Leading hadronic contribution to the muon magnetic moment from lattice QCD

Sz. Borsanyi[1], Z. Fodor[1,2,3,4,5,*], J. N. Guenther[6], C. Hoelbling[1], S. D. Katz[3], L. Lellouch[7], T. Lippert[1,2], K. Miura[7,8,9], L. Parato[7], K. K. Szabo[1,2], F. Stokes[2], B. C. Toth[1], Cs. Torok[2], L. Varnhorst[1]

[1] Department of Physics, University of Wuppertal, D-42119 Wuppertal, Germany

[2] Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52428 Jülich, Germany

[3] Institute for Theoretical Physics, Eötvös University, H-1117 Budapest, Hungary

[4] University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

[5] Pennsylvania State University, Department of Physics, State College, PA 16801, USA

[6] Department of Physics, University of Regensburg, Regensburg D-93053, Germany

[7] Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France

[8] Helmholtz Institute Mainz, D-55099 Mainz, Germany

[9] Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, Nagoya University, Nagoya 464-8602, Japan

The standard model of particle physics describes the vast majority of experiments and observations involving elementary particles. Any deviation from its predictions would be a sign of new, fundamental physics. One long-standing discrepancy concerns the anomalous magnetic moment of the muon, $(g_\mu - 2)$, a measure of the magnetic field surrounding that particle. Indeed, standard model predictions for $(g_\mu - 2)$, reviewed in [1], exhibit disagreement with the measurement [2] that is tightly scattered around 3.7 standard deviations. Today, theory and measurement errors are comparable. However, a new experiment is underway at Fermilab and another is planned at J-PARC, both aiming to reduce the measurement's error by a factor of four. On the theory side, the dominant source of error is the leading-order, hadronic vacuum polarization (LO-HVP) contribution. To fully leverage the upcoming measurements, it is critical to check the prediction for this contribution with independent methods and to reduce its uncertainties. The most precise, model-independent determinations currently rely on dispersive techniques, combined with measurements of the cross-section for electron-positron annihilation into hadrons [3–6]. Here we use *ab initio* simulations in quantum chromodynamics and quantum electrodynamics to compute the LO-HVP contribution with sufficient precision to discriminate between the measurement of $(g_\mu - 2)$ and the dispersive predictions. Surprisingly our result, $[(g_\mu - 2)/2]_{\mathrm{LO-HVP}} = 708.7(5.3) \times 10^{-10}$, eliminates the need to invoke new physics to explain the measurement of $(g_\mu - 2)$. Moreover, the methods used and developed here will allow further increases in precision, as more powerful computers become available.

The muon is an ephemeral sibling of the electron. It is 207 times more massive, but has the same electric charge and spin. Similarly to the electron, it behaves like a tiny magnet, characterized by a magnetic moment. This quantity is proportional to the spin and charge of the muon, and inversely proportional to twice its mass. Dirac's relativistic quantum mechanics predicts that the constant of proportionality, g_μ, should be 2. However, in a relativistic quantum field theory such as the standard model, this prediction receives small corrections due to quantum, vacuum fluctuations. These corrections are called the anomalous magnetic moment and are quantified by $(g_\mu - 2)/2$. They were measured to an exquisite 0.54 ppm at the Brookhaven National Laboratory in the early 2000s [2], and have been calculated with a comparable precision (see [7] for a recent review).

At this level of precision, all of the interactions of the standard model contribute. The leading contributions are electromagnetic and described by quantum electrodynamics (QED), but the one that dominates the theory error is induced by the strong interaction and requires solving the highly non-linear

2

equations of quantum chromodynamics (QCD) at low energies. This contribution is determined by the leading-order, hadronic vacuum polarization (LO-HVP), which describes how the propagation of a virtual photon is modified by the presence of quark and gluon fluctuations in the vacuum. Here we compute this LO-HVP contribution to $(g_\mu - 2)/2$, denoted by $a_\mu^{\mathrm{LO-HVP}}$, using *ab initio* simulations in QCD and QED.

QCD is a generalized version of QED. The Euclidean Lagrangian for this theory is $\mathcal{L} = 1/(4e^2)F_{\mu\nu}F_{\mu\nu} + 1/(2g^2)\mathrm{Tr}G_{\mu\nu}G_{\mu\nu} + \sum_f \bar{\psi}_f[\gamma_\mu(\partial_\mu + iq_f A_\mu + iB_\mu) + m_f]\psi_f$, where γ_μ are the Dirac-matrices, f runs over the flavors of quarks, the m_f are their masses and the q_f are their charges in units of the electron charge e. Moreover, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and $G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + [B_\mu, B_\nu]$ and g is the QCD coupling constant. In electrodynamics, the gauge potential A_μ is a real valued field, whereas in QCD, B_μ is a 3×3 Hermitian matrix field. The different "flavors" of quarks are represented by independent fermionic fields, ψ_f. These fields have an additional "color" index in QCD, which runs from 1 to 3. In the present work, we include both QED and QCD, as well as four non-degenerate quark flavors (up, down, strange and charm), in a lattice formulation taking into account all dynamical effects. We also consider the tiny contributions of the bottom and top quarks, as discussed in the Supplementary Information.

We compute $a_\mu^{\mathrm{LO-HVP}}$ in the so-called time-momentum representation [8], which relies on the following, zero three-momentum, two-point function in Euclidean time t:

$$G(t) = \frac{1}{3} \sum_{\mu=1,2,3} \int d^3x \langle J_\mu(\vec{x}, t) J_\mu(0) \rangle , \tag{1}$$

where eJ_μ is the quark electromagnetic current with $J_\mu = \frac{2}{3}\bar{u}\gamma_\mu u - \frac{1}{3}\bar{d}\gamma_\mu d - \frac{1}{3}\bar{s}\gamma_\mu s + \frac{2}{3}\bar{c}\gamma_\mu c$. u, d, s and c are the up, down, strange and charm quark fields and the angle brackets stand for the QCD+QED expectation value to order e^2. It is convenient to decompose $G(t)$ into light, strange, charm and disconnected components, which have very different statistical and systematic uncertainties. Integrating the one-photon-irreducible ($1\gamma\mathrm{I}$) part of the two-point function (1) yields the LO-HVP contribution to the magnetic moment of the muon [8–11]:

$$a_\mu^{\mathrm{LO-HVP}} = \alpha^2 \int_0^\infty dt\, K(t)\, G_{1\gamma\mathrm{I}}(t) , \tag{2}$$

with the weight function,

$$K(t) = \int_0^\infty \frac{dQ^2}{m_\mu^2}\, \omega\left(\frac{Q^2}{m_\mu^2}\right) \left[t^2 - \frac{4}{Q^2}\sin^2\left(\frac{Qt}{2}\right)\right] , \tag{3}$$

3

and where $\omega(r) = [r + 2 - \sqrt{r(r+4)}]^2/\sqrt{r(r+4)}$, α is the fine structure constant in the Thomson limit and m_μ is the muon mass. Since we consider only the LO-HVP contribution, for brevity we drop the superscript and multiply the result by 10^{10}, ie. a_μ stands for $a_\mu^{\mathrm{LO-HVP}} \times 10^{10}$ throughout this work.

The subpercent precision, that we are aiming for, represents a huge challenge for lattice QCD. To reach that goal, we have to address four critical issues: A. scale determination; B. noise reduction; C. QED and strong-isospin breaking; D. infinite-volume and continuum extrapolations. We discuss these one by one.

ad A. The quantity a_μ depends on the muon mass. When computing (2) on the lattice, m_μ has to be converted into lattice units, am_μ, where a is the lattice spacing. A relative error of the lattice spacing propagates into about a twice as large a relative error on a_μ, so that a has to be determined with a few permil precision. We use the mass of the Ω baryon, $M_\Omega = 1672.45(29)$ MeV [1], to set the lattice spacing. We also use the w_0-scale from [12], in order to define an isospin decomposition of our observables. Though w_0 can be measured with sub-permil precision on the lattice, it is inaccessible experimentally. In this work we determine the physical value of w_0 including QED and strong-isospin-breaking effects: $w_{0,*} = 0.17236(29)(63)[70]$ fm, where the first error is statistical, the second is systematic and the third is the total error. In total we reach a relative accuracy of four permil, which is better than the error of the previous best determination of [13], whose value agrees with ours. There the pion-decay-constant was used as experimental input and the isospin-breaking effects were only included as an estimate.

ad B. Our result for a_μ is obtained as an integral over the conserved current-current correlation function, from zero to infinite time separation, as shown in Equation (2). For large separations the correlator is quite noisy. This noise manifests itself as a statistical error in a_μ. To reach the desired accuracy on a_μ, one needs high-precision at every step. Over 20,000 configurations were accumulated for our 27 ensembles on $L \approx 6$ fm lattices. In addition, we also include a lattice with $L \approx 11$ fm (see point D). The most important improvement over our earlier a_μ determination in [14] is the extensive use of analysis techniques based on the lowest eigenmodes of the Dirac operator, see eg. [15–18]. About an order of magnitude accuracy-gain can be reached using this technique for a_μ [19, 20].

ad C. The precision needed cannot be reached with pure, isopin-symmetric QCD. Thus, we include QED effects and allow the up and down quarks to have different masses. These effects are included both in the scale determination and in the current-current correlators. Note that the separation of isospin symmetric and isospin breaking contributions requires a convention, which we discuss in detail in the Supplementary Information. Strong-isospin breaking is implemented by taking derivatives of QCD+QED

4

expectation values with respect to up/down quark masses and computing the resulting observables on isospin-symmetric configurations [21]. Note that the first derivative of the fermionic determinant vanishes. We also implement derivatives with respect to the electric charge [22]. It is useful to distinguish between the electric charge in the fermionic determinant, e_s or sea electric-charge, and in the observables, e_v or valence electric-charge. The complete list of graphs that should be evaluated are shown in Figure 1 with our numerical results for them.

The final observable is given as a Taylor-expansion around the isospin-symmetric, physical-mass point with zero sea and valence charges. Instead of the quark masses, we use the pseudoscalar meson masses of pions and kaons, which can be determined with high precision. With the expansion coefficients, we extrapolate in the charges, in the strong-isospin breaking parameter and in the lattice spacing and interpolate in the quark masses to the phyiscal point. Thus, we obtain a_μ and its statistical and systematic uncertainties.

ad D. The standard wisdom for lattice calculations is that $M_\pi L > 4$ should be taken, where M_π is the mass of the pion and L is the spatial extent of the lattice. Unfortunately, this is not satisfactory in the present case: a_μ is far more sensitive to L than other quantities, such as hadron masses, and large volumes are needed to reach permil accuracy. For less volume-sensitive quantities in this work, we use well-established results to determine the finite-volume corrections on the pion-decay constant [23] and on charged hadron masses [24–26]. Leading-order chiral perturbation theory [27] or two-loop, partially-quenched chiral perturbation theory [20, 28] for a_μ help, but the non-perturbative, leading-order, large-L expansion of [29] indicates that those approaches still lead to systematic effects which are larger than the accuracy that we are aiming for. In addition to the infinite-volume extrapolation, the continuum extrapolation is also difficult. This is connected to the taste-symmetry breaking of staggered fermions, which we use in this work.

We determine the finite-volume corrections and/or improve the continuum extrapolations by three means: i. we work out the full two-loop, finite-volume, staggered chiral perturbation theory corrections for a_μ; ii. we apply a modified Lellouch-Lüscher-Gounaris-Sakurai technique to calculate the necessary contributions; iii. most importantly we carry out a full lattice simulation on an $L \approx 11$ fm lattice, with highly-suppressed taste violations and with physical, taste-averaged pion masses.

Combining all of these ingredients we obtain, as a final result, $a_\mu = 708.7(2.8)(4.5)[5.3]$. The first, statistical error comes mostly from the noisy, large-distance region of the current-current correlator. The second, systematic error is dominated by the continuum extrapolation and the finite-size effect computa-

tion. The third, total error is obtained by adding the first two in quadrature. In total we reach a relative accuracy of 0.7%. In Figure 2 we show the continuum extrapolation of the light, connected component of the isospin symmetric part of a_μ, which gives the dominant contribution to a_μ.

Figure 3 compares our result with previous lattice computations and also with results from the R-ratio method. In principle, one can reduce the uncertainty of our result by combining our lattice correlator, $G(t)$, with the one obtained from the R-ratio method, in regions of Euclidean time where the latter is more precise [19]. We do not do so here because there is a tension between our result and those obtained by the R-ratio method, as can be seen in Figure 3. For the total, LO-HVP contribution to a_μ, our result is 2.2σ, 2.7σ and 2.6σ larger than the R-ratio results of $a_\mu = 694.0(4.0)$ [3], $a_\mu = 692.78(2.42)$ [4] and $a_\mu = 692.3(3.3)$ [5, 6], respectively. This tension requires further investigation. Note that all three R-ratio determinations are based on the same experimental data set and thus they are strongly correlated.

As a first step in that direction, it is instructive to consider a modified observable, where the correlator $G(t)$ is restricted to a finite interval by a smooth window function [19]. This observable, which we denote by $a_{\mu,\mathrm{win}}$, is obtained much more readily than a_μ on the lattice. Its shorter-distance nature makes it significantly less susceptible to statistical noise and to finite-volume effects. Moreover, in the case of staggered fermions, it has reduced discretization artefacts. This is shown in Figure 4, where the light, connected component of $a_{\mu,\mathrm{win}}$ is plotted as a function of a^2. Because the determination of this quantity does not require overcoming many of the challenges described above, other lattice groups have obtained it with errors comparable to ours [19, 20]. This allows for a sharper benchmarking of our calculation of this challenging, light-quark contribution that dominates a_μ. On average, our $a_{\mu,\mathrm{win}}^{\mathrm{light}}$ is only 1.2σ away from those of [19, 20]. Moreover, $a_{\mu,\mathrm{win}}$ can be computed in the R-ratio approach, and we do so using the data set courteously given to us by the authors of [4]. However, here we find a 3.4σ tension with our lattice result.

To conclude, when combined with the other standard model contributions (see eg. [3, 4]), our result for the leading-order hadronic contribution to the anomalous magnetic moment of the muon, $a_\mu^{\mathrm{LO-HVP}} = 708.7(5.3) \times 10^{-10}$, eliminates the longstanding discrepancy between experiment and theory. However, as discussed above and can be seen in Figure 3, our lattice result exhibits a tension with the R-ratio determinations of [3–6]. Obviously, our findings should be confirmed –or refuted– by other collaborations using other discretizations of QCD. Those investigations are underway.

References

1. Tanabashi, M. *et al.* Review of Particle Physics. *Phys. Rev.* **D98,** 030001 (2018).

2. Bennett, G. W. *et al.* Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL. *Phys. Rev.* **D73,** 072003 (2006).

3. Davier, M., Hoecker, A., Malaescu, B. & Zhang, Z. A new evaluation of the hadronic vacuum polarisation contributions to the muon anomalous magnetic moment and to $\alpha(m_Z^2)$. *Eur. Phys. J. C* **80,** 241 (2020).

4. Keshavarzi, A., Nomura, D. & Teubner, T. $g-2$ of charged leptons, $\alpha(M_Z^2)$, and the hyperfine splitting of muonium. *Phys. Rev.* **D101,** 014029 (2020).

5. Colangelo, G., Hoferichter, M. & Stoffer, P. Two-pion contribution to hadronic vacuum polarization. *JHEP* **02,** 006 (2019).

6. Hoferichter, M., Hoid, B.-L. & Kubis, B. Three-pion contribution to hadronic vacuum polarization. *JHEP* **08,** 137 (2019).

7. Aoyama, T. *et al.* The anomalous magnetic moment of the muon in the Standard Model. arXiv: 2006.04822 [hep-ph] (June 2020).
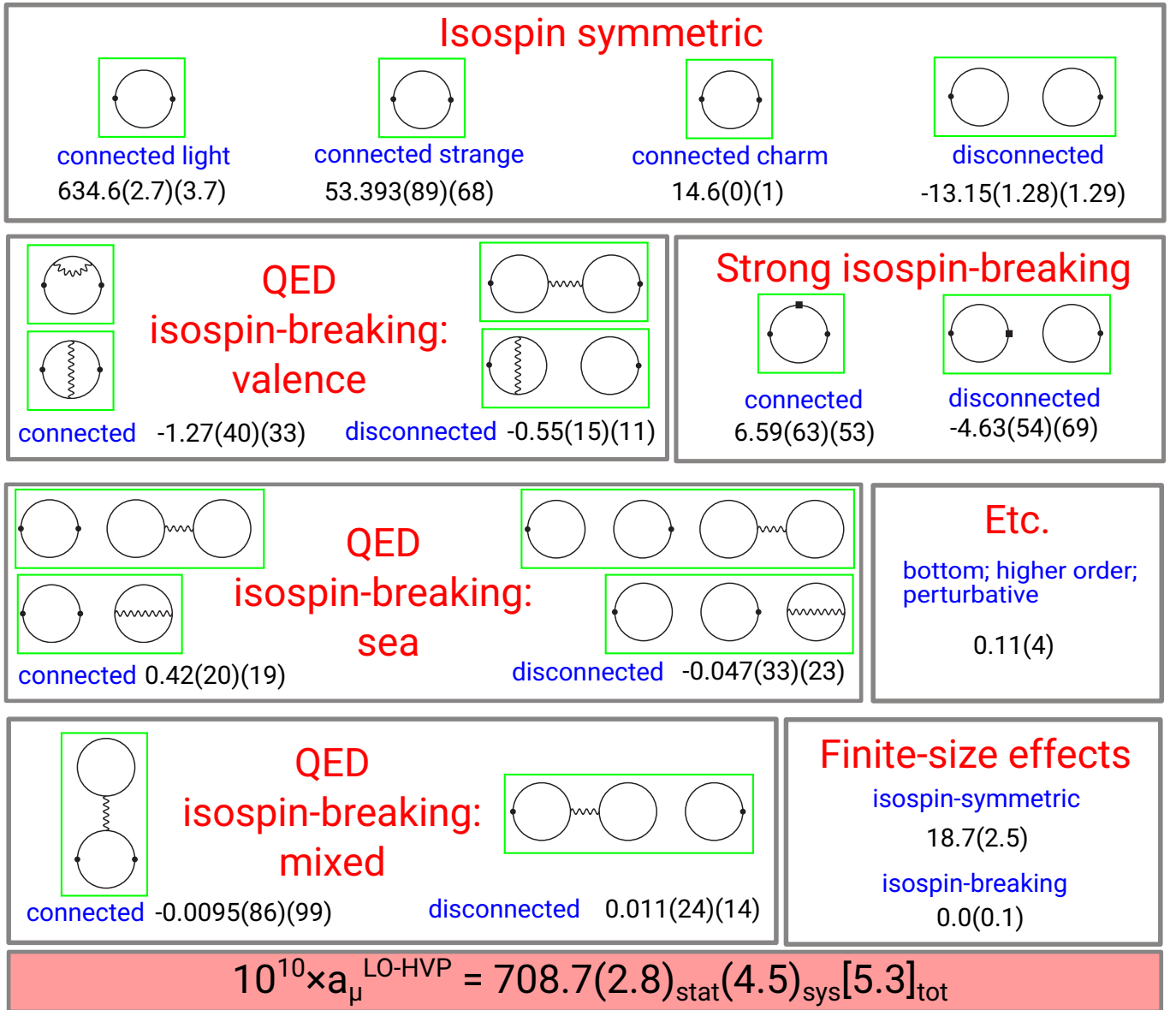
Figure 1: List of the contributions to a_μ, including examples of the corresponding Feynman diagrams. Solid lines are quarks and curly lines are photons. Gluons are not shown explicitly, and internal quark loops, only if they are attached to photons. Dots represent coordinates in position space, a box indicates the mass insertion relevant for strong-isospin breaking. The numbers give our result for each contribution, they correspond to our "reference" system size given by $L_{\rm ref} = 6.272$ fm spatial and $T_{\rm ref} = 9.408$ fm temporal lattice extents. We also explicitly compute the finite-size corrections that must be added to these results, these are given separately in the lower right panel. The first error is the statistical and the second is the systematic uncertainty; except for the contributions where only a single, total error is given.
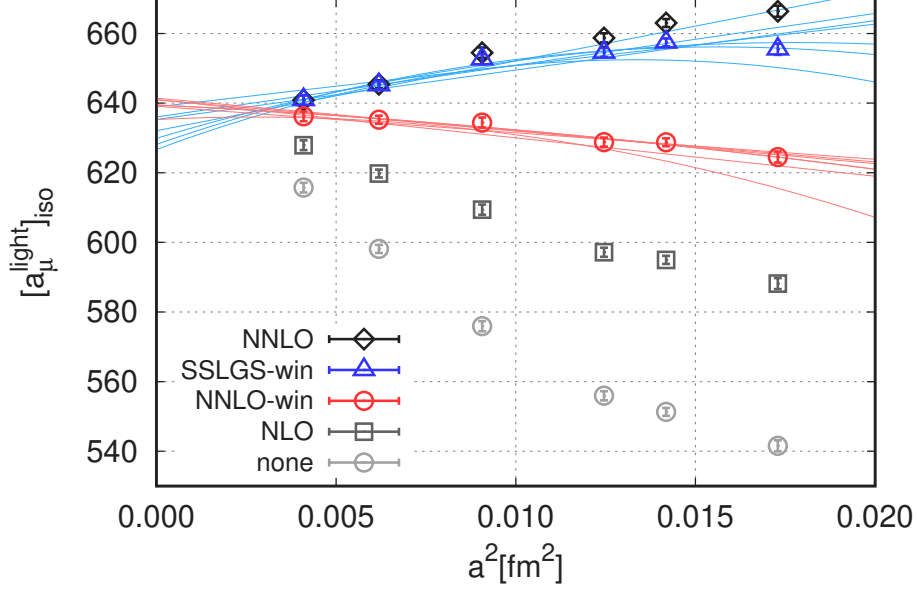
Figure 2: Continuum extrapolation of the isospin-symmetric light connected component of a_μ, denoted by $[a_\mu^{\text{light}}]_{\text{iso}}$. The data points are obtained on lattices of sizes $L \approx 6$ fm. The different colors/symbols correspond to different types of improvement procedures: "none" stands for applying no improvement; "NLO" and "NNLO" refer to improvements based on the next-to-leading and the next-to-next-to-leading orders of finite-volume, staggered chiral perturbation theory; "SLLGS" is an approach based on experimental input parameterized by a Gounaris-Sakurai model combined with the Lellouch-Lüscher formalism (see the Supplementary Information for details). The two methods labeled with 'win' are used to obtain the final results of the paper. The lines show fits using linear and quadratic fits in a^2 with varying number of lattice spacings in the fit.
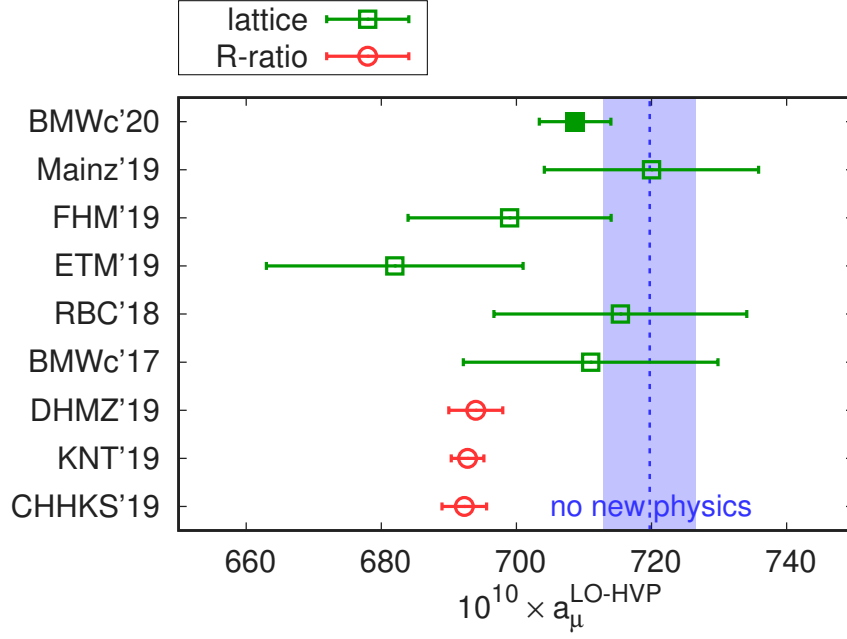
Figure 3: Comparison of recent results for the leading-order, hadronic vacuum polarization contribution to the anomalous magnetic moment of the muon (see [7] for a recent review). Green squares are lattice results: this work's result, denoted by BMWc'20 and represented by a filled symbol at the top of the figure, is followed by Mainz'19 [30], FHM'19 [31], ETM'19 [32], RBC'18 [19] and our earlier work BMWc'17 [14]. Red circles are obtained using the R-ratio method from DHMZ'19 [3], KNT'19 [4] and CHHKS'19 [5, 6]; these results use the same experimental data as input. The blue shaded region is the value that $a_\mu^{\mathrm{LO-HVP}}$ would have to have to explain the experimental measurement of $(g_\mu - 2)$, assuming no new physics.
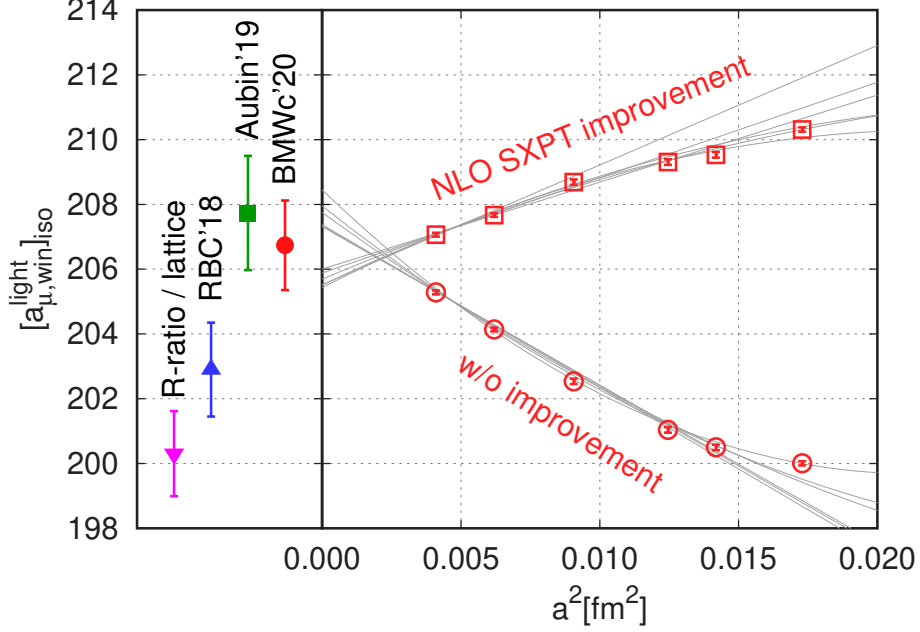
Figure 4: Continuum extrapolation of the isospin-symmetric, light, connected component of the window observable $a_{\mu,\text{win}}$, denoted by $[a_{\mu,\text{win}}^{\text{light}}]_{\text{iso}}$. The data points are extrapolated to the infinite-volume limit. Two different ways to perform the continuum extrapolations are shown: one without improvement, and another with corrections from next-to-leading-order staggered chiral perturbation theory. In both cases the lines show linear and quadratic fits in a^2 with varying number of lattice spacings in the fit. The continuum extrapolated result is shown with the results from other lattice groups, RBC'18 [19] and Aubin'19 [20]. Also plotted is our R-ratio-based determination, obtained using the experimental data compiled by the authors of [4] and our lattice results for the non light connected contributions. This value, denoted by 'R-ratio/lattice', is 3.4σ smaller than our pure lattice result for $[a_{\mu,\text{win}}^{\text{light}}]_{\text{iso}}$.

8. Bernecker, D. & Meyer, H. B. Vector Correlators in Lattice QCD: Methods and applications. *Eur. Phys. J.* **A47,** 148 (2011).

9. Lautrup, B. E., Peterman, A. & de Rafael, E. Recent developments in the comparison between theory and experiments in quantum electrodynamics. *Phys. Rept.* **3,** 193–259 (1972).

10. De Rafael, E. Hadronic contributions to the muon g-2 and low-energy QCD. *Phys. Lett.* **B322,** 239–246 (1994).

11. Blum, T. Lattice calculation of the lowest order hadronic contribution to the muon anomalous magnetic moment. *Phys. Rev. Lett.* **91,** 052001 (2003).

12. Borsanyi, S. *et al.* High-precision scale setting in lattice QCD. *JHEP* **09,** 010 (2012).

13. Dowdall, R., Davies, C., Lepage, G. & McNeile, C. Vus from pi and K decay constants in full lattice QCD with physical u, d, s and c quarks. *Phys. Rev. D* **88,** 074504 (2013).

14. Borsanyi, S. *et al.* Hadronic vacuum polarization contribution to the anomalous magnetic moments of leptons from first principles. *Phys. Rev. Lett.* **121,** 022002 (2018).

15. Neff, H., Eicker, N., Lippert, T., Negele, J. W. & Schilling, K. On the low fermionic eigenmode dominance in QCD on the lattice. *Phys. Rev.* **D64,** 114509 (2001).

16. Giusti, L., Hernandez, P., Laine, M., Weisz, P. & Wittig, H. Low-energy couplings of QCD from current correlators near the chiral limit. *JHEP* **04,** 013 (2004).

17. DeGrand, T. A. & Schaefer, S. Improving meson two point functions in lattice QCD. *Comput. Phys. Commun.* **159,** 185–191 (2004).

18. Shintani, E. *et al.* Covariant approximation averaging. *Phys. Rev.* **D91,** 114511 (2015).

19. Blum, T. *et al.* Calculation of the hadronic vacuum polarization contribution to the muon anomalous magnetic moment. *Phys. Rev. Lett.* **121,** 022003 (2018).

20. Aubin, C. *et al.* Light quark vacuum polarization at the physical point and contribution to the muon $g - 2$. *Phys. Rev. D* **101,** 014503 (2020).

21. De Divitiis, G. M. *et al.* Isospin breaking effects due to the up-down mass difference in Lattice QCD. *JHEP* **04,** 124 (2012).

22. De Divitiis, G. M. *et al.* Leading isospin breaking effects on the lattice. *Phys. Rev.* **D87,** 114505 (2013).

217 23. Colangelo, G., Durr, S. & Haefeli, C. Finite volume effects for meson masses and decay constants.
218   *Nucl. Phys.* **B721,** 136–174 (2005).

219 24. Davoudi, Z. & Savage, M. J. Finite-Volume Electromagnetic Corrections to the Masses of Mesons,
220   Baryons and Nuclei. *Phys. Rev.* **D90,** 054503 (2014).

221 25. Borsanyi, S. *et al.* Ab initio calculation of the neutron-proton mass difference. *Science* **347,** 1452–
222   1455 (2015).

223 26. Fodor, Z. *et al.* Quantum electrodynamics in finite volume and nonrelativistic effective field theories.
224   *Phys. Lett.* **B755,** 245–248 (2016).

225 27. Aubin, C. *et al.* Finite-volume effects in the muon anomalous magnetic moment on the lattice.
226   *Phys. Rev.* **D93,** 054508 (2016).

227 28. Bijnens, J. & Relefors, J. Vector two-point functions in finite volume using partially quenched chiral
228   perturbation theory at two loops. *JHEP* **12,** 114 (2017).

229 29. Hansen, M. T. & Patella, A. Finite-volume effects in $(g-2)_\mu^{\mathsf{HVP,LO}}$. arXiv: 1904.10010 [hep-lat]
230   (2019).

231 30. Gerardin, A. *et al.* The leading hadronic contribution to $(g-2)_\mu$ from lattice QCD with $N_{\mathrm{f}} = 2+1$
232   flavours of O(a) improved Wilson quarks. *Phys. Rev.* **D100,** 014510 (2019).

233 31. Davies, C. T. H. *et al.* Hadronic-Vacuum-Polarization Contribution to the Muon's Anomalous
234   Magnetic Moment from Four-Flavor Lattice QCD. *Phys. Rev.* **D101,** 034512 (2020).

235 32. Giusti, D., Lubicz, V., Martinelli, G., Sanfilippo, F. & Simula, S. Electromagnetic and strong isospin-
236   breaking corrections to the muon $g-2$ from Lattice QCD+QED. *Phys. Rev.* **D99,** 114502 (2019).

Supplementary Information

Leading hadronic contribution to the muon magnetic moment from lattice QCD

Sz. Borsanyi[1], Z. Fodor[1,2,3,4,5,*], J. N. Guenther[6], C. Hoelbling[1], S. D. Katz[3], L. Lellouch[7], T. Lippert[1,2], K. Miura[7,8,9], L. Parato[7], K. K. Szabo[1,2], F. Stokes[2], B. C. Toth[1], Cs. Torok[2], L. Varnhorst[1]

[1] Department of Physics, University of Wuppertal, D-42119 Wuppertal, Germany
[2] Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52428 Jülich, Germany
[3] Institute for Theoretical Physics, Eötvös University, H-1117 Budapest, Hungary
[4] University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[5] Pennsylvania State University, Department of Physics, State College, PA 16801, USA
[6] Department of Physics, University of Regensburg, Regensburg D-93053, Germany
[7] Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France
[8] Helmholtz Institute Mainz, D-55099 Mainz, Germany
[9] Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, Nagoya University, Nagoya 464-8602, Japan

Contents

Figure 1: Position of the 4stout ensembles in the plane of the hadron mass combinations of Equation (1). These correspond approximately to the light and strange quark masses. The lattice spacings are $a = 0.1315, 0.1191, 0.1116, 0.0952, 0.0787$ and 0.0640 fm, respectively. The corresponding β gauge couplings are indicated at the top of each panel.

1   The 4stout action and gauge ensembles

The main part of the simulation effort was carried out using the 4stout lattice action. This discretization is defined through the use of the tree-level Symanzik gauge action [2] and a one-link staggered fermion action with four steps of stout smearing [3]. The smearing parameter was set to $\rho = 0.125$.

We have chosen six gauge coupling parameters, $\beta = 6/g^2$, as shown in Table 1. All of these ensembles were generated using 2+1+1 dynamical flavors with no isospin breaking. The charm mass is set by its ratio to the strange mass, $m_c/m_s = 11.85$, which comes from the spectroscopy of the pseudoscalar charmed mesons in the continuum limit worked out in [4]. This value is within one per-cent of the latest FLAG average [5]. The light and strange quark masses are chosen to scatter around a "physical point" defined by the pseudoscalar masses M_π and M_K and the mass of the Omega baryon, M_Ω, as follows:

$$\frac{M_\pi^2}{M_\Omega^2} = \left[\frac{M_{\pi_0}^2}{M_{\Omega_-}^2} \right]_* , \qquad \frac{M_K^2 - \frac{1}{2} M_\pi^2}{M_\Omega^2} = \left[\frac{M_{K_\chi}^2}{M_{\Omega_-}^2} \right]_* . \qquad (1)$$

where $*$ denotes the experimental value and

$$M_{K_\chi}^2 \equiv \tfrac{1}{2} \left(M_{K^+}^2 + M_{K^0}^2 - M_{\pi^+}^2 \right) . \qquad (2)$$

The latter quantity is designed to be approximately proportional to the strange quark mass with a vanishing leading order sensitivity to strong-isospin breaking.

In Equation (1), the mass of the Omega baryon plays the role of the scale setting variable. It could, in principle, be replaced by any other dimensionful quantity that satisfies the criteria: a) moderate quark

β	a[fm]	$L \times T$	m_s	m_s/m_l	#conf
3.7000	0.1315	48×64	0.057291	27.899	904
3.7500	0.1191	56×96	0.049593	28.038	315
			0.049593	26.939	516
			0.051617	29.183	504
			0.051617	28.038	522
			0.055666	28.038	215
3.7753	0.1116	56×84	0.047615	27.843	510
			0.048567	28.400	505
			0.046186	26.479	507
			0.049520	27.852	385
3.8400	0.0952	64×96	0.043194	28.500	510
			0.043194	30.205	190
			0.043194	30.205	436
			0.040750	28.007	1503
			0.039130	26.893	500
3.9200	0.0787	80×128	0.032440	27.679	506
			0.034240	27.502	512
			0.032000	26.512	1001
			0.032440	27.679	327
			0.033286	27.738	1450
			0.034240	27.502	500
4.0126	0.0640	96×144	0.026500	27.634	446
			0.026500	27.124	551
			0.026500	27.634	2248
			0.026500	27.124	1000
			0.027318	27.263	985
			0.027318	28.695	1750

Table 1: List of 4stout ensembles with gauge coupling, lattice spacing at the physical point, lattice size, quark masses and number of configurations. Two different lines with the same parameters means that the ensembles were generated in two different streams.
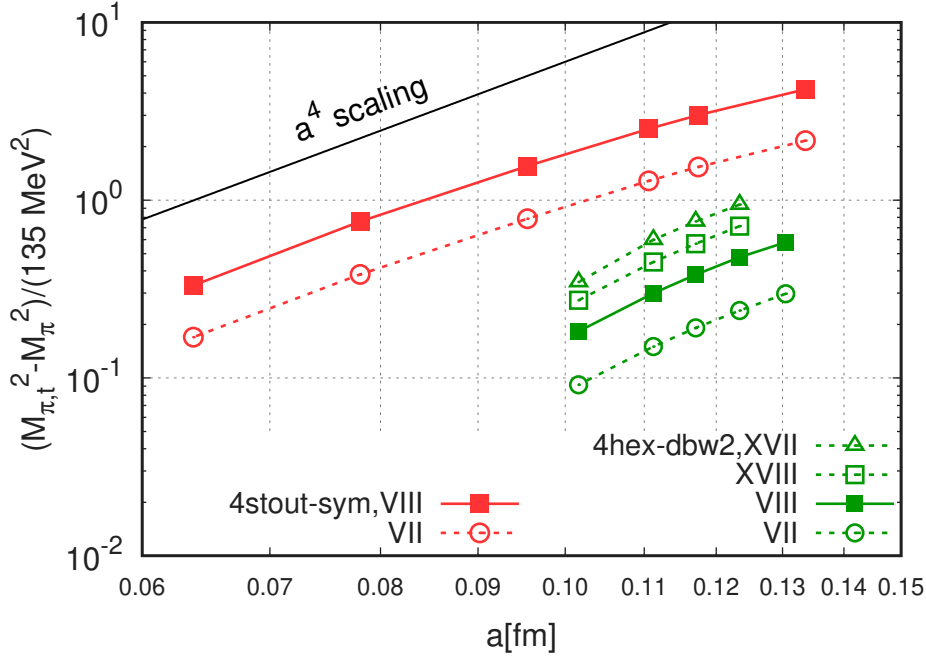
Figure 2: Taste multiplet of staggered pions as a function of lattice spacing, both for the `4stout` and the `4HEX` action. We label the meson operators by Roman numbers as in [1]. The root-mean-square pion mass corresponds approximately to the operator VIII, drawn with a solid line.

mass dependence, b) precisely determined in a lattice simulation, c) known experimental value to an accuracy better than a permil level. The mass M_Ω satisfies all three criteria, see Section 4 for more details. In this work we also use the w_0-scale [6], which is derived from the Wilson-flow of the gauge fields [7]. The main motivation for this scale setting is to define an isospin decomposition, see Section 6. The w_0-scale readily satisfies both the a) and b) criteria but, alas, it is defined in terms of observables in Euclidean space, not by any experiment. In order to use w_0 for scale setting, we first determine its physical value from our simulations, using the accurate M_Ω scale as an input. This is described in Section 21. Evidently whenever we use the w_0 scale setting, both the statistical and the systematic error of w_0, as well as the statistical correlation, will be accounted for. As a by-product of this procedure, we give a physical value for w_0, including dynamical QED effects, for the first time in the literature.

Our main analysis is based on the 27 ensembles shown in Table 1. In Figure 1 we show the "landscape" for each of our lattice spacings: we plot the ensembles in a plane where the x and y axes give the relative deviation of the light and strange quark masses from their physical value. These are defined by the hadronic observables and their experimental values in Equation (1). The simulation parameters are chosen in a way that makes interpolation to the physical point possible. This "bracketing" feature is not available for each lattice spacing, but only if all lattice spacings are considered together. This is not a problem, since in our analyses we apply global fits with all ensembles included. In Figure 1 each ensemble is represented by two points, corresponding to the M_Ω and w_0 scale settings. Although we determined the physical value of w_0 using M_Ω itself, the mass ratios vary with the choice of scale setting. This is because there are discretization effects in the $w_0 M_\Omega$ product on the lattice. Notice that the finer the lattices, the smaller the difference in the respective mass ratios.

We also measured the taste violation for all six lattice spacings. The result is shown in Figure 2. The plot shows the mass-squared difference between a non-Goldstone pion and the Goldstone pion as a function of the lattice spacing. The difference shows a behavior that resembles a^4 in the range of our smallest three lattice spacings. This is much faster than the $\alpha_s a^2$ [8], where α_s is the strong coupling constant at the lattice cutoff scale. The faster falloff is most probably due to higher order terms of the type $\alpha_s^n a^2$ with $n > 1$. At the smallest lattice spacing the root-mean-square pion mass is about $m_{\pi,\mathrm{RMS}} = 155$ MeV.

In Figure 3 we show the topological-charge history in a run on our finest lattice. The charge Q was computed using the standard discretization of the topological charge density at a Wilson-flow time of τ,
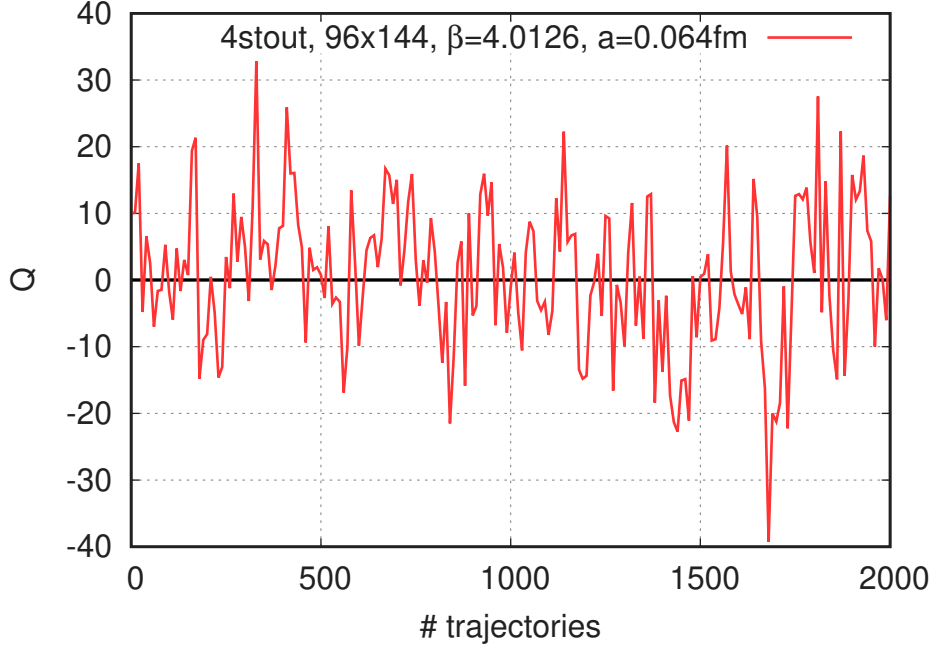
Figure 3: History of topological charge Q, defined from the Wilson-flow in a `4stout` run at the physical point. The lattice spacing is about $a = 0.064$ fm.

β	a[fm]	m_s	m_s/m_l	$L \times T$	# conf	M_π[MeV]
0.7300	0.112	0.060610	44.971	56×84	7709	104
				96×96	962	
			33.728	56×84	8173	121
				96×96	813	

Table 2: List of `4HEX` ensembles with gauge coupling, lattice spacing, quark masses, lattice size, number of configurations and Goldstone pion mass. These masses are chosen so that they bracket the point where a certain taste-average pion mass has the physical value of the pion mass (see text). At that point the Goldstone-pion mass is $M_\pi = 110$ MeV.

which was set to have a smearing radius of about $\sqrt{8\tau} \approx 0.6$ fm. The integrated autocorrelation time of Q is found to be $19(2)$ trajectories.

2 The `4HEX` action and gauge ensembles

A major systematic effect of the `4stout` ensembles is related to their box size, which is about $L \approx 6$ fm. To obtain the infinite-volume result it is desirable to extend the data set with a significantly larger box. A large box is computationally only affordable with a large lattice spacing. On coarse lattices, however, taste violations make the pions too heavy, which completely distorts the finite-volume behavior and makes a finite-size study pointless.

We introduce here a new staggered action, called `4HEX`, to drastically reduce the taste violation. This requires a gauge action that heavily suppresses ultraviolet fluctuations and a fermion action with a more agressive smearing than `4stout`. As a result, taste splitting is reduced by an order of magnitude. Additionally, we lower the Goldstone-pion mass below the physical value, ensuring that the heavier tastes are closer to the physical pion mass. The topological susceptibility is particularly sensitive to taste violations. We show here that the observed reduced taste violation is paired with having the topological susceptibility much closer to the continuum than in `4stout` at the same lattice spacing. We use this 4HEX action to generate the lattices for our finite-size study.

The 4HEX gauge ensembles use the same action both for the valence and the sea quarks. They have
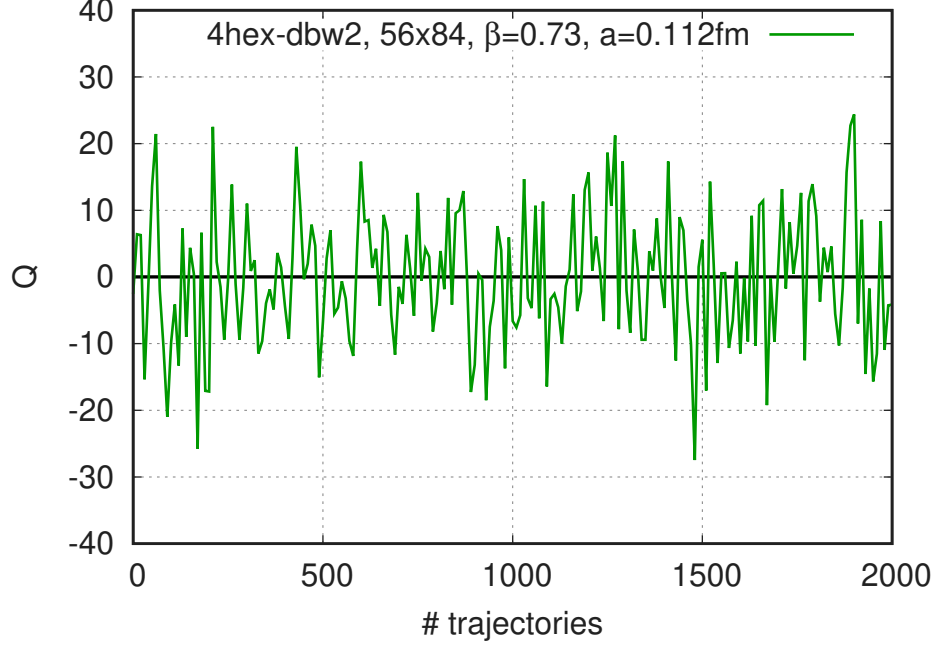
Figure 4: History of topological charge, defined from the Wilson-flow in a 4HEX run at the physical point. The lattice spacing is about $a = 0.112$ fm and the strange-to-light-quark mass ratio is $m_s/m_l = 33.728$.
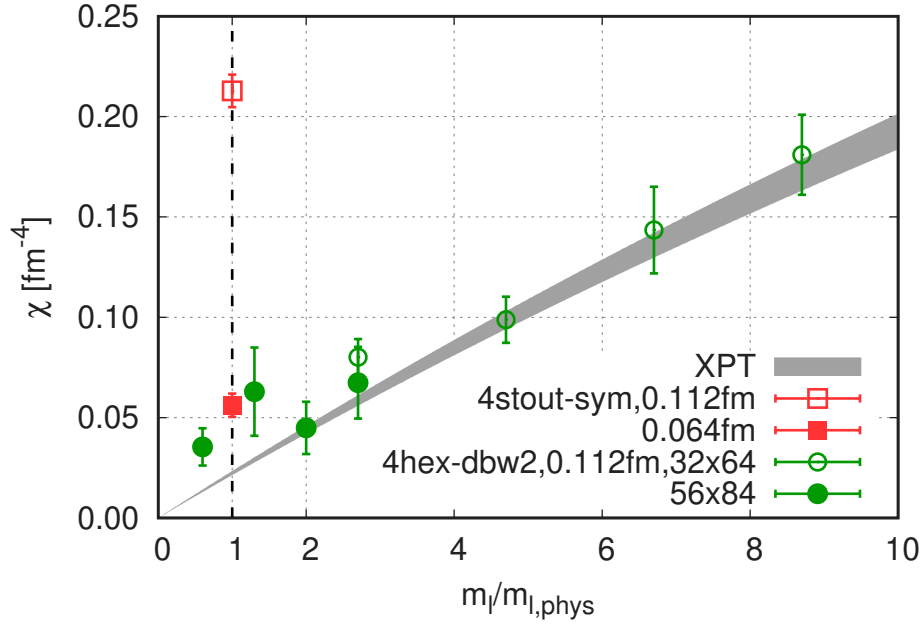
Figure 5: Topological susceptibility as a function of the quark mass with the 4HEX action at $a = 0.112$ fm lattice spacing. Also plotted are 4stout results at the physical point and with two different lattice spacings. The grey band is the prediction of leading order chiral perturbation theory, with parameters taken from [9]. The 4HEX simulation with the lightest quark mass ($m_l = 0.6 \cdot m_{\mathrm{phys}}$) has a topological susceptibility, that is about the same as in the continuum at the physical point.

the following characteristics:

- $n_f = 2 + 1$ flavors of one-link staggered quarks with four steps of HEX smeared [10] gauge links,

- DBW2 gauge action [11], which differs from the Symanzik gauge action used in `4stout` only in the coefficient of the 1×2 Wilson-loop.

The rationale for this choice is to drastically reduce the ultraviolet fluctuations in the gauge configurations, since these directly impact the size of the taste violation.

Four steps of HEX smearing suppresses the ultraviolet fluctuations much more than four steps of stout smearing, and does so without increasing the locality range of the smearing procedure. Though higher numbers of HEX smearing steps would also have been possible, the increasing cost of the smearing and the marginal improvement in the taste violations make an even higher number of steps less practical.

The DBW2 gauge action suppresses taste violations even more than the Symanzik gauge action, as shown eg. in [12]. However, it slows down the decorrelation of the topological charge towards the continuum limit much more dramatically than other gauge actions. We use the 4HEX action only at lattice spacings where sufficient tunnelings in the topological charge Q are observed. This includes the lattice spacing where we carry out the finite-volume study of the hadronic vacuum polarization. Figure 4 shows the history of the charge Q in one of these runs. Again, Q is computed using the standard discretization of the topological charge density at a Wilson-flow time of τ, which was set to have a smearing radius of about $\sqrt{8\tau} \approx 1.1$ fm. The integrated autocorrelation time of Q is found to be $6(1)$ trajectories.

In exploring the parameter space of the action we carried out $n_f = 3$ simulations at five different β values in the range $\beta = 0.70 \ldots 0.75$, corresponding to lattice spacings $a \approx 0.13 \ldots 0.10$ fm. Note that such small β values are typical with the DBW2 gauge action. The quark mass was tuned to the vicinity of the three flavor symmetric point where the quark mass equals the physical value of $\frac{1}{3}(2m_l + m_s)$. The lattice sizes were 32×64. Members of the pion taste multiplet are shown as a function of the lattice spacing in Figure 2, together with 4stout data. The 4HEX taste violations are an order of magnitude smaller than that of the 4stout action. At a lattice spacing of $a = 0.112$ fm, 4HEX is as good as 4stout at $a = 0.064$ fm, which is the finest lattice spacing available.

This reduced taste violation is also reflected in the topological susceptibility. Non-chiral actions, including staggered fermions, typically show large discretization errors in this quantity. Figure 5 shows $n_f = 2 + 1$ 4HEX simulations at $\beta = 0.73$, $a = 0.112$ fm, with the physical strange-quark mass. The light-quark mass was varied from 0.6 to 8.7 times its physical value, set by the Goldstone-pion mass. The lattice sizes are 32×64 and 56×84. Results obtained with the 4stout action at the physical point are also given in the plot. The 4stout result at the same lattice spacing is off by an order of magnitude from the continuum expectation. On the other hand the 4HEX data closely follows the continuum curve almost down to the chiral limit, and is as good as the 4stout result at the finest lattice spacing available.

For the finite-size study of the hadronic vacuum polarization we work at $\beta = 0.73$ and $m_s = 0.06061$. This choice corresponds to $a = 0.112$ fm and about a physical strange-quark mass. The ensembles generated are listed in Table 2. We have two different volumes with the same parameters. We will refer to the smaller volume as the "reference volume". It has a spatial and time extent of

$$L_{\mathrm{ref}} = 6.272 \text{ fm} \qquad \text{and} \qquad T_{\mathrm{ref}} = \frac{3}{2}L_{\mathrm{ref}} \ . \tag{3}$$

This geometry corresponds approximately to the geometry of the lattices in the 4stout data set. The larger lattice extents will be denoted by L_{big} and T_{big} and are given as

$$L_{\mathrm{big}} = T_{\mathrm{big}} = 10.752 \text{ fm} \ . \tag{4}$$

We also use two light-quark-mass values, so that we can bracket the physical point. Here, differently from above and also from the 4stout data set, we set the physical point, not with the Goldstone-pion mass, rather with a prescription that takes into account taste violations. Such a choice is advantegeous for

β	$a[\text{fm}]$	$L \times T$	m_s	m_s/m_l	#conf
3.7000	0.1315	24×48	0.057291	27.899	716
3.7753	0.1116	28×56	0.047615	27.843	887
3.8400	0.0952	32×64	0.043194	28.500	1110
3.9200	0.0787	40×80	0.032000	26.512	559
			0.033286	27.738	364
4.0126	0.0640	48×96	0.026500	27.634	339
			0.027318	27.263	264

Table 3: List of `4stout` ensembles used in a crosscheck with valence overlap quarks. The columns are gauge coupling, lattice spacing at the physical point, lattice size, quark masses and number of configurations.

β	m_l	m_{ov}	Z_V
3.7000	0.0021	0.0164(2)	1.1474(3)
3.7753	0.0017	0.0076(1)	1.1162(3)
3.8400	0.0015	0.0041(1)	1.0981(2)
3.9200	0.0012	0.0021(1)	1.0805(1)

Table 4: Staggered light-quark mass, matched overlap quark mass and vector renormalization constant for different lattice spacings.

studying finite-size effects, which depend strongly on the masses of the pions. The precise definition of the taste-average pion mass will be given in the section on finite-size effects, Section 17. For the Goldstone pion this prescription gives $M_\pi = 110$ MeV. Let us note here also that the topological susceptibility, with such a choice, is about the same as in the continuum limit at the physical point, as shown in Figure 5. The number of configurations saved is also given in Table 2. They are separated by 10 unit-length RHMC trajectories.

3 Overlap action

In order to crosscheck our results for staggered valence quark artefacts, including the normalization of the vector current, we compute $a_{\mu,\text{win}}^{\text{light}}$ in a mixed action setup, with overlap valence quarks on gauge backgrounds generated with the `4stout` staggered action. We work at the isospin-symmetric point in this crosscheck.

For the sea quarks we use the `4stout` staggered action and generate configurations with $L \approx 3$ fm box sizes and at five lattice spacings, given in Table 3. The parameters are chosen to match the parameters of a subset of the $L \approx 6$ fm lattices of the `4stout` data set given in Table 1. Though there are significant finite-size effects in a box of $L \approx 3$ fm for an observable like a_μ^{light}, these are much less severe for the window observable $a_{\mu,\text{win}}^{\text{light}}$, which is our target here. Our setup is appropriate for crosschecking the continuum extrapolation and also the normalization of the vector current.

For the valence quarks we use the overlap fermion formulation [13]. In particular, the overlap Dirac operator D_{ov} is constructed from the sign function of the Wilson Dirac operator D_{W} as

$$D_{\text{ov}} = m_{\text{W}} \left[\text{sgn}(\gamma_5 D_{\text{W}}) + 1 \right] \ , \tag{5}$$

where the Wilson operator has a mass of $-m_{\text{W}}$. We choose $m_{\text{W}} = 1.3$ and use the Zolotarev approximation of the sign function. The gauge fields undergo two steps of HEX smearing [14]. The overlap mass m_{ov} is introduced as

$$D_{\text{ov}}(m_{\text{ov}}) = \left(1 - \tfrac{1}{2} m_{\text{ov}} \right) D_{\text{ov}} + m_{\text{ov}} \ . \tag{6}$$
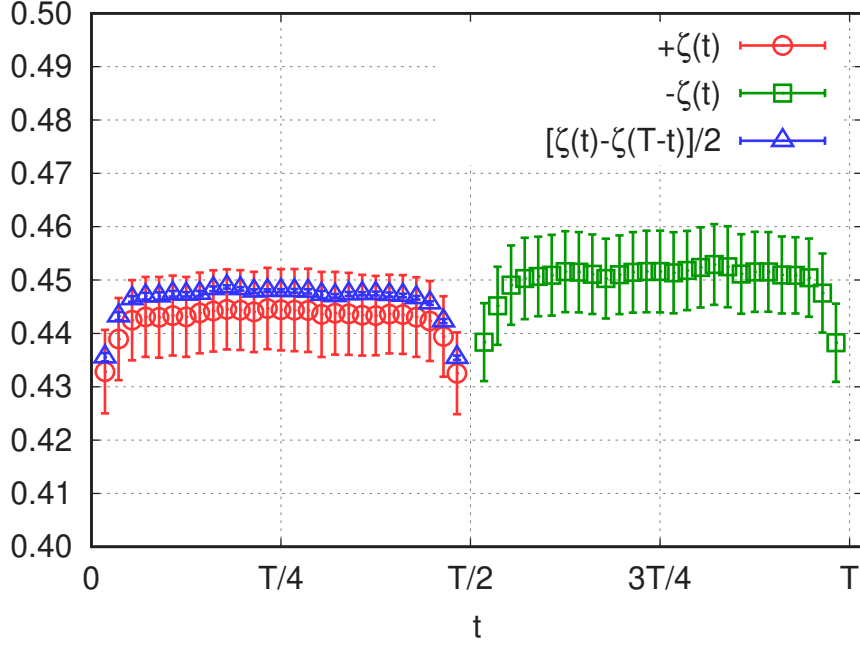
Figure 6: Ratio of three-point and two-point functions $\zeta(t)$ as function of the current insertion time t. This ratio is used to define the local vector current renormalization factor as $Z_V = [\zeta(T/4) - \zeta(3T/4)]^{-1}$. The plot shows data from a $\beta = 3.7753$ ensemble, $T = 56$.

This version of the operator has been extensively used in previous thermodynamical studies [15, 16]. We apply $O(a)$ improvement to the overlap propagator by transforming each instance of $D_{ov}^{-1}(m_{ov})$ as

$$D_{ov}^{-1}(m_{ov}) \rightarrow \left(1 - \tfrac{1}{2}D_{ov}/m_W\right) D_{ov}^{-1}(m_{ov}) . \tag{7}$$

At the same time we also compute the propagators with `4stout` staggered quarks, using the noise reduction technique as on the large volume ensembles.

We set the overlap quark mass by matching the staggered and overlap pion masses. For this purpose, we compute overlap pion masses at four values of the quark mass $m_{ov} = 0.002, 0.005, 0.010, 0.020$ and interpolate the pion mass squared using the form

$$M_{\pi,ov}^2(m_{ov}) = A m_{ov}^B + C m_{ov}^2 \tag{8}$$

with A, B and C fit parameters. This form can capture a possible quenched chiral logarithm [17] typical in mixed action setups. Our matching condition is to set the root-mean-square staggered pion mass equal to the overlap pion mass. Using the RMS pion is more advantageous than using the Goldstone-pion mass, since in the latter case we face drastic increase in the statistical error on our coarsest lattices. For the matched overlap quark masses we get the values given in Table 4.

Our determination of $a_{\mu,\text{win}}^{\text{light}}$ proceeds in a similar way as in the staggered-on-staggered case, described in Sections 10 and 11. A major difference is that we use the local vector current in the overlap case. We thus need to compute the current renormalization constant Z_V, which we get by measuring the electric charge of the pion. For this we compute the ratio of three-point and two-point functions:

$$\zeta(t) = \frac{\langle P(T/2)V_4(t)\bar{P}(0)\rangle}{\langle P(T/2)\bar{P}(0)\rangle} , \tag{9}$$

where the pseudoscalar density P and the local vector current V_μ are given in terms of valence overlap

fermion fields ψ_1 and ψ_2 as:

$$P(t) = \sum_{\vec{x}} (\bar{\psi}_2 \gamma_5 \psi_1)(\vec{x}, t) \ , \quad \bar{P}(t) = \sum_{\vec{x}} (\bar{\psi}_1 \gamma_5 \psi_2)(\vec{x}, t) \ , \quad V_\mu(t) = \sum_{\vec{x}} (\bar{\psi}_1 \gamma_\mu \psi_1)(\vec{x}, t) \ . \qquad (10)$$

In Figure 6 we show the ratio $\zeta(t)$ as a function of the timeslice of the current insertion t. In the case of a conserved current, $\zeta(t) = \frac{1}{2}$ for $t < T/2$ and $\zeta(t) = -\frac{1}{2}$ otherwise. The renormalization factor should be defined, so that the ζ-ratio for the renormalized current $Z_V V_4$ equals to 1 at some physical distance, for which we take $T/4$. We define the renormalization factor as $Z_V = [\zeta(T/4) - \zeta(3T/4)]^{-1}$, which includes a trivial symmetrization in time. The values for the different ensembles are given in Table 4.

	range #1 [fm]	range #2 [fm]
pion	$1.8\ldots3.0$	$2.0\ldots3.8$
kaon	$2.1\ldots3.3$	$2.4\ldots3.6$
$\bar{s}s$	$2.1\ldots3.3$	$2.4\ldots3.6$

Table 5: Fit ranges for extracting pseudoscalar masses on isospin symmetric ensembles.

β	N_{Wptl}	N_{3d}	four-state fit		GEVP fit		
			range #1	range #2	t_a	t_b	range
3.7000	24	32	$7\ldots17$	$6\ldots14$	4	7	$6\ldots14$
3.7500	30	40	$7\ldots19$	$6\ldots16$	4	7	$6\ldots14$
3.7753	34	46	$8\ldots20$	$7\ldots17$	4	7	$7\ldots17$
3.8400	46	62	$9\ldots23$	$8\ldots20$	4	9	$8\ldots16$
3.9200	67	90	$11\ldots28$	$10\ldots24$	6	9	$9\ldots17$
4.0126	101	135	$14\ldots30$	$12\ldots30$	6	11	$11\ldots19$

Table 6: Parameters used for obtaining the Ω mass: number of Wuppertal and gauge-link smearing steps in the Ω operator; fit ranges #1 and #2 for the four-state mass fit in Equation (15); parameters and fit ranges for the GEVP based mass fit.

4   Hadron mass measurements

Pseudoscalar mass measurements

The pseudoscalar propagators are computed with random wall sources and point sinks, using an operator corresponding to the pseudo-Goldstone taste. To extract the mass and the decay constant we performed a correlated $\cosh[M(t - T/2)]$ fit, using sufficiently late time slices to allow for this simple form. To estimate systematic errors, we selected two fit windows which are given for the different pseudoscalars in Table 5. For the kaons we selected the even/odd slices for the first/second fit window, respectively. These fit ranges were chosen by performing a Kolmogorov-Smirnov test, which ascertains whether the fit qualities in all of the fits on the ensembles of Table 1 follow a uniform distribution.

Omega propagators

To extract the mass of the positive-parity, ground-state Ω baryon, a number of different operators are available in the staggered formalism. First, there are two operators from the pioneering work of Golterman and Smit [18]. To label these operators we use the convention of [1]:

$$\Omega_{\mathrm{VI}}(t) = \sum_{x_k\,\mathrm{even}} \epsilon_{abc} \left[S_1\chi_a S_{12}\chi_b S_{13}\chi_c - S_2\chi_a S_{21}\chi_b S_{23}\chi_c + S_3\chi_a S_{31}\chi_b S_{32}\chi_c \right](x), \tag{11}$$

$$\Omega_{\mathrm{XI}}(t) = \sum_{x_k\,\mathrm{even}} \epsilon_{abc}[S_1\chi_a S_2\chi_b S_3\chi_c](x). \tag{12}$$

Here, $\chi_a(x)$ is the strange-quark field with color index a. The operator S_μ performs a symmetric, gauge-covariant shift in direction μ, while $S_{\mu\nu} \equiv S_\mu S_\nu$. Both Ω_{VI} and Ω_{XI} couple to two different tastes of the Ω baryon, which become degenerate in the continuum limit. At finite lattice spacing however, there is a splitting between the two tastes. In principle they could be disentangled by carrying out an analysis involving the correlators of both Ω_{VI} and Ω_{XI} and also their cross terms. Later, Bailey successfully constructed an operator which only couples to a single taste [19]. To achieve this, two additional (valence) strange quarks are introduced. In other words, the strange-quark field gets an additional "flavor" index:
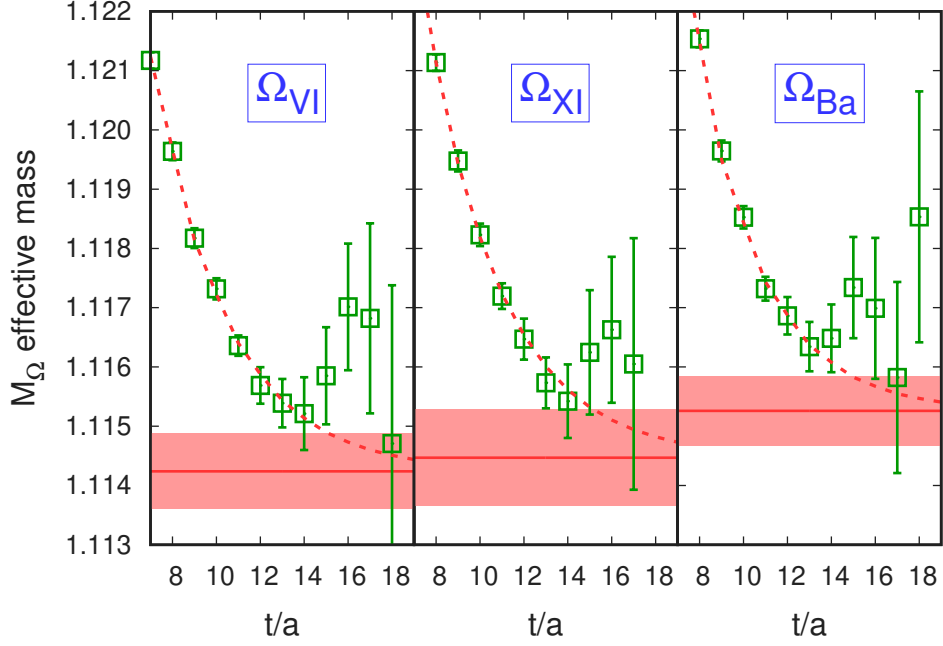
Figure 7: Effective mass of the ground state of the Ω baryon in lattice units on our coarsest ensemble with $\beta = 3.7000$. Results with three different staggered operators, Ω_{VI}, Ω_{XI} and Ω_{Ba} are shown. The horizontal lines and the shaded regions represent the fit values and the errors obtained with a four-state fit, Equation (15), using range #1 from Table 6. The respective χ^2-values including the contribution of the priors are 2.1, 1.6 and 2.7 for 6 degrees of freedom. The dashed lines are the effective masses computed from the fitted functions.

$\chi_{a\alpha}$ with $\alpha = 1, 2, 3$. The operator is then given as

$$\Omega_{\mathrm{Ba}}(t) = [2\delta_{\alpha 1}\delta_{\beta 2}\delta_{\gamma 3} - \delta_{\alpha 3}\delta_{\beta 1}\delta_{\gamma 2} - \delta_{\alpha 2}\delta_{\beta 3}\delta_{\gamma 1} + (\ldots \beta \leftrightarrow \gamma \ldots)] \cdot$$
$$\cdot \sum_{x_k \text{even}} \epsilon_{abc} \left[S_1\chi_{a\alpha}S_{12}\chi_{b\beta}S_{13}\chi_{c\gamma} - S_2\chi_{a\alpha}S_{21}\chi_{b\beta}S_{23}\chi_{c\gamma} + S_3\chi_{a\alpha}S_{31}\chi_{b\beta}S_{32}\chi_{c\gamma} \right](x). \qquad (13)$$

The mass of this state becomes degenerate with the above two taste partners in the continuum limit. We investigated the difference between these three operators on an ensemble with large statistics. At $\beta = 3.7000$, corresponding to our coarsest lattice spacing, we generated about 3000 configurations in addition to the statistics listed in Table 1. Note that only the Ω operators were measured on these extra configurations. The effective masses for the above three operators are shown in Figure 7. In the asymptotic regime we see deviations below 0.1%, which gives an estimate of the taste violation. We expect that these will get smaller as we go to finer lattice spacings, as it does for pions. In this work we chose the Ω_{VI} operator for our scale setting measurements. This is justified, since typical statistical and systematic errors on our ensembles are around 0.1%, and thus cover the taste-violation effects estimated here.

As is usual with staggered fermions, these propagators have an oscillating contribution, corresponding to negative parity states. There are also excited states for both parities in the propagators. We suppress the excited state by a number of Wuppertal smearing steps [20] applied equally to the source and sink. To avoid mixing between time slices, we define the smearing to act in the spatial directions only. Since staggered baryon operators are defined on a coarse lattice with spacing $2a$, our implementation of the Wuppertal smearing connects only sites that reside on the same sublattice with $2a$ lattice spacing. In this way there is no interference with the spin-taste structure of the operators. The action of the smearing
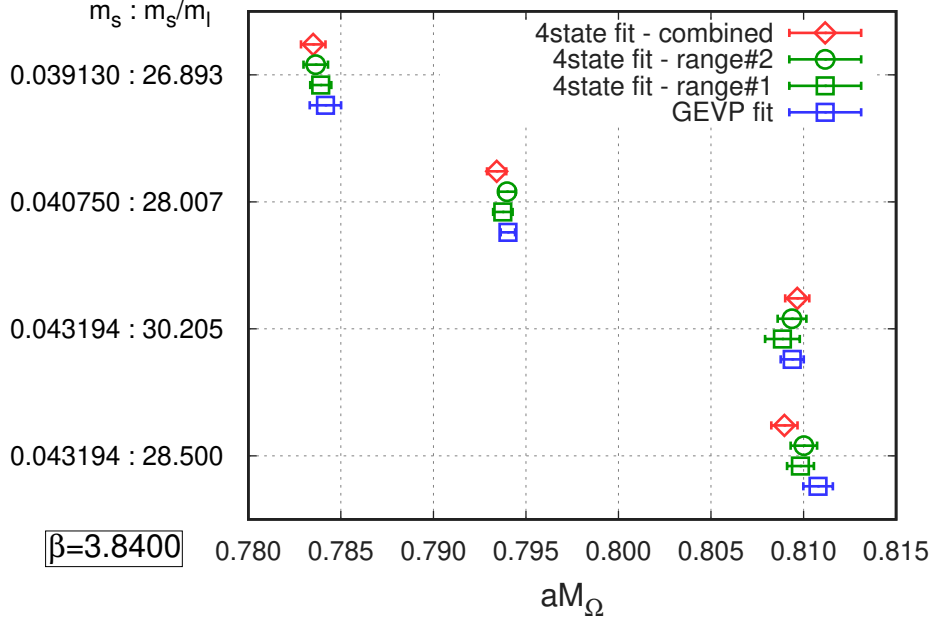
Figure 8: The mass of the Ω baryon in lattice units extracted for four ensembles with $\beta = 3.8400$, which bracket the physical point. We compare four methods. The deviation between them is used to estimate the systematic error (see text for details).

operator $\hat{W}$ on a vector v_x with coefficient σ is given as:

$$[\hat{W}v]_x = (1 - \sigma)v_x + \frac{\sigma}{6} \sum_{\mu=1,2,3} \left(U^{3d}_{\mu,x} U^{3d}_{\mu,x+\mu} v_{x+2\mu} + U^{3d,\dagger}_{\mu,x-\mu} U^{3d,\dagger}_{\mu,x-2\mu} v_{x-2\mu} \right), \tag{14}$$

where the $U^{3d}[U]$ parallel transporters are inserted to keep the recipe gauge invariant: no gauge fixing is needed. These parallel transporters are built from a smeared version of the underlying gauge field configuration U, by applying a number of three dimensional stout smearing steps N_{3d}. For this we use the same $\rho = 0.125$ parameter that we also have in the link smearing of the Dirac operator. Note that the U^{3d} links are also needed to build the baryon operator, as they appear in the S_μ shifts. The smearing operation in Equation (14) is iterated N_{Wptl} times on both source and sink sides on a point vector with a coefficient of $\sigma = 0.5$. In Table 6 we list the number of smearing steps for each lattice spacing. The smearing radii corresponding to the number of smearing steps approximately follow the change in the lattice spacing. That way the smearing radius in physical units is kept constant. Note that we are discussing a three dimensional smearing here: this only effects the overlap of the mass eigenstates with the operator and leaves the masses invariant.

To enhance the signal, we calculate the Ω propagator from 256 different source fields on each gauge configuration listed in Table 1. For each source field we select a random time slice, which, in turn, is populated with eight independent Z_3 random point sources at $(0, 0, 0)$, $(L/2, 0, 0)$, ... and $(L/2, L/2, L/2)$. This formation of sources is usually called a grid source. We also randomize the center of the grid source.

Omega mass determination: four-state fits

Our model for the propagator is a four-state fit function h, with two positive and two negative parity states:

$$h(t, A, M) = A_0 h_+(M_0, t) + A_1 h_-(M_1, t) + A_2 h_+(M_2, t) + A_3 h_-(M_3, t) \tag{15}$$

where the

$$h_+(M, t) = e^{-Mt} + (-1)^{t-1}e^{-M(T-t)} \qquad \text{and} \qquad h_-(M, t) = -h_+(M, T - t) \tag{16}$$

functions describe the time dependence of the positive and negative parity states, see eg. Equation (123) of [21]. Here M_0 and A_0 are the mass and amplitude of the ground state. Our χ^2 function is defined as a sum of the correlated χ^2 of the model h and a prior term:

$$\chi^2(A, M) = \sum_{i,k} [h(t_i, A, M) - H_i] \, \text{Cov}_{ik}^{-1} \, [h(t_k, A, M) - H_k] + \chi_{\text{prior}}^2(M), \tag{17}$$

where H_i is the value of the hadron propagator on the time slice t_i and Cov_{ik} stands for the covariance between H_i and H_k. A prior term was introduced to stabilize the fit, containing priors on the masses except for the ground state. The concrete form is:

$$\chi_{\text{prior}}^2(M) = \sum_{s=1}^{3} \left(\frac{M_s/M_0 - \mu_s}{\delta\mu_s} \right)^2, \tag{18}$$

where the prior parameters are set as follows:

s	$\mu_s \cdot 1672$ MeV	$\delta\mu_s$
1	2012 MeV	0.10
2	2250 MeV	0.10
3	2400 MeV	0.15

The prior for the negative parity ground state, $s = 1$, is motivated by the recent observation from the Belle collaboration [22]. The excited states, $s = 2, 3$, have not been discovered in experiments so far, so their priors follow from the quark model [23]. The existence of these undiscovered states is also motivated by lattice thermodynamics below the chiral transition [24, 25].

The range of time slices, that are included in the χ^2, were chosen by an optimization on the coarsest lattice, $\beta = 3.7000$. As already mentioned, we have around 4000 configurations there, which is about four times larger than on the ensembles at other lattice spacings. In the fit range $[7 \ldots 17]$ we obtained the mass with a relative precision of 0.06% and with fit quality of $Q = 0.55$: $M_{\Omega,\text{VI}} = 1.11424(63)$ in lattice units. The priors did not impose a significant pull on the result, ie. the final values of the fit parameters were well within the prior widths. A different fit range $[6 \ldots 14]$ resulted in a change in M_Ω within a small fraction of the statistical error. These results reassure us that the excited state effects are smaller than the statistical error with these two fit ranges. On the other ensembles, with lesser statistics, we used these two fit ranges, keeping their values in physical units approximately constant upon changing the lattice spacing. The exact fit ranges used are given in Table 6 in lattice units.

In Figure 8 we show a comparison of the Ω masses obtained with the various fits on four ensembles at $\beta = 3.8400$. Besides the four-state fit with two different fit ranges, see Table 6, we also show a value from a "combined" fit. Here we combine the correlators from all of the ensembles at a given β and apply to them a common four-state fit. Assuming that all excited state masses are a linear function of the bare strange mass for a given β, one can fit this linear dependence across the ensembles along with the still independent ground state masses. This reduces the number of fit parameters and results in more stable fits. The bare light-mass dependence of the excited states can be safely ignored, even for the ground state these effects are barely significant. The result of this combined fit agrees well with the ones obtained from the individual fit ranges. We do not use this combined fit in our final analyses though, it would introduce correlations between ensembles, making the analysis procedure more complicated.

Omega mass determination: GEVP method

In addition to the above four-state fit to the Ω propagator we also used a mass extraction procedure proposed in [26], which is based on the Generalized Eigenvalue Problem (GEVP). The method has the advantage of not using priors. We first apply a folding transformation to the original hadron propagator H_t:

$$H_t \rightarrow \begin{cases} \frac{1}{2}[H_t + (-1)^{t+1}H_{T-t}] & 0 < t < \frac{T}{2} \\ H_t & t = 0 \text{ or } t = \frac{T}{2} \end{cases} \tag{19}$$

Then we construct a matrix for each time slice t:

$$\mathcal{H}(t) = \begin{pmatrix} H_{t+0} & H_{t+1} & H_{t+2} & H_{t+3} \\ H_{t+1} & H_{t+2} & H_{t+3} & H_{t+4} \\ H_{t+2} & H_{t+3} & H_{t+4} & H_{t+5} \\ H_{t+3} & H_{t+4} & H_{t+5} & H_{t+6} \end{pmatrix} \tag{20}$$

For a given t_a and t_b let $\lambda(t_a, t_b)$ be an eigenvalue and $v(t_a, t_b)$ an eigenvector solution to this 4×4 generalized eigenvalue problem:

$$\mathcal{H}(t_a)v(t_a, t_b) = \lambda(t_a, t_b)\mathcal{H}(t_b)v(t_a, t_b). \tag{21}$$

Here we select the smallest eigenvalue λ and use the corresponding eigenvector v to project out the ground state:

$$v^+(t_a, t_b)\mathcal{H}(t)v(t_a, t_b), \tag{22}$$

which then can be fitted to a simple $\exp(-Mt)$ type function. This assumes that backward propagating states are negligible between t_a and t_b as well as in the range used to fit Equation (22). The parity partner states inherent in the staggered formulation appear as excited states, that give large contributions to the correlation functions (22) constructed with an eigenvalue λ with non-minimal absolute value. In that case, both the correlation function (22) and the eigenvalue λ exhibit oscillating signs. In the case of the correlation function, this oscillation occurs as a function of t, and in the case of λ as a function of $t_b - t_a$. See [27] for details of the variational method with staggered fermions.

The tuneable parameters of the procedure are t_a and t_b for specifying the GEVP, as well as the fit range for the $\exp$ fitting in the last step; they are given in the last three columns of Table 6. Similar to the pion mass analysis these parameters were chosen by performing a Kolmogorov-Smirnov test across all the ensembles. In Figure 8 we show the fit results obtained with this GEVP procedure for ensembles at $\beta = 3.8400$. They are in good agreement with the four-state fit values.

The mass extracted using the GEVP gives a third M_Ω value for each ensemble, beside the results with the four-state fit procedure with the two fit ranges. We will use the deviation between these three values as a systematic error in the Ω mass determination.

Finite size corrections

In order to determine the finite-volume corrections for the pseudoscalar masses, $M_\pi(L) - M_\pi(\infty)$, and decay constants, $f_\pi(L) - f_\pi(\infty)$, we use the chiral perturbation theory based formulae of Reference [28]. Our pion masses are very close to the physical point, where one obtains a relative correction of 0.022% for the mass and a relative correction of 0.077% for the decay constant. The mass of the kaon also receives a correction due to the finite volume. However, this correction is so small, and any uncertainty related to it is so subdominant, that we ignore it.

We also take into account the effect of the finite time extent T in the decay constants, both for pions and kaons, assuming that they are free particles. This is obtained by noting that the T-dependence of the

free particle propagator is given by $\cosh[M(t - T/2)]/\sinh(MT/2)$. Therefore, we fit our propagators to the form $A\cosh[M(t - T/2)]/\sinh(MT/2)$ and extract the decay constant from the amplitude A.

The finite-size effects on the Ω mass is estimated from next-to-leading order, three-flavor, heavy-baryon chiral perturbation theory. See [29] for the corresponding formulas. To this order the pions give no contribution to the finite-size effects, but only the kaons and the eta do. As a result, the finite-size correction is so tiny that it can be safely neglected.

5 Path integral and expectation values

Our staggered path integral includes four flavors of quarks, $f = \{u, d, s, c\}$, gluon fields U and photon fields A and is given by:

$$Z = \int [dU] \exp(-S_g[U]) \int [dA] \exp(-S_\gamma[A]) \prod_f \det M^{1/4}[V_U \exp(ieq_f A), m_f]. \tag{23}$$

The ensemble specific definition of the gauge action S_g is given in Sections 1 and 2. The photon integral measure $[dA]$ and action S_γ are defined in the $\mathrm{QED_L}$ scheme [30]. The one-hop staggered matrix in a background field W_μ can be written as

$$M[W, m] = D[W] + m = \sum_\mu D_\mu[W_\mu] + m, \tag{24}$$

where D_μ is the covariant differentiation in the μ direction involving W and its adjoint $W^\dagger$ together with the obligatory staggered phases. In the path integral the fermions are coupled to a gauge field that is a product of the exponentiated photon field and of the smeared gluon gauge field V_U. Our smearing recipes are given in Sections 1 and 2. The photon field is not smeared. $q_f \in \{+\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}, +\frac{2}{3}\}$ stand for the quark electric charges in units of the positron charge e, m_f for the quark masses and $\alpha = e^2/(4\pi)$. We use the notation $\delta m \equiv m_d - m_u$ for the difference in the up and down quark masses and $m_l \equiv \frac{1}{2}(m_u + m_d)$ for their average. To simplify later formulas we also introduce the notations

$$M_f \equiv M[V_U e^{ieq_f A}, m_f] \qquad \text{and} \qquad \mathrm{dets}[U, A; \{m_f\}, \{q_f\}, e] \equiv \prod_f \det M_f^{1/4}, \tag{25}$$

where the latter is the product of all fermion determinants.

In this work, isospin-breaking is implemented by taking derivatives with respect to the isospin-breaking parameters and by measuring the so obtained derivative operators on isospin-symmetric configurations [31]. A different approach would be to generate configurations at non-zero values of the isospin breaking parameters and use the same operators as at zero isospin breaking, see eg. [32]. We choose the former approach in this work, so as to optimally distribute the computing resources among the various isospin-breaking contributions.

We introduce a set of notations for isospin-symmetric observables and their isospin-breaking derivatives. Consider the observable $X(e, \delta m)$, which is a function of e and δm. Then we define

$$X_0 \equiv X(0,0), \qquad X'_m \equiv m_l \frac{\partial X}{\partial \delta m}(0,0), \qquad X'_1 \equiv \frac{\partial X}{\partial e}(0,0), \qquad X''_2 \equiv \frac{1}{2}\frac{\partial^2 X}{\partial e^2}(0,0). \tag{26}$$

The isospin-breaking derivatives are denoted by prime(s) and an index. The mass derivative has the index m, it requires no renormalization, since δm and m_l have the same renormalization factor at zero electromagnetic coupling. The electric charge derivatives have a single digit index: 1 or 2. Below, we also define electric charge derivatives with two-digit indices. We take into account only leading-order isospin-breaking in this work, so no higher derivatives are needed.

In the case of the fermion determinant, the isospin-symmetric value is denoted by dets_0. The strong-

isospin-breaking of dets is zero at leading order:

$$\mathrm{dets}'_m = 0, \tag{27}$$

since dets is symmetric under the exchange $u \leftrightarrow d$. The electromagnetic derivatives are

$$\frac{\mathrm{dets}'_1}{\mathrm{dets}_0} = \sum_f \frac{q_f}{4} \mathrm{Tr}\left(M_f^{-1} D[iAV_U] \right),$$

$$\frac{\mathrm{dets}''_2}{\mathrm{dets}_0} = \frac{1}{2}\left[\left(\frac{\mathrm{dets}'_1}{\mathrm{dets}_0}\right)^2 - \sum_f \frac{q_f^2}{4}\mathrm{Tr}\left(M_f^{-1}D[A^2 V_U]\right) - \sum_f \frac{q_f^2}{4}\mathrm{Tr}\left(M_f^{-1}D[iAV_U]M_f^{-1}D[iAV_U]\right) \right], \tag{28}$$

where Tr is trace over color and spacetime indices and the argument of the D operator is a 3×3 complex matrix valued field, eg. $A^2 V_U$ has components $A_{\mu,x}^2 [V_U]_{\mu,x}$. The implementation of these derivatives is given in Section 7.

We also make a distinction between the electric charge in the fermion determinant and in the operator that we measure. We call the former sea electric charge and denote it by e_s, the latter is the valence electric charge and is denoted by e_v. For an observable X that depends on both the valence and sea charges, $X(e_v, e_s)$, the second order electric charge derivatives are defined as follows:

$$X''_{20} \equiv \frac{1}{2}\frac{\partial^2 X}{\partial e_v^2}(0,0), \qquad X''_{11} \equiv \frac{\partial^2 X}{\partial e_v \partial e_s}(0,0), \qquad X''_{02} \equiv \frac{1}{2}\frac{\partial^2 X}{\partial e_s^2}(0,0). \tag{29}$$

For functions that depend on either e_v or e_s, but not on both, we use the single digit notations of Equation (26).

The expectation value of an operator O is calculated by inserting $O[U, A]$ into the integrand of the path integral of Equation (23) and normalizing the integral by Z. Here we consider operators whose photon field dependence arises entirely from the photon-quark interaction, ie. $O = O[U, e_v A]$. The expectation value of this operator depends on δm, e_v and e_s, and the isospin expansion can be written as:

$$\langle O \rangle = [\langle O \rangle]_0 + e_v^2 \langle O \rangle''_{20} + e_v e_s \langle O \rangle''_{11} + e_s^2 \langle O \rangle''_{02} + \frac{\delta m}{m_l}\langle O \rangle'_m. \tag{30}$$

Here, the individual terms can be expressed as expectation values obtained with the isospin-symmetric path integral, which we denote by $\langle \ldots \rangle_0$. The concrete expressions are:

$$
\begin{aligned}
\text{isospin-symmetric:} \quad & [\langle O \rangle]_0 = \langle O_0 \rangle_0 \\
\text{qed valence-valence:} \quad & \langle O \rangle''_{20} = \langle O''_2 \rangle_0 \\
\text{qed sea-valence:} \quad & \langle O \rangle''_{11} = \left\langle O'_1 \frac{\mathrm{dets}'_1}{\mathrm{dets}_0} \right\rangle_0 \\
\text{qed sea-sea:} \quad & \langle O \rangle''_{02} = \left\langle O_0 \frac{\mathrm{dets}''_2}{\mathrm{dets}_0} \right\rangle_0 - \langle O_0 \rangle_0 \left\langle \frac{\mathrm{dets}''_2}{\mathrm{dets}_0} \right\rangle_0 \\
\text{strong-isospin-breaking:} \quad & \langle O \rangle'_m = \langle O'_m \rangle_0
\end{aligned}
\tag{31}
$$

In the derivation of these expressions we use $\left\langle \frac{\mathrm{dets}'_1}{\mathrm{dets}_0} \right\rangle_0 = 0$. In Table 7 we give an overview of the isospin-breaking derivatives for the observables that are computed in this paper.

Note that Equation (30) is an expansion in bare parameters and not what we consider a decomposition into isospin-symmetric and isospin breaking parts. The latter involves derivatives with respect to renormalized observables and our prescription for that is given in Section 6. There is no need to introduce a renormalized electromagnetic coupling though: its running is an $O(e^4)$ effect, ie. beyond the leading order isospin approximation that we consider here.

X	X_{20}''	X_{11}''	X_{02}''	X_m'	Section
$M_\Omega, M_{\pi_\chi}, M_{K_\chi}$	✓	✓	✓	-	9
$\Delta M_K^2, \Delta M^2$	✓	✓	-	✓	9
w_0	-	-	✓	-	8
$\langle JJ \rangle$-light	✓	✓	✓	✓	14
$\langle JJ \rangle$-strange	✓	✓	✓	-	14
$\langle JJ \rangle$-disc.	✓	✓	✓	✓	14

Table 7: Overview of isospin-breaking derivatives computed in this paper. For each observable, we specify the Section in which the implementation details can be found. A dash indicates, that the particular contribution vanishes.

6 Isospin breaking: decomposition

For various purposes it is useful to decompose the observables into isospin-symmetric and isospin-breaking parts. This requires a matching of the isospin symmetric and full theories, in which we specify a set of observables that must be equal in both theories. Of course, different sets will lead to different decompositions, which is commonly referred to as scheme dependence. Only the sum of the components, ie. the result in the full theory, is scheme independent.

A possible choice for the observables are the Wilson-flow–based w_0 scale and the masses of mesons built from an up/down/strange and an anti-up/down/strange quark, $M_{uu}/M_{dd}/M_{ss}$. These mesons are defined by taking into account only the quark-connected contributions in their two-point functions [33]. Their masses are practical substitutes for the quark masses. Also, they are neutral and have no magnetic moment, so they are a reasonable choice for an isospin decomposition. These masses cannot be measured in experiments, but have a well defined continuum limit and thus a physical value can be associated to them.

According to partially-quenched chiral perturbation theory coupled to photons [34], the combination

$$M_{\pi_\chi}^2 \equiv \tfrac{1}{2}(M_{uu}^2 + M_{dd}^2) \tag{32}$$

equals the neutral pion mass, $M_{\pi_\chi} = M_{\pi_0}$, up to terms that are beyond leading order in isospin breaking. Since such terms are beyond the accuracy needed in this work, we use the experimental value of the neutral pion mass as the physical value of M_{π_χ}. Furthermore the difference,

$$\Delta M^2 \equiv M_{dd}^2 - M_{uu}^2 \tag{33}$$

is a measure of strong-isospin-breaking not affected by electromagnetism. According to [34], $\Delta M^2 = 2B_2\delta m$ is valid up to effects that are beyond leading order in isospin breaking, at least around the physical point. Here, B_2 is the two-flavor chiral condensate parameter. For the determination of the physical values of w_0, M_{ss} and ΔM^2, see Section 21.

For the decomposition we start with the QCD+QED theory and parameterize our observable $\langle O \rangle$ with the quantities defined above:

$$\langle O \rangle (M_{\pi_\chi} w_0, M_{ss} w_0, \tfrac{L}{w_0}, \Delta M w_0, e) . \tag{34}$$

Here, the continuum limit is assumed. We can isolate the electromagnetic part by switching off the electromagnetic coupling, while keeping the other parameters fixed:

$$\langle O \rangle_{\text{qed}} \equiv e^2 \cdot \frac{\partial \langle O \rangle}{\partial e^2} \bigg|_{M_{\pi_\chi} w_0, M_{ss} w_0, \frac{L}{w_0}, \Delta M w_0, e=0} . \tag{35}$$

β	a[fm]	$L \times T$	m_s	m_s/m_l	#conf
3.7000	0.1315	24×48	0.057291	27.899	716
		48×64	0.057291	27.899	300
3.7753	0.1116	28×56	0.047615	27.843	887
3.8400	0.0952	32×64	0.043194	28.500	1110
			0.043194	30.205	1072
			0.040750	28.007	1036
			0.039130	26.893	1035

Table 8: List of `4stout` ensembles used for computing dynamical QED effects with gauge coupling, lattice spacing at the physical point, lattice size, quark masses and number of configurations.

The strong-isospin-breaking part is given by the response to the ΔM parameter:

$$\langle O \rangle_{\text{sib}} \equiv (\Delta M w_0)^2 \cdot \left. \frac{\partial \langle O \rangle}{\partial (\Delta M w_0)^2} \right|_{M_{\pi_\chi} w_0, M_{ss} w_0, \frac{L}{w_0}, \Delta M w_0 = 0, e = 0} , \tag{36}$$

and the isospin-symmetric part is just the remainder:

$$\langle O \rangle_{\text{iso}} \equiv \langle O \rangle (M_{\pi_\chi} w_0, M_{ss} w_0, \frac{L}{w_0}, 0, 0). \tag{37}$$

One can also define the decomposition at a finite lattice spacing, for which w_0 in lattice units can be additionally fixed. In doing so the isospin symmetric part $\langle O \rangle_{\text{iso}}$ has to be distinguished from the value of the observable at the bare isospin-symmetric point $[\langle O \rangle]_0$.

In this work we use the above definitions for the isospin decomposition; a similar scheme was put forward in [35]. A different scheme would be to keep the renormalized quark masses and the strong coupling constant fixed as QED is turned on [36]. In case of light quark observables the two schemes supposed to agree well. This can be justified by the smallness of the electromagnetic part of the neutral pion in the scheme of [36]. Reference [35] found $\epsilon_{\pi_0} = 0.03(2)$, where ϵ_{π_0} is the parameter that measures the size of the electromagnetic contribution in the neutral pion mass. In comparison the same quantity for the charged pion was found $\epsilon_{\pi_+} = 1.03(2)$ in [35].

7 Isospin breaking: dynamical QED

In the isospin expansion of an observable $\langle O \rangle$ (see Equation (30)) we refer to the e_s dependent terms as dynamical QED contributions.

The sea-valence contribution is given by Equation (31) as

$$\langle O \rangle''_{11} = \left\langle \left\langle O'_1 \frac{\text{dets}'_1}{\text{dets}_0} \right\rangle_A \right\rangle_U . \tag{38}$$

Here we made explicit that the path integral is carried out over two gauge fields: the index A of the expectation value means averaging over free photon fields with the action S_γ. The rest of the path integral weight is contained in the gluon expectation value, labeled with index U. The trace over coordinate and color space in the first derivative of the fermion determinant (see Equation (28)) is computed exactly in the low-lying eigenmode space of the Dirac operator and with random vectors in the complement. According to our findings, the noise in this term overwhelmingly stems from the random sources, and not from the gauge fields. For each U field we generate one A field and on this (U, A) gauge field pair we use about 10^4 random vectors to estimate the first derivative $\text{dets}'_1/\text{dets}_0$. The first derivative of the observable is estimated by a finite difference $O'_1 \approx \frac{1}{2e_v}(O_+ - O_-)$, see eg. Section 9.

The sea-sea contribution is given by

$$\langle O \rangle_{02}'' = \left\langle [O_0 - \langle O_0 \rangle_U] \left\langle \frac{\mathrm{dets}_2''}{\mathrm{dets}_0} \right\rangle_A \right\rangle_U , \tag{39}$$

where the A-average of the second derivative of the determinant can be done independently from the observable. This is especially useful, since the noise in this term is dominated by fluctuations in the photon field. On each U configuration we use about 2000 photon fields, and on each photon field, 12 random sources to estimate the second derivative $\mathrm{dets}_2''/\mathrm{dets}_0$.

For both contributions we apply the Truncated Solver Method [37, 38]: the matrix inverters are run with a reduced precision most of the time, and the resulting small bias is corrected using occasional, high-precision inversions.

In this work we compute dynamical QED effects on a dedicated set of ensembles using the 4stout action. We have three lattice spacings, with box sizes around $L = 3$ fm with $T = 2L$. Additionally, on the coarsest lattice there is also an ensemble with an $L = 6$ fm box. Table 8 gives the parameters of these ensembles, together with the number of configurations. The chosen parameters in these dedicated runs match the parameters of selected 4stout ensembles. For the observables that we consider, we see no significant difference in the size of dynamical QED contributions between the two different volumes. For the volume dependence of the hadron masses, see Section 9 and Figure 9. The $L = 3$ fm volume simulations need about an order of magnitude less computer time for the same precision. Therefore, on the finer lattices we performed simulations in the smaller volume only.

8 Isospin-breaking: w_0-scale

In this section we derive a formula that gives the electromagnetic correction of the w_0-scale [6]. The starting point is the operator $W_\tau[U]$, which is the logarithmic derivative of the gauge-action density along the gradient flow [7]:

$$W_\tau[U] \equiv \frac{d(\tau^2 E[U, \tau])}{d \log \tau}, \tag{40}$$

where τ is the gradient flow time and E is a suitable discretization of the gluonic gauge action density. The expectation value of this operator defines the w_0-scale via

$$\left\langle W_{\tau = w_0^2(e)} \right\rangle = 0.3 . \tag{41}$$

Since $W_\tau[U]$ is a pure-gauge observable, it neither depends on the valence charge nor on fermion masses: the only isospin-breaking dependence in Equation (41) comes from the electric sea charge. The derivatives with respect to δm and the valence electric charge are zero. The expansion of the expectation value is given by Equations (30) and (31):

$$\langle W_\tau \rangle = \langle W_\tau \rangle_0 + e_s^2 \left\langle (W_\tau - \langle W_\tau \rangle_0) \frac{\mathrm{dets}_2''}{\mathrm{dets}_0} \right\rangle_0 \tag{42}$$

We also have to expand the w_0-scale:

$$w_0(e_s) = w_0 + e_s^2 \delta w_0 \tag{43}$$

and the W operator:

$$W_{\tau = w_0^2(e_s)} = W_{\tau = w_0^2} + e_s^2 \cdot 2 w_0 \delta w_0 \cdot \left. \frac{dW}{d\tau} \right|_{\tau = w_0^2} \tag{44}$$
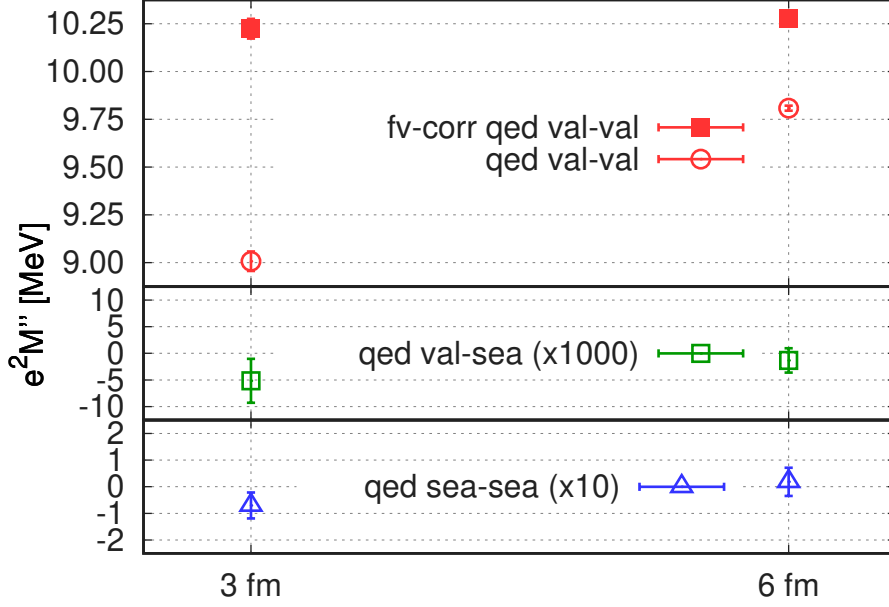
Figure 9: Volume dependence of various electromagnetic contributions to the π_+ mass. For the valence-valence contribution M_{20}'' we apply an infinite-volume correction given by Equation (53). The valence-sea M_{11}'' and sea-sea M_{02}'' contributions are multiplied by 1000 and 10 on the plot, respectively. The results are obtained with the 4stout action at $\beta = 3.7000$.

isospin component	meson #1 [fm]	meson #2 [fm]	omega #1 [fm]	omega #2 [fm]
M_0	$2.0 \ldots 3.5$	$2.5 \ldots 4.0$	$1.4 \ldots 2.0$	$1.5 \ldots 2.1$
M_{20}''	$1.5 \ldots 3.0$	$2.0 \ldots 3.5$	$1.2 \ldots 2.0$	$1.3 \ldots 2.1$
M_m'	$1.5 \ldots 3.0$	$2.0 \ldots 3.5$	$-$	$-$
M_{11}'', M_{02}''	$1.0 \ldots 2.5$	$1.5 \ldots 3.0$	$0.3 \ldots 1.5$	$0.6 \ldots 1.5$

Table 9: Plateau fit ranges for different isospin-breaking components, for mesons and for the Ω baryon.

Here, w_0 denotes the value of the w_0-scale at the isospin-symmetric point, which of course satisfies

$$\left\langle W_{\tau=w_0^2} \right\rangle_0 = 0.3 \tag{45}$$

From these we obtain the following formula for the electromagnetic correction:

$$\delta w_0 = - \left[\frac{1}{2\sqrt{\tau}} \left\langle \frac{dW_\tau}{d\tau} \right\rangle_0^{-1} \left\langle (W_\tau - \langle W_\tau \rangle_0) \frac{\mathrm{dets}_2''}{\mathrm{dets}_0} \right\rangle_0 \right]_{\tau=w_0^2} \tag{46}$$

Section 7 gives details on the fermion-determinant derivative computations. On our coarsest lattice spacing we computed the electromagnetic correction in two different volumes, $L = 3$ fm and $L = 6$ fm (see Table 8 for the ensemble parameters). We obtained for this correction $\delta w_0 = -0.018(2)$ on the small and $\delta w_0 = -0.018(3)$ on the large lattice, which are in perfect agreement. Even the isospin-symmetric values show no significant finite-size effect: we have $w_0 = 1.2899(9)$ on the small and $w_0 = 1.2908(2)$ on the large lattice. In our analyses, we use the isospin symmetric w_0 measured on the large lattices listed in Table 1; whereas for δw_0 we use the small volume ensembles listed in Table 8.

9 Isospin breaking: hadron masses

In this section we describe the procedure to obtain the isospin-breaking derivatives of a hadron mass M. On certain ensembles we measure the hadron propagator H at four different values of isospin breaking:

$$H_0, \quad H_+, \quad H_-, \quad H_{\delta m}. \tag{47}$$

The first is measured at the isospin-symmetric point, the second/third with valence electric charge $e_v = \pm\sqrt{4\pi\alpha_*}$ and zero quark mass difference $\delta m = 0$, and the fourth with $e_v = 0$ and $\delta m = 2m_l \left.\frac{1-r}{1+r}\right|_{r=0.485}$. These allow to calculate finite differences with respect to e_v and δm, whereas the e_s derivatives can be calculated exactly using the formulae in Equation (31). In these measurements both gluon and photon fields are fixed to Coulomb gauge, for the former the gauge fixing procedure is applied after smearing.

We use the notation $\mathcal{M}[\langle H \rangle]$ for the mass that is extracted from the hadron-propagator expectation value $\langle H \rangle$. At the isospin-symmetric point we have

$$M_0 = \mathcal{M}[\langle H_0 \rangle_0]. \tag{48}$$

The QED, sea-sea, isospin-breaking component of the propagator, $\langle H \rangle''_{02}$, is given by Equation (31). Then the derivative of the mass can be obtained by application of the chain rule:

$$M''_{02} = \left.\frac{\delta\mathcal{M}[H]}{\delta H}\right|_{\langle H_0 \rangle_0} \langle H \rangle''_{02} = \left.\frac{\delta\mathcal{M}[H]}{\delta H}\right|_{\langle H_0 \rangle_0} \left\langle (H_0 - \langle H_0 \rangle_0)\frac{\mathrm{dets}''_2}{\mathrm{dets}_0} \right\rangle_0, \tag{49}$$

where we use $\partial\langle H \rangle/\partial e_s = 0$ at the isospin-symmetric point. In the case of the valence-valence QED component we can build the derivative as a finite difference:

$$M''_{20} \approx \frac{1}{2e_v^2}\left(\mathcal{M}[\langle H_+ \rangle_0] + \mathcal{M}[\langle H_- \rangle_0] - 2\mathcal{M}[\langle H_0 \rangle_0]\right) = \frac{1}{e_v^2}\left(\mathcal{M}[\tfrac{1}{2}\langle H_+ + H_- \rangle_0] - \mathcal{M}[\langle H_0 \rangle_0]\right). \tag{50}$$

Here we used $\langle H_+ \rangle_0 = \langle H_- \rangle_0 = \frac{1}{2}\langle H_+ + H_- \rangle_0$. Working with the average propagator has the advantage that the $O(e_v)$ noise is absent in its expectation value [39]. Equation (50) gives an approximation that is valid up to, and including, $O(e_v^2)$ terms. We have confirmed that these terms are on the order of a few percent relative to M''_{20} with our choice of e_v. There is an analogous formula for the strong-isospin-breaking component:

$$M'_m \approx \frac{m_l}{\delta m}\left(\mathcal{M}[\langle H_{\delta m} \rangle_0] - \mathcal{M}[\langle H_0 \rangle_0]\right). \tag{51}$$

The QED valence-sea derivative is available in a mixed form: the derivation with respect to e_s is exact, but with respect to e_v, it is a finite difference. Applying these to $\mathcal{M}[\langle H \rangle]$ we get:

$$M''_{11} \approx \frac{1}{2e_v}\left[\left.\frac{\delta\mathcal{M}[H]}{\delta H}\frac{\partial\langle H \rangle}{\partial e_s}\right|_{e_s=0} - (e_v \to -e_v)\right] = \left.\frac{\delta\mathcal{M}[H]}{\delta H}\right|_{\langle H_+ + H_- \rangle_0}\left\langle \frac{H_+ - H_-}{2e_v}\frac{\mathrm{dets}'_1}{\mathrm{dets}_0} \right\rangle_0. \tag{52}$$

We now specify $\mathcal{M}[H]$, i.e. the way to extract the mass from the hadron propagator. Since we are interested in the small isospin breaking effects here, the choice of the mass extraction procedure is not crucial. We utilize a procedure based on effective masses instead of fitting the propagator to multiple exponentials. This has the advantage over the standard fitting procedure that the derivatives $\frac{\delta\mathcal{M}}{\delta H}$ can be computed easily, in particular they can be given in closed analytic form. Using two/four propagator points, an effective-mass value and its differential can be given in analytical form for mesons/baryons [40]. Then we fit a constant to the plateau of the effective mass. We choose two plateau ranges, so that a systematic error can be associated with finding the plateau. The ranges are given in Table 9.

In this work the strong-isospin-breaking M'_m and valence-valence contributions M''_{20} are evaluated on 4stout configurations with box sizes around $L = 6$ fm. For the sea contributions, M''_{11} and M''_{02}, we use

$X \longrightarrow$	C^{light}	C^{strange}	C^{charm}
X_0	$\frac{5}{9} C^{\text{conn}}(m_l, 0)$	$\frac{1}{9} C^{\text{conn}}(m_s, 0)$	$\frac{4}{9} C^{\text{conn}}(m_c, 0)$
X_1'	$\frac{7}{27} \left[\frac{\partial}{\partial e} C^{\text{conn}} \right] (m_l, 0)$	$-\frac{1}{27} \left[\frac{\partial}{\partial e} C^{\text{conn}} \right] (m_s, 0)$	$\frac{8}{27} \left[\frac{\partial}{\partial e} C^{\text{conn}} \right] (m_c, 0)$
X_2'	$\frac{17}{81} \left[\frac{1}{2} \frac{\partial^2}{\partial e^2} C^{\text{conn}} \right] (m_l, 0)$	$\frac{1}{81} \left[\frac{1}{2} \frac{\partial^2}{\partial e^2} C^{\text{conn}} \right] (m_s, 0)$	$\frac{16}{81} \left[\frac{1}{2} \frac{\partial^2}{\partial e^2} C^{\text{conn}} \right] (m_c, 0)$
X_m'	$-\frac{m_l}{6} \left[\frac{\partial}{\partial m_l} C^{\text{conn}} \right] (m_l, 0)$	$-$	$-$

Table 10: Isospin symmetric value X_0 and isospin-breaking derivatives X_1', X_2', X_m' of various observables X, namely the light, strange and charm connected contractions of the current propagator, in terms of the connected vector meson contraction and its derivatives. See Equation (59) and (61) for the definitions.

also the 4stout action, but on smaller volumes, $L = 3$ fm (see Section 7 for details of these ensembles).

In the presence of the electromagnetic interaction, hadron masses have a finite-size effect that is power-like in the size of the box. In some cases it can be much larger than the exponentially-suppressed, finite-size effect related to the strong interaction. For the QED_L photon action, the effect in the first two orders depends on the hadron only through its electric charge Q and mass M and is known analytically [32, 41]:

$$M(L) - M = -\frac{(Qe)^2 c}{8\pi} \left[\frac{1}{L} + \frac{2}{ML^2} + O(L^{-3}) \right] \quad \text{with} \quad c = 2.837297\ldots . \tag{53}$$

The first two orders of this formula can be used to correct for electromagnetic finite-size effects. Remaining $O(L^{-3})$ effects are beyond the precision of this work and are neglected. Since, for charged hadrons, sea effects are typically much smaller than the valence-valence contribution, we use the universal finite-size formula (53) to correct the valence-valence component and apply no correction to the rest. We can corroborate this choice by looking at the different isospin breaking components of the charged pion mass on two different volumes on our coarsest lattice, as shown in Figure 9. The corrected M_{20}'' values almost agree on the two volumes. On finer lattices we use the results of the $L = 6$ fm runs, correcting them with Equation (53). In the case of the sea contributions the uncorrected M_{11}'' and M_{02}'' data are consistent on the two volumes. On finer lattices we use the small-volume runs to estimate the electromagnetic sea effects without correcting for finite-volume effects.

10 Current propagator $\langle JJ \rangle$

In this section we consider in detail the definition and decomposition of the current propagator:

$$\langle J_{\mu,x} J_{\bar{\mu},\bar{x}} \rangle, \tag{54}$$

where $e J_\mu$ is the quark electromagnetic current. In the continuum limit this propagator can be obtained by coupling the quarks to an external photon field A_μ^{ext} and building the second differential with respect to this field. In our lattice regularization, we use this prescription to define the current propagator. Specifically the partition function in the presence of an external photon field is given by:

$$Z[A^{\text{ext}}] \equiv \int \ldots \ \text{dets}[U, A + A^{\text{ext}}; \{q_f\}, \{m_f\}, e]. \tag{55}$$

The current propagator is then defined as the following second differential:

$$\langle J_{\mu,x} J_{\bar{\mu},\bar{x}} \rangle \equiv \frac{1}{e^2} \frac{\delta^2 \log Z}{\delta A_{\mu,x}^{\text{ext}} \delta A_{\bar{\mu},\bar{x}}^{\text{ext}}} \bigg|_{A^{\text{ext}}=0}. \tag{56}$$

The so defined propagator satisfies current conservation on both the source and sink sides. To compute it, we need the first and second derivatives of the fermion matrix at zero external field:

$$\frac{\delta}{\delta A_{\mu,x}^{\text{ext}}} \bigg|_0 M_f = eq_f \cdot D_\mu[iP_x V_U e^{ieq_f A}],$$
$$\frac{\delta^2}{\delta A_{\mu,x}^{\text{ext}} \delta A_{\bar{\mu},\bar{x}}^{\text{ext}}} \bigg|_0 M_f = -e^2 q_f^2 \cdot \delta_{\mu\bar{\mu}} \cdot D_\mu[P_x P_{\bar{x}} V_U e^{ieq_f A}]. \tag{57}$$

From these we get the current propagator as follows:

$$\langle J_{\mu,x} J_{\bar{\mu},\bar{x}} \rangle = \left\langle \sum_f q_f^2 C_{\mu,x,\bar{\mu},\bar{x}}^{\text{conn}}(m_f, eq_f) + C_{\mu,x,\bar{\mu},\bar{x}}^{\text{disc}} - \sum_f \frac{q_f^2}{4} \text{Tr}\left(M_f^{-1} \ D_\mu[P_x P_{\bar{x}} V_U e^{ieq_f A}] \delta_{\mu\bar{\mu}} \right) \right\rangle \tag{58}$$

where the connected vector meson contraction is defined as

$$C_{\mu,x,\bar{\mu},\bar{x}}^{\text{conn}}(m,e) \equiv -\frac{1}{4} \text{Tr}\left(M^{-1} \ D_\mu[iP_x V_U e^{ieA}] \ M^{-1} \ D_{\bar{\mu}}[iP_{\bar{x}} V_U e^{ieA}] \right) \tag{59}$$

and the disconnected contraction as

$$C_{\mu,x,\bar{\mu},\bar{x}}^{\text{disc}} \equiv \sum_{f,\bar{f}} q_f q_{\bar{f}} I_{\mu,x}(m_f, eq_f) I_{\bar{\mu},\bar{x}}(m_{\bar{f}}, eq_{\bar{f}}) \quad \text{with} \quad I_{\mu,x}(m,e) \equiv \frac{1}{4} \text{Tr}\left(M^{-1} D_\mu[iP_x V_U e^{ieA}] \right). \tag{60}$$

In these formulas, Tr is the trace over color and spacetime indices and the P_x projection operator clears the components of a vector on all sites except for x. Here, the fermion matrix M is understood with mass m and on a gauge background $V_U e^{ieA}$. The M_f notation, defined in Equation (28), stands for the fermion matrix with m_f mass and eq_f charge. Due to gauge invariance, $\langle I_{\mu,x} \rangle = 0$. Equation (58) is our master formula for the current propagator. In the following we decompose it into several pieces. There are three terms. First is the connected contribution, second is the disconnected contribution and the third is a contact term. This last one gives no contribution to the observables that we are interested in and it will be omitted from now on. To obtain the expansion in Equation (30), we have to calculate the Wick-contractions C^{conn} and C^{disc} at the isospin-symmetric point and also their isospin-breaking derivatives.

It is common to split the connected part, $\sum_f q_f^2 C^{\text{conn}}(m_f, eq_f)$, into the contributions of individual flavors:

$$C^{\text{light}} \equiv \frac{4}{9} C^{\text{conn}}(m_u, \frac{2}{3}e) + \frac{1}{9} C^{\text{conn}}(m_d, -\frac{1}{3}e),$$
$$C^{\text{strange}} \equiv \frac{1}{9} C^{\text{conn}}(m_s, -\frac{1}{3}e),$$
$$C^{\text{charm}} \equiv \frac{4}{9} C^{\text{conn}}(m_c, \frac{2}{3}e), \tag{61}$$

where we suppressed Lorentz indices and coordinates for simplicity. The isospin-limit components of these, as defined in Section 5, in terms of C^{conn} and their derivatives are given in Table 10. For the disconnected contribution we give here the formulas for the isospin-symmetric point and for the strong-isospin-breaking term:

$$C_0^{\text{disc}} = \frac{1}{9} \left[I_{\mu,x}(m_l, 0) - I_{\mu,x}(m_s, 0) + 2I_{\mu,x}(m_c, 0) \right] \left[I_{\mu,x}... \to I_{\bar{\mu},\bar{x}}... \right],$$
$$[C^{\text{disc}}]'_m = -\frac{3m_l}{2} \frac{\partial}{\partial m_l} C_0^{\text{disc}} \tag{62}$$

The detailed implementation of these quantities will be given in Sections 12 and 14. From these we then construct the total expectation value as shown in Equations (30) and (31).

It is also common to split the propagator at the isospin-symmetric point into isospin singlet and triplet parts: $[\langle JJ \rangle]_0 = \langle JJ \rangle_{I=0} + \langle JJ \rangle_{I=1}$. These are given by

$$
\begin{aligned}
\langle JJ \rangle_{I=1} &\equiv \left[\langle \tfrac{9}{10} C^{\text{light}} \rangle\right]_0 = \langle \tfrac{1}{2} C^{\text{conn}}(m_l, 0) \rangle_0 \ , \\
\langle JJ \rangle_{I=0} &\equiv \left[\langle \tfrac{1}{10} C^{\text{light}} + C^{\text{strange}} + C^{\text{charm}} + C^{\text{disc}} \rangle\right]_0 = \\
&= \langle \tfrac{1}{18} C^{\text{conn}}(m_l, 0) + \tfrac{1}{9} C^{\text{conn}}(m_s, 0) + \tfrac{4}{9} C^{\text{conn}}(m_c, 0) + C_0^{\text{disc}} \rangle_0 \ .
\end{aligned}
\tag{63}
$$

Finally, we introduce the notation $G(t)$ for the zero-momentum timelike current propagator with averaged Lorentz indices:

$$
G(t) \equiv \frac{1}{3} \sum_{\vec{x},\mu=1,2,3} \frac{1}{2} \langle J_{\mu,t,\vec{x}} J_{\mu,0} + J_{\mu,T-t,\vec{x}} J_{\mu,0} \rangle .
\tag{64}
$$

This is the lattice version of the propagator given in Equation (1) of the main text. $G(t)$ can also be decomposed into connected terms of different flavors and a disconnected part. Note that the imaginary parts of these quantities are zero due to the gauge averaging.

11 Anomalous magnetic moment a_μ

In this section we provide the definition for the central observable of the paper: the leading-order hadronic vacuum polarization (LO-HVP) contribution to the anomalous magnetic moment of the muon, $a_\mu^{\text{LO-HVP}}$. Furthermore, we detail the decomposition of $a_\mu^{\text{LO-HVP}}$. Since we consider only the LO-HVP contribution, we drop the superscript and multiply the result by 10^{10}, ie. a_μ stands for $a_\mu^{\text{LO-HVP}} \times 10^{10}$ throughout this work.

The renormalized scalar hadronic vacuum polarization function (HVP) can be extracted from the zero momemtum current propagator $G(t)$ as [42]:

$$
\Pi(Q^2) \equiv \sum_{t=0}^{T/2} \left[t^2 - \frac{4}{(aQ)^2} \sin^2 \left(\frac{aQt}{2} \right) \right] G(t)
\tag{65}
$$

where t and $G(t)$ are given in lattice units here and a is the lattice spacing. This formula corresponds to a Fourier transformation followed by a division by Q^2, including an explicit removal of: 1. a pure finite-volume effect and 2. the ultraviolet divergence. Renormalization is performed on shell such that $\Pi(0) = 0$. While, in a finite volume, $\Pi(Q^2)$ is only formally defined at momenta with components that are integer mutilples of either $2\pi/L$ or $2\pi/T$, we use (65) to analytically continue $\Pi(Q^2)$ to any real values of Q^2. It is worth noting that this approach is related to the time-moment approach of [43]. Time moments can also be used as input in various approximants that were put forward in [44, 45]. These are based on the application of Mellin–Barnes techniques. They converge rapidly with the number of moments retained [45] and also allow for a systematic matching to perturbation theory at short distances.

In this work we compute all hadronic $O(e^2)$ effects in the vacuum polarization, including ones that are reducible under cutting a photon line. Like $G(t)$, $\Pi(Q^2)$ can also be decomposed into the connected contributions of various quark flavors and a disconnected contribution (Section 10).

The LO-HVP contribution to the anomalous magnetic moment of the muon is computed from the one-photon irreducible part of $\Pi(Q^2)$, denoted by $\Pi_{1\gamma I}(Q^2)$, using the following integral [46]:

$$
a_\mu = 10^{10} \alpha^2 \int_0^\infty \frac{dQ^2}{m_\mu^2} \, \omega \left(\frac{Q^2}{m_\mu^2} \right) \Pi_{1\gamma I}(Q^2)
\tag{66}
$$

where $\omega(r)$ is given after Equation (3) of the main paper, α is the fine structure constant renormalized in the Thomson-limit, and m_μ is the mass of the muon. The difference between having $\Pi(Q^2)$ and $\Pi_{1\gamma I}(Q^2)$ in the integral in (66) is the one-photon-reducible $(1\gamma R)$ contribution denoted by $a_\mu^{1\gamma R}$. It is an $O(e^2)$ effect that is included in the higher-order hadronic vacuum polarization (HO-HVP) contribution. This effect is tiny compared to a_μ and has already been computed on the lattice, as discussed in our final result section, Section 23.

We partition the momentum integral in Equation (66) by cutting it into two contributions at a momentum $Q_{\max}$. Below $Q_{\max}$ we use the lattice and above that perturbation theory. In the two parts the vacuum polarization is renormalized differently: it is renormalized to zero at $Q = 0$ on the lattice and at $Q = Q_{\max}$ in perturbation theory. This requires the introduction of an extra term, that accounts for this difference. The lattice part is then splitup into the contributions of different flavors. In detail, our partitioning takes the following form:

$$a_\mu = a_\mu^{\text{light}} + a_\mu^{\text{strange}} + a_\mu^{\text{charm}} + a_\mu^{\text{disc}} + a_\mu^{\text{pert}} - a_\mu^{1\gamma R} \ . \tag{67}$$

Here, the connected light contribution is given as

$$a_\mu^{\text{light}} = 10^{10}\alpha^2 \left[\int_0^{Q_{\max}^2} \frac{dQ^2}{m_\mu^2} \, \omega\left(\frac{Q^2}{m_\mu^2}\right) \Pi^{\text{light}}(Q^2) + \Pi^{\text{light}}(Q_{\max}^2) \int_{Q_{\max}^2}^\infty \frac{dQ^2}{m_\mu^2} \, \omega\left(\frac{Q^2}{m_\mu^2}\right) \right] , \tag{68}$$

and similarly for the other flavors[1] and the disconnected contributions. The second term accounts for the difference in the lattice and perturbative renormalization points, as explained above. Using Equation (65) we can express all of these contributions as a weighted sum of the corresponding current propagator. For the connected light term we have, for instance:

$$a_\mu^{\text{light}} = 10^{10}\alpha^2 \sum_{t=0}^{T/2} K(t; aQ_{\max}, am_\mu) \, G^{\text{light}}(t) \ , \tag{69}$$

where the kernel $K(t; aQ_{\max}, am_\mu)$ is given by Equations (65) and (68). It depends on the gauge ensemble only through the lattice spacing. The perturbative contribution is given by

$$a_\mu^{\text{pert}} = 10^{10}\alpha^2 \int_{Q_{\max}^2}^\infty \frac{dQ^2}{m_\mu^2} \, \omega\left(\frac{Q^2}{m_\mu^2}\right) \left[\Pi^{\text{pert}}(Q^2) - \Pi^{\text{pert}}(Q_{\max}^2) \right] \ . \tag{70}$$

In Reference [47] we demonstrated on our 4stout data set that switching to the perturbative calculation can be safely done for $Q_{\max}^2 \gtrsim 2$ GeV^2, ie. from this point on a_μ does not depend on the choice of $Q_{\max}$. In this work we use $Q_{\max}^2 = 3$ GeV^2. The perturbative part for this choice was computed in [47] and is given in Section 23, where the final result for a_μ is put together.

We also consider a modification of Equation (65), in which the current propagator is restricted to a certain region in time, from t_1 to t_2. To achieve this, we multiply the propagator by a smooth window function [48]

$$W(t; t_1, t_2) \equiv \Theta(t; t_1, \Delta) - \Theta(t; t_2, \Delta) \quad \text{with} \quad \Theta(t; t', \Delta) \equiv \tfrac{1}{2} + \tfrac{1}{2} \tanh[(t - t')/\Delta] \tag{71}$$

or equivalently we replace the weight factor as $K(t) \to K(t)W(t)$. We will focus on a particular window defined in Reference [48], with parameters $t_1 = 0.4$ fm, $t_2 = 1.0$ fm and $\Delta = 0.15$ fm. The corresponding contribution to the magnetic moment of the muon is denoted by $a_{\mu,\text{win}}^{\text{LO-HVP}}$ and for brevity we use $a_{\mu,\text{win}} = a_{\mu,\text{win}}^{\text{LO-HVP}} \times 10^{10}$. We can do the same partitioning as we did with a_μ in Equation (67). We will use those notations extended by a win subscript.

---

[1]The contributions of the bottom and the top are not shown explicitly here, they will be added in our final result section, Section 23.
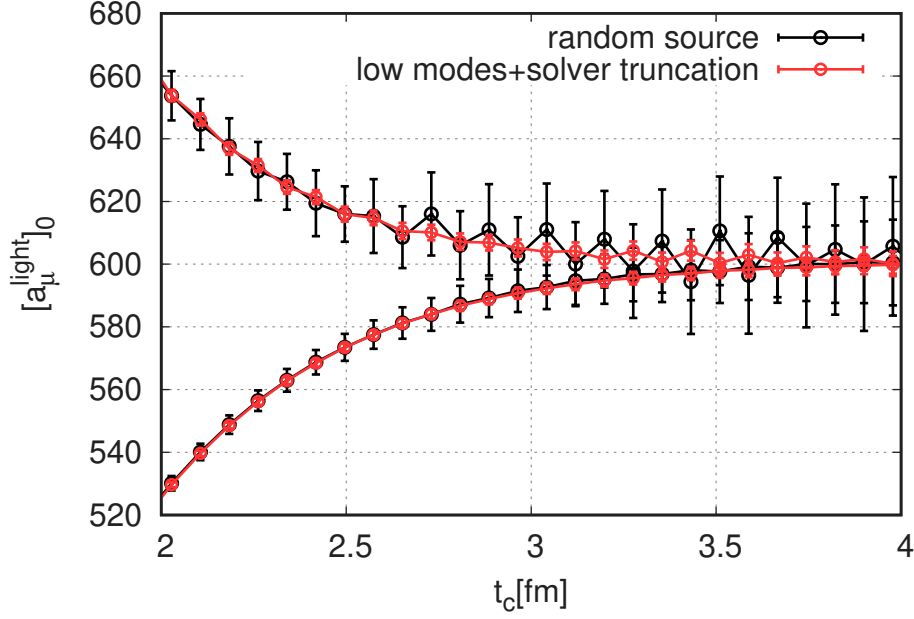
Figure 10: Comparison of a conventional random source based technique, as we applied it in our earlier work [47], and a low mode utilizing technique of this work on a $\beta = 3.9200$ `4stout` ensemble for the case of $[a_\mu^{\text{light}}]_0$ upper and lower bounds (see Section 13).

12   Noise reduction techniques

In this section we consider quantities at the isospin-symmetric point; noise reduction techniques for the isospin-breaking part are discussed in Section 14. For the strange and charm connected contributions, C_0^{strange} and C_0^{charm}, and for the disconnected contribution C_0^{disc} we use the same measurements that are presented in our previous work [47]. A new measurement procedure is implemented for the light connected component C_0^{light}. It is used to reanalyze the old configurations and make measurements on new ensembles. This plays a key role in reducing the final statistical error in a_μ.

The technique utilizes the lowest eigenmodes of the fermion matrix; for an early work with low eigenmodes, see [49]. The way in which we use these modes here is essentially the same as in [50], where it is called Low Mode Substitution. In the space orthogonal to these modes, the computational effort is reduced considerably by applying imprecise (aka. sloppy) matrix inversions. This is called the Truncated Solver Method [37] or All Mode Averaging [38]. Here we describe the technique for the connected part of the current propagator. The same technique was applied recently for magnetic moment computations in [51] also.

We consider the connected propagator of Equation (59) for timelike separation, and perform an averaging over the source positions, together with a zero spatial-momentum projection at the sink:

$$C(t, \bar{t}) \equiv \frac{1}{3L^3} \sum_{\vec{x}, \vec{x}, \mu=1,2,3} C_{\mu,x,\mu,\bar{x}}^{\text{conn}}(m_l, 0) = -\frac{1}{12L^3} \sum_{\mu=1,2,3} \text{ReTr}\left[\mathcal{D}_{\mu,t} M^{-1} \mathcal{D}_{\mu,\bar{t}} M^{-1}\right], \quad (72)$$

where $\mathcal{D}_{\mu,t} = \sum_{\vec{x}} D_\mu[iP_xU]$ is an operator that performs a symmetric, gauge-covariant shift on a vector v_x:

$$[\mathcal{D}_{\mu,t}v]_x = \begin{cases} x_4 = t : & i\eta_{\mu,x}\left(U_{\mu,x}v_{x+\mu} + U_{\mu,x-\mu}^\dagger v_{x-\mu}\right) \\ x_4 \neq t : & 0 \end{cases} \quad (73)$$

where $\eta_{\mu,x}$ are the Dirac-gamma matrices in the staggered representation. We use the simplifying notation

$\mathcal{D} = \mathcal{D}_{\mu,t}$ and $\overline{\mathcal{D}} = \mathcal{D}_{\mu,\bar{t}}$ in the following. In Equation (72), we apply the real part because the imaginary part vanishes anyway after averaging over gauge configurations.

Using the lowest eigenmodes of M we split the quark propagator into an eigenvector part and into its orthogonal complement, denoted by "e" and "r", respectively:

$$M^{-1} = M_e^{-1} + M_r^{-1}, \tag{74}$$

$$M_e^{-1} = \sum_i \frac{1}{\lambda_i} v_i v_i^\dagger \qquad \text{and} \qquad M_r^{-1} = M^{-1}\left(1 - \sum_i v_i v_i^\dagger\right), \tag{75}$$

where v_i/λ_i is the i-th eigenvector/eigenvalue of the operator M. For the projection we used a modified version of the symmetric Krylov-Schur algorithm described in [52]. Correspondingly, C splits into three terms:

$$C = C_{ee} + C_{re} + C_{rr}, \tag{76}$$

with eigen-eigen, rest-eigen and rest-rest contributions:

$$C_{ee} = -\frac{1}{4L^3}\mathrm{ReTr}\left[\mathcal{D}M_e^{-1}\overline{\mathcal{D}}M_e^{-1}\right], \tag{77}$$

$$C_{re} = -\frac{1}{4L^3}\mathrm{ReTr}\left[\mathcal{D}M_r^{-1}\overline{\mathcal{D}}M_e^{-1} + \mathcal{D}M_e^{-1}\overline{\mathcal{D}}M_r^{-1}\right], \tag{78}$$

$$C_{rr} = -\frac{1}{4L^3}\mathrm{ReTr}\left[\mathcal{D}M_r^{-1}\overline{\mathcal{D}}M_r^{-1}\right], \tag{79}$$

where an average over μ is assumed but not shown explicitly. The benefit of this decomposition is that the trace in the **eigen-eigen part** can be calculated exactly, and is thus equivalent to calculating the propagator with all possible sources in position space. This is the main ingredient for the noise reduction. Though no extra inversions are needed in this part, it has to be optimized carefully, since there is a double sum over the eigenmodes, where each term is a scalar product $v_i^\dagger \mathcal{D} v_j$. In the **rest-eigen part** we have terms $v_i^\dagger \mathcal{D} M_r^{-1} \overline{\mathcal{D}} v_i$ and also terms where $\mathcal{D}$ and $\overline{\mathcal{D}}$ are exchanged. Therefore, this part is only a single sum over the eigenmodes, and each term involves one matrix inversion. Note that these inversions are preconditioned by the eigenvectors, so they need much less iterations than standard inversions. Additionally, we speed up the inversions by running them with a reduced precision, and for some randomly selected eigenvectors we correct for the small bias by adding the difference between a high precision solver and the reduced precision one [37, 38]. Finally, the **rest-rest part** is evaluated using random source vectors ξ: we calculate $\xi^\dagger \mathcal{D} M_r^{-1} \overline{\mathcal{D}} M_r^{-1} \xi$, which requires two inversions per random source. The reduced precision inverter technique is used here too.

As an example we give here the algorithm parameters for one of the `4stout` ensembles at $\beta = 3.9200$: 1032 modes of the even-odd preconditioned Dirac operator are projected; the high precision inversion has 10^{-8} accuracy; the reduced precision inverter is capped at 400 conjugate gradient iterations; the bias correction is calculated with a frequency of $1/32$ and 384 random sources are chosen for the rest-rest term. With these choices, the eigen-eigen part is the dominant source of the error, and since it is already evaluated with all possible sources, we have reached the limit where the noise comes from the fluctuations between gauge configurations. Using this technique we achieve a factor of five improvement in the statistical error compared to our previous work [47]. There we applied a random source technique similar to the one that we now use for the rest-rest term. The number of random sources was 768 per configuration. The comparison of the result with the old and new techniques is shown in Figure 10.

The number of projected eigenmodes is around 1000 for all lattice spacings in the $L = 6$ fm boxes. The number of projected modes has to be scaled with the physical four-volume, to keep the magnitude of the eigen-eigen part constant. On the large `4HEX` lattice with $L = 11$ fm we project 6048 eigenmodes.
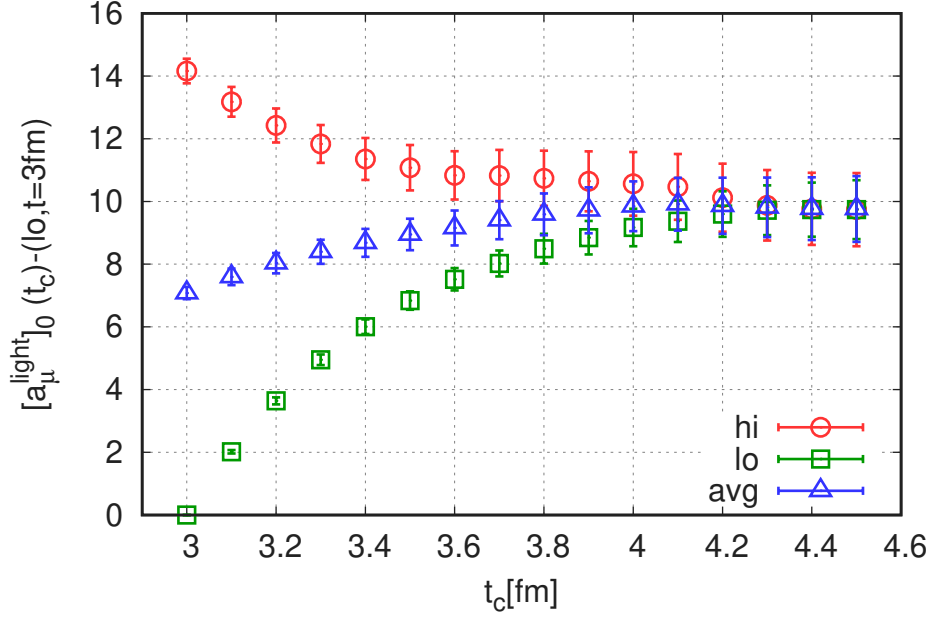
Figure 11: Upper and lower bounds on $[a_\mu^{\text{light}}]_0$ as a function of t_c, ie. the upper limit of the time integration in Equation (69). The lower bound value at 3.0 fm is subtracted and the plot shows the result of a combined fit of all ensembles that is evaluated at $a = 0.064$ fm (see text). Such plots are used to set the value of t_c on our ensembles. We use $t_c = 4.0$ fm everywhere.

13 Upper and lower bounds on $\langle JJ \rangle$

In the case of the light and disconnected contributions to the current propagator, the signal deteriorates quickly as distance is increased. To calculate the HVP, a sum over time of the propagator has to be performed, as Equation (69) shows. As was suggested in [53, 54], we introduce a cut in time t_c, beyond which the propagator is replaced by upper and lower bounds, thereby reducing the statistical noise. Our estimate is given by the average of the bounds at a t_c where the two bounds meet. In this section only isospin-symmetric quantities are considered.

Bounds are derived from the assumption that the current propagator is a sum of exponentials with positive coefficients. In the case of staggered fermions, where opposite parity states with oscillating coefficients give also a contribution, the assumption is only satisfied after some distance and within the statistical error. On our ensembles this is usually the case beyond about $t \sim 2.5$ fm.

For the **light connected propagator** at the isospin-symmetric point the bounds express the positivity (lower bound) and that the propagator should decay faster than the exponential of two pions (upper bound). They are given as

$$0 \leq G^{\text{light}}(t) \leq G^{\text{light}}(t_c) \frac{\varphi(t)}{\varphi(t_c)}, \tag{80}$$

where $\varphi(t) = \exp(-E_{2\pi}t)$. For $E_{2\pi}$ we use the energy of two non-interacting pions with the smallest non-zero lattice momentum $2\pi/L$. The larger the t_c the better the upper bound, but it comes with more statistical noise.

The exponential decay above assumes an infinite time extent, $T = \infty$. We incorporate the effects of a finite-T using next-to-leading-order chiral perturbation theory. There the exponential decay with the two-pion energy gets replaced by the following $\cosh$-type form:

$$\exp(-E_{2\pi}t) \longrightarrow \frac{\cosh[E_{2\pi}(t - T/2)] + 1}{\cosh(E_{2\pi}T/2) - 1} . \tag{81}$$

isospin component	t_c[fm] $X = G^{\text{light}}$	t_c[fm] $X = G^{\text{disc}}$
X'_m	2.5	2.5
X''_{20}	2.5	2.0
X''_{11}, X''_{02}	1.0	1.5

Table 11: Cuts in time for different isospin breaking components of the light and disconnected propagator. For each component we use two different cuts: the one that is given in the Table, and another that is 0.5 fm larger.

This is shown in detail in Section 15, where the above replacement corresponds to $\epsilon \to \epsilon_T$ in Equations (104) and (108). In this case an appropriate upper bound on the propagator is given by using the right hand side of Equation (81) as the bounding function $\varphi(t)$. We use this form of $\varphi(t)$ to compute the upper bounds in this work.

To obtain a suitable t_c on the 4stout ensembles, we combine the propagators of all ensembles in a single analysis. The reason for this is to use all available statistics to analyze the behavior of the bounds on t_c. The a_μ results of the ensembles depend on the pion and kaon masses and the lattice spacing. The first two dependencies can be safely eliminated if we consider only the tail of the propagator. (Remember, we are working close to the physical point.) For this we subtracted the value of the lower bound at 3.0 fm from both bounds on each ensemble. The lattice-spacing dependence was taken into account by making a continuum extrapolation for this "subtracted" a_μ at each value of t_c. The result of the continuum fit for the upper and lower bounds on $[a_\mu^{\text{light}}]_0$ is shown in Figure 11, in which the a^2 function used in the continuum extrapolation is evaluated at our finest lattice spacing, $a = 0.064$ fm. The two bounds meet around 4.0 fm. Here the statistical error of the average covers the central value of both bounds. We will use this average at $t_c = 4.0$ fm as our estimate for $[a_\mu^{\text{light}}]_0$ on the 4stout ensembles. Variation of t_c in the plateau region had negligible effect on the result. Note, that this value of t_c is much larger than the one that we used in our previous work, $t_c = 3.0$ fm. This improvement is made possible by the noise reduction technique of Section 12.

In the case of the isospin-symmetric **disconnected propagator** the bounds are

$$0 \leq -G^{\text{disc}}(t) \leq \tfrac{1}{10} G^{\text{light}}(t_c) \frac{\varphi(t)}{\varphi(t_c)} + G^{\text{strange}}(t) + G^{\text{charm}}(t). \tag{82}$$

Since the strange and charm terms fall off much faster than the light and disconnected one, their contribution does not change the value of t_c obtained. We use the same measurements for G_0^{disc} as in our previous work [47], and take the average of the bounds at a single cut value: $t_c = 2.5$ fm. This choice is in accordance with our findings in [47], that the variation in t_c within the plateau of the average bound has a negligible effect on the result.

14 Isospin-breaking effects in $\langle JJ \rangle$

In this section we describe the procedure that we use to compute the isospin-breaking corrections to the current propagator. We consider the contribution of the light and strange quarks only. The charm quark contribution was computed on the lattice in [55].

We start with the **connected contributions** (see Table 10). The electric derivatives in those formulas, X'_1 and X''_2, are measured by finite differences, as in the case of the isospin-breaking in hadron masses. Specifically, we compute the following contractions

$$C^{\text{conn}}(m_s, 0), \quad C^{\text{conn}}(m_s, +\tfrac{1}{3}e_*), \quad C^{\text{conn}}(m_s, -\tfrac{1}{3}e_*) \tag{83}$$
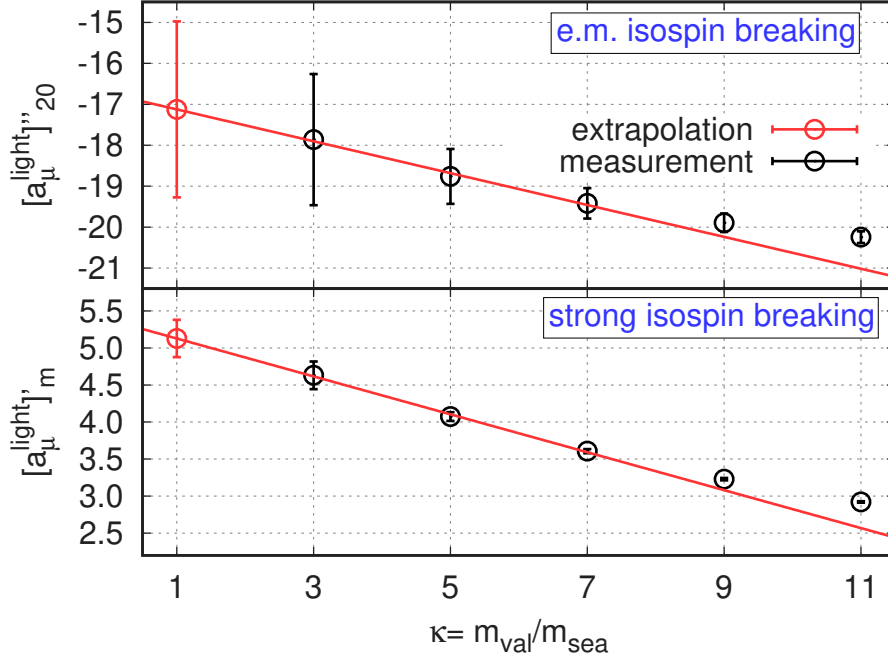
Figure 12: Extrapolation procedure to obtain the electromagnetic and strong-isospin-breaking corrections to the light connected contribution. Measurements are performed with valence over sea quark mass ratios of $\kappa = \{3, 5, 7, 9, 11\}$. We use the three lightest κ values to obtain the value at $\kappa = 1$ with a linear extrapolation shown in the Figure. The data on the plot is taken from our coarsest lattice, corresponding to $\beta = 3.7000$.

for the strange quark and

$$C^{\mathrm{conn}}(\kappa m_l, 0), \quad C^{\mathrm{conn}}(\kappa m_l, +\tfrac{1}{3}e_*), \quad C^{\mathrm{conn}}(\kappa m_l, -\tfrac{1}{3}e_*) \tag{84}$$

for the light quark, where e_* is the physical value of the electric coupling. From these the finite-difference approximators of the first and second derivatives can be built in the standard way. In the light quark case, the $G^{\mathrm{conn}}(m_l)$ propagators are noisy. Instead of the low-mode technique of Section 12, we use a simpler approach to reduce the noise, which is sufficiently accurate. We perform computations with valence quark masses that are some multiple κ of the sea quark mass, m_l, and then implement a chiral extrapolation to the target point at $\kappa = 1$. We measure the contractions in Equation (84) at five different values of κ and use the three lightest of these, $\kappa = 3, 5, 7$, to perform a linear extrapolation gauge-configuration by gauge-configuration. Figure 12 shows the result of this chiral extrapolation procedure on our coarsest ensemble for the case of $[a_\mu^{\mathrm{light}}]''_{20}$. A quadratic extrapolation including all five κ values gives a result that is consistent with the $\kappa = 3, 5, 7$ linear extrapolation within statistical uncertainty.

For the strong-isospin derivative of the connected contraction $[C^{\mathrm{light}}]'_m$ (see last line of Table 10) we implement directly the operator corresponding to the mass derivative. Again we use a chiral extrapolation from non-physical valence quark masses with $\kappa = 3, 5, 7$, similar to the case of the electromagnetic derivative. This extrapolation procedure is also plotted in Figure 12 for $[a_\mu^{\mathrm{light}}]'_m$.

Finally the sea-valence and sea-sea electromagnetic derivatives of the connected part, $[a_\mu^{\mathrm{light}}]''_{11/02}$ and $[a_\mu^{\mathrm{strange}}]''_{11/02}$, are measured on the set of ensembles that are dedicated to dynamical QED effects (see Table 8). There we combine the low-mode, noise-reduction technique with the mass extrapolation procedure above to reach sufficient accuracy.

In order to further reduce the noise in the light quark isospin-breaking corrections, when computing quantities like a_μ, we set the propagator to zero after some cut t_c. The value of t_c is obtained by looking for a plateau in the derivative of a_μ^{light} as a function of t_c. For the different isospin-breaking derivatives these values are given in Table 11. To assess the systematic error of the procedure, we also use a second value of t_c that is 0.5 fm larger. Let us note here that, when computing various derivatives of a_μ, not

D	N_L	N_4	N_5	N_6	
1	0	0	0	0	LO
2	0	1	0	0	NLO
	1	0	0	0	
5/2	0	0	1	0	N$\sqrt{\overline{\text{N}}}$LO
3	0	2	0	0	
	0	0	0	1	NNLO
	1	1	0	0	
	2	0	0	0	

Table 12: Solutions to the power counting formula of Reference [56] for the current-current correlator. For an NNLO calculation, diagrams of dimension $D \leq 3$ will contribute. Each solution describes diagrams of dimension D having N_L loops, along with some specified number of vertices from the Lagrangians of each order denoted by $N_{4,5,6}$. The number of vertices from the leading order Lagrangians N_2 is only limited by the number of loops.

only do we have to consider the derivatives of the propagator, but also those of the lattice scale, which enters in the weighting function $K(t; Q_{\max}a, m_\mu a)$ (see Equation (69)).

Now let us turn to the **disconnected contribution**. The basic operator for this measurement is the trace of the quark propagator, $I(m, e)$ of Equation (60). This is computed with the help of the low-lying eigenmodes of the Dirac operator, in a similar way to the calculation of the connected diagram $C^{\text{conn}}(m, e)$, described in Section 12. In this case, the computation is technically much simpler, since in $I(m, e)$ there is only one quark propagator under the trace, whereas in $C^{\text{conn}}(m, e)$ there are two. The eigenvectors depend on the electric coupling, and we compute them for each value of e that we need. For the electromagnetic derivative it is useful to rewrite the current

$$\sum_{f=\{u,d,s\}} q_f I(m_f, q_f e) = \tfrac{2}{3} I(m_l, \tfrac{2}{3}e) - \tfrac{1}{3} I(m_l, -\tfrac{1}{3}e) - \tfrac{1}{3} I(m_s, -\tfrac{1}{3}e) \tag{85}$$

using a Taylor expansion around $e = 0$ in the following way:

$$\sum_{f=\{u,d,s\}} q_f I(m_f, q_f e) = -2I(m_l, 0) + 2I(m_l, \tfrac{1}{3}e) + \tfrac{1}{3}I(m_l, -\tfrac{1}{3}e) - \tfrac{1}{3}I(m_s, -\tfrac{1}{3}e) + O(e^3). \tag{86}$$

The advantage of this form is that we can compute the first and second derivatives without having to compute the traces with $\pm\tfrac{2}{3}e$ charge. For the strong isospin derivative, where only the first derivative is needed, it suffices to compute only one additional trace at a slightly different mass than m_l. We use $0.9 \cdot m_l$. Altogether we measure traces at the following masses and electric couplings:

$$\begin{gathered} I(m_l, 0), \qquad I(m_s, 0), \\ I(0.9 \cdot m_l, 0), \\ I(m_l, +\tfrac{1}{3}e_*), \qquad I(m_l, -\tfrac{1}{3}e_*), \qquad I(m_s, +\tfrac{1}{3}e_*), \qquad I(m_s, -\tfrac{1}{3}e_*), \end{gathered} \tag{87}$$

from which all the isospin-breaking derivatives can be constructed using finite differences. All these traces are measured using the same set of random vectors, so that the random noise does not wash out the signal in the differences.

15 Staggered chiral perturbation theory

In this section we consider the current propagator in staggered chiral perturbation theory (SXPT) to next-to-next-to-leading order (NNLO). The goal of this effort is to describe the taste violation and finite-size effects in our lattice simulations. We work in the isospin-symmetric limit throughout this section.

In continuum chiral perturbation theory and in momentum space, the NNLO contribution was computed in [57, 58], and the finite-volume corrections are given in [59]. A coordinate space computation including finite-volume effects was given recently in [51]. In staggered chiral perturbation theory, only the next-to-leading order (NLO) has been computed [60]. Here we work out the following order and give the final result in the coordinate space representation. This order requires the NLO contributions to the staggered chiral Lagrangian, which is given in [61]. This Lagrangian has already been used to compute pseudoscalar meson masses to NLO [56]. This result will be an important ingredient here: just as in continuum chiral perturbation theory, the current propagator to NNLO can be considerably simplified if one writes it in terms of masses including NLO corrections.

The SXPT Lagrangian relevant to our computation is given by

$$\mathcal{L} = \mathcal{L}_2 + \mathcal{L}_{2,LS} + \mathcal{L}_4 + \mathcal{L}_{4,SV} + \mathcal{L}_5 + \mathcal{L}_6 \ , \tag{88}$$

with $\mathcal{L}_{2,4,6}$ the standard continuum LO, NLO and NNLO Lagrangians of Gasser and Leutwyler [62] and Bijnens, Colangelo and Ecker [63]. We denote $\mathcal{S}_i = \int dx \ \mathcal{L}_i$ the corresponding actions, $i \in \{2; 2, LS; 4; 4, SV; 5; 6\}$. We use the standard SXPT power counting scheme of Reference [64], whereby the LO contributions are $O(p^2) \approx O(m_q) \approx O(a^2)$, where p stands for a derivative operation, m_q is the light quark mass and a is the lattice spacing. According to this counting there is a LO staggered correction $\mathcal{L}_{2,LS}$ given by the Lagrangian of Lee and Sharpe [64] and a NLO staggered correction $\mathcal{L}_{4,SV}$ given by Sharpe and van de Water [61]. There is also a staggered specific contribution between NLO and NNLO, $\mathcal{L}_5$. More details on these terms are given below. There are further $O(a^2)$ terms in the Lagrangian, which –differently from the staggered corrections– are invariant under the continuum taste symmetry. We are not giving them explicitly here; their effect is to change the low-energy constants of the theory by $O(a^2)$ amounts – at least to the order that we consider here.

A general power counting formula is provided in the Appendix of [56]. Under this scheme, we say that a diagram $\mathcal{M}(p, m_q, a^2)$ has dimension D if it scales as $\mathcal{M}(p, m_q, a^2) \rightarrow \lambda^D \mathcal{M}(p, m_q, a^2)$ under a rescaling of the external momenta, quark masses and lattice spacing by $p \rightarrow \sqrt{\lambda}p$, $m_q \rightarrow \lambda m_q$, and $a^2 \rightarrow \lambda a^2$. In Table 12, we enumerate the contributions that are required for our NNLO computation under this counting scheme. Note that the LO contribution is zero.

The field variables are denoted by ϕ. They describe all the flavors and tastes of staggered pions. They are expressed as linear combinations of the product of the Hermitian generators of the $U(4)$ taste group and of the $U(N)$ replicated flavor group, ie. $\phi \equiv \sum_{\alpha a} \phi_{\alpha a} \xi_\alpha T_a$. Here the taste index α runs over the 16 element set $\{5, \mu5, \mu\nu, \mu, I\}$ with $\mu < \nu$. A possible representation for the taste generators can be built from the ξ_μ Dirac-matrices and the 4×4 identity matrix as:

$$\xi_\alpha \in \{ \quad \xi_5, \quad \xi_{\mu5} = i\xi_\mu\xi_5, \quad \xi_{\mu\nu} = i\xi_\mu\xi_\nu, \quad \xi_\mu, \quad \xi_I = 1 \quad \} \tag{89}$$

For the $U(N)$ generators T_a we use the generalized Gell-Mann matrices and the a index runs from 0 to $N^2 - 1$. We work with N_f degenerate flavors and rooting is implemented with the replica trick, ie. $N = N_f N_r$, with $N_r \rightarrow \frac{1}{4}$ at the end of the computation. In most cases, ϕ appears in an exponential form in the Lagrangian, $U = \exp(i\phi/F)$, where F is the pion decay constant in the chiral limit. Traces are taken over both the taste and replica-flavor indices. Generators are normalized as $\operatorname{tr}(\xi_\alpha T_a \cdot \xi_\beta T_b) = 2\delta_{ab}\delta_{\alpha\beta}$.

The current propagator is obtained by incorporating an external Hermitian vector field $v_\mu = v_\mu^\dagger$ in the Lagrangian, setting $v_\mu = QA_\mu$ (where Q is the charge matrix) and taking the second order functional

derivative of the partition function with respect to A_μ:

$$\langle J_\mu(x) J_{\bar\mu}(\bar x) \rangle \equiv \frac{\delta^2 \log Z[v_\nu = QA_\nu]}{\delta A_\mu(x)\, \delta A_{\bar\mu}(\bar x)} \bigg|_{A=0}$$
$$= \left\langle \frac{\delta \mathcal{S}\,[v_\nu = QA_\nu]}{\delta A_\mu(x)} \frac{\delta \mathcal{S}\,[v_\nu = QA_\nu]}{\delta A_{\bar\mu}(\bar x)} - \frac{\delta^2 \mathcal{S}\,[v_\nu = QA_\nu]}{\delta A_\mu(x)\, \delta A_{\bar\mu}(\bar x)} \right\rangle \tag{90}$$

up to vanishing disconnected terms. In the following we take the current to operate as a taste singlet, so Q is proportional to the identity matrix in taste space, that is $Q = Q_a \xi_I T_a$. This restriction is discussed in more detail at the end of the computation. In flavor space, Q is diagonal but non-singlet. The field-strength tensor of the electromagnetic vector potential is denoted by $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

In the first term of Equation (90), we will refer to the derivative $\frac{\delta \mathcal{S}}{\delta A_\mu}$ as current term. The second term in Equation (90) gives rise to contact terms proportional to $\delta(x - \bar x)$ and derivatives thereof. The anomalous magnetic moment considered in this work is obtained by integrating the propagator with a kernel function, that behaves as $(x - \bar x)^4$ for small differences. This eliminates all contact terms with less than four derivatives. The smallest order they can enter is therefore $O(p^6)$, since the two photon fields add an extra p^2 to the counting.

Leading-order contributions

The $O(p^2)$ and $O(m_q)$ terms are given by the standard Euclidean chiral Lagrangian, with an additional mass term for the taste-flavor singlet

$$\mathcal{L}_2 = \frac{F^2}{4} \mathrm{tr}\left(D_\mu U D_\mu U^\dagger\right) - \frac{F^2}{4} M^2 \mathrm{tr}\left(U + U^\dagger\right) + \frac{m_0^2}{12} (\mathrm{tr}\phi)^2, \tag{91}$$

where M^2 is the tree-level Goldstone-boson mass and $D_\mu = \partial_\mu U - i[v_\mu, U]$ is the covariant derivative including the vector field. Note that the external vector field is $O(p)$ in the chiral power counting scheme. The singlet meson mass has to be sent to infinity at the end of the computation ($m_0 \to \infty$). The functional derivatives described above give the current couplings

$$\frac{\delta \mathcal{S}_2}{\delta A_\mu(x)} = \frac{i}{2} \mathrm{tr}\left(Q\,[\partial_\mu\phi, \phi]\right) - \frac{i}{24F^2} \mathrm{tr}\left([Q, \partial_\mu\phi]\,\phi\phi - 3Q\phi\,[\partial_\mu\phi, \phi]\,\phi\right) + O(\phi^6). \tag{92}$$

The leading order staggered terms have been described by Lee and Sharpe [64] and generalized to multiple flavors by Aubin and Bernard [65]. We write these terms as $\mathcal{L}_{2,LS}$. Since these are $O(a^2)$, they can contribute to our NNLO calculation in diagrams with up to two loops, so when expanding in ϕ, we consider terms up to and including $O(\phi^4)$. The $O(\phi^2)$ terms can be absorbed into the $O(\phi^2)$ mass term from $\mathcal{L}_2$, providing taste-dependent corrections to the tree-level mass. There are also extra terms for the flavor singlets of each taste, shifting these masses separately from the other flavor components. Since $\mathcal{L}_{2,LS}$ does not depend on the external field, there are no contributions to the current terms.

Next-to-leading order contributions

The $O(p^4)$, $O(p^2 m_q)$, and $O(m_q^2)$ vertices have been described by Gasser and Leutwyler [62]. We denote this Lagrangian by $\mathcal{L}_4$, and use the standard notation for the low-energy constants. According to Table 12 they contribute to the calculation in diagrams with up to one loop, so when expanding in ϕ, we only consider terms of $O(\phi^2)$. These terms can mostly be absorbed into redefinitions of the coefficients for the mass term, and a rescaling of the field variable ϕ. Such a rescaling cannot affect the final result, since we are computing the correlation function of an external field. The rescaling of the field variable absorbs the contributions to the current term from the terms in the Lagrangian that are proportional to

the low-energy constants L_4 and L_5 and only the following term remains:

$$\frac{\delta \mathcal{S}_4}{\delta A_\mu(x)} = \frac{2i}{F^2} \, L_9 \, \mathrm{tr} \left(Q \partial_\nu \left[\partial_\mu \phi, \partial_\nu \phi \right] \right) + O(\phi^4). \tag{93}$$

The staggered specific NLO contributions involve $O(a^2 p^2)$, $O(a^2 m_q)$, and $O(a^4)$ terms, which have been described by Sharpe & van de Water [61]. As NLO terms, these can only contribute at one loop, so we only consider their expansion up to $O(\phi^2)$. Similar to $\mathcal{L}_{2,LS}$, most of the terms of $\mathcal{L}_{4,SV}$ can be absorbed into the LO mass as taste-dependent corrections and by field rescaling. There are three groups of terms that cannot be treated this way:

1. In the first group we have $O(a^2 p^2)$ terms that violate the remnant $SO(4)$ taste symmetry of the LO action, for example $\sum_\mu \langle \partial_\mu \phi \xi_\mu \partial_\mu \phi \xi_\mu \rangle$. The net effect of these terms on the current propagator is only to change the mass that appears in the pion propagators by $SO(4)$ violating terms. In our lattice simulations these $SO(4)$ violations are tiny: on our coarsest lattice the pion mass splittings, within the $SO(4)$ multiplets, are about 50 times smaller than the splittings between different multiplets. We will set them to zero in our SXPT computation.

2. In the second group we have $O(a^2 p^2)$ terms that depend explicitly on the external field, and not through the covariant derivatives. Reference [61] calls these "extra source-terms". In principle, these can contribute to the current terms. However, they are all proportional to the commutator of the vector field with a taste matrix ξ_α, and since we take our vector current to be a taste singlet, such commutators vanish. An example for such a term is $\sum_{\mu,\nu} \langle [A_\mu, \xi_\nu] \phi \xi_\nu \partial_\mu \phi \rangle$.

3. Finally, there might be taste-symmetry-breaking terms containing the field-strength tensor $F_{\mu\nu}$. A spurion analysis, similar to the one in Section III of Reference [60], indicates that no such terms are allowed in the $\mathcal{L}_{4,SV}$ Lagrangian.

Contributions from beyond next-to-leading order

In SXPT, it is possible to construct a chiral Lagrangian between NLO and NNLO, as discussed by Bailey, Kim and Lee [56]. We denote this Lagrangian $\mathcal{L}_5$. In it, terms might arise from the dimension-8 or dimension-9 Lagrangians in the Symanzik effective theory, that are either $O(a^4 p)$ or $O(a^5)$. In order to contribute to our calculation, vertices at this order would need to appear in a tree-level diagram (see Table 12, where we denote it N$\sqrt{\text{N}}$LO). As a result, the only terms that could contribute to the current propagator would be contact terms. However, such terms require the square of the vector field $A_\mu \approx O(p)$ in order to have a non-zero second functional derivative and therefore they must be $O(p^2)$. We conclude that $\mathcal{L}_5$ gives no contribution in this calculation.

Finally, we come to the NNLO Lagrangian. Similarly to the Bailey-Kim-Lee Lagrangian discussed above, NNLO terms can only contribute to tree-level diagrams through contact terms at $O(\phi^0)$. The continuum terms are described by Bijnens, Colangelo and Ecker [63], and we write them as $\mathcal{L}_6$. As we already mentioned, in our observables only contact terms that are at least $O(p^6)$ can contribute, and the only such term is:

$$\frac{\delta^2 \mathcal{S}_6}{\delta A_\mu(x) \delta A_{\bar{\mu}}(\bar{x})} = C_{93} \, \mathrm{tr}(Q^2) \, (\delta_{\mu\bar{\mu}} \partial^2 - \partial_\mu \partial_{\bar{\mu}}) \partial^2 \delta(x - \bar{x}) + O(\phi^2) \,. \tag{94}$$

The $O(a^2)$ taste-violating contributions at NNLO must similarly be contact terms at $O(\phi^0)$. As mentioned above, such terms must be at least $O(p^6)$ in order to contribute to the a_μ. However, $O(a^2 p^6)$ terms can enter only beyond NNLO, so they are not relevant in our calculation.

Infinite-L and T result

Performing the computations with the Lagrangians from the previous subsections we obtain the full current propagator. As we mentioned in the beginning, it is also necessary to compute the NLO mass shift δM_α^2 for each taste arising from the terms not absorbed into the tree-level mass M_α^2.

To arrive to the propagator that is used in our computation, we take Equation (90), apply a spatial integral over x and assume that $\mu, \bar{\mu}$ are spatial. With all the terms enumerated above we end up with:

$$
\int d^3x \, \langle J_\mu(\vec{x}, t) J_{\bar{\mu}}(0) \rangle =
$$

$$
= \sum_\alpha \int \frac{d^3p}{(2\pi)^3} 2NQ_{\text{ns}}^2 \left[1 + \frac{16}{F^2} L_9 E_{p,\alpha}^2 - \frac{N}{4F^2} \sum_\beta \mathcal{G}_{0,\beta} + \delta M_\alpha^2 \frac{d}{dM_\alpha^2} \right] \frac{e^{-2E_{p,\alpha}t}}{E_{p,\alpha}^2} p_\mu p_{\bar{\mu}} +
$$

$$
+ \sum_{\alpha,\beta} \int \frac{d^3p}{(2\pi)^3} \frac{d^3r}{(2\pi)^3} \frac{N^2 Q_{\text{ns}}^2}{8F^2} \frac{e^{-2E_{r,\beta}t} E_{p,\alpha} - e^{-2E_{p,\alpha}t} E_{r,\beta}}{E_{p,\alpha}^2 E_{r,\beta}^2 (E_{p,\alpha}^2 - E_{r,\beta}^2)} p_\mu r_{\bar{\mu}} (\vec{p} \cdot \vec{r}) + \text{contact terms},
$$

(95)

where we define the relativistic energy of a free particle with α taste as $E_{p,\alpha} = \sqrt{M_\alpha^2 + \vec{p}^2}$. The summations mean a sum over sixteen tastes of the flavor non-singlet pions, $\alpha \in \{5, \mu 5, \mu \nu, \mu, I\}$. The flavor-singlet pseudoscalars only contribute to the δM_α^2 mass-shift terms. These can be transformed away by switching to the NLO mass everywhere in the formula, ie. applying the shift $M_\alpha^2 \to M_\alpha^2 + \delta M_\alpha^2$. This changes the result by effects that are higher order than the NNLO considered here. The non-singlet charge squared is defined as $Q_{\text{ns}}^2 = Q_a Q_a - Q_0 Q_0$. Since the result is proportional to Q_{ns}^2, the flavor-singlet part of the current gives no contribution at this order.

We use dimensional regularization at scale μ and the $\overline{\text{MS}}$ scheme to work with the ultraviolet divergent loop integrals. Specifically, the one-loop integral $\mathcal{G}_{0,\beta}$ is given by:

$$
\mathcal{G}_{0,\beta} = \int \frac{d^3p}{(2\pi)^3} \frac{1}{2E_{p,\beta}} = \frac{M_\beta^2}{16\pi^2} \left(\log \frac{M_\beta^2}{\mu^2} + R \right)
$$

(96)

where R contains the divergence isolated by the $\overline{\text{MS}}$ prescription, see eg. [66]. Seemingly, there is also a singularity in the integrand of the double loop integral in the second line of Equation (95), when $\vec{p} = \vec{r}$. This singularity is superficial, since there is a zero in the numerator which cancels it. It is useful to work with the terms of the numerator separately, in which case the separated terms are singular. These singularities have to be regulated. We do this by adding an $+i\eta$ into the denominator. The full expression has to be smooth as $\eta \to 0$, and we take this limit at the end of the computation.

The only low-energy constant, that appears in the result, is L_9 from the continuum Gasser-Leutwyler Lagrangian $\mathcal{L}_{4,GL}$. The ultraviolet divergences coming from the single loop integral $\mathcal{G}_{0,\beta}$ and the double loop integral on the second line of Equation (95) can all be absorbed into L_9. This procedure defines a renormalized L_9^r, and also a scheme-independent $\overline{L}_9$ in the standard way [66].

The only contact term that affects our observables is the one proportional to C_{93}; we will ignore this term here, since it has no effect on the finite volume and taste-splitting corrections.

We now move to our specific case of two degenerate light quarks with rooting, so we set $N_f = 2$, $N_r = \frac{1}{4}$, $L_9 = -\frac{1}{2}l_6$ and $Q = (\frac{1}{6} + \frac{1}{2}\sigma_3) \otimes 1$ in flavor-replica space, which give $Q_{\text{ns}}^2 = N_r/4$. One of the two momentum integrals in Equation (95) can be performed analytically, and we also average over the Lorentz indices to arrive to our final expression:

$$
G(t) \equiv \frac{1}{3} \sum_{\mu=1,2,3} \int d^3x \, \langle J_\mu(\vec{x}, t) J_\mu(0) \rangle = \frac{1}{3} \sum_\alpha{}' \int \frac{d^3p}{(2\pi)^3} \frac{e^{-2E_{p,\alpha}t}}{E_{p,\alpha}^2} p^2 \left[1 + \frac{1}{F^2} \sum_\beta{}' \Gamma(p^2; M_\alpha, M_\beta) \right], \quad (97)
$$

where $p^2 = \vec{p}^{\,2}$ and the summation symbol $\sum_\alpha' = \frac{1}{16} \sum_\alpha$ stands for averaging over the taste index α. The first part in the square-bracket is the well-known NLO expression of [60], ie. the taste average of the

continuum NLO result. The second part of the square-bracket contains the NNLO correction and is given explicitly as

$$\Gamma(p^2; M_\alpha, M_\beta) = \frac{p^2 + M_\alpha^2}{12\pi^2}\left(\bar{l}_6 - \log\frac{M_\beta^2}{M_5^2}\right) + \frac{5(p^2 + M_\alpha^2)}{36\pi^2} - \frac{M_\beta^2}{6\pi^2} +$$

$$-\frac{p^2 + M_\alpha^2 - M_\beta^2}{6\pi^2}\begin{cases}\sqrt{\frac{1-x}{x}}\arcsin\sqrt{x} & \text{if } x < 1 \\ \sqrt{\frac{x-1}{x}}\log(\sqrt{x} + \sqrt{x-1}) & \text{if } x > 1\end{cases} \quad \text{with} \quad x = \frac{p^2 + M_\alpha^2}{M_\beta^2}. \qquad (98)$$

The scheme independent $\bar{l}_6$ is introduced with the Goldstone pion mass M_5 as

$$l_6 = -\frac{1}{96\pi^2}\left(\bar{l}_6 + \log\frac{M_5^2}{\mu^2} + R\right). \qquad (99)$$

Equations (97) and (98) reproduce the continuum NNLO result of [51], if we set the masses of all tastes to M_5.

When we use the above result later in this paper, we take the flavor non-singlet pion masses from the simulations and fix the remaining two parameters as: $F = 92.21$ MeV and $\bar{l}_6 = 16$ as in [51].

Finite-L effects

The above computation can also be carried out in finite volume. In the continuum case, this was already done in [51]. Here we generalize those formulas in the presence of taste violations. We use the same techniques that were applied there, and we also correct that computation. A detailed derivation will not be given here, we just briefly describe the main strategy and give the results.

The current propagator $G(t)$ in finite volume can be obtained from the one in infinite volume by replacing the momentum integrals with sums. In the infinite-volume expression, Equation (95), there are terms with single and double integrals. The finite-size correction for those with a single integral can be obtained from a Poisson summation formula, formally:

$$\frac{1}{L^3}\sum_p - \int\frac{d^3p}{(2\pi)^3} = \int\frac{d^3p}{(2\pi)^3}\sum_{n\neq 0}e^{i\vec{n}\vec{p}L} \equiv \sum\!\!\!\!\!\!\!\int_{p,n\neq 0}, \qquad (100)$$

whereas for double integrals the correction involves more terms:

$$\frac{1}{L^3}\sum_p\frac{1}{L^3}\sum_r - \int\frac{d^3p}{(2\pi)^3}\int\frac{d^3r}{(2\pi)^3} = \sum\!\!\!\!\!\!\!\int_{p,n\neq 0}\sum\!\!\!\!\!\!\!\int_{r,m\neq 0} + \sum\!\!\!\!\!\!\!\int_{p,n\neq 0}\int_r + \int_p\sum\!\!\!\!\!\!\!\int_{r,m\neq 0}, \qquad (101)$$

where we also introduced the notation $\int_p \equiv \int\frac{d^3p}{(2\pi)^3}$. In some cases the momentum integrals can be more easily performed if the line of integration is deformed into the complex plane. Specifically we need the

following integrals:

$$I_1[f] \equiv \sum_{p,n\neq 0}\! f(p^2) = \frac{1}{2\pi^2}\sum_{n^2=1}^{\infty}\frac{\nu_n}{nL}\int_0^{\infty}dp\, p\sin(npL)f(p^2)$$

$$I_2(M_\alpha) \equiv -\sum_{p,n\neq 0}\! \frac{1}{E_{p,\alpha}} = -\frac{M_\alpha}{2\pi^2}\sum_{n^2=1}^{\infty}\frac{\nu_n}{nL}\int_1^{\infty}\frac{e^{-ynM_\alpha L}y\,dy}{\sqrt{y^2-1}}$$

$$I_3(p^2;M_\alpha,M_\beta) \equiv \frac{1}{6}\sum_{s=\pm}\sum_{r,m\neq 0}\! \frac{r^2}{E_{r,\beta}(E_{r,\beta}^2-E_{p,\alpha}^2+is\eta)} = \frac{M_\beta}{6\pi^2}\sum_{m^2=1}^{\infty}\frac{\nu_m}{mL}\left[\int_1^{\infty}\frac{e^{-ymM_\beta L}y^3\,dy}{\sqrt{y^2-1}(y^2+x-1)}+\right.$$

$$\left. +\frac{\pi}{2}\frac{x-1}{\sqrt{x}}\begin{cases}\exp(-m\sqrt{1-x}M_\beta L) & \text{if}\quad x<1\\ \cos(m\sqrt{x-1}M_\beta L) & \text{if}\quad x>1\end{cases}\right]\quad\text{with}\quad x=\frac{p^2+M_\alpha^2}{M_\beta^2}.$$

(102)

Here $f(p^2)$ is an arbitrary integrable function, $\nu_\xi = \sum_{\vec{n}^2=\xi}1$ and the $\eta > 0$ regulator was introduced according to our earlier discussion. In the I_3 integral, the second term in the square-bracket comes from poles at $r^2 = (p^2 + M_\alpha^2 - M_\beta^2)/M_\beta^2$. This pole term was dropped in [51] by saying that the pole can be shifted outside a complex contour. We give the result as the finite- minus infinite-volume difference and we split it into five terms, one NLO and four NNLO:

$$G(t;L) - G(t;\infty) = \Delta G(t)_{\text{NLO}} + \sum_{i=1}^{4}\Delta G(t)_{\text{NNLO},i}.$$

(103)

To keep the formulas simple we introduce the notation

$$\epsilon(p^2;M_\alpha,t) \equiv \frac{e^{-2E_{p,\alpha}t}}{E_{p,\alpha}^2}\,p^2\,.$$

(104)

Our result for the finite-volume correction is then:

$$\Delta G(t)_{\text{NLO}} = \frac{1}{3}\sum_{\alpha}{}' I_1[\epsilon(M_\alpha,t)]$$

$$F^2\Delta G(t)_{\text{NNLO},1} = \frac{1}{3}\sum_{\alpha,\beta}{}' I_1[\epsilon(M_\alpha,t)\Gamma(M_\alpha,M_\beta)]$$

$$F^2\Delta G(t)_{\text{NNLO},2} = \frac{1}{3}\sum_{\alpha,\beta}{}' I_1[\epsilon(M_\alpha,t)]\cdot I_2(M_\beta)$$

(105)

$$F^2\Delta G(t)_{\text{NNLO},3} = \frac{1}{3}\sum_{\alpha,\beta}{}'\int_p\epsilon(p^2;M_\alpha,t)\left[I_2(M_\beta)+I_3(p^2;M_\alpha,M_\beta)\right]$$

$$F^2\Delta G(t)_{\text{NNLO},4} = \frac{1}{3}\sum_{\alpha,\beta}{}' I_1[\epsilon(M_\alpha,t)I_3(M_\alpha,M_\beta)].$$

The $\frac{1}{3}$ prefactors arise from the Lorentz-index averaging and $\sum_{\alpha,\beta}' = \frac{1}{256}\sum_{\alpha,\beta}$. In the continuum these formulas agree with the ones in [51], up to the pole contribution that affects the $\Delta G(t)_{\text{NNLO},3}$ and $\Delta G(t)_{\text{NNLO},4}$ terms. The largest NNLO term by far is $\Delta G(t)_{\text{NNLO},1}$, and has the same order of magnitude as the NLO term on our lattices.

Finite-T effects

Until now we assumed that the temporal extent of the lattice is infinite, $T = \infty$. The corrections introduced by a finite T can also be computed in SXPT. The integrals over the time component of the momenta become sums, which are then related to integrals via Poisson's summation formula. Formally:

$$\int \frac{dp_4}{2\pi} \longrightarrow \frac{1}{T} \sum_{p_4} = \int \frac{dp_4}{2\pi} \sum_{n_4} e^{ip_4 n_4 T} \ . \tag{106}$$

We give here the result of this procedure for three integrals that appear in the current-current correlator, in a one-loop and in a two-loop diagram, respectively:

$$\int \frac{dp_4}{2\pi} \frac{e^{ip_4 t}}{p_4^2 + E_p^2} \longrightarrow \frac{\cosh\left[E_p\left(t - T/2\right)\right]}{2E_p \sinh(E_p T/2)}$$

$$\int \frac{dp_4}{2\pi} \frac{dq_4}{2\pi} \frac{e^{ip_4 t}}{p_4^2 + E^2} \frac{e^{iq_4 t}}{q_4^2 + E^2} \longrightarrow \left(\frac{\cosh\left[E_p\left(t - T/2\right)\right]}{2E_p \sinh(E_p T/2)}\right)^2 = \frac{\cosh\left[2E_p\left(t - T/2\right)\right] + 1}{4E_p^2 \left[\cosh(E_p T) - 1\right]} \tag{107}$$

$$\int dt' \int \frac{dp_4}{2\pi} \frac{dq_4}{2\pi} \frac{dr_4}{2\pi} \frac{ds_4}{2\pi} \frac{e^{ip_4(t-t')}}{p_4^2 + E_p^2} \frac{e^{iq_4(t-t')}}{q_4^2 + E_p^2} \frac{e^{ir_4 t'}}{r_4^2 + E_r^2} \frac{e^{is_4 t'}}{s_4^2 + E_r^2} \longrightarrow$$

$$\frac{E_r E_p^2 \cosh\left[2E_r(t - T/2)\right] \sinh(E_p T) - \{p \leftrightarrow r\} + (E_p^2 - E_r^2)\left[E_r \sinh(E_p T) + E_p \sinh(E_r T) + E_p E_r T\right]}{16 E_p^3 E_r^3 (E_p^2 - E_r^2)\left[\cosh(E_p T) - 1\right]\left[\cosh(E_r T) - 1\right]}$$

The computation proceeds as in the case of finite-L. As in Equation (104), we introduce new notations to keep the formulas simple:

$$\epsilon_T(p^2; M_\alpha, t) \equiv \frac{\cosh\left[2E_{p,\alpha}(t - T/2)\right] + 1}{E_{p,\alpha}^2\left[\cosh(E_{p,\alpha} T) - 1\right]} \, p^2 \ , \qquad \sigma_T(p^2; M_\alpha) \equiv \frac{p^2}{E_{p,\alpha}^2\left[\cosh(E_{p,\alpha} T) - 1\right]} \ . \tag{108}$$

For the correction due to finite L and T we get:

$$G(t; L, T) - G(t; \infty, \infty) = \tag{109}$$

$$\left[\frac{1}{3} \sum_\alpha{}' \int_p \epsilon(p^2; M_\alpha, t)\right]_{\epsilon \to \epsilon_T - \epsilon} + \left[\frac{1}{3} \sum_\alpha{}' \int_p \epsilon(p^2; M_\alpha, t) \frac{1}{F^2} \sum_\beta{}' \Gamma(p^2; M_\alpha, M_\beta)\right]_{\epsilon \to \epsilon_T - \sigma_T - \epsilon}$$

$$+ \left[\Delta G(t)_{\mathrm{NLO}}\right]_{\epsilon \to \epsilon_T} + \left[\sum_{i=1}^4 \Delta G(t)_{\mathrm{NNLO},i}\right]_{\epsilon \to \epsilon_T - \sigma_T}$$

$$+ \frac{T}{18F^2}\left[\sum_\alpha{}' \sum\!\!\!\!\!\!\oint_{p,n} \sigma_T(p^2; M_\alpha)\right]^2$$

$$+ \frac{1}{9F^2}\left[\sum_\alpha{}' \sum\!\!\!\!\!\!\oint_{p,n} \sigma_T(p^2; M_\alpha)\right]\left[\sum_\alpha{}' \sum\!\!\!\!\!\!\oint_{p,n} \sigma_T(p^2; M_\alpha) \frac{1 - e^{-E_{p,\alpha} T}}{E_{p,\alpha}}\right]$$

$$- \frac{2}{3F^2}\left[\sum_\alpha{}' \sum\!\!\!\!\!\!\oint_{p,n} \epsilon_T(p^2; M_\alpha, t)\right]\left[\sum_\alpha{}' \sum\!\!\!\!\!\!\oint_{p,n} \frac{1}{E_{p,\alpha}\left(e^{E_{p,\alpha} T} - 1\right)}\right]$$

$$+ \frac{1}{18F^2} \sum_\alpha{}' \sum\!\!\!\!\!\!\oint_{p,n} \left[\epsilon_T(p^2; M_\alpha, t) - \sigma_T(p^2, M_\alpha)\right] \sum_\beta{}' \sum_{s=\pm} \sum\!\!\!\!\!\!\oint_{r,m} \frac{r^2}{E_{r,\beta}(E_{r,\beta}^2 - E_{p,\alpha}^2 + is\eta)} \frac{1 - e^{-E_{r,\beta} T}}{\cosh(E_{r,\beta} T) - 1}$$

The first line is the correction due to finite T at infinite L, ie. $G(t; \infty, T) - G(t; \infty, \infty)$. The rest of the formula gives the correction due to finite L at finite T, ie. $G(t; L, T) - G(t; \infty, T)$. The terms in the second line are obtained by replacing ϵ by ϵ_T or by $\epsilon_T - \sigma_T$ in Equations (105) from the previous subsection. The first and second lines give most of the contribution to the finite-size effect for the cases of interest in this work. The rest, from the third to the sixth line, are genuine finite-T corrections in that they vanish at infinite T. They are all NNLO and are negligible for the lattices considered here.

Taste non-singlet contributions to the current

Here we explore briefly to what extent our previous assumption, that the vector current is a taste singlet, is justified. For this purpose it is useful to work with the valence staggered fermion fields χ. The lattice current propagator, that we introduced in Section 10, can be derived from the following zero-spatial-momentum operator:

$$\sum_{\vec{x}} \left(\overline{\chi}_x U_{\mu,x} \eta_{\mu,x} \chi_{x+\mu} + \overline{\chi}_{x+\mu} U^{\dagger}_{\mu,x} \eta_{\mu,x} \chi_x \right) \ , \tag{110}$$

where the $\eta_{\mu,x}$ phase is the staggered representation of the Dirac matrix and we drop the flavor index for simplicity. Equation (110) is just the conserved vector current of the staggered fermion action.

Staggered bilinears can be assigned with a spin$\otimes$taste structure, see [67] for a modern treatment. In the case of the operator in (110), two such assignments can be given:

$$\gamma_\mu \otimes 1 \qquad \text{and} \qquad \gamma_\mu \gamma_4 \gamma_5 \otimes \xi_4 \xi_5 \ . \tag{111}$$

The conserved current couples to states of both types. The first corresponds to the taste-singlet vector current, the case which we have fully covered previously. The second is a taste non-singlet pseudovector[2]. Its correlator gives the characteristic contribution to the staggered propagator which oscillates in time with a factor $(-1)^{x_4}$. The observables that we consider in this paper are obtained by integrating the propagator over the whole time range, or at least over some physical distance. In the continuum limit such an integration completely eliminates the oscillating contribution. At finite lattice spacing it gives an $O(a^2)$ suppression compared to the taste singlet vector contribution.

It is possible to compute the pseudovector propagator in chiral perturbation theory. This requires the inclusion of an external antisymmetric tensor source $t_{\mu\nu}$. This has been done to NNLO in [68], which has chosen $t_{\mu\nu} \approx O(p^2)$ for the chiral counting of the tensor field. With this choice a vertex with the tensor field first appears in the NLO Lagrangian. The one-loop contribution to the tensor-tensor propagator is zero, due to charge conjugation and parity invariance. The two-loop contribution has two NLO vertices and is thus $O(p^{10})$. This, combined with the suppression explained in the previous paragraph, gives $O(a^2 p^{10})$, which is well beyond the order to which we work.

As explained in [68], there is an ambiguity in the chiral-counting assignment of $t_{\mu\nu}$. A closely related fact is that the tensor-tensor propagator is renormalization scheme dependent, since there is no conservation law for the tensor current. Even if we used a counting $t_{\mu\nu} \approx O(1)$, the oscillating contribution would still be beyond the NNLO to which we work.

Based on these arguments, we ignore the taste non-singlet contribution to the current in our SXPT computation.

16 Lellouch-Lüscher-Gounaris-Sakurai model

In this section we describe a phenomenological model that we use to make predictions for finite-volume corrections in Section 17. We also use it to correct for taste-breaking effects in the $I = 1$ contribution to

---

[2]Pseudovectors are to be distinguished from axial-vectors, the latter have $\gamma_\mu \gamma_5$ spin structure.

the Euclidean, current-current correlation function, in Section 18. We work in the isospin-symmetric limit throughout this section.

Model for the pion form factor

As shown in [42], in infinite volume the Euclidean, current-current correlation function is a Laplace transform of the corresponding spectral function, $\rho(E)$:

$$G(t) = \int_0^\infty dE E^2 \, e^{-E|t|} \rho(E) \,, \tag{112}$$

for $t \neq 0$. Here, $G(t)$ is defined in Equation (1) of the main paper and E is the center-of-mass energy.

For large $|t|$, $G(t)$ is dominated by the low end of the spectrum, which is governed by two-pion, scattering states. In addition, phenomenology indicates that two-pion states, up to $E = 1.8$ GeV [69, 70], are responsible for over 70% of the total a_μ and over 85% of the $I = 1$ contribution computed in this work. Thus, the contribution of two-pion states should not only provide a good description of the long-distance behavior of the current-current correlator, important for understanding finite-volume effects, but also a reasonable model for this correlator at all distances relevant for the determination of a_μ. Now, the two-pion contribution to the spectral function is given by (see eg. [71])

$$\rho(E)|_{\pi\pi} = \frac{1}{6\pi^2} \left(\frac{k}{E}\right)^3 |F_\pi(k)|^2 \,, \tag{113}$$

where $E = 2\sqrt{M_\pi^2 + k^2}$, with k the magnitude of the pions' back-to-back momenta in the center-of-mass frame, and $F_\pi(k)$ is the timelike, pion, electromagnetic form factor.

A good phenomenological description of the pion form factor and the corresponding π-π scattering phase shift is given by the Gounaris-Sakurai (GS) parametrization [72]. In the context of estimating finite-volume effects in $g_\mu - 2$, it was used first in [73]. The GS parametrization describes well the experimental spectral function in the $I = 1$ channel from threshold to E around 1 GeV, thus covering the very important ρ-resonance contribution. This parametrization is given by:

$$F_\pi(k) = \frac{M^2(0)}{M^2(E) - E^2 - iM_\rho\Gamma(E)} \,, \tag{114}$$

with the energy-dependent width

$$\Gamma(E) = \Gamma_\rho \left(\frac{k}{k_\rho}\right)^3 \left(\frac{M_\rho}{E}\right) \,, \tag{115}$$

where $k_\rho = \sqrt{\frac{M_\rho^2}{4} - M_\pi^2}$, and the energy-dependent mass squared:

$$M^2(E) = M_\rho^2 + \frac{\Gamma_\rho M_\rho^2}{k_\rho^3} \left[k^2 \left[h(E) - h(M_\rho)\right] - \left[E^2 - M_\rho^2\right] \frac{k_\rho^2}{2M_\rho} h'(M_\rho)\right] \,. \tag{116}$$

Here, the pion-loop function is

$$h(E) = \frac{2k}{\pi E} \log \frac{E + 2k}{2M_\pi} \tag{117}$$

and $h'(E)$ is its derivative with respect to E. To complete the model, we determine the width of the ρ in terms of the ρ-$\pi\pi$ coupling, g, to leading order in an effective theory where the ρ and π are pointlike

42

particles. It is straightforward to show that

$$\Gamma_\rho = \frac{g^2}{6\pi} \frac{k_\rho^3}{M_\rho^2} \; . \tag{118}$$

From these expressions for $F_\pi(k)$, the phase-shift is simply obtained using Watson's theorem:

$$\delta_{11}(k) = \arg F_\pi(k) \; . \tag{119}$$

$G^{\mathrm{GS}}(t)$ will denote the infinite-volume correlator given by Equations (112), (113), (114) and (118).

The GS model has two free parameters: g and M_ρ. Since our simulations are performed very near the physical mass point, we fix these parameters to their physical value, neglecting their sub-percent uncertainties: $M_\rho = 775$ MeV is the mass of the ρ_0 meson from [74], and $g = 5.95$ is obtained from Equation (118), using the width of ρ_0 and the mass of $\pi_\pm$, also from [74].

Model for finite-volume effects

In a finite spatial volume of size $L \times L \times L$, the two-pion spectrum is discrete because of momentum quantization, and the spectral representation of the current-current correlator becomes a sum, instead of an integral, over two-pion states. Thus, the large-t behavior of the corresponding, finite-volume correlation function, $G(t; L)$, can be written as:

$$G(t; L) \overset{|t| \to \infty}{\longrightarrow} \frac{1}{3} \sum_{n>0} |\vec{A}_n|^2 e^{-E_n |t|} \; , \tag{120}$$

where n labels the energy eigenstates, in order of increasing energy. Below the four-pion, inelastic threshold, the energy of state n is given by $E_n = 2\sqrt{M_\pi^2 + k_n^2}$, with k_n determined by the infinite-volume, $I^G(J^{PC}) = 1^+(1^{--})$, π-π scattering phase shift, $\delta_{11}(k)$, through Lüscher's formula [75, 76]:

$$\phi(q_n) + \delta_{11}(k_n) = n\pi, \qquad n = 1, 2, \dots \; , \tag{121}$$

where $q = kL/2\pi$ and $\phi(q)$ is given in [75] (see also [77]). The amplitude of the n-th state $\vec{A}_n$ is proportional to $\langle 0|\vec{J}_{I=1}(0)|n\rangle$, where $\vec{J}_{I=1}$ collects the spatial components of the isospin $I = 1$ contribution to the quark electromagnetic current defined after Equation (1) of the main paper. This amplitude is determined by the phase-shift and by the timelike, pion, electromagnetic form factor, through a Lellouch-Lüscher (LL) equation [78–80]

$$|\vec{A}_n|^2 = [q\phi'(q) + k\delta'_{11}(k)]^{-1}_{k=k_n} \frac{2k_n^5}{\pi E_n^2} |F_\pi(k_n)|^2 \; , \tag{122}$$

where the primes indicate a derivative of the function with respect to its argument. We assume that Equations (121) and (122) are also approximately true above the inelastic threshold, because the ρ decays almost exclusively into two pions [74]. Equations (119) and (120) then define the Lellouch-Lüscher-Gounaris-Sakurai (LLGS) model for the finite-volume current correlator. It will be denoted $G^{\mathrm{LLGS}}(t; L)$.

It is important to note that the GS parametrization for $\rho(E)|_{\pi\pi}$, obtained from Equations (114) and (118), decreases as $(E \log E)^{-2}$ for large E and becomes smaller than that of free pions after the ρ peak, when $E \gtrsim 1.1$ GeV. Thus, in the sums over two-pion states in finite-volume, one can reasonably neglect terms for which E_n is greater than 1.2 GeV. In the reference volume with $L = L_{\mathrm{ref}} = 6.272$ fm, used in Section 17, this corresponds to $n = 8$ for the Goldstone.

In Section 17, we use this LLGS model to compute the finite-volume correction to the $I = 1$ contribution to a_μ. It is determined in the continuum limit for the reference volume L_{ref} and is given by

integrating the difference of the infinite and finite-volume correlators:

$$a_\mu^{\mathrm{GS}}(L = \infty) - a_\mu^{\mathrm{LLGS}}(L_{\mathrm{ref}}) = 10^{10}\alpha^2 \int_0^\infty dt\, K(t) \left[G^{\mathrm{GS}}(t) - G^{\mathrm{LLGS}}(t; L_{\mathrm{ref}}) \right] \ . \tag{123}$$

In Section 17, this difference is compared to the results of a dedicated lattice study of finite-volume effects in a_μ. The good agreement represents a strong validation of the model.

Model for taste violations

Here we generalize the LLGS model, for the finite-volume correlation function, to include the lattice spacing effects arising from taste breaking. Indeed, the dominant, taste-breaking effects in a_μ are expected to be those associated with the two-pion spectrum: these states give the dominant contribution to a_μ and the masses of pions are significantly affected by taste breaking on coarser lattices. Our conserved, quark electromagnetic current couples, not only to two-Goldstone-pion states, but also to fifteen additional pairs of more massive taste partners of the pion. We label the masses of these states with M_τ, where the τ index runs over the 16 element set $\tau \in \{5, \mu5, \mu\nu, \mu, I\}$ with $\mu < \nu$.

In a description where the pions are free, corresponding to NLO staggered XPT, the two-pion states have energies, $E_{n,\tau}^{(0)} = 2\sqrt{M_\tau^2 + k_n^2}$, with $k_n = |\vec{n}|(2\pi/L)$, $n = \vec{n}^2$ and $\vec{n} \in \mathbb{Z}^3$. In the interacting case, we make the assumption that two-pion-state energies have the same set of taste copies, but with the momentum, $k_{n,\tau}$, given by the Lüscher quantization condition of Equation (121). We further assume that a similar conclusion holds for the amplitudes, i.e. that they satisfy Equation (122) with $k_{n,\tau}$ given by Equation (121). Thus we model the long-distance behavior of the correlator in a finite spatial volume and at finite lattice spacing as:

$$G(t; L, a) \xrightarrow{|t|\to\infty} \frac{1}{3}\sum_{n>0} \frac{1}{16}\sum_\tau |\vec{A}_{n,\tau}|^2 e^{-E_{n,\tau}|t|} \ , \tag{124}$$

which we denote $G^{\mathrm{SLLGS}}(t; L, a)$. In the formula, the lattice spacing dependence arises from the splittings in the pion spectrum. In implementing this model, we assume that all taste-partner pairs of pions couple to the physical, lightest, ρ taste, with the same coupling g. Thus, we implicitly assume that the dependence of Γ_ρ and of $\delta_{11}(k)$ on M_τ is mostly kinematic, an assumption which is borne out by simulations [81]. All of these choices guarantee that our model has the correct continuum limit. Also, we keep states up to $n = 8$, even for the more massive taste partners. However, we only apply the model to simulations in which the physical ρ has suffcent phase space to decay into all sixteen, taste-partner pion pairs. This excludes only our coarset simulation, with $\beta = 3.7000$.

We use the SLLGS model to correct taste-breaking effects in the $I = 1$ contribution to a_μ, simulation by simulation. This significantly reduces the a-dependence of the dominant contribution to a_μ, allowing for a more precise determination of its continuum limit. As shown in Section 18, the optimal way in which to apply taste-breaking corrections is to consider these corrections in different time-windows. Thus, for each simulation we compute,

$$a_{\mu,\mathrm{win}}^{\mathrm{LLGS}}(L_{\mathrm{ref}}) - a_{\mu,\mathrm{win}}^{\mathrm{SLLGS}}(L, a) = 10^{10}\alpha^2 \int_0^\infty dt\, K(t)\, W(t; t_1, t_2) \left[G^{\mathrm{LLGS}}(t; L_{\mathrm{ref}}) - G^{\mathrm{SLLGS}}(t; L, a) \right] \ , \tag{125}$$

where the window function, $W(t; t_1, t_2)$, is defined in Equation (71) and L is the size of the lattice for that simulation. The lattice spacing dependence enters in SLLGS from the taste breaking in the pion masses M_τ and these are also taken from the simulation. This correction is applied as an additive shift on the measured a_μ.

Inclusion of finite-T effects

In this section, we generalize the LLGS and SLLGS models to include the effects of the finite-time extent, T, of the lattice. In our simulations, gauge-boson fields obey periodic boundary conditions in time and quark fields, antiperiodic ones. Finite-T effects are expected to be smaller than those due to finite L, because $T \geq L$ in our simulations and because they occur only in a single spacetime direction.

The LLGS model describes the contributions to the current-current correlator of two-pion states, in the presence of interactions. Thus, in studying finite-T effects, we consider terms in which each current couples only to two pions[3]. The three largest contributions are:

1. The Euclidean propagation of interacting $\pi^+\pi^-$ states between times 0 and $t > 0$, with energies E_n. These contributions fall-off exponentially as $e^{-E_n t}$.

2. The Euclidean propagation of interacting $\pi^-\pi^+$ states from t to T, which increases as $e^{-E_n(T-t)}$ as t approaches $T/2$.

3. The contribution of a single pion that wraps around the time direction. This is at the same order in decreasing exponentials of T as the previous contribution. At any given time, there is only a single pion on the lattice. Therefore, the energy of this state is that of a free pion with at least one unit of momentum, i.e. $E_n^{(0)}/2 = \sqrt{n(2\pi/L)^2 + M_\pi^2}$, with $n \geq 1$ the norm-squared of a three vector in $\mathbb{Z}^3$. This contribution is constant in t and proportional to $e^{-E_n^{(0)}T/2}$. Moreover, since the pion couples to the current without recoiling, the matrix element describing this coupling is also that of the free pion theory. This contribution appears with a factor of 2, because it can be caused by the propagation of either a π^+ or a π^-. Altogether, this contribution has the same form as it does at NLO in XPT.

The next order in decreasing exponentials of T is a term of three pions propagating from 0 to t and a single pion from t to T. We have estimated this contribution and find that it contributes to a_μ at a level that is orders of magnitude smaller than our statistical error. Thus, we neglect it, as well as all higher-order winding terms and, to describe finite-T effects, we replace $G^{\mathrm{LLGS}}(t; L)$ by

$$G^{\mathrm{LLGS}}(t; L, T) \equiv G^{\mathrm{LLGS}}(t; L) + \frac{1}{3}\sum_{n>0}|\vec{A}_n|^2\, e^{-E_n(T-t)} + \frac{2}{3}\sum_{n>0}|\vec{A}_n^{(0)}|^2\, e^{-E_n^{(0)}T/2}, \qquad (126)$$

with the free pion amplitude squared, $L^3|\vec{A}_n^{(0)}|^2 = 4\nu_n n(2\pi/LE_n^{(0)})^2$, and keeping only states up to $n = 8$, as above. Here, ν_n counts the number of vectors of $\mathbb{Z}^3$ that have norm squared n.

Using similar arguments, the inclusion of taste-violations is straightforward. One merely performs the replacement $E_n \to E_{n,\tau}$ and averages over the sixteen tastes. This yields $G^{\mathrm{SLLGS}}(t; L, T, a)$.

Now, using the above correlators, it is straightforward to generalize Equation (123) to also include finite-T corrections. We obtain:

$$a_\mu^{\mathrm{GS}}(L = \infty, T = \infty) - a_\mu^{\mathrm{LLGS}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$$

$$= 10^{10}\alpha^2\left[\int_0^\infty dt\, K(t)G^{\mathrm{GS}}(t) - \int_0^{T_{\mathrm{ref}}/2} dt\, K(t)G^{\mathrm{LLGS}}(t; L_{\mathrm{ref}}, T_{\mathrm{ref}})\right]. \qquad (127)$$

Similarly, for each simulation, we can estimate taste-breaking effects on an $T \times L^3$ lattice, through:

$$a_{\mu,\mathrm{win}}^{\mathrm{LLGS}}(L_{\mathrm{ref}}, T_{\mathrm{ref}}) - a_{\mu,\mathrm{win}}^{\mathrm{SLLGS}}(L, T, a) = 10^{10}\alpha^2\left[\int_0^{T_{\mathrm{ref}}/2} dt\, K(t)\, W(t; t_1, t_2)G^{\mathrm{LLGS}}(t; L_{\mathrm{ref}}, T_{\mathrm{ref}})\right.$$

---

[3]There are also terms in which, four, six, eight, etc. pions couple to the currents. Compared to the two pion terms, these are exponentially suppressed in time but also suppressed by the small coupling of four or more pions to the ρ, which dominates the energy range that we are modelling here.
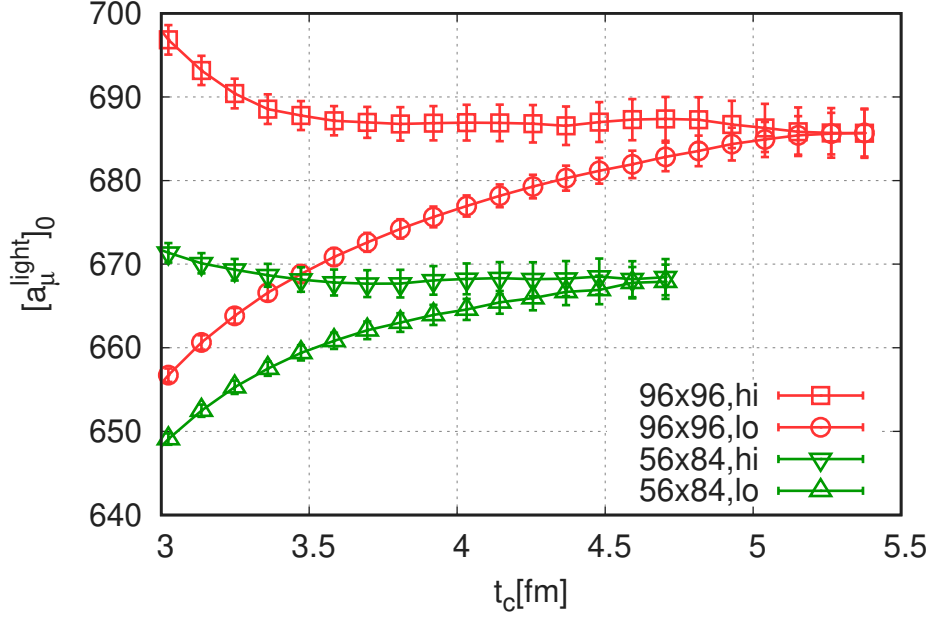
Figure 13: Upper and lower bounds on the light isospin-symmetric component of a_μ. The results shown here are obtained with the 4HEX action on two different volumes at $a = 0.112$ fm lattice spacing and $M_\pi = 121$ MeV Goldstone-pion mass. We also have another simulation with $M_\pi = 104$ MeV mass. From these two we interpolate to $M_\pi = 110$ MeV. This value ensures that a particular average of pion tastes is fixed to the physical value of the pion mass (see text).

$$-\int_0^{T/2} dt\, K(t)\, W(t; t_1, t_2) G^{\mathrm{SLLGS}}(t; L, T, a) \Bigg] \,, \tag{128}$$

thus generalizing Equation (125) to finite T.

17 Finite-size effects in a_μ

Finite-size effects on a_μ were the largest source of uncertainty in our previous work [47]. In this section we present the computation of these effects in a systematic way, which includes dedicated lattice simulations, chiral perturbation theory and phenomenological models. The concrete goal of this section is to provide a single number that is to be added to the continuum-extrapolated lattice result obtained in a reference box, which is defined by a spatial extent of $L_{\mathrm{ref}} = 6.272$ fm and a temporal extent of $T_{\mathrm{ref}} = \frac{3}{2} L_{\mathrm{ref}}$.

First we concentrate on the finite-size effect of the isospin-symmetric part. Section 6 details our isospin decomposition. The isospin-breaking part will be discussed later in the last subsection. The isospin-symmetric part can be further decomposed into an $I = 0$ and an $I = 1$ channel. From these the $I = 1$ is supposed to give the majority of the finite-size effect. We focus on the $I = 1$ first, and give an estimate of the $I = 0$ contribution later. According to Equation (63) the $I = 1$ result is given by the $\left(\frac{9}{10}\right)$'th of the connected light contribution.

We perform dedicated lattice simulations with two different lattice geometries: one on a 56×84 lattice with the reference box size and another on a large 96×96 lattice with box size $L = L_{\mathrm{big}} = 10.752$ fm and $T = T_{\mathrm{big}} = L_{\mathrm{big}}$. Since taste violations severely distort the finite-size behavior, we designed a new action with highly-suppressed taste breaking for these computations. The details of the 4HEX action and the simulation parameters are given in Section 2. Our strategy is then to compute the finite-size correction

as the following sum:

$$a_\mu(\infty, \infty) - a_\mu(L_{\mathrm{ref}}, T_{\mathrm{ref}}) =$$
$$= [a_\mu(L_{\mathrm{big}}, T_{\mathrm{big}}) - a_\mu(L_{\mathrm{ref}}, T_{\mathrm{ref}})]_{\texttt{4HEX}} + [a_\mu(\infty, \infty) - a_\mu(L_{\mathrm{big}}, T_{\mathrm{big}})]_{\mathrm{XPT}}. \quad (129)$$

The first difference on the right hand side is taken from the dedicated 4HEX simulations. The second difference is expected to be much smaller than the first and is taken from a non-lattice approach: chiral perturbation theory.

We consider three non-lattice approaches for both differences on the right hand side of Equation (129). In case of the first difference, they will be compared to our 4HEX simulations. The first is chiral perturbation theory (XPT), discussed in detail in Section 15. The second is the Lellouch-Luscher-Gounaris-Sakurai model (LLGS), with details in Section 16. In this approach we compute values for the reference box only, and not for the large box. This is because L_{big} is relatively large and one would have to deal with a large number of states, which is not practical in that approach. The third approach is that of Hansen and Patella (HP) [82], who use a generic field theory framework to relate the finite-size effect to the electromagnetic form factor of the pion, the latter being determined on the lattice. Note that their first published result does not include effects that are of order $e^{-\sqrt{2}M_\pi L}$. These can be significant and have been added later [83]. Though the latter version also includes finite-T effects, it does so with assumptions that are not applicable to our lattices, where $T < 2L$. Therefore we use the HP approach here in the infinite-T limit.

Results with the 4HEX action

We compute the first difference in Equation (129) using dedicated simulations with the 4HEX action. First we describe the way in which we fix the physical point in these simulations. For this purpose, it is instructive to look at the influence of taste violations on the finite-size effect in NNLO staggered chiral perturbation theory (SXPT). The necessary formulas are given in Section 15. We apply them to various cases that are described below. The following numbers are obtained for the finite-size effect:

NNLO SXPT results for →	continuum	4stout	4HEX	4HEX@110MeV
$a_\mu(L_{\mathrm{big}}, T_{\mathrm{big}}) - a_\mu(L_{\mathrm{ref}}, T_{\mathrm{ref}})$	15.7	1.6	8.1	15.8

The first number gives the continuum prediction, which is about 2% of the total a_μ. The second number stands for the 4stout action at a lattice spacing of $a = 0.112$ fm. Here, most of the pion tastes are too heavy to play any role in the finite-size behavior. According to SXPT the finite-size effect is practically non-existent there. The 4HEX action has much suppressed taste violations, and the corresponding number, the third in the table, is already much closer to the continuum. Until now the Goldstone pion mass is set to the physical value $M_\pi = M_{\pi_0,*}$. This pion is the lightest of the sixteen pions in the taste multiplet. We can get much closer to the size of the continuum finite-size effect if we use Goldstone-pion masses below $M_{\pi_0,*}$. For example one can set a taste-averaged pion mass to the physical value.

In NLO SXPT the slope of the hadronic vacuum polarization is proportional to $\sum_\alpha M_{\pi,\alpha}^{-2}$. This motivates to use the harmonic-mean-square (HMS), defined by

$$M_{\pi,\mathrm{HMS}}^{-2} \equiv \frac{1}{16} \sum_\alpha M_{\pi,\alpha}^{-2} \ ,$$

to average over the tastes. Setting $M_{\pi,\mathrm{HMS}} = M_{\pi_0,*}$ requires lowering the Goldstone-pion mass to $M_\pi = 110$ MeV. With this choice, the finite-size effect is of the same size on the lattice and in the continuum in NNLO SXPT. This gives the fourth number in the table. This choice results in much smaller lattice artefacts than the usual setting with the Goldstone-pion, at least for an observable like the finite-size effect.

To generate the 4HEX data set, we performed simulations with two different Goldstone pion masses: $M_\pi = 104$ MeV and 121 MeV. To set the physical point as described above, we perform an interpolation

from these two pion masses to $M_\pi = 110$ MeV.

To compute a_μ^{light} from the current propagator in our 4HEX simulations we use the upper and lower bounds described in Section 13. The results are plotted in Figure 13 for the $M_\pi = 121$ MeV simulation point. The bounds meet at around 4.2 fm and 4.7 fm on the small and large volumes, respectively. At these distances we take the average of the two bounds as an estimate for a_μ^{light}. The results are given in the table below:

M_π in 4HEX $\rightarrow$	104 MeV	121 MeV	110 MeV
$a_\mu^{\text{light}}(56 \times 84)$	685.9(2.7)	668.3(2.0)	679.5(1.9)
$a_\mu^{\text{light}}(96 \times 96)$	710.7(1.9)	684.3(1.7)	701.1(1.3)

In the last column we also give the interpolated value at the physical point, using the HMS averaged pion-mass prescription defined above.

We only have one lattice spacing with the 4HEX action, so no proper continuum extrapolation of the finite-size effect can be done. We estimate the cutoff effect of the result by comparing the total a_μ with the 4HEX action at this single lattice spacing to the continuum extrapolated 4stout lattice result, both in the L_{ref} volume. The 4HEX result is about 7% larger than the continuum value. Therefore we reduce the measured finite-size effect by 7%, and assign a 7% uncertainty to this correction step. For the difference we get

$$a_\mu(L_{\text{big}}, T_{\text{big}}) - a_\mu(L_{\text{ref}}, T_{\text{ref}}) = 18.1(2.0)_{\text{stat}}(1.4)_{\text{cont}} . \tag{130}$$

The result is obtained from the a_μ^{light} numbers from above including a multiplication by the $\left(\frac{9}{10}\right)$ charge factor. The first error is statistical, the second is an estimate of the cutoff effect.

Results from non-lattice approaches

The table below collects the finite-size effect computed in various non-lattice approaches:

	NLO XPT	NNLO XPT	LLGS	HP
$a_\mu(L_{\text{big}}, T_{\text{big}}) - a_\mu(L_{\text{ref}}, T_{\text{ref}})$	11.6	15.7	17.8	—
$a_\mu(L_{\text{big}}, \infty) - a_\mu(L_{\text{ref}}, \infty)$	11.2	15.3	17.4	16.3

As we mentioned before, the LLGS approach was not used in the large box. The LLGS numbers in the table are actually a difference of the LLGS prediction for $a_\mu(\infty, \infty) - a_\mu(L_{\text{ref}}, T_{\text{ref}})$ and the residual finite-size effects of the big lattice taken from NNLO XPT. We also give results for the case of infinite time extent. We see that, according to the models, the finite-T effect is much smaller than the finite-L effect.

The different models give a finite-size effect of similar size that agrees well with the lattice determination of Equation (130). Only the NLO result differs by about 3σ's. The fact that NLO chiral perturbation theory underestimates the finite-size effect was already shown in [84], at a non-physical pion mass. Using physical pion mass, a dedicated finite-volume study was carried out in [85]. It reaches the same conclusion as we do, albeit with larger errors.

The good agreement for the finite-size effect of the reference box, between the models and the lattice, gives us confidence that the models can be used to reliably compute the very small, residual, finite-size effect of the large box. We get:

	NLO XPT	NNLO XPT	HP
$a_\mu(\infty, \infty) - a_\mu(L_{\text{big}}, T_{\text{big}})$	0.3	0.6	—
$a_\mu(\infty, \infty) - a_\mu(L_{\text{big}}, \infty)$	1.2	1.4	1.4

For an infinite-time extent the NNLO XPT and HP approaches agree nicely. As a final value for the large box finite-size effect we take the NNLO XPT result including the finite-T effects:

$$a_\mu(\infty, \infty) - a_\mu(L_{\text{big}}, T_{\text{big}}) = 0.6(0.3)_{\text{big}} , \tag{131}$$
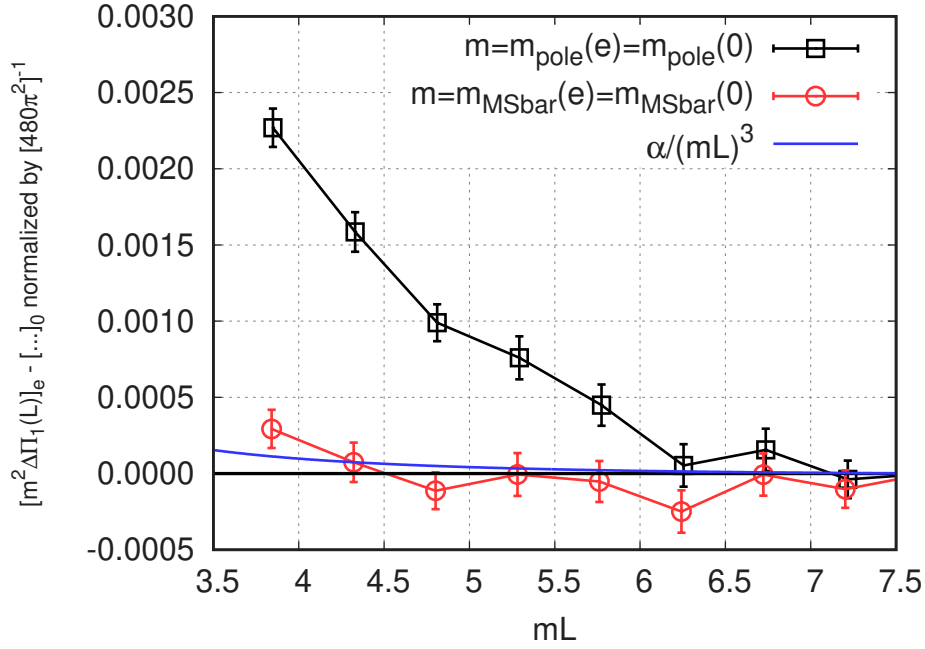
Figure 14: Electromagnetic part of the finite-size effect for scalar QED. Shown is the dimensionless double difference for the slope of the vacuum polarization, ie. $m^2 \left[\Delta\Pi_1(L,e) - \Delta\Pi_1(L,0) \right]$, normalized by the free field value $m^2\Pi_1(\infty,0) = [480\pi^2]^{-1}$ (see text). Two different definitions for the electromagnetic part are shown, defined by matching with two different masses, m_{pole} and $m_{\overline{\text{MS}}}$. A blue curve shows an order of magnitude estimate of the finite-size effect from [86].

where the uncertainty is an estimate of higher-order effects, given here by the difference of the NNLO and NLO values.

Until now we have been discussing the $I = 1$ finite-size effects. Here we make an estimation of the $I = 0$ channel using XPT. The long-distance behavior of the $I = 0$ channel is dominated by three pions. A neccessary photon-pion-pion-pion vertex arises from the Wess-Zumino-Witten term in the chiral Lagrangian, which is NLO [66]. The lowest order diagram involves two such vertices and two loops, thus the contribution is N^{4}LO. From the NLO and NNLO values for $a_\mu(\infty, \infty) - a_\mu(L_{\text{ref}}, T_{\text{ref}})$ we estimate the size of the N^{4}LO term, from which we take $0.0(0.6)_{I=0}$ as our estimate for the $I = 0$ finite-size effect.

Finite-size effect in the isospin-breaking contributions

A comprehensive study of the electromagnetic finite-size effects on the current propagator has appeared recently [86]. The authors conclude that if all particles, except the photon, are treated in infinite volume, then the finite-size effects are of order $\alpha/(M_\pi L)^3$. In practice, however, when all particles reside in the finite box, the usual exponential finite-size effects become dominant over their electromagnetic counterpart suppressed by α. In this case it is useful to separate the electromagnetic contributions from the isospin-symmetric part. The QED part exhibits an $\alpha/(M_\pi L)^3$ behavior. The isospin-symmetric part will have an exponential suppression governed by the neutral-pion mass, $\exp(-M_{\pi_0} L)$. These isospin-symmetric effects are sizeable and discussed earlier in this Section. A subtle point here is the definition of the electromagnetic contribution or equivalently the matching of QCD+QED to QCD.

It is instructive to study, in a simple model, the role of matching in the size of finite-volume effects. For this purpose we carry out lattice simulations in scalar QED. Only quenched QED is implemented, since dynamical QED effects enter at order $O(e^4)$. We perform two sets of simulations:

1. First we perform simulations in QED with a bare scalar mass $m_0 = 0.1210$, a coupling $\alpha = 1/137$ and in L^4 boxes in the range $L = 16 \ldots 32$. The mass of the charged scalar boson extracted from the propagator and extrapolated to infinite volume is $m_{\text{pole}}(e) = 0.2406$. One can also define a

renourmalized mass using the $\overline{\text{MS}}$ prescription [86], and we find $m_{\overline{\text{MS}}}(e) = 0.2401$.

2. We also perform simulations without QED, which is just the free scalar field theory, in the same box sizes and at two bare values of the mass, $m_0 = 0.2405$ and 0.2415. We use these two values to perform the interpolations that are necessary for the different matching conditions. Note that even in the free case, the pole mass is slightly different from the bare mass due to lattice artefacts [87]. We use $m_{\overline{\text{MS}}}(0) = m_{\text{pole}}(0)$.

We choose a simple observable, the slope of the vacuum polarization function $\Pi_1(L, e) \equiv \frac{d\Pi(Q^2)}{dQ^2}\Big|_{Q^2=0}$. This is built from the conserved current of the scalar field theory, which is given as

$$j_{\mu,x}/i = \phi^\dagger_{x+\mu} e^{ieA_{\mu,x}} \phi_x - \phi^\dagger_x e^{-ieA_{\mu,x}} \phi_{x+\mu} . \tag{132}$$

From the measured Π_1, we build the difference $\Delta\Pi_1(L, e) \equiv \Pi_1(32, e) - \Pi_1(L, e)$ and investigate its L dependence. To define the QED part, we compute the difference between $\Delta\Pi_1(L, e)$ and $\Delta\Pi_1(L, 0)$. This can be done in different ways, depending on how the two theories, with and without QED, are matched. One way is to use m_{pole} to match the theories, another is to do the same with the $m_{\overline{\text{MS}}}$. The results are shown in Figure 14. As one can see, the $\overline{\text{MS}}$ matching leads to the expected $\alpha/(mL)^3$ behavior, but pole-mass matching leaves the electromagnetic part with a much larger finite-size effect.

In QCD+QED we can define matching schemes similar to the pole and $\overline{\text{MS}}$ mass matchings of scalar QED. The pole mass is the measured mass of the charged particle, whereas the $\overline{\text{MS}}$ mass is the parameter that appears in the renormalized Lagrangian. In QCD+QED the analogue of m_{pole} is the charged-pion mass, and that of $m_{\overline{\text{MS}}}$ is the renormalized quark mass. If we were to base our matching on M_{π_+}, the finite-size effects would be much larger than the expected $\alpha/(M_\pi L)^3$. Whereas if we use the renormalized quark masses in the matching, we expect to see the $\alpha/(M_\pi L)^3$ finite-volume behavior in the electromagnetic part. In our scheme, as introduced in Section 6, we keep the neutral-pion mass, M_{π_0}, fixed instead of the charged-pion mass M_{π_+}. This should be very close to a scheme where the renormalized quark masses are kept fixed. As such, we expect the QED corrections to a_μ to exhibit an $\alpha/(M_\pi L)^3$ behavior in our scheme too.

The finite-size effect of $O[\alpha/(M_\pi L)^3]$ from the electromagnetic part is very small compared to the precision of our study. The finite-size effect of the strong-isospin-breaking part must also be small: it is actually exactly zero in NLO chiral perturbation theory. In our reference box $L_{\text{ref}} = 6.272$ an $\alpha/(M_\pi L_{\text{ref}})^3$ relative correction corresponds to a finite-size effect of 0.1 in a_μ. We will use a value of $0.0(0.1)_{\text{qed}}$ as an estimate of the finite-size effect of the isospin-breaking in our reference box.

Final result

For our final result for the finite-size effect of the reference box, we add the numbers in Equations (130) and (131), as well as the estimates from the conclusions of the preceeding two subsections, giving:

$$a_\mu(\infty, \infty) - a_\mu(L_{\text{ref}}, T_{\text{ref}}) = 18.7(2.0)_{\text{stat}}(1.4)_{\text{cont}}(0.3)_{\text{big}}(0.6)_{I=0}(0.1)_{\text{qed}}[2.5] . \tag{133}$$

the first error is the statistical uncertainty of our 4HEX computation, the second is an estimate of the 4HEX cutoff effects, the third is the uncertainty of the residual finite-size effect of the "big" lattice, the fourth is a XPT estimate of the $I = 0$ finite size effect, the fifth is an estimate of the isospin-breaking effects. The last, total error in the square-brackets is the first five added in quadrature. The vast majority of the finite-size effect is obtained using the 4HEX lattice computation; for the rest we apply analytic methods. These methods have been validated by the lattice computation: for the majority contribution they give values that are consitent with the lattice.
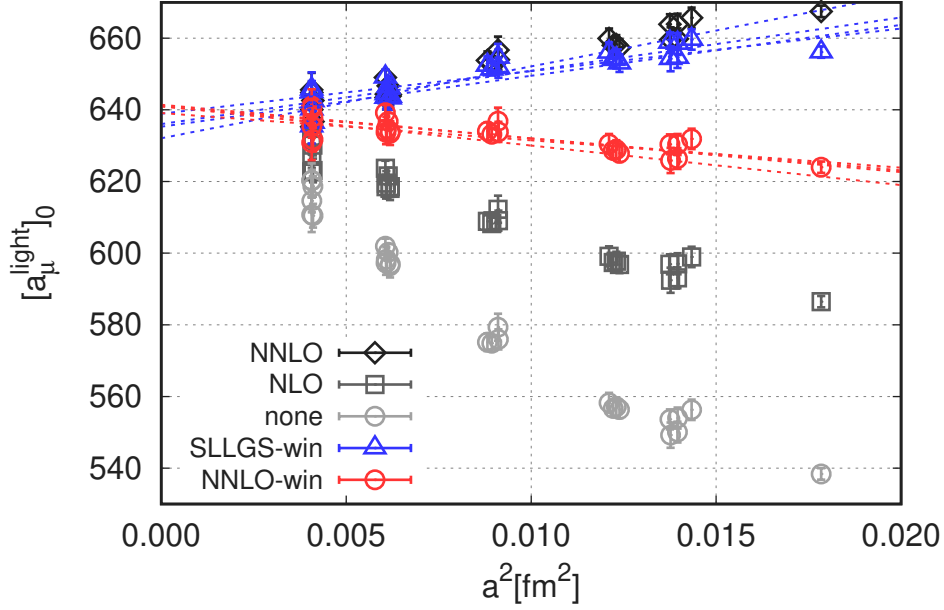
Figure 15: Continuum extrapolation of $[a_\mu^{\text{light}}]_0$ after applying different types of taste violation corrections. NLO and NNLO refer to the orders of staggered chiral perturbation theory, there is no correction at LO. SLLGS corresponds to the model presented in Section 16. The dashed lines show the continuum extrapolations used in our final result.

18 Continuum extrapolation of a_μ^{light} and a_μ^{disc}

In this section we investigate different methods to correct cutoff effects related to taste violations, and argue for two particular approaches that are used to obtain the final result of the paper. We focus on observables at the isospin-symmetric point here, since lattice artefacts in the isospin-breaking parts have negligible effect on the final uncertainties.

The connected light and disconnected components of the current propagator show significant cutoff effects with staggered fermions. They arise due to the well-known taste violation at finite lattice spacing. Not only are they significant, but they appear to decrease much faster than $O(a^2)$ as the continuum limit is approached as shown in Figure 2 for the taste violations in the pion spectrum. These two facts together suggest that the standard a^2 continuum extrapolation from commonly used lattice spacings may not be sufficient.

There is a subtle problem in connection with taste violations and finite-size effects. On coarse lattices, finite-size effects are largely suppressed, since most of the pion taste partners are heavy. The finite-size effects increase gradually for finer lattices, but even on our finest lattice, only about half of the expected effect is present, as discussed in Section 17. A good correction for cutoff effects should also restore the finite-volume dependence that is expected in the continuum limit.

Figure 15 shows the continuum extrapolation of $[a_\mu^{\text{light}}]_0$. Here we consider four linear fits, by leaving out zero, one, two or three of the coarsest lattice spacings from the fit, these are shown by dashed lines on the plot. In our final analysis, described in Section 23, we also include quadratic fits in a^2, and take into account several other systematics together with isospin-breaking effects. These lead to several thousands of fits, from which the final value for a_μ^{light} is obtained. The four fits in this section serve as an illustration.

In Figure 15 different symbols correspond to different approaches to remove taste violations. From the four continuum-extrapolatied values we then make a weighted average using the Akaike Information Criterion (AIC) weights of the fits, as discussed in Section 20. The following results are obtained:

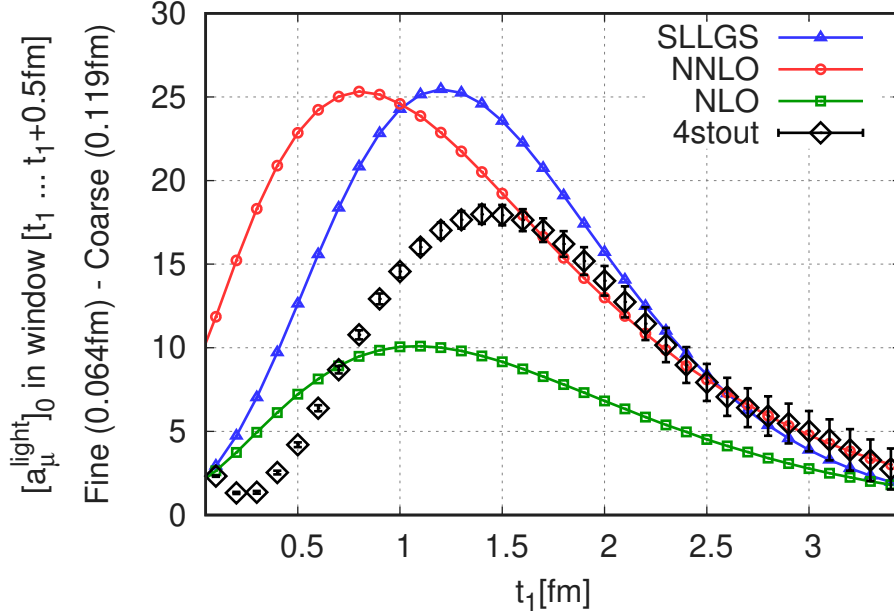| correction type $\rightarrow$	none	NLO SXPT	NNLO SXPT	NNLO SXPT-win	SLLGS-win
$[a_\mu^{\text{light}}]_0(L_{\text{ref}}, T_{\text{ref}})$	650(3)	643(2)	633(2)	641(2)	635(3)

Figure 16: $[a_\mu^{\text{light}}]_0$ computed with a sliding window: the window starts at t_1 and ends 0.5 fm later. The plot shows the difference between a fine and a coarse lattice, the volumes are $L = 6.14$ fm and $L = 6.67$ fm. The black squares with errors are obtained from the simulation. The colored curves are the predictions of NLO and NNLO staggered chiral perturbation theory and the SLLGS model. They are computed at the parameters (pion mass, taste violation, volume) of the simulations.

The different abbreviations and the role of the reference box, given by spatial and time extents $L_{\text{ref}} = 6.272$ fm and $T_{\text{ref}} = \frac{3}{2}L_{\text{ref}}$, are explained below. In the case of a fit without any correction[4], the fit qualities are bad and the continuum extrapolated values differ significantly. The AIC weighted result is dominated by the fit with the three finest points.

Much better fit qualities compared to the uncorrected case can be obtained if we apply a correction, obtained from staggered chiral perturbation theory (SXPT), to the lattice results before the continuum extrapolation. Section 15 provides the necessary formulas. Specifically, each data point in a box with size L and T and a lattice with spacing a gets an additive shift as

$$
\begin{aligned}
[a_\mu^{\text{light}}]_0(L, T, a) \to {}& [a_\mu^{\text{light}}]_0(L, T, a) + \tfrac{10}{9}\left[a_\mu^{\text{XPT}}(L, T) - a_\mu^{\text{SXPT}}(L, T, a) \right] \\
& + \tfrac{10}{9}\left[a_\mu^{\text{XPT}}(L_{\text{ref}}, T_{\text{ref}}) - a_\mu^{\text{XPT}}(L, T) \right] \; .
\end{aligned}
\tag{134}
$$

The first additive correction removes taste violation artefacts. It vanishes as the lattice spacing goes to zero, and therefore provides a valid continuum extrapolation procedure. Note, however, that on our lattice spacings, the correction changes the value of the a^2-continuum extrapolation. This is because it behaves much more like $\sim a^4$ than a pure a^2. The second additive correction in Equation (134) corrects for the small differences between the volumes in our different simulations by shifting them to a common system size given by the reference box. The finite-size effect of the reference box is computed in dedicated simulations in Section 17. There is a charge factor $\left(\frac{10}{9}\right)$ in front of both correction terms. It is required, because the XPT results of Section 15 correspond to the $I = 1$ contribution, not to the light one.

The correction with NLO SXPT already makes fit qualities good, except for the case when all lattice spacings are included. It also decreases the slope of the continuum extrapolation by a factor of two compared to the unimproved case and reduces the variation in the extrapolated values. The AIC weighted result is significantly smaller than the one obtained without corrections.

With one order higher, using NNLO SXPT, all fit qualities are good and the sensitivity to leaving out the

---

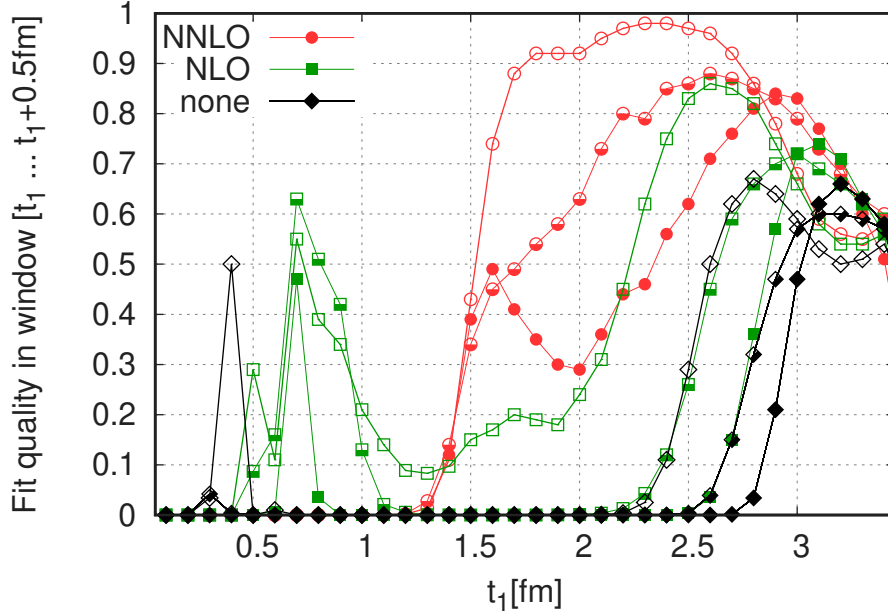[4]This corresponds to the LO of staggered chiral perturbation theory.

Figure 17: Fit qualities of continuum-extrapolation fits for $[a_\mu^{\text{light}}]_0$ as a function of t_1, where the propagator is restricted to a window $[t_1, t_1 + 0.5\ \text{fm}]$. Different colors correspond to corrections obtained using different orders of staggered chiral perturbation theory. LO corresponds to no correction at all. Different symbols with the same color correspond to different number of the coarse lattice spacings ignored in the fit: filled/half-filled/empty for zero/one/two.

coarse lattices is below the statistical errors. However, this order seems to "overdo" the improvement: it turns the extrapolation function, which was increasing towards the continuum limit, to one that approaches it from above. As a result the continuum extrapolated value is further reduced. Also we see no signs of convergence: the NNLO correction is more than twice as large as the NLO.

More can be learned about cutoff effects by looking at the propagator G_0^{light} instead of $[a_\mu^{\text{light}}]_0$. To get rid of the staggered oscillations we slide over the propagator with a smoothing window and compute the $[a_\mu^{\text{light}}]_0$ in those windows. We use the window function from Equation (71) with a step width of $\Delta = 0.15$ fm and a width of $t_2 - t_1 = 0.5$ fm. We then compute the difference between a coarse and a fine lattice, $\beta = 3.7500$ and 4.0126. The lattice result is plotted in Figure 16 as a function of t_1, together with the curves obtained from NLO, NNLO SXPT and SLLGS. These are evaluated at the parameters of the ensembles. We see that NLO SXPT reproduces the cutoff effect only for distances larger than about 3.0 fm, whereas the NNLO/SLLGS starts to work already at about $1.5/2.0$ fm. Another important observation is that both NNLO and SLLGS fail badly in the region below ca. 1.0 fm, there the NLO is much closer to the actual cutoff effect. This is not surprising, since both NNLO and SLLGS have a $t \to 0$ behaviour which does not agree with QCD's $\sim t^{-3}$.

We can corroborate these findings by comparing the fit qualities of the a^2-linear continuum extrapolations for the sliding windows. For each window, starting at t_1 and ending at $t_2 = t_1 + 0.5$ fm, we perform continuum extrapolations with the LO, NLO and NNLO corrections. The fit qualities of these as a function of t_1 are shown in Figure 17. The different colors correspond to different corrections, whereas the different symbols represent different number of coarse lattices dropped in the fit (zero, one or two). Below $t_1 \approx 0.3$ fm the statistical error is so small that none of the fits have a good fit quality. Above this there is a short range where the uncorrected fits are good. Then, in the range $0.5 \lesssim t_1 \lesssim 1.3$ fm, acceptable fits are obtained using the NLO correction, while the other fits are much worse. Between $1.3 \lesssim t_1 \lesssim 3.0$ fm the NNLO corrected fits have the best quality, as the NNLO corrected data is almost perfectly linear in a^2. This agrees with the findings in Figure 16, that above $t_1 \approx 1.5$ fm the NNLO describes the cutoff effects well. Above $t_1 \approx 3.0$ fm the statistical errors become large, and all fits are good. The picture is very similar if we replace the NNLO correction with the one obtained from the SLLGS model.
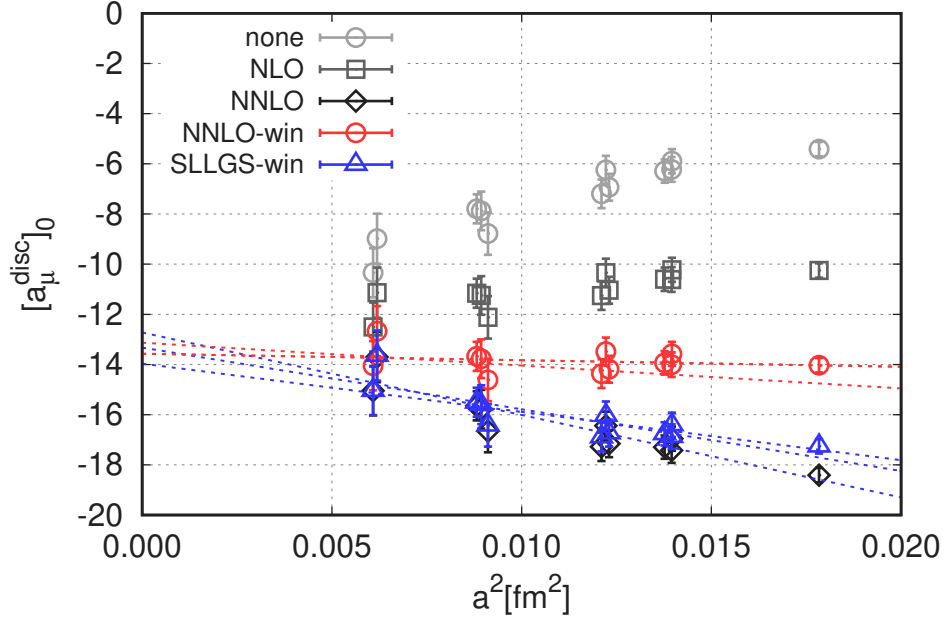
Figure 18: Continuum extrapolation of $[a_\mu^{\text{disc}}]_0$ after applying different types of taste violation corrections. NLO and NNLO refer to the orders of staggered chiral perturbation theory, there is no correction at LO. SLLGS corresponds to the model presented in Section 16. The dashed lines show the continuum extrapolations that are used in our final result.

These findings explain why the continuum extrapolation of $[a_\mu^{\text{light}}]_0$ using the NNLO correction seems to overdo the improvement, as seen in Figure 15. For windows with $t_1 \lesssim 1.0$ fm, the NNLO largely overestimates the size of the cutoff effect. This might not come as a surprise, since the chiral expansion is not expected to work at short distances. This motivates us to apply the NNLO correction only in windows where it provides good fit qualities. Accordingly we propose to use the correction

$$[a_\mu^{\text{light}}]_0(L, T, a) \rightarrow [a_\mu^{\text{light}}]_0(L, T, a) + \tfrac{10}{9}\left[a_{\mu,\text{win1}}^{\text{NLO}-\text{XPT}}(L_{\text{ref}}, T_{\text{ref}}) - a_{\mu,\text{win1}}^{\text{NLO}-\text{SXPT}}(L, T, a)\right] +$$
$$+ \tfrac{10}{9}\left[a_{\mu,\text{win2}}^{\text{NNLO}-\text{XPT}}(L_{\text{ref}}, T_{\text{ref}}) - a_{\mu,\text{win2}}^{\text{NNLO}-\text{SXPT}}(L, T, a)\right], \tag{135}$$

ie. apply the NLO correction in some window-1 and the NNLO in a complementary window-2. In Figure 15 we show the continuum extrapolation obtained by this type of correction with red color and with the label "NNLO-win". The window limits were set based on the fit qualities in Figure 17: window-1 corresponds to the range $[0.5 \ldots 1.3]$ fm, and window-2 to the range above 1.3 fm. No correction is applied for $t < 0.5$ fm. Just as in the case of the NNLO correction over the full t-range, the fit qualities are good even when including the coarsest ensemble, and the sensitivity to dropping coarse ensembles is small. The important difference to the NNLO fit is that this improved correction also gives a good fit in the short-intermediate time range.

We use the taste-violation correction in Equation (135) to improve the continuum limit in our final analysis. Two choices for the window-1 limits are taken: $[0.4 \ldots 1.2]$ fm and $[0.6 \ldots 1.4]$ fm. The variation in the result due to this choice is included in our systematic error.

Analogously we can set up taste violation corrections with the SLLGS model, replacing the NNLO SXPT used in window-2 with SLLGS and keeping the rest of the analysis the same. This gives the blue curves and points in Figure 15, labeled by "SLLGS-win". The AIC weighted result for the continuum extrapolation is somewhat smaller than in the case of NNLO-win. We add the difference between SLLGS-win and NNLO-win to the systematic error associated with the continuum-extrapolation procedure.

A similar investigation can be carried out in the case of $[a_\mu^{\text{disc}}]_0$. There we apply corrections as in Equations (134) or (135), but with a charge factor $\left(-\tfrac{1}{9}\right)$ instead of $\left(\tfrac{10}{9}\right)$. For example the analogoue of

Equation (134) is:

$$[a_\mu^{\text{disc}}]_0(L, T, a) \to [a_\mu^{\text{disc}}]_0(L, T, a) - \tfrac{1}{9} \left[a_\mu^{\text{XPT}}(L_{\text{ref}}, T_{\text{ref}}) - a_\mu^{\text{SXPT}}(L, T, a) \right] \ . \tag{136}$$

The resulting continuum extrapolations are shown in Figure 18. The unimproved data points show severe lattice artefacts, and the NNLO SXPT seems to overdo the improvement again. On the other hand, applying the NNLO SXPT only at large distances results in almost flat continuum extrapolations.

19 Global fit procedure

In this section we describe the procedure that is used to obtain the physical values of a_μ. Two types of fit functions are introduced, Type-I and Type-II, which differ in their input parameters. In Type-I fits these are experimentally measurable quantities. In Type-II fits the inputs are observables that are not directly accessible in experiments. Type-II fits are needed to implement the separation of observables into isospin-symmetric contribution and isospin-breaking corrections that is described in Section 6. We close the section by presenting an alternative fit procedure.

Type-I fits

In the case of Type-I fits we parameterize the quark-mass and electric-charge dependence of an observable Y around the physical point and for small isospin breaking with a linear function f:

$$Y = f(\{X\}; A, B, \dots) \equiv A + BX_l + CX_s + DX_{\delta m} + EX_{vv} + FX_{vs} + GX_{ss} \ . \tag{137}$$

The $X_l, X_s, \dots$ are called independent variables of the fit function, though they can be (statistically) correlated. The $A, B, \dots$ are called the fit coefficients. The Type-I fits have the feature that their independent variables $\{X\}$ are quantities that are experimentally measurable. Here the X_l and X_s variables describe the deviation from the physical light and strange mass

$$X_l = \frac{M_{\pi_0}^2}{M_\Omega^2} - \left[\frac{M_{\pi_0}^2}{M_\Omega^2}\right]_* \ , \qquad X_s = \frac{M_{K_\chi}^2}{M_\Omega^2} - \left[\frac{M_{K_\chi}^2}{M_\Omega^2}\right]_* \tag{138}$$

with $*$ denoting the experimental value. No higher orders in X_l or X_s are needed, since we work close to the physical point. The remaining X variables measure the distance from the isospin-symmetric limit

$$X_{\delta m} = \frac{\Delta M_K^2}{M_\Omega^2} \ , \qquad X_{vv} = e_v^2 \ , \qquad X_{vs} = e_v e_s \ , \qquad X_{ss} = e_s^2 \ , \tag{139}$$

where e_v and e_s are the valence and sea electric charges, respectively. Higher-order isospin-breaking terms are not considered in this work. The meson masses are defined as

$$M_{K_\chi}^2 \equiv \tfrac{1}{2} \left(M_{K_0}^2 + M_{K_+}^2 - M_{\pi_+}^2\right) \ , \tag{140}$$
$$\Delta M_K^2 \equiv M_{K_0}^2 - M_{K_+}^2 \ .$$

In case of the neutral pion we use the combination

$$M_{\pi_\chi}^2 \equiv \tfrac{1}{2} \left(M_{uu}^2 + M_{dd}^2\right) \ , \tag{141}$$

$$\tag{142}$$

where the masses of mesons uu and dd are obtained from contractions involving connected diagrams only. It can be shown in partially-quenched chiral perturbation theory coupled to photons [34], that $M_{\pi_0} = M_{\pi_\chi}$ up to terms that are second order in isospin breaking.

The coefficients $A, B, \dots$ in Equation (137) are specific to the observable Y. They can depend on

the lattice spacing, and also on the X variables defined above, in particular we use:

$$
\begin{aligned}
A &= A_0 + A_2 a^2 + A_4 a^4, \\
B &= B_0 + B_2 a^2, \\
C &= C_0 + C_2 a^2, \\
D &= D_0 + D_2 a^2 + D_4 a^4 + D_l X_l + D_s X_s, \\
E &= E_0 + E_2 a^2 + E_4 a^4 + E_l X_l + E_s X_s, \\
F &= F_0 + F_2 a^2, \\
G &= G_0 + G_2 a^2.
\end{aligned}
\tag{143}
$$

The lattice spacing a is defined in a so-called mass-dependent, scale-setting scheme: for any ensemble, a is given as the ratio of the Ω mass measured in lattice units divided by its experimental value. In the A, D, E coefficients we use both linear and quadratic dependecies in a^2; all other depedencies are assumed to be linear. Depending on the fit qualities, some of these parameters will be set to zero.

The parameters $A_0, A_2, A_4, B_0, \ldots$ can be determined by performing a fit for sufficiently many ensembles that scatter around the physical point. The physical value of Y can then be obtained from this fit as

$$
Y_* = A_0 + D_0 [X_{\delta m}]_* + (E_0 + F_0 + G_0) \cdot e_*^2 ,
\tag{144}
$$

ie. by setting the independent variables X to their physical values, including setting the valence and sea electric charges to the physical value of the coupling e_*. The value e_* is related to the experimental value of the fine structure constant as $e_* = \sqrt{4\pi\alpha_*}$. This choice is valid up to second order in isospin-breaking.

As described in Section 5, isospin-breaking corrections are obtained by measuring derivatives with respect to the δm, e_s and e_v parameters. These can be incorporated into the above procedure by deriving a system of coupled equations: one by taking Equation (137) at the isospin-symmetric point, and the other four by applying the isospin breaking derivatives, see Equations (26) and (29). We then find the following five equations:

$$
\begin{aligned}
[Y]_0 &= [A + BX_l + CX_s]_0 \\
[Y]'_m &= [DX_{\delta m}]'_m \\
[Y]''_{20} &= [A + BX_l + CX_s + DX_{\delta m}]''_{20} + [E]_0 \\
[Y]''_{11} &= [A + BX_l + CX_s + DX_{\delta m}]''_{11} + [F]_0 \\
[Y]''_{02} &= [A + BX_l + CX_s + DX_{\delta m}]''_{02} + [G]_0
\end{aligned}
\tag{145}
$$

where various isospin components of the coefficients $A, B, \ldots$ have to be included, eg. the isospin symmetric value of E is given by:

$$
[E]_0 = E_0 + E_2 [a^2]_0 + E_4 [a^4]_0 + E_l [X_l]_0 + E_s [X_s]_0 .
\tag{146}
$$

The first line in (145) parameterizes the isospin-symmetric data, and is the only equation that depends on the A_0 parameter. The next equation describes strong-isospin-breaking, where the electromagnetic coefficients E, F, G trivially drop out. B and C are also absent here, since they depend symmetrically on the u and d quarks. This equation is the main constraint for D. The final three equations are the electric derivatives; they constrain the E, F and G coefficients.

Note that the derivatives in Equation (145) are with respect to the bare parameters. The strong-isospin-breaking derivative $[\ldots]'_m$ defines a renormalized observable, but the electric charge derivatives do not. This is due to the fact that the electric charge changes the running of the quark masses and the

57

lattice spacing. However, differences like

$$[Y]''_{20} - [A + BX_l + CX_s + DX_{\delta m}]''_{20} \,, \tag{147}$$

which actually appear in (145), are free of divergences. When preparing plots to illustrate the continuum extrapolation, the electric derivatives will always refer to such renormalized combinations.

Type-II fits

We introduce a second type of parametrization, called Type-II, in order to obtain the isospin decomposition described in Section 6. Type-II fits use the w_0-scale for scale setting and are defined through:

$$Y = f(\{\tilde{X}\}; \tilde{A}, \tilde{B}, \dots) \equiv \tilde{A} + \tilde{B}\tilde{X}_l + \tilde{C}\tilde{X}_s + \tilde{D}\tilde{X}_{\delta m} + \tilde{E}\tilde{X}_{vv} + \tilde{F}\tilde{X}_{vs} + \tilde{G}\tilde{X}_{ss} \,, \tag{148}$$

where the independent variables of the fit function are defined as

$$\begin{aligned}
\tilde{X}_l &= M_{\pi_\chi}^2 w_0^2 - [M_{\pi_\chi}^2 w_0^2]_* \,, \qquad \tilde{X}_s = M_{ss}^2 w_0^2 - [M_{ss}^2 w_0^2]_* \,, \\
\tilde{X}_{\delta m} &= \Delta M^2 w_0^2 \,, \qquad \tilde{X}_{vv} = X_{vv} \,, \qquad \tilde{X}_{vs} = X_{vs} \,, \qquad \tilde{X}_{ss} = X_{ss} \,,
\end{aligned} \tag{149}$$

with $\Delta M^2 = M_{dd}^2 - M_{uu}^2$. Some of the $\tilde{X}$ variables contain w_0, M_{ss} and ΔM^2, that cannot be measured experimentally. The physical values of these quantities have to be determined from a Type-I fit of Equation (137) first. $\tilde{A}$, $\tilde{B}$, ... in general depend on hadron masses and on the lattice spacing, analogously to the dependencies in Equation (143). Here the lattice spacing is defined through w_0: it is the physical value of w_0 divided by the one measured in lattice units. The fit procedure is also completely analogous to the one described above, including the coupled equations for the different isospin components. The isospin decomposition can be obtained from the Type-II fit coefficients as

$$[Y]_{\text{iso}} = \tilde{A}_0, \quad [Y]_{\text{sib}} = \tilde{D}_0[\Delta M^2 w_0^2]_*, \quad [Y]_{\text{qed}} = (\tilde{E}_0 + \tilde{F}_0 + \tilde{G}_0) \cdot e_*^2. \tag{150}$$

One can also decompose the electromagnetic contribution further to valence-valence, valence-sea and sea-sea parts:

$$[Y]_{\text{qed}-vv} = \tilde{E}_0 e_*^2, \quad [Y]_{\text{qed}-sv} = \tilde{F}_0 e_*^2, \quad [Y]_{\text{qed}-ss} = \tilde{G}_0 e_*^2. \tag{151}$$

The two fit types, Type-I in Equation (137) and Type-II in Equation (148) have to yield the same physical value Y_* within error bars. This was always the case for the observables considered here. Later, when we discuss the fits, it will be obvious from the text which parametrization we are working with, so we drop the ~ from the coefficients of the Type-II fits for simplicity.

Correlations

In both parametrizations we have to work with a system of equations such as (145), where the unknown parameters are contained in $A, B, \dots$. To obtain these we perform a fit taking $[Y]_0$ and the isospin derivatives from several ensembles. The $[Y]_0, [Y]'_m$ and $[Y]''_{20}$ components are measured on the same $L \approx 6$ fm ensembles of Table 1, and they are therefore correlated. One also has to take into account the correlation between the sea quark derivatives $[Y]''_{11}$ and $[Y]''_{02}$ that are measured on the $L \approx 3$ fm ensembles of Table 8. These correlations have to be properly included in the fit. Also, we have to take into account the correlation of Y and the independent variables $\{X\}$, including the lattice spacing. Specifically we compute and minimize the following function to determine the fit parameters $A_0, \dots$:

$$\chi^2 = \sum_{i,j} (Y_i - f_i) \, \text{Cov}_{ij}^{-1} \, (Y_j - f_j). \tag{152}$$

Here the sums run over all ensembles and Y_i (f_i) are the values of the observable (function) on ensemble i. The matrix Cov_{ij} is the statistical covariance of the residuals $Y_i - f_i$, computed as

$$\mathrm{Cov}_{ij} = \overline{\left[(Y_i - f_i) - \overline{(Y_i - f_i)}\right]\left[(Y_j - f_j) - \overline{(Y_j - f_j)}\right]}, \tag{153}$$

where we denote the statistical average with an overline. Using the jackknife samples this can be obtained as:

$$\mathrm{Cov}_{ij} = \frac{N_J - 1}{N_J} \sum_{J=1}^{N_J} \left[\left(Y_i^{(J)} - f_i^{(J)}\right) - \left(Y_i^{(0)} - f_i^{(0)}\right)\right]\left[\ldots i \to j \ldots\right], \tag{154}$$

where an upper index (J) means that the quantity is computed on the J-th jackknife sample and $J = 0$ stands for the average over all jackknife samples. The minimization of the χ^2-function yields non-linear equations for the parameters, since the Cov matrix depends on them too. To solve the minimization problem numerically, we first guess the minimum by ignoring the parameter dependence of the Cov matrix. In all cases this was already a good starting point, which is related to the fact, that the errors on Y are typically much larger than on X. This guessing can be iterated and after a few iterations we switch to Newton's method to accelerate the convergence.

Alternative fit procedure

In addition to the previously described fit procedure we also use an alternative approach, in which isospin corrections are included in a different way. The idea is to use the fit function, eg. the Type-I function in Equation (137), directly without working with the isospin breaking derivatives of that function. For this purpose we create new, "virtual" ensembles in addition to the already existing isospin-symmetric ones. These virtual ensembles have an isospin breaking with one or more of the e_s, e_v and $(m_d - m_u)/m_l$ parameters set to non-vanishing values, close but not necessarily exactly to their physical values. The observables on the virtual ensembles are computed using the isospin-symmetric values and isopsin breaking derivatives measured on the original ensembles. For the global fit we use the original ensembles together with these newly created ones. Since the virtual ensembles were created from the original isospin-symmetric ensembles, there are strong correlations between them. Computing the covariance matrix is similar to Equation (154), but now the indices i, j run over all ensembles, including the newly created ones.

We used this technique, with the Type-I fit function, to compute the quantities w_0, M_{ss}^2 and ΔM^2. In all three cases the results had similar uncertainties as the original approach, presented earlier in this section, and for which the results can be found in Section 21. Also, the central values agreed within their systematic uncertainty in the two approaches.

20 Uncertainty estimation

Calculation of statistical errors

We use the jackknife method to calculate the statistical errors. To suppress the auto-correlation between data from subsequent configurations we introduce a blocking procedure. It is very convenient to use an equal number of blocks for all ensembles. In this work we use $N_J = 48$ blocks. With this choice we have typically 100 trajectories or more in a block, which is much larger than the autocorrelation time of the topological charge (around 20 on our finest ensembles). For the blocks we apply the delete-one principle, resulting in N_J jackknife samples plus the full sample.

We keep the correlation between all quantities calculated from the same ensemble. For simplicity, we match the jackknife samples between ensembles, too. This means that each global fit using all ensembles at the same time is performed $N_J + 1$ times. The covariance matrix is calculated only for the main sample, there is no need for the errors on the correlations here.
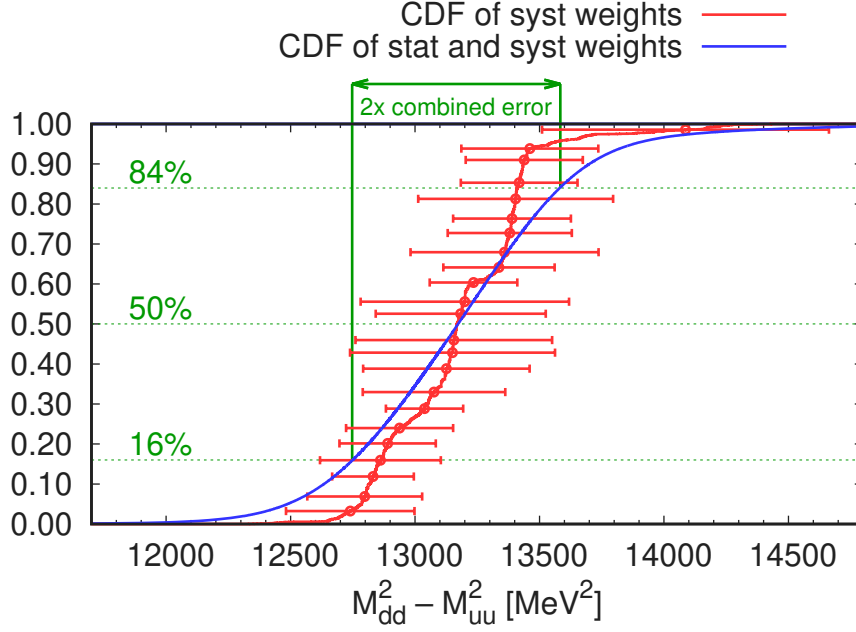
Figure 19: Cumulative distribution functions (CDF) of $\Delta M^2 = M_{dd}^2 - M_{uu}^2$ values. The red curve shows the CDF of about 3k different analyses obtained from their corresponding AIC weights. At a given point along the curve the horizontal band is the statistical error of the analysis at that point. The blue curve shows a CDF corresponding to a combined distribution of the AIC weights and the Gaussian distributions of the statistical errors. This latter curve is obtained from Equation (157) with $\lambda = 1$. We use the median and the width of this curve to define the central value and the total error. For the separation of the total error into a statistical and a systematic part we also use the CDF with $\lambda = 2$.

Estimation of systematic errors

Throughout the chain of analyses many choices are made, ranging from fit windows and mass extractions, through the various Ansätze for the large-time behavior of the JJ correlator, to the various parametrizations of a global fit. We call the global fit with a specific set of such choices an analysis. Each choice of k possible options introduces a factor k in the total number of analyses, which already includes a factor of $N_J + 1$ corresponding to the statistical sampling. Here we describe the procedure to derive a systematic error coming from the ambiguity of these choices. We follow closely the strategy introduced by us in [32] and also extend it by a new method to separate statistical and systematic errors.

For a target observable y we build a histogram from the different analyses. Each analysis gets a weight assigned. This weight is given by the Akaike Information Criterion (AIC). The AIC is derived from the Kullback-Leibler divergence, which measures the distance of the fit function from the true distribution of the points. A derivation of the formula can be found in Section 11 of [32]. Here we use a slightly modified version of the AIC:

$$\text{AIC} \sim \exp\left[-\tfrac{1}{2}\left(\chi^2 + 2n_{\text{par}} - n_{\text{data}}\right)\right], \tag{155}$$

where the χ^2, the number of fit parameters n_{par} and the number of data points n_{data} describe the global fit. The first two terms in the exponent correspond to the standard AIC, the last term is introduced to weight fits with different number of ensembles; this happens when we apply cuts in the lattice spacing. It can be derived from the Kullback-Leibler divergence and arises from the first term in Equation (S41) of [32]. In case of normally distributed errors this term can be computed and one obtains $\tfrac{1}{2}n_{\text{data}}$, which then leads to Equation (155) that we use in this paper.

The analyses differing only in the parametrization of the fit function, or in a cut in the lattice spacing, are weighted with their AIC weights; in the directions corresponding to other systematic variations, a flat

weighting is applied. Finally, the weights are normalized in such a way, that their sum over all analyses equals 1.

Let w_i denote the weight of the i-th analysis for a quantity y, with $\sum_i w_i = 1$. We interpret this weight as a probability. The statistical uncertainties can be included by noting that, due to the central limit theorem, they follow a Gaussian distribution $N(y; m_i, \sigma_i)$ with a central value m_i and a standard deviation σ_i. These parameters are given by the jackknife average and the jackknife error calculated from the jackknife samples in the i-th analysis. We then define a joint probability distribution function of y, including both statistical and systematic uncertainties, as:

$$\sum_i w_i N(y; m_i, \sigma_i). \tag{156}$$

In the following we work with the cumulative distribution function (CDF):

$$P(y; \lambda) = \int_{-\infty}^{y} dy' \sum_i w_i N(y'; m_i, \sigma_i \sqrt{\lambda}). \tag{157}$$

Here, for later use, we introduce a parameter λ that rescales the statistical error.

The median of the CDF is our choice for the central value of y and its total error is given by the 16% and 84% percentiles of the CDF:

$$\sigma_{\text{total}}^2 \equiv \left[\tfrac{1}{2}(y_{84} - y_{16})\right]^2 \qquad \text{with} \qquad P(y_{16}; 1) = 0.16, \qquad P(y_{84}; 1) = 0.84. \tag{158}$$

One could define a systematic error by evaluating the 16% and 84% percentiles of the $P(y; 0)$ function, since here the choice $\lambda = 0$ erases the statistical contribution to the distribution. However, $P(y; 0)$ is a sum of step functions (shown in red in Figure 19), making the percentiles a function that has jumps, which makes the definition of the systematic error highly sensitive to the value of the percentile chosen. Here we make a more robust choice for the systematic error. First we demand, that:

$$\sigma_{\text{stat}}^2 + \sigma_{\text{sys}}^2 \equiv \sigma_{\text{total}}^2. \tag{159}$$

Now, let us note that the rescaling of each jackknife error σ_i^2 with a factor λ is expected to increase the total squared statistical error with the same factor:

$$\lambda \sigma_{\text{stat}}^2 + \sigma_{\text{sys}}^2 \equiv \left[\tfrac{1}{2}(\tilde{y}_{84} - \tilde{y}_{16})\right]^2 \qquad \text{with} \qquad P(\tilde{y}_{16}; \lambda) = 0.16 , \qquad P(\tilde{y}_{84}; \lambda) = 0.84 . \tag{160}$$

Equations (158), (159) and (160) then provide a definition for separate statistical and systematic errors. If the λ is not too small, then the joint CDF is smooth and has no sudden jumps, see Figure 19, and the procedure is insensitive to the choice of λ. We use $\lambda = 2$ in our error estimations.

To understand the composition of the systematic error we calculate the **error budget** for all important quantities in the following way. Imagine that the full analysis uses 9 values of lattice spacing cuts, and we are interested in the corresponding systematic error. We first determine 9 total errors for each possible cuts. From these we construct a second CDF, which is a sum of 9 Gaussians as in Equation (157), with $i = 1 \ldots 9$, the σ_i being the total error and m_i the average of the 16 and 84 percentiles of the fits with the i-th cut, and the w_i the sum of the weights of those fits. From this CDF we derive the systematic error as done above for the original CDF, which is our result for the systematic error corresponding to the 9 cuts. We remark that the systematic errors are correlated within one error budget, distorting the quadratic sum of the components, that ought to sum up to the full systematic error.

Error propagation

Here we describe the way to propagate errors to consecutive analysis steps. Such a case occurs when we perform a Type-II fit using the physical values of w_0, M_{ss} and ΔM^2 that were determined in a Type-I fit.

The statistical errors are taken into account by keeping the jackknife samples throughout the whole analysis and computing the statistical error only in the end, ie. after the Type-II fit. For the systematic error there are certain analysis choices, like hadron mass fit ranges, that are shared between the Type-I and Type-II fits. We carry these over as we do with the jackknife samples.

There are also systematics that are independent in the two types of fits. For those, one would like to combine all of the corresponding analyses of the Type-I fit with all of those of the Type-II fit. The number of individual analyses can already be several thousand for each type of fit, and by mixing each analysis in the first step with each analysis in the second step, the total number of analyses would easily reach a million. These many combinations are unnecessary, since they include many bad fits with tiny weights.

In our approach we select N_I results from the Type-I fit by an "importance sampling": we uniformly split the probability interval $[0, 1]$ into N_I bins and, for each bin, take whichever individual fit corresponds to the midpoint of that bin. This produces a list of N_I Type-I analyses, sampled according to their importance. This selection is then the input into the Type-II fit, and the total number of analyses in the second step will only get multiplied by a factor of N_I instead of several thousands. We choose $N_I = 8$ in our analyses. We ascertained that when using these $N_I = 8$ fits, both the statistical and the systematic errors are approximately the same as when considering all fits.

21 Results for w_0, M_{ss} and ΔM^2

In this section we describe briefly the details of the global fits that are used to obtain the physical values of w_0, M_{ss} and ΔM^2 from the experimental values of hadron masses, including the mass of the Ω baryon. In all three cases we use the Type-I fit function of Equation (137), which can be related to isospin-breaking derivatives as described in Equation (145). The set of parameters, that are used in these fits, can be read off from Table 13. For a given observable some of the parameters are included in all fits, some never, and there are also some that are either included or excluded. A systematic error is associated with the latter and is given in the Table.

In the case of w_0, the observable we fit is $Y = w_0 M_\Omega$. Since $w_0 M_\Omega$ is symmetric under $u \leftrightarrow d$ exchange, no leading order strong-isospin-breaking terms can appear. Thus we can set the strong-isospin-breaking coefficient (D) to zero.

To account for the systematic error due to the different continuum extrapolations we apply both linear and quadratic functions in the isospin-symmetric component, also we skip zero/one/two/three of the coarsest lattice spacings in the linear and zero/one/two lattice spacings in the quadratic fits. For the tiny valence QED component only linear fits are applied, with zero/one/two skips; for the even smaller sea QED contributions we have either constant or linear fit with all lattice spacings.

The systematic error of the hadron mass fits is taken into account by 24 different combinations of the fit ranges: three for the M_Ω mass, two for the pseudoscalars, two for the isospin breaking of the M_Ω and two for the isospin breaking of the pseudoscalars. The pseudoscalar fit ranges are given in Table 5. For the Ω mass we use two fit ranges from the four-state fit and one from the GEVP procedure that are given in Table 6. The fit ranges for the isospin breaking components can be found in Table 9. To account for the experimental error on M_Ω we carry out the analysis with two different experimental values: one that corresponds to the central value plus the experimental error; the other with this error subtracted.

Altogether, these yield a total of 129024 fits. When the different analyses are combined into a histogram to determine the systematic error, the results from different fit functions or lattice spacing cuts are weighted with the Akaike Information Criterion, the rest with flat weighting. We obtain

$$[w_0]_* = 0.17236(29)(63)[70] \text{ fm}, \tag{161}$$

where the first error is statistical, and the second is systematic, the third is the total error; we reach a relative precision of 0.4%. The split up of the error into different sources can be found in Table 13. In Figure 20 we show the various isospin components of $w_0 M_\Omega$ against the lattice spacing squared together with the different continuum extrapolations. For the electric derivatives we took the definition in Equation (147). Our result (161) is in good agreement with earlier four-flavor determinations: $w_0 = 0.1715(9)$ fm of [88] and $w_0 = 0.1714\left({}^{+15}_{-12}\right)$ fm of [89]. In these works the isospin-breaking effects were only estimated, whereas in our case they are fully accounted for.

The same procedure is used for M_{ss} as for w_0. We actually work with $Y = (M_{ss}/M_\Omega)^2$ instead of M_{ss}/M_Ω, since the fit qualities are much better in the first case. The 129024 different fits give

$$[M_{ss}]_* = 689.89(28)(40)[49] \text{ MeV}, \tag{162}$$

with statistical, systematic and total errors as above. The error budget can be found in the second column of Table 13 and the continuum extrapolations for M_{ss}^2 are shown in Figure 21.

Finally we also carry out the analysis for $Y = \Delta M^2/M_\Omega^2$ with $\Delta M^2 = M_{dd}^2 - M_{uu}^2$. Since this observable has no isospin-symmetric part, the A, B and C coefficients are set to zero. Also, since this is an isospin splitting effect, no electromagnetic sea-sea effects can contribute, so the fit function becomes:

$$\frac{\Delta M^2}{M_\Omega^2} = D\left(\frac{\Delta M_K^2}{M_\Omega^2}\right) + E e_v^2 + F e_v e_s \tag{163}$$

Differently from the fits earlier we use a^2-quadratic fits also in the D and E coefficients. We apply four

	w_0[fm]	M_{ss}[MeV]	ΔM^2[MeV^2]
median	0.17236	689.89	13170
total error	70 (0.4%)	49 (0.07%)	420 (3.2%)
statistical error	29	28	320
systematic error	63	40	270
$M_\pi/M_K/M_{ss}$ fit	<1	24	0
$M_\pi/M_K/M_{ss}$ fit QED	4	2	140
M_Ω fit	16	4	0
M_Ω fit QED	9	<1	10
M_Ω experimental	5	1	<10
Continuum limit (beta cuts)	62	24	120
A_0 on/off	on	on	off
A_2 on/off	on	on	off
A_4 on/off	60	38	off
B_0 on/off	2	3	off
B_2 on/off	off	off	off
C_0 on/off	on	on	off
C_2 on/off	9	5	off
D_0 on/off	off	off	on
D_2 on/off	off	off	on
D_4 on/off	off	off	40
D_l on/off	off	off	40
D_s on/off	off	off	40
E_0 on/off	on	on	on
E_2 on/off	18	2	on
E_4 on/off	off	off	40
E_l on/off	7	<1	20
E_s on/off	10	2	70
F_0 on/off	on	on	on
F_2 on/off	<1	1	<10
G_0 on/off	on	on	off
G_2 on/off	2	3	off

Table 13: Physical values and error budgets for w_0, M_{ss} and ΔM^2. The errors are to be understood on the last digits of the central value, as usual. Both statistical and systematic uncertainties of these Type-I fits are propagated to the Type-II fits. The systematic uncertainties below the dashed line are propagated by choosing $N_I = 8$ representative fits, as described in the text.
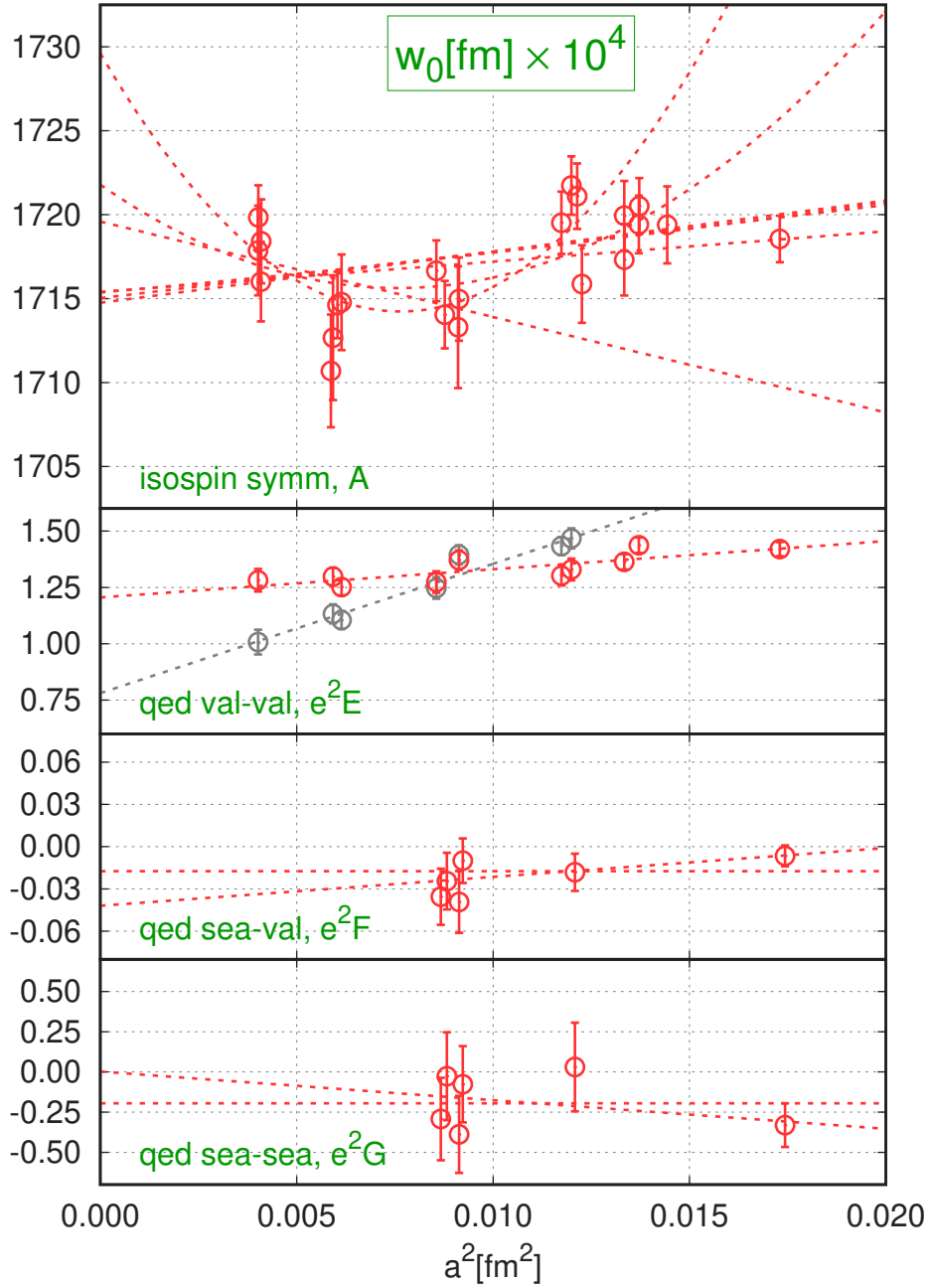
Figure 20: Continuum extrapolations of the contributions to $w_0 M_\Omega$. From top to bottom: isospin-symmetric, electromagnetic valence-valence, sea-valence and sea-sea component. The results are multiplied by $10^4/[M_\Omega]_*$. For the definitions of the components see Equations (145) and (147). The electric derivatives are multiplied by e_*^2. Dashed lines are continuum extrapolations corresponding to the lattice spacing dependent part of the A, E, F and G coefficients. They are illustrative examples from our several thousand fits. Only the lattice spacing dependence is shown: the data points are moved to the physical light and strange quark mass using the X_l and X_s dependent terms in the fit. This adjustment varies from fit to fit, the red datapoints are obtained in an a^2-linear fit to all ensembles. If in a fit the adjusted points differed significantly from the red points, we show them with grey color. The final result is obtained from a weighted histogram of the several thousand fits.
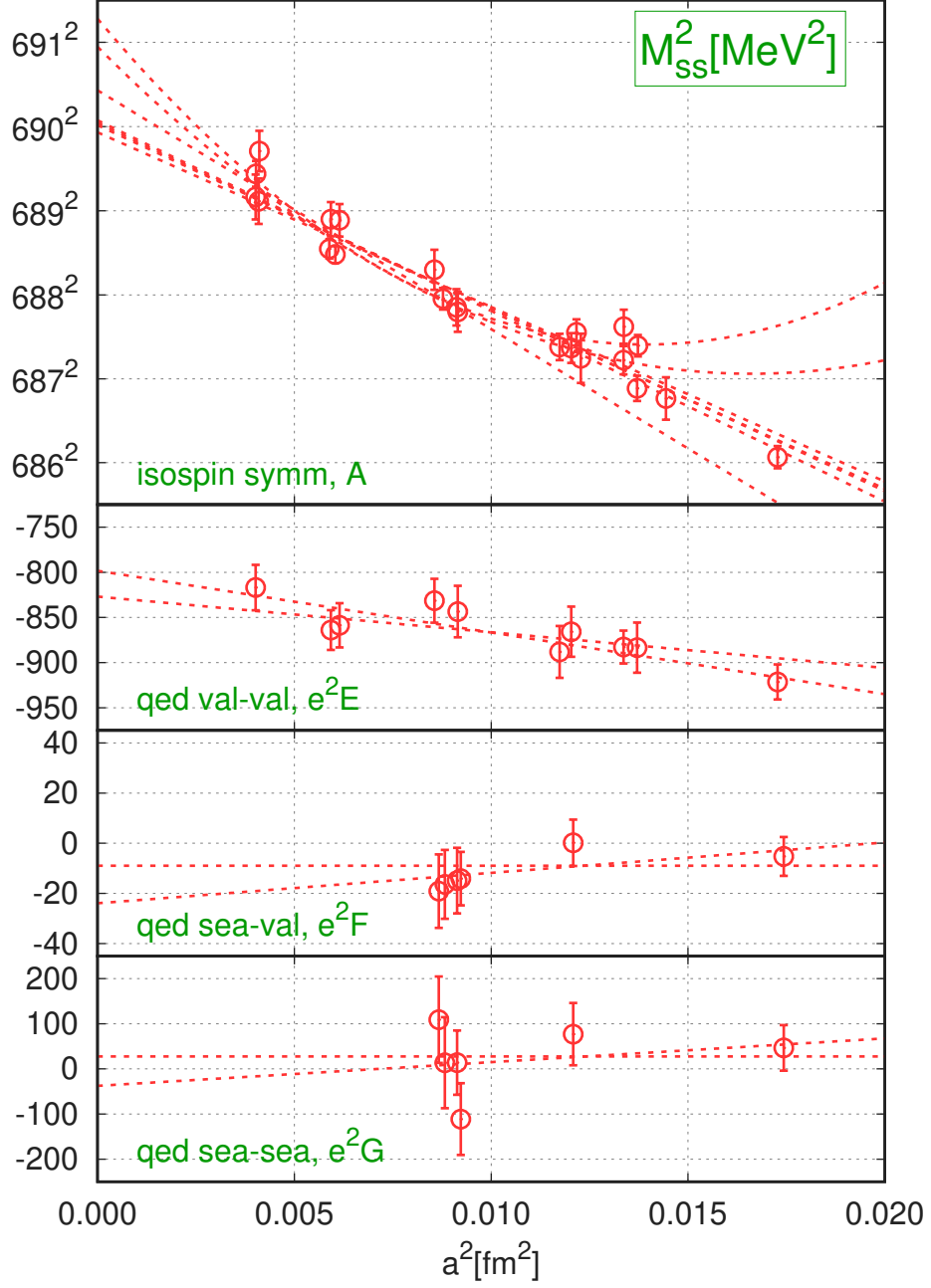
Figure 21: Continuum extrapolations of the contributions to M_{ss}^2. Plotted is the ratio M_{ss}^2/M_Ω^2 multiplied by $[M_\Omega]_*^2$. Other details as in Figure 20.
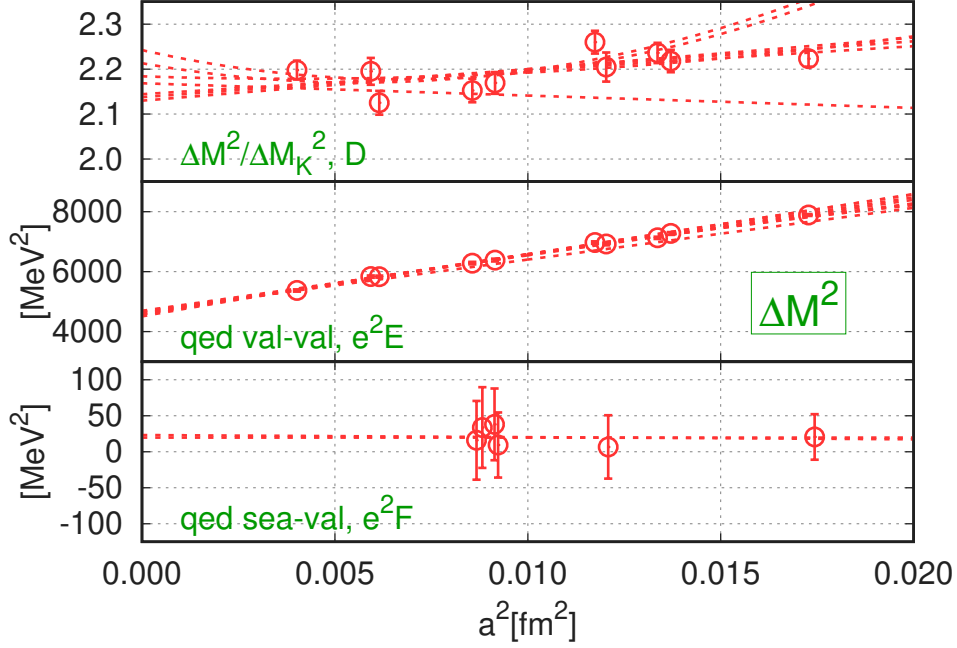
Figure 22: Continuum extrapolations of the contributions to ΔM^2. From top to bottom: $[\Delta M^2]'_m/[\Delta M_K^2]'_m$, electromagnetic valence-valence and sea-valence components of $\Delta M^2/M_\Omega^2$. The electric derivatives are multiplied by $[e^2 M_\Omega^2]_*$. Other details as in Figure 20.

lattice spacing cuts by skipping zero/one/two/three of the coarsest lattices for linear fits and three cuts for quadratic fits with zero/one/two skips. Other systematics were treated as in the above fits. Altogether we have 3328 fits, which give a central value with statistical, systematic and total errors as:

$$[\Delta M^2]_* = 13170(320)(270)[420] \text{ MeV}^2. \tag{164}$$

The corresponding error budget can be found in Table 13. In Figure 22 we show continuum extrapolations for the $\Delta M^2/\Delta M_K^2$ ratio and the valence-valence and sea-valence electric derivatives; these correspond to the D, E and F coefficients in the fit function.

22 Alternatives to the M_Ω scale-setting

To check our determination of the physical value of w_0, we consider two other scale-setting quantities: the pion decay constant f_π and the Wilson-flow scale t_0. Both approaches are independent from the systematics of the M_Ω mass determination. We also investigate a^4 effects by changing the definition of w_0 by lattice artefacts. We do not consider f_π and t_0 in our final analysis, because their relation to experiments is indirect. We work in the isospin-symmetric limit throughout this section.

In current lattice simulations it is common to use the pion decay constant for scale setting. This observable however is well defined only in the absence of electromagnetism, and thus useful only in simulations in the isospin-symmetric point. It is possible to connect the experimental decay rate of the pion to an isospin-symmetric pion decay constant, f_π. Current state-of-the-art uses a chiral perturbation theory based approach, which yields $f_\pi = 130.50(14)$ MeV [74]. There are also computations underway to determine f_π [90–92] on the lattice. In these approaches the isospin-symmetric point is defined using renormalized quark masses, which is different from our hadronic scheme in Section 6. When turning on the electromagnetic interaction, our scheme keeps certain neutral hadron masses and w_0 constant, contrary to the one used in the f_π-based scheme, where the renormalized quark masses and the strong coupling are fixed.

Here we carry out an analysis to determine an isospin-symmetric value of $[w_0]_{\text{isoq}}$ using the above f_π
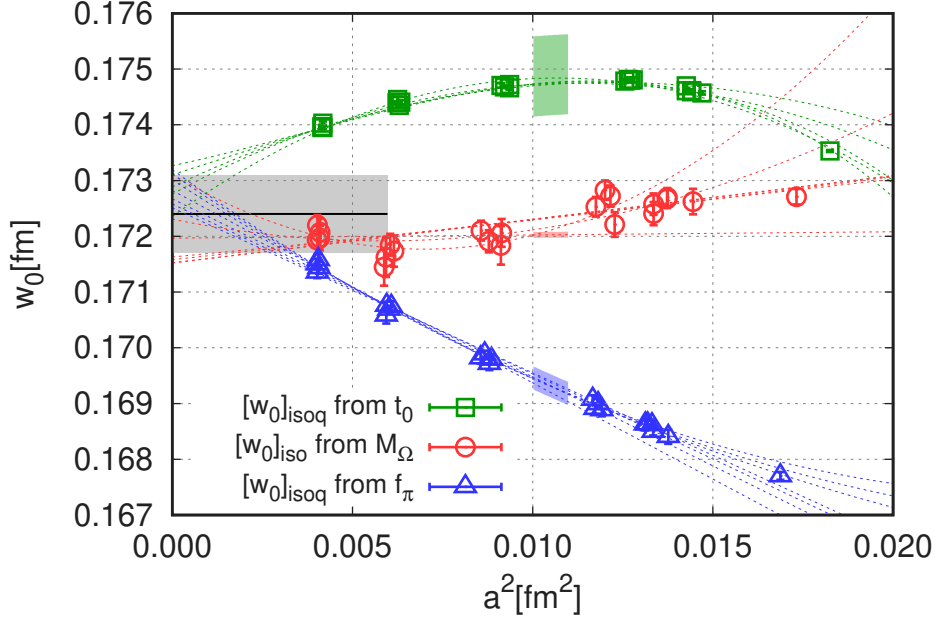
Figure 23: Continuum extrapolation of the isospin-symmetric value of w_0 using three different inputs: t_0 from the lattice work [89], M_Ω from experiment [74] and f_π from a combination of chiral perturbation theory and experiment [74]. The dashed lines are quadratic and cubic functions of a^2 in case of t_0, and linear and quadratic otherwise. The colored shaded regions around $a^2 = 0.010$ fm^2 correspond to the uncertainty in the input quantity. The horizontal grey shaded region is our final w_0 determination from Equation (161). Note, there is a difference in the definition of the isospin-symmetric point in the different inputs.

as input. We introduce the notation isoq to emphasize the difference from our definition of the isospin-symmetric value $[w_0]_{\text{iso}} = [w_0]_*$. To obtain $[w_0]_{\text{isoq}}$ we also need a pion and kaon mass that is purified from isospin-breaking effects. For these we take $M_\pi = 134.8(3)$ MeV and $M_K = 494.2(3)$ MeV [9].

The fit procedure is similar to the Type-I fits that we performed before for $w_0 M_\Omega$. The physical point is given by the f_π, M_π and M_K values above. Since we work with the isospin-symmetric component, only the A, B and C coefficients of Equation (137) are kept. We apply both linear and quadratic fits in a^2, with the usual cuts in the lattice spacing. Figure 23 shows representative fits from this analysis, with good fit qualities. The continuum extrapolated values are consistent with our $[w_0]_*$ from Equation (161). However the spread between the different continuum extrapolations is smaller, since the curvature of $w_0 f_\pi$ in a^2 is smaller than in $w_0 M_\Omega$.

Another way to determine w_0 is to take the t_0-scale, also defined from the Wilson-flow, as input. This determination basically computes the w_0/t_0 ratio. For the physical value of t_0 we use $[t_0]_{\text{isoq}} = 0.1416 \left({}^{+8}_{-5} \right)$ fm from [89], which has a precision of about 0.5%. The same analysis is carried out as before, with the difference that now we also include cubic fits in a^2, since the data shows a very strong curvature and the linear fits have a bad quality. Figure 23 shows representative fits, giving continuum values consistent with using M_Ω as input, Equation (161).

Finally we show here a method to determine $[w_0]_{\text{iso}}$, which is also based on M_Ω as an input parameter, but uses the idea of a t-shift in the Wilson flow [93]. The main reason for this analysis is to determine whether the strong quadratic upward trend in w_0 for small lattice spacings, see top panel of Figure 20, is a genuine cutoff effect? Indeed, the Wilson flow is known to have a transient for small flow times. Although the affected region shrinks as one approaches the continuum limit, the effect might be sizable particularly if we want to reach an accuracy on the few per-mil level.

The t-shift in the Wilson-flow replaces $\langle t^2 E(t) \rangle$ with $\langle t^2 E(t + sa^2) \rangle$, which is essentially applying the flow on a smeared gauge field (pre-smearing). It can be interpreted as an improved operator for the energy density. Obviously, in the continuum limit flows with or without t-shifts are the same. We measured a
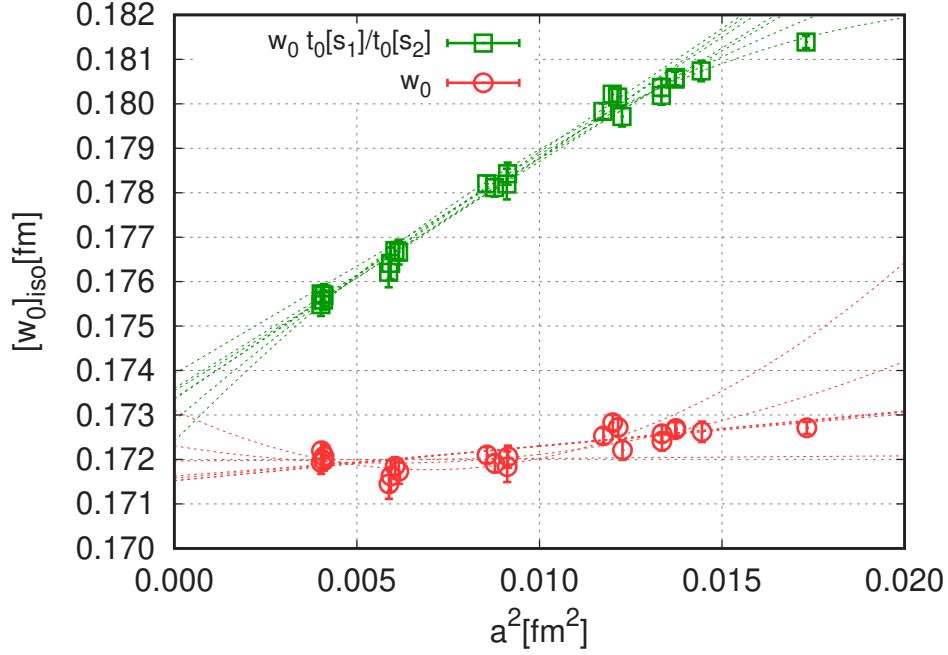
Figure 24: Continuum extrapolation of the isospin symmetric value of w_0 using M_Ω. Two w_0 definitions, the standard one (red circles), and another modified by the $t_0[s_1]/t_0[s_2]$ ratio (green squares) are shown. The ratio approaches 1 in the continuum limit.

combination, $w_0 \cdot t_0(s_1)/t_0(s_2)$, which obviously gives back w_0 in the continuum limit. Clearly, this combination has a different lattice spacing dependence than the original w_0, determined in the previous Section. There are several s_1, s_2 choices, which eliminate the strong a^4 behaviour, the upward turning of w_0 for small lattice spacings. They do so without changing the result in the continuum limit within errors. This finding indicates that the upward trend is indeed a cutoff effect related to the Wilson flow and can be removed by modified operators of the Wilson flow. As an illustration we show $s_1 = 0.35$ and $s_2 = 0.21$ in Figure 24.

The latter procedure reduces the coefficient of the a^4 term in the continuum extrapolation and, as a consequence, could also reduce the error on w_0. Since the t-shift method is a somewhat unconventional way to determine w_0, we leave it as an illustration of how cutoff effects can play a role and we quote our original w_0, Equation (161), with the larger error as our final result.

23   Results for a_μ and its various contributions

In this section we present results for the strange, light and disconnected components of a_μ in the continuum and infinite-volume limits. We perform the two limits in two separate steps. We introduce a reference box with spatial extent $L_{\mathrm{ref}} = 6.272$ fm and time extent $T_{\mathrm{ref}} = \frac{3}{2}L_{\mathrm{ref}}$. These correspond approximately to the size of our boxes in the 4stout ensemble set. In this reference box we perform the continuum extrapolation for each flavor component. The finite-size effect of the reference box is then added in the second step. For this we prepared dedicated lattice simulations, including a large box of size $L_{\mathrm{big}} = T_{\mathrm{big}} = 10.752$ fm, as discussed in Section 17. The simulations give the difference $a_\mu(L_{\mathrm{big}}, T_{\mathrm{big}}) - a_\mu(L_{\mathrm{ref}}, T_{\mathrm{ref}})$, which are in good agreement with non-lattice estimates. For the tiny residual finite-size effect, $a_\mu(\infty, \infty) - a_\mu(L_{\mathrm{big}}, T_{\mathrm{big}})$, the predictions of the non-lattice approaches are taken.

In this section we use both Type-I and Type-II parametrizations from Section 19 to perform the global fits. They give compatible results for the observables, however in most cases the Type-I results are more precise. This can be partly explained by the relatively large error on the w_0 value, Equation (161), which is caused by the strong curvature of $w_0 M_\Omega$ at small lattice spacings. We will take the final result from the Type-I fit, and use the Type-II fit to perform the isospin decomposition. To get the isospin-symmetric

value we take the total result from the Type-I fit and substract the isospin-breaking contributions obtained from the Type-II fit.

Type-II fits require the physical values of w_0, M_{ss} and ΔM^2 as input. These were determined in the previous Section. From those fits we keep the 24 different possibilites related to the hadron mass determinations. The remaining systematic variations of these are represented by $N_I = 8$ suitably chosen fit combinations, as discussed in Section 20.

Some of the fit parameters are included in all fits, some never used, and there are also ones that are included in half of the fits and excluded in the other half. Which of these options is applied for a given parameter is decided by looking at the influence of the parameter on the fit result. The options chosen can be read off from Table 14.

Connected strange contribution

The strange contribution to the connected component of a_μ, denoted by a_μ^{strange}, is obtained from the strange flavor term of the connected contractions C^{strange} given in Equations (59) and (61). Its isospin-symmetric component, as well as its electromagnetic isospin-breaking derivatives are given in Table 10. The propagator is then summed over space to project to zero momentum and in time with a weight factor:

$$a_\mu^{\mathrm{strange}} = 10^{10}\alpha^2 \sum_{t=0}^{T/2} K(t; aQ_{\mathrm{max}}, am_\mu) \frac{1}{6} \sum_{\vec{x},\mu=1,2,3} \left\langle C_{\mu,t,\vec{x};\mu,0}^{\mathrm{strange}} + C_{\mu,T-t,\vec{x};\mu,0}^{\mathrm{strange}} \right\rangle , \tag{165}$$

see Equations (64) and (69). Strong-isospin-breaking does not enter in a_μ^{strange}, so the D coefficient can be set to zero in the Type-I fit function. The systematic error estimation was carried out as in the cases of w_0 and M_{ss} in Section 21. The differences are, that E_2 is always kept and C_2 is not used. Altogether we have 32256 fits: the continuum limit and fit-form-related variations are weighted with AIC, the rest with a flat distribution. In the continuum limit we get

$$a_\mu^{\mathrm{strange}}(L_{\mathrm{ref}}, T_{\mathrm{ref}}) = 53.379(89)(67)[111] , \tag{166}$$

with statistical, systematic and total errors. The result is obtained in a finite box and the finite-size correction term will be added in a later step. The error budget for the total a_μ^{strange} is given in Table 14. We also perform a Type-II fit, from which we obtain the different isospin contributions in Table 15. The corresponding continuum extrapolations are shown in Figure 25.

Connected light contribution

The contribution of the light flavors to the connected part of a_μ, denoted by a_μ^{light}, is given by replacing the strange contraction C^{strange} with the connected light quark contraction C^{light} in Equation (165). In the isospin-symmetric part a bounding procedure is applied on the propagator to reduce the noise as discussed in Section 13. In the isospin-breaking parts we apply a cut in time, beyond which the propagator is set to zero, see Section 14. Two different cuts, given in Table 11, are used to estimate the corresponding systematic error.

As explained in detail in Section 18, the continuum extrapolation is carried out by first applying a correction to a_μ^{light} on each ensemble. This is necessary in order to remove large cutoff effects related to taste violations. Two different procedures are used for this purpose: one based on NNLO staggered chiral perturbation theory (Section 15) and another based on the Lellouch-Lüscher-Gounaris-Sakurai model (Section 16). Both of these provide a good description of the lattice artefacts for the long distance part of the propagator. For shorter distances the lattice artefacts are much smaller in relative size. There a good description of the discretization effects is given by either NLO staggered chiral perturbation theory or by no improvement at all. In particular we use no improvement below $t_1 = 0.4$ fm, NLO SXPT between t_1 and $t_2 = 1.2$ fm and either NNLO SXPT or SLLGS above t_2. To assess the systematics we also use

70

	$a_\mu^{\mathrm{strange}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$	$a_\mu^{\mathrm{light}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$	$a_\mu^{\mathrm{disc}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$
median	53.379	640.3	-18.37
total error	111 (0.2%)	4.4 (0.7%)	1.58 (8.6%)
statistical error	89	2.6	1.15
systematic error	67	3.6	1.09
$M_\pi/M_K/M_{ss}$ fit	5	<0.1	0.01
$M_\pi/M_K/M_{ss}$ fit QED	3	0.1	<0.01
M_Ω fit	56	0.3	0.04
M_Ω fit QED	2	0.1	<0.01
M_Ω experimental	5	0.1	0.01
t_c in Table 11	–	0.3	0.23
NNLO SXPT vs SLLGS	–	3.7	0.02
window borders	–	1.1	0.04
Continuum limit (beta cuts)	47	0.5	0.85
A_0 on/off	on	on	on
A_2 on/off	on	on	on
A_4 on/off	26	0.1	off
B_0 on/off	11	on	0.10
B_2 on/off	off	off	off
C_0 on/off	on	on	on
C_2 on/off	off	off	off
D_0 on/off	off	on	on
D_2 on/off	off	on	on
D_4 on/off	off	off	off
D_l on/off	off	0.2	0.02
D_s on/off	off	on	off
E_0 on/off	on	on	on
E_2 on/off	on	<0.1	0.59
E_4 on/off	off	off	off
E_l on/off	3	<0.1	off
E_s on/off	<1	off	off
F_0 on/off	on	on	on
F_2 on/off	1	<0.1	off
G_0 on/off	on	on	on
G_2 on/off	3	<0.1	off

Table 14: Continuum extrapolated results and error budget for the strange, light and disconnected contributions to a_μ. The errors are to be understood on the last digits of the central value, as usual. The results correspond to a box size $L_{\mathrm{ref}} = 6.272$ fm and $T_{\mathrm{ref}} = \frac{3}{2} L_{\mathrm{ref}}$.

	$a_\mu^{\mathrm{strange}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$	$a_\mu^{\mathrm{light}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$	$a_\mu^{\mathrm{disc}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$
total	53.379(89)(67)	640.3(2.6)(3.6)	-18.37(1.15)(1.09)
iso	53.393(89)(68)	634.6(2.7)(3.7)	-13.15(1.28)(1.29)
qed	-0.0136(86)(76)	-0.92(34)(43)	-0.58(14)(10)
qed-vv	-0.0086(42)(41)	-1.28(40)(33)	-0.55(15)(11)
qed-sv	-0.0014(11)(14)	-0.0080(85)(98)	0.011(24)(14)
qed-ss	-0.0031(76)(69)	0.42(20)(19)	-0.047(33)(23)
sib	–	6.59(63)(53)	-4.63(54)(69)

Table 15: Continuum extrapolated results for the different components of the strange, light and disconnected contributions to a_μ. The results correspond to a box size $L_{\mathrm{ref}} = 6.272$ fm and $T_{\mathrm{ref}} = \frac{3}{2} L_{\mathrm{ref}}$.

strange, $a_\mu^{\text{strange}}(L_{\text{ref}}, T_{\text{ref}})$	53.379(89)(67)	this work, Equation (166)
light, $a_\mu^{\text{light}}(L_{\text{ref}}, T_{\text{ref}})$	640.3(2.6)(3.6)	this work, Equation (167)
disconnected, $a_\mu^{\text{disc}}(L_{\text{ref}}, T_{\text{ref}})$	-18.37(1.15)(1.09)	this work, Equation (168)
finite-size, $a_\mu(\infty, \infty) - a_\mu(L_{\text{ref}}, T_{\text{ref}})$	18.7(2.5)	this work, Equation (133)
charm iso, $[a_\mu^{\text{charm}}]_{\text{iso}}$	14.6(0.0)(0.1)	[47], Table S2
charm qed, $[a_\mu^{\text{charm}}]_{\text{qed}}$	0.0182(36)	[55]
charm effect on a_μ^{disc}	<0.1	[47], Section 4 in Supp. Mat.
bottom, a_μ^{bottom}	0.271(37)	[94]
perturbative, a_μ^{pert}	0.16	[47], Table S5
one-photon-reducible subtraction, $-a_\mu^{1\gamma\text{R}}$	-0.321(11)	[95], Table II

Table 16: List of contributions to a_μ, ie. the leading order hadronic vacuum polarization contribution to the muon anomalous magnetic moment multiplied by 10^{10}.

a (t_1, t_2) pair, where both values are 0.2 fm larger. The corresponding errors can be found in the error budget of Table 14. This improvement is only applied to the isospin-symmetric component.

There is a small variation in the size of the lattices between different ensembles. We correct for this by adding a shift, computed from the finite-L and T versions of NNLO SXPT or the SLLGS model, in such a way that the results correspond to the same box size $L = L_{\text{ref}}$ and $T = T_{\text{ref}}$. This is done simultaneously with the correction for the cutoff effects, see Equation (134).

The above variations in the analysis procedure yield 258048 Type-I fits. We apply the usual weighting, in particular the NNLO SXPT and SLLGS improved results are given the same weight. For the total light connected contribution we get

$$a_\mu^{\text{light}}(L_{\text{ref}}, T_{\text{ref}}) = 640.3(2.6)(3.6)[4.4] \,, \tag{167}$$

with statistical, systematic and total errors. The used fit parameters and the error budget is given in Table 14. Performing analogous Type-II fits we get the breakup into individual isospin contributions, given in Table 15. The corresponding continuum extrapolations are shown in Figure 26.

Disconnected contribution

The disconnected contribution, denoted by a_μ^{disc}, is obtained using Equation (165) with C^{disc}, given in Equation (60), instead of C^{strange}. In our previous work, at the isospin-symmetric point [47], we computed the effect of charm quarks on a_μ^{disc} at the coarsest lattice spacing, and found that it changes the result by a value much smaller than the statistical error. Thus, we perform the current disconnected analysis without taking into account valence charm quarks.

The same analysis procedure is applied as in the case of a_μ^{light}: we use upper and lower bounds for the isospin-symmetric part, a cut in time for the isospin breaking components and we improve the continuum limit either with NNLO SXPT or SLLGS. The NNLO SXPT results are almost completely flat as a function of a^2, as such the quadratic fits give a curvature that is consistent with zero. We therefore use only a^2-linear continuum extrapolations. A further difference is that we have one less lattice spacing as in the case of a_μ^{light}. We end up with 27648 fits in total and a result of

$$a_\mu^{\text{disc}}(L_{\text{ref}}, T_{\text{ref}}) = -18.37(1.15)(1.09)[1.58] \,. \tag{168}$$

The error budget is given in Table 14. The Type-II fit gives the individual contributions in Table 15, and the continuum extrapolations are shown in Figure 27.

Finite-size effects and other contributions

Beside the three contributions that we have presented until now, there are a number of smallers ones that have to be added to get the final value on a_μ. These are listed in Table 16 together with source indications. Several of these have a size that is much smaller than our accuracy. They are given here for completeness. We discuss them, now, one by one.

The previously presented results correspond to the reference box. Its finite-size effect is computed in Section 17. This is the fourth entry in Table 16, which already includes the tiny contribution of the electromagnetic finite-size effects.

The contribution of the connected charm quark, a_μ^{charm}, was computed in the isospin-symmetric limit by many groups. Here we use our own result from [47]. An upper bound on the small effect of the charm on a_μ^{disc} was given in [47]. We use the value as an error here. The result was obtained using a single lattice spacing. The even smaller isospin-breaking corrections on a_μ^{charm} was computed in [55].

Until now we have considered four quark flavors. Obviously, the contributions of the remaining flavors have to be added. For the bottom quark contribution there is a lattice determination available, [94], whose value we use here. The top can be safely neglected at our level of precision.

The a_μ in this work involves an integration in momentum up to $Q_{\mathrm{max}}^2 = 3$ GeV^2, as discussed in Section 11. The integration from this value to infinity can be computed in perturbation theory. We use the value given in [47].

As mentioned in Section 11, we compute the current propagator with all $O(e^2)$ effects included. In this result the one-photon-reducible ($1\gamma R$) contribution belongs to the higher order HVP. This term has to be subtracted if we are interested in the leading order HVP. A recent lattice determination of this term can be found in [95]. The corresponding diagram was labeled by the letter 'c' in their Figure 2.

Summing all the contributions gives

$$a_\mu = 708.7(2.8)(4.5)[5.3] \, , \tag{169}$$

which is our final result for the LO-HVP contribution. The first error is statistical. It includes the statistical errors of the strange, light and disconnected contributions. The latter two are the dominant ones. All other uncertainties are added in quadrature, and this is given as the second, systematic error. Its major sources are the finite-size effect estimation and the continuum extrapolations. The last error in brackets is the combined error, which corresponds to a relative precision of 0.7%.
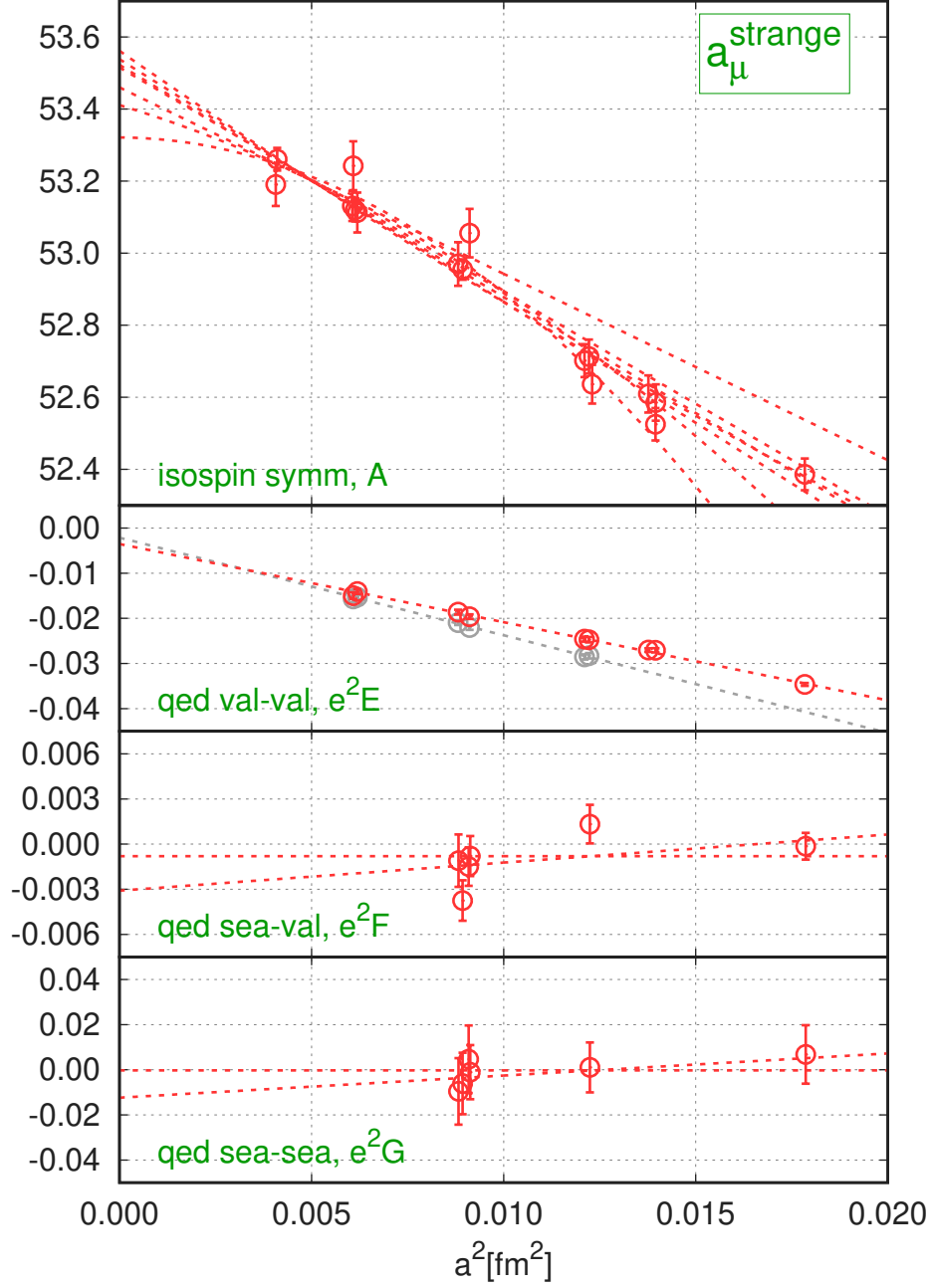
Figure 25: Continuum extrapolations of the contributions to $a_\mu^{\mathrm{strange}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$. For more explanations see the caption of Figure 20.

Figure 26: Continuum extrapolations of the contributions to of $a_\mu^{\mathrm{light}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$. For more explanations see the caption of Figure 20. In case of the light contribution we perform the analysis with two different improvement techniques, the corresponding data sets are denoted by the red and blue colors.

Figure 27: Continuum extrapolations of the contributions to of $a_\mu^{\mathrm{disc}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$.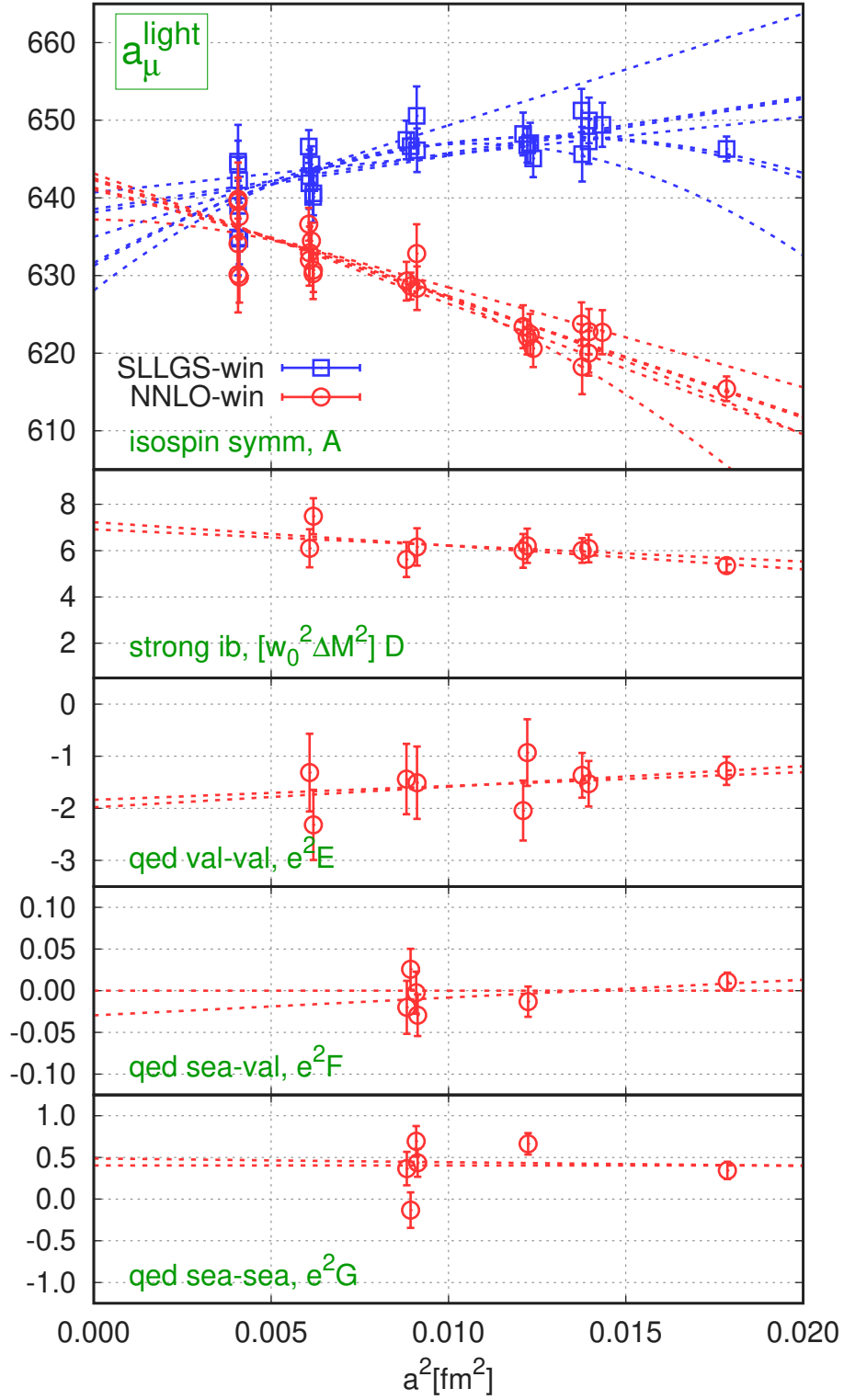 For more explanations see the caption of Figure 20. In case of the light contribution we perform the analysis with two different improvement techniques, the corresponding data sets are denoted by the red and blue colors.
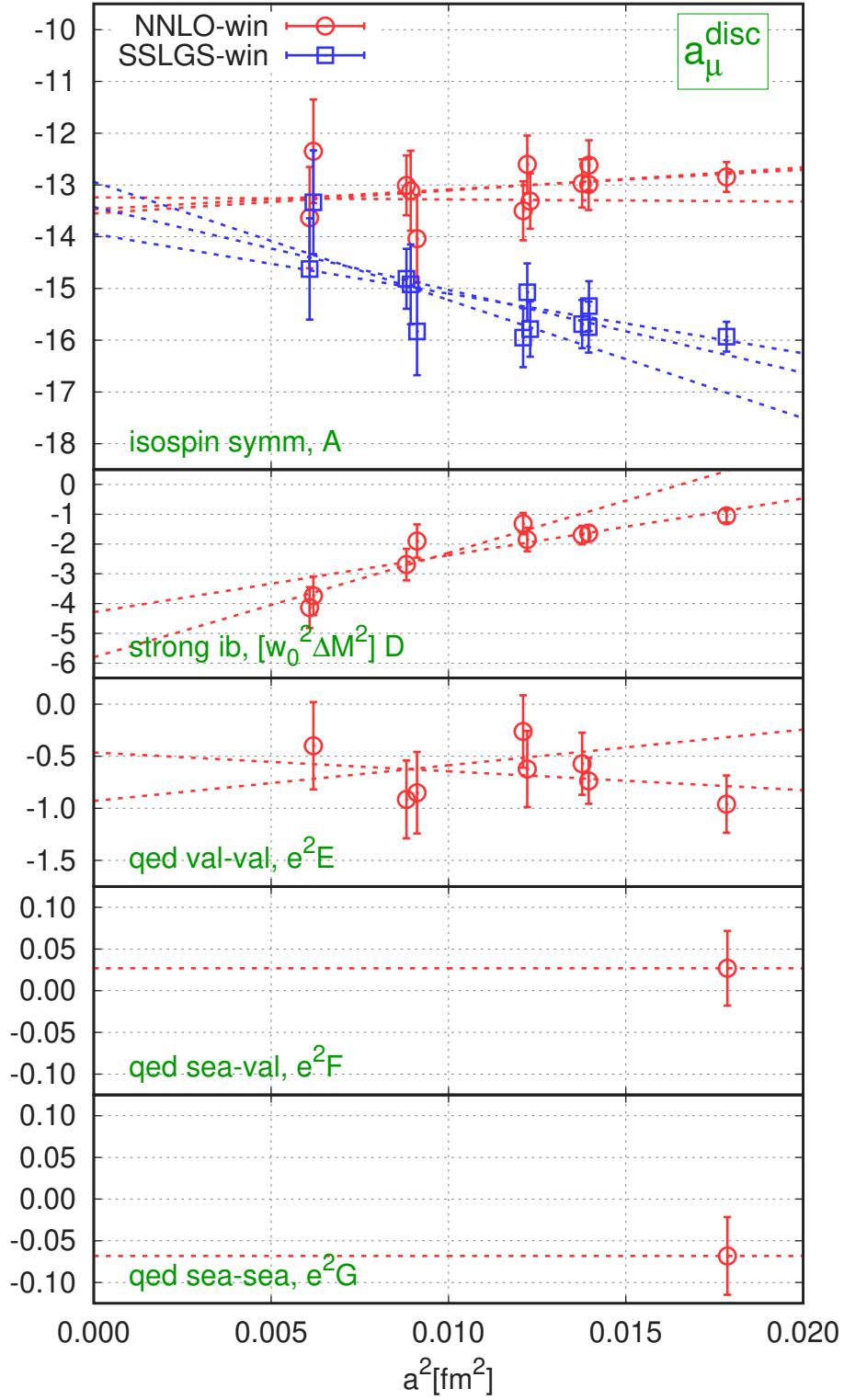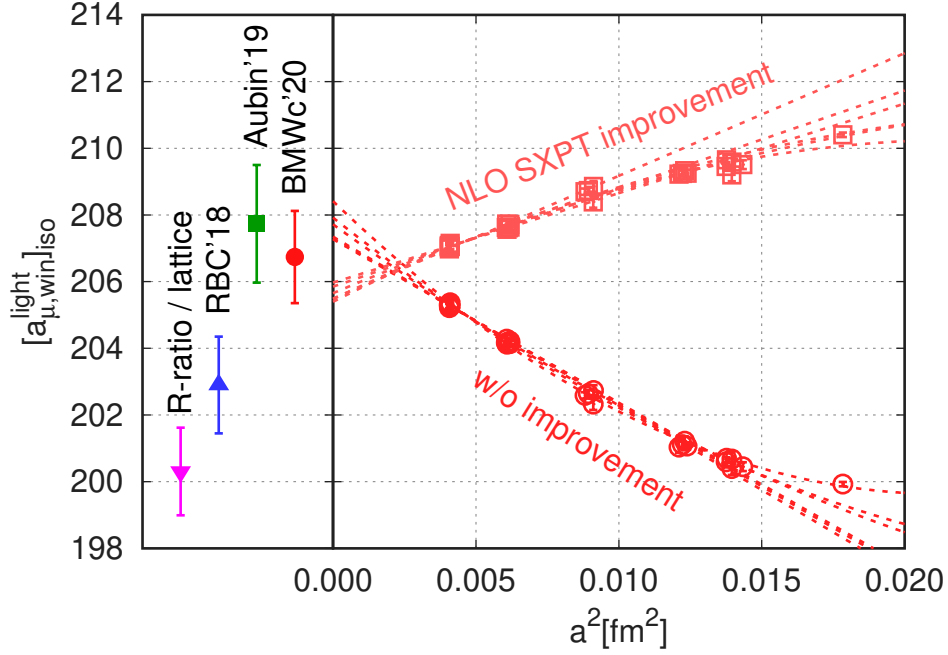
Figure 28: Continuum extrapolation of $[a_{\mu,\mathrm{win}}^{\mathrm{light}}]_{\mathrm{iso}}$. Two types of improvements are shown: one where NLO SXPT is used to improve the lattice result and another where no improvement is performed. For each, some continuum extrapolations are shown as illustration with dashed lines. They include fits linear and quadratic in a^2 and also where different numbers of coarse lattices are skipped in the fit. The data points on the plot are corrected for light and strange quark mass effects, this adjustment is different from fit to fit. The red points belong to a quadratic fit to all ensembles. Our final value in the continuum limit comes from a histogram of several thousand fits and is given by the filled red circle in the left panel. The results are corrected for finite-size effects using Equation (173). Other lattice computations are shown with a green box [51] and a blue triangle [48]. A value computed from the R-ratio method is also given (see text for details).

24   Result for $a_{\mu,\mathrm{win}}$

The work [48] defined a particularly useful observable $a_{\mu,\mathrm{win}}$, in which the current propagator is restricted to a time window $[t_1, t_2]$, using a smooth weight function $W(t; t_1, t_2)$. See Section 11 for the definition of W. The advantage of $a_{\mu,\mathrm{win}}$ over a_μ is that, by choosing an appropriate window, the calculation can be made much less challenging on the lattice than for the full a_μ. Here we will be interested in the window between $t_1 = 0.4$ fm and $t_2 = 1.0$ fm, ie. in an intermediate time range. By this choice we eliminate both the short-distance region, where large cutoff effects are present, and the long-distance region, where the statistical uncertainties, taste violations and finite-size effects are large. Because the determination of $a_{\mu,\mathrm{win}}$ does not require overcoming many of the challenges described in the main paper, other lattice groups have obtained this quantity with errors comparable to ours [48, 51]. This allows for a sharper benchmarking of our calculation. At the same time $a_{\mu,\mathrm{win}}$ can also be computed using the phenomenological approach. This is done in Section 25. Therefore, $a_{\mu,\mathrm{win}}$ is also a powerful tool to compare the results of lattice and phenomenological computations.

To compute $a_{\mu,\mathrm{win}}$ on the lattice we performed similar global fits that were used to get a_μ in Section 23. A major difference is that, in case of the light connected contribution we improve the continuum limit differently. While for a_μ we use corrections from either NNLO SXPT or the SLLGS model, here for $a_{\mu,\mathrm{win}}$ we take either NLO SXPT or no improvement at all. For this window observable these choices are more appropriate, as discussed in Section 18, where we compare the different improvements in different windows. A further difference compared to the a_μ fit procedure is that no cuts are applied on the propagator in time; the window function suppresses the propagator for distances beyond $t_2 = 1.0$ fm.

The results for the strange, light and disconnected contributions and different isospin breaking correc-

	$a_{\mu,\mathrm{win}}^{\mathrm{strange}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$	$a_{\mu,\mathrm{win}}^{\mathrm{light}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$	$a_{\mu,\mathrm{win}}^{\mathrm{disc}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$
median	27.170	207.03	-1.260
total error	30 (0.1%)	1.38 (0.7%)	0.097 (7.7%)
statistical error	28	0.19	0.037
systematic error	13	1.37	0.089
$M_\pi/M_K/M_{ss}$ fit	1	<0.01	0.001
$M_\pi/M_K/M_{ss}$ fit QED	1	<0.01	0.002
M_Ω fit	8	0.02	0.038
M_Ω fit QED	<1	<0.01	<0.001
M_Ω experimental	2	0.01	0.001
Continuum limit (beta cuts)	5	0.98	0.044
none vs NLO SXPT	–	1.28	0.066
A_0 on/off	on	on	on
A_2 on/off	on	on	on
A_4 on/off	<1	1.03	off
B_0 on/off	3	on	on
B_2 on/off	off	off	0.010
C_0 on/off	on	on	on
C_2 on/off	1	0.11	0.045
D_0 on/off	off	on	on
D_2 on/off	off	on	on
D_4 on/off	off	off	off
D_l on/off	off	0.01	off
D_s on/off	off	on	off
E_0 on/off	on	on	on
E_2 on/off	on	on	on
E_4 on/off	off	off	off
E_l on/off	<1	0.01	0.003
E_s on/off	on	off	off
F_0 on/off	on	on	on
F_2 on/off	1	<0.01	off
G_0 on/off	on	on	on
G_2 on/off	2	0.01	off

Table 17:   Continuum extrapolated results and error budget for the strange, light and disconnected contributions to $a_{\mu,\mathrm{win}}$. The errors are to be understood on the last digits of the central value, as usual. The results correspond to a box size of $L_{\mathrm{ref}} = 6.272$ fm and $T_{\mathrm{ref}} = \frac{3}{2}L_{\mathrm{ref}}$.

	$a_{\mu,\mathrm{win}}^{\mathrm{strange}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$	$a_{\mu,\mathrm{win}}^{\mathrm{light}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$	$a_{\mu,\mathrm{win}}^{\mathrm{disc}}(L_{\mathrm{ref}}, T_{\mathrm{ref}})$
total	27.170(28)(13)	207.03(0.19)(1.37)	-1.260(37)(89)
iso	27.175(28)(13)	206.19(0.20)(1.37)	-0.906(42)(90)
qed	-0.0050(35)(37)	0.082(46)(34)	-0.115(17)(07)
qed-vv	-0.0018(14)(10)	0.067(30)(21)	-0.114(18)(05)
qed-sv	-0.00142(60)(30)	-0.0222(91)(52)	0.0013(23)(00)
qed-ss	-0.0018(31)(38)	0.033(30)(39)	-0.0027(26)(22)
sib	–	0.757(40)(16)	-0.238(10)(06)

Table 18:   Continuum extrapolated results for the different isospin components of the strange, light and disconnected contributions to $a_{\mu,\mathrm{win}}$. The results correspond to a box size of $L_{\mathrm{ref}} = 6.272$ fm and $T_{\mathrm{ref}} = \frac{3}{2}L_{\mathrm{ref}}$.

tions are summarized in Table 18 and the error budget for the total values and the fit parameters used, in Table 17. The largest source of error is the continuum extrapolation of the light connected component. However, it is still much smaller than the typical size of uncertainties in the full a_μ determination. In Figure 28 we plot the continuum extrapolation of the isospin-symmetric component of $a_{\mu,\text{win}}$, both for the unimproved and the NLO SXPT corrected results.

The comparison of lattice results for $[a_{\mu,\text{win}}^{\text{light}}]_{\text{iso}}$ is particularly interesting, because it allows to benchmark the leading, light-quark contribution to a_μ through a quantity that can be computed precisely without resorting to highly-advanced techniques. Using our result obtained in the reference box from Table 18 and correcting for finite-size effects (see later) we get:

$$[a_{\mu,\text{win}}^{\text{light}}]_{\text{iso}} = 206.7(0.2)(1.4)[1.4] \ , \tag{170}$$

with statistical, systematic and total uncertainties. This result is 0.4σ smaller than $[a_{\mu,\text{win}}^{\text{light}}]_{\text{iso}} = 207.7(1.8)$ of Aubin'19 [51]. Here we use the continuum-extrapolated value, which Aubin'19 obtain from their two finest lattices in the upper panel of their Figure 7, because its errors bars cover the results of the other continuum extrapolations that they consider. Compared with $[a_{\mu,\text{win}}^{\text{light}}]_{\text{iso}} = 202.9(1.4)$ of RBC'18 [48] our result is 2.0σ larger. These two comparisons yield an average deviation of 1.2σ. These two lattice results are also shown in Figure 28.

Additionally, we also made an analysis of the charm quark contribution. The total a_μ^{charm} was obtained in our previous work [47]. Here we perform a Type-II fit for $a_{\mu,\text{win}}^{\text{charm}}$. Only the isospin-symmetric component is used and we obtain the following result:

$$[a_{\mu,\text{win}}^{\text{charm}}]_{\text{iso}} = 2.7(1) \ . \tag{171}$$

Here the error is the systematic uncertainty: the statistical is an order of magnitude smaller. The isospin breaking of the charm should be well below the uncertainties of the fit. See Table 16 for the case of a_μ^{charm}. Furthermore, in our dedicated finite-size study with the 4HEX action we compute the difference of the light contribution between the "big" and reference boxes and obtain

$$a_{\mu,\text{win}}^{\text{light}}(L_{\text{big}}, T_{\text{big}}) - a_{\mu,\text{win}}^{\text{light}}(L_{\text{ref}}, T_{\text{ref}}) = 0.57(2) \ , \tag{172}$$

where the error is statistical. Applying the same procedure for $a_{\mu,\text{win}}$ as we did for a_μ in the finite-size study of Section 17, we get for the finite-size effect:

$$a_{\mu,\text{win}}(\infty, \infty) - a_{\mu,\text{win}}(L_{\text{ref}}, T_{\text{ref}}) = 0.49(2)(4) \ . \tag{173}$$

The first error is statistical, the second is an estimate of the cutoff effect of the 4HEX action. We find that the finite-T effects are even less important than in the case of the total a_μ, where they were already much smaller than the finite-L effects.

Summing up these contributions we get

$$a_{\mu,\text{win}} = 236.1(0.2)(1.4)[1.4] \ (\text{lattice}) \ , \tag{174}$$

where the first error is statistical, the second is systematic and the third in the square brackets is the first two added in quadrature. Further contributions, that are listed in Table 16, should have an effect much smaller than the uncertainties of this result.

The value can be directly compared to the one obtained from the R-ratio method in Section 25:

$$a_{\mu,\text{win}} = 229.7(1.3) \ (\text{R-ratio}) \ , \tag{175}$$

which is smaller than the lattice result by 3.4σ or 2.8%. We can also derive an R-ratio result for the isospin-symmetric light contribution. From the value in Equation (175) we subtract the lattice results for
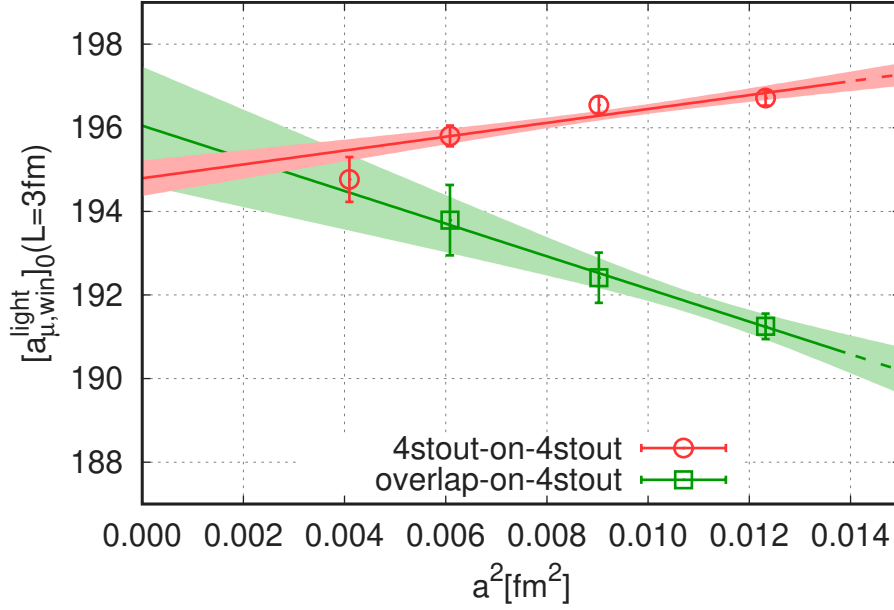
Figure 29: Continuum extrapolation of $[a_{\mu,\text{win}}^{\text{light}}]_0$. The two datasets correspond to the staggered-on-staggered and the overlap-on-staggered simulations.

all contributions, except for $[a_{\mu,\text{win}}^{\text{light}}(L_{\text{ref}}, T_{\text{ref}})]_{\text{iso}}$ and its finite-size correction. We get:

$$[a_{\mu,\text{win}}^{\text{light}}]_{\text{iso}} = 200.3(1.3) \text{ (R-ratio \& lattice)} . \tag{176}$$

This value is compared in Figure 28 to continuum and infinite-volume extrapolated lattice results from this work and from other lattice groups.

Crosscheck with overlap fermions

We perform a crosscheck of the above results with a mixed action formulation: overlap valence and 4stout staggered sea quarks. Our goal is to provide more evidence that the continuum extrapolation and the current renormalization are done correctly in the 4stout case. The overlap fermion action, the matching and the current renormalization are described at length in Section 3. Our target is the isospin symmetric value of the light connected window observable $[a_{\mu,\text{win}}^{\text{light}}]_0$. We use lattices of size $L \approx 3$ fm in this crosscheck, the related finite-volume effects are about five percent.

For the measurement of the overlap current propagator we use 512 random wall sources per configuration, the trace over color indices is performed randomly. For the staggered current propagator we use the same noise reduction technique as on the $L \approx 6$ fm lattices. In the overlap case we find, that the coarsest lattice, corresponding to $\beta = 3.7000$, is outside the a^2-scaling region. Figure 29 shows the results together with continuum extrapolations, they are in good agreement for the two formulations. The scenario, in which replacing staggered by overlap fermions removes the discrepancy between the R-ratio and the lattice result, seems improbable. In order for this to happen, the overlap result would have to get more than 4σ smaller than we determine it here.

25 Phenomenological determination of $a_{\mu,\text{win}}$

The purpose of this section is to describe the computation of the phenomenological result for $a_{\mu,\text{win}}$, which we compare with the corresponding lattice result in Section 24. For this we use the R-ratio from e^+e^- collision experiments and the corresponding covariance matrix from the work of KNT18 [70], courteously

Ref.	a_μ
KNT18 [70]	693.26(2.46)
KNT19 [96]	692.78(2.42)
DHMZ17 [97]	693.1(3.4)
DHMZ19 [69]	693.9(4.0)
CHHKS19 [98, 99]	692.3(3.3)

Table 19: Recent phenomenological determinations of a_μ. KNT stands for Keshavarzi, Nomura and Teubner; DHMZ for Davier, Hoecker, Malaescu and Zhang; CHHKS for Colangelo, Hoferichter, Hoid, Kubis and Stoffer.

given to us by Keshavarzi, Nomura and Teubner.

As Table 19 shows, there is a sizeable difference in the uncertainties on a_μ in the recent phenomenological literature. While the KNT19 result has an error of 2.42, DHMZ19 gives a 65% larger error of 4.0. Since we do not want to risk overstating possible differences between the phenomenological and lattice approaches, we extend the error estimates of KNT18 [70] by two additional sources, bringing them much closer to those of DHMZ.

a. The first source of uncertainty is related to a tension between the two e^+e^- experiments, KLOE and BaBar, which have the smallest uncertainties in the window from 0.6 to 0.9 GeV center-of-mass energy. These two experiments exhibit, for the pion-pion channel in that window, a close to 3σ discrepancy or a 2.7% relative difference [100]. Note that the overlapping energy region of the pion-pion channel for KLOE and BaBar ($0.324 - 0.972$ GeV) provides about 70% of the total a_μ. The discrepancy, fully accounted for in DHMZ19 [69], has a strong impact on the discrepancy between the measurement of $g_\mu - 2$ and theory predictions based on the R-ratio.

In order to address this discrepancy, we follow the prescription of the Particle Data Group (PDG) for similar tensions between experimental results [74]. After calculating the weighted average of all the experimental results in the $0.6 - 0.9$ GeV energy range [100], the PDG prescription tells us to adjust the error by a factor of $S = [\chi^2/(N-1)]^{1/2}$, where N is the number of experiments (in our case $N = 5$). This yields an uncertainty of 1.97 instead of the 1.32 of [70]. This uncertainty is far less than half the difference between BaBar and KLOE, because the other, less precise experiments dilute the discrepancy. We include this increased error estimate as $(1.97^2 - 1.32^2)^{1/2} = (1.46)_{\pi\pi}$, where "$\pi\pi$" denotes the uncertainty coming from the tension between experiments in the pion-pion channel. Note that the other experimental channels may have similar uncertainties, which would increase the error further.

It is worth pointing out that hadronic τ decays can be used, in principle, to provide an independent measurement of the spectral function in this important low-energy region, as first proposed in [101] and updated in [102–104]. However, this requires controlling isospin-breaking corrections [105–109], which is a challenge and has led to putting this approach aside in the last few years.

b. Another possible source of uncertainty comes from the way in which the dispersive integral of the experimental data for the various, final-state channels is performed, including correlations. KNT [70, 96] use a trapezoidal rule and argue that the error resulting from this choice is negligible. They also take into account correlations in systematic uncertainties within the same experiment and between different experiments, as well as within and between different channels, over extended ranges of center-of-mass energy. On the other hand, DHMZ [69] limit the effects of these correlations to small energy bins and use splines for integrating the data, correcting for biases if necessary. The end result is that, despite using the same experimental input, the two teams find results for the various channels which differ, more often outside the error bars of [70] than of [69]. To account for this when using the correlation matrices from [70], we follow a suggestion put forward at the last "Muon g-2 Theory Initiative" meeting [110]. We add, to results obtained with these correlations, an uncertainty obtained by summing, in quadrature, half the differences of the individual channels. This gives an additional error of $(2.26)_{\rm int}$, where "int" stands for the uncertainty related to the integration and correlation procedures.

Thus, using the well known dispersive integral (see eg. Section 5 of [111])

$$a_\mu = 10^{10} \left(\frac{\alpha m_\mu}{3\pi} \right)^2 \int_{s_{\rm th}}^{\infty} \frac{ds}{s^2} \, R(s)\hat{K}(s) \,, \tag{177}$$

the experimental R-ratio data set of [70] and the perturbative R-ratio from the rhad package [112], we obtain, $a_\mu = 693.27(2.46)_{\rm stat}(1.46)_{\pi\pi}(2.26)_{\rm int}[3.65]$, where the first error reproduces the one given in [70], while the second and third errors are computed above. The last error, in brackets, is the quadratically combined error of the first three. It is larger than the one of KNT and is closer to the error of DHMZ17, but still a bit smaller than that of DHMZ19. This enlarged error, as well as the most recent value of 4.0 from DHMZ, reduce a bit the strong tension between the measurement of $g_\mu - 2$ and the theory predictions based on the R-ratio method.

Having checked that we are able to reproduce well known R-ratio results, we repeat the whole procedure for $a_{\mu,\rm win}$ of Equation (71). For this observable, high statistical precision is easier to reach on the lattice and the continuum and infinite-volume extrapolations are less difficult. In addition, we expect that the R-ratio method yields a similar relative error for $a_{\mu,\rm win}$ as for a_μ. These facts make $a_{\mu,\rm win}$ a convenient observable to compare the two approaches and, eventually, to combine them for improved overall precision [48].

To determine $a_{\mu,\rm win}$ from R-ratio data, we transform the latter to Euclidean coordinate space by a Laplace transform [42], where a weighted integral with weight function $K(t)W(t)$ has to be performed, as described in Section 11. One ends up with an integral as in Equation (177), but the kernel $\hat{K}$ replaced by $\hat{K}_{\rm win}$:

$$\hat{K}_{\rm win}(s) = \frac{3s^{5/2}}{8m_\mu^2} \int_0^{\infty} dt \, e^{-\sqrt{s}t} K(t)W(t; t_1, t_2), \tag{178}$$

where $K(t)$ is given by Equation (3) of the main paper. The window parameters $t_1 = 0.4$ fm, $t_2 = 1.0$ fm and $\Delta = 0.15$ fm are the same as in Section 24. We proceed with the computation of the s-integral as in the case of a_μ. In particular, we include the $\pi\pi$ and int errors as follows:

a. Repeating the R-ratio method, with the $\hat{K}_{\rm win}$ kernel, gives the same relative difference between the KLOE and BaBar results, ie. 2.7%, as with the original kernel function $\hat{K}$. Carrying out the PDG procedure for adjusting errors, we obtain a value of 0.5 for the additional "$\pi\pi$" error.

b. Since we do not have the contributions of the individual experimental channels in our window for both the KNT and the DHMZ frameworks, we simply scale down the "int" error of the full a_μ. Thus, instead of 2.26, we obtain 0.8 as an "int" uncertainty.

Putting the above components together and also including the tiny perturbative contribution from [112], we obtain

$$a_{\mu,\rm win} = 229.7(0.9)_{\rm stat}(0.5)_{\pi\pi}(0.8)_{\rm int}[1.3] \,. \tag{179}$$

The last error, in brackets, is all errors added in quadrature. This value is compared with our lattice result in Section 24.

26 Consequences for electroweak precision observables?

In this section we investigate the claim, put forward by Crivellin, Hoferichter, Manzari and Montull (CHMM) [113], that the result of this paper may lead to a significant tension with electroweak precision fits for the hadronic contribution to the running of the electromagnetic coupling, $\Delta\alpha_{\rm had}^{(5)}(M_Z^2)$. Here "(5)" means that we consider five active flavors and M_Z is the mass of the Z-boson. CHMM's analysis is based on the earlier work of [114] that studies the impact, on precision electroweak fits, of increasing the HVP contribution to reproduce the measured value of $(g_\mu - 2)/2$. That work has been recently updated
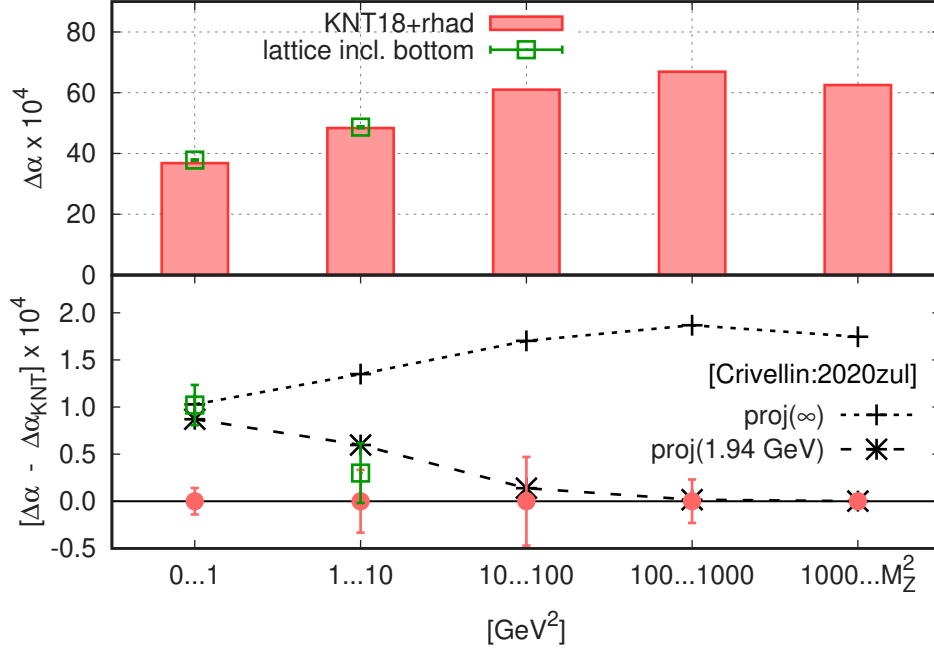
Figure 30: Hadronic contribution to the running of the electromagnetic coupling for Euclidean momenta. For each bin we show $\Delta\alpha^{(5)}_{\text{had}}(-q^2_{\text{max}}) - \Delta\alpha^{(5)}_{\text{had}}(-q^2_{\text{min}})$, where $q^2_{\text{min/max}}$ is the lower/upper end of the bin. The conventional value of $\Delta\alpha^{(5)}_{\text{had}}(M^2_Z)$ can be obtained by summing the results of all bins and performing a rotation to Minkowswki space (see text). The upper panel shows results obtained from the experimental R-ratio of KNT18 [70] and perturbation theory [112], as well as our lattice results. The bottom panel compares the results with the R-ratio as a baseline. The "reference point" scenario of CHMM [113] ("proj(∞)") is shown with crosses. Their "proj(1.94 GeV)" scenario is shown with bursts.

and expanded in [115]. In contrast, [116] shows that the shift in $\Delta\alpha^{(5)}_{\text{had}}(M^2_Z)$, induced by the sole knowledge of the difference between the lattice and the R-ratio prediction for a_μ, is at least nine times smaller than the error in the electroweak fit determination of $\Delta\alpha^{(5)}_{\text{had}}(M^2_Z)$, given in [113] and in Equation (180) below.

Indeed, the authors of [113] perform a global fit to electroweak observables and obtain

$$\Delta\alpha^{(5)}_{\text{had}}(M^2_Z) = 270.2(3.0) \times 10^{-4} \text{ (electroweak)} . \tag{180}$$

Note that this value is somewhat smaller, both in value and in uncertainty, than the latest result of the Gfitter group [117]. The same observable can also be obtained from the experimental R-ratio [69, 96], eg. using the KNT19 result [96] as do CHMM:

$$\Delta\alpha^{(5)}_{\text{had}}(M^2_Z) = 276.1(1.1) \times 10^{-4} \text{ (R-ratio)} , \tag{181}$$

which is 1.8σ higher than the electroweak-fit value.

CHMM then consider a variety of scenarios to estimate the possible value of $\Delta\alpha^{(5)}_{\text{had}}(M^2_Z)$ from the results given in the present paper. In their "reference point" scenario ("proj(∞)"), they assume that the relative difference between the R-ratio and our result for a_μ corresponds to an energy-independent rescaling in the $e^+e^- \to$ hadrons spectral function for all center-of-mass energies, from threshold to infinity. Thus, they obtain

$$\Delta\alpha^{(5)}_{\text{had}}(M^2_Z) = 276.1(1.1) \times 10^{-4} \times \frac{712}{693} = 283.8(1.3) \times 10^{-4} \text{ (CHMM)} , \tag{182}$$

which deviates by 4.2σ from the electroweak-fit value. This leads CHMM to conclude that the result of

	Π^{light}	Π^{strange}	Π^{charm}	Π^{disc}	$\infty - \text{ref}$	Π^{bottom}	$\Pi^{(5)}$
$\Pi(1)$	355.7(2.3)	41.8(0.1)	17.0(0.1)	$-5.1(0.6)$	2.8(0.2)	0.3(0.1)	412.6(2.4)
$\Pi(10) - \Pi(1)$	363.4(2.2)	67.7(0.3)	96.6(2.6)	$-0.2(0.2)$	0.2(0.0)	2.9(0.5)	530.7(3.5)

Table 20: Continuum extrapolated lattice results for the HVP. We give separately the light, strange, charm and disconnected contributions and also the finite-size effects as computed from the 4HEX simulations. The bottom quark contribution was obtained from the work [94]. The last column is the sum of all.

the present paper, while removing the discrepancy with the experimental determination of a_μ, may create a new one, now with electroweak precision measurements.

We now take a closer look at the claim of CHMM and point out problems with their assumption. We will not study the running of α all the way up to M_Z^2 here, because that would take us significantly beyond the scope of the present study of a_μ. However, without too much effort we can investigate the running of α in the Euclidean regime up to scales accessible in our lattice computation.[5] The Euclidean running of the coupling is obtained from the Euclidean HVP as

$$\Delta\alpha_{\text{had}}^{(5)}(-q^2) = e^2 \Pi^{(5)}(q^2) \;, \tag{183}$$

where $\Pi^{(5)}$ is the five-flavor HVP given, with the notations of Section 11, by:

$$\Pi^{(5)} = \Pi^{\text{light}} + \Pi^{\text{strange}} + \Pi^{\text{charm}} + \Pi^{\text{disc}} + \Pi^{\text{bottom}} \;. \tag{184}$$

Here we compute $\Pi^{(5)}$ both on the lattice and from the R-ratio.

To compute the HVP on the lattice we apply the same Type-II fit procedure as we use for the determination of a_μ and $a_{\mu,\text{win}}$ in the previous Sections. We correct for finite-size effects by computing them in our 4HEX simulations. For the bottom quark we combine the first four moments of Π^{bottom}, determined by the HPQCD collaboration [94], into a Padé approximation. For the purpose of the present discussion it suffices to consider two observables: $\Pi^{(5)}(1)$ the value of $\Pi^{(5)}$ at $q^2 = 1$ GeV^2 and the difference in $\Pi^{(5)}$ at $q^2 = 10$ GeV^2 and 1 GeV^2, denoted by $\Pi^{(5)}(10) - \Pi^{(5)}(1)$. Our final continuum extrapolated results are given in Table 20.

In the case of $\Pi^{(5)}(10) - \Pi^{(5)}(1)$ we have departed from our standard procedure in two details. First, one must note that the full dataset gives bad fit qualities in the continuum extrapolation, especially for the charm contribution. This is because the precision of our results is orders of magnitude better than eg. for a_μ. Thus, we fit only a subset of about 50 configurations, maximally spaced along the simulation chain. This increases the statistical error and leads to acceptable fit qualities. Also, we observe lattice artefacts that are much larger than for the other observables in the paper. In the charm case they are on the level of 100%. To reflect this in the uncertainty of the continuum extrapolation, we use fit functions that are quadratic polynomials in a^2 and also apply a flat weighting of the results of different procedures in our determination of the systematic error.

To compute the HVP from the R-ratio we apply a dispersion integral (see eg. Section 3 of [111])

$$\Pi^{(5)}(q^2) = \frac{q^2}{12\pi^2} \int_{s_{\text{th}}}^{\infty} ds \; \frac{R(s)}{s(s + q^2)} \tag{185}$$

to the R-ratio data set of [70]. Uncertainties are computed from the covariance matrix of the data. For energies above the range of this data set we use the perturbative result from the `rhad` package [112].

In Figure 30 we show HVP differences corresponding to five energy bins, starting at zero and ending at the scale M_Z^2. The first bin gives the difference of the HVP between 1 and 0 GeV^2, the second between 10 and 1 GeV^2 and so on. The running to the scale M_Z^2 can then be obtained by summing the values in the five bins. In the top panel we show the result obtained from the R-ratio. For the first two bins, we

---

[5]If we were able to run in the Euclidean up to M_Z^2, the conversion to timelike M_Z^2 can be computed in perturbation theory and the resulting correction is significantly smaller than the present error bars on $\Delta\alpha_{\text{had}}^{(5)}(M_Z^2)$, as shown eg. in [118].

also have lattice results. The bottom panel shows our lattice results with the R-ratio result as a baseline.

In the first bin, corresponding to an energy range from 0 to 1 GeV^2, we see a difference between the lattice and the R-ratio determinations, that is about 1.0 in $\Delta\alpha_{\mathrm{had}}^{(5)}$. This corresponds to a relative deviation of approximately 2.8%, which is about the same as we have in the total a_μ. This fact is not surprising because over 99% of a_μ comes from this spacelike region of momenta. In the second bin, corresponding to the energy range from 1 to 10 GeV^2, the lattice and R-ratio results already agree. For the bins with larger energies we show no lattice results: discretization errors are too large to allow a controlled continuum extrapolation.

In Figure 30 we also show two scenarios from CHMM. The first is their "reference point" projection ("proj(∞)") described above. There, the R-ratio and lattice results are assumed to have the same 2.8% relative discrepancy in all energy bins up to M_Z, as observed in the first bin. The difficulty with this assumption is that CHMM extrapolates over two orders of magnitude in energy, while using data only from the first bin. In addition, our lattice result in the second energy bin is already in clear disagreement with their hypothesis, and invalidates their estimate of Equation (182).

The second of their scenarios considered here is the one in which they assume that the 2.8% rescaling of the spectral function only applies to center-of-mass energies below 1.94 GeV ("proj(1.94 GeV)"). This second scenario agrees much better with our lattice results, as can be seen in the first two bins. If future lattice calculations confirm that the agreement holds in the remaining bins, the tension implied by our lattice calculation of a_μ on $\Delta\alpha_{\mathrm{had}}^{(5)}$ would be below 2.4σ, significantly smaller than the 4.2σ of the "reference point" scenario of [113] and only slightly larger than the 1.8σ already observed with the R-ratio result of Equation (181).

References

1. Ishizuka, N., Fukugita, M., Mino, H., Okawa, M. & Ukawa, A. Operator dependence of hadron masses for Kogut-Susskind quarks on the lattice. *Nucl. Phys.* **B411,** 875–902 (1994).

2. Luscher, M. & Weisz, P. On-Shell Improved Lattice Gauge Theories. *Commun. Math. Phys.* **97.** [Erratum: Commun. Math. Phys.98,433(1985)], 59 (1985).

3. Morningstar, C. & Peardon, M. J. Analytic smearing of SU(3) link variables in lattice QCD. *Phys. Rev.* **D69,** 054501 (2004).

4. Davies, C. T. H. *et al.* Precise Charm to Strange Mass Ratio and Light Quark Masses from Full Lattice QCD. *Phys. Rev. Lett.* **104,** 132003 (2010).

5. Aoki, S. *et al.* FLAG Review 2019. arXiv: 1902.08191 [hep-lat] (2019).

6. Borsanyi, S. *et al.* High-precision scale setting in lattice QCD. *JHEP* **09,** 010 (2012).

7. Luscher, M. Properties and uses of the Wilson flow in lattice QCD. *JHEP* **08,** 071 (2010).

8. Lepage, G. P. Flavor symmetry restoration and Symanzik improvement for staggered quarks. *Phys. Rev.* **D59,** 074502 (1999).

9. Aoki, S. *et al.* Review of lattice results concerning low-energy particle physics. *Eur. Phys. J.* **C77,** 112 (2017).

10. Capitani, S., Durr, S. & Hoelbling, C. Rationale for UV-filtered clover fermions. *JHEP* **11,** 028 (2006).

11. Takaishi, T. Heavy quark potential and effective actions on blocked configurations. *Phys. Rev.* **D54,** 1050–1053 (1996).

12. DeGrand, T. A., Hasenfratz, A. & Kovacs, T. G. Improving the chiral properties of lattice fermions. *Phys. Rev.* **D67,** 054501 (2003).

13. Neuberger, H. Exactly massless quarks on the lattice. *Phys. Lett.* **B417,** 141–144 (1998).

14. Durr, S. Logarithmic link smearing for full QCD. *Comput. Phys. Commun.* **180,** 1338–1357 (2009).

15. Borsanyi, S. *et al.* Calculation of the axion mass based on high-temperature lattice quantum chromodynamics. *Nature* **539,** 69–71 (2016).

16. Borsanyi, S. *et al.* QCD thermodynamics with dynamical overlap fermions. *Phys. Lett.* **B713,** 342–346 (2012).

17. Sharpe, S. R. Quenched chiral logarithms. *Phys. Rev.* **D46,** 3146–3168 (1992).

18. Golterman, M. F. L. & Smit, J. Lattice Baryons With Staggered Fermions. *Nucl. Phys.* **B255,** 328–340 (1985).

19. Bailey, J. A. Staggered baryon operators with flavor SU(3) quantum numbers. *Phys. Rev.* **D75,** 114505 (2007).

20. Gusken, S. *et al.* Nonsinglet Axial Vector Couplings of the Baryon Octet in Lattice QCD. *Phys. Lett.* **B227,** 266–269 (1989).

21. Bazavov, A. *et al.* Nonperturbative QCD Simulations with 2+1 Flavors of Improved Staggered Quarks. *Rev. Mod. Phys.* **82,** 1349–1417 (2010).

22. Yelton, J. *et al.* Observation of an Excited Ω^- Baryon. *Phys. Rev. Lett.* **121,** 052003 (2018).

23. Capstick, S. & Isgur, N. Baryons in a Relativized Quark Model with Chromodynamics. *Phys. Rev.* **D34.** [AIP Conf. Proc.132,267(1985)], 2809 (1986).

24. Bazavov, A. *et al.* Additional Strange Hadrons from QCD Thermodynamics and Strangeness Freeze-out in Heavy Ion Collisions. *Phys. Rev. Lett.* **113,** 072001 (2014).

25. Alba, P. *et al.* Constraining the hadronic spectrum through QCD thermodynamics on the lattice. *Phys. Rev.* **D96,** 034517 (2017).

26. Aubin, C. & Orginos, K. A new approach for Delta form factors. *AIP Conf. Proc.* **1374,** 621–624 (2011).

27. DeTar, C. & Lee, S.-H. Variational method with staggered fermions. *Phys. Rev.* **D91,** 034504 (2015).

28. Colangelo, G., Durr, S. & Haefeli, C. Finite volume effects for meson masses and decay constants. *Nucl. Phys.* **B721,** 136–174 (2005).

29. Ishikawa, K. I. *et al.* SU(2) and SU(3) chiral perturbation theory analyses on baryon masses in 2+1 flavor lattice QCD. *Phys. Rev.* **D80,** 054502 (2009).

30. Hayakawa, M. & Uno, S. QED in finite volume and finite size scaling effect on electromagnetic properties of hadrons. *Prog. Theor. Phys.* **120,** 413–441 (2008).

31. De Divitiis, G. M. *et al.* Leading isospin breaking effects on the lattice. *Phys. Rev.* **D87,** 114505 (2013).

32. Borsanyi, S. *et al.* Ab initio calculation of the neutron-proton mass difference. *Science* **347,** 1452–1455 (2015).

33. Borsanyi, S. *et al.* Isospin splittings in the light baryon octet from lattice QCD and QED. *Phys. Rev. Lett.* **111,** 252001 (2013).

34. Bijnens, J. & Danielsson, N. Electromagnetic Corrections in Partially Quenched Chiral Perturbation Theory. *Phys. Rev.* **D75,** 014505 (2007).

35. Horsley, R. *et al.* QED effects in the pseudoscalar meson sector. *JHEP* **04,** 093 (2016).

36. Gasser, J., Rusetsky, A. & Scimemi, I. Electromagnetic corrections in hadronic processes. *Eur. Phys. J.* **C32,** 97–114 (2003).

37. Bali, G. S., Collins, S. & Schafer, A. Effective noise reduction techniques for disconnected loops in Lattice QCD. *Comput. Phys. Commun.* **181,** 1570–1583 (2010).

38. Blum, T., Izubuchi, T. & Shintani, E. New class of variance-reduction techniques using lattice symmetries. *Phys. Rev.* **D88,** 094503 (2013).

39. Blum, T., Doi, T., Hayakawa, M., Izubuchi, T. & Yamada, N. Determination of light quark masses from the electromagnetic splitting of pseudoscalar meson masses computed with two flavors of domain wall fermions. *Phys. Rev.* **D76,** 114508 (2007).

40. Fleming, G. T., Cohen, S. D., Lin, H.-W. & Pereyra, V. Excited-State Effective Masses in Lattice QCD. *Phys. Rev.* **D80,** 074506 (2009).

41. Davoudi, Z. & Savage, M. J. Finite-Volume Electromagnetic Corrections to the Masses of Mesons, Baryons and Nuclei. *Phys. Rev.* **D90,** 054503 (2014).

42. Bernecker, D. & Meyer, H. B. Vector Correlators in Lattice QCD: Methods and applications. *Eur. Phys. J.* **A47,** 148 (2011).

43. Chakraborty, B. *et al.* Strange and charm quark contributions to the anomalous magnetic moment of the muon. *Phys. Rev. D* **89,** 114501 (2014).

44. De Rafael, E. Moment Analysis of Hadronic Vacuum Polarization - Proposal for a lattice QCD evaluation of $g_\mu - 2$. *Phys. Lett. B* **736,** 522–525 (2014).

45. Charles, J., de Rafael, E. & Greynat, D. Mellin-Barnes approach to hadronic vacuum polarization and $g_\mu - 2$. *Phys. Rev. D* **97,** 076014 (2018).

46. Blum, T. Lattice calculation of the lowest order hadronic contribution to the muon anomalous magnetic moment. *Phys. Rev. Lett.* **91,** 052001 (2003).

47. Borsanyi, S. *et al.* Hadronic vacuum polarization contribution to the anomalous magnetic moments of leptons from first principles. *Phys. Rev. Lett.* **121,** 022002 (2018).

48. Blum, T. *et al.* Calculation of the hadronic vacuum polarization contribution to the muon anomalous magnetic moment. *Phys. Rev. Lett.* **121,** 022003 (2018).

49. Neff, H., Eicker, N., Lippert, T., Negele, J. W. & Schilling, K. On the low fermionic eigenmode dominance in QCD on the lattice. *Phys. Rev.* **D64,** 114509 (2001).

50. Li, A. *et al.* Overlap Valence on 2+1 Flavor Domain Wall Fermion Configurations with Deflation and Low-mode Substitution. *Phys. Rev.* **D82,** 114501 (2010).

51. Aubin, C. *et al.* Light quark vacuum polarization at the physical point and contribution to the muon $g - 2$. *Phys. Rev. D* **101,** 014503 (2020).

52. Hernandez, V., Roman, J. E. & Vidal, V. SLEPc: A Scalable and Flexible Toolkit for the Solution of Eigenvalue Problems. *ACM Trans. Math. Software* **31,** 351–362 (2005).

53. Lehner, C. *RBRC Workshop on Lattice Gauge Theories (2016).*

54. Borsanyi, S. *et al.* Slope and curvature of the hadronic vacuum polarization at vanishing virtuality from lattice QCD. *Phys. Rev.* **D96,** 074507 (2017).

55. Giusti, D., Lubicz, V., Martinelli, G., Sanfilippo, F. & Simula, S. Electromagnetic and strong isospin-breaking corrections to the muon $g - 2$ from Lattice QCD+QED. *Phys. Rev.* **D99,** 114502 (2019).

56. Bailey, J. A., Kim, H.-J. & Lee, W. Taste non-Goldstone, flavor-charged pseudo-Goldstone boson masses in staggered chiral perturbation theory. *Phys. Rev.* **D85,** 094503 (2012).

57. Golowich, E. & Kambor, J. Two loop analysis of vector current propagators in chiral perturbation theory. *Nucl. Phys.* **B447,** 373–404 (1995).

58. Amoros, G., Bijnens, J. & Talavera, P. Two point functions at two loops in three flavor chiral perturbation theory. *Nucl. Phys.* **B568,** 319–363 (2000).

59. Bijnens, J. & Relefors, J. Vector two-point functions in finite volume using partially quenched chiral perturbation theory at two loops. *JHEP* **12,** 114 (2017).

60. Aubin, C. & Blum, T. Calculating the hadronic vacuum polarization and leading hadronic contribution to the muon anomalous magnetic moment with improved staggered quarks. *Phys. Rev.* **D75,** 114502 (2007).

61. Sharpe, S. R. & Van de Water, R. S. Staggered chiral perturbation theory at next-to-leading order. *Phys. Rev.* **D71,** 114505 (2005).

62. Gasser, J. & Leutwyler, H. Chiral Perturbation Theory: Expansions in the Mass of the Strange Quark. *Nucl. Phys.* **B250,** 465–516 (1985).

63. Bijnens, J., Colangelo, G. & Ecker, G. The Mesonic chiral Lagrangian of order p^6. *JHEP* **02,** 020 (1999).

64. Lee, W.-J. & Sharpe, S. R. Partial flavor symmetry restoration for chiral staggered fermions. *Phys. Rev.* **D60,** 114503 (1999).

65. Aubin, C. & Bernard, C. Pion and kaon masses in staggered chiral perturbation theory. *Phys. Rev.* **D68,** 034014 (2003).

66. Scherer, S. Introduction to chiral perturbation theory. *Adv. Nucl. Phys.* **27,** 277 (2003).

67. Follana, E. *et al.* Highly improved staggered quarks on the lattice, with applications to charm physics. *Phys. Rev.* **D75,** 054502 (2007).

68. Cata, O. & Mateu, V. Chiral perturbation theory with tensor sources. *JHEP* **09,** 078 (2007).

69. Davier, M., Hoecker, A., Malaescu, B. & Zhang, Z. A new evaluation of the hadronic vacuum polarisation contributions to the muon anomalous magnetic moment and to $\boldsymbol{\alpha(m_Z^2)}$. *Eur. Phys. J. C* **80,** 241 (2020).

70. Keshavarzi, A., Nomura, D. & Teubner, T. Muon $g - 2$ and $\alpha(M_Z^2)$: a new data-based analysis. *Phys. Rev.* **D97,** 114025 (2018).

71. Jegerlehner, F. & Nyffeler, A. The Muon g-2. *Phys. Rept.* **477,** 1–110 (2009).

72. Gounaris, G. J. & Sakurai, J. J. Finite width corrections to the vector meson dominance prediction for rho to e+e-. *Phys. Rev. Lett.* **21,** 244–247 (1968).

73. Francis, A., Jaeger, B., Meyer, H. B. & Wittig, H. A new representation of the Adler function for lattice QCD. *Phys. Rev.* **D88,** 054502 (2013).

74. Tanabashi, M. *et al.* Review of Particle Physics. *Phys. Rev.* **D98,** 030001 (2018).

75. Luscher, M. Signatures of unstable particles in finite volume. *Nucl. Phys.* **B364,** 237–251 (1991).

76. Luscher, M. Two particle states on a torus and their relation to the scattering matrix. *Nucl. Phys.* **B354,** 531–578 (1991).

77. Lellouch, L. *Flavor physics and lattice quantum chromodynamics* in *Modern perspectives in lattice QCD: Quantum field theory and high performance computing. Proceedings, International School, 93rd Session, Les Houches, France, August 3-28, 2009* (2011), 629–698. arXiv: 1104.5484 [hep-lat].

78. Lellouch, L. & Luscher, M. Weak transition matrix elements from finite volume correlation functions. *Commun. Math. Phys.* **219,** 31–44 (2001).

79. Lin, C. J. D., Martinelli, G., Sachrajda, C. T. & Testa, M. K to pi pi decays in a finite volume. *Nucl. Phys.* **B619,** 467–498 (2001).

80. Meyer, H. B. Lattice QCD and the Timelike Pion Form Factor. *Phys. Rev. Lett.* **107,** 072002 (2011).

81. Metivet, T. Lattice study of $\pi\pi$ scattering using $N_f = 2 + 1$ Wilson improved quarks with masses down to their physical values. *PoS* **LATTICE2014,** 079 (2015).

82. Hansen, M. T. & Patella, A. Finite-volume effects in $(g-2)_\mu^{\text{HVP,LO}}$. arXiv: 1904.10010 [hep-lat] (2019).

83. Hansen, M. T. & Patella, A. Finite-volume and thermal effects in the leading-HVP contribution to muonic $(g-2)$. arXiv: 2004.03935 [hep-lat] (Apr. 2020).

84. Giusti, D., Sanfilippo, F. & Simula, S. Light-quark contribution to the leading hadronic vacuum polarization term of the muon $g-2$ from twisted-mass fermions. *Phys. Rev.* **D98,** 114504 (2018).

85. Shintani, E. & Kuramashi, Y. Study of systematic uncertainties in hadronic vacuum polarization contribution to muon $g-2$ with 2+1 flavor lattice QCD. arXiv: 1902.00885 [hep-lat] (2019).

86. Bijnens, J. *et al.* Electromagnetic finite-size effects to the hadronic vacuum polarization. *Phys. Rev. D* **100,** 014508 (2019).

87. Montvay, I. & Munster, G. *Quantum fields on a lattice* ISBN: 9780521599177, 9780511879197 (Cambridge University Press, 1997).

88. Dowdall, R., Davies, C., Lepage, G. & McNeile, C. Vus from pi and K decay constants in full lattice QCD with physical u, d, s and c quarks. *Phys. Rev. D* **88,** 074504 (2013).

89. Bazavov, A. *et al.* Gradient flow and scale setting on MILC HISQ ensembles. *Phys. Rev. D* **93,** 094510 (2016).

90. Carrasco, N. *et al.* QED Corrections to Hadronic Processes in Lattice QCD. *Phys. Rev.* **D91,** 074506 (2015).

91. Giusti, D. *et al.* First lattice calculation of the QED corrections to leptonic decay rates. *Phys. Rev. Lett.* **120,** 072001 (2018).

92. Di Carlo, M. *et al.* Light-meson leptonic decay rates in lattice QCD+QED. *Phys. Rev.* **D100,** 034514 (2019).

93. Cheng, A., Hasenfratz, A., Liu, Y., Petropoulos, G. & Schaich, D. Improving the continuum limit of gradient flow step scaling. *JHEP* **05,** 137 (2014).

94. Colquhoun, B., Dowdall, R. J., Davies, C. T. H., Hornbostel, K. & Lepage, G. P. Υ and Υ' Leptonic Widths, a_μ^b and m_b from full lattice QCD. *Phys. Rev.* **D91,** 074514 (2015).

95. Chakraborty, B., Davies, C. T. H., Koponen, J., Lepage, G. P. & Van de Water, R. S. Higher-Order Hadronic-Vacuum-Polarization Contribution to the Muon g-2 from Lattice QCD. *Phys. Rev.* **D98,** 094503 (2018).

96. Keshavarzi, A., Nomura, D. & Teubner, T. $g-2$ of charged leptons, $\alpha(M_Z^2)$, and the hyperfine splitting of muonium. *Phys. Rev.* **D101,** 014029 (2020).

97. Davier, M., Hoecker, A., Malaescu, B. & Zhang, Z. Reevaluation of the hadronic vacuum polarisation contributions to the Standard Model predictions of the muon $g-2$ and $\alpha(M_Z^2)$ using newest hadronic cross-section data. *Eur. Phys. J.* **C77,** 827 (2017).

98. Colangelo, G., Hoferichter, M. & Stoffer, P. Two-pion contribution to hadronic vacuum polarization. *JHEP* **02,** 006 (2019).

99. Hoferichter, M., Hoid, B.-L. & Kubis, B. Three-pion contribution to hadronic vacuum polarization. *JHEP* **08,** 137 (2019).

100. Anastasi, A. *et al.* Combination of KLOE $\sigma\left(e^+e^- \to \pi^+\pi^-\gamma(\gamma)\right)$ measurements and determination of $a_\mu^{\pi^+\pi^-}$ in the energy range $0.10 < s < 0.95$ GeV^2. *JHEP* **03,** 173 (2018).

101. Alemany, R., Davier, M. & Hocker, A. Improved determination of the hadronic contribution to the muon (g-2) and to alpha (M(z)) using new data from hadronic tau decays. *Eur. Phys. J. C* **2,** 123–135 (1998).

102. Davier, M. *et al.* The Discrepancy Between tau and e+e- Spectral Functions Revisited and the Consequences for the Muon Magnetic Anomaly. *Eur. Phys. J. C* **66,** 127–136 (2010).

103. Davier, M., Hoecker, A., Malaescu, B. & Zhang, Z. Reevaluation of the Hadronic Contributions to the Muon g-2 and to alpha(MZ). *Eur. Phys. J. C* **71.** [Erratum: Eur.Phys.J.C 72, 1874 (2012)], 1515 (2011).

104. Davier, M., Höcker, A., Malaescu, B., Yuan, C.-Z. & Zhang, Z. Update of the ALEPH non-strange spectral functions from hadronic τ decays. *Eur. Phys. J. C* **74,** 2803 (2014).

105. Cirigliano, V., Ecker, G. & Neufeld, H. Isospin violation and the magnetic moment of the muon. *Phys. Lett. B* **513,** 361–370 (2001).

106. Cirigliano, V., Ecker, G. & Neufeld, H. Radiative tau decay and the magnetic moment of the muon. *JHEP* **08,** 002 (2002).

107. Jegerlehner, F. & Szafron, R. $\rho^0 - \gamma$ mixing in the neutral channel pion form factor F_π^e and its role in comparing e^+e^- with τ spectral functions. *Eur. Phys. J. C* **71,** 1632 (2011).

108. Bruno, M., Izubuchi, T., Lehner, C. & Meyer, A. On isospin breaking in τ decays for $(g-2)_\mu$ from Lattice QCD. *PoS* **LATTICE2018,** 135 (2018).

109. Miranda, J. & Roig, P. New τ-based evaluation of the hadronic contribution to the vacuum polarization piece of the muon anomalous magnetic moment. arXiv: 2007.11019 [hep-ph] (July 2020).

110. https://indico.fnal.gov/event/21626/session/2/contribution/68/material/slides/2.pdf.

111. Jegerlehner, F. The Anomalous Magnetic Moment of the Muon. *Springer Tracts Mod. Phys.* **274,** pp.1–693 (2017).

112. Harlander, R. V. & Steinhauser, M. rhad: A Program for the evaluation of the hadronic R ratio in the perturbative regime of QCD. *Comput. Phys. Commun.* **153,** 244–274 (2003).

113. Crivellin, A., Hoferichter, M., Manzari, C. A. & Montull, M. Hadronic vacuum polarization: $(g-2)_\mu$ versus global electroweak fits. arXiv: 2003.04886 [hep-ph] (Mar. 2020).

114. Passera, M., Marciano, W. & Sirlin, A. The Muon g-2 and the bounds on the Higgs boson mass. *Phys. Rev. D* **78,** 013009 (2008).

115. Keshavarzi, A., Marciano, W. J., Passera, M. & Sirlin, A. The muon g-2 and $\Delta\alpha$ connection. arXiv: 2006.12666 [hep-ph] (June 2020).

116. De Rafael, E. On Constraints Between $\Delta\alpha_{\mathrm{had}}(M_Z^2)$ and $(g_\mu-2)_{\mathrm{HVP}}$. arXiv: 2006.13880 [hep-ph] (June 2020).

117. Haller, J. *et al.* Update of the global electroweak fit and constraints on two-Higgs-doublet models. *Eur. Phys. J. C* **78,** 675 (2018).

118. *Theory report on the 11th FCC-ee workshop* (2019). arXiv: 1905.05078 [hep-ph].