Abstract. Precise determinations of transition frequencies of simple atomic systems are required for a number of fundamental applications such as tests of quantum electrodynamics (QED), the determination of fundamental constants and nuclear charge radii. The sharpest transition in atomic hydrogen occurs between the metastable 2S state and the 1S ground state. Its transition frequency has now been measured with almost 15 digits accuracy using an optical frequency comb and a cesium atomic clock as a reference [1]. A recent measurement of the $2\text{S} - 2\text{P}_{3/2}$ transition frequency in muonic hydrogen is in significant contradiction to the hydrogen data if QED calculations are assumed to be correct [2, 3]. We hope to contribute to this so-called “proton size puzzle” by providing additional experimental input from hydrogen spectroscopy.

1. Introduction

Atomic hydrogen has been the Rosetta Stone for quantum physics. Through the successive refinement of its theoretical description, starting from the phenomenological description of J.J Balmer, it provided a concise argument for the wave nature of matter. N. Bohr’s quantization requirement, that the electrons angular momentum must be an integer multiple of $\hbar$, can be understood such that the corresponding de Broglie wave forms a standing wave. E. Schrödinger was the first to formulate a wave equation for matter wave. With his Schrödinger equation, he put the emerging field quantum mechanics on solid grounds, while P.A.M. Dirac figured out how to include relativity. These refinements were introduced to keep up with increasing experimental accuracy, and to fix the resulting inconsistencies, but continuously complicated the theoretical picture. W.E. Lamb and R.C. Retherford discovered a discrepancy with the Dirac theory in 1947. It turned out that this was due to the fact that quantum mechanics had not yet been applied in the most rigorous way. In particular, the effects of the quantum vacuum, both the electronic and photonic, had been neglected up to then. Their inclusion led to the development of Quantum Electrodynamics (QED) by E.A. Ühling, H.A. Bethe, R.P. Feynman and others. Unlike any of the predecessors, QED survived about six orders of magnitude improvement of the experimental accuracy and is still the current description today. Of course several discrepancies between experiment and theory have shown up in the last 65 years, but thus far they could all traced back to errors in the computation of QED terms, neglected higher order terms or to
underestimated experimental errors. In that sense QED is probably the most successful theory in all of physics. As such it has served as a template for all subsequent quantum field theories.

A serious conceptual drawback of this development is not only that our description of nature has become more complicated but also that a parameter was introduced, the dimensionless fine structure constant \( \alpha \approx \frac{1}{137} \), in order to account for the observed fine structure of hydrogen spectral lines. Since we do not know any way to calculate fundamental constants, such as \( \alpha \), from first principles, the additional constant alpha reduces the predictive power of theory. In practice though, there are even more parameters that enter. The electron to proton mass ratio \( m_e/m_p \) and proton charge radius \( r_p \) take into account the finite mass and size of the nucleus respectively. Finally, most transition frequencies in hydrogen are measured in SI units, so that the Rydberg constant \( R_\infty \) appears as a multiplicative unit converter from atomic units to SI units. In QED the hydrogen energy levels can be expressed, like any other QED result, as a power series in \( \alpha \):

\[
E = R_\infty \left( -\frac{1}{n^2} + A_{20} \alpha^2 + A_{30} \alpha^3 + A_{31} \alpha^3 \ln(\alpha) + A_{40} \alpha^4 + \ldots + \frac{16\pi^2 m_e^2 c^2 \alpha^2}{3n^3 \hbar^2} r_p^2 \right)
\]

In leading order this reproduces the result of N. Bohr and E. Schrödinger. The dependence of \( m_e/m_p \), as well as on the other quantum numbers, is hidden in the coefficients \( A_{mn} \) of which a complete collection of all the known ones are listed in [4]. The last term in eqn.(1) describes the effect of the finite size of the nucleus in terms of the r.m.s proton charge radius \( r_p \) for S-states. For hydrogen this term contributes very little to the total energy\(^1\), but adds the largest part to the uncertainty, if the value of the proton charge radius is taken from elastic electron proton scattering (see for example table 2 of [5]). In muonic hydrogen the electron is replaced by a muon which possesses a \( \approx 200 \) times larger mass. The finite size term increases by more than 4 orders relative to the other terms and by about 7 orders of magnitude in absolute terms, as the Rydberg constant is proportional to the orbiting lepton mass as well. Therefore the measurement of transition frequencies in muonic hydrogen provides a much more sensitive method to determine the proton charge radius compared to regular hydrogen.

2. Determining the Parameters

The theoretical description of the hydrogen energy levels uses four parameters: The fine structure constant \( \alpha \), the electron to proton mass ratio \( m_e/m_p \), the proton charge radius \( r_p \) and the Rydberg constant \( R_\infty \). To determine their values one could use four measured transition frequencies in atomic hydrogen. However, other branches of physics may provide more precise values for some of them. This is the case for the fine structure constant which can be determined with a relative uncertainty of 3.7 parts in \( 10^{10} \) by comparing the measured value of the electron \( g \)-factor with the corresponding QED expression [6]. In contrast to eqn.(1) it is linear in \( \alpha \), so that the hydrogen energy levels are about 137 times less sensitive for the determination of \( \alpha \) than the electron \( g \)-factor.

The other parameter that is advantageously taken from another experiment, given the accuracy of currently available data, is the electron to proton mass ratio. Similar to the fine structure constant, the hydrogen energy levels are not sensitive to this quantity in first-order. On the other hand, mass ratios of particles that can simultaneously be held in a Penning trap can be determined with relative uncertainties in the \( 10^{-10} \) range by comparing their cyclotron frequencies [7]. In this way the electron proton mass ratio has been determined [8] with an uncertainty lower than can be extracted using eqn.(1). This leaves us with two parameters \( r_p \) and \( R_\infty \) to be determined from two measured hydrogen transition frequencies.

\(^1\) For that reason the additional parameters that appear there can be taken with a few digits accuracy from any other experiment.
The proton charge radius obtained from precision spectroscopy of atomic hydrogen. Either radio frequency measurements of the 2S-2P Lamb shift (violet) or optical transition frequencies (blue from the 2S and green from the 1S state) are used. To extract the two parameters, the Rydberg constant and the proton charge radius, each of these measurements needs to be combined with another independent measurement, which is the 1S-2S transition frequency here. The analysis reveals a 4σ discrepancy between the hydrogen mean value (H_{avg}) and the value determined from laser spectroscopy of muonic hydrogen (µ–p). An even larger inconsistency of 7σ is obtained when including proton-electron scattering data. This CODATA analysis also uses deuterium data, that have only limited effect on the proton charge radius, because its nucleus is not a proton. A similar picture is obtained when plotting the corresponding results for the Rydberg constant.

3. Testing QED

While one would use the two hydrogen transition frequencies with the lowest uncertainty and the largest sensitivity to determine the parameters, testing the internal consistency of QED requires more input data. For this purpose one can use different pairs of hydrogen transition frequencies to determine values for \( r_p \) and \( R_\infty \) and verify that all pairs give the same result within their respective uncertainty. The analysis is done by combining the 1S-2S transition frequency, which has by far the lowest uncertainty, with each of the other measured transition frequencies that are listed in table XI of the CODATA 2010 report [4]. In this way one obtains 15 values for \( R_\infty \) and \( r_p \) of which only the latter is shown in fig. 1a. To check for hidden systematic effects in these data the weighted mean quadratic deviation from the weighted mean is used. The square root of this quantity is known as the Birge ratio \( R_B \) which has an expectation value of 1 for Gaussian distributed data and a variance of \( 2/(N-1) \) where \( N = 15 \) is the number of data points. For the proton charge radii determined from hydrogen one finds \( R_B = 0.70(38) \), so that the scattering of the data does not provide evidence for systematic uncertainties. However, systematic effects common to all measured transition frequencies can not be ruled out in this way.

Even though this simple approach uses the same value for the 1S-2S transition frequency several times for averaging, the influence on the overall result is insignificant because the 1S-2S contribution to the uncertainty is essentially zero. The proper method for obtaining best estimates for \( r_p \) and \( R_\infty \) using hydrogen spectroscopy data is a least square adjustment. The result of \( r_p = 0.8764(89) \) fm with a Birge ratio of \( R_B = 0.96 \) is given as “adjustment #8” in table XXXVIII of the CODATA 2010 report [4].
Another value for the proton charge radius is obtained from the measured 2S-2P\(_{3/2}\) transition frequency in muonic hydrogen [2, 3]. Due to the enhanced sensitivity, moderate measurement accuracy and 5 of the 12 known digits of the Rydberg constant are sufficient to obtain the most precise value for the proton charge radius \(r_p = 0.84087(39)\) fm. This value differs from the “hydrogenic” value by 4 combined standard deviations.

An even larger discrepancy of 7 combined standard deviations exists between the “muonic” value (see fig. 1b) and recommended the CODATA value that includes a selection of results from elastic electron proton scattering (“adjustment #3”). In the past electron proton scattering has produced inconsistent results for \(r_p\). Depending on the evaluation method of the same data it may either support the “hydrogenic” [9] or the “muonic” [10] proton charge radius.

The discrepancy persists and QED is no longer consistent with experimental data, even when the electron proton scattering data are ignored. So the problem seems deeper and one may speculate about possible causes. Not only does it affect the value for \(r_p\) but likewise the value for \(R_\infty\). Nevertheless the problem has been referred to as the “proton size puzzle”. From the experimental side one may ask by how many standard deviations certain measurements need to be off the true value to explain the discrepancy. It turns out that the measured 1S-2S transition frequency would have to be wrong by about 4000 standard deviations while the measured 2S-2P transition frequencies in muonic hydrogen would have to be wrong by about 100 standard deviations. Unless a very large but yet unknown systematic effect is effective, it seems unlikely that an experimental problem there to be the cause the “proton size puzzle”.

4. Improving the Hydrogen Input Data

The weakest experimental data are provided by the transitions in regular hydrogen that start from the metastable 2S level. This is because these transitions have larger line widths and/or are much more sensitive to stray electric fields. Still, as evident from fig. 1a, most individual values do not contradict the “muonic” value. Only when averaging the hydrogen data the discrepancy arises. For this reason we are seeking to reduce the uncertainty of at least one of those transition frequencies significantly. Depending on the outcome this may solve or reinforce the “proton size puzzle”.

4.1. 2S-4P Transition

To generate a beam of metastable 2S atoms we are using two-photon laser excitation at 243 nm. Part of the experiment was used before to measure the 1S-2S transition frequency [1]. In all previous experiments that use the 2S as the ground state, electron impact excitation has been used to populate it. This method severely heats the atomic beam and even produces non-Maxwellian velocity distributions. It complicates a detailed line shape analysis which is essential to determine the line center much better than the line width. For the envisioned 2S-4P lines [11] we would like to find unperturbed line center within a few kHz in order to generate a value for the proton charge radius that can be decisive between the “muonic” and the “hydrogenic” value. This will require that we split the line better than parts in \(10^{-3}\) of the natural line width of 12.9 MHz.

Another complication of the line shape arises from the unresolved hyperfine structure of the P states. Previous experiments have relied on selecting the hyperfine components by laser polarization. Thanks to the well resolved hyperfine splitting of the 1S-2S transition we can now selectively populate the 2S\((F = 0)\) hyperfine state. As the \(F = 0 \rightarrow F = 0\) transition is strictly forbidden, independent of polarization, only one hyperfine component is left in the 2S-4P\(_{1/2}\) manifold. Likewise, for the 2S-4P\(_{3/2}\) transition only the \(F = 0 \rightarrow F = 1\) can be excited. Avoiding the electron beam also helps to reduce the dc Stark effect by introducing less charges into the system.
Figure 2. Left: An extended cavity laser diode at 972 nm [12], that is stabilized to a high finesse cavity (not shown) to obtain a 1 Hz line width [13], is frequency doubled twice (243 nm) to optically excite atomic hydrogen, that emerges from a 5.8 K nozzle, to the 2S state. A second similar laser is frequency doubled once (486 nm) and used to excite the 2S-4P transitions by illuminating the atomic beam at an angle as close as possible to 90°. An actively stabilized highly reflecting mirror ensures that the wavefronts of the beam coupled back into the fiber closely matches those of the forward wave. This suppresses the first-order Doppler effect to high degree. Right: Enlarged sketch of the Lyman-γ detector. Atoms in the 4P states quickly release a Lyman-γ photon at 97 nm. To obtain a large solid angle of detection, photons are not detected directly. Instead we detect photo-electron released from a cylindrical graphite coated surface. The photo-electrons are guided by electric fields to two channel electron multiplier (CEM) for detection. The atoms are shielded from the electric fields of the CEMs by a Faraday cage (not shown).

The apparatus that is sketched in fig. 2 uses two ultra stable diode lasers at 972 nm. One of them operates at a fixed frequency that is controlled by a frequency comb. Its 4-th harmonic runs collinearly with the atomic beam and drives the 1S-2S two-photon transition. From the previous 1S-2S experiments we know that the velocities of the resulting 2S atoms are essentially Maxwellian distributed with a temperature that is very close to 5.8 K which is the set temperature of the nozzle from which the 1S atoms emerge. The second laser at 972 nm is detuned by a about 2.3 GHz and scanned in frequency such that its second harmonic probes the 2S-4P transitions.

Transitions to the 4P_{1/2} and 4P_{3/2} levels are detected via their decay to the 1S ground state upon which they emit a Lyman-γ photon at 97 nm. These photons are detected via photo electrons that they generate inside two graphite coated cylinders. The photo electrons are guided to channel electron multipliers (CEM) with a suitable electric field configuration. With this detector we obtain a solid acceptance angle of almost 4π [14] while the large work function of graphite (≈ 5 eV) suppresses the background from the 486 nm laser.

Almost all of the transitions from the 2S states that contribute to the “hydrogenic” proton charge radius are two-photon transitions to another S state or a D state. A dipole allowed one photon transition is subject to another set of possible systematic shifts. Therefore it is most suitable to investigate the “proton size puzzle”. The largest among these systematic shifts is the first-order Doppler effect that is easily avoided in two-photon spectroscopy by using counter propagating laser beams. Similarly to the two-photon case we are using two counter propagating waves that cause opposite Doppler shifts which average to zero.
Taking the laser wavefront curvature into account the local Doppler shift is given by the projection of the velocity vector onto the normal of the local phase front. Therefore the back propagating wave has to match the forward wave in phase fronts and amplitude. To achieve this we couple the laser out of and back into the same single mode fiber with a waist at the flat high reflector whose tip and tilt are actively controlled (see fig. 2). At the same time we do not simply rely on perfectly matched counter propagating waves but measure the residual Doppler effect. To do that we employ time resolved detection: after turning off the 243 nm laser with a chopper we wait an adjustable time before opening the gate of the Lyman-γ detector. By this the fast atoms that could contribute to the signal have escaped and we are left with the slower ones within the Maxwell distribution. Observing the line center as a function of this delay we can set a very tight upper limit for the residual first-order Doppler effect.

We believe we have this effect under control and we are currently analysing other possible systematic shifts. Among them is the line pulling effect due to the quantum interference between different excitation and decay paths via the various Zeeman sublevels of the excited states [15]. As the emission pattern depends on these Zeeman sublevels, the observed line shift depends on the angle of the emitted photons relative to the polarization of the exciting laser. Recently this effect has been demonstrated experimentally with the D lines of Lithium [16]. This angle dependent line pulling effect exactly vanishes when collecting photons in all directions. With our Lyman-γ detector we hope to get close to this ideal situation.

4.2. 1S-3S/3D Transitions
Another transition that is under investigation in our lab for this purpose is the 1S-3S two-photon transition [17]. Its frequency has been measured by the group of F. Biraben with an uncertainty of 13 kHz [18]. Among the challenges here is the required wavelength of 205 nm that is difficult to produce as a continuous wave (cw). The only commercially available crystal that can generate it as a second harmonic is BBO. Not only does the required phase matching angle lie very close to 90° for that crystal, but it also suffers from photochemical reactions on the surfaces and from the photo-refractive effect. By cooling the crystal down to −10° the phase matching angle shifts by about 1° and effective nonlinearity increases [19].

To deal with the remaining crystal issues we employ a mode-locked laser rather than a cw laser whose spectrum can be described as a frequency comb source. It features a high peak intensity and therefore greater nonlinear interactions and does not induce the photo-refractive effect. Thus, a deep ultraviolet frequency comb can, in general, be generated more efficiently than cw radiation. The comb modes can be expressed as \( \omega_m = \omega_{ceo} + m\omega_r \), where \( \omega_{ceo} \) is the carrier envelope offset frequency, \( \omega_r \) the repetition rate of the laser and \( m \) an integer number. To drive a two-photon transition with a frequency comb the modes add pairwise to produce the transition energy \( h\omega_{eg} \). By tuning \( \omega_{ceo} \) such that a comb mode \( \omega_m \) corresponds to half the transition frequency, all mode pairs \( n \) which satisfy the relation \( \omega_{eg} = \omega_{m-n} + \omega_{m+n} \) contribute to the excitation rate. This is sketched for \( n = -1, 0, +1 \) at the left side of fig. 3. The same applies if the two-photon transition occurs exactly between two comb modes thus the signal repeats with half the repetition rate when scanning \( \omega_{ceo} \). For Fourier-limited pulses, all frequency modes are properly phased and the excitation paths add up coherently. In this way a pulsed laser drives a two-photon transition as efficiently as a cw laser with the same average intensity while the spectral line width is limited by the width of a single comb mode rather than by the spectral envelope of the comb [20].

This two-photon direct frequency comb spectroscopy is not exactly free of the first-order Doppler effect. In the frequency domain one would argue that the contributing mode pairs do not have the same frequency in the lab frame. However, for each pair \( n \) there is a pair \( -n \) such that the Doppler shift is balanced provided that the spectral envelopes of the counter propagating pulses are identical. This condition is closely matched by an enhancement cavity.
Figure 3. Left: spectral envelope of the frequency comb tuned to excite a two-photon transition at $\omega_{eg}$. One of the modes is picked to measure the detuning, of which only values modulo the pulse repetition rate $\omega_r$ are known without additional information. On resonance, pairwise addition of properly phased modes provides an efficient excitation of the atoms. Three such mode pairs are shown. Right: A mode-locked Ti:sapphire laser provides Fourier-limited pulses of 1.3 ps pulse duration at $\omega_r = 2\pi \times 82$ MHz repetition frequency with an average output power of 1.8 W. This laser is stabilized to a reference cavity using the radio-frequency side band technique. A double-pass AOM serves for tuning while a self referenced frequency comb is used to measure the frequency of one of the modes of the ps laser. After two resonant frequency doubling stages 10 to 15 mW of power at 205 nm becomes available for spectroscopy. A third cavity inside a vacuum chamber enhances this power to 100 mW for spectroscopy. An interferometer allows to adjust the polarization and the position of the pulse collision point where the Doppler free excitation of the 1S-3S transition takes place. Bare multimode fibers are used to guide the emitted Balmer-α light through an interference filter (IF) onto a photomultiplier tube (PMT).

In the time domain one would argue that the finite excitation region leads to a time of flight broadening. In both cases the first-order Doppler effect gives rise to a line broadening rather than to a line shift. At a thermal velocity of about 350 m/s (5.8 K) a pulse duration of 1 ps gives rise to a time of flight (or Doppler) broadening of 500 kHz. This pulse duration is therefore well adapted to the natural line width of 1 MHz of the 1S-3S transition. A possible frequency chirp of the pulses will reduce the excitation rate but does not induce a line shift [21]. Furthermore, the ac-Stark shift is determined by the average power rather than by the peak power [22].

The experimental realization, that is sketched at the right side of fig. 3, uses a ps Ti:sapphire mode-locked laser at 820 nm that is frequency doubled twice to 205 nm using two enhancement cavities [19]. A Mach-Zehnder type interferometer generates double pulses and allows to adjust their pulse collision point inside the enhancement cavity where the Doppler free excitation takes place. With this interferometer the polarisation of the colliding pulses may be adjusted to enhance or suppress either the 1S-3S or the 1S-3D components. The signal is obtained by detecting the Balmer-α photons due to the decay of the 3S/3D to the 2P state at 656 nm. As shown in fig. 3 these photons are collected with multimode fiber bundles whose end facets are brought into close proximity of the pulse collision point. Using several of these fiber bundles we obtain a solid angle of about 20% of $4\pi$.

Similar to the 2S-4P transitions we are currently investigating possible systematic frequency shifts. Besides the usual suspects we are also working out the effects of quantum interference.
4.3. Outlook

Two-photon direct frequency comb spectroscopy may be pushed to even shorter wavelengths. Pulsed lasers are readily up converted into the XUV spectral region and beyond by employing the process of high harmonic generation (HHG). With this it might become feasible to excite the 1S-2S transition in hydrogen like He\(^+\) at 60.8 nm [23]. Besides proving another independent QED test this system has several advantages: As the power series in eqn.(1) is written for hydrogen, \(\alpha\) has to be replaced by \(Z\alpha\) for hydrogen like systems with nuclear charge \(Z > 1\). Therefore He\(^+\) provides a more sensitive test of the higher order terms in particular for yet uncalculated terms of order \((Z\alpha)^6...7\). On the experimental side He\(^+\) holds the advantage that it can be held in an ion trap and sympathetically cooled by another ion with easy to generate cooling wavelength. This option is not yet available for hydrogen as the cooling wavelength of 121 nm is difficult to generate as a cw. The nuclear charge radius can be measured using muonic He\(^+\). Preparations for such an investigation are under way at the Paul Scherer Institute, Switzerland [24]. A frequency comb at 60.8 nm has already been generated in our lab [25] albeit with power levels that are not yet sufficient to readily perform this spectroscopy.

References

[1] Ch. G. Parthey et al., Phys. Rev. Lett. 107, 203001 (2011).
[2] R. Pohl et al., Nature 466, 213 (2010).
[3] A. Antognini et al., Science 339, 417 (2013).
[4] P.J. Mohr et al., Nature 466, 213 (2010).
[5] D. Hanneke, S. Fogwell, and G. Gabrielse, Phys. Rev. Lett. 100, 120801 (2008).
[6] S. Rainville, J.K. Thompson, and D.E. Pritchard, Science 339, 3598 (1995).
[7] J.C. Bernauer et al., Phys. Rev. Lett. 105, 242001 (2010).
[8] L.T. Lorenz, H.W. Hammer, and U.G. Meißner, Eur. Phys. J. A 48:151 (2012).
[9] A. Beyer et al., Ann. Phys. (Berlin), DOI 10.1002/andp.201300075
[10] N. Kolachevskyy et al., Opt. Lett. 36, 4299 (2011).
[11] J. Alnis, A. Matveev, N. Kolachevsky, Th. Udem, and T.W. Hänsch, Phys. Rev A, 77, 053809 (2008).
[12] A. Matveev et al., Phys. Rev. Lett. 110, 230801 (2013).
[13] M. Horbatsch and E.A. Hessels, Phys. Rev. A 82, 052519 (2010).
[14] R.C. Brown et al., Phys. Rev. A 87, 032504 (2013).
[15] E. Peters, D.C. Yost, A. Matveev, T.W. Hänsch, and Thomas Udem, Ann. Phys. 525, L29 (2013).
[16] O. Arnoult, F. Nez, L. Julien, and F. Biraben, Eur. Phys. J. D 60, 243 (2010).
[17] E. Peters et al., Opt. Express 17, 9183 (2009).
[18] Y.V. Baklanov and V.P. Chebotayev, Appl. Phys. 12, 97 (1977).
[19] S. Reinhardt, E. Peters, T.W. Hänsch, and Th. Udem, Phys. Rev. A 81, 033427 (2010).
[20] P. Fendel, S.D. Bergeson, Th. Udem, and T.W. Hänsch, Opt. Lett. 32, 701 (2007).
[21] M. Herrmann et al., Phys. Rev. A 79, 052505 (2009).
[22] A. Antognini et al., Can. J. Phys. 89, 47 (2011).
[23] A. Ozawa et al., Phys. Rev. Lett. 100, 253901 (2008).