure 37 gives an overall picture of the relative reduction in the gas density of T_2 from the center of the WGTS right through to the start of the cryogenic pumping section.

4.2.2 The cryo-pumping sections CPS1-F and CPS2-F

In the next part of the Transport System, the cryo-sorption section, all remaining traces of tritium will be trapped onto the liquid helium cold surface of a transport tube surrounded by transport magnets. It is considered to cover the tube surface with a condensed argon polycrystal layer (snow) or graphite for better trapping. The cryo-pumping section consists of two sections, CPS1-F and CPS2-F. Each section consists of 3 individual transport elements of again 1 m length and 75 mm diameter within superconducting coils with a central magnetic field of 5.6 T. The different tubes are tilted by 20° to each other, thus prohibiting a direct line of sight.

The trap will accumulate less than 10^{17} molecules/day which is negligible in view of

the analysis of the β spectrum it will lead to a shift of the observable m_ν^2 of $\Delta m_\nu^2 \approx -2\Delta E_{\perp,\mathrm{A}}^2$, which is more than one order of magnitude smaller than the other systematic uncertainties of KATRIN (see sec. 11)

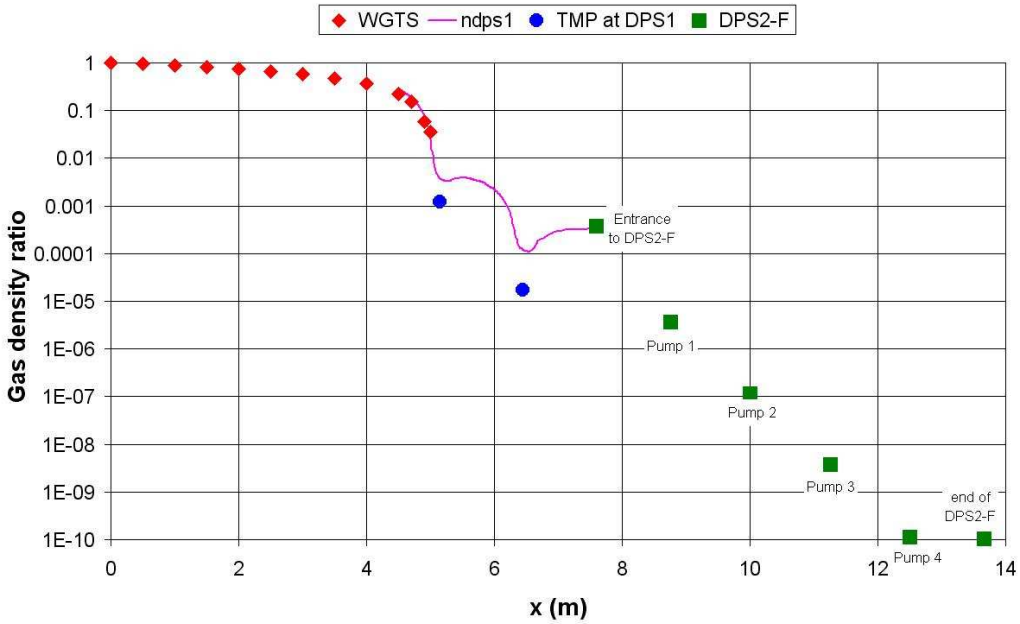

Figure 37: The gas density ratio for tritium (*i.e.* the relative numbers of tritium molecules at any point) along the KATRIN beam tube, between the center of the WGTS ($x = 0$) to the end of DPS2-F. The calculations for the WGTS tube consider the complete gas flow dynamics (diamonds), the calculations for the differential pumping system (solid line) are based on molecular flow approximation.

its huge capacity. The technique of cryo-sorption pumps with a pre-condensed polycrystal layer of gas was developed mostly in 1960 – 1970 [103]. Different studies of this method show that with an appropriate design of the cryo-sorption pump a partial pressure level below 10^{-13} mbar can be achieved for protium and deuterium [104]. Under normal conditions its leakage into the pre-spectrometer should be essentially zero according to the experience of the Troitsk experiment with such cryo-sorption pumps. Safety valves will protect the systems downstream (CPS2-F and pre-spectrometer) in case of failure, *e.g.* a warmup. The reduction factor of the system with/without argon snow at liquid helium temperature can only be estimated. To check these expectations the test experiment TRAP is being performed at the Tritium Laboratory Karlsruhe (see sec. 4.2.3).

In section CPS1-F the middle transport tube is replaced by a split coil magnet with a central magnetic field of 5.6 T. This split coil magnet has access ports for the optional quench-condensed molecular tritium source QCTS as well as for various calibration sources (*e.g.* the condensed krypton conversion electron source QCKrS or the americium cobalt photo-electron source ^{241}Am/Co). The tritium reduction factor downstream of the split-coil is still sufficiently large to allow the evaporation of the complete tritium film of the QCTS in case of a failure of the QCTS cryostat.

Fig. 38 shows the design of the cryo-pumping section CPS. It should be noted that further downstream there is another strong differential pump, the pre spectrometer followed by two transport magnets. With its large pumping capacity and its large volume

it will provide an additional reduction factor for tritium gas.

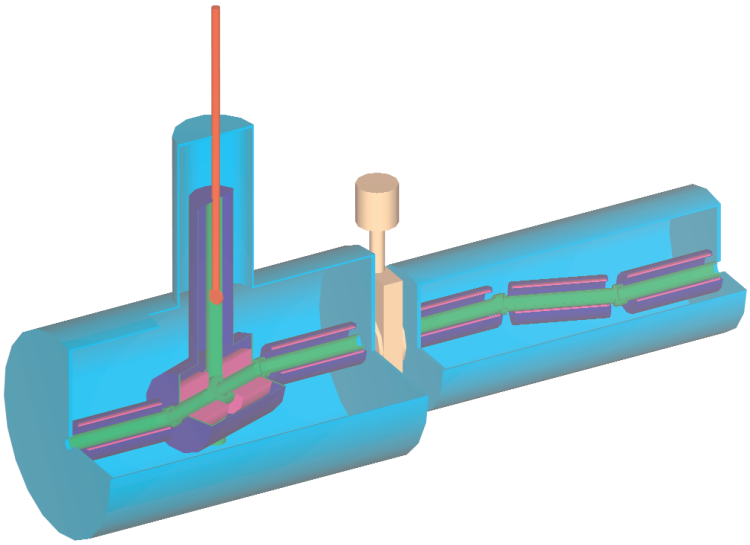

Figure 38: Cryo pumping section consisting of CPS1-F and CPS2-F separated by a valve.

As mentioned above, the tritium flow (T_2 and HT molecules) into the pre-spectrometer should be smaller than 10^{-14} mbar l/s. Since the the overall reduction factor of DPS1-F and DPS2-F amounts to 10^7 to $\approx 5 \cdot 10^7$, the required minimum tritium flow reduction factor of the CPS is in the order of 10^7. Most of the tritium should be pumped with CPS1-F to have a safety margin (*e.g.* in case of a turbomolecular pump failure in DPS1 or DPS2). In the WGTS standard mode about $\approx 2.5 \cdot 10^{19}$ tritium molecules per second (0.9 mbar l/s (STP)) are entering DPS1-F. The cryo-pumping section will therefore accumulate of the order of 10^{17} tritium molecules/day ($\approx 0.4 \cdot 10^{-7}$ mbar l/s) which is negligible in view of its huge capacity.

4.2.3 The TRAP experiment

To demonstrate the required tritium reduction factor for the CPS the test experiment TRAP (TRitium Argon frost Pump) is being performed at the Tritium Laboratory Karlsruhe. The objectives of this experiment are

1. to find an upper limit for a possible migration of tritium through the KATRIN cryo-pump section (see also fig. 39). Such an upper limit corresponds to a lower limit of the tritium reduction factor. Parameters to be checked are *e.g.* the tritium feed flow rate and the amount of sorption material adsorbed on the inner wall of the tube.

2. to test different sorption materials, *e.g.* argon (= reference material) and neon.

3. to test the decontamination procedure for the cryo-pump.

4. to get additional knowledge needed for the specification of the CPS (*e.g.* dimensions of capillaries to inject argon).

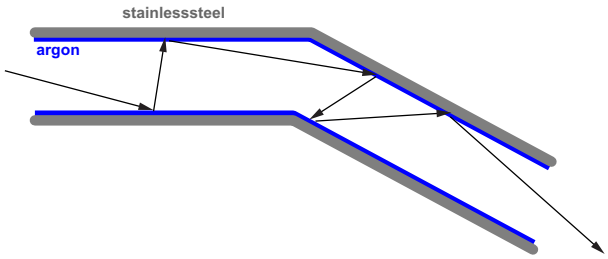

Figure 39: Migration: tritium molecules may pass the whole CPS by multiple sorption - desorption processes. This effect is probably enhanced due to the β-decay, leading to energy deposits inside the argon layer.

A special problem in pumping radioactive gases such as tritium is the possibility of migration due to radioactive decay. It may be caused by the recoiling daughter sputtering some molecules in the vicinity from the bulk or by energy loss of the decay electron somewhere in the system. We thus differentiate between the capture factor for a non radioactive gas such as H_2 and D_2 and the factor for T_2 which will be a bit smaller due to migration.

TRAP can be regarded as a small model of the CPS, which has the following properties:

1. It corresponds to the CPS with regard to the cryogenic properties such as operating temperatures and tube conductance.

2. TRAP has no superconducting magnet system (in contrast to the CPS), since the electron transport through the cryo system is not subject of the investigation.

3. The TRAP experiment is designed such that essential functions of the CPS can be investigated:

   - Preparation of argon frost as sorption material for hydrogen isotopes.
   - Feeding the cryo-pump with KATRIN relevant tritium flow rates
   - Cryo-pumping of tritium and afterwards decontamination of the cryo-pump by warming up of the tube and purging with helium gas.
   - Baking the system at 500 K.

Fig. 40 shows the scheme of the TRAP setup and fig. 41 a photo. A tube with a 20° bend is placed in a liquid helium bath. Diameter and length of the tube have been chosen in a way that the conductance of the tube is similar to the conductance of one section of CPS. The preparation of the argon layer is done by means of a thin capillary with small holes along almost the complete length, which is placed inside the tube. Hydrogen isotopes (in the beginning deuterium, later tritium) from a buffer vessel with calibrated

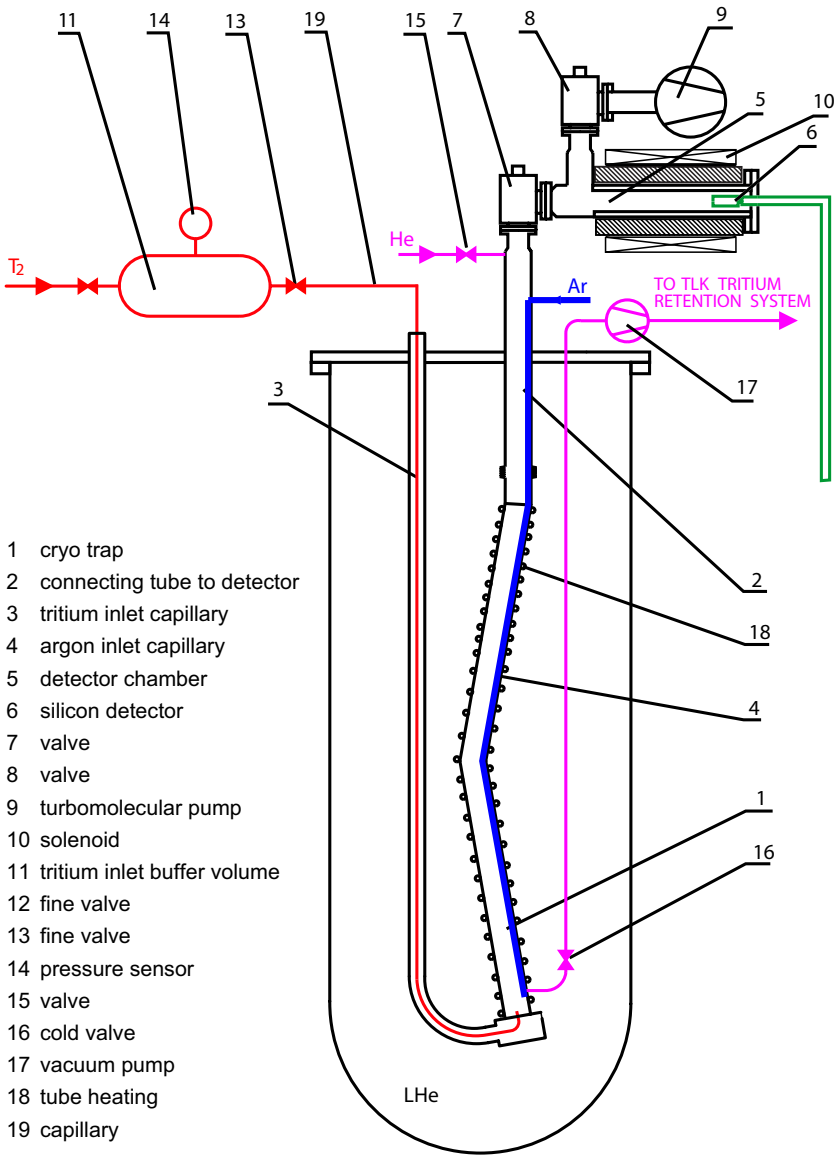

Figure 40: Scheme of the TRAP experiment. Further explanation see text.

volume are fed into the tube through a combination of a valve (for regulation) and a thin capillary. This combination enables the adjustment of a well-defined and KATRIN relevant flow rate. The detection system for deuterium and tritium is placed at the outlet of the cryo-pump (above the cryostat). In case of deuterium the partial pressure is measured with a quadrupole mass spectrometer. The transmitted tritium is measured by the detection of the β decay rate at a silicon detector.

TRAP has been successfully commissioned in summer 2004, a first set of test measurements with deuterium has been performed afterwards. Figure 42 shows the influence of different deuterium feed flow rates on the deuterium partial pressure at the outlet of

Figure 41: TRAP experiment. Shown are the cryostat with inserted tube (left) and the gas handling system (right).

the cryo-pump. During this measurement the thickness of the argon layer was about 2000 monolayers. The deuterium feed flow rates have been increased stepwise from 2.6 to $5.3 \cdot 10^{-7}$ mbar l/s (first group of data points) to about $1 \cdot 10^{-5}$ mbar l/s (last group of data points). The first group represents the KATRIN flow rates in standard mode ($0.4 \cdot 10^{-7}$ mbar l/s, see above). The size of the error bars is dominated by the uncertainty of the calibration of the buffer volume. Figure 42 shows, that within the detection limit of the quadrupole mass spectrometer no influence of the feed flow rate on the corresponding deuterium partial pressure is visible. The deuterium partial pressure was - with one exception - all the time below 10^{-12} mbar. The only increase above this value is due to a pressure increase inside the cryostat, leading to a slightly higher temperature of the pump.

Taking the deuterium partial pressure of 10^{-12} mbar, the maximum deuterium feed flow rate of about 10^{-5} mbar l/s and the conductance of the tube between cryo-pump and mass filter (about 20 l/s) one gets a lower limit of the deuterium retention factor of $5 \cdot 10^5$ for the 1–bend TRAP cryo-pump. This value is only a rough estimate, a more detailed

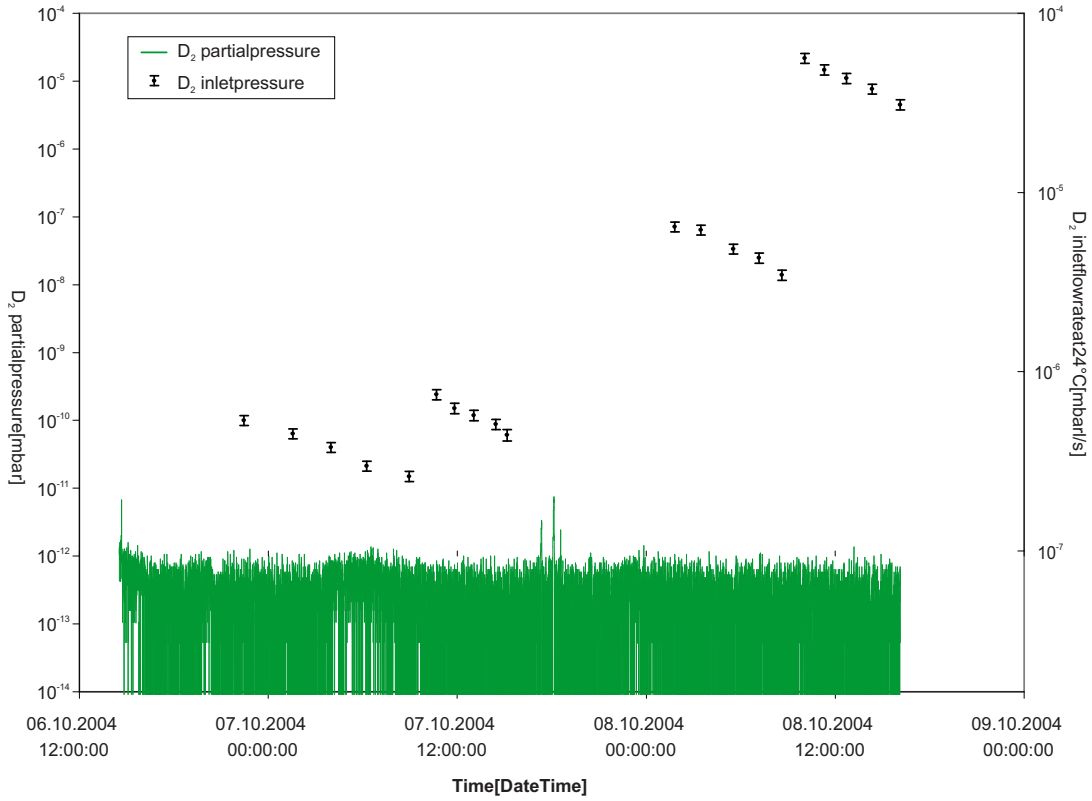

Figure 42: Cryo-sorption of deuterium. Shown is the deuterium feed flow rate (data points with error bars, right axis) and the corresponding D_2 partial pressure at the outlet of the cryo-pump (green line, left axis) with time. The partial pressure was measured at room temperature, the slight increase at 18:04 was due to a pressure increase inside the liquid helium cryostat. The reference feed flow rate for KATRIN is 10^{-7} mbar l/s (at 273 K). Further discussion see text.

investigation needs to be done.

4.2.4 Transport System summary

The electron transport system has to adiabatically guide β decay electrons from the tritium sources to the spectrometer, while at the same time eliminating any tritium flow towards the spectrometer, which has to be kept practically free of tritium for background and safety reasons. It consists of a differential pumping section (DPS2-F) and two cryogenic pumping sections (CPS1-F & CPS2-F)

- **DPS2-F:**
  This section has been specified and ordered in 2003 and will be delivered end of 2005. Monte Carlo simulations predict a tritium flow reduction factor of about 10^5, an experimental verification will be done after delivery.

- **CPS1-F & CPS2-F:**
  Most of the physical and technical numbers are fixed (magnetic field, transported flux, temperature etc.), specification and tender action will be done in 2005. Delivery is expected for 2007.

- **Test experiment TRAP:**
  TRAP has been successfully commissioned in summer 2004, a first set of test measurements with deuterium resulted in a lower limit of the deuterium reduction factor of $5 \cdot 10^5$. The necessary modifications of TRAP to adapt tritium handling requirements and first measurements with tritium will be done in 2005.

The required flow rate reduction factor of about 10^{14} combined with an estimated reduction factor of 100 to 500 for the DPS1-F leads to a reduction factor of about 10^{12} which has to be realized by the Transport System. The KATRIN collaboration is quite confident to reach this value with the above discussed system.

4.3 The Rear System

As already mentioned above a so-called rear system will be located at the rear side of the WGTS (see fig. 43). This rear system consists of the differential pumping section DPS2-R and the Calibration and Monitoring System CMS. The DPS2-R will have 2 pumping ports and has to protect the CMS from a to high partial tritium pressure. Tritium gas pumped through these ports will be sent into the Outer Loop of KATRIN (fig. 20. DPS2-R is separated to the WGTS by a room temperature gate valve (V1-R) and to the CMS by the gate valve V2-R.

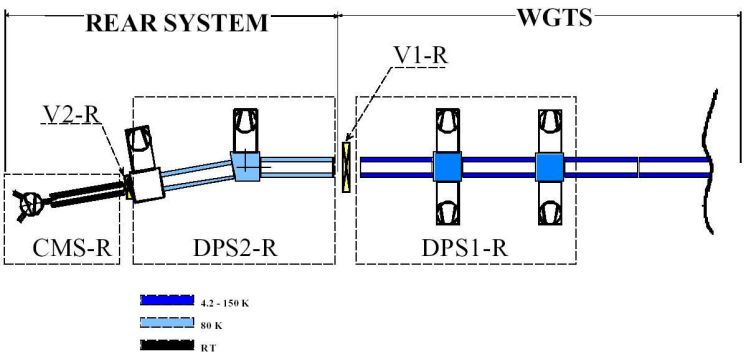

Figure 43: Connection of the calibration and monitoring system CMS to the WGTS.

The CMS will be located at the rear end of KATRIN. It comprises essentially electron guns and electron detectors for monitoring of the flux of the β-electrons and of calibration sources (see fig. 44). Secondly it houses the rear plate to ensure the well-defined and constant electrical potential over the full WGTS (see sec. 4.1.5). Additionally, calibration sources like one of the ^{83m}Kr conversion electron sources or the ^{241}Am/Co photoelectron

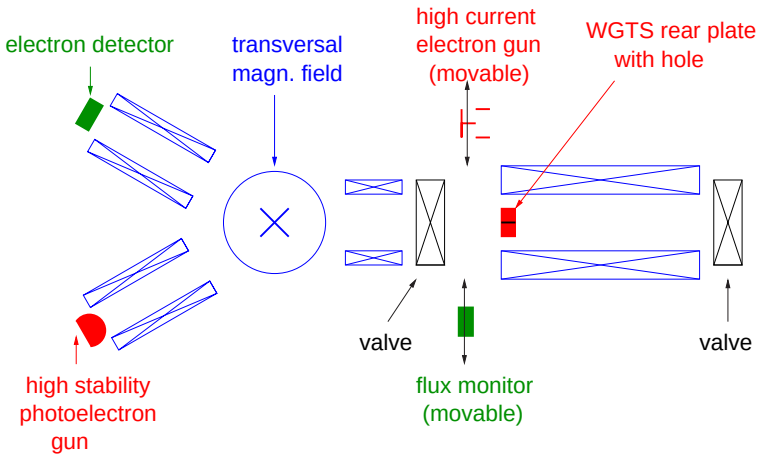

Figure 44: Schematic view of the essential components of the Rear System (for details see text). Optionally calibration sources could be installed, which are not shown here.

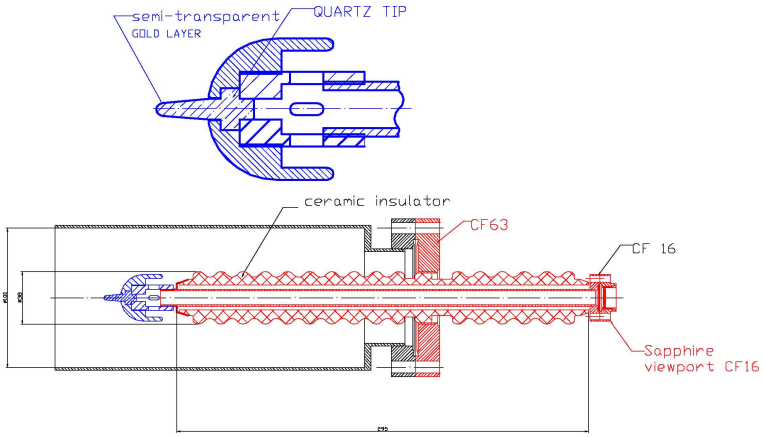

Figure 45: Photoelectron gun for tests of the KATRIN pre-spectrometer.

source (see section 9) could be installed as well. The CMS has to fulfill the following requirements:

1. To investigate the transmission function of the main spectrometer, and to determine the energy dependence of the inelastic scattering cross section, we need an energetically very sharp electron gun ($\Delta E \approx 0.2\,\mathrm{eV}$) with well defined angular emittance. As a prototype of such a source a photoelectron gun already realized at the Troitsk experiment can be used in which ultra-violet light from a deuterium lamp is focussed on a small photocathode at high voltage. The shape of the photocathode and the surrounding metal electrodes as well as the strength of the magnetic field at the source with respect to the pinch magnetic field define the angular emittance of the source. Such a photoelectron source will be used for the pre spectrometer tests (see fig. 45). This electron gun could be fixed in transversal position. The point-like electron flux will be scanned through the cross section of the WGTS and the main

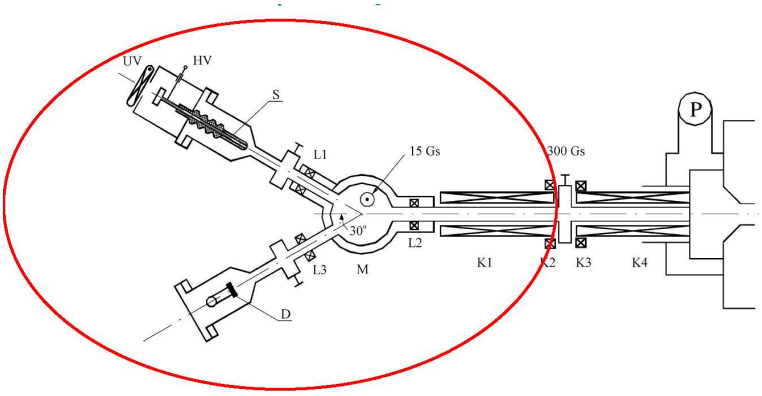

Figure 46: Schematic view of the Troitsk electron gun system composed of a electron gun, a transversal field dipole magnet which bends the electron (counterclockwise) into the beam line direction. Back coming electrons will get the same direction change and will be led to an electron detector. This transversal field magnet is connected to the axial fields of the solenoids by a non-adiabatic transition.

spectrometer with the help of the 2 scanning magnets in the DPS1 sections (see above). This electron gun is used in regular intervals to monitor the column sensitivity of the WGTS as well. The required electron intensity is $10^4 - 10^6$ electrons per second. The necessary short-term stability (on the scale of a few minutes) needs to be better than 10^{-3}. The Troitsk group has made the experience that the electron flux coming back from the spectrometer to the photoelectron source deteriorates the stability of the photoelectron intensity. Therefore a non-adiabatic transition to a transversal magnetic field region was installed to differentiate between the electron flux from the photoelectron gun and the back coming electron flux (see fig. 46). A similar system is considered for the KATRIN experiment.

2. A rear plate of well-defined electrical potential will be the effective rear end of the WGTS [12]. It will define the electrical potential of the source, since the electrical potential along a magnetic field line is short-cut by the high axial charge conductivity (see sections 4.1.5 and 11). It might be, that this rear plate needs to be illuminated by light to create additional photoelectrons. A little hole within the rear plate allows the electrons from the electron guns to be led through. The rear plate may also be a thin foil stopping only low energy electrons but allowing the high energy fraction to pass onto a faraday cup used as monitor detector. Note that in such a foil electrons close to the endpoint will loose a few keV only.

3. A monitor detector to determine the β-electron flux from the WGTS could be brought into the beam line. This detector will be either installed behind the hole of the rear plate or instead of the rear plate. The high energy part of the electron spectrum is a good measure of the WGTS β activity.

4. A high-current source with electron intensities of $10^{10} - 10^{13}$ electrons per second

[12] In terms of the experimental components, the WGTS ends at the pumping port DPS1-R (see fig. 20).

could be used to investigate the filtering properties of the pre spectrometer. In an ideal case a high electron current being reflected in the pre spectrometer does not lead to an increase of the background.

5. Optionally one or more of the conversion electron or photoelectron calibration sources should be connected at the position of the CMS to the KATRIN setup.

4.4 Monitoring of source parameters

Very important for the determination of the neutrino mass from the tritium β spectrum measured by the KATRIN experiment is the continuous precise knowledge of the β activity and column density of the WGTS –or even better a high stability of these parameters. The knowledge and stability of the β activity of the source is needed to combine data taken at different retarding voltage settings, whereas the knowledge of the column density is required to apply the energy loss correction. Several means will be used to obtain a high stability of the WGTS:

1. The tritium inlet pressure is stabilized and monitored with the help of the pressure controlled tritium buffer vessel (see fig. 29).

2. The temperature of the WGTS is stabilized to $\pm$ 0.03 K. The high stability can be achieved by circulating liquid neon gas in a small pipe thermally coupled to the tritium tube or by using a LNe bath. If the temperature of the WGTS is stable, the critical column density ρd of the WGTS can be controlled and stabilized by the pressure of the pressure controlled tritium buffer vessel alone.

3. A fraction of the circulating gas is sent to the TLK infrastructure (Outer Loop) for detritiation, tritium removal and isotope separation and is replaced by T_2 gas of highest purity to keep the tritium concentration ε_T constant.

The following dedicated measurements will be used to monitor the strength and column density of the WGTS (see also section 11.4.3):

1. The stability of the β activity of the tritium sources will be monitored by the rate of the β-electrons. The total β activity of the WGTS of 10^{11} decays per second corresponds to a β-electron current of about 16 nA, which could be easily measured with the help of a Faraday cup. But this method can not be applied in our case. In section 11.4.8 it is explained that at the rear end an electron emitting plate is needed to compensate the WGTS charging-up. Therefore, in our case the net current to this plate corresponds only to the rate of β-electrons, which are able to pass the electrostatic potential barrier of the main spectrometer. This rate depends strongly on the setting of the retarding voltage of the main spectrometer and is so small that it can not be measured in a Faraday cup mode.

   One solution for a permanent monitoring would be to measure the rate of high energy β-electrons through a hole in the rear plate with a semiconductor detector.

Another possibility is a very thin foil used at the rear end with an energy loss of a few keV for electrons in the upper part of the β spectrum. Thus, a considerable fraction of decay electrons could pass the foil and be detected with a Faraday cup just behind it. To compensate the loss at the end of the source an additional current of photoelectrons has to be generated from the foil. The rate of electrons passing the hole in the rear foil and the Faraday cup can be measured by a semiconductor detector.

In addition to this permanent monitoring we will use a different way to determine the β activity of the tritium source, which already has been successfully applied to the Troitsk experiment. Repeatedly we will lower the retarding voltage of the main spectrometer significantly to measure a larger fraction of the β spectrum yielding a high rate at the detector. A possible scenario would be to apply every 1000 s a setting of the retarding voltage of $qU = E_0 - 500\,\mathrm{eV}$, *i.e.* $U = 18075\,\mathrm{V}$. The signal rate at the detector will then be about $\dot{N}_{18.075} = 100\,\mathrm{kHz}$. Therefore within less than 1 minute $N_{18.075}$ will reach about 10^6 and thus the β activity can be determined with a relative precision of 1 per mill. If the fluctuations of the β activity have smaller time scales shorter than 1000 s this scheme has to be changed accordingly. In principle by using lower retarding voltages any rate can be achieved, which the detector is able to take.

2. The isotopic content and thus the purity ε_T of the WGTS gas mixture will be monitored quantitatively in real time by Laser Raman Spectroscopy in a small optical cell near the pressure controlled tritium buffer vessel (see fig. 47). Laser Raman spectroscopy is particularly advantageous for monitoring mixtures of hydrogen isotopomers containing tritium, deuterium and protium because of its excellent differentiation among species (see fig. 48). This technique has been developed and successfully tested at TLK [102] (see e.g. fig.49) and should allow to monitor the gas composition in the WGTS in-situ during the long-term measurements. By this method ε_T can be determined to a 0.2 % level.

   In parallel the high resolution mass spectrometry method will be applied in one of the differential pumping ports to measure the isotopic content of the circulated gas. These two methods (Laser Raman and mass spectrometry) applied at the entrance and the outlet of the WGTS will provide us with a reliable and accurate knowledge about the tritium purity.

3. The source column density is monitored by sending monoenergetic electrons from an electron gun with an energy above the tritium β endpoint E_0 through the WGTS and measuring their rate at different retarding potentials (see sect. 11). By comparing the rates at different retarding potentials the column density ρd as fraction of the free column density $(\rho d)_\mathrm{free}$ can be determined with 2 per mill precision. For this measurement the electron gun intensity does not need to be known absolutely, it only needs to be stable over a few minutes measurement time needed for the different retarding voltage settings of this monitoring measurement.

These three measurements of $N_{18.075}$, ε_{T} and ρd determine the two necessary parameters β activity and column density ρd redundantly since they are related by

$$N_{18.075} \propto \varepsilon_{\mathrm{T}} \cdot \rho d \tag{42}$$

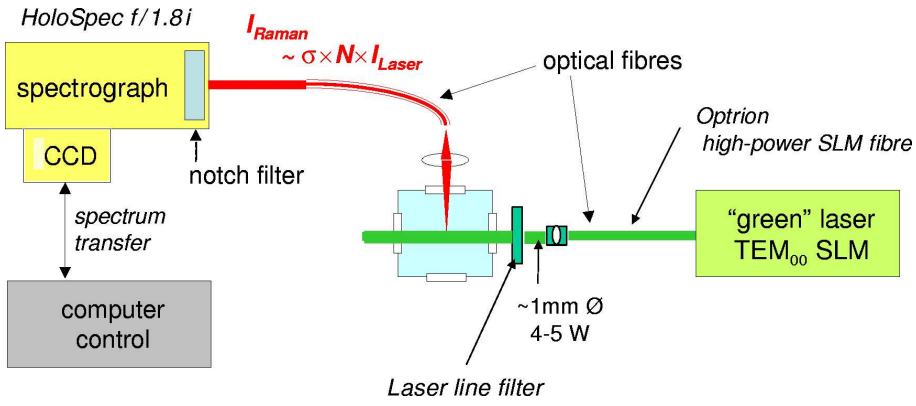

Figure 47: Schematic view of the Laser Raman Spectroscopy setup. The optical cell containing the tritium gas at a pressure of about 10 mbar is connected to the high-power NdYag laser by an optical fibre. The Raman scattered light is coupled by another fibre to an holographic spectrograph.

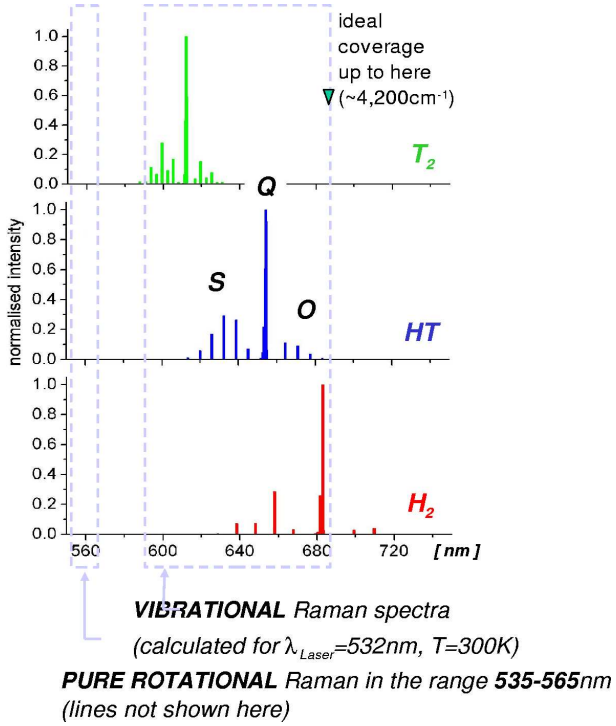

Figure 48: Schematic picture of the Raman vibration spectra for the three molecules T_2, HT and H_2.

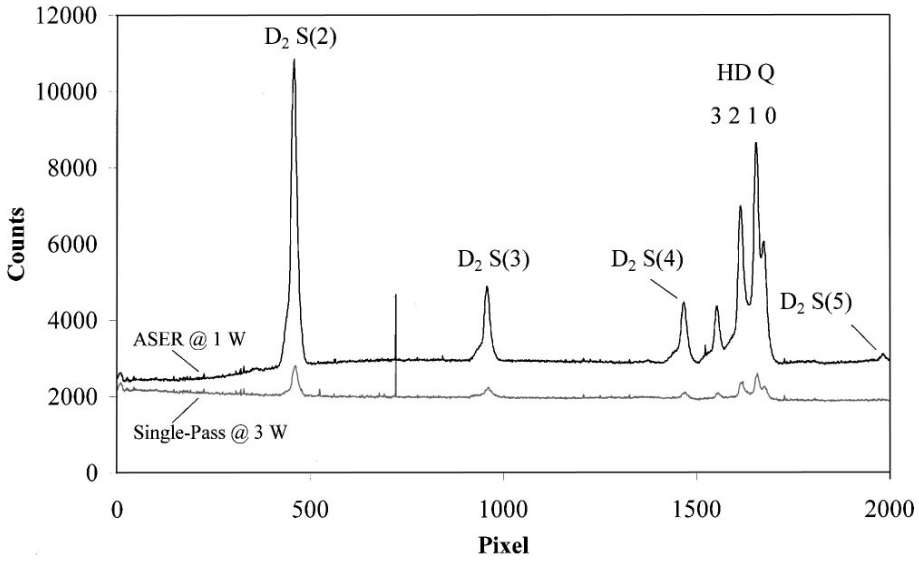

Figure 49: Experimental vibrational Raman spectra of a HD - D_2 mixture [102].

4.5 Quench condensed tritium source

As discussed above we are considering the option of installing a molecular quench condensed tritium source (QCTS) in addition to the WGTS for important self-consistent checks with a tritium source, which has independent and complementary systematics. The design of such a source –a thin film of molecular tritium quench condensed onto a graphite substrate at temperatures of $T \approx 1.6 - 1.8\,\mathrm{K}$– will follow largely the source concept of the Mainz experiment [87].

The QCTS consists mainly of the graphite substrate with the T_2 film mounted on a cold head of a continuous flow cryostat with 1.6 K base temperature. This low temperature is required to suppress surface diffusion and film roughening [105, 106]. The whole setup is rather compact and does not emit tritium gas in standard operation. Therefore it does not need strong differential pumping stations on both sides. On the other hand condensation of residual gas onto the T_2 film has to be prohibited. Therefore the ideal location of this source is in the center of a split-coil magnet within the differential cryo-pumping section. Still having most of the cryo-pumps on the way to the spectrometer avoids tritium contamination of the spectrometer, also in the case of a failure of the QCTS with evaporation the whole T_2 film in very short time. For technical reasons –in contrast to the Mainz experiment– we aim for a vertical cryostat of a total height of about $4-5\,\mathrm{m}$.

Fig. 50 shows the proposed setup of the QCTS inside the split coil magnet within the cryo-pumping beam line. The superconducting split coil will provide a magnetic field at the QCTS, which is significantly higher than the magnetic field of the WGTS:

$$B_{\mathrm{s,QCTS}} = 5.3\ \mathrm{T} \quad \Rightarrow \quad \begin{cases} \theta_{\max} = 70^\circ \\ A_{\mathrm{s,QCTS}} = 36\ \mathrm{cm}^2 \end{cases} \tag{43}$$

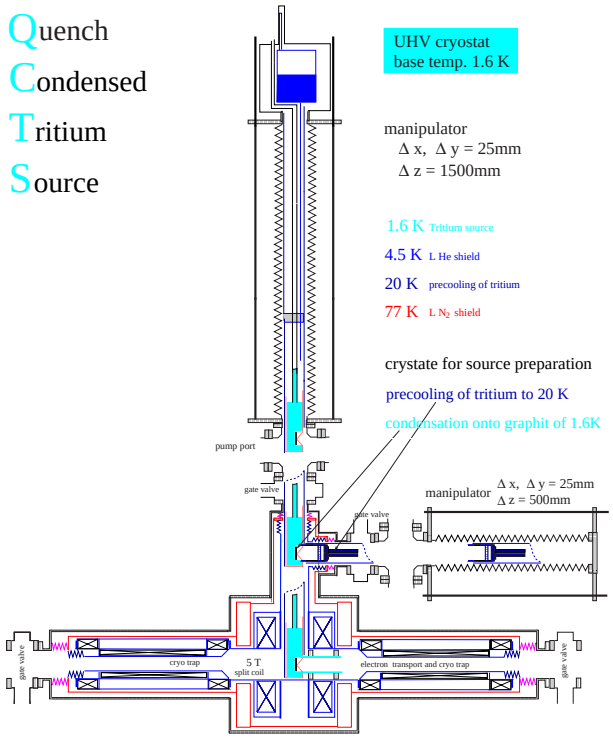

Figure 50: Schematic view of the quench condensed tritium source (QCTS) at 3 different positions for measurement, for preparation, analysis and desorption of the T_2 film and for service (from bottom to top).

The graphite substrate diameter of 70 mm provides a reserve of 2 mm in each radial direction. The higher magnetic field $B_{\mathrm{s,QCTS}} = 5.3\,\mathrm{T}$ is required to reach a higher signal rate for smaller film thicknesses (see fig. 15), since the self-charging effect of the QCTS forbids the use of thick T_2 films. We aim for a QCTS film thickness of about 200 monolayers (1 monolayer is equivalent to a film thickness of about 3.4 Å and corresponds to a column density of $\rho d \approx 0.9 \cdot 10^{15}\,\mathrm{molecules/cm^2}$) which would still provide a signal rate which is about 2/3 of the signal rate of the WGTS (see fig. 15) and would increase the signal intensity by a factor 66 compared to the previous experiments at Mainz and Troitsk. Since the self-charging effect will result for this thick films in a potential difference of 4 V between top and bottom film layer the use of such a condensed tritium source makes only sense if the self-charging of the T_2 film can be significantly reduced (see sec. 4.5.2).

The split coil magnet could also serve as the insertion position for electron sources such as a condensed $^{83\mathrm{m}}$Kr conversion electron source or the ^{241}Am/Co photo electron source (see sec. 9) for energy calibration purposes.

4.5.1 Tritium film preparation

To allow the preparation and analysis of the T_2 film as well as to perform the measurement with the QCTS a large bellows will be used to put the cold head of the QCTS at three

different positions (see fig. 50):

- the *measuring position* inside a 5.3 T split coil magnet of 200 mm diameter and LHe cold bore.
- the *source preparation area* facing a second cryostat. This area is equipped with a facility to pre-cool the tritium to 20 K before quench condensing. It contains a laser system for surface treatment (surface cleaning will be done in-situ by laser ablation) as well as standard diagnostics instruments, such as laser ellipsometer for film thickness measurements.
- the *service position* for maintenance of the QCTS at room temperature.

One of the latter two positions will also be used when performing a standard tritium measurement with the WGTS.

The T_2 film will be prepared in the following way:

1. The cold head and substrate are baked out at a temperature of about 450 K in the *service position* and then moved to the *source preparation area*.
2. With the help of a high-power ablation laser the substrate will be cleaned further at the *source preparation position*. Pre-cooled tritium gas from a capillary will first hit a reflecting plate to randomize the directions of the T_2 molecules and then quench-condense onto the graphite substrate. The growing T_2 film thickness will be monitored by a spatially resolving laser ellipsometer. The use of additional methods to analyse the T_2 film is considered.
3. The cold head with the fresh T_2 film is transported into the *measuring position*.

This whole technique and film preparing procedure will be developed and tested for the condensed ^{83m}Kr source CKrS.

4.5.2 Properties of quench-condensed tritium films

At Mainz the properties and systematics of quench condensed tritium films have been studied intensively with the help of the solid state physics group at University Konstanz [97, 105, 106, 94, 107]. Although a quench condensed tritium film has the advantage of short term stability of source strength and source column density [13] compared to a windowless gaseous tritium source, major additional sources of systematic uncertainties contribute when using the QCTS instead of the WGTS:

[13] The effective lifetime of the QCTS will be determined by the rate of tritium losses caused by $^3HeT^+$ recoil. This has been found to be of the order of 0.16 monolayer per day at Mainz[99]. The evaporating tritium is of no concern for KATRIN as it is trapped by the LHe cold cryotraps between the QCTS and the pre spectrometer.

1. A quench condensed T_2 film is not in thermal equilibrium. Temperature activated surface diffusion will lead to a roughening of the tritium film and to the build up of small micro-crystals [105, 106]. This effect will increase the inelastic scattering in an unreproducable manner. However, intensive studies [106] have shown that this effect follows an Arrhenius law with an activation energy of $\Delta E_{\mathrm{act}} = 45 \cdot \mathrm{k_B\,K}$. Therefore, this surface diffusion process can be slowed down to time scales of decades when using T_2 film temperatures below 2 K. Therefore, the roughening transition of the QCTS is not of concern for KATRIN.

2. Concerning the electric potential, the situation of the QCTS is somehow similar to that of the WGTS: While the β-electrons are leaving the source the remaining positive ions will stay and build up a positive charge within the QCTS [97, 107]. This positive charge will build up until the electrical field –produced by this positive charge– is sufficiently high to help the charge carriers to overcome the potential barriers inside the insulating T_2 film. This equilibrium state corresponds to an electric field of 63 MV/m which corresponds to an electrical potential difference of about 20 mV per monolayer. A charge carrier hopping model describes the time dependent and equilibrium phase of the T_2 film in good agreement with the measurements [107].

   This equilibrium electric field means that there will be an electric potential difference as big as 4 V between top and bottom of a 200 monolayer thick T_2 film broadening significantly the energy resolution with which the tritium β spectrum could be investigated with the KATRIN experiment. Of course, this potential difference could be decreased by using a much smaller thickness of the T_2 film of the QCTS, but then the source strength will be reduced too much (see fig. 15) to allow a sensitive check of the results obtained with the WGTS. When using a 200 monolayer thick T_2 film, the statistical uncertainty of m_ν^2 for a 3 years measurement will be a factor 2.1 larger than for the WGTS[14]. If the self-charging effect was not present a quench-condensed tritium source can be made as thick as a gaseous tritium source and would allow similar sensitivity.

   Therefore the use of the QCTS will need the self-charging effect to be decreased significantly. Currently a group at University Konstanz/Germany [101] is investigating possibilities to decrease this effect by offering the quench condensed T_2 film additional low energy electrons from a photo cathode and whether the properties and the conductance of the quench condensed films could be changed by the quench condensation parameters as gas temperature and gas flow.

3. The electronic final states of solid molecular tritium show some differences compared to the gaseous molecular tritium [95]. In addition to an energy shift of -0.88 eV due to the polarizability of quench condensed tritium there are small changes in

[14] The factor 2.1 is based on a preliminary simulation of the QCTS without the optimization of the measuring interval as applied for the WGTS (see sec. 11.6). Such an optimization depends on the source strength, the background rate and the energy resolution.

the electronic wave functions by electrostatic repulsion and Pauli blocking due to the presence of the neighboring molecules [108], as well as the probability to excite neighboring molecules during the β decay. The former is only relevant when analysing considerably larger energy intervals below the endpoint than the excitation energy of the first excited electronic level. The latter has been calculated [95] and experimentally checked by the Mainz experiment [96].

4. The inelastic scattering of electrons within a solid hydrogen film is somewhat different to the inelastic scattering with hydrogen molecules in the gas phase, again due to the change of the electronic final states. The difference of the total and the energy differential cross section has been measured and shown to be in good agreement with quantum-chemical calculations [94].

4.5.3 QCTS summary

The KATRIN collaboration is considering to build a quench-condensed tritium source QCTS as an additional option to the windowless gaseous tritium source WGTS, if the self-charging effect can be significantly reduced or avoided totally. The QCTS measurement would provide results with *independent* and *complementary* systematic uncertainties. A comparison of the results from both sources would help to investigate systematic effects in the KATRIN measurements. Of course the additional systematic effects described above clearly show that the WGTS will be the standard tritium source of the KATRIN experiment. On the other hand the good understanding of these systematics obtained by the Mainz experiment will allow the use of the QCTS as an important tool to check the results obtained with the WGTS with only slightly lower sensitivity as long as the self-charging effect can be reduced significantly. Since the split coil magnet system –possibly housing the QCTS– will be needed and built for calibration purpose the option to build a QCTS remains open.

4.6 The Tritium Laboratory Karlsruhe (TLK)

The Tritium Laboratory Karlsruhe (TLK) is a scientific-engineering facility which was founded with the aim to establish tritium handling techniques in Europe and to develop the technologies for the fuel cycle of future fusion reactors [109, 110]. Laboratory scale and technical scale experiments are conducted at TLK within the framework of the European Fusion Technology Program with realistic concentration of tritium, tritiated gases and tritiated materials which are typical for future fusion devices (*e.g.* ITER).

The TLK was commissioned with tritium in 1994 and since then has continuously improved its infrastructure and has expanded its experimental activities. With a fully closed tritium cycle and the license to handle 40 g of tritium TLK is an almost unique experimental facility. The TLK is therefore the only facility which can provide the infrastructure and the experience necessary for KATRIN.

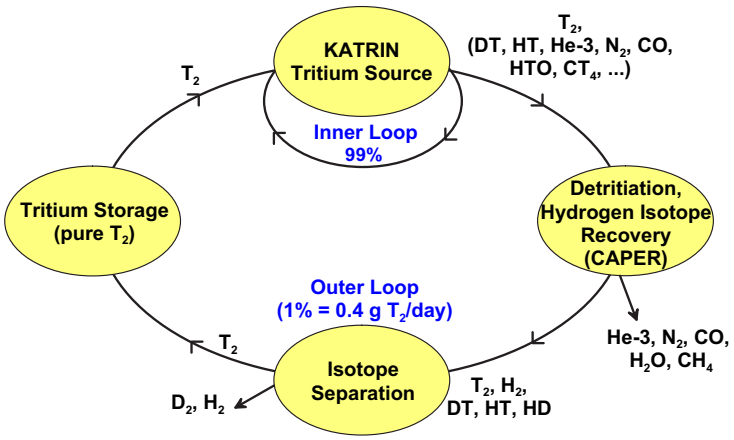

Figure 51: KATRIN and the closed tritium cycle at TLK (simplified scheme)

4.6.1 Embedding of KATRIN in the closed tritium cycle of TLK

All tritium related parts of KATRIN will be installed inside the TLK and will have a lot of interfaces to the TLK infrastructure (see fig. 20). Fig. 51 shows in a simplified scheme how KATRIN will be incorporated in the closed tritium cycle of TLK [111]. The above mentioned Outer Loop of KATRIN is almost completely covered by the existing closed tritium cycle: Tritium supplied from external sources is introduced into the cycle and either stored in the Tritium Storage system or immediately used for experiments (e.g. KATRIN). During these experiments the tritium invariably becomes diluted with other hydrogen isotopes or is even converted to tritiated water vapor or tritiated hydrocarbons such as methane. The experimental facility CAPER is operated to demonstrate the technology for the ITER Tokamak Exhaust Processing system and simultaneously - as part of TLK infrastructure - recovers hydrogen isotopes from all primary off-gases from TLK. The pure hydrogen isotopes are transferred to the Isotope Separation System (ISS) of TLK, while the remaining gases are detritiated, typically down to ppm levels of tritium concentration, and then sent to the central tritium retention system. The tritium loop of TLK is closed with the return of the pure tritium from the ISS to the Tritium Storage System. During standard operation of KATRIN about 0.4 g tritium per day will be sent into the Outer Loop for impurity processing and isotopic enrichment.

Fig. 52 shows the planned setup of KATRIN with regard to the TLK. Rear system and WGTS will be installed in the existing TLK building (left side, grey), the Transport System (DPS2-F, CPS1-F and CPS2-F) will be installed in a new building (red section up to wall) which will become part of TLK.

4.6.2 Equipment of TLK

The laboratory is equipped with ample infrastructure facilities, like a Tritium Storage System, a Tritium Transfer Station, an Isotope Separation System, diverse analytical equip-

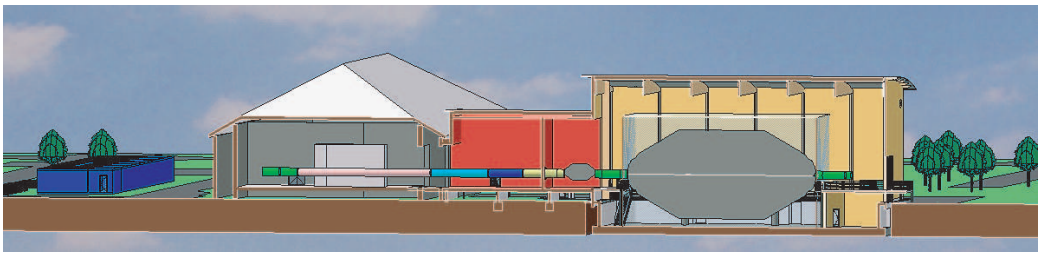

Figure 52: View on the existing TLK building (grey) and the planned transport hall (red) and spectrometer hall.

ment and one Central and numerous Local Tritium Retention Systems [109, 110, 112]. In addition, cooling water circuits, a ventilation system and all the necessary power supply systems including a non-interruptible power supply are available. As a key experiment, the plasma exhaust clean-up facility CAPER has been developed for the recovery of unburned deuterium and tritium from fusion reactor exhaust gases and other sources. CAPER is also used to process tritiated gas streams arising in TLK from infrastructure facilities or other experiments.

1. **Building**
   The Tritium Laboratory (building 452) has a total area of 841 m^2 for experiments and 615 m^2 for infrastructure. Currently a total glove box volume of 125 m^3 is available. The pressure inside the building is kept below ambient pressure by a ventilation system. The atmosphere of the air inside the building is operated in a once-through mode (60000 m^3/h) and has no tritium retention. Tritium is handled in primary systems which are enclosed in secondary systems (glove boxes).

2. **Tritium Retention System**
   The TLK confinement concept is illustrated in Fig.53 It has two major features, namely a two barrier design as well as an atmosphere and gaseous waste treatment. The two barriers preventing tritium release are the primary system (comprising all the tritium containing components and pipework of the experiment or process) and the secondary system (also called secondary containment) being in most cases glove boxes.

   Primary exhaust gases and a small volume flow from the secondary containments are passed through the Central Tritium Retention System (ZTS). Its main tasks are

   - to maintain negative pressure in the glove boxes
   - to detritiate the glove box gases before release via stack
   - to pre-process highly contaminated gases from primary systems before discharge.

Additionally each glove box is equipped with a Local Tritium Retention System for the removal of tritium permeated or released from primary systems into the glove box. The tritium retention in both systems are based on catalytic oxidation followed by adsorption of water vapor on molecular sieve beds.

3. **Tritium Transfer System (TTS)**
   All major systems of the laboratory and experiments are linked together via the TTS to allow transfers of hydrogen isotopes from one system to another and to verify by pVT-c measurements[15] the amount of tritium transferred. Imported tritium is generally received by the Tritium Transfer System.

4. **Tritium Storage System (TLG)**
   Most of the tritium on site is stored in metal getter beds in the TLG. The Tritium Storage System consists of nine uranium and one ZrCo getter beds for the storage of mainly pure tritium. The total allowed tritium storage capacity is 30 g.

5. **Isotope Separation System (ISS)**
   The Tritium Laboratory is equipped with a technical-sized gas chromatographic Isotope Separation System with a column containing α-alumina oxide coated with

[15] (separate measurements of the presseure p and temperature T in a calibrated volume V and the tritium concentration c)

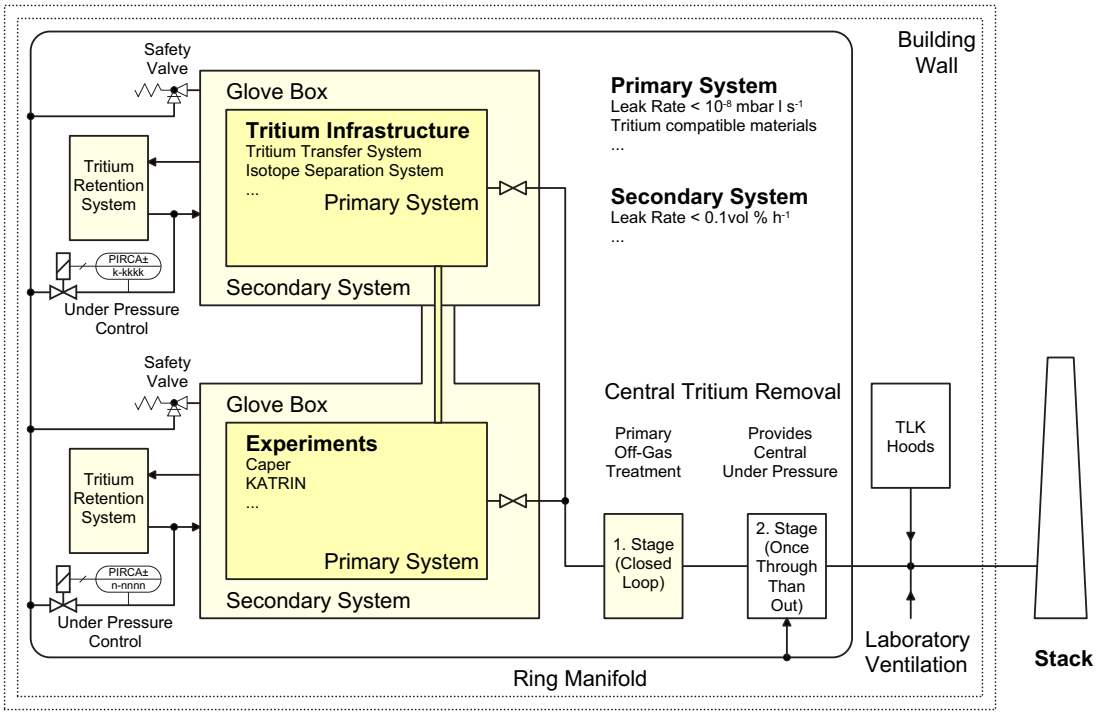

Figure 53: The TLK confinement concept

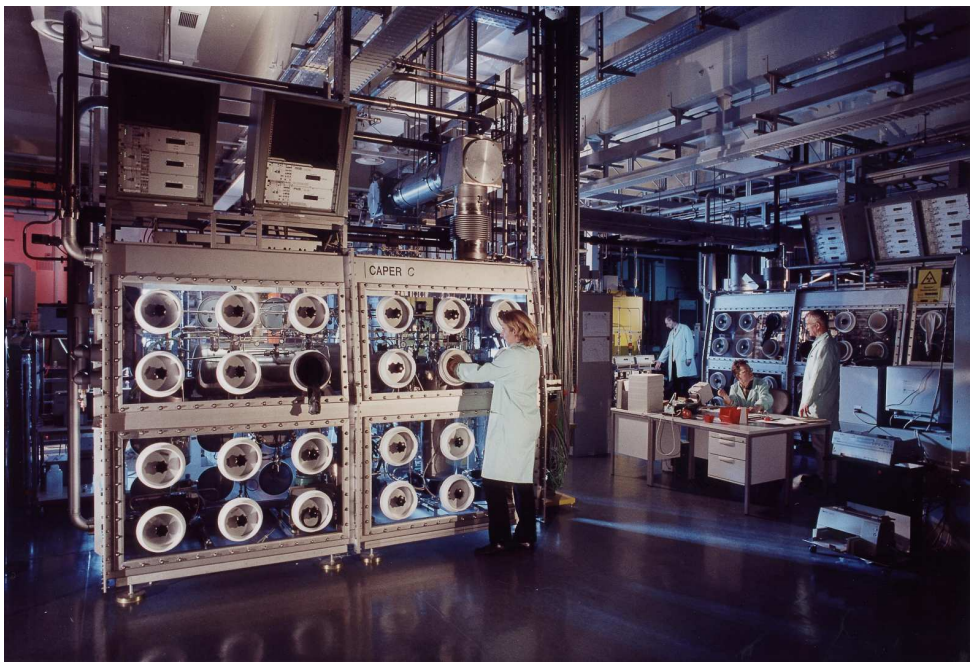

Figure 54: CAPER facility of TLK. The glove box at the front is a standard box with two modules and contains the third step of the CAPER process. First and second step are located in the 4-module glove box at the right. The tritium inventory during experimental runs is typically 3 g $\approx$ 30 kCi.

20% by weight palladium. The column operates at room temperature and atmospheric pressure on the so-called displacement gas chromatography principle [113]. The Isotope Separation System accepts clean hydrogen isotopes and separates the tritium from other hydrogen isotopes. A tritium purity up to 99% can be achieved.

6. **Tritium Analytics**
   To measure the six different hydrogen isotopomers (H_2, HD, HT, D_2, DT and T_2) and other gaseous components typically for future fusion reactors (such as CO, CO_2, He) the TLK is equipped with quite a lot of different analytical tools. Several radio gas chromatographs equipped with thermal conductivity detectors and small ionization chambers are on site. Additionally, an Omegatron mass spectrometer, quadrupole mass spectrometers and liquid scintillation counters are used for the qualitative and quantitative analysis of tritiated gases. A Laser Raman system is being setup, too.

7. **Detritiation and hydrogen isotope recovery (experiment CAPER)**
   The experimental facility CAPER is an essential system for the R&D programme at TLK and is also part of the closed tritium cycle of TLK. All primary gaseous wastes arising from the different experiments at TLK (with tritium concentrations up to $2.5 \cdot 10^6$ Ci/m^3) are detritiated with CAPER and afterwards the pure hydrogen mixtures are sent to the ISS. The detritated impurity stream with less than 1 Ci/m^3

is sent into the Central Tritium Retention System for additional processing before exhausted to stack. More details can be found in [114, 115].

8. **Tritium Accountancy**
   Calorimetry is a technique for measuring the thermal power of heat producing samples, in the case of tritium it can measure the decay heat and hence the tritium activity of a sample. In addition to pVT-c measurements, the Tritium Laboratory Karlsruhe uses calorimetry for the accountancy of tritium.

9. **Process Control System**
   The systems of the tritium infrastructure (Tritium Transfer System, Tritium Storage System, Isotope Separation System, Central Tritium Retention System, Local Tritium Retention Systems and Analytical System) are controlled by a process control system 'Siemens TELEPERM M'. The system consists of components for automation and display of process parameters. Control and coordination of measurements, PID controllers, actuators and sequential controls is performed with the system. Safety interlocks are implemented in the system where it supervises processes for deviations from normal operation (over-temperature, over-pressure etc.) and in such cases it automatically trips the system into a well-defined safety state to avoid an overload or damage of the plant equipment. Experiments are operated by local Process Control Systems. Additionally there is a hard wired Safety Circuit independent from the various Process Control Systems.

5 Electrostatic Spectrometers

The electrostatic filter system of KATRIN consists of two spectrometers of MAC-E-Filter type [16]: the *pre-spectrometer*, working at a fixed retarding energy of approximately 300 eV below the endpoint of the β-spectrum allows only electrons with the highest energies to pass into the *main spectrometer*, where the kinetic energy of those electrons is analyzed with a resolution of 0.93 eV. The main task of the pre-spectrometer is to limit the number of electrons, which might scatter on residual gas molecules in the main spectrometer, thereby increasing the rate of background events. The high resolution of the main spectrometer requires large dimensions. It has a diameter of 10 m and a length of 23.3 m. Both spectrometers are connected by two 1 m long superconducting transport magnets (see fig. 55).

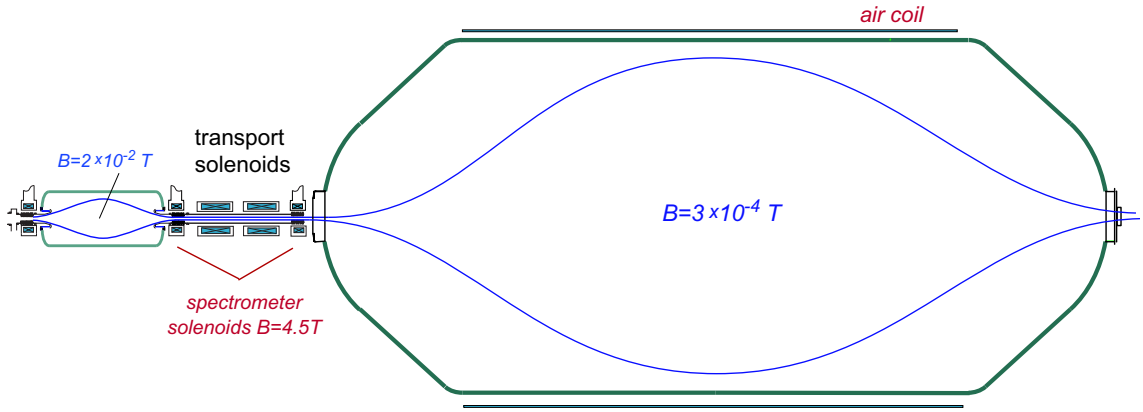

Figure 55: The electrostatic filter system of KATRIN consists of a pre-spectrometer to reject all low energy electrons, followed by a large main spectrometer with an energy resolution of 0.93 eV . Both spectrometers are connected by two superconducting transport solenoids.

5.1 Pre-spectrometer

The pre-spectrometer (see fig. 56) plays a major role both in the R&D phase and during tritium measurements. It fulfills several tasks:

- It serves as a prototype for the main spectrometer, verifying the XHV[17] concept, testing the reliable operation of the heating/cooling system and investigating the performance and properties of the new electro-magnetic design.
- During normal tritium measurements it will be operated as a pre-filter, reducing the incoming flux of β-electrons from $10^{10} s^{-1}$ to $10^{4} s^{-1}$ in the main spectrometer.

[16] The third MAC-E-Filter, the monitor spectrometer, will be discussed in sect. 9.
[17] XHV = eXtreme High Vacuum

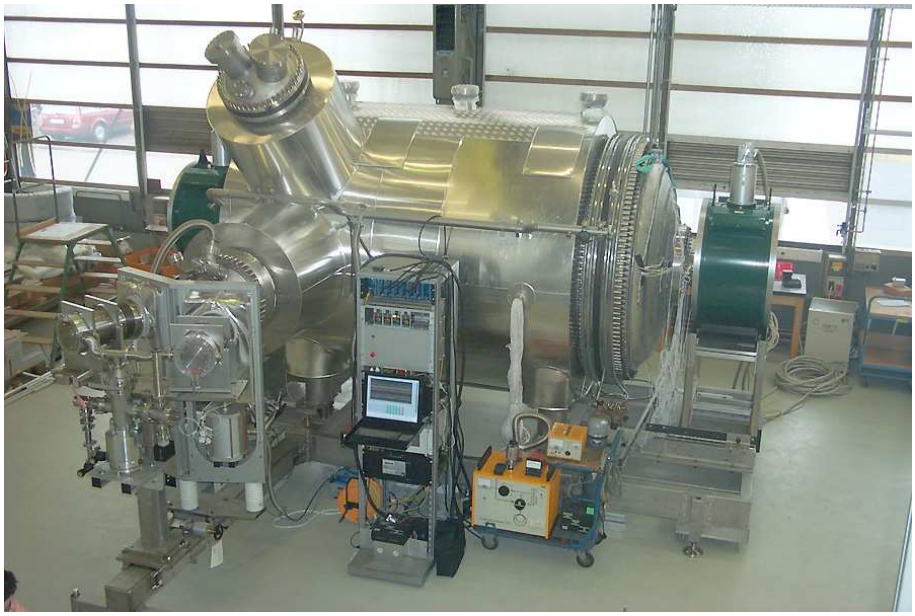

Figure 56: Photograph of the KATRIN pre-spectrometer vessel with tubes of the heating-cooling system, thermal insulation, vacuum system and two superconducting magnets.

- In a second application, the pre-spectrometer will act as a fast switch for running the main spectrometer in the non-integrating MAC-E-TOF mode (see section 2.1.1).

As a pre-filter the requirements on the energy resolution of the pre-spectrometer are rather moderate. A resolution of $\Delta E \approx 100\,\mathrm{eV}$ is sufficient to reduce the flux of β-electrons by a factor of 10^6. This reduction factor can be achieved by fixing the retarding energy of the pre-spectrometer at about 300 eV below the tritium endpoint energy E_0 and let only the last 200 eV of the β spectrum pass undistorted into the main spectrometer. The lower flux minimizes the chances of causing background by ionization of residual gas molecules and building up a trapped plasma in the spectrometers. The moderate energy resolution of the pre-spectrometer considerably relaxes the demands with regard to homogeneity of the magnetic field and electrostatic potential in the central plane of the vessel. The flux tube in the analyzing plane will have a diameter of about 1 m.

The conservation of the magnetic flux (eq. 38) and the definition of the energy resolution of a MAC-E-Filter (eq. 17) require a magnetic field of about 300 Gauss in the analyzing plane of the pre-spectrometer and a diameter of the flux tube of about 1 m. As a result of detailed electromagnetic calculations and technical considerations, fig. 58 shows the basic configuration of the pre-spectrometer, its shape, the inner electrodes and the superconducting solenoids of the spectrometer magnets[18]. Figure 59 shows some electric equipotential lines of the pre-spectrometer. The transmission function for electrons is

[18]The design and the stray fields of the two superconducting magnets does not only take into account the necessary magnetic field in the analyzing plane of the pre-spectrometer but also the fact, that upstream and downstream of the pre-spectrometer the full magnetic flux has to be transported through 200 mm wide UHV-valves, which are necessary in order to be able to separate the different subsystems.

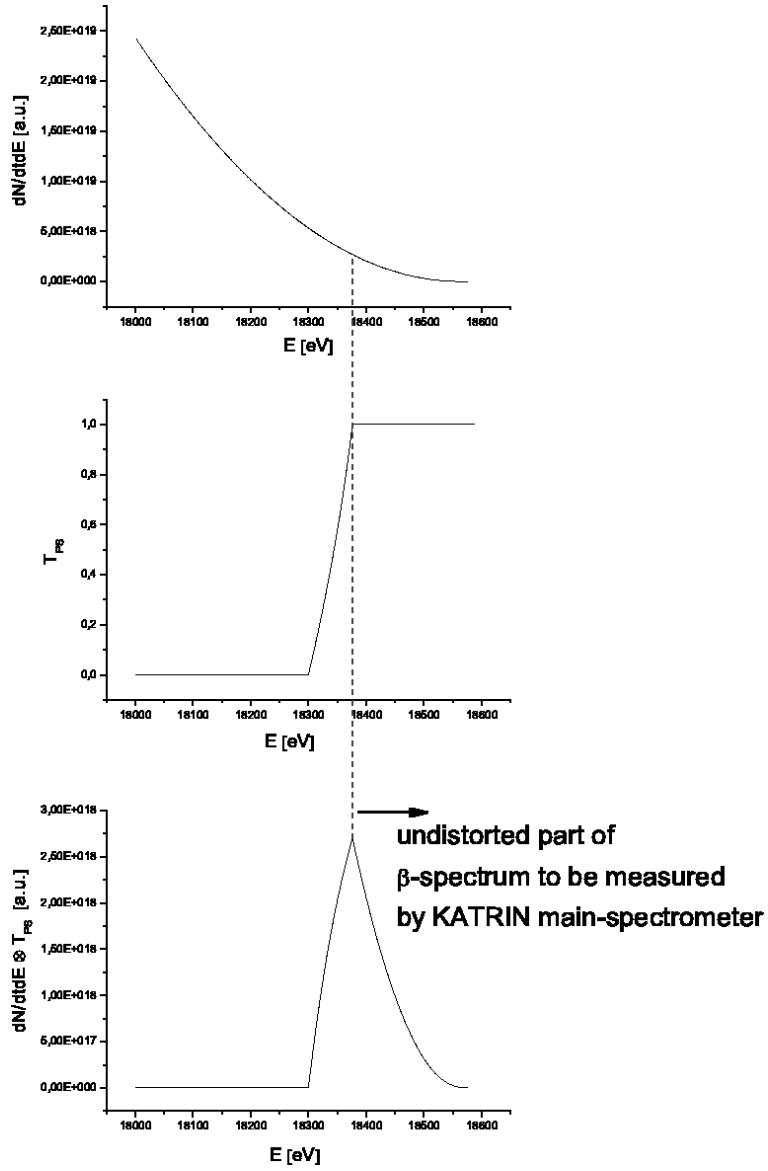

Figure 57: Principle of the pre-spectrometer filter. Top: schematic plot of the β spectrum quadratically rising from the endpoint at about 18575 eV. Centre: transmission function of the pre-spectrometer with a hypothetical energy resolution of $\Delta E = 75$ eV. Bottom: The multiplication of the β spectrum (top) with the transmission function (centre) shows that the last 200 eV of the β spectrum are admitted into the main spectrometer without spectral distortion.

shown in fig. 60, clearly fulfilling the requirement of an energy resolution of $\Delta E \leq 100$ eV. Slight differences in the average transmission energy are due to radial inhomogeneities of the magnetic fields and the electric potentials in the analyzing plane.

The pre-spectrometer (see fig. 61) is a cylindrical tank with a length of 3.38 m and an inner diameter of 1.68 m. The dimensions are comparable to the existing MAC-E-Filters at Mainz and Troitsk. The walls of the vessel are made of 10 mm thick 1.4429

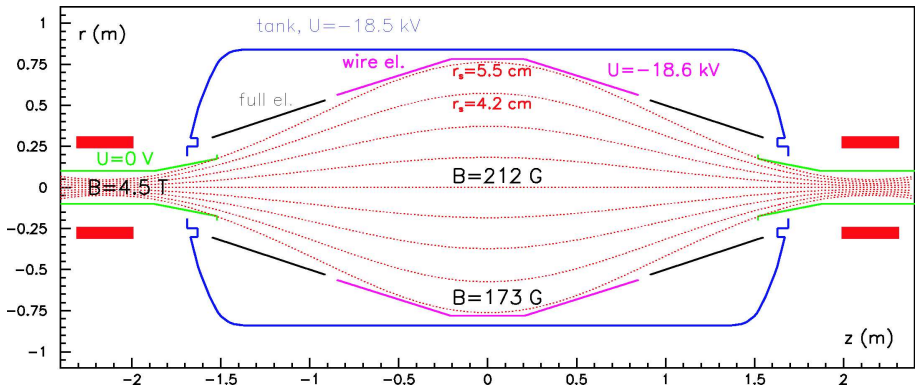

Figure 58: Electromagnetic design of the KATRIN pre-spectrometer: two large warm-bore superconducting magnets at a distance of 4.30 m create a stray field of about 200 Gauss in the analyzing plane. Taking into account stray fields from all other superconducting magnets of the KATRIN setup, the magnetic field in the analyzing plane amounts to about 270 Gauss on the axis. Magnetic field lines are shown in red. The hull of the pre-spectrometer vessel is held at high voltage and houses an inner electrode system with a cylindrical and two conically shaped sections. The inner electrode is on a slightly more negative potential in order to reject low energy electrons from the wall (see sect. 8). The central part of the inner electrode consists of a nearly massless wire frame. The end cones are made of sheet metal in order to avoid strong electric fields at this section where the inner electrodes are close to the ground-electrodes.

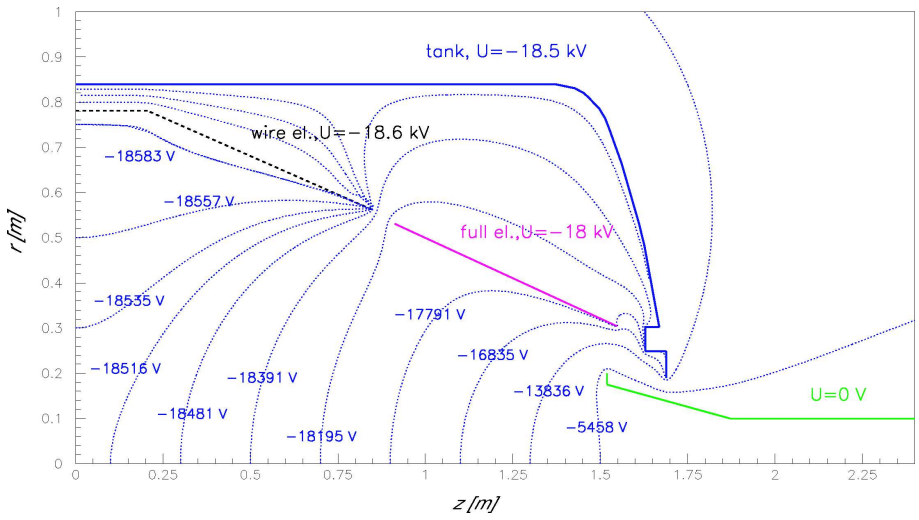

Figure 59: Electric equipotential lines of the pre-spectrometer.

(316 LN) stainless steel. At one end of the cylindrical section a large DN 1680 flange with a differentially pumped all-metal double sealing allows installation of the inner electrode system. At the other end two cylindrical pump ports with a diameter of 50 cm and a length of 1 m are attached to the vessel 75 cm off centre, leaving enough space for an optional air coil. One cylinder, pointing upwards at an angle of 45°, is equipped with vacuum gauges, while the other, horizontal cylinder has two VAT DN 200 CF all-metal valves and

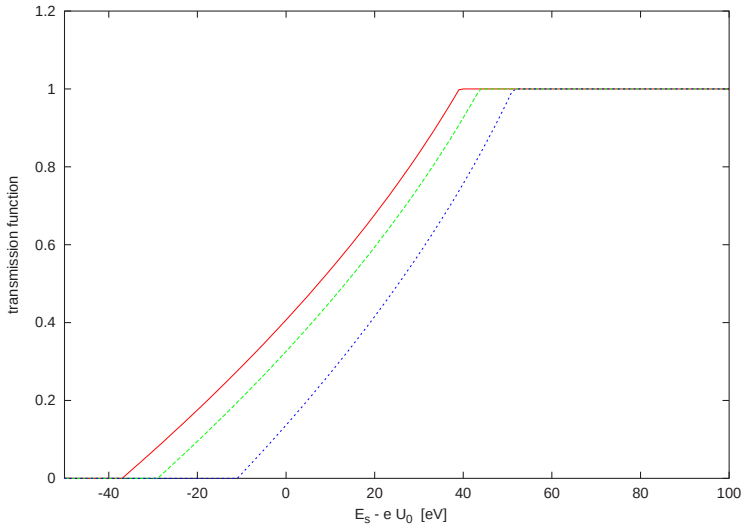

Figure 60: Transmission function of the KATRIN pre-spectrometer for different distances of the trajectory from the central axis (radii r at the centre of the solenoid: r=0 mm (red), r=21 mm (green), r=42 mm (blue, corresponding to full magnetic flux)).

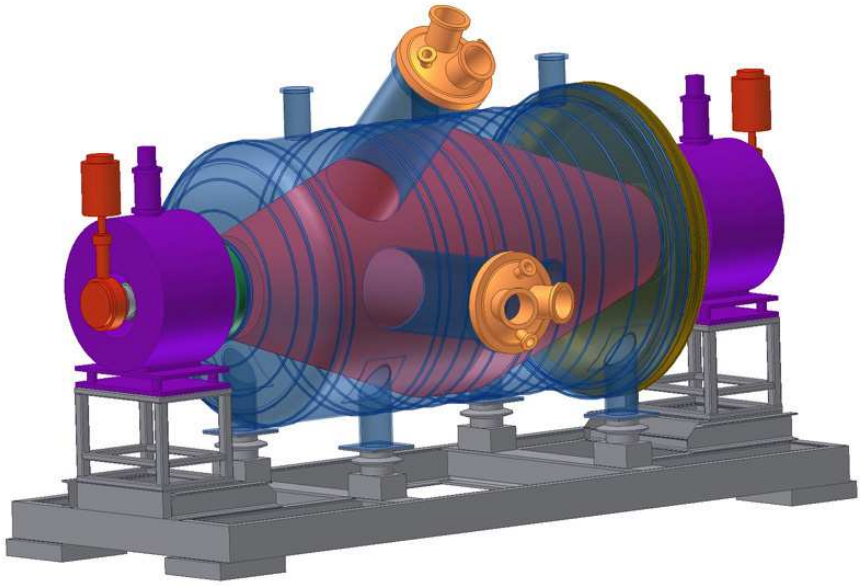

Figure 61: The KATRIN pre-spectrometer is a 3.38 m long stainless steel vessel (316LN) with an outer diameter of 1.70 m. It is equipped with a system of inner electrodes for fine tuning the shape of the electrostatic field and suppressing low energy electrons from the walls, which are a potential source of background.

two TMPs[19] mounted on the lid. Each pumping port is designed to house a tailor-made

[19]TMP = Turbo-Molecular Pump

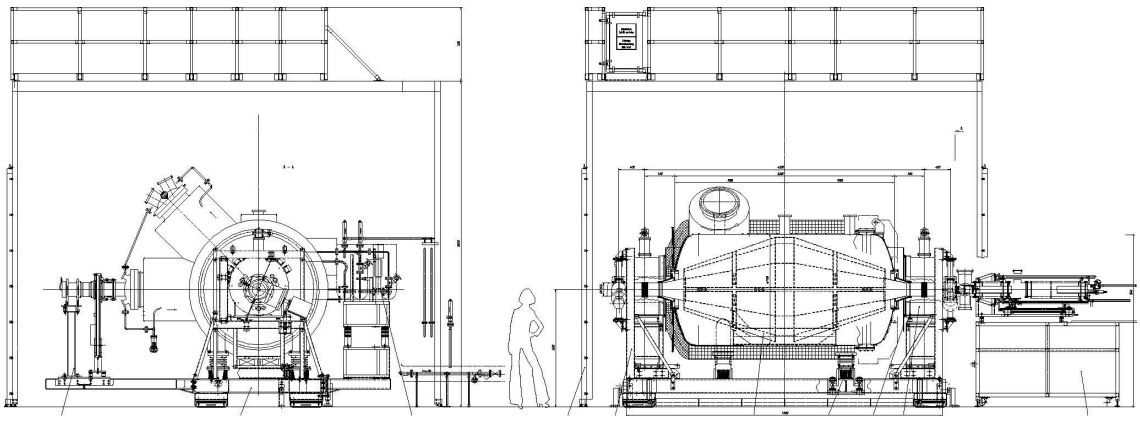

Figure 62: Drawing of the KATRIN pre-spectrometer with the prototype detector attached on one end. The setup stands in a compartment with a continuous flow of de-humidified air.

NEG[20]-pump cartridge to provide the necessary pumping speed for the ultimate vacuum inside the tank. The length of the pumping ports is determined by the maximum magnetic field tolerated by the TMPs. The whole vacuum vessel is electrically insulated and can be put on high voltage (HV) up to 35000 V. It is placed upon ceramic insulators on a stainless steel support structure (see fig. 61). The TMPs and the heating-cooling system are also connected via ceramic insulators.

The whole pre-spectrometer is therefore a small copy of the KATRIN main spectrometer. All essential technical challenges of the KATRIN main spectrometer (*e.g.* XHV $\leq 10^{-11}$ mbar, high-voltage on the hull of the vessel ≤ 35 kV and the nearly massless inner electrode system) have also to be met for the pre-spectrometer. Therefore the KATRIN collaboration decided to build the pre-spectrometer at an early stage of the experiment to test the new ideas and their technical solutions. The pre-spectrometer has been the first major hardware component operated at FZK. After manufacturing and initial vacuum-tests at SDMS (France) the vessel was delivered to FZK in autumn 2003 (see fig. 56). Since then, a detailed experimental programme has been performed to test the KATRIN vacuum concept (see fig. 63). The vacuum measurements have been completed at the end of 2004, and electromagnetic properties of the pre-spectrometer will be investigated starting in early 2005.

5.1.1 Vacuum system

As noted earlier, the pre-spectrometer needs to achieve good XHV conditions in order that the main spectrometer may operate at its design pressure of 10^{-11} mbar or less. It is regarded as a test-bed for the necessary preparation techniques of surfaces, the design and reliable handling of large vacuum flanges and for the pumping systems to be used in the main spectrometer. The vacuum design is largely based on clean and well-prepared vacuum surfaces pumped with a combination of TMP's and NEG-pumps.

[20]NEG = Non-Evaporable Getter

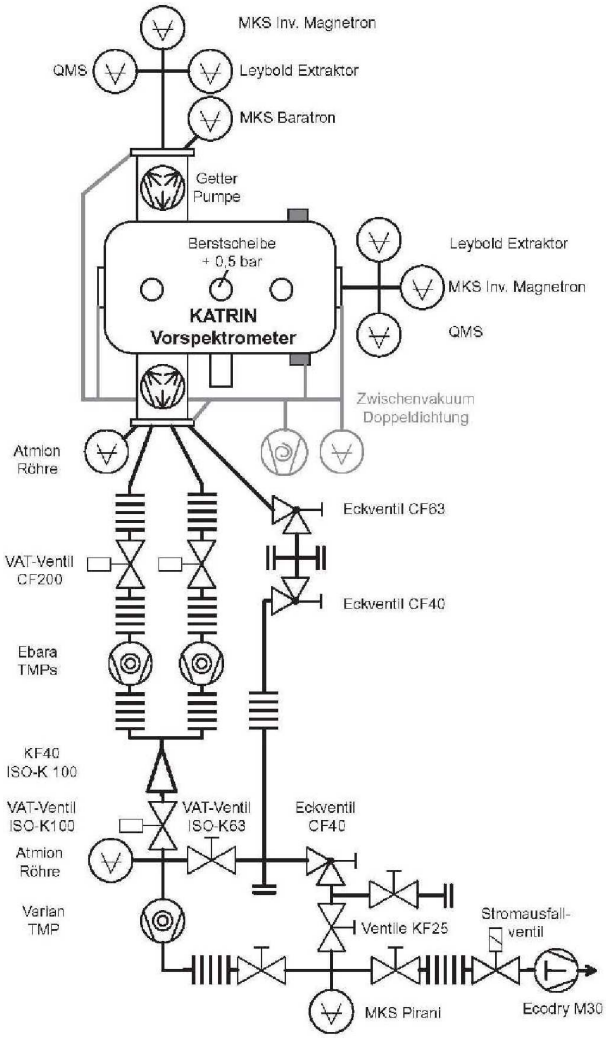

Figure 63: Flow scheme of the pre-spectrometer vacuum setup.

Initial tests of this approach were successful and provided excellent XHV conditions well below 10^{-11} mbar in a small test recipient (volume: 300 l, inner surface: 2.5 m^2, length: 100 cm, diameter: 50 cm) made of the same type of stainless steel (316LN) as the pre-spectrometer.

There are three different types of flanges on the pre-spectrometer:

- At the end caps and at the pumping ports flange rings with an inner diameter of 500 mm, an outer diameter of 610 mm and a height of 60 mm are used. They are made of 1.4429 (ESU) stainless steel with low inclusion content. Flange rings welded to the tank have a flat sealing surface. The counter flanges incorporate differentially pumped seals. A groove holds two spring-loaded all-metal gaskets (HTMS®)with the space in between both gaskets connected by stainless steel tubes to a clean vacuum pump. The interspace is pumped to about 10^{-3} mbar.

- The large vessel flange is based on the same design as the DN 500 flanges. It has an inner diameter of 1680 mm, an outer diameter of 1860 mm and a height of 115 mm. Differentially pumped metal seals are used here, too.
- For smaller flanges commercially available flange rings will be used, ranging from DN 40CF to DN 250CF.

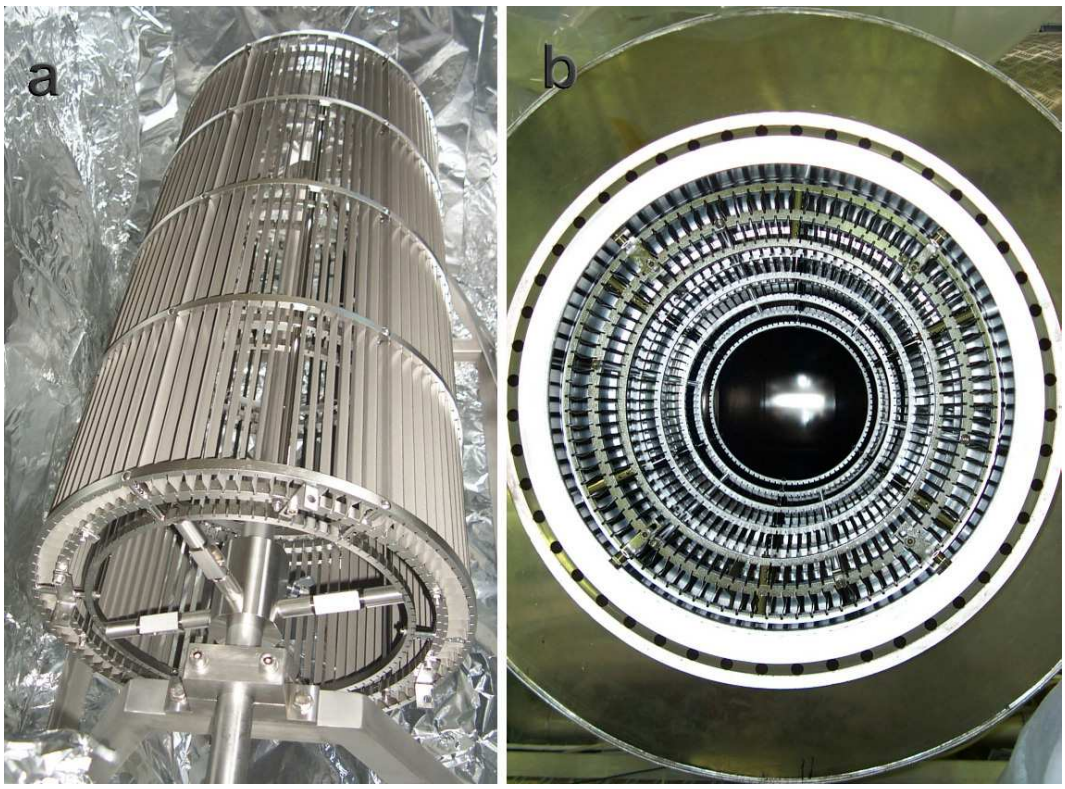

Figure 64: Picture of the 1 m long and 50 cm diameter NEG pump with 90 m of SAES® ST707 getter strips installed. This system has a pumping speed of approximately 10000 l/s and can be upgraded to a total of 180 m of getter strips in each pumping port.

Two turbo-molecular pumps with a pumping speed of 800 l/s each are attached to the horizontal pumping port. The other pumping port has been equipped with a large NEG pumping cartridge (see fig. 64), comprising a total length of 90 m SAES® type ST707 getter strips[21], yielding an estimated pumping speed of 10000 l/s for hydrogen. All inner surfaces of the pre-spectrometer were carefully cleaned and electro-polished following the procedure outlined in section 7.2.

A heating-cooling-system, based on temperature-controlled thermal oil circulating in a closed loop (see fig. 65), allows the pre-spectrometer to operate in a temperature range from -20 °C to +230 °C. An additional electrical heating around the pump ports is used to reach a temperature of 350 °C, which is needed to activate the NEG-pumps. In order

[21] If necessary, a total of 360 m of getter strips can be installed in both pumping parts.

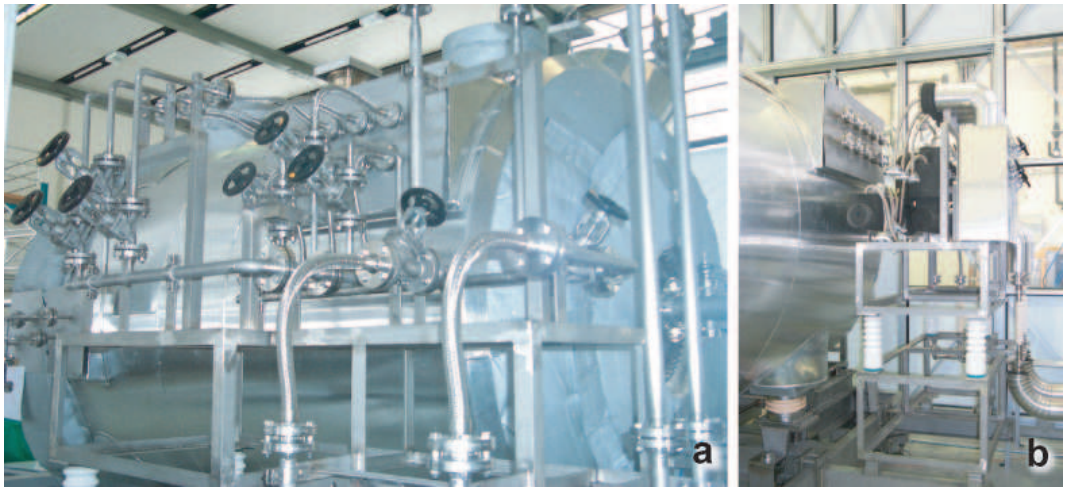

Figure 65: a: The tubes and valves distribute the oil of the heating-cooling system through different circuits of the pre-spectrometer. b: The tubes coming from the heating-cooling system are made of ceramic insulators, separating the vessel, which is on high voltage, from the rest of the system.

to avoid break down of the high voltage on the pre-spectrometer due to the humidity of the air at lower temperatures, the whole pre-spectrometer is surrounded by a dry-air compartment. During low temperature operation the air is kept at a constant dew-point of -25 °C using an adsorption de-humidifier.

Figure 66: The dry-air compartment around the KATRIN pre-spectrometer.

The vacuum tests with the pre-spectrometer addressed several questions. Some preliminary results are listed here[22]:

- A test of the handling procedures and leak-tightness of the new flange design with differentially pumped double sealing was successful. No leak in the inner gasket was detected. Improved procedures for the installation of the outer gasket have been adopted.

- Cleaning procedures and surface preparation were tested by measuring the outgassing rate of the vessel before the first vacuum bake. The methods are used in the specification of the main spectrometer.

- The outgassing rate of the vessel was determined for different baking temperatures, using the *pressure rise method* (separate all pumps from the tank volume by closing valves and measure the rise of the pressure inside versus time). The final step was a vacuum bake at 230 °C. Afterwards the outgassing rate was measured for different wall temperatures (outgassing from vacuum gauges was taken into account and subtracted before calculating the outgassing rate of the stainless steel):
  - +20 °C: $5.3 \cdot 10^{-13}$ $\frac{\text{mbar l}}{\text{s cm}^2}$
  - -0 °C: $1.3 \cdot 10^{-13}$ $\frac{\text{mbar l}}{\text{s cm}^2}$
  - -20 °C: $0.7 \cdot 10^{-13}$ $\frac{\text{mbar l}}{\text{s cm}^2}$

  The specified outgassing rate for the main spectrometer of 10^{-12} $\frac{\text{mbar l}}{\text{s cm}^2}$ has been reached already at room temperature. As a consequence of this result the specifications for the main spectrometer with regard to cooling could be relaxed, leading to a considerable reduction in costs.

- The outgassing from various vacuum gauges (Extractor gauge, mass spectrometer) was determined by measuring the final pressure with a cold cathode gauge with and without these hot cathode gauges on.

- The outgassing and back-streaming from the turbo-molecular pumps was determined by comparing the final pressure with both pumps running with the pressure measured with only one pump connected to the tank.

- The pumping speed of a NEG pump was determined from the final pressure reached after activation of the NEG strips at 350 °C. Unfortunately the final pressure could not be determined, because it was below the sensitivity of the inverted magnetron gauges, which is 10^{-11} mbar. An Extractor gauge, which in principle has a better sensitivity, could not be used, since the outgassing induced by the hot filament completely dominated the local pressure around the gauge. Therefore the pumping speed of the getter pump is only a rough estimate. Assuming the final pressure

[22]The absolute pressure has been measured with an inverted magnetron gauge and was cross-checked with an Extractor gauge. All pressure values here are given in nitrogen equivalent values. No gas correction factors have been applied.

to be just about 10^{-11} mbar and comparing this value with the final pressure of $2.7 \cdot 10^{-10}$ mbar reached with only TMPs (effective pumping speed for hydrogen: 900 l/s) attached to the pre-spectrometer one can estimate a pumping speed for the getter pump of the order of 10^4 l/s. The optimized geometry of the NEG pump has been subject to detailed Monte Carlo simulations showing that this pumping speed correlates with a sticking coefficient on the strips of about 0.5% [117]. This value of the sticking coefficient will be an essential input parameter for the design of the NEG strips of the main spectrometer.

More accurate measurements are planned in 2005, after the installation of the inner electrode system and re-arranging of the Extractor gauges to a position where the distance to the surrounding walls is larger, thereby reducing the outgassing rate.

5.1.2 Electrode system

The retarding field of a MAC-E-Filter is typically created by a system of cylindrical electrodes. For the KATRIN spectrometers we want to avoid a complicated inner electrode system in order to minimize the surface area in the ultra high vacuum and to simplify the construction and assembly. Therefore the retarding potential will be connected directly to the vacuum vessel. For background reduction and fine-tuning of the electric field a very simple and lightweight inner electrode system will be installed inside the vessel serving several tasks:

1. Decoupling of the retarding potential from electric noise on the spectrometer vessel. If we put the vacuum vessel on high voltage the electronic instruments and pumps connected to the vessel might be sources of electronic noise on the retarding potential. If the retarding potential is stabilized by the potential of the additional inner electrode system, noise on the hull of the vessel can be efficiently suppressed.

2. Electrical screening (monopole mode).
   Cosmic muons and ambient or intrinsic radioactivity can induce electrons (see section 8) originating from the electrodes or walls. In general they should not be able to reach the detector, because the Lorentz force bends these electrons back to the wall. In addition the magnetic field gradient creates a magnetic mirror for electrons starting from the central part of the MAC-E-Filter towards the detector. This inherent feature of a MAC-E-Filter is called electromagnetic shielding. However, in the Mainz II setup a small fraction of electrons created at the electrodes ($\approx 10^{-6}$) still reached the detector. This can happen either by non-adiabatic motion, or if the electrons have large cyclotron radii. Both possibilities require large surplus energies, only occurring in a small fraction of electrons. If the wire electrode is put at a slightly more negative potential than the tank, only electrons with enough energy can pass this electric barrier. Experiments at Mainz have tested this idea very successfully (see section 8). It is important to construct inner electrodes with very small geometrical coverage f_{gc} of the vessel, in order to prevent the electrodes themselves to become a significant source of electrons.

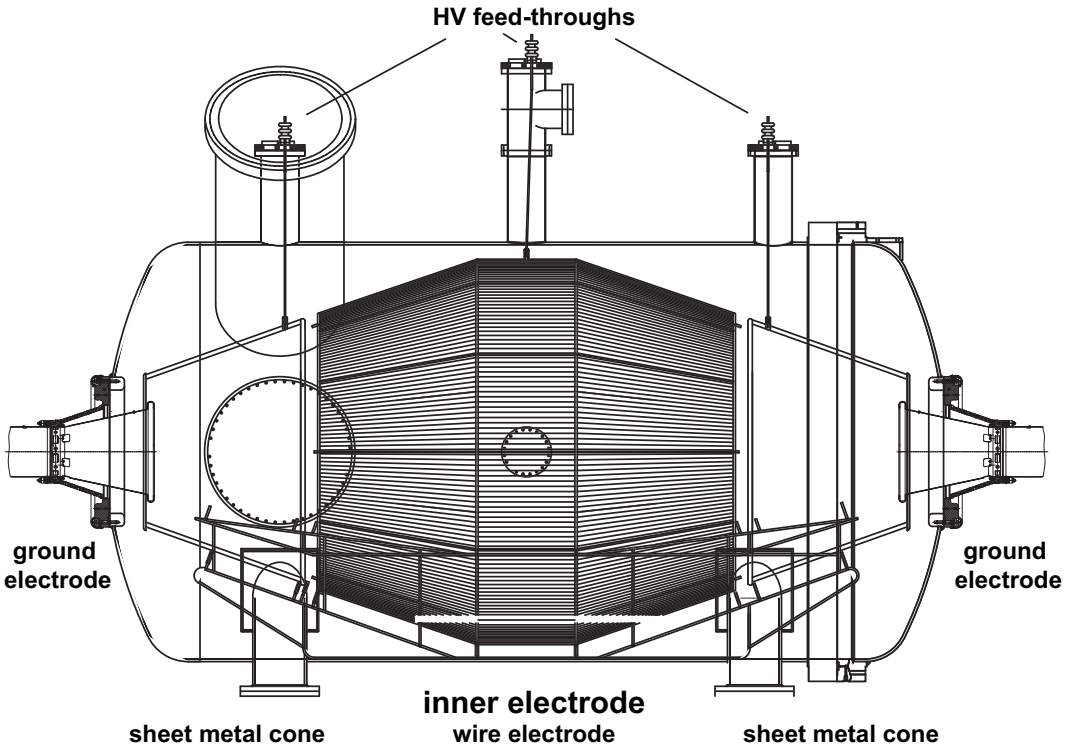

Figure 67: Electrode system of the pre-spectrometer.

3. Trapped particle removal (dipole mode).
   If the additional electrodes are split into two halves and put on different potentials, electrons passing through this region will feel a drift perpendicular to the electric dipole field $\vec{E}$ and to the magnetic field $\vec{B}$ (for this $\vec{E} \times \vec{B}$ drift see also section 8).

4. Shaping of the electric field to avoid Penning traps.
   Simulations showed the need for some shaping of the electric field to avoid Penning traps in corners. This can be achieved by either a complicated shape of the vacuum vessel or a shaping electrode system. In case of the KATRIN main spectrometer, the vacuum vessel has already an optimized shape, whereas for the cylindrical KATRIN pre-spectrometer vessel two conical ground-electrodes at the entrance and exit flanges and an inner electrode system will give the right shape to avoid the Penning traps (see fig. 67).

The very encouraging results obtained with the wire electrode system in the Mainz spectrometer (see sect. 8) led to the design of a similar electrode system for the KATRIN pre-spectrometer. The main parameters of the wire electrodes are described by the following formulae for a cylindrical electrode of wires with a diameter d and wire spacing s placed at a distance l inside a cylinder of diameter r (see fig. 68):

- The geometrical coverage f_{gc} is simply given by (neglecting the frame supporting

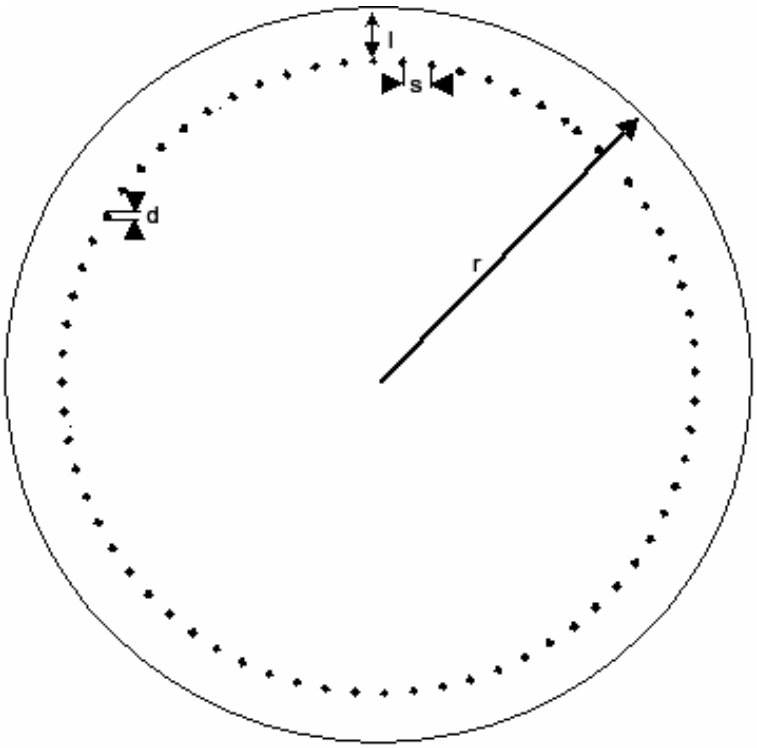

Figure 68: Symmetric setup of wires with diameter d and spacing s placed inside a cylindrical electrode of diameter r at a distance l from the outer wall.

the wires):

$$f_{gc} = \frac{d}{s} \tag{44}$$

The electric potential inside the wire electrode U_{inner} with respect to the vessel is very homogeneous (see fig. 69) and is defined by a screening factor S

$$S = \frac{U_{wire} - U_{vessel}}{U_{wire} - U_{inner}} \approx 1 + \frac{2\pi\ l/s}{\ln(s/\pi d)}\,. \tag{45}$$

With a large screening factor S the requirements on stability of the potential of the vessel U_{vessel} can be relaxed. Distortions on the retarding high voltage U_{inner} are dominated by the stability of the potential of the wire electrodes:

$$\delta U_{inner} = \delta U_{wire} \cdot (1 - 1/S) + \delta U_{vessel} \cdot (1/S)\,. \tag{46}$$

Although the stability of the retarding potential of the KATRIN pre-spectrometer is not as crucial as for the main spectrometer, we plan to test this concept with the pre-spectrometer as a test-bed for the main spectrometer.

- The electric field close to a wire strongly depends on the diameter of the wire (see fig. 68 and 70). It is approximately

$$E_{wire} = \frac{\Delta U}{l} \cdot \frac{s}{\pi d} \tag{47}$$

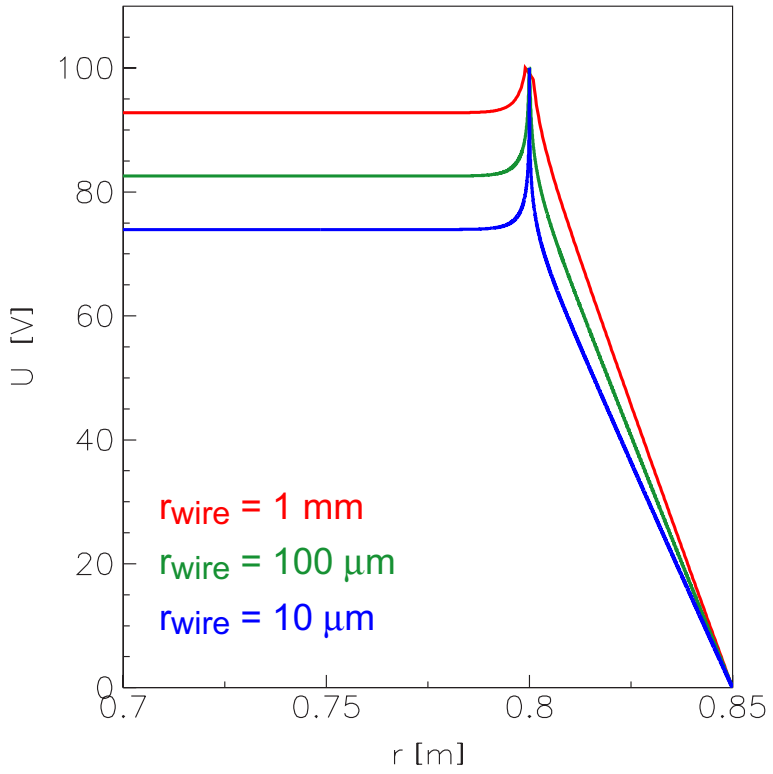

Figure 69: Potential near the wires of a cylindrical wire electrode system as shown in fig. 68. As for the central part of the pre-spectrometer wire electrode the parameters are $r = 0.85\,\mathrm{m}$, $s = 2\,\mathrm{cm}$, $l = 5\,\mathrm{cm}$ and the relative wire potential is $\Delta U = 100\,\mathrm{V}$. The wires of the pre-spectrometer electrode have a radius of $250\,\mu\mathrm{m}$.

with $\Delta U = U_{vessel} - U_{wire}$. The electric field E_{wire} has to be small, since the experiments at Mainz showed that field emission starts above $0.4\,\mathrm{MV/m}$.

The requirements for the inner electrode system are a low geometrical coverage f_{gc}, a high screening factor S and a small electric field E_{wire}. Since these requirements have an opposite dependence on the diameter of the wire d and the spacing s, a compromise has to be found. The inner electrode system of the pre-spectrometer will have a wire diameter of $d = 0.5\,\mathrm{mm}$ with a wire spacing in the cylindrical part of $2\,\mathrm{cm}$, yielding a geometrical coverage of $f_{gc} = 2.5\,\%$. Having a distance to the vessel of $l = 5$ cm, voltages up to several kV can be applied and the screening factor amounts to $S \approx 7$. Both ends of the conical parts are made of sheet metal to reduce the electric fields near the ground electrodes. All three sections of the inner electrodes will be split vertically to apply an electric dipole voltage using the $\vec{E} \times \vec{B}$ drift to eject trapped charged particles. The construction of the electrode system is nearly finished (see figure 71) and installation is scheduled for January 2005.

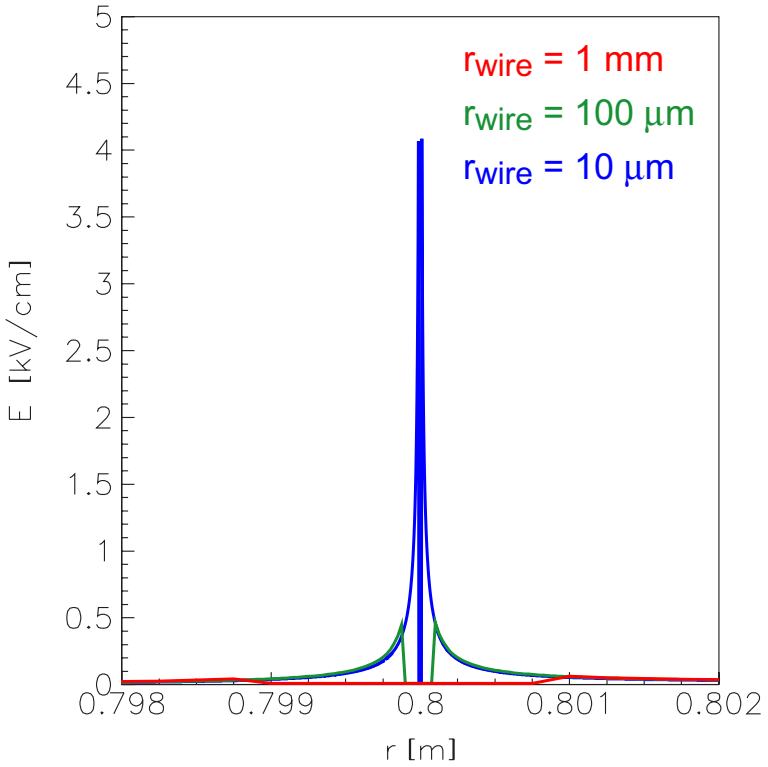

Figure 70: Electric field near the wires shown in fig. 68.

5.1.3 Test programme

The electromagnetic properties of the KATRIN pre-spectrometer will be investigated with an electron gun (see sect. 9) and an electron detector array (see sect. 6). The test programme aims at a verification of the new electromagnetic design concepts, e.g. high voltage directly applied to the spectrometer vessel, the wire electrode and background suppression with the monopole and dipole mode. The results may influence the concept of the electrodes for the main spectrometer. On the other hand this set-up will be useful as a test facility for other components under real conditions, e.g. the slow control system and the DAQ system. Presently, about 10 different measurements are planned, grouped into several questions:

- How can the requirements for the High Voltage stability be fulfilled?
  This question stands at the beginning and the end of the pre-spectrometer test measurements since the retarding HV is applied directly to the spectrometer vessel. The first step of these measurements has been made at the end of 2004. As an essential part of the set-up, the electrode system will be installed in January 2005. When a stability of the retarding HV of better then 10^{-4} is reached, the measurements will be interrupted to perform the next steps of the test programme. At the end of the test programme, the HV stability tests will be continued. Together with the precision High Voltage equipment (see sec. 9.1), we aim for a stability of the retarding

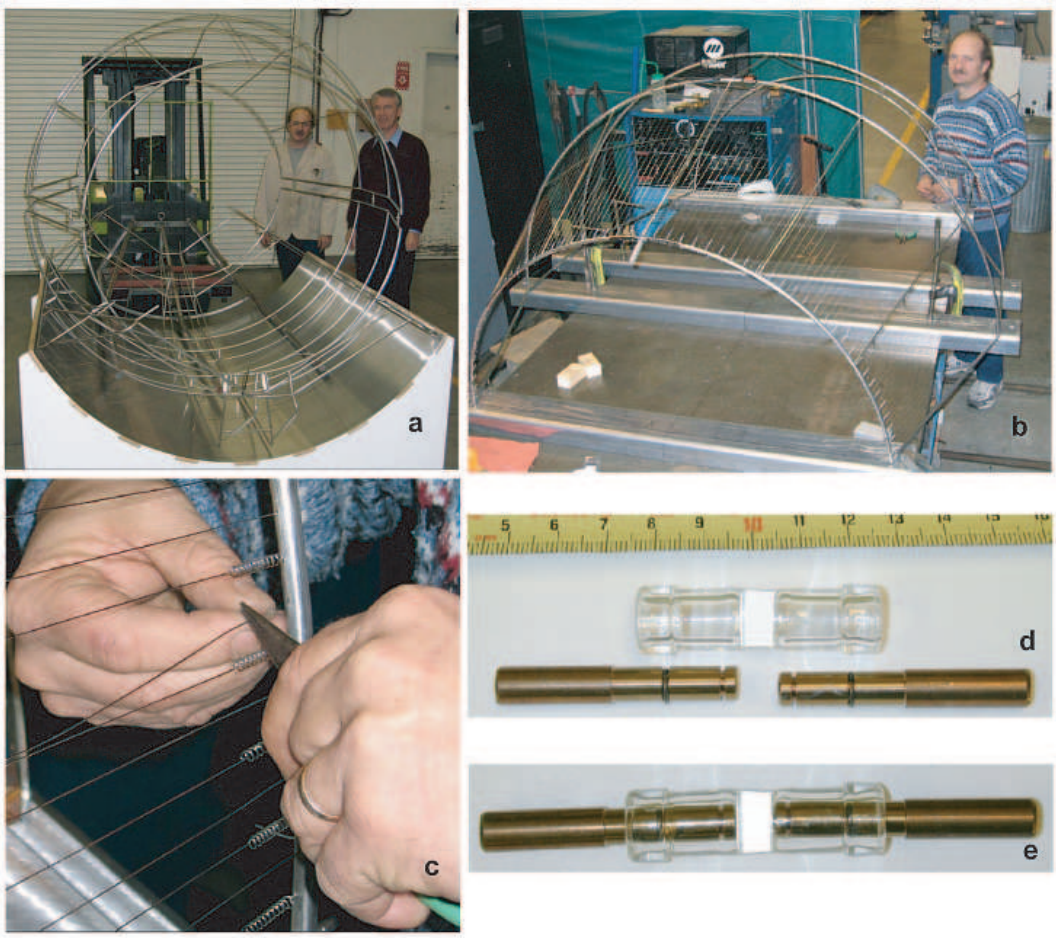

Figure 71: Construction of the wire electrode system for the KATRIN pre-spectrometer. a: wire frame, b: part of the frame equipped with wires, c: wires with springs to keep the wires straight, d and e: glass insulators hold together different parts of the inner electrode.

HV of better then 10^{-6}. Recently the discussion about the High Frequency stability of the spectrometer High Voltage system (pre-spectrometer as well as main spectrometer) has come up. The necessity of test measurements and the installation of appropriate HF filter systems were identified. These measurements will be included into the test programme.

- What is the background level of the pre-spectrometer?
  The measurement of the background level of the pre-spectrometer will be the most time consuming part of the test programme since the better the results are the longer one has to wait for an acceptable statistic (for questions concerning the origin of background in MAC-E-Filters and previous test measurements with the Mainz set-up see sec. 8). Also necessary for these measurements are the pre-spectrometer magnets and the electron detector array. Beside simple measurements of the background level, these tests will also consist of measurements performed under various

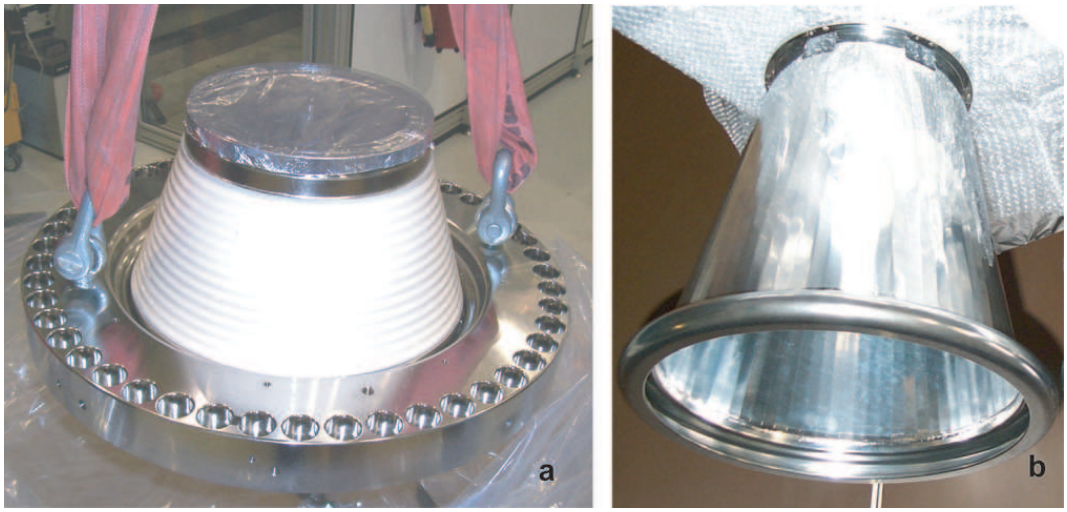

Figure 72: Ground electrode of the KATRIN pre-spectrometer. a: 500 mm flange with ceramic insulator, b: titanium ground electrode which will be attached to smaller end of the ceramic insulator.

conditions of *e.g.* pressure, retarding potential or magnetic field. Further on, background from the walls of the pre-spectrometer vessel will be investigated, *e.g.* by using an external gamma- or X-ray source. In all these experiments, the shielding of the electrode will be tested and the results will be compared to calculations and previous measurements at Mainz. Finally, the storage and the removal of charged particles in the pre-spectrometer will be tested. To expel stored charged particles from the spectrometer volume, the electrode will also run in an electric dipole field mode.

- What are the characteristics of the pre-spectrometer in the MAC-E-Filter mode?
  To answer this question, a couple of typical parameters of a MAC-E-Filter will be measured, *e.g.* the transmission function. For these measurements the above mentioned photoelectron gun is needed. Its features will be an electron rate of about $10^4 - 10^5\,\mathrm{s}^{-1}$ and an energy spread of $< 0.5\,\mathrm{eV}$ (at $18.6\,\mathrm{keV}$). Since the angular spread of the emitted electrons is small, the photoelectron gun will be installed on a 2-dim movable support frame to enable coverage of the total "transported magnetic flux" of the KATRIN experiment. The results will be compared with simulations performed in the design phase of the pre-spectrometer. This measurement will give additional information about the homogeneity of the electric retarding potential and the magnetic field in the pre-spectrometer. Closely connected to this topic is the measurement of the validity of the adiabatic momentum transformation of the decay electrons on their way through the pre-spectrometer. This will be also investigated without retarding potential.

- What is the quality of pre-spectrometer as pre-filter for β-electrons?
  An essential measurement is the test of the pre-spectrometer as a pre-filter for decay electrons with kinetic energies less then $18.4\,\mathrm{keV}$. This measurement is similar to a

combination of a background measurement with the measurement of the transmission function. However, in contrast to the latter one, the retarding potential will block the incoming electrons. So the count rate will be measured as a function of the energy of the photoelectrons. To simulate the electron intensity of a normal tritium measurement a photoelectron gun of a much higher electron emission ($\approx 10^{12}\,\mathrm{s}^{-1}$) is needed requiring modifications of the photoelectron gun.

The test measurements have recently started with the first HV test, background tests will follow in spring 2005, after the electrode is installed and the magnets are available. The estimated measurement time is about one year depending on the results of the first tests. Having finished the pre-spectrometer test measurements, the current set-up will be decommissioned and the pre-spectrometer will be installed at its final position in the new KATRIN building.

5.2 Main spectrometer

The main spectrometer will be a key component of the KATRIN experiment. It is a large electrostatic spectrometer (see fig. 73) with a diameter of about 10 m and an overall length of 23.3 m. Two superconducting solenoids will generate a strongly inhomogeneous magnetic field guiding the β-electron through the spectrometer. The flux tube expands to a diameter of 9 m (see eq. (38)) in the central analyzing plane. Additional air coils will allow fine-tuning of the magnetic field in the centre and compensate the earth magnetic field. Like the pre-spectrometer, the outer hull of the main spectrometer will be on high potential, serving as a *guard electrode* for the more accurate high voltage on the inner wire electrodes. This high resolution MAC-E-Filter will allow the tritium β decay endpoint to be scanned with increased luminosity and a resolution of 0.93 eV, which is a factor of 5 better than previous MAC-E-Filters.

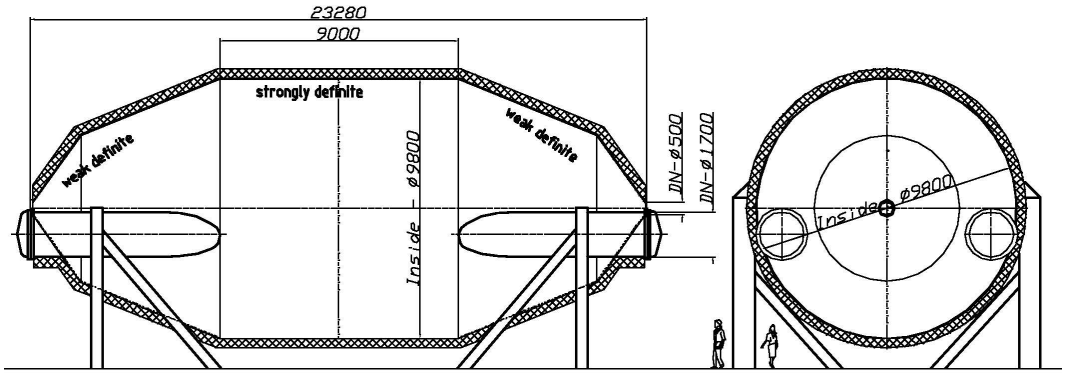

Figure 73: Schematic view of the main spectrometer vessel.

5.2.1 Main spectrometer vessel

Since the background rate has a direct influence on the sensitivity of KATRIN on the neutrino mass (see section 8) vacuum requirements for this big vessel are very demanding. The combination of a large recipient and stringent XHV requirements represents a technological challenge, since XHV vessels of this size have not been built before. Therefore, the technical feasibility of the main spectrometer had to be investigated in a study conducted in collaboration with several industrial partners. Combined with the vacuum experience gained with the pre-spectrometer it demonstrated the feasibility to build a spectrometer vessel of this size, meeting all requirements within a concise construction schedule.

In order to achieve the requisite vacuum performance, a careful cleaning and surface preparation protocol will have to be invoked (see section 7.2.2). Since the surface has to be prepared in such a way as to minimize field emission, surface smoothing techniques, which are often not required for pure vacuum reasons, will need to be employed. In particular, electropolishing of the finished vessel will be specified. A final vacuum bake at 350 °C of the entire vessel will remove water and reduce the outgassing rate of hydrogen from the stainless steel surface below $10^{-12}\,(\mathrm{mbar\,l})/(\mathrm{s}^{-1}\,\mathrm{cm}^{-2})$ at room temperature. This requires a heating system with a total power of 400 kW. The heat will be transmitted by circulating thermal oil through a tube system around the tank.

The vacuum issues of the main spectrometer are described in section 7.2.2. The pumping system is based on NEG-pumps for getterable gases (mainly hydrogen) and TMPs for other, non-getterable gases (mainly noble gases). This combination has been tested successfully with the pre-spectrometer. Depending on the actual outgassing rate that will be achieved for the main spectrometer, the pumping speed of the installed NEG pumps will be in the order of $5 \cdot 10^5$ l/s (3–6 km NEG strips). The effective pumping speed of the TMPs will be in the order of 10000 l/s. While the necessary amount of getter strips depends on the outgassing rate, the number of TMPs will be determined by the large amount of hydrogen set free during activation of the getter strips.

The main vessel will be manufactured from stainless steel sheets, type 1.4429 (316LN) selected both for its strength and excellent magnetic properties, especially in weld regions. The weight will be approximately 200 tons. There are limits on the amount of cobalt permitted in the steel and on its residual radioactivity to minimize background in the spectrometer. The thickness will vary from about 25 mm to 32 mm. The properties of the vessel are:

- inner diameter of the cylindrical section: 9.8 m
- total length: 23.28 m
- inner surface: 650 m^2
- volume: 1400 m^3
- three pump ports with a diameter of 1.7 m and a length of ≈ 3 m. Each pump port will terminate in a large flange, similar to the one tested in the pre-spectrometer.

- the electron beam will enter the spectrometer on axis through a ground electrode attached to a ceramic insulator on a 500 mm flange, similar to the design used for the pre-spectrometer (see fig. 72).

- during bake-out at 350 °C the vessel will expand by 20 cm along the axis. In order to prevent damage to the pre-spectrometer, the differential pumping section and the source, it will be fixed to the floor at its upstream end. The downstream end will push back the transport section and the detector attached to the ground electrode. Therefore all downstream installations have to be mounted on rails.

The vessel has been ordered in December 2004 and will be delivered to FZK at the end of 2006. The final acceptance procedure includes dedicated tests of the heating system and of the vacuum properties of the vessel.

5.2.2 Electrode system

Since the main spectrometer has to analyse the endpoint region of the tritium β spectrum with an unprecedented precision, all electromagnetic properties have to be known very precisely and kept under strict control. Especially the transmission function of the main spectrometer should fully comply with the expectations for perfect adiabatic transformation according to eq. 16. Fig. 75 shows the electromagnetic design of the main spectrometer with a 2–layer inner electrode system.

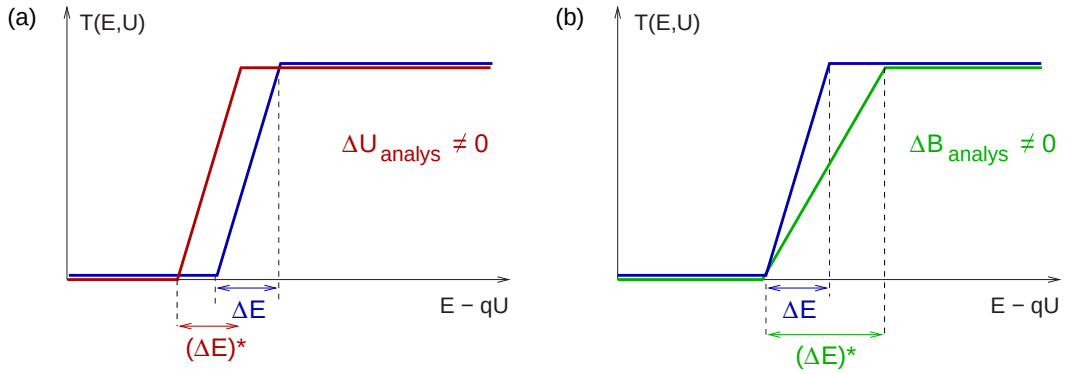

Figure 74: Influence on the transmission function of the inhomogeneities of the electric retarding potential (left) and of the magnetic field (right).

The electrode design is based on the concept that the vacuum tank itself serves as a *guard electrode* for the more accurate high voltage on the inner wire electrode system. The electron beam enters and leaves the vessel through conical ground electrodes attached to the spectrometer via ceramic insulators (see fig. 72).

The actual design of the inner electrode system and of the ground electrodes is currently being optimized. Stringent requirements on the homogeneity of the electric retarding potential and the magnetic field in the analyzing plane are necessary (see fig. 74) to reach the ambitious energy resolution of 0.93 eV and to suppress systematic errors. Small

inhomogeneities of the electric retarding potential and of the magnetic field in the analyzing plane will be compensated to first order by the good spatial resolution of the detector. The electron flux through the analyzing plane will be projected on a segmented electron detector, which has a resolution of about 400 pixels (20×20). Each pixel corresponds to a certain area on the analyzing plane. The inhomogeneities of the electric potential and of the magnetic fields have been calculated in figures 76 and 77 and can be used to correct deviations for each pixel.

The inner electrode system will allow to be operated in the *monopole mode* during normal measurements (see sec. 5.1.2), suppressing electrons from the spectrometer wall. It can be also operated in *dipole mode*, throwing out trapped electrons. Since the surface of the spectrometer is about 2 orders of magnitude larger as compared to the previous MAC-E-Filters a very low geometrical coverage of the wire electrode of $f_{gc} \leq 1$ % is required, which calls for a small wire diameter d and large wire spacing s (see fig. 68). On the other hand the large electrical potential of up to 1 kV needed in the dipole mode and large electrical screening factor $S \geq 30$ (see eq. (45)) required to achieve a very stable retarding potential ask for just the opposite, which cannot be fulfilled by a single wire layer design. Therefore the current design favours a two-layer wire design for optimal suppression of background by a small geometrical coverage f_{gc} of the second wire layer, for the possibility of the dipole operation mode by a large diameter first wire layer and for an improved stabilization of the retarding high voltage by a huge screening factor as

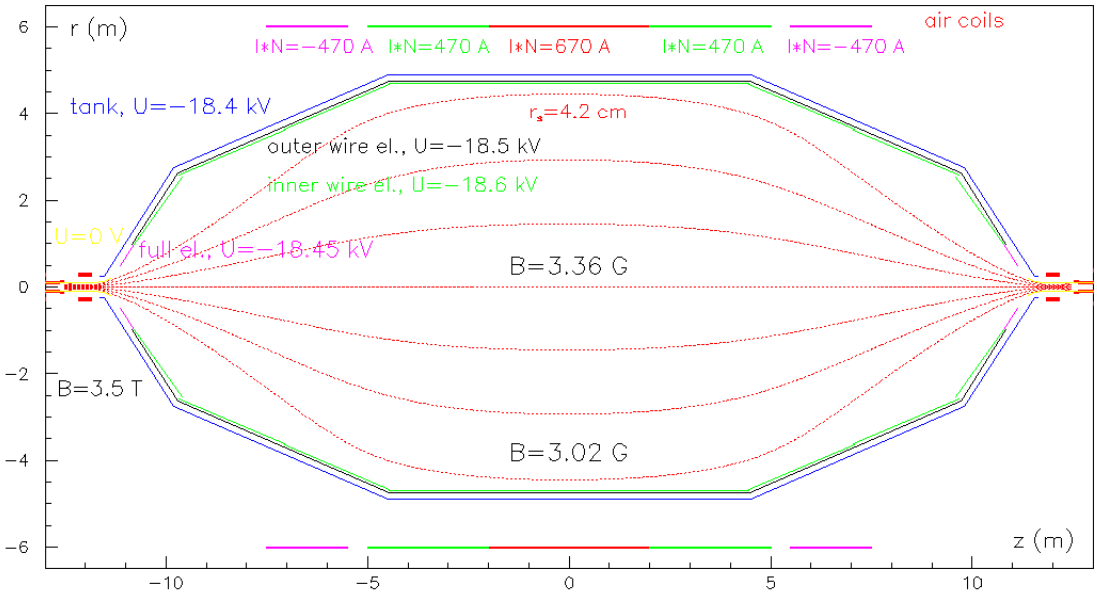

Figure 75: Electromagnetic design of the KATRIN main spectrometer. The field shaping air coils around the analyzing plane are shown, whereas the earth field compensation air coils are not plotted.

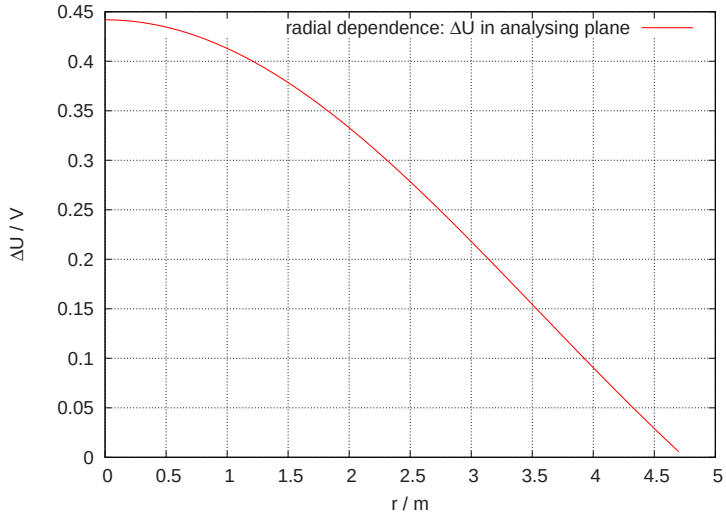

Figure 76: Electric retarding potential inhomogeneity in the analysing plane of the main spectrometer

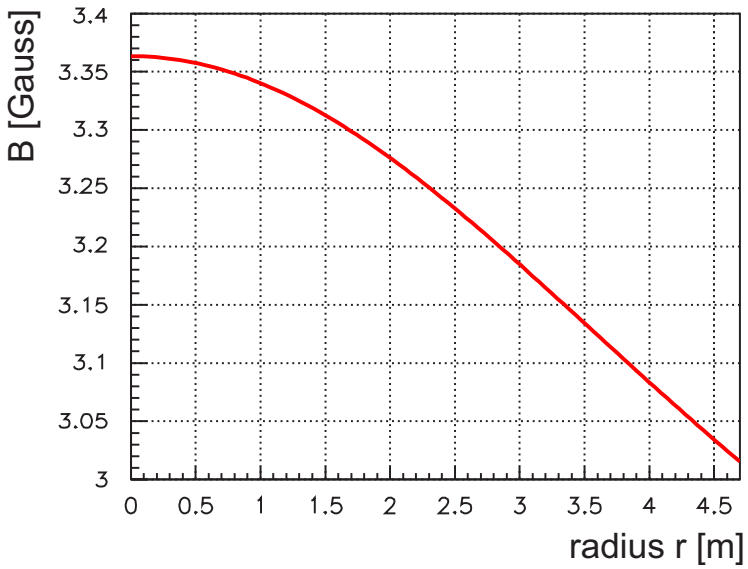

Figure 77: Magnetic field inhomogeneity in the analysing plane of the main spectrometer

a product of the screening factors of both layers.

The inner electrode system will consist of about 200 (double layer) wire modules with a weight of 20–30 kg each (see fig. 78). These modules will be brought into the spectrometer through one of the large diameter flanges and mounted via electrical insulators onto the walls. An ultra-high precision high voltage system (see fig. 79) will feed the different wire layers and wire sections as well as the outer vessel with the necessary voltages for the normal monopole and the dipole mode. High precision variation of the main spectrometer potentials is possible for scanning the β spectrum near its endpoint.

In the analysing volume the magnetic field is at its minimum value $B_A = 3\,\mathrm{Gauss}$

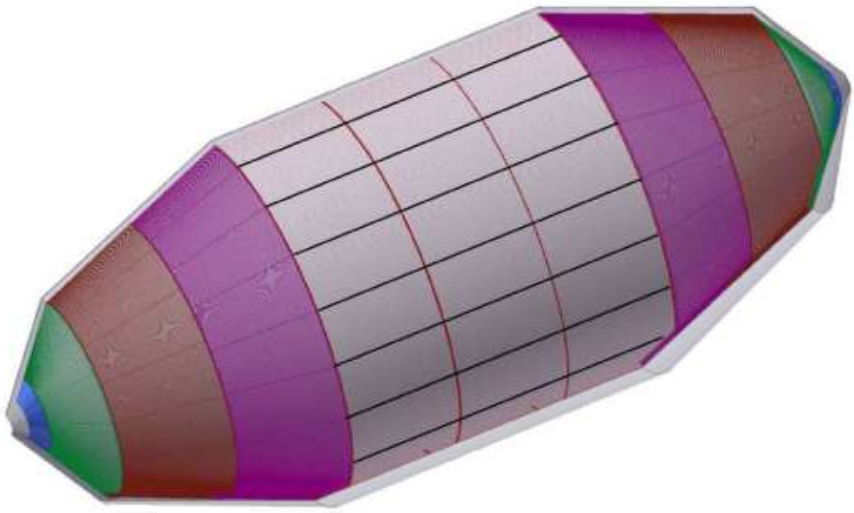

Figure 78: Subdivision of the inner wire electrode system of the main spectrometer into modules.

(see fig. 77). The influence of low transverse magnetic stray fields and especially of the geomagnetic field ($B_E = 0.65\,\mathrm{Gauss}$, inclination 64°) result in transverse shifts of electron tracks which result in a distortion of the radial symmetry of the transmission function and which may cause an increase of background. The compensation in transversal direction will be provided by a generalized Helmholtz coil arrangement of ellipsoidal shape [118] (see also fig. 80) placed close to the outer surface of the main spectrometer. A set of 20 current loops ($I = 50\,\mathrm{A}$) produces a sufficiently constant magnetic field in the analyzing plane. One set is needed for the z (horizontal) and y (vertical) direction respectively. The compensation of the axial (x–direction) geomagnetic field component is achieved by adjusting the currents in the coaxial large diameter low field coils at the centre of the main spectrometer.

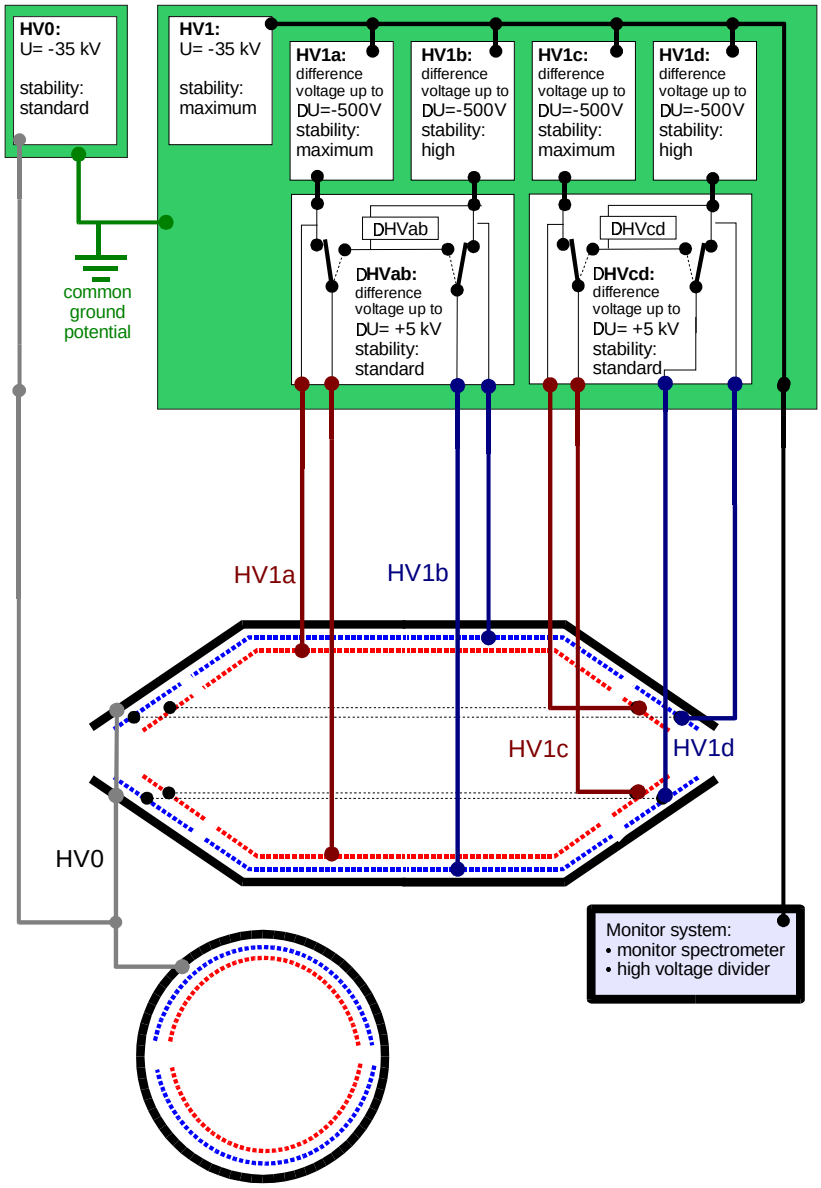

Figure 79: High voltage supply system for the main spectrometer. A base high voltage $HV1$ supplies 4 difference voltage power supplies $HV1a$ – $HV1d$ to feed the 2 layers of the central wire modules ($HV1a$, $HV1b$) and the conical wire modules ($HV1c$, $HV1d$). The inner wire modules will be supplied with the maximum possible stable voltages. High voltage switches allow to split the upper and lower part of the wire electrodes with a difference of up to 5 kV for the dipole mode. The base high voltage $HV1$ is also supplied to the main spectrometer and measured by the high precision high voltage divider. The monitor spectrometer vessel is supplied with the voltage $HV0$. A safety system (not shown in the figure) takes care that the allowed voltage differences do not exceed given limits and shuts the voltage down in a controlled way in case of emergency.

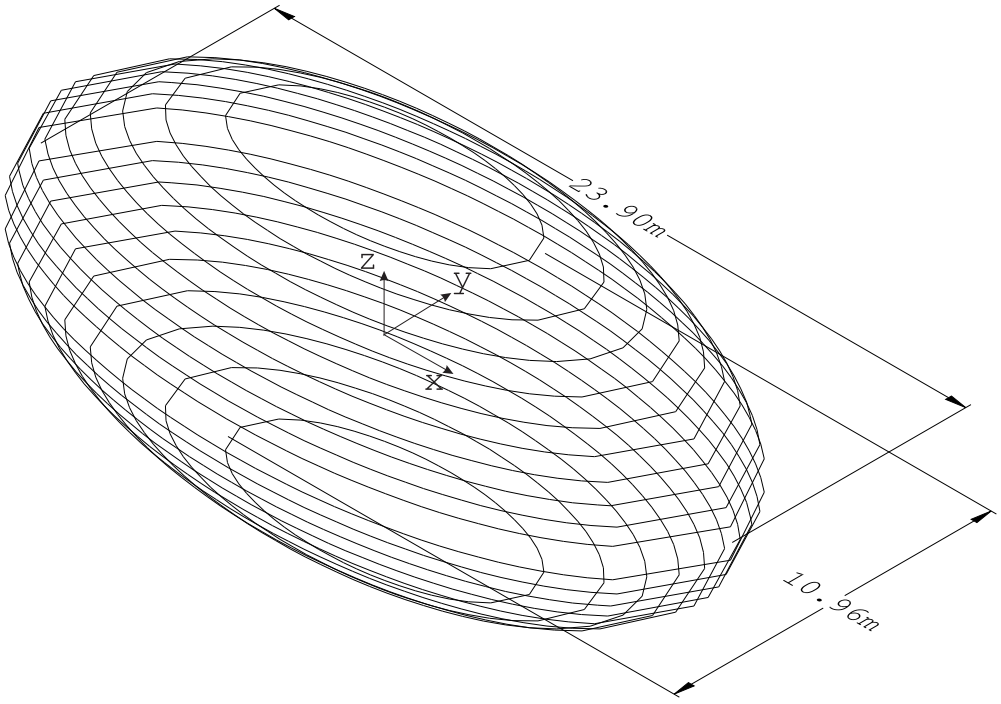

Figure 80: Arrangement of transversal ellipsoidal air coils to compensate transversal magnetic fields.

6 Electron detector system

All β-electrons passing the retarding potential of the main spectrometer are re-accelerated to their initial energy and magnetically guided by the 2-solenoid detector transport system (DTS) to the final plane detector (FPD). The FPD is located inside a separate superconducting solenoid with large warm bore. The final plane detector has to detect

- electrons from tritium β-decay with energies up to 18.6 keV
- conversion electrons from ^{83}Kr with energies from 17.8 keV up to 32 keV
- electrons from a high rate electron gun ($\sim$ 100 kHz)

It is advantageous to post-accelerate electrons by an additional potential of up to 30 kV, implying that the FPD performance has to be optimized over the energy range of 5-50 keV. There are two detector technologies, which are well suited to count electrons in this low-energy regime: semiconductor detectors as well as bolometers. As the operation of cryogenic bolometers in the 'open' KATRIN geometry is technically very challenging, the reference design of the FPD is based on a semiconductor detector (technical details of both options are discussed in sect. 6.3).

6.1 Geometry and layout

The FPD is located at the center of the detector magnet, which generates a magnetic field B_{det}. The maximum angle $\theta_{det,max}$ of incoming electrons from the source is given (without post-acceleration) by the ratio of magnetic field strengths at the detector and the pinch magnet B_{max}:

$$\theta_{\mathrm{det,max}} \approx \arcsin\left(\sqrt{\frac{B_{det}}{B_{max}}}\right)$$

For the reference design with $B_{det} = 3$T the maximum angle of incidence with respect to the surface of the detector amounts to $\theta_{det,max} = 45°$. In fact, a smaller value of B_{det} would reduce θ_{max} and hence reduce backscattering from the detector surface, but it would require also a larger detector area, which would lead to a higher background contribution from the detector environment. Consequently, a higher magnetic field would decrease this background contribution as the sensitive detector area scales with the inverse of the magnetic field strength. However, for $\theta_{det,max} = 45°$ the backscattering probability for a 20 keV electron already amounts to $P_{back} = 0.33$, so for even higher expected incidence angles one would have to take counteractive measures. This option is discussed in subsection 6.2.4. As the FPD is designed to cover a magnetic flux $\phi = 190$ T$\times$ cm^2, the reference value of $B_{det} = 3$ T implies that a circular area of $A_{det} = 63$ cm^2 has to be instrumented. This corresponds to a detector diameter $d_{det} \approx 9$ cm.

To reduce besides backscatter probabilities the detector sensitivity to γ-induced background, a low atomic number Z is required, making silicon-based detectors a suitable choice. This background is further minimized by using a thin detector as electrons with

energies of a few tens of keV have track lengths of up to a few μm only (e.g. 4.4 μm in silicon for E = 18.6 keV). However, a minimum thickness of $d_{det} > 150$ μm ensures that the energy loss signal of minimum ionizing particles such as cosmic muons lies above the electron energy window from 5-50 keV. The short electron track length in turn requires an entrance windows thickness of $\lambda \ll 1\,\mu$m to minimize the energy losses in the insensitive part of the detector (see below).

The detector must be surrounded by an active and passive shielding to fulfill the background goals (see sec. 6.4). Currently, options on the magnet design are evaluated to find a compromise between effective shielding (tendentially needs large bores of magnet) and minimization of stray fields influencing the homogeneity of the analyzing plane of the main spectrometer (small bores of magnet).

6.2 Design requirements

The primary design goal for the FPD is to achieve a very high detection efficiency for electrons in the energy range from a few keV at threshold up to several tens of keV. The reference design thus assumes an overall detection efficiency for electrons of $\varepsilon > 0.9$, allowing efficiency losses of up to 5% due to insensitive detector areas and up to 5% due to an off-line energy cut to suppress specific background classes.

6.2.1 Position sensitivity

A key task of the final plane detector is to record the position of the incident particle in the detector plane (x,y) perpendicular to the magnetic flux tube. This measurement is important, as for an adiabatic MAC-E filter a measurement of the (x,y) coordinates of the particle at the detector plane allows the localization of its entire track coordinates in the (x',y') planes. In particular, the $(\bar{x}, \bar{y})$ coordinates at the tritium source and the $(\tilde{x}, \tilde{y})$ coordinates at the analyzing plane of the spectrometer can be calculated. Thus the position information from the FPD can be used in the off-line data analysis to

- map the inhomogeneities of the electric retarding potential in the analyzing plane (electron gun runs)
- monitor the homogeneity of the radial source potential (tritium runs)
- suppress background originating from outside the guided magnetic flux tube

Electromagnetic design considerations as well as estimates of the potential source charging show that these demands can be met if the analyzing plane can be segmented into ≈ 20 divisions along the diameter in arbitrary direction of the analyzing plane. For these reasons, the FPD will consist of an array of several hundred small independent semiconductor detectors, each with an area in the range of 10-20 mm^2. This also has the beneficial side effect of a minimized noise contribution for each sub-unit and hence of a superior energy resolution of the array.

6.2.2 Energy resolution

The energy resolution of the FPD has to be optimized for low-energy electrons in the 5-50 keV range[23]. This allows the definition of a narrow energy interval for signal electrons in the off-line analysis and discrimination against background reactions leading to a *flat* energy spectrum at the detector. In this case, a background reduction proportional to the inverse of the detector energy resolution can be achieved. A continuum type background distribution in the energy range from 5-50 keV is expected for example from the β- and γ-induced background from the areas surrounding the detector (see section 6.4). In [119] it was shown that the level of this background continuum could potentially limit the ν-mass sensitivity of KATRIN. The MC simulations indicate that an energy resolution $\Delta E < 1$ keV (FWHM) is sufficient to discriminate against this type of background if special emphasis is put on careful selection of building materials.

A significantly higher energy resolution would allow discrimination against other types of background reactions, such as beam correlated background. A major source of potential background are low-energy electrons with a few eV of surplus energy originating from the large area close to the analyzing plane of the main spectrometer, as they arrive with essentially the same energy as the β-electrons ('full energy peak'). To discriminate this type of background from signal electrons would require a detector with an energy resolution well below 10 eV and able to identify a potential micro-structure within this 'full energy' peak. This is a very challenging benchmark even for a high resolution bolometer array. For this reason the KATRIN collaboration has developed and successfully tested other methods to suppress this background class, i.e. by incorporating massless inner wire electrodes into the pre- and main spectrometer (see sections 5 and 8.3)

Based on the above considerations, the design criterion for the energy resolution of the final plane detector for low-energy electrons is $\Delta E < 600$ eV (FWHM). If this demand can met by the FPD during measurements, it would correspond to more than a factor of 2 improvement compared to the performance of the Si-detectors presently in use at the Mainz and Troitsk experiments, which have achieved resolutions of 1.4 keV [86, 87].

The total energy resolution $\Delta E_{tot.}$ of a detector is given by the quadratic sum of the intrinsic resolution ΔE_{Fano} and the equivalent noise contributions (ENC) of the electronics ΔE_{ENC}. In silicon, the intrinsic Fano noise for an electron of 18.8 keV energy amounts to $\Delta E_{Fano} = 220$ eV (FWHM). This resolution is slightly broadened to 230 eV due to energy loss and path length fluctuations in the dead layer of the detector. The electronic noise contribution strongly depends on the detector type. For a PIN-diode and an Avalanche Photo-Diode (APD) with a sensitive area of several mm^2, an electronic noise contribution of $\Delta E_{ENC} \approx 600$ eV has been assumed (see Fig.81 a). By contrast, a high resolution detector such as a silicon drift diode (SDD) achieves a value of $\Delta E_{ENC} < 100$ eV.

[23]Note that the energy measurement of the incoming electrons at the detector is *not* used for β-spectroscopy, this task is performed at the analyzing plane of the main spectrometer.

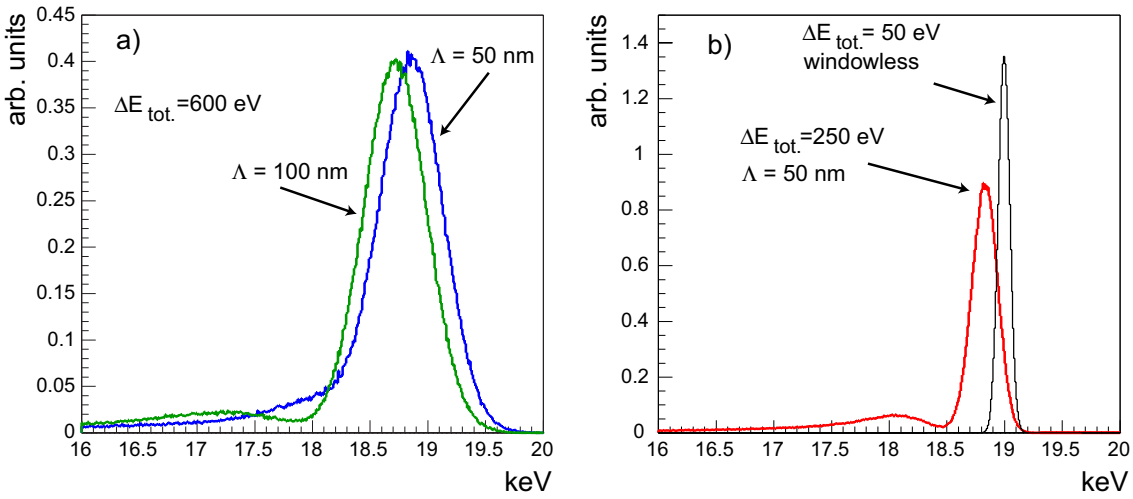

Figure 81: Simulated detector response functions of various detector types for mono-energetic electrons with $E_0 = 19\,\mathrm{keV}$. a) response function expected for PIN diodes or APDs with resolutions $\mathrm{E_{tot}}$(FWHM)=600 eV and different dead layers (Λ = 50,100 nm), b) response function expected for high resolution devices such as SDDs, DEPFETs (red curve) or bolometers (black curve, scaled by factor 1/2). Curves in figure a) and b) have the same normalization.

6.2.3 Detector response function

The overall detector response to low-energy electrons is governed not only by the energy resolution considered above but also by energy losses in the dead layer, in particular if backscatter processes occur. An optimization of the FPD performance thus requires careful modelling of electron backscattering. This process was investigated for different electron impact angles by using a MC code, which combines a theoretical model for backscatter probabilities with experimental data for backscatter energies and angles. The model of [120] was used to calculate backscattering probabilities $P_\mathrm{B}(E_0, \alpha)$ with $P_\mathrm{B}(19\,\mathrm{keV}, 0°) = 0.18$, with α being the angle of incidence with respect to the detector surface. Then experimental data [121] for the distribution $f(E_\mathrm{back}, \alpha_\mathrm{back})$ of the backscattered electrons in energy E_back and backscattering angle α_back were used, yielding a mean energy $\langle E \rangle_\mathrm{back} = 0.6\,E_0$ and a mean angle of 45° for the backscattered electron.

Electromagnetic particle tracking demonstrates that most of the backscattered electrons are reflected back onto the detector so that they pass the dead layer at least twice more. Thus they deposit less energy in the sensitive bulk material of the detector, resulting in a shift of the full energy peak. The energy losses of backscattered electrons in the dead layer are larger than average due to their larger angles of incidence, and, owing to their lower energy, their larger differential energy loss dE/dx.

The results of the MC simulations for the response functions of different types of detectors with various dead layer parameters are shown in figure 81 for mono-energetic electrons ($\mathrm{E_0} = 19\,\mathrm{keV}$). While the shape of the detector response function is dominated by the energy resolution, the dead layer and electron backscattering visibly affect the response function. If the energy losses in the dead layer exceed the value of the energy resolution

of the detector, a low energy peak of backscattered electrons is visible. Otherwise the dead layer manifests itself as low energy tail. This behavior has also been observed in the Mainz and Troitsk experiments. For an efficient detection of the backscattered electrons a detection threshold for electrons of $E_{thr} > 5$ keV should be achieved. Furthermore, it is dominantly the low energy tail, and thus the detector dead layer rather than electronic noise, which defines the width of region of interest, e.g. to achieve a detection efficiency of $\varepsilon = 0.95$.

The size and the shape of the low energy tail of the detector response function have to be known with high precision to reliably calculate the efficiency of an energy window of interest for signal electrons. For the FPD, the lower boundary of this window has to be adjusted on the basis of the MC code so that the required efficiency of the energy cut (95% for the reference design) is guaranteed.

6.2.4 Post-Acceleration

An important issue for the optimization of the detector response function is the post-acceleration of electrons prior to their impact on the detector. As shown in section 6.4, the post-acceleration of electrons by up to 30 keV is an option to reduce the level of background induced by β- and γ- emitting nuclei close to the detector. Post-acceleration of tritium β–electrons also has the beneficial result of reducing the backscatter effects due to high angles of incidence and thus allows to operate the electron detector in higher magnetic fields. As already mentioned, a higher magnetic field at the detector decreases the magnetic flux and thus the required sensitive detector area. Again, this would lower the intrinsic detector background. Finally, post-acceleration of particles is equivalent to lowering the detection threshold for electrons, which is of great interest for the study of systematic effects with low-energy electrons at the source or spectrometer. Monte Carlo simulations have been performed to estimate the background reduction by the usage of post-acceleration. As input the Monte Carlo uses the simulation of the detector response function (see sec. 11.4.2) as well as GEANT4 simulations on the expected cosmic and environmental background. According these simulations a post acceleration of $U \approx 15\,\mathrm{keV}$ with the detector in a magnetic field of B=4 T is necessary to obtain a background reduction by a factor 4-5 in comparison to the reference design (B=3 T, no post acceleration).

However, the technical realization is non-trivial under the aspect of the involved high magnetic and electric fields as well as the space restriction. Large $E \times B$ values can cause discharges due to vacuum breakdown mechanism. Furthermore, when the magnetic field and the electrical field are not properly aligned Penning traps can build up. Currently, the application is investigated in detail to prevent these effects. Therefore, post acceleration of the electrons is at the moment treated as an option in the design of the focal plane detector, i.e. the detector system shall meet all named requirements also without the post acceleration mechanism.

To implement the concept of post-acceleration, the detector section including the read-out electronics and its interconnection to the data acquisition system must work under high voltage bias. A detailed description of the overall concept of the electronics using

fiber optics for galvanic separation of frontend electronics and data acquisition is given in section 6.3.

6.2.5 Timing

The intrinsic timing ΔT_{det} of the FPD has to be short enough so that no pile-up effects deteriorate the measurements at maximum signal rate. The normal tritium mode with an event rate of a few up to several hundred of mHz (depending on the actual setting of the retarding voltage) requires only a moderate FPD timing of ΔT_{det} <100 μs. In case of a segmentation into smaller sub-units, e.g. 400 pixels, this value for ΔT_{det} could be further relaxed to 40 ms.

However, the highest signal rate can be expected for the short yet frequent runs with the high-intensity electron gun to determine the actual value of the WGTS column density $\rho\, d$ precisely (see section 11.4.3). During these short measurements the rate of signal electrons increases to a maximum value of 1 MHz, distributed over the entire guided magnetic flux of 192 T $\times$ cm^2.

The time precision of the detector system should be in the range $\delta t < 500$ ns. This ensures to correlate events in the active veto system to events in the focal plane detector.

The time information is also used to check the stability of the apparatus. It enables one to inspect the time difference between two events and to compare it with the mean time expected from average count rate. This check is useful in the search for bunches of events arising from charged particles in magnetic traps.

6.2.6 Ultra high vacuum conditions

An important further demand is that the FPD can be operated under ultra high vacuum conditions (p $< 10^{-10}$ mbar). This imposes stringent upper limits on the tolerable outgassing rate of the detector and electronics material. Especially materials commonly used in detector mounting systems like glue and solders have to be reviewed on their compatibility with the aimed for vacuum pressure. The maintenance of UHV conditions also requires minimized leakage rates of the signal cable feedthroughs as well as installation of a differential vacuum pumping between detector and main spectrometer, where extreme high vacuum conditions are obtained.

6.2.7 Intrinsic detector background

An important design criterion of the FPD is its intrinsic background rate, resulting from decays of β and γ-unstable isotopes in the detector itself and from its close surroundings. This rate should be low enough not to deteriorate the ν-mass sensitivity of KATRIN. This goal is met if the intrinsic detector background rate is $R_{BG} \leq$ 1mHz, corresponding to less than 10 % of the overall reference background rate of 10 mHz (see section 6.4).

	Performance goal
Electron energies E	$5 - 50\,\mathrm{keV}$
Detection threshold	$5\,\mathrm{keV}$
Sensitive area A	$> 8 \times 10^3\,\mathrm{mm}^2$
Detection efficiency ε	> 0.9
Energy resolution $\Delta E_{Det.}$	< 600 eV
Position resolution $(\delta x \cdot \delta y)$	$0.3\,\mathrm{mm} < \delta x, \delta y < 5$ mm
Time resolution	$\delta t < 0.5\,\mu s$
Detector thickness	$d = 150 - 300\,\mu m$
Dead layer entrance window	$\lambda < 50$ nm (Si)
Operation in magnetic fields	2.5-4 T
Outgassing rate	UHV compatibility
ENC of electronics	< 68 e$^-$(rms)

Table 2: Summary of detector requirements.

6.3 Detector options

The basic design requirements for the KATRIN detector, which are listed in table 2, can be met, in principle, by several different detector techniques. In the following, semiconductor based detectors and bolometers are compared with regard to their performance (energy, time and position resolution, dead layer) as well as their technical feasibility (e.g. can a large array with minimized insensitive areas be realized in the near future).

At first conventional detector types with moderate resolution such as PIN-diode arrays and multi-element avalanche photodiodes are specified. These detectors can reliably be upscaled to the KATRIN geometry. Then detector technologies offering superior resolution like silicon drift diodes (SDD's), DEPFETS and bolometers are discussed. However, in general these advanced detector types require significant R&D work for upscaling to a larger size.

6.3.1 Monolithic PIN-Diode arrays

A well-established and rather economic detector technique for measurements of low-energy electrons are PIN diodes, to be more specific diodes with similar doping structure. Arrays of PIN-diodes can be manufactured in monolithic style with a large range of sizes feasible, ranging from a few μm (pixel detectors) up to several mm for a single diode. The geometric arrangement of the single elements is flexible, so that many different detector shapes can be implemented. At present, monolithic PIN-diode arrays with pixel sizes in the range of several mm^2 e.g. $5 \times 5\,\mathrm{mm}^2$ and total array sizes in the range of several tens of cm^2 e.g. $5 \times 5\,\mathrm{cm}^2$ are used in hybrid systems for the read-out of scintillator crystal arrays for X-ray spectroscopy [122]. PIN-diode arrays are available from several industrial manufacturers. An important advantage of monolithic diode arrays in comparison to arrays of independent single diodes is the common, uniform entrance window with minimized insensitive areas.

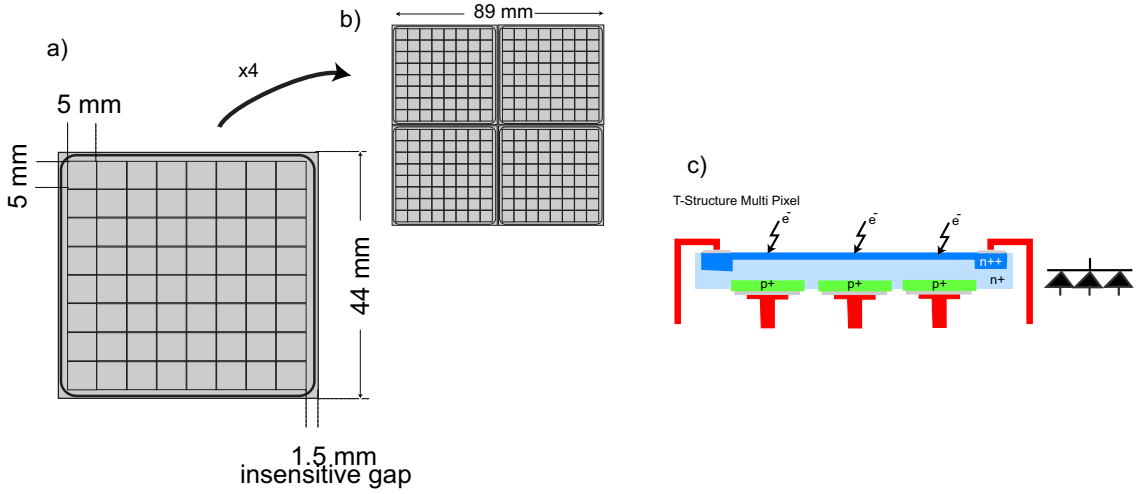

Figure 82: a)Schematic drawing of a commercial PIN-diode array with 64 pixels. b) Possible arrangement of 4 detectors for a final KATRIN design. c) Schematic cross section of the diode array.

The detector is fixed on a board, which carries on its backside transistors (FETs) as the first stage of the analog readout electronics. The board is enclosed by a copper ring, which is used for the cooling of the detector and the FETs. The cooling system will allow the free choice of operating temperature down to lN_2–temperatures. The signals of the detector channels are fed from the vacuum to the ambient side by a matrix of spring loaded metal pins. The pins are integrated into either a glass or a ceramic substrate. The use of spring loaded metal pins for the connection to the electronic board minimizes the use of soldering material for background reasons. On the ambient side, directly behind the end flange of the beam-pipe the next stage of the analog electronics with the preamplifiers is situated.

Figure 82 shows the design of a 8×8 array with $5 \times 5\,\mathrm{mm}^2$ pixels, which will be tested at the pre-spectrometer and is used as testbed to face also the detector technology independent challenges like UHV compatibility, space restriction and so on.

The schematic cross section in fig. 82(c) shows that, strictly speaking, the diodes are not PIN type as their bulk is already weakly n-doped. The pixelated side will be mounted directly on a printed circuit board. The connection from the pixels to the FETs on the electronic board, will be made by bonding directly to the p-doped pixels through the board. The entrance side of the detector is the heavily n-doped side of the wafer. The thickness of the dead layer in the detector chip amounts to 100 nm. Depending on the architecture of the diode this value can be reduced to 50 nm. Insensitive areas of the detector occur only due to a guard ring at the outer edge of the detector chip (see fig. 83) In an quadratic arrangement of four PIN-diode arrays a sensitive area of $80\,\mathrm{cm}^2$ would be covered while insensitive areas in the range of $5 - 10\%$ occur. This ratio can be improved by larger detector chips, which are technically feasible. Typical leakage currents for such detectors are between 0.1 and 1.5 nA per cm^2 and per $100\,\mu$m detector thickness. With

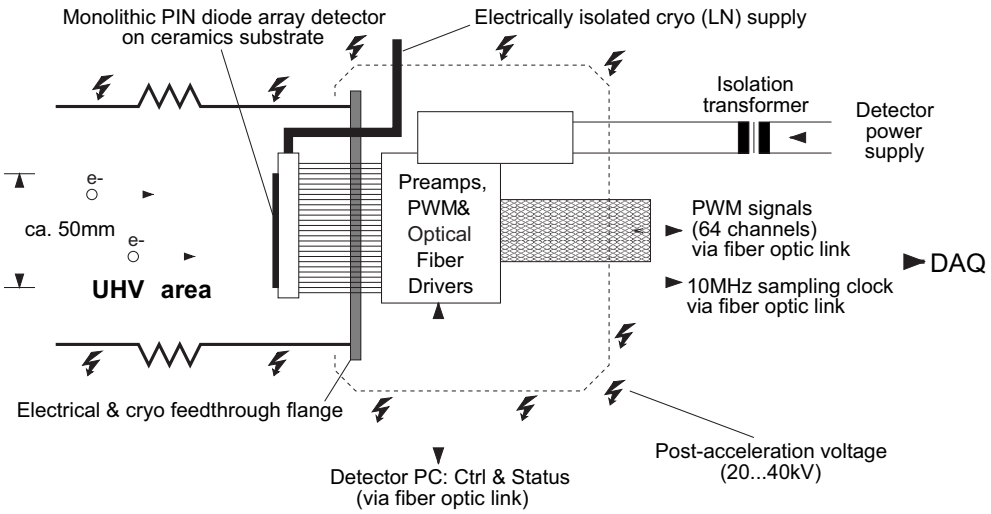

Figure 83: Concept of electronic readout

the help of a detector cooling system, which ranges down to lN_2–temperatures, any noise contribution from leakage current are eliminated. The energy resolution of the detector is then limited by the noise contribution of the analog readout circuit.

Electronics

The electron detector for the pre-spectrometer will also serve as a prototype for the electronics. The electron detector for the pre-spectrometer is a special monolithic 8x8-element silicon PIN-diode manufactured by CANBERRA, Olen, Belgium. Each pixel has an area of 5x5mm^2, the whole chip is 44x44mm^2. Each diode element has a junction capacitance of 16pF when fully depleted. The recommended reverse bias voltage needed to deplete the diodes is 20V. The readout electronics of the pre-spectrometer detector could be used later in the main spectrometer by increasing the number of channels. The detector chip is mounted on a Al_2O_3 ceramic substrate which also carries the first amplifier stage JFETs. The JFETs are operated in source follower configuration (see Front End Design Considerations below). The ceramics substrate is mounted to a Cu cooling ring which is cooled by lN_2 and can be reheated electrically in case the lN_2 cooling is too strong. The JFETs can operate down to $-100°C$, but for a sufficient reduction of detector leakage current $-50°C$ at the detector chip will be enough. The JFET output signals are brought out of the vacuum area by spring-loaded needles located on the pins of a glass-metal feedthrough flange. The needle points touch contact pads on the ceramic substrate that carries the detector chip. In order to avoid more than one feedthrough pin per detector element/channel, no charge feedback to the JFET gates is possible. Therefore the noise advantage of the common source configuration (which needs feedback) was traded off against the stable gain and low output impedance of the source follower configuration, which needs no feedback (see Front End Design Considerations below). Outside the

vacuum area, further amplification of the detector signal of about 60 dB is needed. This is performed by a multi-stage preamplifier. Low-gain DC coupling allows monitoring of the leakage current of individual detector elements. In order to allow post-acceleration of the electrons before striking the detector, the detector electronics have to be galvanically isolated from ground. The post-acceleration voltage is envisaged to be up to ≈ 30 kV. As standard photocouplers are only available with isolating voltages of up to 15 kV, plastic optical fibers are used to obtain HV isolation. The optical separation is also advantageous to avoid EMC (electro-magnetic compatibility) problems. To become independent from LED photoemitter ageing and nonlinearities, pulse width modulation (PWM) is used for transmission via the optical fibers. This implies that the sampling of the analog detector signal has to be done in the front end electronics before the PWM. The analog detector signal leaving the multi-stage preamplifier is integrated and sampled at a rate of 10MHz which has to be phase-locked to the sampling clock of the ADCs in the DAQ rack. An additional optical fiber link is needed for this. The PWM signals of the individual channels are then transmitted via one fiber each by IR LED photoemitters. On the receiving end of the optical fibers, the PWM is demodulated into analog voltage samples which are then fed to the ADCs. It is planned to use three ADC board assemblies from the Auger neutrino experiment, with 22 channels each. The analog sub boards of these units have to be modified accordingly. In order to accommodate the fiber optic connectors, additional connector boards may become necessary. Power to the detector electronics is supplied via a HV isolation transformer. A shielding cage enclosing the detector electronics is connected to the same electric potential as the detector (i.e. the post-acceleration voltage) to prevent dust from collecting within the electronics. The optical link components are standard consumer digital audio parts, which greatly reduces the costs of the post-acceleration option. All parts of the detector frontend electronics (ceramics substrate for the detector, preamplifiers, optical link) will be designed and manufactured by FZK IPE. For the first pre-spectrometer measurements, a modified variant of the preamplifiers has been designed, which has 50 Ohm coaxial cable outputs connecting to the existing ORCA shaper subrack (from SNO experiment, University of Washington at Seattle).

Front End Design Considerations

Available first stage JFETs exhibit input related noise voltages of $1 \ldots 5\,\mathrm{nV}/\sqrt{Hz}$, even in types with gate capacitances larger than 16 pF (the optimum gate capacitance for 16 pF of detector junction capacitance). In a source follower configuration, a bipolar second stage with an even lower voltage noise is needed in order not to deteriorate the noise performance. Although it was not possible to find such transistors/integrated amplifiers, the noise advantage of the common source configuration was set aside in favor of a range of other advantages of the source follower configuration. This configuration offers a stable (unity) gain without needing feedback (extra feedthrough pins) and at the same time has a low output impedance. This low output impedance is essential in the intended setup due to its insensitivity to crosstalk and capacitive loading - potential pitfalls when having to feed a high number of very small signals out of the vacuum. The source fol-

lower front end provides high bandwidth and stability almost independent of the length of the feedthrough lines, even in the case when flexible cable extensions between detector ceramics and feedthrough flange are needed (University of Washington at Seattle electron gun with movable/tiltable detector for dead layer measurements). First measurements showed that not amplifier noise but microphonics is the limiting factor for energy resolution. Therefore, further design effort will be put into minimizing microphonics rather than amplifier noise.

6.3.2 Avalanche Photodiodes (APD)

Avalanche photodiodes are comparable to PIN diodes with regard to the energy resolution and their typical sizes. They have an internal gain of up to 10^4, and are thus used for the detection of small signals. Recently, large area APDs (diameters up to 20 mm) and segmented APDs have been developed. APD pixels require a $0.5 - 1.0$ mm annular guard ring, which would lead to $10 - 20$ % dead space. However, it is possible to read out larger pixels using 4 corner contacts and one common contact which results, by charge division, in position resolutions of $\Delta x, \Delta y < 1$ mm. Thus to satisfy KATRIN one could envision a 100 mm x 100 mm wafer divided into 16 pixels, each 25 mm x 25 mm. Though it is not expected from this technique to improve significantly the benchmarks given by a PIN diode design, APDs are distinct from PIN diodes by their small active region. The diodes are sensitive to radiation in a $20 - 30\mu$m thick region. Thus, the expected γ– background rate is one order of magnitude lower, when compared to PIN diodes.

6.3.3 Monolithic SDD arrays

Silicon drift diodes (SDDs) [123] are high resolution room temperature detectors which are used in X-ray spectroscopy [124] and as tracking devices in high energy physics [125]. This detector type achieves a good performance even at very high count rates. The SDD consists of a volume of fully depleted silicon in which an electric field with a strong component parallel to the surfaces drives the signal electrons towards a small size collecting anode with an extremely small anode capacitance. For the counting of β-electrons one has to select SDD-types with entrance windows on uniform electrical potential to avoid position dependent detection efficiencies. For commercially available single SDD's an energy resolution of the order of 150 eV (FWHM) or better has been achieved at the Mn Kα energy of E=5.9 keV. This is already close to the Fano noise for a 5.9 keV energy deposit in silicon which corresponds to 120 eV.

The instrumentation of the entire KATRIN flux tube with an area of $\sim 10^4$ mm^2 requires the implementation of a monolithic multi-element SDD array. With individual SDD areas of 5×5 mm^2 this would correspond to an array 400 single SDD elements. Up to now, the largest monolithic array in routine operation consists of 7 SDDs with single sensitive areas of 5mm^2 [126]. Presently, a monolithic 61-element SDD array is under development in an European project [127]. The construction of a significantly larger monolithic SDD array with several hundred elements would require a substantial R&D

effort. The option of a monolithic SDD-array will be further investigated by a feasibility study in close collaboration with an industrial partner. First tests at FZK with a single SDD element are reported below.

6.3.4 DEPFET pixel matrices

The DEpleted P-channel Field Effect Transistor (DEPFET) [128] is a low noise pixel detector developed at the MPI Halbleiterlabor München with the first stage amplifier integrated into the detector. Typical pixel sizes are of the order of $50 - 150\,\mu$m. At the University of Bonn a 64×64 pixel matrix has been developed [129], which operates as Bioscope-System used for biomedical applications using samples with tritium markers. Measurements with a DEPFET pixel show a very good intrinsic energy resolution of 158 eV (FWHM) at $E_\gamma = 6$ keV [129], equivalent to an electronic noise contribution of ENC $= (12 \pm 1)\,\mathrm{e}^-$. The DEPFET principle allows novel pixel detectors to be built with large areas for applications in high energy physics experiments and X-ray astronomy [130, 131]. Due to the small capacitance of each pixel these devices could also be produced much thinner than the current standard value of $300\,\mu m$. One of the challenges of this technology, if applied and scaled to the β-counting requirements of KATRIN, would be the read-out scheme for the $\sim 10^4$ channels per cm^2 and much more the handling of MHz rates during calibration measurements.

6.3.5 Bolometers

Thermal microcalorimeters (bolometers) offer the possibility to achieve a superior energy resolution of about 20-30 eV for β-electrons in the energy range above a few keV. This type of detector is in use for neutrinoless double beta decay searches [132] ((Mi-Beta [133], Cuoricino [134], Cuore [57]), dark matter (WIMP) searches and β-decay spectroscopy to measure the electron neutrino mass (Munu2 [135]). The absorber material of the bolometers can be a low-Z material which minimizes background from environmental gammas. A feasible absorber material would be Be.

There are two types of semiconductor based thermistors which can be used as sensors for the read-out of microcalorimeters in strong magnetic fields of several Tesla: a) silicon implanted and b) NTD germanium thermistors. Silicon implanted thermistors have been developed in the framework of $0\nu\beta\beta$ searches [136]. Their major advantages are reproducibility and the possibility of large scale production. Work is in progress to implement micromachining techniques to build fully integrated thermistors. For KATRIN, a thin Si-implanted thermistor would have to be used, with the absorber dominating the heat capacity. The second type of thermistors, so-called NTD-germanium thermistors, will be used for the $0\nu\beta\beta$-experiments Cuoricino and CUORE. Their production process is however rather sophisticated. The characteristics of this thermistor type are described in more detail in [137].

The implementation of a bolometer array would be technologically challenging, given the requirements of microcalorimeters with regard to their operating temperatures in the

mK range. A dilution refrigerator with horizontal cold finger would provide the required operational temperature, however, the 'open' geometry required for the counting of the β-electrons would be challenging, taking into account the incoming radiation flux from the main spectrometer which is operated at a temperature of 250 K. A suppression of this heat load to the bolometer would require at least two additional transport elements with the lHe bore. Nevertheless, a bolometric read-out for KATRIN seems to be feasible. In particular, one could make ideal use of existing thermistor techniques, which have been developed for double beta decay searches [57], as well as of recently reported improvements with regard to noise suppression [138] and read-out stabilization [139].

6.4 Detector background investigations

The detector background is defined as the remnant background, when there is no direct connection between the detector and the main spectrometer. This background comes from the following components:

- Background due to radioactive contaminations in the materials used, in particular due to events from the ^{238}U and ^{232}Th decay chain.
- Cosmic induced background from prompt events (e.g. muons, neutrons) and from events arising from activated long-living products (cosmogenics).

Detailed Monte Carlo Simulations within the framework of GEANT 4 [140] have been performed [119] using a 3-dimensional working model of the detector and its environment to obtain a reliable estimate with regard to these background components.

The simulation results, i.e. the rate of background events depositing energy in the region of interest, rely on the contamination levels used for the different radioactive isotopes, contamination levels of cosmogenic products and on the flux of cosmic muons. For the flux of muons, data from [141] has been used, whereas for the contamination levels of cosmogenics the COSMO[142] code has been applied. For contamination levels of ^{238}U, ^{232}Th, ^{40}K, ^{210}Pb, ^{210}Po and ^{137}Cs, a literature study [143, 144, 145, 146, 147, 148, 149] on their abundances in the different materials of the detector components has been performed. In total, 33 different radioactive isotopes including their complete decay chains, started in 16 different materials of the detector model, have been simulated, leading to over 900 single simulations of different radioactive elements and host materials.

The simulations clearly identify the critical components in the preliminary detector setup. Using the average values of contamination levels determined by the literature study, the simulations find the ceramic carrier of the silicon detector, with the installed electronics, as the main background source. Depending on the energy region of interest, up to 96 % of all background events originate from contaminations located in the ceramics. Thus, a more careful design of the detector holder is mandatory. Moreover, the simulations strongly suggest that materials and detector components should be subject to a low level survey before using them in the setup. On the other hand, if one succeeds with a careful selection of ceramic parts and electronics including suitable solder materials, the major

background then arises from cosmic muons and neutrons. Thus, neutron shielding and an active muon veto is foreseen in the detector setup. Figure 84 summarizes in a breakdown

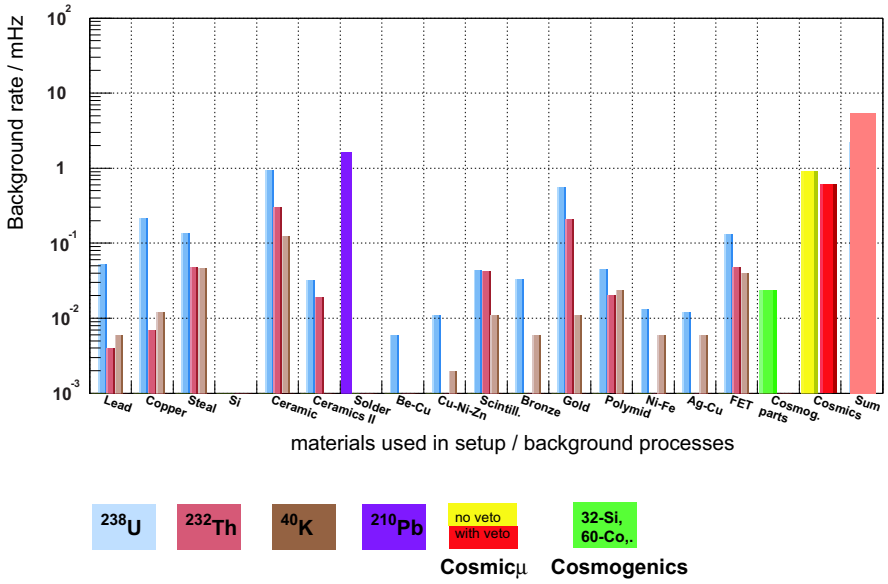

Figure 84: Breakdown of background contributions (3σ upper limits), originating from different materials and physics processes used in the simulated detector setup. The simulation assumed a post acceleration of electrons by $U_{bias} = +30\,\mathrm{kV}$.

chart the detector components contributing to the background.

The simulations were done for various assumptions on the contamination levels, as well as for different strengths of a possible post acceleration of the signal electrons. The analysis, which uses the average values of the contamination levels, determines as statistical upper limit (3σ)

$$R < 5.5\,\mathrm{mHz} \tag{48}$$

for the expected background rate R, if one uses an energy region of interest, corresponding to a post acceleration by $U_{Bias} = +30\,\mathrm{kV}$. The systematic uncertainty of the result amounts to a factor of 2. Figure 85 shows the simulated background spectrum. The spectrum is composed of a continuous background increasing towards lower energies superimposed by lines originating from atomic shell processes and some nuclear decays (e.g the 46 keV line from ^{210}Pb).

The results of the MC simulations show that a background rate in the required range below 1 mHz can only be achieved after careful selection of materials, in particular of the carrier ceramics, the nearby electronics components and the soldering. Apart from passive shielding, consisting of an inner passive copper (OFHC free) tube and an outer 10-15 cm thick low-level lead shielding, the active vetoing of cosmic muons by anti-coincidence counters is required.

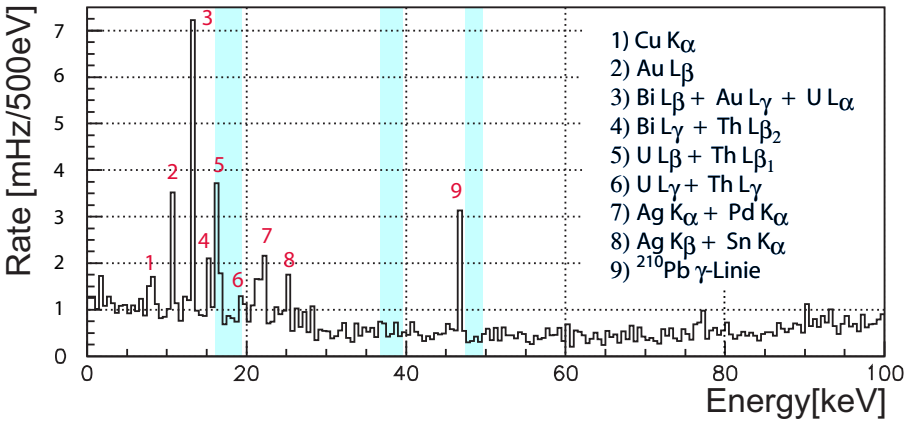

Figure 85: Simulated detector background spectrum with identified X-rays, respectively gamma lines. The blue areas show different energy regions of interest in the case of post acceleration (no post acceleration, +20 keV, +30 keV). The different regions of interest yield as $3\,\sigma$ upper limits on the detector background rate $R < 15.4; R < 7.4; R < 5.5$ mHz.

7 Magnet- and Vacuum-System

The KATRIN beam line comprises about 30 superconducting magnets with a magnetic field strength ranging from 3–6 T which adiabatically guide the β-decay electrons from the WGTS to the detector. In addition, stringent ultra high vacuum conditions have to be fulfilled for the tandem electrostatic spectrometer system. In the following we discuss the electromagnetic design considerations for the magnet system and the spectrometers and give a short review of the required magnet technology and cryo supply. Finally we discuss the challenging vacuum aspects of KATRIN associated with achieving an UHV in large volume recipients.

7.1 Magnet system and cryogenic supply

The magnet and cryo system of KATRIN has been investigated with respect to design, operation, operational safety questions and economic optimization. This process has first results with an industrial study [150], the order of two pre-spectrometer solenoids for early tests of the pre-spectrometer, and the selection of the cooling concept for the KATRIN magnets. The magnetic fields required for electron transport and spectrometer resolution have a level between 5 and 6 T and must be constant with time over months. Such fields can reasonably be generated only with superconducting magnets. In the aforementioned study a first attempt has been made for the design of the complete magnet system. Based on this first result, further studies and optimization calculations with respect to technology and economics need to be performed, taking into account the requirements resulting from electron transport, particle trapping, tritium source, pumping ports, cryopumping section and beam tube temperature variations. The main winding geometry is simply solenoidal, but field correction coils are needed for the enhancement of the field level between any two transport magnets in order to avoid strong magnetic flux line opening guiding electrons to hit the walls.

A further requirement originates from the need of long term current and field stability, with $\leq 2\%$ field drop over 3 months required, leading us to use a low operating current and a superconducting permanent current switch. The low current has the further advantage of low losses in the current leads (from ambient temperature down to operating temperature), but also the disadvantage of high voltages and high hot spot temperatures in the windings during quench, unless special measures of subdivision of the magnet coils and shunting with diodes and resistors are taken. The fail-safe operation of all magnets needs therefore to be tested in a special test bed (KAMAT, KATRIN magnet test stand) at the Institute for Technical Physics (ITP) well ahead of integration into the complete KATRIN system. The current leads should apply a HTS (High Temperature Superconductor) section for the reduction of heat load onto the cold part of the magnets, but also conventional technology can be applied. The heat input through the leads is lowered also in the ON state of the integrated permanent current switch and zero current flow in the leads. A retractable current lead would further reduce the heat input, but has disadvantages in operation in view of the great number of magnets when any change of current needs much handling

Central magnetic field	4.5 T
Maximum magnetic field at the winding	5.6 T
Winding inner diam.	0.48 m
Length of winding pack	0.32 m
Operating current	156 A
Persistent mode stability	5×10^{-6}/h
Cryostat warm bore diameter	0.40 m
Cryostat horizontal overall length	0.56 m

Table 3: Pre-spectrometer magnet parameters

effort, and is therefore excluded.

The choice of the superconductor is dominated by economical assessments. Due to the high fields involved, HTS could only be applied at very low operating temperatures near that of liquid helium, a temperature well suited for the standard NbTi multifilament wires, which are much cheaper than the HTS. Therefore, NbTi and helium cooling will be used for all magnets with central fields of 3.6 or 5.6 T. The pinch magnet will have a maximum field of 10 T to allow special calibration measurements with increased energy resolution compared to the standard value of 0.93 eV. The parameters of the pre-spectrometer magnets have been fixed and are listed in table 3.

The conceivable concepts for cooling the magnet system have been studied and evaluated. The tasks of the cryogenic supply are the controlled cool-down of the superconducting magnets and the tritium tube and the compensation of the heat loads during the run-time of the experiment. The appropriate cooling concept should ensure an economic, safe and continuous operation over a period of time of several years. Due to the increase of the efficiency of refrigerators and liquefiers with the refrigeration power or the liquefaction rate, respectively, and due to the underproportional scaling of the capital costs with the refrigeration power ($\frac{d\,cost}{d\,P_{refr}} < 1$), a single fully automatic standard refrigerator for the whole magnet system and the tritium tube is favored. A rough estimate of the required refrigeration power leads to a plant with a refrigeration power in the order of 300 W. Since the WGTS has special requirements concerning the stability of the temperatures a special cryostat will be developed to ensure a temperature stability of better 0.1 % in the 30 K region. This cryostat requires a pressure controlled vessel with liquid neon which has a boiling point in this temperature range. An additional supply of helium at 27 K is needed to condensate the Neon gas. Further requirements lead to a cryogenic plant shown schematically in figure 86.

Before the magnet system and the tritium tube can be cooled down, the impurities like moisture, oxygen and nitrogen have to be removed from all He-lines and cryostats. The helium-purifier existing in the ITP can be applied for this purpose. In addition, the gaseous helium storage system of the ITP can be made available. Mainly two cooling concepts have been investigated, on the one hand helium bath cooling and on the other hand forced flow cooling. For the magnet system of KATRIN, helium bath cooling offers

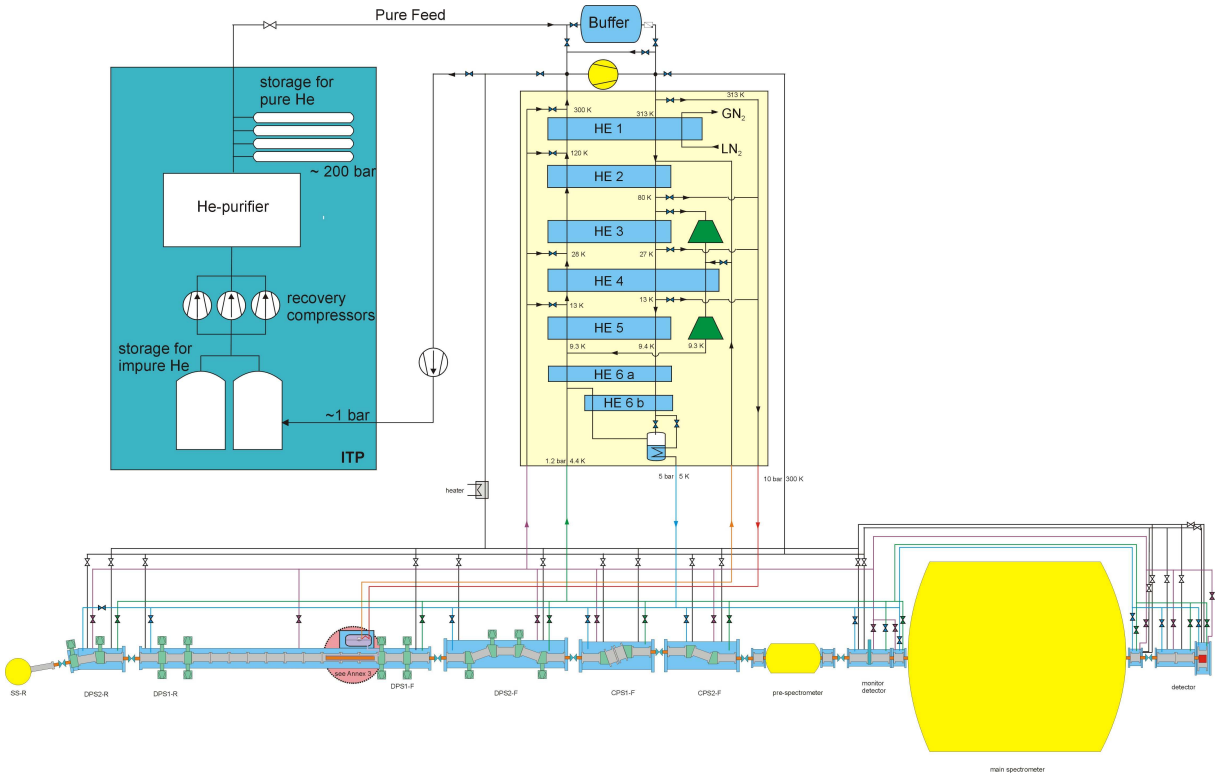

Figure 86: Schematic view of the cryo supply system for the KATRIN experiment

several advantages. For example, the bath cooling concept implies liquid helium storages within the magnet cryostats. In case of an interruption of the cryogenic supply, this liquid helium storage gives time to initiate appropriate measures to avoid an untimely warm-up of the magnets and the tritium tube. Detailed studies and investigations of both magnet and cryo systems are foreseen in order to enable the selection of optimized solutions.

An exception to this cooling concept are the two pre-spectrometer magnets which are also used for electromagnetic tests of the pre-spectrometer. The two magnets are cryogen free systems since the current pre-spectrometer experimental hall has no helium recovery infrastructure. Figure 87 shows the magnets at the force test in KAMAT. Currently, the DPS2-F is ordered and in the production phase (see figure 88 for a design drawing as well as sec. 4.2.1 for more details), the order of the WGTS is placed and the contract was signed in December 2004.

7.2 Vacuum systems

The vacuum system of the entire KATRIN beam line can be subdivided into the following three subsystems:

- The WGTS and the differential pumping stages,
- the cryogenic pumping section CPS,

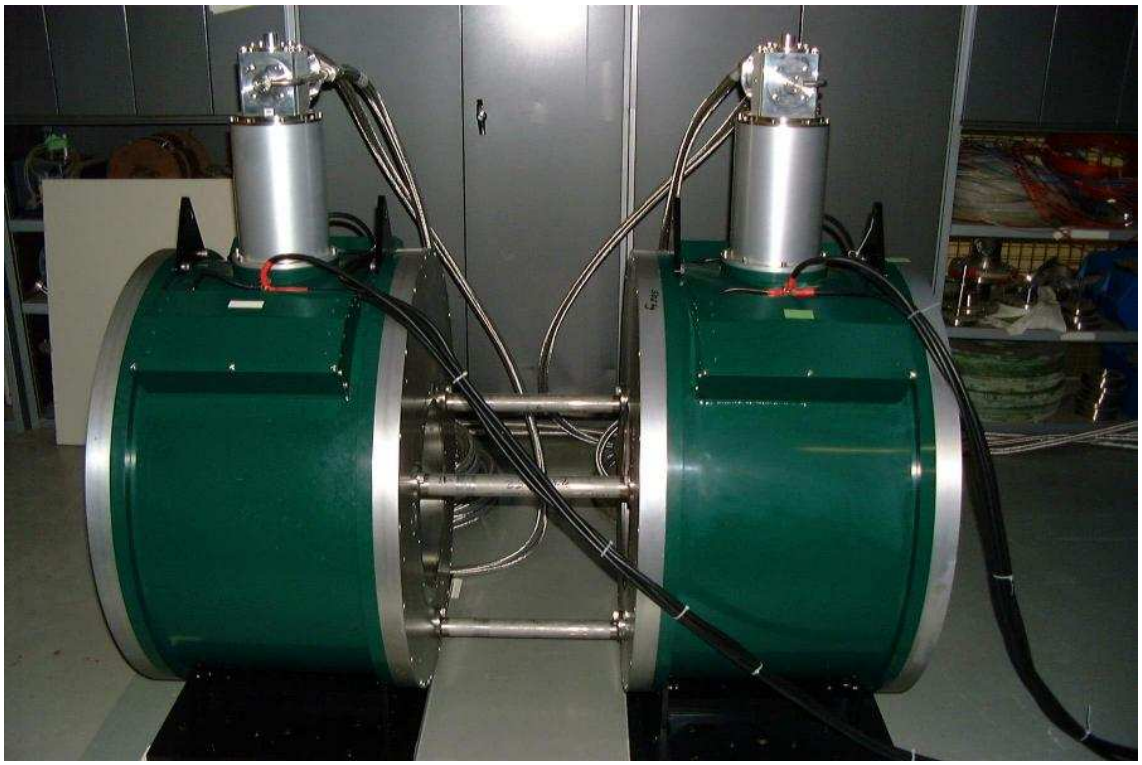

Figure 87: The two pre-spectrometer magnets at the force test. In normal operation mode attractive forces up to 100 kN are acting on the magnets.

- the XHV (Extreme High Vacuum) system of the pre- and main spectrometer.

In each system, the following items need to be considered:

- the materials used in constructing the vacuum chambers,
- the construction techniques to be employed,
- the surface preparation techniques to be used,
- the amount of vacuum pumping (usually referred to as pumping speed) and its location required to meet the necessary vacuum levels,
- the types of pump to be used and any limitations to be taken into account,
- the measurement of total and partial pressures.

In KATRIN, each of the three vacuum subsystems has to fulfill specific tasks and faces technological challenges which will be outlined in more detail below. The basis of the KATRIN vacuum system has been comprehensively discussed and agreed to at a dedicated workshop [151].

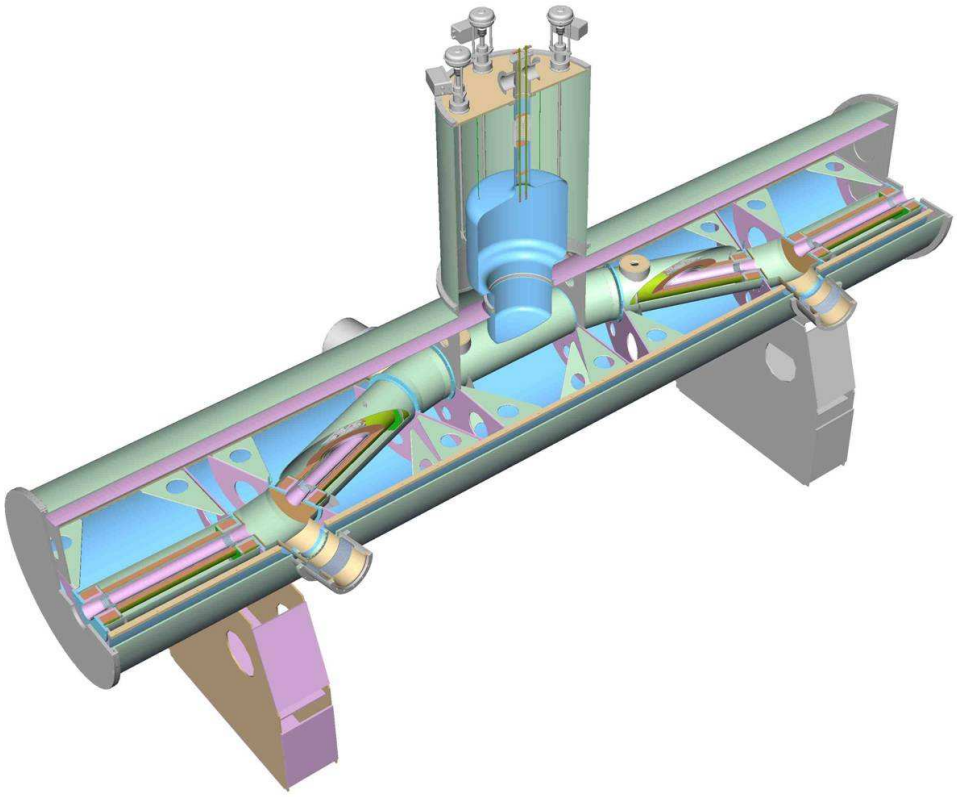

Figure 88: Isometric view of the DPS2-F. The rigid cold mass is required to withstand the forces and torques resulting from the nonlinear arrangement and neighboring magnet systems

7.2.1 The WGTS, differential and cryogenic pumping subsystems

Differential pumping systems are located both at the rear and front end of the tritium tube of the WGTS (see section 4.1). The cryopumping section (CPS1 and CPS2 divided by a gate valve, see section 4.2.2) is located as last pumping section directly upstream of the pre-spectrometer.

These systems have been discussed in some detail earlier, but the overall description and performance is repeated here for completeness. The differential pumping section combines two sub-structures. First, it comprises a loop system, which has to be designed so that molecular tritium of very high purity is circulating under stable conditions in the WGTS. Second, a system of cascaded pumps has to provide further reduction of tritium partial pressure and recover the active gas. Of equal importance is the minimization of the outflow of tritium molecules after the last active pumping port into the cryogenic pumping sections and into the spectrometers.

The layout of the KATRIN vacuum system will be based on the design of the UHV systems in Mainz and Troitsk as proposed in [1]. Both experiments have also developed efficient cleaning procedures for their spectrometers in case of migration of T_2 molecules in trace amounts into the spectrometers. For the KATRIN system, the amount of T_2

present at a given point under standard operation is well known and it is anticipated that it will be continuously recorded in the active differential pumping system.

The differential pumping system must provide a tritium reduction factor of about 10^7 (a tritium flow reduction from 2.4 Ci/s at the WGTS injection tube down to 1 Ci/60 d). Assuming molecular flow downstream the injection tube, Monte Carlo simulation of the actual design result in a reduction factor of $2 \cdot 10^8$ [153]. In the case of an increase of the tritium flow to a non-acceptable level, a computer controlled guard gate valve in front of the cryogenic pumps will be closed. To prevent T_2 migration into the pre-spectrometer, the temperature inside the cryosection has to be kept below a well-defined value. In case of a warm up, the gate valve to the pre-spectrometer will be closed and the cryogenic pumps will be cleaned.

The cryogenic pumps that have to safely trap the remaining tritium flow after the active pumping section will contain two parts. The first containing the split coil magnet is a cryosorption pump using Ar frost at LHe temperatures (4.5 K) as the cryosorbent. The frost will be produced with a nozzle device. It is expected that practically all the remaining tritium molecules will end up incorporated in the Ar frost. The flow of tritium into CPS1 results in covering this part of the system with a few monolayers of T_2 at maximum (the maximum permissible total load is < 1 Ci, corresponding to about 1 monolayer on one m^2). The Ar frost will provide an effective surface much larger than the surface of the polished tube. The downstream end of the CPS1 section to the pre-spectrometer will be essentially free of tritium as long as the whole system is kept at the nominal operating temperature. After warming up, the T_2 will begin to migrate. To reduce migration as far as possible, the tritium will be blown out through the upstream end of the cryosorption section by a stream of He. The remaining traces of T_2 will then be removed by baking the system at 500 K.

In the current reference design of KATRIN, the second part of the cryogenic pumping section will be similar to the first one, but without the split coil magnet. However, consideration will be given to replacing the second cryogenic stage with a simple NEG coated tube, not covered with argon, as the downstream pumping systems of the spectrometers are not designed to pump the potential load of Ar. Operating this system under room temperature would have the advantage that once activated it is relatively fail safe. It is anticipated that over the lifetime of the experiment only a very small quantity of tritium will actually be pumped at this point, although this will need to be verified by calculation. CPS2 is thought to provide a safety function and is not planned to be regularly regenerated. This section will further reduce the flow of tritium.

The overall performance of these beam line vacuum stages must be such that the required partial pressure of tritium in the main spectrometer vessel must be kept below 10^{-20} mbar, which is well below the detection limit of the best quadrupole mass analysers.

The differential pumping stages will use large, tritium compatible magnetic bearing turbomolecular pumps. Where necessary these will be boosted by dry molecular drag pumps. Backing pumps will also be oil-free. Pressure will be measured by inverted magnetron gauges and quadrupole residual gas analysers will be available. Since the differential pumping stages are essentially cryogenic beam transport lines, they can be

manufactured by conventional techniques and much of the detailed design work has been carried out by industry.

Prepumping of the cryogenic stages will be similar to the pumping systems employed in the differential pumping stages. Pressure can only effectively be measured at the warm ends of these devices. Again, these stages are relatively conventional in construction and design will largely be by industry.

7.2.2 Spectrometer XHV system

Undoubtedly, the greatest challenge for the vacuum design is the large spectrometer vessel. The size of the vessel is such that manufacturing is beyond the scope of conventional UHV vacuum system specialists. Consequently, the task will have to be undertaken in a heavy industrial machine shop and the successful contractor will need to develop techniques for handling, assembly, cleaning and vacuum testing of such a vessel. However, the actual techniques which will need to be used are simply extensions of best practice found elsewhere. This leads to a high degree of confidence that a successful result will be obtained. Nevertheless, it will be imperative that a thorough degree of inspection, monitoring and control is implemented by KATRIN so that potential problems may be identified at an early stage.

Vacuum performance

The vacuum performance criteria are straightforward. The vessel must reach a pressure $< 10^{-11}$ mbar in a reasonable time in the first instance (*i.e.* weeks).

Since this is a static vacuum system (*i.e.* there is no external gas load and little stimulated desorption from the surfaces) the pressure achievable will be determined simply by thermal outgassing from the vacuum surfaces and the pumping speed applied. The outgassing will be dominated by the vessel walls. The total surface area of the two layer inner electrode system may sum up to about 10 % of the tank surface, so special consideration has to be given to sufficient cleaning in order to keep the outgassing contribution of the electrode system small.

With a total surface area of about $650\,\mathrm{m}^2$ and an anticipated outgassing rate of better than $10^{-13}\,\mathrm{mbar\,l\,s^{-1}\,cm^{-2}}$, a total gas load of around $10^{-6}\,\mathrm{mbar\,l/s}$ is expected. Therefore, reaching $10^{-11}\,\mathrm{mbar}$ requires an applied pumping speed of about $10^5\,\mathrm{l/s}$. This gas load will be almost completely hydrogen. Hence, gettering is a suitable pumping technique, using non evaporable getter strip. Non-getterable gases, e.g. methane and helium will be pumped by turbomolecular pumps. It is tried to achieve the target without ion pumps as much care would be needed to site these because of their magnetic fields. Pressure will be measured by inverted magnetron gauges, extractor gauges and residual gas analysers.

The main vessel

The main vessel will be manufactured from stainless steel sheet, type 1.4429 (316LN) selected both for its strength and excellent magnetic properties, especially in weld regions. There will be limits on the amount of cobalt permitted in the steel and on its residual radioactivity to minimize background in the spectrometer. The thickness will vary from about 25 mm to 32 mm. This means that welds must be built up from many passes. The vacuum sealing weld will be on the inside of the vessel and this will be subject to rigorous leak testing before the full thickness is built up. Detailed procedures are available for this and for other aspects of vacuum quality control [152].

The main pumping ports will terminate in DN 1700 flanges, similar to that on the pre-spectrometer, and will incorporate differentially pumped metal seals. The main axial beam tube flanges will be DN 500. There will also be many conventional CF type flanges on the vessel. The heating/cooling system will use oil and is again similar to that on the pre-spectrometer. In the initial phase of the project, cooling will not be implemented and the vessel will operate at room temperature.

Cleaning and surface preparation

In order to achieve the requisite vacuum performance, a careful cleaning and surface preparation protocol will have to be invoked. The size of the vessel is such that conventional techniques cannot be implemented, for example traditional solvent cleaning or alkaline bath degreasing. Additionally, since the surface has to be prepared in such a way as to minimize field emission, surface smoothing techniques, which are often not required for pure vacuum reasons, will need to be employed. In particular, electropolishing of the finished vessel will be specified. The regime to be used is therefore:

- hot rolled plates will be dry surface ground under clean conditions to a finish of approximately $R_a < 1\,\mu m$ surface roughness
- weld areas will be thoroughly degreased before welding
- weld seams will be ground to the same finish
- degreasing of the entire vessel by pressure washing (steam cleaning) with detergent and hot demineralized water
- electropolishing ($R_a < 0.6\,\mu m$)
- pressure spray washing with hot demineralized water
- drying in a dust free atmosphere
- final vacuum bake of the entire vessel to 350 °C

Vacuum Testing

As mentioned earlier, careful testing of the vacuum integrity of the vessel will be required. Each weld seam will be tested after the first pass sealing weld (on the internal surface of the vessel) to a leak rate of better than $1.0 \cdot 10^{-9}$ mbar l/s.

Following final assembly and cleaning, the entire vessel will be flanged up and pumped using clean, dry pumps. The entire vessel will be leak tested by flooding a suitable containment system around the vessel with helium before, during and after bake out. An integrated helium leak rate of $< 5 \cdot 10^{-9}$ mbar l/s (at room temperature) will be demonstrated in each case.

Before final acceptance of the vessel, a total integrated outgassing rate of $< 10^{-6}$ mbar l/s will be demonstrated and the overall cleanliness of the vessel shall be assessed using the ASTeC procedures [152].

8 Background

The sensitivity simulations for KATRIN (see sec. 12.1) are based on a gain in signal rate by more than a factor of 100 with respect to the existing spectrometers at Mainz and Troitsk but with the same background level which is of order 10 mHz at both instruments. The question arises, of course, whether this low level can be kept when the instrument is blown up by a factor of 10 in linear dimensions. Do we have to face a much higher background, or do we have measures to fight it back or even beat it below the present level?

At Mainz and Troitsk by far the largest fraction of the background rate originates from the spectrometers itself and is not background at the detector caused by its sensitivity to cosmic rays and environmental background. That the background of a MAC-E-Filter is not just scaling with the dimensions can be seen by the fact, that the Troitsk spectrometer provides about the same background rate as the Mainz spectrometer, although it is 4 times larger in volume. In addition, at Troitsk the residual gas pressure is 10 times higher. This shows, that the background rate also depends on other critical parameters, such as *e.g.* the shape and strength of the electric and magnetic fields. This fact and the detailed background investigations at Mainz and Troitsk under different conditions make us confident, that we will be able to reach about the same level of background rate of 10 mHz or below also with the large KATRIN spectrometer.

In this chapter we will discuss these questions based on the knowledge which has been gained over the years at Mainz and Troitsk and decisively enhanced by recent dedicated experiments and calculations. The chapter will be arranged in 4 sections, the first dealing with the influence of the background on the mass sensitivity, the second with a review on background phenomena observed at MAC-E-Filters, the third with recent experiments to actively reduce background further and the last with conclusions for KATRIN.

8.1 Background contribution to the statistical error of m_ν^2

Although the precise dependence of the mass sensitivity on background can be determined only by extensive simulations of spectra (see sec. 12.1), we can gain useful qualitative insight into the problem from some simple analytical formulae [154].

A MAC-E-Filter integrates the β spectrum above the analyzing energy qU. Hence the number of signal events within a time t for $m_\nu = 0$ is given to first order by

$$N_s(E) = r(E_0 - qU)^3 t\,. \tag{49}$$

A constant background rate b yields within the time t an event number N_b of

$$N_b = bt\,. \tag{50}$$

Let us assume for the moment that ${m_\nu}^2$ is the only unknown parameter. Then one could in principle reach from a single data point taken at the potential U a statistical sensitivity

on the mass squared of

$$\delta m_\nu^2 c^4 = \frac{2}{3}(E_0 - qU)^2 \frac{\sqrt{N_s + N_b}}{N_s} \tag{51}$$

which shows a minimum for

$$N_s(qU) = 2N_b\,. \tag{52}$$

Measured at the optimal point defined by eq. (52), we can now express δm_ν^2 as function of these parameters

$$\delta m_{\nu,min}^2 c^4 = \left(\frac{16}{27}\right)^{1/6} r^{-2/3} b^{1/6} t^{-1/2}\,. \tag{53}$$

We learn from Eq. (53) that the statistical uncertainty fortunately rises more slowly with the background rate than it decreases with the signal rate. Inserting for KATRIN $r \simeq 1.1 \cdot 10^{-3}$ Hz/eV^3 and assuming a background rate of 10 mHz the statistical sensitivity obtained from the optimum measuring point would be

$$\delta m_{\nu,min}^2 c^4 \approx \frac{40}{\sqrt{t/\mathrm{s}}}\ \ \mathrm{eV}^2\mathrm{s}\,. \tag{54}$$

Monte-Carlo simulations of experimental spectra confirm all functional dependencies of the above formulae. However, expecting a $\delta {m_\nu}^2$ of 0.02 $(\mathrm{eV})^2/c^4$ from eq. (53) after a counting time of $4 \cdot 10^6$ s only is by far too optimistic. In reality one has to scan a spectrum of many data points in order to fix all the other parameters like r, b, E_0 and last not least the spectral shape as such. In addition, the correlation between fit parameters leads to an increase of the individual fit errors. Consequently it takes two years of data taking to reach that level of statistical uncertainty.

8.2 Background phenomena at MAC-E-Filters

In this section we will summarize how MAC-E-Filters *suppress* background from the surrounding walls and from the T_2 source and how they can *produce* background by their very mechanisms. The former is easily said, the latter is a complex field.

The advantage of a MAC-E-Filter of very large energy resolution at very high luminosity is connected with a principle disadvantage: The collimating magnetic field transports all electrons from a large volume towards the detector. Secondly if these electrons are born in the large volume of the spectrometer they gain the energy corresponding to the electrical potential difference between this volume and the detector. Thus, low energy electrons originating by ionization processes within a large fraction of the spectrometer volume will have an energy at the detector of about the analysing potential. Therefore, they can hardly be distinguished from the signal events.

Of course, this disadvantage of a MAC-E-Filter only counts, if there is a source of electrons within the spectrometer volume. Consequently a very good vacuum inside the spectrometer is required to suppress any kind of interaction of tritium β-electrons with residual gas molecules acting as a source of secondary electrons.

Two other features of a MAC-E-Filter suppress background very successfully:

1. The magnetic field shields the detector from any electron which starts from the walls and electrodes of the spectrometer, provided that

   (a) the magnetic flux connecting source and detector does not strike any of this material in between

   (b) the particles are guided strictly adiabatically along the field lines.

   Under these conditions particles stemming from the walls are guided past the detector and necessarily miss it[24].

2. Secondly, there is the electric potential barrier: any spectral background from the tritium source by β-particles which are lower in energy than the analysing potential qU is *totally* cut off by energy conservation.

These two cut-off properties of MAC-E-Filters are of utmost importance in this game of low level counting.

As mentioned before, experiments at Mainz and Troitsk ran eventually with similar background rates of order $10-20$ mHz, these final levels being achieved in the course of extensive research and development work. Part of this has been reported in the publications of the two groups, the latest being ref. [96]. Background studies at Mainz have been written up in detail in a number of theses [98, 155, 156, 157, 158, 159, 160, 161, 162, 165, 166]. In the following we will discuss in detail background spectra and sources.

Background spectra

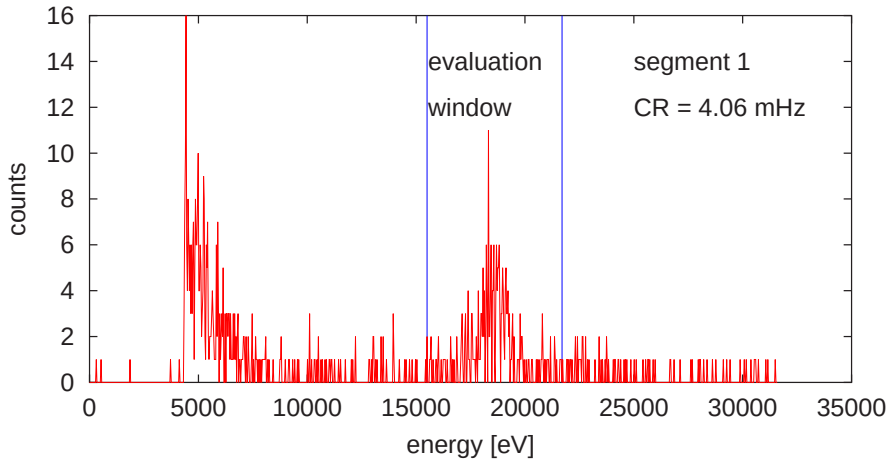

Figure 89: Energy spectrum of background electrons at Mainz detector.

[24] Even high energy electrons from the electrodes/walls, whose cyclotron radii are large enough to touch the magnetic flux will not reach the detector when their motion is adiabatic due to the magnetic mirror effect (see below).

Fig. 89 shows a typical background spectrum from a run at Mainz [96, 98]. It is characterized by a single peak whose mean energy agrees with the analyzing potential U within quite narrow statistical limits of about 30 eV. Hence, it falls into the signal window and cannot be separated energetically by present detector technology. As to the origin of this background one concludes that it has to stem from electrons which are produced at low energy somewhere in the large analyzing volume and then accelerated by the analyzing potential U towards the detector. Various production processes are discussed below. Troitsk observes in addition a significant structure at lower energy (mainly outside of the signal window) which is attributed to H^--ions from the analyzing volume[25]; they are also accelerated onto the detector. The H^--background has virtually disappeared at Mainz after NEG-getters have been installed which are quite effective in pumping out residual hydrogen [158, 159].

A second structure a few keV above the analyzing energy qU has been observed earlier at Mainz and probably traces back to a three step mechanism: electrons emitted from an electrode at high negative potential are accelerated onto an electrode at lower potential at which they generate characteristic X-rays of energies up to 8 keV. These X-rays in turn produce keV photo electrons from the electrode at high potential; the latter are then eventually accelerated to the detector having this surplus energy [158]. Note that keV electrons will have large cyclotron radii in the low field region of the analyzing volume and hence may reach the detector via a non-adiabatic trajectory [156]. During the improvement programme 1996/97 the electrode system has been redesigned to suppress this complex effect successfully [98].

To allow a more detailed discussion, the different sources of background are described in the following:

1. from direct environmental radiation onto the detector
2. from the analyzing volume through tritium decay
3. from the analyzing volume through ionization by trapped electrons
4. from walls and electrodes
5. from the Penning trap between the pre– and main spectrometers
6. from plasmas.

We have also observed strong dependencies on spectrometer parameters like B-field, vacuum, high voltage, surface conditions and the electromagnetic design.

[25] H^--ions appear at significantly lower energies in the detector spectrum with respect to electrons, since they have a much higher energy loss in the insensitive dead layer of the detector.

8.2.1 Environmental radioactivity and cosmic rays around the detector

This background source can be reduced by using selected materials, by active and passive shielding and by a detector with good intrinsic energy resolution. At Mainz this background contributes with a rate of about 1 mHz in the analyzing energy window between 12 keV and 21 keV. The same level can be obtained with the KATRIN detector. Although this will be 20 times larger in area, its background rate can be reduced by the same factor, namely a 2 times better energy resolution, a 2 times thinner active layer and a 5 times better shielding (see section 6, especially 6.4).

8.2.2 Tritium decay in the main spectrometer

A T_2 decay in the analyzing volume will be accompanied in 15 % of all cases by a secondary electron at low energy shaken off the valence shell. With 50 % probability this low energy electron will be accelerated and collimated towards the detector, where it will be detected with an energy close to that of the retarding energy qU. In addition, the more energetic β particles have a certain chance to get trapped electromagnetically, also leading to background in the detector, which will be discussed below.

In order to limit the background rate from tritium decay inside the KATRIN main spectrometer to not more than 1 mHz a very stringent limit on the tritium partial pressure of $p(T_2) \leq 10^{-20}$ mbar ($\approx 4 \cdot 10^5$ T_2 molecules in total) has to be required. The differential pumping and prior trapping sections as well as the pumping speed in pre- and main spectrometer have to guarantee this value.

8.2.3 Background from trapped electrons in the main spectrometer

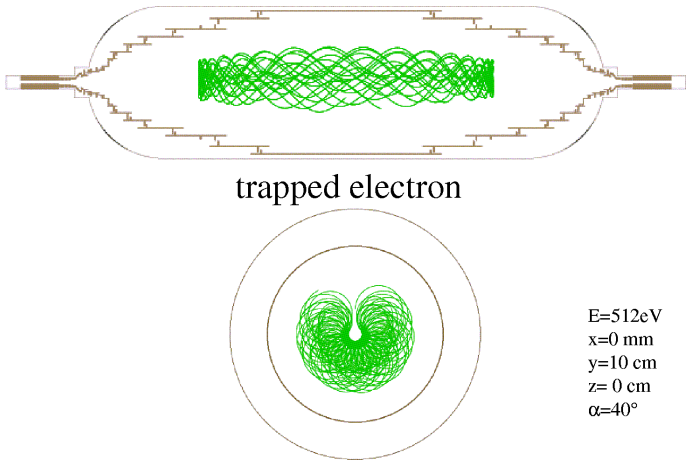

Figure 90: Computed track of a stored electron inside the Mainz spectrometer, for details of the starting conditions see the numbers given in the figure. The electron is stored radially by the magnetic field and axially by the magnetic field gradient ("magnetic mirror"). A slow rotation around the symmetry axis, the magnetron drift, is visible, which is due to the curvature of the magnetic field (see curvature drift (59)) and a residual transversal component of the electric field (see $\vec{E} \times \vec{B}$ drift (57)).

Electrons can be trapped within the main spectrometer. One possibility is complete magnetic trapping in the B-field minimum by magnetic mirroring in the field gradients on both sides (see fig. 90). This occurs for electrons born with a cyclotron energy larger than the width of the MAC-E-Filter. The other possibility is that the electron is restricted to one side by the electric potential barrier and by the magnetic mirror effect to the other one. Then it is trapped within one half of the spectrometer (in both cases the radial storage is provided by the cyclotron motion within the magnetic field.).

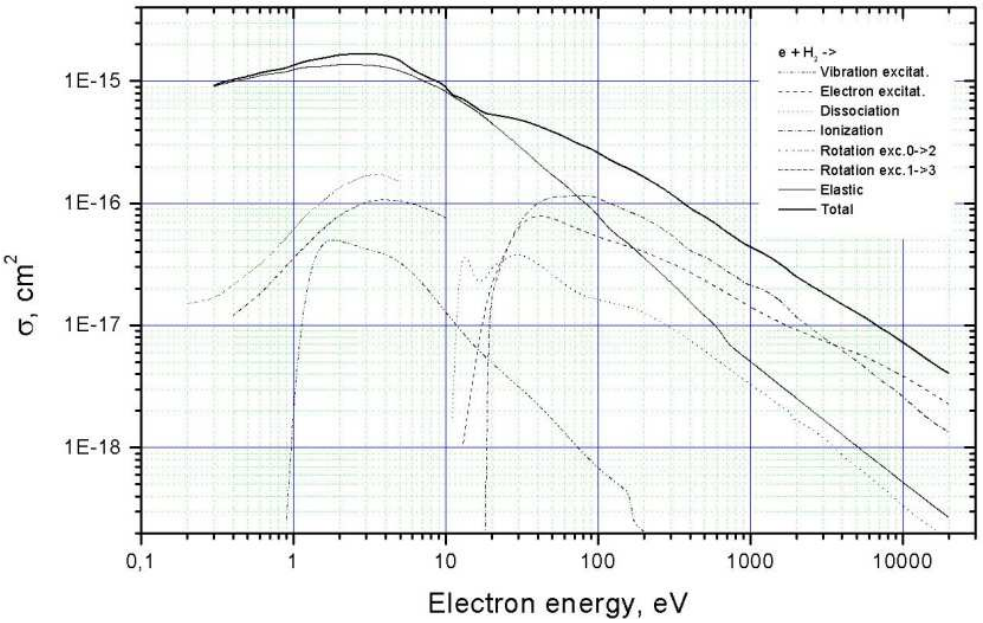

Figure 91: Cross sections for various e^- scattering processes on H_2, adapted from [163].

An electron, trapped in the main spectrometer, will produce a chain of ionizations at a pressure-dependent rate until its energy is used up or the trapping condition is lost. Fig. 91 shows among others the cross section for ionization (σ_{ion}) by electrons on H_2. It exhibits a maximum at an electron energy of about 100 eV. Typical energies of stored electrons of a few keV will yield at a H_2 pressure of 10^{-10} mbar an ionization rate of about 0.1 Hz. The secondary electrons will typically have low energy and end up sooner or later either on the detector or in the source as discussed above. Hence, on average less than one energetic electron, trapped in the sensitive flux volume, would be sufficient to account for the total observed background rate at Mainz. These chains have been clearly identified at Troitsk (at somewhat higher rest gas pressure) and could be removed in the critical part of the spectrum near the endpoint where the signal rate is definitely smaller. At the XHV at KATRIN, however, the time between subsequent ionizations will be of order minutes and the discriminating signature of correlated background events will be gone.

8.2.4 Background from walls and electrodes

Transiting cosmic rays or environmental radioactivity can create secondary charged particles like δ-electrons. Also field emission at edges of the electrode system could lead to the emission of electrons.

The majority of secondary electrons emitted from a surface into vacuum do so at low energy at which they should be magnetically guided safely past the detector along an adiabatic trajectory. Energetic electrons from walls, however, have a chance to reach the detector on a more or less direct trajectory. But they can also be trapped within the spectrometer after scattering or due to non-adiabatic motion (see section 8.2.3). Otherwise their trajectory will become unstable at some time and they may find their way to the detector along a more or less chaotic path (for experimental evidence see below).

We have also to consider the role of residual gas collisions on the transport of wall electrons to the detector. Elastic collisions give rise to a diffusion of trapped electrons from the outside into the sensitive magnetic flux tube. An upper limit for the diffusion coefficient perpendicular to the magnetic field lines can be estimated by

$$D_\perp \leq \rho^2/\tau \tag{55}$$

where ρ is the cyclotron radius and τ the mean collision time. Even at $\rho = 10$ cm (corresponding to a cyclotron energy of 100 eV at $3 \cdot 10^{-4}$ T) it takes many collisions to reach the magnetic flux tube guiding to the detector. Another collision may then kick the momentum into the longitudinal direction and drive it out of the magnetic trap. Of course, this kick-out may happen already in the early diffusion phase outside the sensitive flux and hence leave particles with small cyclotron radii little chance to diffuse far inside.

The electron may also ride on a heavier particle after electron capture, say as H_2^--ion, at a large ρ into the center. The negative ion would not be trapped anymore but immediately accelerated towards the detector and would contribute to the above discussed ion signal.

In another scenario a fairly energetic electron starting from the electrodes may reach the sensitive flux volume right away and then be scattered into the trapped phase space or overcome the magnetic trapping condition by a non-adiabatic trajectory.

Another possibility is the magnetron drift: the electrons are able to move perpendicularly to the magnetic field lines due to the $\vec{E} \times \vec{B}$ and the gradient-B drifts (see sec. 8.3.1). In an axially symmetric field (where both the electric and magnetic field are parallel with the meridian plane) this magnetron drift is in the azimuthal direction (perpendicular to the meridian plane), and it has no radial component. It is a general property of the axially symmetric field system that radial drift motion of charged particles is strictly prohibited, even if their motion is non-adiabatic; this property is a consequence of the conservation of the axial (generalized) angular momentum, which is an absolut constant there [168, 169, 170].

The only remaining possibility for the radial electron motion is due to deviation from axisymmetry, which is to some extent always present. Trajectory calculations show that in this case some electrons starting from the electrodes are really able to penetrate into

the flux tube. Nevertheless, in case of small deviation from axisymmetry only a small part of the electrons can go into the flux tube: namely those electrons, which have small azimuthal magnetron drift velocity [165]. The trajectories of the electrons with large azimuthal magnetron drift velocity are only slightly effected by the drift motion that breaks the axisymmetry [164].

Obviously, the above explanation for the background electrons is only a hypothesis, and it should be tested by experiments. First, one observes that the number of the electrons with small azimuthal drift velocity is small (if the deviation from axisymmetry is small). This fact is in qualitative agreement with the experimental result of [161], namely that only a very small portion (about 10^{-5}) of the electrons created at the inner electrode surfaces is able to reach the detector. According to the theory presented above, one expects a large background in the presence of large electric dipole field (which is an example for the deviation from axisymmetry). Some experiments in Mainz [162, 166] have really seen this increased background rate with large electric dipole field. The connection between the deviation from axisymmetry and radial electron drift was observed in non-neutral plasma experiments with Penning traps. For example, in [171] the substantial improvement of the axisymmetry reduced the radial transport rate of the electrons by a factor 20.

The message for the electromagnetic design of the main spectrometer from the point of view of background minimization is the following: both the electric and magnetic field inside the main spetrometer should be as axisymmetric as possible in order to minimize the ratio of radial to azimuthal magnetron drifts.

Whereas the theoretical concepts for an electron transfer from walls and electrodes to the detector are still in a qualitative phase, an experimental transfer efficiency $\varepsilon =$ (number of background electrons)/(number of ionizing particles) could be determined by placing a source of 511 keV γ's from a ^{22}Na source, onto the spectrometer walls which gives rise to fast Compton or photoelectrons [161]. At standard conditions of the Mainz spectrometer an ε value around 10^{-5} was found. Similar measurements at Troitsk pointed to a value smaller by a factor 10. In 2002, muon paddles of the former KARMEN experiment were installed at the Mainz experiment [119]. Coincident events within a time window of 10 μs in the muon paddles and in the Mainz detector appeared only in 10^{-4} of all background events, although the coincidence of the muon paddles covered a few percent of the full solid angle. This small number is not in contradiction with a large contribution of background electrons from the walls or electrodes, since we expect only a small fraction of the electrons from the walls to be able to directly reach the detector. All the others require intermediate trapping and a scattering process, which will take seconds at least.

In view of a cosmic muon rate of about 10^5/s crossing the KATRIN spectrometer, a transfer efficiency of secondaries of about 10^{-6} would be of concern. However, the much better vacuum foreseen at KATRIN as well as its superior adiabatic magnetic shielding in addition to an electrostatic shielding by a wire grid (see below) will probably more than compensate the enhanced muon rate. The Troitsk spectrometer has a two times larger surface than the one at Mainz and a 10 times higher residual gas pressure. Still Troitsk

finds a smaller ε value and about the same total background. Hence, one may draw the conclusion that the definitely better adiabaticity of the Troitsk instrument actually overcompensates the background dependence on surface and residual gas (see also below).

8.2.5 Penning trap between pre- and main spectrometer

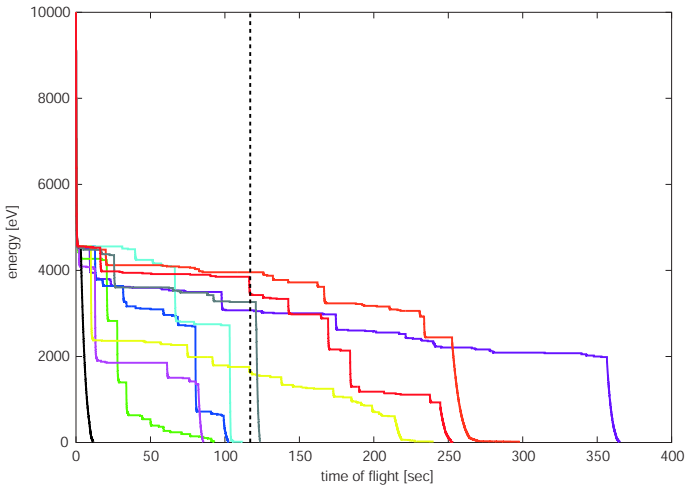

Figure 92: Simulated energy loss of 10 electrons within the Penning trap in between the KATRIN pre and main spectrometers. Energy losses by synchrotron radiation as well as by elastic and and inelastic scattering processes on residual H_2 molecules have been considered. The electrons started with energy $E(0) = 10\,\mathrm{keV}$ in the middle between the 2 spectrometers at a radius of $r = 0.5\,\mathrm{cm}$ and an angle $\theta = 45°$ with respect to the magnetic field lines. The simulations were performed at a residual gas pressure of $p(H_2 = 10)^{-10}$ mbar –10 times higher than expected for KATRIN. After each collision event the transverse energy is lost by synchrotron radiation within a few sec, in addition to the energy transfer to the collision partner (steps in figure). It takes about 100 s (the median of the 10 electrons is 120 s) to cool the electrons down to the ionization threshold of 10 eV.

The large Penning trap for electrons, which is composed by the negative retarding potentials of the pre- and main spectrometers and the magnets in between (see section 7) poses a particular problem. This Penning trap is not a direct source of background, but can produce background in a two step process: Positively charged ions, which are generated by ionization processes of trapped electrons with the residual gas could be accelerated towards the main spectrometer. There, they have a small probability of about 10^{-8} to cause another ionization on the rest gas. The resulting electrons have a large chance to reach the detector and to contribute to the background rate.

One may estimate that this Penning trap is filled from the outside at a similar rate as the total background rate on the detector, which may be in the range of 200 mHz [26] at KATRIN. Considering the very low probability of the multi-step background process

[26] At Mainz the background rate in the energy window of interest is of order 10 mHz, whereas the background over the full energy range amounts to about 100 mHz. We are assuming that the KATRIN pre and main spectrometers will emit at maximum these rates of electrons towards both of their ends.

described above, the Penning trap should not be of any concern unless a chain reaction occurs, leading to an exponential growth of the number of trapped electrons.

We have investigated these processes by trajectory simulations in this Penning trap including cooling processes by synchrotron radiation as well as elastic and inelastic scattering of electrons on the residual gas [167]. Fig. 92 shows the example of energy loss simulated for 10 trapped electrons. It is clearly visible, that these electrons lose their cyclotron energy after each collision event by synchrotron radiation on a time scale of less than a second due to the high magnetic field of 3.6 T in the transport magnets. The corresponding rate constant is

$$\frac{1}{\tau_{Sy}} = \frac{\dot{E}_\perp}{E_\perp} = 0.4\ \mathrm{s}^{-1} \cdot \left(\frac{B}{1\ \mathrm{T}}\right)^2 \tag{56}$$

At a rest gas pressure of 10^{-10} mbar, (chosen for technical reasons 10 times higher than expected for KATRIN) the scattering processes on the rest gas occur at a rate of about 1 per 4 s per stored electron (see fig. 92). As fig. 91 already indicated most processes are elastic or excitation processes without ionization. Like synchrotron radiation, elastic scattering and excitation processes cool the trapped electron without creating dangerous electron ion pairs. It should be noted that elastic scattering plays an important role in this cooling process although it transfers little energy to the partner. However, the scattering angle might be large, reshuffling energy from longitudinal to transverse motion, which is finally radiated away.

Of major concern are the ionization processes, which occur with a rate of 1 per 10 s per stored electron at the considered rest gas pressure of 10^{-10} mbar. They can start a chain reaction if the secondary electrons get sufficient energy for further ionization. Fig. 93 shows the position on the axis at which the ionization processes took place for the 10 simulated trapped electrons of fig. 92. It is clearly seen, that most ionization processes take place at ground potential, where no potential energy is picked up. However, a small fraction of the ionizations take place at the end of the Penning trap. In this case, the secondary electrons gain energy from the electrical potential and become able to further feed the ionization chain process. For the start parameters chosen in the simulations of fig. 92 and 93, just one secondary electron is created at a high electrical potential per initial trapped electron in the average. This number becomes smaller for larger starting angles and larger for small starting angles of the initial electron.

Since the critical number of Penning chain amplification is just one we have to worry about this Penning trap. We have investigated several solutions to eject trapped electrons at a time scale, which is smaller than the time scale of ionization processes (100 s at a residual gas pressure of 10^{-11} mbar). One simple way would be the use of a mechanical wire which is rapidly scanned every few seconds through the magnetic flux during measurement pauses. One way to perform this mechanical scan, which avoids the usual problem of mechanically movable parts inside the UHV, would be to move a wire loop in the strong magnetic axial field by running a current pulse through it.

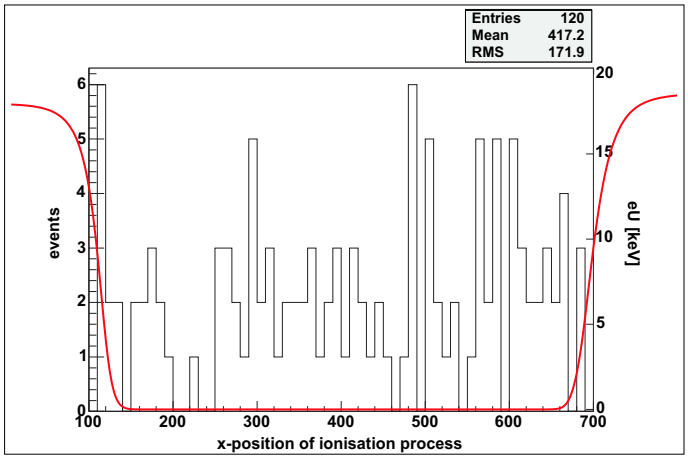

Figure 93: Position of simulated ionization events in between pre- and main spectrometer on the axis. The full line illustrates the corresponding electrical potential right scale.

8.2.6 Penning plasmas from electrodes in high B-field

At the start of their experiments both Mainz and Troitsk have observed tremendous background rates which could result in strong plasma currents. At Mainz this dramatic situation was remedied by lowering the field in the solenoids from the design limit of 8 Tesla to below 3 Tesla. Then the background rate dropped from several kHz to a value of 0.1 Hz which seemed acceptable at that time. Over the years this number was lowered by another factor of 10 to the present 10 mHz level.

Troitsk has suppressed such strong plasmas by a radical redesign of their electrode system. Originally it extended into the high B-field region, as at Mainz, in order to retard electrons early and to optimize the adiabatic passage of β-particles. These high field electrodes were discarded and retardation was achieved essentially in the fringe field of a specially shaped large central electrode at analyzing potential U. Troitsk could then run at full central B-field in spite of their higher residual gas pressure. It is supposed that electrodes in high B-fields provide favourable conditions for the build up of Penning-like plasmas known from high vacuum gauges, for instance. In view of the more critical adiabaticity conditions in the smaller Mainz spectrometer one has dispensed there with the radical Troitsk solution (but see below), but was content with a lower B-field, sufficient to stop the plasma.

8.2.7 On the influence of vacuum and surface conditions

The contribution of the gas pressure to background production through scattering and ionization has been pointed out in various contexts already. In general the superior vacuum of the Mainz spectrometer was a great help; it allowed, for instance, use of the more critical electrode system in the high B-field as discussed above. It might also be the ultimate reason why the spectral irregularities close to the endpoint reported by Troitsk have rarely occurred at Mainz at a significant level. However, a clear cut explanation for

these irregularities in terms of a well-defined mechanism is still missing. Nevertheless, there is empirical evidence that an optimal status of the apparatus is important in this respect. For instance, after the 1998/99–runs at Mainz which were totally free from spectral anomalies, measurements were resumed in late 2000 without re-baking the apparatus beforehand. The following run period was characterized then by minute and variable spectral anomalies which resulted in small negative m^2-values of order of $-10\,\mathrm{eV}^2$. These anomalies also depended on the sense in which the spectrum was scanned. Again, the source of these tiny hysteresis or memory effects could not be traced definitively, but it is probably connected to the time constants of some trapped particles.

Mainz has investigated, therefore, whether the background level changes inside the signal interval below the endpoint by applying a positive pulse to the source which pulled the entire β-spectrum below the analyzing potential. During this pulse the count rate falls back to the background level. Within a time window of $15\,\mu$s to several minutes after pulling back the β-spectrum, the background did not show any memory effect on the previous state within an uncertainty of about 1 mHz [161].

The 2001 running period at Mainz was preceded by a re-baking of the spectrometer and a careful maintenance in many respects. This was rewarded by a further reduction of the background down to 13 mHz and a totally clean β-spectrum showing no anomaly whatsoever. Also Troitsk has improved its vacuum and running conditions in this period and could thus reduce their background level to a total of 10 mHz. Troitsk has then also experienced anomaly free runs but not in all cases (see chapter 2.2.1).

Besides vacuum, surface conditions have also shown a large influence on background. Although all critical parts are electro-polished, optimal conditions are only obtained after careful conditioning of the high voltage system up to voltages 50 % above working level. Adding some helium residual gas during this procedure gives the sparks the necessary power to remove hot spots on the surface from which field emission starts. We expect this problem to be much more relaxed at the large dimensions of KATRIN where the electric field strength is much lower.

8.3 Active measures against background

8.3.1 Traps and trapped electron removal by $\vec{E} \times \vec{B}$ drift

The trapping of particles has been studied in more detail by Monte Carlo simulations. In addition to the commercial program "Simion 7.0" the program "Adipark" was developed [164], which tracks only the guiding center of a gyrating particle to allow faster simulations. To check the long-term behavior, the adiabatic drifts were also taken into account to first order:

$$\vec{E} \times \vec{B} \text{ drift} \qquad \vec{u}_E = \frac{c}{B^2}\vec{E} \times \vec{B} \tag{57}$$

$$\text{gradient drift} \qquad \vec{u}_G = -\frac{cE_\perp}{eB^3}\vec{B} \times \nabla_\perp \vec{B} \tag{58}$$

$$\text{curvature drift} \qquad \vec{u}_C = -\frac{2cE_\parallel}{eB^3}\vec{B} \times \nabla_\perp \vec{B} \tag{59}$$

By these new methods it was found, for instance, that the old Mainz setup was not optimized concerning electron traps: For low-energy electrons local Penning-like trapping conditions were identified when a magnetic field line crosses an electrical equipotential surface twice in the corners of the Mainz electrodes. Electrons with energies of a few eV up to a few keV can be stored by the magnetic mirror effect in the centre of any axially symmetric MAC-E-Filter. Figure 94 shows the calculated "trapping volume" of electrons starting with 16 eV in the Mainz spectrometer. Any electron which starts inside the contour lines under an angle with respect to the local magnetic field line, which is larger than the value specified for this contour line, will be trapped.

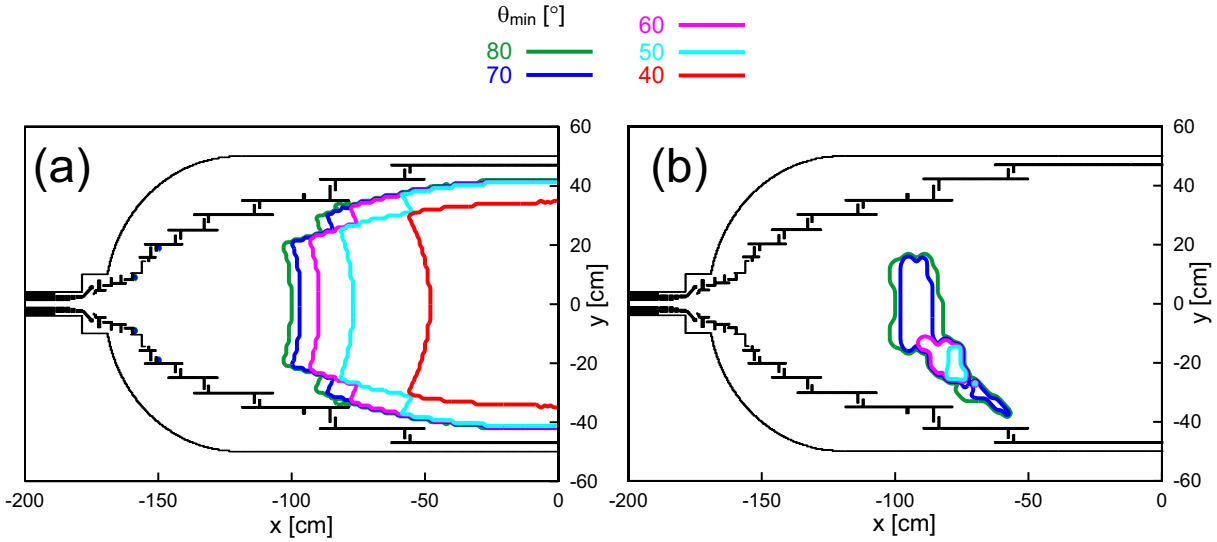

Figure 94: Trapping volume for electrons with a start kinetic energy of 16 eV and different starting angles with respect to the local magnetic field lines, calculated for the Mainz spectrometer. (a): standard MAC-E-Filter. (b): superimposed electric dipole with field strength of about 100 V/m.

To remove stored electrons, the transverse drift terms of adiabatic motion in first order can be used. For the case of a MAC-E-Filter it is much more efficient to apply a perpendicular electric field $\vec{E}$ (see fig. 95) and to use the $\vec{E} \times \vec{B}$ drift (57) than to bend the spectrometer and to use the curvature drift (59), an idea also discussed earlier for KATRIN. Figure 96 shows how the trapping condition is cancelled by a transversal electric field composed of a central dipole electrode within the Mainz spectrometer. The electron starts in the centre and is finally driven onto the wall.

In principle this method should work well and remove stored electrons even using tiny transverse fields, if it were not for the fact that the magnetron motion (see figure 90) would not average out small values of the $\vec{E} \times \vec{B}$ drift and thus stabilize the motion of the trapped particle. Therefore the trapping condition is violated only for electric dipole fields above a critical value. A good estimate is that the $\vec{E} \times \vec{B}$ drift has to be faster than the magnetron motion. The program "Adipark" was used to study this in more detail. Figure 94 (right) shows a strongly reduced trapping volume for electrons starting with a

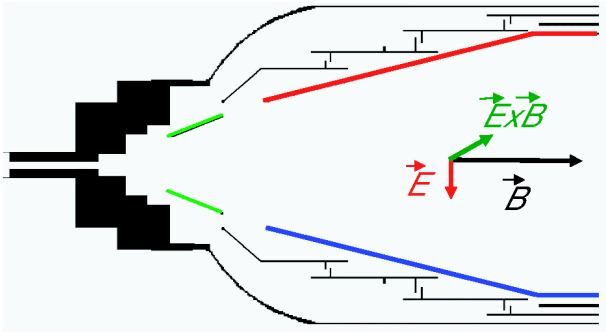

Figure 95: $\vec{E} \times \vec{B}$ drift in a MAC-E-Filter induced by a transverse electric field created by a dipolar electric potential distribution on the wire electrode system of the Mainz V setup (see below).

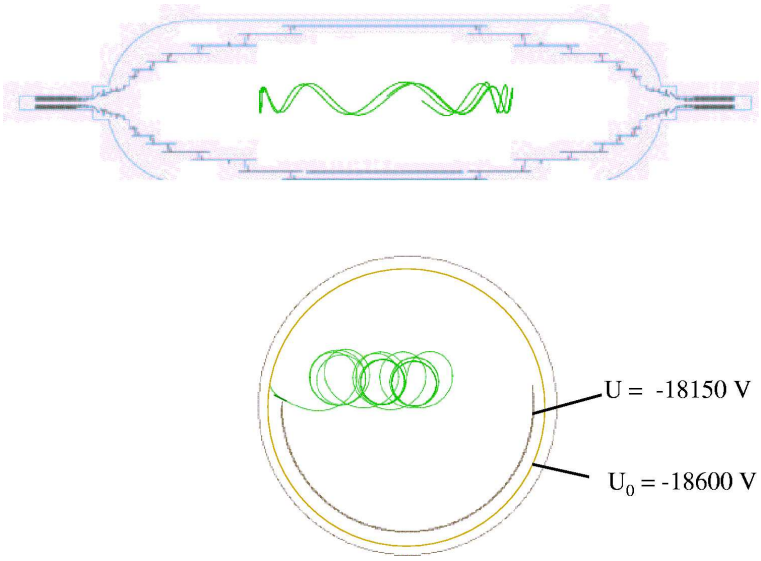

Figure 96: Trajectory of an electron stored in the magnetic field minimum of the Mainz spectrometer (top: side view, bottom: cross section view). By a perpendicular electric dipole field the axial symmetry is broken and the storage condition for the trapped particle is cancelled.

kinetic energy of 16 eV calculated for the Mainz spectrometer for the case of a slightly under-critical dipole field, whereas fig. 96 shows how the particle is driven to the wall in case of an over-critical field.

8.3.2 Shielding electrons from the walls and electrodes by a repelling grid

Electrons originating from the electrodes or walls (induced by radioactivity, cosmic muons or field emission) are normally not able to reach the magnetic flux tube which is imaged onto the detector, because the Lorentz force bends these electrons back to the wall. Moreover, the magnetic field gradient provides a magnetic trap for the majority of electrons starting from the central part of the MAC-E-Filter, as discussed above. But, in spite of this inherent magnetic shielding, one observed in the Mainz II setup that a small frac-

tion of electrons emitted from the electrode still reached the detector. This can usually happen either by non adiabatic motion or in case the electrons have large cyclotron radii. The latter case corresponds to large emission energies, which is the case only for a small fraction of the electrons originating at the electrodes/walls.

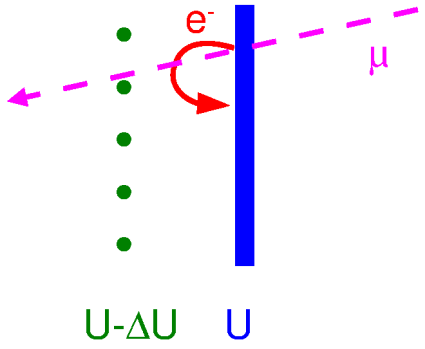

Figure 97: Principle of electrical screening. Secondary electrons emitted from an electrode at potential U are repelled by a wire electrode in front, set to a slightly more negative potential $U - \Delta U$.

A new idea to also get rid of those electrons which evade the magnetic shielding is to install an additional electrical shielding. An electrode at a slightly more negative potential is installed in front of the electron emitting surface (see Fig. 97). In order to avoid this electrode itself becoming a major source of secondary electrons it has to be a wire electrode of small geometrical coverage with moderate electric field strength at its surface.

8.3.3 Experimental tests of active background reduction at Mainz

By the end of 2001 tritium experiments have been stopped at Mainz in favor of dedicated experiments checking new ideas for background suppression for KATRIN. They were preceded by extensive simulations of electromagnetic field configurations and particle trajectories. Not only trapping conditions were studied in detail (see above) but it was also found out that the Mainz spectrometer would still function with a single large electrode at the retarding potential U à la Troitsk, although its shorter length and reduced magnetic flux render adiabaticity rather critical. Calculations also showed that trapped electrons can be removed by three pairs of electric dipoles, one in the analyzing volume, and two more in the fringe field on both sides. However, voltages up to 200 V are necessary to brake up the traps (see above).

According to these calculations the electrode system has been redesigned as shown in fig. 98 [162]. The old electrodes E0 to E3 are still in use but now all at the same analyzing potential U of the central electrode E0. Electrodes G1 and G2 shape the fringe field and should be on ground potential whilst measuring the β-spectrum. But they are split in order to provide the transverse electric dipole field for removing trapped particles during short measuring pauses –say every 30 s. Such periodical cleaning has had the effect

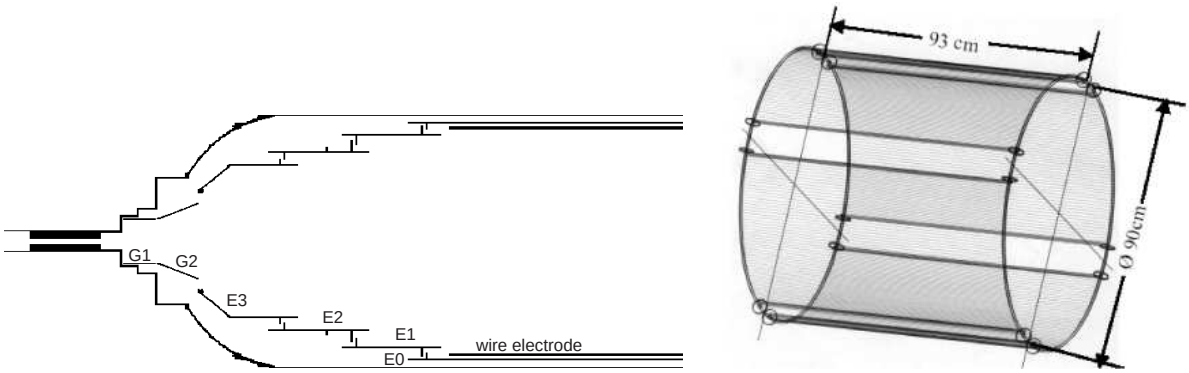

Figure 98: New Mainz electrode system with additional wire electrode in the middle [162]. Left: The electrodes E0 - E3 are usually at the same potential, whereas electrodes G0 and G1 are at ground potential. The electrodes E2, E3, G0 and G1 are split in order to allow the application of an electric dipole field. Right: Wire electrode inside electrode E0. It consists of wires with a diameter of 0.2 mm at a spacing of 1 cm, resulting in a geometrical coverage of 2 %. The wire electrode is also split to allow operation as dipole.

of stabilizing the background rate to a residual hard core value of about 15 mHz, as in earlier Mainz runs [160, 161, 87]. At that time no dipoles were installed yet, but the particles could be heated out of a particular (not clearly identified) trap by applying rf between electrodes 8 and 9, which excites the axial motion.

Background screening was first tried out only for the central part of the spectrometer by covering the electrode E0 by a wire grid (see fig. 98, right). Putting the grid to the potential of the electrodes E1–E3, and E0 to a positive potential with respect to these, the grid should repel low energy electrons released from the electrode surface. The grid has been divided into four sectors in order to possibly serve as dipole or quadrupole in addition.

Magnetic fields, adiabaticity and magnetic shielding

First background studies have been performed with this new system in 2002. First the spectrometer was set to the standard field values used before, namely 1.7 T in the solenoids and $5 \cdot 10^{-4}$ T in the analyzing plane; additional dipole and monopole fields were not applied. A background rate of about 700 mHz in the usual energy window between 15 and 21 keV was measured altogether in the three inner sectors of the detector which cover the usual flux. This is a factor of 60 above the best value measured with the old system. The increase of the background rate from the axis towards larger radii was also much more pronounced than earlier. The magnetic fields were then raised stepwise and with each step the background decreased, down to a total of 12 mHz at four times the standard flux. So the minimal background from before was reached again and the suspicion confirmed that magnetic adiabatic shielding of background required higher fields than before. Trajectory simulations proved that the new single Mainz electrode system violates adiabaticity more readily, at the low magnetic fields, than the old Mainz system

with 27 electrodes. Moreover, these calculations showed, that a 3 times higher magnetic field for the new Mainz setup gives about the same adiabaticity violation as the old Mainz setup with lower field. This underlines that adiabaticity conservation and magnetic shielding are strongly connected.

Function of the electrical-shielding wire electrode

The repelling wire electrode turned out to be an important step forward to further background rejection. Fig. 99 shows that above a repelling voltage of $\Delta U = 50\,\mathrm{V}$ the

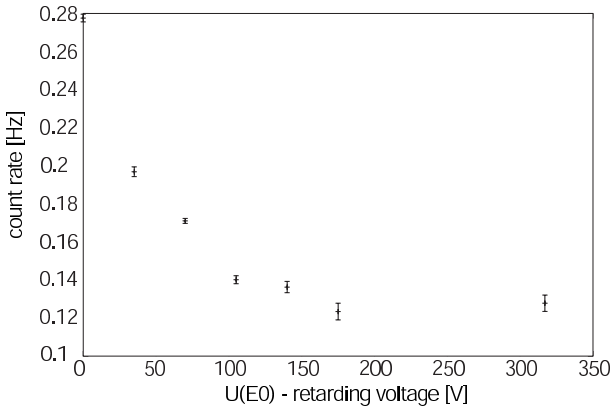

Figure 99: Background rate as function of repelling voltage on the central wire electrode.

background started to drop until it reached eventually half of the starting value at 170 V. This factor of two was gained for the higher rates at the standard field of 1.7 T as well as for the lower ones at 5.1 T. Thus, a new record minimum of 5 mHz was measured within the sensitive flux tube. Since the wire grid covers only 40 % of the inner surface of the electrodes at high voltage, a reduction factor of two is compatible with the following assumptions:

- Electron emission from the high voltage electrodes is the primary source of residual background of a MAC-E-Filter.
- This background can be cut off by covering the electrodes with an electrostatically repelling wire electrode.

These assumptions were further corroborated by placing a ^{22}Na source on top of the spectrometer (compare section 8.2.4). The additional background rate induced by this radiation dropped also by a factor of 2 with the wire grid on. At this point we should make clear that a coverage of only 40 % may well result in a background reduction by 50 % because the electrode surface does not necessarily contribute everywhere equally. These questions have been investigated in more detail with a powerful and well-collimated x-ray tube (see fig. 100).

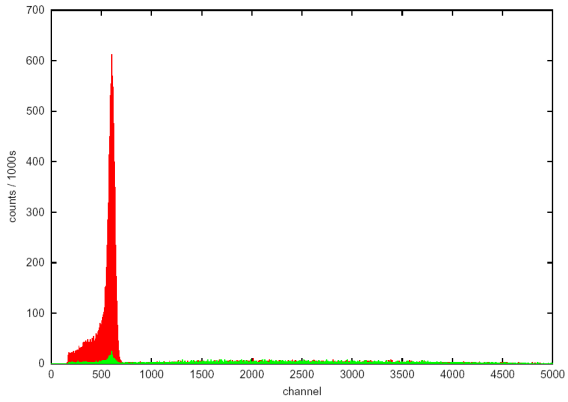

Figure 100: Background spectra at Mainz obtained whilst irradiating the electrodes with a collimated x-ray source, without active screening electrode (red) and with active screening electrode (green). The count rate reduction is about a factor 50.

We have also observed, that the repelling wire electrode reduced the background rate coincident with hits in the muon paddles nearly completely [119], again supporting its importance.

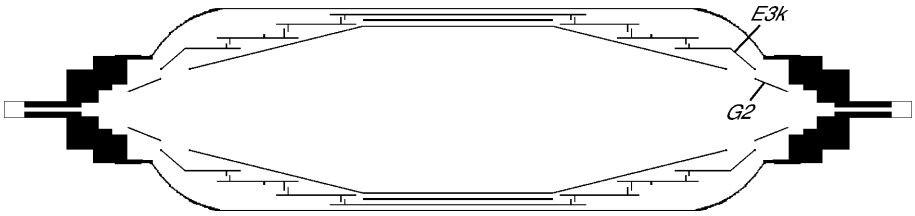

Figure 101: The Mainz V setup with an inner wire electrode system, which covers the full inner electrode system (inner double conical and cylindrical shape). The former wire electrode shown in fig. 98 is still kept in place.

Finally the Mainz electrode system was completely covered by a wire grid electrode This Mainz V setup is shown in fig. 101. This setup resulted in the lowest background rate ever reported for a MAC-E-Filter. Fig. 102 presents the background rate as function of the repelling wire grid potential. The total background rate at a wire potential which is 200 V more negative than the electrodes is 2.8 mHz, out of which 1.6 mHz is due to the environmental background of the detector without spectrometer. The pure spectrometer connected background is 1.2 mHz which is about an order of magnitude lower than during the previous standard operation of the Mainz and Troitsk spectrometers. These Mainz V results clearly show that this new idea of a repelling wire grid is extremely important and that the background rate of KATRIN can be kept low even with the large KATRIN main spectrometer.

Rest gas pressure dependence

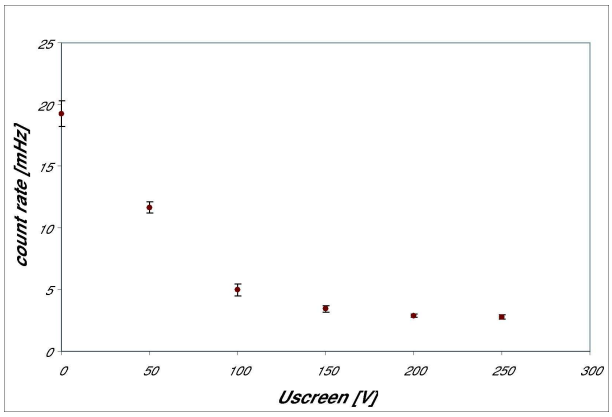

Figure 102: Background rate as a function of repelling voltage on the wire electrodes at the Mainz V setup. It should be noted that the detector background rate (without spectrometer) already amounts to 1.6 mHz.

An unexpected finding was that the background hardly changed upon the addition of a helium partial pressure up to 10^{-8} mbar[27] (see fig. 103). The pressure dependence

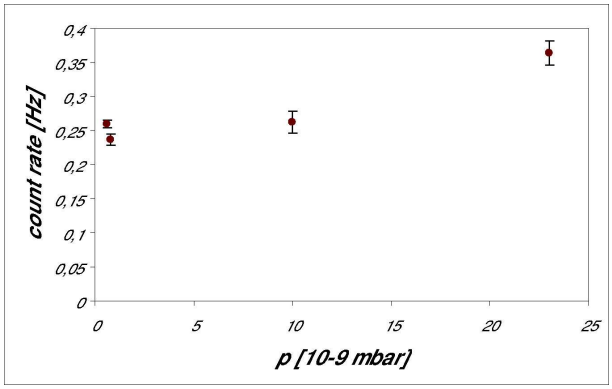

Figure 103: Background rate at the Mainz V setup as function of the helium pressure without repelling potential at the wire electrodes.

for normal residual gas (measured when turning off the turbomolecular pumps with NEG pumps already removed in the Mainz V setup) is still visible probably due to the much lower ionization threshold (see fig. 104 left), but it gets much weaker with the screening potential on (see fig. 104 right). Also the rise of background with analyzing potential was found to be much slower than before. Both findings indicate a greatly reduced sensitivity of the new design to firing and feeding any kind of Penning plasma. One reason for this new behavior could be the absence of the little Penning traps in the corners of the new electrode system. Another would be that strong electric field gradients no longer occur in regions of high B field which favor Penning discharges.

[27]The Troitsk setup also does exhibit a strong background rate dependence on the residual gas pressure.

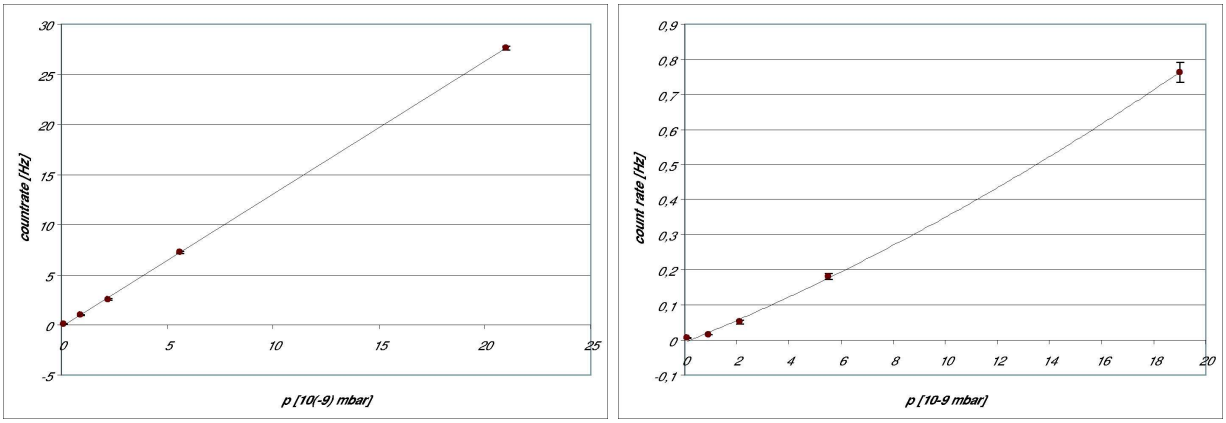

Figure 104: Background rate at the Mainz V setup as function of the residual gas pressure without (left) and with repelling potential at the wire electrodes (right).

Function of electric dipole fields

First tests with transverse electric dipole fields confirmed that they do remove trapped electrons from the spectrometer. To that end isomeric ^{83m}Kr was let into the spectrometer together with some pbar of residual gas. Trapped conversion electrons from krypton decay then produced very prominent chains of correlated background events stemming from consecutive ionizations. The chains disappeared completely when the wire grid was used as a dipole with a voltage of 200 V across, as predicted by simulations (compare section 8.3.1).

As the detector is radially segmented, the length of the chains of correlated events could be investigated as function of its radial position. In the innermost sector 1 up to 25 correlated events could be observed on the oscilloscope without dipole field. Further out (sector 2) the chains were much shorter and in the outer sectors 3 and 4 no chain could be seen. There are apparently loss mechanisms at work that do not allow energetic electrons to be trapped for a longer period of time.

The measurements with ^{83m}Kr decaying in the spectrometer indicate that background induced by trapped particles has to be studied also experimentally as it is quite difficult to calculate an electron trajectory over many seconds, including non-adiabatic processes in the calculation[28].

The electric dipole field used at Mainz to remove trapped electrons from the spectrometer has been observed to introduce some additional background by itself depending upon the amplitude and polarity of the potential applied. The effect will be studied in more detail in the near future.

At first sight enhanced background during operation of the dipoles is not harmful for the neutrino mass measurement since they would be switched on only during pauses

[28]However, the influence of the electric dipole fields on electrons which are not trapped for a long time but which are flying directly from the source through the KATRIN spectrometers to the detector can be easily calculated.

between measuring points. However, it is important to assess their after-effect: Will they have removed all of the trapped electrons or will some of them get trapped in the transient caused by the dipoles shut-off. Such time-dependent studies, as well as experiments with rf pulses, have still to be done.

8.4 Conclusions for background rejection at KATRIN

Although we do not have quantitative theories to calculate the background rates to be expected from the various sources, experience collected over the years at Troitsk and Mainz, as well as from recent dedicated calculations and experiments, is sufficient to show a clear route of success for KATRIN in this respect:

- The differential pumping and vacuum system has to guarantee that the number of T_2 molecules present in the analyzing volume of the spectrometer does not exceed about 10^4.
- The vacuum in the spectrometer vessel should be pushed to the XHV in order to minimize the chance of residual gas ionization and trapping of intruding charged particles by scattering them into the trapped phase space.
- Surfaces should be prepared by electro-polishing, baking in vacuum and conditioning at access voltage in order to minimize the risk of field emission.
- Electrodes at high voltage should be kept out of high B-field regions in order to eliminate the build up of any kind of Penning plasma.
- The design value of the maximum magnetic flux of $191\,\mathrm{T\,cm^2}$ should not be lowered for economic reasons, but be maintained in order to obtain an optimal magnetic adiabatic shielding of the sensitive flux volume from intruding electrons.
- The entire inner surface at high voltage should be covered with a wire electrode which repels electrons emitted at low energies.
- Methods for emptying traps by pulsed static or rf electric dipole fields have to be explored further including tests with the pre spectrometer.

It may well be that only some of the precautions listed above are necessary to meet the design goal of a 10 mHz background at KATRIN. But there is also a good chance that their cumulative effect will eventually lead to even lower background.

9 Energy stability, monitoring and calibration

The determination of the neutrino mass from the measurement of the tritium β-spectrum near the endpoint by the KATRIN experiment[29] requires for each event the precise knowledge of the energy $E_{\rm ret}$ retarded in the analyzing plane of the main spectrometer. This energy is determined by the retarding electrostatic potential at the analyzing plane and the scanning potential U_s applied to the source.

To illustrate the precision needed for $E_{\rm ret}$ we can use the following estimate. An unknown smearing of $E_{\rm ret}$ with the Gaussian variance σ^2 will result in a systematic shift of the squared neutrino mass m_ν^2 [185]

$$\Delta m_\nu^2 = -2 \cdot \sigma^2 \quad . \tag{60}$$

Therefore, requiring the uncertainty of m_ν^2 caused by this smearing to be less than $0.005\,\mathrm{eV}^2$, we need not only a very high short-term stability of the retarding voltage but also a method to measure it with a 50 mV precision for at least three years data taking. For a retarding voltage of 18.6 kV this corresponds to a long-term relative precision better than 3 ppm. The precise measurement of the scanning voltage U_s, which will not exceed 100 V, does not represent any problem.

A more detailed numerical study on the energy scale imperfections can be found in [226]. E.g. it was shown that an *unrecognized shift* of $E_{\rm ret}$ by 0.05 eV would result in the systematic error of the fitted neutrino mass as large as 0.04 eV which is a substantial part of the expected KATRIN sensitivity to m_ν.

We will utilize several methods to monitor the retarding potential in the KATRIN experiment to achieve a high degree of redundancy:

- Direct retarding voltage measurement
  A high-precision voltage divider will divide the retarding voltage U down to about 10 V, which is then measured with a high-precision digital voltmeter.

- Monitor spectrometer
  Calibration measurements in the main system cannot take place in parallel to tritium measurements. But equation (60) shows the necessity to monitor the stability of the retarding voltage during tritium runs. The retarding voltage of the main spectrometer will also be supplied to a third electrostatic analyzer of the MAC-E type. The task of this monitor spectrometer is to measure the energy of a well-defined sharp photoelectron- or conversion electron line which is compared to the retarding energy of the KATRIN main spectrometer. The use of monitoring data for the absolute calibration of retarding voltage of the main spectrometer will be also considered.

[29] The spectrum of β-electrons is described by eq. (10) for an infinitely heavy isolated β emitter. In the case of T_2 we need also to consider the recoil energy and the rotational-vibrational as well as the electronic excitations of the daughter molecular ion (see sec. 11).

- Direct calibration of the main spectrometer
  Frequently we will do spectroscopy of photoelectrons from a ^{241}Am/Co source or of conversion electrons from a ^{83m}Kr source with the KATRIN main spectrometer to absolutely calibrate the retarding energy E_{ret} under measurement conditions.

All three methods will be used not only to monitor the stability of the KATRIN retarded energy, but also to perform an absolute energy calibration repeatedly. The absolute calibration is necessary not only to check the stability of all monitor systems, but also to compare the endpoint energy E_0 obtained by fitting the measured tritium β spectrum with the helium-tritium mass difference $\Delta\mathrm{m}(^3\mathrm{He} - ^3\mathrm{H})$, determined by cyclotron resonance measurements in Penning traps. Any significant difference would point towards an unrecognized systematic error.

9.1 Precise retarding voltage measurement

The retarding voltage of the KATRIN main spectrometer will be reduced by a precision high-voltage divider down to a voltage below 10 V, which is ideally suited for state-of-the-art high precision digital voltmeters. Whereas suitable voltmeters are commercially available with a precision and a long-term stability at the ppm-level, voltage dividers are commercially available only with a stability and precision of 10^{-5}, too low to be suited for KATRIN.

Since the most precise range setting is 10 V for a state of the art high precision digital voltmeter, the divider has to reduce the retarding voltage of up to 35 kV down to a voltage below 10 V in order to keep the precision at the ppm level. Hence, the divider ratio has to be stable at the same level. Therefore, with support from the German Physikalische-Technische Bundesanstalt (PTB) in Braunschweig, we are developing a high precision high voltage divider aiming at a long-term stability and precision at the 1 ppm level for a maximum voltage of 35 kV.

A screening of different high precision resistors from different brands has been performed and resistors with the lowest warm-up and temperature dependent resistance change at the ppm level (type VHA518 from the company Vishay/France) have been selected and bought.

The construction of two identical high voltage dividers has been started, based on the successful design of the 100 kV (2 ppm) divider at PTB [181] (see fig. 105).

To decouple the divider setup from the environment, there will be a metal shielded housing with internal temperature control at a level less than 1 K. The whole setup will have cylindrical shape and the high precision resistors will be mounted on a helix structure and be kept dry under N_2 gas as insulation gas.

The high voltage section will consist of 100 precision resistors with a total resistance of 184 $M\Omega$. Every group of 25 resistors will be shielded by electrodes creating a smooth potential distribution at the mounting position of the resistors, in order to reduce leak currents across the teflon insulators and to prevent sparking. The electrodes will be

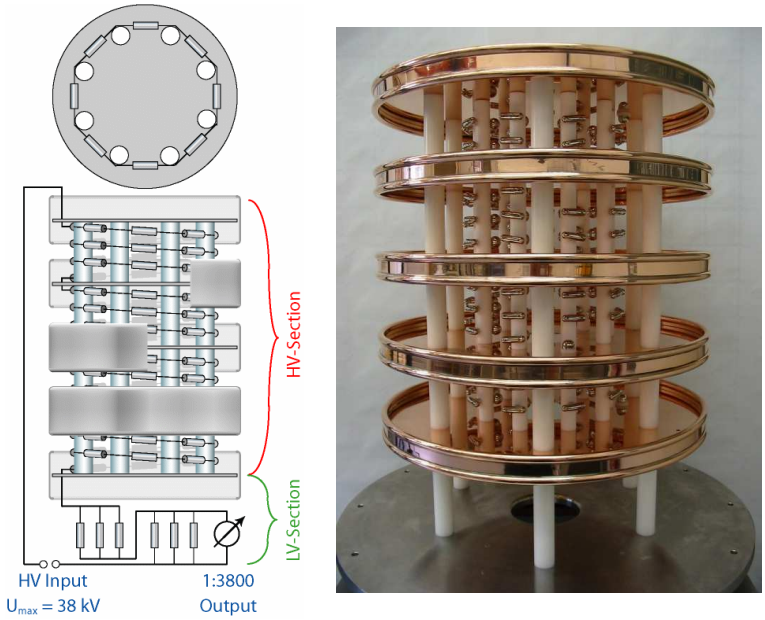

Figure 105: High voltage divider sketch and photo of the realized setup. The diameter of the vessel, which houses the high voltage divider, is 60 cm. The high voltage divider is sitting on a 70 cm high 19 inch rack, which houses the digital voltmeter (DVM) and the temperature control system.

driven by a second independent high voltage divider with a maximum voltage drop of about 10 kV each step, without any connection to the precision resistors circuit.

The low voltage section consists of the same type of resistors as the high voltage section but with lower resistance to provide a dividing ratio of 1:3800 and 1:1900 for the different operation modes[30]. To avoid electromagnetic noise from the temperature stabilization system, the heating or cooling, respectively, of the high voltage divider inside the vessel, which is acting as a Faraday cup, is done by a water circuit. To avoid damage of the high precision resistors when instantaneously switching on and off the high voltage (*e.g.* for calibration purposes) a third, purely capacitive high voltage divider was designed, which is connecting the electrodes by large capacitance high voltage capacitors.

Figure 105 shows a sketch of the setup and a picture of the present status in December 2004. Further screening of the high precision resistors is still going on, in order to find best matching samples according to their warm-up resistance change and with the lowest temperature coefficient. First tests of the high voltage divider are planned at the Mainz spectrometer and at the KATRIN pre-spectrometer for spring/summer 2005.

[30] For some calibration purposes we will run the KATRIN main spectrometer at energies around 30 – 32 keV, *e.g.* in order to measure the line position and line width of L-32, M-32 und N-32 ^{83m}Kr conversion electrons from the WGKrS. The standard retarding energy corresponds to the tritium endpoint energy $E_0 = 18.6$ eV.

9.2 Energetically well-defined and sharp electron sources

In order to calibrate the KATRIN retarded energy absolutely and to check the stability of the high-voltage measurement setup, energetically well-defined and sharp electron sources are needed. The electron energy should be defined by atomic or nuclear standards. We will apply three kinds of sources:

- conversion electrons from ^{83m}Kr in different physical states
- photoelectron sources with well-defined photon energies
- Auger electrons from ^{109}Cd

For measurements to probe energy losses and to test adiabaticity, we will in addition use electrons at energies of about the retarding potential by photoemission with UV light.

9.2.1 Conversion electrons from ^{83m}Kr

The K-conversion electron line of the 32 keV transitions in ^{83}Kr (denoted K-32) has an energy of 17.8 keV and a natural width of 2.8 eV (FWHM) [182]. As this energy differs by only 0.8 keV from the endpoint energy of the tritium β-spectrum, the K-32 line is well suited for the tasks of absolute calibration and monitoring of the spectrometer energy scale. Although the strong L_3-32 line has an energy of 30.5 keV, well above the tritium β-decay endpoint, it may be advantageous for special investigations due to its natural width of 1.2 eV [186]. The half life of ^{83m}Kr is only 1.83 h. On one side this avoids any danger of a long-term contamination of the apparatus, but this also means that replenishment of ^{83m}Kr is necessary for studies exceeding several hours. It has been verified that the mother isotope ^{83}Rb exhibiting a half-life of 86 days can be produced at the cyclotron facilities at Řež/Czech Republic and Bonn/Germany in a sufficient amount.

In the first application we will utilize gaseous ^{83m}Kr to check the properties of the WGTS as well as for absolute energy calibration of the whole apparatus. For the latter task the energy of this conversion line has to be known with high precision. A previous measurement [172] of the K-32 conversion electron energy of $17\,821.4 \pm 2.0$ eV is not sufficiently accurate for the KATRIN experiment. In the following we outline an experimental programme to improve the precision of this quantity.

For ^{83m}Kr atoms in gaseous form, the kinetic energy E_{kin} of the K-32 conversion electrons is given by

$$E_{\mathrm{kin}} = E_\gamma + E_{\gamma,\mathrm{rec}} - E_b^{\mathrm{vac}} - E_{e,\mathrm{rec}} - (\phi_{\mathrm{spectr}} - \phi_{\mathrm{source}}) - C \quad . \tag{61}$$

Here, E_γ is the γ-ray energy, E_b^{vac} is the binding energy of K-shell electrons related to the vacuum level, $E_{\gamma,\mathrm{rec}} = 0.0067$ eV is the energy of the recoil atom after γ-ray emission and $E_{e,\mathrm{rec}} = 0.120$ eV is energy of the recoil atom after emission of the conversion electron. ϕ_{spectr} is the work function of a retarding electrode of the main spectrometer and ϕ_{source} is the work function of the source. In the case of a mixture of both ^{83m}Kr and tritium

Figure 106: The γ-spectrometer with a Si(Li) detector.

within the WGTS, the term C will account in addition for possible space and surface charges within the gaseous source.

The experimental precision for the above parameters is the following: the γ-ray energy E_γ has been determined by semiconductor detectors [173, 174] with an uncertainty of $\pm 1.6\,\mathrm{eV}$ relative to the ^{241}Am standard of [175]. We have re-calibrated the results obtained in [173, 174] according to the new ^{241}Am standard of [176] and calculated a new weighted average of $E_\gamma = 32\,151.55 \pm 0.64\,\mathrm{eV}$. Beside this, the authors of [182] reported the value $E_\gamma = 32\,151.5 \pm 1.1\,\mathrm{eV}$ based on a measurement of seven conversion electron lines and experimental binding energies established relative to vacuum. The correction for surface shifts of binding energies in a quench condensed layer was calculated theoretically.

Since the change of the energy of the 33 keV ^{241}Am calibration line going from the old to the new standard was rather large (4.3 eV) and the recalibration of [173, 174] with original data was impossible, we started with additional measurements of E_γ using a Si(Li) detector (see fig. 106). The current value from our measurements amounts $E_\gamma = 32\,151.24 \pm 0.62\,\mathrm{eV}$. We expect that the final result, i.e. the weighted mean value from all four measurements, will reach a precision of about 0.4 – 0.5 eV.

Two different methods were employed for the determination of the binding energy of K-shell electrons E_b^{vac}: a) the photo-absorption of monochromatized synchrotron radiation

[177], yielding $E_b^{\mathrm{vac}} = 14\,327.2 \pm 0.8\,\mathrm{eV}$, and b) a combination of X-ray diffraction studies with X-ray excited photoelectron spectroscopy (XPS) on higher atomic shells. In this way, the authors [187] obtained $E_b^{\mathrm{vac}} = 14\,327.19 \pm 0.13\,\mathrm{eV}$. Recently, we reanalyzed all available experimental data from relevant electron and photon spectroscopies and came to $E_b^{\mathrm{vac}} = 14\,327.26 \pm 0.04\,\mathrm{eV}$ [188]. Inserting $E_\gamma = 32\,151.55 \pm 0.64$ into eq. (61) we derive the current value for $E_{\mathrm{kin}} + (\phi_{\mathrm{spectr}} - \phi_{\mathrm{source}}) + C = 17\,824.18 \pm 0.64\,\mathrm{eV}$. In addition, we have increased precision of binding energies for the L_2 and L_3 subshells in gaseous krypton, $1\,731.91 \pm 0.03$ and $1\,679.21 \pm 0.03\,\mathrm{eV}$, respectively. These values will allow to calibrate the KATRIN energy scale extending up to 30 keV.

Gaseous $^{83\mathrm{m}}$Kr source

Even more important than the absolute energy calibration is the use of gaseous $^{83\mathrm{m}}$Kr within the gaseous tritium determining the central value and the distribution of the electric potential within the WGTS precisely. For this purpose we will run the WGTS with a $^{83\mathrm{m}}$Kr admixture to the gaseous T_2. To avoid freeze-out of the krypton we need to increase the WGTS temperature to 120 – 150 K. To determine the electric potential within the WGTS by comparing the electron line position with and without tritium we are not restricted to the electron energy close to the β endpoint. We therefore can profit from the sharpness of the high energy L_3-32 or N-32 conversion lines. Simulations have shown that the fluctuations of the electric potential within the WGTS –over its volume or with time– of about 5 mV (see section 11.4.8) will cause observable broadening of the electron line. The space charge potential variation inside the WGTS is expected to be proportional to the tritium gas temperature, therefore the measured result at 120 – 150 K will give an upper limit on this potential variation at the 27 K standard operational temperature.

Condensed $^{83\mathrm{m}}$Kr source

A quench-condensed film of sub-monolayer thickness on a graphite substrate has been used at the Mainz Neutrino Mass Experiment very successfully for different systematic investigations [195, 182, 106, 94] as well as for calibration purposes.

A small vacuum vessel containing ^{83}Rb implanted into a polyamide foil generates gaseous $^{83\mathrm{m}}$Kr emanating from the foil. The krypton gas is purified by a cold trap and quench-condensed on a highly oriented pyrolytic graphite (HOPG) substrate. By laser ellipsometry the substrate and film properties were checked. The advantages of this source are its very easy maintenance, the isotropic electron emission, the possibility of very high count rates, the constant electron emission rate, which is only governed by the 1.83 h half life and compatibility with UHV requirements.

With regard to calibration and monitoring the Mainz setup showed one major disadvantage: From one $^{83\mathrm{m}}$Kr film to another, the reproducibility of the line position was about $\pm 0.2\,\mathrm{eV}$ (short-term) and $\pm 0.6\,\mathrm{eV}$ (long-term) only. The short-term (day-to-day) fluctuations were attributed to variations of the different substrate and film conditions, the long-term (over 3 years) stability seems to be mainly governed by instabilities of the high-voltage system.

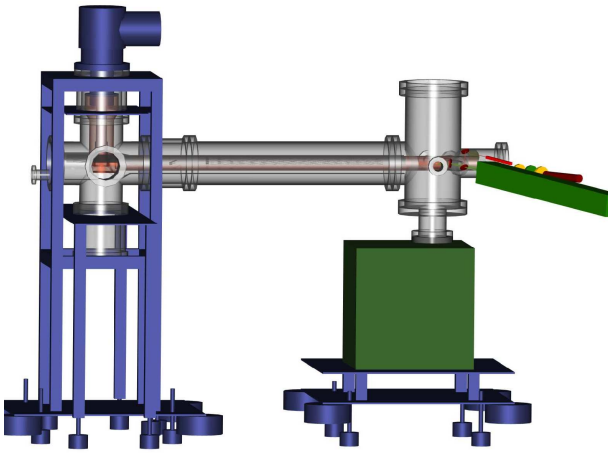

Figure 107: Sketch of the setup of the Condensed Krypton Source. The vibrating cold head sits on the outer left stand. It is mechanically decoupled from the cold finger with the substrate which is suspended from the inner left stand. A long bellow allows to move the substrate from the condensing position with the ablation and the ellipsometry lasers (vacuum vessel with windows on the right stand) to the measurement position inside the monitor spectrometer (not shown). The whole setup can be put on high voltage to adopt the K-32 line position to the retarding voltage of the main spectrometer.

Currently we are constructing a new quench-condensed krypton source system based on the Mainz design but with much better substrate and film control (see fig. 107). A NdYAG laser system will anneal the graphite substrate and thus provide reproducible substrate conditions before each film preparation. The whole gas inlet and purification system will be built with UHV components. We consider to preplate the substrate by a film out of neon, which can be condensed very cleanly at temperatures below 10 K due to its lower sublimation temperature compared to other vacuum components. Laser ellipsometry will be used to measure the thickness of the pre-plated films (see fig. 108). The whole system will be setup in such a way, that the annealing and film preparation can be performed remotely, repeatedly and with high reproducibility.

Solid ^{83}Rb/^{83m}Kr source

To avoid the necessity to repeat the ^{83m}Kr film quench-condensation every few hours due to the short half life of ^{83m}Kr, we are considering to build a solid ^{83}Rb/^{83m}Kr electron source. ^{83}Rb will be evaporated in vacuum onto a HOPG or metal backing. ^{83}Rb decays with a half-life of 86 days, thereby continuously generating ^{83m}Kr. The aim is to capture the ^{83m}Kr within the ^{83}Rb sample till its decay. Moreover, such a source would be more convenient for handling due to its compactness. The analogous ^{83}Rb evaporated source was prepared in works [183, 191]. Currently we are investigating, whether such a mixed source has reliable long-term stability and whether it is compatible with UHV conditions. In particular, we examine a possible release of rubidium compounds out of the source.

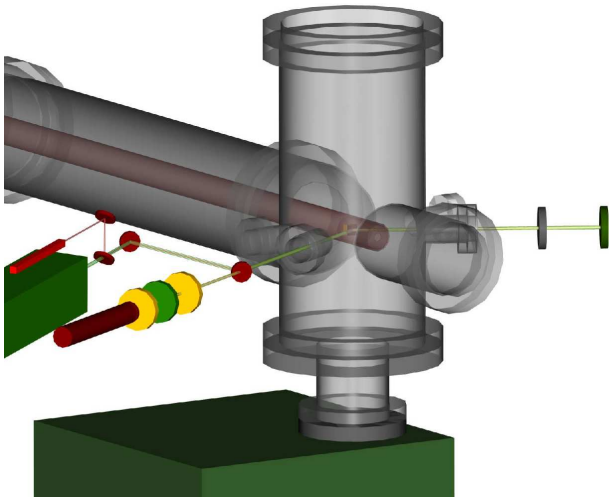

Figure 108: Scheme of the Gas inlet and Control System. The ^{83m}Kr and other gases can be let directly to the substrate. Alternatively a way through a monitor cell with all standard analysing instruments is possible. All valves are computer controlled.

Production of ^{83}Rb

As the ^{83m}Kr has a short half life of 1.83 h it can be obtained for practical use only from the decay of its mother isotope ^{83}Rb which has a half life of 86.2 d. The production of ^{83}Rb is currently tested at Bonn/Germany and Řež/Czech Republic cyclotron facilities.

In the Bonn feasibility study, a beam of alpha particles with an energy of 55 MeV on liquid bromine (natural isotopic abundance) in a quartz tube target was used. After the irradiation, the ^{83}Rb activity of 1.7 kBq, produced via the ^{81}Br(α, 2n) reaction, was indicated as the main component in the gamma spectrum of the target. In a further step the top of the quartz tube was opened and the bromine was allowed to vaporize. An analysis of the gamma spectrum of the quartz tube showed that the produced ^{83}Rb remained completely in the tube, presumably as RbBr. It is expected that the ^{83m}Kr produced by decay of rubidium absorbed on the glass wall will freely escape for the use in WGKrS or QCKrS. Using the gain factors –beam intensity, weight of bromine in the target, irradiation time, and isotopic enrichment (the abundance of ^{81}Br in nature amounts to 49.3 %)– the ^{83}Rb activity of several GBq could be reached.

The classical method for the ^{83}Rb production [192], namely (p, xn) reaction on a target filled with natural gaseous krypton was used at Řež cyclotron. The primary beam energy of H^- amounted to 27 MeV. Routinely two 10 hours irradiations were accomplished—in each the ^{83}Rb was produced with activity of about 85 MBq. By enhancement of the irradiation time to 100 h and using the isotopically enriched ^{84}Kr (57.0 % in nature), an activity of 1 GBq can be obtained. The produced ^{83}Rb is washed out of the target walls with 90 % elution probability into 25 cm^3 of distilled water. After that, the solid source for the gamma spectroscopy measurement was prepared by passing the rubidium water stock solution through cation exchange paper. At standard conditions, 80 – 90 % of rubidium is absorbed in cation exchange matrix. Unfortunately, exchange paper is able to release

gaseous ^{83m}Kr only at rather high humidity levels totally disabling the use of it in the vacuum chamber. The idea of [193] to prepare the ^{83m}Kr gaseous source by absorbtion of ^{83}Rb from aqueous solution in molecular sieve (zeolite) was successfully tested. One milliliter of rubidium solution (3 MBq of ^{83}Rb) was absorbed in 2 grams of zeolite. Then the zeolite was dried to remove the solvent. Such a sample was a subject of several tests in order to check it for easy release of ^{83m}Kr into the vacuum and good retainment of ^{83}Rb in the zeolite. The gamma spectroscopy was used for detection of ^{83m}Kr (32 keV γ-line) and the ^{83}Rb (520 keV γ-line). Even severe tests like heating the sample for 4 hours at 300 °C in a vacuum of 10^{-6} mbar showed no release of ^{83}Rb at the level of 0.7 %. On the other side, measurements of the 32 keV line showed that the zeolite matrix allows a complete release of ^{83m}Kr into the vacuum.

9.2.2 ^{241}Am/Co photoelectron source

Precise energy calibration of electron spectrometers at energies up to several keV is often performed by means of photoelectrons induced by X-rays. However, increasing the natural width of exciting X-radiation from heavier elements decreases the accuracy of this method at higher electron energies. To provide a sharp electron source with an energy very close to the endpoint energy of the tritium β spectrum we are constructing (inspired by [189]) a ^{241}Am/Co photoelectron source: γ radiation from a ^{241}Am-source hits a thin cobalt foil. The photoelectrons ejected by $26\,344.6 \pm 0.2$ eV [176] γ-ray photons of ^{241}Am from the atomic K-shell of metallic cobalt with binding energy $E_{b,F}$ of $7\,708.78 \pm 0.02$ eV [187] have a kinetic energy $E_{\rm kin}$ close to the endpoint of the tritium β-spectrum, in particular,

$$E_{\rm kin} = E_\gamma - E_{b,F} - E_{e,\rm rec} - \phi_{\rm spectr} \quad , \tag{62}$$

where E_γ is the γ-ray energy, $E_{b,F}$ is the binding energy of K-shell electrons related to the Fermi level, $E_{e,\rm rec} < 0.2$ eV is the energy of the recoil atom after photoelectron emission, and $\phi_{\rm spectr}$ is the work function of a retarding electrode of the main spectrometer. Indeed, such a ^{241}Am/Co photoelectron source would be suitable for our purpose:

- the energy of monitoring photoelectrons, 18.636 keV differs from the tritium endpoint only by about 60 eV and the calibration line would be *above* the β–spectrum
- the natural width of exciting γ–rays, its Doppler broadening at 300 K and recoil energy are less than 0.02 eV, i.e. completely negligible for our purpose
- the natural width of atomic K-shell in cobalt is 1.3 eV
- the ^{241}Am half life of 432 y is practical for long term monitoring
- a ^{241}Am/Co source may suffer by physical-chemical changes of the binding energy. Differences of the binding energies of the Co metal component and possible Co oxides are in the range of 1.9 – 2.1 eV. Photoelectrons corresponding to the metal Co component originate with higher energy than the ones corresponding to Co

Figure 109: Electrostatic electron spectrometer ESA12 at NPI Řež. A double-pass cylindrical mirror analyzer is combined with a preacceleration/preretardation system.

oxide states. The differences are both well described and observable in the monitor spectrometer. Moreover, the effect may be suppressed by ion etching

A ^{241}Am/Co photoelectron spectrum up to 14 keV was measured at the ESA12 electrostatic spectrometer in Řež (see fig. 109) with 5, 3, and 0.1 μm thick cobalt foils. The photoelectron lines corresponding to several ^{237}Np L–X-ray transitions (originating in ^{241}Am to ^{237}Np decay) are shown in fig. 110. There is displayed also the corresponding Monte Carlo simulated distribution of photoelectrons emitted from the surface of a cobalt foil. In our Monte Carlo calculation (model approach described in [194]), the arrangement of the ^{241}Am emitter, angular distribution of photoelectrons, their scattering and energy losses in a cobalt layer, and also the setup of the ESA 12 spectrometer were taken into account. A reasonable conformity of the MC simulation with the experimental data is apparent. Some deviations may result from uncertainties in the structure of the ^{241}Am emitter, ESA 12 transmission and detector efficiency.

First tests of ^{241}Am/Co in the Mainz spectrometer were performed. A basic structure of the integral photoelectron spectra at 18 keV was observed. High background caused by direct incidence of γ-rays onto the detector and scattered electrons was substantially reduced by both the enhanced setup of an ^{241}Am/Co source and a proper thickness of the

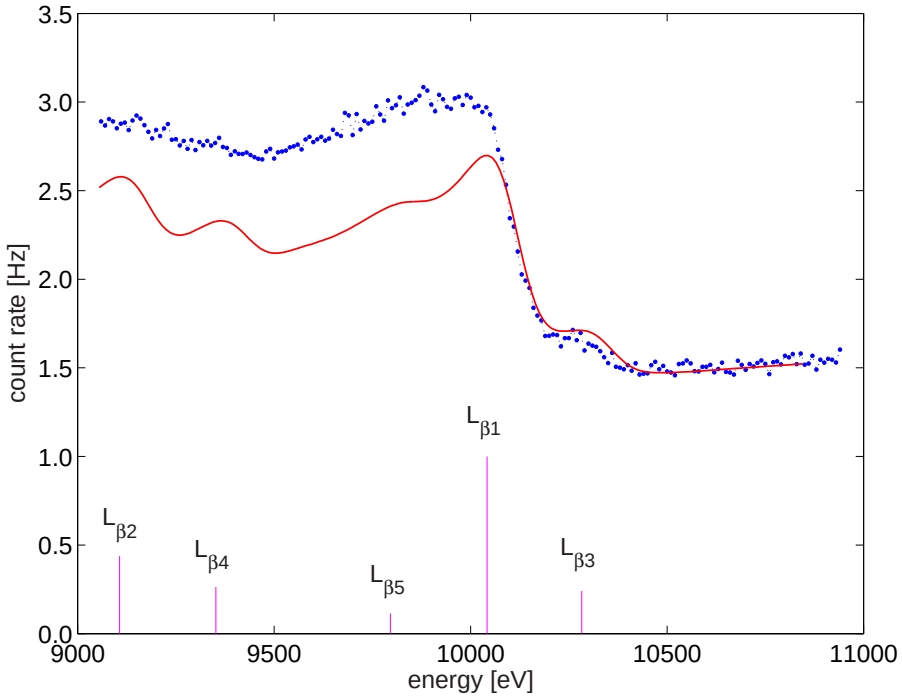

Figure 110: A part of the ^{241}Am/Co photoelectron spectrum taken in the ESA12 spectrometer with a 5 μm foil. The solid line represents a Monte Carlo simulation of ^{237}Np X-ray photoelectron energy distribution in this region (see text). The instrumental energy resolution and transmission of the spectrometer are 100 eV and 0.74 % of 4π, respectively. The vertical bars show the positions of X-ray photoelectrons. The heights of the bars are proportional to the tabulated ^{237}Np X-ray intensities.

Co foil. Further improvement of the ^{241}Am/Co arrangement in the Mainz spectrometer will be performed. The final feasibility test is planned in 2005.

9.2.3 ^{109}Cd Auger electron source

Another useful monitoring source could be ^{109}Cd. The relevant properties of this nucleus are:

- electron capture decay into ^{109}Ag [196]
- almost no gamma emission especially no hard gammas, but low energy X-rays
- emits in 9 % of decays KL_2L_3 ($^1D_2$) Auger electrons of $18\,511.7 \pm 1.3$ eV energy for ^{109}Cd in a particular chemical state [197]
- the natural width of the KL_2L_3 line is 11.2 eV [186]
- half life of 463 days.

Although the line width of the KL_2L_3 ($^1D_2$) is rather broad compared to 50 mV, the required long term stability of the KATRIN filter voltage, it can be used at least complementarily to the other envisaged nuclear standards. It is expected that thin metallic sources of more than 200 kBq can be made. The count rate can be estimated as:

$$\text{Counts/s} = A \cdot \eta_s \cdot I(\text{KLL}) \cdot I(\text{KL}_2\text{L}_3\ (^1\text{D}_2)) \quad , \tag{63}$$

where A is the source activity, $\eta_s = 0.25$ is the assumed efficiency of the spectrometer, $I(\text{KLL}) = 0.21$ is the total intensity of KLL—Auger electrons, $I(\text{KL}_2\text{L}_3(^1\text{D}_2)) = 0.41$ is the fraction of the $^1D_2$ fine structure component of the KLL line [197]. This gives an estimated count rate of 4300 counts/s.

^{109}Cd can be produced by thermal neutron capture on ^{108}Cd (σ = 1 barn). The natural isotopic abundance of ^{108}Cd is only 0.89 % so that an isotopic enriched target for neutron irradiation has to be used. Since a source thickness of more than 3–5 mono layers is not useful, it is planned to produce the target by ion implantation of ^{108}Cd into either titanium or vanadium. Both metals are suitable since radioactive isotopes produced in a thermal neutron flux have half lives shorter than 6 minutes.

The ion implantation can be done with the ECR-implanter at the former Cyclotron department of the Forschungszentrum Karlsruhe. A Cd-beam with an intensity of 100 nA (6×10^{11} ^{108}Cd/s) has been produced before. Since a source of low energy electrons should be very thin, the ion beam needs to be decelerated, so that it hits the target with an energy of a few hundred eV only. An appropriate decelerating electrode needs to be built, which is not very difficult and costly.

Even though the Auger line width is 12 eV, it is still possible to use it for monitoring purposes, since rather strong sources (200 kBq/cm^2) can be produced. This should enable the monitor spectrometer to detect a shift of 50 mV of the retarding voltage within few hundred seconds. Since the source is of metal, it is suitable for UHV applications, although it should not be heated to more than 100 – 200 °C since cadmium melts at 320 °C and is easy to oxidize.

9.2.4 High-voltage photoelectron source

To measure the transmission function of the main spectrometer we can use a photoelectron source illuminated by ultraviolet light. By applying a high voltage to the photocathode such a suitable energy and a narrow width can be achieved. The angular emission characteristics can be modified by the shape of the photocathode and the surrounding electrodes. The high voltage of the source will be the retarding voltage plus a small scanning voltage. The excess energy of the photo electrons is therefore very easy to control.

9.3 Monitor spectrometer

The spectrometer of the Mainz Neutrino Mass Experiment is available for KATRIN. Currently it is being used for background and electromagnetic design investigations. The energy resolution of the Mainz spectrometer is 4.8 eV, a value which has been chosen as

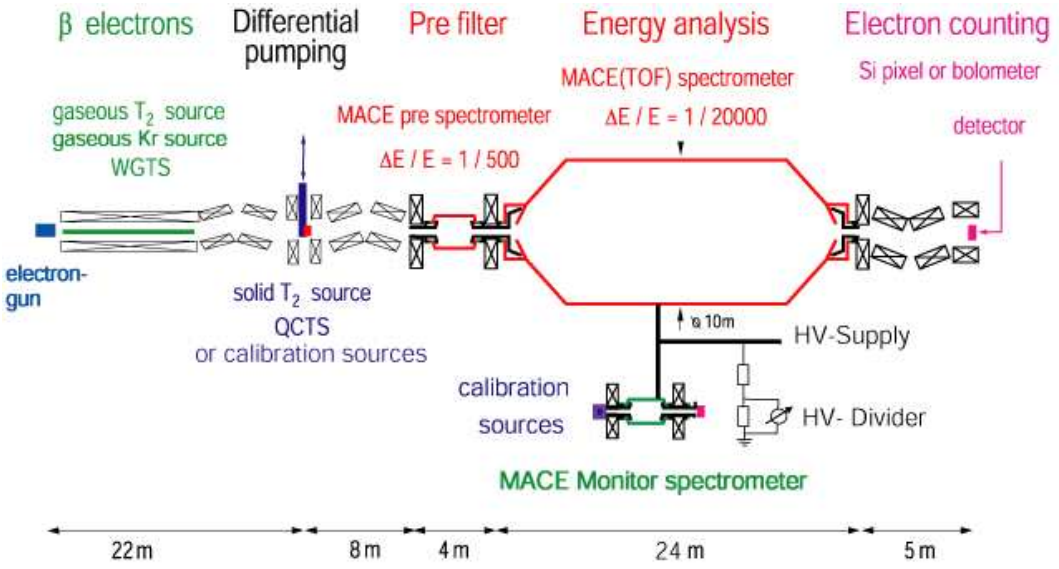

Figure 111: Setup of the monitor spectrometer beam line with a calibration source following an atomic/nuclear standard and fed with the main spectrometer retarding potential.

a compromise between energy resolution and luminosity for tritium β spectrum measurements

For KATRIN we will upgrade the existing Mainz spectrometer into a high resolution spectrometer. We will improve the energy resolution by a factor of 5 down to 1 eV. The corresponding reduction in luminosity does not play a role for calibration and monitoring purposes. The idea is also to apply the retarding voltage of the KATRIN main spectrometer to this new monitor spectrometer (see fig. 111). A well-defined electron source will then be measured by varying the voltage of the electron source allowing practically continuous monitoring of retarding voltage. Such low voltages well below 1 kV can be measured very precisely and reliably. The work functions of the monitor spectrometer and the main spectrometer are very similar (practically identical) as in both cases stainless steal under XHV conditions is used. Absolute calibration additionally requires the precise knowledge of the parameters of both sources and the potential distribution in the analyzing plane of both spectrometers. Electron sources which may fulfill the above requirements are the ^{241}Am/Co photoelectron source, the quench-condensed ^{83m}Kr source or the ^{83}Rb/^{83m}Kr source described above.

9.4 Absolute energy calibration

We will use the gaseous ^{83m}Kr source in the WGTS as well as the condensed ^{83m}Kr source and the ^{241}Am/Co source at the position of the alternate QCTS in the split coil magnet to calibrate the main spectrometer directly. Although in contrast to calibrations of the high voltage at the monitor spectrometer these measurements require parts of the time available for tritium β decay measurements, they are more direct and give redundant information,

e.g. they probe also time variations of the work function of the main spectrometer.

As pointed out above we will use the conversion and photoelectron sources not only to monitor the stability of the high-voltage systems and work functions but also use them to do an absolute calibration of the retarding energy to compare the fitted β endpoint with the tritium-helium mass difference $\Delta m(^3He -^3 H)$, determined by cyclotron resonance measurements in Penning traps. Any significant difference will point to unaccounted systematic errors. The precision of the last helium-tritium mass difference measurement of $\Delta m(^3He -^3 H) = 18\,590.1 \pm 1.7\,eV$ [179] corresponding to a $5 \cdot 10^{-10}$ relative uncertainty in the absolute mass determination is precise enough only to find gross errors. Meanwhile the technique of mass measurement in Penning traps has been improved by one order of magnitude. New attempts are being started to improve the precision available for $\Delta m(^3He -^3 H)$ [180].

An ideal scenario for KATRIN would be to use the external endpoint energy from a very precise $\Delta m(^3He -^3 H)$ determination as a fixed input into the fits of the β spectrum. To achieve an improvement upon the usual fit with free endpoint E_0, two conditions have to be fulfilled: first, two orders of magnitude better precision of the $\Delta m(^3He -^3 H)$ measurement; second, a possibility to calibrate the retarding potential to the same precision.

9.5 Electron flux monitors

As mentioned in section 4 we will frequently monitor the WGTS strength by lowering the retarding potential to facilitate rapid determination of the electron flux from the source with precision better than 10^{-3}. Additionally we will install electron flux monitors behind the source to monitor the electron flux over wide ranges of β energies and to check the stability of the source operation.

9.6 Consistency tests and simulations

Concurrently with the experimental investigation, extensive Monte Carlo simulations of the tritium β-spectra are performed. The aim is to estimate quantitatively the influence of various possible uncertainties in the energy calibration as well as possible imperfections in the long-term HV stability on the neutrino mass deduced. All these studies stem from many years experience in β-decay experiments at Mainz and at Troitsk and on electron spectroscopy at Řež/Prague.

To check the stability of the KATRIN measurement conditions –in particular, to determine whether the conditions are changing or not and, if so, to determine at which minimum level the change can be detected– we utilize statistical tests. Our previous tests [184] applied to β-spectra, simulated as if measured under slightly different conditions, revealed the discrepancies with great sensitivity. (Typically, a shift of energy scale as small as 40 meV may be discovered after several hour measurement of the tritium spectrum near its endpoint. For the case of the calibration conversion line in ^{83m}Kr, the time needed is of order of minutes.) Additionally we adopted five other statistical tests (in particular,

χ^2, mean-value, sign, normal-distribution-of-residuals, and Kolmogorov-Smirnov ones) to study the consistency of β-spectra. It turned out that these tests are very powerful means.

An alternative approach is to fit various simulated β-spectra and study the consistency of the resulting parameters (*e.g.* neutrino mass and endpoint E_0, see also section 11.6). This method is used, too, which gives us a desirable redundancy of conclusions.

10 Slow Control and Data Acquisition System

10.1 Slow Control

The KATRIN Slow Control (SCS) and Data Acquisition Systems handle the management of all measurement data of the experiment other than the event data from the detector. This includes both hardware and software. Main characteristic of the KATRIN SCS is its homogeneous structure with its code being spread over several sub-systems. The deployment of the system is done from a single notebook. Implications of this distributed method are system stability, independency of sub-systems and fast and comfortable maintenance.

The KATRIN SCS is a new generation control system being based on following concepts:

- acquisition and control algorithms homogeneously spread over sub-systems,
- consequent use of model-view-controller architecture,
- use of new web technologies with world-wide access on experimental data,
- use of high level interface descriptions and data protocols and
- use of software components with graphical programming

SCS is not concerned with the detector data acquisition which is being undertaken by the CENPA-group of the University of Seattle (see section 10.2).

For system design both the view of a user onto the KATRIN system and the view of the developers play decisive roles. Main aspects for the user system view are (a) ease of use, (b) distributed data access, and (c) cost-effective solution. (a) Ease of use means intuitive operation of control and data handling, and flexible presentation of system information. The basis of this ease of use is the storing of all data (measurement data as well as user specified data) in a database. A second basis is the use and creation of standards concerning interfaces, protocols and architectures. (b) Distributed access to the system is convenient for a user because of the more wide-spread availability of cell phones and PDAs. On the system part, data acquisition and slow control is distributed on many computers and hardware devices. Newest internet technologies like web-services, XML and XQuery provide the possibility of high level standards and platform-independence. The trend of future control systems and therefore as well for the KATRIN SCS is the accessibility of each (even small) device by a web interface. Besides these requirements it is a fact that data is produced by several sub-systems. One philosophy of the present SCS approach is to gain a stable system by using components able to operate stand-alone. (c) A cost effective solution for control systems for large experiments as KATRIN is only possible if common and application independent parts of the system can be identified and can be encapsulated by modules and hard- and software components. Only then is a short design period achievable.

This view of a general system user has implications on the priorities of a system designer. For him also (a) a short design period and (b) high reusability is desirable.

(a) The design period may be reduced drastically when using graphical programming together with component software. Here both commercial and open source or proprietary development tools are useful. Especially the combination of both guarantees short cycles of development. The necessity of using proprietary tools is the lack of diagnosis, self-tests and documentation of today's commercial software component systems that is highly desirable for dynamically plugable system components. (b) These component systems also provide good reusable soft- and hardware. New internet technologies like web-services aim in the same direction. But still today several standards are missing: for KATRIN and other control systems the most crucial spot is the lack of common data types for data acquisition and control. Here a sound basis should be the new XSchema data specifications that may be enlarged by control system specific requirements.

More specific requirements are the:

- highly reliable, robust construction of hard- and software modules to work in extreme environment,
- ability to operate in different phases of the experiment (calibrating, baking, cooling, running etc),
- high demands to network safety with flexible communication watchdog system,
- handling of alarms and warnings and secure logging of slow control events and
- solving time-critical responds in sub-system parts of the experiment.

For a stable system design efforts are concentrated both on proven, reliable, commercial hardware and software technology (here mainly from National Instruments and Oracle) and proprietary concepts.

10.1.1 System architecture

For the KATRIN SCS a modern architecture was chosen with clear structural distinctions. The first distinction is the local distribution. Close to the measurement devices other real time processing units are located. Fast decisions can only be done within these sub-systems. SCS cares about data acquisition and control of the following sub-systems: (see fig. 112)

- pre-spectrometer
- main spectrometer
- detector
- high voltage system
- ultra high vacuum system

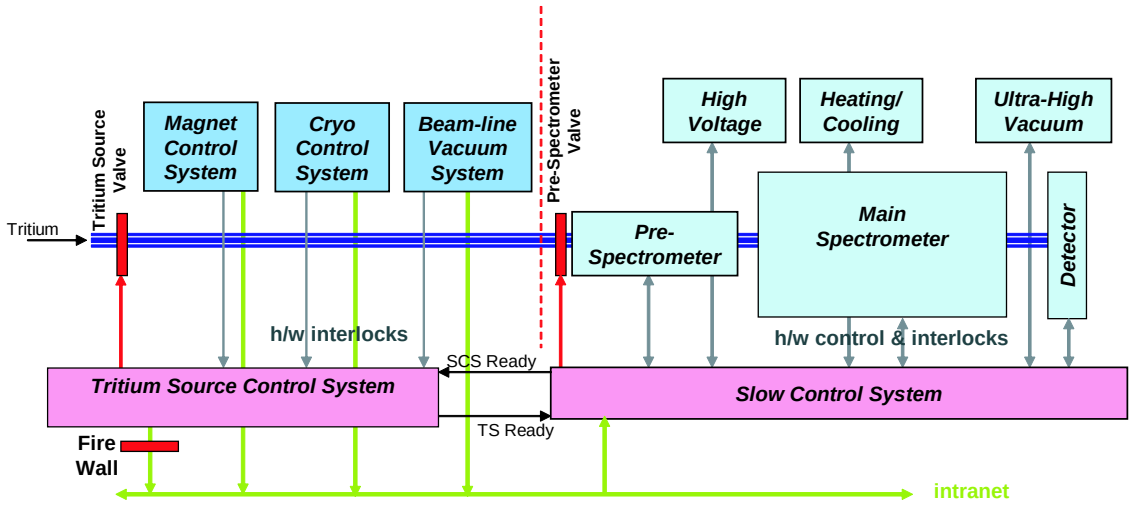

Figure 112: Separation of duties between Tritium Control and Slow Control systems

- heating/cooling equipment (via serial interface from heating/cooling Control System)

All above mentioned systems will stay under h/w and s/w support, including control, monitoring, documentation and description of slow control channel. Meanwhile, due to the high safety and security demands, the following subsystems will stay under control and monitoring of the tritium control system of TLK (see section 4.6), including data taking and h/w interlocks:

- super-conducting magnet system
- cryo supply
- vacuum system for beam-line (up to the entrance valve to the pre-spectrometer)

The SCS task there is the visualization of their status on the standard interfaces used by the supervisor.

Very special attention must be paid to Tritium Source (TS) because of high safety and security demands. TS acquire all necessary data from experiment using independent from SCS special control system and only send to the Supervisor "Source Status" information via intranet from TS' database-server. Only limited h/w signals can be exchanged between TS and SCS. TS control system can use "SCS Ready" flag-signal from SCS to control source output shutter. SCS will use "TS Ready" signal to send to the TS Control system command to open Pre-spectrometer entrance valve. This operation will be successfully completed only with the following necessary preconditions:

1. "TS Ready" signal is active (h/w interlocks from the magnet, cryo and beam-line systems are active) AND

2. H/w interlocks from SCS side (HV, UHV, Heating/Cooling systems are active) AND

3. operator action from the Supervisor Computer (password protected)

Such mutual cooperation of two independent control systems will provide high safety and reliable operation of the experiment, keeping both spectrometer vessels from the penetration of Tritium.

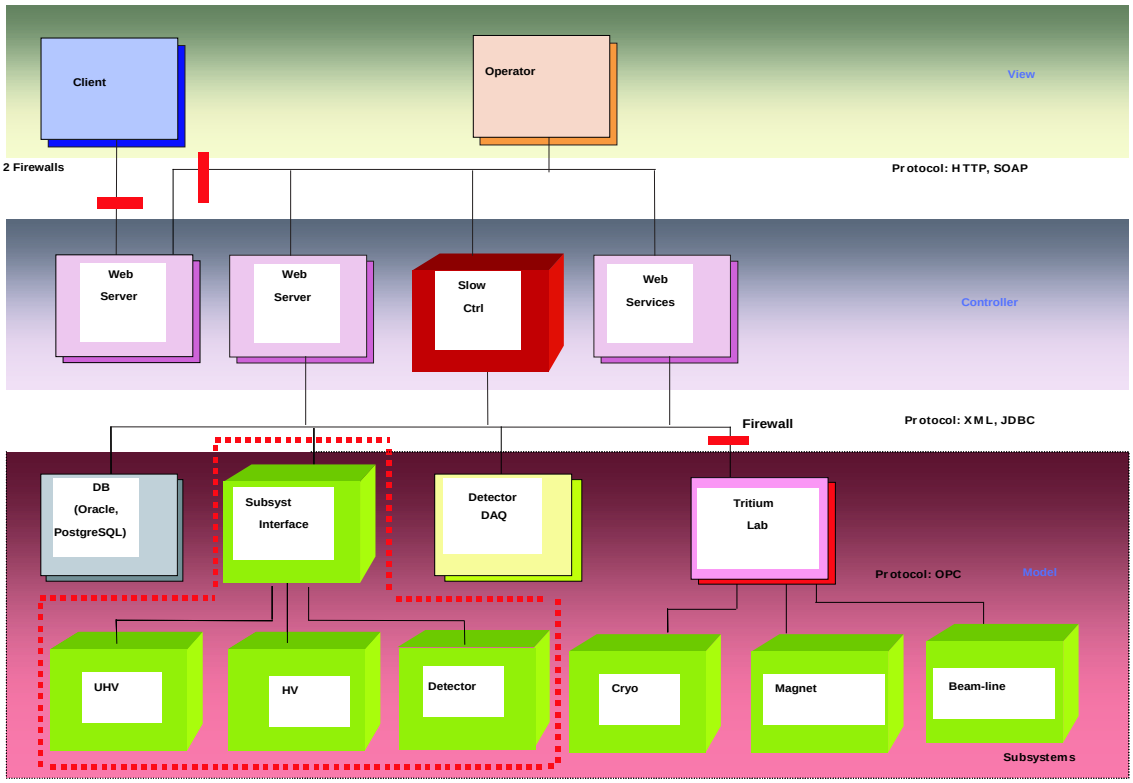

Figure 113: KATRIN Slow Control Data Flow

The Data Flow diagram of the KATRIN experiment is presented in fig. 113. On the lowest h/w level, SCS controls and monitors the subsystems despite of their different complexity via FieldPoint banks from National Instruments. Such devices provide control facilities, a real time data buffer and a connection to the intranet. Intranet connections include access to the central database, other central control tasks and the graphical user interface of the operators.

In this domain of the intranet commercial software (LabView from NI) is combined with proprietary JAVA procedures and web-services. The data logging and supervisory control module (DSC) from NI is used for the operator graphical user interface. The third stage is the internet with world wide access to enabled data. Security aspects are taken very seriously and, if possible, excluded by architectural considerations like double firewall, double data base with replicated data and access restrictions for unknown users.

The second structural element is a basic software engineering paradigm, the model-view-controller (MVC) design pattern. It is extremely useful to provide a clear distinction between data model, presentation of the data for the user, and handling of the user input. MVC may be combined with the new standardized web-services and their proper structure of service requester, service provider, and service broker.

Associated with these architectural elements is a consequent modularization. To create reusable and application-independent components the interfaces and protocols used between the modules are most important. Different proposals were made over the last few years [198, 199, 200, 201] concerning the structural coupling of modules. Also for a flexible plug-in and plug-off a standard data type definition is necessary. Such specification is application-dependent. There is some hope that we can find a solution for the domain of data acquisition and control systems. First steps were made with a proposal based on the new XSchema recommendations of the world wide web consortium W3C. The big advantage when using such general concepts is their applicability to both data base access and measurement device access. In general, measurement device data are logged chronologically to provide a history of the experiment. Therefore it is an essential issue to have a scheme to identify all sources of slow control information in a unique way. We propose to distinguish such slow channels (SC), into dynamic and static data. This was first done in the H1-experiment [202]. The dynamic data are measurement data like Value, Status and Time-of-Measurement. Static attributes are user defined limits for significant data changes, warnings, alarms and channel descriptions. A very efficient data reduction may be implemented in the distributed slow control units by comparing dynamic with static data. Only dynamic data not filtered out are stored in the real time buffer. They are called slow events (SE), in contrast to physical events. The SE's are generated when passing a filter. The function of the filter is twofold: creating Alarms and Warnings, and passing through only data fulfilling special requirements. Fig. 114 shows two examples with different situations where different thresholds Significant Change limits. Any exceed of threshold creates a SE with correspondent Time stamp, Status, Value and SC identifier data.

A flexible control system designed and used by many world-wide distributed collaborators should be independent of a special database. The present standard accessing databases SQL is not enough to ensure database independency. Additional drivers (ODBC, JDBC) are required as well as additional conventions for stored procedures and SQL extensions. A recent approach to overcome these problems again comes from W3C with their XPath and XQuery specification recommendations. In KATRIN these preliminary specifications at least will be used for web access.

For large experiments with international collaborators web access to experimental data is mandatory. A good example for how to do that is the TOSKA experiment [203]. In general, access to the data is read only in order not to influence the course of the running experiment. Nevertheless, for the operating staff it is extremely convenient having some remote control access by web to the experiment. Considering all security problems there might be a solution to give some privileged persons access rights to toggle few command flags.

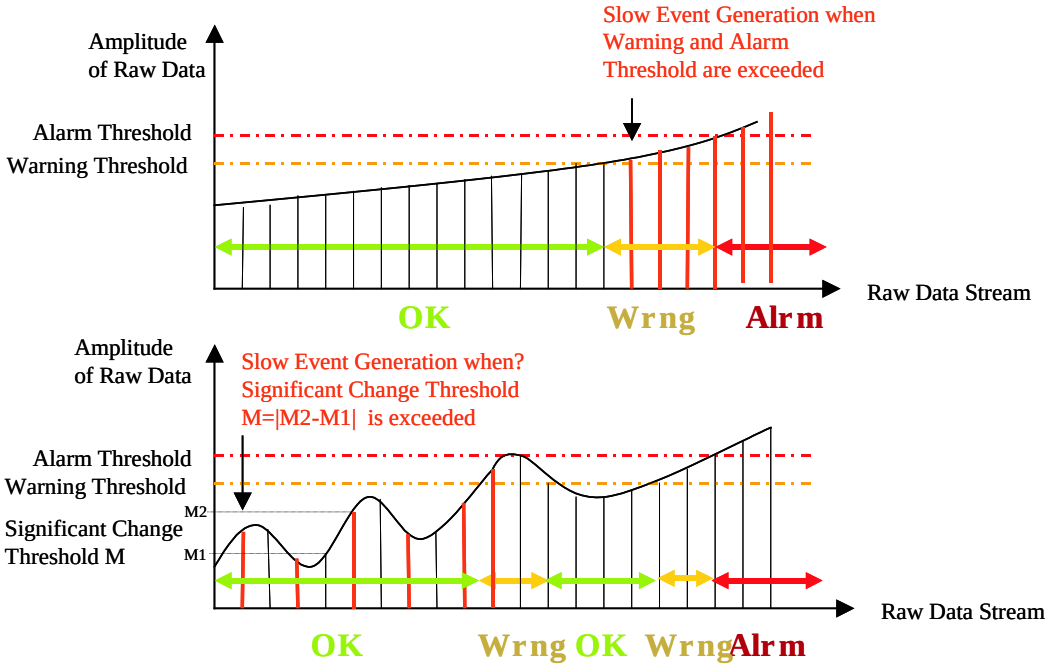

Figure 114: Slow Event Generation Mechanism

10.1.2 Hardware

Due to the large size and distributed architecture of the experiment centralized control hardware would be inappropriate. The FieldPoint Distributed I/O from National Instruments, based on the LabVIEW Real Time software, seems to be a qualified solution to that problem. All Slow Control components will break up into the few local FieldPoint (FP) stations along the experiment. They could have different setup settings dependent on their control task, but their main structure will be the same: the FP processor with a LabVIEW Real Time system running inside to collect data into the slow event real time buffer before transferring them to the next processing layer. Such stations could be created at different points around the experiment for all subsystems. Also, there might be a case that some neighbouring subsystems, which have only few channels or interfaces, will merge into one FP station. Such architecture allows us to have a universal, flexible and reliable hardware, running under commercial software (see fig. 115).

For the pre-spectrometer and test cylinder a rack-mounted station has already been created and tested. This setup could be a prototype for future FieldPoint stations for the experiment. The prototype consists of a 19-inch crate divided into the two segments: the FieldPoint hardware, a DIN rail mounted at the back side and all manual control buttons and indicators on the front. Again it should be mentioned that they are subsystem-dependent.

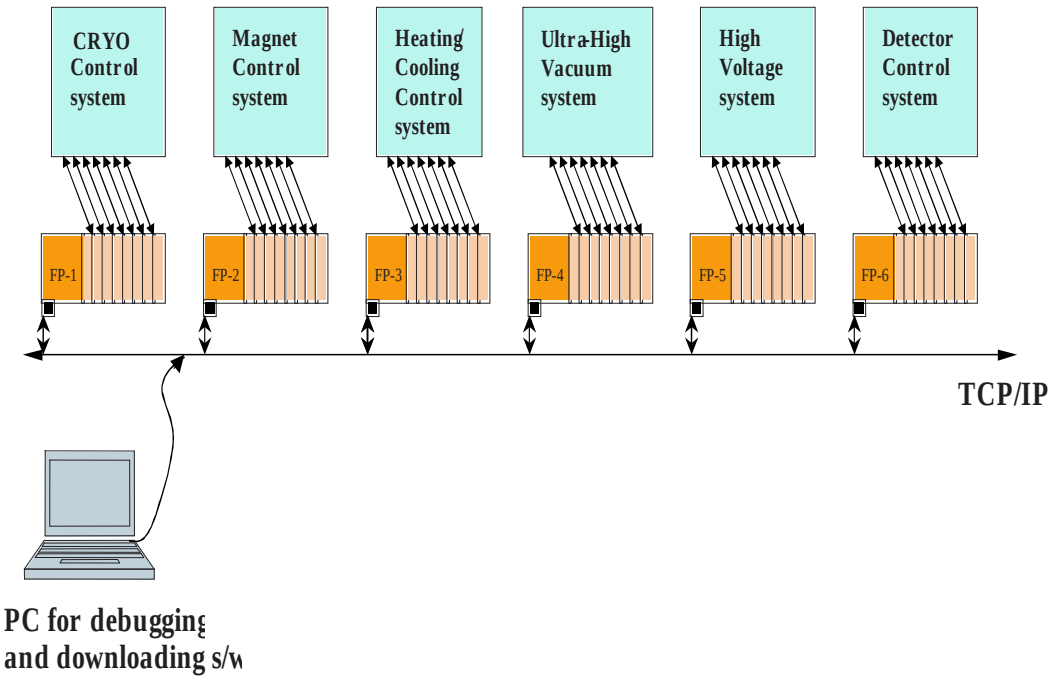

Figure 115: Distributed Slow Control with FieldPoint devices

10.2 Data Acquisition System

The data acquisition system (DAQ) is responsible for the processing and storage of the analog signals of the different particle detectors within the KATRIN setup, especially the focal plane detector. The information stored by the data acquisition system is merged with the data from the Slow Control System. The softwares of data acquisition and slow control system must provide all necessary tools for efficient and reliable operation of a large scale experiment.

10.2.1 Requirements

The requirements on the KATRIN data acquisition system are defined by a wide range of different measurement modi for calibration purposes and investigations of systematics. As it turns out (see table 4), the data acquisition system must be capable to handle event rates spanning over nearly nine orders of magnitude from few mHz during neutrino mass measurements up to Megahertz rates during Krypton measurements. The number of data acquisition channels is determined by the design of the focal plane detector. Depending on the final chosen detector architecture total channel numbers between 400 and 900 channels are expected.

The most important information to be stored and analyzed is the pulse height of the focal plane detectors preamplifiers output, which reflects the deposited energy in the detector. Considering an energy range to be surveyed of $5-50\,\mathrm{keV}$ together with the aimed for energy resolution of $600\,\mathrm{eV}$ (FWHM), a 12 bit analog-digital converter (ADC)

for the digitization of the pulse height value is adequate.

As already outlined the position information is of crucial importance in the offline data analysis and is just defined by the position of the triggered pixel.

A time information of each event is needed to synchronize data from the focal plane detector with data from the active veto system, which surrounds the detector area for suppression of cosmic induced background. For that purpose, a timing precision of $\delta t < 500\,\text{ns}$ is sufficient. Each event must have a time stamp, allowing to synchronize and merge the data acquisition data with the slow control data. The data acquisition should be expandable and must feature the possibility to process signals from additional detector systems, like photomultiplier signals from the active muon shielding or signals from tritium activity monitor detectors.

10.2.2 Hardware

The DAQ system under investigation will essentially consist of VME based custom-made hardware and will be similar in its hardware design and logic to the data acquisition hardware used in the AUGER experiment [204]. To account for the wide range of event rates the data acquisition system is foreseen to be designed in such a way that one can easily and fast switch from *single event mode* to *histogramming mode*. In single event mode the full information (see below) for each channel is stored, whereas in the histogramming mode just the trigger information is used and the DAQ channels functionality is reduced to a scaler function. A very flexible re-programmable 2-level hardware trigger introduced by a field programmable gate array (FPGA) technique is able to accommodate for the large variety of KATRIN measurement modes. The geometry and design of the electron detector system sets the number of required channels of the DAQ hardware. The analogue frontend is foreseen to be readout by a 10 MHz 12 bit ADC. The ADC delivers its data directly to a 1000 samples deep 16 bit ring-buffer recording a film around the trigger (see fig. 116a). 32 different buffers per pixel are sufficient for queuing of events until the result of 1st respectively 2nd level trigger validate or clear an event. All functions of the digital electronics are implemented in re-programmable FPGA logic in order to achieve not only high flexibility but also cost-effectiveness and ease of maintenance. The FPGA logic allows for an automated fast switch from single event mode to histogramming mode.

The sampling of the ADC values with a $100\,\mu\text{s}$ long history around the trigger (see

Measurement	max. total rate	Rate per channel	see sec.
ν-mass	141 Hz.	0.34 Hz.	11.6
WGTS activity	20 kHz.	48 Hz.	4.4
spectrometer response func. with e^--gun	100 kHz	100 kHz	11.4.2
^{83}Kr measurement	< 1.2 MHz	< 3 kHz.	9

Table 4: Summary of data acquisition requirements, assuming 420 detector pixels for the coverage of the magnetic flux tube. For details on the measurement see referenced sections.

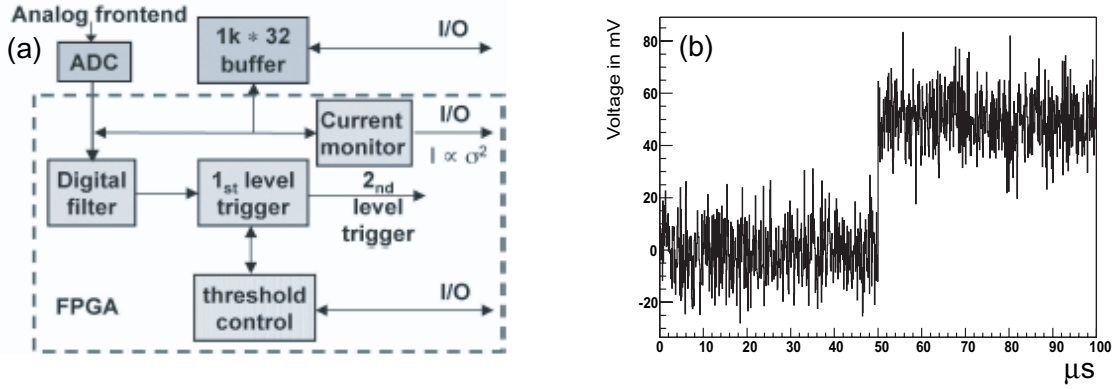

Figure 116: a) schematics of trigger logic b) simulation of sampling the output of the analog frontend for a typical event with a sampling rate of 10 MHz. A signal to noise ratio of 5:1 of the analog frontend has been assumed.

fig. 116b), offers a couple of possibilities. First of all, it allows the determination of the pulse height and thus the energy in the offline analysis, giving the opportunity to match the filter (shaping) times of the algorithms to the actual measurement. Furthermore, the baseline is steadily monitored and thus energy resolution interfering effects (e.g floating dark currents, low frequency interferences) are suppressed. If required, the thresholds of the trigger levels could be steadily adjusted to keep the trigger rate and thus the amount of transferred data constant.

10.2.3 Software

The VME hardware is read out by the ORCA (Object oriented Real-time Control and Acquisition) software. It is a software application tool kit that can be used for building flexible data acquisition systems. It uses a client/server model to totally separate the hardware controlling computers (the servers) from the user interface computers (the clients). So the system is able to have multiple server computers taking and processing data while being controlled and/or viewed by multiple remote clients. The framework is also suited for the inclusion of data analysis modules, which is a very important aspect for the online control and checks of a running experiment.

ORCA utilizes MacOS X as a development platform and the client provides an easy-to-use graphic user interface to set up a hardware configuration. The system provides many options for run control of a large scale experiment, like:

- multi-user compatibility
- data storage via network
- Status and error logging with alarm distribution
- Security and authorization checks

- online monitoring and visualization of data

ORCA provides several options to store data. The ORCA software adapts currently the requirements (non-binary records, XML format) from SCS architecture to guarantee a proper merging of detector data with SCS data.

ORCA is currently in use for several different experiments including the SNO NCD system [205] and the electro-magnetic test measurements of the KATRIN prespectrometer. Currently drivers are evaluated for Firewire connections between PC and VME crates to accommodate the expected data rates during KATRIN measurements.

11 Systematic and statistical uncertainties

In this chapter we will discuss the real experimental spectrum expected for KATRIN taking into account the detailed theoretical β spectrum as well as the KATRIN specific experimental parameters. We will investigate the systematic uncertainties arising from both the description of the β spectrum and variations and limitations of experimental parameters. The dependance of the statistical accuracy of m_ν^2 on measuring intervals and background rates is presented. As we will see, the lower limit of the energy interval to be analyzed, where the sum of systematic and statistical uncertainties becomes minimal, is about 30 eV below the endpoint energy E_0.

11.1 The molecular T_2 β spectrum

11.1.1 Nuclear recoil and radiative corrections

Equations (10) and (13) hold for an infinitely heavy, bare tritium nucleus whereas in reality, the $(^3\mathrm{HeT})^+$ daughter molecule gets a finite recoil energy E_{rec}. If we consider only a small region below the β endpoint, where the electron energy is large, but the neutrino energy negligible, then the recoil of the molecule E_{rec} balances the momentum of the β-electron:

$$E_{rec} = \frac{p_{rec}^2}{2M_{^3\mathrm{HeT}}} = \frac{p_e^2}{2M_{^3\mathrm{HeT}}} = E \cdot \frac{m_e}{M_{^3\mathrm{HeT}}} + \frac{E^2}{2M_{^3\mathrm{HeT}}} \approx E \cdot \frac{m_e}{M_{^3\mathrm{HeT}}} \tag{64}$$

Taking the recoil energy into account essentially means a reduction of the electron energy by E_{rec}. This reduction amounts at the endpoint to 1.7 eV and changes only by about 6 meV for a reasonable energy range of 50 eV and can be precisely considered in the analysis of a measured β spectrum.

The Fermi function $F(Z, E)$ in (10) can be described [206] as

$$F(Z, E) = \frac{x}{1 - \exp(-x)} \left(a_0 + a_1 \cdot \beta\right); \qquad x = \frac{2\pi Z \cdot \alpha}{\beta} \tag{65}$$

with the charge of the daughter nucleus $Z = 2$, the fine-structure constant α, $\beta = v_e/c$ including the empirical values $a_0 = 1.002037$ and $a_1 = -0.001427$. This leads to an asymptotic value for electron energies near the endpoint of $F(2, E \approx E_0) = 1.1875$ (see fig. 117a). Further radiative corrections imply contributions from virtual as well as real photons. These effects are rather small, but should be incorporated [207] to obtain a most accurate β spectrum description near the endpoint. Figure 117(b) shows the applied correction based on a compilation by [208] leading to a small reduction of the original spectral rate Γ_0 near the endpoint $E_0 = 18575$ eV. Note that the relative change of Γ/Γ_0 in a 50 eV wide interval of $[18524, 18574]$ eV is less than $5 \cdot 10^{-4}$.

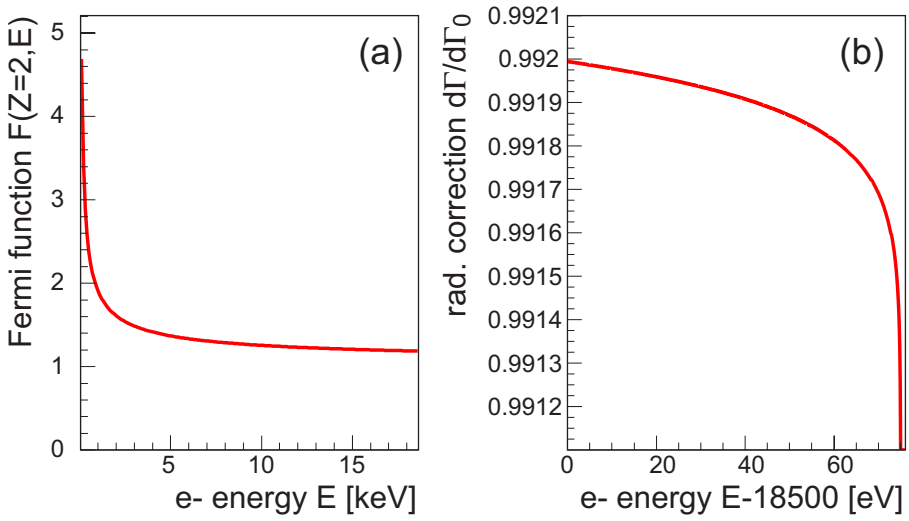

Figure 117: (a): Fermi function $F(Z, E)$ over the entire spectrum; (b): radiative correction to the β spectrum near the endpoint E_0.

11.1.2 Final state distribution

The first electronic excited state of the $(^3\mathrm{HeT})^+$ daughter molecule has an excitation energy of 27 eV [209]. Therefore excited states play almost no role for the energy interval considered for the KATRIN experiment, only the decay to the ground state of the $(^3\mathrm{HeT})^+$ daughter molecule, which is populated with 57 % probability, has to be taken into account (see figure 118). Due to the nuclear recoil, however, a large number of rotational-vibrational states with a mean excitation energy of 1.7 eV and a width of 0.36 eV are populated. This distribution follows the Franck-Condon principle; its precision depends on the knowledge of the ground state wave function, which is extremely well known [210, 211]. Therefore, no significant uncertainty arises from the rotational-vibrational excitation of the final ground state. Also, a contamination of the T_2 molecules by DT or HT molecules does not matter in first order: The shift of the mean rotational-vibrational excitation of HT with respect to T_2 is compensated by a corresponding change of the nuclear recoil energy of HT with respect to the 1.5 times heavier T_2 molecule [209]. This has been included in fig. 118, where the excitation energies of the $(^3\mathrm{HeT})^+$ and the $(^3\mathrm{HeH})^+$ daughter molecules are given *after* correcting for the different recoil energy.

However, this distribution ultimately limits the resolution which can be obtained in T_2 β decay. As a test of the KATRIN sensitivity of m_ν^2 on these final states, the distribution of the rovib. excitations of the electronic ground state was described as a Gaussian with varying width $\sigma = f \cdot \sigma_0$ with $\sigma_0 = 0.36\,\mathrm{eV}$ and $f = 1.0/1.01/1.05$. The resulting systematic shift in m_ν^2 corresponds to $\Delta m_\nu^2 \approx 6 \cdot 10^{-3}\,\mathrm{eV}^2$ $(2.7 \cdot 10^{-2}\,\mathrm{eV}^2)$ for $f = 1.01$ (1.05), respectively, and is nearly independent of the energy interval analysed. As will be discussed in section 11.5, the shift Δm_ν^2 due to an error of the width σ of 1 % corresponds to almost the maximal tolerable shift and therefore underlines the necessity to describe the final state distribution in detail.

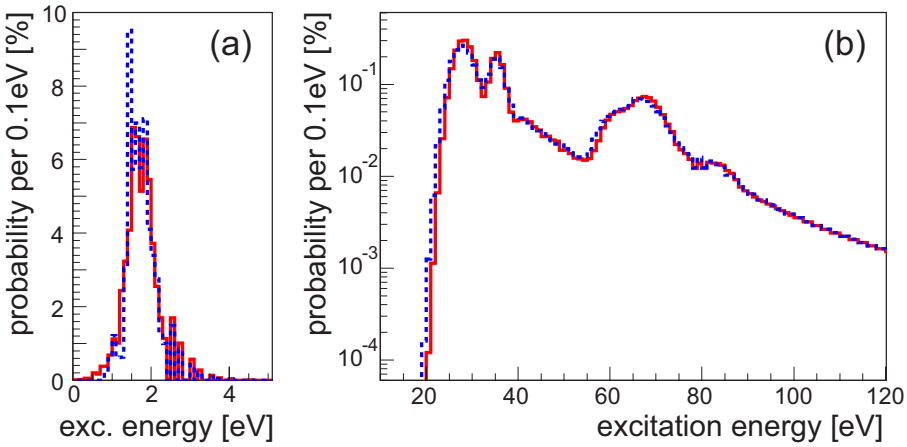

Figure 118: Probability distribution of final state population for the daughter molecules $(^3\mathrm{HeT})^+$ (red solid curves) and $(^3\mathrm{HeH})^+$ (blue dashed lines) based on calculations from [209]. The rotational-vibrational excitations of the electronic ground state (a) have an overall probability of 57 %. The electronic excited states of the daughter molecule (b) are well separated from the electronic ground state.

11.2 Tritium ions within the WGTS

In principle, other molecules or ions containing tritium could occur in the WGTS, giving rise to contributions to the overall β spectrum with different endpoint energies E_0. The tritium β decay itself as well as the β-electrons interacting with source molecules lead to the creation of several types of tritium ions and atomic tritium: T, T^+, T_3^+, T_5^+ or T^-. Primary β decay leads to T^+ formation

$$\mathrm{T}_2 \;\rightarrow\; e^- + (^3\mathrm{HeT})^+ \;\rightarrow\; e^- + T^+ + {}^3\mathrm{He}. \tag{66}$$

The $(^3\mathrm{HeT})^+$ ions transform quickly into T_3^+ ions by the reaction

$$(^3\mathrm{HeT})^+ \; + T_2 \rightarrow \; {}^3\mathrm{He} + T_3^+ \,. \tag{67}$$

Scattering of energetic electrons leads mainly to the formation of T_2^+, but also at a lower rate to T^+ through [212]

$$e^- + T_2 \;\rightarrow\; T_2^+ + 2e^- \qquad (\approx 96\,\%) \tag{68}$$

$$\rightarrow\; T + T^+ + e^- \quad (\approx 4\,\%) \tag{69}$$

with subsequent production of T_3^+ through

$$T_2^+ + T_2 \rightarrow \; T_3^+ + T + 1.17\,\mathrm{eV}. \tag{70}$$

T_5^+ ions could be created further by ternary collisions of T_3^+ with neutral gas molecules [213]. The T^- formation from the ground state T_2, namely the dissociative attachment

$$e^- + T_2 \rightarrow \; T^- + T \tag{71}$$

has a threshold of 3.75 eV [214] and a relatively small cross section compared to vibrational excitation of the T_2 molecule by electrons with energies of a few eV.

Table 5 lists the endpoint energy E_0 for various tritium decays relative to the atomic mass difference $\Delta M(^3\mathrm{He},^3\mathrm{H})$. Only the atomic decay of T^- ions has a higher endpoint

decay process (with $\bar{\nu}_e$)	$\tilde{Q} = E_0 - \Delta M(^3\mathrm{He},^3\mathrm{H})$	comment
$^3\mathrm{H} \rightarrow\ ^3\mathrm{He}^+ + e^-$	-24.6 eV	atomic decays
$^3\mathrm{H}^- \rightarrow\ ^3\mathrm{He} + e^-$	-0.75 eV	
$^3\mathrm{H}^+ + e^- \rightarrow\ ^3\mathrm{He}^{++} + 2e^-$	-65.4 eV	
$^3\mathrm{H}_2 \rightarrow (^3\mathrm{He}^3\mathrm{H})^+ + e^-$	-16.5 eV	molecular decays
$^3\mathrm{H}_2^+ + e^- \rightarrow (^3\mathrm{He}^3\mathrm{H})^{++} + 2e^-$	-48.9 eV	
$^3\mathrm{H}_3^+ +\ ^3\mathrm{H} + e^-$ $\rightarrow (^3\mathrm{He}^3\mathrm{H}_2)^{++} + 2e^-$	-35.1 eV[31]	

Table 5: Endpoint energies $\tilde{Q}$ with respect to the mass difference $\Delta M(^3\mathrm{He},^3\mathrm{H})$ for different atomic and molecular tritium decays

energy than the decay of T_2 molecules. Due to the cubic rise $(E_0 - qU_{ret})^3$ of the count rate of the integral spectrum and the small energy interval analysed by KATRIN, apart from T^-, all other tritium decays but the atomic T decay can be neglected. Figure 119 shows a simulation of the systematic effect of unaccounted contributions of atomic tritium $^3\mathrm{H}$ and $^3\mathrm{H}^-$ as a function of the analysing interval below the endpoint. The curves indicate

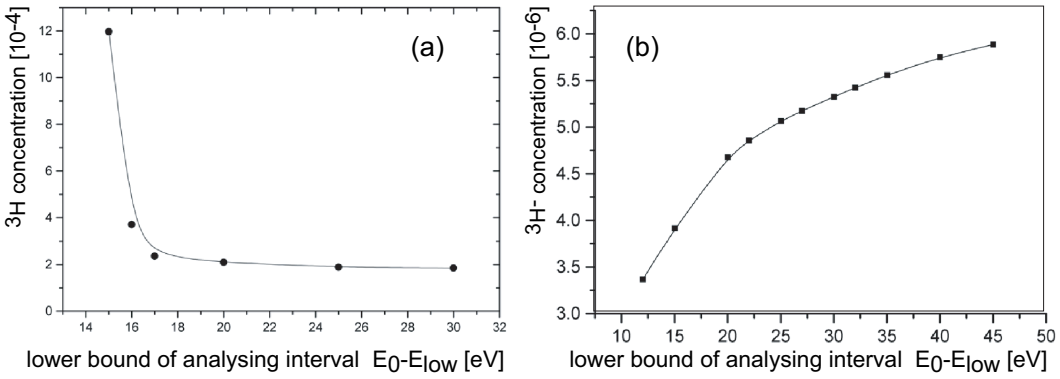

Figure 119: Unaccounted concentrations of atomic $^3\mathrm{H}$ (a) and $^3\mathrm{H}^-$ ions (b) within the WGTS leading to a systematic shift of the observable m_ν^2 of $\Delta m_\nu^2 = 0.01\,\mathrm{eV}^2$ as a function of the energy interval analysed in a MC simulation with KATRIN parameters.

concentrations leading to a systematic shift of the observable m_ν^2 of $\Delta m_\nu^2 = 0.01\,\mathrm{eV}^2$. Due to the higher Q-value, an unaccounted $^3\mathrm{H}^-$ concentration of a few 10^{-6} relative to $^3\mathrm{H}_2$

[31] Due to the uncertainty in the binding energy $\mathrm{E}_B\{(^3\mathrm{He}^3\mathrm{H})^{++} +\ ^3\mathrm{H} \rightarrow (^3\mathrm{He}^3\mathrm{H}_2)^{++}\}$ the accuracy of this endpoint energy is limited to about ± 2 eV [215].

would therefore limit the KATRIN sensitivity whereas for atomic ^{3}H, this concentration would be larger by about two orders of magnitude.

We will deduce the anticipated $^3H^-$ concentration within the WGTS in the following conservative estimation: With a decay density of $\Gamma_\beta \approx 2 \cdot 10^6/\mathrm{cm}^3/\mathrm{s}$ (corresponding to an average T_2 density of $n_{T_2} = 5 \cdot 10^{14}/\mathrm{cm}^3$) and a production of 15 secondary electrons per β-electron, the e^- production rate within the WGTS is $\Gamma_{e^-} \approx 3 \cdot 10^7/\mathrm{cm}^3/\mathrm{s}$. Taking conservatively for all secondaries an energy of $E_e > 3.75\,\mathrm{eV}$, these electrons may produce T^- via equ. (71) with a cross section of $\sigma(T^-) \approx 10^{-20}\,\mathrm{cm}^2$. However, the cross section for vibrational excitation of T_2 for electron energies around 4 eV is much higher, $\sigma_{vib} \geq 10^{-17}\,\mathrm{cm}^2$ [216, 217] with an average electron energy loss of 0.3 eV. The mean free path is therefore $\lambda_{vib} = 1/(n \cdot \sigma) \approx 200\,\mathrm{cm}$. After about 20 vibrational excitations, i.e. 40 m path length of diffusion, most relevant electrons with $E_e < 10\,\mathrm{eV}$ have energies below the Q-value for T^- production. The time needed is $t_{20} \approx 4 \cdot 10^{-5}\,\mathrm{s}$. Thus, the electron density with $3.75 < E_e < 10\,\mathrm{eV}$ is $n_e = \Gamma_{e^-} \cdot t_{20} \approx 10^3\,\mathrm{cm}^{-3}$. The production rate of T^- is then given to be

$$R(T^-) = \sigma(\mathrm{T}^-) \cdot (v \cdot n)_e \cdot n_{T_2} \approx 6 \cdot 10^5\,\mathrm{cm}^{-3}\mathrm{s}^{-1}. \tag{72}$$

Finally, the T^- density is given by

$$n(T^-) = R(T^-) \cdot \tau_{T^-} \tag{73}$$

with τ_{T^-} denoting the storage time within the source-transport system.

There are 3 potential processes limiting the storage time τ_{T^-}: gas flow, recombination and extraction by electric fields. The differential gas pumping leads to a gas flow corresponding to $\tau_{T_2} \approx 1\,\mathrm{s}$, but the tritium ions follow the magnetic field lines and do not leave the system at the pump ports. Whereas the T^- in the rear part of the WGTS are quickly absorbed on the rear plate, the T^- in the front part are guided into the transport system where they are finally reflected back to the front end of the WGTS by a negative dipolar potential applied to drift out charged particles (see sections 4.1.5 and 4.2). Thus these ions are trapped in the front transport system due to the magnetic field, the gas flow and the negative potential.

The main process limiting τ_{T^-} is the recombination of T^- with T_3^+ or T^+,

$$T^- + T^+ \rightarrow\ T_2\,, \qquad T^- + T_3^+ \rightarrow\ 2T_2 \tag{74}$$

with a recombination coefficient of $\alpha(T^-T^+) \approx 5 \cdot 10^{-7}\,\mathrm{cm}^3\mathrm{s}^{-1}$ [218]. To get an estimate on the recombination rate and hence the storage time τ_{T^-}, the concentration of positive ions has to be known.

The main processes driving the T^+ density are the production rate $\mathcal{B}_{T^+}$ from β decay and the processing rate to T_3^+ via

$$T^+ + T_2 + T_2 \rightarrow\ T_3^+ + T_2 \tag{75}$$

which has a reaction rate coefficient $k_{T^+} = 2 \cdot 10^{-28}\,\mathrm{cm}^6/\mathrm{s}$ [213, 219, 220, 221]. In equilibrium, the density n_{T^+} can be derived from

$$\mathcal{B}_{T^+} \cdot n_{T_2} \approx k_{T^+} \cdot n_{T^+} \cdot n_{T_2}^2 \tag{76}$$

with $\mathcal{B}_{T^+} \approx (0.3 + 15 \cdot 0.04) q_\beta$ where the first contribution comes from direct β decay [210] and the latter from ionization by electrons (see equ. (69) assuming 15 secondary electrons per β decay). With a decay rate $q_\beta = 3.6 \cdot 10^{-9}\mathrm{s}^{-1}$ per T_2 molecule, the T^+ density is estimated to be $n(T^+) \approx 10^4 \cdots 10^5\,\mathrm{cm}^{-3}$ depending on the position in the source tube.

Thus, the storage time is $\tau_{T^-} = 1/[n(T^+) \cdot \alpha] \leq 20\,\mathrm{s}$ which will be even further reduced taking into account the recombination with T_3^+ where $n(T_3^+) > n(T^+)$. Therefore, using equ. (73), the relative concentration is conservatively estimated to $n(T^-)/n(T_2) \leq 2 \cdot 10^{-8}$ which is 2 orders of magnitude smaller than the concentrations in fig. 119(b).

If, contrary to the above considerations, the storage time would become much longer, say $\tau_{T^-} \geq 1000\,\mathrm{s}$, the third process, a temporary dipolar field with positive potential of 100 V within the transport section applied *e.g.* every 100 s between two measuring intervals would also efficiently eliminate these negative tritium ions (see section below).

11.3 Tritium ions in the front transport system

A large fraction of the tritium ions leave the WGTS tube together with the gas flow (following the magnetic field lines, which are inside the WGTS tube parallel to the gas flow), and they enter the rear and front transport system. At the rear side they hit the rear wall, where they lose their charge due to surface recombination. The situation at the front side is more complicated. One has to emphasize first that in the transport system all charged particles follow strictly the magnetic field lines, so they are not able to leave the system at the pumping ports; the differential pumping is effective only for the neutral molecules. Second, the ions are not able to return from the transport system to the WGTS tube, because they cannot move against the gas flow (due to their strong contact with the neutral molecules). If the ions cannot leave the system in the direction of the spectrometers, they will be trapped in the front transport system.

If the potential in the whole transport system is the same as the WGTS and rear wall potential, the positive tritium ions (T^+, T_3^+, T_5^+, etc.) are able to enter the pre-spectrometer. The flow rate of these ions is somewhere between 10^{11}/s and 10^{12}/s, which has to be compared with the 10^5/s to 10^6/s maximal allowed flow rate of the neutral tritium molecules entering the pre-spectrometer. Obviously, the positive tritium ion rate is too large: they could substantially increase the background level of the order of a few mHz. In order to prevent these ions to reach the spectrometers, it is planned to apply a positive potential (100 V) at the downstream end of the DPS2-F.

As mentioned above, in this case the positive ions are trapped in the front transport system. They can in principle recombine with electrons, but it might happen that some positive ions have rather small recombination coefficients in the region of low gas density. In this case, a rather large density plasma could be created in the front transport system, and plasma instability could occur creating a time-dependent electric field. This would then change the energy of the β-electrons flying through the front transport system. In order to prevent the accumulation of the positive ions in the front transport system, one may apply a dipole electrode in the DPS2-F, upstream from the positive potential electrode (see section 4.1.5).

The T^- negative ions are also trapped in the front transport system. Their creation rate is rather small (see section 11.2), and they can recombine with the positive ions, so their steady-state density in the front transport system is probably small. Nevertheless, if the recombination rate of the negative ions turns out to be smaller than expected from literature data, one can remove them by changing, for a small time interval, the dipole electrode to positive potential.

There is the plan to install a Penning trap FT-ICR (Fourier Transform Ion Cyclotron Resonance) device into the DPS2-F, with the purpose to measure the densities of the different types of ions in the front transport system. The same trap can be used with different potentials to clean the volume by driving out the ions.

11.4 Experimental parameters and systematic effects

For tritium β decay experiments like KATRIN based on the MAC-E- Filter technique, the ratio of magnetic field strengths in different parts of the experiment (tritium source, magnetic pinch, analyzing plane of the spectrometer and detector) determines several key experimental parameters. In addition, the scattering probability for β-electrons within the source has to be known precisely to determine the overall response function of the system.

11.4.1 Transmission and response function

The transmission function T of a MAC-E-Filter given in equation (19) is fully analytical and depends only on the two field ratios $B_\mathrm{A}/B_\mathrm{max}$ and $B_\mathrm{S}/B_\mathrm{A}$. The total width ΔE of the transmission function from $T = 0$ to $T = 1$ is given by $B_\mathrm{A}/B_\mathrm{max}$ (eq. 17). The shape of T in this interval is determined by $B_\mathrm{S}/B_\mathrm{A}$, as this ratio defines the maximum accepted electron starting angle θ_max (eq. 18). The transmission function T does not take into account the interactions of electrons in the source. Electrons which have undergone inelastic collisions with T_2 molecules in the source have suffered an energy loss and hence have a modified response function. These processes can be described by folding of the corresponding inelastic cross section [94] with the distribution of electron path lengths through the source (including multiple scattering). Folding this energy loss distribution of electrons in the source with the transmission function (19) defines the so-called response function f_{res} of the experiment.

In fig. 120 the response function f_{res} of KATRIN is shown for isotropically emitted monoenergetic particles (with fixed energy E) as a function of the retarding energy qU. The figure is based on the following given standard parameters: a) energy resolution $\Delta E = 0.93\,\mathrm{eV}$, b) WGTS column density $\rho d = 5 \cdot 10^{17}/\mathrm{cm}^2$ and c) maximum accepted starting angle $\theta_\mathrm{max} = 50.77°$. The influence of the electric potential drop and the inhomogeneity of the magnetic field in the analyzing plane has not been considered, since these effects can be compensated by a detector with a radial position sensitivity. The average probability of β-electrons to enter the spectrometer without scattering on tritium is $P_0 = 41.3\,\%$ and defines the plateau in fig. 120. Due to the sharpness of ΔE of KATRIN (note that the

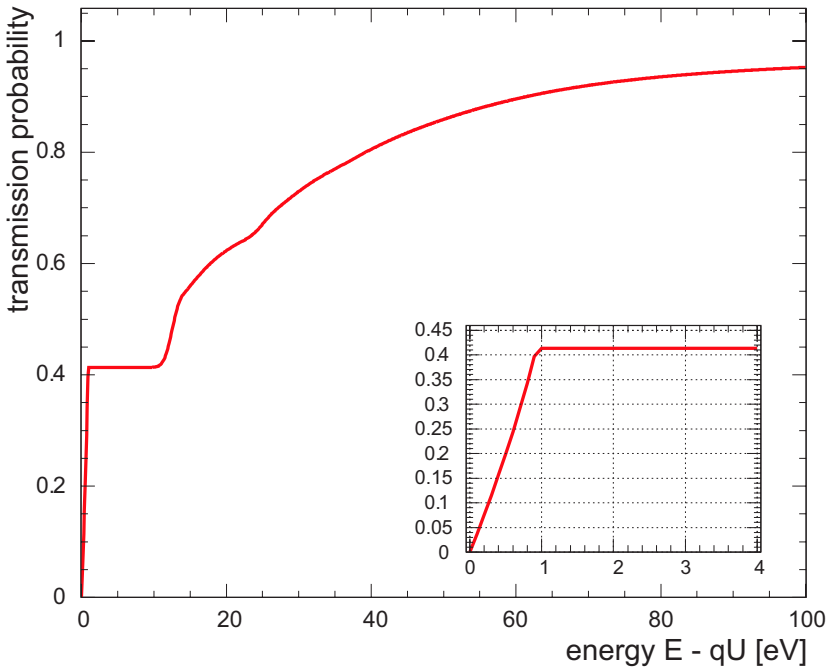

Figure 120: Response function of the KATRIN spectrometer for isotropically emitted electrons with fixed energy E as a function of the retarding energy qU. The energy loss of electrons in the WGTS source (column density $\rho d = 5 \times 10^{17}\,/\,\mathrm{cm}^2$, maximum starting angle $\theta_{max} = 50.77°$) is folded into the response function. The inlay shows the transmission of unscattered electrons to the plateau of $P_0 = 41.33\,\%$.

transmission function of a MAC-E-Filter has no high energy tail as can be seen in the inlay of fig. 120) and due to the high threshold of the T_2 excitation, the "no energy loss" fraction of transmitted electrons can clearly be separated from those electrons which have undergone inelastic collisions. The former fraction corresponds to the sharp rise from 0 to the flat plateau, whereas the latter part represents the second feature at the abscissa of about 12 eV. The relative amounts of each fraction (*i.e.* the relative height of the elastic plateau in fig. 120) is determined by the column density of the source.

For the case of the continuous β electron spectrum from tritium decay, the function f_{res} of KATRIN implies that the last ≈ 10 eV below the endpoint E_0 are fully covered by the elastic plateau of the response function. This means that the region with the highest sensitivity to m_ν is not affected by inelastic processes. Even with a larger measuring interval of 20 eV (25 eV) below E_0, inelastic events contribute to only about 1 % (5 %) of the total signal rate due to steeply rising spectrum and the integrating character of the MAC-E filter (compare fig. 120).

Despite this contribution of scattered electrons on a percentage level to the total count rate, the correct determination of the response function is of crucial importance for the KATRIN sensitivity on m_ν^2. Sections 11.4.2 and 11.4.3 demonstrate how f_{res} will be determined in a special pre-measurement and then monitored via repeated reference measurements of the column density.

11.4.2 Determination of the response function f_{res}

The 2 % accuracy of the total inelastic scattering cross section of electrons on T_2, $\sigma_{\text{inel}} = 3.456 \cdot 10^{-18}\,\text{cm}^2$ as well as the description of the energy loss function of electrons from literature [94] are not sufficient to be used as input to compile the KATRIN response function. Therefore, we have developed a method to deduce the energy loss function as well as the scattering probabilities from an *in situ* measurement within KATRIN.

The principal approach is to determine the response function once before the complete scanning procedure starts. For this initial measurement, an electron gun is used without and with various column density values of T_2 in order to measure

- the transmission function of the KATRIN main spectrometer for all starting points at the WGTS in two spatial dimensions (all starting points on the same magnetic field line within the WGTS will have practically the same transmission function due to the homogeneity of the magnetic field and the electric potential in the axial direction).
- the energy differential inelastic cross section by unfolding the energy loss function
- the column density of the source $\rho\, d$ in units of the mean free path with respect to inelastic scattering $\rho\, d_{\text{free}} = 1/\sigma_{\text{inel}}$. This ratio corresponds to $\rho\, d \cdot \sigma_{\text{inel}}$.

Requirements on an electron emitter

Since we want to use the standard KATRIN detector for our measurements we assume a maximum detected rate of $\Gamma_e = 100\,\text{kHz}$ per pixel. This rate represents no problem for a single detector pixel nor the fast DAQ. This maximum rate per pixel corresponds to an e-gun intensity of $I = 1.6 \cdot 10^{-2}\,\text{pA}$ per equivalent surface ΔA_e, which corresponds to the size of one pixel at the detector ΔA_D. The corresponding areas at the e-gun position ΔA_e and at the detector ΔA_D are connected by

$$\Delta A_e \cdot B_e = \Delta A_D \cdot B_D \tag{77}$$

We are considering an e-gun based on the photoelectric effect. The e-gun emission area can be chosen accordingly, e.g. a source size which corresponds to a few pixels (individual pixel size $5{\times}5\,\text{mm}^2$) at the detector in order to increase the allowed total rate and to shorten the necessary measurement time. We will not lose the spatial resolution of a single pixel as long as we drive the KATRIN experiment in a mode where we have a point-to-point mapping of electrons from the source through the transport channel and through both spectrometers to the detector.

An electron gun ideal for KATRIN would exhibit no energy spread and an isotropic angular distribution. More realistically, we require an energy spread of $\sigma_e < 0.2\,\text{eV}$. How to determine the energy and angular distributions of the electron gun is discussed in the next section.

For determining the transmission function, 2-dimensional scanning of the electron gun over the cross section of the WGTS is needed either by mechanical movement of the electron gun or by 2 scanning dipole coils on top of the solenoidal coils at the DPS1-F and DPS1-R. Although the main components of KATRIN are rotationally symmetric, a full 2 dimensional scanning is required since there are magnetic bends in the transport sections breaking the rotational symmetry.

Determination of the transmission function

To determine the transmission function the electron gun will be used placed at the CMS (see section 4.3 and fig. 44). Thus, the emitted electrons go through the entire system including the main spectrometer, but in this special measurement without filling the WGTS with T_2. The electrons will be counted at the detector for various voltage differences $U_s - U_0$ between analysing voltage U_0 and voltage at the electron gun U_s. We want to determine the transmission function for as many positions in the 2–dimensional transversal space as we have detector pixels (≈ 400). For an isotropically emitting source (*e.g.* tritium), the transmission function T of a MAC-E-Filter is fully analytical as described in eq. (19). However, an electron gun will not have a fully isotropic emission and not a sharp energy E_e. One therefore has to disentangle these two effects of the electron gun (energy distribution, angular distribution), by determining the response function of the electron gun at different energy resolutions ΔE, which one can adjust by the ratio of the magnetic fields $B_{\mathrm{A}}/B_{\mathrm{max}}$. In a simplified picture, one can describe the angular and energetic spectrum of the electron at the electron gun as follows: The energy at the electron gun is given by a mean energy $E'_e = h\nu - E_B$ which has some smearing σ_E. The starting angles are described by a function $f(\theta)$ (*e.g.* $f(\theta) = sin(\theta)$ for $0 \leq \theta \leq \theta_{\mathrm{max,e}}$, $f(\theta) = 0$ else). Then, the electrons are accelerated by the potential at the electron gun to an energy $E_e = E'_e + qU_s$ and the transformation of transversal energy is given by the adiabatic rule

$$p_\perp^2/B = const. \tag{78}$$

which corresponds to the well-known non-relativistic formula $E_\perp/B = const$ (see eq. 16).

Unfolding the energy loss function

The strategy for unfolding the energy loss function from measurements of the response function under different densities $\rho\, d$ is summarized briefly in the following: We assume

- that the response function of the electron gun is known for zero T_2 column density $T_e := R_e(\rho\, d = 0) = T \otimes S_e$ from the measurements described above. The convolution of the transmission function T with the electron source S_e has to be done in energy and in angular coordinates, since the electron gun is usually not isotropically emitting (equation (19) only holds for an isotropically emitting source).

- that the angular emission of the electron source $f(\theta)$ is known.

- that the electron source is stable within intervals of 10 minutes at the 10^{-3} level.

Knowing $f(\theta)$, one can use the adiabatic transformation rule (78) to calculate the angular distribution $f'(\theta)$ at the WGTS. Integrating over this angular distribution one can easily calculate the n-fold scattering probabilities P'_0, P'_1, P'_2, ... as a function of $\mu = \rho\, d/\rho\, d_{\text{free}}$.

We will perform measurements of the response function of the electron gun for different column densities of the WGTS as a function of the difference voltage $\Delta U = U_s - U_0$ (see figure 121). The deconvolution makes use of the fact that the transmission function T is a

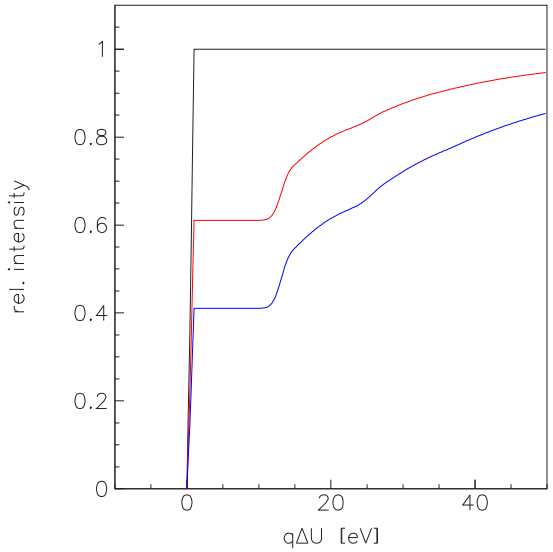

Figure 121: Schematic view of the energy loss measurements for different column densities of the WGTS (red, blue, and black: $\rho\, d = 0$) as a function of the retarding energy $\Delta E = q\Delta U = q(U_s - U_0)$, which is defined by the voltage difference of the electron gun voltage and the retarding voltage of the main spectrometer.

monotonically increasing function and this feature remains when n-fold convoluting with the energy differential cross section $d\sigma/d\Delta E$. To illustrate the deconvolution, figure 122 shows the normalized response functions for different numbers of scattering processes and their differences from 1, $\varepsilon_n(\Delta E)$ $(n = 0, 1, 2, \ldots)$. For a real measurement, all scattering orders contribute and the difference of the measured and normalized response functions from 1, $\varepsilon(\Delta E)$, is given by

$$\varepsilon(\Delta E) = P_0 \cdot \underbrace{\varepsilon_0(\Delta E)}_{=0} + P_1 \cdot \varepsilon_1(\Delta E) + P_2 \cdot \varepsilon_2(\Delta E) + P_3 \cdot \varepsilon_3(\Delta E) + \ldots \tag{79}$$

For a few measurements a, b, c, ... we can write a system of equations

$$\begin{aligned} \varepsilon^a(\Delta E) &= +P_1^a \cdot \varepsilon_1(\Delta E) + P_2^a \cdot \varepsilon_2(\Delta E) + P_3^a \cdot \varepsilon_3(\Delta E) + \ldots \\ \varepsilon^b(\Delta E) &= +P_1^b \cdot \varepsilon_1(\Delta E) + P_2^b \cdot \varepsilon_2(\Delta E) + P_3^b \cdot \varepsilon_3(\Delta E) + \ldots \\ \varepsilon^c(\Delta E) &= +P_1^c \cdot \varepsilon_1(\Delta E) + P_2^c \cdot \varepsilon_2(\Delta E) + P_3^c \cdot \varepsilon_3(\Delta E) + \ldots \\ &\ldots \end{aligned} \tag{80}$$

By simply inverting the linear system of equations (81) one obtains $\varepsilon_1(\Delta E)$. It is im-

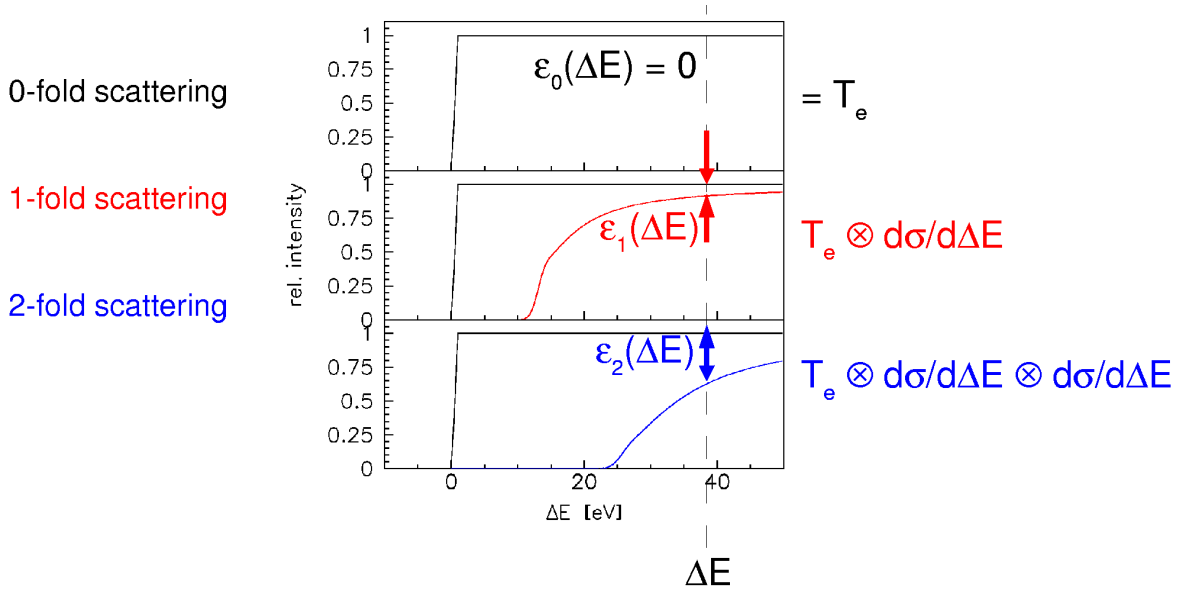

Figure 122: Schematic view of the normalized response functions for different numbers of scattering processes as function of the surplus energy $\Delta E = q(U_s - U_0)$ (for this illustration, we neglect the primary photoelectron energy $E'_e = h\nu - E_B$). In case of 0-fold scattering, the response function is only given by the response function of the electron gun T_e. In case of n-fold scattering, T_e has to be convoluted n-times with the energy differential cross section $d\sigma/d\Delta E$. The difference of the n-fold response function to 1 is defined as $\varepsilon_n(\Delta E)$.

portant to note that this deconvolution works reliably since the response functions are monotonically increasing. Finally, a simple deconvolution of $1 - \varepsilon_1(\Delta E)$ is performed to obtain the energy differential cross section since:

$$T_e \otimes \frac{d\sigma}{d\Delta E} = 1 - \varepsilon_1(\Delta E) \tag{81}$$

In an analogous way one can deconvolute the functions $\varepsilon_2(\Delta E)$, $\varepsilon_3(\Delta E)$, ... for further checks of the energy differential cross section:

$$T_e \otimes \frac{d\sigma}{d\Delta E} \otimes \frac{d\sigma}{d\Delta E} = 1 - \varepsilon_2(\Delta E). \tag{82}$$

Results of the unfolding procedure

In a feasibility study, the following parameters of the e-gun and the spectrometer system were used which finally determined the statistical accuracy of the unfolding procedure:

- constant e-gun intensity of Γ_e,

- Gaussian e^- energy spectrum with $\sigma(E_e) = 0.5\,\text{eV}$,
- isotropically emitting electron source (for simplification of this first test),
- transmission function with width of 0.5 eV for an isotropically emitting source,
- a scanning interval of 100 eV below the electron energy E_e,
- e-gun measurements with 2 different column densities as well as without tritium gas and
- a total measurement time $t = 1.7 \cdot 10^{10}/\Gamma_e$, *e.g.* $t = 2\,\text{d}$ for $\Gamma_e = 100\,\text{kHz}$.

In the first analysis step, the measured response functions are used to unfold the underlying energy differential cross section for electrons undergoing an inelastic scattering process according to the deconvolution procedure described in the previous section (solution to the system of linear equations in (81)). The second deconvolution was done according to equation (81). Here, the precise knowledge of the response function of the electron gun for zero column density was assumed, which in the later experiment can be determined with sufficiently high precision. Figure 123 shows the result of this unfolding procedure. The main structure of the response function up to energy losses of $\Delta E = 20\,\text{eV}$ can be reproduced with high accuracy whereas at higher energy losses statistical fluctuations

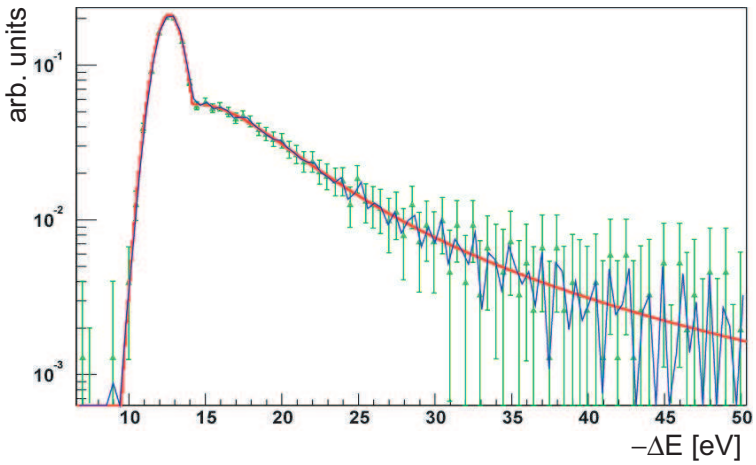

Figure 123: Result of an unfolding procedure of the energy loss function from e-gun measurements. The red solid line shows the input model of energy loss, triangles with error bars indicate the unfolding result with 0.5 eV steps, the blue lines are the result of a spline fit to the unfolding results.

become more prominent. Since the unfolding grid size was 0.5 eV, but the bin size of the analysis of the β spectrum is 0.1 eV, a quintic natural spline fit was applied to the unfolding values to obtain an energy loss function with 0.1 eV resolution.

As the result of the unfolding procedure of the energy loss function, a response function for this system can be reconstructed (taking the scattering probabilities for multiple inelastic electron scatterings) and compared to the one deduced directly by the energy

loss input model. Figure 124(a) shows the relative difference of these response functions. For the tritium case the response function is different, since the β-electrons are not always

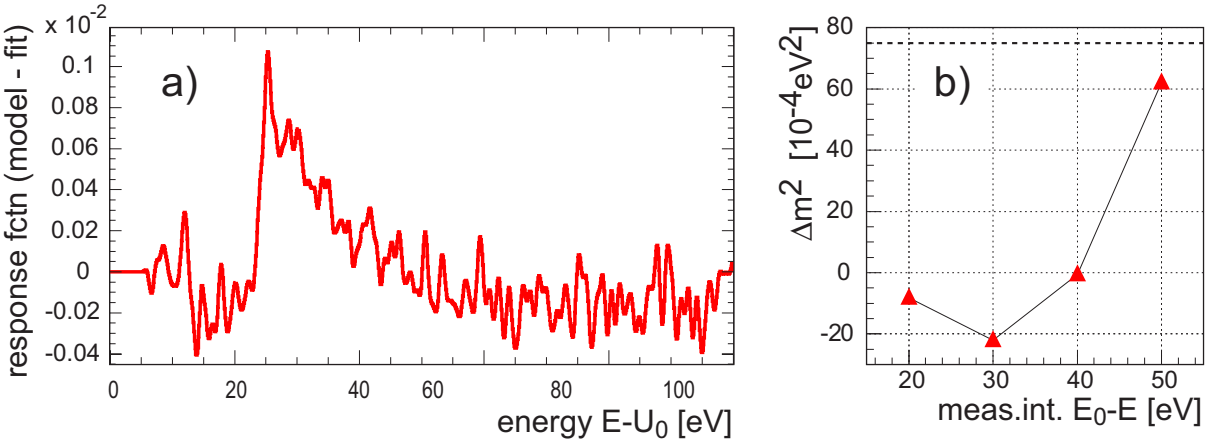

Figure 124: (a) Relative difference between model input and unfolding result of the response function. (b) Systematic shift of m_ν^2 as a function of the width of the analysed interval below the kinetic endpoint. The dashed line shows the maximally allowed shift of $\Delta m^2 = 0.0075\,\mathrm{eV}^2$ for an individual effect.

going through the whole source, but are emitted within the source: on average they are passing through half of the source column density. Also B_s could be different as it is for the case of the electron gun. One therefore has to use the measured energy differential cross section $d\sigma/d\Delta E$ and the measured relative column density $\rho\,d/\rho\,d_{\mathrm{free}} = \rho\,d \cdot \sigma_{\mathrm{free}}$ to reconstruct the response function. Additional inputs are the transmission function T from equation (19) and the n-fold scattering probabilities P_0, P_1, P_2, ...

$$R = T \otimes \left(P_0 + P_1 \frac{d\sigma}{d\Delta E} + P_2 \frac{d\sigma}{d\Delta E} \otimes \frac{d\sigma}{d\Delta E} + \ldots \right). \tag{83}$$

The deconvoluted energy differential cross section with its statistical and systematic uncertainties leads to a systematic uncertainty on the observable m_ν^2 shown in figure 124(b) as a function of the analysed interval below the kinetic endpoint E_0. Compared to the maximally tolerated individual systematic error of $\Delta m^2 < 0.0075\,\mathrm{eV}^2$ (see section 11.5), the deduced accuracy of the energy loss function in this study is sufficient. For a more realistic scenario, characteristics of the e-gun measurement will be taken into account, e.g. potential tails of the energy distribution of the e-gun electrons, fluctuations of the e-gun intensity and HV fluctuations during the relatively long measurements, but the feasibility of this method has been demonstrated.

11.4.3 Monitoring of the column density $\rho\,d$

Having measured the KATRIN response function once in detail in a special pre-measurement and thereby the energy loss function, the actual column density $\rho\,d$ of the source determines the scattering probabilities P_i. Thus, the actual response function depends on $\rho\,d$

which has to be monitored via repeated measurements. The column density of the WGTS and the QCTS is monitored in two ways:

- Special short time measurements with the electron gun to control the column density of the WGTS from the ratio of the scattered to non-scattered electron fraction. The column density of the QCTS is determined by laser ellipsometry as applied in the Mainz experiment [107].
- The tritium count rate determined by the detector, in combination with online mass spectrometry, gives online a spatially resolved tritium column density measurement. For this purpose a particular measurement point further below the tritium endpoint with enhanced count rate will be chosen.

In the following, the two methods to achieve a monitoring accuracy of the WGTS column density of $1 - 2 \cdot 10^{-3}$, relative, are explained in more detail.

Repeated e-gun measurements to determine $\rho\, d$

To a good approximation, the column density is rather uniform over the cross section of the WGTS, or at least that fluctuations in the column density exhibit no significant radial dependence. $\rho\, d$ is determined by measuring the rate of electrons at different analysing energies $q\Delta U = E_e - \delta E_i$ (*e.g.* i=1,2,3) below the electron starting energy E_e. Figure 125 shows the expected electron rate as function of δE. By comparing the rates at different retarding potentials, the column density $\rho\, d$ as fraction of the free column density $\rho\, d_{\mathrm{free}}$ can be determined with $1 - 2 \cdot 10^{-3}$ precision (see below). For this measurement the electron gun intensity does not need to be known absolutely, it only needs to be stable over a few minutes measurement time needed for the different retarding voltage settings of this monitoring measurement.

For $1\,\mathrm{eV} < \delta E_1 < 10\,\mathrm{eV}$, one determines the rate of electrons, which have not undergone an inelastic scattering process. Reasonable values for δE_2 and δE_3 are 20 eV and 40 eV. To achieve a relative statistical accuracy of the count rate of $1 - 2 \cdot 10^{-3}$ one needs $N_e > 10^6$ electrons which corresponds to an e-gun measuring time of $t_1 \geq 25\,\mathrm{s}$ for a zero scattering probability of $P_0 \approx 40\%$ and $t_{2,3} \approx 20\,\mathrm{s}$. With some time to adjust the different voltages within this procedure, we take as total e-gun reference time

$$t_{\mathrm{e-gunref}} \approx 2\,\mathrm{min}. \tag{84}$$

Since any special measurement should not significantly reduce the measuring time in standard β spectrum scanning mode, this e-gun reference measurement should not be repeated more often than once per ≈ 2 hours. Therefore one would have to ensure that the column density fluctuations have time constants of 2 h and longer. If this condition cannot be fulfilled one has to perform these measurements more often. Furthermore, if there is a way to stabilize or measure the electron gun intensity with sufficient precision one could determine $\rho\, d$ by one retarding potential setting $q\Delta U = E_e - \delta E_1$ alone yielding a 3 times faster $\rho\, d$ determination.

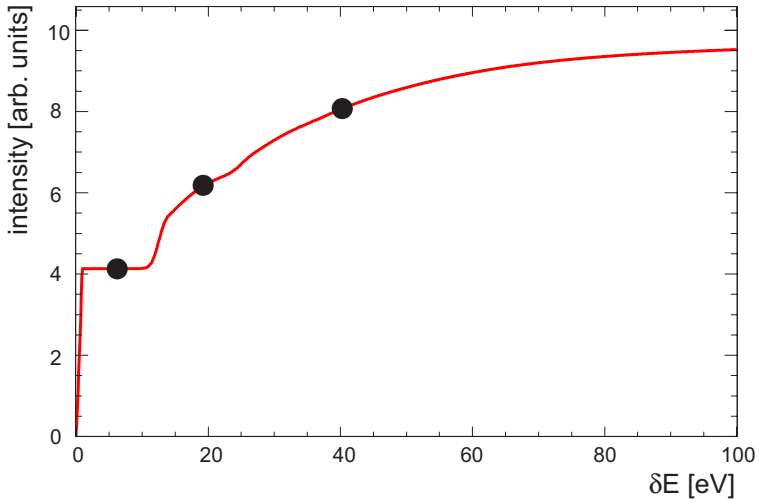

Figure 125: Schematic view of the column density determination by sending electrons from an energetically sharp electron gun of energy E_e from the rear through the WGTS and measuring the count rate at the detector at 3 different retarding energies $qU = E_e - \delta E$ (dots). Since the transmission function of the spectrometer and the energy differential cross section will be known well enough, the ratio of 2 count rates determines the column density $\rho\, d$ in units of the free column density $\rho\, d_{\mathrm{free}} = 1/\sigma_{\mathrm{inel}}$

Monitoring $\rho\, d$ during T_2 measurements

Having subdivided the T_2 measuring time into many intervals of a few hours with subsequent reference measuring to "re-adjust" the column density needed to get a correct description of the response function, the problem of $\rho\, d$ monitoring reduces from an "unaccounted shift" in the long term to "statistical fluctuations" around a well known and monitored mean value, the "2 hour"–values. As seen from figure 126, unaccounted shifts in $\rho\, d$ of a few 10^{-3} over the *entire* measuring period easily lead to an unacceptable bias in the observable m_ν^2. For example, the acceptable systematic shift of $\sigma_{\mathrm{syst}} = 0.0075\,\mathrm{eV}^2$ (to be defined in detail in section 11.5) for a single experimental parameter such as $\rho\, d$ would almost be reached for $\Delta\rho\, d/\rho\, d = 2 \cdot 10^{-3}$ measuring in an interval of 40 eV below E_0.

This very stringent requirement on the monitoring accuracy of $\rho\, d$ might be loosened by a factor of ~ 10 considering "statistical fluctuations" around a well known mean as it is the case for the above measuring procedure, but for a conservative treatment of systematic uncertainties, the accuracy requirement of $2 \cdot 10^{-3}$ is kept.

Monitoring the column density during regular T_2 measurements consists of a two-fold process: measuring the β–activity of the source $\varepsilon_T \cdot \rho\, d$ as well as the isotopical content of the WGTS ε_T. If both parameters are known with sufficient accuracy with respect to the values during the e-gun measurement, variations of $\rho\, d$ can be extracted and the registered data corrected for.

Measuring the source activity can be done by lowering the electric potential barrier between source and spectrometer and using the main detector. For a potential $U_{\mathrm{spec}} =$

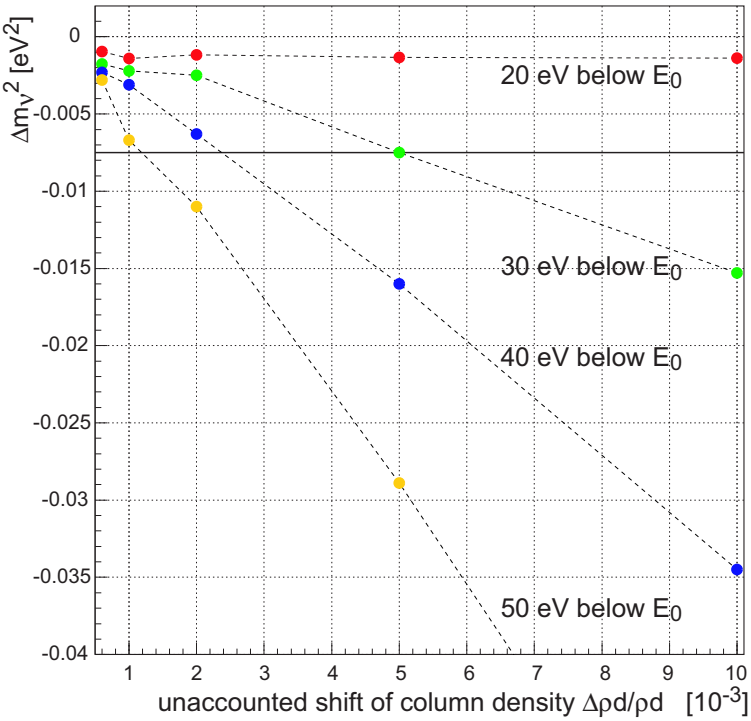

Figure 126: Systematic bias Δm^2 as a function of unaccounted shifts in $\rho\, d$ for different measuring intervals.

$E_0 - 50(100, 500)\,\mathrm{V}$, the count rate is

$$\Gamma_{\mathrm{T}_2} = 140\,\mathrm{Hz}(\approx 1\,\mathrm{kHz}, 100\,\mathrm{kHz}.) \tag{85}$$

For the latter potential, the count rate is sufficient to get an accuracy of 10^{-3} after $t_{mon} = 10\,\mathrm{s}$.

To keep variations of the count rate due to deviations of the HV value from $U_{\mathrm{spec}} = E_0 - 500\,\mathrm{V}$, this 500 V interval has to be adjusted to an accuracy of $\leq 0.17\,\mathrm{V}$ or $\leq 3 \cdot 10^{-4}$ relative. This HV drop at the spectrometer electrode system is difficult to perform in the presence of the high accuracy needed for the overall U_0 precision of a few ppm. However, a method exists whereby a very precise and stable base HV can be added to another voltage which can be remotely varied up to 500 V. This additional HV is foreseen to be applied at the WGTS and therefore incorporated into the technical specification of the tritium source section.

Another method without any interruption of the normal scanning mode would be measuring the total activity with a rear monitor detector. This detector could have a 4 keV threshold thereby providing a continuous high count rate. A severe challenge for a rear monitor detector is the simultaneous requirement for an electron-emitting rear plate to compensate the net charging of the WGTS plasma due to the β-electrons leaving the WGTS towards the main detector. A technical solution for a rear monitor well within the flux tube as well as a complete geometrical coverage of the e^--emitter has not yet been

incorporated as it requires a low energy electron emitting source at well-defined potential and a detector for high energy electrons at the same time. An alternative is a ring monitor which is sensitive to those outer segments of the source not transmitted magnetically to the main detector also represents disadvantages: Placed in between the pre-spectrometer and the main spectrometer, the count rate would be a fraction (scaled from geometry) of the activity of a few kHz of the high energy part of the β spectrum, therefore requiring much longer accumulating times to define an "activity monitoring point". In addition, such a ring monitor could act as a diaphragm scattering β-electrons in or out the magnetic flux tube.

The second part of the two-fold monitoring procedure is the continuous measurement of the isotopical content of the tritium gas ε_{T}. This will be determined by Laser Raman spectroscopy at the inner loop inlet under a gas pressure of 10 mbar with a Nd:YAG or Yb:YAG laser of 10 W power (section 4.4).

As discussed above, the initial determination as well as the continuous monitoring of the response function requires special measurements before and during the scanning of the β spectrum. We present a potential measuring scheme in section 11.7.

11.4.4 Parameters influencing the column density $\rho\, d$

As seen from figure 126, the accurate knowledge of the scattering probabilities P_i is indispensable to accurately describe the response function. These probabilities are defined by the column density $\rho\, d$ as well as the β-electron track length within the WGTS. Various parameters, such as the WGTS tube temperature, the gas influx or the gas density at the first pumping port influence $\rho\, d$. On the other hand, variations in the magnetic field strength of the WGTS lead to variations of the maximal opening angle θ_{max} and thereby to different average track lengths of the source electrons.

The tritium gas flow in the KATRIN WGTS tube is realized at a low pressure so that neither a pure hydrodynamic framework can be used nor intermolecular collisions can be neglected in a free molecular regime. Therefore, the molecular density along the tube axis has been calculated based on a numerical model of rarefied gas flows [222, 223]. Figure 23 shows the density profile normalized to the influx density n_{inj} with the boundary condition $n_{ex} = n(x = L) = 0$ at the first pumping port at $L = 5\,\mathrm{m}$ from the injection point. Note that the boundary condition $n_{ex} = n(x = L) = 0$ is not fulfilled in reality but applied here for the purpose of calculating the influence of different experimental parameters on the column density. See figure 23 for a calculated profile of the WGTS including the pump port DPS1-F.

Whereas the shape of the profile does not depend strongly on the influx pressure p_{inj} (as can be seen from fig. 23), the integrated density profile, i.e. the column density $\rho\, d$, depends on p_{inj}: with a coefficient of [100]

$$\alpha_{p_{inj}} = \frac{p_{inj}}{\rho\, d} \left(\frac{\partial \rho\, d}{\partial p_{inj}} \right)_T = 1.1 \tag{86}$$

for $T = 27\,\mathrm{K}$, the necessary stability for the column density between e-gun measurements

of $\pm 2 \cdot 10^{-3}$ translates into the same input pressure monitoring accuracy for an envisaged value of $p_{inj} = 3.35\,\mu$bar leading to the nominal column density $\rho\, d = 5 \cdot 10^{17}\,\mathrm{cm}^2$.

Given a constant influx density n_{inj}, variations of the gas density at the first pumping port n_{ex} also change the column density. The corresponding coefficient is

$$\alpha_{n_{ex}} = \frac{n_{ex}}{\rho\, d} \left(\frac{\partial \rho\, d}{\partial n_{ex}} \right)_{T, n_{inj}} \approx 0.03, \tag{87}$$

so the required column density stability translates into a stability of $\Delta n_{ex}/n_{ex} < 6\,\%$. With 4 turbomolecular pumps at the first pump port of the DPS1-F (see also section 4), the pumping efficiency can be stabilized at the required level.

The column density is also influenced by the temperature of the WGTS tube. This correlation also depends on the temperature dependence of the gas viscosity η_{T_2}. Whereas the absolute value of η_{T_2} is not accurately known but extrapolated from η_{D_2} at the appropriate temperatures [224, 225], $\eta_{T_2} \approx \sqrt{3/2}\eta_{D_2}$, the temperature dependence $\eta_{D_2}(T)$ can safely be taken to describe $\eta_{T_2}(T)$. The overall temperature dependence of the column density is then

$$\alpha_T = \frac{T}{\rho\, d} \left(\frac{\partial \rho\, d}{\partial T} \right)_{n_{inj}} \approx -1.2 \tag{88}$$

for the operating temperature $T = 27\,\mathrm{K}$ leading to a stabilization requirement of the WGTS tube temperature of $\Delta T < 60\,\mathrm{mK}$.

As mentioned above, also a change of the average track length of β-electrons changes the scattering probabilities and thereby the response function. Magnetic field inhomogeneities in the WGTS as well as a potential drift of the overall magnetic field strength B_S lead to variations or a general drift, respectively, of the maximal opening angle θ_{max}. Both effects result in different track lengths of source electrons. The drift of $B_\mathrm{S} = 3.6\,\mathrm{T}$ is limited to a fraction of 10^{-4} within a period of 3 months due to the persistent mode of the superconducting magnets in their short circuit operation and can be easily monitored. To account properly for inhomogeneities of B_S, a 3-dim map of the WGTS, the resulting track lengths and scattering probabilities will be modelled. As a conservative estimate, a fraction of $2 \times 400\mathrm{mm}/10000\mathrm{mm} = 0.08$ of β-electrons will be emitted in a magnetic field of $(1 - 0.02)B_\mathrm{S}$ leading, if unaccounted, to an average change of the magnetic field of $\Delta B_\mathrm{S}/B_\mathrm{S} \approx 2 \cdot 10^{-3}$. MC simulations showed that an unaccounted shift of $\Delta B_\mathrm{S}/B_\mathrm{S} = 6 \cdot 10^{-3}$ corresponds to the same shift in the analysis of Δm_ν^2 as an unaccounted shift of $\Delta\rho\, d/\rho\, d = 2 \cdot 10^{-3}$. In this oversimplified but conservative approach, the estimated accuracy of the magnetic field strength within the WGTS exceeds the required one already by more than a factor 3.

11.4.5 Corrections to the transmission function T

Since the transmission function T (eq. 19) depends only on the relative field values and potentials, it is insensitive to mechanical adjustment. The inhomogeneity of the magnetic field and the electric potential can be calculated precisely. In addition, the

shape of the transmission function will be checked by a point-like test source of K-32 conversion electrons from ^{83m}Kr at the position of the QCTS or by an electron gun, which is moved across the magnetic flux tube together with the individual pixel information of the detector.

Fluctuations of the absolute position of the transmission function are more critical. A simple relation connects an additional unknown Gaussian broadening of width σ_g (*e.g.* caused by fluctuations of the absolute value of the retardation potential U) to a shift of m_ν^2 (see eq. (60) in section 9). Therefore, the noise and drifts of the high voltage have to be below 70 mV to limit its contribution to Δm_ν^2 to a maximum value of 0.01 eV^2.

11.4.6 Variations of the retarding high voltage

The scanning of the β spectrum is performed by varying the potential difference between the WGTS and the main spectrometer. Any inaccuracy of the potential difference which can be described as uncertainty in the position of the transmission function, is a source of systematic uncertainties on the observable m_ν^2. Since the analysis of the β spectrum takes the absolute energy scale, in other words the endpoint energy, as a free parameter, it is in fact variations of the high voltage scale which are a source of systematic uncertainties. For KATRIN, these effects have been studied in detailed simulations based on the specific

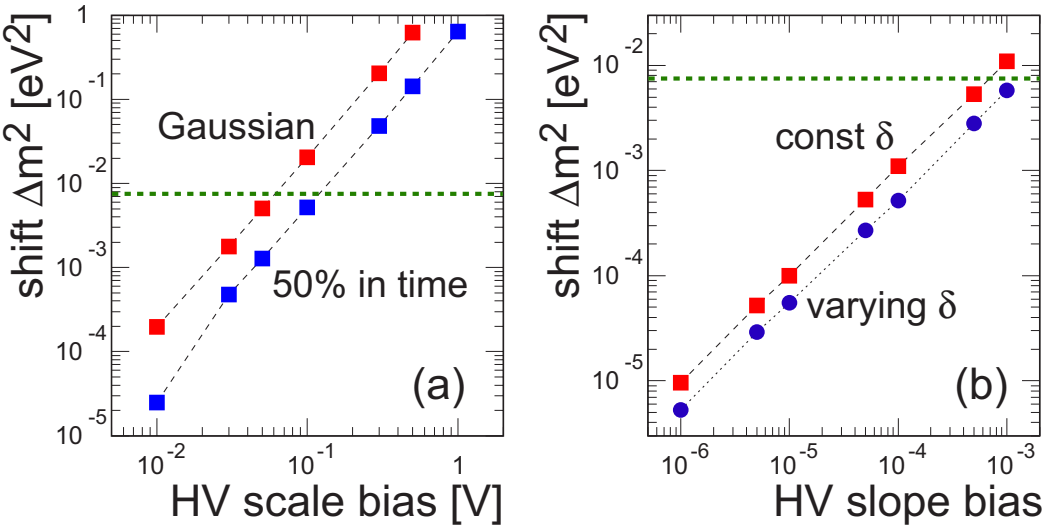

Figure 127: Systematic shift Δm_ν^2 of the observable m_ν^2 due to HV uncertainties in the HV scale (a) or in the HV slope (b) [226]. The horizontal dotted line corresponds to the maximal individual systematic shift of $\Delta m_\nu^2 = 0.0075\,\mathrm{eV}^2$ (see text for details).

experimental parameters [226]. Figure 127 shows the systematic shift of the observable m_ν^2 in case of various imperfections of the HV settings. In fig. 127(a), the absolute scale of the HV varies over time in two exemplified schemes: having an unaccounted shift of the HV dividers in 50 % of the total measuring time or having a Gaussian variation of the energy scale with an appropriate value of σ. Taking the maximally allowed systematic

shift of $\Delta m_\nu^2 = 0.0075\,\mathrm{eV}^2$ (see horizontal dotted line), Gaussian variations of the high voltage have to be limited to $\sigma < 60\,\mathrm{mV}$ in good agreement with the analytical description given in eq. (60) corresponding to a relative stabilization of the HV scale of 3 ppm at $U_{ret} = 20000\,\mathrm{V}$. Another potential source of HV imperfections could be inaccuracies δ of the HV slope of the calibration line of the voltmeter along the energy interval under analysis (see fig. 127b). For the two cases of an unaccounted constant or a Gaussian varying HV slope error δ, inaccuracies have to be limited to the 10^{-3} level, i.e. less than 50 mV along an analyzing interval of 50 eV below the endpoint energy.

11.4.7 Description of background

Looking at the count rate spectrum from MAC-E filters, e.g. fig. 12, the background contribution is considered to be energy-independent, or more specifically, to be independent of the applied retarding potential between the WGTS and the analyzing plane of the spectrometer vessel. In the analysis of the β spectrum, the background is therefore introduced as a constant component with a background rate Γ as a free parameter. However, due to the various sources of background (see section 8), this assumption does not have to hold *a priori*. A series of analytical tests have been performed to check the influence of a non-constant background on the deduced value of m_ν^2. Figure 128 shows the systematic shift

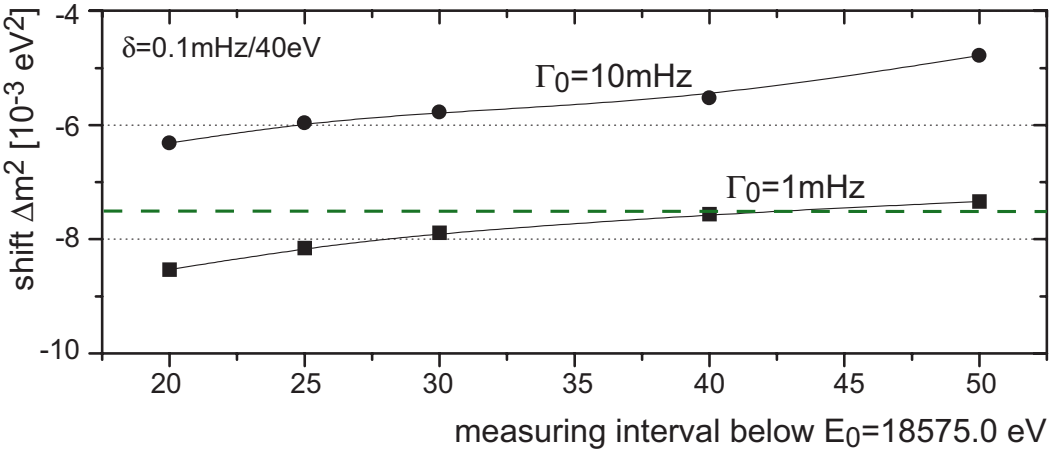

Figure 128: Systematic shift Δm_ν^2 of the observable m_ν^2 due to an unaccounted background slope of $\delta = 0.1\,\mathrm{mHz/40eV}$ for different lengths of the interval under analysis as well as different background rates Γ (see text for details).

of m_ν^2 introduced by an unaccounted slope of the background rate $\Gamma = (1+\delta\cdot(E_0-qU))\Gamma_0$ with $\delta = 2.5\,\mathrm{mHz/keV}$ for two different values of Γ_0 and various intervals (E_0-qU). With the dashed line indicating the maximally allowed individual systematic error, a slope of $\delta = 2.5\,\mathrm{mHz/keV}$ becomes critical for a background rate considerably less than the envisaged $\Gamma_0 = 10\,\mathrm{mHz}$. However, one should note that there is the possibility to measure a constant background slope directly, just above the endpoint energy. This has been done

in the Troitsk experiment leading to an upper bound of $\delta < 0.5\,\mathrm{mHz/keV}$ underlying the approach of a constant background description[32].

11.4.8 Effects from WGTS space charging

The β decay activity of the tritium gas inside the WGTS is about 10^{11} Bq. The high energy β-electrons, following the magnetic field lines, leave the WGTS tube at the rear side (at the front side they are reflected back by the high negative potential of the spectrometers). During their motion through the tritium gas they produce ionization events, one β-electron creates on average 15 secondary electron-ion pairs. Various primary positive ions are are formed by β decay and ionization processes (T^+, He^+, $(^3HeT)^+$, $T_2{}^+$), followed by chemical reactions which produce T_3^+, T_5^+ and even larger cluster ions.

The diffusion of the ions to the wall of the tube is reduced by 4 orders of magnitude due to the transverse magnetic confinement of the plasma. Hence, their storage within the WGTS is determined by the longitudinal pump out time of about 1 s. Therefore, a large positive space charge and potential would build up within the WGTS, if this charge is not compensated or driven out faster by some other mechanism. Fortunately, this positive space charge can be compensated by the secondary electrons stemming from ionization. The positive ions and the secondary electrons form a quasi-neutral plasma within the WGTS. The density of this plasma, mainly determined by gas flow and electron-ion recombination, amounts to a few times 10^6 electron-ion pairs per cm^3. The magnitude of the space charge potential variation inside the WGTS tube is determined by the electron temperature, or more generally by the electron energy distribution.

This energy distribution was computed by a Monte Carlo programme including the most important $e^- - T_2$ scattering cross-sections: ionization, electronic excitation, vibrational and rotational excitation and elastic scattering (see fig. 91). We have used published values for these total and differential cross sections. The β decay events were generated according to the Fermi spectrum with isotropic angular distribution. Then, the path of these electrons inside the WGTS system was followed, and their ionization processes were recorded. The paths and interactions of the secondary electrons created by ionization were also recorded until they left the WGTS system.

The main results of these computations are the following: The largest part of the secondary electrons cools down to gas temperature (27 K, corresponding to 2.4 meV) within a time interval of 1-2 ms, and during this cooling time the electrons diffuse along the magnetic field lines about 1 m distance. Since the length of the WGTS system is 10 m, and the average escaping time of the electrons from the system is about 100 ms (without space charge potential), these Monte Carlo computations indicate that most of the electrons being present inside the system have an energy distribution with a characteristic energy of a few meV (see fig. 129).

[32]If needed, a background slope δ as a further free parameter in the analysis of the β spectrum introduces a larger statistical uncertainty on Γ_0 and hence on m_ν^2 which can be counterbalanced by a larger interval *above* the β endpoint.

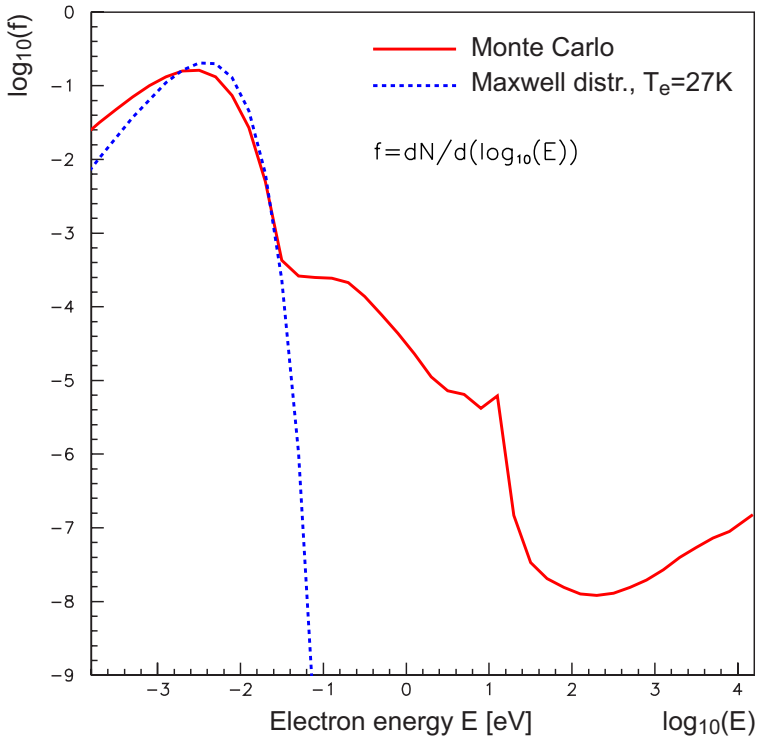

Figure 129: Energy distribution f (solid line, see text for details) of secondary electrons within the WGTS, in comparison with a Maxwell distribution for a temperature $T = 27\,\mathrm{K}$ (dotted line).

The above computations assume that there is no essential heating mechanism (frequently present in plasma physics phenomena), which could prevent the electron cooling process. In the WGTS system of KATRIN the β decay is the main primary source of energy production; there is no external heating process of the plasma. Our investigations show that the effect of superelastic collisions to the electron cooling is negligible. The effect of some possible instability mechanism to the electron energy distribution is planned to be studied in the future.

Since the axial conductivity of the electrons along the magnetic field lines is still very high –in contrast to the very small radial conductivity perpendicular to the magnetic field lines– they will immediately compensate any positive residual space charge within the WGTS. This simple picture is confirmed by our simulations in the frame of an ambipolar diffusion and transport model [227]. Therefore, the inhomogeneity of the electric potential along a magnetic field line within the WGTS is expected to be of the order of a few electron characteristic energies only with a thermal electron energy of $kT = 2.4\,\mathrm{meV}$, based on the above Monte Carlo computations. The high energy tail of the electron energy distribution could slightly change the potential variation (this effect will be studied later). The potential itself is determined by the boundary conditions. Placing a conducting plate into the magnetic flux tube with a well-defined potential at the rear end leads to a well defined potential within the flux tube of the WGTS. Between this plate and the rear end

of the WGTS tube, a small potential drop occurs which is expected to be a few times kT.

If the number of secondary electrons would not be sufficient to ensure quasi neutrality (for instance extremely large recombination unlikely at T=27 K), or if some presently unknown heating mechanism would prevent the electrons to cool to almost gas temperature, then it is considered to offer additional low energy electrons axially from the rear plate, for instance by photo-emission. In addition, the β-electrons impinging on this plate will reemit secondary electrons into the plasma. These may be even sufficient to compensate fully the β current. In case of electron emission from the rear plate, the potential drop between this plate and the WGTS will decrease, and it could even change sign in case of overcompensation.

For special electron gun measurements, this rear plate needs a hole of a few mm diameter to allow the electrons from the electron gun to enter the WGTS. To ensure a constant electrical potential over the WGTS during the measurement of the tritium β spectrum, this hole should be closed. In addition to the rear plate we plan further measures to monitor and manipulate charged particles which were described in section 4.1.5.

Based on the above discussions, potential variations within the WGTS should be limited to about 10 mV. Assuming these variations to be Gaussian, a systematic shift on m_ν^2 can be extracted using eq. (60) leading to shifts of

$$\Delta m_\nu^2(\text{WGTS potential variations}) < 2 \cdot 10^{-4}\,\text{eV}^2. \tag{89}$$

These effects are therefore more than an order of magnitude smaller than the envisaged KATRIN systematic uncertainty.

11.4.9 Synchrotron radiation

The electrons, performing fast gyrational motion in the high magnetic field of the source and transport system, can lose energy due to synchrotron radiation. Such an energy loss has a strong dependence on the starting angle of the electron, therefore it changes the shape of the transmission function.

The energy loss due to the synchrotron radiation can be calculated by using the following formula (valid in the non-relativistic approximation):

$$\Delta E = 0.4\, E_\perp B^2 t\,, \tag{90}$$

where $E_\perp$ denotes the transversal energy component, B is the magnetic field strength in Tesla, t is the time (in second) that the electron spends in the region of B. Both the transversal energy and the time depend on the starting angle θ of the electron direction, therefore the synchrotron radiation energy loss is also sensitive to this parameter: it is zero for $\theta = 0$, and it is maximal for the largest starting angle in the WGTS ($\theta_{max} = 51°$).

The electrons suffer most of their synchrotron radiation energy loss inside the front transport (FT) system. For the 18.6 keV electrons this energy loss can be estimated by

$$\Delta E_{FT} = 0.095\, B_{FT}^2 l_{FT} \frac{\sin^2\theta_{FT}}{\cos\theta_{FT}}\,. \tag{91}$$

Here the energy loss ΔE_{FT} is in meV, B_{FT} and l_{FT} denote the magnetic field strength and the length of the front transport system (in T and m), θ_{FT} is the angle between the electron momentum and magnetic field direction inside the front transport system ($\sin\theta_{FT} = \sin\theta \cdot \sqrt{B_{FT}/B_{WGTS}}$). Using B_{FT}=5.6 T and l_{FT}=12 m, we obtain the following maximal synchrotron radiation energy loss for $\theta = 51°$:

$$\Delta E_{FT}^{max} = 130\,\mathrm{meV}\,. \tag{92}$$

The electrons radiate some energy also inside the WGTS tube. In this case the energy loss can be estimated as

$$\Delta E_{WGTS} = 0.095\,B_{WGTS}^2 l_{WGTS}\frac{\sin^2\theta}{\cos\theta}\,. \tag{93}$$

Here l_{WGTS} denotes the axial distance between the decay point of the electron and the front end of the WGTS tube. The maximal energy loss here is given by using the parameter values B_{WGTS}=3.6 T, l_{WGTS}=10 m, $\theta = 51°$:

$$\Delta E_{WGTS}^{max} = 12\,\mathrm{meV}\,. \tag{94}$$

The electron radiates some energy also inside the pinch magnet. Nevertheless, this is planned to be installed downstream from the main spectrometer (before the detector), *i.e.* after the energy analysis, therefore this energy loss has no effect to the transmission function.

The change of the transmission function due to the synchrotron radiation within the whole system can be accurately computed, since the input parameters of this calculation (magnetic field, geometry, tritium gas density in WGTS) are expected to be determined rather precisely.

11.4.10 Doppler effect

The Doppler broadening of the spectral lines is an unpleasant circumstance in precision spectroscopy. Owing to the Doppler effect, the random motion of the atoms or molecules in a gas causes the spectral lines emitted by the gas to be broadened.

In β decay, the motion of the decaying particle also changes the observed energies of the outgoing particles, relative to the energies within the center-of-mass system (CMS) of the decaying particle. The theoretical spectrum refers to the electron energy in the CMS of the T_2 molecule (E_{CMS}); on the other hand, for the transmission function calculation the energy value in the laboratory system (E_{LAB}) is required. The difference (in the non-relativistic approximation) is given by

$$\varepsilon = E_{LAB} - E_{CMS} = \frac{1}{2}m[(V_\parallel + v)^2 + V_\perp^2] - \frac{1}{2}mv^2 \approx m\,v\,V_\parallel\,, \tag{95}$$

where m is the electron mass, v the electron velocity in the CMS of the molecule, $V_\parallel$ denotes the velocity component of the decaying molecule in the direction of the outgoing

electron, $V_\perp$ is the transversal velocity component of the molecule[33]. Using $m = 10^{-30}\,\mathrm{kg}$, $v = 8 \cdot 10^7\,\mathrm{m/s}$, and a typical value $V_\parallel = 200\,\mathrm{m/s}$, we get $\varepsilon \approx 100\,\mathrm{meV}$. Note that the tritium molecules have much lower energy (2–3 meV).

In order to compute the distribution of the ε values, as a function of the decay point and the β-electron direction, one has to use the velocity distribution of the molecules (parallel and perpendicular to the magnetic field), including the gas flow and the isotopic composition of the gas ($T_2, TD, \ldots$).

11.4.11 Elastic $e^- - T_2$ scattering

The β-electrons undergo not only inelastic, but also elastic scattering with the tritium molecules. The total elastic cross section of the 18.6 keV electrons with T_2 molecules is about 12 times smaller than the inelastic (ionization+excitation) cross section. In addition, the elastic collision of the high energy electrons is very forward peaked, the average scattering angle is about 3° [228]. The energy loss of the electron due to elastic scattering is

$$\Delta E = 2\frac{m}{M}E(1 - \cos\theta_s)\,, \tag{96}$$

where m and M denote the electron and molecule masses, E the electron energy, and θ_s is the polar scattering angle. The maximal energy loss is 6.8 eV (with T_2 molecules), nevertheless the average energy loss is only 16 meV, since most of the events have small scattering angle values, therefore the $1 - \cos\theta_s$ factor is very small. The elastic $e^- - T_2$ scattering can be rather reliably calculated (especially for large scattering angle).

This kind of collision has two different effects to the response function: first, the electron loses some energy, and second, the angle between the electron momentum direction and the magnetic field also changes after the elastic scattering event (for example: a small part of the electrons having large starting angle in the WGTS (above 51°) can reach the detector, after an elastic collision with large scattering angle). At the small energy region of the response function (below 1 eV) the energy loss effect is important, at the high energy region (above 20 eV) the angular scattering effect is dominant. Nevertheless, the change of the KATRIN response function due to the elastic $e^- - T_2$ scattering is rather small: a few times 10^{-4}. If not accounted for in the analysis of the β spectrum, this systematic effect, with a 35 eV analysis window and a background rate of $\Gamma = 1\,\mathrm{mHz}$ (100 mHz), would amount to a shift on the fitted m_ν^2 value of about $\Delta m_\nu^2 = 0.0025\,\mathrm{eV}^2$ ($0.005\,\mathrm{eV}^2$).

11.4.12 Trapped electrons in the WGTS

Each β-electron which is trapped in a local magnetic minimum due to eq. (18) may undergo inelastic scattering. In the source–transport system there are regions where the magnetic field has a local minimum. Let us assume that an electron is created at a point inside this region with magnetic field B_{in}, the field outside (in both directions) is B_{out}; we

[33]This energy difference is due to the coordinate transformation from the CMS to the laboratory system. It is completely different from the kT gas temperature energy.

define the trapping angle θ_{trap} by the equation $\sin^2\theta_{trap} = B_{in}/B_{out}$. Then, if the starting angle θ of the electron is larger than θ_{trap}, the electron is trapped inside of this region by the magnetic mirror effect. This trapped electron can lose energy by inelastic scattering, and it can have elastic collision with large polar scattering angle. After these collisions it can reach the detector. Thus, these local magnetic field regions provide an additional systematic effect for the neutrino mass determination.

In the Troitsk system, the tritium source is in a magnetic field region with B_{in} =0.8 T, with B_{out} =5 T on both sides; therefore, most of the β-electrons are created in a large trapping region. In the KATRIN design, this kind of huge electron trapping is avoided: the magnetic field in the whole rear transport system is smaller than the 3.6 T field inside the WGTS tube, therefore the β-electrons reflected back from the front transport system by the 5.6 T field (see fig. 26) are not reflected back again by the rear transport system; they go to the rear wall, where they are subject to surface recombination.

Nevertheless, even in the KATRIN design there are some small trapping regions, and one has to estimate the systematic effect due to the trapped electrons inside these regions. First, there are local field minima inside the pumping ports (with $B_{in} \approx 0.5$ T, $B_{out} \approx 3.6$ T or 5.6 T, see fig. 34). Therefore, many of the electrons created by β decay inside these pumping port regions are trapped by magnetic mirror effect. However, the number of the tritium molecules inside these pumping ports is about 50 – 100 times smaller than the total number of tritium molecules in the whole source–transport system, therefore the overall trapping effect from these regions is $\Delta m_\nu^2 < 10^{-4}\,\mathrm{eV}^2$.

In addition, there are some small trapping regions inside the WGTS tube. There the B_{in} field is only by 1–2 % smaller than B_{out} (see also section 11.4.4), thus only those electrons are trapped, whose starting angle is close to 90°. First, the number of these trapped electrons is small (a few % of all electrons) due to the small phase space; second, they need a large angle elastic scattering (about 40°), in order that they could go through the pinch magnet and thus they reach the detector. The differential elastic scattering cross section for large angle is very small, therefore the probability that an electron suffers such kind of large angle elastic scattering, before losing substantial energy (larger than 50 eV) due to inelastic scattering, is extremely small. Numerical calculations show that the systematic effect due to these trapped electrons inside the WGTS tube (with 1–2 % magnetic field inhomogeneity) is much smaller than the overall elastic scattering effect considered above.

11.4.13 Backscattering

There are two sources for β-electrons being scattered back into the flux tube with an opening angle $\theta < \theta_{\max}$, either scattering on T_2 molecules in the WGTS or on the rear plate behind the DPS1-R. The scattering cross section on T_2 is dominated by inelastic scattering and is highly forward peaked [228]. Therefore, backscattering within the WGTS does not play any significant role for the narrow energy interval below the β endpoint considered for the KATRIN experiment.

The rear plate will consist of a high Z surface material which enhances the probability

for backscattering compared to a graphite substrate for the Mainz QCTS which had a backscattering coefficient of $3 \cdot 10^{-5}$/eV. The exact backscattering probability as well as the energy loss have to be determined in MC simulations taking into account the exact geometry and material composition of the rear plate. Note that these backscattered electrons then have to pass through the total WGTS, i.e. a column density of $\rho\, d = 5 \cdot 10^{17}/\mathrm{cm}^2$. With a scattering cross section on T_2 of $\sigma_{\mathrm{inel}} = 3.456 \cdot 10^{-18}\,\mathrm{cm}^2$, these electrons will most likely undergo at least one inelastic scattering process (see also section 11.4.2) and hence will be suppressed considerably in the energy regime near E_0.

11.4.14 Systematic effects of the QCTS

The KATRIN standard tritium source will be the WGTS. However, there is the option to use also a quench-condensed Tritium source, QCTS. Several additional systematic uncertainties are connected with the QCTS which are summarized in the following:

- Neighbor excitation
  The sudden change of nuclear charge in β decay can excite even neighboring T_2 molecules. According to [95] the probability is $a_{nex} = 5.9\,\%$ in a closely packed single crystal and the mean excitation energy is given as 14.6 eV, based on the spectrum of free hydrogen molecules. For the analysis of the Mainz data the former number has been lowered to $a_{nex} = 4.6\,\%$ and the latter raised to 16.1 eV [87]. The changes account for the reduced density of the quench condensed film [87] and for the observed shift of the energy loss spectrum of 18 keV electrons passing gaseous and solid hydrogen, respectively [94]. This shift is also corroborated by quantum-chemical calculations [229]. In the final analysis of the Mainz data, the fraction of neighbor excitations has been introduced as a free parameter, giving consistent best fit values of $a_{nex} = (5.0 \pm 1.6)\%$ (in good agreement with theoretical expectations) and $m_\nu^2 = 0 \pm 3\,\mathrm{eV}^2$ [96]. To improve on this situation a quantum chemical calculation of the neighbor excitation for the quench condensed T_2 case would be useful. For the KATRIN data this effect will play a marginal role, since inelastic events will not contribute to the signal in a significant way.

- Self-charging
  Due to the continuous radiation of β particles a quench condensed T_2 film –being an excellent insulator– charges up. This effect has been studied in detail and explained by a satisfactory model [97, 107]. In temporal equilibrium, self-charging generates a linear drop of the electric potential across the T_2 film of about 20 mV per monolayer. For the Mainz analysis [99] 20 % of the total self-charging effect was taken into account as systematic uncertainty. Refined measurements could probably reduce this uncertainty by another factor of 2. More serious than systematic uncertainties is the 2 eV broadening of the spectrum due to self-charging of the QCTS since it reduces the sensitivity on m_ν.

- Long term behavior
  A tritium loss of about 0.16 monolayer per day was observed at the Mainz QCTS.

The effect is due to sputtering of T_2 molecules by nuclei recoiling from β decays. It cannot be avoided but has been monitored precisely by measuring the long term decrease of the count rate. In parallel a condensation of H_2 from the residual gas was observed by ellipsometry at the end of the run. This effect will be avoided in the KATRIN experiment by providing more effective cryogenic vacuum conditions at the QCTS.

In summary, self-charging of a high-intensity QCTS, its static potential shift as well as the broadening of the spectrum are the limiting systematic uncertainties of the QCTS. With the substantially reduced KATRIN systematic uncertainties based on

- the very small energy interval of interest below the endpoint E_0,
- additional control measurements at ultrahigh resolution in both the MAC-E- and MAC-E-TOF mode,
- a ppm stability of the high voltage,
- higher T_2 concentration and
- the improved vacuum conditions

with respect to former experiments, self-charging of the QCTS would be the dominant systematic error and therefore clearly favors the WGTS as the prime KATRIN tritium source.

11.5 Expected total systematic uncertainty

In the previous sections of chapter 11, the main contributions to systematic uncertainties on the analysis of m_ν^2 have been described. Based on the knowledge of the theoretical description of the β spectrum as well as the knowledge and stabilization of experimental parameters, there is no systematic effect giving rise to more than $\Delta m_\nu^2 = 0.0075\,\mathrm{eV}^2$, in fact, most investigated sources of systematics turned out to be considerably smaller. In table 6, the effects discussed in the previous sections are summarized. Adding all systematic uncertainties quantified so far quadratically (see last line of table 6) results in a systematic uncertainty of $\sigma_{\mathrm{syst,tot}} \approx 0.01\,\mathrm{eV}^2$. However, a final and complete analysis of all systematic effects cannot be performed in this early stage of the experiment. We therefore anticipate as the total systematic uncertainty of KATRIN a value of

$$\sigma_{\mathrm{syst,tot}} \leq 0.017\,\mathrm{eV}^2 \tag{97}$$

for the observable m_ν^2 corresponding to 5 individual systematic uncertainties of maximal value $\Delta m_\nu^2 = 0.0075\,\mathrm{eV}^2$ added quadratically. As shown before, studies of systematics have demonstrated so far that the limit in eq. (97) seems rather conservative and achievable.

source of systematic shift	achievable/projected accuracy	systematic shift $\sigma_{\mathrm{syst}}(m_\nu^2)[10^{-3}\mathrm{eV}^2]$
description of final states	$f < 1.01$	< 6
T^- ion concentration $n(T^-)/n(T_2)$	$< 2 \cdot 10^{-8}$	< 0.1
unfolding of the energy loss function (determination of f_{res})		< 2 < 6 (including a more realistic e-gun model)
monitoring of $\rho\, d$ $[E_0 - 40\,\mathrm{eV}, E_0 + 5\,\mathrm{eV}]$	$\Delta\epsilon_T/\epsilon_T < 2 \cdot 10^{-3}$ $\Delta T/T < 2 \cdot 10^{-3}$ $\Delta\Gamma/\Gamma < 2 \cdot 10^{-3}$ $\Delta p_{\mathrm{inj}}/p_{\mathrm{inj}} < 2 \cdot 10^{-3}$ $\Delta p_{\mathrm{ex}}/p_{\mathrm{ex}} < 0.06$	$< \frac{\sqrt{5}\cdot 6.5}{10}$
background slope	$< 0.5\,\mathrm{mHz/keV}$ (Troitsk)	< 1.2
HV variations	$\Delta\mathrm{HV}/\mathrm{HV} < 3\,\mathrm{ppm}$	< 5
potential variations in the WGTS	$\Delta U < 10\,\mathrm{meV}$	< 0.2
magnetic field variations in WGTS	$\Delta B_S/B_S < 2 \cdot 10^{-3}$	< 2
elastic $\mathrm{e}^- - \mathrm{T}_2$ scattering		< 5
identified syst. uncertainties	$\sigma_{\mathrm{syst,tot}} = \sqrt{\sum \sigma_{\mathrm{syst}}^2} \approx 0.01\,\mathrm{eV}^2$	

Table 6: Summary of sources of systematic errors on m_ν^2, the achievable or projected accuracy of experimental parameters (stabilization) and the individual effect on m_ν^2 for an analysis interval of $[E_0 - 30\,\mathrm{eV}, E_0 + 5\,\mathrm{eV}]$ if not stated otherwise (for details see individual chapters in section 11).

11.6 Expected statistical uncertainty

Since the MAC-E filter acts as an integrating spectrometer, the spectrum recorded by varying the retarding potential U is an integral β decay spectrum $N(qU, E_0, m_\nu)$ including experimental effects such as the total number N_{tot} of tritium nuclei in the source, the measuring time t_U at the retarding potential U and the response function f_{res}. Therefore, the simulation of β spectra consists of creating integral spectra

$$N(qU, E_0, m_\nu) = N_{tot} \cdot t_U \int_0^{E_0} \frac{dN_\beta}{dE}(E_0, m_\nu) \cdot f_{\mathrm{res}}(E, qU) dE \tag{98}$$

and adding a background component $N_b = \Gamma \cdot t_U$ considered to be a constant rate Γ independent of the variation of the retarding potential U (see section 11.4.7):

$$N_{th}(qU, R_s, R_b, E_0, m_\nu) = R_s \cdot N(qU, E_0, m_\nu) + R_b \cdot N_b\,. \tag{99}$$

The parameters R_s and R_b denote relative norms of the signal and the background. For the simulation of a KATRIN-like spectrum, the expected value N_{th} is randomly smeared according to a Gaussian distribution with $\sigma_{th}(U) = \sqrt{S + N_b}$:

$$N_{exp}(qU) = N(qU, E_0, m_\nu) + N_b + Rnd(\mathrm{Gauss}, \sigma_{th}(U))\,. \tag{100}$$

parameter	value
column density	$\rho d = 5 \cdot 10^{17}\,\mathrm{cm}^{-2}$
analysing plane	$A_A = (\pi \cdot 450^2)\,\mathrm{cm}^2 = 63.6\,\mathrm{m}^2$
source cross section	$A_S = (\pi \cdot 4.11^2)\,\mathrm{cm}^2 = 53\,\mathrm{cm}^2$
magnetic field strengths	$B_S = 3.6\,\mathrm{T}$, $B_{max} = 6\,\mathrm{T}$, $B_A = 3 \cdot 10^{-4}\,\mathrm{T}$
Tritium fraction in the WGTS	0.95
maximal opening angle	$\theta_{max} = 50.77°$
Tritium endpoint energy	$E_0 = 18575.0\,\mathrm{eV}$
energy resolution	$\Delta E/E = 1/20000 = 0.93/18575$
zero-loss probability	$P_0(\rho d, \theta_{max}) = 0.413$
background rate per interval U	$\Gamma_b = 0.01\,\mathrm{s}^{-1}$
detection efficiency	$\epsilon_{\mathrm{Detector}} = 0.9$

Table 7: Parameter values used for the standard KATRIN simulations.

Table 7 summarizes the input parameters used for the simulation of KATRIN-like spectra. If not stated explicitly otherwise, these values are common for all simulations deducing systematic shifts as well as statistical uncertainties of m_ν^2. Figure 130 shows two typical integral count rate spectra for retarding potentials near the endpoint energy E_0=18575 eV as expected from the KATRIN experiment including the spectrometer response function and background. From first sight already, most count rates for $m_\nu = 0.5\,\mathrm{eV}$ in the β spectrum region are lower than for $m_\nu = 0\,\mathrm{eV}$. This is further illustrated in the inset figure where the relative difference $\Delta N/\sigma = [N(m_\nu = 0) - N(m_\nu = 0.5)]/\sqrt{N(m_\nu = 0)}$ is plotted as a function of the retarding potential U. Whereas there is a clear difference $< \Delta N/\sigma >= 1.5$ for the β spectrum region, the average $< \Delta N/\sigma >$ is compatible with zero for the region dominated by background. Note that a significant difference due to a non-vanishing neutrino mass clearly extends a few tens of eV below the endpoint demonstrating the necessity to analyse an interval of such size below E_0.

The analysis of simulated experiment-like spectra is based on a minimization of the function

$$\chi^2(R_s, R_b, E_0, m_\nu^2) = \sum_i \left(\frac{N_{exp}(qU_i) - N_{th}(qU, R_s, R_b, E_0, m_\nu^2)}{\sigma_{th}(U_i)} \right)^2 \qquad (101)$$

by varying 4 free parameters independently, the signal and background overall strength R_s and R_b, the endpoint energy E_0 as well as the square of the neutrino mass m_ν^2. To avoid any bias, the fitting procedure allows for negative, i.e. unphysical values of m_ν^2. We deduce limits of confidence on m_ν^2 or m_ν in terms of frequency of occurrence by repeating large samples of simulated experiment-like integral β spectra.

11.6.1 Optimization of the measuring time

As can be seen from fig. 130, the effect of a non-vanishing neutrino mass in KATRIN-like spectra is maximal a few eV below the endpoint where the signal to background ratio

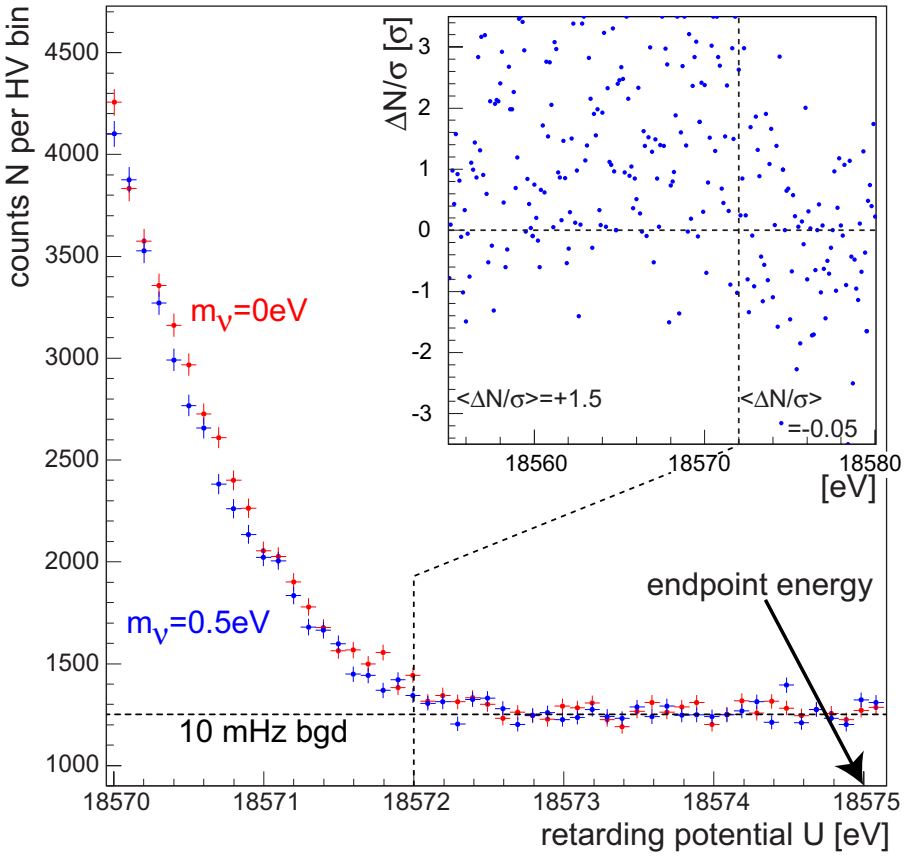

Figure 130: (a) Typical integral β spectra including a constant background of $\Gamma = 10\,\mathrm{mHz}$ per measuring point for two values of the neutrino mass $m_\nu = 0\,\mathrm{eV}$ (red; upper dots) and $m_\nu = 0.5\,\mathrm{eV}$ (blue; lower dots) for a measuring time of 1 year uniformly distributed within [18555,18580] eV.

is about 2:1. Therefore, for a given total measuring time and energy interval, the time spent per retarding potential to minimize the statistical error on m_ν^2 strongly depends on U as well as the background rate Γ and the width of the transmission function (*i.e.* the energy resolution). We optimized the measuring time distribution for different intervals below the endpoint for typical KATRIN count rate spectra and background levels [226] (*e.g.* fig. 131a). These distributions are shown in figure 131(b) for a total measuring time of 3 years. The measuring interval can be described as a combination of three regions: The one at low values of U (high count rate) mainly determines the endpoint from extrapolating the β spectrum, the region around a signal-to-background $\approx$2:1 is the most sensitive to a potential non-vanishing neutrino mass. At retarding potentials above the endpoint the actual background rate per HV bin is deduced. Compared to a uniform measuring time distribution, an improvement in the statistical accuracy of 40-50 % can be achieved with an optimized distribution as shown in fig. 131(b).

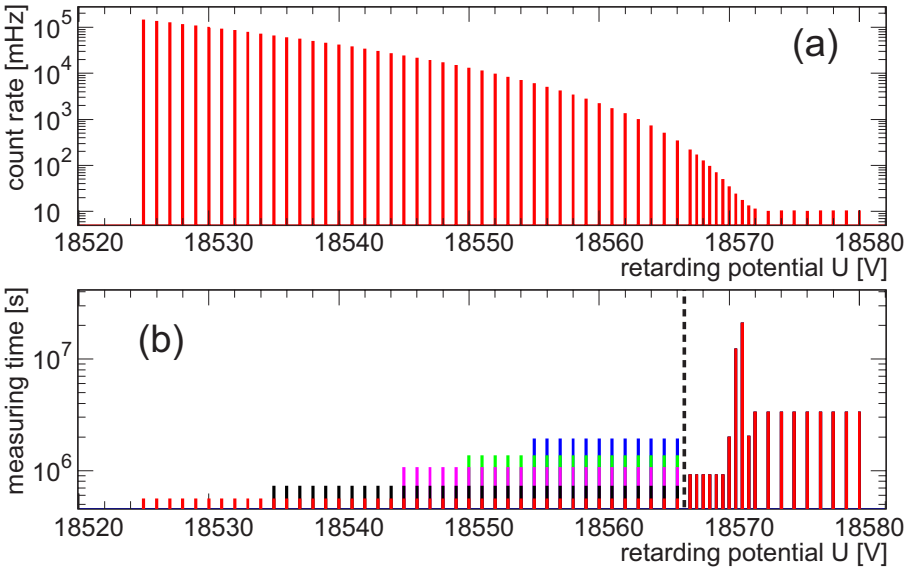

Figure 131: (a): Simulated count rates at various values U of the retarding potential for a given background rate of $\Gamma = 10\,\mathrm{mHz}$. (b): optimized measurement time per retarding potential of the above spectrum for widths from 50 (red) to 20 eV (blue) below E_0 of analyzed intervals, total measuring time 3 years. The measurement time for $U > 18566\,\mathrm{V}$ (dotted line) does not depend on the width of the analyzed interval.

11.6.2 Analyzed energy interval and background rate

It is obvious that the statistical error on m_ν^2 decreases with the width of the analyzed energy interval near the β endpoint E_0 as long as there is enough time spent where the spectrum is most sensitive to the neutrino mass. In a uniform distribution of measurement time per HV bin, this fact limits the statistical accuracy as can be seen from figure 132. Shown are the statistical errors σ as a function of the width of the measurement interval for a total of 3 years measuring time in different experimental configurations. The upper curve (a) shows σ for a 7 m diameter main spectrometer and a background rate of $\Gamma = 10\,\mathrm{mHz}$ as given in [1]. Improving the experimental parameters (increase in source and spectrometer diameter, increase in isotopical purity of T_2 in the WGTS) leads to the curve (b) with a limited accuracy of $\sigma \approx 0.024\,\mathrm{eV}^2$ for measuring intervals of 25 to 45 eV below E_0. With the optimization of the measuring time as discussed in the previous chapter, the statistical error is further reduced (c). For this KATRIN reference configuration and an interval of $[E_0 - 30\mathrm{eV}, E_0 + 5\mathrm{eV}]$, the statistical uncertainty amounts to

$$\sigma_{\mathrm{stat}} = 0.018\,\mathrm{eV}^2. \tag{102}$$

The bottom curve (d) in fig. 132 shows the statistical accuracy which could be reached with an order of magnitude lower background rate[34].

Note that the statistical uncertainty for the reference measuring interval with the anticipated background rate of $\Gamma = 10\,\mathrm{mHz}$ is almost equal to the total systematic uncer-

[34] As a rule of thumb, the statistical accuracy of m_ν^2 scales with the background rate as $\Gamma^{1/5}$.

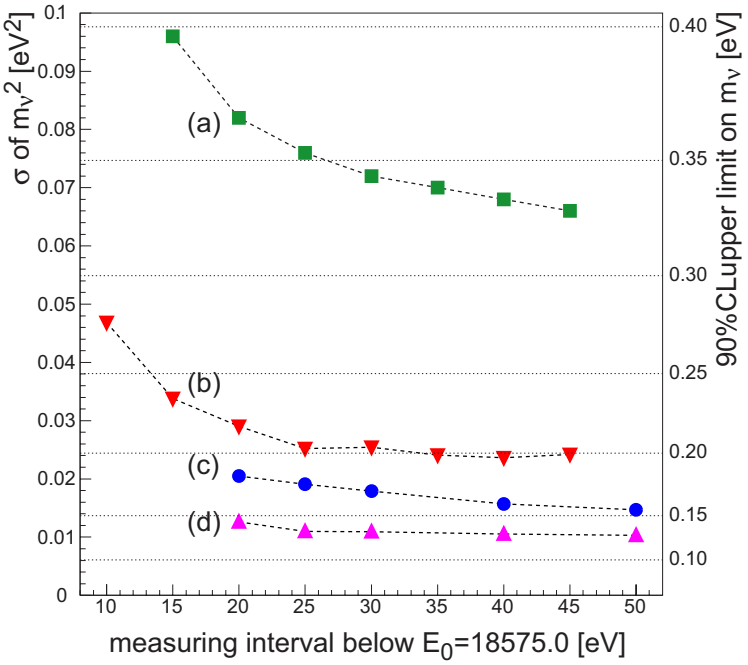

Figure 132: Statistical uncertainty σ (3 years measurement time) of the observable m_ν^2 and corresponding 90% C.L. upper limit on m_ν as a function of the analyzed interval for different configurations and background rates: (a) values stated in the Letter of Intent [1]; (b) 10m configuration, uniform measurement time; (c) 10m configuration, optimized measurement time; (d) 10m configuration, optimized measurement time, but background rate $\Gamma = 1\,\mathrm{mHz}$ instead of 10 mHz as for (a) through (c) (see text for details).

tainty given in eq. (97). A larger measuring interval would slightly improve the statistical uncertainty, but systematic uncertainties such as from scattering of β-electrons rise with lower limits $E_0 - qU_{\mathrm{min}}$ of the interval under analysis. Therefore, we expect a minimum in the total uncertainty on m_ν^2 and thus a maximum in sensitivity on m_ν for measuring intervals of about $[E_0 - 30\mathrm{eV}, E_0 + 5\mathrm{eV}]$. The axis to the right in fig. 132 indicates the 90% C.L. upper limit on m_ν assuming the corresponding statistical uncertainty only for $m_\nu = 0\,\mathrm{eV}$ as $L(90\%\,\mathrm{C.L.}) = \sqrt{1.64 \cdot \sigma_{\mathrm{stat}}}$.

11.7 Measurement time for monitoring and T_2 scanning

In section 11.4.2, the scheme of how to determine the KATRIN response function and its monitoring were presented. This procedure as well as other calibration measurements and maintenance intervals limit the measurement time in the β spectrum scanning mode. In the following, a measurement scheme is given which is not yet fully defined in detail but underlines the possibility to achieve an efficient β spectrum scanning time of more than 60 % per year.

Figure 133 shows the proposed KATRIN measurement scheme. The total measuring time per year in terms of tritium runs is 240 days consistent with the values stated in the Addendum to the Letter of Intent [2]. There will be a period of initial measurements such

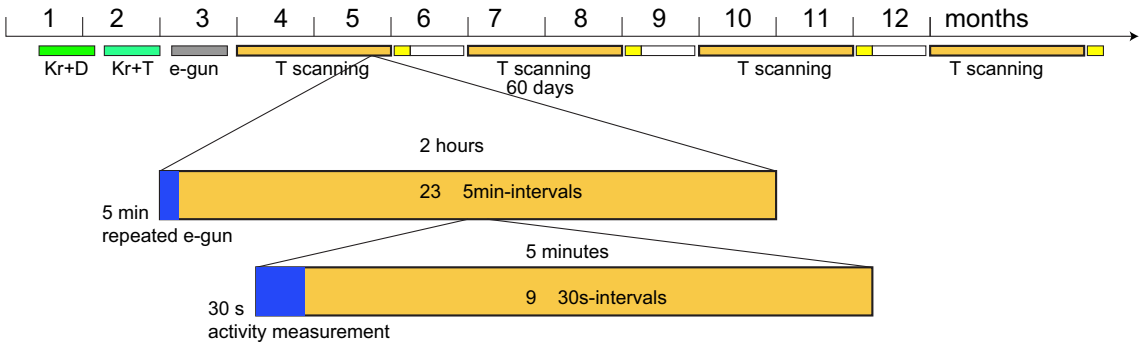

Figure 133: Measuring time distribution for a normal Tritium measuring mode of 60 days including monitoring and scanning phases. Also shown are initial measurements with Kr and e-gun as well as intermediate periods of about 30 days for maintenance (CPS Tritium outbaking etc.).

as calibration with gaseous krypton in deuterium and in the final T_2 configuration as well as the initial e-gun measurements which will take about 30 days. Each tritium run will be separated into ≈ 2 hours intervals starting with the 5 min e-gun reference measurement of the relative $\rho\, d$ value. The following 115 minutes are then separated by 23 intervals of 30 s each for determining the actual tritium β activity. These special measurements account for a total fraction of 9% of the tritium runs, i.e. we have a T_2 scanning efficiency during tritium runs of $\epsilon(T_2 \text{ scanning}) = 91\%$. A continuous monitoring of the tritium activity (by rear or ring monitor) would further increase the efficiency to $\epsilon(T_2 \text{ scanning}) = 95.8\,\%$. A T_2 β spectrum scanning period would last about 60 days followed by a maintenance period of about one month. These intervals would then comprise e.g. outbaking of the cryogenic pumping system and special calibration measurements.

12 KATRIN Sensitivity and Objectives

12.1 Neutrino mass sensitivity

The sensitivity of KATRIN to the neutrino mass m_ν is defined by the systematic and statistical uncertainties on m_ν^2. As discussed in detail in section 11, the total systematic uncertainty and the statistical uncertainty for a 3 year measurement are expected to be of the same size. Adding the values from equations (97) and (102) quadratically leads to an uncertainty of $\sigma_{\mathrm{tot}} \approx 0.025\,\mathrm{eV}^2$. Assuming a vanishing neutrino mass, this uncertainty

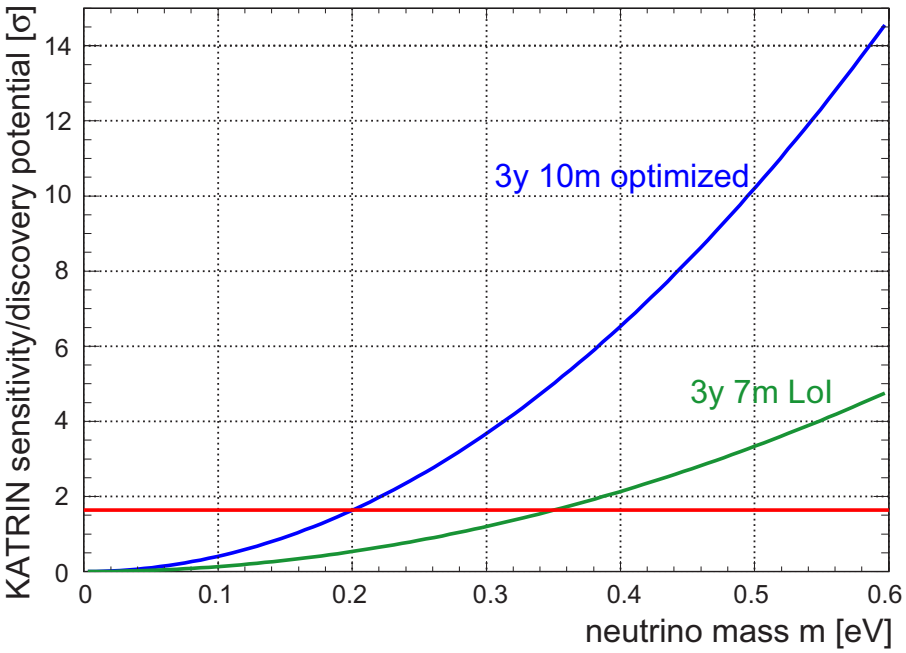

Figure 134: Discovery potential of the KATRIN experiment. Upper curve after improving the experimental set-up and optimizing the measuring time compared to the expectation stated in [1] (lower curve).

translates into an upper limit L(90% C.L.) which is connected to the error on m_ν^2 via $L(90\%\,\mathrm{C.L.}) = \sqrt{1.64 \cdot \sigma_{\mathrm{tot}}}$. After three years of measuring time, this limit becomes

$$m(\nu_e) < 0.2\,\mathrm{eV} \qquad (90\%\,\mathrm{C.L.}) \tag{103}$$

with no finite neutrino mass being observed. This sensitivity improves the existing limits by more than one order of magnitude and is almost 50 % better than the formerly envisaged sensitivity in the Letter of Intent [1]. Figure 134 shows this improvement in sensitivity. The horizontal line represents the 1.64σ or 90% C.L. limit where the crossing with the upper curve as well as the older one (3y 7m LoI) indicate the neutrino mass sensitivity. It also demonstrates the discovery potential of KATRIN as significance in units of σ_{tot} for an electron neutrino mass in the sub-eV range. For example, a potential neutrino mass $m(\nu_e) = 0.4\,\mathrm{eV}$ could be seen by KATRIN with a $6.5\,\sigma$ significance.

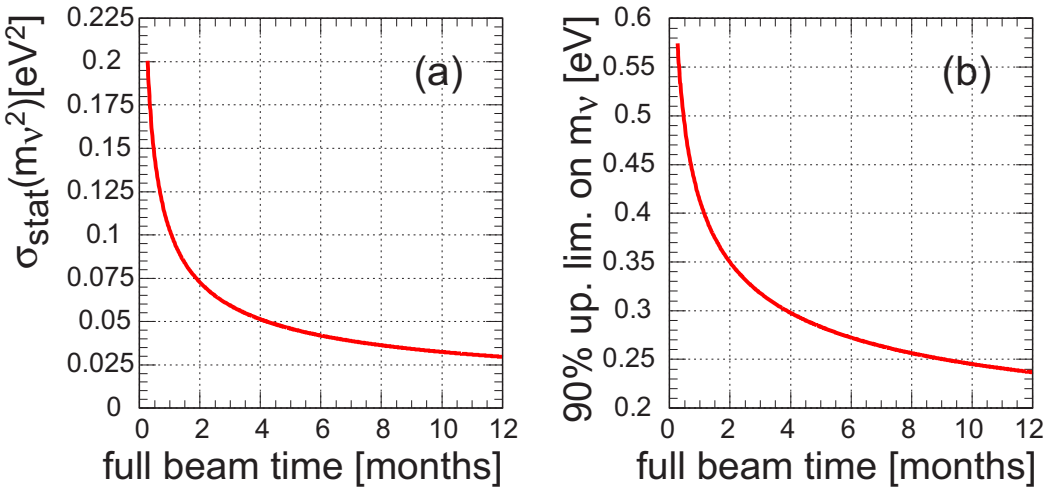

Figure 135: Development in time of the statistical uncertainty (a) and the projected sensitivity (b) under the assumption of a systematic uncertainty of $\sigma_{\mathrm{syst,tot}} = 0.017\,\mathrm{eV}^2$.

Figure 135 shows the statistical uncertainty σ_{stat} as a function of the effective measuring time. Within a few weeks of effective β spectrum scanning, the KATRIN statistical accuracy will exceed the accuracy of the Mainz and Troitsk experiments by more than an order of magnitude. The evolution of the sensitivity curve (fig. 135b) relies on the accurate knowledge of the systematic uncertainty. In the plot, we assumed $\sigma_{\mathrm{syst,tot}} = 0.017\,\mathrm{eV}^2$ which requires extensive measurements, e.g. of the response function, before measuring the β spectrum in the optimized scanning mode. The starting point of the time axis therefore does not include special initial calibration measurements of KATRIN (see section 11.7).

12.2 Impact on Astroparticle Physics and Cosmology

Massive neutrinos with a mass of the order of 0.1 eV or more may play an important role in the energy balance of the present-day universe as well as in the formation of large scale structures in the early universe.

From observations of the fluctuations of the cosmic microwave background radiation (CMBR), the overall energy density of the universe is very close to the critical density

$$\rho_c = \frac{3H^2}{8\pi G_N} = 1.88 \cdot 10^{-26} \cdot h^2\ \mathrm{kg/m^3} = 1.05 \cdot 10^4 \cdot h^2\,\mathrm{eV/cm^3} \tag{104}$$

with the present-day scaled Hubble parameter $h_0 = 0.71$ [25].

The density of primordial neutrinos $n_{\nu 0}$ can be derived from the measured density of the cosmic microwave background photons using the neutrino freeze out in the early universe

$$n_{\nu 0} = \frac{3}{4}\frac{g_\nu}{g_\gamma}\frac{4}{11} n_{\gamma 0} = 336\ \mathrm{cm^{-3}} \tag{105}$$

with the density $n_{\gamma 0} = 411/\mathrm{cm}^3$ for the photons of the CMBR and the relativistic degrees of freedom $g_\nu = 6$ and $g_\gamma = 2$ for the sum of all light neutrino flavors and the photons, respectively.

Using the results from oscillation experiments on the mass differences of the neutrino mass eigenvalues, an upper limit on the electron neutrino mass $m(\nu_e) < 0.2\,\mathrm{eV}$ from KATRIN would also apply to the neutrinos ν_μ and ν_τ and therefore correspond to a density limit

$$m(\nu_e) < 0.2\,\mathrm{eV} \quad \Rightarrow \quad \Omega_\nu = \frac{\rho_\nu}{\rho_c} < \frac{6.4 \cdot 10^{-3}}{h^2} \tag{106}$$

taking the present-day density of primordial neutrinos from eq. (105). With $h_0 = 0.71$, the neutrino density would be limited to about 1 % of the critical density in the universe. Thus, KATRIN would clearly rule out a substantial contribution of primordial neutrinos to the mass and energy density of the universe.

As Hot Dark Matter in the early universe, neutrinos with their free streaming would limit structure formation on smaller scales in contrast to Cold Dark Matter. Observations of large scale structures (CMBR fluctuations, galaxy surveys) are typically analysed with multidimensional fits of cosmological parameters with the neutrino mass being one of these free parameters. Instead of deducing a neutrino mass indirectly from these analyses, the KATRIN result itself will be an independent laboratory input into cosmological models.

In the case of a positive neutrino mass signal from KATRIN, this result would not only determine the νHDM contribution to the matter density of the universe. Such a signal would also be of a different quality than the cosmological multi-parameter fits based on the so-called "concordance model" (see section 1.5). In case of non-compatible neutrino mass results from laboratory and cosmos, this could even point to physics beyond the "concordance model" (sec. 1.5.4).

Since the KATRIN sensitivity on m_ν will be significantly better than what can be achieved by time of flight measurements of neutrinos from a Supernova explosion, the KATRIN result will define the maximal dispersion of the neutrino signal. This can then be used to deduce more reliable information on the evolution of the neutrino emission once the neutrino signal and its time spread of a Supernova explosion will be recorded [65, 66, 67, 68].

One explanation for ultra high energy cosmic rays above the so-called GZK cut-off is the existence of clouds of massive primordial neutrinos on which UHE cosmic ray neutrinos would annihilate to produce a Z-boson on its mass shell, the so-called Z-burst scenario [48, 49, 50, 51]. With the results from KATRIN, the neutrino mass as free parameter in such models of UHE cosmic rays will be much better defined than nowadays, thereby restricting or favoring these models.

12.3 Impact on Particle Physics

A detailed overview of the Particle Physics motivation for the KATRIN experiment has been given in sections 1.3 and 1.11. Here, we shortly summarize the KATRIN impact on

this field, starting with a summary of KATRINs impact on different schemes of neutrino masses and its connection to $0\nu\beta\beta$ decay experiments and their results on m_{ee}. In the following sections, a list of hypothetical effects beyond the Standard Model, which could distort the β spectrum near its endpoint is given. All effects, in general, compete with a non-vanishing neutrino mass in an otherwise SM description of the β decay. Therefore, a detailed analysis of the experimental KATRIN spectrum towards a statement on $m(\nu_e)$ has to take into account potential non-SM effects.

12.3.1 Neutrino masses

The KATRIN sensitivity of $m(\nu_e) = 0.2\,\mathrm{eV}$ (90% C.L.) separates the two possible neutrino mass schemes, *i.e.* quasi-degenerate neutrino mass scenarios from hierarchical ones (fig. 2). Thus, KATRIN is able to solve this long-standing open question.

There is an evidence for $0\nu\beta\beta$ decay of ^{76}Ge with a best fit value on the effective Majorana mass of $m_{ee} = 0.44\,\mathrm{eV}$ [59]. KATRIN will be able to derive such an electron neutrino mass $m(\nu_e)$ at a level of $6.5\,\sigma$ (fig. 134). As pointed out in sec. 1.11.2, the direct kinematic investigation of single β decay is complementary to the search for $0\nu\beta\beta$ decay. Whereas the sensitivity of future experiments on $0\nu\beta\beta$ decay of ^{76}Ge will reach a sensitivity of of $T_{1/2} > 2{\cdot}10^{26}$ years (90% C.L.) within 3 years of measurement time [63], the corresponding limit on the effective mass will range from $m_{ee} < 0.09 - 0.29\,\mathrm{eV}$ depending on the uncertainty in the nuclear matrix element as well as a potential cancellation effect due to the Majorana CP phases. KATRIN does not rely on the latter parameters, thus providing independent information on the neutrino mass scheme. In case of a positive mass signal from KATRIN and a stringent upper limit from $0\nu\beta\beta$ experiments the Dirac nature of neutrino masses would be established.

12.3.2 Non V-A contributions

In the SM, the tritium β decay is governed by a V-A current. Other Lorentz current structures, even weaker than V-A, might produce interferences with the SM interaction which result in effects near the β endpoint different from those produced by the purely kinematic effects of massive neutrino [230]. For T_2 β decay, these new interactions can appear as effective currents in the four fermion formulation with the standard space time structures S, P, T, V or A. Based on the dominant SM structures, these can be regrouped as S_L, S_R, T, R or L with $R \sim (V{+}A)$ and $L \sim (V\text{-}A)$ and similar construction for S_L and S_R from S and P [231]. The various non-SM currents may now be generated by different models of physics beyond the SM, e.g. the exchange of a charge-changing scalar, the existence of a vector boson coupling to right-chiral fermions which might mix with the SM W bosons. Fig. 136 shows the SM W^- exchange as well as these potential new exchange schemes. The influence of such new interactions on the shape of the β spectrum near its endpoint have been studied theoretically [232] for a KATRIN sensitivity as well as with data from the Mainz experiment where no indications for non V-A contributions could be extracted [233].

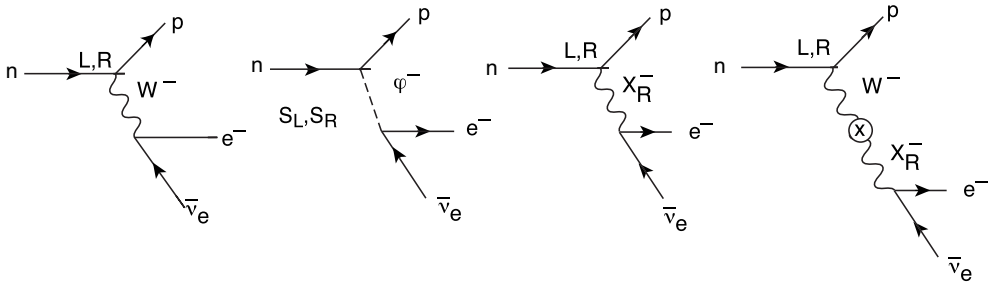

Figure 136: SM and non V-A contributions to tritium β decay via scalar exchange, direct right-chiral vector couplings and possible mixings between the X^- and W^-, adapted from [231]. For all vector couplings, hadronic renormalization leads to both L and R effective hadronic currents.

12.3.3 Violation of Lorentz invariance

The idea that Lorentz invariance is not an exact symmetry, but an approximate one which works extremely well in low-energy experiments, is not new. A plausible dispersion relation for the neutrino is the following one [234]:

$$E_\nu^2 = \mathbf{p}_\nu^2 + m_\nu^2 + 2\lambda|\mathbf{p}_\nu|, \tag{107}$$

where $|\boldsymbol{p}|$ means the module of the three-momentum $\boldsymbol{p}$, m_ν is the neutrino mass, and λ is some mass scale to be determined with $\lambda > 0$ in order to have a positive contribution to the energy squared. The additional term dominates over the standard kinetic term ($\boldsymbol{p}^2$) when $|\boldsymbol{p}| < 2\lambda$ with the non-relativistic kinematics drastically changed. Therefore this type of generalized dispersion relation has to be excluded, except just for one case. The neutrino has two characteristic properties: it has a very small mass, and it interacts only weakly. As a result of this combination, there is no experimental result on its non-relativistic physics. Therefore, the presence of Lorentz invariance violations affecting the non-relativistic limit cannot be excluded a priori in the neutrino case. For details of obtaining the β spectrum with terms violating Lorentz invariance we refer to [234]. In fig. 137 one can clearly distinguish two cases: $\lambda > m$ and $\lambda < m$. Near the end of the spectrum, the Kurie function $K(E)$ for $\lambda > m$ is above a linear approximation $(Q - E)$, which corresponds to an *apparent* excess of electrons at high energies. Indeed it is only apparent, because the curve lies always below the corresponding curve of a relativistic dispersion relation for a massless neutrino, which is also indicated in the figure. In the $\lambda < m$ case, one gets the opposite situation: the effect due to the neutrino mass dominates over the λ term (responsible for the "apparent excess") and results in a reduction of the number of electrons at high energies. Thus, KATRIN will be sensitive to contributions of the order of $\lambda \approx 0.2\,\mathrm{eV}$ for $\lambda \neq m$. If $\lambda = m \neq 0$, the β spectrum again behaves linearly $(K(E)\tilde{Q} - E)$, so KATRIN together with results from cosmology will be able to set a stringent limit on λ.

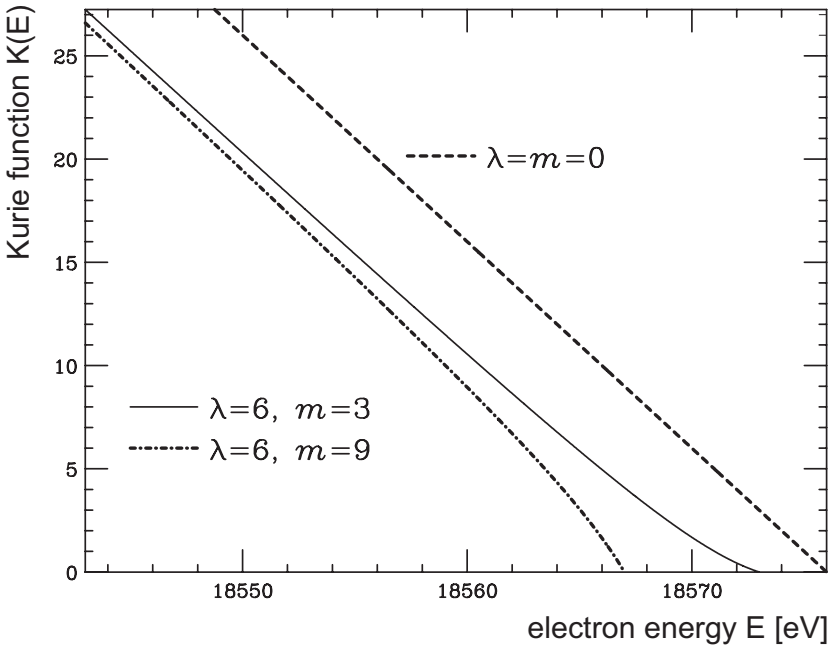

Figure 137: Influence on the tritium β decay from Lorentz non-invariant terms of the neutrino dispersion relation (107), units in eV, in comparison to the undistorted spectrum ($\lambda = m = 0$), adapted from [234].

12.3.4 Sterile neutrinos and extra dimensions

The kinematics of the β decay near its endpoint can also be changed by coupling of the active neutrino ν_{eL} to so-called bulk neutrinos propagating in extra dimensions through which the active neutrino gets a small neutrino mass [21]. Thereby, an extra right-handed neutrino that is a standard model gauge singlet may be a probe of these extra dimensions. Extending the field into the compactified extra dimensions, Kaluza-Klein excitations may arise which may have detectable effects in KATRIN. The couplings of the bulk neutrino states in the extra dimensions with the active neutrino will then result in a massive active neutrino due to the spreading in the bulk volume.

To illustrate the effects of the bulk neutrinos, fig. 138 shows the difference in the Kurie plots of the tritium β decay arising from different scenarios of massive neutrinos. The dashed line shows the deviation due to a single SM massive neutrino of $m = 2.3\,\mathrm{eV}$, the dot-dashed line due to a two neutrino mixing scheme with masses of 2.3 and 25 eV and the solid line shows the scenario with one extra dimension with a radius R of $1/R = 25\,\mathrm{eV}$ and a Dirac mass of the active neutrino of $m = 2.3\,\mathrm{eV}$. The couplings of the active neutrino to the Kaluza-Klein excitations are assumed to be very small. The so-called Kaluza-Klein tower can be seen as various bumps at multiples of 25 eV below E_0. With the parameter values assumed for fig. 138, KATRIN would clearly identify this scenario of extra dimensions. The features of the extra dimensional model become less prominent for a small admixture of a more massive neutrino. However, more realistic models with larger mixings with the active neutrinos and e.g. $1/R = 1\,\mathrm{eV}$ may be constructed [21]

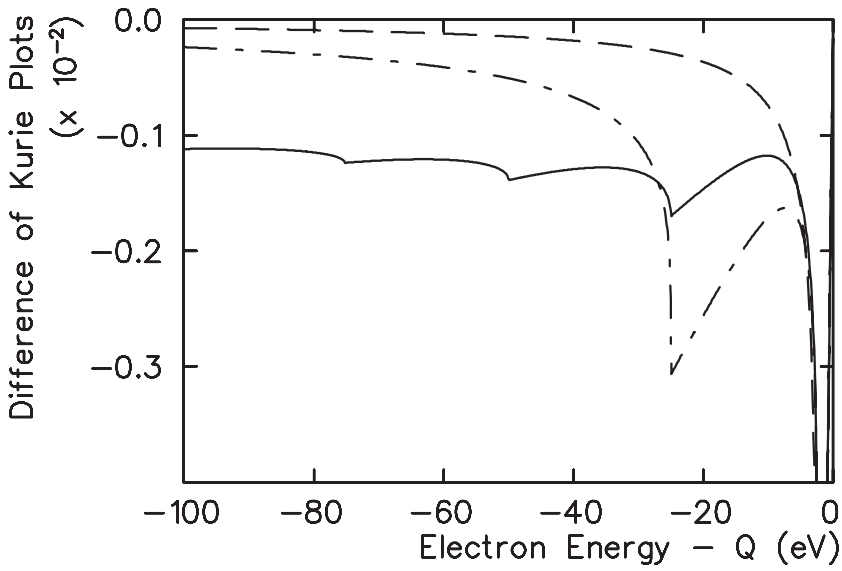

Figure 138: Effects on the β spectrum compared to a standard Kurie plot, for a single massive neutrino of $m = 2.3\,\mathrm{eV}$ (dashed line), for a two neutrino mixing scenario with masses of 2.3 and 25 eV (dot-dashed line) and for a scenario with one extra dimension with a radius R of $1/R = 25\,\mathrm{eV}$, adapted from [21].

which can then be probed by KATRIN.

12.3.5 Models with tachyonic neutrinos

A tachyonic neutrino is a particle which moves with velocities always greater than the velocity of light[35]. The energy-momentum relation is then

$$E^2 - \mathbf{p}_\nu^2 = -\kappa^2 \tag{108}$$

where κ denotes the tachyonic neutrino mass. Such tachyonic massive neutrinos would enhance the count rate near the endpoint E_0 in contrast to a normal massive neutrino (see fig. 139, for details we refer to [235]). This feature was one possible explanation of the excess of counts observed in the Troitsk tritium experiment in the vicinity of the end point [235]. As can be seen from fig. 139, a standard massive neutrino would distort the β spectrum in the opposite direction and hence be clearly distinguished from a potential tachyonic neutrino.

12.3.6 Testing the Troitsk anomaly

The Troitsk experiment has observed a small anomaly in the experimental spectrum a few eV below the β endpoint E_0, which corresponds to a sharp step of the count rate (see section 2.2.1). Since a MAC-E-Filter is integrating, a sharp step corresponds to a narrow line in the primary spectrum. The origin of such an anomaly is not known. Furthermore,

[35]Note that the creation of a tachyonic neutrino does not violate causality [235].

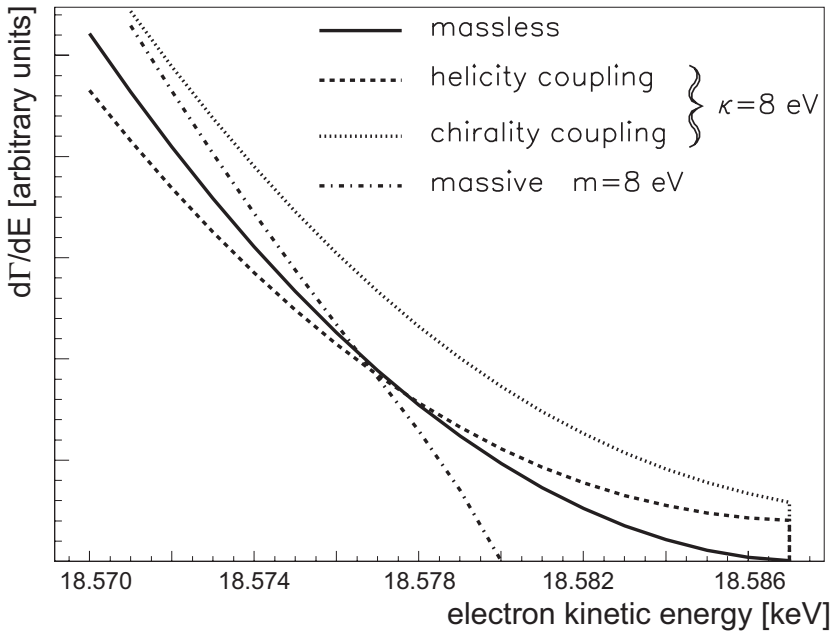

Figure 139: Theoretical tritium β decay spectrum with tachyonic neutrinos of mass $\kappa = 8\,\mathrm{eV}$ as well as standard neutrinos of mass $m_\nu = 0(8)\,\mathrm{eV}$, adapted from [235].

this effect could not be confirmed by Mainz measurements. Experimental upgrades which lowered the background rate at Troitsk reduced also the size of the anomaly.

Nevertheless, if an indication for a step-like structure in the integrated β spectrum would be seen with KATRIN, two independent strategies for a detailed investigation are possible: using a quench-condensed Tritium source or running the system in a non-integrating MAC-E-TOF mode (see section 2.1.1) with the WGTS. This latter non-integrating mode could also be used to further disentangle non-SM effects from pure kinematical effects of a massive neutrino.

13 KATRIN organization and schedule

The initial plans for a next-generation tritium β decay experiment were presented at an international workshop at Burg Bad Liebenzell (Germany) in mid January 2001. Following the very positive response from the international community there with regard to the scientific case for a sub-eV direct neutrino mass experiment, the KATRIN Collaboration was founded in June 2001. Shortly thereafter, the initial layout of the KATRIN experiment was published in a Letter of Intent [1]. Based on this design, the experiment received start-up funding by the German ministry BMBF for the build-up the electrostatic pre-spectrometer as well as for R&D work and specific test measurements with the Mainz spectrometer.

To further strengthen the case for a next-generation direct mass experiment, an updated project status was published in an Addendum to the LoI [2]. In May 2002 an international panel reviewed the physics objectives and project status of KATRIN. The report issued by the review panel strongly supported the physics case for KATRIN, and also encouraged the Collaboration to investigate the possibility of a further increase of the experimental ν-mass sensitivity.

Based on these recommendations, the Collaboration worked out a new reference design with significantly enhanced neutrino mass sensitivity (this report), based on an enhanced source luminosity and enlarged main spectrometer. In parallel, the set-up of hardware components started with the on-site delivery of the pre-spectrometer in September 2003 and the two superconducting guiding magnets shortly thereafter. Over the following months in-depth investigations of the UHV properties of the pre-spectrometer were performed, together with extensive studies at the Mainz and Troitsk spectrometers and other on– and off–site R&D works. These works have formed the basis for the new reference design and also the specifications of the major system components. At the end of 2003, the contract for the first transport element DPS2-F was given to an industrial partner. The DPS2-F assembly is progressing well, and the first magnet winding tests have been performed successfully in November 2004. Earlier in 2004 the detailed specifications for three major system components were worked out: the WGTS, the main spectrometer as well as the 300 W helium refrigerator, forming the basis for the tender procedures. At the end of 2004 the contracts for all three system components were given to industrial partners. In 2005 the components of the outer loop, the cryogenic pumping sections CPS1/CPS2 as well as the element DPS2-R will be specified and ordered together with other components.

The details of the current schedule of the KATRIN project are shown in fig. 141 with a list of milestones in tables 8 through 10. The entire project schedule is focused on a start of first tritium measurements at the end of 2008. Here, the schedule has been sub-divided into three major parts comprising the planning and construction of a) the source and transport (STS) section and b) the spectrometer and detector section as well as of c) the commissioning and test phase. The first on–site element of the STS section of the beam line will be the active pumping element DPS2-F to be delivered at the end of 2005, so that in 2006 an FT-ICR system can be integrated into it. Later, specific test

measurements on the DPS2-F pumping characteristics will be performed.

In mid 2007 the WGTS cryostat will arrive. At this time the planning and setup of the inner loop will be completed so that the WGTS can be connected to the TLK infrastructure. This will be followed by a combined commissioning and initial test phase until mid 2008. In parallel, works on the layout of the outer loop will be performed and finished, as well as the manufacture of the outer loop elements CPS1/2 and DPS2-R, so that the commissioning of the outer loop and the rear system, including element CMS, can be performed in the the first half of 2008.

On the spectrometer side, the pre-spectrometer will serve as a viable test bed for electromagnetic as well as UHV measurements until 2007, when it will be moved from its current location to its final position at the beam line. After a recommissioning period it will be connected to the inner and outer loop in the second half of 2008. The main spectrometer will arrive on-site at FZK in the second half of 2006 and undergo initial leak and UHV testing. Then, extensive installation works will start, culminating in the installation of the large area inner electrode system under clean room conditions. After the completion of these works and UHV recommissioning in early 2008 the main spectrometer will be connected to the upstream end of the beam line via transport magnets to allow detailed tests of its novel electromagnetic design. In parallel, the separate monitor

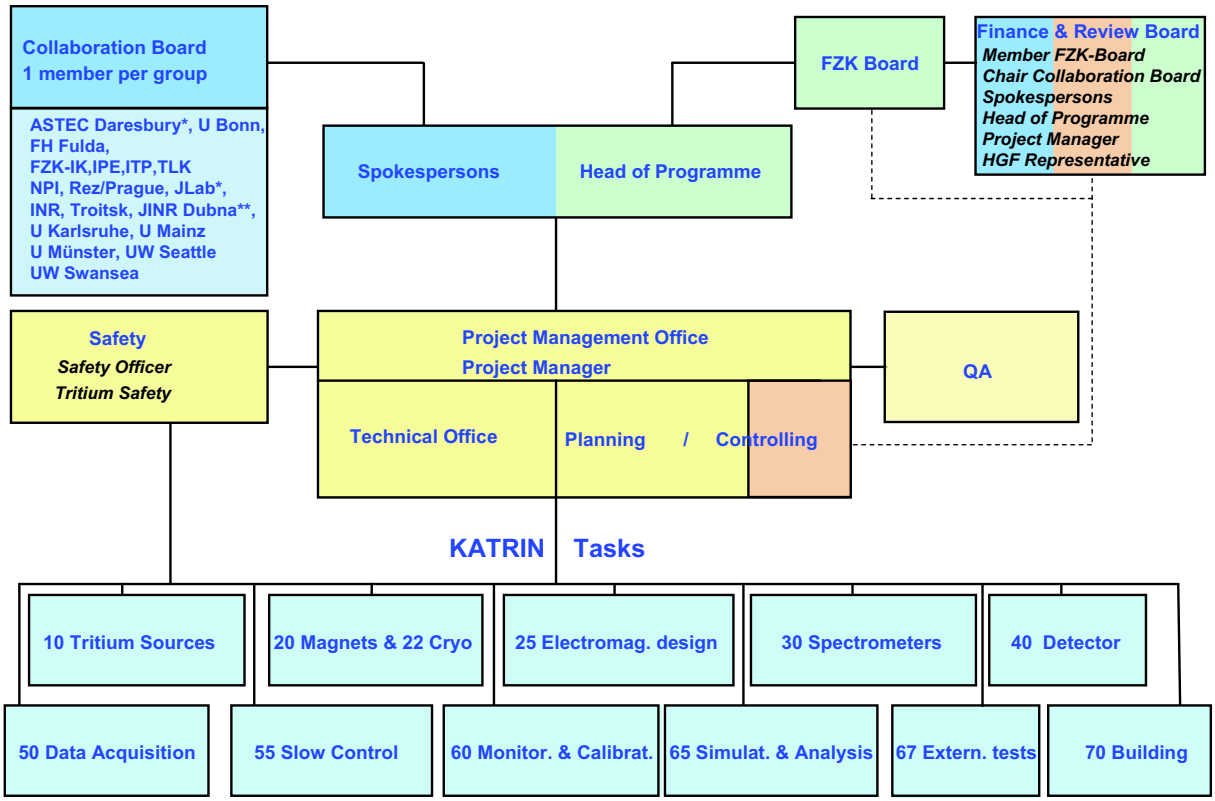

Figure 140: KATRIN organizational chart. The KATRIN collaboration board includes full members, associated groups (**) and expert groups (*).

beam line will be set up, starting with modifications of the Mainz spectrometer and its shipping to FZK. The beam line will be set up in a separate monitor spectrometer hall. Initial functional tests with sources will be completed by mid–2007. In parallel to the spectrometer works, the detector design will be finalized by mid–2005, so that production and testing can be completed by 2007. After shipping and recommissioning of the detector system, the schedule foresees the detector system to be integrated into the beam line in early 2008.

After a 3 months system integration phase comprising the entire beam line from the rear system CMS to the detector, initial tritium measurements are scheduled for the last quarter 2008. These runs will be alternated with run phases dedicated to calibration and systematics. Regular tritium runs are scheduled over a period of 5 calendar years, assuming a 'beam on' fraction per year of about 0.6, so that after this period three 'full beam' years have been accumulated, which is required to achieve the design sensitivity of $0.2\,\mathrm{eV/c^2}$.

Finally, fig. 140 shows the organizational chart of the KATRIN Collaboration with its governing body, the Collaboration Board, where each member institution is represented, and the executive organ, consisting of the spokespersons and the head of programme. The task allocation shown will be adjusted to the actual project phase. The task of the Project Management Office is to plan, monitor and control the project execution and also to organize regular project status meetings and reports for the overview boards. This structure will ensure a tight project execution on schedule with first tritium measurements in 2008.

Figure 141: KATRIN schedule as of January 2005, including milestones described in more detail in tables 8 through 10

Source & transport system planning & construction	
1.1 Troitsk test measurements	
M 1.1.1	Source charging determined by ^{83m}Kr to be below 1 eV
M 1.1.2	Source charging determined by ^{83m}Kr to be below 0.1 eV
1.2 WGTS studies & specification	
M 1.2.1	Source charging estimation
M 1.2.2	WGTS tender action finished
1.3 WGTS technical design & manufacturing	
M 1.3.1	WGTS technical design report
M 1.3.2	Beam tube cooling verification test
1.4 Gaseous ^{83m}Kr source	
M 1.4.1	Conceptual design finished
M 1.4.2	Safety report finished
1.5 Inner loop pre-tests, design & set-up	
M 1.5.1	Demonstration of stabilized gas injection at RT temperature
M 1.5.2	Demonstration of stabilized gas injection at LN2 temperature
1.6 Outer loop set up	
M 1.6.1	Conceptual design fixed
M 1.6.2	Installation finished (including secondary containment)
1.7 Transport system – DPS2-F	
M 1.7.1	DPS2-F tender action finished
M 1.7.2	First HTS module tested
1.8 Transport system - CPS1 / CPS2	
M 1.8.1	Test experiment TRAP commissioned
M 1.8.2	CPS tender action finished
1.9 Rear system – DPS2-R	
M 1.9.1	DPS2-R tender action
M 1.9.2	HTS modules finished and tested
1.10 Rear system – CMS	
M 1.10.1	CMS design finished
1.11 Refrigerator	
M 1.11.1	Refrigerator tender action finished
M 1.11.2	Ready for commissioning
1.12 TLK modification and refrigerator hall	
M 1.12.1	Refrigerator hall ready and TLK modifications finished

Table 8: KATRIN milestones

Spectrometer & detector planning & construction	
2.1 Mainz test measurements	
M 2.1.1	First wire electrode to reduce background as a test for KATRIN spectrometers
M 2.1.2	Considerable background improvement (1.3 mHz compared to ≈15 mHz with solid electrodes)
2.3 Main spectrometer industry studies & specification	
M 2.3.1	Vacuum vessel manufacturing study
M 2.3.2	Main spectrometer tender action finished
2.4 Main spectrometer technical design & manufacturing	
M 2.4.1	Main spectrometer technical design report
M 2.4.2	Successful leak test at company
2.5 Inner electrode design, manufacturing & installation	
M 2.5.1	Detailed technical design of inner electrode
M 2.5.2	First module of inner electrode
M 2.5.3	All modules of inner electrode installed
2.6 Calibration sources & monitor spectrometer set up	
M 2.6.1	Modification of Mainz spectrometer to test calibration sources
M 2.6.2	First calibration source ready
2.7 Detector prototype & tests	
M 2.7.1	Geant4 simulation of detector area
M 2.7.2	Pre-spectrometer detector
2.8 Detector design & manufacturing	
M 2.8.1	Detailed technical design report
2.9 Buildings – planning & licensing	
M 2.9.1	Green map finished
M 2.9.2	Licensing finished
2.10 Buildings – spectrometer hall	
M 2.10.1	Spectrometer hall and infrastructure ready for main spectrometer
2.11 Buildings - other halls & infrastructure	
M 2.11.1	All halls and all infrastructure finished

Table 9: KATRIN milestones (*continued*)

Commissioning & Tests	
3.1 Pre-spectrometer commissioning & systematic studies	
M 3.1.1	XHV conditions reached and start of modifications for electromagnetic tests
M 3.1.2	End of electro-magnetic measurements
3.2 DPS2-F extended tests & investigation	
M 3.2.1	Magnet and vacuum tests finished
M 3.2.2	Electro-magnetic and plasma investigation finished
3.3 WGTS & inner loop – combined commissioning & tests	
M 3.3.1	Successful commissioning
3.4 Transport system & outer loop – combined commissioning & tests	
M 3.4.1	Successful commissioning
3.6 Main spectrometer vacuum test	
M 3.6.1	XHV conditions reached (without inner electrode)
M 3.6.2	XHV conditions reached (with inner electrode)
3.7 Spectrometer electromagnetic tests	
M 3.7.1	Successful electro-magnetic tests
Operation	
4.1 System integration	
M 4.1.1	System aligned and functional with tritium
4.2 Calibration & systematics	
M 4.2.1	Calibration measurement with Krypton sources
M 4.2.2	Transmission function and energy loss

Table 10: KATRIN milestones (*continued*)

14 Acknowledgements

The funds for the major capital investment for the KATRIN experiment are provided by the Helmholtz Association of National Research Centres, Germany, within its research field "Structure of Matter" with the key scientific programme "Astroparticle Physics". We gratefully acknowledge this substantial support.

The KATRIN collaboration gratefully acknowledges support from the German Bundesministerium für Bildung und Forschung under contracts 05CK1VK1/7, 05CK1UM1/5 and 05CK2PD1/5.

We would like to acknowledge the support of KATRIN by the board of directors of Forschungszentrum Karlsruhe, in particular providing access to the large scale infrastructure at Forschungszentrum Karlsruhe.

We acknowledge the financial help of the Deutsche Forschungsgemeinschaft (DFG) with the visitor programmes 436 RUS 17/ and 436 TSE 17/ to allow scientists from Russia and the Czech Republic to visit German universities and the Forschungszentrum Karlsruhe.

The work was supported in part by the Grant Agency of the Czech Republic under contract No. 202/03/0889 and by the Academy of Science of the Czech Republic, project K 1048102.

The test measurements on background identification and suppression were performed with the Mainz spectrometer setup, making use of the technical infrastructure of the Johannes Gutenberg University Mainz. We are grateful for this help.

The University of Wales Swansea wishes to acknowledge PPARC support for KATRIN under project number FD30401.

The Institute of Nuclear Research (INR) activities were supported by the Russian Basic Research Foundation under grant no. 02–02–16908 and 04–02–26085–g as well as by the Russian Academy of Sciences under 'Elementary Particles and Fundamental Nuclear Physics Programme'.

The University of Karlsruhe wishes to acknowledge support by CETA (Centrum für Elementarteilchenphysik und Astroteilchenphysik) and by the Graduiertenkolleg "Hochenergiephysik und Teilchenastrophysik".

Work at the Center for Experimental Nuclear Physics and Astrophysics of the University of Washington is supported under U.S. Department of Energy Grant DE–FG03–97ER41020.

We wish to thank the president of the Helmholtz Association of National Research Centres, Germany, for his support through the President's Initiative and Networking Fund (Impuls- und Vernetzungsfonds). In particular, promotion of young scientists within KATRIN was supported by the Virtual Institute for Dark Matter and Neutrino physics, VIDMAN (VH-VI-033).

15 References

1) A. Osipowicz *et al.*, `hep-ex/0109033`, FZKA scientific report 6691
2) T. Thümmler *et al.*, FZKA scientific report 6752
3) Y. Fukuda *et al.*, Phys. Rev. Lett. **81** (1998) 1562,
   Y. Fukuda *et al.*, Phys. Rev. Lett. **85** (2000) 3999
4) S. Fukuda *et al.*, Phys. Rev. Lett. **86** (2001) 5656
5) Q.R. Ahmad *et al.*, Phys. Rev. Lett. **87** (2001) 071301
6) A. Suzuki *et al.*, Nucl. Phys. **B77** (Proc. Suppl.) (1999) 171
7) K. Nakamura *et al.*, Nucl. Phys. **A663-664** (2000) 795c
8) A. Aguilar *et al.*, Phys. Rev. **D64** (2001) 112007
9) B. Armbruster *et al.*, Phys. Rev. **D65** (2002) 112001
   E.D. Church *et al.*, Phys. Rev. **D66** (2002) 013001
10) plots courtesy H. Murayama, `http://hitoshi.berkeley.edu/neutrino/`,
    see also S. Eidelman *et al.*, Phys. Lett. **B592** (2004) 1 (PDG)
11) M. Ishitsuka (Super-Kamiokande Collaboration) `hep-ex/0406076`
12) Y. Ashie *et al.*, Phys. Rev. Lett. **93** (2004) 101801
13) M. C. Gonzalez-Garcia, `hep-ph/0410030`
14) B.T. Cleveland *et al.*, Astrophys. J. 496 (1998) 505
15) Q.R. Ahmad *et al.*, Phys. Rev. Lett. **89** (2002) 011301
16) Q.R. Ahmad *et al.*, Phys. Rev. Lett. **89** (2002) 011302
17) S.N. Ahmed *et al.*, Phys. Rev. Lett. **92** (2004) 181301
18) J.N. Bahcall *et al.*, JHEP**08** (2004) 016
19) R. Barbieri *et al.*, Phys. Lett. **B90** (1980) 91,
    T. Yanagida, Prog. Theor. Phys. **64** (1980) 1103,
    R.N. Mohapatra *et al.*, Phys. Rev. Lett. **44** (1980) 912
20) A. Zee, Phys. Lett. **B93** (1980) 389,
    L. Wolfenstein, Nucl. Phys. **B175** (1980) 93,
    D. Chang, A. Zee, Phys. Rev. **D61** (2000) 071303
21) G.C. McLaughlin, J.N. Ng, Phys. Rev. **D63** (2001) 053002
22) Y. Farzan *et al.*, Nucl. Phys. **B612** (2001) 59
23) H. Georgi, S.L. Glashow, Phys. Rev. **D61** (2000) 097301,
    V. Barger *et al.*, Phys. Lett. **B437** (1998) 107,
    J.A. Casas *et al.*, Nucl. Phys. **B569** (2000) 82
24) S. Hannestad, `astro-ph/0404239`
25) S. Eidelman *et al.*, Phys. Lett. **B592** (2004) 1 (PDG)
26) X. Wang *et al.*, Phys. Rev. **D65** (2002) 123001
27) D. Spergel *et al.* (WMAP), Astrophys. J. Suppl. **148** (2003) 175
28) R.A.C. Croft *et al.*, Phys. Rev. Lett. **83** (1999) 1092
29) M. Fukugita *et al.*, Phys. Rev. Lett. **84** (2000) 1082
30) M. Tegmark *et al.*, Astrophys. J. **571** (2002) 191,
    M. Tegmark *et al.*, Phys. Rev. **D69** (2004) 103501
31) Ø. Elgarøy *et al.*, Phys. Rev. Lett. **89** (2002) 061301

32) V. Barger *et al.*, Phys. Lett. **B595** (2004) 55
33) S.W. Allen *et al.*, Mon. Not. Roy. Astron. Soc. **346** (2003) 593
34) A. Blanchard *et al.*, Astron. Astrophys. **412** (2003) 35
35) S. Hannestad, JCAP**05** (2003) 004,
S. Hannestad, New J. Phys. **6** (2004) 108
36) J.H. Goldstein *et al.*, Astrophys. J. **599** (2003) 773
37) J.L. Sievers *et al.*, Astrophys. J. **591** (2003) 599
38) S. Tegmark *et al.*, Astrophys. J. **606** (2004) 702
39) W.J. Percival *et al.*, Mon. Not. Roy. Astron. Soc. **337** (2002) 1068
40) H. Päs, P.Q. Hung, `astro-ph/0311131`, report WUE-ITP-2003-012
41) M. Tegmark *et al.*, `astro-ph/0304536`
42) J. Beacom *et al.*, Phys. Rev. Lett. **93** (2004) 121302
43) M. Kaplinghat *et al.*, Phys. Rev. Lett. **91** (2003) 241301
44) D.J. Eisenstein *et al.*, Astrophys. J. **518** (1999) 2
45) `http://www.lsst.org/lsst_home.shtml`
46) W. Buchmüller *et al.*, `hep-ph/0401240`
47) S. Antusch, S.F. King, Phys. Lett. **B597** (2004) 199
48) T. Weiler, Phys. Rev. Lett. **49** (1982) 234
49) Z. Fodor *et al.*, JHEP **06** (2002) 046
50) A. Ringwald, `hep-ph/0111112`, DESY 01-189
51) Z. Fodor *et al.*, Phys. Rev. Lett.**88** (2002) 171101
52) K.S. Capelle et al., Astropart. Phys. **8** (1998) 321
53) S.R. Elliott, J. Engel, J. Phys. G: Nucl. Part. Phys. **30** (2004) R183
54) H.V. Klapdor-Kleingrothaus, *"60 Years of Double Beta Decay - From Nuclear Physics to Beyond the Standard Model"*, World Scientific, Singapore (2001)
55) H.V. Klapdor-Kleingrothaus *et al.*, Eur. Phys. J. **A12** (2001) 147
56) C.E. Aalseth *et al.*, Phys. Rev. **D65** (2002) 092007
57) A. Alessandrello *et al.*, Phys. Atom. Nucl. **66** (2003) 452, `hep-ex/0201038`
E. Fiorini, Phys. Rep. **307** (1998) 309
58) H.V. Klapdor-Kleingrothaus *et al.*, Mod. Phys. Lett. **A16** (2002) 2409
59) H.V. Klapdor-Kleingrothaus *et al.*, Phys. Lett. **B586** (2004) 198
60) R. Arnold *et al.*, JETP Lett. **80** (2004) 429, `hep-ex/0410021`
61) F. Feruglio, *et al.*, Nucl. Phys. **B637** (2002) 345, Nucl. Phys. **B659** (2003) 359
62) C.E. Aalseth *et al.*, Mod. Phys. Lett. **A17** (2002) 1475
63) I. Abt *et al.*, `hep-ex/0404039`
64) `http://www-project.slac.stanford.edu/exo/`
65) P. Vogel, Prog. Part. Nucl. Phys. **48** (2002) 29
66) J.F. Beacom *et al.*, Phys. Rev. **D63** (2001) 073011
67) J.F. Beacom *et al.*, Phys. Rev. Lett. **85** (2000) 3568
68) N. Arnaud et al., Phys. Rev. **D65** (2002) 033010
69) K. Assamagan *et al.*, Phys. Rev. **D53** (1996) 6065
70) R. Barate *et al.*, Eur. Phys. J. **C2** (1998) 395
71) C. Arnaboldi *et al.*, Phys. Rev. Lett. **91**, 16802 (2003)

72) M. Sisti *et al.*, Nucl. Instr. and Meth. **A520** (2004) 125
73) M. Galeazzi *et al.*, Phys. Rev. **C63** (2001) 014302
74) S. Hannestad, `hep-ph/0412181`
75) A.D. Dolgov *et al.*, Nucl. Phys. **B632** (2002) 363
76) S.M. Bilenky *et al.*, Phys. Rev. **D64** (2001) 113003
77) S. Pascoli *et al.*, Phys. Lett. **B524** (2002) 319
78) S. Pascoli, S.T. Petcov, `hep-ph/0111203`
79) M. Czakon, *et al.*, Phys. Rev. **D65** (2002) 053008
80) S. Pascoli, S.T. Petcov, `hep-ph/0205022`
81) R.G.H. Robertson *et al.*, Phys. Rev. Lett. **67** (1991) 957
82) E. Holzschuh *et al.*, Phys. Lett. **B287** (1992) 381
83) H. Kawakami *et al.*, Phys. Lett. **B256** (1991) 105
84) H.C. Sun *et al.*, CJNP **15** (1993) 261
85) W. Stoeffl, D.J. Decman, Phys. Rev. Lett. **75** (1995) 3237
86) V.M. Lobashev *et al.*, Phys. Lett. **B460** (1999) 227
87) C. Weinheimer *et al.*, Phys. Lett. **B460** (1999) 219
88) J. Beamson *et al.*, J. Phys. **E13** (1980) 64
89) V.M. Lobashev, P.E. Spivak, Nucl. Instr. and Meth. **A240** (1985) 305
V.M. Lobashev *et al.*, Nucl. Instr. Meth. **A238** (1985) 496
90) A. Picard *et al.*, Nucl. Instr. Meth. **B63** (1992) 345
91) J. Bonn *et al.*, Nucl. Instr. and Meth. **A421** (1999) 256
92) A.I. Belesev *et al.*, Phys. Lett. **B350** (1995) 263
93) H. Backe *et al.*, Proc. of Neutrino 96, Helsinki/Finland, June 1996, World Scientific/Singapure
94) V.N. Aseev *et al.*, Eur. Phys. J. **D10** (2000) 39
95) W. Kolos *et al.*, Phys. Rev. **A37** (1988) 2297
96) C. Kraus *et al.*, `hep-ex/0412056`
97) H. Barth *et al.*, Prog. Part. Nucl. Phys. **40** (1998) 353
98) B. Bornschein, PhD thesis, Mainz university (2000)
99) J. Bonn *et al.*, Nucl. Phys. **B** (Proc. Suppl.) **91** (2001) 273
100) F. Sharipov, KATRIN internal report `10-ME-2102`
101) J. Angrik, PhD thesis in preparation, Konstanz university
102) D.J. Taylor *et al.*, Rev. Sci. Instrum. **72** (2001) 1970.
103) R.A. Haefer, 'Kryo Vakuumtechnik', Springer Verlag, 1981.
104) C. Benvenuti *et al.*, J. Vac. Sci. Technol. **13,6** (1976) 1172
105) L. Fleischmann *et al.*, J. Low Temp. Phys. **119** (2000) 615
106) L. Fleischmann *et al.*, Eur. Phys. J. **B16** (2000) 521
107) B. Bornschein *et al.*, J. Low Temp. Phys. **131** (2003) 69
108) A. Saenz, *private communication*
109) H.J. Ache *et al.*, Fusion Eng. Design **12** (1990) 331
110) P. Schirra *et al.*, Fusion Eng. Design **18** (1991) 19
111) L. Dörr *et al.*, Proceedings of the 7th Int. Conf. on Tritium Science and Technology (Tritium2004), Baden-Baden, to be published

112) T. Vollmer *et al.*, Fusion Technol. **28** (1995) 998
113) L. Dörr *et al.*, Fusion Science and Technology **41** (2002) 1155
114) M. Glugla, R.-D. Penzhorn, Fusion Eng. Design **28** (1995) 348
115) B. Bornschein *et al.*, Fusion Eng. Design **68** (2003) 51
116) Ansaldo Superconduttori S.p.A., 16152 Genova, Italy, `http://www.as-g.it/`
117) X. Luo *et al.*, KATRIN Internal Report 95-TRP-1751-Luo, 2004
118) J.E. Everett and J.E. Osemeikhian J. Sci. Instrum. **43** (1966) 470
119) F. Schwamm, PhD thesis, Karlsruhe university (2004), FZKA scientific report 7067
120) Love and Scott, J. Phys. D, Appl. Phys. 11 (1678) 1369
121) P. Gerard *et al.*, SCANNING Vol. 17 (1995) 377-386
122) C. Fiorini *et al.*, Nucl. Instr. Meth. **A512** (2003) 265
123) E. Gatti, P. Rehak, Nucl. Instr. Meth. **A225** (1984) 608
124) P. Leutenegger *et al.*, Nucl. Instr. Meth. **A439** (2000) 458
125) P. Holl *et al.*, Nucl. Instr. Meth. **A377** (1996) 367
126) P. Kappen *et al.*, Nucl. Instr. Meth. **A467** (2001) 1163
127) K. Hansen and L. Tröger, IEEE Trans. on Nucl. Sci. **47**, 6 (2000)
128) J. Kemmer, Nucl. Instr. Meth. **A253** (1987) 365
129) W. Neeser *et al.*, Nucl. Instr. Meth. **A477** (2002) 129
130) G. Lutz *et al.*, Nucl. Instr. Meth. **A461** (2001) 393
131) P. Holl *et al.*, IEEE Trans. Nucl. Sci., Vol. **47**, No. 4 (2000)
132) O. Cremonesi, in Proc. of the 9th Int. Workshop on Low Temperature Detectors (LTD-9), Ed. F.S. Porter, AIP Conf. Proceed. Vol. **65** (2002) 443
133) S. Pirro *et al.*, *ibid*, p. 449
134) A. Alessandrello *et al.*, Nucl. Phys. **B87** (2000) 78
135) F. Gatti, in Proc. of the 9th Int. Workshop on Low Temperature Detectors (LTD-9), Ed. F.S. Porter, AIP Conf. Proceed. Vol. **605** (2002) 429
136) A. Alessandrello *et al.*, J. Phys. D: Appl. Phys. **32** (1999) 3099
137) J. Beeman, in Proc. of the 9th Int. Workshop on Low Temperature Detectors (LTD-9), Ed. F.S. Porter, AIP Conf. Proceed. Vol. **605** (2002) 211
138) S. Pirro *et al.*, Nucl. Instr. Meth. **A444** (2000) 331
139) A. Alessandrello *et al.*, Nucl. Instr. Meth. **A412** (1998) 454
140) GEANT4 MC simulation code `http://geant4.web.cern.ch/geant4/`
141) O.C. Allkofer, Introduction to Cosmic Radiation, Verlag K. Thiemig, München (1975)
142) J. Formaggio, University of Washington, Seattle, private communication (2003)
143) G. Heusser, Ann. Rev. Nucl. Part. Sci **45** (1995) 543
144) P. Jagam and J.J. Simpson, Nucl. Instr. Meth. **A324** (1993) 389
145) C.Arpesella *et al.*, Astrop. Phys. 18 (2002) 1
146) J.H. Reeves *et al.*, IEEE Trans. Nucl. Sci. NS-31 1 (1984) 697
147) K. Bunzl *et al.*, Nucl. Instr. Meth. **A238** (1985) 191
148) D.O. Caldwell *et al.*, Phys. Rev. Lett. 65 11 (1990) 1305
149) R. Plaga, Nucl. Instr. Meth. **A309** (1991) 598
150) ACCEL Instruments GmbH, D-51429 Bergisch Gladbach, Germany,

http://www.accel.de/
151) C. Day and G. Drexlin, Extreme High Vacuum - Application and Technology (X-VAT), Proceedings of the International Workshop, Report # FZKA 6886, Forschungszentrum Karlsruhe, Sept. 2003
152) ASTeC Vacuum Quality Control Documentation, available from ASTeC, CCLRC Daresbury Laboratory, Warrington WA4 4AD, UK
153) X. Luo *et al.*, Monte-Carlo simulation of the KATRIN differential pumping system, International Vacuum Congress, Venice, Italy, June/July 2004, to be published in: Vacuum
154) E.W. Otten, Prog. Part. Nucl. Phys. **32** (1994) 153
155) A. Picard, PhD thesis, Mainz university, 1990
156) J.U. Groos, diploma thesis, Mainz university, 1992
157) M. Przyrembel, PhD thesis, Mainz university, 1995
158) D. Goldmann, diploma thesis, Mainz university, 1995
159) H. Barth, PhD thesis, Mainz university, 1997
160) H. Ulrich, diploma thesis, Mainz university, 2000
161) J.P. Schall, diploma thesis, Mainz university, 2001
162) B. Müller, diploma thesis, Mainz university, 2002
163) R.K. Janev *et al.*, Elementary Processes in Hydrogen-Helium Plasmas. Cross Sections and Reaction Rate Coefficients. Springer 1987
164) T. Thümmler, diploma thesis, Mainz university, 2002
165) S. Sanchez, diploma thesis, Mainz university, 2003
166) B. Flatt, PhD thesis, Mainz university, 2005
167) K. Essig, diploma thesis, Bonn university, 2004
168) P.W. Hawkes and E. Kasper, Electron Optics, **1** (Academic Press, 1989)
169) J. Delcroix, Plasma Physics, **1** (John Wiley and Sons, 1968)
170) M.P. Srivastava and P.K. Bhat, J. Phys. Soc. Japan, **32** (1972) 831
171) C.F. Driscoll *et al.*, Phys. Fluids **29** (1986) 2015
172) W. Stoefflˆ, D.J. Decman, Phys. Rev. Lett. **75** (1995) 3237
173) S.T. Staggs *et al.*, Phys. Rev. **C39** (1989) 1503
174) V.M. Gorozhankin *et al.*, Proc. 43^{rd} Int. Conf. Nucl. Spectrosc. Struct. At. Nuclei, Dubna, (1993) p. 181
175) G.C. Nelson, B.G. Saunders, Nucl. Instr. Meth. **84** (1970) 90
176) R.G. Helmer, C. van der Leun, Nucl. Instr. Meth. **A450** (2000) 35
177) M. Breinig *et al.*, Phys. Rev. **A22** (1980) 520
178) R. D. Deslattes *et al.*, Int. Tables for Crystallography, 2^{nd} Edition, Ed. A.J.C. Wilson and E. Prince, Kluwer Acad. Publishers, Norwell, Mass. (1999) p. 200
179) R.S. van Dyck *et al.*, Phys. Rev. Lett. **70** (1993) 2888
180) K. Blaum, *private communication*
181) R. Marx, IEEE Trans. Instr. Meas. Vol. 50, No. 2, 2001
182) A. Picard *et al.*, Z. Phys. **A342** (1992) 71
183) A. Kovalik *et al.*, J. Electron Spectr. Relat. Phenom. **58** (1992) 49
184) O. Dragoun *et al.*, Nucl. Instr. Meth. **A391** (1997) 345

185) R.G.H. Robertson, D.A. Knapp, Ann. Rev. Nucl. Part. Sci. **38** (1988) 185
186) J.L. Campbell, T. Papp, Atom. Data Nucl. Data Tables **77** (2001) 1
187) R.D. Deslattes *et al.*, Rev. Mod. Phys. **75** (2003) 3
188) O. Dragoun *et al.*, Czech. J. Phys. **54** (2004) 833
189) E.G. Kessler, Jr., private commun., 2002
190) K.D. Sevier, Atom. Data Nucl. Data Tables **24** (1979) 323
191) A. Kovalík, V.M. Gorozhankin, J. Phys. G: Nucl. Part. Phys. **19** (1993) 1921
192) D. Vénos *et al.*, Report of NPI ASCR Rez, EXP-07, 2004
193) V.M. Lobashev (2003) *private communication*
194) O. Dragoun *et al.*, Nucl. Instr. Meth. **B 194** (2002) 112
195) A. Picard, Dissertation, Mainz university, 1990
196) R. R. Kinsey *et al.*, The NUDAT/PCNUDAT Program for Nuclear Data, paper submitted to the 9th International Symposium of Capture Gamma-Ray Spectroscopy and Related Topics, Budapest, Hungary, October 1996. Data extracted from the NuDat2—Decay Radiation database version of 12/10/2004 `http://www.nndc.bnl.gov/nudat2/dec_searchi.jsp`
197) H. Kawakami *et al.*, Physics Letters **A 121** (1987) 414
198) ICE: `http://hpesun16.fzk.de/projekt/med/kompo/e_index.html`
199) LabView: `http://amp.ni.com/niwc/labview/resource.jsp`
200) CORBA: `http://www.corba.org/`
201) EJB: `http://java.sun.com/products/ejb/`
202) I. Abt *et al.*, Nucl. Instr. Meth. **A386** (1997) 310
I. Abt *et al.*, Nucl. Instr. Meth. **A386** (1997) 348
203) TOSKA: `http://hikwww4.fzk.de/itp/toska/e_index.html`
204) H. Gemmeke, Proceedings of ICRC 2001, Copernicus Gesellschaft (2001)
205) M.C. Browne *et al.*, IEEE Transactions in Nuclear Science **46** (1998) 873
206) J.J. Simpson, Phys. Rev. **D23** (1981) 649
207) S. Gardner *et al.*, Phys. Lett. **B598** (2004) 188
208) W.W. Repko, C. Wu, Phys. Rev. **C28** (1983) 2433
209) A. Saenz *et al.*, Phys. Rev. Lett. **84** (2000) 242
210) S. Jonsell *et al.*, Phys. Rev.**C60** (1999) 034601
211) S. Jonsell, H.J. Monkhorst, Phys. Rev. Lett.**76** (1996) 4476
212) R.K. Janev *et al.*, Elementary Processes in Hydrgen-Helium Plasmas (1987)
213) R. Johnsen *et al.*, Jornal of Chemical Physics **65**(4), 1539 (1976)
214) M. Bacal *et al.*, Rev. Sci. Instruments **67**(3), 1138 (1996)
215) J. Tennyson (2004) *private communication*
216) S. Trajmar *et al.*, Phys. Rep. **97** (1983) 219
217) H. Tawara *et al.*, J. Phys. Chem. Ref. Data **19** (1990) 617
218) R. Olson *et al.*, J. Chem. Phys. **53** (1970) 3391
219) T.M. Miller *et al.*, Phys. Rev. **173** (1968) 115
220) E. Graham *et al.*, J. Chem. Phys. **59** (1973) 4648
221) D Gerlich and S. Horning, Chem. Rev. **92** (1992) 1509
222) F. Sharipov, J. Vac. Sci. Technol. A**12** (1994) 2933

223) F. Sharipov and V. Seleznev, J. Phys. Chem. Ref. Data **27** (1998) 657
224) M.J. Assael *et al.*, J. Phys. Chem. Ref. Data **15** (1986) 1315
225) M.J. Assael *et al.*, J. Phys. Chem. Ref. Data **16** (1987) 189
226) J. Kaspar *et al.*, Nucl. Instr. and Meth. **A527** (2004) 423
227) A.F. Nastoyashchii *et al.*, Proceedings of the 7th Int. Conf. on Tritium Science and Technology (Tritium2004), Baden-Baden, to be published
228) J. Geiger, Zeitschr. f. Physik, **181** (1964) 413
229) A. Saenz (2004) *private communication*
230) G.J. Stephenson *et al.*, Phys. Lett. **B440** (1998) 89
231) G.J. Stephenson *et al.*, Phys. Rev. **D62** (2000) 093013
232) B.H.J. McKellar *et al.*, Proc. XX Intern. Symp. on Lepton and Photon Inter. at High Energies, Rome, July 2001, `hep-ph/0106122`
233) C. Kraus, diploma thesis, Mainz university, (2000)
234) J.M. Carmona, J.L. Cortés, Phys. Lett. **B494** (2000) 75
235) J. Ciborowski *et al.*, Eur. Phys. J. **C8** (1999) 157