Colloquium: The neutron lifetime

Fred E. Wietfeldt*

Department of Physics, Tulane University, New Orleans, Louisiana 70118, USA

Geoffrey L. Greene†

Department of Physics, University of Tennessee, Knoxville, Tennessee 37996
and Physics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

(published 3 November 2011)

The decay of the free neutron into a proton, electron, and antineutrino is the prototype semileptonic weak decay and is the simplest example of nuclear beta decay. It played a key role in the early Universe as it determined the ratio of neutrons to protons during the era of primordial light element nucleosynthesis. Neutron decay is physically related to important processes in solar physics and neutrino detection. The mean neutron lifetime has been the subject of more than 20 major experiments done, using a variety of methods, between 1950 and the present. The most precise recent measurements have stated accuracies approaching 0.1%, but are not in good agreement as they differ by as much as 5σ using quoted uncertainties. The history of neutron lifetime measurements is reviewed and the different methods used are described, giving important examples of each. The discrepancies and some systematic issues in the experiments that may be responsible are discussed, and it is shown by means of global averages that the neutron lifetime is likely to lie in the range of 880–884 s. Plans and prospects for future experiments are considered that will address these systematic issues and improve our knowledge of the neutron lifetime.

DOI: 10.1103/RevModPhys.83.1173 PACS numbers: 14.20.Dh, 23.40.−s

CONTENTS

I. HISTORICAL INTRODUCTION

A. Early work

The suggestion that the atomic nucleus contains a heavy neutral particle along with the positively charged proton is generally attributed to Ernest Rutherford (1920). This was largely motivated by the observation that atomic mass differs from atomic number and by the fact that an ensemble of positively charged protons alone would not be bound electromagnetically. In suggesting the existence of such a "neutron,"[1] Rutherford clearly had in mind the concept that this heavy neutral particle was a tightly bound combination of an electron and proton:

> Under some circumstances, however, it may be possible for an electron to combine much more closely with the H nucleus, forming a kind of neutral doublet. Such an atom would have very novel properties. Its external field would be practically zero, and in consequence it should be able to move freely through matter...

An important consequence of this picture is that such a neutron, being a bound state, would have a mass less than the sum of the masses of the proton and electron. Of course, if this was true, it would be energetically impossible for the neutron to spontaneously decay into an electron and proton (let alone with an additional antineutrino). As long as this model for the neutron was accepted, the notion of neutron decay would not be considered. Thus, the idea of neutron decay thus depended on the following: (1) a sufficiently precise determination of the neutron mass indicating that the process was energetically allowed, and (2) a theoretical

*few@tulane.edu
†greenegl@ornl.gov

[1]The term neutron was variously used before its current definition became accepted. According to the Oxford English Dictionary, the word neutron was introduced by W. Sutherland (1899), who used it to describe a "molecule" consisting of an electron and a, then extremely hypothetical, positive electron. It is interesting to note that Rutherford does not use the word neutron in the now famous Bakerian Lecture of 1920.

context for the understanding of beta decay. Interestingly, both emerged at nearly the same time.

Chadwick, in presenting his discovery of the neutron in 1932 (Chadwick, 1932), provided the first determination of the neutron mass by noting that it was ''probably between 1.005 and 1.008'' atomic mass units. This result is reasonably consistent with the current accepted value of about 1.0087 amu. However, Chadwick's suggested error bar included the then current value of about 1.007 amu for the mass of the proton. As a result, it was not inconsistent with the Rutherford picture.

Heisenberg (1932), as well as Iwanenko (1932), anticipated that the Rutherford picture was incorrect by suggesting that the neutron was not a bound state, but rather an elementary spin $1/2$ particle, and that nuclei were made of neutrons and protons, rather than protons and electrons. However, they did not appear to consider the possibility of neutron decay. It is noteworthy that Chadwick considered the same idea but dismissed it noting that, ''It is, of course, possible to suppose that the neutron may be an elementary particle. This view has little to recommend it at present.'' (1932).

The first ''accurate'' determination of the neutron mass was made in 1934 by Chadwick and Goldhaber (1934) who compared the binding energy of the deuteron, determined by photo disassociation, with an accurate mass spectroscopic measurement of the hydrogen and deuterium atomic masses (Bainbridge, 1933) Their result of 1.0080(5) amu indicated that the neutron mass was probably greater than that of the proton, but the accuracy was such that it did not convincingly contradict the Rutherford picture.

In 1935, a more accurate result for the neutron mass (Chadwick and Goldhaber, 1935), based largely on improved mass spectroscopic data and further data on photodiassociation of light nuclei, was presented by the same authors. That result of $M_n = 1.0090$ amu clearly showed that the neutron mass exceeded the mass of the hydrogen atom, then estimated at $M_H = 1.00081$ amu. Chadwick and Goldhaber certainly recognized that their result was inconsistent with Rutherford's conception of the neutron as a bound state and observed:

> If the neutron is definitely heavier than the hydrogen atom, then one must conclude that a free neutron is unstable, i.e., it can change spontaneously into a proton + electron + neutrino

This appears to be the first serious suggestion that a free neutron would decay.

Clearly, the first observation of neutron decay and the subsequent accurate measurement of its lifetime would be notable achievements. However, prior to the development of particle accelerators, the only available neutron beams were derived from reactions based on alpha active radioisotopes such as radium-beryllium sources. It is notable that Chadwick's initial observation was performed with a neutron beam having a flux of order 1 neutron per cm^2 per s. Such sources were quite incapable of producing enough neutrons to allow the observation of neutron beta decay. Even the development of the cyclotron, which increased neutron intensities approximately a million fold, was not adequate to allow the detection of neutron decay. This is understandable when one considers that neutrons produced from nuclear reactions typically have energies on the order of MeVs. With an anticipated lifetime of tens of minutes, it would suggest the probability of decay of such a neutron over a flight distance of a few meters would be of order 10^{-10}. The observation of such a rare event would be precluded by both limited statistical sensitivity and background considerations. Of course, Fermi had demonstrated, in his ground breaking neutron research, that it was possible to slow down or ''thermalize'' neutrons by allowing them to make many elastic collisions with light nuclei. Such neutrons, if fully thermalized, would have energies characteristic of the temperature of the moderator material and velocities of a few $\times 10^3$ m/s. Unfortunately, the process of thermalization randomizes the neutrons' directions, so what initially may have been a well collimated beam became a broadly diffuse source. Limitations to neutron source intensity as well as other consideration discouraged any serious attempt to detect neutron decay in the 1930s.

The development of the nuclear reactor provided the prospect of significantly higher neutron flux. According to George Gamow (1966), Fermi suggested an ingeniously simple experiment to detect neutron decay using the first fission reactor (known as Chicago Pile 1 or CP-1) under Stagg Field at the University of Chicago. This involved inserting an evacuated bottle into a reactor in a region of high neutron flux and leaving it for a long period of time. Neutrons traversing the bottle have a small probability of decaying in the vacuum producing a proton and an electron (the antineutrino would of course escape). Ultimately, these electrons and protons would lead to an accumulation of hydrogen in the bottle,[2] some of which might reside in the walls of the bottle but could, in principle, be recovered. Upon removal from the reactor, the amount of hydrogen could be compared with estimates of the neutron flux (from the reactor power), and the estimated thermal velocity of the neutron to provide a determination of the neutron lifetime. While Gamow suggests that the experiment was seriously considered, it is doubtful that it was actually attempted. Indeed, while a detectable amount of hydrogen could be formed in this fashion, it seems dubious that it could have been observed above background in a real experiment. The suggestion does, however, indicate that the possibility of the observation of neutron decay using a fission reactor was recognized quite early on. In fact, the detection of free neutron decay did require the development of reactors that incorporated the extraction of beams of thermal neutrons.

B. First observation

In November of 1943, less than a year after CP-1 achieved criticality for the first time, the world's first dedicated nuclear reactor began operation at Oak Ridge, Tennessee. In contrast to CP-1 which was in reality a rather daring experiment, the Graphite Reactor at Oak Ridge was a very

[2]This bottle, which was to contain hydrogen and be transparent to neutrons, should not be confused with the neutron bottles used in modern determinations of the neutron lifetime.

serious engineering project. Unlike CP-1, which operated only marginally above criticality, the Graphite Reactor operated at the thermal power of ≈ 3.5 MW and had a peak thermal neutron flux in its core of $\approx 10^{12}$ cm^{-2} s^{-1}. The Graphite Reactor was an extraordinary engineering achievement. From a modern perspective, perhaps its most extraordinary aspect is the fact that a 3.5 MW reactor was designed, built, and brought into operation in less than 11 months.

The primary mission of the Graphite Reactor was as a pilot plant for the production of plutonium. However, with considerable foresight, the reactor design incorporated several channels through the concrete radiological shielding allowing access to the high neutron flux at the core. These channels provided the first intense beams of low energy neutrons. It was on one of these beams that the first observation of free neutron decay was carried out by Arthur Snell and collaborators in the 1940s (Snell and Miller, 1948; Snell, Pleasonton, and McCord, 1950).

The detection of neutron decay in a beam of neutrons requires (1) that the neutron flux be sufficient to have a reasonable average neutron density[3] (and thus a reasonable activity per unit volume of beam) and (2) that background events be reduced or rejected. The thermal neutron flux at the Graphite Reactor indeed offered the expectation of a detectable rate of in-flight decays within a beam directed into an external decay detector. Background was a more troubling issue, not only must the measurement be carried out in the close proximity to the core of a nuclear reactor, but the background generated by the beam itself is considerable. This is due to the fact that the probability for a thermal neutron to decay over a flight path on the order of 10 cm (comparable to the length of the Snell *et al.* experiments) is on the order of a few parts in 10^8. However, all neutrons that do not decay will eventually be absorbed producing ionizing radiation. Thus, even if the beam related background is significantly reduced, a simple beam-on versus beam-off comparison is not likely to provide complete background suppression. It is worth considering the Snell *et al.* experiment in some detail as it exhibited many features employed in later neutron decay experiments.

Snell and colleagues choose to address the background issue by coincident detection of decay protons and electrons. Their apparatus is shown schematically in Fig. 1. In the figure, the neutron beam is normal to the page. Electrons are detected by passage through two beta sensitive gas detectors (A and B in the figure). The use of two beta detectors provided rough directional tracking. The protons were more challenging to detect as the end point of the proton energy spectrum is only 752 eV. In order to allow detection of these low energy particles, the neutron beam was largely enclosed in an electrode structure that effectively biased the neutron beam at a positive potential of 8 kV with respect to the proton detector. The proton detector itself was a negatively biased electron multiplier. The positive electrode was designed to efficiently focus the decay protons onto the first plate of the electron multiplier as well as providing sufficient acceleration to allow detection.

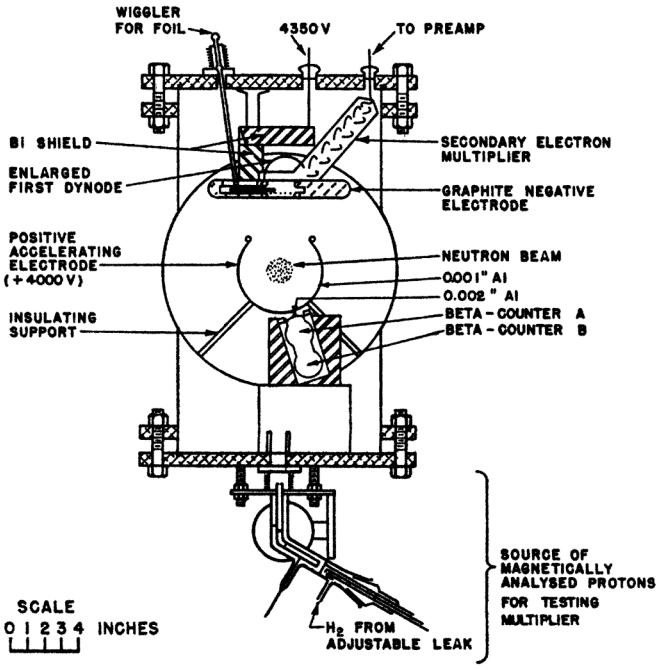

FIG. 1. The apparatus that made the first detection of coincident electrons and protons from neutron decay. The neutron beam passes through the page in the center. Protons are directed upward by the accelerating electrode. The two beta counters are below the beam. From Snell, Pleasonton, and McCord, 1950.

The singles counting rate in each of the beta detectors was 75 000 counts per min. Counting only coincident events in both beta detectors reduced this to $\approx$ 1500 cpm. Considering electron-proton coincidences only further reduce this to about 1 cpm. But this coincidence rate included random events as well as real decays. These two classes of coincident events were distinguished by inserting a 25 μs delay, corresponding to the anticipated flight time of the decay proton, into the coincidence logic. The final result of their analysis was a decay rate of 0.67 ± 0.05 cpm, a clear observation of free neutron decay.

Extraction of a neutron lifetime from an in-beam decay rate requires two addition quantities: the density of neutrons in the beam and the overall efficiency of the decay detection system. Because of uncertainties in the evaluation of the detector efficiencies, Snell, *et al.* were not able give an accurate value for the neutron lifetime, but rather provided an estimate of 10–30 min for its half-life. Independently and nearly simultaneously,[4] Robson, working at the Chalk River Reactor (Robson, 1950), described the observation of neutron decay and provided an estimate of 9–25 min for its half-life. It is noteworthy that both results are consistent with the current value for the neutron half-life of 10.2 min.

[3]Snell *et al.* do not give an explicit value for the flux or density in their beam. A rough estimate suggests they had about 10^3 cm^{-3} s^{-1} in the active volume of the detector.

[4]The 1950 papers of Snell *et al.* (1950) and Robson (1950) appear adjacent to one another, with identical titles, Radioactive Decay of the Neutron, in volume 78 of the Physical Review. Neither paper cites the earlier work of Snell and Miller (1948).

C. First measurement

While both Snell *et al.* and Robson convincingly demonstrated that free neutrons do in fact beta decay, neither claimed to have carried out a measurement of the neutron's mean lifetime. Rather, recognizing the uncertainties in their apparatus, they preferred to set rough limits on its value. The first experiment to be properly called a measurement of the neutron lifetime was performed by Robson (1951) at the NRX reactor at Chalk River, Canada. Like Snell *et al.*, Robson detected decay electrons and protons in coincidence. Also like Snell *et al.*, decay protons were accelerated (in this case to 14 keV) and focused onto an electron multiplier. Electrons in the Robson apparatus were directed into a beta spectrometer. This allowed the first determination of the end point of the neutron beta decay energy spectrum.

An essential feature of the Robson experiment was the care with which the density of the neutron beam in the active decay volume was determined. In principle, it is possible to determine the density of neutrons in a beam by a measurement of both the spectrum of the beam and the integrated flux. In practice, this is difficult to do accurately. However, under conditions that are satisfied by many neutron absorbers, the probability of neutron capture is proportional to $1/v$, where v is the neutron velocity. Robson cleverly recognized that a detector based on such a $1/v$ efficiency would effectively provide (within a multiplicative constant) the density in the neutron beam. The experiment is discussed further in Sec. III.A.

Robson concluded that the half-life of the neutron was 12.8 min with a probable error of ± 2.5 min. This corresponds to a mean lifetime of 1110 ± 220 s. The determination of the neutron flux contributed 8% to the total uncertainty. Robson was also able to provide the first direct determination of the end point of the beta spectrum, which he found to be 782 ± 13 keV. These pioneering measurements agree with current measurements.

II. THEORETICAL ASPECTS OF THE NEUTRON LIFETIME

A. Neutron decay and the weak coupling constants

The decay of the free neutron is the prototype for nuclear beta decay, the process by which a neutron or proton in the nucleus can change identity, via the charged-current weak interaction, to produce the nucleus of another element. Beta decay can occur whenever it is energetically possible, i.e., when the mass of the parent system is larger than the mass of the daughter system. This is why, for example, ^{14}C will beta decay to ^{14}N by conversion of a neutron to a proton. In contrast, ^{14}N is stable because its mass is smaller than that of both ^{14}C and ^{14}O.

A neutron (n) can decay to a proton (p), electron (e^-), and antineutrino ($\bar{\nu}$)

$$n \rightarrow p + e^- + \bar{\nu}. \tag{1}$$

The electron and antineutrino are needed because the weak interaction conserves both charge and lepton number. The free neutron mass (939.57 MeV) exceeds the mass sum of the proton (938.27 MeV), electron (0.511 MeV), and antineutrino (negligible) by 0.782 MeV so it will decay and the excess mass is converted to kinetic energy of the final state particles. It is interesting to note that the neutron-proton mass difference originates from isospin breaking caused by the different masses and charges of their constituent light quarks. If the light quark masses, which make a very minor contribution to the total nucleon mass, were equal, isospin symmetry would tend to equalize the strong interaction contribution to the neutron and proton masses. The proton would presumably have a slightly larger mass due the difference in electrostatic binding energy, in which case the free neutron would be stable while the hydrogen atom would be unstable: it would decay by electron capture to a neutron and neutrino.

The Hamiltonian for beta decay was first written by Fermi (1934), in analogy to the quantum electrodynamic theory, as a four-fermion vector interaction involving the neutron, proton, electron, and antineutrino. Gamow and Teller (1936) later proposed that all Lorentz-invariant interaction forms should be considered. They generalized the Fermi Hamiltonian to include scalar (S), pseudoscalar (P), tensor (T), axial vector (A), and vector (V) terms based on how each transforms under spatial rotations and reflections. The P interaction is strongly suppressed in beta decay because the neutron and proton are nonrelativistic, so we do not consider it here (the pseudoscalar interaction is important in other systems such as muon decay). The S, V interactions create the beta electron and antineutrino as a spin zero state (Fermi decay) while the A, T interactions create them as a spin one (Gamow-Teller decay). This leads to different selection rules for the nuclear states in allowed beta decay: $\Delta J = 0$ for the Fermi case and $\Delta J = 0, \pm 1$ for Gamow-Teller. Because both cases were observed, it was clear that both Fermi (S, V) and Gamow-Teller (A, T) terms must be present in the Hamiltonian. Experiments and theory later showed that the V and A interactions were responsible and, in particular, that the weak current is V–A (Feynman and Gell-Mann, 1958; Sudarshan and Marshak, 1958). To date, there has been no evidence for S or T. This is the basis of the V–A standard electroweak model (SM) of Weinberg, Salam, and Glashow (Weinberg, 1967; Salam, 1968; Glashow and Iliopoulos, 1971). In the SM, the charged weak interaction transforms quark states within the nucleon and is mediated by the W boson, but because the energy released in beta decay is much smaller than the W mass, these details are not important here. The Fermi-Gamow-Teller formulation is quite adequate for describing beta decay. Using modern notation, the matrix element of the Hamiltonian for allowed beta decay can be written[5]

$$\mathcal{M} = [G_V \bar{p}\gamma_\mu n - G_A \bar{p}\gamma_5\gamma_\mu n][\bar{e}\gamma_\mu(1+\gamma_5)\nu]. \tag{2}$$

Here $\bar{p}$, n, $\bar{e}$, and ν are the spinors for the initial and final state neutron, proton, electron, and antineutrino. The term $1 + \gamma_5$ provides the maximal parity violation of the weak interaction. The coupling constants G_V and G_A give the strengths of the vector and axial vector forces. At the quark level, they are equal in magnitude but the strong interaction within the

[5]We choose the sign convention $1 + \gamma_5$ for historical consistency, but note that it is opposite to the more modern $1 - \gamma_5$ convention of Bjorken and Drell.

nucleon can alter the values that appear in Eq. (2). Conservation of vector current, the SM principle that the vector weak current forms a conserved triplet with the electromagnetic current, guarantees that the vector strength remains unchanged. There is no such guarantee for G_A and in fact neutron decay experiments have shown that $|G_A/G_V| > 1$ (see below). The strong interaction also induces other form factors in beta decay, such as weak magnetism, that will be neglected here. There is some inconsistency in the use of the symbols for beta decay coefficients in the literature. We follow the predominant usage $G_V = G_F V_{ud} C_V$ and $G_A = G_F V_{ud} C_A$, where $G_F = 1.16637(1) \times 10^{-5}$ GeV^{-2} is the Fermi weak coupling constant, V_{ud} is the first element of the Cabbibo-Kobayashi-Maskawa matrix, and C_V, C_A are the constants in the Hamiltonian of Jackson, Trieman, and Wyld (1957). In the standard model $C_V = 1$. G_V and G_A do not include radiative or recoil order corrections.

The neutron decay probability per unit time can be computed from Fermi's golden rule:

$$dW = (2\pi)^{-5}\delta(E_e + E_\nu - \Delta)\frac{1}{2E_e}\frac{1}{2E_\nu}d^3\boldsymbol{p}_e d^3\boldsymbol{p}_\nu|\mathcal{M}|^2. \tag{3}$$

Here E_e, $\boldsymbol{p}_e$, E_ν, and $\boldsymbol{p}_\nu$ are the electron and antineutrino total energy and momentum and Δ is the neutron-proton mass difference: 1.293 332 14(43) MeV (Mohr, Taylor, and Newell, 2008). Neutron decay is a spin-1/2 to spin-1/2 decay, so both Fermi and Gamow-Teller decays are allowed. Integrating Eq. (3) over the antineutrino momentum and electron solid angle gives the beta electron energy spectrum

$$\frac{dW}{dE_e} = \frac{\xi}{2\pi^3}E_e|\boldsymbol{p}_e|(\Delta - E_e)^2, \tag{4}$$

where $\xi = G_V^2|\langle 1\rangle|^2 + G_A^2|\langle\sigma\rangle|^2 = G_V^2 + 3G_A^2$ for the neutron. Further integration over electron energy gives the exponential decay constant

$$W = \frac{(G_V^2 + 3G_A^2)}{2\pi^3}f_R. \tag{5}$$

The factor f_R is the value of the integral over the Fermi energy spectrum, including Coulomb, recoil order, and radiative corrections. The neutron lifetime τ_n is the inverse of W

$$\tau_n = \frac{2\pi^3}{(G_V^2 + 3G_A^2)f_R} \tag{6}$$

in natural units. Including physical constants explicitly

$$\tau_n = \frac{2\pi^3\hbar^7}{(G_V^2 + 3G_A^2)m_e^5c^4f_R}. \tag{7}$$

We see that the neutron lifetime provides a means to measure the weak coupling constants G_A and G_V.

A number of theoretical corrections are important to consider at the current 10^{-4} precision level of the neutron lifetime. At tree level, we have recoil effects on the phase space integral and induced hadronic currents at recoil order ($E_{\max}/M_n \approx 8 \times 10^{-4}$). Of the latter, the largest effect is weak magnetism which, as was shown using the conservation of vector current hypothesis, can be computed from the neutron and proton magnetic dipole moments (Gell-Mann, 1958; Holstein, 1974). Radiative corrections are traditionally separated into the *outer* and *inner* corrections. The outer corrections, which total about 1.5% of the neutron lifetime, are the long range QED corrections for both real (bremsstrahlung) and virtual photons, including the infrared divergence and Coloumb corrections to the electron wave function. The inner corrections, about 2%, are the model-dependent short range electroweak corrections. All of these are included in the corrected phase space factor f_R. With recent updates on the radiative corrections, Eq. (7) can be written as (Marciano and Sirlin, 2006)

$$\tau_n = \frac{G_F^2}{G_V^2 + 3G_A^2}4908.7(1.9)\text{ s}. \tag{8}$$

The vector coupling constant G_V is found accurately from the ft values of superallowed $0^+ \to 0^+$ nuclear beta decays. These are pure Fermi (pure vector) decays where there is maximum overlap between the initial and final nuclear states, so the nuclear structure uncertainties in calculating the transition matrix elements are very small. The transition ft value for a particular decay system depends on three quantities that can be measured accurately: the decay energy Q_{EC}, the beta decay lifetime of the parent nucleus, and the branching ratio. Hardy and Towner (2005) evaluated and combined the results of hundreds of such measurements for 12 different nuclei. From this, the value $G_V = 1.1358(5) \times 10^{-5}$ GeV^{-2} is obtained.

Because the axial current is not a conserved quantity, we cannot obtain a reliable and precise value of the axial vector coupling constant G_A from other nuclear decay systems. Therefore, G_A is best measured by neutron decay, either from the neutron lifetime via Eq. (8) or from a beta decay correlation, such as the beta asymmetry that depends on the ratio G_A/G_V.

In addition to neutron decay, the weak coupling constants G_A and G_V govern other important processes that involve charged weak interactions between neutrons and protons, for example (Dubbers, 1991),

$n + e^+ \leftrightarrow p + \bar{\nu}$ (big-bang nucleosynthesis)
$p + e^- \leftrightarrow n + \nu$ (big-bang nucleosynthesis, neutron star formation)
$p + p \rightarrow {}^2H + e^+ + \nu$ (solar fusion)
$p + p + e^- \rightarrow {}^2H + \nu$ (solar fusion)
$\nu + n \rightarrow e^- + p$ (neutrino detection)
$\bar{\nu} + p \rightarrow e^+ + n$ (antineutrino detection).

A key motivation for precision measurement of the neutron lifetime is to obtain precise and reliable values of G_A and G_V. Furthermore, the neutron lifetime can be used, along with other beta decay observables, to place limits on physics beyond the SM. Physics at a large mass scale such as electroweak symmetry breaking or hypothetical leptoquark interactions can cause effective weak scalar and tensor interactions at low energy and manifest as small departures from SM predictions for low energy processes like beta decay. Adding scalar and tensor interactions to neutron decay [Eq. (2)] would put the additional term $G_S^2 + 3G_T^2$ in the denominator of Eq. (7) and therefore affect the neutron lifetime. Important limits on scalar and tensor weak currents can be obtained from a combined fit to experimental beta decay

observables, including the neutron lifetime. By comparing the value of the neutron lifetime to other neutron decay parameters, such as the beta asymmetry A and the neutrino asymmetry B, one can produce interesting limits on the mixing of hypothetical right-handed weak currents and test the unitarity of the Cabbibo-Kobayashi-Maskawa matrix. These various limits on new physics are reviewed in detail in a recent paper (Severijns, Beck, and Naviliat-Cuncic, 2006).

B. The neutron lifetime and big bang nucleosynthesis

The big bang nucleosynthesis model (BBN) predicts the primordial abundances of the light nuclei believed to be produced in the early Universe (Gamow, 1946; Alpher, Bethe, and Gamow, 1948): H, D, ^{3}He, ^{4}He, and ^{7}Li. These predictions are in fairly good agreement with astrophysical observations (Olive, Steigman, and Walker, 2000; Cyburt, Fields, and Olive, 2003; Izotov and Thuan, 2010). At a time of 0.1–1.0 s, the Universe consisted of a hot plasma of protons, neutrons, electrons, positrons, neutrinos, antineutrinos, and photons kept in thermal equilibrium by the weak, strong, and electromagnetic interactions. In particular, the temperature was high enough for the following reactions to remain in equilibrium:

$$n + e^+ \leftrightarrow p + \bar{\nu} \qquad p + e^- \leftrightarrow n + \nu.$$

At that time, the ratio of neutrons to protons was given by a simple Boltzmann factor

$$n/p = e^{-\Delta/T}, \tag{9}$$

with T the temperature of the Universe and $\Delta = m_n - m_p = 1.293$ MeV. The Universe was expanding and cooling at a decreasing rate ($\propto T^2$) but the weak interaction rates were decreasing more rapidly ($\propto T^5$), so there came a time, at $t \approx 1$ s and $T \approx 1$ MeV, where these reactions fell out of equilibrium. This is called ''nucleon freeze-out'' and the time is denoted t_{freeze}. We see from Eq. (9) that the neutron to proton ratio at freeze-out was about $1/6$. For several minutes following t_{freeze} the neutron to proton ratio gradually decreased due to neutron decay, until the Universe was cool enough ($t \approx 100$ s, $T \approx 0.1$ MeV) for the neutrons and protons to combine and form stable deuterium and, ultimately, helium nuclei. By $t = 3$ min, virtually all neutrons in the Universe were bound into light nuclei. The neutron to proton ratio in this period is a key parameter for determining the resulting element abundances.

The neutron lifetime is significant here. First, and most importantly, the reaction rates $n \leftrightarrow p$ prior to t_{freeze} depend on the strength of the charged weak interaction, and, in particular, they are proportional to the combination of coupling constants $G_V^2 + 3G_A^2$ that appear in Eq. (7) for the neutron lifetime. These reaction rates determine t_{freeze} and hence the neutron to proton ratio at t_{freeze} via Eq. (9). Second, the neutron lifetime governs the change in the neutron to proton ratio between t_{freeze} and nucleosynthesis. The precise value of the neutron lifetime is an important input for theoretical BBN calculations of primordial element abundances. In fact, the theoretical uncertainty in the predicted ^{4}He abundance is dominated by the experimental uncertainty in the neutron lifetime (Burles *et al.*, 1999; Lopez and Turner, 1999). We note that t_{freeze} also depends on the number of light neutrino species N_ν, and the BBN abundances depend on the baryon to photon ratio of the Universe η. By combining BBN calculations with precise astronomical observations of primordial element abundances, one can use the neutron lifetime to place constraints on N_ν and η. If we assume the standard model value of $N_\nu = 3$, then a range of $5.1 \times 10^{-10} \leq \eta \leq 6.5 \times 10^{-10}$ is obtained (Cyburt, Fields, and Olive, 2008). The uncertainty is dominated by systematic effects in the astronomical data. This agrees very well with the Wilkinson Microwave Anisotropy Probe result from temperature fluctuations in the cosmic microwave background (Dunkley *et al.*, 2009): $\eta = 6.23 \pm 0.17 \times 10^{-10}$, a remarkable fact considering that the two methods rely on completely different physics occurring at very different time periods in the early Universe.

III. REVIEW OF NEUTRON LIFETIME MEASUREMENTS

Three distinct experimental strategies have been used to measure the free neutron lifetime: the beam method, the bottle method, and the magnetic trap method. We discuss these methods individually, describe their important features, and give important examples of each.

A. Beam method

In the beam method, the specific activity (decay rate per neutron) of a neutron beam is measured. This is the oldest method, having been used in the first true measurement of the neutron lifetime by Robson at the Chalk River pile reactor in 1950 (Robson, 1951). A beam of slow neutrons with density ρ_n passes through a known decay volume V. The instantaneous neutron decay rate $\dot{N}$ in the decay volume is given by the differential equation that governs exponential decay

$$\dot{N} = \frac{dN}{dt} = -\frac{N}{\tau_n} = -\frac{\rho_n V}{\tau_n}. \tag{10}$$

As noted in Sec. I, Robson discovered a method to measure the neutron beam density, which has been employed by many subsequent beam neutron lifetime experiments. It exploits the fact that the slow neutron absorption cross section is inversely proportional to neutron velocity [the $1/v$ law (Fermi *et al.*, 1934)]. It is not an exact law, but it holds to excellent approximation, with a relative error of less than 10^{-4} in many materials, the exceptions being very strong neutron absorbers and isotopes with low energy neutron capture resonances. For a $1/v$ absorber we can write the absorption cross section as

$$\sigma_{\text{abs}}(v) = \frac{\sigma_{\text{th}} v_{\text{th}}}{v}, \tag{11}$$

where σ^{th} is the thermal neutron absorption cross section at the reference thermal velocity $v_{\text{th}} = 2200$ m/s. Typically a very thin absorbing foil is placed in the neutron beam and the reaction products from neutron absorption are detected with efficiency ε_n. The neutron beam normally contains a broad spectrum of velocities. If the differential flux spectrum is $d\phi(v)/dv$, the count rate R_n will be

$$R_n = \varepsilon_n a \rho_{\text{foil}} \int \sigma_{\text{abs}}(v) \frac{d\phi(v)}{dv} dv, \tag{12}$$

where ρ_{foil} is the areal density of the foil (atoms per m^2) and a is the cross sectional area of the neutron beam, which is assumed to be smaller than the foil area. Combining Eqs. (11) and (12),

$$R_n = \varepsilon_n a \rho_{\text{foil}} \sigma_{\text{th}} v_{\text{th}} \int \frac{1}{v} \frac{d\phi(v)}{dv} dv. \tag{13}$$

Noticing that the integral in Eq. (13) is equal to the neutron density ρ_n (neutrons per m^3) in the beam we have

$$\rho_n = \frac{R_n}{\varepsilon_n a \rho_{\text{foil}} \sigma_{\text{th}} v_{\text{th}}}. \tag{14}$$

The neutron decay products (protons and/or electrons) are counted at rate R_p with efficiency ε_p in the decay volume V. This, along with Eqs. (10) and (14), gives an expression for the neutron lifetime in terms of experimental quantities,

$$\tau_n = \left(\frac{L}{\rho_{\text{foil}} \sigma_{\text{th}} v_{\text{th}}}\right)\left(\frac{R_n}{R_p}\right)\left(\frac{\varepsilon_p}{\varepsilon_n}\right), \tag{15}$$

where we have substituted $V = aL$, L being the length of the decay volume.

In the 1950 Chalk River experiment (Robson, 1951), protons and electrons from neutron decay were detected in coincidence in order to reduce detector backgrounds. While effective for that purpose, it had the disadvantage of making the efficiency ε_p strongly dependent on the decay position (it had to be estimated from the acceptances of the electron and proton detectors), which led to a large systematic uncertainty. Robson measured the neutron density by activating a series of manganese foils and determining their activity using classical counting techniques. This required knowledge of (1) the mass of the manganese foils, (2) the neutron capture cross section, and (3) the efficiency of his counting system. Rather than determining each of these quantities separately, he directly measured the overall efficiency by comparison with an independent, well calibrated detector. As we shall see, Robson thus anticipated a technique used in the most recent determination of the neutron lifetime using in-beam decay.

In another early experiment (D'Angelo, 1959) the neuton beam passed through a diffuse cloud chamber. This produced a 100% efficiency ε_p, except for end effects, in a well-defined volume. Unfortunately, neutron absorption in the gas gave very large gamma ray backgrounds. In subsequent experiments the neutron beam path was maintained in high vacuum to help minimize backgrounds.

A major advance in measurement of the neutron lifetime occurred in 1971 at the Risö reactor in Denmark (Christensen *et al.*, 1972); see Fig. 2. The decay volume was defined by two parallel paddles of plastic scintillator placed between the poles of a large (60 cm diameter) uniform electromagnet. The neutron beam passed between the paddles. When a neutron decayed within this region, the beta electron was transported in cyclic orbits along the magnetic field line to one of the paddles and detected. If the electron backscattered, it would be transported to the other paddle, so effectively 100% of the electron energy was detected for each decay. Therefore, the efficiency ε_p was very uniform in the decay region, as verified using a ^{198}Au calibration source. The systematic uncertainty from the beta spectrometer, which

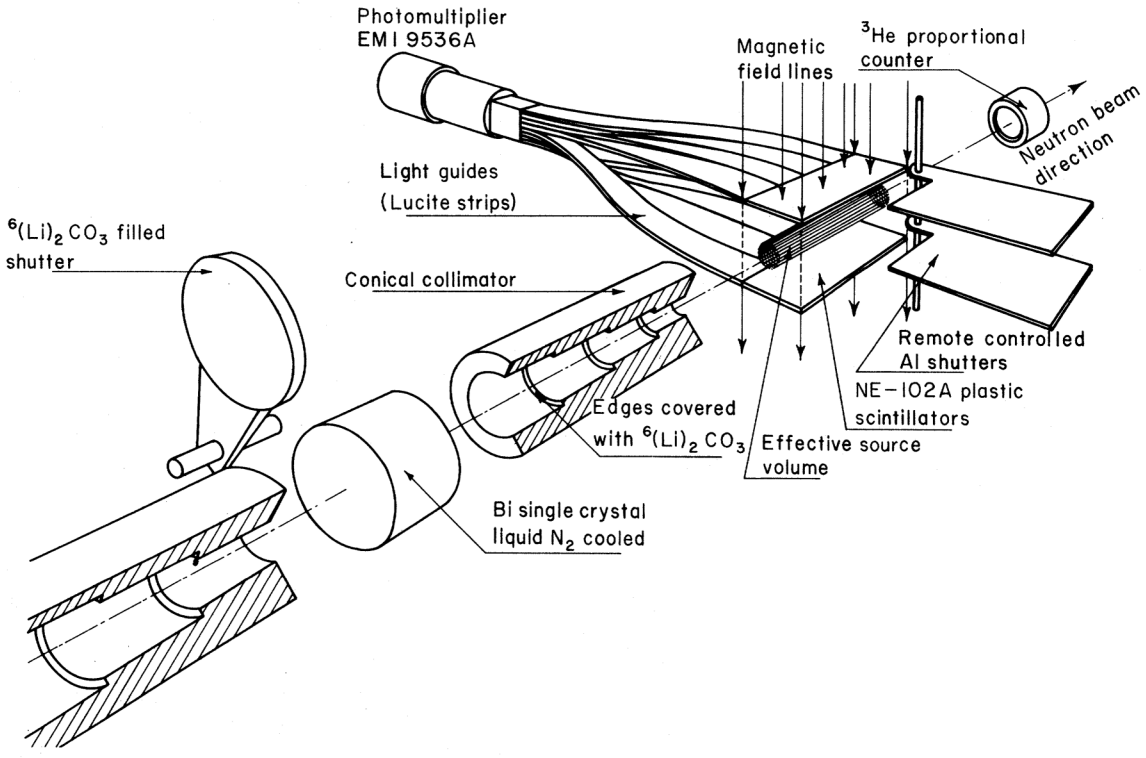

FIG. 2. The Risö neutron lifetime experiment . The decay volume is defined by the beta spectrometer, which consists of two parallel plastic scintilator paddles and a perpendicular magnetic field. From Christensen *et al.*, 1972.

included edge effects associated with magnetic field inhomogeneities, determination of the beta threshold, and scattering of partially trapped electrons by residual gas, totalled about 1%. Background gamma radiation from neutron capture was another significant source of error (0.8%). The neutron beam density was measured using a large area ^{3}He proportional counter with gold foil activation intercomparisons.

Beginning in the mid- 1950s, a program of neutron lifetime measurements led by P. E. Spivak was conducted for more than three decades at the Kurchatov Institute in Moscow (Spivak *et al.*, 1956, 1988; Sosnovsky *et al.*, 1959; Bondarenko *et al.*, 1978). The scheme of the experiment is shown in Fig. 3. Here neutron decay protons were extracted through a precision collimator that defined the beam volume, accelerated to 25 keV, and counted by a CH_4 proportional counter. The efficiency ε_p for counting protons from the decay region was very uniform and nearly 100%. This experiment had an advantage over previous work in that it employed a method to subtract neutron capture background radiation. Periodically, an 800 V potential was placed on a retarding grid in front of the proton detector. The maximum decay proton kinetic energy is 751 eV, so this prevented all protons from entering the detector but did not significantly change the background radiation spectrum and rate, so it could be easily measured and subtracted. The later runs of this experiment obtained the first 1% precision measurement of the neutron lifetime.

The PERKEO experiment, well known for a series of precision measurements of the beta asymmetry in neutron decay, also made a measurement of the neutron lifetime (Last *et al.*, 1988) at the Institut Laue-Langevin (ILL) reactor in France. In this experiment, the neutron beam was pulsed using a neutron-absorbing chopper drum such that the neutrons were bunched inside the decay volume within a 4π beta spectrometer. This provided two advantages: (1) there were no edge effects that make the decay volume uncertain; and (2) the number of decays was proportional to the total number of neutrons in the bunch, not the neutron density as in Eq. (10), and this could be measured using precision neutron counting techniques. The chief disadvantage was the large reduction in average neutron flux caused by the pulsing, which led to a relatively large counting statistics uncertainty (10 s). In another neutron lifetime experiment at the ILL (Kossakowski *et al.*, 1989), a pulsed neutron beam was passed through a helium time projection chamber enriched with ^{3}He. This detector could simultaneously count neutron decays by observing the beta electron tracks and measure neutron flux by observing protons from $^3\mathrm{He}(n, p)^3\mathrm{H}$ reactions in the decay volume. Both of these ILL experiments exploited novel and interesting methods, but neither was able to match the precision of the previous Risö and Kurchatov experiments.

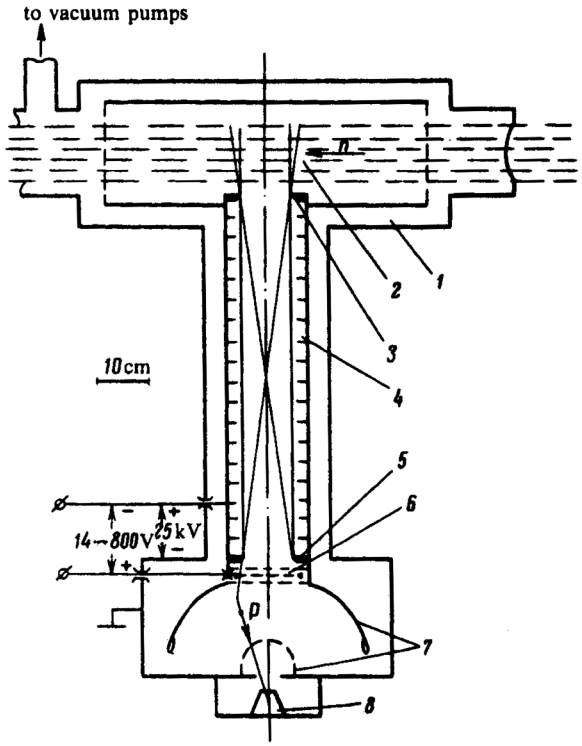

FIG. 3. The Kurchatov neutron lifetime experiment: (1) vacuum chamber, (2) neutron beam, (3,5) precision proton collimator, (6) retarding grid, (7) focusing electrode, (8) proton detector. From Bondarenko *et al.*, 1978.

A substantial improvement in accuracy of beam neutron lifetime measurements was obtained by the Sussex-ILL-NIST series of experiments (Byrne *et al.*, 1990, 1996; Dewey *et al.*, 2003; Nico *et al.*, 2005). A neutron beam was passed through a segmented quasi-Penning trap where the decay protons were trapped and counted; see Fig. 4. The trap consisted of 16 individual electrode segments. In the trapping state, the first three segments (the "door") were held at +800 v, a variable number (3–10) of trap segments were held at ground, and the following three segments (the "mirror") were at +800 v. When a neutron decayed inside the grounded region, the proton was trapped electrostatically in the axial direction and by the 4.6 T magnetic field in the transverse direction. After some period of time in the trapping state, typically 10 ms, the door electrodes were briefly lowered to the ground and a small ramped potential was applied to the trap electrodes. Trapped protons were flushed through the door and transported by a 9.5° bend in the magnetic field to a silicon surface barrier detector where they were electrostatically accelerated and counted (Byrne *et al.*, 1986). The trap was then returned to the trapping state to repeat. When the neutron beam exited the trap, it passed through a very thin deposit of ^{6}LiF on a silicon substrate. A set of four surface barrier detectors with very well characterized detection solid angle and efficiency counted the alpha and triton reaction products from $^6\mathrm{Li}(n, \alpha)^3\mathrm{H}$ reactions, thus giving an *in situ* measurement of the neutron density.

If the neutron decayed in the region of the grounded electrodes, the proton was trapped with 100% efficiency. In the region of the door and mirror electrodes, the trapping efficiency was less and depended on the proton momentum and the magnetic field shape and therefore was complicated to calculate. This is the reason for segmenting the trap. The trap length L can be written as $L = nl + L_{\mathrm{end}}$, where n is the number of grounded trap electrodes, l is the known electrode length, and L_{end} is the *effective* length of the door and mirror trapping regions, which is unknown but by symmetry was approximately constant for all trap lengths. The ratio of proton count rate to neutron count rate is measured as a function of trap length n. From Eq. (15),

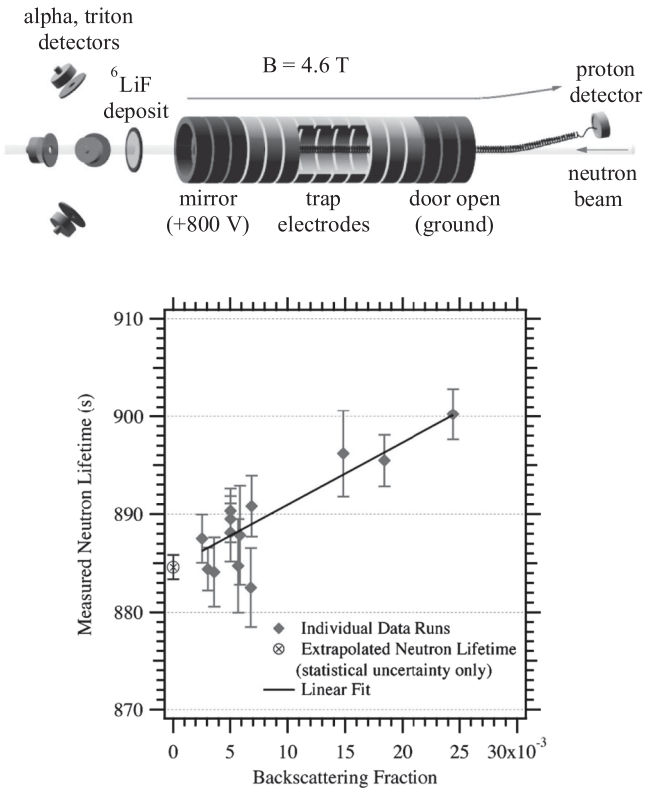

FIG. 4. The NIST beam neutron lifetime experiment. Top: The neutron beam passes through a 16 segment quasi-Penning trap where neutron decay protons are trapped and later extracted through a 9.5° bend in the magnetic field to the proton detector. The neutron beam density is measured by counting ^{6}Li$(n, \alpha)^3$H reactions in a thin ^{6}LiF deposit. Bottom: The measured neutron lifetime vs the calculated proton backscattering fraction for different detectors and acceleration voltages. A linear extrapolation to zero backscattering gives the result. From Nico *et al.*, 2005.

$$\frac{R_p}{R_n} = \tau_n^{-1}\left(\frac{\varepsilon_p}{\rho_{\text{foil}}\sigma_{\text{th}}v_{\text{th}}\varepsilon_n}\right)(nl + L_{\text{end}}). \tag{16}$$

From this, R_p/R_n vs n is fit to a straight line and τ_n is determined from the slope and therefore is independent of L_{end}. The proton counting efficiency ε_p was high ($> 97\%$) but less than 1 mainly due to Rutherford scattering of the proton from the dead surface layer of the silicon detector. This could be calculated by Monte Carlo methods, but it was complicated by the fact that the acceleration field of the detector will redirect some, but not all, scattered protons back into the detector. To account for this, several surface barrier detectors with different dead layer thicknesses and a variety of acceleration potentials were used. The measured neutron lifetimes were extrapolated to zero backscattering to obtain the neutron lifetime result of $\tau_n = 886.3 \pm 3.4$ s (see Fig. 4, bottom).

The largest systematic uncertainties in the NIST experiment (Nico *et al.*, 2005) came from neutron counting. The areal densities of the deposits ρ_{foil} were determined by optical, mechanical profilometry, and mass spectroscopic methods in measurements performed over several years (Pauwels *et al.*, 1995; Scott *et al.*, 1995). The uncertainty in the ^{6}Li thermal neutron absorption cross section σ^{th} was also significant. Probably the best avenue for improving the precision of future beam neutron lifetime experiments will be via improved absolute neutron counting methods (see Sec. IV).

B. Bottle method

In the bottle method, ultracold neutrons are stored in a material bottle for a time comparable to the neutron lifetime and the surviving neutrons that did not decay are counted. Ultracold neutrons [Golub, Richardson, and Lamoreaux (1991) give a thorough introduction] are neutrons whose kinetic energy is less than about 100 neV (temperature less than 1 mK, de Broglie wavelength greater than 100 nm). At this energy, a neutron can be trapped and stored for long periods. The Fermi effective potential, which originates from coherent nuclear scattering, of many materials is in the range 100–300 neV, so ultracold neutrons can be totally reflected from a suitably prepared surface and stored in a material bottle. The $\mu \cdot B$ potential of the neutron is 60 neV/T, so ultracold neutrons can be trapped in a strong inhomogeneous laboratory magnetic field. The mgh potential of the neutron at the Earth's surface is 100 neV/m, so ultracold neutrons can be confined vertically by the Earth's gravity.

The basic idea of this method is simple. Ultracold neutrons are produced and admitted into a bottle whose walls have high effective potential and therefore are totally reflecting for neutrons. If the bottle is sufficiently tall, then the neutrons are vertically trapped by gravity. Following some storage time Δt the surviving neutrons are extracted and counted. Two (or more) different storage times are used: Δt_1 and Δt_2, with the longer time comparable to the neutron lifetime. If the only loss mechanism is due to neutron beta decay, the ratio of neutron count rates for the two storage times N_1/N_2 gives the neutron lifetime

$$\tau = \frac{\Delta t_2 - \Delta t_1}{\ln(N_1/N_2)}. \tag{17}$$

In practice, there are always competing loss mechanisms that must be accounted and/or corrected for. In particular, total reflection of neutrons predicted by the bulk effective potential of the wall material is not fully achieved because impurities at the surfaces cause inelastic upscattering and absorption. The temperature of the bottle wall is always much higher than the 1 mK effective temperature of the stored ultracold neutrons, so inelastic scattering leaves the neutron with a kinetic energy higher than the Fermi effective potential of the wall material and it will quickly escape. Hydrogen has a large neutron scattering cross section (82 barns) and it is ubiquitous on solid surfaces; it is often the chief culprit in wall losses of ultracold neutrons. In the reflection process, the evanescent wave penetrates the wall surface, so neutron capture in the wall is also possible. The probability, per bounce, of inelastic scattering or capture is relatively small ($\approx 10^{-5}$) so the neutron storage time associated with wall losses tends to be hundreds of seconds, comparable to the beta decay lifetime. In general, there is a twofold strategy to mitigate wall losses: (1) carefully prepare the surfaces by polishing, etching, or coating to minimize the loss probability per reflection, and (2) vary the rate of reflections inside the bottle in a systematic

way so that the neutron storage time measurements can be extrapolated to zero wall losses. The measured neutron storage lifetime τ in Eq. (17) can be written as

$$\frac{1}{\tau} = \frac{1}{\tau_n} + \frac{1}{\tau_{\text{inel}}} + \frac{1}{\tau_{\text{cap}}} + \frac{1}{\tau_{\text{other}}}, \tag{18}$$

where τ_n is the beta decay lifetime, τ_{inel} is loss due to inelastic scattering at the walls, τ_{cap} is neutron capture at the walls, and τ_{other} accounts for any other loss mechanisms such as residual gas interactions or gaps, seams, or holes in the bottle that allow ultracold neutrons to escape. It is important to note that Eq. (18) assumes that all loss mechanisms proceed at a constant rate. As we discuss, this is not strictly true in practice as loss mechanisms may be dependent on the neutron velocity spectrum, which changes during the neutron storage time.

Ultracold neutron bottle experiments have two important advantages: (1) The neutron count rates N_1 and N_2 in Eq. (17) are relative measurements and the detector efficiencies cancel to first order, unlike beam measurements which require absolute counting of both neutrons and their decay products; and (2) radiation backgrounds are much lower in an ultracold neutron experiment, which need not be in the environment of an intense neutron beam.

The prospect of measuring the neutron lifetime using ultracold neutrons stored in a bottle was a strong motivating factor in the development of the first ultracold neutron sources at research reactors in the 1970s. The earliest bottle neutron lifetime experiments were performed at the SM-2 reactor in Dimitrovgrad, Russia, which obtained ultracold neutrons by vertical extraction from a cooled zirconium hydride moderator. A simple aluminum bottle with ports and flapper valves at the bottom to admit and extract neutrons was used (Kosvintsev *et al.*, 1980). A set of 30 removable aluminum fins were used to vary the internal surface area and hence the reflection rate. The measured lifetimes were fit to a calculated function for the wall loss rate, with the loss probability per reflection assumed to be constant and extracted from the fit. This assumption is not precisely valid because (1) the loss probability depends on the neutron energy spectrum which changes with time as neutrons are lost to the walls, and (2) the different surfaces were not identical and could have different loss probabilities. However, the experiment was more limited by poor counting statistics; it gave the result $\tau_n = 875 \pm 95$ s. A second, similar experiment used a larger aluminum bottle whose walls were etched, cleaned, and cooled to 80 K for the measurements (Kosvintsev, Morozov, and Terekhov, 1986). With these improvements, the observed wall loss rates were reduced by about 40%, although the main improvement was in counting statistics. The result was $\tau_n = 903 \pm 13$ s. This experiment was repeated with a thin layer of solid D_2O at 80 K covering the aluminum walls, which made another factor of 3.5 reduction in wall losses, presumably because the D_2O prevented neutron scattering and absorption on hydrogen at the surfaces. It gave $\tau_n = 893 \pm 20$ s (Morozov, 1989). All of these results were consistent with previous beam lifetime experiments.

The ultracold neutron turbine at the ILL, Grenoble (Steyerl *et al.*, 1986) provided a large gain in usable ultracold neutron densities. A mechanical turbine cannot increase the phase space density of ultracold neutrons beyond that which is produced in the cold source, but it enables neutrons to be vertically extracted at somewhat higher energy, which is more efficient, and converted down to ultracold energy at the turbine which is close to the experimental area. This, coupled with the fact that the ILL had the world's most intense cold neutron source, made a dramatic improvement; densities of up to 40 ultracold neutrons per cm^2 were measured in bottles fed by this source, more than 1000 times higher than the SM-2 source. The first ultracold neutron lifetime experiment to run at the ILL source was the MAMBO experiment, led by Walter Mampe (Mampe *et al.*, 1989).

MAMBO used a novel fluid-wall bottle based on Fomblin oil, first developed by J. C. Bates at the Risley reactor in England (Bates, 1983). Fomblin Y Vac18/8 is a viscous fluorinated polyether, well known for its use in diffusion vacuum pumps. It contains no hydrogen, has very low vapor pressure, and it forms a stable, smooth, renewable surface on glass. It has an effective Fermi potential of 106.5 neV so it is suitable for storing ultracold neutrons. A sketch of the MAMBO apparatus is shown in Fig. 5 (top). The walls were glass and covered by a thin coat of Fomblin that could be renewed *in situ* by means of a spray nozzle. The rear wall could be moved by a piston in order to vary the surface to volume ratio of the bottle. If the trapped neutrons are monoenergetic and their trajectories are isotropic, the wall loss rate can be calculated by kinetic theory

$$\frac{1}{\tau_{\text{wall}}} = \frac{\mu v}{\lambda}, \tag{19}$$

with μ the loss probability per bounce, v the neutron velocity, and λ the mean free path given by $\lambda = 4V/S$, where V/S is the volume to surface ratio in the bottle. The movable wall contained sinusoidal corrugations to randomize the trajectories of reflected neutrons, making them approximately isotropic. A problem arises from the fact that the neutrons are not monoenergetic and μ is velocity dependent, so as in the SM-2 experiments τ_{wall} varies in a complicated way with time. MAMBO employed a clever technique to circumvent this problem. The initial neutron velocity spectrum immediately after filling the bottle will be the same for all measurements. If the storage times Δt_1 and Δt_2 are both chosen to be proportional to V/S, then for two sets of measurements with different V/S the average number of wall reflections during Δt_1 and Δt_2 will be the same for both sets. Furthermore, the neutron velocity spectrum will evolve in the same way for both sets, each at a rate scaled by a time factor proportional to V/S. Therefore, both the initial and final neutron spectra are the same, in the isotropic approximation, for all Δt_1 measurements as V/S is varied; the same is true for all Δt_2 measurements. Because each set experienced the same number of wall reflections, the data can be linearly extrapolated to a zero wall reflection rate, as shown in Fig. 5 (bottom). MAMBO set a new standard of precision in neutron lifetime measurements, obtaining $\tau_n = 887.6 \pm 3.0$ s. A second version of the experiment, MAMBO II, was built. The main improvement was the addition of a neutron prestorage volume that is filled first. It has a moveable absorbing roof that removes the most energetic neutrons, thus ''cleaning'' the

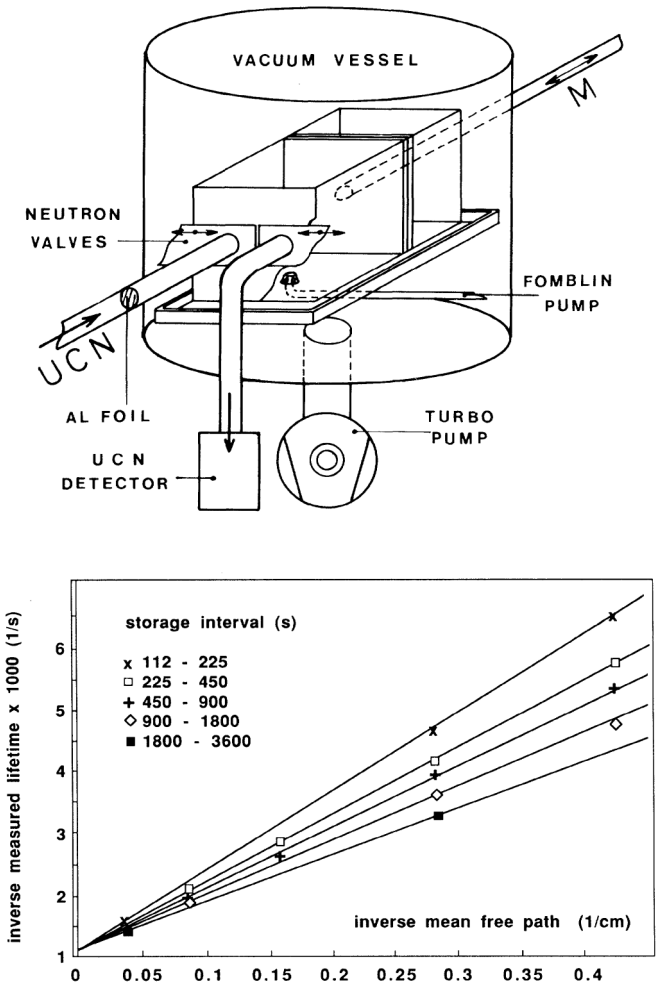

FIG. 5. Top: The ultracold neutron bottle used in the MAMBO experiment. Neutrons are admitted through the straight tube in front during filling; they exit through the bent tube during the counting phase. The rear wall can be moved using the rod M to change the volume to surface ratio of the bottle, and hence the wall loss rate. Bottom: Data from the MAMBO experiment; the measured inverse bottle lifetimes as a function of bottle inverse mean free path, for different storage intervals corresponding to different positions of the rear wall. The data extrapolate to a single point at zero inverse mean free path (infinite volume to surface ratio), giving the neutron beta decay lifetime. From Mampe *et al.*, 1993.

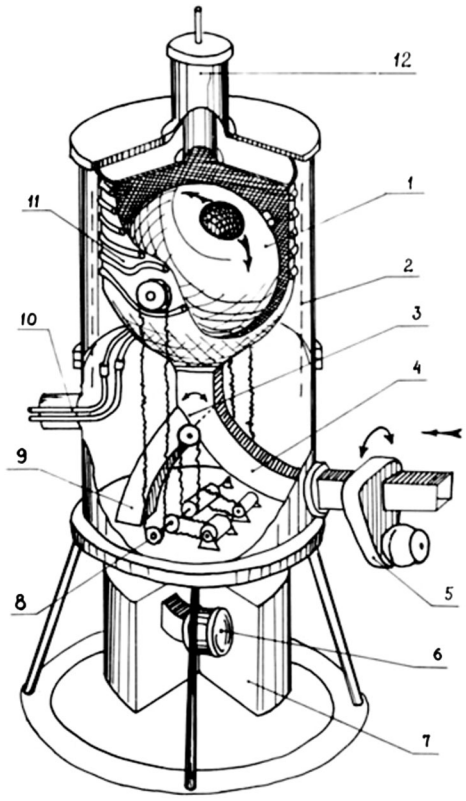

FIG. 6. The original Gravitrap neutron lifetime experiment: (1) ultracold neutron storage bottle, (2) liquid nitrogen screen, (3) neutron distribution valve, (4) neutron guide, (5) neutron filling valve, (6) ^{3}He neutron counter, (7) radiation shielding, (8) rotational control linkage, (9) neutron guide, (10) cryogen lines, (11) cryostat, (12) frozen oxygen system. From Kharitonov *et al.*, 1989.

neutron spectrum prior to admission into the main storage volume. A preliminary result was published in 2000 (Pichlmaier *et al.*, 2000), but it was later retracted (Pichlmaier *et al.*, 2001) citing systematic problems in determining the bottle volume. MAMBO II ran again and recently published a new result: $\tau_n = 880.7 \pm 1.3_{\text{stat}} \pm 1.2_{\text{sys}}$ s (Pichlmaier *et al.*, 2010).

An intense ultracold neutron source was developed at the Petersburg Nuclear Physics Institute (PNPI, formerly the Leningrad Nuclear Physics Institute) in Gatchina, Russia, using vertical extraction from a liquid hydrogen moderator placed near the center of the reactor core. The strength of this source profited from the very high thermal neutron flux at this location; usable ultracold neutron density of 16 per cm^2 was achieved. The first (and only) neutron lifetime experiment to use this source was the Gravitrap (Kharitonov *et al.*, 1989).

The neutron bottle was a 75 cm diameter spherical container coated with beryllium and cooled to below 15 K to minimize wall losses due to scattering; see Fig. 6. The sphere had a small hole at its top and could be rotated about a horizontal axis to vary the height of the hole. For each measurement, after the initial filling, the sphere was rotated by a known angle to allow some neutrons, those whose mechanical energy exceeded the gravitational potential at the height of the hole, to escape. The sphere was then rotated back to return the hole to the top and the neutron storage lifetime was measured. Each rotation angle corresponded to a different starting neutron velocity spectrum and, because the wall loss rate is proportional to neutron velocity, resulted in a different wall loss rate. This procedure also helped to "clean" the trap of marginally trapped neutrons (see Sec. III.D). The relationship between the rotation angle and the wall loss rate was calculated from the assumed initial spectrum after filling and a simple model of neutron trajectories in the sphere. The storage time was measured as a function of an average wall loss rate and extrapolated to zero wall losses. The experiment was repeated with a solid oxygen coating, which exhibited a significantly lower neutron loss probability per reflection on the bottle wall. Both versions of the experiment gave consistent results and yielded a combined neutron lifetime of $\tau_n = 888.4 \pm 3.3$ s (Alfimenkov *et al.*, 1992).

The first generation ultracold neutron bottle lifetime experiments described above share some general features. The measured neutron storage lifetimes were significantly shorter than the beta decay lifetime and as a result large corrections were needed. In each case, the correction was done by systematically varying some feature of the experiment, namely, the volume to surface ratio of the bottle or the starting neutron velocity spectrum. A theoretical model, based on assumptions that were not exact but valid to some approximation, was used to calculate the wall loss rate as a function of the feature that was varied. With this calculation, the measured storage times were linearly extrapolated to a zero wall loss rate, yielding the beta decay lifetime. The dependence of these experiments on such calculations was a weakness. They cannot be done exactly, and any discrepancies between the theoretical models and the actual conditions of the experiments that were unaccounted for could have led to errors in the extrapolation and hence the final results. The main goal of a second generation of experiments was to address and reduce this problem.

One of these experiments employed a new feature: an array of thermal neutron detectors was placed around the outside of the bottle to detect, with some efficiency ε^{th}, neutrons that escaped the bottle due to inelastic scattering at the wall. This gave an independent measure of wall losses. In the hypothetical case of a monoenergetic population of trapped neutrons, this feature leads to a particularly elegant resolution of the neutron lifetime. The number of neutrons J counted by the thermal neutron detectors during a storage time $\Delta t = t_f - t_i$ is given by

$$J = \varepsilon_{\text{th}}(n_f - n_i)\left(\frac{\tau}{\tau_{\text{inel}}}\right), \tag{20}$$

where n_i and n_f are the initial and final number of trapped neutrons and τ is the measured storage lifetime as in Eq. (18). As usual, two storage times Δt_1 (short) and Δt_2 (long) are used, but now four quantities are measured: J_1, J_2, the number of scattered thermal neutrons counted during the two storage times; and N_1, N_2, the number of ultracold neutrons counted when the bottle is opened. With $\Delta J = J_2 - J_1$ and $\Delta N = N_2 - N_1$, we have

$$\Delta J = \Delta N\left(\frac{\varepsilon_{\text{th}}}{\varepsilon_{\text{UCN}}}\right)\left(\frac{\tau}{\tau_{\text{inel}}}\right). \tag{21}$$

Here ε_{UCN} is the efficiency for counting ultracold neutrons when the bottle is opened. The experiment is repeated using two different bottle geometries, A and B, having different volume to surface ratios and hence different wall loss rates. Assuming that the counting efficiencies and all other conditions are the same, one can determine the ratio ξ of the inelastic upscattering lifetimes from measured quantities:

$$\xi \equiv \frac{\tau^A_{\text{inel}}}{\tau^B_{\text{inel}}} = \frac{J^A \tau^B \Delta N^B}{J^B \tau^A \Delta N^A}. \tag{22}$$

The storage lifetimes τ^A and τ^B are found as usual from Eq. (17). Up to this point neutron capture at the walls has been neglected. If the neutrons are monoenergetic, the neutron capture rate, a smaller effect, should be exactly proportional to the inelastic scattering rate, as both are proportional to the rate of neutrons hitting the walls. Therefore, the neutron capture lifetime τ_{cap} can be absorbed into the efficiency factor ε_{th} and it cancels in Eq. (22). In light of this, one can write

$$\frac{1}{\tau^{A(B)}} = \frac{1}{\tau_n^{A(B)}} + \frac{1}{\tau_{\text{inel}}^{A(B)}}, \tag{23}$$

and combining Eqs. (22) and (23)

$$\tau_n = \frac{\xi - 1}{\xi/\tau^B - 1/\tau^A}. \tag{24}$$

The neutron beta decay lifetime can be found directly from measured quantities without the need for a linear extrapolation.

A number of difficulties arise relative to the ideal case just discussed. First, the neutrons are not monoenergetic. Even if monoenergetic neutrons were admitted into the bottle, the gravitational potential would cause their kinetic energies to be nonuniform, so neutrons striking the wall at different heights would have different incident velocities. Consequently, the loss rates are time dependent and depend on the initial neutron spectrum. Because neutron capture is velocity dependent ($1/v$ law), the proportionality factor between the capture and inelastic scattering rates is time dependent. The differences of geometry in measurements A and B cause the efficiencies ε^{th} and ε_{UCN} to be different, so they do not exactly cancel. Thus, a number of important corrections must be made when analyzing the experimental data.

The first experiment based on this method ran at the ILL in 1990. The ultracold neutron bottle was a vertical, stainless steel, Fomblin-coated cylinder. A concentric, internal thin cylinder, also Fomblin coated, provided a second geometry with lower volume to surface ratio. Outside the bottle an array of 28 ^{3}He proportional counters detected inelastically scattered neutrons. The result was $\tau_n = 882.6 \pm 2.7$ s (Mampe *et al.*, 1993). A second, refined version ran at the ILL eight years later. The apparatus is shown in Fig. 7 (top). This time, the cylindrical bottle was mounted horizontally and constructed of aluminum, but the construction was similar. The inner cylinder could be rotated *in situ* on its axis through a puddle of Fomblin at the bottom to refresh its liquid coating. The experiment was run at a variety of wall temperatures from -26 to $+20\,^\circ$C, in order to vary the overall inelastic scattering rates. The data are shown in Fig. 7 (bottom). The result of this experiment was $\tau_n = 885.4 \pm 0.9_{\text{stat}} \pm 0.4_{\text{sys}}$ s (Arzumanov *et al.*, 2000). This was the first neutron lifetime to report an overall uncertainty below 1 s. Recently they discovered two systematic corrections that were not properly accounted for in the original analysis. In an upcoming publication, this result will be corrected to a lower value (Bondarenko, 2011).

A second version of the Gravitrap experiment was developed by the PNPI group and run at the ILL. The chief improvement was the wall coating, a fluorinated polyether similar to Fomblin that can be evaporated onto a surface and forms a stable coating at cryogenic temperatures, in this case $-160\,^\circ$C. With this coating, the inelastic scattering probability per bounce was observed to be about 2×10^{-6}, a factor of 10 smaller than liquid Fomblin oil. Two nested neutron bottles were used, a narrow cylinder and a larger quasispherical bottle, both rotatable on a horizontal axis in the Gravitrap

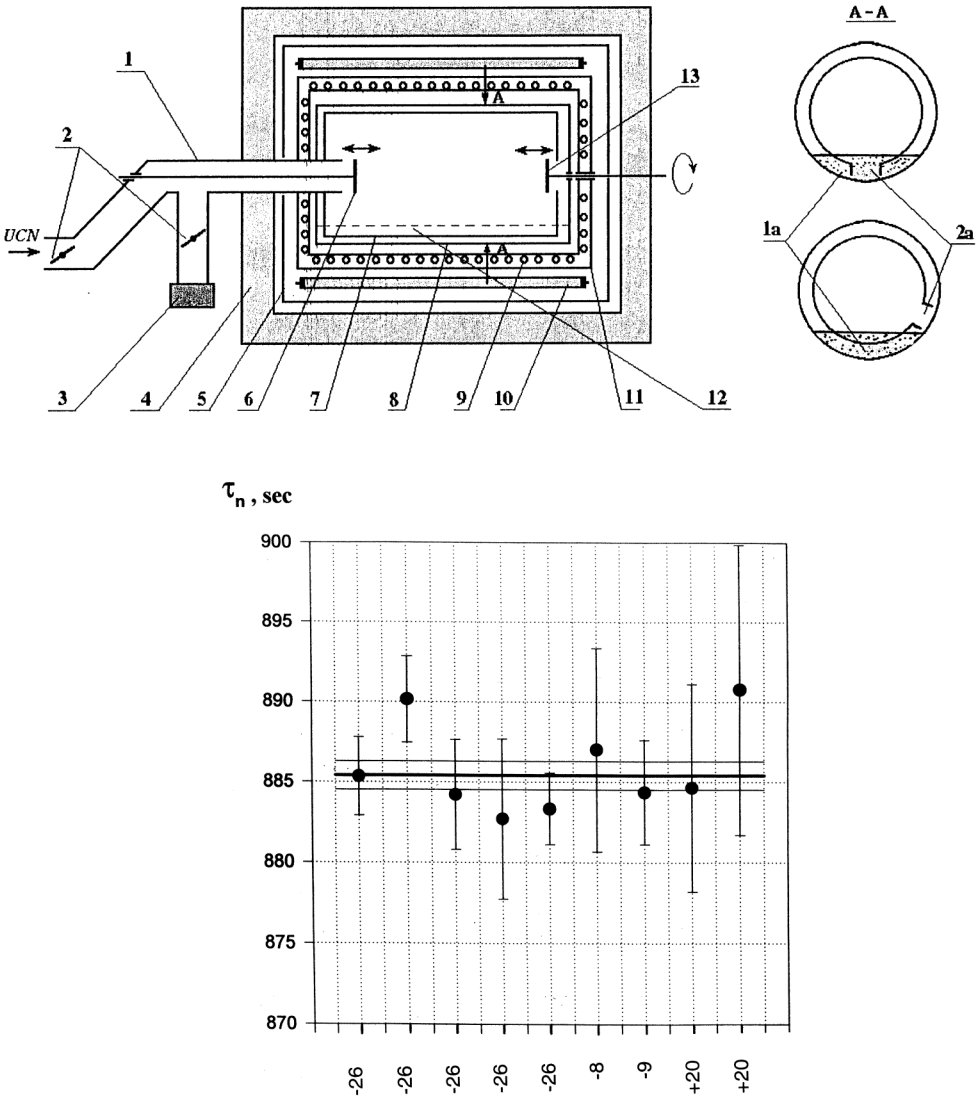

FIG. 7. Top: The ultracold neutron bottle in Arzumanov *et al.* (2000) used to measure the neutron lifetime. (1) neutron guide, (2) neutron shutters, (3) neutron detector, (4) polyethylene shielding, (5) cadmium neutron absorber, (6) neutron entrance shutter, (7) inner storage vessel, (8) outer storage vessel, (9) cooling coil, (10) ^{3}He proportional counters, (11) vacuum vessel, (12) Fomblin oil puddle, (13) entrance shutter for the annular vessel, (1a) Fomblin oil puddle, (2a) slit. Bottom: The extrapolated neutron lifetime for nine different runs. The bottle wall temperature (°C) is given on the bottom axis. Note that the collaboration will soon publish a correction to the analysis that will lower the neutron lifetime result. From Bondarenko, 2011.

scheme. In this way, both the neutron velocity spectrum and the volume to surface ratio could be varied to change the wall loss rate in the measurements. Finally, the counting statistics were much higher than in the previous experiment.

The data, showing the linear extrapolation to zero wall loss rate, are shown in Fig. 8. With this new cryogenic wall coating, the total extrapolation from the measured points to the final result is only 5–15 s, much smaller than in previous experiments, and is consistent with the lower wall loss rates. As a further test of the effectiveness and integrity of the coating, additional measurements were made using a titanium bottle coated with the cryogenic oil. Titanium is one of the few elements with a negative Fermi effective potential, so it cannot reflect ultracold neutrons at any incident angle. With this bottle, the neutron storage time was measured to be 869 s, just slightly shorter than in the main experiment where the bottles had a beryllium coating (a good neutron reflector) under the cryogenic oil. This demonstrated that the fraction of the bottle surface area that was not effectively coated with the oil was less than 10^{-6}. The neutron lifetime result from this experiment was $\tau_n = 878.5 \pm 0.7_{\text{stat}} \pm 0.3_{\text{sys}}$ s (Serebrov *et al.*, 2005, 2008). The combined quadrature uncertainty is 0.76 s. When this result was announced, it immediately created a ''neutron lifetime problem'' as it was about 7 s below the previous world average and disagreed with the only other subsecond neutron lifetime result by more than 5 sigma (see Sec. III.D).

C. Magnetic traps

The magnetic trap method is similar to the bottle method in that ultracold neutrons are stored for some period of time and then counted. The difference is that rather than relying on neutron reflection from a material wall, where neutron

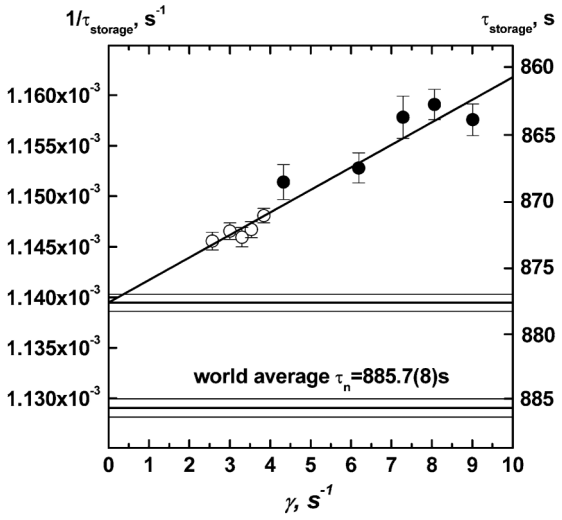

FIG. 8. Data from the Gravitrap II experiment showing the extrapolation of the neutron storage lifetimes to zero wall loss rate. Dark circles are from the smaller cylindrical bottle and open circles from the larger bottle. From Serebrov *et al.*, 2005.

upscattering and capture cause significant wall losses, neutrons are trapped by magnetic field gradients alone and therefore do not interact with matter during the storage interval. The magnetic dipole moment of the neutron is 60 neV/T, so in the field of a strong magnet the kinetic energy of a neutron in the ultracold regime can be less than the $\mu \cdot B$ potential energy. If a set of such neutrons are contained in a magnetic field minimum surrounded by a field of several tesla, the gradient will provide a force that returns one spin state toward the center (low field seekers) and pushes the other spin state out of the trap (high field seekers). Neutrons with 100% spin polarization can be trapped indefinitely until they beta decay.

In practice, there are two important loss mechanisms to contend with. The first is depolarization. If, in the rest frame of the neutron, the time rate of change of the magnetic field is comparable to or larger than its Larmor precession frequency, there is a significant probability for spin flip. The Larmor precession frequency is proportional to B, the magnitude of the magnetic field: $\nu_{\mathrm{Larmor}} = \gamma_n B$, where $\gamma = 1.83 \times 10^8\ \mathrm{s}^{-1}\,\mathrm{T}^{-1}$ is the gyromagnetic ratio of the neutron. A related phenomenon is the Majorana spin flip which can occur if the neutron passes through a zero magnetic field point, where the two spin states are degenerate. To minimize depolarization one must maintain the condition

$$\frac{dB}{dt} \ll \gamma_n B^2 \tag{25}$$

throughout the trap by avoiding large gradients and zero field points. If neutrons are slowly depolarized by either mechanism, some become high field seekers and are ejected from the trap. This loss can compete with beta decay. The second important loss mechanism is marginally trapped neutrons, neutrons whose mechanical energy exceeds the maximum trap potential but whose trajectories are such that their escape time is comparable to the neutron beta decay lifetime (see Sec III.D).

The first neutron lifetime experiment using a magnetic trap was the hexapole torus storage ring NESTOR at the ILL (Kugler, Paul, and Trinks, 1978; Anton *et al.*, 1989; Paul *et al.*, 1989). A simple hexapole torus magnet consists of six current loops of alternating direction arranged around a torus. Hexapole magnets have been widely used to store and focus neutral atomic beams. They are also suitable for neutrons; although the much smaller magnetic moment of the neutron means that a larger magnetic field is required. A true hexapole has a zero magnetic field point at its center. To avoid neutron depolarization, the NESTOR current loops were distorted to eliminate the zero points. For a neutron in a circular orbit inside the torus, the magnetic gradient force provides the centripetal force, so only neutrons within a certain velocity range have stable orbits. Neutron trajectories are actually not circles, but exhibit small betatron oscillations around a circular orbit. Therefore, NESTOR was not a true three-dimensional trap as the neutron axial kinetic energy was much larger than the magnetic potential energy. A stable trap is realized only insofar as the transverse and orbital kinetic energy modes are independent. This independence is only approximate; betatron oscillations can couple them and cause a gradual transfer of orbital energy to transverse energy, leading to a loss of neutrons. In this sense, the hexapole torus is similar to a charged particle storage ring. In NESTOR, betatron losses were reduced by limiting the momentum range of neutrons in the trap using beam scrapers. Neutrons were admitted to the trap and stored for time intervals ranging from 20 to 3500 s. The final result was limited by counting statistics: $\tau_n = 877 \pm 10$ s.

The NIST ultracold neutron lifetime experiment is the first to employ a true three-dimensional magnetic trap: an Ioffe-type trap consisting of a magnetic quadrupole to trap neutrons radially and a pair of solenoids to trap them axially. The trap is loaded with ultracold neutrons using the superthermal method in liquid helium (Golub and Pendlebury, 1975; Golub *et al.*, 1983; Doyle and Lamoreaux, 1994). In this method, a beam of cold neutrons interacts in a liquid helium target creating phonons and downscattering to near zero kinetic energy. The use of *in situ* downscattering eliminates the need to open the trap to fill with ultracold neutrons. The phonon-downscattering reaction is most efficient for the single phonon process, which conserves energy and momentum only at a particular neutron incident energy: 1.03 meV ($\lambda = 0.89$ nm, $v = 445$ m/s), so typically a monoenergetic neutron beam at this energy is used. Unlike mechanical ultracold neutron converters such as gravity or the ILL turbine source, the superthermal source can increase the neutron phase space density because the resulting net decrease in neutron entropy is compensated by the positive entropy of the phonons. So there is no fundamental limit to the ultracold neutron density that can be achieved. A simplified rendering of the apparatus is shown in Fig. 9. Cold neutrons at 1.03 MeV pass through a collimator and enter the trapping region, which contains a cell filled with liquid helium cooled to below 250 mK by a dilution refrigerator. Approximately 1% of these neutrons produce a phonon in the liquid and convert to ultracold neutrons below 94 neV, the height of the magnetic trap. Of those, one spin state is trapped and the other is quickly ejected and absorbed at the walls of the cell. The

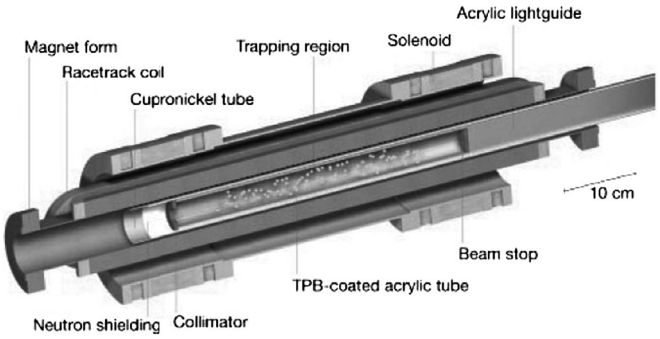

FIG. 9. The NIST ultracold neutron lifetime apparatus. Cold neutrons with energy 1.03 MeV enter from the left and are converted to ultracold neutrons in liquid helium, by the superthermal method, in the trapping region. One spin state is trapped there by the quadrupole racetrack coils and two solenoids and the other spin state is ejected. Trapped neutrons that beta decay ionize the liquid helium, producing scintillation light that is downshifted to visible light and transported by the acrylic light guide to a pair of photomultiplier tubes. From Huffman *et al.*, 2000.

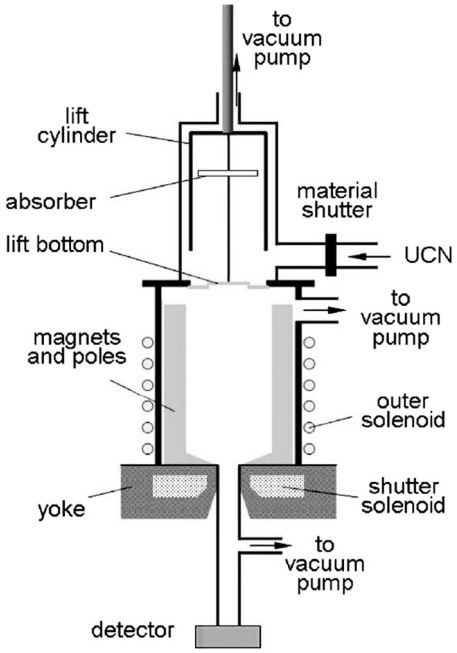

FIG. 10. The magnetic bottle apparatus of Ezhov *et al.* (2009). A cylindrical array of permanent magnets lining the bottle wall traps neutrons magnetically such that they cannot touch the material walls. The shutter solenoid forms the bottom of the trap, and gravity prevents trapped neutrons from escaping at the top. A moveable lift bottle is used to load the trap. The outer solenoid eliminates zero field points that can cause neutron depolarization and leakage.

trapped neutrons cannot reach the walls or any other materials (other than liquid helium) so there are no wall losses in this experiment. There are, however, two neutron loss mechanisms associated with the liquid helium. One is inelastic scattering from phonons. This is reduced to minimal levels by cooling the liquid below 250 mK so that the population of sufficiently energetic phonons is very small. The other is absorption by ^{3}He, which is present at the 10^{-6} level in natural helium and has a very large (5333 barns) thermal neutron absorption cross section. To eliminate this, isotopically pure ^{4}He purified using the heat flush technique (McClintock, 1978), is used. The experiment has successfully demonstrated trapping of ultracold neutrons and detection of their beta decay (Huffman *et al.*, 2000; Brome *et al.*, 2001) and made a preliminary measurement of the neutron lifetime, limited by statistics: $\tau_n = 831(+58, -51)$ s (Yang, 2006). A larger version of the apparatus with a stronger magnet has been built (O'Shaughnessy *et al.*, 2009) and is currently running at NIST.

A rather different magnetic bottle neutron lifetime experiment has recently run at the ILL (Ezhov *et al.*, 2009). A cylindrical, 20-pole permanent magnet array forms the main part of the bottle (see Fig. 10). A solenoid is used to close the bottom. The top of the bottle is open and neutrons are trapped at the top by gravity. A mechanical lift, a separate moveable cylinder lined with Fomblin oil, is used to fill the trap. With the lift in the upper position, a shutter is opened to admit ultracold neutrons. During this time, the bottom solenoid remains off so that neutrons that leak into the main bottle quickly escape. The shutter is then closed and the lift is moved down slowly at an adiabatic speed of 5 cm/s. When the lift is fully inserted into the lower bottle, the solenoid is energized and, with the lift bottom open, the lift is raised leaving the neutrons in the magnetic bottle where one spin state is trapped by the magnetic field and gravity. It is important to note that trapped neutrons cannot reach the material surfaces of the bottle, and therefore do not suffer wall losses. The bottle is coated with Fomblin so that the high field seeking neutrons reflect from the walls and escape through the hole in the bottom where they are counted. This serves as an *in situ* monitor of marginally trapped neutrons leaking from the bottle. After a variable storage period of up to 2200 s, the solenoid is switched off and the trapped neutrons exit through the bottom and are counted. An additional outer solenoid surrounding the bottle is used to add a bias magnetic field that eliminates zero points and therefore reduces losses due to depolarization. The current in the outer solenoid can be reversed to "force" depolarization, causing a deliberate leakage of neutrons as a systematic check. A comparison of storage times with and without forced depolarization also provides a measurement of the efficiency of the neutron detector. A result from this experiment has been presented at conferences: $\tau_n = 878.2 \pm 1.9$ s (Ezhov, 2009) but as of this writing has not yet been published. The quoted error is statistical only.

D. Discussion of experimental results

Figure 11 and Table I contain 21 measurements of the neutron lifetime covering a period of 60 years. Early experiments obtained lifetimes in excess of 1000 s and had large uncertainties. The first precision experiments were published in the 1970s (Christensen *et al.*, 1972; Bondarenko *et al.*, 1978). For historical purposes, we also mention here the first result from the Sussex-ILL-NIST beam experiments: 937 ± 18 s (Byrne *et al.*, 1980), which was later retracted due to an error in the calibration of the neutron counting foils (Byrne *et al.*, 1990). These three measurements, all using the beam

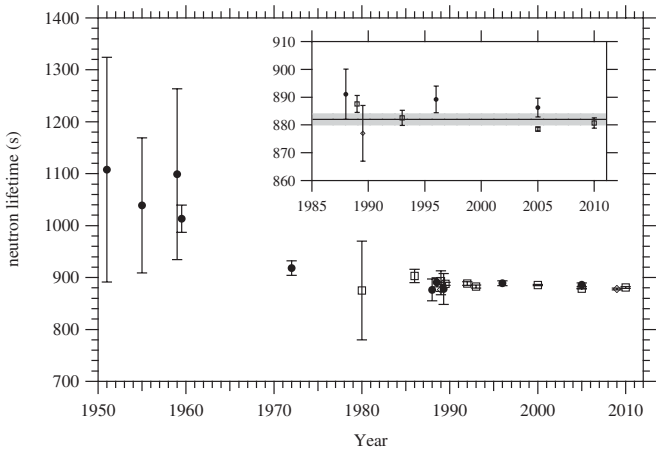

FIG. 11. A summary of neutron lifetime measurements. Solid circles are beam experiments, open squares are bottle experiments, and diamonds are magnetic trap experiments. The inset shows the eight experiments included in our global averages. See Table I for references.

method, span a range of 881–937 s and are in poor agreement relative to their quoted uncertainties. They comprise the first ''neutron lifetime problem.'' This troubling disagreement motivated concerted programs at neutron sources around the world to produce more measurements using novel techniques such as the bottle method and magnetic traps. The fruit of this substantial effort can be seen in Fig. 11 as the cluster of 10 precision results in the period of 1986–1993. These are all in good agreement and they confirmed the lowest of the three disagreeing numbers (Bondarenko *et al.*, 1978). Thus, the problem seemed to be solved. Subsequent experiments through 2004 gave additional confirmation. The 2004 Review of Particle Properties (Eidelman *et al.*, 2004) used a weighted mean of the seven most recent measurements to date with quoted errors less than 10 s to obtain a recommended world average for the neutron lifetime of $\tau_n = 885.7 \pm 0.8$ s, with a chi squared of 3.5 for 6 degrees of freedom, a very comfortable agreement.

In 2004, Serebrov and collaborators from PNPI announced the result of the Gravitrap II experiment: $\tau_n = 878.5 \pm 0.76$ s, in serious conflict with the existing world average. Not unexpectedly, this result was treated with some skepticism at first, but since then it has been widely discussed and is now taken seriously by scientists in this field. The PNPI group is very experienced and the experiment was done carefully. Because the new cryogenic oil coating gave a much smaller probability for inelastic scattering at the wall, the measured neutron storage times were much longer, and the extrapolation time to the neutron lifetime much shorter than previous bottle experiments. It is often claimed, as a general argument, that experiments with smaller extrapolations from their measurements to the physics result tend to be more reliable. As a guiding principle that is no doubt true, but of course it does not prove the validity of a particular experiment. We are now faced with a second neutron lifetime problem. The new result from the magnetic bottle experiment (Ezhov, 2009) adds support to the lower number. The Particle Data Group chose not to include these new results in their most recent

TABLE I. A summary of neutron lifetime measurements. When applicable, statistical and systematic errors have been added in quadrature. Asterisks indicate the 8 experiments included in our global averages.

Reference	Neutron lifetime (s)	Uncertainty (s)
Beam Experiments		
Robson, 1951	1110	220
Spivak *et al.*, 1956	1040	130
D'Angelo, 1959	1100	160
Sosnovsky *et al.*, 1959	1013	26
Christensen *et al.*, 1972	918	14
Last *et al.*, 1988	876	21
Spivak, 1988*	891	9
Kossakowski *et al.*, 1989	878	30
Byrne *et al.*, 1996*	889.2	4.8
Nico *et al.*, 2005*	886.3	3.4
Bottle Experiments		
Kosvintsev *et al.*, 1980	875	95
Kosvintsev, Morozov, and Terekhov, 1986	903	13
Morozov, 1989	893	20
Mampe *et al.*, 1989*	887.6	3.0
Alfimenkov *et al.*, 1992	888.4	3.3
Mampe *et al.*, 1993*	882.6	2.7
Arzumanov *et al.*, 2000	885.4	0.98
Serebrov *et al.*, 2005*	878.5	0.76
Pichlmaier *et al.*, 2010*	880.7	1.8
Magnetic Trap Experiments		
Paul *et al.*, 1989*	877	10
Ezhov *et al.*, 2009	878.2	1.9

evaluated average, nor to expand the uncertainty in the usual prescription. Instead, they maintained their 2004 recommended value, noting (Nakamura, 2010) the following:

> The most recent result, that of Serebrov *et al.* (2005, 2008), is so far from other results that it makes no sense to include it in the average. It is up to workers in this field to resolve this issue. Until this major disagreement is understood, our present average of 885.7 ± 0.8 s must be suspect.

In a recent and significant development, Arzumanov *et al.* stated that they reanalyzed the experiment and found two important new corrections: (1) a previous correction for the geometry dependence of the thermal neutron detector efficiencies had the wrong sign, and (2) an ultracold neutron heating effect that had not been previously accounted for. The combination of these is expected to lower their neutron lifetime result significantly (Bondarenko, 2011), bringing it much closer to the Gravitrap II number. We note, however, that this will not completely solve the problem. If we take the set of experiments used for the Particle Data Group 2004 average and omit the Arzumanov *et al.* (2000) number, the average becomes 886.4 ± 1.4 s, still 5 standard deviations above the Gravitrap II result. The consensus in the field is that

the problem is probably due to unaccounted systematic effects in several of the experiments. It is useful to consider some possibilities.

The problem of marginally trapped neutrons has had a lot of attention in recent experiments, especially for magnetic traps where they tend to be a large effect. They are also present in the bottle experiments. Marginally trapped neutrons are neutrons whose mechanical energy exceeds the maximum trap potential but whose trajectories are such that their escape time is comparable to the neutron beta decay lifetime. The term ''marginally trapped'' is a misnomer because technically these neutrons are not trapped. Perhaps a better name would be ''persistent untrapped neutrons,'' but we use the prevalent terminology here. Ultracold neutrons inside a bottle with kinetic energy less than the effective Fermi potential (V_{eff}) of its walls cannot escape other than by inelastic scattering, absorption, or beta decay. Neutrons with kinetic energy $V_{\mathrm{eff}} + \epsilon$, slightly in excess of the bottle potential, can otherwise escape, but only if they strike the wall at an angle $\phi < \sqrt{\epsilon/V_{\mathrm{eff}}}$ from normal incidence. For ideal, specular, symmetric bottles such as spheres, rectangular boxes, and cylinders, clearly there exist stable trajectories for marginally trapped neutrons that will never satisfy that criterion, and others that will satisfy it only after many reflections. Irregular surfaces, which were a design feature of some bottle experiments, produce randomized reflections that tend to reduce but not eliminate marginally trapped neutrons. Most experiments employed some means of cleaning the neutron spectrum, i.e., removing the most energetic part by lowering an absorbing material (Arzumanov *et al.*, 2000; Ezhov *et al.*, 2009; Pichlmaier *et al.*, 2010), rotating the bottle to expel energetic neutrons (Alfimenkov *et al.*, 1992; Serebrov *et al.*, 2005), or lowering the magnetic field for a brief time before the storage interval (Yang, 2006). Cleaning mitigates but does not eliminate the problem and has the unfortunate side effect of reducing counting statistics. In general, it is very difficult to reliably calculate the population and time evolution of marginally trapped neutrons as they depend very sensitively on details of the bottle and trap geometry and how it is filled. Because they tend to occupy small and particular regions of phase space, the loss of these neutrons does not vary in an easily predictable way as a function of volume to surface ratio or kinetic energy. It seems likely that systematic errors due to marginally trapped neutrons have not been completely understood in the bottle and magnetic trap experiments.

The beam neutron lifetime experiments are immune from the problem of marginally trapped neutrons. The chief systematic difficulty in these experiments has been the absolute calibration of the neutron counting system. In the Sussex-ILL-NIST experiments, this was done by an intricate chain of measurements involving destructive and nondestructive characterizations of absorbing foils to obtain their areal density, precision measurements of particle detector solid angles, and the neutron capture cross section of the foil, the largest source of systematic error. The $^6\mathrm{Li}(n, \gamma)$ cross section used in the most recent experiment (Nico *et al.*, 2005) has not been directly measured to sufficient precision; rather it came from an hierarchical global evaluation of many related neutron reactions (Carlson *et al.*, 1993). As the standard cross section shifts with new evaluations, the neutron lifetime result will vary with it, not a satisfactory situation. An independent, absolute calibration of the neutron counter used in the experiment will be essential to improve its precision and reliability.

The weighted average of all 21 measurements listed in Table I is 881.6 ± 0.5 s, with a chi squared of 90.0 for 20 degrees of freedom. If we exclude experiments prior to 1975, it improves somewhat: 881.5 ± 0.5 s, with a chi squared of 52.3 for 15 degrees of freedom. The excess chi squared is not associated with only one or two outliers, it is broadly distributed, indicative of a general problem of underestimating systematic errors. We believe that marginally trapped neutrons in the bottle and magnetic trap experiments and absolute neutron counting calibration in the beam experiments are the most likely suspects, but of course other effects may be important.

We can explore some options for a more refined average and evaluation of the neutron lifetime based on this body of experiments. For this purpose, we include only experiments with a net quoted uncertainty of 10 s or less. This is a natural dividing line between the more recent precision experiments and the older experiments with large uncertainties, some of which are probably not reliable from a modern view. We exclude the Gravitrap I result (Alfimenkov *et al.*, 1992), because the Gravitrap II result is almost 3 standard deviations lower, and the group responsible has stated that they believe the latter is more reliable (Serebrov and Fomin, 2010). We exclude, for now, the magnetic trap of Ezhov *et al.* as only a preliminary result has been presented at conferences; there is as of yet no published account of the data analysis, and the result does not include an estimated systematic error. We also exclude the Arzumanov *et al.* result pending an upcoming publication of their revised number. This leaves eight experiments to be included, indicated with asterisks in Table I. The weighted average of these is 880.0 ± 0.64 s, with a chi squared of 20.1. As discussed, we believe the disagreement is due to a general underestimation of systematic errors by a number of experiments, but we cannot say which and by what amounts. A typical procedure in cases such as this is to expand the error in the weighted average by a factor of $\sqrt{\chi^2/\nu}$, where ν is the number of degrees of freedom in the fit (Cohen and Taylor, 1973; Nakamura, 2010). With this we obtain $\tau_n = 880.0 \pm 1.1$ s. A downside of this method is that the few most precise measurements dominate the average value; the others carry relatively little weight.

An alternate approach is to assume that the statistical errors on all experiments have been correctly evaluated but that each has neglected some source of systematic error and these errors were equal in size. We add this error in quadrature to the quoted errors. While this assumption is not quite true, and it neglects the possibility of correlated errors, it has the effect of reducing the bias toward experiments with small statistical error. Since the discrepancy between the values is almost certainly systematic, this may provide a more fair estimate. An additional error of 2.8 s added in quadrature to each of the eight selected experiments produces an average with a reduced chi square of 1.0, and $\tau_n = 882.8 \pm 1.5$ s.

Finally, we consider a more radical step in equalizing the weight of the experiments. We effectively assume that all

TABLE II. Six different weighted global averages of the neutron lifetime.

Method	Average (s)	Error (s)	χ^2	ν
All 21 experiments	881.6	0.5	90.0	20
All experiments after 1975	881.5	0.5	52.3	15
8 "precision" experiments	880.0	0.64	20.1	7
With expanded error	880.0	1.1	...	...
With 2.8 s additional error	882.8	1.5	7.0	7
Average with equal weights	884.0	1.8	...	...

measured values are valid and only the errors are in question, and because the agreement is poor we regard all the stated errors to be equally unreliable. We disregard the errors and average the eight measurements with equal weight. This gives $\tau_n = 884.0 \pm 1.8$ s.

The results are summarized in Table II. These methods differ in how the included experiments are weighted, in effect providing different distributions of the "blame" for the disagreement. Choosing the best method would require more detailed information than we have about the origins of systematic errors in these experiments. We note that the resulting averages all lie within a range of 880–884 s, as do the weighted averages of all 21 experiments and those performed after 1975. Because this range is consistent, we believe it is a fair representation of our current knowledge of the neutron lifetime. A single new experiment is unlikely to aid this situation; a new generation of experiments with improved methodology is needed.

We can make an indirect determination of the neutron lifetime using Eq. (8) and the value of G_V from superallowed beta decay: $G_V = 1.1358(5) \times 10^{-5}$ GeV^{-2} (Hardy and Towner, 2005) along with the value of $\lambda = G_A/G_V$ from neutron decay correlations: $\lambda = -1.2694(28)$ (Nakamura, 2010) and G_F from muon decay. This gives $\tau_n = 887.3 \pm 3.3$ s. The uncertainty is dominated by the expanded error in λ owing to a long-standing discrepancy between older and newer experimental results. We note that the two most recent and sophisticated neutron beta asymmetry measurements are in good agreement and their weighted average gives a significantly larger value with a smaller uncertainty: $\lambda = -1.2741(18)$ (Abele *et al.*, 2002; Liu *et al.*, 2010). This gives $\tau_n = 881.9 \pm 2.1$ s. Both of these values are consistent with our preferred range, but the second is in better agreement.

IV. SUMMARY AND FUTURE PROSPECTS

There are several emerging approaches that provide optimism about an improved determination of the neutron lifetime. Given the nonstatistical character of the discrepancy between the current results, it is clear that any new experiment must take particular cognizance of its sensitivity to systematic effects. For bottle experiments, as discussed, there are two particularly important concerns: (1) a good understanding of the details of the neutron-trap interaction, and (2) a robust method for the estimation of the required spectral cleaning to ensure that marginally trapped neutrons are adequately accounted for. For beam experiments, the critical point is a determination of the neutron density in the beam.

Given the concerns about the details of the microscopic interactions that occur when a neutron "bounces" off a material surface, there is an obvious appeal to the use of magnetic and gravitational confinement, where, at least in principle, the dynamics of the neutron-trap interaction is straightforward. The PENeLOPE project (Materne *et al.*, 2009), under development at the Technische Universität München, represents an important second generation magnetic bottle experiment. Like the magnetic bottle of Ezhov *et al.*, PENeLOPE will use a magnetic trap made from an array of permanent magnets that is "closed" by gravity at the top.

While magnetic traps offer the attraction of a well understood confinement mechanism, the calculation of particle trajectories, as well as the associated detailed modeling of loss mechanism, is complicated by the presence of chaotic orbits which, in principle, cannot be completely modeled numerically. A magnetic bottle deliberately designed to be fully chaotic would provide a nearly ergodic system whose losses may be estimated through the use of well understood scaling rules (Bowman and Penttila, 2005). This suggestion was the impetus for a new magnetic bottle experiment at Los Alamos National Laboratory (Walstrom *et al.*, 2009).

Until recently, the very low errors claimed by neutron bottle experiments have discouraged the pursuit of further neutron beam experiments. Now, however, the clear discrepancy between the bottle measurements has highlighted the importance of independent lifetime measurements subject to very different systematic effects. In light of this, the NIST group and its collaborators are currently carrying out a recalibration of the neutron flux monitor used to determine the neutron density in the most recent, and most accurate, beam experiment. In some ways, this new procedure recalls the Robson experiment in that the neutron detector used in the experiment is compared with a second detector that is configured in such a way that it is capable of an *a priori* absolute calibration. At least three such detector schemes have been explored: a ^{6}Li neutron calorimeter (Chowdhuri *et al.*, 2003), a ^{10}B alpha-gamma spectrometer (Gilliam, Greene, and Lamaze, 1989), and a ^{3}He gas scintillation chamber (Wietfeldt, 2008). Preliminary results from the alpha-gamma device have been reported (Yue *et al.*, 2010). These recalibrations will directly result in a more accurate, adjusted value for the lifetime that avoids the largest systematic effects in the previous result. It is noteworthy that yet another neutron beam experiment is being developed at Japan Proton Accelerator Research Complex (Shimizu, 2009) that will use a pulsed beam of neutrons to simplify the determination of the neutron density in the target.

When considering future measurements, it is important to recognize that when there is a significant number of individually discrepant data, the addition of further measurements whose errors are comparable does not easily improve the situation. This is because, as discussed, there is no rigorous statistical method that can justify the rejection of a specific experiment or set of experiments. Progress will only come from improved methods with substantially smaller errors or by a detailed reanalysis of systematic effects in existing measurements. We note that attempts at such reanalysis

have been made (Serebrov and Fomin, 2010; Steyerl *et al.*, 2010). Unfortunately, these reanalyses reach rather contrary conclusions. Notwithstanding the current unsatisfactory situation, we hope and expect that in the next few years there will be significant progress both in better experiments and in a better understanding of current experiments.

ACKNOWLEDGMENTS

We thank M. Scott Dewey, Susan Gardner, Vladimir Gudkov, and Jeffrey Nico for carefully reading the manuscript and providing helpful comments. We are grateful for support from the National Institute of Standards and Technology (NIST), U. S. Department of Commerce. F. E. Wietfeldt acknowledges support from the National Science Foundation Grant No. PHY-0855310. G. L. Greene acknowledges support from the U. S. Department of Energy Grant No. DE-FG02-03ER41258.

REFERENCES

Abele, H., M. Astruc Hoffmann, S. Baeßler, D. Dubbers, F. Glück, U. Müller, V. Nesvizhevsky, J. Reich, and O. Zimmer, 2002, Phys. Rev. Lett. **88**, 211801.

Alfimenkov, V. P., V. V. Nesvizhevsky, A. P. Serebrov, A. V. Strelkov, R. R. Taldaev, A. G. Kharitonov, and V. N. Shevtsov, 1992, Zh. Eksp. Teor. Fiz. **102**, 740.

Alpher, R. A., H. Bethe, and G. Gamow, 1948, Phys. Rev. **73**, 803.

Anton, F., W. Paul, W. Mampe, L. Paul, and S. Paul, 1989, Nucl. Instrum. Methods Phys. Res., Sect. A **284**, 101.

Arzumanov, S., L. Bondarenko, S. Chernyavsky, W. Drexel, A. Fomin, P. Geltenbort, V. Morozov, Y. Panin, J. Pendlebury, and K. Schreckenbach, 2000, Phys. Lett. B **483**, 15.

Bainbridge, K., 1933, Phys. Rev. **44**, 57.

Bates, J., 1983, Nucl. Instrum. Methods Phys. Res. **216**, 535.

Bondarenko, L., 2011 (private communication).

Bondarenko, L. N., V. V. Kurguzov, Y. A. Prokofiev, E. V. Rogov, and P. E. Spivak, 1978, JETP Lett. **28**, 303.

Bowman, J. D., and S. I. Penttila, 2005, J. Res. Natl. Inst. Stand. Technol. **110**, 361.

Brome, C. R., *et al.*, 2001, Phys. Rev. C **63**, 055502.

Burles, S., K. M. Nollett, J. W. Truran, and M. S. Turner, 1999, Phys. Rev. Lett. **82**, 4176.

Byrne, J., P. G. Dawber, C. Habeck, S. Smidt, J. Spain, and A. Williams, 1996, Europhys. Lett. **33**, 187.

Byrne, J., P. G. Dawber, R. D. Scott, J. M. Robson, and G. L. Greene, 1986, in *NBS Special Publication 711: The Investigation of Fundamental Interactions with Cold Neutrons*, edited by G. L. Greene (National Bureau of Standards, U. S. Department of Commerce, Washington, DC), pp. 48–53.

Byrne, J., *et al.*, 1990, Phys. Rev. Lett. **65**, 289.

Byrne, J., J. Morse, K. F. Smith, F. Shaikh, K. Green, and G. L. Greene, 1980, Phys. Lett. B **92**, 274.

Carlson, A., W. Poenitz, G. Hale, R. Peelle, D. Doddler, C. Fu, and W. Mannhart, 1993, National Institute of Standards and Technology Internal Report No. 5177.

Chadwick, J., 1932, Proc. R. Soc. A **136**, 692.

Chadwick, J., and M. Goldhaber, 1934, Nature (London) **134**, 237.

Chadwick, J., and M. Goldhaber, 1935, Proc. R. Soc. A **151**, 479.

Chowdhuri, Z., *et al.*, 2003, Rev. Sci. Instrum. **74**, 4280.

Christensen, C. J., A. Nielsen, A. Bahnsen, W. K. Brown, and B. M. Rustad, 1972, Phys. Rev. D **5**, 1628.

Cohen, E. R., and B. N. Taylor, 1973, J. Phys. Chem. Ref. Data **2**, 663.

Cyburt, R. H., B. D. Fields, and K. A. Olive, 2003, Phys. Lett. B **567**, 227.

Cyburt, R. H., B. D. Fields, and K. A. Olive, 2008, J. Cosmol. Astropart. Phys. 11, 012.

D'Angelo, N., 1959, Phys. Rev. **114**, 285.

Dewey, M. S., *et al.*, 2003, Phys. Rev. Lett. **91**, 152302.

Doyle, J. M., and S. K. Lamoreaux, 1994, Europhys. Lett. **26**, 253.

Dubbers, D., 1991, Nucl. Phys. **A527**, 239.

Dunkley, J., *et al.*, 2009, Astrophys. J. Suppl. Ser. **180**, 306.

Eidelman, S., *et al.*, 2004, Phys. Lett. B **592**, 1, Review of Particle Physics.

Ezhov, V., 2009, *Seventh International Workshop on Cold and Ultracold Neutrons, St. Petersburg Russia*.

Ezhov, V., *et al.*, 2009, Nucl. Instrum. Methods Phys. Res., Sect. A **611**, 167, particle physics with slow neutrons.

Fermi, E., 1934, Z. Phys. **88**, 161.

Fermi, E., E. Amaldi, O. D'Agostino, F. Rasetti, and E. Segre, 1934, Proc. R. Soc. A **146**, 483.

Feynman, R. P., and M. Gell-Mann, 1958, Phys. Rev. **109**, 193.

Gamow, G., 1946, Phys. Rev. **70**, 572.

Gamow, G., 1966, *Thirty Years That Shook Physics* (Doubleday, Garden City, New York).

Gamow, G., and E. Teller, 1936, Phys. Rev. **49**, 895.

Gell-Mann, M., 1958, Phys. Rev. **111**, 362.

Gilliam, D., G. Greene, and G. Lamaze, 1989, Nucl. Instrum. Methods Phys. Res., Sect. A **284**, 220.

Glashow, S. L., and J. Iliopoulos, 1971, Phys. Rev. D **3**, 1043.

Golub, R., C. Jewell, P. Ageron, W. Mampe, B. Heckel, and I. Kilvington, 1983, Z. Phys. B **51**, 187.

Golub, R., and J. M. Pendlebury, 1975, Phys. Lett. A **53**, 133.

Golub, R., D. Richardson, and S. K. Lamoreaux, 1991, *Ultra-Cold Neutrons* (Adam Hilger, Bristol, England).

O'Shaughnessy, C., *et al.*, 2009, Nucl. Instrum. Methods Phys. Res., Sect. A **611**, 171, particle physics with slow neutrons.

Hardy, J. C., and I. S. Towner, 2005, Phys. Rev. C **71**, 055501.

Heisenberg, W., 1932, Z. Phys. **77**, 1.

Holstein, B. R., 1974, Rev. Mod. Phys. **46**, 789.

Huffman, P., *et al.*, 2000, Nature (London) **403**, 62.

Iwanenko, D., 1932, Nature (London) **129**, 798.

Izotov, Y. I., and T. X. Thuan, 2010, Astrophys. J. Lett. **710**, L67.

Jackson, J. D., S. B. Treiman, and H. W. Wyld, 1957, Phys. Rev. **106**, 517.

Kharitonov, A., V. Nesvizhevsky, A. Serebrov, R. Taldaev, V. Varlamov, A. Vasilyev, V. Alfimenkov, V. Lushchikov, V. Shvetsov, and A. Strelkov, 1989, Nucl. Instrum. Methods Phys. Res., Sect. A **284**, 98.

Kossakowski, R., P. Grivot, P. Liaud, K. Schreckenbach, and G. Azuelos, 1989, Nucl. Phys. **A503**, 473.

Kosvintsev, Y. Y., Y. A. Kushnir, V. I. Morozov, and G. I. Terekhov, 1980, JETP Lett. **31**, 236.

Kosvintsev, Y. Y., V. I. Morozov, and G. I. Terekhov, 1986, JETP Lett. **44**, 571.

Kugler, K. J., W. Paul, and U. Trinks, 1978, Phys. Lett. B **72**, 422.

Last, J., M. Arnold, J. Döhner, D. Dubbers, and S. J. Freedman, 1988, Phys. Rev. Lett. **60**, 995.

Liu, J., *et al.* (UCNA Collaboration), 2010, Phys. Rev. Lett. **105**, 181803.

Lopez, R. E., and M. S. Turner, 1999, Phys. Rev. D **59**, 103502.

Mampe, W., P. Ageron, C. Bates, J. M. Pendlebury, and A. Steyerl, 1989, Phys. Rev. Lett. **63**, 593.

Mampe, W., L. N. Bondarenko, V. I. Morozov, Y. N. Panin, and A. I. Fomin, 1993, JETP Lett. **57**, 82.

Marciano, W. J., and A. Sirlin, 2006, Phys. Rev. Lett. **96**, 032002.

Materne, S., R. Picker, I. Altarev, H. Angerer, B. Franke, E. Gutsmiedl, F. Hartmann, A. Müller, S. Paul, and R. Stoepler, 2009, Nucl. Instrum. Methods Phys. Res., Sect. A **611**, 176, particle physics with slow neutrons.
McClintock, P., 1978, Cryogenics **18**, 201.
Mohr, P. J., B. N. Taylor, and D. B. Newell, 2008, Rev. Mod. Phys. **80**, 633.
Morozov, V., 1989, Nucl. Instrum. Methods Phys. Res., Sect. A **284**, 108.
Nakamura, K. (Particle Data Group), 2010, J. Phys. G **37**, 075021.
Nico, J. S., *et al.*, 2005, Phys. Rev. C **71**, 055502.
Olive, K. A., G. Steigman, and T. P. Walker, 2000, Phys. Rep. **333**, 389.
Paul, W., F. Anton, L. Paul, S. Paul, and W. Mampe, 1989, Z. Phys. C **45**, 25.
Pauwels, J., R. D. Scott, R. Eykens, P. Robouch, J. V. Gestel, J. Verdonck, D. M. Gilliam, and G. Greene, 1995, Nucl. Instrum. Methods Phys. Res., Sect. A **362**, 104.
Pichlmaier, A., J. Butterworth, P. Geltenbort, H. Nagel, V. Nesvizhevsky, S. Neumaier, K. Schreckenbach, E. Steichele, and V. Varlamov, 2000, Nucl. Instrum. Methods Phys. Res., Sect. A **440**, 517.
Pichlmaier, A., P. Geltenbort, K. Schreckenbach, and V. Varlamov, 2001, *Meeting of the American Physical Society, Washington, D.C.*.
Pichlmaier, A., V. Varlamov, K. Schreckenbach, and P. Geltenbort, 2010, Phys. Lett. B **693**, 221.
Robson, J., 1950, Phys. Rev. **78**, 311.
Robson, J. M., 1951, Phys. Rev. **83**, 349.
Rutherford, E., 1920, Proc. R. Soc. A **97**, 374.
Salam, A., 1968, *Nobel Symposium, No. 8*, edited by N. Svartholm (Almquist and Wiksell, Stockholm), p. 367.
Scott, R. D., P. D'hondt, R. Eykens, P. Robouch, D. F. G. Reher, G. Sibbens, J. Pauwels, and D. M. Gilliam, 1995, Nucl. Instrum. Methods Phys. Res., Sect. A **362**, 151.
Serebrov, A., *et al.*, 2005, Phys. Lett. B **605**, 72.
Serebrov, A. P., and A. K. Fomin, 2010, Phys. Rev. C **82**, 035501.
Serebrov, A. P., *et al.*, 2008, Phys. Rev. C **78**, 035505.
Severijns, N., M. Beck, and O. Naviliat-Cuncic, 2006, Rev. Mod. Phys. **78**, 991.
Shimizu, H., 2009, *Presented at the Seventh International Workshop on Cold and Ultracold Neutrons, St. Petersburg Russia, 2009.*
Snell, A. H., and L. Miller, 1948, Phys. Rev. **74**, 1217.
Snell, A. H., F. Pleasonton, and R. V. McCord, 1950, Phys. Rev. **78**, 310.
Sosnovsky, A. N., P. E. Spivak, Y. A. Prokofiev, I. E. Kutikov, and Y. P. Dobrinin, 1959, Nucl. Phys. **10**, 395.
Spivak, P. E., 1988, Zh. Eksp. Teor. Fiz. **94**, 1.
Spivak, P. E., A. N. Sosnovsky, Y. A. Prokofiev, and V. S. Sokolov, 1956, *Proceedings of the International Conference on the Peaceful Uses of Atomic Energy, Geneva, 1955* (United Nations, New York), Vo. 2, p. 33.
Steyerl, A., S. S. Malik, A. M. Desai, and C. Kaufman, 2010, Phys. Rev. C **81**, 055505.
Steyerl, A., *et al.*, 1986, Phys. Lett. A **116**, 347.
Sudarshan, E. C. G., and R. E. Marshak, 1958, Phys. Rev. **109**, 1860.
Sutherland, W., 1899, Philos. Mag. **47**, 273.
Walstrom, P., J. Bowman, S. Penttila, C. Morris, and A. Saunders, 2009, Nucl. Instrum. Methods Phys. Res., Sect. A **599**, 82.
Weinberg, S., 1967, Phys. Rev. Lett. **19**, 1264.
Wietfeldt, F. E., 2008, Absolute neutron fluence measurement using ^{3}He gas scintillation, unpublished report.
Yang, L., 2006, Ph.D. thesis (Harvard University).
Yue, A., G. Greene, M. Dewey, D. Gilliam, J. Nico, and A. Laptev, 2010, in *Bulletin of the American Physical Society*, Vol. 15–4, p. 30.