It is now recognized that the International System of Units (SI units) will be redefined in terms of fundamental constants, even if the date when this will occur is still under debate. Actually, the best estimate of fundamental constant values is given by a least-squares adjustment, carried out under the auspices of the Committee on Data for Science and Technology (CODATA) Task Group on Fundamental Constants. This adjustment provides a significant measure of the correctness and overall consistency of the basic theories and experimental methods of physics using the values of the constants obtained from widely differing experiments. The physical theories that underlie this adjustment are assumed to be valid, such as quantum electrodynamics (QED). Testing QED, one of the most precise theories is the aim of many accurate experiments. The calculations and the corresponding experiments can be carried out either on a boundless system, such as the electron magnetic moment anomaly, or on a bound system, such as atomic hydrogen. The value of fundamental constants can be deduced from the comparison of theory and experiment. For example, using QED calculations, the value of the fine structure
constant given by the CODATA is mainly inferred from the measurement of the electron magnetic moment anomaly carried out by Gabrielse’s group. (Hanneke et al. 2008 Phys. Rev. Lett. 100, 120801) The value of the Rydberg constant is known from two-photon spectroscopy of hydrogen combined with accurate theoretical quantities. The Rydberg constant, determined by the comparison of theory and experiment using atomic hydrogen, is known with a relative uncertainty of \(6.6 \times 10^{-12}\). It is one of the most accurate fundamental constants to date. A careful analysis shows that knowledge of the electrical size of the proton is nowadays a limitation in this comparison. The aim of muonic hydrogen spectroscopy was to obtain an accurate value of the proton charge radius. However, the value deduced from this experiment contradicts other less accurate determinations. This problem is known as the proton radius puzzle. This new determination of the proton radius may affect the value of the Rydberg constant \(R_\infty\). This constant is related to many fundamental constants; in particular, \(R_\infty\) links the two possible ways proposed for the redefinition of the kilogram, the Avogadro constant \(N_A\) and the Planck constant \(h\). However, the current relative uncertainty on the experimental determinations of \(N_A\) or \(h\) is three orders of magnitude larger than the ‘possible’ shift of the Rydberg constant, which may be shown by the new value of the size of the proton radius determined from muonic hydrogen. The proton radius puzzle will not interfere in the redefinition of the kilogram. After a short introduction to the properties of the proton, we will describe the muonic hydrogen experiment. There is intense theoretical activity as a result of our observation. A brief summary of possible theoretical explanations at the date of writing of the paper will be given. The contribution of the proton radius puzzle to the redefinition of SI-based units will then be examined.

Keywords: proton radius; muonic hydrogen; Rydberg constant; Planck constant; Avogadro constant

1. The proton

Even though the proton is one of the most abundant constituents of the visible Universe, some of its properties are not well known. The study of its properties is important for our deep understanding of matter. The proton is made up of three valence quarks (up, up, down), held together by strong interactions. Ab initio calculations can be made on this structure using quantum chromodynamics (QCD) theory.

An important step has been the calculation of the mass of the proton using QCD theory with a relative uncertainty better than 4 per cent [1,2]. Experimentally, the absolute mass of the proton is known with a relative uncertainty of \(5 \times 10^{-8}\) [3].

The ab initio calculation of the spin of the proton (i.e. 1/2) has also been undertaken. The present state of the problem is known as the ‘spin crisis’ [4]. This refers to the experimental finding that only a small contribution of the spin of the proton seems to be carried by the quarks [5–7]. Indeed, the study of the spin structure of the nucleon is an important task in particle physics, either theoretically [8] or experimentally [9].

Some attempts have also been made to estimate the proton charge radius with QCD theory. Some values of \(r_p\) have been published [10,11]. New and more
reliable values of the proton radius should be available in the future as calculation
capabilities are increasing with time.

2. Experimental determinations of the proton charge radius

(a) Scattering experiments

In fact, the best knowledge of the proton charge radius comes from experiments. First determinations were given by electron–proton elastic scattering experiments. The principle of these experiments is to measure the differential cross section of a scattered electron beam sent onto a thin hydrogen (H₂) target. The relevant parameter is the space-like momentum transfer $-Q²$. The proton mean square charge radius is given by the slope of the Sachs form factor ($G_E$) of the proton at $Q² = 0$,

$$\langle r_p^2 \rangle = -6 \frac{\partial G_E(Q^2)}{Q^2} \bigg|_{Q^2=0}. \quad (2.1)$$

The proton radius is defined as

$$r_p = \sqrt{\langle r_p^2 \rangle}. \quad (2.2)$$

The main contribution to $r_p$ comes from low-momentum measurements. Even if conceptually the experiment is simple, practically it is not. The space-like momentum transfer cannot be made arbitrarily small. The electron must cross the target in which a single elastic scatter has to occur. Measurements near to $Q^2 = 0$ must not be affected by the direct electron beam. Therefore, the proton radius can only be obtained from extrapolation of the differential cross section at $Q^2 = 0$. This extrapolation is strongly model dependent. The first determination of $r_p$ was obtained in 1955 [12]. In the 1980s, an experiment was specially designed in Mainz to obtain an accurate value of $r_p$ from low-momentum transfers. The analysis of the data obtained in Mainz gives $r_p = 0.862(12)\text{fm}$ [13], which is in strong disagreement with an accurate previous determination [14]. During the 2000s, no new scattering experiment were carried out at low-momentum transfers. All the different values of $r_p$ which have been published come from a re-analysis of the low-momentum scattering data, associated or not to high-momentum scattering data (e.g. [15,16]). The re-analysis considered by the Committee on Data for Science and Technology (CODATA) task group has been obtained by Sick [17]. This work takes into account the world data on elastic electron–proton scattering, the coulomb distortion, and uses a parametrization that allows us to deal properly with the higher moments. In the 2000s, a sophisticated experiment was carefully designed at Mainz to measure $r_p$ accurately [18]. Taking advantage of three high-resolution spectrometers, it was possible to measure the elastic electron–proton scattering cross section with a statistical precision of better than 0.2 per cent. The value of $r_p$ published in 2010, deduced from an analysis done in Mainz with their data, is $0.879(8)\text{fm}$ [19]. Another value of $r_p$ from scattering experiments has been published very recently [20]. This value of the proton radius is deduced from the global analysis reported in Arrington et al. [21], in which the new measurements done at ‘high’ $Q^2$ [20] are included.
At short distance, the coulombian electrostatic potential is ‘shielded by the finite size of the proton’. Consequently, the energy levels of atomic hydrogen are slightly shifted. Therefore, a value of $r_p$ can be obtained from high-resolution spectroscopy of hydrogen. A simplified but powerful analysis of the main hydrogen data has been presented in de Beauvoir et al. [22] by using only the most accurate experimental data: the frequency of the 1S–2S transition measured in Garching, Germany [23], the 2S–8S/8D transitions measured in Paris, France [24], and the $1/n^3$ dependence of quantum electrodynamics (QED) corrections [25,26] (where $n$ is the principal quantum number of hydrogen theory). A precise value of the 1S Lamb shift of hydrogen can be extracted from a proper linear combination of these three quantities. Assuming the exactness of QED calculations, an accurate value of $r_p$ can be deduced. A complete analysis of all the spectroscopic measurements is also regularly done by the CODATA Task Group on Fundamental Constants (e.g. [3]). The last resulting spectroscopic value of $r_p$ was 0.8760(78) fm.

The aim of the muonic hydrogen spectroscopy of the 2S–2P splitting was to determine the proton radius more accurately. Muonic hydrogen ($\mu$-p) is an exotic atom in which the muon ($\mu^-$) replaces the electron ‘orbiting around’ the proton in normal atomic hydrogen. Like the electron, the muon is a lepton, but it is 207 times heavier and its lifetime is only 2.2 $\mu$s. Because of the mass dependence, the Bohr radius of $\mu$-p is around 200 times smaller than that of electronic hydrogen, and the sensitivity to the finite size of the proton is then largely enhanced. The contribution owing to the finite size of the proton is around 2 per cent to the 2S–2P splitting of muonic hydrogen, whereas it is only $1.4 \times 10^{-4}$ for the $2S_{1/2}$–$2P_{1/2}$ of electronic hydrogen. Another important issue in the comparison between muonic and electronic hydrogen is the vacuum polarization, which is the dominant contribution to the 2S–2P splitting of $\mu$-p. As the Bohr radius is 200 times smaller, the overlap of S state wave functions with the distribution of virtual electron–positron pairs is more important; consequently, the contributions of the vacuum polarization corrections are larger. The wavelength of the 2S–2P transition in $\mu$-p is 6 $\mu$m and the oscillator strength is $10^{-7}$ of that of electronic hydrogen. Moreover, the population of muonic hydrogen in the 2S state is small. Because of those considerations this experiment has been really challenging.

3. Muonic hydrogen experiment

(a) Principle of the experiment

The principle of the experiment has been described in many papers [27].

Let us recall the main step. The muon ($\mu^-$) beam is sent into a H$_2$ gas target. Muonic hydrogen atoms are formed at the principal quantum number of around $n = 14$: the muon goes from the continuum to a bound state by ejection of an Auger electron. Subsequently, the H$_2$ molecule breaks up and the muonic cascade of the neutral $\mu$-p atom begins. Radiative and collisionally induced de-excitation processes bring it down to the 2S or 1S state within a few nanoseconds (figure 1).
This cascade has been carefully studied [28], especially the collisional quenching of the 2S state [29–31]. Ninety-nine per cent of atoms decay to the 1S state, producing a large 2keV fluorescence early peak. About 1 per cent of muonic hydrogen is formed in the long-lived 2S state [32]. A laser pulse triggered by muons (see below) is sent into the H2 target. Because of the short delay between the laser trigger and the output of the laser, a time-delayed 2keV fluorescence peak is observed if the 2S–2P transition has been excited by the 6µm light.

(b) Apparatus

The challenges of the muonic hydrogen experiment were the production of long-lived muonic hydrogen in the 2S state, the development of the 6µm laser that can be randomly triggered with a short delay and the analysis of the small signal produced (few counts/hours).

To reduce collisional quenching of the 2S state, muons are stopped in H2 gas at a pressure of 1hPa. To do this, a special low-energy muon beam (approx. keV) has been designed and built at the Paul Scherrer Institute to efficiently stop muons at this ultra-low gas pressure within the small target volume required for efficient laser excitation. At 1hPa, the lifetime of the 2S metastable state is around 1µs [29,30]. Many details of this muon line can be found in [28,33–37]. We want to emphasize that a lot of effort has been put into the quality of the signal triggering the laser. The detection of keV-muons without stopping them is not a simple task. It is done with a thin carbon foil stack which simultaneously acts as the muon detector and improves the beam quality by frictional cooling [38].

The design of the laser is dictated by the need for a tunable light source at 6µm that can be triggered within 1µs after a random trigger by incoming muons. The laser chain used in the 2009 run derived from development of the initial laser...
Proton radius to redefine SI units

chain used in 2003 [34]. The fast and powerful pulsed laser that can be triggered at 515 nm is used to pump an oscillator–amplifier titanium sapphire (TiSa) laser. Three consecutive vibrational Raman scatterings in H$_2$ in a multiple-pass cell are used to convert the 708 nm light into a 6 µm pulse. The 6 µm light is then sent, 20 m away, in a specially design multi-pass non-resonant cavity surrounding the H$_2$ target in which µ-p atoms are formed. The latter cavity is used to illuminate all the stopping volume of the muon in the target (5 × 15 × 190 mm$^3$). The frequency of the 6 µm light is driven by the TiSa oscillator, which is seeded by a cw-TiSa laser. The frequency of the cw-laser is permanently controlled with two wave metres, a very stable Fabry–Perot cavity and atomic/molecular lines (I$_2$, Cs, Rb). The cw-laser is permanently locked on a Fabry–Perot fringe. The step of scanning the laser frequency is a multiple of the free spectral range of this Fabry–Perot interferometer [39]. Whereas the design of the TiSa ensemble has remained the same between the first and the last runs [40], the design of the green pump laser has had to be changed drastically. A thin disc laser had to be built specially for our experiment [41].

The analysis of small signal requires very good knowledge of the background. Since the early stages of the experiment, the 2 keV detectors (large area avalanche photodiodes) have been well studied in order to maximize their time and energy resolutions [35,36,42]. During the data analysis, background events were efficiently rejected. The final rate of about one background event per hour originated mainly from electrons from muon decay which were wrongly identified as 2 keV Lyman-α X-rays.

(c) Results

Many muonic hydrogen lines were observed during the fourth beam time period. The first line which was observed and analysed was the most intense one: 2S$_{1/2}$(F = 1)–2P$_{3/2}$(F = 2) of muonic hydrogen (figure 2). The signal is clearly visible above the noise. A total of 550 events were measured in the resonance, where 155 background events were expected. A Lorentzian profile has been fitted to the data to give the frequency of the line versus the Fabry–Perot fringe. The absolute frequency of the line is determined from the absolute calibration of the light at 6 µm using the well-known H$_2$O spectroscopy and the free spectral range of the Fabry–Perot interferometer (figure 3). A careful and detailed analysis of the uncertainty budget of the centroid position of this line has been done in Nebel [37] and Pohl et al. [39].

The quantity determined by our experiment is the absolute frequency of the 2S$_{1/2}$(F = 1)–2P$_{3/2}$(F = 2) resonance of muonic hydrogen: 49 881.88(76) GHz. The main contribution to the uncertainty is statistical; it is given by the fit of the Lorentzian shape on the experimental data. The relative uncertainty of the centroid position (1.6 × 10$^{-5}$) corresponds to 4 per cent of the line width (approx. 18 GHz), which shows that our experiment is far away from a high-resolution spectroscopy one but which also indicates a very weak dependence of the centroid of the resonance on the line shape model.

Nevertheless, the present result is good enough to extract the proton radius with the smallest uncertainty to date. The value of the proton radius is obtained from the comparison of our determination of the frequency of the 2S$_{1/2}$(F = 1)–2P$_{3/2}$(F = 2) muonic line with the theoretical prediction. The theoretical
Figure 2. $2S_{1/2}(F = 1) - 2P_{3/2}(F = 2)$ resonance. The discrepancy with other determinations is clearly visible. The frequency gap is around 75 GHz with the position of the line expected with electronic hydrogen. The inset calibration $H_2O$ line was recorded at 30 hPa. (Online version in colour.)

Figure 3. Principle of the determination of the absolute frequency of the muonic hydrogen line. The absolute calibration has been made with five $H_2O$ lines, 17 $I_2$ lines, five Cs two-photon atomic lines and three Rb lines. Absolute calibration between 5.49 µm and 6.04 µm was performed with $H_2O$ lines with the pulsed laser source while the free spectral range of the Fabry–Perot interferometer was determined in the cw regime in the 695–780 nm range with well-known atomic/molecular lines. To detect an eventual drift of the Fabry–Perot interferometer, this was regularly compared with iodine lines before, during and after the beam time. The integer numbers $(N,N')$ can be determined accurately using the two wavemeters. (Online version in colour.)

prediction derives for a large part from the bound state QED of electronic hydrogen scaled with the mass of the muon. These predictions account for radiative, recoil, proton structure fine and hyperfine contributions. A detailed
Proton radius to redefine SI units

Figure 4. Comparison of the various determinations of the proton radius. Defining the proton radius as the slope of the Sachs form factor, experimental results are sorted from low-momentum transfer (electronic hydrogen) to high-momentum transfer (scattering experiments) [44]. This presentation emphasizes the surprising result of muonic hydrogen, which can be viewed as a ‘dip in the form factor slope’ at the corresponding $Q^2$ momentum of the muonic hydrogen experiment. (Online version in colour.)

The predicted frequency of the $2S_{1/2}(F = 1) - 2P_{3/2}(F = 2)$ transition of muonic hydrogen is

$$\nu (\text{GHz}) = 50772.43(1.18) - 1263.69 \times r_p^2 + 8.39 \times r_p^3.$$  \hspace{1cm} (3.1)

The deduced value of $r_p$ is $0.84184(36)(56)$ fm, where the first and second uncertainty originate, respectively, from experimental uncertainty and theoretical uncertainty. This result ($0.84184(68)$ fm) differs significantly (approx. 5 s.d.) from other determinations (figure 4).

4. The proton radius puzzle

The confusing situation about $r_p$ is known as the proton radius puzzle. It has stimulated intense activity in the community. It is impossible to make an exhaustive list of all the contributions already available on the arxiv base; we apologize to authors whose papers are not cited. We can only point out some of the searches at the present date.

(a) Definition of the proton radius

An obvious concern in the comparison is to make sure that the proton radius extracted from various experiments has the same meaning. The first positive answer was given in Jentschura [45]: ‘a conceivable accidental incompatibility of
the conventions used in references [...] for the proton radius therefore cannot be the reason for the observed discrepancy’, where [...] refers to all the proton radius determinations.

(b) Charge distribution in the proton

The charge radius distribution is related to $G_E$ by the Fourier transformation. One attempt has been made to re-evaluate the third Zemach moment of the proton [46] $\langle r_p^3 \rangle$ to solve the proton radius puzzle. However, the re-evaluation of $\langle r_p^3 \rangle$ with the charge distribution presented in De Rujula [47] is in contradiction to the electron–proton scattering data [48,49].

(c) Quantum electrodynamics corrections

The contradiction between the proton radius extracted from electronic hydrogen and that from muonic hydrogen is intriguing for QED specialists. Because of the mass scaling law, the QED of muonic hydrogen has always been considered to be simpler than that of the hydrogen atom. Nevertheless, the proton radius puzzle has stimulated either careful checks of present QED corrections or evaluation of new corrections for hydrogen atoms [50–52]. Up to now, the proton radius puzzle cannot be solved with new or wrong QED corrections.

(d) Rydberg constant

The Rydberg constant is the scaling factor of the atomic level. Assuming the correctness of the QED calculations in the electronic hydrogen atom, the problem can be solved by shifting by 5 s.d. the value of the Rydberg constant, known with an uncertainty of 22 kHz, using the most accurate frequency measured in hydrogen [23]. However, the CODATA Rydberg constant is determined not only with 1S–2S and 2S–8S frequency transitions but also with other frequency measurements that are certainly less accurate but that are all highly consistent with each other and with theory [3]. Moreover, the recent measurement of the 1S–3S transition in hydrogen [53] is in very good agreement with the theoretical estimate [54]. At the present time, there are no indications of a disagreement between theory and experiments in electronic hydrogen. But it is clear that the accuracy of the hydrogen experiments other than the 1S–2S one has to be improved to clarify the situation. Figure 5 shows all the values of $r_p$ deduced from the 1S–2S transition frequency, the $1/n^3$ law and only one of the 2S–n(S,P,D) transition frequencies that have been measured. Individually, there is not a large discrepancy with the muonic determination of $r_p$. On the other hand, most of the values are pointing in the same direction.

Many ongoing experiments are being carried out to test the bound state QED. Results for the 2S–6S/D transitions measured at the National Physical Laboratory [62], which are currently being analysed, may bring an important and independent contribution to the hydrogen spectroscopy and so to the proton radius puzzle. Combined with the ongoing experiment at Garching, which is aimed at measuring the 1S–2S transition in He$^+$ [63], the experiment planned at the Paul Scherrer Institute to measure the muonic helium ion Lamb shift ($\mu$He$^+$) may illuminate the proton radius conundrum [64]. A determination of the
Proton radius to redefine SI units

4073

CODATA 2006

Figure 5. Comparison of various determinations of the proton radius from hydrogen spectroscopy. Each value is obtained from the 1S–2S transition frequency, the $1/n^3$ law and one of the other hydrogen experimental data from 2S–$n$(S,P,D). ((a) From Lundeen & Pipkin [55], (b) from Hagley & Pipkin [56], (c) from Newton et al. [57], (d) from Weitz et al. [58], (e) from Berkeland et al. [59], (f) from Bourzeix et al. [60] combined with Arnoult et al. [53], (g) from de Beauvoir et al. [24], (h) from Schwob et al. [61], and (i) from Arnoult et al. [53]). The double line corresponds to the uncertainty of the proton radius determination obtained from muonic hydrogen spectroscopy. (Online version in colour.)

Rydberg constant nearly independent of the nucleus structure has also started at NIST Gaithersburg, with the study of circular states of $^{20}$Ne$^{9+}$ [65].

(e) New physics

A possible way to solve the proton radius puzzle would be to introduce new ‘particles’ which are coupled differently to the muon compared with the electron. However, this new physics search is already constrained by many low-energy data [44,66,67]. Indeed, for example, recent laboratory-sized experiments are sufficiently accurate to set a limit on the internal structure of the electron or on the existence of new dark matter particles assuming the exactness of QED calculations or testing for the first time the muon and hadron contributions to the electron anomaly $\alpha_e$ [68,69]. A deviation from Coulomb’s Law in muonic hydrogen is also ruled out by measurements in electronic hydrogen [70].

5. The proton and the redefinition of the International System of Units

As discussed above, the new determination of the proton radius may affect the value of the Rydberg constant $R_\infty$. This constant is related to many fundamental constants. For example, the estimate of the mass of the electron $m_e$ in the CODATA adjustment is derived from the relation

$$m_e = \frac{2R_\infty h}{c\alpha^2},$$

(5.1)
where $\alpha$ is the fine structure constant, $c$ is the velocity of light and $h$ is the Planck constant. The Rydberg constant also links the two ways proposed for the redefinition of the kilogram, the Avogadro constant $N_A$ and the Planck constant $h$,

$$N_A \times h = \frac{c}{2} \frac{A_e \alpha^2 M_u}{R_{\infty}}. \quad (5.2)$$

However, the current relative uncertainty on the experimental determinations of $N_A$ [71] or $h$ [72] is of the order of a few parts of $10^8$. This is three orders of magnitude larger than the ‘possible’ shift of the Rydberg constant that may be disclosed by the new value of the proton size radius from muonic hydrogen. The proton radius puzzle will not interfere in the redefinition of the kilogram.

For many years, the hydrogen atom has been advocated as an ‘ideal’ clock (e.g. [73]) as, in principle, it can be calculated in contrary to the other clocks. Should the second be redefined by the hydrogen atom in terms of fundamental constants? Today, calculations for hydrogen are about 10,000 times less accurate than current experimental caesium clock precision, owing to the growing complexity of the calculations of higher order corrections. The best experimental determination is also $10^3$ times worse than the best optical clocks, which are front-runners to redefining the second in several years’ time (see Gill [74]). Improved theory with higher order corrections to two or three more orders is needed before the calculable clock approach for hydrogen becomes competitive.

Nevertheless, finding the solution of the proton radius puzzle would be a step in the long quest to decode the hydrogen spectrum: the ‘Rosetta stone of modern physics’ [75].

References

Proton radius to redefine SI units


18 Bernauer, J. C. 2010 Measurement of the elastic electron–proton cross section and separation of the electric and magnetic form factor in the Q^2 range from 0.004 to 1 (GeV/c)^2. PhD thesis, Johannes Gutenberg-Universität Mainz. (http://ubm.opus.hbz-nrw.de/volltexte/2010/2442/).


Proton radius to redefine SI units


