Calculation of the axion mass based on high temperature lattice quantum chromodynamics

Sz. Borsanyi[1], Z. Fodor[1,2,3], J. Gunther[1], K.-H. Kampert[1], S. D. Katz[3,4], T. Kawanai[2], T. G. Kovacs[5], S. W. Mages[2], A. Pasztor[1], F. Pittler[3,4], J. Redondo[6,7], A. Ringwald[8], K. K. Szabo[1,2]

[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] Institute for Nuclear Research of the Hungarian Academy of Sciences, H-4026 Debrecen, Hungary
[6] University of Zaragoza, E-50009 Zaragoza, Spain
[7] Max Planck Institut für Physik, D-80803, Germany
[8] Deutsches Elektronen-Synchrotron DESY, D-22607 Hamburg, Germany

The theories of the electroweak and the strong interactions determine the equation of state (EoS) of the early universe. Here we present a result of it. Quantum Chromo Dynamics (QCD), unlike the rest of the Standard Model, is surprisingly symmetric under time reversal, leading to a serious fine tuning problem. The most attractive solution for this [1] leads to a new particle, the axion [2, 3] –a promising dark matter candidate. Assuming that axions are the dominant component of dark matter we determine the axion mass. The key quantities of the calculation are the previously mentioned EoS and the temperature (T) dependence of the topological susceptibility ($\chi(T)$) of QCD, a quantity notoriously difficult to calculate [4–8]. Determining $\chi(T)$ was considered to be difficult in the most relevant high temperature region (T up to several GeV), however an understanding of the deeper structure of the vacuum by splitting it into different sectors and re-defining the fermionic determinants has led to its controlled calculation. Thus, our two-fold prediction helps most cosmological calculations [9] to describe the evolution of the early universe by using the EoS and may be decisive for guiding experiments looking for dark matter axions. In the next couple of years, it should be possible to confirm or rule out post-inflation axions experimentally if the axion's mass is or is not found to be as predicted here. Alternatively, in a pre-inflation scenario our calculation determines the universal axionic angle that corresponds to the initial condition of our universe.

In this paper, we use the lattice formulation of QCD [10], i.e. we discretize space-time on a four dimensional lattice with N_t and N_s points in the temporal and spatial directions. The lattice spacing is denoted by a, the box size by $L = N_s a$, the temperature by $T = (aN_t)^{-1}$ and the four-volume by $V = N_s^3 N_t a^4$.

Our most important qualitative knowledge about the QCD transition is that it is an analytic crossover [11], thus no cosmological relics are expected. Outside the narrow temperature range of the transition we know that the Hubble rate and the relationship between temperature and the age of the early universe can be described by a radiation-dominated EoS. The calculation of the EoS is a challenging task, the determination of the continuum limit at large temperatures is particularly difficult.

In our lattice QCD setup we used 2+1 or 2+1+1 flavours of staggered fermions [12] with four steps of stout-smearing [13]. The quark masses are set to their physical values, however we use degenerate up and down quark masses and the small effect of isospin breaking is included analytically. The continuum limit is taken using three, four or five lattice spacings with temporal lattice extensions of N_t=6, 8, 10, 12 and 16. In addition to dynamical staggered simulations we also used dynamical simulations with 2+1 flavours of overlap quarks [14] down to physical masses. The inclusion of an odd number of flavours was a non-trivial task, however this setup was required for the determination of $\chi(T)$ at large temperatures in the several GeV region.

Charm quarks start to contribute to the equation of state above 300 MeV. Therefore up to 250 MeV we used 2+1 flavours of dynamical quarks. Connecting the 2+1 and the 2+1+1 flavour results at 250 MeV can be done smoothly. For large temperatures the step-scaling method for the equation of state of Ref. [15] was applied. We determined the EoS with complete control over all sources of systematics all the way to the GeV scale.

Two different methods were used to set the overall scale in order to determine the equation of state. One of them took the pion decay constant the other applied the w_0 scale [16]. 32 different analyses (e.g. the two different scale setting procedures, different interpolations, keeping or omitting the coarsest lattice) entered our histogram method [17, 18] to estimate systematic errors. We also calculated the goodness of the fit Q and weights based on the Akaike information criterion AICc [18] and we looked at the unweighted or weighted results. This provided the systematic errors on our findings. In the low temperature region we compared our results with the prediction of the Hadron Resonance Gas (HRG) approximation and found good agreement (within errorbars). This HRG approach is used to parameterize the equation of state for small temperatures. In addition, we used the hard thermal loop approach [19] to extend the EoS to high temperatures.

In order to have a complete description of the thermal evolution of the early universe we supplement our QCD calculation for the EoS by including the rest of the Standard Model particles (leptons, bottom

and top quarks, W, Z, Higgs bosons) and results on the electroweak transition. As a consequence, the final result on the EoS covers four orders of magnitude in temperature from MeV to several hundred GeV.

Figure 1 shows the EoS. The widths of the lines represent the uncertainties. Both the figure and the data can be used (similarly to Figure 22.3 of Ref. [20]) to describe the Hubble rate and the relationship between temperature and the age of the universe in a very broad temperature range.

We now turn to the determination of another cosmologically important quantity, $\chi(T)$. In general the action of QCD should have a term proportional to the topological charge of the gluon field, Q. This term violates the combined charge-conjugation and parity symmetry (CP). The surprising experimental observation is that the proportionality factor of this term θ is unnaturally small. This is known as the strong CP problem. A particularly attractive solution to this fundamental problem is the so-called Peccei-Quinn mechanism [1]. One introduces an additional pseudo-scalar U(1) symmetric field. The underlying Peccei-Quinn U(1) symmetry is spontaneously broken –which can happen pre-inflation or post-inflation– and an axion field A acts as a massless Goldstone boson of the broken symmetry [2, 3]. The symmetry breaking scale f_A is a free parameter. Due to the chiral anomaly the axion is coupled to the topological charge density. As a consequence, the original potential of the axion field with its U(1) symmetry breaking gets tilted and has its minimum where $(\theta + A/f_A) = 0$. This sets the proportionality factor of Q in the QCD action to zero and solves the strong CP problem. Furthermore, the axion acquires a mass m_A, which is given by $m_A^2 = \chi/f_A^2$. Here $\chi = \langle Q^2 \rangle / V$ is the susceptibility of the topological charge normalized by the four-volume. We determined its value at $T = 0$, which turned out to be $\chi(T = 0) = 0.0245(24)(12)/\mathrm{fm}^4$ in the isospin symmetric case, where the first error is statistical, the second is systematic. Isospin breaking results in a small, 12% correction, thus the physical value is $\chi(T = 0) = 0.0216(21)(11)/\mathrm{fm}^4 = [75.6(1.8)(0.9)\mathrm{MeV}]^4$.

On the lattice χ can be conveniently calculated using a Q defined along the Wilson-flow [21]. In an earlier study of ours [5] we looked at $\chi(T)$ in the quenched approximation. We provided a result within the quenched framework and reached a temperature about half to one third of the necessary temperatures for axion cosmology (a similar study with somewhat less control over the systematics is [4]). To obtain a complete result one should use dynamical quarks with physical masses. Dynamical configuration production is, however, about three orders of magnitude more expensive and the $\chi(T)$ values are several orders of magnitude smaller than in the quenched case. Due to cutoff effects the continuum limit is far more difficult to carry out in dynamical QCD than in the pure gauge theory [5]. All in all we estimate that the brute-force approach to provide a complete result on $\chi(T)$ in the relevant temperature region would be at least ten orders of magnitude more expensive than the result of [5].

The huge computational demand and the physics issue behind the determination of $\chi(T)$ has two main sources. a.) In high temperature lattice QCD the most widely used actions are based on staggered quarks. When dealing with topological observables staggered quarks have very large cutoff effects and b.) The tiny topological susceptibility needs extremely long simulation threads to observe enough changes of the topological sectors.

We solve both problems and determine the continuum result for $\chi(T)$ for the entire temperature range of interest. For the a.) problem we call our proposed solution "eigenvalue reweighting". The method is based on substituting the topology related eigenvalues of the staggered quark operator with the eigenvalues of the quark operator in the continuum. For the b.) problem we propose to measure the logarithmic differential of the susceptibility instead of the susceptibility itself, which is related to quantities, that are to be measured in fixed topological sectors. The final result is obtained with an integral, we call our method "fixed sector integral technique". Both techniques are explained in detail in the Methods.

The CPU costs of the conventional technique scale as T^8, whereas the new "fixed sector integration" method scales as T^0. The gain in CPU time is tremendous. This efficient technique is used to obtain the final result for $\chi(T)$. Since we work with continuum extrapolated quantities both for the ratios in the starting-point as well as for their changes, one can in principle use any action in the procedure, we will use here overlap and/or staggered actions.

Through combining these methods one can determine $\chi(T)$ (see Figure 2). The several thousand percent cutoff effects of staggered fermions are removed, leaving a very mild $\mathcal{O}(10\%)$ continuum extrap-

olation to be performed. In addition, the direct determination of $\chi(T)$ all the way up to 3 GeV means that one does not have to rely on the dilute instanton gas approximation (DIGA). Note that *a posteriori* the exponent predicted by DIGA turned out to be compatible with our finding but its prefactor is off by an order of magnitude, similar to the quenched case. Though some of the simulations are already carried out with chiral (overlap) fermions, where large cutoff effects are a-priori absent, it is an important task for the future to crosscheck these results with a calculation using chiral fermions only.

As a possible application for these two cosmologically relevant lattice QCD results, we show how to calculate the amount of axionic dark matter and how it can be used to determine the axion's mass. $\chi(T)$ is a rapidly decreasing function of the temperature. Thus, at high temperature m_A (which is proportional to $\chi(T)^{1/2}$) is small. In fact, much smaller than the Hubble expansion rate of the universe at that time or temperature ($H(T)$). The axion does not feel the tilt in the Peccei-Quinn Mexican hat type potential yet and it is effectively massless and frozen by the Hubble friction. As the Universe expands the temperature decreases, $\chi(T)$ increases and the axion mass also increases. In the meantime, the Hubble expansion rate –given by our equation of state– decreases. As the temperature decreases to $T_{\rm osc}$ the axion mass is of the same order as the Hubble constant ($T_{\rm osc}$ is defined by $3H(T_{\rm osc}) = m_A(T_{\rm osc})$). Around this time the axion field rolls down the potential, starts to oscillate around the tilted minimum and the axion number density increases to a nonzero value, thus axions as dark matter are produced. The details of this production mechanism, usually called misalignment, are quite well known (see e.g. [9]).

In a post-inflationary scenario the initial value of the angle θ takes all values between -π and π, whereas in the pre-inflationary scenario only one θ_0 angle contributes (all other values are inflated away). One should also mention that during the U(1) symmetry breaking topological strings appear which decay and also produce dark matter axions. In the pre-inflationary scenario they are inflated away. However, in the post-inflationary framework their role is more important. This sort of axion production mechanism is less well-understood and in our final results it is necessary to make some assumptions.

The possible consequences of our results on the predictions of the amount of axion dark matter can be seen in Figure 3. Here we also study cases, for which the dark matter axions are produced from the decay of unstable axionic strings (see the discussion in the figure's caption). For the pre-inflationary Peccei-Quinn symmetry breaking scenario the axion mass determines the initial condition θ_0 of our universe.

Acknowledgments We thank M. Dierigl, M. Giordano, S. Krieg, D. Nogradi and B. Toth for useful discussions. This project was funded by the DFG grant SFB/TR55, and by OTKA under grant OTKA-K-113034. The work of J.R. is supported by the Ramon y Cajal Fellowship 2012-10597 and FPA2015-65745-P (MINECO/FEDER). The computations were performed on JUQUEEN at Forschungszentrum Jülich (FZJ), on SuperMUC at Leibniz Supercomputing Centre in München, on Hazel Hen at the High Performance Computing Center in Stuttgart, on QPACE in Wuppertal and on GPU clusters in Wuppertal and Budapest.

Author Contributions SB and SM developed the fixed sector integral, TGK and KS the eigenvalue reweighting, FP the odd flavour overlap techniques, respectively. SB, JG, SK, TGK, TK, SM, AP, FP and KS wrote the necessary codes, carried out the runs and determined the EoS and $\chi(T)$. JR, AP and AR calculated the DIGA prediction. K-HK and JR and AR worked out the experimental setup. ZF wrote the main paper and coordinated the project.

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to ZF (fodor@bodri.elte.hu).

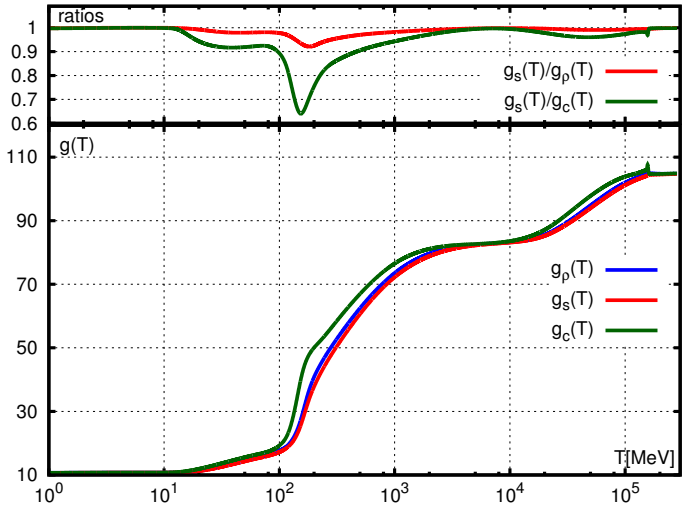


Figure 1: The effective degrees of freedom g_ρ for the energy density ($\rho = g_\rho \frac{\pi^2}{30} T^4$), g_s for the entropy density $s = g_s \frac{2\pi^2}{45} T^3$, and g_c for the heat capacity ($c = g_c \frac{2\pi^2}{15} T^3$). Neglecting the cosmological constant, the time dependence of the temperature in the early universe is given by these factors as: $\frac{dT}{dt} = -\frac{2\pi^{3/2}}{3\sqrt{5}} \frac{T^3}{M_{Pl}} \frac{\sqrt{g_\rho} g_s}{g_c}$, where M_{Pl} is the Planck mass. The line width is chosen to be the same as our error bars (s.e.m.) at the vicinity of the QCD transition, where we have the largest uncertainties. At temperatures $T < 1$ MeV the equilibrium equation of state becomes irrelevant for cosmology, because of neutrino decoupling. The EoS comes from our calculation up to $T = 100$ GeV. At higher temperatures the electroweak transition becomes relevant and we use the results of Ref. [22]. Note that for temperatures around the QCD transition non-perturbative QCD effects modify the EoS significantly, compared to the ideal gas limit, an approximation which is often used in cosmology, e.g. g_s/g_c is reduced from the SB limit by about 35%. Also note that g_s/g_c has four local minima: near the muon threshold, the QCD transition, the W, Z-boson thresholds and the electroweak transition. For parameterizations for the QCD regime or for the whole temperature range see the Supplementary Information.

References

1. Peccei, R. D. & Quinn, H. R. CP Conservation in the Presence of Instantons. *Phys. Rev. Lett.* **38,** 1440–1443 (1977).
2. Weinberg, S. A New Light Boson? *Phys. Rev. Lett.* **40,** 223–226 (1978).
3. Wilczek, F. Problem of Strong p and t Invariance in the Presence of Instantons. *Phys. Rev. Lett.* **40,** 279–282 (1978).
4. Berkowitz, E., Buchoff, M. I. & Rinaldi, E. Lattice QCD input for axion cosmology. *Phys. Rev.* **D92,** 034507 (2015).
5. Borsanyi, S. *et al.* Axion cosmology, lattice QCD and the dilute instanton gas. *Phys. Lett.* **B752,** 175–181 (2016).
6. Trunin, A., Burger, F., Ilgenfritz, E.-M., Lombardo, M. P. & Müller-Preussker, M. Topological susceptibility from $N_f = 2 + 1 + 1$ lattice QCD at nonzero temperature. *J. Phys. Conf. Ser.* **668,** 012123 (2016).
7. Bonati, C. *et al.* Axion phenomenology and θ-dependence from $N_f = 2 + 1$ lattice QCD. *JHEP* **03,** 155 (2016).
8. Petreczky, P., Schadler, H.-P. & Sharma, S. The topological susceptibility in finite temperature QCD and axion cosmology. arXiv: `1606.03145 [hep-lat]` (2016).
9. Wantz, O. & Shellard, E. P. S. Axion Cosmology Revisited. *Phys. Rev.* **D82,** 123508 (2010).

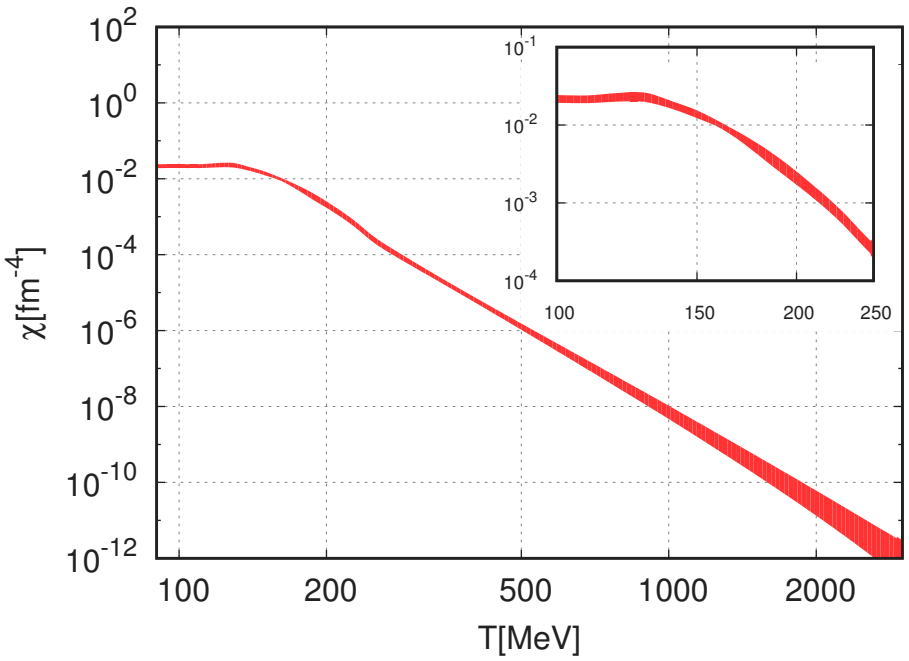


Figure 2: Continuum limit of $\chi(T)$. The insert shows the behaviour around the transition temperature. The width of the line represents the combined statistical and systematic errors (s.e.m.). The dilute instanton gas approximation (DIGA) predicts a power behaviour of T^{-b} with b=8.16, and the lattice results is close to this value.

10. Wilson, K. G. Confinement of Quarks. *Phys. Rev.* **D10.** [,45(1974)], 2445–2459 (1974).
11. Aoki, Y., Endrodi, G., Fodor, Z., Katz, S. D. & Szabo, K. K. The Order of the quantum chromodynamics transition predicted by the standard model of particle physics. *Nature* **443,** 675–678 (2006).
12. Kogut, J. B. & Susskind, L. Hamiltonian Formulation of Wilson's Lattice Gauge Theories. *Phys. Rev.* **D11,** 395–408 (1975).
13. Morningstar, C. & Peardon, M. J. Analytic smearing of SU(3) link variables in lattice QCD. *Phys. Rev.* **D69,** 054501 (2004).
14. Neuberger, H. Exactly massless quarks on the lattice. *Phys. Lett.* **B417,** 141–144 (1998).
15. Borsanyi, S., Endrodi, G., Fodor, Z., Katz, S. D. & Szabo, K. K. Precision SU(3) lattice thermodynamics for a large temperature range. *JHEP* **07,** 056 (2012).
16. Borsanyi, S. *et al.* High-precision scale setting in lattice QCD. *JHEP* **09,** 010 (2012).
17. Durr, S. *et al.* Ab-Initio Determination of Light Hadron Masses. *Science* **322,** 1224–1227 (2008).
18. Borsanyi, S. *et al.* Ab initio calculation of the neutron-proton mass difference. *Science* **347,** 1452–1455 (2015).
19. Andersen, J. O., Leganger, L. E., Strickland, M. & Su, N. NNLO hard-thermal-loop thermodynamics for QCD. *Phys. Lett.* **B696,** 468–472 (2011).
20. Olive, K. A. *et al.* Review of Particle Physics. *Chin. Phys.* **C38,** 090001 (2014).
21. Lüscher, M. Properties and uses of the Wilson flow in lattice QCD. *JHEP* **08.** [Erratum: JHEP03,092(2014)], 071 (2010).
22. Laine, M. & Meyer, M. Standard Model thermodynamics across the electroweak crossover. *JCAP* **1507,** 035 (2015).
23. Aoki, Y. *et al.* The QCD transition temperature: results with physical masses in the continuum limit II. *JHEP* **06,** 088 (2009).

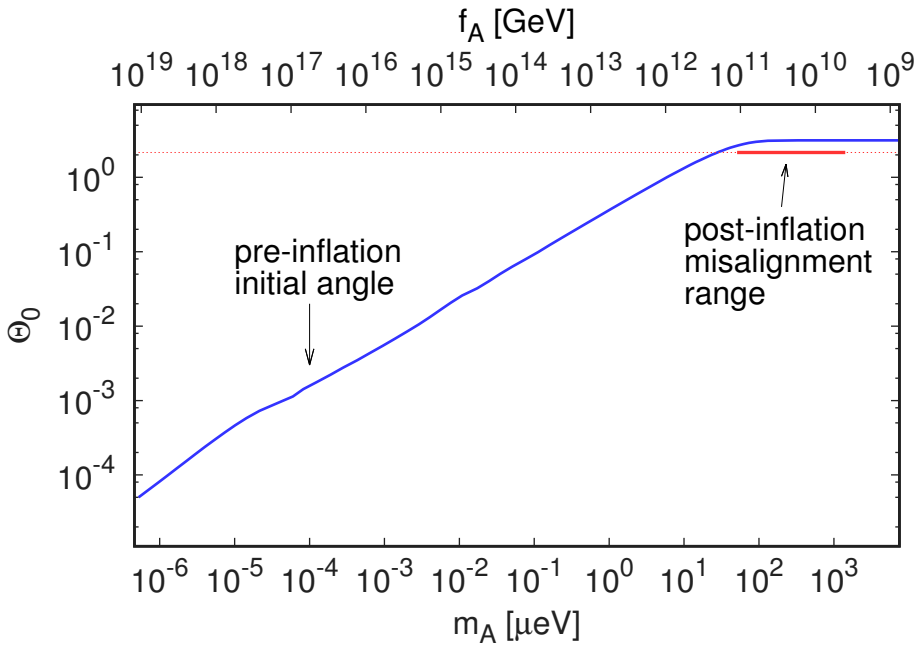


Figure 3: The relation between the axion's mass and the initial angle θ_0 in the pre-inflation scenario. The result is shown by the blue line, the error (s.e.m.) is smaller than the line width. The post-inflation scenario corresponds to $\theta_0 = 2.155$ with a strict lower bound on the axion's mass of m_A=28(2)μeV. The thick red line shows our result on the axion's mass for the post-inflation case. E.g. m_A=50(4)μeV if one assumes that axions from the misalignment mechanism contributes 50% to dark matter. Our final estimate is m_A=50-1500μeV (the upper bound assumes that only 1% is the contribution of the misalignment mechanism the rest comes from other sources e.g. topological defects). For an experimental setup to detect post-inflationary axions see the Supplementary Information. The slight bend around $m_A \sim 10^{-5}$ μeV corresponds to an oscillation temperature at the QCD transition [23, 24].

24. Borsanyi, S. *et al.* Is there still any T_c mystery in lattice QCD? Results with physical masses in the continuum limit III. *JHEP* **09,** 073 (2010).

Methods

Eigenvalue reweighting technique

Here we show how cut-off effects in χ arise with staggered quarks and propose a new method to efficiently suppress them.

The cut-off effects are strongly related to the zero-modes. To understand their importance, we first note that in the quark determinant every zero-mode for each dynamical flavor contributes a factor m_f, the corresponding quark mass. In this way gauge configurations with zero modes are strongly suppressed in the path integral, especially if the quark masses are small. Due to the index theorem, this also implies that light dynamical quarks strongly suppress higher topological sectors and thus χ itself.

On the lattice, however, there can be strong cut-off effects in this suppression. This is because the suppression factor is not m_f but $m_f + i\lambda_0$, where λ_0 is an eigenvalue corresponding to the would-be zero mode of the staggered Dirac operator, D_{st}. The lack of exact zero modes can thus introduce strong cut-off effects and slow convergence to the continuum limit. Indeed, as long as the typical would-be zero eigenvalues are comparable to or larger than the lattice bare quark mass m_f, higher sectors are much less suppressed on the lattice than in the continuum.

To improve the situation, even at finite lattice spacing we can identify the would-be zero modes and restore their continuum weight in the path integral. In case of rooted staggered quarks this amounts to a reweighting of each configuration with a weight factor

$$w[U] = \prod_f \prod_{n=1}^{2|Q[U]|} \prod_{\sigma=\pm} \left(\frac{m_f}{\sigma i \lambda_n[U] + m_f} \right)^{n_f/4} \tag{1}$$

where the second product runs over the would-be zero eigenvalues of the staggered Dirac operator with positive imaginary part. The third product takes into account the $i\lambda \to -i\lambda$ symmetry of the eigenvalue spectrum. The $n_f/4$ factor takes rooting into account, the factor 2 next to $|Q|$ together with the $\pm$ symmetry make up for the fact that in the continuum limit the staggered zero modes become four-fold degenerate [25].

Let us now turn to the most important part of the reweighting: the definition of the would-be zero modes. Since we are interested in χ, we identify the number of these modes with the magnitude of the topological charge $2|Q|$ as obtained from the gauge field after using the Wilson flow, see the Supplementary Information. We investigated two specific choices for the would-be zero modes. In the first approach we took the $2|Q|$ eigenmodes that have the largest magnitude of chirality among the eigenmodes with the appropriate sign of chirality, positive if $Q < 0$ and negative if $Q > 0$. In the second approach we took the $2|Q|$ eigenmodes with smallest magnitude. These two approaches are equivalent in the continuum limit, where zero-modes are exactly at zero and their chirality is unity. In practical simulations they give very similar results, we use the second approach in our analysis.

Since in the continuum limit the would-be zero eigenvalues get closer to zero, the reweighting factors tend to unity and in the continuum limit we recover the original Dirac operator. This way, however, even at finite lattice spacings the proper suppression of higher sectors is restored and cut-off effects are strongly reduced resulting in much faster convergence in the continuum limit. For completeness let us note, that the above modification corresponds to a non-local modification of the path integral[1]. In the following we provide several pieces of numerical evidence for the correctness of the approach.

In Extended Data Figure 1 we plot the distribution of the eigenvalues corresponding to the would-be zero modes at a temperature of $T = 240$ MeV for different lattice spacings. The distributions get narrower

[1]In this respect it stands on a footing similar to another method, which also modifies the quark determinant and which we also use in our staggered simulations: determinant rooting. As of today there is ample theoretical and numerical evidence for the correctness of the staggered rooting. See [26] and its follow ups.

and their center moves towards zero as the lattice spacing is decreased. In Extended Data Figure 2 we show the expectation value of the reweighting factors in the first few sectors. In the continuum limit $\langle w \rangle_Q = 1$ should be fulfilled in each sector. The results nicely converge to 1.

In most of our runs, especially at large temperatures and small quark masses, the weights were much smaller than 1. As a result there are orders of magnitude differences between χ with and without reweighting. It is therefore important to illustrate how the standard approach breaks down if the lattice spacing is large and how the correct result is recovered for very small lattice spacings. In the following we show two examples, Extended Data Figures 3 and 4, where the standard method produces cut-off effects so large, that a reliable continuum extrapolation is not possible. In contrast, the lattice spacing dependence of the reweighted results is much milder. To make sure that the reweighted results are in the a^2-scaling regime, for both cases we present a non-standard approach to determine χ and compare them to reweighting.

In the first case (Extended Data Figure 3) the temperature is just at the transition point, $T = 150$ MeV, where we expect to get a value close to the zero temperature susceptibility. This suggests that in this case the cut-off effects of the standard method can be largely eliminated by performing the continuum limit of the ratio $\chi(T, a)/\chi(T = 0, a)$, where the finite temperature result is divided by the zero temperature one at the same lattice spacing. We call this approach "ratio method", see e.g. [7]. As it can be seen in Extended Data Figure 3, this is indeed the case. The so obtained continuum extrapolation is nicely consistent with reweighting.

In the second case, Extended Data Figure 4, we have a temperature well above the transition, $T = 300$ MeV. We see again, that the standard method produces results with large cut-off effects. The ratio method seems to perform better, however the apparent scaling is misleading. Although a nice continuum extrapolation can be done from lattice spacings $N_t = 8, 10$ and 12, the $N_t = 16$ result is much below the extrapolation curve. The reweighting produces a result that is an order of magnitude smaller. Below we introduce a new method, called "integral method", which is tailored for large temperatures. The so obtained result, where no reweighting is applied, agrees reasonably with the reweighted one in the continuum limit.

These results provide numerical evidence for our expectations: reweighting not only produces a correct continuum limit, it also eliminates the large cut-off effects of staggered fermions.

Fixed sector integral technique

There are many proposals to increase the tunneling between the topological sectors, see e.g. [27–29]. Here we forbid sector changes and determine the relative weight of the sectors by measuring the Q dependence of certain observables[2]. We illustrate the method in the quenched theory, for the extension in the case of dynamical fermions see the Supplementary Information. The gauge configurations are generated with a probability proportional to $\exp(-\beta S_g)$, where β is the gauge coupling parameter and S_g is the gauge action. Let us consider the following differentials:

$$b_Q \equiv -\frac{d \log Z_Q/Z_0}{d \log T} = -\frac{d\beta}{d \log a} \langle S_g \rangle_{Q-0}, \tag{2}$$

where Z_Q is the partition function of the system restricted to sector Q. In the continuum limit the sectors are unambiguously defined, however, on the lattice several different definitions are possible, our choice is given later on. In Equation (2) we introduced the notation $\langle O \rangle_{Q-0} = \langle O \rangle_Q - \langle O \rangle_0$ for the difference of the expectation values of an observable between the sectors Q and 0. Equation (2) gives a renormalized quantity, the ultraviolet divergences cancel in the difference of the gauge actions. The important observation is that the necessary statistics to reach a certain level of precision on b_Q's does not depend on the temperature.

[2]A few hours after the submission of the present paper to the arXiv, a paper appeared by J. Frison et al. discussing essentially the same method [30], though only in the quenched approximation using coarse lattices.

To obtain the relative weights Z_Q/Z_0, we just have to integrate Equation (2) in the temperature. For that we start from a temperature T_0, where the standard approach is still feasible and determine Z_Q/Z_0. Then by measuring the b_Q's for higher temperatures, we can use the following integral to obtain Z_Q/Z_0

$$Z_Q/Z_0|_T = \exp\left(-\int_{T_0}^{T} d\log T'\ b_Q(T')\right) Z_Q/Z_0|_{T_0}. \tag{3}$$

Let us make a remark about the volume dependence. As we increase the temperature, the ratios Z_{Q+1}/Z_Q get smaller. This effect is in competition with the infinite volume limit, which brings these ratios closer to 1. The question is how many sectors are needed to determine χ reliably. χ is an intensive quantity, and as such, its finite volume effects can be neglected, if the box size is large enough to accommodate all correlation lengths in the system. In our quenched study [5] we found, that for $LT_c \gtrsim 2$ the finite size effects on χ are negligible, where T_c denotes the phase transition temperature.

For high temperatures only the $Q = 0$ and 1 sectors remain relevant and $\langle Q^2\rangle = \chi V$ becomes small. Using the data of our quenched simulations [5] we found that in a box size of $L = 2/T_c$ the contribution of $Q \geq 2$ sectors to χ and also χV are on the percent level at $1.7T_c$ and they decrease rapidly with the temperature.[3] In this case it is appropriate to write $\chi = 2Z_1/Z_0/(1 + 2Z_1/Z_0)/V$. To the accuracy of $\mathcal{O}(\chi V)$ one can also use $\chi \approx 2Z_1/Z_0/V$, and then the decay exponent of the susceptibility b can be simply obtained as

$$b \equiv -\frac{d\log\chi}{d\log T} \approx b_1 - 4. \tag{4}$$

Here the term -4 reflects, that the physical volume also changes with the temperature. To derive the Stefan-Boltzmann limit of Equation (4), we can use that for large temperatures $\beta = 33\log a/(4\pi^2)$. The gauge action difference is given by the classical action of one instanton $\langle S_g\rangle_{1-0} = 4\pi^2/3$. Up to lattice artefacts we get $b = 7$ in the Stefan-Boltzmann limit.

As we have already mentioned, to reach the same level of precision on b_1 as the temperature increases, the statistics can be kept constant. However, with increasing the spatial size N_s, the statistics has to be increased as N_s^3, and as a result the computer time goes as N_s^6. This can be understood as follows: the gauge action difference between sectors 1 and 0 will be approximately given by the action of one instanton, which remains constant with increasing volume. The gauge action S_g itself, however, increases with the volume and the cancellation in Equation (2) gets more severe. This volume squared scaling problem can be mildened by putting more and more topological charge into the box with increasing box size. If the topological objects are localized, well separated, then for large volumes the action difference between sectors 1 and 0 can be obtained from the difference between sectors Q and 0:

$$\langle S_g\rangle_{1-0} = \langle S_g\rangle_{Q-0}/|Q|. \tag{5}$$

It can be used to achieve a Q-fold increase in the signal-to-noise ratio, which translates into a Q^2-fold decrease in the necessary computing time. We are going to check this relation in our numerical simulations.

Numerical illustration

We have carried out several numerical simulations to test the new approach. Details on the algorithm and on definition of Q can be found in the Supplementary Information.

At finite lattice spacing Q is not necessarily an integer, thus there is a certain degree of ambiguity in defining the sectors, this ambiguity disappears in the continuum limit. First we looked at simulations, where we constrained Q to be larger than 0.5. The parameters can be found in Extended Data Table

[3] Similar was found with dynamical fermions, see the Supplementary Information.

1 under the label "N_t-scan". The results can be seen in Extended Data Figure 5, where the charge distributions for four different lattice spacings are plotted. Since the temperature was high, the system did not explore configurations with $Q > 1$. The non-zero width of the distributions is a lattice artefact. We can clearly observe that the peaks get sharper towards the continuum. Also the center of the peak gets closer to 1. This is expected, since our definition of Q, which is evaluated along the Wilson-flow, is renormalized. These center values are also given on the plot, and are denoted by z. We found z is compatible with a $1+c/N_t^2$ behaviour. Thus these peaks at finite lattice spacing correspond to the $Q = 1$ sector in the continuum. The z factors can be used as an $\mathcal{O}(1/N_t^2)$ correction to move the peaks of the Q distribution closer to integer values. This correction is optional for the standard evaluation of χ but becomes useful for identifying higher Q sectors especially on coarse lattices. Inclusion of this z-factor corresponds to a $\mathcal{O}(a^2)$ improvement on renormalized quantities.

Making the peaks sharper can be also achieved by using improved gauge actions, like the tree-level Symanzik, Iwasaki or DBW2 actions. They suppress the topological dislocations and produce less tunneling events. For a comparison of the topological properties of these gauge actions see [31]. It can also be useful to improve the definition of Q along the lines presented in [32], which pushes Q towards integer values.

To explore sectors with higher Q, we defined the boundaries of the intervals as $\left(Q - \frac{1}{2}\right) \cdot z$. In the distribution of Q we found sharp peaks, for an example see Extended Data Figure 6, where Q-histograms corresponding to the "Q-scan" simulations are shown. For the parameters see Extended Data Table 1. The peaks are centered approximately around $Q \cdot z$, using the z factor found in the $Q = 1$ simulations. In these runs we went up to $Q = 8$. As it can be seen with increasing Q the charge distributions get broader. We can also observe, that changing the volume at this particular temperature does not have a large effect on the distribution. Note, that the relative weights between the histograms are not included in the plot. These can be determined from the fixed sector integral technique.

It can also happen, that a simulation gets trapped on the predefined sector boundary with a small acceptance ratio. It can be interpreted as a topological dislocation that is trying to disappear in each update, but is not allowed to disappear due to the Metropolis-step. In the presented simulations this happened in one out of 16 simulation streams, on a 8×64^3 lattice at $\beta = 7.30$ in sector $Q = 8$. In the gauge action difference the result was consistent with the untrapped streams, thus the inclusion or non-inclusion of this stream did not change the value of b_8. Nevertheless, we discarded this stream from our final analysis, because due to the small acceptance it had a large autocorrelation time and was obviously non-ergodic.

An important issue in fixed topology simulations is ergodicity. We used 16 streams, the starting configurations were picked from a low temperature simulation where topology decorrelated on a timescale of few updates. Therefore the streams can be regarded as independent. After a few thousands of updates the gauge action was consistent among the different streams. As an example, in Extended Data Figure 7 we show the results of the Q-scan runs, see Extended Data Table 1. Plotted is b_Q from Equation (2). The odd-Q sectors are not shown for clarity. The results obtained from different streams are all consistent with one another.

It is also interesting to investigate the Q-dependence of b_Q. As we explained before, a naive expectation is, that sector Q contains Q localized objects, each of them independently contributing b_1 to the total b_Q, thus $b_Q = |Q| \cdot b_1$, see Equation (5). With increasing volume the corrections to this equation should get smaller, due to the cluster decomposition principle. We found that even on 8×16^3 lattices upto $Q = 8$ the gauge action differences are consistent with a linear increase with Q. The lines in Extended Data Figure 7 represent the fit to all streams and charges assuming Equation (5). A good fit quality can be obtained. Based on this finding we used the $8 - 0$ difference in a large volume run 8×64^3, for which measuring the $1 - 0$ difference would have been much more expensive.

Let us now investigate the cutoff and finite size effects in b_1 at a temperature of $T \approx 6T_c$. As we already discussed, at such a high temperature the contributions from $b_{Q\geq 2}$ can be safely neglected when calculating the full susceptibility and the decay exponent is given by $b = b_1 - 4$. The upper plot in Extended Data Figure 8 shows b as a function of the lattice spacing squared in a fixed physical volume,

whereas the middle panel shows it as a function of the aspect ratio N_s/N_t. The parameters of the runs are listed in Extended Data Table 1 under the labels N_s-scan and N_t-scan. Starting from aspect ratio ≈ 3, we see no significant finite size effects. Note, that starting from aspect ratio 6, the boxes are large enough to accommodate non-perturbative length scales, i.e. $LT_c \geq 1$. We see no difference between boxes with perturbative and non-perturbative size.

In the second set of simulations we investigated the temperature dependence of the decay exponent. Based on the above results we took a lattice size of 8×32^3, the parameters can be found in Extended Data Table 1 under the label "temperature scan". The upper panel of Extended Data Figure 9 shows the results for $b = b_1 - 4$. Again we find agreement between the new data and the direct approach. At one temperature we did a simulation on an 8×64^3 lattice, where the exponent was obtained from measuring the difference between the $Q = 8$ and 0 sectors, $b = b_8/8 - 4$. We see no significant finite size effect.

To get the Z_1/Z_0 ratio we performed a direct simulation at a temperature of $T_0 = 1.2T_c$. From this temperature we used a trapezoidal integration of b_1 to obtain the Z_1/Z_0 ratio as the function of temperature, up to $7T_c$. In the lower panel of Extended Data Figure 9 we plot $\chi = 2Z_1/Z_0/(1 + 2Z_1/Z_0)/V$, which can be compared to the lattice result obtained from the direct method [5]. As we already discussed starting from a temperature of $1.7T_c$ the contribution of $Q \geq 2$ can be neglected to obtain the susceptibility. We find a good agreement both for the exponent and the susceptibility itself.

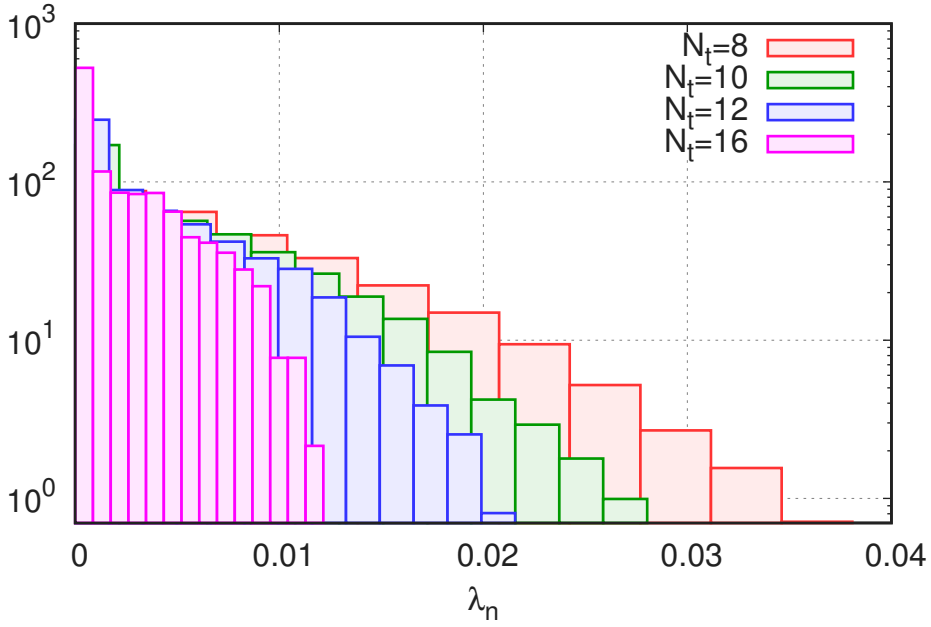
Extended Data Figure 9 was made using 30 million 8×32^3 and 1 million 8×64^3 update sweeps. The cost of a simulation at $T = 7T_c$ using the standard method can be estimated from [5]: it would require about 250 million updates on 8×64^3 lattices or about 2 billion on 8×32^4, two orders of magnitude more, than with the novel method.

Code Availability

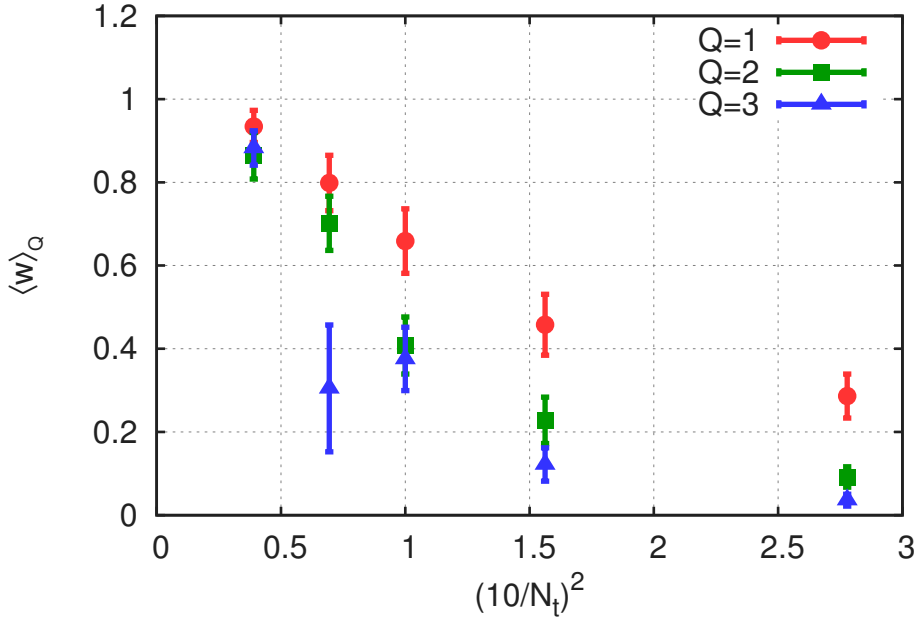
A CPU-code for configuration production can be obtained from the corresponding author upon request. The Wilson flow evolution code, which was used to determine Q, can be downloaded from https://arxiv.org/abs/1203.4469

References

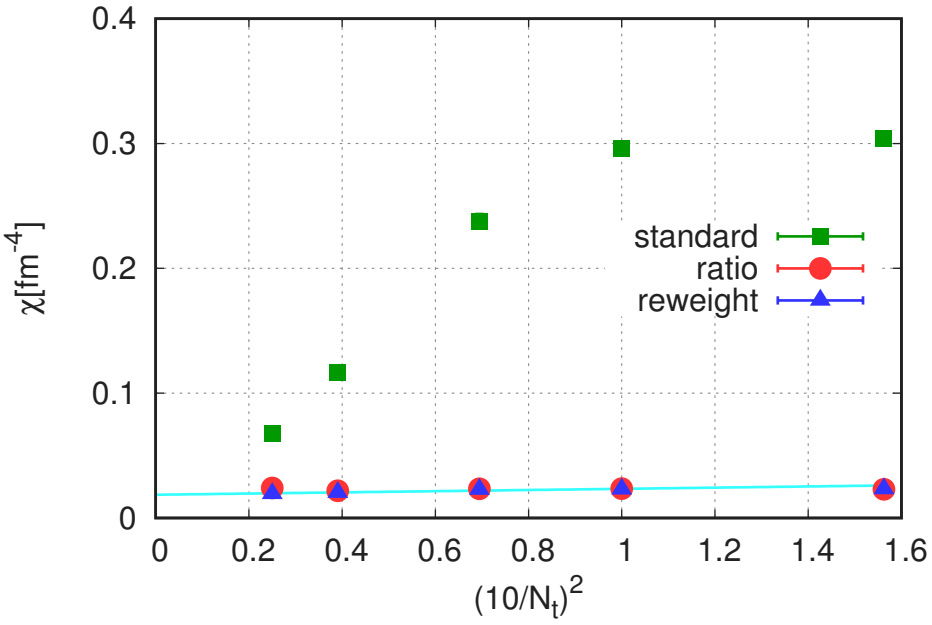
25. Durr, S., Hoelbling, C. & Wenger, U. Staggered eigenvalue mimicry. *Phys. Rev.* **D70,** 094502 (2004).
26. Durr, S. Theoretical issues with staggered fermion simulations. *PoS* **LAT2005,** 021 (2006).
27. Luscher, M. & Schaefer, S. Lattice QCD without topology barriers. *JHEP* **07,** 036 (2011).
28. Laio, A., Martinelli, G. & Sanfilippo, F. Metadynamics Surfing on Topology Barriers: the CP(N-1) Case. arXiv: 1508.07270 [hep-lat] (2015).
29. Mages, S. *et al.* Lattice QCD on Non-Orientable Manifolds. arXiv: 1512.06804 [hep-lat] (2015).
30. Frison, J., Kitano, R., Matsufuru, H., Mori, S. & Yamada, N. Topological susceptibility at high temperature on the lattice. arXiv: 1606.07175 [hep-lat] (2016).
31. DeGrand, T. A., Hasenfratz, A. & Kovacs, T. G. Improving the chiral properties of lattice fermions. *Phys. Rev.* **D67,** 054501 (2003).
32. DeGrand, T. A., Hasenfratz, A. & Kovacs, T. G. Topological structure in the SU(2) vacuum. *Nucl. Phys.* **B505,** 417–441 (1997).



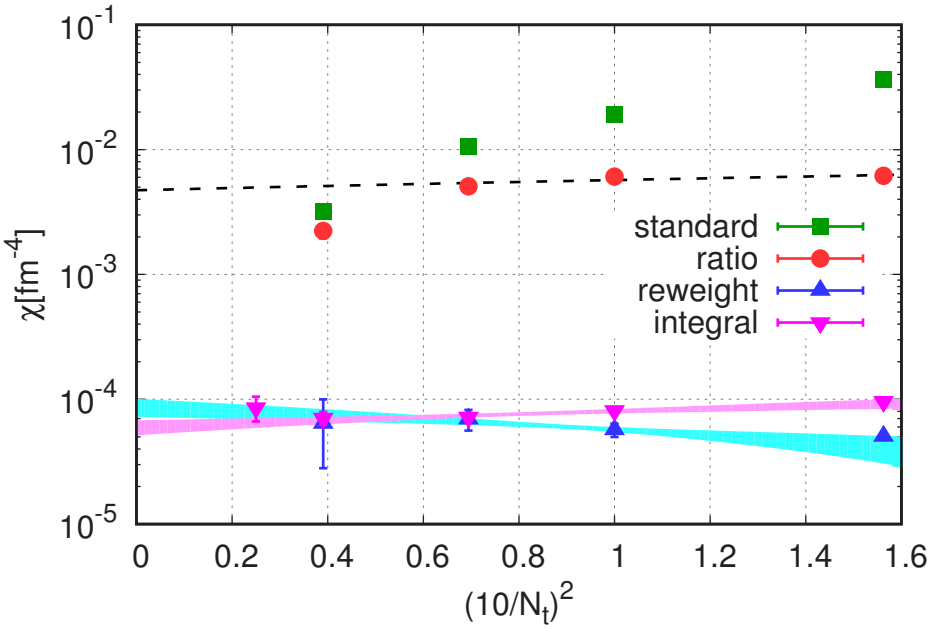
Extended Data Figure 1: The probability distribution of the eigenvalues corresponding to the would-be zero modes. The result is obtained using the chirality method described in the text. The different colors refer to different lattice spacings. The plot shows $n_f = 2 + 1 + 1$ flavor staggered simulations at a temperature of $T = 240$ MeV.



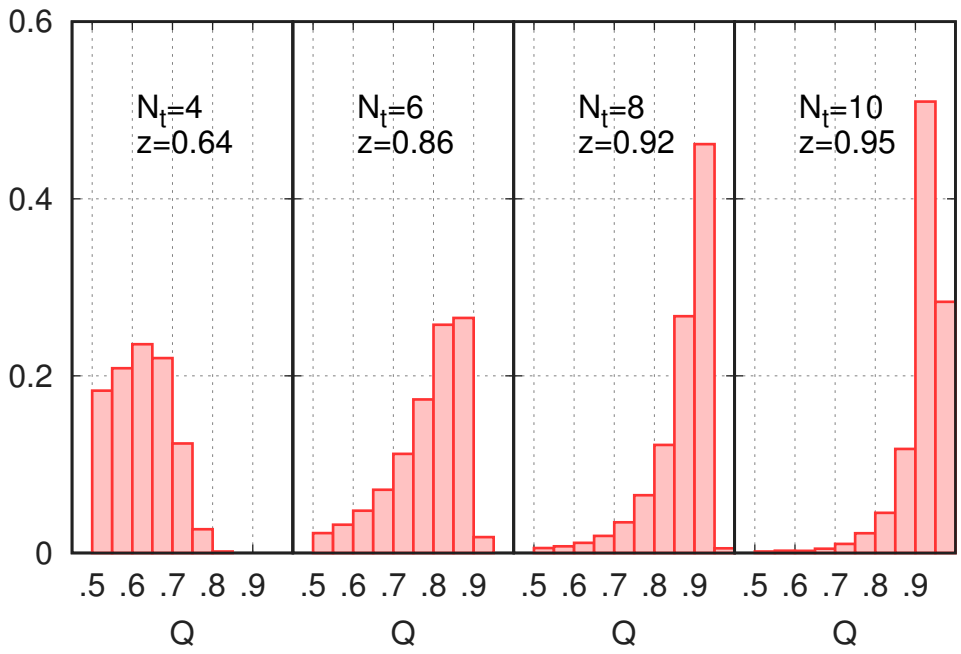
Extended Data Figure 2: Expectation value of the weight factors in different topological sectors, $\langle w \rangle_Q$, as the function of the lattice spacing squared. The plot shows $n_f = 3 + 1$ flavor staggered simulations at a temperature of $T = 300$ MeV. Errorbars are s.e.m.



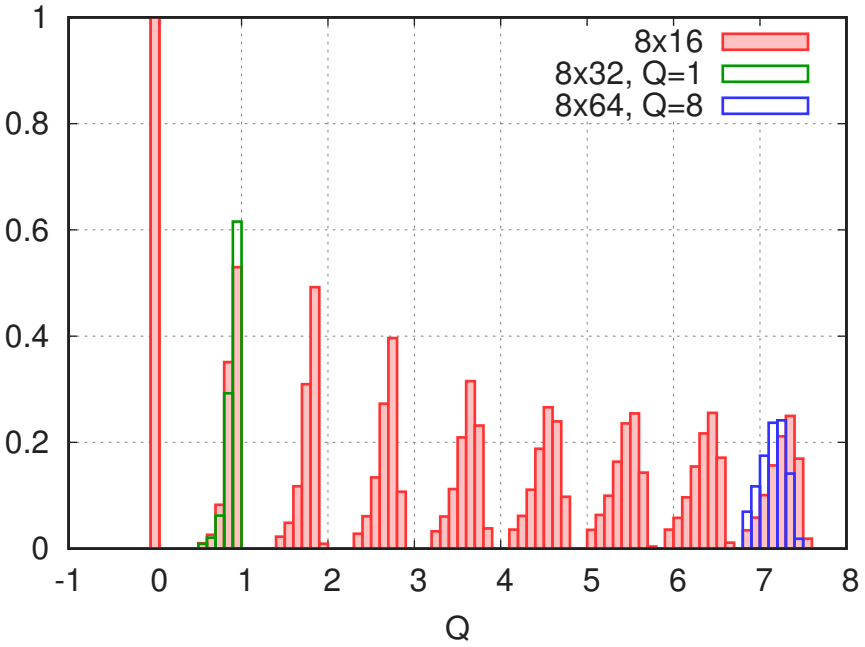
Extended Data Figure 3: Lattice spacing dependence of the topological susceptibility obtained from three different methods described in the text: standard, ratio and reweighting. For the last method a continuum extrapolation is also shown. At this relatively small temperature the standard ("brute force") method still cannot provide three lattice spacings, which extrapolate to the proper continuum limit, though they correspond to very fine lattices with $N_t = 12, 16$ and 20. The plot shows $n_f = 2+1+1$ flavor staggered simulations at a temperature of $T = 150$ MeV. Errorbars (s.e.m.) are smaller than the symbols.



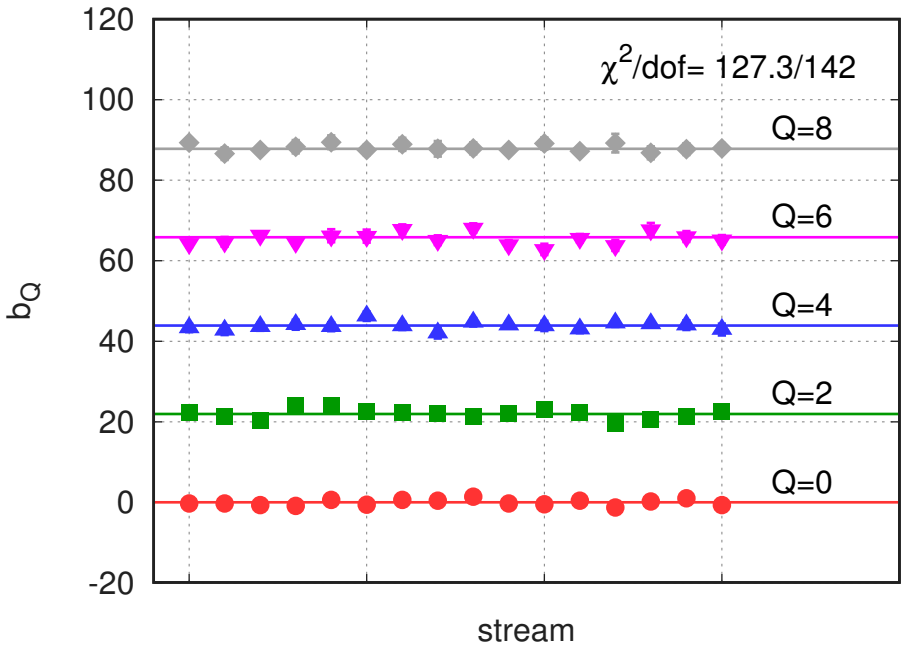
Extended Data Figure 4: Lattice spacing dependence of the topological susceptibility obtained from four different methods described in the text: standard, ratio, reweighting and integral. For the ratio method a misleading continuum extrapolation using $N_t = 8, 10$ and 12 is shown with dashed line. For the reweighting and integral methods continuum extrapolations are shown with bands. The plot shows $n_f = 2+1+1$ flavor staggered simulations at a temperature of $T = 300$ MeV. Errorbars are s.e.m.



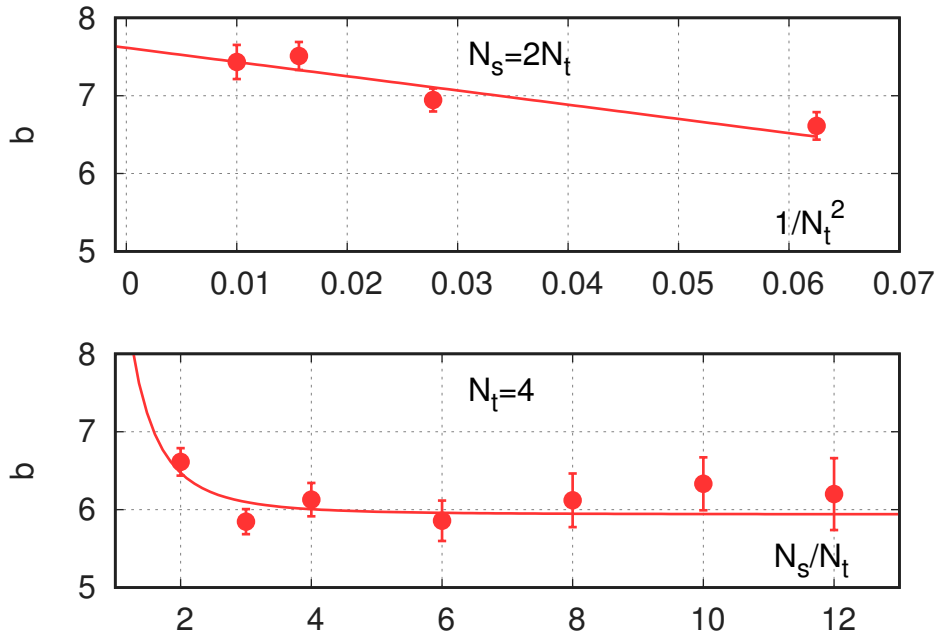
Extended Data Figure 5: Lattice spacing dependence of the charge distribution in simulations under the constraint $Q > 0.5$. The center of the peaks, denoted by z, is also given. The plot shows pure gauge theory simulations at $T \approx 6T_c$.



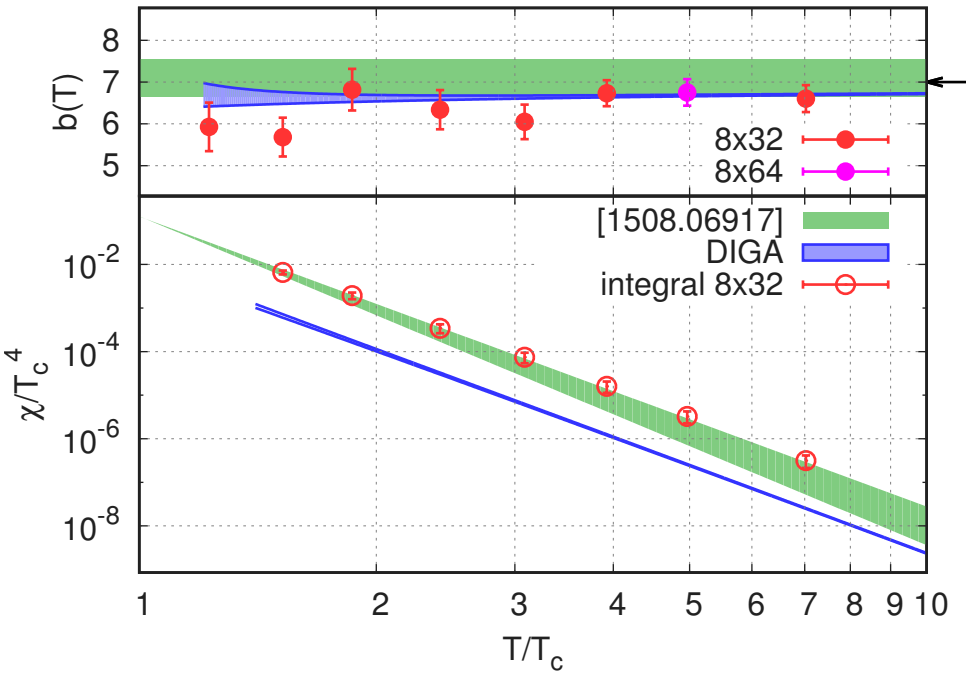
Extended Data Figure 6: Histograms of the topological charge from fixed sector simulations for $Q = 0 \ldots 8$. The sector boundaries are defined using a z-factor as described in the text. Note, that the relative weights between the histograms are not included in the plot. These can be determined from the fixed sector integral technique. The plot shows pure gauge theory simulations at a temperature of $T = 5T_c$.



Extended Data Figure 7: Gauge action difference as defined in Equation (2). The different points correspond to independent simulations and different topological sectors. A good fit can be obtained assuming ergodicity and that the action difference scales linearly with the topological charge, see Equation (5). The plot shows pure gauge theory simulations on 8×16 lattices at $T = 5T_c$ temperature. Errorbars (s.e.m.) are smaller than the symbols.



Extended Data Figure 8: Lattice spacing (top) and finite volume (bottom) dependence of the decay exponent of the topological susceptibility b. The lines are obtained from a joint fit, which takes into account both finite spacing and finite size effects. For the exponent we obtain $b = 7.1(3)$ in the continuum and infinite volume limit at this particular temperature. This is in good agreement with our previous estimate from the direct method [5], where we obtained $b = 7.1(4)$. The plot shows pure gauge theory simulations at $T = 6T_c$ temperature. Errorbars are s.e.m.



Extended Data Figure 9: Topological susceptibility in the pure gauge theory. Results shown from an earlier direct simulation [5] and from the novel fixed Q integral method. Upper plot is the decay exponent b, the lower the susceptibility itself. The arrow indicates the Stefan-Boltzmann limit. We also show the result from the DIGA. The necessary formulas can be found in eg. [33]. To convert the result into units of T_c we used $T_c/\Lambda_{\overline{MS}} = 1.26$ from [15]. Three different renormalization scales were used to test the scheme dependence: 1, 2 and $1/2$ times πT. For the exponent b we see a good agreement for temperatures above $\sim 4T_c$, for smaller temperatures the lattice tends to give smaller values than the DIGA. In case of the susceptibility the DIGA underestimates the lattice result by about an order of magnitude, this was already observed in [5]. The ratio at $T = 2.4T_c$ is $K = 11.1(2.6)$, where the error is dominated by the lattice calculation. Errorbars are s.e.m.

β	T/T_c	$N_s \times N_t$	Mupdates	Q	acc. $Q \neq 0$
N_s-scan					
6.90	6.2	12×4	1.3	0,1	73%
		16×4	1.7	0,1	73%
		24×4	4.3	0,1	73%
		32×4	5.8	0,1	73%
		40×4	24	0,1	73%
		48×4	28	0,1	73%
N_t-scan					
6.90	6.2	8×4	0.3	0,1	72%
7.23	6.1	12×6	1.3	0,1	92%
7.46	6.0	16×8	2.9	0,1	92%
7.65	6.0	20×10	4.1	0,1	98%
Q-scan					
7.30	5.0	16×8	0.7	0	-
			0.7	1	92%
			0.7	2	88%
			0.7	3	85%
			2.6	4	83%
			2.4	5	81%
			2.4	6	78%
			2.3	7	76%
			2.3	8	73%
temperature scan					
6.20	1.2	32×8	3.7	0,1	88%
6.35	1.5	32×8	3.7	0,1	93%
6.50	1.9	32×8	3.7	0,1	94%
6.70	2.4	32×8	3.7	0,1	94%
6.90	3.1	32×8	3.7	0,1	94%
7.10	3.9	32×8	3.7	0,1	94%
7.30	5.0	32×8	3.7	0,1	94%
7.30	5.0	64×8	1.2	0,8	64%
7.60	7.0	32×8	3.7	0,1	94%

Extended Data Table 1: Parameters of fixed sector simulations in the pure gauge theory. In the last column acceptance rates in the $Q > 0$ sectors are given. In the trivial sector we always achieved an acceptance of 100%, which just reflects the fact, that at high temperatures the probability of exploring non-trivial topologies is very small. In the $Q = 1$ sector the acceptance was about 70% on the coarsest lattice. For this we had to switch off the overrelaxation step, which makes large moves in the configuration space, and would have almost always resulted in a topology change. On finer lattices the acceptance was around 90% or better even in the presence of overrelaxation. In the $Q > 1$ sectors the acceptance was gradually decreasing with increasing charge, for which a simple explanation is, that the disappearance probability of multiple instantons is approximately the sum of the individual disappearance probabilities. The worst acceptance was around 65% on a 8×64^3 lattice in sector $Q = 8$.

Supplementary Information

Calculation of the axion mass based on high temperature lattice quantum chromodynamics

Sz. Borsanyi[1], Z. Fodor[1,2,3], J. Gunther[1], K.-H. Kampert[1], S. D. Katz[3,4], T. Kawanai[2], T. G. Kovacs[5], S. W. Mages[2], A. Pasztor[1], F. Pittler[3,4], J. Redondo[6,7], A. Ringwald[8], K. K. Szabo[1,2]

[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] Institute for Nuclear Research of the Hungarian Academy of Sciences, H-4026 Debrecen, Hungary
[6] University of Zaragoza, E-50009 Zaragoza, Spain
[7] Max Planck Institut für Physik, D-80803, Germany
[8] Deutsches Elektronen-Synchrotron DESY, D-22607 Hamburg, Germany

In the following sections we provide details of the work presented in the main paper. We start with giving the details of our pure gauge simulations, Section S1. In Section S2 we summarize the simulation setup for our staggered lattice QCD calculations. This is the basis for the determination of the equation of state and one of the two key elements of the topological susceptibility calculations. In Section S3 we discuss our simulations with overlap fermions using even and odd number of flavors. The line of constant physics is determined.

Section S4 contains the technical details for the EoS focusing on the lattice QCD part, Section S5 presents the perturbative methods to determine the QCD equation of state. Section S6 lists the complete results for the equation of state both in the $n_f = 2+1+1$ and in the $n_f = 2+1+1+1$ frameworks. For cosmological applications we give the effective degrees of freedom for a wide temperature range, starting from 1 MeV all the way up to 500 GeV.

In Sections S7, S8 we apply the fixed sector integral and reweighting methods for the real physical case using staggered and overlap fermions. It is interesting to mention that for the determination of the topological susceptibility overlap fermions turned out to be the less CPU-demanding fermion formulation. In Section S9 we present the details of the analysis procedure that was used to obtain the result for the topological susceptibility. The non-perturbative lattice findings are compared with those of the dilute instanton gas approximation. Section S10 contains the comparison with other recent works.

In Section S11 we use the equation of state and the topological susceptibility results to make predictions for axion cosmology. Both the pre-inflation and the post-inflation Peccei-Quinn symmetry breaking scenarios are discussed. In Section S12 we present experimental setups, which could explore axions in the predicted post-inflation mass range.

S1 Quenched simulations

We used the Wilson-plaquette action. For scale setting we took the Sommer-scale, r_0, parameterized according to the formula of [S1]. To convert the temperatures into units of T_c we took $r_0 T_c = 0.75$. Configuration update was made by alternating overrelaxation and heatbath steps. One update sweep is defined as a heatbath step plus a certain number of overrelaxation steps, the number depending on the lattice spacing. The topological charge was defined using the standard clover definition after applying a Wilson-flow for a flow time of $(8T^2)^{-1}$.

To implement fixed sector simulations we applied a constraint on Q: we added a Metropolis step to the end of each update, that rejected configurations if Q escaped from a pre-defined interval. At high temperatures it was enough to fix only that end of the interval, which was closer to 0, since the system did not attempt to cross the other end. Since the Metropolis step is a global update, one has to make sure that the configuration update satisfies detailed balance, see [S2]. This can be achieved by symmetrizing the ordering of the site and direction loops in the overrelaxation and heatbath steps.

S2 Staggered simulations

For the majority of the results in this paper we use a four flavor staggered action with 4 levels of stout smearing. The action parameters are given in [S3]. The quark masses and the lattice spacing are functions of the gauge coupling:

$$m_s = m_s^{st}(\beta), \quad m_{ud} = R \cdot m_s^{st}(\beta), \quad m_c = C \cdot m_s^{st}(\beta), \quad a = a^{st}(\beta), \tag{S1}$$

these sets of functions are called the Lines of Constant Physics (LCP), and are also given in [S3]. For the quark mass ratios we use $1/R = 27.63$ and $C = 11.85$, which are in good agreement with recent large scale lattice QCD simulations [S4, S5].

In addition to the algorithms mentioned in [S3] we now used a force gradient time integrator [S6, S7] to generate gauge configurations.

Throughout this paper the index f labels the quark flavors, $f = \{ud, s, c\}$, and N_t and N_s are the number of lattice points in the temporal and spatial directions. The temperature T is introduced in the fixed-N_t approach of thermodynamics, ie. $T = (aN_t)^{-1}$, which can be changed by the parameter β while N_t and N_s are fixed. The dimensionful four-volume is given by $V = N_s^3 N_t a^4$. The quark masses m_f are given in lattice units, if not indicated otherwise.

Two different sets of staggered ensembles are used in this paper: a physical $n_f = 2 + 1 + 1$ flavor simulation set and a three flavor symmetric $n_f = 3 + 1$ simulation set. In the following we describe them in more detail.

S2.1 Physical point $n_f = 2 + 1 + 1$ simulations

The lattice geometries and statistics of the $n_f = 2 + 1 + 1$ simulations at zero and finite temperature, are described in [S3]. The quark masses are set to their physical values. In a few cases we increased the statistics of the existing ensembles.

On these ensembles we evaluated the clover definition of the topological charge Q after applying a Wilson-flow [S8]. We used an adaptive step size integration scheme to reduce the computational time. For the finite temperature ensembles we chose a flow time of $(8T^2)^{-1}$, whereas at zero temperature $t = w_0^2$, where the w_0 scale is defined in [S9]. In the systematic error analyses we usually allow for a variation in the flow time. As was shown in our quenched study at finite temperature [S10] the susceptibility reaches a plateau for large flow times. The above choices are nicely in this plateau region even on coarse lattices. The so obtained charge is not necessarily an integer number. To evaluate the topological susceptibility we usually considered both definitions: with and without rounding the topological charge, the difference

between the two is used to estimate systematic errors. When the $\langle Q^2 \rangle$ was large, or close to the continuum limit the rounding did not change the results significantly.

For some of the finite temperature ensembles we also calculated the eigenvalues λ and eigenvectors of the staggered Dirac-operator, D_{st}. To solve the eigenproblem of D_{st} we used a variant of the symmetric Krylov-Schur algorithm described in [S11]. The chirality of these eigenmodes was determined using the staggered taste-singlet γ_5-matrix, which is described in [S12].

S2.2 Topological susceptibility in the vacuum

Due to the index theorem, in the background of a gauge field with non-zero topological charge Q, the Dirac operator has at least $|Q|$ exact zero eigenvalues [S13]. These zero-modes have chirality $+1$ or -1. This is, however, true only in the continuum theory. On the lattice, due to cut-off effects, a non-chiral Dirac operator, like the staggered operator, does not have exact zero eigenvalues, only close to zero small eigenvalues. Also the magnitude of the chirality of these would-be zero modes is smaller than unity. In the continuum limit of the lattice model the would-be zero modes become zero modes with chirality of unit magnitude. However, at any nonzero lattice spacing the lack of exact zero modes results in cut-off effects that can be unexpectedly large for some observables. This is shown for the topological susceptibility in Figure S1, where χ is plotted as function of the lattice spacing squared. The result changes almost an order of magnitude by moving from the coarsest to the finest lattice spacing on the plot.

At zero temperature χ is proportional to the pion mass squared in the continuum. On the lattice with staggered fermions it is expected, that χ will be proportional to the mass squared of the taste singlet pion. For the staggered chiral perturbation theory analysis of χ, see [S14]. Thus it is natural to rescale χ with the pseudo-Goldstone mass squared over the taste-singlet pion mass squared. Since in our $n_f = 2 + 1 + 1$ simulations physical pion pseudo-Goldstone masses are used, in Figure S1 we plot χ multiplied by $(m_{\pi,phys}/m_{\pi,ts})^2$, where $m_{\pi,ts}$ is the taste-single pion mass [S15, S16]. The data shows much smaller cut-off effects, than without multiplication and a nice a^2-scaling sets in starting from a lattice spacing of about 0.1 fm. The continuum extrapolated value is

$$\chi(T=0) = 0.0245(24)_{\rm stat}(03)_{\rm flow}(12)_{\rm cont}/{\rm fm}^4, \tag{S2}$$

the first error is statistical. The second, systematic error comes from varying the definition of the charge, i.e. the flow time at which the charge is measured. The third error comes from changing the upper limit of the lattice spacing range in the fit. According to leading order chiral perturbation theory [S17]

$$\chi = \Sigma \left(\frac{1}{m_u} + \frac{1}{m_d} + \frac{1}{m_s} + \dots \right)^{-1}, \tag{S3}$$

where Σ is the condensate in the chiral limit and the ellipses stand for higher order terms. Using the values for quark masses and the condensate from [S4, S18] we obtain $\chi_{\rm LO} = 0.0224(12)/{\rm fm}^4$ in the isospin symmetric limit, which is in good agreement with Equation (S2). For the size of the isospin corrections see Section S9.

For the high temperature region the cut-off effects are not supposed to be described by chiral perturbation theory. Other techniques are required to get the large cut-off effects under control, such a technique is presented in the Methods Section.

S2.3 Three flavor symmetric $n_f = 3 + 1$ simulations

As it will be described in later Sections as an intermediate step we used results from simulations at the three flavor symmetric point, i.e. where all the light-quark masses are set to the physical strange quark mass $m_s^{st}(\beta)$ and the ratio of the charm and the 3 degenerate flavour masses is $C = 11.85$.

In this theory for each gauge coupling β one can measure the pseudo-scalar mass m_π and the Wilson-flow based w_0-scale. The dimensionless product $m_\pi w_0$ as well as the w_0 in physical units, i.e. $w_0 a^{st}(\beta)$,

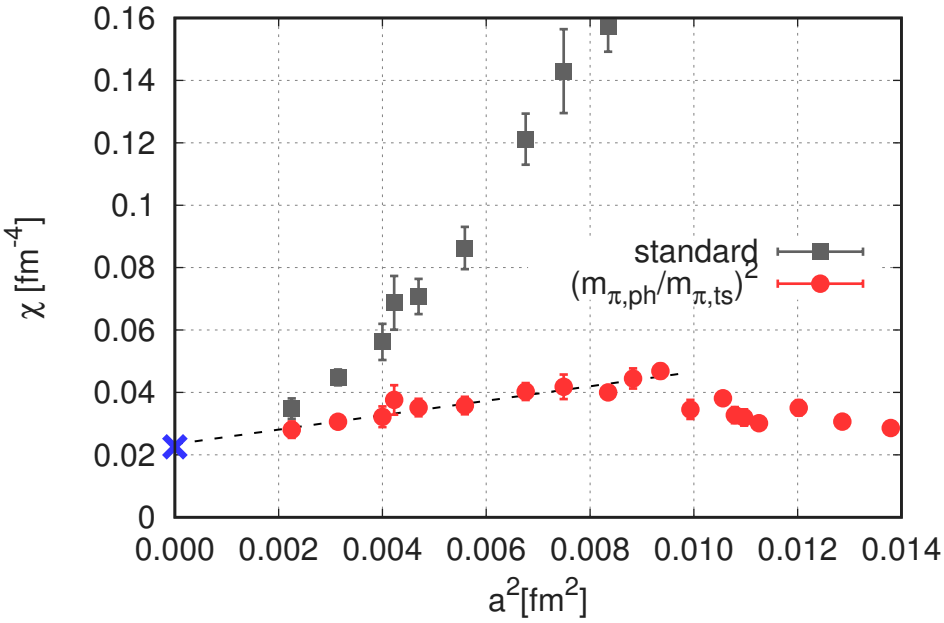


Figure S1: Lattice spacing dependence of the zero temperature topological susceptibility. The grey squares are obtained with the standard approach, the red circles after dividing by the taste singlet pion mass squared. The line is a linear fit. The blue cross corresponds to leading order chiral perturbation theory. The plot shows $n_f = 2+1+1$ flavor staggered simulations at zero temperature.

have well defined continuum limits, since the $n_f = 3+1$ and $n_f = 2+1+1$ theories differ only in the masses of quarks, that do not play a role in the ultraviolet behaviour of those theories. So we end up with the continuum value of the three flavor pseudo-scalar mass $m_\pi^{(3)}$, and that of the w_0-scale $w_0^{(3)}$. We performed $n_f = 3+1$ flavor zero temperature simulations in 64×32^3 volumes for seven lattice spacings ranging between $a = 0.15$ and 0.06 fm. We measured w_0 and $m_\pi w_0$ and performed a continuum extrapolation. This is shown in Figure S2. We obtain the continuum values

$$w_0^{(3)} = 0.153(1) \text{ fm} \quad \text{and} \quad m_\pi^{(3)} = 712(5) \text{MeV}. \tag{S4}$$

In the $n_f = 3+1$ theory we performed finite temperature simulations and measured the same observables as in the $n_f = 2+1+1$ case. Additionally, for the new integral method described in Section S7, we generated configurations at fixed topology. This was achieved by measuring the topological charge after each Hybrid Monte-Carlo trajectory and rejecting the configuration in case of topology change. The typical acceptance rates were 40% on the coarsest lattices and higher on the finer ones. The finite temperature ensembles, unconstrained and fixed topology, are listed in the analysis section, Section S9.

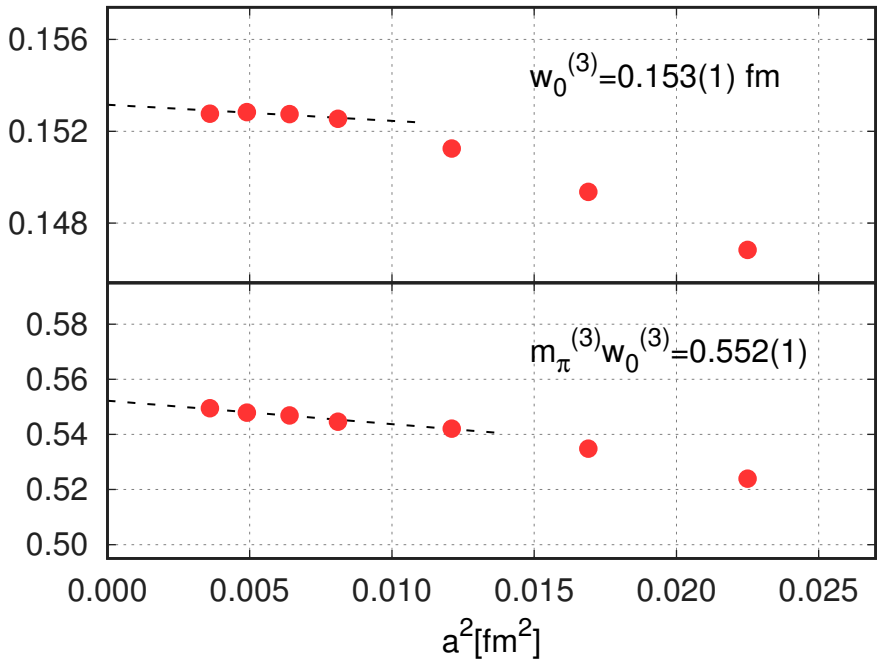


Figure S2: Continuum extrapolations of the w_0-scale in lattice units (up) and $m_\pi w_0$ (down). The plots show $n_f = 3 + 1$ flavor staggered simulations at zero temperature.

S3 Overlap simulations

One of the results of this paper, namely the mass dependence of the topological susceptibility, is obtained using overlap fermions. In this Section we give a short summary of the numerical simulations with overlap quarks used in this work. We also describe our method to determine the Lines of Constant Physics with overlap fermions.

Our setup is based on the one used in Refs. [S19, S20]. For completeness we give a brief summary here:

- tree level Symanzik improved gauge action with gauge coupling parameter β.
- three flavors of overlap quarks. The sign function in the Dirac operator D_{ov} is computed using the Zolotarev approximation. The Dirac operator is constructed from a Wilson operator D_W with mass -1.3. The quark fields are coupled to two step HEX smeared [S21] gauge fields.
- two flavors of Wilson fermions using the above D_W operator.
- two scalar fields with mass 0.54.

The extra Wilson-fermion fields are required to fix the topological charge [S22] and to avoid difficulties when topology changing is required [S23, S24]. These fields are irrelevant in the continuum limit. Note, that their action does not constrain the topology but suppresses the probability of the low lying modes of D_W. In a continuous update algorithm, no eigenmode of D_W can cross zero, which is equivalent to having a fixed topology. The role of the boson fields is to cancel the ultraviolet modes of the extra Wilson fermions. These boson fields are also irrelevant in the continuum limit.

S3.1 Odd flavor algorithm

The main difference to the works in [S19, S20] is, that here we use $n_f = 2 + 1$ flavors instead of $n_f = 2$. The simulations are done using the standard Hybrid Monte Carlo (HMC) algorithm. For the strange quark we use the chiral decomposition suggested first in Ref. [S25] and later in [S26]. The square of the Hermitian Dirac operator $H_{ov}^2 = (\gamma_5 D_{ov})^2$ can be decomposed as $H_{\pm}^2 = P_{\pm} H_{ov}^2 P_{\pm}$ with $P_{\pm} = (1 \pm \gamma_5)/2$. It is trivial to show that the determinant of the one flavor Dirac operator is $\det D_{ov} \sim \det H_{\pm}^2$ where the proportionality constants depend only on the topological charge. Since we do our runs at fixed topology, these constants can be factored out from the partition function and become irrelevant. Therefore a straightforward HMC with either of the $H_{\pm}^2$ operators corresponds to simulating a single flavor at fixed topology. The actual implementation is very simple: the pseudo-fermion generated at the beginning of each trajectory of the HMC has to be projected to one of the chirality sectors.

We tested the simulation code by comparing its results to a brute force update, in which the configurations are picked from a pure gauge heatbath and subsequently reweighted by the exact determinant. Since the determinant calculation is expensive, we ran the test on a small lattice, 4^4. We found a perfect agreement between the two updating algorithms, as shown in Figure S3. We did tests by running the code with two copies of $n_f = 1$ fields and comparing the results obtained with our previous code for $n_f = 2$ and agreement was found in this case, too.

S3.2 Determination of the LCP

We now show how to determine the LCP for physical quark masses in the overlap formalism. It usually requires zero temperature simulations at the physical point, which is for overlap quarks prohibitively expensive with todays computer resources. Fortunately, these are not needed. Here we present an alternative, cost efficient strategy to determine the physical LCP with overlap quarks, based on the physical LCP, that is already known in a different, less expensive fermion formulation. In our case this will be an LCP with staggered fermions, which we have described in the previous Section, see Equation (S1).

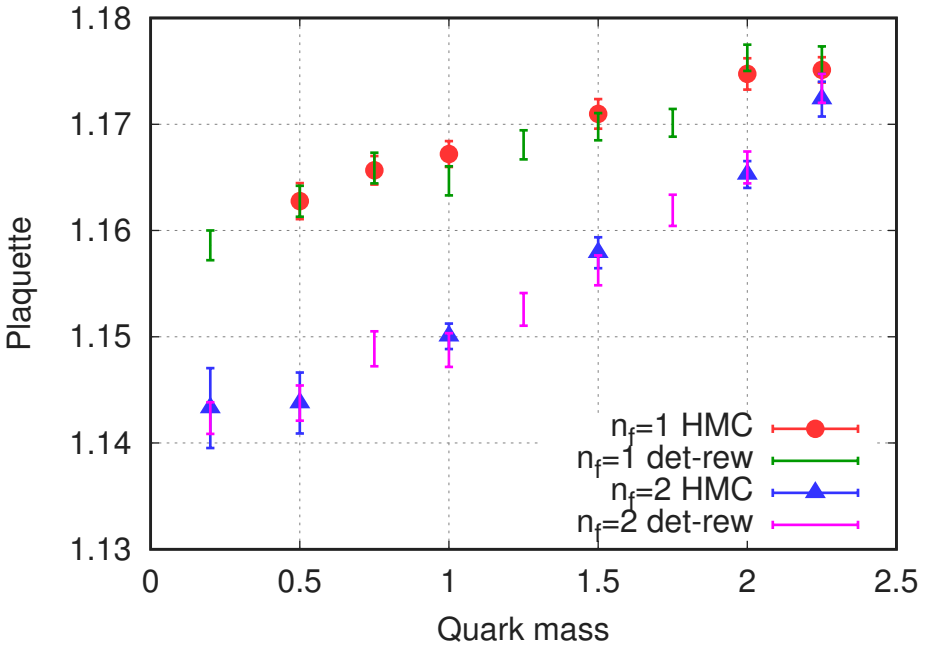


Figure S3: Comparison of the plaquette obtained by two different updating algorithms: the HMC algorithm and determinant reweighting. Shown are the average plaquette in the simulation as a function of the quark mass with both updating algorithms. The plot shows $n_f = 1$ and $n_f = 2$ flavor overlap simulations on a test lattice of size 4^4.

The idea is to use the inexpensive three-flavor symmetric $n_f = 3$ theory as a bridge between the staggered and the overlap LCP's. In Subsection S2.3 we already determined the pion mass and the w_0 scale in the $n_f = 3$ theory in the continuum limit using staggered quark simulations[4]. This can be used to construct an $n_f = 3$ overlap LCP, by tuning the quark mass for each gauge coupling so that $m_\pi w_0 \equiv m_\pi^{(3)} w_0^{(3)}$, and in this way we get the quark mass function $m_s^{ov}(\beta)$. This function is of course not the same as $m_s^{st}(\beta)$, as they are obtained with different fermion discretizations. The important point is, that both define the same physics, e.g. they both give the same $m_\pi w_0$ in the continuum limit. To close the definition of this three-flavor overlap LCP one has to measure $w_0^{ov}(\beta)$, the w_0-scale as a function of the coupling. From this non-physical LCP one can get a physical $n_f = 2 + 1$ LCP with overlap fermions as follows

$$m_s = m_s^{ov}(\beta), \quad m_{ud} = R \cdot m_s^{ov}(\beta), \quad a = w_0^{(3)} / w_0^{ov}(\beta). \tag{S5}$$

Here the value of the lattice spacing at the physical point was obtained by dividing the three-flavor continuum value of w_0 in physical units by the dimensionless three-flavor w_0^{ov}-scale measured in the overlap simulations.

For the $n_f = 3$ flavor overlap LCP we performed simulations at parameters listed in Table S1. From those we determined the quark mass by interpolating to the point, where $m_\pi w_0 = 0.552$. Then we used the formulas of Equation (S5) to obtain the lattice spacing and quark mass parameters for each β. The results are shown in Figure S4. Finally we fitted a three-parameter curve to these points to interpolate to β values, where no simulations were performed. These interpolations are also shown in Figure S4.

[4]Note, that the staggered theory also contained the charm quark, whereas the overlap simulations not. The continuum values of $m_\pi^{(3)}$ and $w_0^{(3)}$ are expected to be insensitive to the presence of the charm.

β	m	$N_s \times N_t$	ntraj
3.80	0.150,0.130	16×32	1000
3.90	0.120,0.100	16×32	1500
4.00	0.090,0.070	16×32	2000
4.05	0.070,0.055	16×32	1200
4.10	0.042,0.032	24×48	2200
4.20	0.042,0.032	24×48	2000
4.30	0.030,0.025	32×64	1600
4.40	0.020,0.030	32×64	1400

Table S1: Gauge coupling parameter, quark mass, lattice size and number of trajectories for the three flavor overlap simulations at zero temperature, that were used to determine the LCP.

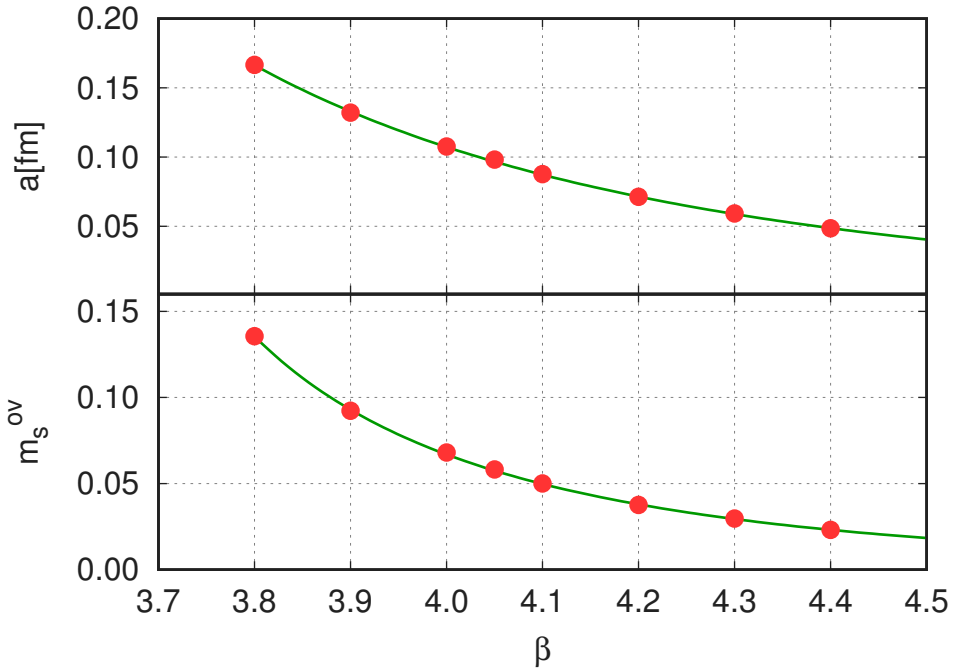


Figure S4: Lines of constant physics in the physical $n_f = 2 + 1$ flavor theory with overlap fermions. The upper plot shows the lattice spacing, the lower plot the strange mass parameter, both as the function of β. The light mass can be obtained as $m_{ud} = R \cdot m_s^{ov}$. The errors are smaller than the symbol size, the lines are smooth interpolations between the points.

S4 Lattice methods for the equation of state

For 2+1 dynamical flavors with physical quark masses we calculated the equation of state in Refs. [S27, S28]. This result was later confirmed in Ref. [S29]. The additional contribution of the charm quark has since been estimated by several authors using perturbation theory [S30], partially quenched lattice simulations [S31, S32] and from simulations with four non-degenerate quarks, but unphysical masses [S33, S34]. However, the final results must come from a dynamical simulation where all quark masses assume their physical values [S35].

We meet this challenge by using the 2+1+1 flavor staggered action (with 4 levels of stout smearing) that we introduced in Section S2. The action parameters, the bare quark masses and the tuning procedure, as well as the lattice geometries and statistics have been given in detail in Ref. [S3].

We calculate the EoS, ie. the temperature dependence of the pressure p, energy density ρ, entropy density s and heat capacity c, from the trace anomaly $I(T)$. This latter is defined as $I = \rho - 3p$, and on the lattice it is given by the following formula:

$$\frac{I(T)}{T^4} = \frac{\rho - 3p}{T^4} = N_t^4 R_\beta \left[\frac{\partial}{\partial \beta} + \sum_f R_f m_f \frac{\partial}{\partial m_f} \right] \frac{\log Z[m_u, m_d, m_s, ...]}{N_t N_s^3} \tag{S6}$$

with

$$R_\beta = -\frac{d\beta}{d \log a} \qquad \text{and} \qquad R_f = \frac{d \log m_f}{d\beta}, \quad f = u, d, s, \ldots . \tag{S7}$$

The derivatives of $\log Z$ with respect to the gauge coupling β and the quark masses m_f are easily accessible observables on the lattice: they are the gauge action S_g and the chiral condensate, respectively.

The R_β and R_f functions describe the running of the coupling and the mass. They can be obtained by differentiating the LCP in Eq. (S1). Since R_β appears as a factor in front of the final result, the systematics of the determination of the LCP directly distorts the trace anomaly. To estimate this uncertainty we use two different LCP's, one based on the w_0-scale and another other on the pion decay constant f_π, which are supposed to give the same continuum limit, but can differ by lattice artefacts. We calculate R_β both from the w_0 and the f_π-based LCP and include the difference in the systematic error estimate. Let us mention, that the R_β and R_f functions are universal at low orders of perturbation theory: e.g. for QCD with n_f flavors we have $R_\beta = 12\beta_0 + 72\beta_1/\beta + \mathcal{O}(\beta^{-2})$ with $\beta_0 = (33 - 2n_f)/48\pi^2$ and $\beta_1 = (306 - 38n_f)/768\pi^2$.

There is an additive, temperature independent divergence in the trace anomaly. In the standard procedure, that we also followed in Ref. [S27], each finite temperature simulation is accompanied by a zero temperature ensemble. The trace anomaly is then calculated on both sets of configurations, their difference is the physical result. This defines a renormalization scheme where the zero temperature pressure and energy density both vanish. Using the configurations already introduced in Ref. [S3] and applying the standard method that we also described in Ref. [S28] we calculated the trace anomaly, as shown in the left panel of Fig. S5.

This strategy is bound to fail at high temperatures. Since the temperature on the lattice is given by $T = (aN_t)^{-1}$, high temperatures can only be reached with very fine lattices. As the lattice spacing is reduced, the autocorrelation times for zero temperature simulations rise, and the costs of these simulations explode beyond feasibility. Notice, however, that the renormalization constant is accessible not only from zero temperature simulations, but from any finite temperature data point that was taken using the same gauge coupling and quark masses.

In our approach we generate a renormalization ensemble for each finite temperature ensemble at exactly half of its temperature with the same physical volume and matching bare parameters. The trace anomaly difference is then divergence-free. We have tested this idea in the quark-less theory in Refs. [S36, S37].

Thus we determine the quantity $[I(T) - I(T/2)]/T^4$ in the temperature range 250...1000 MeV for four resolutions $N_t = 6, 8, 10$ and 12. The volumes are selected such that $LT_c \gtrsim 2$. In Fig. S5 we show this subtracted trace anomaly and its continuum limit.

We now turn to our systematic uncertainties. We use the histogram method to estimate the systematic

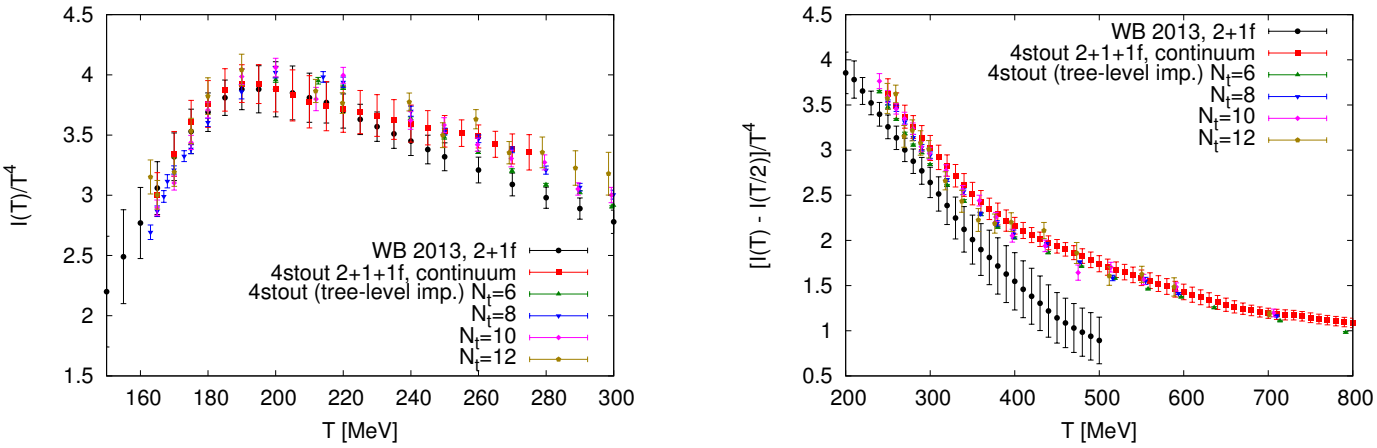


Figure S5: The trace anomaly renormalized with zero temperature simulations (left panel) and the subtracted trace anomaly (right panel) in the 2+1+1 and 2+1 flavor theories. For $T < 300$ MeV the two results agree within our uncertainty. We also show the individual lattice resolutions $(N_t = 6 \ldots 12)$ that contribute to the continuum limit.

errors, this means that we analyze our data in various plausible ways and form a histogram of the results. The median gives a mean, the central 68% defines the systematic error [S38]. Here we use uniform weights dropping the continuum results where the fit quality of the continuum limit was below 0.1. For the details of the analysis we largely follow our earlier work in Ref. [S3]: we interpolated the lattice data in temperature and then we performed a continuum extrapolation in $1/N_t^2$ temperature by temperature. The error bars in Fig. S5 combine the statistical and systematic errors, the latter we estimate by varying the scale setting prescription (w_0-based or f_π-based scale setting), the observable (subtracted trace anomaly or its reciprocal), the interpolation, and whether or not we use tree-level improvement prior to the continuum extrapolation [S27].

Then we use the trace anomaly data with zero temperature renormalization from the left panel of Fig. S5 and extend it towards lower temperatures from our already established 2+1 flavor equation of state result. Then we can calculate $I(T)/T^4$ from the continuum limit of $[I(T) - I(T/2)]/T^4$ using the formula:

$$\frac{I(T)}{T^4} = \sum_{k=0}^{n-1} 2^{-4k} \, \frac{I\left(T/2^k\right) - I\left(T/2^{k+1}\right)}{\left(T/2^k\right)^4} + 2^{-4n} \, \frac{I\left(T/2^n\right)}{\left(T/2^n\right)^4} \;, \tag{S8}$$

where n is the smallest non-negative integer with $T/2^n < 250$ MeV.

The temperature integral of $I(T)/T^5$ gives the normalized pressure $p(T)/T^4$. The energy density, entropy density and heat capacity are then obtained from the thermodynamic relations: $\rho = I + 3p$, $sT = \rho + p$ and $c = d\rho/dT$ respectively.

S5 The QCD equation of state in the perturbative regime

S5.1 Massless perturbation theory

Recent progress in Hard Thermal Loop (HTL) perturbation theory has provided for a next-to-next-to-leading-order (NNLO) result for the free energy, which is in fair agreement with our data both for the 2+1 theory [S39] and also for the 2+1+1 flavor theory for high enough temperatures. The trace anomaly and the pressure are shown in Fig. S6 for both cases.

We also show a comparison to the results of conventional perturbation theory with four massless quarks in Fig. S7. Here $g = \sqrt{4\pi\alpha_s}$ is the QCD coupling constant. The completely known fifth order [S40] result is in good agreement with the lattice data. Note, that whereas the perturbative result treats even the charm quark massless, the lattice EoS includes the mass effects correctly.

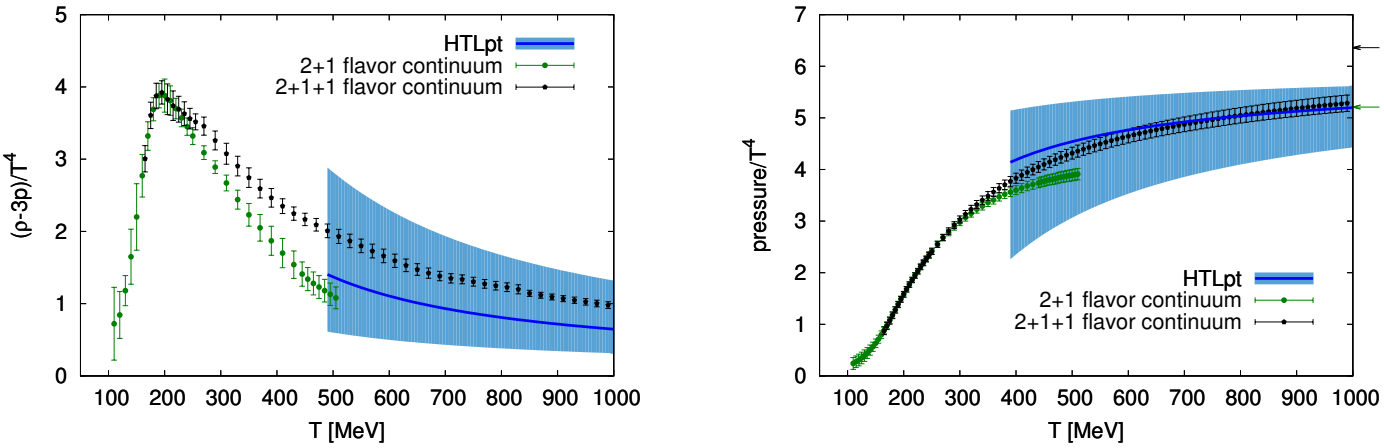


Figure S6: The QCD trace anomaly and pressure in the 2+1+1 and 2+1 flavor theories. We also give the four flavor NNLO HTL result at high temperatures [S39].

S5.2 Charm mass threshold in the QCD equation of state

Thanks to the lattice data that we have generated, we can present non-perturbative results for the charm quark contribution. It is instructive to study the inclusion of the charm quark in detail. This way we can design an analytical technique for the inclusion of the bottom quark, for which the standard formulation of lattice QCD is computationally not feasible.

The quark mass threshold for the charm quark entering the EoS has already been estimated in Ref. [S30]. There, the effect of a heavy quark was calculated to a low order of perturbation theory. This effect was expressed as a pressure ratio between QCD with three light and one heavy flavor and QCD with only three light flavors. When that paper was completed the lattice result for the QCD equation of state was not yet available, but the perturbative methods were already in an advanced state.

Despite the known difficulties of perturbation theory the estimate of Ref. [S30] is very close to our lattice result if we plot the ratio of the pressure with and without the charm quark included. We show our lattice data together with the perturbative estimate in Fig. S8.

Though the individual values for the 2+1+1 and 2+1 flavor pressures of [S30] are not very accurate, their ratio describes well the lattice result. This is true both for the leading and for the next-to-leading order results (See Fig. S8).

The tree-level charm correction is given by

$$\frac{p^{(2+1+1)}(T)}{p^{(2+1)}(T)} = \frac{SB(3) + F_Q(m_c/T)}{SB(3)} \tag{S9}$$

where $SB(n_f)$ is the Stefan Boltzmann limit of the n_f flavor theory, and $F_Q(m/T)T^4$ is the free energy density of a free quark field with mass m. In this paper we used the $\overline{MS}$ mass $m_c(m_c) = 1.29$ GeV [S41].

Order g^2 in the ratio of Fig. S8 starts to be important correction below a temperature of about $2 - 3T_c^{QCD}$ temperature. Near $2T_c$ the difference between the two approximations is 3%. The difference reduces to 0.2% at 1 GeV up to which point we have lattice data.

S5.3 Bottom mass threshold in the QCD equation of state

In the previous discussion we saw that even the tree-level quark mass threshold gives a correct estimate for the equation of state. This allows us to introduce the bottom threshold along the same lines.

First, we remark that one can write the charm threshold relative to the $2 + 1 + 1$ flavor theory:

$$\frac{p^{(2+1+1)}(T)}{p^{(2+1+1)}(T)|_{m_c=0}} = \frac{SB(3) + F_Q(m_c/T)}{SB(4)}\,. \tag{S10}$$

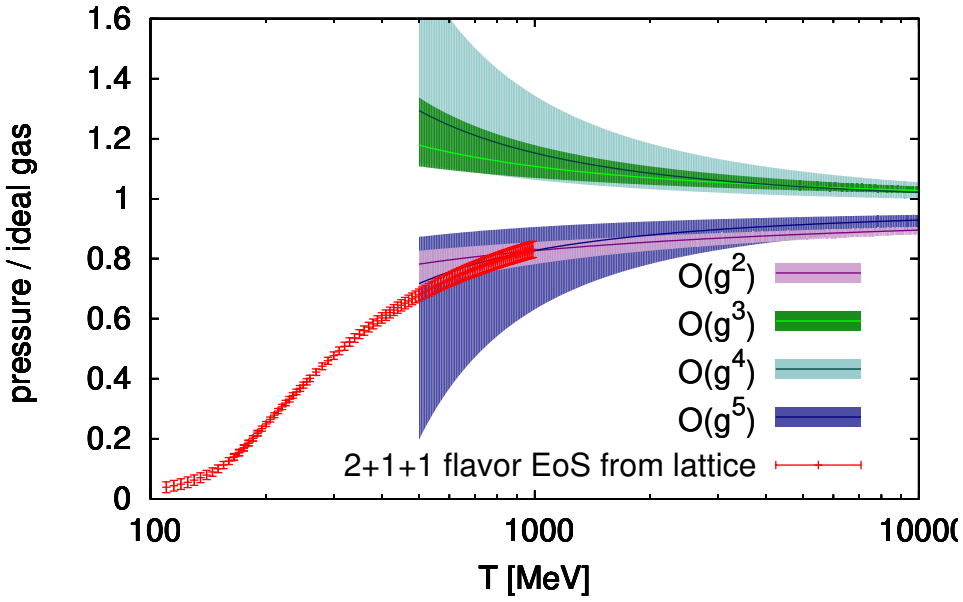


Figure S7: The QCD pressure for 2+1+1 flavors together with various orders of conventional perturbation theory. The renormalization scale is varied in the range $\mu = (1 \dots 4)\pi T$. The middle lines correspond to $\mu = 2\pi T$.

The error of not using the g^2 order is about 0.2% of the total QCD contribution at 1 GeV.

From the lattice data we have $p^{(2+1+1)}(T)$. Using Eq. (S10) we can calculate the QCD pressure for the theory with four light quarks. Perturbation theory can provide just that, at least, if the temperature is high enough. In the following we demonstrate that it is possible to use a perturbative formula that matches our lattice-based $p^{(2+1+1)}(T)\big|_{m_c=0}$ already from approx. 500 MeV, i.e. below the bottom threshold.

The perturbative results have a mild dependence on the choice of the Λ parameter, here we use the standard $\Lambda_{\overline{MS}} = 290$ MeV value for $n_f = 4$ [S41].

The highest fully known order for the QCD pressure is $g^6 \log g$ [S42]. The coefficient of the g^6 order is not known analytically, but the missing term, q_c (following the notation of Ref. [S42]) can be fitted against lattice data. We fix the renormalization scale to $\mu = 2\pi T$. This fitting method has already been applied for the Yang-Mills theory [S37, S42]. The order g^6 result describes our pressure data at 1 GeV within errors if $-3400 < q_c < -2600$, for the trace anomaly we have $-3200 < q_c < -2800$. We propagate this uncertainty into the perturbative result, keeping the range $2700 < -q_c < 3200$. We show the fitted curves for the central choice, $q_c = -3000$, in Fig. S9. The result in the plot has already been converted to the case of a massive charm. Notice, that both for the trace anomaly and for the pressure the $\mathcal{O}(g^6)$ perturbative result follows the lattice data already from 500 MeV.

Having a pressure and trace anomaly function that is valid for the 2+1+1 flavor theory, that agrees with the lattice data below 1 GeV and provide a perturbatively correct continuation towards high temperatures we can proceed to include the effect of the bottom quark. The tree-level correction for the bottom quark reads

$$p^{(2+1+1+1)}(T) = p^{(2+1+1)}(T)\frac{SB(4) + F_Q(m_b/T)}{SB(4)} \tag{S11}$$

where $m_b(m_b) = 4.18$ GeV is the bottom mass [S41]. Eq. (S11) works beyond the bottom threshold, too, since the ratio of the perturbative massless 4 and 5 flavor pressure is, to a very good approximation, equal to the ratio of the Stefan-Boltzmann limits. Comparing the free energy up to order g^5 this statement holds to 0.3% accuracy in the entire relevant temperature range. (For earlier formulations of this idea see Refs. [S43, S44].)

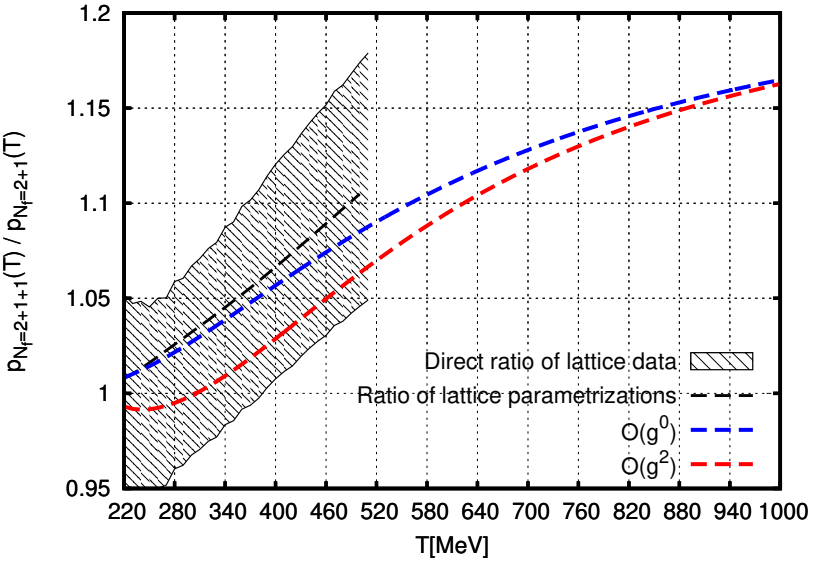


Figure S8: The ratio of the pressure between 2+1+1 flavor and 2+1 flavor theories from our lattice simulations. Note that the parameterization smoothly approaches 1 as we further decrease the temperature (see Eq. (S12)). We also show two perturbative estimates for the ratio of the pressure functions. The tree-level ($\mathcal{O}(g^0)$) estimate (see text), and the $\mathcal{O}(g^2)$ estimate of Ref. [S30]. Both of these estimates show agreement with the lattice data within our accuracy.

Since the massless four flavor perturbative result is used as a starting point for both heavy quarks we have a fair approximation between 500 MeV and 10 GeV with one analytical formula.

S6 Equation of state for 2+1+1 flavor and 2+1+1+1 flavor QCD and for the whole Standard Model

It has been a longstanding challenge to determine the pressure, energy density, and the number of effective degrees of freedom as a function of the temperature from first principles. This is the equation of state of the universe. Cosmology requires this information over a temperature range of many orders of magnitude, ranging from beyond the electroweak scale down to the MeV scale [S44, S45].

As the Universe covers this broad temperature range it passes though several epochs, each with a different dominant interaction. We restrict our study here to the Standard Model of particle physics. At the high end of the temperature range of interest there is the electroweak phase transition at a temperature of about 160 GeV [S46]. The equation of state of the electroweak transition has been worked out in Refs. [S30, S47, S48] with the intent to provide a description for the entire Standard Model. For this reason the contribution from all other degrees of freedom (i.e. up to the bottom quark) had to be estimated. While the photon, neutrinos and leptons can easily be described as practically free particles, the QCD part requires a non-perturbative approach. This was not available when Refs. [S30, S47] were published.

In this paper we add the last missing piece to the cosmological equation of state: the QCD contribution. In this section we give the results of our efforts for the 2+1+1 and 2+1+1+1 flavor theories separately. Finally, we combine all the elements of the Standard Model and present the number of effective degrees of freedom from the energy density and entropy ($g_\rho(T)$ and $g_s(T)$) in the full temperature range.

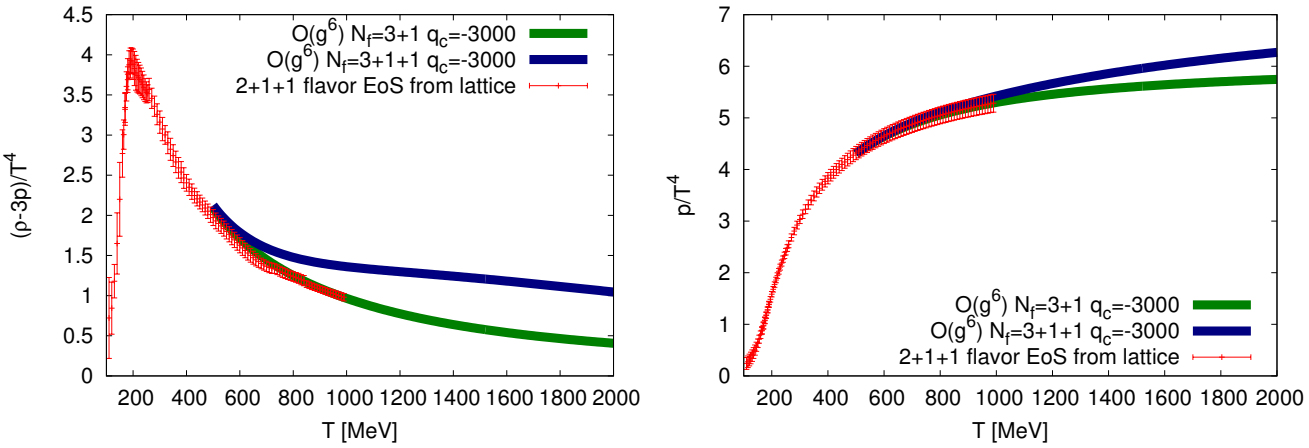


Figure S9: The lattice result for the 2+1+1 flavor QCD pressure together with the fitted value of the g^6 order. We included the charm mass at tree-level. The perturbative result agrees with the data from about 500 MeV temperature. Using the same fitted coefficient we also calculated the effect of the bottom quark with the same method. The blue curve shows the EoS including the bottom contribution.

S6.1 The 2+1+1 flavor QCD equation of state

Now we show the complete result obtained from $n_f = 2+1+1$ lattice QCD. Figure S10 depicts the trace anomaly (left panel) and pressure (right panel). For comparison the 2+1 flavor results are also shown.

Plotting p/T^4 (which is the normalized free energy density), we can compare our result to other approaches. At low temperatures the Hadron Resonance Gas model (using the 2014 PDG spectrum) gives a good description of the lattice data. This was already observed in Ref. [S28].

In Ref. [S28] we gave a simple parameterization for the 2+1 flavor equation of state. Here we update the 2+1 flavor parameters and provide a parameterization that covers the 100-1000 MeV temperature range and describes the 2+1+1 lattice data, i.e. including the effect of the charm quark. As before, the parameterizing formula reads

$$\frac{I(T)}{T^4} = \exp(-h_1/t - h_2/t^2) \cdot \left(h_0 + f_0 \frac{\tanh(f_1 \cdot t + f_2) + 1}{1 + g_1 \cdot t + g_2 \cdot t^2} \right), \tag{S12}$$

with $t = T/200\ \mathrm{MeV}$. The parameters are given in Table. S2, the resulting curves are shown in Fig. S10. For completeness the $n_f = 2+1$ parameterization is also shown.

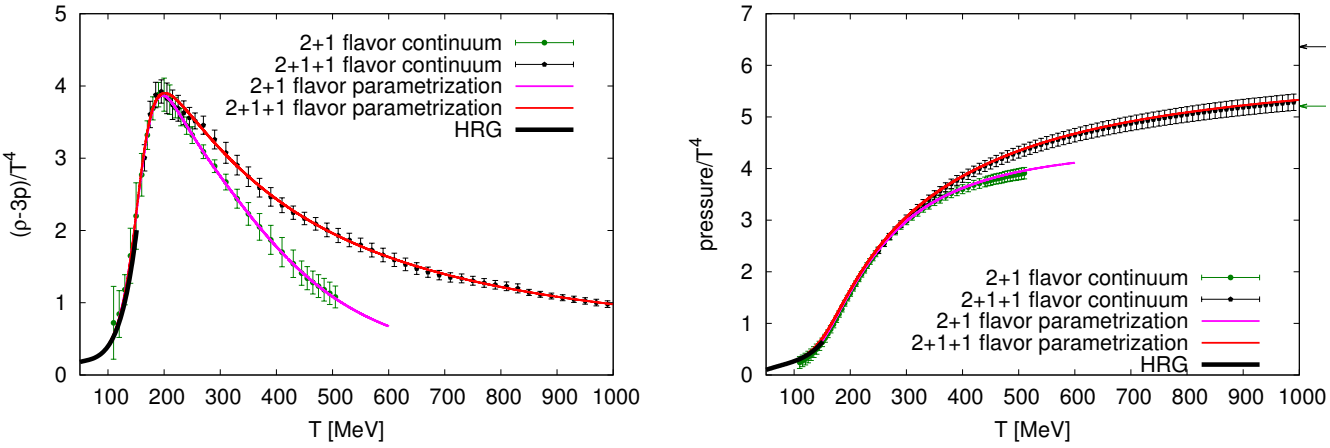


Figure S10: The QCD trace anomaly and pressure in the 2+1+1 and 2+1 flavor theories in our parameterization Eq. (S12). We also show the Hadron Resonance Gas model's prediction for comparison.

	h_0	h_1	h_2	f_0	f_1	f_2	g_1	g_2
2+1+1 flavors	0.353	-1.04	0.534	1.75	6.80	-5.18	0.525	0.160
2+1 flavors	-0.00433	-1.00	-0.288	0.293	6.10	-4.90	-0.787	0.289

Table S2: Constants for our parameterization of the trace anomaly in Eq. (S12).

S6.2 The 2+1+1+1 flavor QCD equation of state

Here we present our final result on 2+1+1+1 flavor QCD. The bottom threshold has been added as described in Sec. S5.3. We use the 2+1 flavor lattice results up to 250 MeV, 2+1+1 flavor data up to 500 MeV. In the range 500...1000 MeV we observed that our $\mathcal{O}(\alpha_s^3)$ order perturbative result agrees very well with the 2+1+1 flavor lattice data. This justifies the use of the $\mathcal{O}(\alpha_s^3)$ formula to include the effect of the bottom quark as described in Sec. S5.3. The effect of the bottom quark starts at a temperature of about 600 MeV. See Fig. S9.

Because of its large mass the top quark can only be included in the framework of the electroweak theory. Thus the calculation of the bottom quark's effect completes the discussion of the QCD contribution to the equation of state. The resulting thermodynamic functions are shown in Fig. S11.

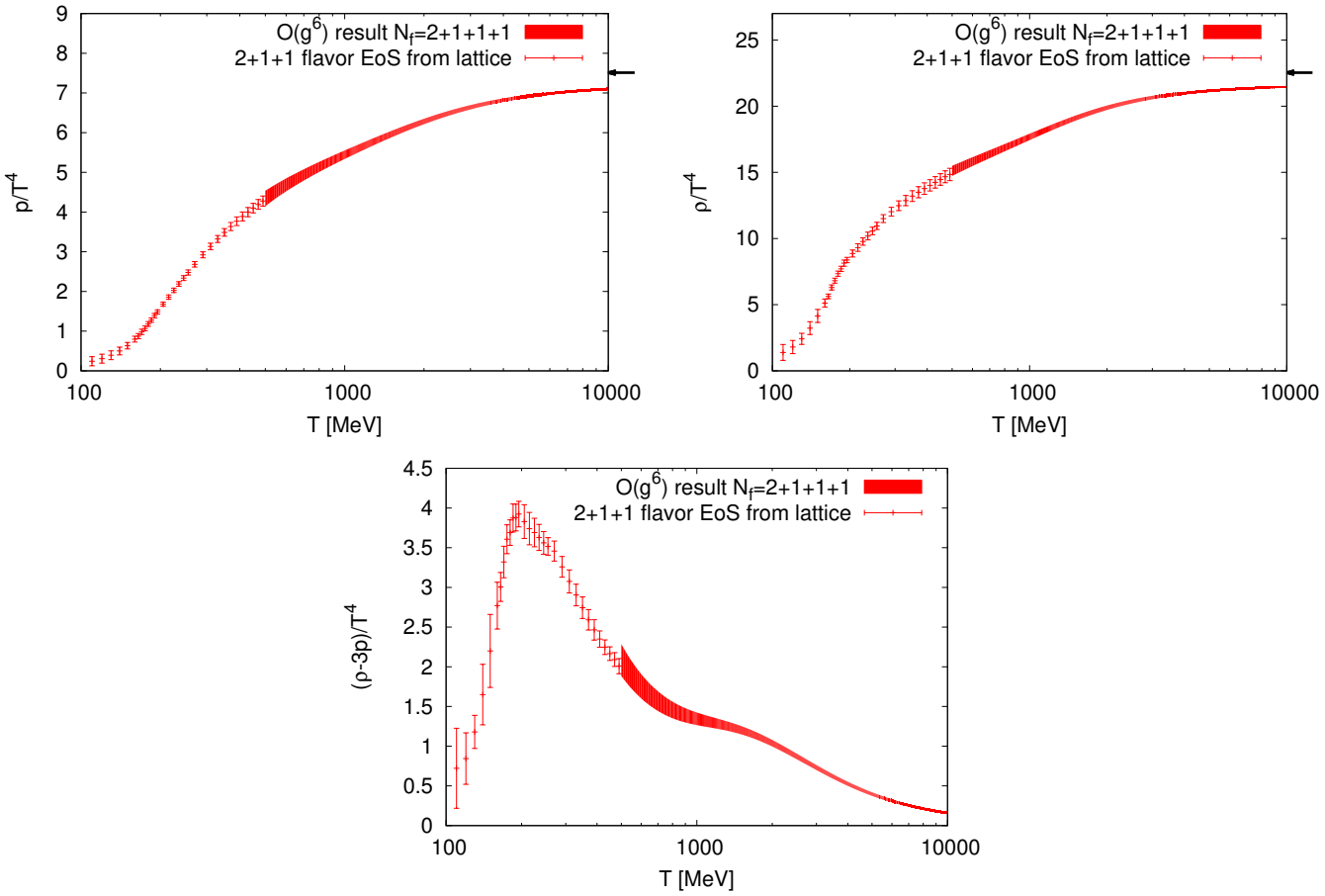


Figure S11: The QCD pressure, energy density and the trace anomaly in the 2+1+1+1 theory.

S6.3 The Standard Model

Now we are in the position to construct an equation of state that gives a good description over the entire temperature range of the Standard Model.

In particular, we compute the effective number of degrees of freedom. This is defined by the energy density or the entropy normalized by the Stefan-Boltzmann limit of a single scalar field:

$$g_\rho(T) = \rho(T)\frac{30}{\pi^2 T^4}\,, \qquad g_s(T) = s(T)\frac{45}{2\pi^2 T^3}\,, \qquad g_c(T) = c(T)\frac{15}{2\pi^2 T^3}. \tag{S13}$$

(These quantities are not to be confused with the strong coupling constant of the previous section.)

The final results for the energy and entropy densities and for the heat capacity can be obtained from the following the thermodynamical identities

$$\rho(T) = 3p(T) + T^5\frac{dp(T)/T^4}{dT}\,, \qquad s(T)T = \rho(T) + p(T)\,, \qquad c(T) = \frac{d\rho}{dT}\,. \tag{S14}$$

a) Photons and neutrinos

We treat these light particles in their Stefan-Boltzmann limit, assuming three generations of left-handed neutrinos. In this paper we work out the equilibrium equation of state, the neutrinos give a trivial contribution of $p_\nu/T^4 = \frac{7}{8}\cdot 2\cdot 3\cdot \pi^2/90$, for the photons we have $p_\gamma/T^4 = \pi^2/45$.

b) Charged leptons

We sum the free energy of the non-interacting leptons with the formula

$$p/T^4 = \frac{1}{2\pi^2}\sum_i g_i \left(\frac{m_i}{T}\right)^2 \sum_{k=1}^{\infty}\frac{(-1)^k}{k^2}K_2\left(\frac{m_i k}{T}\right)\,, \tag{S15}$$

where K_2 is a modified Bessel function of the second kind, and g_i is the spin degeneracy factor, $g_i = 4$ for leptons, and m_i is its mass. The right hand side of Eq. (S15) with $g_i = 12$ gives the free quark contribution $F_Q(m_i/T)$ already introduced in Eq. (S9).

c) Light hadrons

In Refs. [S3, S28] we have tested the Hadron Gas Model's prediction in detail. Here we use this model's prediction up to a temperature of 120 MeV. From that point we switch over to the lattice result.

d) QCD

We have continuum extrapolated lattice results for the equation of state up to a temperature of 1 GeV. We have found these as a combination of 2+1 and 2+1+1 flavor simulations. The lattice data can be smoothly continued with α_s^3 order perturbative result, where one analytically unknown parameter was fitted to our data. In Fig. S9 we showed that the sixth order (highest order with one fitted coefficient) perturbative result gives a good description of both the pressure and the trace anomaly. This pressure function is the basis of our result at high temperatures. We included the bottom threshold using the method described in Sec. S5.3. The full 2+1+1+1 flavor QCD contribution we show in Fig. S11.

e) $W^\pm, Z^0$ and the Higgs boson The bosonic version of Eq. (S15) can be used as a first estimate:

$$p/T^4 = \frac{1}{2\pi^2}\sum_i g_i \left(\frac{m_i}{T}\right)^2 \sum_{k=1}^{\infty}\frac{1}{k^2}K_2\left(\frac{m k}{T}\right)\,. \tag{S16}$$

Ref. [S30] goes beyond this and adds the one-loop electroweak corrections. The one loop corrections become noticeable at the temperature of approximately $T \gtrsim 90\ \mathrm{GeV}$. For this correction we use the data of Ref. [S30].

f) The electroweak transition For the electroweak epoch we quote the results of Ref. [S48]. They use perturbation theory, dimensional reduction [S49] and the results of 3D simulations to estimate the equation of state of the Standard Model near the electroweak transition. This is a continuation of the earlier work [S30]. Although at the time continuum extrapolated electroweak lattice input was not yet available [S50], the final continuum extrapolation shows a very mild lattice spacing dependence [S46].

Adding all components from a) to f) we arrive at our final result for $g_\rho(T)$ and $g_s(T)$ that we plot in the main text. Here we give a cubic spline parameterization for $g_\rho(T)$ and the ratio of $g_\rho(T)$ and $g_s(T)$, see Table S3.

$\log_{10}(T/\mathrm{MeV})$	g_ρ	g_ρ/g_s	$g_s/g_c = 3v_s^2$
0.000	10.681	1.00264	0.99418
0.300	10.727	1.00071	0.99846
0.700	10.751	1.00016	0.99976
1.050	10.786	1.00055	0.99506
1.100	10.822	1.00103	0.99141
1.250	11.100	1.00518	0.95992
1.537	13.125	1.01986	0.91001
1.800	15.469	1.01997	0.92730
2.000	17.599	1.02267	0.88343
2.070	19.275	1.03161	0.81345
2.178	26.529	1.06705	0.64528
2.250	35.759	1.08461	0.69624
2.350	44.400	1.07079	0.79693
2.500	52.855	1.04679	0.86461
3.000	73.713	1.01900	0.94167
3.500	81.460	1.00524	0.98787
4.000	83.254	1.00143	0.99286
4.300	85.527	1.00374	0.97957
4.600	92.144	1.00876	0.96057
5.000	101.937	1.00755	0.97431
5.180	104.508	1.00510	0.98139
5.190	104.664	1.00513	0.97557
5.196	104.816	1.00524	0.96942
5.202	104.909	1.00524	0.98837
5.210	104.919	1.00499	0.99671
5.300	104.781	1.00244	0.99838
5.450	104.802	1.00090	0.99832

Table S3: Data set on a logarithmic scale that can be used with simple cubic spline interpolation to find a parameterization for the entire Standard Model. The spline's typical deviation from g_ρ is about 1%, and about 0.3% for the two ratios.

Ref. [S48] has constructed an equation of state of the universe. In their work only the electroweak theory was based on lattice simulations. Here we replace the earlier perturbative deliberations on the QCD epoch by fully controlled lattice QCD result, which we conveniently parameterize. In Fig. S12 we show this non-perturbative effect by comparing our result to the published data set in Ref. [S48].

Our final result for the quantities in Eq. (S13), and their ratios is shown in Fig. 1 of the main paper.

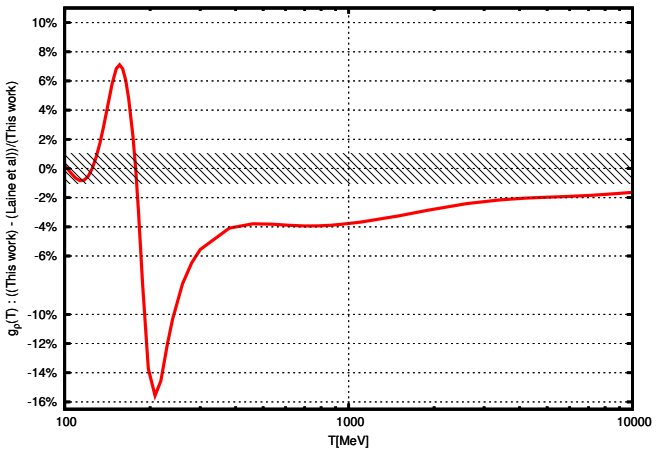


Figure S12: Our results and a previous estimate [S48]. We computed the relative difference of the two results. Before and after the QCD epoch there is agreement. The discrepancy in the range between 0.1 and 10 GeV is explained by the lack of lattice QCD input in [S48]. The dashed region around $\pm 1\%$ indicates the systematics of our parameterization.

Let us comment on the equation of state and cosmology. In the previous subsection we calculated the EoS up to 10 GeV, and here we provided a parameterization up to 300 GeV. In Section S11 we use this EoS to describe both pre- and post-inflation axion generation. Post-inflation axion cosmology is not very sensitive to the 16% difference between our result and that of e.g. Ref. [S48], see Fig. S12. This is easy to see since the oscillation temperature is well above the temperatures at which the difference is most pronounced. The 16% difference is around 200 MeV, thus somewhat above the QCD transition. Nevertheless, there are cosmological questions, e.g. even the pre-inflation axion generation, which are sensitive to this temperature region. The presented unified parameterization over five orders of magnitude could be useful for many cosmological studies.

S7 Fixed sector integral with staggered fermions

The integral method presented in Methods can be trivially generalized in the presence of fermions. The definition of b_Q is still given by Equation (2). In the fixed-N_t approach changing the temperature is achieved via changing the lattice spacing, which requires a simultaneous change of β and the mass parameters m_f, to keep the system on the LCP, see Equation (S1). Then for b_Q we obtain:

$$-b_Q \equiv \frac{d\log Z_Q/Z_0}{d\log T} = \frac{d\beta}{d\log a}\langle S_g\rangle_{Q-0} + \sum_f \frac{d\log m_f}{d\log a} m_f \langle\overline{\psi}\psi_f\rangle_{Q-0}. \tag{S17}$$

Besides the gauge action S_g, we also have to measure the chiral condensate $\overline{\psi}\psi_f$ of each flavor. The full expression is a renormalized quantity, and so is the chiral condensate difference multiplied by the quark mass. To obtain the susceptibility we have to apply the same integral as in the pure gauge case, see Equation (3).

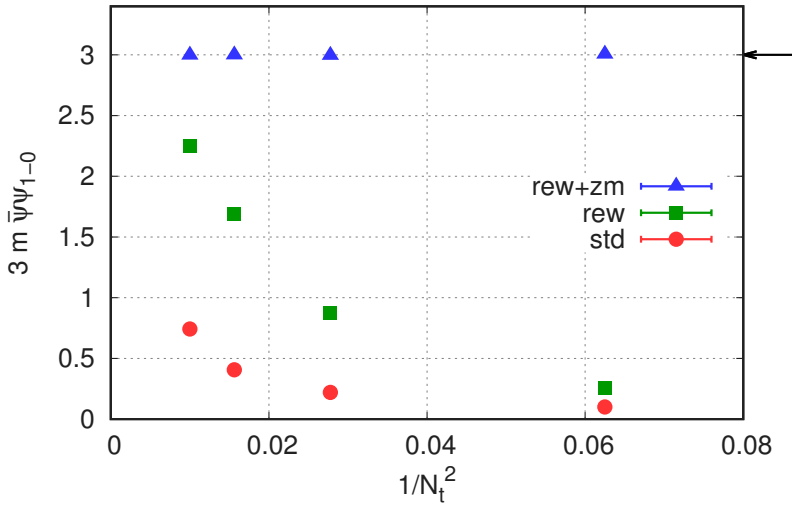
Now let us look at the Stefan-Boltzmann limit of the decay exponent of the susceptibility. We can neglect the contribution of the $Q > 1$ sectors, so the decay exponent is $b = b_1 - 4$. The gauge action difference is the same as in the pure gauge case $4\pi^2/3$. The gauge parameter depends on the lattice spacing as $\beta = (33 - 2n_f)\log a/(4\pi^2)$ and the mass parameter as $\log m_f = \log a$ up to logarithmic corrections in a. The difference in the chiral condensate between sectors $Q = 1$ and 0 comes entirely from the presence of the zero mode, which gives a $\langle\overline{\psi}\psi_f\rangle_{1-0} = 1/m_f$. Altogether we have $b = (33 - 2n_f)/3 - 4 + n_f$ in the high temperature limit.

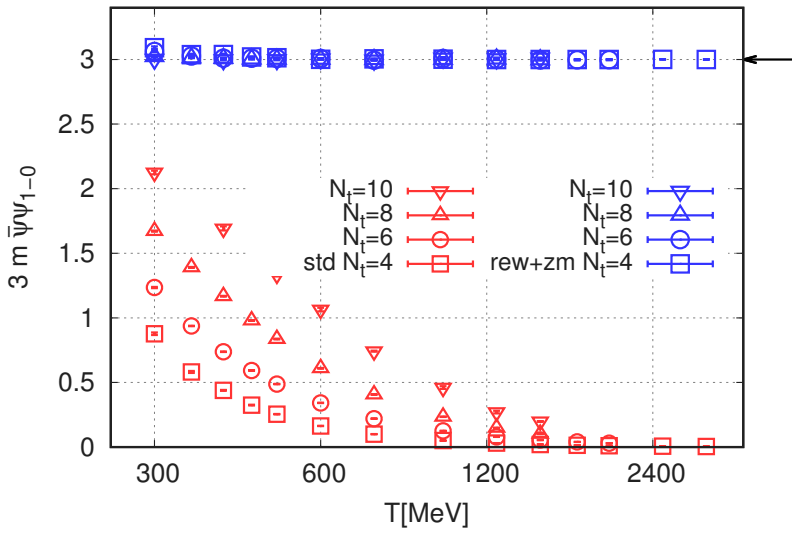
The statements of the Methods Section about the computer time scaling with the volume and the possibility of using $Q > 1$ sectors also apply in the case of dynamical fermions. We used $Q = 0$ and 1 in this work, this is sufficient, since the topological susceptibility with dynamical fermions is tiny.

In numerical simulations the statistical noise on the gauge action difference is much larger than on the chiral condensate difference. This is very similar to, what was already observed in the context of the equation of state [S27]. This inspired us to use the following strategy: evaluate the b_Q and the susceptibility at a quark mass, where the simulation is less expensive than at the physical point. We choose a point, the so-called three-flavor symmetric point, where the two light-quark masses were set to the physical strange mass: $m_{ud} \equiv m_{s,\mathrm{phys}}$. At this point we determined χ using the eigenvalue reweighting method (see later). Then we carried out an integration in the light-quark mass from $m_{s,\mathrm{phys}}$ down to the physical light-quark mass $m_{ud,\mathrm{phys}} = m_{s,\mathrm{phys}}/R$. In this way we could avoid calculating the expensive gauge action difference at the physical point.

We observed that there are huge lattice artefacts on the chiral condensate contribution, if a non-chiral fermion discretization is used. In the absence of exact zero modes the chiral condensate difference needs very fine lattices to reach the continuum limit. The lattice spacing dependence of the three flavor chiral condensate difference is shown on the data labeled by "std" in the upper panel Figure S13. We used $3 + 1$ flavor staggered quarks in the simulation at a temperature of $T = 750$ MeV. There is an order of magnitude increase in the condensate by going from the coarsest to the finest lattice spacing. In the middle panel of Figure S13 the temperature dependence of the three flavor condensate is shown for different lattice spacings. As the temperature increases the condensate, which is calculated in the standard way, approaches zero contrary to the expectation in the high temperature limit. This also happens for the charm quark condensate, although at a somewhat smaller pace, see lower panel. The vanishing of the chiral condensate difference decreases the decay exponent of the susceptibility by n_f in the standard approach, which largely explains the unexpectedly small exponent obtained in the recent lattice calculation [S51].

We present two independent approaches to solve this problem. One is to use a chiral fermion discretization to evaluate the chiral condensate difference, this is explained in a separate section, Section S8. The other is the modification of the path integral by the reweighting technique, that we already introduced in the Methods Section. Let us present the details of the reweighting here. The introduction of the reweighting factors $w[U]$ in Equation (1) means, that our simulation corresponds to a modified





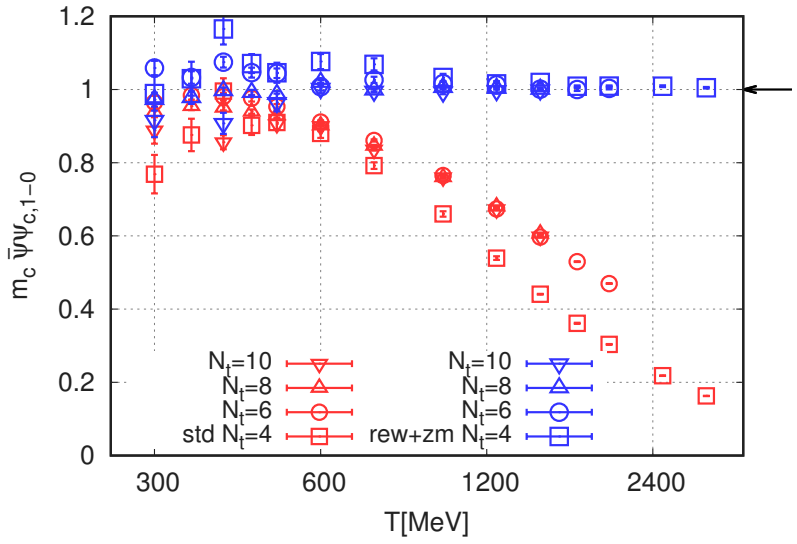


Figure S13: Chiral condensate difference between sectors $Q = 1$ and $Q = 0$ multiplied by the quark mass. The data labeled by "std" denotes the value calculated from the standard chiral condensate. The "rew" data is obtained by reweighting with the weights in Equation (1). The "rew+zm" data includes the contribution of the zero modes, i.e. the mass dependence of the weight factors, see Equation (S20). The arrows indicate the Stefan-Boltzmann limit. The upper plot shows the difference as a function of the lattice spacing squared at $T = 750$ MeV temperature. The middle plot shows the difference as a function of the temperature, whereas the lower plot is the same for the charm quark. The plots show $n_f = 3 + 1$ flavor staggered simulations on $N_t = 4, 6, 8$ and 10 lattices.

partition function:

$$Z^{\mathrm{rw}} = \int [dU] \ \exp(-\beta S_g) \cdot \prod_f \det(D_{st} + m_f)^{1/4} \cdot w[U]. \tag{S18}$$

Note, that reweighting affects only the sectors with non-trivial topologies. This results in a modification of the expression for b_Q in Equation (S17) as:

$$-b_Q^{\mathrm{rw}} \equiv \frac{d \log Z_Q^{\mathrm{rw}}/Z_0^{\mathrm{rw}}}{d \log T} = \frac{d\beta}{d \log a} \langle S_g \rangle_{Q-0}^{\mathrm{rw}} + \sum_f \frac{d \log m_f}{d \log a} m_f \langle \overline{\psi}\psi_f \rangle_{Q-0}^{\mathrm{rw+zm}}, \tag{S19}$$

where $\langle . \rangle_Q^{\mathrm{rw}}$ denotes the fixed Q expectation value including the weights $w[U]$. The fermionic contribution has two parts, one coming from the reweighted chiral condensate and another from the mass dependence of the weight factors, which we call zero mode contribution:

$$\langle \overline{\psi}\psi_f \rangle_{Q-0}^{\mathrm{rw+zm}} = \langle \overline{\psi}\psi_f \rangle_{Q-0}^{\mathrm{rw}} + \frac{|Q|}{m_f} - \left\langle \frac{1}{2m_f} \sum_{n=1}^{2|Q|} \frac{m_f^2}{\lambda_n^2[U] + m_f^2} \right\rangle_Q^{\mathrm{rw}}. \tag{S20}$$

In the end we have to measure three observables, the gauge action, the chiral condensate and the zero mode contribution, on the reweighted configurations.

The effects of reweighting and including zero mode contribution can also be seen in the panels of Figure S13. In the upper panel we see, that reweighting already improves the estimate of the fermionic contribution significantly and including the additional zero mode contribution reduces the lattice artefacts even further. In the middle panel of the figure we see that the reweighted condensate together with the zero mode contribution approaches the Stefan-Boltzmann limit for high temperatures, as expected. Indeed for the strange quark at our smallest temperature $T = 300$ MeV the Stefan-Boltzmann value is already reached, whereas the charm contribution is about 20% lower. The temperature dependence of the latter is plotted in the lower panel of Figure S13.

The lattice artefacts and the finite size effects on the decay exponent turned out to be remarkably small. These will be shown in the analysis section, Section S9, where we discuss our procedure for the continuum and infinite volume extrapolations.

S8 Fixed sector integral with overlap fermions

As demonstrated in the Methods Section and in Section S7 staggered fermions produced huge lattice artefacts in the topological susceptibility and in the high temperature chiral condensate. An obvious remedy is to carry out simulations in a chirally symmetric discretization.

Determining the topological susceptibility with overlap fermions has a long history. The direct measurement is numerically difficult, since one has to deal with the non-analyticity of the overlap operator on the topological sector boundary. Though solutions exist [S23, S24, S52], they are somewhat cumbersome. Alternatively one can perform simulations in fixed topology [S22] and determine the topological susceptibility from the long distance behaviour of the topological charge correlator [S53]. This is a viable approach at zero temperature, but for high temperatures, where the susceptibility is small, one would need to measure the correlator with a very high precision.

Our new approach, presented in Section S7, also requires simulations with fixed topology. However we need to determine only the chiral condensate difference, the rest can be taken from direct simulations at parameters, where the direct approach is feasible. As we have seen the use of staggered fermions is complicated and difficult for this purpose. The difference, as we will show in this section, can be nicely measured in the overlap formulation. For algorithmic and other technical details we refer the reader to Section S3. Before showing results for the chiral condensate, we start with a previously unknown subtlety in fixed topology simulations with overlap fermions, which is related to configurations with a pair of an instanton and an anti-instanton.

S8.1 Instanton–anti-instanton (IA) configurations

The overlap topological charge is defined as the difference of the number of left and right handed zero modes of the overlap Dirac operator. A smooth instanton/anti-instanton produces a left/right handed zero mode in the spectrum. In practical simulations, one never encounters a configuration, where simultaneously left and right handed zero modes are present. This of course does not mean, that configurations with an IA pair are not allowed. If we look at smooth configurations, which contain a well-separated instanton and an anti-instanton, the overlap operator has a complex conjugate pair of overlap modes with very small but non-zero eigenvalues[5].

For general configurations the definition of an IA pair or the number of IA pairs is of course not unambiguous. However, for sufficiently high temperatures we observe very small modes, that are well separated from the rest of the non-zero modes, the latter being on the scale of the temperature. We looked at the topological charge distribution of such configurations and indeed observed the concentration of the charge into a positive and a negative lump, see Figure S14. Since such objects produce small complex conjugate pairs in the overlap operator spectrum, the value of the chiral condensate depends strongly on the presence of IA pairs. Therefore it is important to know their weight in the path integral.

For an overlap fermion with a topology fixing term, configurations with a well-separated IA pair pose the following problem. Annihilating such a pair cannot proceed by simply removing the instanton and the anti-instanton one-by-one, since this would change the topological sector. Either they have to be removed simultaneously or they have to be brought to the same position, where they can annihilate. If the volume is large this latter can be difficult to achieve. In unfortunate cases we are stuck with an IA pair, and do not sample the probability distribution correctly. In our concrete numerical simulations we encountered this problem at only one parameter set: at the mass of the strange quark, $T = 300$ MeV and an aspect ratio of $N_s/N_t = 4$. For the history of the lowest eigenvalue in three different Monte-Carlo streams see Figure S15. "stream-1" contains no, "stream-3" one IA pair, in "stream-2" there was an IA annihilation after about 900 trajectories. For smaller masses and larger temperatures the runs always ended up without having IA pairs after a short thermalization time.

[5]One can prepare an artificial configuration with simultaneous left and right handed overlap zero modes, where an instanton is placed in the first half of the volume and the second half is obtained by CP transforming the first. Such configurations are expected to be a zero measure subset in the configuration space.

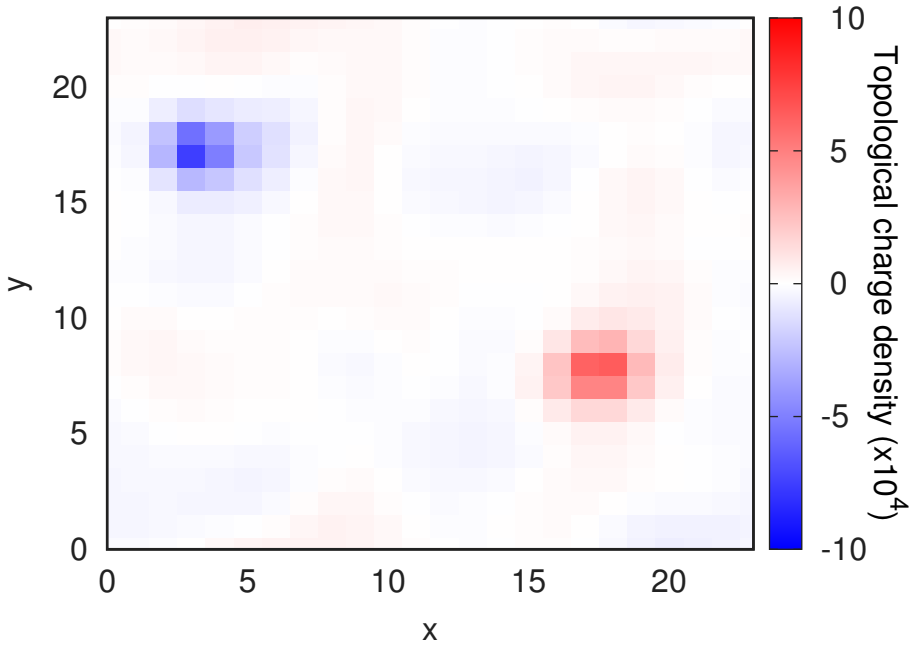


Figure S14: Topological charge density $q(x)$ distribution on an IA configuration. The $q(x)$ is averaged over the z and t coordinates and scaled by 10^4. The plot shows a configuration from an $n_f = 3$ flavor overlap simulation on a 6×24^3 lattice at $T = 300$ MeV temperature.

It is interesting to look at the eigenvalues of the overlap kernel operator on IA configurations. In our case the kernel is a Wilson-Dirac operator with a negative mass: $D_W - m_W$. In the presence of an IA pair D_W will have two real modes lying between $0 \ldots m_W$ and with opposite chiralities. To annihilate the IA pair, the two real modes have to be placed to a different region in the spectrum. For an illustration see the plot in Figure S16. The simplest way, i.e. to move them towards larger real values, is not possible. This is because, the effect of the topology fixing term is to forbid real modes to go through the point m_W. So the only way they can disappear is, to move into the complex plane. However due to the γ_5-hermiticity of the D_W operator, complex eigenvalues have to come in complex conjugate pairs. So in order to leave the real axis, the two real eigenvalues first have to become degenerate, it is only then possible for them to go into the complex plane.

For the single problematic overlap run, mentioned above, we calculate the contribution of the IA configurations as follows. We measure the weight of such configurations with staggered fermions in the continuum limit and with the same physical parameters ($m = m_s$, $T = 300$ MeV and $LT = 4$). Here the configurations were generated without fixing the topology using the direct approach, which is still efficient at these temperatures. To measure the number of topological objects we use a smeared overlap operator with kernel mass parameter $m_W = -1.3$. To define the number of IA pairs, IA, we counted the number of complex conjugate pairs, for which the eigenvalue satisfied $|\lambda|^2 < 10^{-4}$. We checked that the results are not sensitive to small variations of the upper bound exponent. The results as a function of the lattice spacing can be seen in Figure S17. The probability of $IA > 0$ configurations is non-negligible, somewhat less than 10% at this particular parameter set. An important observation is, that the probability is independent of the topological charge of the configuration, i.e.

$$\left.\frac{Z(IA=1)}{Z(IA=0)}\right|_{Q=1} = \left.\frac{Z(IA=1)}{Z(IA=0)}\right|_{Q=0}. \tag{S21}$$

We also find in general, that the weights of sectors depend only on $Q + 2 \cdot IA$, which is the total number of topological objects. This finding is in accordance with these objects being independent of each other, which is a key assumption of the DIGA. The consequence is, that the IA contribution drops out of the

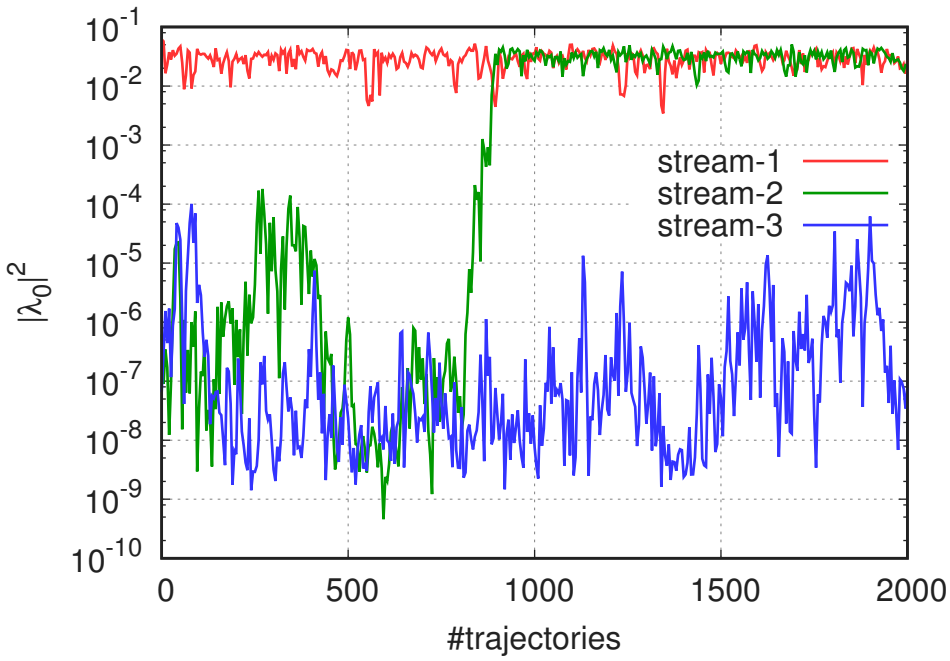


Figure S15: Lowest eigenmode squared $|\lambda|^2$ of the overlap Dirac operator in three different Monte-Carlo streams in the trivial topological sector. The plot shows $n_f = 3$ flavor overlap simulations on a 6×24^3 lattice at $T = 300$ MeV temperature.

ratio Z_1/Z_0 and

$$\frac{Z_1}{Z_0} = \frac{Z_1}{Z_0}\bigg|_{IA=0} \tag{S22}$$

holds to a very good accuracy in the continuum limit. Based on these findings, as we increase the temperature or decrease the mass, we expect that in a given topological sector the IA pairs will have a decreasing contribution, like the contribution of non-trivial topological sectors decreases compared to the trivial sector. Since all our overlap runs use either a larger temperature or a smaller mass, than this particular simulation, we can safely neglect the contribution of IA configurations in all of our overlap simulations and use Equation (S22) to calculate sector weights.

When using staggered fermions, similarly to the lack of exact zero modes, also the eigenvalues corresponding to IA pairs lie higher in the spectrum than in the overlap case. We explicitely studied the consequences of these IA pairs on the reweighting procedure on the $n_f = 3 + 1$ flavor $T = 300$ MeV staggered ensemble. The zero modes corresponding the topological sectors and the approximate zero modes corresponding to the IA pairs were determined with the overlap operator, as we described before. Carrying out the eigenvalue reweighting also for the IA pairs the value of χ turned out to be consistent with the one, where only the zero modes were reweighted. This was true not only in the continuum limit but also for each lattice spacing. Since both for higher temperatures and smaller quark masses IA pairs become less frequent we conclude that such a reweighting is not necessary.

Let us note, that the DIGA also suggests, that adding IA pairs into the reweighting has no effect on the result. For this let us calculate the ratio Z_1/Z_0 and let us work with only two sectors for brevity. If we only reweight with zero modes, than we get wZ_1/Z_0, where w is the typical weight factor of a configuration (w approaches 1 in the continuum limit). If we add the IA pairs into the reweighting, then a configuration with $N = Q + 2 \cdot IA$ topological objects gets a weight factor of w^N on average, thus the ratio is $[wZ_1(IA = 0) + w^3 Z_1(IA = 1)]/[Z_0(IA = 0) + w^2 Z_0(IA = 1)]$. Using Equation (S21) this equals to wZ_1/Z_0, thus the two reweightings give the same result.

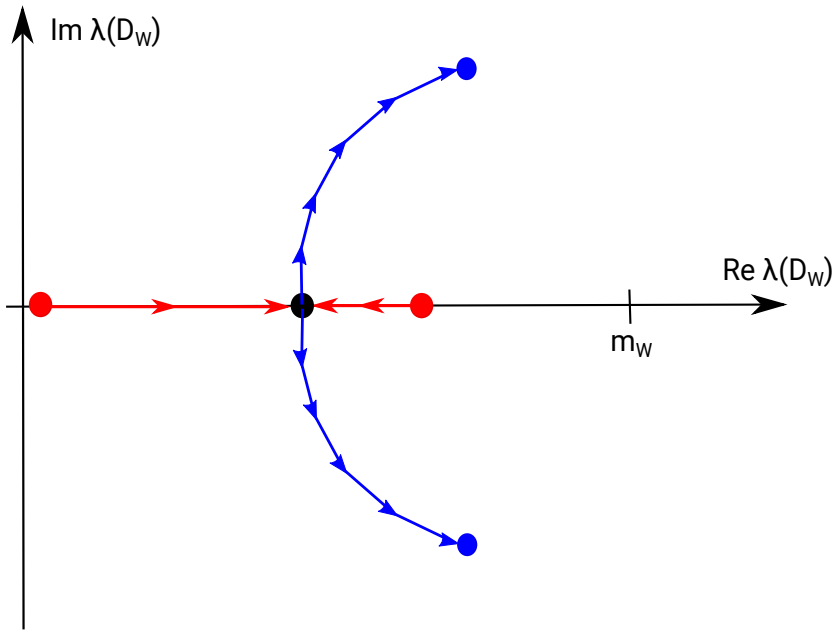


Figure S16: Illustration of an IA annihilation in the spectrum of the Wilson-Dirac operator, which is used in the kernel of the overlap Dirac operator.

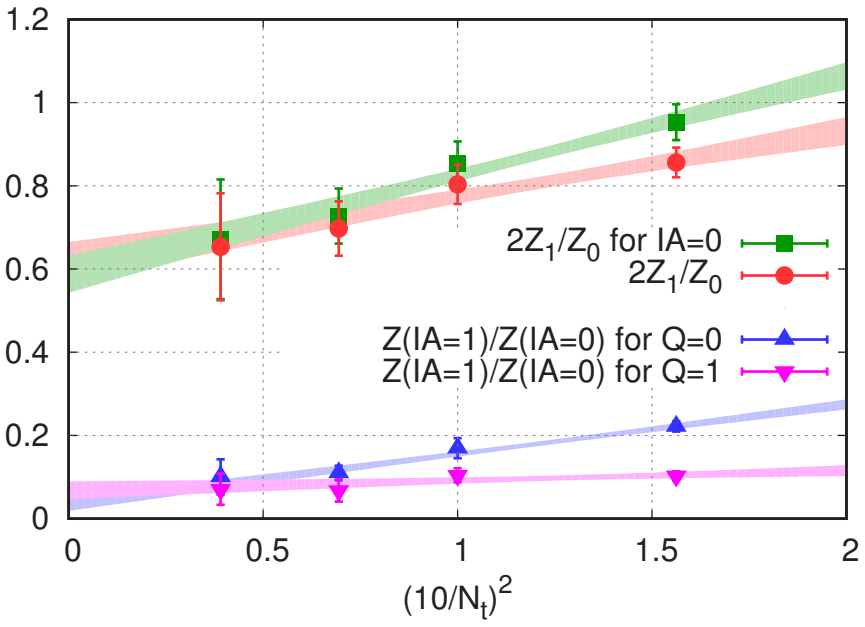


Figure S17: Continuum extrapolation of partition function ratios. The lower index in Z_Q stands for the topological charge, IA for the number of IA pairs. The plot shows $n_f = 3+1$ flavor staggered simulations on lattices with $N_s/N_t = 4$ at $T = 300$ MeV temperature.

S8.2 The chiral condensate difference

The chiral condensate can be decomposed into the eigenmodes of the Dirac-operator. The contribution of each topological mode is $1/m_f$. In the infinite temperature limit the rest of the spectrum drops out from the chiral condensate difference and therefore

$$m_f \langle \overline{\psi}\psi_f \rangle_{1-0} = 1. \tag{S23}$$

For finite temperatures we expect corrections to this result. We also expect, that with decreasing quark mass the topological contribution will dominate and the corrections to Equation (S23) will get smaller. To investigate the size of these corrections we have carried out overlap simulations for a wide range of parameters. These are given in Table S4.

β	$N_s \times N_t$	m_{ud}	m_s	# ktraj	$\frac{1}{2} m_{ud} \langle \overline{\psi}\psi_{ud} \rangle_{1-0}$
m_{ud}-scan at $T = 300$ MeV					
3.99	12×6	0.0690	0.0690	10	1.00(1)
3.99	12×6	0.0460	0.0690	5	0.99(1)
3.99	12×6	0.0172	0.0690	8	1.00(1)
3.99	12×6	0.0069	0.0690	10	1.00(1)
m_{ud}-scan at $T = 450$ MeV					
4.19	12×6	0.0389	0.0389	10	1.00(1)
4.19	12×6	0.0291	0.0389	6	1.00(1)
4.19	12×6	0.0259	0.0389	3	1.00(1)
4.19	12×6	0.0195	0.0389	3	1.00(1)
4.19	12×6	0.0097	0.0389	3	1.00(1)
4.19	12×6	0.0049	0.0389	3	1.00(1)
m_{ud}-scan at $T = 650$ MeV					
4.38	12×6	0.0242	0.0242	5	1.00(1)
4.38	12×6	0.0181	0.0242	5	1.00(1)
4.38	12×6	0.0161	0.0242	3	1.00(1)
4.38	12×6	0.0121	0.0242	2	1.00(1)
4.38	12×6	0.0060	0.0242	2	1.00(1)
N_t-scan					
3.99	12×6	0.0690	0.0690	12	1.00(1)
4.13	16×8	0.0458	0.0458	29	1.02(2)
4.24	20×10	0.0342	0.0342	80	1.00(1)
N_s-scan					
3.99	12×6	0.0690	0.0690	12	1.00(1)
3.99	16×6	0.0690	0.0690	20	1.00(1)
3.99	20×6	0.0690	0.0690	32	1.02(1)
3.99	24×6	0.0690	0.0690	48	1.00(1)

Table S4: Gauge coupling parameter, lattice size, quark masses and number of thousand trajectories for the 2+1 flavor overlap simulations at finite temperature. Last column contains the chiral condensate difference.

First we calculated the corrections to Equation (S23) for light-quark masses in the range $m_{ud}/m_s = 0.1 \ldots 1$, we refer to it as "$m_{ud}$-scan" in the table. We fixed the strange mass and used lattices of fixed size 6×12 and used three different temperatures. Then we looked at the lattice spacing dependence of the results, at $T = 300$ MeV at the three flavor point, we call it "N_t-scan". Finally at the same temperature and quark mass we investigated the finite size effects, these runs are called "N_s-scan". In all cases we found, that Equation (S23) holds with an accuracy of about one percent. The results are given in the last column of Table S4.

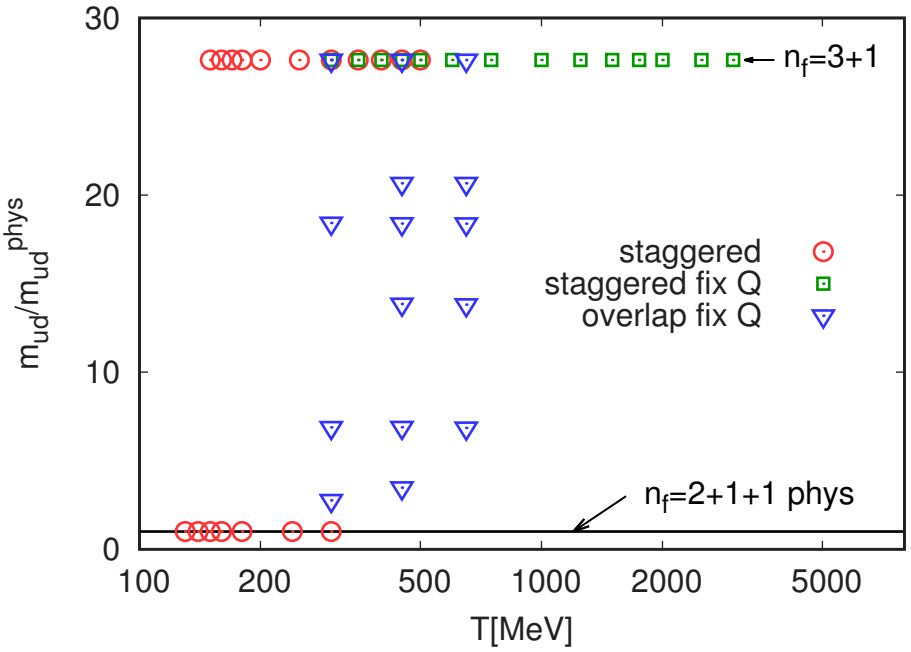


Figure S18: Summary plot of simulation points to determine χ. There are staggered simulations in the $n_f = 3 + 1$ and $n_f = 2 + 1 + 1$ flavor theories, which are then connected by overlap simulations.

S9 Analysis for the topological susceptibility

We combine all approaches developed in the other sections to obtain our final result for the continuum extrapolated topological susceptibility at the physical point. Figure 2 in the main text shows this result including statistical and systematic error estimates.

In Figure S18 we show the simulation points that were used in the analysis. The plot shows the temperature – light-quark mass (m_{ud}) plane. Four simulation sets can be distinguished:

1. $n_f = 3 + 1$ flavor staggered simulations in the region $T = 150 \ldots 500$ MeV
2. $n_f = 3 + 1$ flavor staggered simulations at fixed topology in the region $T = 300 \ldots 3000$ MeV
3. $n_f = 2 + 1$ flavor overlap simulations at fixed topology for temperatures $T = 300, 450$ and 650 MeV building a bridge between the three flavor and the physical theories
4. $n_f = 2 + 1 + 1$ flavor staggered simulations at the physical point for temperatures $T = 130 \ldots 300$ MeV

The main feature of our strategy is that the the starting points for the fixed sector integral method are taken in the three flavor symmetric theory (1. and 2.) instead of gathering statistics at the physical point. Starting from a temperature of $T \sim 300$ MeV the difference between the two can be taken into account by rescaling the topological sector weights, which is justified by the results of the overlap simulations (3.) connecting the two theories. Note that before the connection is done we carried out the continuum extrapolation, thus physical results are connected with physical results. The main observation is, as expected, that for large temperatures χ scales with the mass. Since it works on the one percent level already at 300 MeV, there is no need to go beyond 650 MeV for these bridging simulations. In the transition region the scaling behavior of the susceptibility with the quark mass is expected to change, so for temperatures $T \lesssim 300$ MeV we still resort to direct simulations at the physical point (4.).

T[MeV]	$N_s \times N_t$	ktraj	$\langle Q^2 \rangle$	$\langle \mathrm{round}(Q)^2 \rangle$	$\langle Q^2 \rangle_{rw}$	$\langle \mathrm{round}(Q)^2 \rangle_{\mathrm{rw}}$
180	24×6	90	$4.49(0.11) \times 10^{+1}$	$4.50(0.11) \times 10^{+1}$	–	–
200		170	$2.89(0.04) \times 10^{+1}$	$2.90(0.04) \times 10^{+1}$	$2.44(0.03) \times 10^{+1}$	$2.44(0.03) \times 10^{+1}$
250		260	$4.01(0.04) \times 10^{+0}$	$4.12(0.05) \times 10^{+0}$	$2.07(0.02) \times 10^{+0}$	$2.10(0.02) \times 10^{+0}$
300		370	$7.25(0.10) \times 10^{-1}$	$9.22(0.10) \times 10^{-1}$	$2.74(0.05) \times 10^{-1}$	$3.29(0.05) \times 10^{-1}$
350		380	$1.76(0.03) \times 10^{-1}$	$2.64(0.05) \times 10^{-1}$	$4.32(0.13) \times 10^{-2}$	$5.66(0.14) \times 10^{-2}$
400		490	$5.34(0.13) \times 10^{-2}$	$8.58(0.21) \times 10^{-2}$	$1.04(0.04) \times 10^{-2}$	$1.36(0.05) \times 10^{-2}$
450		490	$1.89(0.08) \times 10^{-2}$	$3.12(0.12) \times 10^{-2}$	$2.50(0.13) \times 10^{-3}$	$3.26(0.20) \times 10^{-3}$
500		490	$6.90(0.45) \times 10^{-3}$	$1.13(0.08) \times 10^{-2}$	$8.40(0.96) \times 10^{-4}$	$9.49(1.26) \times 10^{-4}$
150	32×8	70	$1.18(0.04) \times 10^{+2}$	$1.18(0.04) \times 10^{+2}$	–	–
160		60	$9.71(0.18) \times 10^{+1}$	$9.72(0.18) \times 10^{+1}$	–	–
170		60	$7.68(0.15) \times 10^{+1}$	$7.69(0.15) \times 10^{+1}$	–	–
180		70	$6.07(0.15) \times 10^{+1}$	$6.07(0.16) \times 10^{+1}$	–	–
200		100	$4.02(0.10) \times 10^{+1}$	$4.03(0.10) \times 10^{+1}$	$3.40(0.07) \times 10^{+1}$	$3.41(0.07) \times 10^{+1}$
250		70	$4.54(0.29) \times 10^{+0}$	$5.03(0.29) \times 10^{+0}$	$2.97(0.22) \times 10^{+0}$	$3.33(0.23) \times 10^{+0}$
300		100	$7.54(0.16) \times 10^{-1}$	$9.56(0.18) \times 10^{-1}$	$3.75(0.10) \times 10^{-1}$	$4.60(0.12) \times 10^{-1}$
350		140	$1.62(0.09) \times 10^{-1}$	$2.15(0.11) \times 10^{-1}$	$6.39(0.55) \times 10^{-2}$	$7.96(0.65) \times 10^{-2}$
400		140	$4.27(0.38) \times 10^{-2}$	$5.94(0.48) \times 10^{-2}$	$1.47(0.23) \times 10^{-2}$	$1.85(0.28) \times 10^{-2}$
450		210	$1.45(0.16) \times 10^{-2}$	$1.97(0.20) \times 10^{-2}$	$4.50(1.04) \times 10^{-3}$	$5.44(1.25) \times 10^{-3}$
500		440	$5.81(1.09) \times 10^{-3}$	$8.11(1.38) \times 10^{-3}$	$1.09(0.46) \times 10^{-3}$	$1.36(0.56) \times 10^{-3}$
150	40×10	50	$1.34(0.03) \times 10^{+2}$	$1.35(0.03) \times 10^{+2}$	–	–
160		50	$1.13(0.03) \times 10^{+2}$	$1.13(0.03) \times 10^{+2}$	–	–
170		50	$8.16(0.24) \times 10^{+1}$	$8.16(0.24) \times 10^{+1}$	–	–
180		50	$6.53(0.22) \times 10^{+1}$	$6.56(0.22) \times 10^{+1}$	–	–
200		150	$3.86(0.22) \times 10^{+1}$	$3.86(0.22) \times 10^{+1}$	$3.32(0.15) \times 10^{+1}$	$3.32(0.15) \times 10^{+1}$
250		90	$5.11(0.29) \times 10^{+0}$	$5.72(0.31) \times 10^{+0}$	$3.90(0.21) \times 10^{+0}$	$4.35(0.23) \times 10^{+0}$
300		160	$6.91(0.41) \times 10^{-1}$	$8.12(0.46) \times 10^{-1}$	$4.68(0.33) \times 10^{-1}$	$5.37(0.38) \times 10^{-1}$
350		400	$1.27(0.07) \times 10^{-1}$	$1.55(0.09) \times 10^{-1}$	$7.49(0.52) \times 10^{-2}$	$8.65(0.59) \times 10^{-2}$
400		410	$3.36(0.36) \times 10^{-2}$	$4.11(0.42) \times 10^{-2}$	$1.79(0.21) \times 10^{-2}$	$2.07(0.24) \times 10^{-2}$
450		860	$1.11(0.17) \times 10^{-2}$	$1.34(0.20) \times 10^{-2}$	$6.49(1.10) \times 10^{-3}$	$7.39(1.25) \times 10^{-3}$
500		1300	$3.20(0.46) \times 10^{-3}$	$4.17(0.56) \times 10^{-3}$	$1.22(0.26) \times 10^{-3}$	$1.43(0.29) \times 10^{-3}$

Table S5: $n_f = 3 + 1$ flavor staggered simulation points for direct measurements of χ on $N_t = 6, 8$ and 10 lattices. The columns are: temperature, lattice geometry, number of trajectories in thousands, and the variance of the topological charge $\langle Q^2 \rangle$ calculated in four different ways (standard, rounded to nearest integer, reweighted, reweighted and rounded).

T[MeV]	$N_s \times N_t$	ktraj	$\langle Q^2 \rangle$	$\langle \mathrm{round}(Q)^2 \rangle$	$\langle Q^2 \rangle_{rw}$	$\langle \mathrm{round}(Q)^2 \rangle_{\mathrm{rw}}$
150	48×12	30	$1.43(0.06) \times 10^{+2}$	$1.43(0.06) \times 10^{+2}$	–	–
160		30	$1.03(0.04) \times 10^{+2}$	$1.03(0.04) \times 10^{+2}$	–	–
170		30	$7.44(0.28) \times 10^{+1}$	$7.44(0.28) \times 10^{+1}$	–	–
180		30	$5.88(0.25) \times 10^{+1}$	$5.89(0.25) \times 10^{+1}$	–	–
200		120	$3.46(0.31) \times 10^{+1}$	$3.53(0.30) \times 10^{+1}$	$3.10(0.25) \times 10^{+1}$	$3.17(0.24) \times 10^{+1}$
250		210	$4.50(0.12) \times 10^{+0}$	$4.94(0.13) \times 10^{+0}$	$3.89(0.10) \times 10^{+0}$	$4.27(0.11) \times 10^{+0}$
300		270	$6.70(0.56) \times 10^{-1}$	$7.49(0.62) \times 10^{-1}$	$5.39(0.48) \times 10^{-1}$	$5.92(0.53) \times 10^{-1}$
350		410	$1.09(0.07) \times 10^{-1}$	$1.24(0.07) \times 10^{-1}$	$8.10(0.60) \times 10^{-2}$	$8.93(0.66) \times 10^{-2}$
400		470	$2.50(0.27) \times 10^{-2}$	$2.88(0.29) \times 10^{-2}$	$1.68(0.25) \times 10^{-2}$	$1.85(0.27) \times 10^{-2}$
450		620	$7.37(1.10) \times 10^{-3}$	$8.50(1.20) \times 10^{-3}$	$4.69(0.96) \times 10^{-3}$	$5.21(1.04) \times 10^{-3}$
500		350	$4.12(2.25) \times 10^{-3}$	$4.53(2.47) \times 10^{-3}$	$2.76(1.79) \times 10^{-3}$	$2.99(1.94) \times 10^{-3}$
150	64×16	20	$1.32(0.06) \times 10^{+2}$	$1.32(0.06) \times 10^{+2}$	–	–
160		20	$9.29(0.65) \times 10^{+1}$	$9.36(0.64) \times 10^{+1}$	–	–
170		20	$6.53(0.51) \times 10^{+1}$	$6.67(0.51) \times 10^{+1}$	–	–
180		20	$6.02(0.62) \times 10^{+1}$	$6.17(0.62) \times 10^{+1}$	–	–
200		20	$3.72(0.61) \times 10^{+1}$	$3.86(0.62) \times 10^{+1}$	$3.52(0.54) \times 10^{+1}$	$3.67(0.55) \times 10^{+1}$
250		40	$4.94(0.82) \times 10^{+0}$	$5.19(0.86) \times 10^{+0}$	$4.74(0.78) \times 10^{+0}$	$4.97(0.82) \times 10^{+0}$
300		180	$6.71(0.79) \times 10^{-1}$	$7.08(0.82) \times 10^{-1}$	$6.30(0.76) \times 10^{-1}$	$6.62(0.79) \times 10^{-1}$
150	80×20	20	$1.37(0.18) \times 10^{+2}$	$1.39(0.18) \times 10^{+2}$	–	–
160		20	$9.26(0.87) \times 10^{+1}$	$9.50(0.89) \times 10^{+1}$	–	–
170		20	$7.52(1.14) \times 10^{+1}$	$7.78(1.16) \times 10^{+1}$	–	–
180		20	$6.28(0.93) \times 10^{+1}$	$6.51(0.96) \times 10^{+1}$	–	–
200		20	$3.90(0.88) \times 10^{+1}$	$4.02(0.91) \times 10^{+1}$	$3.78(0.83) \times 10^{+1}$	$3.90(0.85) \times 10^{+1}$

Table S6: $n_f = 3 + 1$ flavor staggered simulation points for direct measurements of χ on $N_t = 12, 16$ and 20 lattices. The columns are: temperature, lattice geometry, number of trajectories in thousands, and the variance of the topological charge $\langle Q^2 \rangle$ calculated in four different ways (standard, rounded to nearest integer, reweighted, reweighted and rounded).

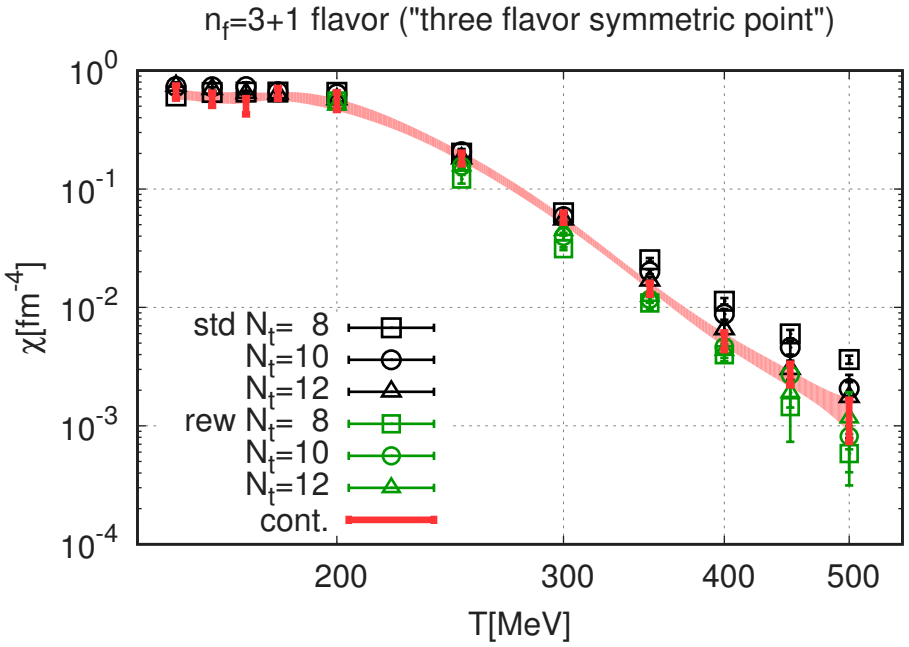


Figure S19: Topological susceptibility as the function of the temperature. Results for different lattice spacings are shown using two different methods (standard and reweighted). The continuum extrapolation, which is described in the text, is also shown. The band is obtained from quadratic spline interpolations from the continuum extrapolated reweighted data. The plot shows $n_f = 3{+}1$ flavor staggered simulations.

S9.1 Results for $n_f = 3 + 1$ flavors - direct approach

In the region between $T = 150$ and 500 MeV we performed simulations at six different lattice spacings, $N_t = 6, 8, 10, 12, 16$ and 20, for direct measurements of χ. The simulation points together with the statistics are given in Tables S5 and S6. On these configurations we measured the Wilson flow based topological charge at a flow-time of $(8T^2)^{-1}$, and for temperatures above $T \geq 200$ MeV we also calculated the low-lying eigenvalues of the Dirac operator. We employed four different ways to compute the susceptibility, the corresponding values are all given in the tables above. The first is the standard way to calculate the topological charge. In the second approach we rounded the charge to the nearest integer value. This should give the same result in the continuum limit as the first approach, the difference between them is a $\mathcal{O}(a^2)$ lattice artefact, see the discussion in the Methods Section. In the third approach we applied reweighting, as described in the Methods Section, and in the fourth we combined reweighting with rounding.

In Figure S19 we plot the susceptibility for different lattice spacings. At a temperature of about 200 MeV, there is an abrupt change in the behaviour of the susceptibility. Below this temperature the result is flat, above it there is a rapid decrease. The turning point can be associated with the transition from the hadron to the quark gluon plasma phase. The transition temperature, $T_c^{(3)} \approx 200$ MeV, turns out to be somewhat higher than at the physical point. The two approaches, standard and reweighted, have very similar lattice artefacts. This is due to the fact, that the temperature is not too high and that the pion mass is rather heavy, $m_\pi^{(3)} \approx 710$ MeV, see Section S2.

The lattice volume was set to $LT = 4$ in all these runs. In our previous quenched study [S10], we already saw that the finite volume effects on the total susceptibility were small, if the box size satisfied $LT_c \gtrsim 2$. We expect the finite size effects in our $n_f = 3 + 1$ flavor simulations to be similar to the quenched case, since the pion mass is rather heavy here. Thus the results should be safe from finite size effects up to a temperature of $T \sim 400$ MeV. Figure S19 shows results for even higher temperatures, but these were not used in subsequent analyses.

If we look at the charge distributions at different temperatures, then one can notice, that above $T \gtrsim 400$ MeV the distributions contain a very small number of configurations from sectors $Q > 1$. This

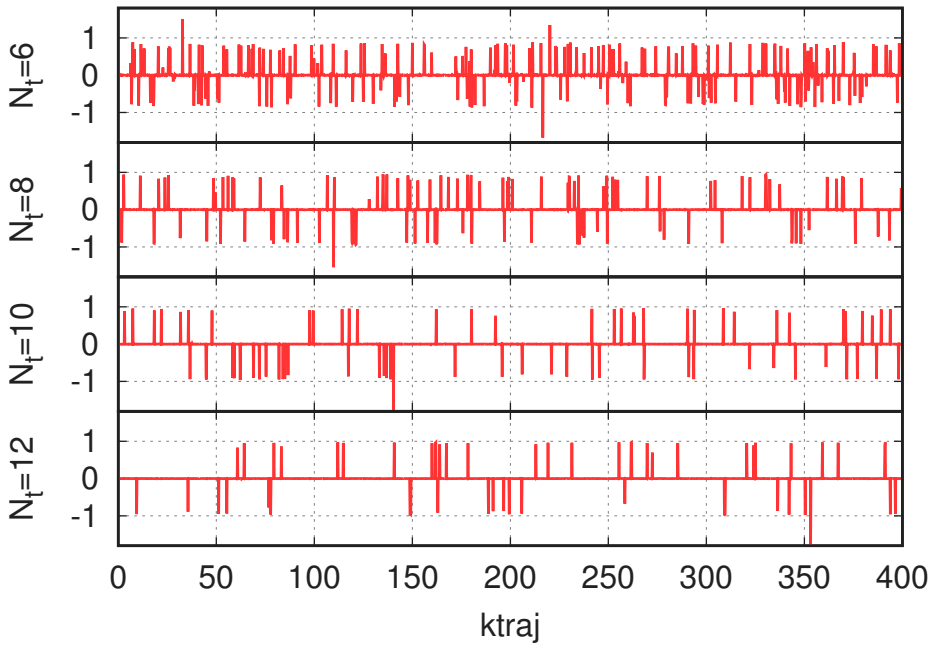


Figure S20: Simulation time history of the topological charge for four different lattice spacings, $N_t = 6, 8, 10$ and 12. Note, that the cutoff effect, i.e. the $N_t = 6$ simulation fluctuates more than the $N_t = 12$, is largely reduced by reweighting. The plot shows $n_f = 3+1$ flavor staggered simulations at a temperature of $T = 400$ MeV.

is also indicated by the smallness of the second moment $\langle Q^2 \rangle \ll 1$. For an illustration see Figure S20, where the simulation time history of the topological charge is plotted for four different lattice spacings at $T = 400$ MeV. Starting from this temperature the total susceptibility is statistically consistent with the one, where we restrict the calculation to the $Q = 0$ and 1 sectors only.

We carried out a continuum extrapolation for each temperature separately, for which several different extrapolation procedures were used. For the lattice spacing dependence we used either purely $1/N_t^2$ dependence or we also added higher order terms. The different fits were obtained from changing the definition of the charge (standard or rounded), from excluding the first 0, 1 or 2 coarsest lattice spacings from the fit and from including/excluding higher order terms in $1/N_t$. The systematic uncertainties were then estimated using our histogram method [S38, S54]. In Figure S19 we show the continuum extrapolated results using the reweighted data. The statistical and systematic errors were added in quadrature. The continuum limit presented in Figure S19 was calculated using the reweighting method. As a consistency check we also calculated the continuum limit using the standard method, and found a consistent continuum limit, however the standard method has larger errors for the largest 2-3 temperatures due to the fact, that the continuum extrapolation is getting steeper in the standard case. In Figure S19 we also give an interpolation between the continuum extrapolated data points, which was obtained from quadratic splines with randomized nodepoints, see e.g. [S20].

S9.2 Results for $n_f = 3 + 1$ flavors - fixed sector integral approach

To reach the temperature region that is needed for axion phenomenology we employ the fixed sector integral method introduced in the Methods Section. This way no extrapolation in the temperature is needed. For this we generated configurations in topological sectors $Q = 0$ and 1, at four different lattice spacings $N_t = 4, 6, 8$ and 10 and aspect ratios $N_s/N_t = 2 \ldots 6$. The simulated temperature values, lattice geometries, the number of trajectories and acceptance rates of the Metropolis step in the $Q = 1$ sector are given in Table S7. The acceptance ratios are somewhat worse, than in the quenched case (see Methods). Note, that the the simulation algorithm for dynamical fermions was different from the one, that we used in quenched. They decorrelate the topological charge differently, which results in different acceptance

T[MeV]	$N_s \times N_t$	Mtraj	acc.	$b_1 - 4$
300	8×4	0.4	24%	6.8(0.5)
350		0.4	22%	5.9(0.5)
400		0.4	21%	6.9(0.7)
450		0.4	20%	5.4(0.7)
500		0.4	20%	6.3(0.6)
600		0.4	20%	5.9(0.4)
750		0.4	20%	7.2(0.5)
1000		0.4	20%	7.0(0.6)
1250		0.4	20%	8.3(0.6)
1500		1.7	20%	6.8(0.2)
1750		1.3	20%	7.1(0.4)
2000		1.7	20%	7.2(0.2)
2500		1.7	20%	7.3(0.2)
3000		1.2	20%	7.4(0.2)
750	12×4	1.9	19%	7.1(0.4)
750	16×4	3.2	19%	6.9(0.4)
750	24×4	8.9	18%	6.8(0.5)
1500	12×4	1.9	19%	6.6(0.5)
1500	16×4	3.2	19%	7.5(0.5)
1500	24×4	8.9	18%	5.9(0.5)

T[MeV]	$N_s \times N_t$	Mtraj	acc.	$b_1 - 4$
300	12×6	2.5	51%	7.4(0.3)
350		2.5	49%	6.9(0.2)
400		2.5	47%	7.5(0.3)
450		2.5	45%	6.9(0.3)
500		2.5	44%	7.5(0.5)
600		2.5	42%	6.3(0.4)
750		2.5	40%	7.4(0.4)
1000		2.5	38%	6.8(0.5)
1250		2.5	37%	6.9(0.7)
1500		2.5	36%	7.6(0.8)
1750		2.5	36%	7.6(0.7)
2000		2.5	35%	7.4(0.8)
300	16×8	5.2	67%	7.2(0.4)
350		5.6	65%	7.7(0.3)
400		5.7	62%	6.8(0.3)
450		5.7	59%	7.5(0.4)
500		5.7	57%	6.7(0.4)
600		5.7	53%	7.4(0.4)
750		5.7	49%	7.0(0.5)
1000		5.7	45%	7.2(0.4)
1250		5.7	43%	6.9(0.5)
1500		5.5	41%	7.2(1.0)
300	20×10	1.4	79%	8.1(1.1)
400		1.3	73%	7.1(1.0)
500		1.4	68%	7.5(1.4)
600		1.4	63%	6.2(0.9)
750		5.8	57%	7.0(0.5)
1000		1.4	51%	10.7(1.4)
1250		1.4	47%	5.8(1.8)
1500		5.7	45%	8.2(1.1)

Table S7: $n_f = 3{+}1$ flavor staggered simulation points with fixed topology, $Q = 0, 1$, for measuring χ with the fixed sector integral method. The columns are: temperature, lattice geometry, number of trajectories in millions, acceptance rate in the $Q = 1$ sector. The last column contains the decay exponent, with the statistical error in the parentheses.

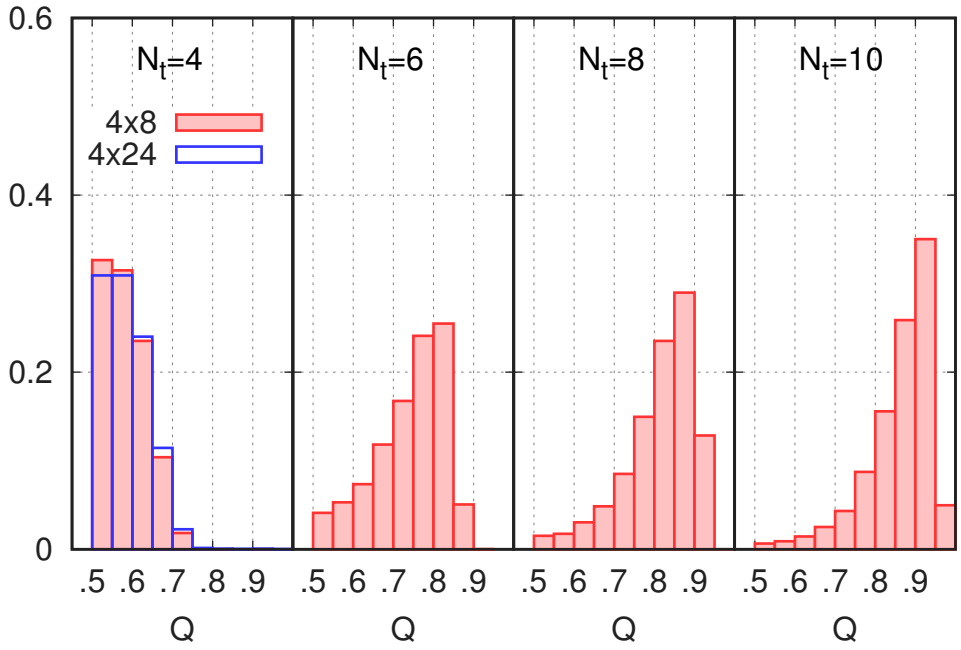


Figure S21: Lattice spacing and volume dependence of the charge distribution in fixed sector simulations using the constraint $Q > 0.5$. The plot shows $n_f = 3 + 1$ flavor staggered simulations at $T = 750$ MeV.

rates.

In Figure S21 we plot the charge distribution in the $Q = 1$ sector simulations, which were obtained by running simulations using the constraint $Q > 0.5$. The plot is very similar to the quenched case: we see practically no volume dependence and the charge peaks shift towards 1 in the continuum limit.

To obtain the susceptibility we first determined the continuum extrapolation of the coefficient b_1. We used reweighting to improve the continuum extrapolation, as described in Section S7. From the direct simulations we found, that starting from $T = 400$ MeV the $Q > 1$ sectors give negligible contribution to the susceptibility. Therefore the decay exponent of the susceptibility can be calculated as

$$b = -d \log \chi / d \log T = b_1 - 4 \tag{S24}$$

. In Table S7 we also give the measured values of the exponent for all of the ensembles.

In order to obtain the continuum extrapolation we performed several fits to the data points. In the temperature range we consider, the exponent is expected to be a regular function of the temperature. We took this into account by using a polynomial up to third order in the temperature. Beside the T-dependence we also added terms to describe possible lattice spacing and finite size effects. These fits can be described by the following form:

$$b(\tau = 1000/T[\text{MeV}], N_t, N_s) = b_{T0} + b_{T1} \cdot \tau + b_{T2} \cdot \tau^2 + b_{T3} \cdot \tau^3 + b_a / N_t^2 + b_L \cdot (N_t/N_s)^3 \tag{S25}$$

We took 4 different fit ansätze by keeping only $1 \ldots 4$ out of the four b_{Tn} coefficients describing the temperature dependence, thus performing the fit with less parameters. We achieved reasonable fit qualities: for example a fit with all parameters included gave a $\chi^2/\text{dof} = 50/44$. The different fits were then combined with the Akaike Information Criterion (AIC), which penalizes not only bad fits but also overfitting. For more details, see e.g. [S54]. In order to test the $\mathcal{O}(N_t^{-2})$ scaling hypothesis included in the Equation (S25), we also made four additional fits with excluding the $N_t = 4$ and another four with excluding $N_t = 4, 6$ data points, these were also weighted with their corresponding AIC weights. Finally we combined the 12 fits uniformly. From these we computed an average and a width, the first is our final result for the exponent, the latter is used as an estimate of the systematic error. The statistical error was estimated by repeating the above procedure on the jackknife samples.

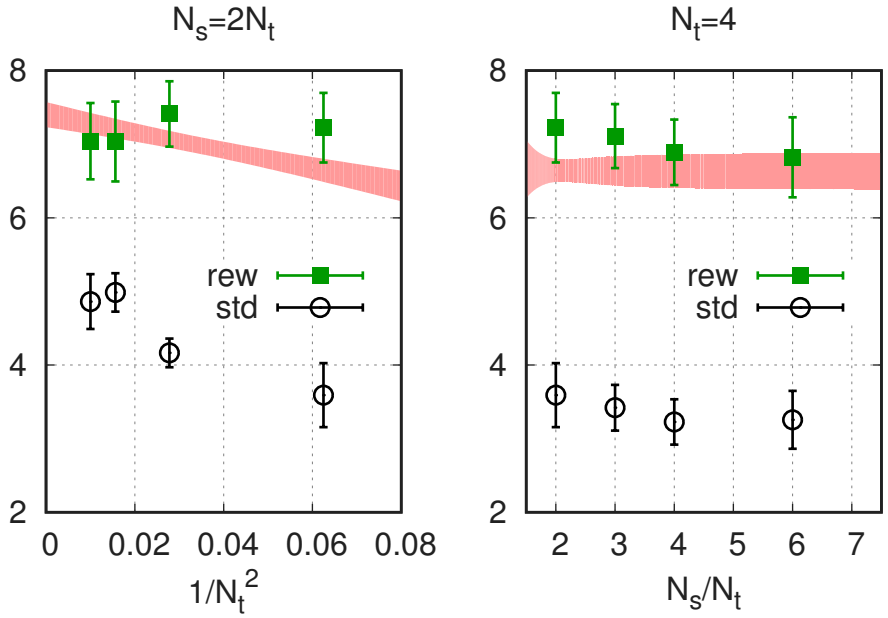


Figure S22: Lattice spacing (left) and finite size (right) dependence of the decay exponent b. The lattice spacing dependence is shown for $N_s = 2N_t$, while the finite volume dependence for $N_t = 4$. The labels "std" and "rew" indicate, whether the data set was obtained with the standard or with the reweighting method including the zero mode contribution. The red band is the result of the fit in Equation (S25) to the full dataset. The plot shows $n_f = 3 + 1$ flavor staggered simulations at $T = 750$ MeV temperature.

In Figure S22 we show the lattice artefacts and finite size corrections on the exponent at a temperature of $T = 750$ MeV, where we made dedicated simulations to study these effects. The left panel shows the lattice spacing dependence with and without reweighting. The lattice artefacts are larger without reweighting and the continuum extrapolations of the two data sets differ significantly. This difference is due to the problematic behaviour of the fermionic contribution without reweighting, as we explained before in Section S7. The result of the fit procedure, Equation (S25), is given with a red band, where the width of the band was obtained from adding the statistical and systematic error in quadrature. We observe a slight increase of b towards the continuum limit.

The right panel of Figure S22 shows the finite size dependence of b. We did simulations with an aspect ratio up to and including 6. The box size corresponding to the largest aspect ratio is 1.6 fm, which is large enough to accommodate non-perturbative length scales. The coefficient of the finite volume term, b_L, is consistent with zero in the fits even including aspect ratios as small as $LT = 2$. Again we give the result from the fit procedure as a red band.

The temperature dependence of the exponent is shown in Figure S23. The lattice data are plotted with green points, using different symbols for the different lattice spacings. The continuum and infinite volume extrapolation is shown by the red band.

We also calculated the prediction of the DIGA for $n_f = 3+1$ flavors. For this we took the strange mass from [S4] and $m_c/m_s = C$ from Equation (S1), since this ratio was used in the simulations. We took $\Lambda^{(4)}_{\overline{\text{MS}}}$ from [S41] to convert the results to physical units. The renormalization scale dependence was estimated by using three different scales: 1, $1/\sqrt{2}$ and $\sqrt{2}$ times πT. The continuum extrapolated lattice result gives somewhat smaller exponents than the DIGA, but the difference is less than 2σ. Similar behaviour is seen in the quenched case for smaller temperatures.

To obtain χ an integration in the temperature has to be performed, see Equation (3). We start with a continuum extrapolated value from the low temperature region and then use the continuum extrapolation of the data for the exponent b to perform the integration. The systematic error is derived from the systematic error at the starting point of the integration, from three choices of the starting point (300,

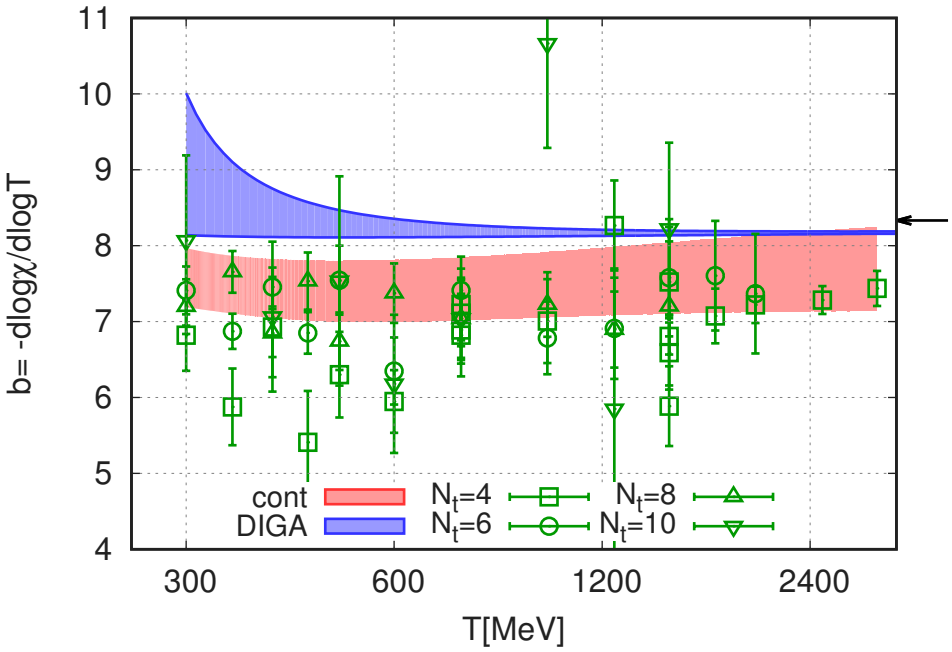


Figure S23: Temperature dependence of the exponent $b = -d\log\chi/d\log T$. The plot shows $n_f = 3+1$ flavor staggered simulations on $N_t = 4, 6, 8$ and 10 lattices. The red band is the continuum extrapolation. The prediction of the $n_f = 3+1$ flavor DIGA is given by the blue band. The arrow shows the Stefan-Boltzmann limit.

350 and 400 MeV) and from the systematic error of the exponent fit. The result is then plotted together with the direct simulations at lower temperature in Figure S24 and also with the prediction of the DIGA.

Let us compare the lattice result of χ with the DIGA prediction. The overall picture turned out to be quite similar for both findings (of course only a posteriori): a strong drop as a function of the temperature. There are, however, sizable differences. Similarly to the quenched case [S10] the prefactor of the DIGA result is off by an order of magnitude also in the present case (full dynamical result with physical quark masses). This is not the only difference. A more subtle effect is related to the temperature dependence. The lattice exponent b is all the way from 300 MeV to 3000 MeV somewhat below the DIGA prediction, see Figure S23. This means that the topological susceptibility decreases faster in the dilute instanton gas approximation than in the lattice picture. When the temperature reaches 3000 MeV (which is a relevant temperature for the post-inflationary axion scenario) the central value of χ is more than twenty times smaller in the DIGA framework than on the lattice. This difference in χ would mean a factor of five difference in the axion's mass m_A. For the final prediction of the axion's mass this is partly compensated by a change in the oscillation temperature.

S9.3 Topological susceptibility for $n_f = 2+1+1$ flavors

For the zero temperature susceptibility we applied a method based on leading order chiral perturbation theory to remove the lattice artefacts. This is described in Section S2.

At finite temperature we can start from the continuum extrapolated topological sector weights in the $n_f = 3+1$ theory. The result for $n_f = 2+1+1$ flavors is obtained by performing an integration in the light-quark mass. For example, the relative weight of sectors $Q = 0$ and 1 can be calculated as:

$$\left.\frac{Z_1}{Z_0}\right|_{2+1+1} = \exp\left(\int_{m_{ud}^{phys}}^{m_s^{phys}} d\log m_{ud}\; m_{ud}\langle\overline{\psi}\psi_{ud}\rangle_{1-0}\right)\cdot\left.\frac{Z_1}{Z_0}\right|_{3+1} \tag{S26}$$

The overlap simulations in Section S8 provided ample evidence, that above $T = 300$ MeV, to a very good precision the integrand is given by the number of light flavors $m_{ud}\langle\overline{\psi}\psi_{ud}\rangle_{1-0} = 2$. Thus the sector

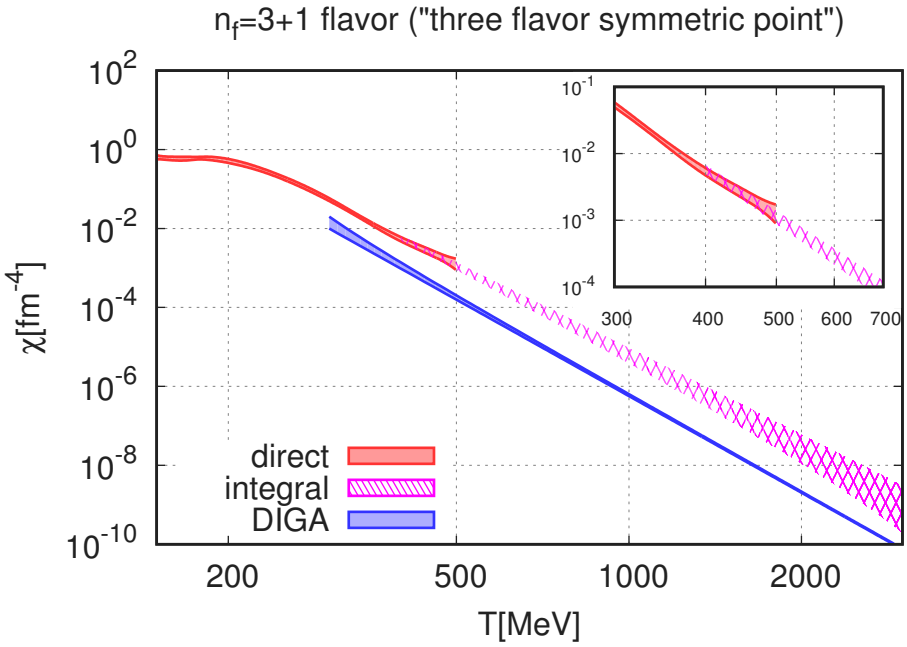


Figure S24: Continuum extrapolated topological susceptibility as a function of the temperature in the three flavor symmetric theory. The result was obtained from $n_f = 3+1$ flavor staggered simulations after performing the continuum limit. The result of the direct method is shown below $T \leq 500$ MeV and for the fixed sector integral above $T \geq 400$ MeV. Also shown is the prediction of the DIGA.

weights at the physical point are given by the following scaling:

$$\left.\frac{Z_1}{Z_0}\right|_{2+1+1} = R^2 \cdot \left.\frac{Z_1}{Z_0}\right|_{3+1}, \tag{S27}$$

with R given in Equation (S1). The error on the result of the overlap measurement propagates to an error at the physical point, which is below the statistical uncertainty.

To obtain the susceptibility we have to take into account the contribution of $Q \geq 2$ sectors, too. Starting from $T = 400$ MeV even in the three flavor symmetric theory only $Q = 0$ and 1 sectors contribute. At $T = 300$ MeV the ratio $Z_{Q\geq 2}/Z_0$ is around ~ 0.1, however the mass integration suppresses it by a factor R^4 and thus becomes negligible at the physical point compared to Z_1/Z_0. It is therefore safe to work only with $Q = 0$ and 1 configurations in the $T \geq 300$ MeV region at the physical point.

Equation (S26) can of course be used at any temperature, but the simple scaling with R^2 will not work e.g. at low temperatures or in the transition region. For temperatures below $T = 300$ MeV we decided to fall back on the direct measurement at the $n_f = 2+1+1$ physical point, which turned out to be feasible even on the already existing statistics from [S3]. The lattice geometries, number of trajectories and the measured charge variances are given in Table S8.

In Figure S25 we show the results for the topological susceptibility. The zero temperature value is given by a red horizontal band at the left hand side of the plot. Black and green points show the direct measurement of χ using the standard and reweighting approaches. As was described in the Methods Section the lattice artefacts are much larger with the standard approach and reliable continuum extrapolations could only be performed after applying the reweighting procedure. The continuum extrapolation was performed separately for each temperature between 130 and 300 MeV in a similar way as in the case of the three flavor symmetric theory. These are also shown in the plot. Finally for $T \geq 300$ MeV we also give the result of the mass integration, where for the starting point the $n_f = 3+1$ result was used. It is consistent with the direct measurement after applying reweighting.

T[MeV]	lattice	ktraj	$\langle Q^2 \rangle$	$\langle \mathrm{round}(Q)^2 \rangle$	$\langle Q^2 \rangle_{rw}$	$\langle \mathrm{round}(Q)^2 \rangle_{\mathrm{rw}}$
130	$32^3 \times 8$	54	$9.74(0.18) \times 10^{+1}$	$9.75(0.18) \times 10^{+1}$	$1.37(0.02) \times 10^{+1}$	$1.37(0.02) \times 10^{+1}$
140		51	$7.77(0.18) \times 10^{+1}$	$7.78(0.18) \times 10^{+1}$	$8.45(0.17) \times 10^{+0}$	$8.36(0.17) \times 10^{+0}$
150		52	$5.82(0.09) \times 10^{+1}$	$5.84(0.09) \times 10^{+1}$	$4.53(0.09) \times 10^{+0}$	$4.45(0.08) \times 10^{+0}$
160		50	$4.38(0.09) \times 10^{+1}$	$4.40(0.09) \times 10^{+1}$	$2.10(0.05) \times 10^{+0}$	$2.00(0.05) \times 10^{+0}$
180		70	$2.03(0.05) \times 10^{+1}$	$2.04(0.05) \times 10^{+1}$	$3.66(0.14) \times 10^{-1}$	$2.95(0.11) \times 10^{-1}$
240		50	$2.24(0.04) \times 10^{+0}$	$2.59(0.05) \times 10^{+0}$	$2.86(0.13) \times 10^{-2}$	$9.74(0.55) \times 10^{-3}$
300		52	$3.25(0.10) \times 10^{-1}$	$4.35(0.13) \times 10^{-1}$	$2.94(0.19) \times 10^{-3}$	$6.03(0.36) \times 10^{-4}$
130	$40^3 \times 10$	51	$1.16(0.03) \times 10^{+2}$	$1.16(0.03) \times 10^{+2}$	$1.39(0.02) \times 10^{+1}$	$1.38(0.02) \times 10^{+1}$
140		54	$8.11(0.18) \times 10^{+1}$	$8.13(0.18) \times 10^{+1}$	$8.26(0.15) \times 10^{+0}$	$8.22(0.15) \times 10^{+0}$
150		50	$5.67(0.09) \times 10^{+1}$	$5.67(0.10) \times 10^{+1}$	$4.45(0.08) \times 10^{+0}$	$4.41(0.08) \times 10^{+0}$
160		51	$3.83(0.08) \times 10^{+1}$	$3.84(0.08) \times 10^{+1}$	$2.08(0.05) \times 10^{+0}$	$2.07(0.05) \times 10^{+0}$
180		50	$1.59(0.03) \times 10^{+1}$	$1.60(0.03) \times 10^{+1}$	$3.91(0.09) \times 10^{-1}$	$3.54(0.09) \times 10^{-1}$
240		53	$1.35(0.04) \times 10^{+0}$	$1.62(0.04) \times 10^{+0}$	$1.57(0.07) \times 10^{-2}$	$1.13(0.06) \times 10^{-2}$
300		50	$1.83(0.07) \times 10^{-1}$	$2.31(0.09) \times 10^{-1}$	$9.51(1.17) \times 10^{-4}$	$6.80(0.83) \times 10^{-4}$
130	$48^3 \times 12$	54	$1.02(0.01) \times 10^{+2}$	$1.02(0.01) \times 10^{+2}$	$1.34(0.02) \times 10^{+1}$	$1.34(0.02) \times 10^{+1}$
140		50	$6.63(0.12) \times 10^{+1}$	$6.63(0.12) \times 10^{+1}$	$7.98(0.17) \times 10^{+0}$	$8.04(0.16) \times 10^{+0}$
150		50	$4.57(0.12) \times 10^{+1}$	$4.58(0.11) \times 10^{+1}$	$4.32(0.08) \times 10^{+0}$	$4.40(0.08) \times 10^{+0}$
160		50	$2.79(0.04) \times 10^{+1}$	$2.81(0.04) \times 10^{+1}$	$1.96(0.04) \times 10^{+0}$	$2.02(0.04) \times 10^{+0}$
180		55	$1.07(0.02) \times 10^{+1}$	$1.13(0.02) \times 10^{+1}$	$3.68(0.10) \times 10^{-1}$	$3.71(0.13) \times 10^{-1}$
240		103	$7.86(0.23) \times 10^{-1}$	$9.19(0.26) \times 10^{-1}$	$1.06(0.06) \times 10^{-2}$	$9.56(0.61) \times 10^{-3}$
300		83	$1.04(0.05) \times 10^{-1}$	$1.26(0.06) \times 10^{-1}$	$8.15(1.60) \times 10^{-4}$	$8.28(1.56) \times 10^{-4}$
130	$64^3 \times 16$	51	$5.97(0.12) \times 10^{+1}$	$6.03(0.12) \times 10^{+1}$	$1.19(0.01) \times 10^{+1}$	$1.23(0.01) \times 10^{+1}$
140		50	$3.62(0.08) \times 10^{+1}$	$3.71(0.08) \times 10^{+1}$	$6.80(0.13) \times 10^{+0}$	$7.15(0.13) \times 10^{+0}$
150		51	$2.13(0.04) \times 10^{+1}$	$2.23(0.05) \times 10^{+1}$	$3.68(0.05) \times 10^{+0}$	$3.90(0.06) \times 10^{+0}$
160		50	$1.26(0.03) \times 10^{+1}$	$1.34(0.03) \times 10^{+1}$	$1.78(0.05) \times 10^{+0}$	$1.89(0.05) \times 10^{+0}$
180		44	$4.63(0.09) \times 10^{+0}$	$5.02(0.09) \times 10^{+0}$	$4.03(0.10) \times 10^{-1}$	$4.26(0.12) \times 10^{-1}$
240		52	$2.78(0.14) \times 10^{-1}$	$3.11(0.15) \times 10^{-1}$	$8.36(1.35) \times 10^{-3}$	$8.53(1.43) \times 10^{-3}$
300	$64^2 \times 128 \times 16$	50	$6.72(0.93) \times 10^{-2}$	$7.67(0.98) \times 10^{-2}$	$1.45(0.82) \times 10^{-3}$	$1.53(0.86) \times 10^{-3}$
130	$80^3 \times 20$	10	$3.43(0.17) \times 10^{+1}$	$3.58(0.18) \times 10^{+1}$	$1.01(0.04) \times 10^{+1}$	$1.06(0.05) \times 10^{+1}$
140		51	$2.04(0.05) \times 10^{+1}$	$2.14(0.06) \times 10^{+1}$	$6.16(0.08) \times 10^{+0}$	$6.46(0.08) \times 10^{+0}$
150		10	$1.24(0.07) \times 10^{+1}$	$1.30(0.07) \times 10^{+1}$	$3.55(0.16) \times 10^{+0}$	$3.70(0.17) \times 10^{+0}$
160		10	$5.50(0.44) \times 10^{+0}$	$5.77(0.46) \times 10^{+0}$	$1.71(0.10) \times 10^{+0}$	$1.79(0.10) \times 10^{+0}$
180		10	$1.94(0.15) \times 10^{+0}$	$2.06(0.15) \times 10^{+0}$	$4.20(0.61) \times 10^{-1}$	$4.39(0.63) \times 10^{-1}$

Table S8: $n_f = 2 + 1 + 1$ flavor staggered simulation points for direct measurements of χ on $N_t = 8, 10, 12, 16$ and 20 lattices. The columns are: temperature, lattice geometry, number of trajectories in thousands, and the variance of the topological charge $\langle Q^2 \rangle$ calculated in four different ways (standard, rounded to nearest integer, reweighted, reweighted and rounded).

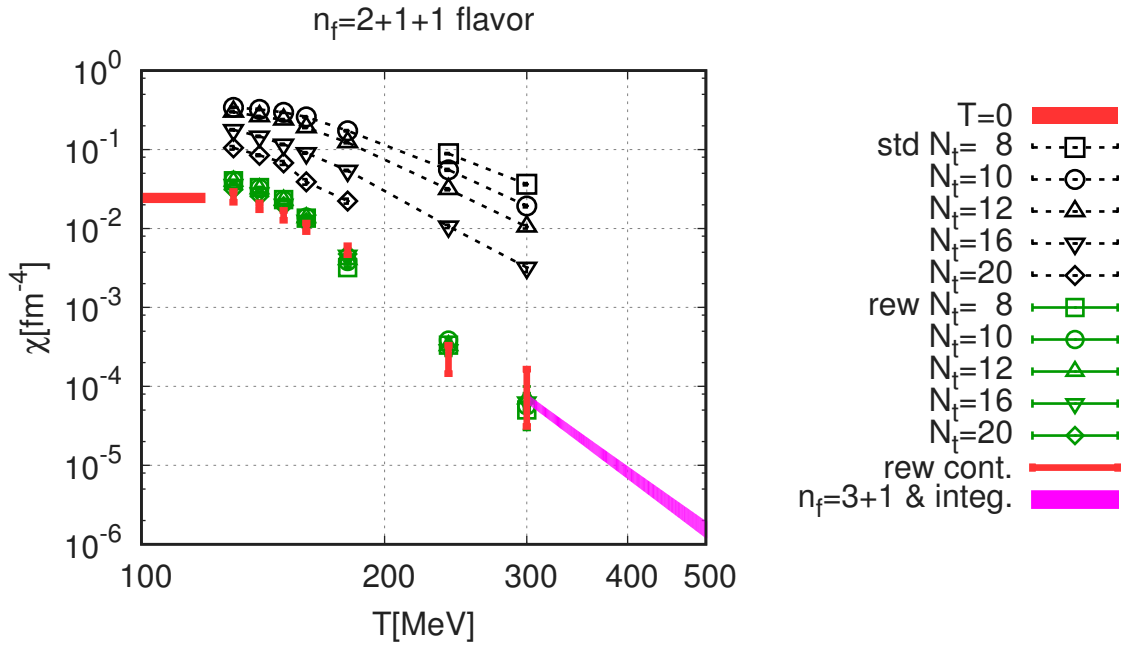


Figure S25: Topological susceptibility as the function of the temperature. Result for different lattice spacings are shown using two different methods, rounded and reweighted-rounded, with black and green points respectively. The red band on the left hand side is from our zero temperature simulations. The red points are the continuum extrapolations from the reweighted data. The integral method is shown with a pink band on the right. The plot shows $n_f = 2 + 1 + 1$ flavor staggered simulations.

S9.4 Topological susceptibility - full result

There are two effects, that are missing in the $n_f = 2 + 1 + 1$ flavor theory, and have to be taken into account to obtain the full result for χ: the presence of the bottom quark and the mass difference between the up and the down quarks.

As we have seen in Section S7, the charm contribution to the decay exponent b has almost reached the high temperature limit at $T = 300$ MeV. We also found that the charm starts to contribute to the equation of state at $T \sim 250$ MeV (see Section S6). We therefore expect that the bottom contribution starts to be appreciable at temperatures above $\sim m_b/m_c \times 250$ MeV. To take into account the bottom contribution, we added $1/3$ to the $n_f = 2 + 1 + 1$ flavor exponent for temperatures higher than some threshold temperature. The value $1/3$ is the contribution of an extra flavor to the high temperature limit. Then we integrated the so obtained b and finally rescaled the results according to Equation (S27). We have chosen three different threshold temperatures: $T = 1.0, 1.5$ and 2.0 GeV. The resulting variation in χ was added to the systematic error.

Although isospin violating effects are typically on the level of 1%, the topological susceptibility is a notable exception. This is because, the susceptibility is proportional to the product of the quark masses. Therefore the topological susceptibility is a factor of

$$\frac{4m_u m_d}{(m_u + m_d)^2} \approx 0.88 \tag{S28}$$

smaller than in the isospin symmetric, $n_f = 2+1+1$ flavor case. The quark mass values were taken from [S55]. To take isospin violation into account we scaled the isospin symmetric results by this factor for all temperatures.

To provide a continuous result for all temperatures we applied the following strategy. Since for high temperatures we expect $\chi \propto T^{-b}$ or $\log \chi = -b \log T$, it is convenient to parameterize $\log \chi$ as the function of $\log T$. For temperatures below 300 MeV we used a spline interpolation with randomized

T[MeV]	$\log_{10}(\chi[\mathrm{fm}^{-4}])$	T[MeV]	$\log_{10}(\chi[\mathrm{fm}^{-4}])$
100	$-1.66(5)$	600	$-6.47(10)$
120	$-1.64(6)$	720	$-7.06(12)$
140	$-1.75(5)$	860	$-7.63(15)$
170	$-2.15(5)$	1000	$-8.12(17)$
200	$-2.67(8)$	1200	$-8.72(19)$
240	$-3.42(8)$	1500	$-9.46(23)$
290	$-4.12(7)$	1800	$-10.07(26)$
350	$-4.74(6)$	2100	$-10.60(29)$
420	$-5.33(7)$	2500	$-11.20(32)$
500	$-5.89(8)$	3000	$-11.82(36)$

Table S9: Final result for the topological susceptibility of QCD taking into account the effect of the up, down, strange, charm and bottom quarks.

nodepoints [S20]. The fits with different nodepoints are weighted according to their AIC weights. We fitted jointly the following two continuum extrapolated datasets: the $n_f = 2 + 1 + 1$ reweighted sets for temperatures $T \leq 180$ MeV and the $n_f = 3 + 1$ reweighted sets rescaled to the physical point for $T \geq 300$ MeV. The systematic errors of the data points, coming from different continuum extrapolations procedures, are used to estimate the systematic error of the spline interpolation. We use the so obtained curve below 300 MeV. Above 300 MeV we have a continuous result integrating $d \log \chi / d \log T$ using the continuum and infinite volume parameterization in Equation (S25). The complete procedure was repeated on jackknife samples to obtain the statistical error. The total error is the statistical and systematic error combined in quadrature.

Our final result for $\chi(T)$ is shown in Figure 2 of the main text. We also tabulate the base-10 logarithm $\log_{10} \chi(T)$ for a couple of temperature values in Table S9.

S10 Comparison with other recent works

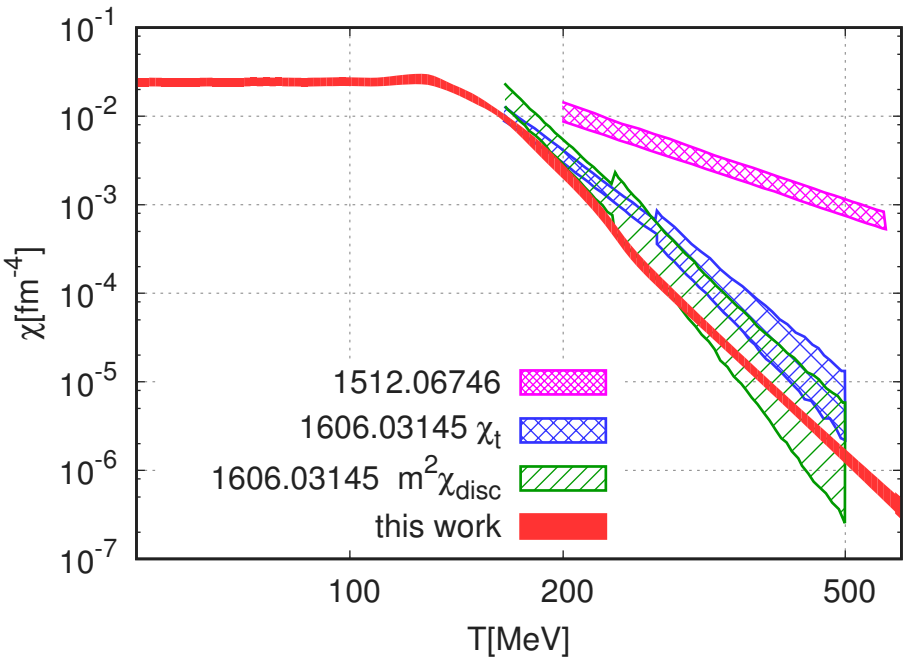


Figure S26: Comparison of our result for the $n_f = 2+1+1$ susceptibility with two recent works. "1512.06746" corresponds to Reference [S51]. "1606.03145" corresponds to Reference [S56], where two different definitions, a gauge (χ_t) and a fermionic one ($m^2\chi_{disc}$), are used.

In Figure S26 we compare our results for χ with other findings in the literature. Since our highest temperature was 3000 MeV, whereas in the literature it is 500 MeV, we focus on the temperature range between 200 and 500 MeV. Reference [S51] used an approach in which staggered fermions were taken and the ratio of the $T>0$ and $T=0$ susceptibilities were determined. This brute force approach turned out to be quite difficult. As we have presented before, our result for χ is many orders of magnitude smaller than that of [S51] in the cosmologically relevant temperature region. One understands why the staggered artefacts mimic such a behaviour and we showed how to remove these artefacts by the eigenvalue reweighting technique. Similarly large values for χ were obtained in a study using twisted mass fermions [S57]. Since in Reference [S57] no continuum extrapolation was carried out, we did not put their finite a lattice spacing results on our figure.

Whilst writing up our results, a paper [S56] appeared with findings similar to ours. The authors used two techniques to determine χ. The first technique is based on the gradient flow the second one on the disconnected chiral condensate. The two methods give compatible result (within their errorbars). Between 250 to 330 MeV they were even able to carry out a controlled continuum extrapolation. Their decay exponent is in good agreement with the DIGA prediction, which means that their findings are consistent with ours, albeit with a larger errorbar.

Independent analyses, e.g. where all simulations are carried out with chiral fermions, could further clarify the present situation.

S11 Axion dark matter from misalignment

The details of axion production via the misalignment mechanism are well described in the literature (see e.g. [S58]) but for completeness we briefly discuss our calculations.

In order to calculate the amount of axions produced we have to solve the equation of motion for the $A(x)$ axion field or equivalently for the $\theta(x) = A(x)/f_A$ axionic angle in an expanding universe:

$$\frac{d^2\theta}{dt^2} + 3H(T)\frac{d\theta}{dt} + \frac{dV(\theta)}{d\theta} = 0, \tag{S29}$$

where $V(\theta) = m_A^2(T)(1-\cos\theta) = \chi(T)/f_A^2(1-\cos\theta)$ is the temperature dependent axion potential. Since we focus on the misalignment mechanism, we assume that θ changes slowly in space on the relevant scales. Spatial fluctuations and defects lead to a string contribution which we do not discuss here.

Note that in addition to the Hubble friction there is also a thermal friction term in Equation (S29) caused by sphaleron-like transitions [S59]. This term can be neglected if $\frac{\Gamma_{sph}}{f_A^2 T} \ll 3H$, where Γ_{sph} is the $SU(3)$ sphaleron rate. The sphaleron rate, being the diffusion coefficient for the topological charge, is very difficult to estimate. At weak coupling, using the effective classical field theory of [S60], the leading log sphaleron rate was calculated in [S61, S62]. Using our results for the energy density and the results of [S62] (which also extends the leading log approximation by classical lattice simulations) for Γ_{sph}, we find that for our entire f_A range the thermal friction can be neglected as long as the temperature is below $10GeV$, which is the largest temperature considered in our calculation.

The expansion is governed by the Friedmann equations:

$$H^2 = \frac{8\pi}{3M_{Pl}^2}\rho \tag{S30}$$

$$\frac{d\rho}{dt} = -3H(\rho + p) = -3HsT \tag{S31}$$

where ρ, p and s are the energy density, pressure and entropy density of the early universe and M_{Pl} is the Planck mass. At the temperatures where axion production happens, the contribution of axions to these densities can be neglected. ρ and s can be expressed as:

$$\rho = \frac{\pi^2}{30}g_\rho T^4 \qquad s = \frac{2\pi^2}{45}g_s T^3 \tag{S32}$$

using the effective number of degrees of freedom of Figure 1. Since we determined the energy density and entropy density for a wide temperature range, the solution of these equations yields the following relation between the age of the universe (t) and its temperature:

$$\frac{dt}{dT} = -M_{Pl}\sqrt{\frac{45}{64\pi^3}}\frac{1}{T^3 g_s(T)\sqrt{g_\rho(T)}}\left(T\frac{dg_\rho(T)}{dT} + 4g_\rho(T)\right) \tag{S33}$$

With the help of this expression, Equation (S29) can be rewritten in terms of temperature derivatives:

$$\frac{d^2\theta}{dT^2} + \left[3H(T)\frac{dt}{dT} - \frac{d^2t}{dT^2}\Big/\frac{dt}{dT}\right]\frac{d\theta}{dT} + \frac{\chi(T)}{f_A^2}\left(\frac{dt}{dT}\right)^2\sin\theta = 0 \tag{S34}$$

We solve this equation by numerical integration with some initial angle θ_0 and vanishing first derivative. When the temperature is large, the θ angle is frozen to its initial value. It starts to roll down the potential around $T_{\rm osc}$ which is defined as $3H(T_{\rm osc}) = m_A(T_{\rm osc})$. At the same time the axion number density (n_A) starts to increase. After a few oscillations its ratio to the entropy density (n_A/s) converges to a finite value which is then conserved for the rest of the evolution. Figure S27 shows $T_{\rm osc}$ for a large range of axion masses/couplings. Note that a coupling close to the Planck scale results in a $T_{\rm osc}$ below the QCD phase

transition which emphasizes the need for the equation of state and $\chi(T)$ even for these low temperatures. Below the transition $T_{\rm osc} \propto m_A^{0.47}$ while above the transition $T_{\rm osc} \propto m_A^{0.17}$.

We start the numerical solution at $T = 5T_{\rm osc}$. The oscillation starts around $T_{\rm osc}$. We detect this by looking for the first sign change of θ which happens at T_s. We then extract n_A/s by averaging

$$\frac{n_A}{s}(T) = \frac{45}{2\pi^2}\frac{f_A^2}{m_A g_s T^3}\left[\frac{1}{2}\left(\frac{d\theta}{dT}/\frac{dt}{dT}\right)^2 + \frac{\chi(T)}{f_A^2}(1-\cos\theta)\right] \tag{S35}$$

for the temperature range $0.8 - 0.2T_s$. Throughout the solution of Equation (S34) we fix f_A (or equivalently m_A) and use our results for $\chi(T)$ and $\rho(T)$. The present axion energy density is obtained by using the conservation of n_A/s:

$$n_{A;\rm today} = \frac{n_A(T)}{s(T)} s_{\rm today} \qquad \rho_{A;\rm today} = m_A n_{A;\rm today} \tag{S36}$$

The current entropy of the universe is dominated by photons and neutrinos:

$$s_{\rm today} = \frac{2\pi^2}{45}(2T_\gamma^3 + 6\frac{7}{8}T_\nu^3) = \frac{2\pi^2}{45}\frac{43}{11}T_\gamma^3 \tag{S37}$$

where $T_\gamma = 2.725K$ is the cosmic microwave background temperature. This axion energy density has to be compared to the critical density or its dark matter component:

$$\Omega_{\rm axion} = \frac{\rho_{a;\rm today}}{\rho_{\rm crit}} \qquad R_A = \frac{\Omega_{\rm axion}}{\Omega_{DM}} \tag{S38}$$

In the pre-inflation scenario a single θ_0 and m_A (or f_A) determines R_A uniquely. Assuming $R_A = 1$ results in the curve in Figure 3 of the main text. In the post-inflation scenario all θ_0 angles are present with equal probabilities after the Peccei-Quinn transition and we have to average over them:

$$\overline{R_A}(m_A) = \frac{1}{2\pi}\int_{-\pi}^{\pi} R_A(\theta_0, m_A) d\theta_0 \tag{S39}$$

For the whole m_A range which is relevant for the post-inflation scenario, we found that up to a few per mil the average angle of $\theta_0 = 2.155$ can be used:

$$\overline{R_A}(m_A) = R_A(\theta_0 = 2.155, m_A) \tag{S40}$$

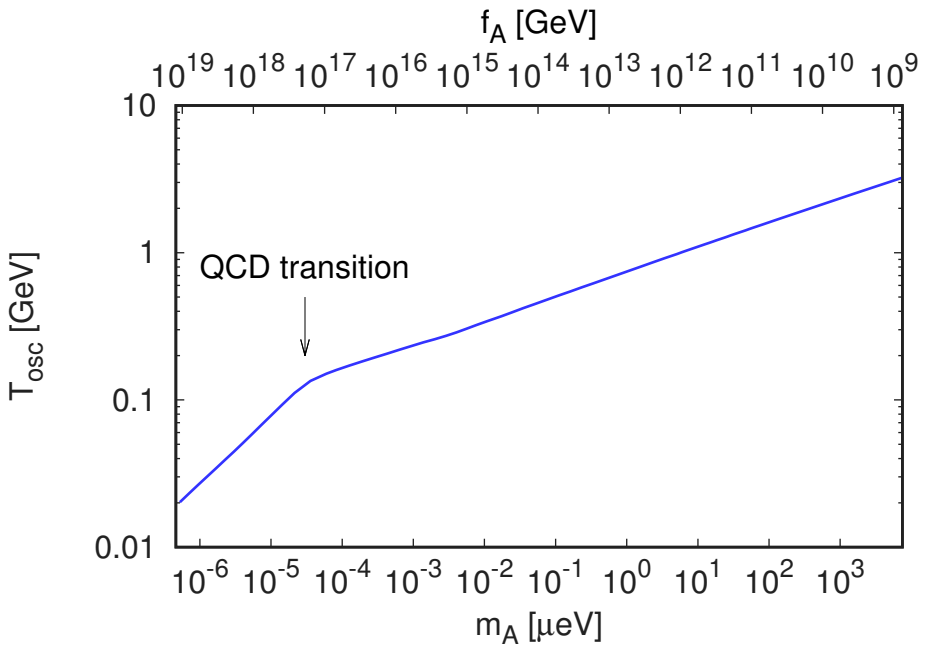


Figure S27: The oscillation temperature as a function of the axion mass. For this figure we assume that all the observed dark matter comes from the axions. Within this pre-inflation scenario the roll-down comes from a single θ_0 angle. The bend on the figure represents the QCD transition temperature. It reflects the very different behaviour of $\chi(T)$. Above the QCD transition χ rapidly drops, whereas below the QCD transition it has a much milder –almost constant– behaviour.

S12 Experimental searches for the axion in the predicted mass region

Using lattice QCD to determine the cosmological equation of state (EoS) and the temperature dependence of the topological susceptibility $\chi(T)$ the present paper showed that post-inflation dark matter axions have a mass between 50 and 1500 μeV (see Figure 3). In this section we show what should be done experimentally to approach and explore this mass region and detect dark matter axions. We conclude that, though presently operating and planned next generation experiments are not able to cover the predicted mass region, it is possible to design experiments, which offer discovery potential already in the near future.

Axions in the predicted 50-1500 μeV mass range are extremely challenging to detect. Recently, their theoretical appeal has been increasingly recognized and new techniques and experiments have been proposed. These include axion dark matter searches exploiting the excitation of atomic transitions in tuneable Rydberg atoms [S63] and electron spin precession [S64], but also purely laboratory searches for virtual axion long-range mediated forces [S65, S66]. Unfortunately, these are model dependent. The success of the former depends on a large axion-electron coupling and the latter upon the existence of new sources of CP violation beyond the Standard Model.

The most promising venue is to exploit the axion coupling to photons, $\mathcal{L}_{a\gamma} = -\theta \mathbf{E}\cdot\mathbf{B}\alpha C_{A\gamma}/(2\pi)$, with $C_{A\gamma}$ being an $\mathcal{O}(1)$ model-dependent constant. The simplest model compatible with the post-inflation scenario is the Kim-Shifman-Vainshtein-Zakharov (KSVZ) axion [S67, S68], which has $C_{A\gamma} = -1.92$ and we take it as our benchmark. The local axion dark matter field oscillations $\theta(t) \sim \theta_0 \cos(m_a t)$ in a homogeneous magnetic field $\mathbf{B}_e$ generate an electric field $\mathbf{E}_\theta = \theta(t)\mathbf{B}_e C_{A\gamma}\alpha/(2\pi)$.

The haloscope experiment of Sikivie [S69] uses this field to drive the resonant mode of a microwave cavity when the oscillation frequency ($\nu_A = m_A/2\pi$) coincides with its resonant frequency. Several collaborations have already employed this technique. The local dark matter density $\rho_{\rm dm} \simeq 0.3\,\mathrm{GeV/cm^3}$ fixes the amplitude of the oscillations $\theta_0 = \sqrt{2\rho_{\rm dm}/\chi(0)} \simeq 3.6\times10^{-19}$ and the electric field $|\mathbf{E}_\theta| = 1.2\times10^{-12}(|\mathbf{B}_e|/10\,\mathrm{T})$ V/m. Since the precise axion mass still remains unknown, the cavity has to be tuned to scan over the desired mass range. The bandwidth of the signal follows from the velocity dispersion of dark matter particles in the galactic halo $\Delta\nu \sim \nu_A/Q_A$ with $Q_A^{-1} = \langle(v/c)^2\rangle/2 \sim 10^{-6}$. In the mass range of interest, there are still $10^6\log(1500/50) \sim 3.4\times10^6$ channels to be explored in the frequency range ν_a =12-363 GHz. The power extracted from the cavity on resonance is given by

$$P_A = \kappa V m_A \max\{Q, Q_A\}\mathcal{G}|\mathbf{E}_\theta|^2/2, \tag{S41}$$

where $\mathcal{G} = \left(\int dV \mathbf{E}_m\cdot\mathbf{B}_e\right)^2/(|\mathbf{B}_e|^2 V\int dV|\mathbf{E}_m|^2)$ is the geometric overlap between the electric field of the cavity mode $\mathbf{E}_m$ with the background B-field, Q the quality factor, V the volume of the cavity, and κ the coupling coefficient (ratio of the power extracted to the full cavity losses), optimally set to ~ 0.5. The integration time required to find this signal with a given signal-to-noise-ratio S/N within the thermal and amplifier noise fluctuations is given by Dicke's radiometer equation

$$\Delta t = \Delta\nu\left(\frac{T_{\rm sys}}{P_A}\frac{S}{N}\right)^2 \tag{S42}$$

where $T_{\rm sys}$ is the system noise.

As an example, ADMX is a state of the art and only fully commissioned experiment [S70]. It utilizes a cylindrical cavity (1 m long, 0.5 m diameter, $Q\sim10^5$) in an 8 T magnetic solenoid in a dilution refrigerator reaching 100 mK and SQUID amplifiers with noise close to the quantum limit. A measurement campaign of three years is being started and has the sensitivity to find dark matter axions in the pre-inflation scenario in the region labeled ADMX in Fig. S28. Generation 2 (G2) experiments to reach higher frequencies are currently under preparation by the ADMX HF-group and the Center of Axion and Precision Physics (CAPP) in South Korea. Our estimated sensitivities with $Q\sim10^6$ and with cavities operated in fields of

up to 20 T may discover axions in the pre-inflation scenario up to $m_A = 30\,\mu$eV, see G2 region in Fig. S28. The post-inflation scenario predicted in this paper may only be partially explored by presently envisaged Generation 3 (G3) experiments. Still, this would require magnetic fields as strong as 40 T and combining signals of several tuneable cavities. At this moment it is not clear at all if the required technologies will ever be available for such a search. Nevertheless we include this G3 region in Fig. S28.

An alternative method was proposed in a recent paper [S71]. A spherical mirror in a strong magnetic field was shown to emit electromagnetic waves of frequency ν_A that focus at the center of curvature in response to the oscillating axion dark matter field. The power per unit dish area is

$$\frac{P_A}{A} = \frac{|\mathbf{E}_\theta|^2}{2} = 2.2 \times 10^{-27} \frac{\mathrm{W}}{\mathrm{m}^2} \left(\frac{C_{A\gamma} |\mathbf{B}_e|}{10\,\mathrm{T}} \right)^2 , \tag{S43}$$

too small for a wide-band search. However, it has been pointed out that the power can be enhanced by exploiting a dielectric planar mirror made of a sequence of N dielectrics [S72]. An equivalent power per area is emitted by each dielectric interface and can be added up coherently. This increases the power by a factor $4N^2$ (using a mirror at one end) which can be focused into a microwave receiver by a parabolic dish like the one used in [S73]. Detuning the dielectric thickness from $\lambda/2$, the dielectrics become partially reflecting, the power stored builds up like in a resonant cavity and the boost factor can be increased significantly so that a realistic axion search becomes feasible [S74].

Here, we envision a variable set of 20-40 sapphire dielectric slabs of ~ 1 mm thickness and 1 m^2 transverse area placed in a 10 T magnetic field with a planar mirror at one side. The distance between the dielectrics can be adjusted to have boost factors of order $\beta \sim 10^5$ in a relatively broad band $\mathcal{O}(50)$ MHz [S74]. Typically, one needs $d \sim \lambda/2$, which ranges between 3.1 cm and 2.4 mm in the axion mass range $40 - 250\,\mu$eV. For $m_A = 250\,\mu$eV, the coherence length of the axion field reaches $(m_A v)^{-1} \sim 1$ m. Thus, coherent detection with such large axion-photon transducers can be severely hampered for larger masses. A 3 year measurement campaign with such an apparatus may scan the 50-100 μeV range with sensitivity to KSVZ axions with commercial low-noise high-electron-mobility transistor (HEMT) amplifier technology and up to 250 μeV with quantum limited detection. The reach of such a tunable dielectric mirror is shown in Fig. S28 as a green-yellow band. The feasibility of an experiment of this type is currently being assessed at the Max-Planck-Institute for Physics in Munich.

This problem of coherence suggests us to reconsider the spherical dish antenna idea. Reaching sensitivity to KSVZ axions at 1 meV with a plain dish in a magnetic field of 10 T requires a sensitivity of 2.2 photons/(m^2 day). Again, the yield can be increased by a factor $4N^2$ when a few dielectric slabs are mounted (at adjustable relative distances) in front of each (planar) mirror element. With a boost of only $\mathcal{O}(100)$ and a total mirror area of 5 m^2 the photon rate would increase to 10^{-2} Hz with appears technically feasible for devices cooled down to temperatures of 10 mK and operating near the quantum limit [S64]. Still, it will be very challenging to shield the entire setup sufficiently against thermal noise.

In summary, we have shown that the region of high axion masses predicted by the QCD lattice calculations in this paper remains largely unexplored by presently operating and planned next generation experiments. This is mostly because of technical and practical limitations, particularly when searches over large ranges of axion masses are attempted. However, new experimental directions, some of which have been discussed in this paper and still being very challenging, may offer discovery potential already in the near future.

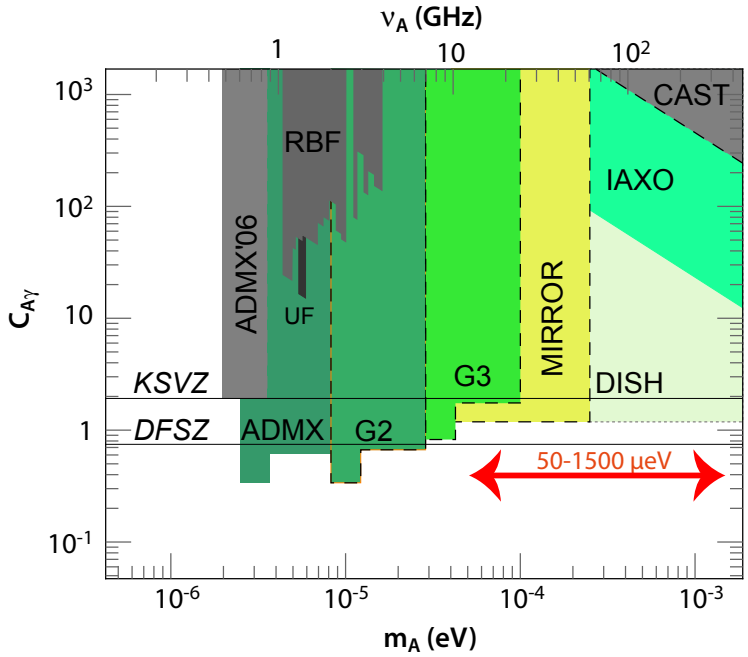


Figure S28: Sensitivity reach of the experiments discussed in the text as well as the International Axion Observatory (IAXO) [S75] (colored regions) with the current exclusion limits from previous cavity experiments: ADMX, RBF, UF and CAST (grey regions). We also show $C_{A\gamma}$ for the KSVZ and Dine-Fischler-Srednicki-Zhitnitsky (DSFZ) [S76, S77] models most interesting in the post-inflation and pre-inflation scenarios, respectively, and the range of m_A that could fit the dark matter abundance in the post-inflation scenario as follows from this work.

References

S1. Durr, S., Fodor, Z., Hoelbling, C. & Kurth, T. Precision study of the SU(3) topological susceptibility in the continuum. *JHEP* **04,** 055 (2007).

S2. Hasenbusch, M. Speeding up finite step size updating of full QCD on the lattice. *Phys. Rev.* **D59,** 054505 (1999).

S3. Bellwied, R. *et al.* Fluctuations and correlations in high temperature QCD. *Phys. Rev.* **D92,** 114505 (2015).

S4. Durr, S. *et al.* Lattice QCD at the physical point: light quark masses. *Phys. Lett.* **B701,** 265–268 (2011).

S5. Durr, S. *et al.* Lattice QCD at the physical point: Simulation and analysis details. *JHEP* **08,** 148 (2011).

S6. Clark, M. A., Joo, B., Kennedy, A. D. & Silva, P. J. Improving dynamical lattice QCD simulations through integrator tuning using Poisson brackets and a force-gradient integrator. *Phys. Rev.* **D84,** 071502 (2011).

S7. Yin, H. & Mawhinney, R. D. Improving DWF Simulations: the Force Gradient Integrator and the Móbius Accelerated DWF Solver. *PoS* **LATTICE2011,** 051 (2011).

S8. Lüscher, M. Properties and uses of the Wilson flow in lattice QCD. *JHEP* **08.** [Erratum: JHEP03,092(2014)], 071 (2010).

S9. Borsanyi, S. *et al.* High-precision scale setting in lattice QCD. *JHEP* **09,** 010 (2012).

S10. Borsanyi, S. *et al.* Axion cosmology, lattice QCD and the dilute instanton gas. *Phys. Lett.* **B752,** 175–181 (2016).

S11. Hernandez, V., Roman, J. E. & Vidal, V. SLEPc: A Scalable and Flexible Toolkit for the Solution of Eigenvalue Problems. *ACM Trans. Math. Software* **31,** 351–362 (2005).

S12. Durr, S. Taste-split staggered actions: eigenvalues, chiralities and Symanzik improvement. *Phys. Rev.* **D87,** 114501 (2013).

S13. Atiyah, M. F. & Singer, I. M. The Index of elliptic operators. 1. *Annals Math.* **87,** 484–530 (1968).

S14. Bazavov, A. *et al.* Topological susceptibility with the asqtad action. *Phys. Rev.* **D81,** 114501 (2010).

S15. Golterman, M. F. L. Staggered Mesons. *Nucl. Phys.* **B273,** 663–676 (1986).

S16. Ishizuka, N., Fukugita, M., Mino, H., Okawa, M. & Ukawa, A. Operator dependence of hadron masses for Kogut-Susskind quarks on the lattice. *Nucl. Phys.* **B411,** 875–902 (1994).

S17. Leutwyler, H. & Smilga, A. V. Spectrum of Dirac operator and role of winding number in QCD. *Phys. Rev.* **D46,** 5607–5632 (1992).

S18. Durr, S. *et al.* Lattice QCD at the physical point meets SU(2) chiral perturbation theory. *Phys. Rev.* **D90,** 114504 (2014).

S19. Borsanyi, S. *et al.* QCD thermodynamics with dynamical overlap fermions. *Phys. Lett.* **B713,** 342–346 (2012).

S20. Borsanyi, S. *et al.* QCD thermodynamics with continuum extrapolated dynamical overlap fermions. arXiv: `1510.03376 [hep-lat]` (2015).

S21. Durr, S. Logarithmic link smearing for full QCD. *Comput. Phys. Commun.* **180,** 1338–1357 (2009).

S22. Fukaya, H. *et al.* Lattice gauge action suppressing near-zero modes of H(W). *Phys. Rev.* **D74,** 094505 (2006).

S23. Fodor, Z., Katz, S. D. & Szabo, K. K. Dynamical overlap fermions, results with hybrid Monte Carlo algorithm. *JHEP* **08,** 003 (2004).

S24. Egri, G. I., Fodor, Z., Katz, S. D. & Szabo, K. K. Topology with dynamical overlap fermions. *JHEP* **01,** 049 (2006).

S25. Bode, A., Heller, U. M., Edwards, R. G. & Narayanan, R. *First experiences with HMC for dynamical overlap fermions* in *Lattice fermions and structure of the vacuum. Proceedings, NATO Advanced Research Workshop, Dubna, Russia, October 5-9, 1999* (1999), 65–68. arXiv: `hep-lat/9912043 [hep-lat]`.

S26. DeGrand, T. & Schaefer, S. Simulating an arbitrary number of flavors of dynamical overlap fermions. *JHEP* **07,** 020 (2006).

S27. Borsanyi, S. *et al.* The QCD equation of state with dynamical quarks. *JHEP* **11,** 077 (2010).

S28. Borsanyi, S. *et al.* Full result for the QCD equation of state with 2+1 flavors. *Phys. Lett.* **B730,** 99–104 (2014).

S29. Bazavov, A. *et al.* Equation of state in (2+1)-flavor QCD. *Phys. Rev.* **D90,** 094503 (2014).

S30. Laine, M. & Schroder, Y. Quark mass thresholds in QCD thermodynamics. *Phys. Rev.* **D73,** 085009 (2006).

S31. Cheng, M. Charm Quarks and the QCD Equation of State. *PoS* **LAT2007,** 173 (2007).

S32. Levkova, L. Effects of the charm quark on the QCD equation of state. *PoS* **LAT2009,** 170 (2009).

S33. Burger, F., Hotzel, G., Müller-Preussker, M., Ilgenfritz, E.-M. & Lombardo, M. P. Towards thermodynamics with $N_f = 2 + 1 + 1$ twisted mass quarks. *PoS* **Lattice2013,** 153 (2013).

S34. Bazavov, A. *et al.* Update on the 2+1+1 flavor QCD equation of state with HISQ. *PoS* **LATTICE2013,** 154 (2014).

S35. Borsanyi, S. *et al.* The QCD equation of state and the effects of the charm. *PoS* **LATTICE2011,** 201 (2011).

S36. Endrodi, G., Fodor, Z., Katz, S. D. & Szabo, K. K. The Equation of state at high temperatures from lattice QCD. *PoS* **LAT2007,** 228 (2007).

S37. Borsanyi, S., Endrodi, G., Fodor, Z., Katz, S. D. & Szabo, K. K. Precision SU(3) lattice thermodynamics for a large temperature range. *JHEP* **07,** 056 (2012).

S38. Durr, S. *et al.* Ab-Initio Determination of Light Hadron Masses. *Science* **322,** 1224–1227 (2008).

S39. Andersen, J. O., Leganger, L. E., Strickland, M. & Su, N. NNLO hard-thermal-loop thermodynamics for QCD. *Phys. Lett.* **B696,** 468–472 (2011).

S40. Zhai, C.-x. & Kastening, B. M. The Free energy of hot gauge theories with fermions through g**5. *Phys. Rev.* **D52,** 7232–7246 (1995).

S41. Olive, K. A. *et al.* Review of Particle Physics. *Chin. Phys.* **C38,** 090001 (2014).

S42. Kajantie, K., Laine, M., Rummukainen, K. & Schroder, Y. The Pressure of hot QCD up to g6 ln(1/g). *Phys. Rev.* **D67,** 105008 (2003).

S43. Karsch, F., Laermann, E. & Peikert, A. The Pressure in two flavor, (2+1)-flavor and three flavor QCD. *Phys. Lett.* **B478,** 447–455 (2000).

S44. Hindmarsh, M. & Philipsen, O. WIMP dark matter and the QCD equation of state. *Phys. Rev.* **D71,** 087302 (2005).

S45. Srednicki, M., Watkins, R. & Olive, K. A. Calculations of Relic Densities in the Early Universe. *Nucl. Phys.* **B310,** 693 (1988).

S46. D'Onofrio, M. & Rummukainen, K. Standard model cross-over on the lattice. *Phys. Rev.* **D93,** 025003 (2016).

S47. Laine, M., Nardini, G. & Rummukainen, K. First order thermal phase transition with 126 GeV Higgs mass. *PoS* **LATTICE2013,** 104 (2014).

S48. Laine, M. & Meyer, M. Standard Model thermodynamics across the electroweak crossover. *JCAP* **1507,** 035 (2015).

S49. Kajantie, K., Laine, M., Rummukainen, K. & Shaposhnikov, M. E. Generic rules for high temperature dimensional reduction and their application to the standard model. *Nucl. Phys.* **B458,** 90–136 (1996).

S50. D'Onofrio, M., Rummukainen, K. & Tranberg, A. Sphaleron Rate in the Minimal Standard Model. *Phys. Rev. Lett.* **113,** 141602 (2014).

S51. Bonati, C. *et al.* Axion phenomenology and θ-dependence from $N_f = 2 + 1$ lattice QCD. *JHEP* **03,** 155 (2016).

S52. Cundy, N. *et al.* Numerical methods for the QCD overlap operator IV: Hybrid Monte Carlo. *Comput. Phys. Commun.* **180,** 26–54 (2009).

S53. Aoki, S., Fukaya, H., Hashimoto, S. & Onogi, T. Finite volume QCD at fixed topological charge. *Phys. Rev.* **D76,** 054508 (2007).

S54. Borsanyi, S. *et al.* Ab initio calculation of the neutron-proton mass difference. *Science* **347,** 1452–1455 (2015).

S55. Fodor, Z. *et al.* Up and down quark masses and corrections to Dashen's theorem from lattice QCD and quenched QED. arXiv: 1604.07112 [hep-lat] (2016).

S56. Petreczky, P., Schadler, H.-P. & Sharma, S. The topological susceptibility in finite temperature QCD and axion cosmology. arXiv: 1606.03145 [hep-lat] (2016).

S57. Trunin, A., Burger, F., Ilgenfritz, E.-M., Lombardo, M. P. & Müller-Preussker, M. Topological susceptibility from $N_f = 2 + 1 + 1$ lattice QCD at nonzero temperature. *J. Phys. Conf. Ser.* **668,** 012123 (2016).

S58. Wantz, O. & Shellard, E. P. S. Axion Cosmology Revisited. *Phys. Rev.* **D82,** 123508 (2010).

S59. McLerran, L. D., Mottola, E. & Shaposhnikov, M. E. Sphalerons and Axion Dynamics in High Temperature QCD. *Phys. Rev.* **D43,** 2027–2035 (1991).

S60. Bodeker, D. On the effective dynamics of soft nonAbelian gauge fields at finite temperature. *Phys. Lett.* **B426,** 351–360 (1998).

S61. Moore, G. D. *Do we understand the sphaleron rate?* in *Strong and electroweak matter. Proceedings, Meeting, SEWM 2000, Marseille, France, June 13-17, 2000* (2000), 82–94. arXiv: hep-ph/0009161 [hep-ph].

S62. Moore, G. D. & Tassler, M. The Sphaleron Rate in SU(N) Gauge Theory. *JHEP* **02,** 105 (2011).

S63. Sikivie, P. Axion Dark Matter Detection using Atomic Transitions. *Phys. Rev. Lett.* **113,** 201301 (2014).

S64. Barbieri, R. *et al.* Searching for galactic axions through magnetized media: the QUAX proposal. arXiv: 1606.02201 [hep-ph] (2016).

S65. Arvanitaki, A. & Geraci, A. A. Resonantly Detecting Axion-Mediated Forces with Nuclear Magnetic Resonance. *Phys. Rev. Lett.* **113,** 161801 (2014).

S66. Crescini, N. *et al.* The QUAX-g_p g_s experiment to search for monopole-dipole Axion interaction. arXiv: 1606.04751 [physics.ins-det] (2016).

S67. Kim, J. E. Weak Interaction Singlet and Strong CP Invariance. *Phys. Rev. Lett.* **43,** 103 (1979).

S68. Shifman, M. A., Vainshtein, A. I. & Zakharov, V. I. Can Confinement Ensure Natural CP Invariance of Strong Interactions? *Nucl. Phys.* **B166,** 493–506 (1980).

S69. Sikivie, P. Experimental Tests of the Invisible Axion. *Phys. Rev. Lett.* **51.** [Erratum: Phys. Rev. Lett.52,695(1984)], 1415–1417 (1983).

S70. Asztalos, S. J. *et al.* A SQUID-based microwave cavity search for dark-matter axions. *Phys. Rev. Lett.* **104,** 041301 (2010).

S71. Horns, D. *et al.* Searching for WISPy Cold Dark Matter with a Dish Antenna. *JCAP* **1304,** 016 (2013).

S72. Jaeckel, J. & Redondo, J. Resonant to broadband searches for cold dark matter consisting of weakly interacting slim particles. *Phys. Rev.* **D88,** 115002 (2013).

S73. Suzuki, J., Inoue, Y., Horie, T. & Minowa, M. *Hidden photon CDM search at Tokyo* in *Proceedings, 11th Patras Workshop on Axions, WIMPs and WISPs (Axion-WIMP 2015)* (2015). arXiv: 1509.00785 [hep-ex].

S74. Millar, A., Raffelt, G., Redondo, J. & Steffen, F. *Layered dielectric haloscopes: a new way to detect axion dark matter,* MPP-2016-141, in preparation.

S75. Armengaud, E. *et al.* Conceptual Design of the International Axion Observatory (IAXO). *JINST* **9,** T05002 (2014).

S76. Dine, M., Fischler, W. & Srednicki, M. A Simple Solution to the Strong CP Problem with a Harmless Axion. *Phys. Lett.* **B104,** 199–202 (1981).

S77. Zhitnitsky, A. R. On Possible Suppression of the Axion Hadron Interactions. (In Russian). *Sov. J. Nucl. Phys.* **31.** [Yad. Fiz.31,497(1980)], 260 (1980).