Peer Review File

Manuscript Title: Comparison of the Antiproton-to-Proton q/m Ratios at 16 Parts per Trillion Precision

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

This work reports about a high-precision measurement of the ratio of the mass-to-charge-ratio of an antiproton and a negatively charged hydrogen atom.
After correction for atomic physics, the corresponding ratio for antiproton to proton is derived and as such, its measurement represents a test of CPT invariance. The experiment shows that this invariance principle holds, at the level of 1.6E-11.
The improvement compared to the previous best experiment is more than a factor of 4.

A second result of the team is a search for a sidereal variation of this ratio. No effect was found and this can be interpreted in terms of a test of Local Position Invariance (this is how such tests are conventionally referred to), of anomalous gravitational coupling of antimatter to matter, and of the standard model extension.

In an experiment like this, a large effort must be made to control systematic effects and to discern them in the data. The authors have done a thorough job, here. The supplemental material gives a detailed report of all aspects of data evaluation and systematic corrections applied.

Overall, this work is very rich and deep. That the improvement in the precision is "only" a factor 4 should not be seen as a weakness - these measurements are at the limit of feasibility and require enormous statistical averaging; this is a natural experimental limitation.
On the contrary, this factor represents a notable progress in the field.

Because of the fundamental importance of verifying symmetry principles in Physics, I believe that the result is of interest to a very wide audience.

The manuscript is well written, and in my opinion certainly deserves publication in a top journal such as Nature.

Detailed comments on the presentation.
=======================================

In Fig. 1, emphasize in the caption that (d) shows the sideband method and (e) the peak method. There is mention of "upper" and "lower" potential configuration, but in the figure the trap is oriented horizontal?

Line 38: perhaps related to the above: the magnet is oriented horizontally. Why is this information important?

The notation for the campaigns is not always consistent throughout the manuscript and supplemental material (SM). I suggest to stick to the naming in Tab. I, SM. (note that there, the "1" turned into a "I")

Just above Fig.3 "d Phi_E" is called the change of the gravitational potential, but later it is the amplitude of the modulation. That is not very consistent. In the axis label in Fig.3 the subscript E is missing.

I think it may be more appropriate to use the subscript "S" for "sun", since it is the solar gravitational potential that is of interest here (of course, at the location of the Earth), but the point of the authors is that it is not the galactic potential discussed earlier.

In line 156, it is not immediately apparent why a transportable trap should lead to better results. Maybe one can mention that then one will not be limited by the disturbances in the CERN experimental hall.

Delete lines 174-5

Line 188: typos in the paper title

SM, p.1. Ref. 6: from the title, it seems that that is vibrational, not rotational spectroscopy.

SM, p.2, Eq.(8). I may suggest to write "\sigma_Q(t_{avg})=....", since the AD is actually a function of the averaging time, and not of Q.
In that and in many more equations, "t_avg" should actually be "t_avg/(1 s)", right?

SM, p.2. Is the qualitfactor really only ~ 200? How does that match the eq.(8) for the Allan deviation?

SM, Tab. 3. Nomenclature "2019 SB-SB" does not agree with Tab. 1.

SM, Fig. 1: the coloured items are not described. "dB/dt" does not carry the right units. "D_\nu" in the right plot lacks the subscript of "\nu". The right plot is related to the left one. Still, the y-axis label has "\nu_+" instead of "\nu_c"? One issue that is recurring is the use of "time". In the right plot, the x-label should be "averaging time".

SM, p.4, after Fig. 3. In the text, there appears a "mUnit"?

p. 5, Fig.4: there appears a "Spine"

p.6, right col., the statement "the ... green data points are close to the mean of the upper and the lower ratio" is very difficult to understand. Are these ratios the red and blue curves? If so, then the improvement is not so convincing.

p.6, caption of Fig. 6: "t" ->"t_{avg}"

Fig.9, caption: delete an extra "with"

Fig. 10, right. What are the green lines? If they are the standard deviation of the shown data, I am not sure this makes sense, because this is not data, but the results of different analyses.

Fig. 10, left. The "blue" mentioned in the text below the figure is not really visible?

Fig. 11. It would be nice to see the centra dip expanded. Perhaps a logarithmic scale can be used for the frequency? And the difference data-fit shown in an additional plot below?

p.9, after eq.(28). "A" does not carry the right units.

p.14 bottom right. The statement of the accuracy makes little sense. As stated, that seems to indicate a typical drift rate of the clock frequency. That would not be "accuracy". Please check

p.15, col. 1: "if" -> "of"

Table 7, caption. Here the naming "dip method" is introduced. Please maintain consistency: is this the same as "sideband method"?

Referee #2 (Remarks to the Author):

Physical laws have primarily been established through experimentation with matter, and have then been extrapolated to antimatter using CPT symmetry. This symmetry essentially arises as a consequence of special relativity and quantum physics, and is thus inextricably linked to two cornerstones of physics. Experimentation with antimatter can therefore place the aforementioned extrapolation on solid observational grounds and at the same time probe the foundations of physics. Such efforts assume particular urgency in light of the observed baryon asymmetry in the universe and various theoretical ideas for physics beyond the SM and GR that can accommodate departures from CPT invariance. The present manuscript reports the best CPT test of a certain type for baryons and will therefore be met with definite interest by the community.

More specifically, the manuscript describes an experimental comparison of the charge-to-mass ratio for the proton and the antiproton. The basic idea behind this experiment involves single particles confined in a state-of-the-art cryogenic Penning-trap set-up. The possibility of high-sensitivity measurements of the particle's axial and modified cyclotron frequencies in such a set-up permits the extraction of its free cyclotron frequency. The comparison of free cyclotron frequencies of different particles in the same magnetic field then directly translates into an ultrahigh-precision comparison of their charge-to-mass ratios. The experiment utilizes antiprotons and H^{-} ions for the actual measurements. The H^{-} ion offers substantial benefits regarding trapping and is sufficiently well understood to serve as an experimentally advantageous substitute for the proton.

Relative to the previous best charge-to-mass ratio comparison, which was also performed by this group, the apparatus has undergone various upgrades to its cryogenic, shielding, and detection systems; an additional measurement methodology with enhanced sensitivity has been employed; and the data set has been enlarged allowing studies at different gravitational potentials. The upgrades have led to an overall gain in sensitivity by about a half an order of magnitude. This translates into an almost 5-fold improvement in the limit on the WEP-violating $|\alpha_g-1|$, and the construction and measurement of a differential WEP-violating $|\alpha_{g,D}-1|$, which may offer interpretational advantages. These WEP studies provide physical insight into the interaction of antimatter with gravity, a topic that---with a few notable exceptions---remains largely unexplored experimentally. Interpreted as a flat-space CPT test in effective field theory, these measurements also improve constraints on 27 Lorentz- and CPT-violating SME coefficients by factors up to 4. I regard these results as major achievements with the potential to attract a broad readership, and I endorse their publication in Nature subject to addressing a few minor comments:

(1) The font size in some figures (e.g. 1c, 1d, 1e, ...) as displayed in the main text is quite small.

(2) Figure 1c: What is PDF (a.u.)?

(3) Figures 2b and 2c plot different quantities, but the vertical axis in both plots carry the same label.

(4) The notation for the gravitational potential in Fig. 3a appears to be different from the text (see e.g. line 135)

(5) Line 148: What are ``DC-coefficients of the SME?"

Referee #3 (Remarks to the Author):

The manuscript addresses the fundamental problem of matter-antimatter asymmetry by a „direct“ measurement of the q/m ratio for proton and antiproton which is related to CPT invariance. Furthermore also WEP is considered in the long term studies of oscillation frequencies at various gravitational potentials. The key results are a remarkable reduction of the constrain for CPT violation in the proton-antiproton system and a low limit for WEP violation.

The manuscript is well written and it includes all details to follow the extraction of the results. The measurement procedure and data evaluation is carefully described and the improvement in precision due to upgrades in the apparatus is clearly demonstrated.
The use of H- for the proton measurement allows a comparison under identical conditions reducing systematic uncertainties and the consequential systematic error is in detail discussed in the supplementary material. In case of the WEP studies measurements during apsides was used for a comparison of different gravitational interactions.

The manuscript fits well into the antimatter studies which have been published in Nature up to now. In view of the previous Nature articles concerning the comparison of the proton/antiproton charge/mass ratio the manuscript will be of interest for the readership to follow the development in this field.

The manuscript is very well structured and should be published in the present form.
A small remark concerns figure 1d showing the dip spectra. For the single particle dip the resulting fit is shown but the corresponding data points are not given. Here also the data points should be included.
Apart from this I don’t see any necessary modification.

Author Rebuttals to Initial Comments:

Dear Editors of Nature,

we would like to thank the reviewers very much for all the time they've invested to review our manuscript 2020-02-03244B, for their positive evaluation and for their valuable comments. Please find below our point-by-point reply to their comments.

Sincerely

Stefan Ulmer for the BASE collaboration

Referee #1 (Remarks to the Author):

This work reports about a high-precision measurement of the ratio of the mass-to-charge-ratio of an antiproton and a negatively charged hydrogen atom.
After correction for atomic physics, the corresponding ratio for antiproton to proton is derived and as such, its measurement represents a test of CPT invariance. The experiment shows that this invariance principle holds, at the level of 1.6E-11.
The improvement compared to the previous best experiment is more than a factor of 4.
A second result of the team is a search for a sidereal variation of this ratio. No effect was found and this can be interpreted in terms of a test of Local Position Invariance (this is how such tests are conventionally referred to), of anomalous gravitational coupling of antimatter to matter, and of the standard model extension.
In an experiment like this, a large effort must be made to control systematic effects and to discern them in the data. The authors have done a thorough job, here. The supplemental material gives a detailed report of all aspects of data evaluation and systematic corrections applied.

Overall, this work is very rich and deep. That the improvement in the precision is "only" a factor 4 should not be seen as a weakness - these measurements are at the limit of feasibility and require

RIKEN
2-1 Hirosawa, Wako, Saitama, 351-0198, Japan
Tel: +81-48-462-1111
Fax: +81-48-462-1554
http://www.riken.jp/en/

enormous statistical averaging; this is a natural experimental limitation.
On the contrary, this factor represents a notable progress in the field.
Because of the fundamental importance of verifying symmetry principles in Physics, I believe that the result is of interest to a very wide audience.
The manuscript is well written, and in my opinion certainly deserves publication in a top journal such as Nature.

We would like to thank you very much for carefully reading our manuscript, for your positive evaluation of our work, and for all the very detailed and constructive comments, which improved to quality of our manuscript. Please find below a point-by-point reply to the made suggestions.

Detailed comments on the presentation.
======================================

In Fig. 1, emphasize in the caption that (d) shows the sideband method and (e) the peak method. There is mention of "upper" and "lower" potential configuration, but in the figure the trap is oriented horizontal?

Thank you very much for the comment. We have added to Fig. 1 b.) the sub-definitions "b.1" and "b.2" and modified the text to "we shuttle the particles between the potential configurations b.1 and b.2.". We also followed your suggestion to emphasize "sideband method" in (d) and "peak method" in (e). We have added labels to the plots and have modified the figure caption

Line 38: perhaps related to the above: the magnet is oriented horizontally. Why is this information important?

The horizontal orientation of the magnet is important for sensitivity to SME coefficients. In addition, in the accelerator hall the magnetic field changes imposed by the accelerator are vertical and a horizontal experiment is less sensitive to field changes imposed by the accelerator. We would like to keep this description as is, especially since most precision Penning trap experiments are operating magnets with vertical bores.

The notation for the campaigns is not always consistent throughout the manuscript and supplemental material (SM). I suggest to stick to the naming in Tab. I, SM. (note that there, the "1" turned into a "I")

We followed the suggestion and carefully checked the manuscript and the supplementary information for consistent notation. Thank you for spotting our mistakes.

Just above Fig.3 "d Phi_E" is called the change of the gravitational potential, but later it is the amplitude of the modulation. That is not very consistent. In the axis label in Fig.3 the subscript E is missing.

I think it may be more appropriate to use the subscript "S" for "sun", since it is the solar gravitational potential that is of interest here (of course, at the location of the Earth), but the point of the authors is that it is not the galactic potential discussed earlier.

Thank you for these two comments. As you suggest, we have consistently changed the index from “E” to “S”. We have updated the figure label and have defined “d Phi_S” as the amplitude of the gravitational potential.

In line 156, it is not immediately apparent why a transportable trap should lead to better results. Maybe one can mention that then one will not be limited by the disturbances in the CERN experimental hall.

As you suggest, we have modified the statement to “…and by the development of transportable antiproton traps, to move precision antiproton experiments from the accelerator environment to calm laboratory space, as anticipated by BASE-STEP.”

Delete lines 174-5

Sorry, our mistake, lines were deleted.

Line 188: typos in the paper title

Fixed

SM, p.1. Ref. 6: from the title, it seems that that is vibrational, not rotational spectroscopy.

This has been corrected as you suggest

SM, p.2, Eq.(8). I may suggest to write "\sigma_Q(t_{avg})=....", since the AD is actually a function of the averaging time, and not of Q.
In that and in many more equations, "t_avg" should actually be "t_avg/(1 s)", right?

Modified as suggested, thank you.

SM, p.2. Is the quality factor really only ~ 200? How does that match the eq.(8) for the Allan deviation?

The quality factors of the AXIAL detectors are described by Eq. (8), and are summarized in table 2 to the supplementary material. The quality factor of the cyclotron resonator is as stated in the text. Because of the low intrinsic inductance L quality factors of cyclotron detectors are systematically smaller than those of the superconducting axial detectors.

SM, Tab. 3. Nomenclature "2019 SB-SB" does not agree with Tab. 1.

RIKEN
2-1 Hirosawa, Wako, Saitama, 351-0198, Japan
Tel: +81-48-462-1111
Fax: +81-48-462-1554
http://www.riken.jp/en/

Thank you, this has been changed

SM, Fig. 1: the coloured items are not described. "dB/dt" does not carry the right units. "D_\nu" in the right plot lacks the subscript of "\nu". The right plot is related to the left one. Still, the y-axis label has "\nu_+" instead of "\nu_c"? One issue that is recurring is the use of "time". In the right plot, the x-label should be "averaging time".

Sorry for causing confusion, in our daily laboratory language "magnetic field", "cyclotron frequency", and "modified cyclotron frequency" are in some sense equivalent. We have modified the figure as you suggest and extended the caption describing the coloured items.

SM, p.4, after Fig. 3. In the text, there appears a "mUnit"?

The trap tuning ratio V_ce/V_r is a dimensionless value, typical changes in C_4 and C_6 are optimized in steps 0.001 = 1mUnit, which is the common practical Unit in all precision Penning trap experiments.

p. 5, Fig.4: there appears a "Spine"

Thank you, the figure has been updated.

p.6, right col., the statement "the ... green data points are close to the mean of the upper and the lower ratio" is very difficult to understand. Are these ratios the red and blue curves? If so, then the improvement is not so convincing.

Thank you for the comment. A result of a much more detailed statistical analysis of the full data-set shows that the polynomial method efficiently suppresses non-linear contributions. A manuscript comparing the different evaluation approaches is in preparation. Still the qualitative result displayed by the plot is and instructive representation of removed non-linearities. For our experiment, the present drifts and the data-sampling interval the spline approach and the polynomial fit approach lead, within the assigned systematic uncertainties, to consistent results.

p.6, caption of Fig. 6: "t" ->"t_{avg}"

Thank you, changed.

Fig.9, caption: delete an extra "with"

Changed as suggested

Fig. 10, right. What are the green lines? If they are the standard deviation of the shown data, I am not sure this makes sense, because this is not data, but the results of different analyses.

Fig. 10, left. The "blue" mentioned in the text below the figure is not really visible?

RIKEN
2-1 Hirosawa, Wako, Saitama, 351-0198, Japan
Tel: +81-48-462-1111
Fax: +81-48-462-1554
http://www.riken.jp/en/

Fig. 11. It would be nice to see the centra dip expanded. Perhaps a logarithmic scale can be used for the frequency? And the difference data-fit shown in an additional plot below?

p.9, after eq.(28). "A" does not carry the right units.

Since marginal in the final analysis, this entire section has been removed in the updated version of the resubmitted manuscript.

p.14 bottom right. The statement of the accuracy makes little sense. As stated, that seems to indicate a typical drift rate of the clock frequency. That would not be "accuracy". Please chec

We have modified this statement to

“The used FFT analyzer shows a weak scaling of measured frequency ν_{exp} as a function of reference window center ν_{ref} with a slope of $\nu_\text{ref}-\nu_\text{exp}=0.2574(32)\,\mu$Hz/Hz. The resulting frequency offsets are being corrected. For the peak run the related uncertainty is at a level of $3.69\,$p.p.t$.$, accounted to the uncertainty budget, for the sideband runs the uncertainty of 1.24$\,$p.p.t$.$ is absorbed in the line-shape correction. Both quoted uncertainties contain the calibration as well as the intrinsic short term stability of the used Rb clock. ”

p.15, col. 1: "if" -> "of"

Changed

Table 7, caption. Here the naming "dip method" is introduced. Please maintain consistency: is this the same as "sideband method"?

As you say, this is indeed the sideband method, we have changed the caption of the table as you propose.

Referee #2 (Remarks to the Author):

Physical laws have primarily been established through experimentation with matter, and have then been extrapolated to antimatter using CPT symmetry. This symmetry essentially arises as a consequence of special relativity and quantum physics, and is thus inextricably linked to two cornerstones of physics. Experimentation with antimatter can therefore place the aforementioned extrapolation on solid observational grounds and at the same time probe the foundations of physics. Such efforts assume particular urgency in light of the observed baryon asymmetry in the universe and various theoretical ideas for physics beyond the SM and GR that can accommodate departures from CPT invariance. The present manuscript reports the best CPT test of a certain type for baryons and will therefore be met with definite interest by the community.

More specifically, the manuscript describes an experimental comparison of the charge-to-mass ratio for the proton and the antiproton. The basic idea behind this experiment involves single particles confined in a state-of-the-art cryogenic Penning-trap set-up. The possibility of high-sensitivity measurements of the particle's axial and modified cyclotron frequencies in such a set-up permits the extraction of its free cyclotron frequency. The comparison of free cyclotron frequencies of different particles in the same magnetic field then directly translates into an ultrahigh-precision comparison of their charge-to-mass ratios. The experiment utilizes antiprotons and H^{-} ions for the actual measurements. The H^{-} ion offers substantial benefits regarding trapping and is sufficiently well understood to serve as an experimentally advantageous substitute for the proton.

Relative to the previous best charge-to-mass ratio comparison, which was also performed by this group, the apparatus has undergone various upgrades to its cryogenic, shielding, and detection systems; an additional measurement methodology with enhanced sensitivity has been employed; and the data set has been enlarged allowing studies at different gravitational potentials. The upgrades have led to an overall gain in sensitivity by about a half an order of magnitude. This translates into an almost 5-fold improvement in the limit on the WEP-violating $|\alpha_g-1|$, and the construction and measurement of a differential WEP-violating $|\alpha_{g,D}-1|$, which may offer interpretational advantages. These WEP studies provide physical insight into the interaction of antimatter with gravity, a topic that---with a few notable exceptions---remains largely unexplored experimentally. Interpreted as a flat-space CPT test in effective field theory, these measurements also improve constraints on
27 Lorentz- and CPT-violating SME coefficients by factors up to 4. I regard these results as major achievements with the potential to attract a broad readership, and I endorse their publication in Nature subject to addressing a few minor comments:

We would like to thank you very much for taking your precious time to review our article, for your very positive comments and your suggestions to improve our article. Please find below point-by-pointe replies to your comments and questions.

(1) The font size in some figures (e.g. 1c, 1d, 1e, ...) as displayed in the main text is quite small.

As suggested, we have increased all the font sizes in Figures 1c, 1d, 1e.

(2) Figure 1c: What is PDF (a.u.)?

PDF stands for probability density function. To clarify we have replaced “PDF” by “fraction of total counts”

(3) Figures 2b and 2c plot different quantities, but the vertical axis in both plots carry the same label.

Thank you very much for this important comment. Fig. 2b displays the experiment fluctuation “sigma(R)/R” for individual sample lengths of 1.5 h. Fig 2c displays the statistical frequency ratio uncertainty \Delta R_{stat}/R as a function of averaging time. We have modified the figure and

RIKEN
2-1 Hirosawa, Wako, Saitama, 351-0198, Japan
Tel: +81-48-462-1111
Fax: +81-48-462-1554
http://www.riken.jp/en/

extended the figure caption writing "…b.) experiment stability $\sigma(R)/R$ as a function of time, each data-point contains and individual data-window of 1.5$\,$h. c.) statistical frequency ratio precision $\Delta R_{\text{stat}}/R$ as a function of averaged measurements…."

(4) The notation for the gravitational potential in Fig. 3a appears to be different from the text (see e.g. line 135)

Thank you. We have modified the text to "dSigma_{S}" and "Sigma_{S,avg}" and have updated the Figure accordingly.

(5) Line 148: What are ``DC-coefficients of the SME?"

We define as "DC" coefficients the ones that lead to constant stationary shifts in measured fundamental constants. The "AC" coefficients add time dependent phenomena, such as e.g. an oscillation of the antiproton-to-proton charge-to-mass ratio. We have removed the expression "DC coefficient".

Referee #3 (Remarks to the Author):

The manuscript addresses the fundamental problem of matter-antimatter asymmetry by a „direct" measurement of the q/m ratio for proton and antiproton which is related to CPT invariance. Furthermore also WEP is considered in the long term studies of oscillation frequencies at various gravitational potentials. The key results are a remarkable reduction of the constrain for CPT violation in the proton-antiproton system and a low limit for WEP violation.
The manuscript is well written and it includes all details to follow the extraction of the results. The measurement procedure and data evaluation is carefully described and the improvement in precision due to upgrades in the apparatus is clearly demonstrated.
The use of H- for the proton measurement allows a comparison under identical conditions reducing systematic uncertainties and the consequential systematic error is in detail discussed in the supplementary material. In case of the WEP studies measurements during apsides was used for a comparison of different gravitational interactions.
The manuscript fits well into the antimatter studies which have been published in Nature up to now. In view of the previous Nature articles concerning the comparison of the proton/antiproton charge/mass ratio the manuscript will be of interest for the readership to follow the development in this field.
The manuscript is very well structured and should be published in the present form.
A small remark concerns figure 1d showing the dip spectra. For the single particle dip the resulting fit is shown but the corresponding data points are not given. Here also the data points should be included.

Apart from this I don't see any necessary modification.

RIKEN
2-1 Hirosawa, Wako, Saitama, 351-0198, Japan
Tel: +81-48-462-1111
Fax: +81-48-462-1554
http://www.riken.jp/en/

We would like to thank you very much for taking your precious time to review our article, for your very positive comments and your suggestions to improve our article. Thank you very much for the comment on Fig. 1, and for your careful reading, during one of the last Figure edits the single dip data points were mistakenly deleted. We have added the single-dip data again and have updated the Figure as you suggest.

In summary, we would like to thank you, Federico, and all the reviewers again, for all the time and work in context of the evaluation of our manuscript, and for the very positive evaluation.

Sincerely,

Stefan Ulmer, for the BASE collaboration

RIKEN
2-1 Hirosawa, Wako, Saitama, 351-0198, Japan
Tel: +81-48-462-1111
Fax: +81-48-462-1554
http://www.riken.jp/en/