Ab initio calculation of the neutron-proton mass difference

Ab initio calculation of the neutron-proton mass difference

Sz. Borsanyi[1], S. Durr[1,2], Z. Fodor[1,2,3], C. Hoelbling[1], S. D. Katz[3,4], S. Krieg[1,2], L. Lellouch[5], T. Lippert[1,2], A. Portelli[5,6], K. K. Szabo[1,2], B. C. Toth[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] Lendület Lattice Gauge Theory Research Group, Magyar Tudományos Akadémia–Eötvos Loránd University, H-1117 Budapest, Hungary
[5] CNRS, Aix-Marseille Université, Université de Toulon, CPT UMR 7332, F-13288, Marseille, France
[6] School of Physics and Astronomy, University of Southampton, SO17 1BJ, UK

The existence and stability of atoms rely on the fact that neutrons are more massive than protons. The measured mass difference is only 0.14% of the average of the two masses. A slightly smaller or larger value would have led to a dramatically different universe. Here, we show that this difference results from the competition between electromagnetic and mass isospin breaking effects. We performed lattice quantum-chromodynamics and quantum-electrodynamics computations with four nondegenerate Wilson fermion flavors and computed the neutron-proton mass-splitting with an accuracy of 300 kilo-electron volts, which is greater than 0 by 5 standard deviations. We also determine the splittings in the Σ, Ξ, D and Ξ_{cc} isospin multiplets, exceeding in some cases the precision of experimental measurements.

The mass of the visible universe is a consequence of the strong interaction (*1*), which is the force that binds together quarks into protons and neutrons. To establish this with percent-level accuracy, very precise calculations based on the lattice formulation of quantum chromodynamics(QCD), the theory of the strong interaction, were needed. Going beyond such calculations to control much finer effects that are at the per mil (‰) level is necessary to, for instance, account for the relative neutron-proton mass difference which was experimentally measured to be close to 0.14% (*2*). Precisely, this difference is needed to explain the physical world as we know it today (*3*). For example, a relative neutron-proton mass difference smaller than about one third of the observed 0.14% would cause hydrogen atoms to undergo inverse beta decay, leaving predominantly neutrons. A value somewhat larger than 0.05% would have resulted in the Big Bang Nucleosynthesis (BBN), producing much more helium-4 and far less hydrogen than it did in our universe. As a result, stars would not have ignited in the way they did. On the other hand, a value considerably larger than 0.14% would have resulted in a much faster beta decay for neutrons. This would have led to far fewer neutrons at the end of the BBN epoch and would have made the burning of hydrogen in stars and the synthesis of heavy elements more difficult. We show here that this tiny mass splitting is the result of a subtle cancellation between electromagnetic and quark mass difference effects.

The Standard Model of Particle Physics is a $SU(3) \times SU(2) \times U(1)$ gauge theory with massless fermions. During the expansion of the early universe, the Higgs mechanism broke this symmetry down to $SU(3) \times U(1)$ and elementary particles acquired masses proportional to their couplings to the Higgs field. As the universe continued to expand, a QCD transition took place, confining quarks and gluons into hadrons and giving those particles most of their mass. This same theory today is believed to be responsible for the tiny isospin splittings which are the subject of this paper. At the level of precision that we aim for here, the effects of the weak interaction, of leptons, and of the two heaviest quarks can either be neglected or absorbed into the remaining parameters of the theory. The resulting theory is one of u, d, s and c (up, down, strange and charm) quarks, gluons, photons and their interactions. 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 four flavors of quarks, the m_f are their masses and the q_f are their charges in units of the electron charge e. Moreover, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + [B_\mu, B_\nu]$ and g is the QCD coupling constant. In electrodynamics, the gauge potential A_μ is the real valued photon field, whereas in QCD, B_μ is a Hermitian 3 by 3 matrix field. The ψ_f are Dirac-spinor fields representing the quarks and carry a "color" index, which runs from 1 to 3. In the present work, we consider all of the degrees of freedom of this Lagrangian; that is, we include quantum electrodynamics (QED) and QCD, as well as the four nondegenerate quark flavors, in a fully dynamical formulation.

The action S of QCD+QED is defined as the spacetime integral of $\mathcal{L}$. Particle propagators are averages of products of fields over all possible field configurations, weighted by the Boltzmann factor $\exp(-S)$. A notable feature of QCD is asymptotic freedom, which means that the interaction becomes weaker and weaker as the relative momentum of the interacting particles increases (*4*, *5*). Thus, at high energies the coupling constant

is small, and a perturbative treatment is possible. However, at energies typical of quarks and gluons within hadrons, the coupling is large, and the interactions become highly nonlinear. The most systematic way to obtain predictions in this nonperturbative regime of QCD involves introducing a hypercubic spacetime lattice with lattice spacing a (*6*) on which the above Lagrangian is discretized, numerically evaluating the resulting propagators and extrapolating the results to the continuum ($a \to 0$). The discretization procedure puts fermionic variables on the lattice sites, whereas gauge fields are represented by unitary 3 by 3 matrices residing on the links between neighboring sites. The discretized theory can be viewed as a four-dimensional statistical physics system.

Calculating the mass differences between the neutral and charged hadron partners by using lattice techniques has involved different levels of approximation. In the pioneering work of (*7*), the quenched approximation was used both for QCD and QED. Recent studies (*8–10*) have typically performed dynamical QCD computations with quenched QED fields. Another quenched QED approach, in which the path integral is expanded to $O(\alpha)$, has also recently been implemented (*11*). In all such calculations, the neglected terms are of the same leading order in α as the isospin splittings of interest (*10*). To have a calculation that fully includes QED effects to $O(\alpha)$ requires including electromagnetic effects in the quark sea. Three exploratory studies have attempted to include these effects. The first two used reweighting techniques in $N_f = 2 + 1$ QCD simulations (*12*, *13*). Beyond the difficulty of estimating the systematic error associated with reweighting, the computation in (*12*) was carried out with a single lattice spacing in a relatively small $(3\,\mathrm{fm})^3$ spatial volume and the one in (*13*) on a single, much coarser and smaller lattice, with pion masses larger than their physical value. In the third study (*14*), real dynamical QCD and QED simulations were performed, albeit on a single lattice at unphysical quark mass values.

Here, we provide a fully controlled ab initio calculation for these isospin splittings. We used 1+1+1+1 flavor QCD+QED with 3HEX (QCD) and 1 APE (QED) smeared clover improved Wilson quarks. Up to now, the most advanced simulations have included up, down, and strange quarks in the sea but neglected all electromagnetic and up-down mass difference effects. Such calculations have irreducible systematic uncertainties of $O(1/N_c/m_c^2, \alpha, m_d - m_u)$, where $N_c = 3$ is the number of colors in QCD. This limits their accuracy to the percent level. We reduced these uncertainties to $O(1/N_c/m_b^2, \alpha^2)$, where m_b is the bottom quark mass, yielding a complete description of the interactions of quarks at low energy, accurate to the per mil level.

In our parameter set, we have four lattice spacings ranging from $0.06\,\mathrm{fm}$ to $0.10\,\mathrm{fm}$. We observed very small cutoff effects in our results, which is in good agreement with our earlier spectrum determination (*15*, *16*). Nevertheless, these small cutoff effects are accounted for in our systematic error analysis as $g^2 a$ or a^2 corrections in the histogram method described in (*17*).

We performed computations with four values of the bare fine structure constant: 0, a value close to the physical value of $1/137$, and two larger values, approximately $1/10$ and $1/6$. Most of our runs were carried out at $\alpha = 0$ and $\sim 1/10$. Because QED effects in typical hadron masses are small (around or below the $1‰$ level), statistical noise in the splittings can be reduced by interpolating between results obtained with the larger value of α and those obtained with $\alpha = 0$. We then confirmed and improved this interpolation with simulations near the physical value of the coupling. The actual interpolation to the physical point is performed in terms of a renormalized fine structure constant defined via the QED Wilson flow (*18*). Within the precision reached in our work, the splittings studied show no deviation from linear behavior in the range of couplings studied. Our largest value of the fine structure constant was chosen so as to increase the signal for the mass splittings, while keeping under control large finite-volume corrections of the kind discussed below.

Our smallest pion mass is about $195\ \mathrm{MeV}$ (with more than 20,000 trajectories), and our largest lattice has a spatial extent of 8 fm. These parameters were carefully chosen to allow for a determination of the neutron-proton mass splitting that is ~ 5 standard deviations (SDs) from 0, with currently available computing resources. This is a challenge because accounting for isospin breaking effects increases the cost (*17*) of the calculation compared with computations with two degenerate light flavors used typically in recent works (*19–27*).

We produced gauge configurations with an improved version of the Hybrid Monte Carlo algorithm and

checked, a posteriori, that the probability weights are always positive in the region of the parameter space used in our simulations.

We used two previously suggested frameworks for the photon fields. These correspond to a nonlocal modification of the action that vanishes in the infinite-volume limit. As we argue in (*17*), these nonlocalities do not generate new ultraviolet divergences at one-loop order in α. The final analysis is performed in the framework of (*28*), which respects reflection positivity and has a well-defined, large-time limit, unlike previously used techniques (*17*). Generically, the photon fields show very large autocorrelation times of several thousand trajectories. We designed a Fourier accelerated algorithm within this QCD+QED framework that dramatically reduces these large autocorrelation times.

The long-range nature of the electromagnetic interaction poses one of the most serious difficulties of the present work. It induces finite-volume corrections that only fall off like inverse powers of the linear extent of the system. These are far more severe than the QCD finite-volume corrections, which are exponentially suppressed in these dimensions. Exponential corrections can easily be included in large scale spectrum studies ((*16*)). We performed an extensive study of the much larger power-suppressed finite-volume corrections using both one-loop analytical QED calculations and high-precision QED simulations (*17*). The size and volume behavior of these corrections in our full QCD+QED calculation are illustrated in Fig. 1.

Statistical errors on the mass splittings are calculated by using 2000 bootstrap samples. The systematic uncertainties on the final results are determined with our histogram method (*16*). We considered a wide range of analyses, each of which provides a valid approach to obtain the physical splittings from our simulation results, and calculated the associated goodness of fit. Because these procedures have different numbers of free parameters, we combined them using the Akaike information criterion (AIC) (*29*) and obtained a distribution for each splitting. The means of these distributions are our central values, whereas the widths of the distributions provide estimates of systematic uncertainties. This procedure yields conservative errors.

Our final results for the mass splittings are shown in Fig. 2. A comparison with the results of (*10*) indicates that the precision of the signal for ΔM_N (thus the splitting being non-zero) increased from $\sim 1\sigma$ to 5σ. For the other channels, the improvement is even more pronounced. In addition, the present work represents a fully-controlled approach, whereas (*10*) was based on the electroquenched approximation with degenerate light quarks in the sea. The nucleon, Σ, and Ξ splittings are consistent with the Coleman-Glashow relation $\Delta_{\mathrm{CG}} \equiv \Delta M_N - \Delta M_\Sigma + \Delta M_\Xi = 0$ (*30*). According to our calculation, this relation is fulfilled with an accuracy of $130\,\mathrm{keV}$. We also computed the individual contributions to the splittings coming from mass isospin breaking effects ($\alpha = 0$, $m_d - m_u \neq 0$) and electromagnetic effects ($m_d - m_u = 0$, $\alpha \neq 0$), as defined in (*17*). The numerical results for all of our results are given in Table 1. Because the precision of the experimental result for the nucleon is far greater than ours, we additionally give the QED and QCD separation obtained using the experimental value of $M_n - M_p$: $(M_n - M_p)_{\mathrm{QCD}}/(M_n - M_p)_{\mathrm{QED}} = -2.49(23)(29)$. Last, we used this number in Fig. 3 to plot the result of the neutron-proton mass splitting as a function of quark-mass difference and electromagnetic coupling. In combination with astrophysical and cosmological arguments, this figure can be used to determine how different values of these parameters would change the content of the universe. This in turn provides an indication of the extent to which these constants of nature must be fine-tuned to yield a universe that resembles ours.

Acknowledgments:

This project was supported by the Deutsche Forschungsgemeinschaft grant SFB/TR55, the Partnership for Advanced Computing in Europe (PRACE) initiative, the Gauss Centre for Supercomputing e.V, the European Research Council grant (FP7/2007-2013/ERC No 208740), the Lendület program of the Hungarian Academy of Sciences (LP2012-44/2012), "Origines, Constituants et Évolution de l'Univers" (OCEVU) Labex (ANR-11-LABX-0060), the A*MIDEX project (ANR-11-IDEX-0001-0) funded by the "Investissements d'Avenir" French government program and managed by the Agence Nationale de la Recherche (ANR), and the Grand Équipement National de Calcul Intensif–Institut du Développement et des Ressources en Informatique Scientifique (IDRIS) Grand Challenge grant 2012 "StabMat" as well as grant No. 52275. The computations were performed on JUQUEEN and JUROPA at Forschungszentrum Jülich (FZJ), on Turing at IDRIS in Orsay, on SuperMUC at Leibniz Supercomputing Centre in München, on Hermit at the High Performance Computing Center in Stuttgart and on local machines in Wuppertal and Budapest. The data described in the paper are 60 TB and archived in FZJ.

References:

(1) A. Kronfeld, *in 100 Years of Subatomic Physics, ed. by E. Henley and S. Ellis, Chap. 18, World Scientific* (2013).

(2) J. Beringer, *et al.*, *Phys.Rev.* **D86**, 010001 (2012).

(3) R. L. Jaffe, A. Jenkins, I. Kimchi, *Phys.Rev.* **D79**, 065014 (2009).

(4) D. J. Gross, F. Wilczek, *Phys.Rev.Lett.* **30**, 1343 (1973).

(5) H. D. Politzer, *Phys.Rev.Lett.* **30**, 1346 (1973).

(6) K. G. Wilson, *Phys.Rev.* **D10**, 2445 (1974).

(7) A. Duncan, E. Eichten, H. Thacker, *Phys.Rev.Lett.* **76**, 3894 (1996).

(8) T. Blum, *et al.*, *Phys.Rev.* **D82**, 094508 (2010).

(9) S. Basak, *et al.*, *PoS* **CD12**, 030 (2013).

(10) S. Borsanyi, *et al.*, *Phys.Rev.Lett.* **111**, 252001 (2013).

(11) G. de Divitiis, *et al.*, *Phys.Rev.* **D87**, 114505 (2013).

(12) S. Aoki, *et al.*, *Phys.Rev.* **D86**, 034507 (2012).

(13) T. Ishikawa, *et al.*, *Phys.Rev.Lett.* **109**, 072002 (2012).

(14) R. Horsley, *et al.*, *PoS* **Lattice2013**, 499 (2013).

(15) S. Durr, *et al.*, *Phys.Rev.* **D79**, 014501 (2009).

(16) S. Durr, *et al.*, *Science* **322**, 1224 (2008).

(17) Supplementary online material.

(18) M. Luscher, *JHEP* **1008**, 071 (2010).

(19) S. Aoki, *et al.*, *Phys.Rev.* **D79**, 034503 (2009).

(20) H. Ohki, *et al.*, *Phys.Rev.* **D78**, 054502 (2008).

(21) H.-W. Lin, *et al.*, *Phys.Rev.* **D79**, 034502 (2009).

(22) A. Bazavov, *et al.*, *Rev.Mod.Phys.* **82**, 1349 (2010).

(23) S. Durr, *et al.*, *JHEP* **1108**, 148 (2011).

(24) R. Baron, *et al.*, *JHEP* **1006**, 111 (2010).

(25) W. Bietenholz, *et al.*, *Phys.Rev.* **D84**, 054509 (2011).

(26) R. Arthur, *et al.*, *Phys.Rev.* **D87**, 094514 (2013).

(27) P. Fritzsch, *et al.*, *Nucl.Phys.* **B865**, 397 (2012).

(28) M. Hayakawa, S. Uno, *Prog.Theor.Phys.* **120**, 413 (2008).

(29) H. Akaike, *IEEE Transactions on Automatic Control* **19**, 716 (1974).

(30) S. R. Coleman, S. L. Glashow, *Phys.Rev.Lett.* **6**, 423 (1961).

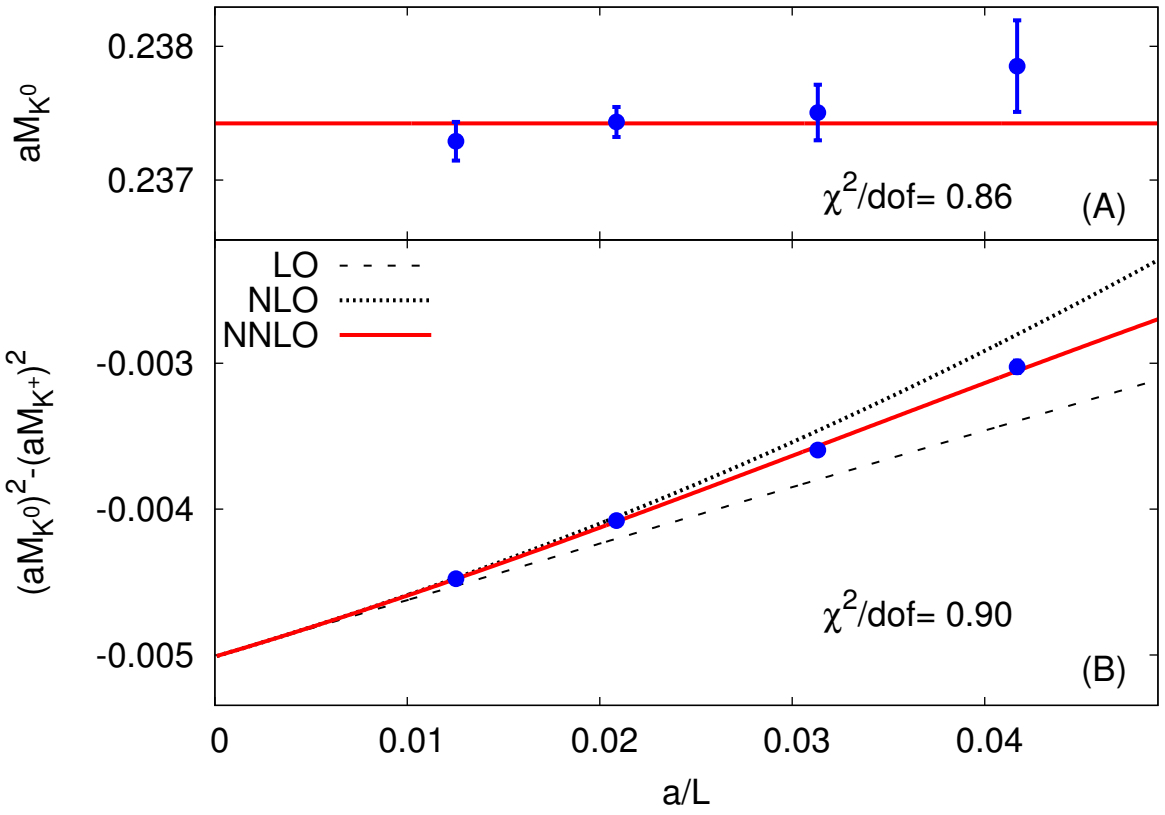

Figure 1: **Finite-volume behavior of kaon masses.** (**A**) The neutral kaon mass, M_{K^0}, shows no significant finite volume dependence; L denotes the linear size of the system. (**B**) The mass-squared difference of the charged kaon mass, M_{K^+}, and M_{K^0} indicates that M_{K^+} is strongly dependent on volume. This finite-volume dependence is well described by an asymptotic expansion in $1/L$ whose first two terms are fixed by QED Ward-Takahashi identities (*17*). The solid curve depicts a fit of the lattice results (points) to the expansion up to and including a fitted $O(1/L^3)$ term. The dashed and dotted curves show the contributions of the leading and leading plus next-to-leading order terms, respectively. The computation was performed by using the following parameters: bare $\alpha \sim 1/10$, $M_\pi = 290\,\mathrm{MeV}$, and $M_{K^0} = 450\,\mathrm{MeV}$. The mass difference is negative because a larger-than-physical value of α was used. The lattice spacing a is $\sim 0.10\,\mathrm{fm}$.

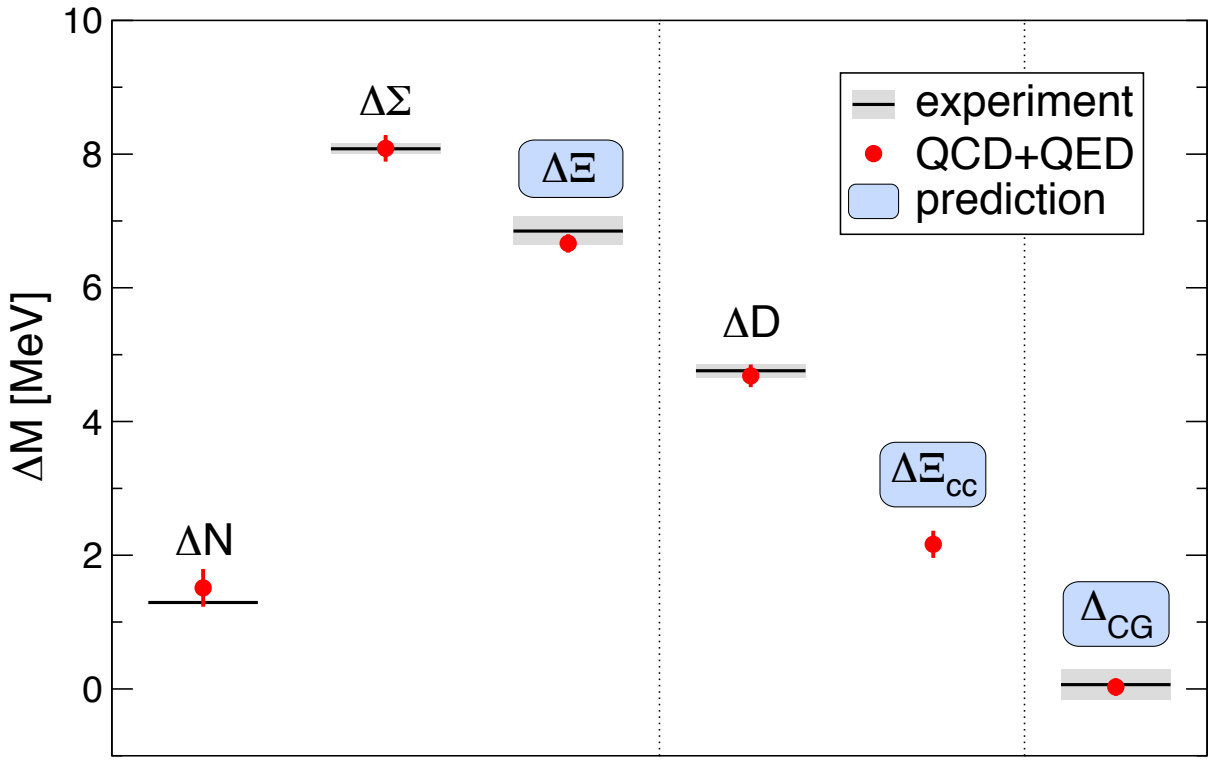

Figure 2: **Mass splittings in channels that are stable under the strong and electromagnetic interactions.** Both of these interactions are fully unquenched in our 1+1+1+1 flavor calculation. The horizontal lines are the experimental values and the grey shaded regions represent the experimental error (*2*). Our results are shown by red dots with their uncertainties. The error bars are the squared sums of the statistical and systematic errors. The results for the ΔM_N, ΔM_Σ, and ΔM_D mass splittings are post-dictions, in the sense that their values are known experimentally with higher precision than from our calculation. On the other hand, our calculations yield ΔM_Ξ, $\Delta M_{\Xi_{cc}}$ splittings, and the Coleman-Glashow difference $\Delta_{\rm CG}$, which have either not been measured in experiment or are measured with less precision than obtained here. This feature is represented by a blue shaded region around the label.

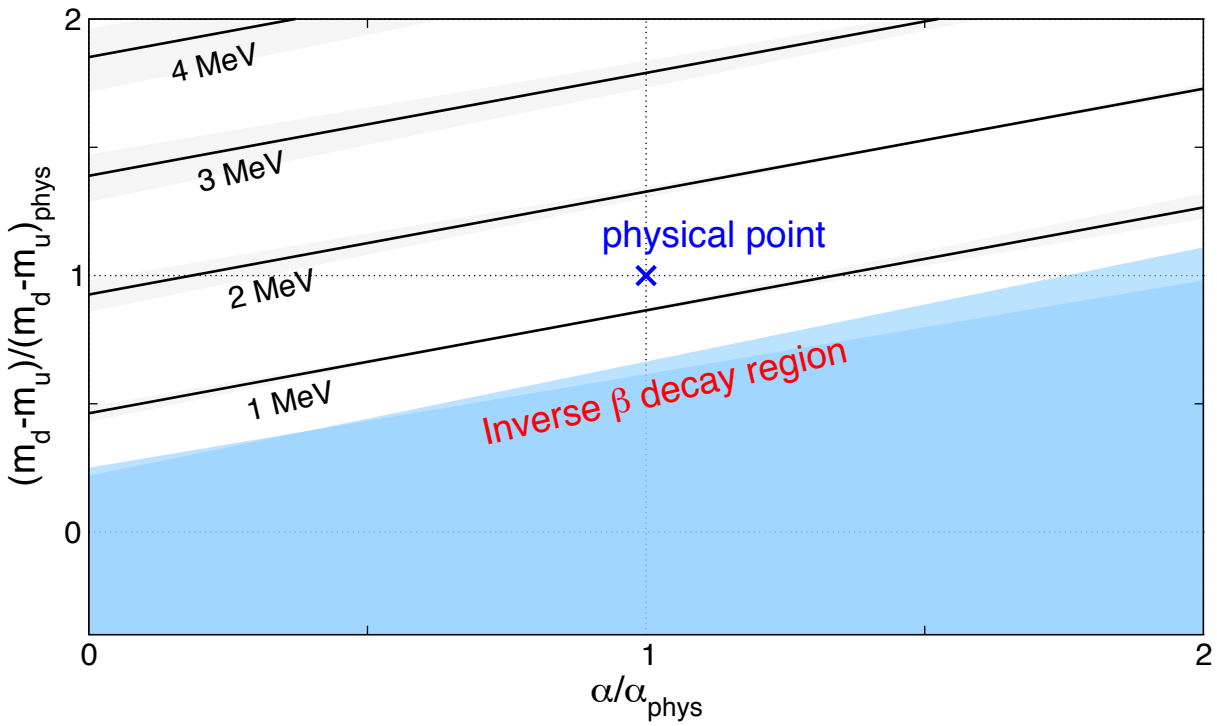

Figure 3: **Contour lines for the neutron-proton mass splitting.** The contours are shown as a function of the quark mass difference and the fine structure constant, both normalized with their real world, physical value. Because these two effects compete, by increasing α at fixed quark mass difference one can decrease the mass difference between the neutron and the proton to 0.511 MeV, at which inverse β-decay sets in, as depicted by the blue region. The blue cross shows the physical point. The shaded bands around the contours represent the total statistical and systematic uncertainties on these predictions. A constraint on the neutron-proton mass difference obtained from other considerations leads to a constraint on $m_d - m_u$ and/or α, which can be directly read off from the figure.

	mass splitting [MeV]	QCD [MeV]	QED [MeV]
$\Delta N = n - p$	1.51(16)(23)	2.52(17)(24)	-1.00(07)(14)
$\Delta\Sigma = \Sigma^- - \Sigma^+$	8.09(16)(11)	8.09(16)(11)	0
$\Delta\Xi = \Xi^- - \Xi^0$	6.66(11)(09)	5.53(17)(17)	1.14(16)(09)
$\Delta D = D^\pm - D^0$	4.68(10)(13)	2.54(08)(10)	2.14(11)(07)
$\Delta\Xi_{cc} = \Xi_{cc}^{++} - \Xi_{cc}^{+}$	2.16(11)(17)	-2.53(11)(06)	4.69(10)(17)
$\Delta_{\rm CG} = \Delta N - \Delta\Sigma + \Delta\Xi$	0.00(11)(06)	-0.00(13)(05)	0.00(06)(02)

Table 1: **Isospin mass splittings of light and charm hadrons.** Also shown are the individual contributions to these splittings from the mass difference $(m_d - m_u)$ (QCD) and from electromagnetism (QED). The separation requires fixing a convention, which is described in (*17*). The last line is the violation of the Coleman-Glashow relation (*30*), which is the most accurate of our predictions.

Supplementary Materials for

Ab initio calculation of the neutron-proton mass difference

Sz. Borsanyi[1], S. Durr[1,2], Z. Fodor[1,2,3], C. Hoelbling[1], S. D. Katz[3,4], S. Krieg[1,2], L. Lellouch[5], T. Lippert[1,2], A. Portelli[5,6], K. K. Szabo[1,2], B. C. Toth[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] MTA-ELTE Lendület Lattice Gauge Theory Research Group, H-1117 Budapest, Hungary

[5] CNRS, Aix-Marseille Université, Université de Toulon, CPT UMR 7332, F-13288, Marseille, France

[6] School of Physics & Astronomy, University of Southampton, SO17 1BJ, UK

Correspondence to: fodor@physik.uni-wuppertal.de

This PDF File includes:

Methods in Sections 1 to 12
Figures S1 to S12
Tables S1 to S5

1 Outline

In the following sections we provide details of the work presented in the main paper. In Sec. 2 we define the theory that we use, namely QCD and QED with four quark flavors on a four dimensional lattice. After fixing our notations we discuss the action for the photon field in Sec. 2.1. There are many subtleties here, such as gauge fixing and zero-mode subtraction. In Sec. 2.2 we define our Dirac operator, which contains both photon and gluon fields. We apply one step of APE smearing to the electromagnetic field and 3 steps of HEX smearing to the $SU(3)$ field. In Sec. 2.3 we discuss in detail what the advantages of these choices are and how we optimized the smearing parameters.

The determination of the mass splittings for the isospin multiplets needs a careful treatment of QED on the lattice. In Sec. 3 we discuss in detail the differences between two possible formulations of lattice QED (*28*) and illustrate numerically the disadvantages of the one used in all previous numerical studies in Sec. 4. We also compare our numerical implementation of lattice QED with lattice perturbation theory up to $\mathcal{O}(e^2)$ order (*31*) and recover the tiny $\mathcal{O}(e^4)$ corrections.

The photon fields are long-ranged and the mass corrections are proportional to $1/L$, $1/L^2$, ... types, where L is the spatial size of the system. These are much larger corrections than those in QCD, which are exponentially suppressed for stable particles. Actually they are of the same order as the mass-splittings themselves. In Sec. 3 we determine the finite-volume corrections analytically for point particles. In Sec. 5 the study is generalized to the case of composite particles. Using the Ward-Takahashi identities we show that the coefficients of the $1/L$ and $1/L^2$ terms are universal [1] and we compare these analytical findings to our numerical QCD+QED simulation results.

Starting with Sec. 6 we present the details of the many simulations that are performed and summarized here. The use of Rational Hybrid Monte-Carlo method is discussed with a special emphasis on the lowest eigenvalues of the Dirac operator. Autocorrelations are under control for our choice of parameters in the QCD part of our work. However, due to the zero mass of the photon and the correspondingly large correlation lengths, a standard Hybrid Monte-Carlo integration of the photon fields results in large autocorrelation times. We show how we solved this problem by developing a Fourier accelerated algorithm. For the propagator calculations we used a 2-level multi-grid approach to have several hundred source positions and significantly improve our statistics. We present the ensembles generated for this project in Sec. 7. We use four different lattice spacings in the range of 0.06 to $0.10\,\mathrm{fm}$, and pion masses down to $195\,\mathrm{MeV}$. We have runs with zero electromagnetic coupling and with non-zero ones. Altogether we have accumulated 41 ensembles, whose parameters are detailed in that section. In Sec. 8 we present our renormalization prescription for the electric charge.

In the final sections, Secs. 10, 11 and 12, we detail the procedure that is used to extract mass splittings. We explain how we separate the QED and quark-mass-difference contributions to these differences. In obtaining our final results, we conduct a thorough investigation of systematic uncertainties. To determine these, we use the histogram method (*16*). We extend it by using the Akaike's information criterion (AIC).

2 Lattice and action details

The elementary particles that we consider in this paper are photons (A), gluons (U), and the up, down, strange and charm quarks (ψ_f with $f = u, d, s, c$). The contributions from other known elementary particles to the isospin splittings can be neglected given the accuracy required for our study. The following action describes

[1] While we were writing up the results of the present paper, an analytical calculation of finite-volume effects in a non-relativistic effective field theory framework was presented (*32*). We agree on the universality of the $1/L$ and $1/L^2$ coefficients and on their values. However, simplifying the composite-particle results of (*32*) to the point-particle case leads to a coefficient of the $1/L^3$ term which differs from the one that we obtain in Sec. 3. We confirm the validity of the latter with high-precision QED simulations in that section.

the interactions the degrees of freedom kept here:

$$S[U, A, \overline{\psi}, \psi] = S_g[U; g] + S_\gamma[A] + \sum_f \overline{\psi}_f D[U, A; e, q_f, m_f]\psi_f.$$

where S_g and S_γ are the gluon and photon actions, respectively, and D is the Dirac operator. This action has the following parameters: the gluon gauge coupling g, the electromagnetic coupling e, the four quark masses m_f and the four charge parameters $q_u = q_c = 2/3$ and $q_d = q_s = -1/3$.

We work on a four dimensional, cubic, Euclidean lattice with L points in the spatial and T points in the time direction. The boundary condition is periodic for the photon and gluon fields. For the quark fields it is periodic in the spatial directions and anti-periodic in time. Lattice fields in coordinate space f_x can be transformed into momentum space f_k and back by Fourier transformation. To avoid cumbersome notations we use the same symbol for the Fourier-transformed fields. For periodic fields

$$k_\mu = \begin{cases} 2\pi/aT \cdot \{0, 1, \ldots, T-1\} & \mu = 0, \\ 2\pi/aL \cdot \{0, 1, \ldots, L-1\} & \mu = 1, 2, 3. \end{cases}$$

The discrete differential of a lattice field f_x is defined as $\partial_\mu f_x = (f_{x+a\mu} - f_x)/a$, and its adjoint is given by $\partial_\mu^\dagger f_x = (f_x - f_{x-a\mu})/a$. These differentials are given by multiplications in momentum space: $\hat{k}_\mu f_k$ and $-\hat{k}_\mu^* f_k$, where the components of the complex $\hat{k}$ vector are defined as

$$\hat{k}_\mu = \frac{\exp(iak_\mu) - 1}{a}. \tag{S1}$$

For the discretized gluon action S_g we choose the tree-level improved Symanzik action (*33*). Its properties are well known and will not be discussed here.

In order to eliminate discretization artefacts, we carry out a continuum limit. In QED this is a subtle issue, although without practical relevance here. When we talk about continuum limit in this work, we always mean a limiting procedure where the lattice spacing does not go exactly to zero, but to a minimal value. This value can be chosen to be extremely small for the coupling values considered in this work, and the remaining lattice artefacts are completely negligible. The uncertainties associated with extrapolating results to this minimal value are orders of magnitude larger and are duly accounted for in our analysis.

2.1 Photon action

This section gives the derivation of the photon action S_γ. We use a non-compact formulation. The naive photon action

$$S_\gamma^{naive}[A] = \frac{a^4}{4} \sum_{\mu,\nu,x} (\partial_\mu A_{\nu,x} - \partial_\nu A_{\mu,x})^2$$

is left invariant by gauge transformations with a field f_x:

$$A_{\mu,x} \to A_{\mu,x} - \partial_\mu f_x.$$

Field modes that are generated by these transformations, appear neither in the naive photon action nor in the part of the action which describes the coupling to the quarks. As a result, in the non-compact formulation gauge-variant observables, such as charged particle propagators, are ill-defined. To avoid this, one chooses a gauge (this would also be needed in the compact formulation). We use the Coulomb gauge for the photon field in this work.

There is another set of symmetry transformations of the naive photon action that shift the photon field by a constant c_μ:

$$A_{\mu,x} \to A_{\mu,x} + c_\mu.$$

Because of our use of periodic boundary conditions, this symmetry is not a gauge symmetry. However, in the infinite-volume limit it becomes a gauge symmetry with $f_x = -c_\mu x_\mu$. The treatment of this symmetry requires special attention, which we detail presently.

Zero mode subtraction

We eliminate the shift symmetry of the naive photon action by removing the zero-momentum mode of the photon field from the path integral. This step is not strictly necessary, since only a discrete subset of the shift transformation is a symmetry of the quark action. However significant complications were observed in simulations with a non-vanishing zero-mode (*34*). Additionally the zero-mode removal makes the theory well-defined perturbatively. The removal of modes, that form a set of measure zero in the infinite-volume limit, is a legitimate procedure, since it does not effect the path integral in this limit. There are different proposals in the literature for removing the zero mode. These correspond to different realizations of the theory in finite volume.

The simplest procedure is to set

$$a^4 \sum_x A_{\mu,x} = 0 \text{ for all } \mu. \tag{S2}$$

The sum runs for the temporal and spatial directions, so we will denote this choice QED_{TL}. This setup is the one used in all previous studies which include QED corrections to hadronic observables in lattice QCD (*7*, *10*, *11*, *35*). The disadvantage of this choice is, that it violates reflection positivity. This can be seen by adding the zero-mode constraints to the path integral in the following form:

$$\lim_{\xi \to 0} \exp\left[-\sum_\mu (a^4 \sum_x A_{\mu,x})^2 / \xi^2 \right].$$

A $\left(a^4 \sum_x A_{\mu,x}\right)^2$ term in the action, which connects fields at arbitrary positive and negative times, spoils reflection positivity. It has a serious consequence: charged particle propagators are ill-behaved, if the time extent of the box is sent to infinity while keeping its spatial size fixed. This is demonstrated both analytically and numerically in Secs. 3 and 4.

Another choice, proposed by Hayakawa and Uno (*28*), is to remove the zero mode of the field on each time slice separately:

$$a^3 \sum_{\vec{x}} A_{\mu,x_0,\vec{x}} = 0 \text{ for all } \mu \text{ and } x_0. \tag{S3}$$

The sums run here only over the spatial directions and this prescription will be denoted QED_L. This constraint can be shown to be reflection positive and charged-particle propagators are well behaved in this case. Thus particle masses can be extracted from the large time behavior of these propagators. We study both prescriptions in detail in this paper and compare them. The main results of the paper are obtained in the QED_L formulation.

Although QED_L represents a nonlocal modification of the path-integral and violates hypercubic symmetry, these effects vanish in physical quantities in the infinite volume limit. In finite volume an important issue is the renormalizability of the theory. We will show in Sec. 3 for the case of point-like particles and argue in Sec. 5 for the case of composite ones, that the divergences in the one-loop self-energy are the same in QED_L and in the infinite volume theory. No new counterterms are required for this particular diagram, which is the relevant one, when determining electromagnetic corrections to the masses. We expect, that this property holds for other quantities as well, although this has to be checked by explicit calculations, similar to the ones, that are presented in this paper.

Coulomb gauge via Feynman gauge

In order to ensure the existence of a transfer matrix it is convenient and usual to choose the Coulomb gauge. After removing the zero mode on each time slice, the Coulomb-gauge fixing condition $\vec{\nabla}^\dagger \cdot \vec{A}_x = 0$ defines a unique operator P_C, that transforms a field configuration into Coulomb gauge. The transformation in momentum space is given by:

$$A \to A' = P_C A \quad \text{with} \quad P_{C,\mu\nu} = \delta_{\mu\nu} - |\vec{\hat{k}}|^{-2} \hat{k}_\mu (0, \vec{\hat{k}}^*)_\nu, \tag{S4}$$

with $\hat{k}_\mu$ given in Eq. (S1). Generating field configurations in Coulomb gauge in the full dynamical case would be somewhat cumbersome. We therefore decided to generate configurations in Feynman gauge using the action

$$S_\gamma[A] = \frac{1}{2TL^3} \sum_{\mu,k} |\hat{k}|^2 |A_{\mu,k}|^2 \tag{S5}$$

and then transform them into Coulomb gauge using the P_C operator. It can be proven that this is equivalent to using Coulomb gauge directly.

2.2 The Dirac operator

The Wilson operator with tree-level clover improvement (*36*) is chosen as our lattice Dirac operator. The starting point is the gauge-covariant Dirac operator, which acts on a spinor field, ψ, as follows:

$$\begin{aligned}
&\left(D[U, A; e, q, m]\psi\right)_x = \\
&\left(\frac{4}{a} + m\right)\psi_x - \\
-\frac{1}{2a} \sum_\mu &\left[(1+\gamma_\mu) \exp(ieqa\tilde{A}_{\mu,x}) \tilde{U}_{\mu,x} \psi_{x+\mu} + (1-\gamma_\mu) \exp(-ieqa\tilde{A}_{\mu,x-\mu}) \tilde{U}^\dagger_{\mu,x-\mu} \psi_{x-\mu}\right] + \\
&-\frac{ia}{4} \sum_{\nu>\mu} \left(F^{(\tilde{U})}_{\mu\nu,x} + eqF^{(\tilde{A})}_{\mu\nu,x}\right) [\gamma_\mu, \gamma_\nu] \psi_x.
\end{aligned}$$

The gauge-invariance of the quark action is ensured by exponentializing the non-compact photon fields. We use the MILC convention for the gamma matrices γ_μ (*37*). Note also that this Dirac operator differs from the usual definition by an extra minus sign in front of the γ's. $F^{(U)}_{\mu\nu,x}$ is the usual discretization of the gluon field strength tensor and is built up from the products of the gluon links along the "clover" path. $F^{(A)}_{\mu\nu,x}$ is a discretization of the electromagnetic field strength tensor. It is chosen as the sum of the photon fields around a two-by-two plaquette centered at x in the $\mu - \nu$ plane. Gluon and photon fields ($\tilde{U}$ and $\tilde{A}$) that enter the Dirac operator are obtained by smearing the original gluon and photon fields. In this work the gluon fields have undergone three levels of HEX smearing. The parameters of the HEX smearing procedure are chosen with care, as described in Sec. 2.3. We smear the photon fields with the following transformation

$$A_{\mu,x} \to \tilde{A}_{\mu,x} = 0.9 \cdot A_{\mu,x} + 0.1 \cdot \sum_{\pm\nu\neq\mu} \left(A_{\nu,x} + A_{\mu,x+\nu} - A_{\nu,x+\mu}\right),$$

where $A_{-\nu,x} = -A_{\nu,x-\nu}$.

The most important advantage of smearing is the reduction of the additive quark mass renormalization. In our case, it has two contributions: one stems from non-trivial gluon fields, the other is due to the presence of photons. Although a large additive renormalization is not a problem of principle, a small one facilitates tuning the parameters in dynamical simulations. For illustration of the effect we define the electromagnetic

mass renormalization δ by a neutral mesonic state, which is obtained by dropping the disconnected part of the propagator of the quark–anti-quark system. δ is defined as the shift in the bare quark mass to get the same meson mass as in the $e = 0$ case, see also Ref. (*38*). At a lattice spacing of $0.10\,\mathrm{fm}$ and a coupling of $e = 1$ the shift for the up quark is about $\delta_u = -0.070$ without photon smearing, which is about the same size as the additive mass renormalization coming from the gluons. With our smearing recipe we have a four times smaller value $\delta_u = -0.017$.

As any smearing, our choices for the photon and gluon fields change the results by effects that disappear in the continuum limit. In Secs. 2.3 and 4 we demonstrate the advantages of our smearing choices.

2.3 HEX smearing

In our study, the gluon fields U have undergone three levels of HEX smearing (*39*). The smearing procedure replaces the original gluon fields U with the HEX smeared $\tilde{U}$:

$$\begin{aligned}
V_{\mu,\nu\rho,x} &= \exp\Big(\frac{\rho_1}{2}\sum_{\pm\sigma\neq\mu,\nu,\rho}\Big\{\big[U_{\sigma,x}\,U_{\mu,x+\sigma}\,U^\dagger_{\sigma,x+\mu}\,U^\dagger_{\mu,x} - \mathrm{h.c.}\,\big] - \frac{1}{3}\mathrm{Tr}[.]\Big\}\Big)U_{\mu,x},\\
W_{\mu,\nu,x} &= \exp\Big(\frac{\rho_2}{2}\sum_{\pm\rho\neq\mu,\nu}\Big\{\big[V_{\rho,\mu\nu,x}\,V_{\mu,\nu\rho,x+\rho}\,V^\dagger_{\rho,\mu\nu,x+\mu}\,U^\dagger_{\mu,x} - \mathrm{h.c.}\,\big] - \frac{1}{3}\mathrm{Tr}[.]\Big\}\Big)U_{\mu,x},\\
\tilde{U}_{\mu,x} &= \exp\Big(\frac{\rho_3}{2}\sum_{\pm\nu\neq\mu}\Big\{\big[W_{\nu,\mu,x}\,W_{\mu,\nu,x+\nu}\,W^\dagger_{\nu,\mu,x+\mu}\,U^\dagger_{\mu,x} - \mathrm{h.c.}\,\big] - \frac{1}{3}\mathrm{Tr}[.]\Big\}\Big)U_{\mu,x}. \qquad \text{(S6)}
\end{aligned}$$

For further details on our implementation see (*23*).

Here, we illustrate our procedure of iterated HEX smearings on lattices generated in the pure-gauge case with a Wilson action. We generated sets of matched lattices with a fixed box length in units of the Sommer scale $L/r_0 = 3$. We used the formula for r_0/a as a function of the gauge coupling from (*40*), which is based on data from (*41*). A wide range of lattice spacings were covered from $a = 0.245$ fm down to $a = 0.046$ fm. We consider the three-fold HEX smeared plaquette as a function of g^2 on these lattices with special attention to the $g^2 \to 0$ behavior. Since the last step in Eq. (S6) is standard stout smearing, we restrict ourselves to $\rho_3 = 0.12$. This value has been used in many studies based on stout smearing in the past and is considered safe by perturbative considerations (*39*). Hence, only two parameters need to be tuned to optimize the scaling to the continuum limit. We selected the value $\rho_{\mathrm{HEX}} = (0.22, 0.15, 0.12)$ as our preferred HEX-smearing parameter. These parameters correspond to $\alpha_{\mathrm{HYP}} = (0.44, 0.60, 0.72)$ in the HYP-smearing scheme. In Fig. S1 we show results for the average plaquette, $\langle U_\Box\rangle$, for three different smearing levels. $\langle U_\Box\rangle$ approaches 1 monotonically in the continuum limit.

3 Analytical studies of various QED formulations in finite volume

In this section we derive a number of important results concerning the properties of the pole mass of a charged particle in various formulations of QED on a finite spacetime volume. We focus here on point particles, because the main features are already present in this simpler situation. For the formulation of QED that we use in our simulations, in Sec. 5 we investigate the modifications to these calculations which result from the fact that mesons and baryons have internal structure.

It is important to have a solid analytical handle on QED finite-volume (FV) corrections, because they are expected to be large due to the long-range nature of the electromagnetic interaction. Unlike QCD, QED has no gap and the photon remains massless even in the presence of interactions. While the gap in QCD guarantees that FV corrections fall off exponentially in LM_π for sufficiently large LM_π (*42*), in the presence of QED, quantities are much more sensitive to the volume and topology of spacetime. It is the main characteristics of

this sensitivity which concerns us in this section. We use the computed analytical expressions in two important ways. The first is to decide on the finite-volume formulation of QED to use in our numerical work. The second is to test our implementation of QED and the corresponding codes.

The work presented in this paper is concerned with spin-$1/2$ baryons and spin-0 pseudoscalar mesons. Thus we compute the FV corrections in spinor and scalar QED. Our photon field has periodic boundary conditions, while the quark fields are periodic in space and antiperiodic in time. Therefore, baryon fields are antiperiodic in time and periodic in space, while meson fields are periodic in all directions. As a result, the topology of our spacetime is the four-torus, $\mathbb{T}^4$, up to a twist for baryons in the time direction. Note that for corrections in inverse powers of the torus size, only the photon boundary conditions are relevant.

As discussed in Sec. 2.1, we consider two different versions of FV QED:

- the first where only the four-momentum zero-mode of the photon field is eliminated, i.e. $A_\mu(k=0) \equiv 0$, which we denote QED_{TL};
- the second where all three-momentum zero-modes of the photon field are eliminated, i.e. $A_\mu(k_0, \vec{k} = \vec{0}) \equiv 0$ for all k_0, which we denote QED_L.

Power-like FV corrections arise from the exchange of a photon around the torus. They are obtained by comparing results obtained in FV with those of our target theory, QED in infinite volume (IV), that is in $\mathbb{R}^4$. Here we are interested in the FV corrections to a charged particle's pole mass. This is the physical mass of the particle, as obtained by studying the Euclidean time-dependence of a relevant, zero three-momentum, two-point correlation function. This mass is gauge invariant and we use this freedom to work in the simpler Feynman gauge.

The FV corrections to the mass m of a point particle of spin J and of charge q in units of e, on a torus of dimensions $T \times L^3$, is given by the difference of the FV self energy, $\Sigma_J(p, T, L)$, and its IV counterpart, $\Sigma_J(p)$, on shell:

$$\begin{aligned} \Delta m_J^{n_J}(T,L) \equiv m_J^{n_J}(T,L) - m^{n_J} &= (qe)^2 \Delta\Sigma_J(p = im, T, L) \\ &\equiv (qe)^2 \left[\Sigma_J(p = im, T, L) - \Sigma_J(p = im)\right] , \end{aligned} \tag{S7}$$

where $n_J = 1$ (resp. $n_J = 2$) for spin $J = 1/2$ fermions (resp. spin $J = 0$ bosons) and $p = im$ is a shorthand for $p = (im, \vec{0})$ (with $\not{p} \to im$ for spin-$1/2$ fermions). Here and below, quantities without the arguments L and T are infinite spacetime-volume quantities.

Because we only work in a regime where electromagnetic effects are linear in the fine structure constant α, we evaluate the self-energy difference in Eq. (S7) at one loop. At this order, we generically write differences of self energies or of contributions to self energies as

$$\Delta\Sigma(p, T, L) = \left[\sum_k\!\!\!\!\!\!\int^{\prime} - \int \frac{d^4k}{(2\pi)^4} \right] \sigma(k, p) , \tag{S8}$$

where k is the momentum of the photon in the loop and $\sigma(k, p)$ is the appropriate, IV self-energy integrand, a number of which are defined below. The individual FV and IV terms in Eq. (S8) are generally UV and possibly IR divergent. Thus, individually they should be regularized, e.g. with dimensional regularization. However, on shell the IV integral is IR finite and in finite volume, the sums are IR finite because the FV formulations of QED that we consider are regulated by the space or spacetime volume. Moreover, for large k^2, the functions $k^3\sigma(k, p)$ that arise in Eq. (S8) are strictly monotonic. Therefore, the proof of Cauchy's integral criterion (*43*) guarantees that the difference of the FV sums and IV integrals is also UV finite. Thus, to compute the FV corrections at one loop, no explicit regularization is required.

In Eq. (S8), the information about the topology of the finite volume and the specific formulation of QED is contained in the definition of $\not{\sum}'_k$. In addition to the case of $\mathbb{T}^4$ already discussed, we will also consider the four-cylinder $\mathbb{R} \times \mathbb{T}^3$. This is a useful intermediate step computationally, because it allows single particle propagators to develop a pole at the particle's energy.

For the cases of interest here, we have the following definitions for $\not{\sum}'_k$:

- QED_L on $\mathbb{R} \times \mathbb{T}^3$:

$$\not{\sum_k}' \equiv \int_{-\infty}^{+\infty} \frac{dk_0}{2\pi} \frac{1}{L^3} \sum_{\vec{k} \in \mathrm{BZ}_L^{3*}} , \tag{S9}$$

  with $\mathrm{BZ}_L^{3*} \equiv \frac{2\pi}{L}\mathbb{Z}^{3*}$ and the star, as usual, indicates the removal of the zero element;

- QED_L on $\mathbb{T}^4$:

$$\not{\sum_k}' \equiv \frac{1}{TL^3} \sum_{k_0 \in \mathrm{BZ}_T} \sum_{\vec{k} \in \mathrm{BZ}_L^{3*}} , \tag{S10}$$

  with $\mathrm{BZ}_T \equiv \frac{2\pi}{T}\mathbb{Z}$;

- QED_{TL} on $\mathbb{T}^4$:

$$\not{\sum_k}' \equiv \frac{1}{TL^3} \sum_{k_\mu \in \mathrm{BZ}_{TL}^{4*}} , \tag{S11}$$

  where $\mathrm{BZ}_{TL}^{4*} \equiv [\frac{2\pi}{T}\mathbb{Z} \times \frac{2\pi}{L}\mathbb{Z}^3]^*$.

The last ingredient of a general nature, needed to study the FV corrections in the three cases of interest, is the integrand of the self-energy $\sigma_J(k,p)$ for fermions ($J = 1/2$) and bosons ($J = 0$). These are obtained from the usual one-loop spinor and scalar self-energy Feynman diagrams, yielding the following expressions:

$$\sigma_{\frac{1}{2}}(k,p) = (2i\not{p} + 4m)\sigma_{S_1}(k,p) + 2i\not{\sigma}_{S_2}(k,p) \tag{S12}$$

$$\sigma_0(k,p) = 4\sigma_T(k) - \sigma_{S_0}(k,p) - 4p^2\sigma_{S_1}(k,p) - 4p_\mu\sigma_{S_2,\mu}(k,p) , \tag{S13}$$

with,

$$\begin{aligned} \sigma_T(k) &= \tfrac{1}{k^2} , & \sigma_{S_0}(k,p) &= \tfrac{1}{[(p+k)^2+m^2]} , \\ \sigma_{S_1}(k,p) &= \tfrac{1}{k^2[(p+k)^2+m^2]} , & \sigma_{S_2,\mu}(k,p) &= \tfrac{k_\mu}{k^2[(p+k)^2+m^2]} . \end{aligned} \tag{S14}$$

In the following subsection we provide a brief description of the methods used to obtain the FV corrections in inverse powers of the volume for the three formulations of FV QED described above. We summarize the results and discuss their consequences in Sec. 3.2.

3.1 Computation of finite-volume corrections in various QED formulations

Finite-volume corrections in QED_L on $\mathbb{R} \times \mathbb{T}^3$

QED_L on $\mathbb{R} \times \mathbb{T}^3$ is not a setup that can be considered directly in lattice simulations, as it describes a spacetime with an infinite time direction. However, it is a useful first step for computing FV corrections to masses analytically. These corrections are obtained from Eqs. (S7) and (S8) with the FV self-energy sum defined through Eq. (S9). To evaluate the resulting expressions and obtain an asymptotic expansion in powers of $1/L$, we apply the Poisson summation formula to the sum over the three-momentum $\vec{k}$, subtracting appropriately the

$\vec{k}=0$ modes. Then, using techniques from (*28*, *44*) and carrying out the asymptotic expansions to the end, up to exponentially small corrections in mL, we obtain:

$$\left[\Delta^{\mathrm{QED}\to\mathrm{QED}_L}_{\mathbb{R}^4\to\mathbb{R}\times\mathbb{T}^3}\right]\begin{Bmatrix}\Sigma_T(L)\\ \Sigma_{S_0}(im,L)\\ \Sigma_{S_1}(im,L)\\ \Sigma_{S_{2,\mu}}(im,L)\end{Bmatrix} = \left[\sum_{\vec{x}\in L\mathbb{Z}^{3*}} -\frac{1}{L^3}\int d^3x\right]\int\frac{d^4k}{(2\pi)^4}\begin{Bmatrix}\frac{1}{k^2}\\ \frac{1}{2imk_0+k^2}\\ \frac{1}{k^2[2imk_0+k^2]}\\ \frac{k_\mu}{k^2[2imk_0+k^2]}\end{Bmatrix}e^{i\vec{k}\cdot\vec{x}}$$
$$\underset{L\to+\infty}{\sim}\begin{Bmatrix}-\frac{\kappa}{4\pi L^2}\\ -\frac{1}{2mL^3}\\ -\frac{\kappa}{16\pi mL}+\frac{1}{8m^3L^3}\\ im\delta_{\mu0}\left(\frac{\kappa}{8\pi m^2L^2}-\frac{1}{4m^3L^3}\right)\end{Bmatrix}, \tag{S15}$$

where

$$\begin{aligned}\kappa &\equiv \int_0^\infty\frac{d\lambda}{\lambda^{3/2}}\left\{\lambda^{3/2}+1-\left[\theta_3(0,e^{-\frac{\pi}{\lambda}})\right]^3\right\}\\ &= 2.837297(1)\,,\end{aligned} \tag{S16}$$

and where $\theta_3(u,q)=\sum_{n\in\mathbb{Z}}q^{n^2}e^{i2nu}$ is a Jacobi theta function. In Eq. (S15), $\Sigma_i(im,L)$ is the on-shell self energy corresponding to $\sigma_i(k,im)$, $i=T,\cdots,S_{2,\mu}$, and the notation $\left[\Delta^{\mathrm{QED}\to\mathrm{QED}_L}_{\mathbb{R}^4\to\mathbb{R}\times\mathbb{T}^3}\right]$ indicates that these corrections must be added to the relevant quantity determined in QED on $\mathbb{R}^4$ (i.e. standard IV QED) to obtain the quantity appropriate for QED_L on $\mathbb{R}\times\mathbb{T}^3$. A similar notation is used below for other corrections, with a meaning which is a straightforward generalization of the one described here.

Finite-volume corrections in QED_L on $\mathbb{T}^4$

The FV corrections to the mass of a point particle are obtained from Eqs. (S7) and (S8) with the FV self-energy sum defined through Eq. (S10). Instead of performing an asymptotic expansion for $T,L\to\infty$ directly on this expression, it is easier to compute the corrections to the results obtained for QED_L on $\mathbb{R}\times\mathbb{T}^3$ that result from compactifying the time direction to a circle of circumference T. In that case, instead of the expressions in Eq. (S8), we must compute:

$$\left[\Delta^{\mathrm{QED}_L\to\mathrm{QED}_L}_{\mathbb{R}\times\mathbb{T}^3\to\mathbb{T}^4}\right]\Sigma(im,T,L) = \left[\frac{1}{T}\sum_{k_0\in\mathrm{BZ}_T}-\int\frac{dk_0}{2\pi}\right]\sum_{\vec{k}\in\mathrm{BZ}^{3*}_L}\sigma(k,im) \tag{S17}$$
$$= \sum_{x_0\in T\mathbb{Z}^*}\sum_{\vec{k}\in\mathrm{BZ}^{3*}_L}\int\frac{dk_0}{2\pi}\sigma(k,im)e^{ik_0x_0}\,, \tag{S18}$$

where, again, we have used Poisson's summation formula. Inspection of Eq. (S14) indicates that the functions $\sigma(k,im)$ have no poles on the real k_0-axis, are infinitely differentiable and all of their derivatives are integrable. Therefore, their Fourier transform in Eq. (S18) vanishes faster than any power of $1/T$ as $T\to\infty$. This means that the FV corrections to the on-shell self-energy in QED_L on the four-torus of dimensions $T\times L^3$, are the same as those on the four-cylinder $\mathbb{R}\times\mathbb{T}^3$, up to corrections that vanish faster than any inverse power of T , i.e.:

$$\left[\Delta^{\mathrm{QED}\to\mathrm{QED}_L}_{\mathbb{R}^4\to\mathbb{T}^4}\right]\Sigma(im,T,L)\underset{T,L\to+\infty}{\sim}\left[\Delta^{\mathrm{QED}\to\mathrm{QED}_L}_{\mathbb{R}^4\to\mathbb{R}\times\mathbb{T}^3}\right]\Sigma(im,L)\,. \tag{S19}$$

Finite-volume corrections in QED_{TL} on $\mathbb{T}^4$

In this setup, the FV corrections to the self-energy of a point particle are obtained from Eq. (S8) with the FV sum defined through Eq. (S11). As in the previous section, instead of performing a $T, L \to \infty$ asymptotic expansion directly on this expression, we compute the corrections to the results obtained for QED_L on $\mathbb{T}^4$ which we obtain from reinstatement, as dynamical variables, the photon field modes $\tilde{A}_\mu(k_0, \vec{0})$ with $k_0 \neq 0$. These corrections require computing:

$$\left[\Delta^{\mathrm{QED}_L \to \mathrm{QED}_{TL}}_{\mathbb{T}^4 \to \mathbb{T}^4}\right] \Sigma(p, T, L) = \frac{1}{TL^3} \sum_{k_0 \in \mathrm{BZ}^*_T, \vec{k}=\vec{0}} \sigma(k, p) \,, \tag{S20}$$

for the various self-energy integrands $\sigma(k, p)$ of Eqs. (S12)-(S14).

The functions which appear in $\sigma(k, p)$, for $\vec{k} = \vec{0}$, are rational functions of k_0. There are known systematic methods to sum series of such functions. These involve performing partial fraction decompositions and then exploiting the properties of the polygamma functions to sum the individual terms in these decompositions. Using these methods, we obtain:

$$\left[\Delta^{\mathrm{QED}_L \to \mathrm{QED}_{TL}}_{\mathbb{T}^4 \to \mathbb{T}^4}\right] \begin{Bmatrix} \Sigma_T(T, L) \\ \Sigma_{S_0}(im, T, L) \\ \Sigma_{S_1}(im, T, L) \\ \Sigma_{S_2,\mu}(im, T, L) \end{Bmatrix} \underset{T,L\to+\infty}{\sim} \begin{Bmatrix} \frac{T}{12L^3} \\ \frac{\coth(mT)}{4mL^3} - \frac{1}{4m^2TL^3} \\ \frac{T}{48m^2L^3} - \frac{\coth(mT)}{16m^3L^3} + \frac{1}{16m^4TL^3} \\ -2im\delta_{\mu 0}\, \Sigma_{S_1}(im, T, L) \end{Bmatrix} \tag{S21}$$

which are the corrections that must be added to the self-energy contributions in QED_L on $\mathbb{T}^4$ to obtain the FV contributions in QED_{TL} on $\mathbb{T}^4$.

3.2 Results for finite-volume corrections to the pole mass and consequences for the various QED formulations

In this subsection we combine the results of the previous subsection to obtain the FV corrections to the pole masses of point spinor and scalar particles for two versions of FV QED of interest for lattice calculations:

- QED_L on $\mathbb{T}^4$, which is the formulation used in the present study;
- QED_{TL} on $\mathbb{T}^4$, which is the formulation used in previous lattice studies of isospin breaking effects.

While some of the details of the results obtained in this section are specific to point particles, the general conclusions also carry over to the case of composite particles, which is discussed in Sec. 5.

Finite-volume corrections in QED_L on $\mathbb{T}^4$

QED_L on $\mathbb{T}^4$ is the formulation used in the present study of isospin breaking effects. Combining the results of Eqs. (S15) and (S19) and putting everything together, we find that the mass of a point-like fermion of spin $1/2$, of charge q in units of e, on the four-torus $\mathbb{T}^4$ of dimensions $T \times L^3$ in QED_L, is at one-loop, in terms of its infinite-volume mass m:

$$m_{\frac{1}{2}}(T, L) \underset{T,L\to+\infty}{\sim} m \left\{1 - q^2\alpha \left[\frac{\kappa}{2mL}\left(1 + \frac{2}{mL}\right) - \frac{3\pi}{(mL)^3}\right]\right\} \,, \tag{S22}$$

up to terms which are exponentially suppressed in mL and terms which fall faster than any power in $1/(mT)$, with κ given in Eq. (S16). Similarly, the FV corrections to the mass of a point-like boson of spin 0 are, in terms of its infinite-volume mass m:

$$m_0^2(T, L) \underset{T,L\to+\infty}{\sim} m^2 \left\{1 - q^2\alpha \left[\frac{\kappa}{mL}\left(1 + \frac{2}{mL}\right)\right]\right\} \,. \tag{S23}$$

Four important comments are in order. The first is that the finite-volume pole masses in both cases have a well defined $T, L \to \infty$ limit and converge onto their infinite-volume counterparts. The second is that the coefficient of the leading $1/L$ and $1/L^2$ corrections to the mass m of a particle of charge qe is the same for spin-$1/2$ fermions and spin-0 bosons at $O(\alpha)$. In Sec. 5 we show that these coefficients are always the same, independent of the spin and point-like nature of the particle: they are fixed by QED Ward-Takahashi identities. Moreover, as suggested in (*38*) and worked out explicitly in (*32*), the leading $1/L$ term is the FV correction to the classical, electrostatic potential of a point charge on $\mathbb{T}^3$, with the spatial zero-modes removed from Gauss' law. The third comment is that the dimensionless, relative FV corrections must be functions of the only dimensionless parameter, mL, in the two theories considered. This will no longer be the case when we consider physical mesons and baryons, as these particles are not point-like and therefore have relevant scales other than their mass. The final remark is that we find a coefficient for the $1/(mL)^3$ term in Eq. (S22) which is twice the one found in (*32*), when the result for composite fermions in that paper is reduced to the point-like case. As shown in Sec. 4, this factor of 2 is confirmed by direct simulation of QED_L on $\mathbb{T}^4$.

Finite-volume corrections in QED_{TL} on $\mathbb{T}^4$

The setup considered in this section, QED_{TL} on the four-torus $\mathbb{T}^4$ of dimensions $L^3 \times T$, is the one used in all previous studies which include QED corrections to hadronic observables in lattice QCD (*7*,*8*,*10–13*,*35*,*38*,*45*). As discussed in Sec. 2.1, it violates reflection positivity. Here we show that it has another problem: it does not have a well defined $T \to \infty$ limit for fixed L. It is these reasons which have led us to choose to simulate QED_L instead of QED_{TL} for the precision computation presented in the present paper.

The finite-volume corrections to the masses of point particles in QED_{TL} on $\mathbb{T}^4$ are obtained by adding, to those in QED_L on $\mathbb{T}^4$ (Eqs. (S22) and (S23)), the corrections on the self-energy components determined in Eq. (S21). This yields the following result for the mass of a spin $J = 1/2$ point-particle, of charge q in units of e, in QED_{TL} on the four-torus of dimensions $T \times L^3$, in terms of its infinite-volume counterpart, m:

$$m_{\frac{1}{2}}(T,L) \underset{T,L\to+\infty}{\sim} m\left\{1 - q^2\alpha\left[\frac{\kappa}{2mL}\left(1 + \frac{2}{mL}\left[1 - \frac{\pi}{2\kappa}\frac{T}{L}\right]\right) - \frac{3\pi}{(mL)^3}\left[1 - \frac{\coth(mT)}{2}\right] - \frac{3\pi}{2(mL)^4}\frac{L}{T}\right]\right\}, \tag{S24}$$

up to terms which are exponentially suppressed in mL and terms which fall faster than any power in $1/(mT)$, with κ given in Eq. (S16). Similarly, the FV corrections to a point-like boson of spin 0 are, in terms of the infinite-volume mass m:

$$m_0^2(T,L) \underset{T,L\to+\infty}{\sim} m^2\left\{1 - q^2\alpha\left[\frac{\kappa}{mL}\left(1 + \frac{2}{mL}\left[1 - \frac{\pi}{2\kappa}\frac{T}{L}\right]\right)\right]\right\}. \tag{S25}$$

A number of important remarks about these results deserve to be made. We begin by considering T/L as being $O(1)$, as it is usually in lattice simulations. The first remark is that the leading $1/L$ and $1/L^2$ contribution are identical for both spins, as was the case in QED_L. However, only the $1/L$ terms here are equal with those found in QED_L. The reinstatement of the spatially-uniform photon modes for $k_0 \neq 0$ reduces the coefficient of the subleading $1/L^2$ contributions for $T \sim L$, compared to what it is in QED_L.

It should also be noted that both masses acquire new T-dependent, $1/L^3$ and $1/L^4$ contributions, which remain under control even when $T \gg L$. And, as in the case of QED_L, the dimensionless, relative, FV mass corrections of Eqs. (S24) and (S25) can be written in terms only of dimensionless quantities. However, while in QED_L there is only one dimensionless variable, mL, here there is another, the aspect ratio of $\mathbb{T}^4$, $\xi = T/L$. This will no longer be the case when we consider physical mesons and baryons, as these particles are not point-like and therefore have relevant scales other than their mass.

The problem in QED_{TL} on the four-torus arises if one considers the limit $T \to \infty$ for fixed L. In that case the $1/L^2$ corrections blow up linearly in T/L. In particular, this means that the conventional mass extraction procedure, which relies on examining the asymptotic time behavior of Euclidean propagators, is not well-defined at finite L. This justifies, for the precision study presented in this paper, our choice of working with QED_L instead of QED_{TL} that was used in previous studies. It should be noted, however, that this problem only arises at $O(1/L^2)$ and remains mild for typical T/L used in lattice computations. Thus, for lower precision studies with $T/L \sim 1$ such as in (*10*), where no sensitivity to terms of higher order than a fitted $1/L$ was found, the effect of using QED_{TL} will not significantly distort results.

4 Finite-volume and lattice-spacing corrections in QED: numerical investigations

The results of this section serve to verify the implementation of the QED part of our simulation, which is one of the novel features of this work. The infinite volume and continuum extrapolations will be shown to be under control. Both extrapolations from the numerical data yield results that are consistent with analytical calculations.

Our QED setup is obtained from the full case by setting the gluon fields to $U = 1$. Since we focus on the order e^2 behavior, we will use the quenched QED approximation in this illustrative section, which is correct up to e^4 effects. We take $e = \sqrt{4\pi/137}$ and a fermion charge of $q = 1$ unless stated otherwise. In some cases we switch off the clover improvement and/or the photon field smearing.

The field configuration generation in the absence of fermion loops can be done by a heatbath algorithm. Using a normal Gaussian distributed random vector η_μ, a Feyman-gauge field configuration with the right distribution is given by $A_{\mu,k} = |\hat{k}|^{-1}\eta_{\mu,k}$. This is then fixed to Coulomb-gauge by the operator in Eq. (S4).

Let us first take a look at the effective mass of a fermion in the two different realizations of finite volume QED: QED_{TL} and QED_L (see Fig. S2). The difference is in the zero-mode removal prescription. The following parameters were used: spatial box size $L = 4$, bare Wilson mass $m = 0.4$, no clover improvement, no smearing. In QED_{TL} no clear plateaus can be seen, moreover increasing the time extent also increases the mass values. This is in complete agreement with the analytical finding of Sec. 3: the mass diverges in the infinite time extent limit, if the spatial box size is kept fixed. This of course makes the conventional mass extraction procedure, which examines the asymptotic behavior of the propagator, ambiguous in this scheme. If instead we work with the zero-mode prescription of QED_L, we get masses that are insensitive to the time extent of the lattice, as predicted.

Now let us consider the finite size dependence of the fermion mass (see Fig. S3). For the plot we used a bare Wilson mass of $m = 0.2$, clover improvement and smearing. The points with filled symbols are obtained in the QED_L formulation. We also extract masses using data in the QED_{TL} setup, though identification of the plateaus is sometimes unclear, as discussed above. These are the open symbols on the plot and correspond to three different aspect ratios. The masses change with T at a fixed spatial volume. We fit the results of QED_L and QED_{TL} together to the analytical formulae of Eqs. (S22) and (S24) with a single fit parameter, which is the infinite volume fermion mass. A good quality fit can be achieved, if we restrict the boxes to $a/L < 0.04$. The results are the solid lines on the plot. The small disagreements for smaller boxes can probably be explained by exponentially small finite size corrections, which are neglected in the formulae. The data has sufficient precision to determine the coefficient of the $1/L^3$ term. It strongly favors the one that we find by analytic calculation in Eq. (S24) over the one obtained by reducing the composite-particle result of (*32*) to the point-particle case. [2] Fitting the numerical results to the formula of (*32*) leads to a $\chi^2/\mathrm{dof} = 1.4$. A fit of the

[2] We perform this reduction by setting the charge radius and all other coefficients which reflect the composite nature of the particle to zero.

reduced, NNLO expression of (*32*) to the QED_L points gives a $\chi^2/\mathrm{dof}= 15$. These fits are shown in Fig. S3 with solid/dotted lines. Including also the small volumes in the analysis (dashed lines) the χ^2/dof is increased by a factor of 2.5 for Eq. (S24), whereas the formula of (*32*) provides a χ^2/dof well over 100.

Now let us turn our attention to the lattice artefacts of our QED lattice action. In order to carry out a lattice spacing scaling study, we need two observables: one which sets the lattice spacing and another which has a non-trivial continuum limit[3]. For the first the obvious choice is the fermion mass, in a given finite box. We denote it m_L. For the second we take the finite volume correction: the difference of m_L to the mass measured on a lattice with twice the spatial extent, denoted by m_{2L}. We set the box size as $Lm_L = 2$ and work in the QED_L formulation. The result is shown in Fig. S4. The filled/open points are obtained with/without clover improvement and smearing. In the unimproved case we use two functions to extrapolate into the continuum: one is linear and the other is linear-plus-quadratic in the lattice spacing. For the improved case a quadratic function is plotted, but the data is even consistent with a constant after ignoring the result from the coarsest lattice. A continuum value can also be obtained analytically using the results of the previous section. At a mass of $Lm_L = 2$ the exponential corrections cannot be neglected. Thus we compute the $L(m_{2L} - m_L)$ using Eq. (S7) and evaluating the relevant integrals and sums in Eq. (S15) numerically instead of using the asymptotic formulae. This is the crossed circle on the plot. Both the unimproved and the improved results agree with the analytical formula, though the unimproved has a much larger systematic error coming from the continuum extrapolation.

Finally in order to further support our QED setup, we compare the infinite volume extrapolated fermion masses at finite lattice spacing to analytical results from the literature. In (*31*) Mertens, Kronfeld and El-Khadra calculated the pole position of the fermion propagator for arbitrary Wilson mass, clover coefficient and to leading order in e^2. For the comparison we used the QED_L formulation, clover improvement and unsmeared photons. The bare Wilson mass was set to $am = 0.2$. We carried out the infinite volume extrapolation using the analytical finite volume formula, Eq. (S22). There is a small discrepancy at the standard coupling value, $e = \sqrt{4\pi/137}$. To demonstrate, that it is an $O(e^4)$ effect, we carried out simulations at two other coupling values. The infinite volume extrapolated results are recorded in Tab. S1. The difference to the analytical result is consistent with being an e^4 correction.

5 Finite-volume effects for composite particles

As discussed in Sec. 3, finite-volume (FV) effects are particularly important, because of the presence of the massless photon. While in that section we computed these effects for point-like scalar and spinor particles, here we generalize the discussion to composite particles, which can be hadrons, nuclei, etc. We consider only the QED formulation that was actually used in our numerical work, namely QED_L on the four-torus $\mathbb{T}^4$ of dimensions $T \times L^3$ and use the fact that it differs from QED_L on $\mathbb{R} \times \mathbb{T}^3$ by corrections in T which fall of faster than any power in $1/T$ (see Sec. 3).

As shown in Sec. 3 (Eqs. (S22) and (S23)), the leading and next-to-leading FV corrections to the mass of a point particle of charge qe in scalar and fermion QED are identical at $\mathcal{O}(\alpha)$:

$$m(T,L) \underset{T,L\to+\infty}{\sim} m\left\{1 - q^2\alpha\frac{\kappa}{2mL}\left[1+\frac{2}{mL}\right] + \mathcal{O}\left(\frac{\alpha}{L^3}\right)\right\}\,, \tag{S26}$$

with κ given in Eq. (S16). Here we show that both corrections remain unchanged for composite particles. This result follows from the QED Ward-Takahashi identities (WTIs) and from the assumptions that the photon is the only massless asymptotic state, and that the charged particle considered is stable and non-degenerate in mass with any other particle which shares its quantum numbers.

[3] Since we work in the quenched approximation in this section, the triviality issue is absent. This is because the renormalized coupling equals the bare coupling and it can be kept fixed for arbitrary small lattice spacings.

While the project described in the present paper was being finalized, a determination of FV corrections to the masses of composite particles was presented in (*32*), up to $\mathcal{O}\left(1/L^4\right)$. In that paper, the authors use an elegant approach based on heavy-hadron effective field theories. Nevertheless, because of the importance of the infinite-volume (IV) extrapolation in studies involving QED and because the precision of our results depends critically on knowing the value of the coefficients of the $1/L$ and $1/L^2$ corrections, we have chosen to outline our independent derivation of Eq. (S26) in the present section. Although we work out the corrections only up to $\mathcal{O}\left(1/L^2\right)$, our approach can be considered more general, as it is based on general field theoretical considerations and not on a specific effective theory.

Note that Eq. (S26) implies that, to $\mathcal{O}\left(1/L^2\right)$ at least, composite, neutral particles receive no FV corrections. According to (*32*), these only appear at $\mathcal{O}\left(1/L^4\right)$ for such particles.

Here we consider a scalar particle of mass m and charge qe. The fermionic case is entirely analogous, but is technically more involved because of spin. As in the point-particle case, power-like FV corrections to the mass of composite particles arise from the exchange of a massless photon around the spacetime torus. The modification in the case of composite particles comes from the fact that the photons couple to a state which has internal structure. The coupling, that enters in the self-energy calculation, is described by the connected scalar-scalar-photon-photon vertex $G(k';p,k)$, with the momentum of the outgoing scalar equal to $p+k-k'$. At $\mathcal{O}\left(\alpha\right)$ it can be decomposed as shown in Fig. S5. $D(p)$ is the full, re-summed scalar propagator. $\Gamma_\mu(p,k)$ is the one-particle irreducible (1PI) scalar-scalar-photon, three-point vertex function, with the momentum of the outgoing scalar equal to $p+k$. $\tilde{\Gamma}_{\mu\nu}(k';p,k)$ is the 1PI scalar-scalar-photon-photon, four-point vertex function, with the momentum of the outgoing scalar equal to $p+k-k'$. These functions contain no photon lines and they are $\mathcal{O}\left(\alpha^0\right)$. They are all infinite-volume, renormalized quantities. They satisfy the following WTIs:

$$k_\mu\Gamma_\mu(p,k)/q = D^{-1}(p+k)-D^{-1}(p) \tag{S27}$$

$$-k_\mu k_\nu\tilde{\Gamma}_{\nu\mu}(k;p,k)/q^2 = D^{-1}(p+k)+D^{-1}(p-k)\,, \tag{S28}$$

where again q is the charge of the particle in units of e.

The FV corrections to the mass are obtained by taking the difference of the particle's on-shell self-energy in finite and infinite volumes. For a scalar particle at $\mathcal{O}\left(\alpha\right)$, this yields:

$$\Delta m^2(T,L)\equiv m^2(T,L)-m^2=-\frac{e^2}{2}\left[\sum\!\!\!\!\!\!\int_k^{\prime}-\int\frac{d^4k}{(2\pi)^4}\right]\frac{1}{k^2}G(k;p=im,k)= \tag{S29}$$

$$=-\frac{e^2}{2}\left[\sum\!\!\!\!\!\!\int_k^{\prime}-\int\frac{d^4k}{(2\pi)^4}\right]\frac{1}{k^2}\{2\Gamma_\mu(p+k,-k)D(p+k)\Gamma_\mu(p,k)+\tilde{\Gamma}_{\mu\mu}(k;p,k)\}_{p=im},$$

where the sum is the one defined by Eq. (S9). Since the contributions to the pole mass are gauge independent, we work in the simpler Feynman gauge.

In Eq. (S29) an important step has already been taken: the propagator D and the vertices Γ_μ and $\tilde{\Gamma}_{\mu\nu}$ have been replaced by their IV counterparts. This is because, for the kinetics of the self-energy diagrams these functions differ only by terms that fall off exponentially with the size of the box. For the propagator of a stable particle with an inflowing energy smaller than that of any possible on-shell, multi-particle state, this was shown by Lüscher (*42*). In the same paper the analiticity of the three-point and four-point 1PI vertex functions and the associated Feynman-integrands is proven (Theorem 2.3), assuming again, that the inflowing energy is smaller than that of any possible multi-particle states. The FV vertices are obtained by replacing the momentum integrals, which appear in the loop expansions of their IV counterparts, by momentum sums. The analyticity proof applies also in FV. Combined with the well known properties of Fourier transforms, this guarantees that the FV corrections to the vertex functions, computed using Poisson's formula, fall off faster than any inverse power of size of the box. The same arguments apply to the 1PI vertices Γ_μ and $\tilde{\Gamma}_{\mu\nu}$ considered here.

No regularization is required in Eq. (S29), because as we now argue, the ultraviolet divergences cancel in the difference of the sum and the integral. First, if we replaced the momentum sum by the one, which corresponds to QED_{TL}, the difference must be finite. This is because in QED_{TL} only the single $k = 0$ mode is removed from the momentum sums, and these sums define the integral in the $L, T \to \infty$ limit. Furthermore the difference between QED_L and QED_{TL} one loop self-energies is a one-dimensional integral over k_0 with $\vec{k} = 0$, which is ultraviolet convergent in the point-like case, because of the $1/k_0^2$ suppression from the photon propagator. Since the form factors, which arise in the vertex functions of composite particles are expected to suppress ultraviolet modes, this integral will be ultraviolet convergent also for composite particles. From this reasonable assumption it follows that the difference of the one-loop self-energies in IV QED and QED_L in Eq. (S29) is free of ultraviolet divergences.

To obtain the desired asymptotic expansion of $\Delta m^2(T, L)$ in powers of $1/L$, we apply the same Poisson formula as the one used in Eq. (S15). Power-law volume corrections result from infrared singularities in the integrand. We saw that $(p^2 + m^2)D(p)$, $\Gamma_\mu(p, k)$ and $\tilde{\Gamma}_{\nu\mu}(k'; p, k)$ are analytic functions of their arguments in the domain of integration of Eq. (S29). Thus, the only singularities that can arise are those corresponding to the poles of the free photon propagator, $1/k^2$, and of the free scalar propagator, $D_0(p + k) = 1/(2k{\cdot}p + k^2)$, for $p^2 = -m^2$. These singularities show up in the limit $\vec{k} \to 0$. It should be noted that there is an additional source of power-law corrections in Eq. (S29): the subtraction of the $\vec{k} = 0$ contribution in the sum over photon momenta.

Dimensional arguments suggest that corrections up to and including $1/L^2$ can be obtained by expanding the numerator of the $\Gamma_\mu \cdot D \cdot \Gamma_\mu$ term in Eq. (S29) to $\mathcal{O}(k_\mu)$ and that of the $\tilde{\Gamma}_{\mu\mu}$ term to $\mathcal{O}(k_\mu^0)$. Setting $p^2 = -m^2$, we perform this expansion around the on-shell point $k^2 = 0$ and $(p + k)^2 + m^2 = P{\cdot}k = 0$, with $P \equiv 2p + k$, considering only terms up to the desired order in k_μ at the end of the calculation. This expansion can be performed because these functions are analytic in the domain of integration of Eq. (S29). For the derivation, it is actually convenient to push the expansion to one order higher in k_μ, as it allows to include the full point-particle contribution. Note that we are expanding here the IV functions, and can therefore use all of the properties of these functions in infinite volume, including Lorentz covariance.

We begin with the propagator

$$D(p + k) \;=\; \frac{1}{P{\cdot}k}\left[1 - (P{\cdot}k)\delta D(P{\cdot}k)\right]^{-1}, \tag{S30}$$

where $\delta D(P{\cdot}k)$ is the full self-energy of the composite scalar at $e = 0$. It can be expanded as

$$D(p + k) = \frac{1}{P{\cdot}k} + \delta D(0) + (P{\cdot}k)\{\delta D(0)^2 + \delta D'(0)\} + \mathcal{O}\left((P{\cdot}k)^2\right) . \tag{S31}$$

The two vertex functions are then decomposed into linear combinations of invariant functions using Lorentz covariance and the properties of these functions under discrete spacetime symmetry transformations. Substituting the invariant decompositions into the WTIs of Eqs. (S27) and (S28), expanding the resulting equations to the desired order in k_μ and matching order by order yields:

$$\Gamma_\mu(p, k)/q \;=\; P_\mu\left[F(k^2) - \delta D(0)(P{\cdot}k) - \delta D'(0)(P{\cdot}k)^2\right] - k_\mu F'(0)(P{\cdot}k) + \mathcal{O}(k^3) , \tag{S32}$$

$$\tilde{\Gamma}_{\nu\mu}(k; p, k)/q^2 \;=\; -2\delta_{\mu\nu} + 8p_\mu p_\nu \delta D(0) + \mathcal{O}(k) , \tag{S33}$$

where $F(k^2)$ is the on-shell electromagnetic form-factor of the scalar. Its first derivative gives the electromagnetic radius $\langle r^2\rangle = -6F'(0)$. Then, replacing the propagator and vertex functions in Eq. (S29) by the above expansions, one finds that the terms in $\delta D(0)$ and $\delta D'(0)$ cancel and one obtains the point-particle result of Eq. (S26) plus a finite term proportional to $\langle r^2\rangle/L^3$ and a residual

$$\Delta m^2_{\mathrm{res}}(T, L) = -\left[\sum_k\!\!\!\!\!\!\int^{\prime f} - \int \frac{d^4k}{(2\pi)^4}\right] \frac{1}{m^2k^2}\left\{\frac{k^2(p{\cdot}k)f_1 + (p{\cdot}k)^3 f_2/m^2}{2k{\cdot}p + k^2} + k^2 f_3 + (p.k)^2 f_4/m^2\right\} . \tag{S34}$$

In this expression, the dimensionless, nonperturbative functions $f_i = f_i(p{\cdot}k, k^2)$, $i = 1, \cdots, 4$ must fall off fast enough as $|k_\mu| \to \infty$ for this residual to be finite. Indeed, as discussed above, $\Delta m^2(T, L)$ is finite. Moreover, the terms already computed in the $1/L$ expansion are also finite and by definition these functions are such that $f_i(p{\cdot}k, k^2) \underset{k_\mu \to 0}{\longrightarrow} \text{constant}$. Note that we have used crossing symmetry to guarantee that contributions from $\tilde{\Gamma}_{\mu\mu}(k; p, k)$ in the brackets begin at $\mathcal{O}(k^2)$.

Now, performing the change of variables $\vec{k} = 2\pi\vec{n}/L$, $\vec{n} \in \mathbb{Z}^3$, and taking the IV limit, we find:

$$\Delta m^2_{\text{res}}(T, L) \underset{T,L\to+\infty}{\sim} \frac{1}{m^2 L^3} \left[\int \frac{dk_0}{2\pi} \left\{ \frac{m\, g_1(k_0)}{k_0 + 2im} + g_2(k_0) \right\} \right] \times \left[\sum_{\vec{n}\in\mathbb{Z}^{3*}} - \int d^3n \right] 1 \,, \tag{S35}$$

where the dimensionless, nonperturbative functions $g_i(k_0)$, $i = 1, 2$, are derived from the $f_j(p{\cdot}k, k^2)$, $j = 1, \cdots, 4$, with $\vec{k} = 0$. These functions have a finite limit as $k_0 \to 0$ and they fall off fast enough at large k_0 to guarantee that the integral over k_0 is a finite number. Moreover, the difference between the sum and integral of 1 over $\vec{n}$ is -1. Therefore Δm^2_{res} is $\mathcal{O}\left(1/L^3\right)$.

The spinor case proceeds in an entirely analogous way. The end result is that the first two FV corrections to the mass of a composite particle are identical for spin-0 and $1/2$ particles, and are equal to those of point particles, as given in Eq. (S26). We use this result throughout our numerical analysis of mass splittings.

In addition, in performing the expansions in Eqs. (S32) and (S33), it is clear that the next terms leading to corrections of order $1/L^3$ or higher depend on the structure and spin of the particles. This confirms the picture obtained in (*32*).

5.1 Comparison to numerical QCD+QED results

Here we compare the finite-volume behavior of hadron masses to the analytical findings of the previous subsections.

Let us begin with the kaon mass, whose volume dependence is shown in Fig. 1 of the main text. Panel A shows the neutral kaon mass as a function of the box size. For the plot we used four ensembles which differ only in box size, all other lattice parameters being the same (see Sec. 7). There is no significant volume dependence in accordance with Eq. (S26), since $q = 0$. In contrast, the charged kaon mass is strongly dependent on the volume as can be seen in panel B which displays the mass squared difference of the two particles. The four different volumes cannot be described solely by the point-particle behavior, which includes up to $1/L^2$ terms. This is a consequence of the compositeness of the kaon and is nicely shown by our data. If we add an $1/L^3$ term to the formula, a nice fit quality can be achieved, shown as the solid line on the plot. The $1/L$ and $1/L^2$ terms of the fit correspond to dashed and dotted lines.

Fig. S6 shows the finite-volume dependence of various baryon mass splittings. The precision in these channels is considerably smaller than for the case of the kaon. All of them can be nicely described by the universal behavior of Eq. (S26) and the higher order term $1/L^3$ is always insignificant. The first panel is the splitting of the Σ baryons, $\Delta M_\Sigma = M_{\Sigma^-} - M_{\Sigma^+}$. There is no significant finite-volume dependence at all, since the absolute value of the two particles' charges are the same. The other panels show splittings, where the charged squared difference is non-zero. The dashed curves correspond to the universal fits. Since the mL combination is rather large for baryons, even the $\mathcal{O}(1/L^2)$ part is negligibly small.

6 Algorithms

In this Section we describe the implementation of our updating algorithm. The main novelty is the acceleration of the HMC by means of Fourier transformation, which is essential to obtain a reliable update of the

photon field. Another algorithmic feature, which is taken from the literature, is the combined use of multi-grid techniques and all-mode averaging (this will be discussed in Sec. 9).

6.1 Hybrid Monte Carlo

We use standard techniques to generate gauge configurations, including the hybrid Monte-Carlo (HMC) algorithm (*46*) with even-odd preconditioning (*47*), multiple time-scale integration aka. "Sexton-Weingarten scheme" (*48*), mass preconditioning aka. "Hasenbusch trick" (*49*), Omelyan integrator (*50*), Rational Hybrid Monte Carlo acceleration with multiple pseudofermions (*51*). Configurations are separated by ten unit-length trajectories.

In previous isospin symmetric simulations, the two lightest quarks could be bundled together and a standard Hybrid Monte Carlo was applicable. Now that all the quarks have different charges and masses, each of them has to be included with a separate fermion determinant. This causes an extra difficulty, because the sign of $\mathrm{det}D$ can become negative for small quark masses when Wilson fermions are used. If this happens, the common update algorithms break down. However if the spectral gap of the hermitian $\gamma_5 D$ operator is large enough, then the negative sign configurations play no role in the simulation. This is the case, as one approaches the continuum or the infinite volume limits (*52*). In practice, if the lowest eigenvalue of the operator $\sqrt{D^\dagger D}$ has a non-vanishing separation from zero, then the number of negative-weight configurations is exponentially small in that separation and the standard algorithms can be safely used. If there are no negative sign configurations, then the relation

$$\mathrm{det}D = \mathrm{det}\sqrt{D^\dagger D}$$

holds and the simulation can be straightforwardly done with the RHMC algorithm. The RHMC also needs the lowest eigenvalue of the $\sqrt{D^\dagger D}$ operator to be determined, in order to get a suitable rational approximation for the (inverse) square root function. The small eigenvalues are obtained using the Krylov-Schur algorithm. Our implementation is based on the program package, which is described in Ref. (*53*). In order to split the determinant of the fermion matrix D into an infrared and ultraviolet part, we combine the RHMC algorithm with the Hasenbusch trick:

$$\mathrm{det}D = \mathrm{det}\frac{1}{\sqrt{1+\rho^2/(D^\dagger D)}} \cdot \mathrm{det}\sqrt{D^\dagger D + \rho^2},$$

and integrate the first/second determinant on a coarse/fine timescale. A similar combination of the two techniques was already proposed in the staggered fermion formulation (*54*).

The absence of the negative-weight configurations and the validity of the absolute-value approximation can only be justified *a posteriori*. This validation is shown in Figure S7. The eigenvalue distributions are sufficiently bounded away from zero, from which we conclude that our simulations are safe from problems induced by eigenmodes becoming too small or even changing signs during the RHMC integration, for the parameter values considered in this work. We also monitored the fermion force for dangerous spikes, which would signal eigenmodes approaching zero and being repelled by a large force. We do not see any such spikes in the time history even in our potentially most affected ensemble.

We check that the violation of the HMC energy ($\mathcal{H}$) conservation is small and that the condition $\langle \exp(-\Delta\mathcal{H}) \rangle = 1$ is always true within errors.

Another important aspect that requires special attention is autocorrelations of physical observables. Since we generate a Markov Chain with our HMC simulation algorithm, subsequent measurements are correlated. The most strongly autocorrelated quantity is typically the topological charge. Its autocorrelation becomes very large for lattice spacings much smaller than ≈ 0.05 fm and our choice of boundary conditions. We calculated the autocorrelation of the topological charge using the openly available package UWerr (*55*). To define the charge the Wilson-flow was used to smoothen the configuration. Even on the potentially most affected ensemble with smallest lattice spacings the integrated autocorrelation time was about 50 trajectories.

6.2 Updating the photon field

The HMC has to generate photon fields that fulfill the zero mode constraint, Eq. (S3). Since the constraint is linear in the fields, the implementation of such HMC is straightforward: one has to require that the HMC momenta also satisfy the same constraint. This has to be enforced by removing the zero-mode part of the usual HMC force in each microcanonical step.

Due to the vanishing mass of the photon and the correspondingly large correlation lengths a standard HMC integration of the photon fields results in large autocorrelation times. Even for a pure photon simulation we observed that the 1x1 compact plaquette does not thermalize after several thousand trajectories, see Fig. S8. To eliminate this we use a Fourier accelerated HMC. Let us first discuss the case of the pure photon theory. In the standard HMC we introduce momenta $\Pi_{\mu,k}$ for all $A_{\mu,k}$'s and add a kinetic term $|\Pi_{\mu,k}|^2/(2m)$ to the photon action, Eq. (S5). The mass parameter is usually chosen as $m = 1$, but one can also allow it to be k-dependent, leading to the following "Hamiltonian":

$$\mathcal{H} = \frac{1}{2TL^3} \sum_{\mu,k} \left\{ |\hat{k}|^2 |A_{\mu,k}|^2 + \frac{|\Pi_{\mu,k}|^2}{m_k} \right\}.$$

At the beginning of an HMC trajectory we generate new momenta with a distribution

$$\exp(-|\Pi_{\mu,k}|^2/(2m_k))$$

and then integrate the canonical equations obtained from the above Hamiltonian until time t. The Hamiltonian is just a set of uncoupled harmonic oscillators, therefore after time t the photon field will be

$$A_{\mu,k}(t) = A_{\mu,k}(0)\cos(\omega_k t) + \frac{\Pi_{\mu,k}(0)}{m_k \omega_k}\sin(\omega_k t)$$

with $\omega_k = |\hat{k}|/\sqrt{m_k}$. For any oscillator if we let it evolve for a quarter period ($\omega_k t = \pi/2$) then the first term becomes zero, so the $A_{\mu,k}$ field "forgets" its previous ($t = 0$) value. Since we have the freedom to choose the parameters m_k arbitrarily, setting $m_k = 4|\hat{k}|^2/\pi^2$ will eliminate the first term for all oscillators at $t = 1$, therefore producing an exact heatbath step with zero autocorrelation. This is demonstrated in Fig. S8, where the plaquette has the correct distribution starting with the first trajectory. The implementation of this algorithm requires Fourier transformations of the photon fields and momenta for every step of the HMC, but this turned out to be computationally negligible compared to Dirac operator inversions.

After including the coupling to quarks, the autocorrelation does not vanish any more, but since the coupling is small it is much smaller than with a standard HMC algorithm. In our runs the thermalization of the photon field always occurred after a few trajectories.

6.3 Cost issues

Compared to a calculation with physical quark masses in the degenerate 2 light flavor setup the 1+1 light flavor framework doubles the computational costs. Using two different electromagnetic couplings increases the cost by another factor of two. Fixing the up-down quark mass difference to near its physical value reduces the mass of the up quark by up to 40%, which slows down the inversions. The small mass region also requires large volumes to avoid deceleration due to the fluctuations of the smallest Dirac eigenvalue toward the origin. To guarantee that the mean of the distribution of this eigenvalue remains as many standard deviations away from zero as in our earlier studies, while decreasing the mass by 40%, requires increasing the volume by a factor of 2.8. About a factor of five larger statistics are needed to have the desired significance on the small neutron-proton mass difference. In addition a more costly analysis is required (the inversions were about twice

as expensive as the configuration generation, which gives a factor of three). All in all one has an increase in numerical cost by a factor of $2 \times 2 \times 1.7 \times 2.8 \times 5 \times 3 \sim 300$. This increase by two to three orders of magnitude is the reason why we did not attempt to extend our simulations all the way down to the physical value of the pion mass as we did in (*23*). However, as shown in (*16*) and confirmed here, simulations with pion masses down to $195\ \mathrm{MeV}$, as reached in the present work, are sufficient to reliably determine hadron masses and isospin splittings.

7 Ensembles

We generated gauge ensembles at several different lattice parameter values to be able to extrapolate to the physical quark mass point. The following strategy was applied to find parameter values in the desired range, following closely the method of Ref. (*56*). First we tuned, in $n_f = 3 + 1$ simulations, the common light quark (q) mass and the charm quark mass, so that the pseudo-scalar meson masses approximately took the following values:

$$M_{\bar{q}q} = 410\,\mathrm{MeV} \qquad \text{and} \qquad M_{\bar{c}c} = 2980\,\mathrm{MeV}.$$

The value for $M_{\bar{q}q}$ is approximately the average of the physical masses of the mesons in the meson-octet with non-vanishing third component of isospin. For the parameter tuning, we set the lattice spacing using the recently proposed w_0 scale (*57*), which is based on the Wilson-flow (*18*). The latter can be measured easily and with high precision even on a handful of configurations. In the exploratory runs, we took the value $w_0 = 0.1755$ fm, which was obtained using pure QCD simulations (*57*). For the final results, we used the physical Ω mass for setting the scale. In addition, we have redetermined the value of w_0 in our QCD+QED setup. Afterwards we applied the method of Ref. (*56*): in a series of $n_f = 1 + 1 + 1 + 1$ simulations we decreased the masses of the up and down quarks and simultaneously increased the mass of the strange quark, while keeping the sum of the three masses constant, introducing a small splitting in the mass of the up and down quarks. This line points towards the physical point with a very good precision. In this way we obtained 27 ensembles with no electromagnetic interaction, $e = 0$. These are the **"neutral ensembles"**.

At the bare parameters of some of these "neutral ensembles" we switched on the electromagnetic interaction. Since this step introduces an extra additive renormalization, the spectrum always changed significantly. We tuned the masses by δ_u, δ_d and δ_s additive shifts, so that the connected meson masses for each quark took approximately the same values as in the starting "neutral ensemble" (*10*, *38*). In this way we obtained our 14 **"charged ensembles"**. Four of these ensembles have four different volumes, while all the other parameters are the same. They were used to study the finite size effects (see Sec. 5). Similarly there are four ensembles that have four different e values and approximately equal pseudoscalar masses. They can be used to test the e-dependence of the results (see Sec. 8).

Altogether we have 41 ensembles at four different gluon gauge couplings. Tab. S2 lists the gauge couplings, the corresponding lattice spacings and the bare charm mass. The lattice spacings are calculated using the Ω mass as an input. The detailed lists of our neutral and charged ensembles are given in Tabs. S3 and S4.

8 Electric charge renormalization

For small electromagnetic couplings there is a signal/noise problem. Operators that are symmetric under charge conjugation, including the particle propagators in this study, can only depend on e^2. However on a given gauge configuration the operators in general have a linear contribution in e, which vanishes after doing the ensemble average. The remaining signal is proportional to e^2, whereas the noise is of $O(e)$. This linear noise term can be set to zero in the case of quenched QED: one averages the observables at couplings $+e$ and $-e$, canceling

the linear term configuration by configuration (*35*). However in the dynamical case, this trick does not work, because the sign of the charge in the sea cannot be flipped while maintaining correlations in the gauge fields. Instead we decided to simulate at couplings that are larger than the physical one, so that the signal outweighs the noise. Then we interpolate the results, between those obtained with $e > 0$ and $e = 0$, to the physical value of the coupling.

Here we have to discuss another issue: how to define the physical value of the electromagnetic coupling? Ideally one would like to tune e so that, in the Thomson limit, i.e. at the scale of the electron mass, the renormalized coupling $\alpha = e_R^2/(4\pi)$ takes on its experimental value, $\alpha^{-1} = 137.036\ldots$. Evidently our lattices are not large enough to make measurements in this limit. Instead we define the renormalized coupling at a hadronic scale. The difference between the two is of order α^2. If α in the Thomson-limit takes the experimental value, the relative difference is on the percent level and can be neglected in our results, since it is much smaller than other errors.

For the definition of a renormalized coupling we use the Wilson flow (*18*). Now the photon fields are evolved with the gradient of the photon action. Similarly to the gluonic case, we measure along the flow:

$$E(\tau) = \frac{\tau^2 a^4}{TL^3} \sum_{\mu,\nu,x} F^{(A)}_{\mu\nu,x}(\tau) F^{(A)}_{\mu\nu,x}(\tau),$$

where $F^{(A)}_{\mu\nu,x}(\tau)$ is the photon field-strength tensor at a Wilson-flow time τ. Because $e^2 E(\tau)$ does not renormalize[4] and is proportional to e^2, we can use it to define the renormalized coupling at scale τ:

$$e_R^2(\tau) = Z(\tau) \cdot e^2 \quad \text{with} \quad Z(\tau) = \langle E(\tau) \rangle / E_{\text{tree}}(\tau)$$

The proportionality constant requires the tree-level value $E_{\text{tree}}(\tau)$. In infinite volume and in the continuum limit, the tree-level value does not depend on the flow time: it is $E_{\text{tree}} = 3/(32\pi^2)$, see Ref. (*58*). In finite volume and at finite lattice spacing with our definition of $F^{(A)}_{\mu\nu,x}$ the tree-level value becomes:

$$E_{\text{tree}}(\tau) = \frac{\tau^2}{TL^3} \sum_k \frac{\exp(-2|\hat{k}|^2\tau)}{a^2|\hat{k}|^2} \left[\sum_{\mu\neq\nu} (1 + \cos ak_\nu) \sin^2 ak_\mu \right]. \tag{S36}$$

Fig. S9 shows the renormalization constant $Z(\tau)$ as a function of flow time for the runs of the finite-volume scaling study. We plot $Z(\tau)$ using with the infinite-volume, continuum tree-level value with dashed lines and with Eq. (S36) with solid lines. Carrying out this step for all four volumes ($L = 24, \ldots, 80$) the resulting lines lie on top of each other. As one can see all volume dependence is removed by the improved tree-level formula.

The difference of the renormalized coupling at different flow times is an $O(\alpha^2)$ effect, which at the physical value of α is negligible in our calculation. However, in order to be conservative, we carry out the final analysis using two different τ's, so that we can estimate systematic effects arising from this choice. Since in the fermion-photon coupling the photon fields have undergone smearing, we also use the smeared photon field in the definition of $E(\tau)$. This changes the renormalized coupling only by lattice artefacts.

The renormalized coupling defined above has an important benefit in combination with simulating at larger than physical couplings. For the sake of the illustration let us consider the mass difference

$$\Delta M_\pi^2 = M_{\bar{d}u}^2 - \frac{1}{2}\left(M_{\bar{u}u}^2 + M_{\bar{d}d}^2\right),$$

which we call pion splitting. Here $M_{\bar{q}q}$ are masses of neutral mesons, whose propagator includes only quark-connected diagrams. Using partially quenched chiral perturbation theory one can show that, at leading order

[4]We use a normalization of the photon field where e is not in the photon action but in the covariant derivative.

in isospin breaking, this quantity is purely electromagnetic: there is no contribution from the quark mass difference (*10*, *59*). At the same time this is the isospin splitting most precisely obtained from our simulations and therefore any deviation from the linear behavior in α can be observed. Fig. S10 shows the data of the four charged ensembles, that have four different e values and approximately equal pseudoscalar masses. The solid lines are polynomial fits to the data without a constant term. As a function of the bare coupling (open symbols) the splitting has a significant curvature. If we plot it as a function of the renormalized coupling (filled symbols), we see no deviation from linear behavior. Here an illustrative scale was chosen at $(8\tau)^{-1/2} = 400\,\mathrm{MeV}$. This observation is true for other observables considered in this paper: no significant $O(\alpha^2)$ contribution can be seen for the whole coupling range, which is also true for the other scale choices used in the final analysis.

9 Extracting hadron masses

To calculate hadron propagators we fix the smeared photon and gluon fields to Coulomb gauge. We use local hadron operators, in which the quark fields are Gaussian smeared, with a smearing radius of about 0.3 fm. We observe, as in many other studies that reducing the noise in hadron propagators can be efficiently done by using many different quark sources per gauge configuration. In the case of the nucleon propagator we use several hundred different source positions. For these propagator calculations we use the 2-level multi-grid approach of Ref. (*60*) and also the variance reduction technique of Ref. (*61*).

We fold the propagators around the midpoint of the lattice in time and fit them with standard one-state functions. To obtain the mass splittings our method is to jointly fit the isospin partners with the mass difference, the average mass and the two amplitudes as fit parameters (*10*). Fitting the propagators separately and subtracting the fitted values yields results that are consistent with the joint fit method and the statistical errors are of similar size. All plateau fits take into account the correlation between the timeslices and isospin partners. In order to have a stable correlation matrix we only fit ten time slices at once.

Selecting the fit interval for a given hadronic channel is a highly non-trivial task. In order to get rid of the excited states one wants to take large values for the timeslice, where the mass fit starts (t_{min}). However, taking values that are too large results in suboptimal signal/noise ratios. Since our goal is to reach precisions on the sub-permil level, a reliable, robust and controlled method is needed to choose the proper fit intervals.

The most popular way to select a good fit range is to calculate the correlated χ^2 of a mass fit and use ranges for which the χ^2/dof is about one. Note that this procedure is not fully satisfactory. The values of the χ^2/dof for a given hadron channel should follow instead the χ^2 distribution. Thus we want to determine the probability that from a given t_{min} the χ^2/dof values follow the expected distribution (or equivalently the corresponding goodnesses of the fits are uniformly distributed). Since we have 41 independent ensembles it is possible to look at the distributions of the χ^2/dof for various t_{min} values. Fig. S11 shows the expected cumulative probability distributions of the fit qualities (which has to be linear) and the measured distribution. Both the nucleon and Ξ channels are depicted. As expected the larger the t_{min} the closer the measured distribution is to the expected one.

The standard way to statistically interpret Fig. S11 is to carry out a Kolmogorov-Smirnov analysis (*62*). Thus one determines the maximum value of the absolute difference between the expected and measured cumulative probability distributions (D). This is then used to define a significance level or probability (P) that the measured distribution can indeed be one originating from the expected uniform distribution. We start our fits from t_{min} values, for which this probability is larger than 0.3. In addition, to estimate the possible contribution of excited states, we use another fit range starting from one timeslice later. (Since the excited states have even smaller contributions for fit ranges which start even later, one might use them too. However, due to the increasing noise this provides no new information). Obviously, for different hadronic channels the proper t_{min} values are different. They are collected in Tab. S5.

10 Reaching the physical point

Our present procedure closely follows our earlier studies of the hadron spectrum (*16*). We consider two different paths, in bare parameter space, to the physical mass point and continuum limit. These two techniques correspond to two different ways of normalizing hadron masses and their isospin splittings for a set of fixed parameters. For both methods we follow several strategies for the extrapolation to the physical mass point and to the continuum limit. The physical mass point is defined by $M_{\pi^+} = 139.570\,\mathrm{MeV}$, $M_{K^+} = 493.68\,\mathrm{MeV}$, $M_{K^0} = 497.61\,\mathrm{MeV}$, $M_{D^0} = 1864.9\,\mathrm{MeV}$ and the electromagnetic coupling in the Thomson-limit $\alpha^{-1} = 137.036$ (*2*). A particularly important quantity of the analysis is the kaon mass splitting. Its experimental value is well known and it can be measured with high accuracy on the lattice. We use the value ΔM_K^2=$M_{K^0}^2$-$M_{K^+}^2 = 3896\,\mathrm{MeV}^2$. For both methods the lattice spacing at a given g^2 was determined by the Omega baryon mass at the physical mass point (its experimental value is $M_\Omega = 1672.4\,\mathrm{MeV}$). The values for a are given in Tab. S2.

10.1 Normalizing hadron masses

We call the two methods of normalizing the hadron masses and their splittings: a. "mass-independent scale setting" and b. "ratio method".

a. The mass-independent scale setting is the more traditional technique. One takes the lattice spacing from Tab. S2 and expresses all the hadron masses and their isospin splittings in physical units using this a. Thus, the mass values measured on the lattice are divided/normalized by the lattice spacing. The splittings are then extrapolated to the physical mass point and continuum limit as explained below. Since this is the more traditional technique we explain the extrapolations and interpolations using this normalization prescription (for the ratio method the corresponding formulae can be obtained straightforwardly).

b. The ratio method is motivated by the fact that in QCD+QED one can calculate only dimensionless combinations of observables, e.g. mass ratios (or isospin splittings normalized by some hadron mass). Furthermore, in such ratios cancellations of statistical uncertainties and systematic effects may occur. The method uses the mass ratios M_{π^+}/M_Ω, M_{K^+}/M_Ω, M_{K^0}/M_Ω, M_{D^0}/M_Ω as input parameters and expresses the splittings normalized by M_Ω.

10.2 Determining the isospin splittings

There are two sources of isospin violation: electromagnetism and the mass difference of the up and down quarks. The isospin splittings are expanded in powers of the renormalized coupling $\alpha = e_R^2/(4\pi)$ and the quark mass difference $\delta m = m_d - m_u$. This expansion is expected to converge rapidly at the physical values of these parameters. However we work at somewhat larger electromagnetic coupling values. As was shown in Sec. 8, higher-order terms in α are negligible if we use a coupling defined at a hadronic scale. So we work only with the linear terms, i.e. in $\mathcal{O}(\delta m, \alpha)$. In linear order of isospin breaking, an arbitrary mass splitting ΔM_X can be written as a sum of two terms:

$$\Delta M_X = F_X(M_{\pi^+}, M_{K^0}, M_{D^0}, L, a) \cdot \alpha + G_X(M_{\pi^+}, M_{K^0}, M_{D^0}, a) \cdot \Delta M_K^2, \tag{S37}$$

where F_X, G_X are functions of pseudoscalar meson masses and the lattice spacing. The QED part of ΔM_K^2 has an L dependence but this can be absorbed into the L dependence of F_X. The charged particle masses that enter in Eq. (S37), are already corrected for the universal finite-size effect Eq. (S26). Higher order polynomial finite-size effects, starting with $1/L^3$, are still allowed in the electromagnetic part F_X. Since ΔM_K^2 is the mass squared difference of a neutral and charged particle, the second term of Eq. (S37) mixes the strong and electromagnetic isospin breaking effects. This is no problem of principle and good quality fits can be achieved

in practice. An alternative procedure, where ΔM_Σ was used instead of ΔM_K^2 was also performed and yielded compatible results.

We choose several different functional forms for F_X and G_X. The difference in the results coming from different choices will be part of the systematic error (see Sec. 12). In all cases we have in F_X and G_X a constant term and a pion-mass dependence parametrized by M_π^2. F_X always contains a term proportional to $1/L^3$ and G_X a term parameterizing the lattice spacing dependence, either with g^2a or a^2. Note that our action has a leading cutoff dependence of $O(g^2a)$, but in order to give a more conservative estimate of the pertaining systematic error we added the alternative ansatz. Optionally we add $M_{K^0}^2$, $M_{\pi^+}^4$ and M_{D^0} dependencies. In many cases, especially for the neutron-proton mass difference, the coefficients corresponding to all but the constant terms are consistent with zero.

10.3 Separating QED and QCD effects

Here we show how to separate the electromagnetic and strong contributions in an isospin splitting, i.e.

$$\Delta M_X = \Delta_{\mathrm{QED}} M_X + \Delta_{\mathrm{QCD}} M_X,$$

where the first and second terms are proportional to α and δm, respectively. It is clear, that it is sufficient to decompose the kaon mass squared difference,

$$\Delta M_K^2 = \Delta_{\mathrm{QED}} M_K^2 + \Delta_{\mathrm{QCD}} M_K^2,$$

since the separation for the other splittings can be obtained from this decomposition through Eq. (S37). Any such decomposition has an ambiguity of $\mathcal{O}(\alpha\delta m)$, which is NLO in isospin breaking, i.e. one order higher than the one at which we work. This arises from the scheme dependence in the definition of δm. In (*10*) we proposed a separation based on the difference of connected meson masses, $\Delta M^2 = M_{dd}^2 - M_{\bar{u}u}^2$. Using chiral perturbation theory (*59*), ΔM^2 can be shown to be proportional to δm up to corrections of $\mathcal{O}(\alpha^2, \alpha\delta m, (\delta m)^2, \alpha m_{ud})$, which are also NLO if we make the very reasonable assumption that $\mathcal{O}(m_{ud}) \sim \mathcal{O}(\delta m)$. Though this technique is sufficient for the electroquenched approximation used in previous work, it has a drawback here, in that these connected mesons are not in the spectrum of the full theory. One could use this approach in a partially quenched framework, but we want to avoid this in the present *ab initio* setup.

Here we show that, within the precision of our calculation, an alternative separation can be applied. The method is based only on observed particles. Since the Σ^+ and Σ^- have the same charge squared and the same spin, their electromagnetic self-energies are identical at $\mathcal{O}(\alpha)$, if these particles are assumed to be point-like. In that case, $\Delta M_\Sigma = M_{\Sigma^-} - M_{\Sigma^+}$ comes purely from δm. Because these particles are composite, there are corrections to this result. We now argue that they are no larger than the $0.2\,\mathrm{MeV}$ overall uncertainty on our determination of ΔM_Σ. In the seminal quark mass review (*63*), Gasser and Leutwyler find $\Delta M_\Sigma = 0.17(30)\,\mathrm{MeV}$ by evaluating the Cottingham formula (for a recent discussion see (*64*)). Using $\Delta M^2 = 0$ to define the electromagnetic contribution to splittings we found $\Delta_{\mathrm{QED}} M_\Sigma = -0.08(12)(34)\,\mathrm{MeV}$ in our electroquenched lattice computation (*10*), which is consistent with 0, albeit within relatively large uncertainties. Repeating this analysis in the fully unquenched theory we find $\Delta_{\mathrm{QED}} M_\Sigma = 0.18(12)(6)\,\mathrm{MeV}$, again consistent with 0 within a little more than one standard deviation. All of these determinations provide quantitative evidence that our proposed QCD-QED separation convention, i.e. using

$$\Delta_{\mathrm{QED}} M_\Sigma = 0,$$

is adequate within the precision of the present paper. This choice establishes a first benchmark against which future results can be compared. It can certainly be refined when calculations, such as the one presented here, become more precise.

Applying the fit formula Eq. (S37) to the ΔM_Σ splitting and carrying out the fit procedure as described in the next section we obtain

$$\Delta_{QED} M_K^2 = -2250(80)(90)\ \text{MeV}^2, \tag{S38}$$

where the first error is statistical, the second is the systematic error. The separation for the other isospin splittings is based on this finding. The results are given in Table 1 in the main text.

10.4 The Coleman-Glashow relation

In addition to the mass splittings, we also compute $\Delta_{\rm CG} = \Delta M_N - \Delta M_\Sigma + \Delta M_\Xi$, which parametrizes the violation of the Coleman-Glashow relation, $\Delta_{\rm CG} = 0$ (*30*,*65*). From quark exchange symmetries we know that at vanishing α the leading contribution to $\Delta_{\rm CG} \propto (m_s - m_d)(m_s - m_u)(m_d - m_u)$, see e.g. Ref. (*66*). At non-vanishing α there is a remnant $d \leftrightarrow s$ exchange symmetry yielding the leading order contributions

$$\Delta_{\rm CG} = \hat{F}_{\rm CG} \cdot \alpha \cdot (m_s - m_d) + \hat{G}_{\rm CG} \cdot (m_s - m_d)(m_s - m_u)(m_d - m_u). \tag{S39}$$

We now trade the quark mass differences for mass-squared differences of suitable pseudoscalars. We define $M_{sd}^2 = (M_{K+}^2 - M_{\pi+}^2)/2$, which to leading order is proportional to $(m_s - m_d)$. For a quantity, that is proportional to $(m_d - m_u)$ we take $M_{du}^2 = \Delta_{QCD} M_K^2$, where we use the convention for the separation of QCD and QED effects from the previous subsection. Finally we define $M_{su}^2 = M_{sd}^2 + M_{du}^2$ and fit

$$\Delta_{\rm CG} = F_{\rm CG}(L, a) \cdot \alpha \cdot M_{sd}^2 + G_{\rm CG}(a) \cdot M_{sd}^2 M_{su}^2 M_{du}^2, \tag{S40}$$

where we can include discretization terms $g^2 a$ or a^2 and finite-volume terms into the fit functions $F_{\rm CG}$ and $G_{\rm CG}$. A fit of our results to Eq. (S40) yields $\Delta_{\rm CG} = 0.00(11)(06)\,\text{MeV}$: the Coleman-Glashow relation is satisfied to high accuracy.

11 Akaike's information criterion

When we have several candidate models to fit our lattice results, e.g. several fit functions with a different number of parameters, or several fit ranges in Euclidean propagator fits, Akaike's information criterion (AIC) gives a guideline on how to weigh the results of the various models.

Let $\Gamma_1, \ldots, \Gamma_n$ denote the result of n independent measurements from the same (unknown) probability distribution $g^{(1)}(\Gamma)$, e.g. of the Euclidean propagator of a hadron measured on n independent configurations. Let $f^{(1)}(\Gamma|\theta)$ be the model which is used to approximate the true distribution $g^{(1)}(\Gamma)$, depending on p parameters, the parameter vector being denoted by θ.

Akaike proposed (*67*) to measure the distance of the model f from the true distribution g using the Kullback–Leibler divergence (*68*)

$$I(g, f(\theta)) = \int \mathrm{d}\Gamma\ g(\Gamma) \log\left(\frac{g(\Gamma)}{f(\Gamma|\theta)}\right) = \int \mathrm{d}\Gamma \left\{ g(\Gamma) \log\left[g(\Gamma)\right] - g(\Gamma) \log\left[f(\Gamma|\theta)\right] \right\}, \tag{S41}$$

where $g(\Gamma) = \prod_{k=1}^n g^{(1)}(\Gamma_k)$ and $f(\Gamma|\theta) = \prod_{k=1}^n f^{(1)}(\Gamma_k|\theta)$ denote the joint distributions of the n trials and $\mathrm{d}\Gamma = \prod_{k=1}^n \mathrm{d}\Gamma_k$ is the joint integration measure. When the n independent samples $\Gamma_1, \ldots, \Gamma_n$ are drawn from the model probability distribution $f(\Gamma|\theta)$, then in the $n \to \infty$ limit, the negative of this divergence gives the logarithm of the probability that the sample becomes distributed according to the distribution $g(\Gamma)$ (*69*). Therefore, when we have several different models, $f_m(\Gamma|\theta_m)$, where m is the index and θ_m are the parameters of the given model, then the weight

$$w_m = \frac{\exp\left[-I(g, f_m(\theta_m))\right]}{\sum_{m'} \exp\left[-I(g, f_{m'}(\theta_{m'}))\right]} \tag{S42}$$

for model m emerges naturally. That is, we are considering each model with a weight proportional to the probability that the given model reproduces the measured data.

Once we have the outcome Γ of a series of n independent measurements, in order to test the data against our models, we need to perform two steps. First, for each model m we need to find the parameter $\hat{\theta}_{m,\Gamma}$ that minimizes the function $I(g, f_m)$ in Eq. (S41). In the large n limit, under mild regularity conditions, the maximum likelihood estimate minimizes Eq. (S41) with probability one (*70*), therefore we take $\hat{\theta}_{m,\Gamma}$ to be the maximum likelihood estimate.

The second step is to estimate $I(g, f_m(\hat{\theta}_{m,\Gamma}))$. The first term of Eq. (S41) does not depend on the model, therefore it cancels out when the weights in Eq. (S42) are calculated. Thus, the quantity to be estimated is

$$J_m(\Gamma) = \int \mathrm{d}\Gamma' \; g(\Gamma') \; \log\left[f(\Gamma'|\hat{\theta}_{m,\Gamma})\right]. \tag{S43}$$

The integral cannot be directly performed, since the true probability distribution g is unknown. Nevertheless, Akaike (*67*) suggests to estimate Eq. (S43) with

$$-\frac{1}{2}\,\mathrm{AIC}_m(\Gamma) = \log\left[f(\Gamma|\hat{\theta}_{m,\Gamma})\right] - p_m, \tag{S44}$$

where p_m is the number of parameters in model m. If the model distribution is close to the true distribution, then in the large n limit

$$\int \mathrm{d}\Gamma \; g(\Gamma)\left[-\frac{1}{2}\,\mathrm{AIC}_m(\Gamma)\right] \approx \int \mathrm{d}\Gamma \; g(\Gamma)\left[J_m(\Gamma)\right], \tag{S45}$$

that is, in expectation Eq. (S43) and Eq. (S44) agree. Therefore, the estimation of $J_m(\Gamma)$ with $-\frac{1}{2}\,\mathrm{AIC}_m(\Gamma)$ is justified. For a concise derivation of Eq. (S45) see Ref. (*71*), or see Ref. (*72*) for a more detailed exposition. Finally, the results obtained from the various models have to be weighed using the Akaike weights (*73*)

$$w_m = \frac{\exp\left(-\frac{1}{2}\,\mathrm{AIC}_m(\Gamma)\right)}{\sum_{m'}\exp\left(-\frac{1}{2}\,\mathrm{AIC}_{m'}(\Gamma)\right)}. \tag{S46}$$

For models with normally distributed errors, Akaike's information criterion takes the simple form

$$\mathrm{AIC}_m(\Gamma) = \chi^2_m + 2p_m, \tag{S47}$$

where all additive constants independent of the models were dropped. That is, the AIC weight prefers the models with lower χ^2 values, but punishes the ones with too many fit parameters.

12 Final results and systematic errors

Hadron masses or mass splittings in the continuum limit at the physical point are extracted from lattice data following a two step procedure. First one extracts the hadron masses and splittings from Euclidean propagators which are then, in a second step, extrapolated and interpolated to the physical point in the continuum limit. The first step is already discussed in Sec. 9.

The second step of the analysis consists in performing the extra- and interpolations to the physical point. As described in the previous sections, we have some freedom in carrying out this procedure. First we can choose between mass-independent or ratio method (see Sec. 10.1). Second we can choose different parameterizations for ΔM_X (see Sec. 10.2). Third there are two different minimum times for the fits of the relevant Euclidean correlators (see Sec. 9). Fourth we use two different scales $(8\tau)^{-1/2} = 280\ \mathrm{MeV}$ and $(8\tau)^{-1/2} = 525\ \mathrm{MeV}$ to extract the renormalized electromagnetic coupling (see Sec. 8). Altogether, we obtain about $\mathcal{O}(500)$ fits which

are combined into a distribution using the Akaike's information criterion AIC (see Sec. 11). The systematic error is obtained from the variance of the distribution of different fit possibilities. Its mean gives the central value. The final numbers and systematic errors of Table 1 and Figure 2 of the main text are obtained using this procedure. Since estimating the systematic uncertainty is a non-trivial task, we repeated the analysis using instead of the AIC weights the fit-qualities as a weight or no weights at all. In all but one case, the results were within 0.2σ of the central values in Table 1 of the main text For $\Delta\Xi_{cc}$ without using any weights the difference was 0.7σ.

The complete procedure was repeated on 2000 bootstrap samples, which were then used to determine the statistical error.

In Fig. S12 we show the three mass differences of the light hadrons in one particular fit as the function of $g_r^2(a)a$, where $g_r(a)$ is the four-loop running coupling (*74*) at the scale of the lattice spacing. The points are the results of the different ensembles averaged over a common lattice spacing, which were transformed to the physical point in all parameters except the lattice spacing using the fit function. The line is the result of the fit.

In addition, three other, independent analyses were performed by different team members to ensure a robust determination of the results and the error estimation. We do not give the details of these procedures, we just mention that differences were e.g. one fits directly propagator ratios instead of the propagators themselves, jackknife is used instead of bootstrap to estimate statistical errors, one uses w_0 (*57*) instead of the Ω mass to set the scale, one assumes the validity of the Coleman-Glashow relation instead of extracting it from our results, one uses a statistical test based on the binomial distribution instead of the Kolmogorov-Smirnov criterion to decide on the correlator fit time-ranges, the different models are weighed uniformly instead of using the fit quality (like AIC), etc. One of the procedures was a completely blind analysis: the extracted mass differences were multiplied by a random number between 0.7 and 1.3, and the analysis was carried out by a group member who did not know the value of this transformation factor. At the end of his analysis the transformation factor was removed. Comparing the results of these procedures with the main one, we found complete agreement.

References:

(1) A. Kronfeld, *in 100 Years of Subatomic Physics, ed. by E. Henley and S. Ellis, Chap. 18, World Scientific* (2013).

(2) J. Beringer, *et al.*, *Phys.Rev.* **D86**, 010001 (2012).

(3) R. L. Jaffe, A. Jenkins, I. Kimchi, *Phys.Rev.* **D79**, 065014 (2009).

(4) D. J. Gross, F. Wilczek, *Phys.Rev.Lett.* **30**, 1343 (1973).

(5) H. D. Politzer, *Phys.Rev.Lett.* **30**, 1346 (1973).

(6) K. G. Wilson, *Phys.Rev.* **D10**, 2445 (1974).

(7) A. Duncan, E. Eichten, H. Thacker, *Phys.Rev.Lett.* **76**, 3894 (1996).

(8) T. Blum, *et al.*, *Phys.Rev.* **D82**, 094508 (2010).

(9) S. Basak, *et al.*, *PoS* **CD12**, 030 (2013).

(10) S. Borsanyi, *et al.*, *Phys.Rev.Lett.* **111**, 252001 (2013).

(11) G. de Divitiis, *et al.*, *Phys.Rev.* **D87**, 114505 (2013).

(12) S. Aoki, *et al.*, *Phys.Rev.* **D86**, 034507 (2012).

(13) T. Ishikawa, *et al.*, *Phys.Rev.Lett.* **109**, 072002 (2012).

(14) R. Horsley, *et al.*, *PoS* **Lattice2013**, 499 (2013).

(15) S. Durr, *et al.*, *Phys.Rev.* **D79**, 014501 (2009).

(16) S. Durr, *et al.*, *Science* **322**, 1224 (2008).

(17) Supplementary online material.

(18) M. Luscher, *JHEP* **1008**, 071 (2010).

(19) S. Aoki, *et al.*, *Phys.Rev.* **D79**, 034503 (2009).

(20) H. Ohki, *et al.*, *Phys.Rev.* **D78**, 054502 (2008).

(21) H.-W. Lin, *et al.*, *Phys.Rev.* **D79**, 034502 (2009).

(22) A. Bazavov, *et al.*, *Rev.Mod.Phys.* **82**, 1349 (2010).

(23) S. Durr, *et al.*, *JHEP* **1108**, 148 (2011).

(24) R. Baron, *et al.*, *JHEP* **1006**, 111 (2010).

(25) W. Bietenholz, *et al.*, *Phys.Rev.* **D84**, 054509 (2011).

(26) R. Arthur, *et al.*, *Phys.Rev.* **D87**, 094514 (2013).

(27) P. Fritzsch, *et al.*, *Nucl.Phys.* **B865**, 397 (2012).

(28) M. Hayakawa, S. Uno, *Prog.Theor.Phys.* **120**, 413 (2008).

(29) H. Akaike, *IEEE Transactions on Automatic Control* **19**, 716 (1974).

(30) S. R. Coleman, S. L. Glashow, *Phys.Rev.Lett.* **6**, 423 (1961).

(31) B. P. Mertens, A. S. Kronfeld, A. X. El-Khadra, *Phys.Rev.* **D58**, 034505 (1998).

(32) Z. Davoudi, M. J. Savage, *Phys.Rev.* **D90**, 054503 (2014).

(33) M. Luscher, P. Weisz, *Commun.Math.Phys.* **97**, 59 (1985).

(34) M. Gockeler, R. Horsley, P. E. Rakow, G. Schierholz, R. Sommer, *Nucl.Phys.* **B371**, 713 (1992).

(35) T. Blum, T. Doi, M. Hayakawa, T. Izubuchi, N. Yamada, *Phys.Rev.* **D76**, 114508 (2007).

(36) B. Sheikholeslami, R. Wohlert, *Nucl.Phys.* **B259**, 572 (1985).

(37) Milc collaborations public lattice gauge theory code. see http://physics.utah.edu/˜ detar/milc.

(38) A. Portelli, *et al.*, *PoS* **LATTICE2010**, 121 (2010).

(39) S. Capitani, S. Durr, C. Hoelbling, *JHEP* **0611**, 028 (2006).

(40) S. Durr, Z. Fodor, C. Hoelbling, T. Kurth, *JHEP* **0704**, 055 (2007).

(41) S. Necco, R. Sommer, *Nucl.Phys.* **B622**, 328 (2002).

(42) M. Luscher, *Commun.Math.Phys.* **104**, 177 (1986).

(43) E. W. Weisstein, *Integral Test. From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/IntegralTest.html* .

(44) P. Hasenfratz, H. Leutwyler, *Nucl.Phys.* **B343**, 241 (1990).

(45) A. Duncan, E. Eichten, H. Thacker, *Phys.Lett.* **B409**, 387 (1997).

(46) S. Duane, A. Kennedy, B. Pendleton, D. Roweth, *Phys.Lett.* **B195**, 216 (1987).

(47) T. A. DeGrand, P. Rossi, *Comput.Phys.Commun.* **60**, 211 (1990).

(48) J. Sexton, D. Weingarten, *Nucl.Phys.* **B380**, 665 (1992).

(49) M. Hasenbusch, *Phys.Lett.* **B519**, 177 (2001).

(50) T. Takaishi, P. de Forcrand, *Phys.Rev.* **E73**, 036706 (2006).

(51) M. Clark, A. Kennedy, *Phys.Rev.Lett.* **98**, 051601 (2007).

(52) L. Del Debbio, L. Giusti, M. Luscher, R. Petronzio, N. Tantalo, *JHEP* **0602**, 011 (2006).

(53) V. Hernandez, J. E. Roman, V. Vidal, *ACM Trans. Math. Software* **31**, 351 (2005).

(54) M. Clark, A. Kennedy, *Phys.Rev.* **D75**, 011502 (2007).

(55) U. Wolff, *Comput.Phys.Commun.* **156**, 143 (2004).

(56) W. Bietenholz, *et al.*, *Phys.Lett.* **B690**, 436 (2010).

(57) S. Borsanyi, *et al.*, *JHEP* **1209**, 010 (2012).

(58) M. Luscher, *PoS* **LATTICE2010**, 015 (2010).

(59) J. Bijnens, N. Danielsson, *Phys.Rev.* **D75**, 014505 (2007).

(60) A. Frommer, K. Kahl, S. Krieg, B. Leder, M. Rottmann, *SIAM J.Sci.Comput.* **36**, A1581 (2014).

(61) T. Blum, T. Izubuchi, E. Shintani, *Phys.Rev.* **D88**, 094503 (2013).

(62) W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing, Cambridge University Press* (2002).

(63) J. Gasser, H. Leutwyler, *Phys.Rept.* **87**, 77 (1982).

(64) A. Walker-Loud, C. E. Carlson, G. A. Miller, *Phys.Rev.Lett.* **108**, 232301 (2012).

(65) G. Zweig, *in Developments in the Quark Theory of Hadrons, ed. by Lichtenberg and Rosen, pp 22-101, Hadronic Press* (1980).

(66) R. Horsley, *et al.*, *Phys.Rev.* **D86**, 114511 (2012).

(67) H. Akaike, *2nd International Symposium on Information Theory*, B. Petrov, F. Csaki, eds. (Akademiai Kiado, Budapest, 1973), pp. 267–281.

(68) S. Kullback, R. Leibler, *Ann. Math. Stat.* **22**, 79 (1951).

(69) I. Sanov, *Sel. Transl. Math. Stat. Probab.* **1**, 213 (1961).

(70) P. Huber, *Proceedings of the fifth Berkley Symposium on Statistics* (1967), pp. 221–233.

(71) G. Kitagawa, W. Gersch, *Smoothness Priors Analysis of Time Series*, vol. 116 of *Lecture Notes in Statistics* (Springer, New York, 1996).

(72) S. Konishi, G. Kitagawa, *Information Criteria and Statistical Modeling*, Springer Series in Statistics (Springer, New York, 2008).

(73) H. Akaike, *The Statistician* **27**, 217 (1978).

(74) T. van Ritbergen, J. Vermaseren, S. Larin, *Phys.Lett.* **B400**, 379 (1997).

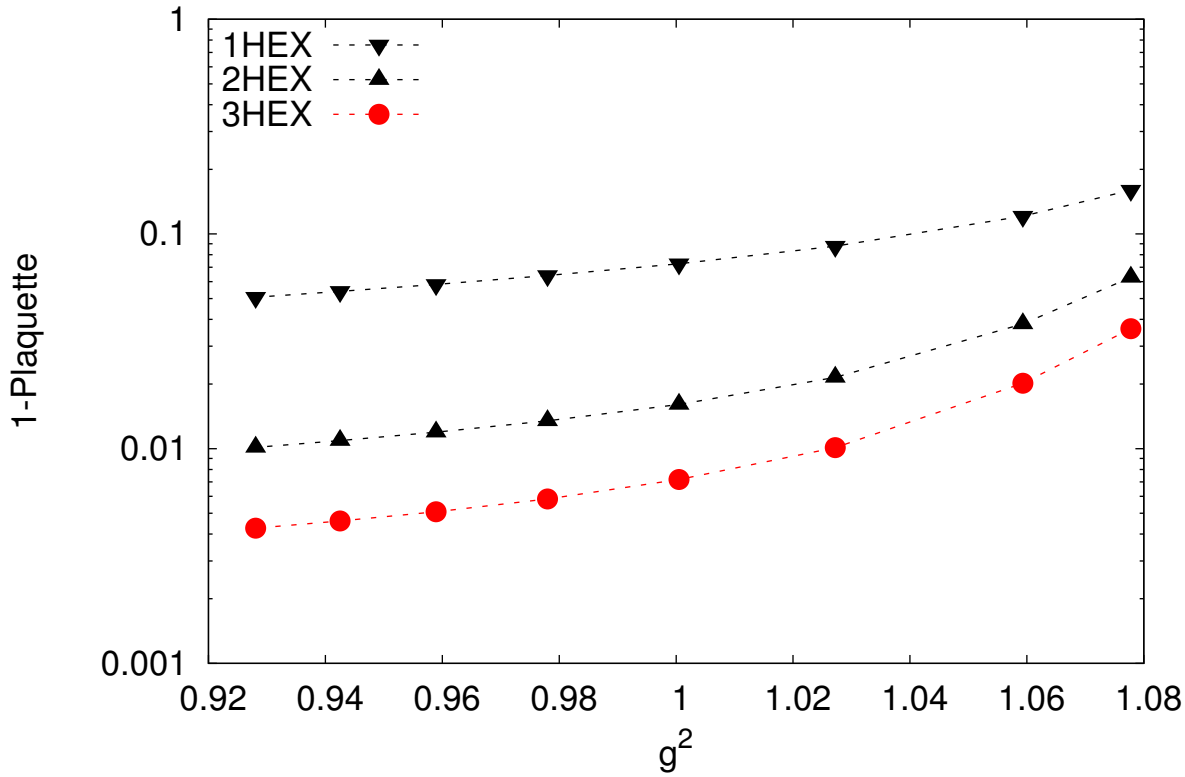

Figure S1: $1 - \langle U_{\Box} \rangle$ *for 1, 2, and 3 HEX smearing steps, using our preferred values for the parameters* $[\rho_1, \rho_2, \rho_3] = (0.22, 0.15, 0.12)$. *The lines connecting the points are to guide the eye.*

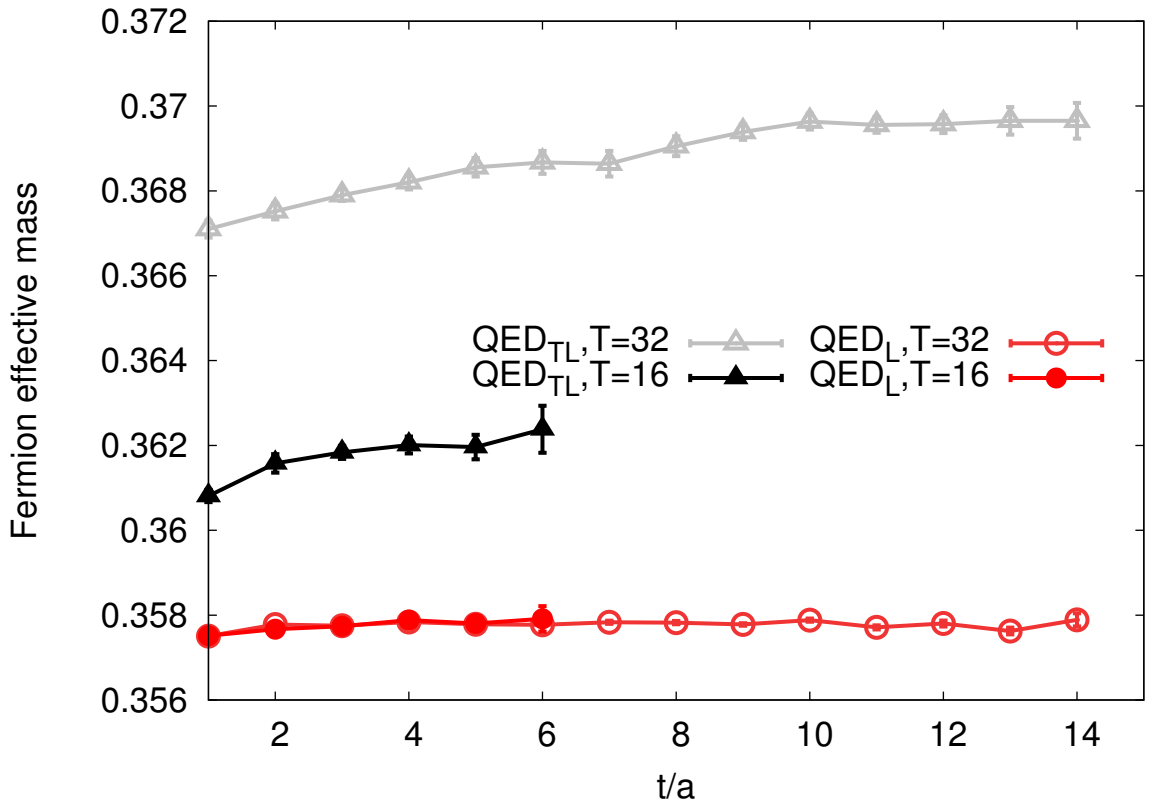

Figure S2: Quark effective masses in QED using different zero-mode subtractions. The four-dimensional zero-mode subtraction (QED_{TL}) produces masses that depend on T. The time-slice by time-slice removal (QED_{L}) makes effective masses well behaved.

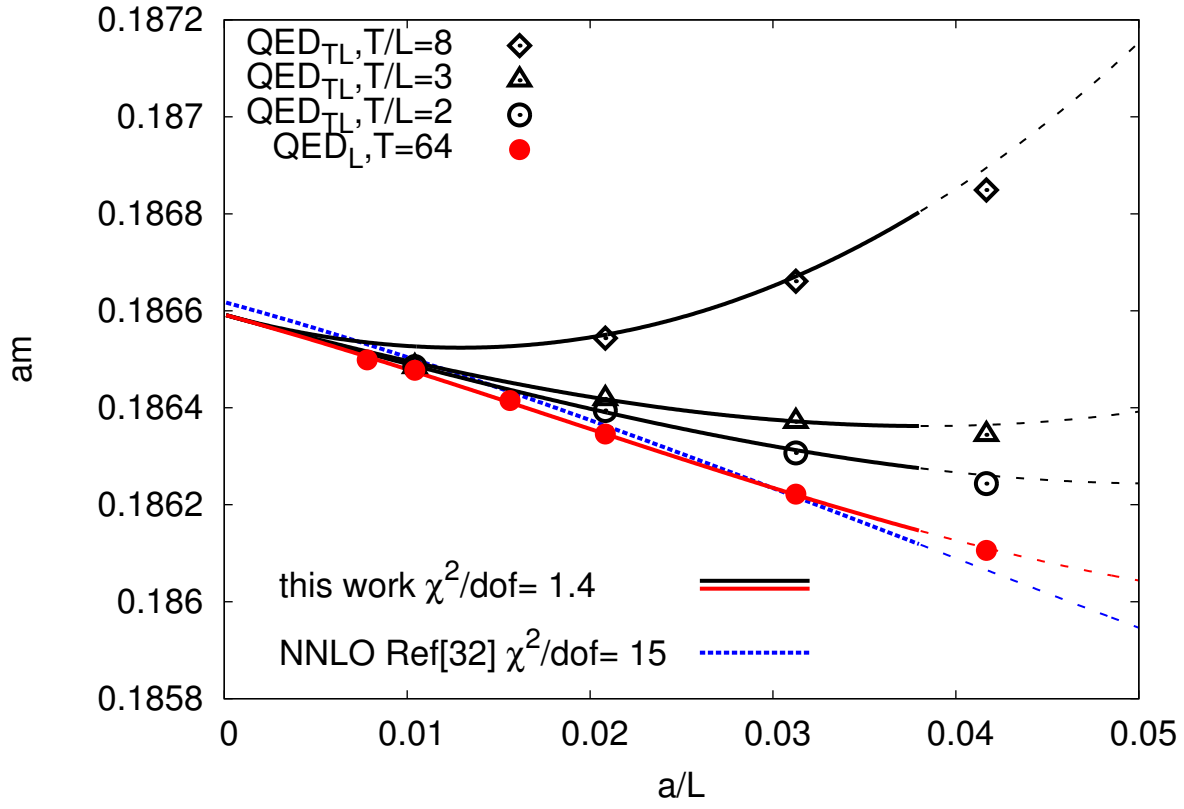

Figure S3: Finite-size dependence of the fermion mass in two different realizations of finite-volume QED, that differ in the zero-mode subtraction. The open/filled symbols are obtained using QED_{TL}/QED_{L} prescription. The curves are the analytical formulae from Eqs. (S22) and (S24). The dotted curve is the NNLO formula from Ref. (*32*).

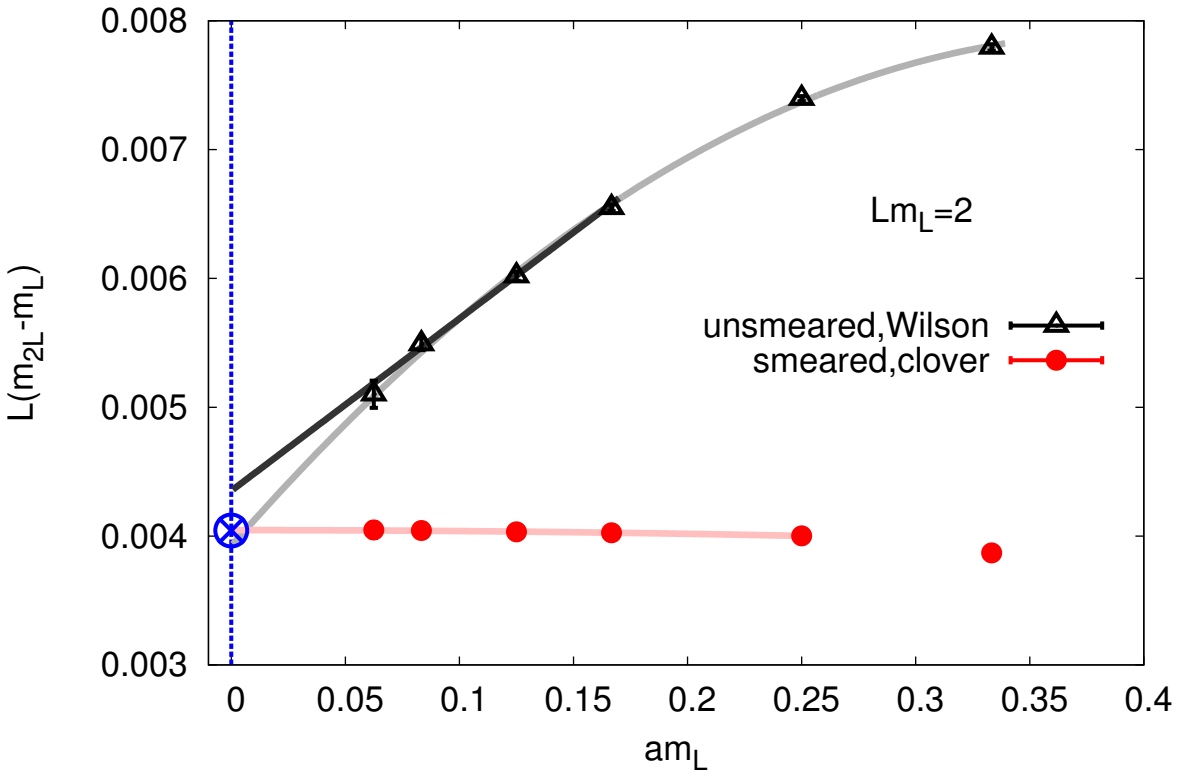

Figure S4: Continuum scaling of a mass difference in QED_L using two different lattice actions. The crossed circle in the continuum limit is the analytical result from the expressions of Sec. 3 (see text). The fitted curves are described in the text.

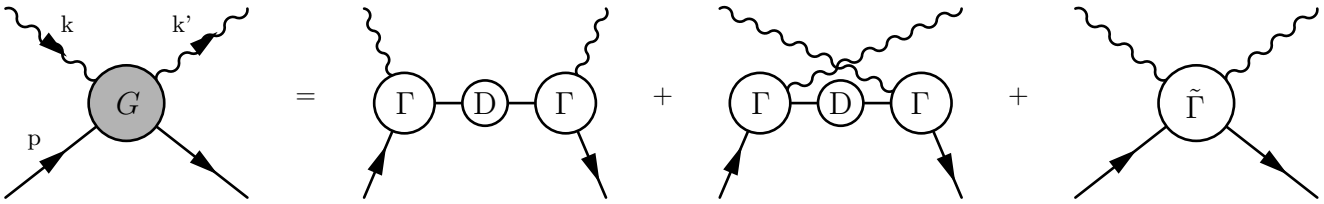

Figure S5: Decomposition of the four-point vertex into 1PI parts.

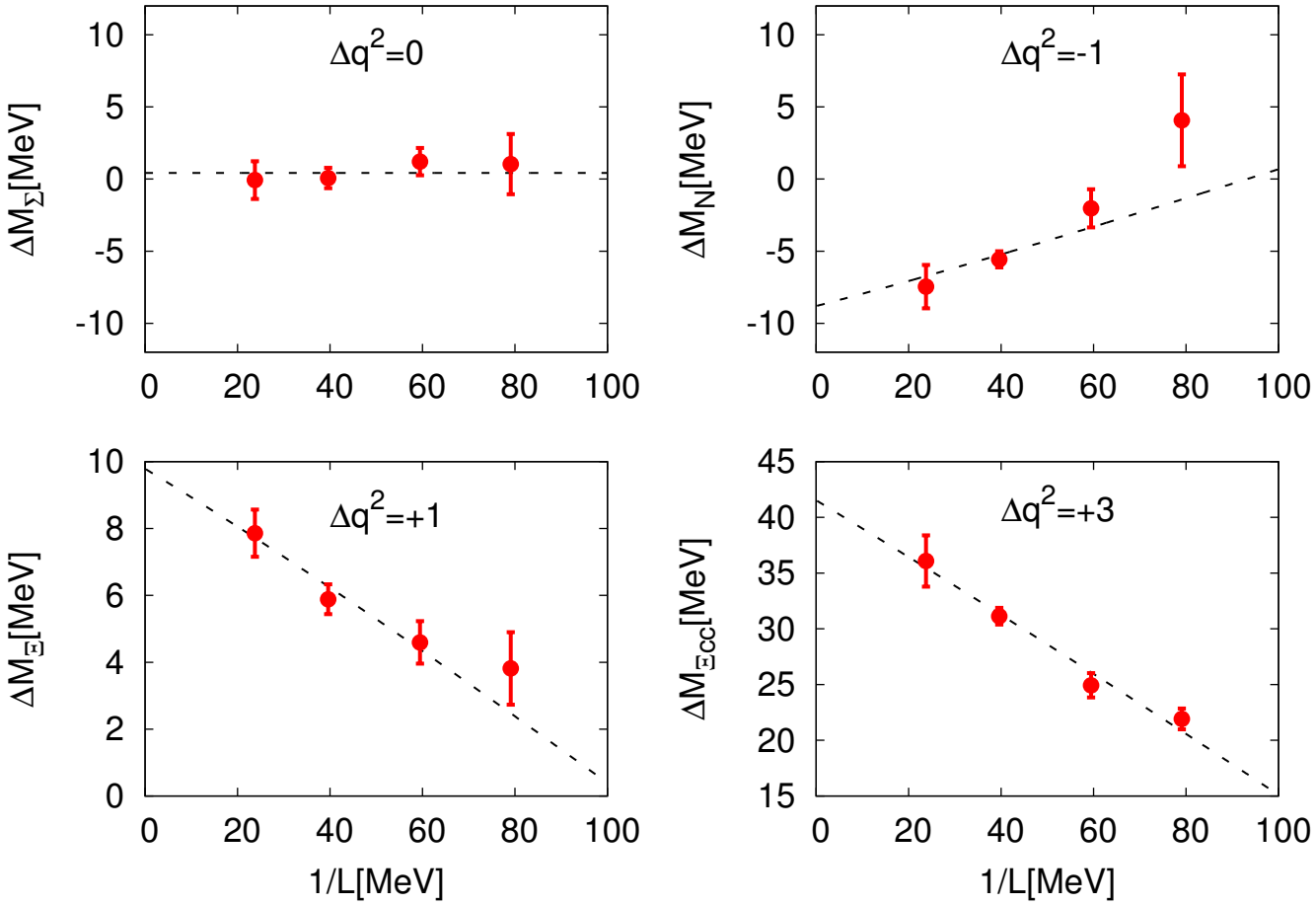

Figure S6: Finite-volume effects in baryon isospin splittings. The dependence is always consistent with the universal behavior of Eq. (S26) (dashed lines).

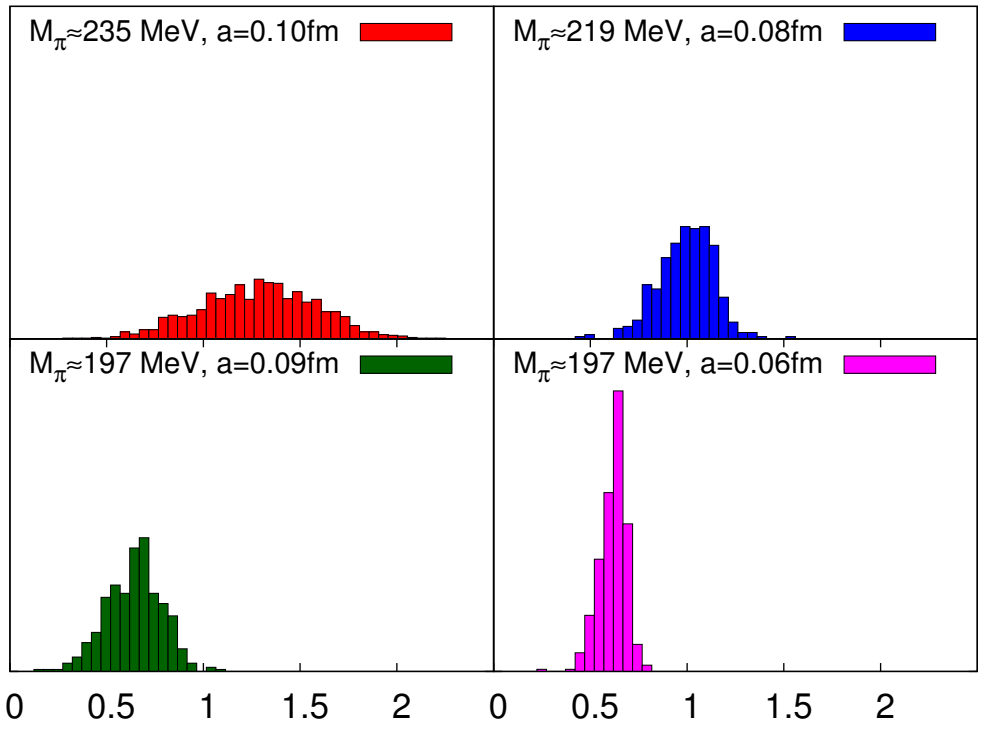

Figure S7: Lowest eigenvalue of the $\sqrt{D^\dagger D}$ operator on the ensembles with the smallest pion mass at each lattice spacing demonstrating the algorithmic stability.

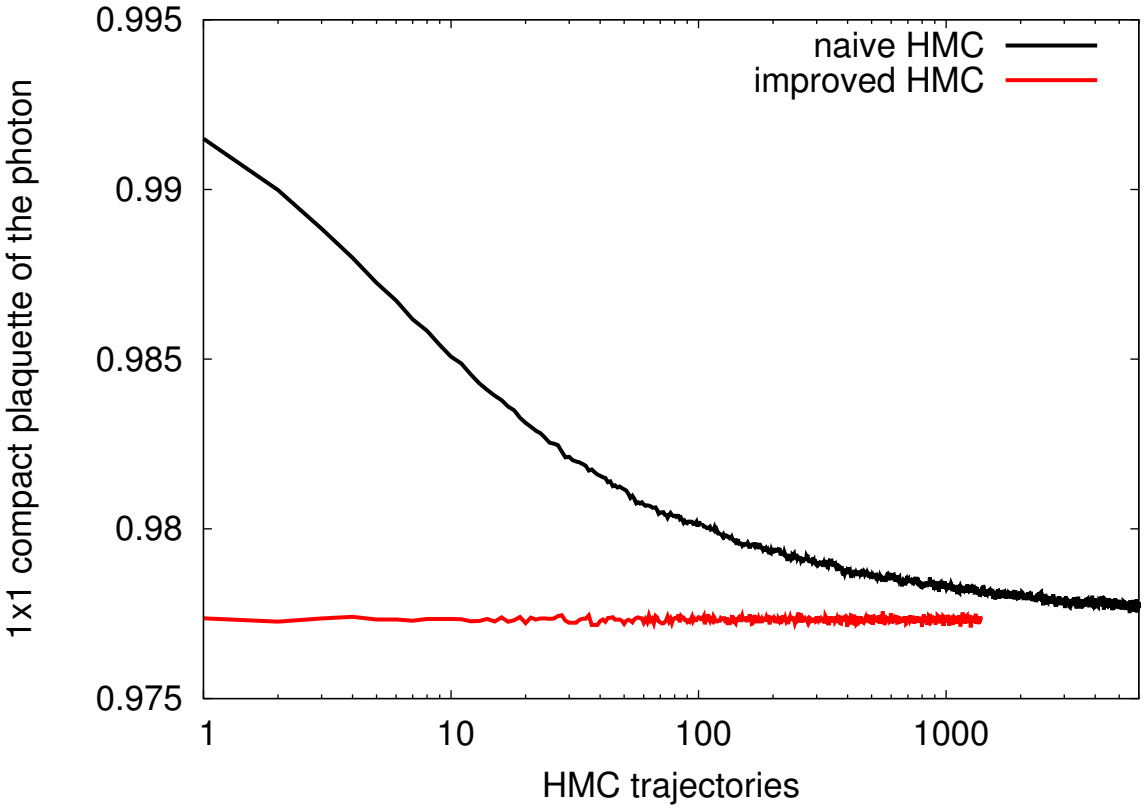

Figure S8: Thermalization of the compact plaquette in the pure photon theory using the standard HMC and our improved variant.

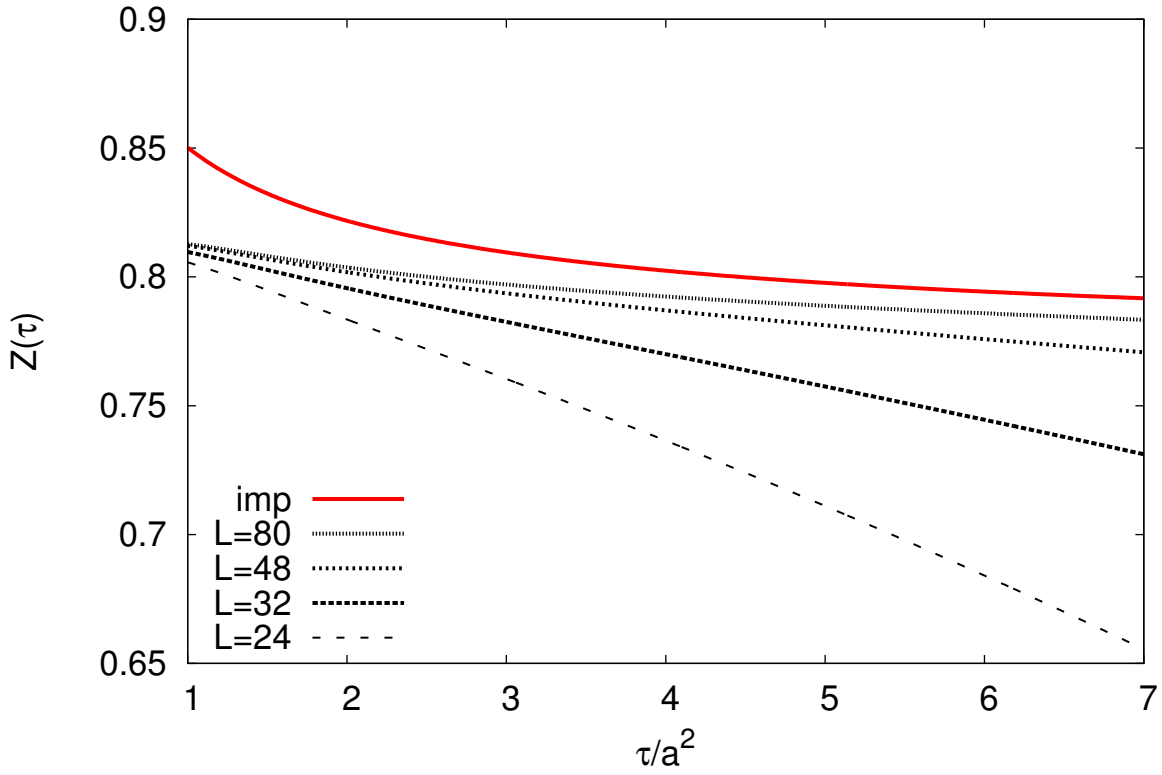

Figure S9: Renormalization factor of the electric charge as the function of flow time on four ensembles with fixed bare parameters. The dashed curves show $Z(\tau)$'s using $E_{\text{tree}} = 3/(32\pi^2)$ and have sizeable finite-volume dependence. Using $E_{\text{tree}}(\tau)$ from Eq. (S36) eliminates volume dependence completely: the curves corresponding to the four volumes lie on top of one another (solid line).

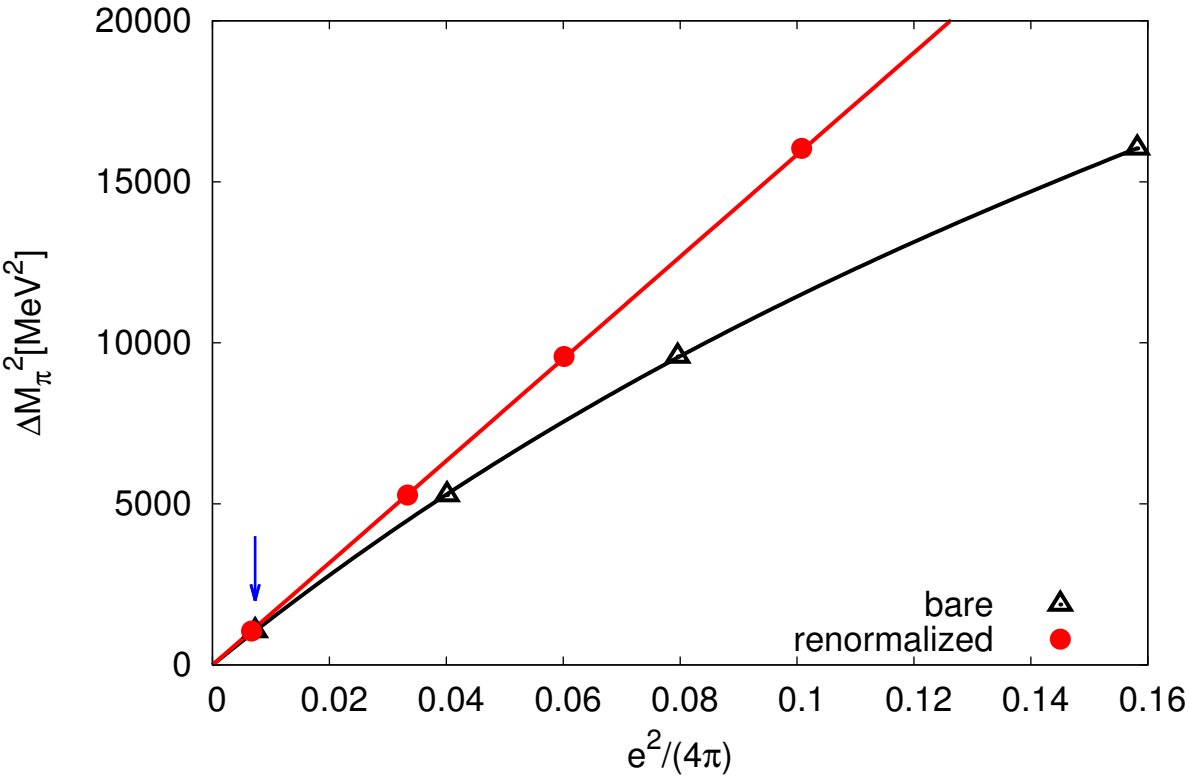

Figure S10: Pion splitting as a function of the bare and renormalized couplings for four charged ensembles. The solid lines are linear/cubic fits to the renormalized/bare data points. The arrow indicates the physical value of α.

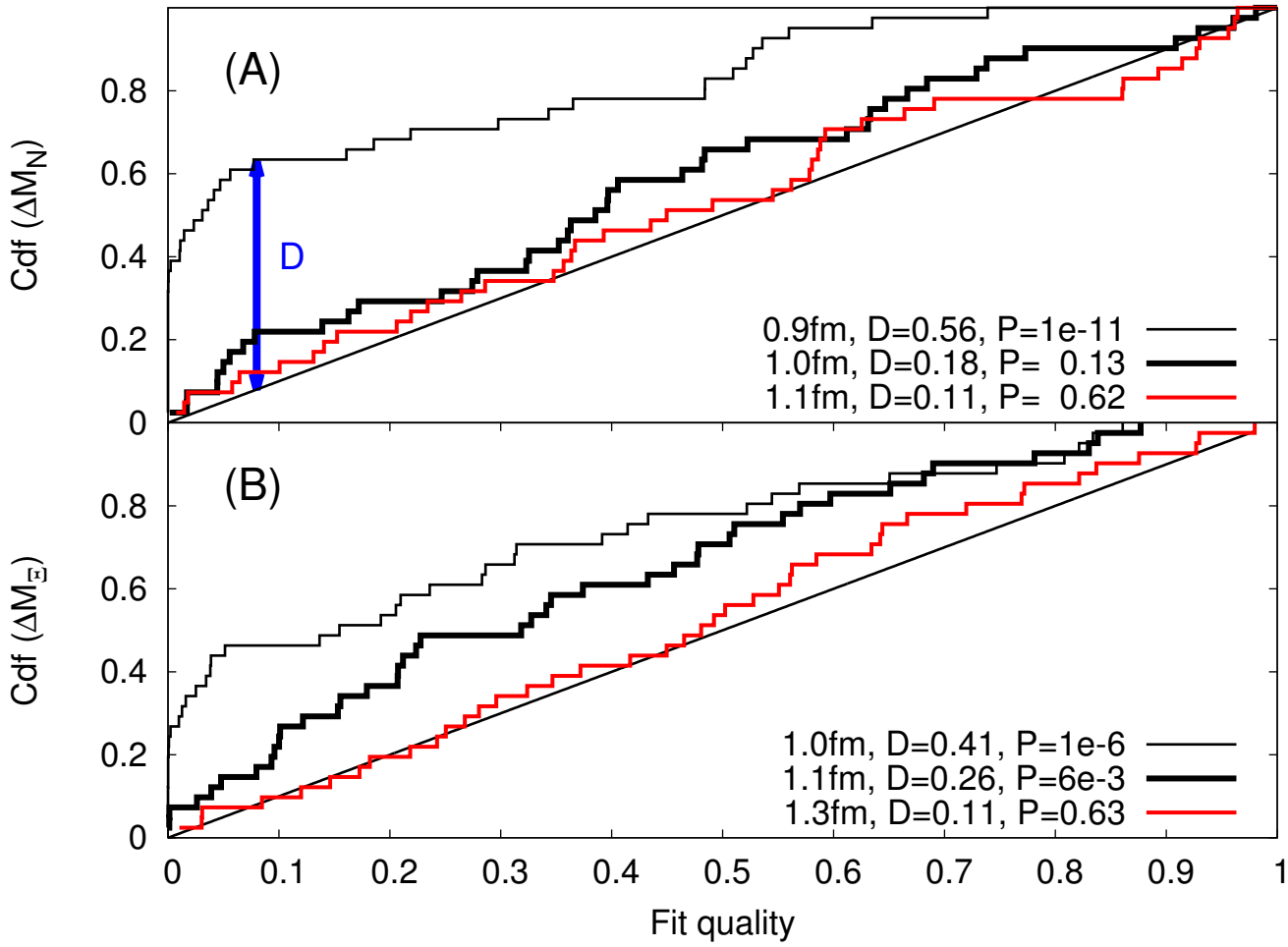

Figure S11: Cumulative distribution functions of the fit qualities. The distribution is obtained using the fits for the ΔM_N (panel A) and ΔM_Ξ (panel B) mass difference on the 41 ensembles. Three different measured distributions are plotted using three different t_{min} values, the largest ones correspond to our choices in the final analysis. The straight line corresponds to the expected uniform distribution. An arrow shows the maximum distance D for one of the distributions. For each value of t_{min} we provide the distances D and the probabilities P that the observed distributions are uniform.

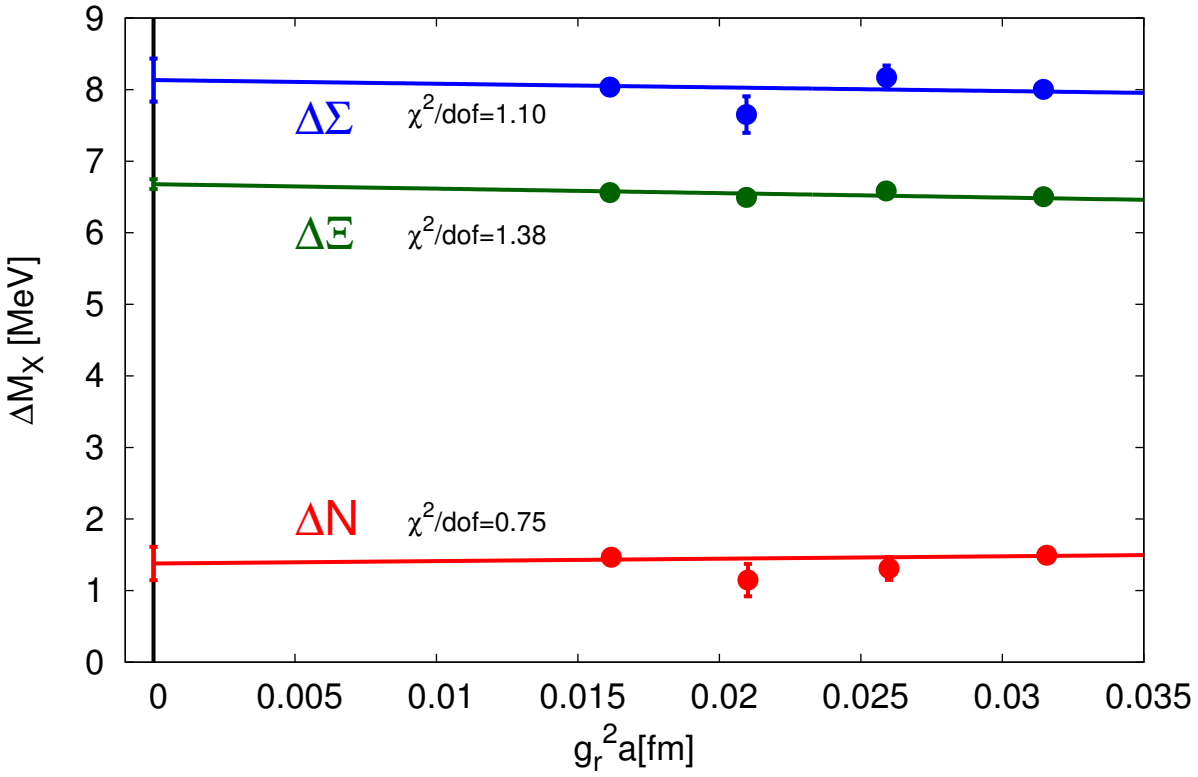

Figure S12: Lattice spacing dependence of the baryon octet isospin splittings in one particular fit.

e	analytical (*31*)	numerical	difference
0.21	0.191391	0.191378(05)	0.000013(05)
$\sqrt{4\pi/137}$	0.201230	0.201155(10)	0.000075(10)
0.40	0.215411	0.215161(10)	0.000250(10)

Table S1: Comparison of analytical and infinite volume extrapolated numerical results.

$6/g^2$	a[fm]	am_c
3.2	0.102	0.71
3.3	0.089	0.58
3.4	0.077	0.47
3.5	0.064	0.35

Table S2: Lattice spacing and charm quark mass at the four gauge couplings used in this work.

$6/g^2$	am_u	am_d	am_s	$L^3 \times T$	m_π[MeV]	$m_\pi L$	$\times 1000$ trajectories
3.2	-0.0686	-0.0674	-0.068	$32^3 \times 64$	405	6.9	1
3.2	-0.0737	-0.0723	-0.058	$32^3 \times 64$	347	5.9	4
3.2	-0.0733	-0.0727	-0.058	$32^3 \times 64$	345	5.8	1
3.2	-0.0776	-0.0764	-0.05	$32^3 \times 64$	289	4.9	4
3.2	-0.0805	-0.0795	-0.044	$32^3 \times 64$	235	4.0	12
3.2	-0.0806	-0.0794	-0.033	$32^3 \times 64$	256	4.4	12
3.2	-0.0686	-0.0674	-0.02	$32^3 \times 64$	440	8.1	4
3.2	-0.0737	-0.0723	-0.025	$32^3 \times 64$	377	6.8	4
3.2	-0.0776	-0.0764	-0.029	$32^3 \times 64$	317	5.6	4
3.2	-0.077	-0.0643	-0.0297	$32^3 \times 64$	404	7.3	4
3.2	-0.073	-0.0629	-0.0351	$32^3 \times 64$	435	7.8	4
3.2	-0.077	-0.0669	-0.0391	$32^3 \times 64$	378	6.7	4
3.3	-0.0486	-0.0474	-0.048	$32^3 \times 64$	407	6.1	1
3.3	-0.0537	-0.0523	-0.038	$32^3 \times 64$	341	5.1	2
3.3	-0.0535	-0.0525	-0.038	$32^3 \times 64$	340	5.0	2
3.3	-0.0576	-0.0564	-0.03	$32^3 \times 64$	269	4.0	12
3.3	-0.0576	-0.0564	-0.019	$32^3 \times 64$	281	4.2	12
3.3	-0.0606	-0.0594	-0.024	$48^3 \times 64$	197	4.3	20
3.4	-0.034	-0.033	-0.0335	$32^3 \times 64$	403	5.0	4
3.4	-0.0385	-0.0375	-0.0245	$32^3 \times 64$	318	4.0	4
3.4	-0.0423	-0.0417	-0.0165	$48^3 \times 64$	219	4.1	4
3.5	-0.0218	-0.0212	-0.0215	$32^3 \times 64$	420	4.4	4
3.5	-0.0254	-0.0246	-0.0145	$48^3 \times 64$	341	5.4	4
3.5	-0.0268	-0.0262	-0.0115	$48^3 \times 64$	307	4.8	8
3.5	-0.0269	-0.0261	-0.0031	$48^3 \times 64$	306	4.9	8
3.5	-0.0285	-0.0275	-0.0085	$48^3 \times 64$	262	4.1	8
3.5	-0.0302	-0.0294	-0.0049	$64^3 \times 96$	197	4.1	4

Table S3: List of "neutral ensembles".

$6/g^2$	e	am_u	am_d	am_s	$L^3 \times T$	m_π[MeV]	$m_\pi L$	$\times 1000$ trajectories
3.2	1.00	-0.0819	-0.0752	-0.0352	$32^3 \times 64$	373	6.6	4
3.2	$\sqrt{4\pi/137}$	-0.07788	-0.07722	-0.05022	$32^3 \times 64$	290	4.9	4
3.2	1.00	-0.0859	-0.0792	-0.0522	$32^3 \times 64$	290	4.9	4
3.2	1.41	-0.0943	-0.0812	-0.0542	$32^3 \times 64$	290	4.9	4
3.2	0.71	-0.0815	-0.0781	-0.0511	$32^3 \times 64$	290	4.9	4
3.2	1.00	-0.0889	-0.0822	-0.0462	$32^3 \times 64$	236	4.0	4
3.2	1.00	-0.0859	-0.0792	-0.0522	$24^3 \times 48$	292	3.7	5
3.2	1.00	-0.0859	-0.0792	-0.0522	$48^3 \times 96$	290	7.3	4
3.2	1.00	-0.0859	-0.0792	-0.0522	$80^3 \times 64$	289	12.2	1
3.3	1.00	-0.063	-0.0555	-0.0405	$48^3 \times 96$	335	7.4	4
3.3	1.00	-0.0666	-0.0592	-0.0329	$48^3 \times 96$	270	6.0	4
3.5	1.00	-0.034	-0.02575	-0.02575	$32^3 \times 64$	411	4.3	4
3.5	1.00	-0.0359	-0.0277	-0.0173	$48^3 \times 96$	362	5.7	4
3.5	1.00	-0.0389	-0.0307	-0.0111	$48^3 \times 96$	283	4.5	4

Table S4: List of "charged ensembles".

	t_{min}[fm]
ΔM_N	1.1
ΔM_Σ	1.1
ΔM_Ξ	1.3
ΔM_D	1.1
$\Delta M_{\Xi cc}$	1.2

Table S5: Starting time of the fit-intervals for different mass-splittings.