Lattice QCD at the physical point: Simulation and analysis details

S. Dürr[1,2], Z. Fodor[1,2,3], C. Hoelbling[1], S.D. Katz[1,3], S. Krieg[1,2], T. Kurth[1], L. Lellouch[4], T. Lippert[1,2], K.K. Szabó[1], G. Vulvert[4]

[1]Bergische Universität Wuppertal, Gaussstr. 20, D-42119 Wuppertal, Germany.
[2]Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany.
[3]Institute for Theoretical Physics, Eötvös University, H-1117 Budapest, Hungary.
[4]Centre de Physique Théorique*, Case 907, Campus de Luminy, F-13288 Marseille, France.

[Budapest-Marseille-Wuppertal Collaboration]

We give details of our precise determination of the light quark masses $m_{ud} = (m_u + m_d)/2$ and m_s in $2+1$ flavor QCD, with simulated pion masses down to 120 MeV, at five lattice spacings, and in large volumes. The details concern the action and algorithm employed, the HMC force with HEX smeared clover fermions, the choice of the scale setting procedure and of the input masses. After an overview of the simulation parameters, extensive checks of algorithmic stability, autocorrelation and (practical) ergodicity are reported. To corroborate the good scaling properties of our action, explicit tests of the scaling of hadron masses in $N_f = 3$ QCD are carried out. Details of how we control finite volume effects through dedicated finite volume scaling runs are reported. To check consistency with SU(2) Chiral Perturbation Theory the behavior of M_π^2/m_{ud} and F_π as a function of m_{ud} is investigated. Details of how we use the RI/MOM procedure with a separate continuum limit of the running of the scalar density $R_S(\mu, \mu')$ are given. This procedure is shown to reproduce the known value of $r_0 m_s$ in quenched QCD. Input from dispersion theory is used to split our value of m_{ud} into separate values of m_u and m_d. Finally, our procedure to quantify both systematic and statistical uncertainties is discussed.

*CPT is research unit UMR 6207 of the CNRS and of the universities Aix-Marseille I, Aix-Marseille II and Sud Toulon-Var, and is affiliated with the FRUMAM.

Contents

1 Introduction

The goal of this paper is to give technical details of our calculation of the average light quark mass $m_{ud} = (m_u + m_d)/2$ and of the strange quark mass m_s at the physical mass point and in the continuum [1]. This calculation is from first principles and sets new standards in terms of controlling all systematic aspects of a direct calculation of quark masses. Because the values m_u and m_d are also of fundamental importance, we determine them by combining our results for m_{ud} and m_s with dispersive information based on $\eta \to 3\pi$ decays. A summary of recent determinations of light quark masses in $N_f = 2$ and $N_f = 2+1$ QCD is found in [1].

Let us begin by stating the minimum requirements for a first-principles determination of m_{ud} and m_s with fully controlled systematics:

1. The action should belong to the universality class of QCD according to standard arguments, based on locality and unitarity, and an exact algorithm should be used.
2. The light quark masses should be sufficiently close to their physical values such that an extrapolation, if necessary, can be performed without adding non-trivial assumptions. Our simulations are performed "*at the physical mass point*", i.e. with values of M_π and M_K which bracket the physical values; this eliminates the need for a "chiral extrapolation".
3. Simulations should be performed in volumes large enough to ensure that finite-volume effects are well under control (we use box sizes up to $L \simeq 6\,\mathrm{fm}$).
4. Simulations should be performed at no less than three lattice spacings a to make sure that a controlled extrapolation of all results to the continuum, $a \to 0$, can be performed.
5. All renormalizations should be performed nonperturbatively, and the final result should be given in a scheme which is well-defined beyond perturbation theory (we will give our results in the RI/MOM scheme).
6. The scale and other input masses should be set by quantities whose relation to experiment are direct and transparent (we use the masses of the Ω baryon, the pion and the kaon).

The present work contains additional innovative features which are not required to give an ab-initio result, but help to keep all systematic errors small:

7. We devise a method, tailored to needs of studies with Wilson-like fermions, to reconstruct the renormalized quark masses m_{ud} and m_s from the much simpler renormalization of the quantities m_s/m_{ud} and $m_s - m_{ud}$. We call it the "ratio-difference method".
8. We propose an approach which overcomes the RI/MOM "window-problem". It is based on taking a separate continuum limit of the evolution function $R_S(\mu, \mu')$ of the scalar density S from a scale $\mu \sim 2\,\mathrm{GeV}$, where the RI/MOM procedure yields reliable results, to a scale $\mu' \sim 4\,\mathrm{GeV}$ where one may make (controlled) contact with perturbation theory.
9. We use an advanced analysis procedure to assess the size of both the statistical and the systematic uncertainties (the same one as in [2]).

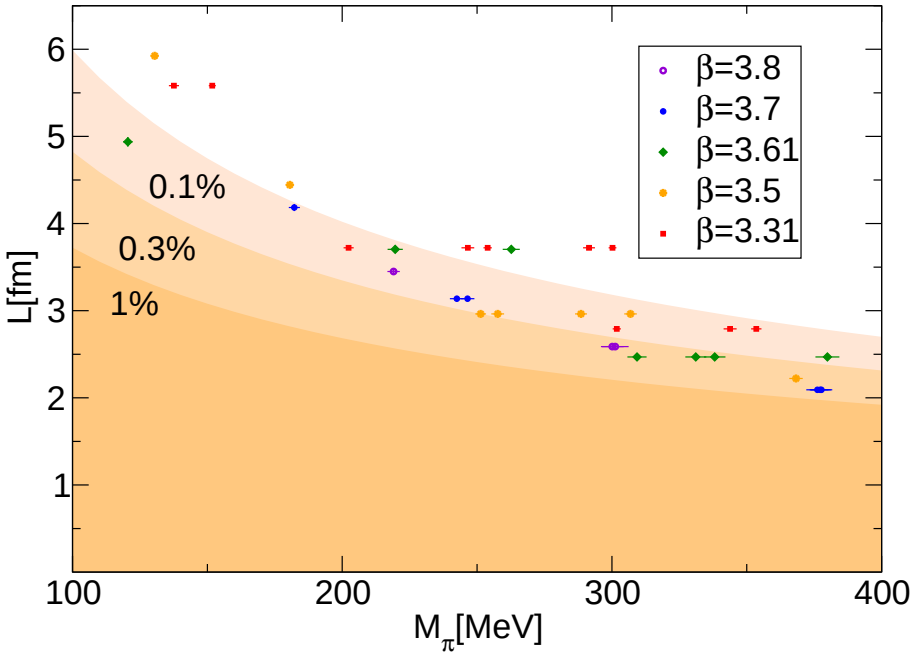

Figure 1: *Summary of our simulation points. The pion masses and the spatial sizes of the lattices are shown for our five lattice spacings. The percentage labels indicate regions, in which the expected finite volume effect [3] on M_π is larger than 1%, 0.3% and 0.1%, respectively. In our runs this effect is smaller than about 0.5%, but we still correct for this tiny effect.*

In our view, item 2 marks the beginning of a new era in numerical lattice QCD, because it avoids an extrapolation in quark masses which, generically, requires strong assumptions, thus relinquishing the first-principles approach (see the discussion in [1]).

To give the reader an overview of where we are in terms of simulated pion masses M_π and spatial box sizes L, a graphical survey of (some of) our simulation points is provided in Fig. 1 (with more details given in Sec. 5). We have data at 5 lattice spacings in the range $0.054 - 0.116\,\mathrm{fm}$, with pion masses down to $\sim 120\,\mathrm{MeV}$ and box sizes up to $\sim 6\,\mathrm{fm}$. Comparison with Chiral Perturbation suggests that our finite volume effects are typically below 0.5%, and close to the physical mass point (which is the most relevant part) even smaller. Still, we correct for them by means of Chiral Perturbation Theory [3], and test the correctness of this prediction through explicit finite volume scaling runs (see below).

The remainder of this paper is organized as follows. In Sec. 2 details are given concerning the action and algorithm employed, while Sec. 3 specifies how one determines the HMC force with HEX smeared clover fermions. Our choice of the scale setting procedure and of the input masses is discussed in Sec. 4, with simulation parameters tabulated in Sec. 5. Checks of algorithmic stability are summarized in Sec. 6, while autocorrelation and (practical) ergodicity issues are reported in Sec. 7. To corroborate the good scaling properties of our action, explicit tests of the scaling of hadron masses in $N_f = 3$ QCD are carried out, see Sec. 8. Details of how

we control finite volume effects through dedicated finite volume scaling runs are reported in Sec. 9. To test consistency with SU(2) Chiral Perturbation Theory the behavior of M_π^2/m_{ud} and F_π as a function of m_{ud} is investigated in Sec. 10. Details of how we use the RI/MOM procedure with a separate continuum limit of the running of the scalar density $R_S(\mu,\mu')$ are given in Sec. 11. To show the reliability of this procedure the known value of $r_0 m_s$ in quenched QCD is reproduced, see Sec 12. In Sec. 13 it is discussed how one may use input from dispersion theory to split our value of m_{ud} into separate values of m_u and m_d. Sec. 14 specifies our procedure to quantify both systematic and statistical uncertainties. Our final result is summarized in Sec. 15.

2 Action and algorithm details

We use a tree-level $O(a^2)$-improved Symanzik gauge action [4] together with tree-level clover-improved Wilson fermions [5], coupled to links which have undergone two levels of HEX smearing. The latter derives from the HYP setup [6], but with stout/EXP smearing [7] as the effective ingredient – see [8] for details. In terms of the original $[U_\mu(x)]$ and smeared $[V_\mu(x)]$ gauge links (see below) our action takes the form [4, 5]

$$\begin{aligned} S &= S_g^{\rm Sym} + S_f^{\rm SW} \\ S_g^{\rm Sym} &= \beta\,\Big[\frac{c_0}{3}\sum_{\rm plaq}\mathrm{Re\,Tr}\,(1-U_{\rm plaq}) + \frac{c_1}{3}\sum_{\rm rect}\mathrm{Re\,Tr}\,(1-U_{\rm rect})\Big] \\ S_f^{\rm SW} &= S_f^{\rm W}[V] - \frac{c_{\rm SW}}{2}\sum_x\sum_{\mu<\nu}(\overline{\psi}\,\sigma_{\mu\nu}F_{\mu\nu}[V]\,\psi)(x) \end{aligned} \tag{1}$$

with $\sigma_{\mu\nu}=\frac{\mathrm{i}}{2}[\gamma_\mu,\gamma_\nu]$ and $S_f^{\rm W}$ denoting the standard Wilson action, and where the expression for the field strength can be found in [5]. In $S_g^{\rm Sym}$ only the original thin links $U_\mu(x)$ are used. The parameters c_0, c_1 [4] and $c_{\rm SW}$ [5] are set to their tree-level values

$$c_1 = -1/12\,, \quad c_0 = 1-8c_1 = 5/3\,, \quad c_{\rm SW} = 1\,. \tag{2}$$

Note that both the hopping part and the clover term in $S_f^{\rm SW}$ use the same type of HEX-smeared links $V_\mu(x)\equiv V_\mu^{(2)}(x)$. Those are constructed from the thin links $V_\mu^{(0)}(x)\equiv U_\mu(x)$ via [8]

$$\begin{aligned} \Gamma^{(1,n)}_{\mu,\nu\rho}(x) &= \sum_{\pm\sigma\neq(\mu,\nu,\rho)} V_\sigma^{(n-1)}(x)V_\mu^{(n-1)}(x+\hat\sigma)V_\sigma^{(n-1)}(x+\hat\mu)^\dagger \\ V^{(1,n)}_{\mu,\nu\rho}(x) &= \exp\Big(\frac{\alpha_3}{2}P_{\rm TA}\Big\{\Gamma^{(1,n)}_{\mu,\nu\rho}(x)V_\mu^{(n-1)}(x)^\dagger\Big\}\Big)V_\mu^{(n-1)}(x) \\ \Gamma^{(2,n)}_{\mu,\nu}(x) &= \sum_{\pm\sigma\neq(\mu,\nu)} V^{(1,n)}_{\sigma,\mu\nu}(x)V^{(1,n)}_{\mu,\nu\sigma}(x+\hat\sigma)V^{(1,n)}_{\sigma,\mu\nu}(x+\hat\mu)^\dagger \\ V^{(2,n)}_{\mu,\nu}(x) &= \exp\Big(\frac{\alpha_2}{4}P_{\rm TA}\Big\{\Gamma^{(2,n)}_{\mu,\nu}(x)V_\mu^{(n-1)}(x)^\dagger\Big\}\Big)V_\mu^{(n-1)}(x) \\ \Gamma^{(3,n)}_{\mu}(x) &= \sum_{\pm\nu\neq(\mu)} V^{(2,n)}_{\nu,\mu}(x)V^{(2,n)}_{\mu,\nu}(x+\hat\nu)V^{(2,n)}_{\nu,\mu}(x+\hat\mu)^\dagger \end{aligned}$$

$$V_\mu^{(3,n)}(x) \;=\; \exp\Big(\frac{\alpha_1}{6}\,P_{\rm TA}\Big\{\Gamma_\mu^{(3,n)}(x)V_\mu^{(n-1)}(x)^\dagger\Big\}\Big)V_\mu^{(n-1)}(x) \;\equiv\; V_\mu^{(n)}(x) \tag{3}$$

without summation over repeated indices. Here

$$P_{\rm TA}\{M\} = \frac{1}{2}[M-M^\dagger] - \frac{1}{6}{\rm Tr}[M-M^\dagger] \tag{4}$$

denotes the traceless anti-hermitean part of the 3×3 matrix M. We use the parameters $\alpha_1 = 0.95$, $\alpha_2 = 0.76$ and $\alpha_3 = 0.38$. In terms of the standard stout/EXP smearing convention [7]

$$\begin{aligned}\Gamma_\mu^{(n)}(x) &= \sum_{\pm\nu\neq\mu} V_\nu^{(n-1)}(x)V_\mu^{(n-1)}(x+\hat\nu)V_\nu^{(n-1)}(x+\hat\mu)^\dagger \\ V_\mu^{(n)}(x) &= \exp\Big(\rho\,P_{\rm TA}\Big\{\Gamma_\mu^{(n)}(x)V_\mu^{(n-1)}(x)^\dagger\Big\}\Big)V_\mu^{(n-1)}(x)\end{aligned} \tag{5}$$

the values of α_i, $i = 1, 2, 3$, above correspond to ρ=0.158333, 0.19 and 0.19, respectively. This smearing differs from the one we used in [2, 9] in that the fermions interact even more locally with the gauge fields here (cf. the discussion in the supplementary online material of [2] and the appendix of [8]). Note, finally, that our action is exactly as ultralocal (in position space) as the original Wilson/clover action, since $D(x, y) = 0$ for $|x-y| > 1$.

As we use the hybrid Monte-Carlo (HMC) algorithm [10], a non-trivial ingredient with this action is the coding of the molecular dynamics (MD) force, which will be discussed in the next section. The MD updates are performed in quadruple precision, to ensure exact reversibility in our target (double precision) accuracy. Further particulars of our implementation – even-odd preconditioning [11], multiple time-scale integration ("Sexton-Weingarten scheme") [12], mass preconditioning ("Hasenbusch trick") [13], Omelyan integrator [14], RHMC acceleration with multiple pseudofermions [15] and mixed-precision solver [9] – have been described in [9]. As has been noted in the literature [16, 17], combining several of these ingredients yields a dramatic reduction of the critical slowing down that has traditionally been observed for light quark masses. As we show in this paper, the thorough combination of all these ingredients allows for simulations directly at the physical mass point, in large volumes, with several lattice spacings, such that a controlled extrapolation to the continuum can be performed.

3 HMC force with smeared links

We use two steps of HEX smearing [8] in our fermion action S_f, both for the Wilson and for the clover term. Our S_f depends on the thin (unsmeared) links only through the smeared links

$$S_f = S_f(V^{(2)}(V^{(1)}(V^{(0)} = U))) \tag{6}$$

where $V^{(n)}$ denotes the HEX smeared links, with n indicating the smearing level. Generically, the fermionic contribution to the HMC force is given by the gauge derivative $\delta S_f/\delta U$. In order

to obtain the derivative of S_f with respect to U for our two-step smearing recipe, we will apply the chain rule twice, which leads us to the following scheme. First one calculates

$$\frac{\delta S_f}{\delta V^{(2)}} \tag{7}$$

which encodes the details of how the fermions are coupled to the smeared gauge fields. This part of the calculation is not related to the smearing, one just takes $\delta S_f[U]/\delta U$ and replaces $U \to V^{(2)}$. The main consequence of the nested dependence (6) is the recursion formula

$$\frac{\delta S_f}{\delta V^{(n-1)}} = \frac{\delta S_f}{\delta V^{(n)}} \star \frac{\delta V^{(n)}}{\delta V^{(n-1)}} \tag{8}$$

where the proper definition of the star-product and of the second term will be given below. Thinking in terms of routines of the computer code, one such step takes the previous derivative $\delta S_f/\delta V^{(n)}$ and the links $V^{(n)}, V^{(n-1)}$ as input and yields the next derivative $\delta S_f/\delta V^{(n-1)}$. This procedure needs to be called n times, at the end we obtain the final fermion force

$$\frac{\delta S_f}{\delta U} = \frac{\delta S_f}{\delta V^{(0)}} \ . \tag{9}$$

Since the extension to a second and possibly more steps is straightforward, we will only consider the derivation for one level of HEX smearing in the following.

We now specify the two main ingredients in the derivation of the HMC force for fat-link actions, that is the gauge derivative and the pertinent chain rule. Since an $SU(3)$ matrix in the fundamental representation is a complex 3×3 matrix with only 8 independent degrees of freedom, it is a structured matrix and the derivative has to be defined properly.

The Lie algebra. For $U \in SU(3)$ an infinitesimal change can be written as

$$U \to U' = \exp\Big(\sum_A u_A T_A\Big) U \tag{10}$$

with real parameters u_A and the set $\{T_A | A = 1 \ldots 8\}$ forming a basis in the space of the traceless, anti-hermitian matrices, i.e. of the tangent space of the group. These matrices are normalized through $\mathrm{Tr}(T_A T_B) = -\delta_{AB}$. Using the trace, one can define a scalar product on this vector space; for $X = \sum_A x_A T_A$ and $Y = \sum_A y_A T_A$ in the Lie algebra the scalar product

$$\langle X, Y \rangle = -\mathrm{Tr}(XY) = \sum_A x_A y_A \tag{11}$$

allows one to build a projector which restricts an arbitrary 3×3 matrix, M, to the tangent space $P_{\mathrm{TA}}(M) = -\sum_A T_A \mathrm{Re}\,\mathrm{Tr}(T_A M)$, with $P_{\mathrm{TA}}(M)$ defined in (4). Furthermore, it is easy to show that for arbitrary matrices M and N

$$\mathrm{Re}\,\mathrm{Tr}(P_{\mathrm{TA}}(M)N) = \mathrm{Re}\,\mathrm{Tr}(P_{\mathrm{TA}}(N)M) \ . \tag{12}$$

Complex valued functions. Let f be a complex valued function on the group $SU(3)$. Its derivative with respect to the group element is a vector in the Lie algebra

$$\frac{\delta f}{\delta U} = \sum_A T_A \left[\frac{\delta f}{\delta U}\right]_A \tag{13}$$

where the components are defined as

$$\left[\frac{\delta f}{\delta U}\right]_A = \frac{\partial f(U')}{\partial u_A} = \lim_{u_A \to 0} \Big[f(\exp(u_A T_A)U) - f(U)\Big]/u_A = \mathrm{Tr}\Big(T_A U \frac{\partial f}{\partial U^T}\Big) \ . \tag{14}$$

Throughout, the partial derivative with respect to U under the trace is to be understood as a derivative with respect to unconstrained matrix elements. In particular, this means that

$$\frac{\partial U_{cd}}{\partial U_{ab}} = \delta_{ca}\delta_{bd} \ . \tag{15}$$

If f depends on the adjoint matrix $U^\dagger$, then using the identity $U^\dagger = U^{-1}$ this dependence is converted into a dependence on U, with the consequence that

$$\frac{\partial U^\dagger_{cd}}{\partial U_{ab}} = -U^\dagger_{ca} U^\dagger_{bd} \ . \tag{16}$$

Real valued functions. For a real valued function f the group derivative takes the form

$$\frac{\delta f}{\delta U} = \sum_A T_A \left[\frac{\delta f}{\delta U}\right]_A = \sum_A T_A \mathrm{Tr}\Big(T_A U \frac{\partial f}{\partial U^T}\Big) = -P_{\mathrm{TA}}\Big(U \frac{\partial f}{\partial U^T}\Big) \ . \tag{17}$$

SU(3) valued functions. Let $V \in SU(3)$ be a function of U, such that $V : SU(3) \to SU(3)$ is a mapping on the group space. If U changes as in (10), with small parameters $u_A = O(\epsilon)$, then V will also undergo a small change, which may be written as

$$V \to V' = V(U') = \exp(\sum_B v_B T_B) V \tag{18}$$

where the small parameters $v_B = O(\epsilon)$ are real-valued functions of the original parameters u_A. Below we shall need the u_A derivative of V', that is

$$\frac{\partial V'}{\partial u_A} = \frac{\partial(\sum_B v_B(u) T_B)}{\partial u_A} V \tag{19}$$

which, upon taking the limit $u_A \to 0$, implies

$$\left[\frac{\delta V}{\delta U}\right]_A = \sum_B \frac{\partial v_B}{\partial u_A}\Big|_{u=0} T_B V \ . \tag{20}$$

Chain rule. Let the function f depend on U only via V, and let us calculate its derivative with respect to U. Again, U transforms with infinitesimal parameters u_A, resulting in an infinitesimal change of the variables $v_B = v_B(u)$ of V. The usual chain rule yields

$$\frac{\partial f}{\partial u_A} = \sum_B \frac{\partial f}{\partial v_B} \frac{\partial v_B}{\partial u_A} \tag{21}$$

and, after taking the limit $u_A \to 0$, we arrive at

$$\left[\frac{\delta f}{\delta U}\right]_A = \sum_B \left[\frac{\delta f}{\delta V}\right]_B \frac{\partial v_B}{\partial u_A}\Big|_{u=0} . \tag{22}$$

With (11) and the definition of the gauge derivative from (13, 20), this may be rewritten as

$$\left[\frac{\delta f}{\delta U}\right]_A = -\mathrm{Tr}\Big(\frac{\delta f}{\delta V}\left[\frac{\delta V}{\delta U}\right]_A V^\dagger\Big) . \tag{23}$$

This formula is the chain rule for the gauge derivative, which can formally be stated as

$$\frac{\delta f}{\delta U} = \frac{\delta f}{\delta V} \star \frac{\delta V}{\delta U} . \tag{24}$$

With these ingredients we can now specify the HMC force for a fermion action with one step of HEX smearing. In the following we will simplify our notation by replacing $V^{(i,n)}$ with $V^{(i)}$. One HEX smearing, $U_\mu \to V_\mu^{(3)}$, is built from three substeps (cf. Eq. 3)

$$\begin{aligned}
V^{(1)}_{\mu,\nu,\rho}(x) &= \exp\Big(P_{\mathrm{TA}}[C^{(1)}_{\mu,\nu,\rho}(x)U^\dagger_\mu(x)]\Big)\cdot U_\mu(x) \\
V^{(2)}_{\mu,\nu}(x) &= \exp\Big(P_{\mathrm{TA}}[C^{(2)}_{\mu,\nu}(x)U^\dagger_\mu(x)]\Big)\cdot U_\mu(x) \\
V^{(3)}_{\mu}(x) &= \exp\Big(P_{\mathrm{TA}}[C^{(3)}_{\mu}(x)U^\dagger_\mu(x)]\Big)\cdot U_\mu(x)
\end{aligned} \tag{25}$$

where P_{TA} has been defined in (4) and $C^{(1)}$, $C^{(2)}$, $C^{(3)}$ represent staples constructed via

$$\begin{aligned}
C^{(1)}_{\mu,\nu,\rho}(x) &= \frac{\alpha_3}{2}\sum_{\pm\sigma\neq(\mu,\nu,\rho)} U_\sigma(x)U_\mu(x+\hat\sigma)U^\dagger_\sigma(x+\hat\mu) \\
C^{(2)}_{\mu,\nu}(x) &= \frac{\alpha_2}{4}\sum_{\pm\sigma\neq(\mu,\nu)} V^{(1)}_{\sigma,\mu,\nu}(x)V^{(1)}_{\mu,\sigma,\nu}(x+\hat\sigma)V^{(1)\dagger}_{\sigma,\mu,\nu}(x+\hat\mu) \\
C^{(3)}_{\mu}(x) &= \frac{\alpha_1}{6}\sum_{\pm\sigma\neq\mu} V^{(2)}_{\sigma,\mu}(x)V^{(2)}_{\mu,\sigma}(x+\hat\sigma)V^{(2)\dagger}_{\sigma,\mu}(x+\hat\mu)
\end{aligned} \tag{26}$$

with the factors $\alpha_i/(2(d-1))$ included (for reasons to become obvious below). In the following we will drop Lorentz indices and space-time arguments for simplicity.

The task is to calculate the HMC force [with $V = V^{(3)}$]

$$\frac{\delta S_f}{\delta U} = \frac{\delta S_f}{\delta V} \star \frac{\delta V}{\delta U}, \tag{27}$$

where $\delta S_f/\delta V$ is already known. The A component of the star product reads [see (23)]

$$\left[\frac{\delta S_f}{\delta U}\right]_A = -\Big[V^{(3)\dagger}\frac{\delta S_f}{\delta V^{(3)}}\Big]_{ab}\left[\frac{\delta V^{(3)}_{ba}}{\delta U}\right]_A = -\Sigma^{(3)}_{ab}\mathrm{Tr}\Big(T_A U\frac{\partial V^{(3)}_{ba}}{\partial U^T}\Big) \tag{28}$$

where $\Sigma^{(3)}$ contains the part of the force which is already known

$$\Sigma^{(3)} = V^{(3)\dagger}\frac{\delta S_f(V^{(3)})}{\delta V^{(3)}}\ . \tag{29}$$

Since S_f is a real-valued function we can write this in the compact form

$$\frac{\delta S_f}{\delta U} = P_{\mathrm{TA}}\Big(U\Sigma^{(3)}_{ab}\frac{\partial V^{(3)}_{ba}}{\partial U^T}\Big)\ . \tag{30}$$

To improve readability, let us temporarily denote $V^{(3)}$ by V. The last substep is then $V = \exp(A)U$, where $A = P_{\mathrm{TA}}(CU^\dagger)$ and thus

$$\frac{\delta S_f}{\delta U} = P_{\mathrm{TA}}\Big(U\Sigma_{ab}\frac{\partial V_{ba}}{\partial U^T}\Big) = P_{\mathrm{TA}}\Big(U\Sigma\exp(A)\Big) + P_{\mathrm{TA}}\Big(U\Sigma_{ab}\frac{\partial\exp(A)_{bc}}{\partial U^T}U_{ca}\Big)\ . \tag{31}$$

The derivative of the exponential of a traceless anti-hermitian A reads [see Eq. (68) of [7]]

$$d(\exp(A)) = \mathrm{Tr}(A\cdot dA)B_1 + \mathrm{Tr}(A^2\cdot dA)B_2 + f_1 dA + f_2 A\cdot dA + f_2 dA\cdot A \tag{32}$$

with $B_{1,2}$ being second-order polynomials of A and $f_{1,2}$ complex constants which depend on the trace and determinant of A. Using the cyclicity of the trace in the color indices, we arrive at

$$\begin{aligned} &P_{\mathrm{TA}}\Big(U\Sigma_{ab}\frac{\partial V_{ba}}{\partial U^T}\Big) = P_{\mathrm{TA}}\Big(U\Sigma\exp(A)\Big) + \\ &P_{\mathrm{TA}}\Big(U\frac{\partial A_{ab}}{\partial U^T}\cdot[\mathrm{Tr}(U\Sigma B_1)A + \mathrm{Tr}(U\Sigma B_2)A^2 + f_1 U\Sigma + f_2 U\Sigma A + f_2 AU\Sigma]_{ba}\Big) \end{aligned} \tag{33}$$

where the second term contains the derivative of $A = P_{\mathrm{TA}}(CU^\dagger)$. We use the identity

$$P_{\mathrm{TA}}\Big(U\frac{\partial P_{\mathrm{TA}}(M)_{ab}}{\partial U^T}N_{ba}\Big) = P_{\mathrm{TA}}\Big(U\frac{\partial M_{ab}}{\partial U^T}P_{\mathrm{TA}}(N)_{ba}\Big) \tag{34}$$

valid for arbitrary M and N [see (12)] to shuffle the projector in $A = P_{\mathrm{TA}}(CU^\dagger)$ to the matrix in square brackets in (33). Next, we use that the derivative of $CU^\dagger$ can be written as

$$U\frac{\partial(CU^\dagger)_{ab}}{\partial U^T}P_{\mathrm{TA}}[...]_{ba} = U\frac{\partial C_{ab}}{\partial U^T}(U^\dagger P_{\mathrm{TA}}[...])_{ba} - P_{\mathrm{TA}}[...]\cdot CU^\dagger \tag{35}$$

and introduce a convenient notation for the expression in square brackets in (33) by means of

$$Z = U^\dagger P_{\mathrm{TA}}\Big[\mathrm{Tr}(U\Sigma B_1)A + \mathrm{Tr}(U\Sigma B_2)A^2 + f_1 U\Sigma + f_2 U\Sigma A + f_2 AU\Sigma\Big]\ . \tag{36}$$

With this at hand, and using the relation $\exp(A) = VU^\dagger$, we arrive at the compact expression

$$P_{\mathrm{TA}}\Big(U\Sigma_{ab}\frac{\partial V_{ba}}{\partial U^T}\Big) = P_{\mathrm{TA}}\Big(U(\Sigma V - ZC)U^\dagger + U\frac{\partial C_{ab}}{\partial U^T}Z_{ba}\Big)\,. \tag{37}$$

Finally, we reintroduce the superscript (3) and note that the U dependence of $C^{(3)}$ comes exclusively from $V^{(2)}$. With these adjustments, relation (37) takes the form

$$P_{\mathrm{TA}}\Big(U\Sigma^{(3)}_{ab}\frac{\partial V^{(3)}_{ba}}{\partial U^T}\Big) = P_{\mathrm{TA}}\Big(U(\Sigma^{(3)}V^{(3)} - Z^{(3)}C^{(3)})U^\dagger\Big) + P_{\mathrm{TA}}\Big(U\Sigma^{(2)}_{ab}\frac{\partial V^{(2)}_{ba}}{\partial U^T}\Big) \tag{38}$$

where $\Sigma^{(2)}$ is defined as

$$\Sigma^{(2)}_{ab} = \frac{\partial C^{(3)}_{cd}}{\partial V^{(2)}_{ba}}Z^{(3)}_{dc} \tag{39}$$

meaning that the term $\Sigma^{(2)}$ can be calculated in the similar way as $\Sigma^{(3)}$. This procedure can be iterated, and we find for an action with one step of HEX smearing

$$\frac{\delta S_f}{\delta U} = \sum_{i=3,2,1} P_{\mathrm{TA}}\Big(U(\Sigma^{(i)}V^{(i)} - Z^{(i)}C^{(i)})U^\dagger\Big) + P_{\mathrm{TA}}\Big(U\Sigma^{(0)}\Big) \tag{40}$$

where $\Sigma^{(i)}$ is defined as

$$\Sigma^{(i)}_{ab} = \frac{\partial C^{(i+1)}_{cd}}{\partial V^{(i)}_{ba}}Z^{(i+1)}_{dc} \quad \text{for} \quad i = 0, 1, 2\,. \tag{41}$$

The object $\partial C^{(i+1)}/\partial V^{(i)}$ is a straightforward staple derivative, where some care needs to be taken w.r.t. the Lorentz indices. With this formula, one may implement a routine which calculates the HMC force for a fermion action with one step of HEX smearing. The extension to a second smearing step is realized through a second call to this routine as shown in (6).

This calculation of the HMC force with HEX smeared fermion actions extends the results of [7, 18]. An early treatise of the HMC force for fat-link fermion actions is found in [19].

4 Scale setting and input masses

To set the scale and to adjust the light and strange quark masses $m_{ud} = (m_u + m_d)/2$ and m_s to their physical values, we need to identify three quantities which can be precisely computed on the lattice and measured in experiment. We will use the mass of the Ω baryon and of the pseudoscalar mesons π, K for this purpose, in the latter case with a small correction for isospin-breaking and electromagnetic effects (see below).

In other words, at the point where M_π/M_Ω and M_K/M_Ω agree with the physical values of these ratios, the measured value of aM_Ω is identified with the lattice spacing times $1.672\,\mathrm{GeV}$ [20], and this yields a^{-1}. In [2] we used also the Ξ baryon to set the scale. As discussed there,

correlation functions of spin $3/2$ baryons are somewhat noisier than those of spin $1/2$ baryons. On the other hand, the more light valence quarks present in a baryon, the larger the fluctuations. In [2] with M_π down to $190\,\mathrm{MeV}$ these two effects balanced each other, rendering the Ω and the Ξ equally good choices. In the present paper we go down to $M_\pi = 120\,\mathrm{MeV}$, and the Ω with no light valence quark is the better choice. We use the standard mass-independent scale-setting scheme, in which this lattice spacing is subsequently attributed to all ensembles with the same coupling $\beta = 6/g_0^2$ and N_f.

Our ensembles bracket the physical quark masses m_{ud} and m_s in the sense that the span of simulated M_π^2 and $M_{s\bar{s}}^2 \equiv 2M_K^2 - M_\pi^2$ encompass the physical values given below. As a result, it suffices to use a parametrization of aM_Ω as a function of $(aM_\pi)^2$ and $(aM_{s\bar{s}})^2$ which describes the entire data set. We find that the Taylor ansatz $aM_\Omega = c_0 + c_1(aM_\pi)^2 + c_2(aM_{s\bar{s}})^2 + c_3(aM_{s\bar{s}})^4$ works perfectly.

Since our lattice simulations are performed in the isospin symmetric limit $m_d = m_u$ and do not account for electromagnetic interactions, the physical input meson masses must be corrected for these effects. The account of this as given by FLAG [21] is essentially a refined version of the analysis presented by MILC [22] some time ago. The bottom line is that in the framework mentioned above one should use the input masses $M_\pi = 134.8(3)\,\mathrm{MeV}$, $M_K = 494.2(5)\,\mathrm{MeV}$, which means that the electromagnetically corrected, isospin-averaged pseudoscalar input meson masses essentially agree with the PDG values of M_{π^0} and $\sqrt{\frac{1}{2}(M_{K^+}^2 + M_{K^0}^2)}$, respectively.

5 Simulation parameters

An overview of our $N_f = 2+1$ QCD simulations is presented in Tab. 1. For each ensemble we indicate the bare parameters, the lattice geometry, and the ensemble length in $\tau = 1$ units (after thermalization). In addition, the pion mass for the given parameters (determined with a specific choice of the fitting interval) is given. Note that the quoted error is only statistical – a detailed account of our procedure to keep track of the systematic uncertainties is given in Sec. 14. With Wilson fermions negative bare masses are not disturbing; after renormalization they will evaluate to positive quark masses (see Sec. 11). We work with spatial volumes as large as $L^3 \simeq (6\,\mathrm{fm})^3$ and temporal extents up to $T \simeq 8\,\mathrm{fm}$. Besides reducing finite-volume corrections and excited-state contaminations, large (four-dimensional) volumes tend to decrease statistical uncertainties to the same extent as increasing the number of trajectories (in a fixed volume) would do. For instance, in a L^4 box 1300 trajectories at $M_\pi L = 4$ are approximately equivalent to 4100 trajectories at $M_\pi L = 3$. With an HMC-type algorithm, the effort (at fixed pion mass) grows like L^5. Nevertheless, in view of the increased algorithmic stability (see below), choosing large four-dimensional volumes is a beneficial strategy.

The integrated autocorrelation times of the smeared plaquette and of the number of conjugate gradient iterations in the HMC accept/reject step are at most $O(10)$ trajectories. Depending on this autocorrelation time, the gauge field after every fifth or every tenth trajectory is stored as a configuration to be used for calculating hadronic observables.

β	$am_{ud}^{\rm bare}$	$am_{s}^{\rm bare}$	volume	# traj.	aM_π	$M_\pi L$
3.31	-0.07000	-0.0400	$16^3 \times 32$	1650	0.3530(12)	5.61
	-0.09000	-0.0400	$24^3 \times 48$	1600	0.2083(08)	5.00
	-0.09300	-0.0400	$24^3 \times 48$	4350	0.1775(06)	4.30
	-0.09300	-0.0400	$32^3 \times 48$	2500	0.1771(05)	5.65
	-0.09530	-0.0400	$32^3 \times 48$	1225	0.1500(13)	4.81
	-0.09756	-0.0400	$32^3 \times 48$	2600	0.1202(11)	4.00
	-0.09900	-0.0400	$48^3 \times 48$	1700	0.0887(06)	4.26
	-0.09933	-0.0400	$48^3 \times 48$	1240	0.0804(13)	3.94
	-0.09000	-0.0440	$24^3 \times 64$	1065	0.2024(10)	4.86
	-0.09300	-0.0440	$32^3 \times 64$	1030	0.1717(08)	5.50
	-0.09530	-0.0440	$32^3 \times 64$	1300	0.1457(09)	4.66
3.5	-0.04800	-0.0023	$32^3 \times 64$	1500	0.1354(06)	4.33
	-0.02500	-0.0060	$16^3 \times 32$	12000	0.2898(07)	4.62
	-0.03100	-0.0060	$24^3 \times 48$	3000	0.2535(05)	6.07
	-0.03600	-0.0060	$24^3 \times 48$	1800	0.2250(08)	5.41
	-0.04100	-0.0060	$24^3 \times 48$	4000	0.1921(05)	4.61
	-0.04370	-0.0060	$24^3 \times 48$	3900	0.1725(04)	4.13
	-0.04900	-0.0060	$32^3 \times 64$	1100	0.1212(08)	3.90
	-0.05294	-0.0060	$64^3 \times 64$	1100	0.0613(06)	3.92
	-0.04100	-0.0120	$24^3 \times 64$	1020	0.1884(08)	4.52
	-0.04630	-0.0120	$32^3 \times 64$	1065	0.1445(06)	4.62
	-0.04900	-0.0120	$32^3 \times 64$	1000	0.1174(06)	3.76
	-0.05150	-0.0120	$48^3 \times 64$	1200	0.0835(07)	4.01
3.61	-0.02000	0.0045	$32^3 \times 48$	2100	0.1993(3)	6.36
	-0.02800	0.0045	$32^3 \times 48$	3910	0.1488(4)	4.75
	-0.03000	0.0045	$32^3 \times 48$	2000	0.1322(4)	4.24
	-0.03121	0.0045	$32^3 \times 48$	2200	0.1211(2)	3.87
	-0.03300	0.0045	$48^3 \times 48$	2100	0.1026(4)	4.93
	-0.03440	0.0045	$48^3 \times 48$	1100	0.0864(4)	4.15
	-0.03650	-0.0030	$64^3 \times 72$	1004	0.0468(5)	3.00
	-0.02000	-0.0042	$32^3 \times 48$	1750	0.1969(4)	6.30
	-0.03000	-0.0042	$32^3 \times 48$	1450	0.1297(5)	4.17
3.7	-0.00500	0.0500	$32^3 \times 64$	1000	0.2227(04)	7.13
	-0.01500	0.0500	$32^3 \times 64$	1170	0.1711(03)	5.48
	-0.02080	0.0010	$32^3 \times 64$	1150	0.1251(04)	4.00
	-0.01500	0.0000	$32^3 \times 64$	1115	0.1644(04)	5.26
	-0.02080	0.0000	$32^3 \times 64$	1030	0.1245(06)	3.98
	-0.02540	0.0000	$48^3 \times 64$	1420	0.0821(03)	3.94
	-0.02700	0.0000	$64^3 \times 64$	1045	0.0603(03)	3.86
	-0.02080	-0.0050	$32^3 \times 64$	1405	0.1249(04)	4.00
	-0.02540	-0.0050	$48^3 \times 64$	1320	0.0806(03)	3.87

... to be continued on next page ...

... continued from previous page ...

3.8	-0.01400	0.0300	$32^3 \times 64$	1325	0.1242(5)	3.97
	-0.01900	0.0300	$48^3 \times 64$	1045	0.0830(4)	3.99
	-0.00900	0.0000	$32^3 \times 64$	2280	0.1523(4)	4.87
	-0.01400	0.0000	$32^3 \times 64$	1055	0.1249(5)	4.00
	-0.01900	0.0000	$48^3 \times 64$	1080	0.0836(4)	4.01
	-0.02100	0.0000	$64^3 \times 144$	1200	0.0598(2)	3.83

Table 1: *Overview of our $N_f = 2+1$ simulations. The scales at $\beta = 3.31, 3.5, 3.61, 3.7, 3.8$ are $a^{-1} = 1.697(6), 2.131(13), 2.561(26), 3.026(27), 3.662(35)\,\mathrm{GeV}$, respectively. Accordingly, the minimum pion mass per coupling is $M_\pi = 136(2), 131(2), 120(2), 182(2), 219(2)\,\mathrm{MeV}$.*

We put sources for the correlation functions on up to eight timeslices. For the precise form of the meson and baryon interpolating operators see e.g. [23]. To reduce the relative weight of excited states in correlation functions Gaussian sources and sinks are used, with a radius of about $0.32\,\mathrm{fm}$, which was found to be a good choice [2].

6 Algorithm stability

To detect potential instabilities of the HMC algorithm, different stability tests need to be performed. A rather complete battery of such tests was described in [9]. The pion masses used in this work are considerably smaller than those encountered in [2, 9]. For this reason, we repeat the relevant stability tests for the smallest-mass ensemble at each β, in particular for those with the physical pion mass.

With D the Wilson or clover operator, the spectrum of $A = D^\dagger D$ has no strict lower bound, i.e. the operator A is positive semi-definite, but not positive definite. If one could integrate the HMC trajectories exactly, this would not cause any problem, since an eigenvalue λ of A approaching zero would introduce an unbounded back-driving force in the HMC, so that the exact zero would be avoided. In practice, the trajectories are generated with a finite step-size integrator. Therefore, a very small λ_{min} in the MD evolution may experience a smaller back-driving force than it would in an exact evolution scheme, and this may trigger an instability.

If a particularly small eigenvalue appears during the molecular dynamics (MD) evolution, the solver in the MD force calculation will require more iterations to arrive at its target precision. More precisely, the inverse of the iteration count N_{CG} is closely related to the smallest eigenvalue of A. In a given ensemble, N_{CG}^{-1} shows an approximately Gaussian distribution [9]. As long as its median is away from zero by several standard deviations, the simulation is deemed safe [9, 24]. In Fig. 2 we show the "worst case scenario", i.e. the situation for the smallest quark masses in our set of ensembles. As one can see, even for pion masses as small as $120{-}135\,\mathrm{MeV}$ the inverse iteration count and hence the spectrum is bounded away from zero.

Alternatively, one can monitor the magnitude of the various contributions to the MD force

in the MD evolution. This is done in Fig. 3, where the maximum (over space-time) of each individual contribution to the total MD force during the MD integration is shown over a period of $10,000$ MD steps. As usual, the maximum of each contribution to the MD force fluctuates. However, it is important that the magnitude of these fluctuations is not too large. The more frequent spikes with large magnitude occur at any given MD step-size, the lower is the HMC acceptance rate, to the point where the algorithm becomes unstable. For small pion masses and coarse lattice spacings, the situation becomes even worse. This is why we show the ensemble with our smallest pion mass (around the physical pion mass) at the coarsest lattice spacing in Fig. 3. As one can see there are no dangerously large fluctuations present.

Finally, it is good practice to monitor the violation ΔE_{MD} of the MD energy conservation. In Fig. 4 we show again the "worst case scenario", that is the ensemble with our smallest pion mass at the coarsest lattice spacing. As one can see, the typical ΔE_{MD} in this simulation is small. For most of our simulations the acceptance rate is above 90%. Since the acceptance probability is given by $p_{\mathrm{acc}} = \min(1, \exp(-\Delta E_{\mathrm{MD}}))$, it is reasonable to use the data accumulated in the monitoring of the MD energy violation to check that $\langle \exp(-\Delta E_{\mathrm{MD}}) \rangle = 1$ within errors.

In summary, because (*a*) our algorithm is free of dangerous fluctuations of the clover eigen-

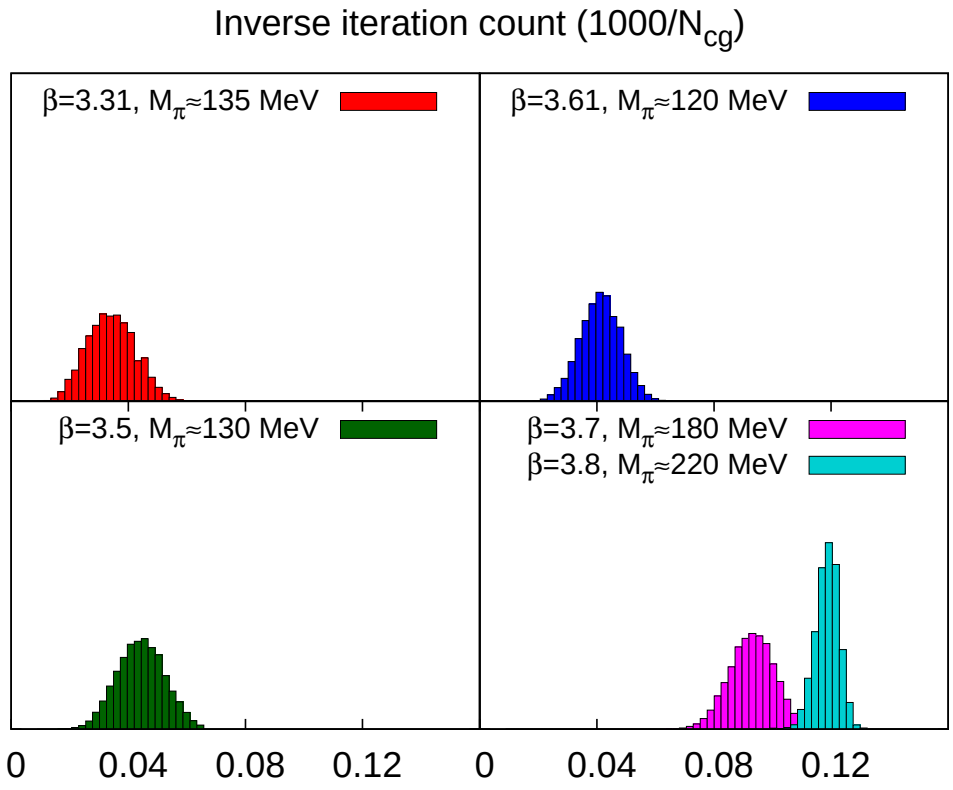

Figure 2: *Histogram of the inverse iteration count (N_{CG}^{-1} of the lightest pseudofermion) in the ensemble with the lightest quark mass per β. There is no danger of a tail stretching out to zero.*

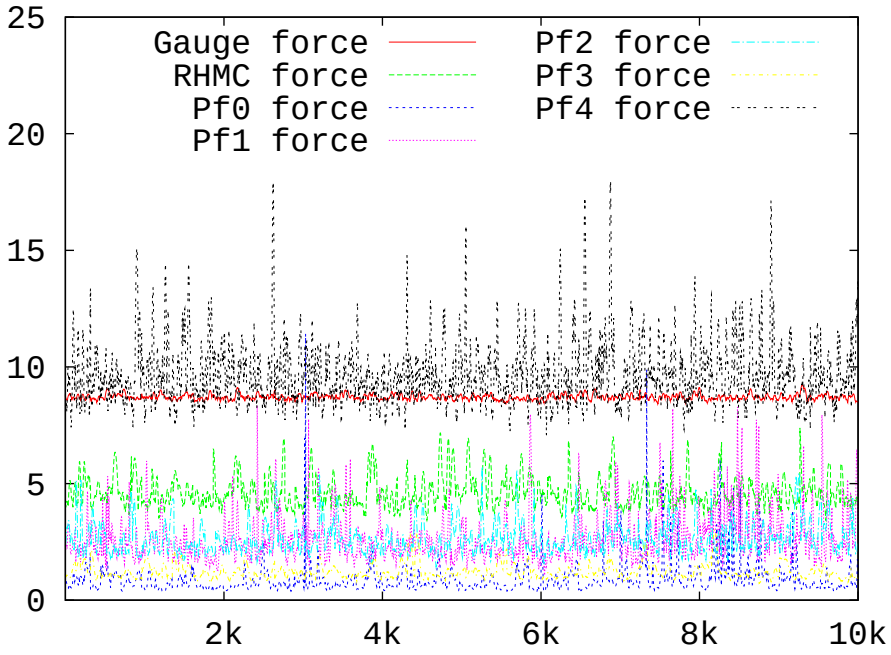

Figure 3: *Evolution of the maximum of each MD force during the MD integration. 256 steps correspond to one $\tau = 1$ trajectory. Shown is one production stream of our "physical pion mass" ensemble at $\beta = 3.31$. The other streams with the same parameters give a similar picture.*

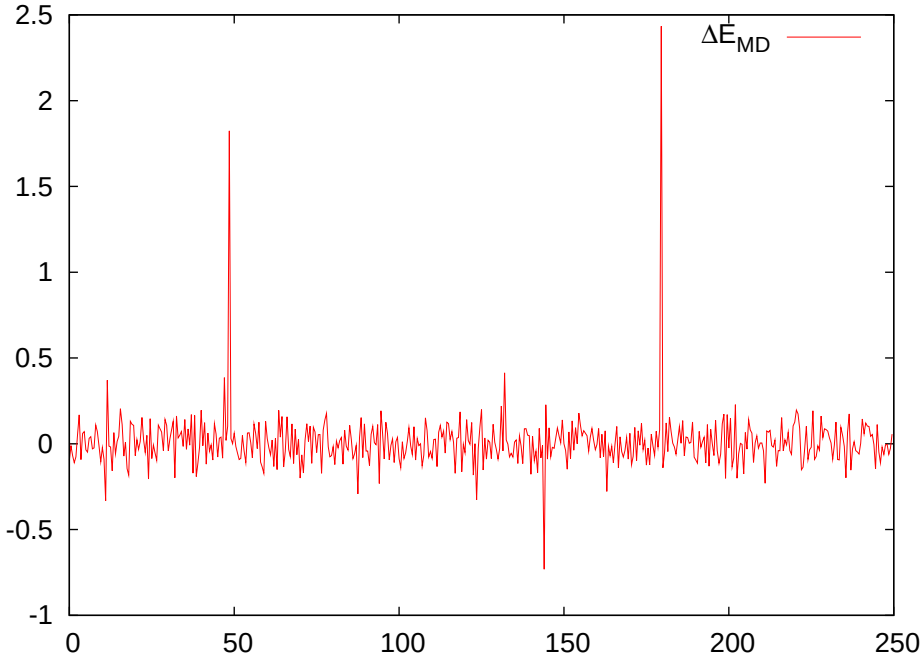

Figure 4: *Evolution of the energy violation over 250 units of MD time in the same simulation as in Fig. 3. All other simulations show similar or smaller energy violations.*

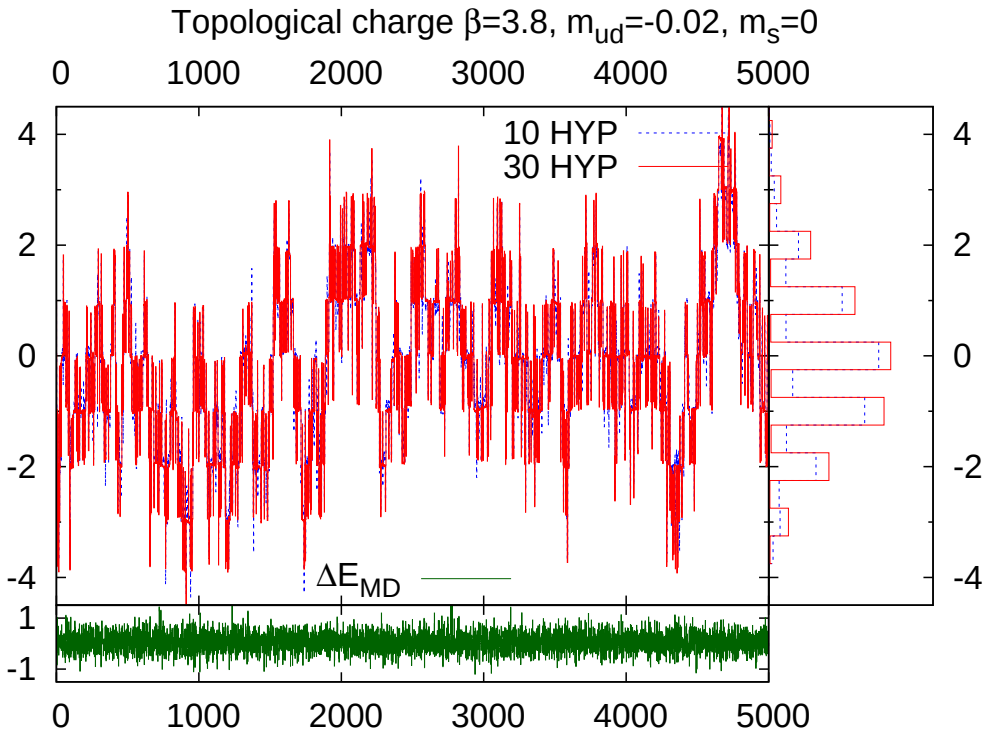

Figure 5: *Topological charge history at our finest lattice spacing ($\beta = 3.8$ corresponds to $a^{-1} \sim 3.7\,\mathrm{GeV}$) using two vigorous smearings (10 or 30 HYP steps) in the gluonic charge definition.*

value spectrum, (*b*) there are no dangerous fluctuations in the MD forces and (*c*) we therefore see that large violations of the MD energy conservation are absent in the simulation (resulting in high acceptance rates), we conclude that our setup is safe.

7 Autocorrelation and ergodicity checks

A known source of concern about HMC simulations in the regime of light quark masses and/or small lattice spacings is whether the Markov chain manages to sample configuration space sufficiently well, i.e. whether the algorithm is (in practical terms) ergodic [25–27].

We monitor two cheap gluonic quantities which are supposed to signal suspicious behavior, if there is any. The first one is the plaquette and/or Symanzik gauge action. With the plaquette it makes sense to consider smeared varieties, too, i.e. $\mathrm{Re}\,\mathrm{Tr}\,V_{\mathrm{plaq}}$ where V is a smeared gauge link, as described in Sec. 2. We find integrated autocorrelation times of such quantities to be at most $O(10)$ in units of unit-length ($\tau = 1$) trajectories. The second quantity is the bare field-theoretical (global) topological charge $q = a^4/(16\pi^2)\sum_{x,\mu<\nu}\mathrm{Tr}[F_{\mu\nu}(x)\tilde{F}_{\mu\nu}(x)]$ where $F_{\mu\nu}(x)$ is constructed from links which have undergone 10 or 30 steps of HYP smearing [6]. The result for our finest lattice spacing which, according to [25–27] represents the worst-case

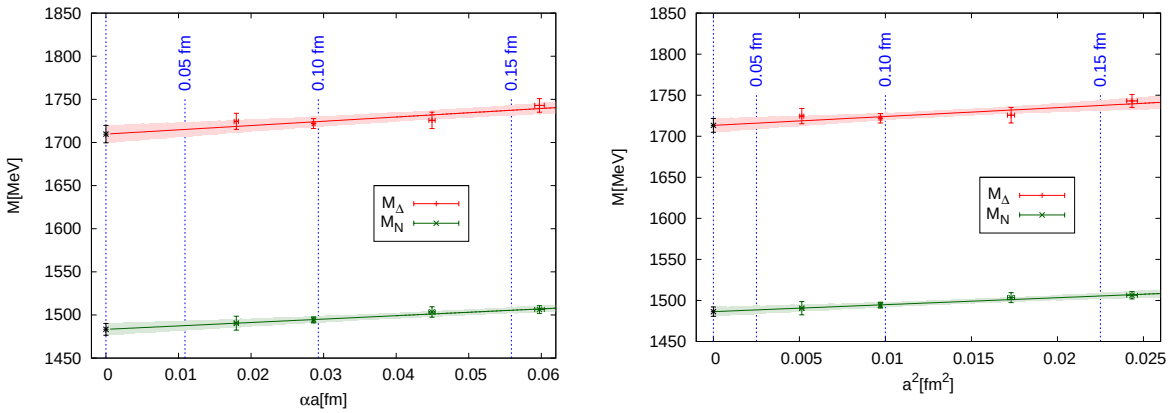

Figure 6: *Scaling of the nucleon and delta mass, at fixed $M_\pi/M_\rho = 0.67$, versus αa and a^2.*

scenario, is shown in Fig. 5. First of all, the 10 HYP and 30 HYP charges are always close to each other. Second, binning them with bin boundaries at $\mathbf{Z}/2+1/4$ yields a clear abundance of integer centered bins over half-integer centered bins, and this effect is more pronounced with the more vigorous smearing recipe. Last but not least, the histogram of either charge is reasonably symmetric after about 5000 trajectories, and the integrated autocorrelation time is $O(10)$ trajectories. Since this is the highest β simulated, we see no reason for concerns about the (practical) ergodicity of our simulations.

One should keep in mind that the topological tunneling rate may depend sensitively on the details of the action (e.g. whether Wilson, Symanzik or Iwasaki glue is used, whether smeared or unsmeared links are used in the fermionic part) and on the algorithm (e.g. the number of time scales and the specific choice of Hasenbusch masses).

8 $N_f = 3$ scaling test for hadron masses

Since the link smearing of the 2HEX action used in the present work differs from the smearing used in [2, 9], we decided to repeat the entire scaling test, as presented in [9], in all its detail for the new action.

To this end, we run a number of $N_f = 3$ simulations at various lattice spacings and various M_π/M_ρ ratios. For each β we interpolate the (common) octet and decuplet baryon mass, i.e. M_N/M_ρ and M_Δ/M_ρ, to the point where M_π/M_ρ assumes the value 0.67. The latter value coincides with $[2(M_K^{\rm phys})^2-(M_\pi^{\rm phys})^2]^{1/2}/M_\phi^{\rm phys}$, hence providing a way to tune to a quark mass which roughly corresponds to the physical strange quark mass. The results for M_N/M_ρ and M_Δ/M_ρ at this interpolation point are then extrapolated, linearly in αa and a^2, to the continuum. Throughout this report, $\alpha = g^2/(4\pi)$ denotes the strong coupling constant. For g^2 we use the perturbative values, at our lattice spacings, at 4-loop order (see below). The result of this procedure is shown in Fig. 6. Three points are worth emphasizing. First of all, the data are consistent with either scaling hypothesis over a large range of lattice spacings (out to $a \simeq$

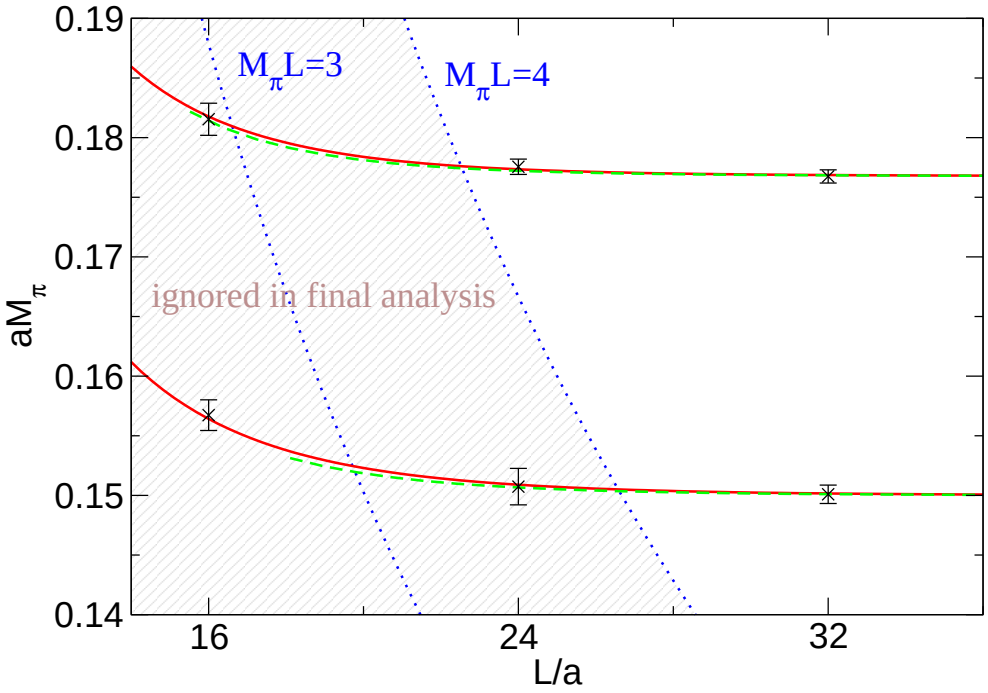

Figure 7: *Dedicated finite-volume analysis at* $\beta = 3.31$*, with* $M_\pi \simeq 250\,\mathrm{MeV}$ *(lower set of data) and* $M_\pi \simeq 300\,\mathrm{MeV}$ *(upper set). Results are compared to the prediction from Chiral Perturbation Theory. The fit to (42) is shown by solid red curves and the prediction of ChPT [3] is the green set of dashed curves. The steep dotted lines indicate the boundaries* $M_\pi L = 3$ *and* $M_\pi L = 4$*.*

$0.15\,\mathrm{fm}$), with a slight preference for $O(a^2)$ over $O(\alpha a)$ scaling, and this suggests that our tree-level value of c_{SW} (see Sec. 2 for the definition and details) is close to the nonperturbative value (which is not known for our action). This finding is in accordance with the results of [8]. Next, the continuum extrapolated values shown in Fig. 6 are in perfect agreement with the continuum extrapolated baryon masses found in [9] with a different action. Last but not least, the slope in either panel of Fig. 6 is small[1], and an action which shows generically a flat slope in scaling quantities is useful for obtaining precise predictions in the continuum.

In summary we find that both the 6stout action used in [2, 9] and the 2HEX action used in the present work exhibit small cut-off effects on standard hadron masses over a broad range of lattice spacings.

9 Finite volume corrections

For a fixed set of bare parameters, β, m_{ud}, m_s, energies and matrix elements of hadronic states depend on the spatial size L of the lattice. Typically, the finite volume tends to increase the

[1]The deviation of the result on the coarsest lattice from the continuum is 2.0% at most [Δ with $O(\alpha a)$ ansatz].

effective mass and to decrease the decay constant, relative to their infinite volume counterparts. As a first step it is important to assess by how much such effects would affect the data. For the theoretical treatment of these finite-volume effects it makes a difference whether the state considered is stable under strong interactions (despite the fact that, in a finite volume, the energy spectrum is discrete and all states are stable). The respective frameworks have been established by Lüscher, both for stable [28, 29] and unstable [30, 31] states. They allow us to quantify and eventually correct for finite-volume effects in a self-consistent manner (i.e. in a way in which only the results of our calculations and axioms of quantum field theory are used).

The structure of these corrections is most transparent for the case of a pion at 1-loop order in Chiral Perturbation Theory (ChPT) [32–34]. Up to higher order terms, the relative shift is

$$R_\pi(L) \equiv \frac{M_\pi(L)}{M_\pi} - 1 = \mathrm{const} \cdot M_\pi^2 \cdot \tilde{g}_1(M_\pi L) \tag{42}$$

where the shape function $\tilde{g}_1(x)$ has a well behaved expansion in terms of a Bessel function of the second kind, which itself has a large-x expansion of the form

$$\tilde{g}_1(x) = \frac{24K_1(x)}{x} + \frac{48K_1(\sqrt{2}x)}{\sqrt{2}x} + ... \tag{43}$$

$$K_1(x) = \Big(\frac{\pi}{2x}\Big)^{1/2} \exp(-x)\Big[1 + \frac{3}{8x} + ...\Big] \tag{44}$$

implying that finite volume corrections are exponentially suppressed at large L [28]. Higher loop orders for $R_\pi(L)$ have been worked out in [3]. For completeness we mention that analytic results for finite volume corrections of the nucleon are given in [35, 36]. The second reference predicts that for physical quark masses and $L = 5\,\mathrm{fm}$ box size (which is what we use in our smallest box at the physical mass point, the one at $\beta = 3.61$, cf. Tab.1) the nucleon experiences just a 2 permil finite-size shift. The point is that ChPT predicts the numerical value of the coefficient "coeff" in (42). In the chiral literature, the low-energy constants that enter "coeff" are pinned down from experiment (at leading order it is F_π).

To avoid using external input, we decide to stay content with just using the functional form of (42). This is permissible, since the shape function $\tilde{g}_1(x)$ is just the free Green's function of a massive scalar particle, summed over all spatial mirror copies (due to periodic boundary conditions in the spatial directions) [32], see also the discussion in [3]. We find that we can establish a global fit to all of our data in various volumes if we adjust the free coefficient in (42). A similar conclusion is reached for other stable hadrons, albeit with a different numerical value of the constant. For the case of the pion, we test the fitting ansatz and the analytic prediction [3] by comparing them to dedicated finite-volume scaling runs, as shown in Fig. 7. Both the fit (full line) and the prediction from ChPT [3] (dashed curve) agree with the data. The latter prediction has a limited range, since ChPT becomes questionable in boxes with $M_\pi L < 3$ [3]. It is important to emphasize that the data with $M_\pi L < 4$ in Fig. 7 have been generated to *test* our treatment of finite-volume effects, they do not enter the final analysis.

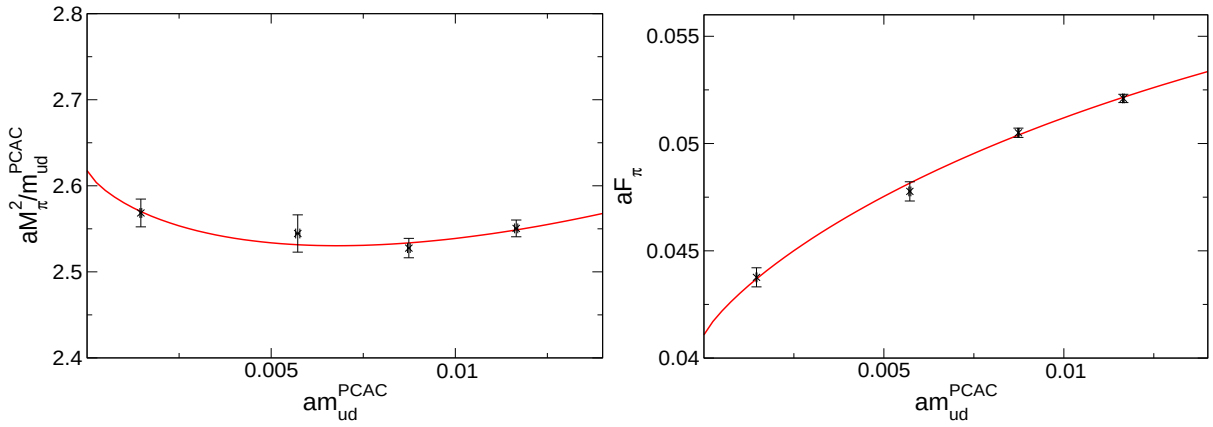

Figure 8: $M_\pi^2/m_{ud}^{\rm PCAC}$ *(left) and* F_π *(right) versus* $m_{ud}^{\rm PCAC}$ *(cf. Sec. 11) for our 4 lightest ensembles at* $\beta = 3.5$*, at fixed* $am_s = -0.006$*, which is close to* $m_s^{\rm phys}$*. A joint fit to the NLO chiral ansatz (45, 46) yields reasonable values of the low-energy constants. Error bars are statistical.*

These results confirm our rule of thumb that simulations with $M_\pi L \geq 4$ and/or $L \gtrsim 5\,\mathrm{fm}$ yield infinite-volume masses within statistical accuracy. An overview of the expected size of R_{M_π} in our simulations is given in Fig. 1. In all of these points the mass correction is less than about 5 permil, and for points close to $M_\pi^{\rm phys}$ (which dominate our analysis) it is even smaller. Nevertheless, we include these (tiny) shifts into our global analysis (cf. Sec. 14).

10 Chiral behavior of pion mass and decay constant

To illustrate the quality of our results obtained in lattice QCD calculations with physical or larger than physical values of the quark mass $m_{ud} = (m_u + m_d)/2$, we briefly investigate here whether the m_{ud} dependence of the pion mass and decay constant can be described by ChPT [37, 38] in this range of quark masses.

To this end we compare our results for M_π^2 and F_π versus m_{ud} at fixed (nearly physical) m_s (cf. Tab. 1) to the NLO predictions of the $SU(2)$ framework. The latter read [37]

$$M_\pi^2 = M^2\Big[1 + \frac{1}{2}x\log(\frac{M^2}{\Lambda_3^2})\Big] \tag{45}$$

$$F_\pi = F\ \Big[1 - x\log(\frac{M^2}{\Lambda_4^2})\Big] \tag{46}$$

with $x = M^2/(4\pi F)^2$ and $M^2 = 2Bm_{ud}$ a shorthand expression for the light quark mass (up to the factor $2B$, with $B = \Sigma/F^2$). The NNLO expressions can be found in [39]. In all of these expressions F, Σ, B refer to the pion decay constant, the absolute value of the quark condensate and the condensate parameter in the 2-flavor chiral limit $m_{ud} \to 0$ with m_s held

fixed. The terms in the square brackets proportional to x^0, x^1 represent the tree-level and 1-loop contributions respectively, and Λ_3, Λ_4 encode low-energy constants (LECs).

In Fig. 8 the quantities $M_\pi^2/m_{ud}^{\rm PCAC}$ and F_π are plotted versus the PCAC quark mass $m_{ud}^{\rm PCAC}$ (see below) at an intermediate lattice spacing ($\beta = 3.5$), where we reach down to $M_\pi \simeq 130\,\mathrm{MeV}$ (cf. Tab. 1). We find that our results with $M_\pi < 400\,\mathrm{MeV}$ can be jointly fitted with the NLO chiral formulae (45, 46), with acceptable $\chi^2/\mathrm{d.o.f.}$ and reasonable values of the low-energy constants. However, the extraction of Gasser-Leutwyler coefficients is beyond the scope of this paper and will be left for future publications.

11 RI/MOM renormalization of quark masses

Our primary goal is to determine the average up-down quark mass $m_{ud} = (m_u + m_d)/2$ and the strange quark mass m_s at the physical mass point in a "continuum" renormalization scheme, such as RI/MOM or RGI, using first-principles lattice computations.

In a lattice study quark masses and the running coupling have a different status than other observables, such as hadron masses and decay constants, since they are input parameters to the simulation. Consequently, one has to tune these parameters until the low-energy spectrum of the theory agrees with experiment (cf. [2] and Sec. 4), before one may read them off from the results of the simulation. To turn them into observables, one has to convert them from the original cut-off scheme (which is specific to the gluon and quark action combination used) to a scheme where the scale μ is not tied to the lattice spacing a.

The remainder of this section is organized as follows. In 11.1 our "ratio difference method" for extracting quark masses in the theory with Wilson fermions is explained, using standard terminology for the renormalization and improvement coefficients. It is important to notice that in the dynamical theory there is a subtlety in the renormalization pattern, due to quark line disconnected diagrams [40–42], but our "ratio difference method" steers around this complication, as explained in 11.2. In subsection 11.3 details of how we determine the flavor non-singlet scalar renormalization constant $Z_S(\mu)$ via the Rome-Southampton RI/MOM method [43] are given. In 11.4 it is specified how we control the systematics that arise from the dedicated $N_f = 3$ computations needed in the RI/MOM procedure. In 11.5 a summary is given.

11.1 Ratio-difference method in a nutshell

With Wilson-type fermions there are two options for obtaining the renormalized quark mass. On the one hand, one may start from the mass parameter $am^{\rm bare}$, as present in the Lagrangian, and apply both additive and multiplicative renormalization to build the VWI quark mass

$$m^{\rm VWI} = \frac{1}{Z_S}\left[1 - \frac{1}{2} b_S a m^{\rm W} + O(a^2)\right] m^{\rm W} \qquad \text{where} \qquad m^{\rm W} = m^{\rm bare} - m^{\rm crit} \tag{47}$$

and VWI means[2] "vector Ward identity". Here $Z_S = Z_S(\mu)$ denotes the lattice-to-continuum renormalization factor of the scalar density (it depends on the chosen scheme and scale, e.g. $\overline{\rm MS}$ and $\mu = 2\,{\rm GeV}$), b_S is an improvement coefficient (see below), and $m^{\rm crit}$ specifies the additive mass renormalization, i.e. the bare quark mass at which the pion becomes massless.

The other way to obtain the renormalized quark mass is as follows. For a pseudoscalar meson made out of valence quarks ψ_1, ψ_2 with Lagrangian masses $m_i^{\rm bare}$ or Wilson masses $m_i^{\rm W}$ ($i = 1, 2$), respectively, the sum of the (unrenormalized) PCAC quark masses is defined as

$$m_1^{\rm PCAC} + m_2^{\rm PCAC} = \frac{\sum_{\vec{x}} \langle \bar{\partial}_\mu [A_\mu(x) + a c_A \bar{\partial}_\mu P(x)] O(0) \rangle}{\sum_{\vec{x}} \langle P(x) O(0) \rangle} \tag{48}$$

where A_μ and P denote the axial current and the pseudoscalar density, respectively, O represents an arbitrary operator which couples to the meson (usually $O = P$ to maximize the signal), and $\bar{\partial}_\mu = (\partial_\mu + \partial_\mu^*)/2$ is the symmetric derivative with $(\partial_\mu \phi)(x) = (\phi(x + a\hat{\mu}) - \phi(x))/(2a)$. Then only a multiplicative renormalization is needed to form the (renormalized) "AWI quark mass"

$$m^{\rm AWI} = \frac{Z_A}{Z_P} \frac{1 + b_A a m^{\rm W} + O(a^2)}{1 + b_P a m^{\rm W} + O(a^2)} m^{\rm PCAC} \tag{49}$$

where AWI stands for "axial Ward identity". Here Z_A and $Z_P = Z_P(\mu)$ denote the lattice-to-continuum renormalization factor of the axial current and the pseudoscalar density, respectively.

The coefficients b_S, b_A, b_P, c_A in (47 - 49) are part of the improvement program. If properly set, $O(a^2)$ scaling of phenomenological quantities can be achieved, but they may be set to zero if one is content with $O(a)$ scaling. We use (47 - 49) with tree-level values of the improvement coefficients, that is $b_S = b_A = b_P = 1$ and $c_A = 0$. Formally, our results thus show cut-off effects proportional to αa, but in the scaling tests presented above cut-off effects proportional to a^2 seem to be numerically dominant. At this point we cannot anticipate whether a similar statement holds true for renormalized quark masses, and we shall thus consider both possibilities (i.e. leading cut-off effects proportional to αa or a^2). In any case the difference (in a given scheme, at a given μ) scales away with $a \to 0$, hence $m^{\rm AWI} = m^{\rm VWI}$ in the continuum.

In principle, $m_s^{\rm phys}$ and $m_{ud}^{\rm phys}$ may be determined using either definition of the quark mass, but in practice it proves beneficial to combine the specific advantages of the VWI and AWI approaches. Let us assume, for a moment, that we were to set all improvement coefficients to zero. Since the Lagrangian quark mass $m^{\rm bare}$ is an exact input quantity which, after a universal shift has been applied, multiplicatively renormalizes with the unproblematic scalar density [cf. (47)], it is natural to use $m^{\rm W}$ for quark mass *differences*, where the additive renormalization term $m^{\rm crit}$ drops out. On the other hand, the PCAC quark mass $m^{\rm PCAC}$ is perfectly suited to measure quark mass *ratios*, since in the ratio the multiplicative renormalization constants cancel [cf. (49)]. It is thus natural to measure the difference $m_s - m_{ud}$ via the Wilson or Lagrangian

[2]Strictly speaking the vector Ward identity constrains only quark mass differences. Below we will use $m^{\rm VWI}$ only in such differences, and by doing so the dependence on $m^{\rm crit}$ will persist only in an $O(a)$-suppressed term.

mass difference $d \equiv am_s^{\rm W} - am_{ud}^{\rm W} = am_s^{\rm bare} - am_{ud}^{\rm bare}$ and the ratio m_s/m_{ud} via the PCAC mass ratio $r \equiv m_s^{\rm PCAC}/m_{ud}^{\rm PCAC}$. In this case one obtains the renormalized masses through

$$am_{ud}^{\rm scheme} = \frac{1}{Z_S^{\rm scheme}} \frac{d}{r-1} \,, \qquad am_s^{\rm scheme} = \frac{1}{Z_S^{\rm scheme}} \frac{rd}{r-1} \tag{50}$$

and we shall refer to this strategy as the "ratio-difference method". The renormalization scheme will be specified below. In practice, things are slightly more involved, as we intend to maintain tree-level improvement. Setting $c_A = 0$ and $b_A = b_P = 1$ does not change anything in (48, 49), but having a quadratic term in (47) through $b_S = 1$ means that the difference $m_1^{\rm W} - m_2^{\rm W}$ does no longer coincide with $m_1^{\rm bare} - m_2^{\rm bare}$. In the next subsection we will show that even with improvement, the renormalized quark masses are given by (50) with $d \to d^{\rm imp}, r \to r^{\rm imp}$, where the latter quantities are defined in (60, 61).

11.2 Ratio-difference method in full QCD with improvement

In the dynamical theory, the renormalization pattern of quark masses is slightly more involved than the familiar equations (47, 49) suggest [42], but it turns out that our "ratio difference method" gets rid of these complications and the final relation is unchanged.

We now discuss how the findings of [42] apply to our method, using their notation, except that we do not use a "hat" to denote renormalized quantities, since they will come with a superscript "VWI" or "AWI", just as in the previous subsection. Equations (26, 48) of [42] read

$$m_j^{\rm VWI} = \frac{1}{Z_S} m_j^{\rm W} \Big[1 - \frac{1}{2} b_S am_j^{\rm W} - \bar{b}_S a {\rm Tr}(M) + O(a^2) \Big] + ... \tag{51}$$

$$m_j^{\rm AWI} = \frac{Z_A}{Z_P} m_j^{\rm PCAC} \Big[1 + (b_A - b_P) am_j^{\rm W} + (\bar{b}_A - \bar{b}_P) a {\rm Tr}(M) + O(a^2) \Big] \tag{52}$$

where Z_J is the flavor *non-singlet* renormalization constant ($J = S, A, P$) , and $b_J = 1 + O(\alpha)$, $\bar{b}_J = O(\alpha^2)$, $c_A = O(\alpha)$ denote improvement coefficients (which now depend on N_f). Finally, $m_j^{\rm W} = m_j^{\rm bare} - m^{\rm crit}$, with $m^{\rm crit}$ defined as the $N_f = 3$ critical mass (i.e. in the *unitary* direction), $m^{\rm PCAC}$ just as in (48), and the ellipses in (51) denote terms which depend on the quark masses only through ${\rm Tr}(M), {\rm Tr}(M^2), {\rm Tr}^2(M)$, where M is the (flavor diagonal) quark mass matrix. The new feature of formulas (51, 52) is the terms proportional to m_j times ${\rm Tr}(M) = \sum_f m_f^{\rm W}$. These terms make the renormalized quark mass of flavor j depend on all other quark masses, too. Evidently, these terms come from quark loops in the functional determinant, and the perturbative expansion of the new improvement coefficients $\bar{b}_S, \bar{b}_A, \bar{b}_P$ shows that they start out at order g_0^4, which means that they come through two-loop effects (one quark loop and a gluon loop which attaches it to the quark line whose renormalization is studied).

Upon considering the difference of two VWI masses and the ratio of two AWI masses

$$m_j^{\rm VWI} - m_k^{\rm VWI} = \frac{1}{Z_S} (m_j^{\rm W} - m_k^{\rm W}) \Big[1 - \frac{1}{2} b_S a (m_j^{\rm W} + m_k^{\rm W}) - \bar{b}_S a {\rm Tr}(M) + O(a^2) \Big] \tag{53}$$

$$\frac{m_j^{\rm AWI}}{m_k^{\rm AWI}} = \frac{m_j^{\rm PCAC}}{m_k^{\rm PCAC}}\Big[1 + (b_A - b_P)a(m_j^{\rm W} - m_k^{\rm W}) + O(a^2)\Big] \tag{54}$$

the term proportional to $a\mathrm{Tr}(M)$ disappears from the second relation, and the term proportional to b_S involves only the sum of the Wilson masses. Applying these formulas to m_s and m_{ud} in $N_f = 2+1$ QCD, with $d \equiv am_s^{\rm bare} - am_{ud}^{\rm bare}$ and $r \equiv m_s^{\rm PCAC}/m_{ud}^{\rm PCAC}$ defined as before, one has

$$am_s^{\rm VWI} - am_{ud}^{\rm VWI} = \frac{1}{Z_S}\, d\,\Big[1 - \frac{1}{2}b_S a(m_s^{\rm W} + m_{ud}^{\rm W}) - \bar{b}_S a(m_s^{\rm W} + 2m_{ud}^{\rm W}) + O(a^2)\Big] \tag{55}$$

$$\frac{m_s^{\rm AWI}}{m_{ud}^{\rm AWI}} = r\,\Big[1 + (b_A - b_P)a(m_s^{\rm W} - m_{ud}^{\rm W}) + O(a^2)\Big] \tag{56}$$

where we have used d and r only in the leading term, so far. The point is that

$$am_s^{\rm W} + am_{ud}^{\rm W} = (am_s^{\rm W} - am_{ud}^{\rm W})\,\frac{m_s^{\rm W}/m_{ud}^{\rm W} + 1}{m_s^{\rm W}/m_{ud}^{\rm W} - 1} \simeq d\,\frac{r+1}{r-1} \tag{57}$$

$$am_s^{\rm W} + 2am_{ud}^{\rm W} = (am_s^{\rm W} - am_{ud}^{\rm W})\,\frac{m_s^{\rm W}/m_{ud}^{\rm W} + 2}{m_s^{\rm W}/m_{ud}^{\rm W} - 1} \simeq d\,\frac{r+2}{r-1} \tag{58}$$

where the approximately equal sign means "up to terms of order $O(a^2)$". Accordingly, we can express the difference of the VWI masses and the ratio of the AWI masses through d and r as

$$am_s^{\rm VWI} - am_{ud}^{\rm VWI} = \frac{1}{Z_S}\, d^{\rm imp} \quad , \qquad \frac{m_s^{\rm AWI}}{m_{ud}^{\rm AWI}} = r^{\rm imp} \tag{59}$$

where $d^{\rm imp}$ and $r^{\rm imp}$ are defined as

$$d^{\rm imp} = d\,\Big[1 - \frac{1}{2}b_S d\frac{r+1}{r-1} - \bar{b}_S d\frac{r+2}{r-1} + O(a^2)\Big] \tag{60}$$

$$r^{\rm imp} = r\,\Big[1 + (b_A - b_P)d + O(a^2)\Big]\,. \tag{61}$$

In total this means that one finds $am_{ud}^{\rm scheme}$ and $am_s^{\rm scheme}$ via (50) with $d \to d^{\rm imp}$, $r \to r^{\rm imp}$.

In our analysis, the tree-level improvement strategy makes all subleading terms in the square brackets of (60, 61) disappear, except for the one proportional to b_S (with $b_S = 1$).

11.3 Determination of the scalar RI/MOM renormalization factor

Having laid out our overall strategy for obtaining the renormalized quark masses $m_{ud}^{\rm phys}(\mu)$, $m_s^{\rm phys}(\mu)$ at the physical mass point in a standard scheme at a given scale μ, we now give details of how we compute the single renormalization factor needed, $Z_S(\mu)$. We implement the nonperturbative Rome-Southampton method which defines the regularization-independent (RI/MOM) scheme [43], with several practical refinements (see below). In the terminology of [40–42] the result is the non-singlet renormalization factor $Z_S^{\rm NS}(\mu)$. In the RI/MOM scheme

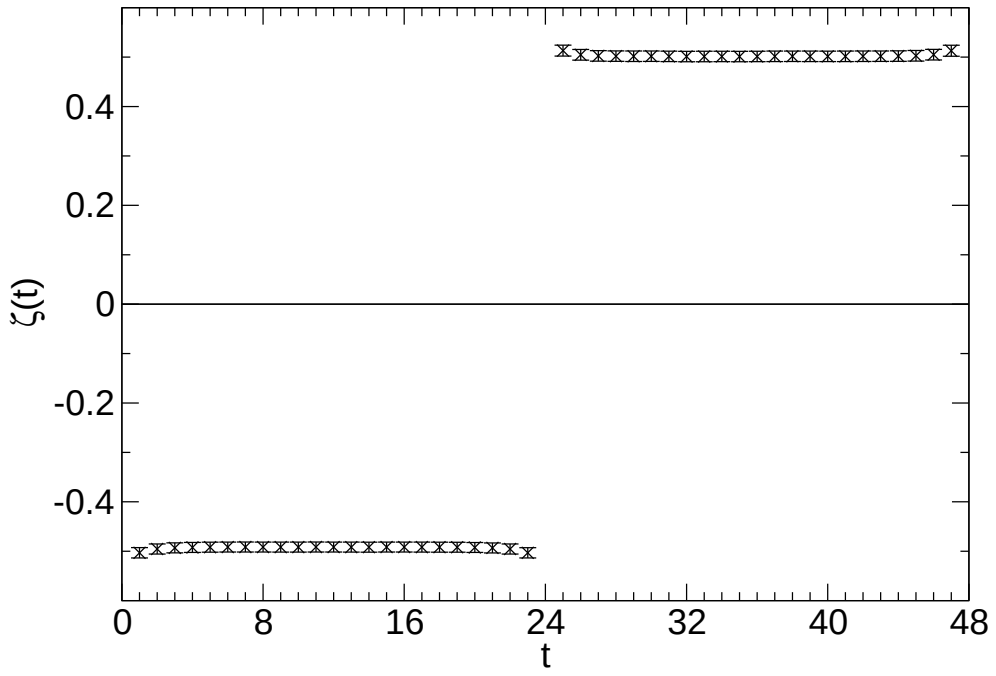

Figure 9: *The ratio $\zeta(t)$ as defined in (66). The gauge coupling in this $N_f = 3$ run is $\beta = 3.61$, the quark mass is $am = -0.0045$. This procedure yields a stable plateau for Z_V.*

the running of $Z_S(\mu)$ is known perturbatively to 4-loop order [44]. However, this is only relevant for the conversion to other schemes, e.g. $\overline{\rm MS}$ at $\mu = 2\,{\rm GeV}$. Our main result, m_{ud} and m_s in the RI/MOM scheme at $\mu = 4\,{\rm GeV}$ is derived without reference to perturbation theory.

In the RI/MOM scheme, renormalization conditions are defined in Landau gauge and require the forward, truncated quark Green's function of an operator to be equal to its tree-level value at a Euclidean four-momentum p, whose magnitude is chosen to be the renormalization scale. Given a quark bilinear operator $O^\Gamma_{12} = \bar\psi_2 \Gamma \psi_1$, where ψ_1 and ψ_2 are mass-degenerate quark fields and Γ is a Dirac matrix, the relevant Green's function is

$$\Lambda_\Gamma(p) \equiv \langle S(p)\rangle^{-1} \left\{ \int d^4x d^4y\, e^{ip(x-y)}\, \langle \psi_2(y) O^\Gamma_{12}(0) \bar\psi_1(x)\rangle \right\} \langle S(p)\rangle^{-1} \,. \tag{62}$$

In this equation, $S(p)$ is the propagator of quark flavors 1 and 2. Now, defining a projector P_Γ such that ${\rm tr}\{P_\Gamma \Gamma\} = 1$ (the trace is over spin$\times$color), the renormalization condition reads

$$Z_\Gamma(\mu) = Z_\psi(\mu) \,/\, \Gamma_\Gamma(p)|_{p^2=\mu^2} \tag{63}$$

where Z_ψ is the wavefunction renormalization factor and

$$\Gamma_\Gamma(p) \equiv {\rm tr}\{P_\Gamma \Lambda_\Gamma(p)\} \,. \tag{64}$$

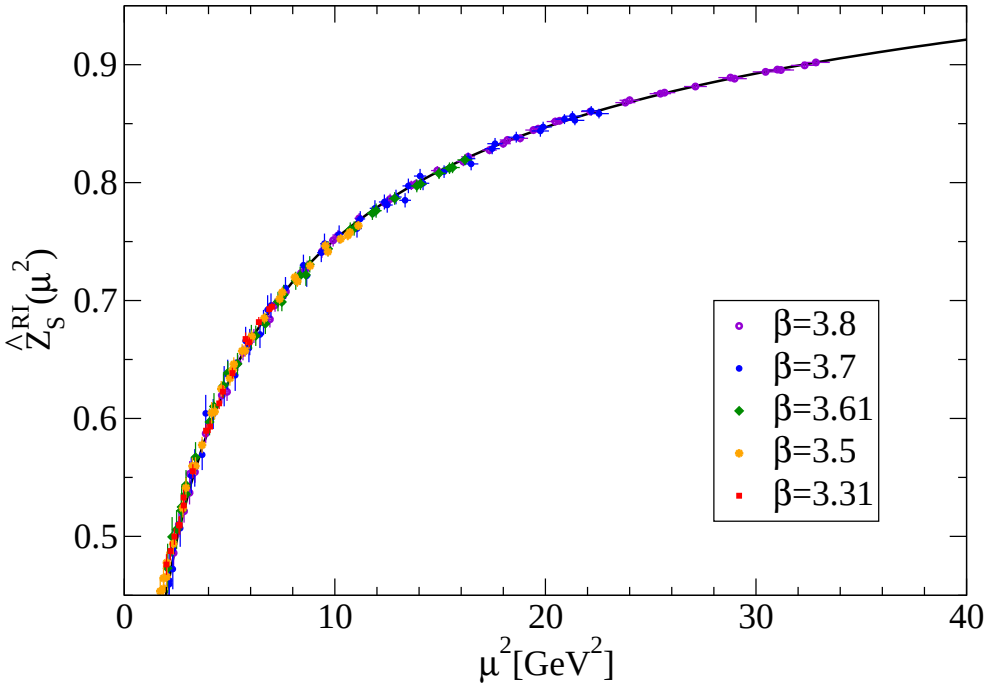

Figure 10: *Renormalization factors $Z^{\rm RI}_{S,\beta}(\mu^2)$ as a function of the bosonic momentum squared. For each β momenta $\mu \leq \pi/(2a)$ are included. The data from the coarser lattices have been multiplied by a μ-independent factor to match those at $\beta = 3.8$. The solid line represents a Pade ansatz where the 1-loop anomalous dimension is built in as a constraint.*

To determine Z_S from the RI/MOM condition (63) with $\Gamma = I$, one needs to know Z_ψ. In the original publication [43] the procedure was supplemented with a recipe to obtain Z_ψ from the momentum dependence of the trace of the inverse propagator. Here we avoid the determination of this wave-function renormalization constant all together, by calculating the ratio $Z_S(\mu)/Z_V$ via the RI/MOM procedure and by combining it with Z_V from the 3-point function with a vector-current insertion. In other words, on each ensemble we compute $Z_S(\mu)/Z_V$ using

$$\frac{Z_{S,\beta,m}(\mu)}{Z_{V,\beta,m}} = \frac{\Gamma_V(p)}{\Gamma_S(p)}\Big|_{\hat{p}^2=\mu^2} \tag{65}$$

where the dependence on the coupling and the $N_f = 3$ quark mass is indicated with subscripts. The bosonic momentum definition $\hat{p}_\nu = (2/a)\sin(ap_\nu/2)$ is used, and a standard cylinder cut around hyperdiagonal momenta is applied [45]. In addition, we determine Z_V from the ratio

$$\zeta(t) \equiv \frac{\langle P(T/2)V_4(t)\bar{P}(0)\rangle}{\langle P(T/2)\bar{P}(0)\rangle} \tag{66}$$

where

$$P(t) = \sum_{\vec{x}}(\bar{\psi}_2\gamma_5\psi_1)(\vec{x},t)\ , \quad \bar{P}(t) = \sum_{\vec{x}}(\bar{\psi}_1\gamma_5\psi_2)(\vec{x},t)\ , \quad V_4(t) = \sum_{\vec{x}}(\bar{\psi}_1\gamma_4\psi_1)(\vec{x},t) \tag{67}$$

3.31	$16^3\times 32$	3.5	$24^3\times 48$	3.61	$24^3\times 48$	3.7	$32^3\times 64$	3.8	$32^3\times 64$
-0.04	4780	-0.006	2560	-0.0045	4620	-0.0060	1010	0.000	505
-0.06	3320	-0.010	3140	-0.0085	3680	-0.0085	1050	-0.004	635
-0.07	2420	-0.012	2580	-0.0100	4140	-0.0110	1020	-0.008	500
-0.08	2500	-0.020	2700	-0.0200	3140	-0.0140	1290	-0.012	1030
		-0.035	1090	-0.0250	1230	-0.0160	1020	-0.014	1000

Table 2: *Bare masses and number of trajectories of our dedicated $N_f = 3$ simulations for RI/MOM renormalization. The β-values are the same as in our phenomenological runs, cf. Tab. 1.*

and T denotes the temporal extent of the lattice. With tree-level improvement one has [42, 46]

$$Z_{V,\beta,m}\,(1 + am^{\mathrm{W}}) = [\zeta(t_1) - \zeta(t_2)]^{-1} \quad \text{for} \quad 0<t_1< T/2<t_2<T \tag{68}$$

where $b_V = 1, \bar{b}_V = 0$ have been used, and Fig. 9 shows the plateau from which we extract $Z_{V,\beta,m}$. Combining this factor with the result of (65) yields $Z_{S,\beta,m}(\mu)$, much in the spirit of [47, 48].

11.4 Controlling the systematics in the RI/MOM procedure

RI/MOM is a mass independent scheme. Applied to the numerical data for $Z_S^{\mathrm{RI}}(\mu)$ this means that we have to extrapolate all three flavors to vanishing sea and valence quark mass. For this reason, we have generated a series of dedicated $N_f = 3$ lattices (i.e. with three degenerate quarks), where the action $S = S_g^{\mathrm{Sym}} + S_f^{\mathrm{SW}}$ and the couplings $\beta = 6/g^2$ are the same as in the phenomenological ensembles. The bare parameters and statistics of these runs are summarized in Tab. 2. The specifics of the extrapolation will be discussed below.

In order to obtain tree-level $O(a)$-improved results with Wilson fermions, one has to improve not only the action, but also the interpolating fields. For standard correlators this has been discussed in the previous two subsections. In addition, in the RI/MOM procedure, one has to remove an $O(a)$ contact term in the quark propagator [43]. We apply here the trace subtraction described in [49–52], which has the added benefit of greatly improving the signal to noise ratio. This subtraction is implemented by replacing the condition (63) by one in which the modified propagator $\bar{S}(p) = S(p) - \mathrm{Tr}(S(p))/4$ is used to define the amputated Green's function, where the trace is in spinor space.

In order to reliably extract the renormalization constants and to convert the resulting quark masses $m^{\mathrm{RI}}(\mu)$ to other schemes without loosing precision, several conditions should be met:

- (a) the scale μ at which we take the continuum limit of the RI/MOM renormalized masses needs to be substantially below the momentum cutoff of the coarsest lattice $\mu \ll 2\pi/a$,
- (b) the conversion to a perturbative scheme has to be done at a scale μ' which is sufficiently large, such that perturbation theory is reliable, i.e. at $\mu' \gg \Lambda_{\mathrm{QCD}}$,
- (c) the effect of the chiral extrapolation $m \to 0$ needs to be fully controlled.

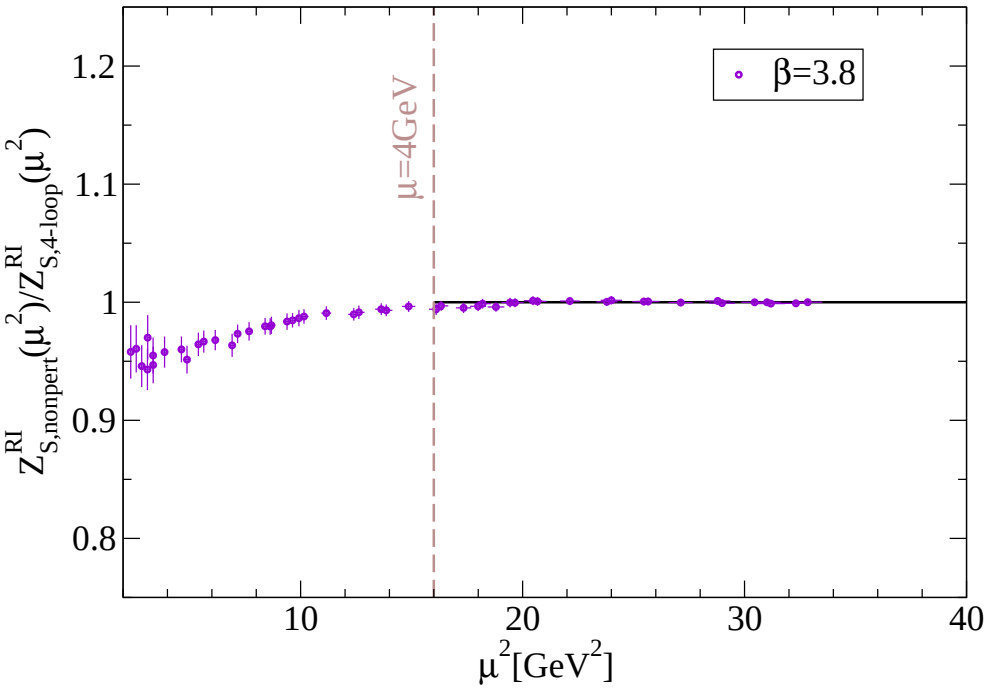

Figure 11: *Ratio of the nonperturbative $Z_S^{\rm RI}$ to the perturbative prediction at 4-loop level. The momentum range shown extends to $\mu_{\rm max}=\pi/(2a)$ at $\beta=3.8$. For $\mu\geq 4\,{\rm GeV}$ the data agree with the plateau within errors.*

The difficulty to fulfill, in one simulation, the first two conditions is sometimes referred to as the "window problem" of the RI/MOM procedure. In the following we show how we can simultaneously satisfy all three requirements.

Ad (a): To renormalize our quark masses and to extrapolate them to the continuum we choose a convenient renormalization scale $\mu=2.1\,{\rm GeV}$. This scale satisfies $\mu<\pi/(2a)$ for all our lattices (on the coarsest one this figure is about $2.7\,{\rm GeV}$). When plotting the running of $Z_{S,\beta}(\mu)$ at different β on top of each other (see Fig. 10), one finds that discretization effects are below our statistical accuracy in this region, and that the form of the running is almost identical for our five β values.

Ad (b): With the procedure described above and by taking the continuum limit we obtain a fully nonperturbative determination of the quark masses in the RI/MOM scheme at $\mu=2.1\,{\rm GeV}$. In principle, we could stop here, quoting this as our main result. However, if one wants to convert this result to another scale or another scheme, it is evident from Fig. 12 that doing so perturbatively would introduce an uncertainty in the $1-2\%$ range. Therefore we use our renormalization data to run our quark mass results, nonperturbatively, to the scale $\mu'=4\,{\rm GeV}$, where this perturbative uncertainty is in the 0.5% range and hence subdominant. At $\mu'=4\,{\rm GeV}$ we still have 3 different β values which satisfy the condition $\mu'<\pi/(2a)$. More specifically, we use our data to extrapolate the ratio $Z^{\rm RI}_{S,\beta}(\mu)/Z^{\rm RI}_{S,\beta}(\mu')$ to the continuum, with an extremely

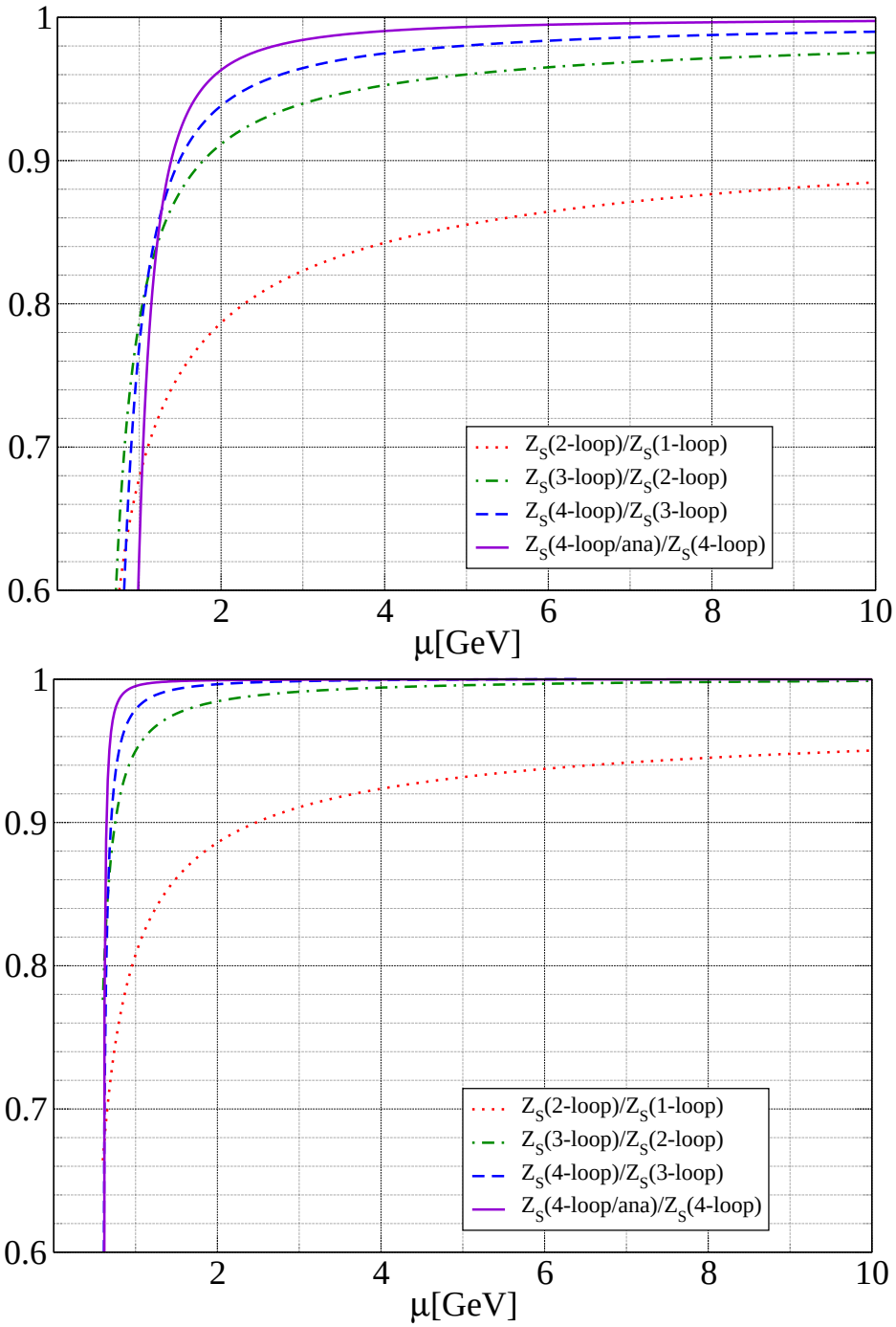

Figure 12: *Ratio of the perturbatively evaluated $Z_S^{\rm RI}(\mu)$ (top panel) and $Z_S^{\overline{\rm MS}}(\mu)$ (bottom panel) at different loop orders. The renormalization group equations have been numerically integrated, using 1-loop through 4-loop anomalous dimensions. To estimate the remaining uncertainty in the 4-loop running, we employ the analytic expression at 4-loop level [44], which differs from the numerically integrated one by 5-loop effects. In the labels this is called "4-loop/ana".*

mild effect (as one can see from Fig. 10, in this interval the three curves lie essentially on top of each other), and the resulting ratio provides us with the nonperturbative running of the scalar renormalization constant, in the continuum, between $\mu = 2.1\,\mathrm{GeV}$ and $\mu' = 4\,\mathrm{GeV}$. Accordingly

$$R_S^{\mathrm{RI}}(\mu, 4\,\mathrm{GeV}) = \lim_{\beta\to\infty} \frac{Z_{S,\beta}^{\mathrm{RI}}(4\,\mathrm{GeV})}{Z_{S,\beta}^{\mathrm{RI}}(\mu)} \tag{69}$$

is the continuum extrapolated ratio which allows us to evolve data from *all* our lattices, including the coarser ones, to $\mu' = 4\,\mathrm{GeV}$, where we perform the final continuum extrapolation. Through this procedure we obtain fully nonperturbatively renormalized quark masses in the RI/MOM scheme at $\mu' = 4\,\mathrm{GeV}$, which represent our main result. For the reader's convenience we also convert them to other schemes. To this end we use 4-loop perturbative running to convert to the RGI framework (where we use the conventions of [53] with b_0, d_0 adjusted to $N_f = 3$), and subsequently to the $\overline{\mathrm{MS}}$ scheme (which is perturbatively defined).

Ad (c): As mentioned above, the RI/MOM scheme is a massless renormalization scheme. Since the dedicated $N_f = 3$ simulations as listed in Tab. 2 use finite quark masses (roughly in the range $m_s^{\mathrm{phys}}/3 < m < m_s^{\mathrm{phys}}$), we have to perform a chiral extrapolation at some point. In the procedure described in the previous paragraph, the numerical data for $Z_{S,\beta,m}^{\mathrm{RI}}(\mu)$ were first extrapolated to the chiral limit to give $Z_{S,\beta}^{\mathrm{RI}}(\mu)$. Based on this the renormalized quark masses $m_\beta^{\mathrm{RI}}(\mu)$ and the ratios $Z_{S,\beta}(\mu')/Z_{S,\beta}(\mu)$ were extrapolated to the continuum, as detailed in (a) and (b), respectively. To give a reliable estimate of the systematic uncertainties involved, we supplement this procedure with a second one where we interchange the order of limits. Technically, this means that one defines an intermediate MOM scheme, which is not a massless one, but instead based on a fixed reference quark mass. We use $m_{\mathrm{ref}}^{\mathrm{RGI}} = 70\,\mathrm{MeV}$, since, for all β, this value can be reached by interpolation. In this scheme the renormalized light and strange quark masses are determined at the scales $\mu \in \{1.3, 2.1\}\,\mathrm{GeV}$, and extrapolated to the continuum. This yields $m_{ud,m_{\mathrm{ref}}}^{\mathrm{MOM}}(\mu)$ and ditto for m_s. Staying in this massive scheme, these quark masses are evolved to the scale $\mu' = 4\,\mathrm{GeV}$. In this step a fully controlled continuum extrapolation can be performed, since we have three lattice spacings satisfying $\mu' < \pi/(2a)$. At this point we have the renormalized quark mass in the form

$$m_{ud,m_{\mathrm{ref}}}^{\mathrm{MOM}}(\mu') = m_{ud,m_{\mathrm{ref}}}^{\mathrm{MOM}}(\mu) \cdot \frac{Z_{S,m_{\mathrm{ref}}}(\mu)}{Z_{S,m_{\mathrm{ref}}}(\mu')} \tag{70}$$

where either factor has been extrapolated to the continuum. In the last step, we switch from the intermediate massive MOM scheme to the massless RI/MOM scheme by multiplying (70) with the continuum extrapolated ratio $Z_{S,m_{\mathrm{ref}}}(\mu')/Z_S(\mu')$. This yields the same $m_{ud}^{\mathrm{RI}}(\mu'), m_s^{\mathrm{RI}}(\mu')$ as before, except that the order of limits has been interchanged. Note that all continuum extrapolations are entirely flat and the effect of the mass extrapolation is about 1%, implying that all limiting procedures are fully controlled. Having obtained our main result, the RI/MOM masses at $\mu' = 4\,\mathrm{GeV}$, we can transform them to other schemes as described under (b).

β	$L^3 \times T$	$(m_s + m_{ud}) r_0$
5.7366	$12^3 \times 24$	0.3070(50)
5.8726	$16^3 \times 32$	0.2801(50)
5.9956	$20^3 \times 40$	0.2758(52)
6.1068	$24^3 \times 48$	0.2654(42)
6.3000	$32^3 \times 64$	0.2685(29)

Table 3: *Details of the quenched overall test. The quark masses are in the* $\overline{\mathrm{MS}}$ *scheme at* $2\,\mathrm{GeV}$.

11.5 Summary of RI/MOM renormalization

Let us summarize this section. We compute the quark masses m_{ud}^{phys} and m_s^{phys} through the "ratio-difference method" in the RI/MOM scheme at the scale $\mu' = 4\,\mathrm{GeV}$, nonperturbatively and with extrapolation to the continuum.

The mild quark mass dependence of the renormalization factors is eliminated through a chiral extrapolation. Also cut-off effects are removed through a continuum extrapolation. In this step we are extremely conservative – we do not only consider the formally leading cut-off effects $O(\alpha a)$, but also subleading effects proportional to $O(a^2)$, counting the spread towards the final systematic error (see Sec. 14). We think this is necessary, since even with a set of 5 lattice spacings, we cannot exclude the possibility that the subleading $O(a^2)$ cut-off effects largely affect the continuum extrapolation. If we were to consider only the leading $O(\alpha a)$ cut-off effects, our systematic error would be significantly smaller.

The quark masses in the RI/MOM scheme at the scale $\mu' = 4\,\mathrm{GeV}$ are our main result, obtained in a way which guarantees that they are truly nonperturbative. Using perturbation theory in a regime where it is well behaved, we convert them to the universal RGI prescription and subsequently to the perturbatively defined $\overline{\mathrm{MS}}$ scheme at the scale $\mu = 2\,\mathrm{GeV}$.

12 Quenched overall check

To demonstrate that our 2-step HEX smeared clover action (1) and the nonperturbative renormalization of the quark mass yield reliable results in the continuum, we repeat the quenched benchmark calculation [53] of the quantity $m_s + m_{ud}$, using our setup.

We use pure Wilson glue at five couplings between $\beta = 5.7366$ and $\beta = 6.3$, each time saving about 600 well-decorrelated configurations for the analysis (i.e. 200 for the Z-factors and 400 for the masses). The couplings and geometries have been chosen to realize a fixed physical box size of about $L = 1.84\,\mathrm{fm}$, see Tab. 3 for details. On each set at least 4 quark masses are used to safely interpolate to the point $M_P\, r_0 = 1.229$, where M_P is the pseudoscalar meson mass and where the numerical value has been chosen to match $M_K^{\mathrm{phys}} r_0$ with $r_0 = 0.49\,\mathrm{fm}$ [54].

The computation closely follows the dynamical case. We renormalize the VWI quark mass sum with the methods described in the previous section, and we use the same procedure to convert to the $\overline{\mathrm{MS}}$ scheme. In more detail, we begin with measuring m^{PCAC} as a function of

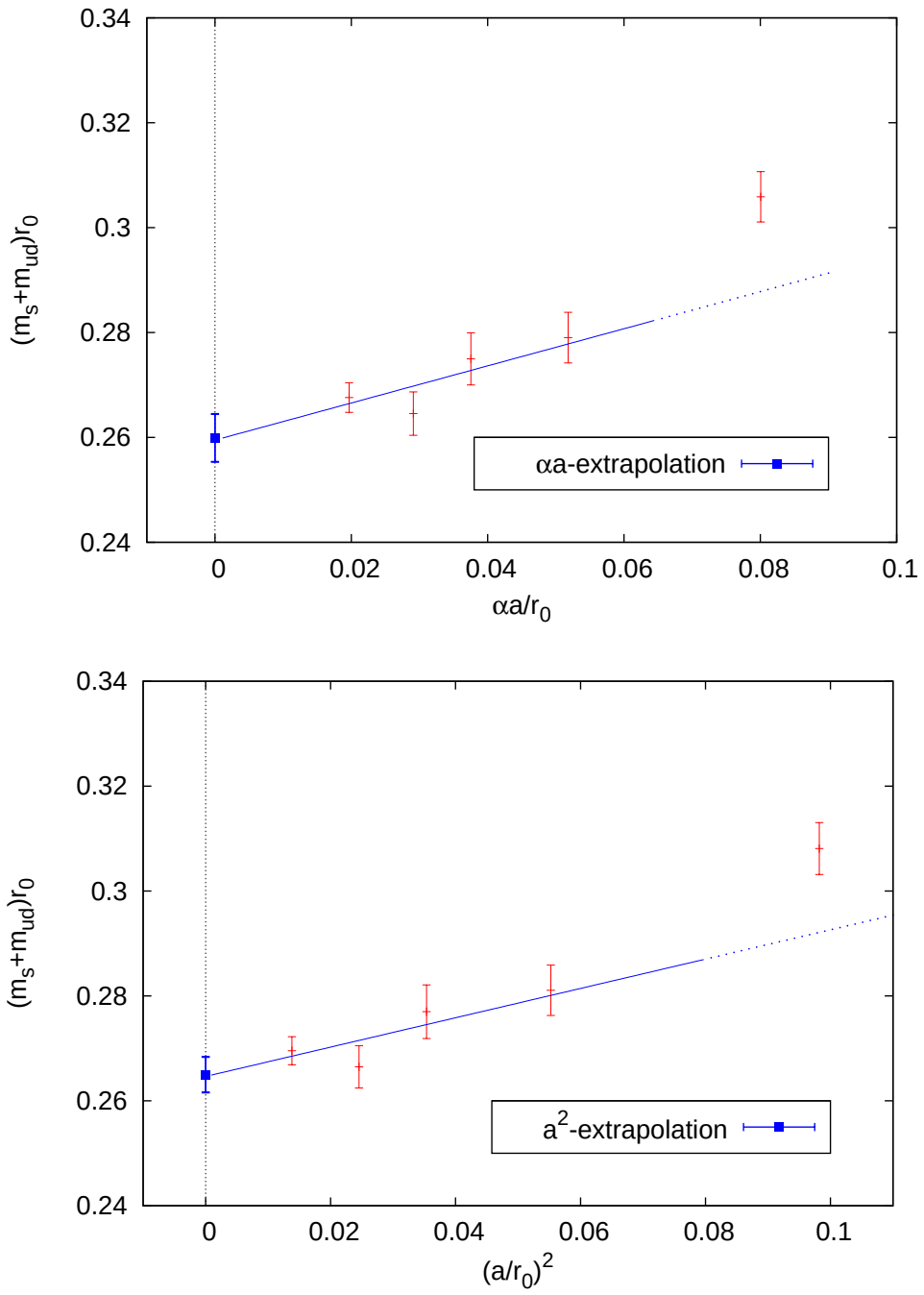

Figure 13: *Quenched continuum extrapolation of* $(m_s+m_{ud})r_0$ *in the* $\overline{\rm MS}$ *scheme at* $\mu = 2\,\mathrm{GeV}$*, assuming* $O(\alpha a)$ *[top] or* $O(a^2)$ *[bottom] scaling. One data-point outside the scaling regime (*$\beta = 5.7366$*) is shown. The difference counts towards the systematic error (see text for details).*

the bare mass $m^{\rm bare}$, which shows a linear relationship. The intercept with the x-axis yields $m^{\rm crit}$ and thus $m^{\rm W}$ as defined in (47). Next, we determine $Z_S(\mu)$ via RI/MOM [43] to obtain $m^{\rm VWI}$ according to (47). It is easy to see that this is the flavor non-singlet $Z_S(\mu)$, since all quark disconnected contributions vanish in the quenched theory. In close analogy with our phenomenological analysis, we choose the matching scales $\mu = 2.1\,\mathrm{GeV}$ and $\mu' = 3.5\,\mathrm{GeV}$. Combining the continuum extrapolated ratio $Z_S(\mu')/Z_S(\mu)$ with $Z_{S,\beta}(\mu)$, we obtain $Z_{S,\beta}(\mu')$ and the renormalized mass $m^{\rm VWI}(\mu')$ in the RI/MOM scheme at the scale $\mu' = 3.5\,\mathrm{GeV}$. Finally, we use perturbation theory to convert to the $\overline{\rm MS}$ scheme at $2\,\mathrm{GeV}$ scale. The result is identified with $m_s + m_{ud}$ in this scheme, at the given lattice spacing, and multiplied with r_0 to obtain a dimensionless quantity (cf. Tab. 3). We find that we can extrapolate these values linearly in αa or a^2, with the data showing a slight preference for the latter option, as can be seen from the two panels in Fig. 13. Using the machinery for propagating both statistical and systematic errors that will be described in Sec. 14, the combined result in the continuum reads $(m_s + m_{ud})r_0 = 0.2609(39)(28)$ in the $\overline{\rm MS}$ scheme at $\mu = 2\,\mathrm{GeV}$.

Our result is in perfect agreement with the continuum value $(m_s + m_{ud})r_0 = 0.261(9)$ quoted by the ALPHA collaboration [53]. It is consistent, within less than 1σ, with the result $0.274(18)$ given by JLQCD [55] and, within less than 2σ, with the value $0.312(28)$ obtained in a computation with quenched overlap fermions that includes a continuum extrapolation [56]. There is some tension with the result $0.293(6)$ by CP-PACS [57], but one should keep in mind that the systematic uncertainty due to the perturbative renormalization is not included in their error. In short, we find good agreement with the most precise results in the literature. We take this as evidence that the renormalization procedure described in Sec. 11 yields reliable results.

13 Using dispersive input to obtain m_u and m_d

For decades the most reliable source of information on light quark masses has been current algebra, in particular in its modern form, known as Chiral Perturbation Theory (ChPT). A major drawback of this framework is that only information on quark mass ratios can be extracted, not on absolute values. This is a consequence of the fact that in the chiral Lagrangian all quark masses appear in the combination $B_0 m_q$ and the condensate parameter B_0 does not occur in any other instance. We have determined $m_{ud} \equiv (m_u + m_d)/2$ and m_s in MeV units. Accordingly, by comparing our value of the ratio m_s/m_{ud} to theirs, we can learn something about the convergence pattern of SU(3) ChPT. Furthermore, one may combine our values for m_{ud} and m_s with the best available information on another ratio (Q, see below) to obtain a result for the individual quark masses m_u, m_d.

13.1 Comparing our value of m_s/m_{ud} to the one in ChPT

As a starting point one might ignore higher-order terms in the chiral expansion and electromagnetic corrections all together. Upon identifying the left-hand sides in

$$M_\pi^2 = B_0(m_u + m_d) \quad (71)$$
$$M_{K^\pm}^2 = B_0(m_u + m_s) \quad (72)$$
$$M_{K^0}^2 = B_0(m_d + m_s) \quad (73)$$
$$M_\eta^2 = B_0(m_u + m_d + 4m_s)/3 \quad (74)$$

with the experimentally measured meson masses[3], one obtains three predictions. On the one hand, the Gell-Mann–Okubo relation

$$3M_\eta^2 + M_\pi^2 \simeq 2M_{K^\pm}^2 + 2M_{K^0}^2 \quad (75)$$

evaluates to $0.919\,\mathrm{GeV}^2 \simeq 0.983\,\mathrm{GeV}^2$, which amounts to a 7% accuracy. On the other hand

$$(M_{K^\pm}^2 + M_{K^0}^2)/M_\pi^2 = (m_s + m_{ud})/(m_{ud}) \quad (76)$$
$$M_\eta^2/M_\pi^2 = (2m_s + m_{ud})/(3m_{ud}) \quad (77)$$

yield $m_s/m_{ud} \simeq 25.1$ and $m_s/m_{ud} \simeq 23.4$, respectively. This spread suggests again a precision of a few percent. Upon noticing that the η undergoes significant mixing with the η' and, as a result, that (76) should be preferred over (77), one arrives at the estimates

$$\frac{m_u}{m_d} = \frac{M_{K^\pm}^2 - M_{K^0}^2 + M_\pi^2}{M_{K^0}^2 - M_{K^\pm}^2 + M_\pi^2} \simeq 0.66 \quad (78)$$
$$\frac{m_s}{m_d} = \frac{M_{K^\pm}^2 + M_{K^0}^2 - M_\pi^2}{M_{K^0}^2 - M_{K^\pm}^2 + M_\pi^2} \simeq 20.8 \quad (79)$$

which do not take into account electromagnetic contributions to isospin breaking.

The chiral framework may be extended to include interactions with photons. At leading order in α_{em} and in the 3-flavor chiral limit the electromagnetic contribution to the excess of the charged kaon mass squared is the same as for the pion, i.e. $[M_{\pi^\pm}^2 - M_{\pi^0}^2]_{\mathrm{em}} = [M_{K^\pm}^2 - M_{K^0}^2]_{\mathrm{em}}$, known as "Dashen's theorem" [58]. This leads to the improved relations[4] [59]

$$\frac{m_u}{m_d} = \frac{M_{K^\pm}^2 - M_{K^0}^2 + 2M_{\pi^0}^2 - M_{\pi^\pm}^2}{M_{K^0}^2 - M_{K^\pm}^2 + M_{\pi^\pm}^2} \simeq 0.56 \quad (80)$$
$$\frac{m_s}{m_d} = \frac{M_{K^\pm}^2 + M_{K^0}^2 - M_{\pi^\pm}^2}{M_{K^0}^2 - M_{K^\pm}^2 + M_{\pi^\pm}^2} \simeq 20.2 \quad (81)$$

which account for electromagnetism at leading order (LO) in the chiral expansion. From this one obtains $m_s/m_{ud} = 2/(m_d/m_s + m_u/m_d \cdot m_d/m_s) \simeq 25.9$ as the LO result in ChPT. Comparing this to our value (85) [see below] indicates that – for this quantity – subleading contributions yield only about 6% of the total result.

[3] Pseudoscalars without superscript refer to isospin averages: $M_\pi^2 = \frac{1}{2}(M_{\pi^\pm}^2 + M_{\pi^0}^2)$, $M_K^2 = \frac{1}{2}(M_{K^\pm}^2 + M_{K^0}^2)$.

[4] The numerical values are based on the latest edition of the PDG [20] and differ from those given in [59].

13.2 Using dispersive information on Q to split m_{ud} into m_u and m_d

As mentioned in the previous subsection, ChPT is well suited to address the ratios m_s/m_d and m_u/m_d. A way to encode such information on quark mass ratios which, from the ChPT viewpoint, is particularly robust is to introduce the double ratio

$$Q^2 \equiv \frac{m_s^2 - m_{ud}^2}{m_d^2 - m_u^2} \tag{82}$$

since this quantity is unaffected by next-to-leading order (NLO) effects in the chiral expansion. Modulo a tiny correction, (82) can be put into a form known as "Leutwyler's ellipse" [59]

$$\frac{1}{Q^2}\Big(\frac{m_s}{m_d}\Big)^2 + \Big(\frac{m_u}{m_d}\Big)^2 = 1 \tag{83}$$

and relying on Dashen's theorem [58] or refinements thereof (see e.g. [21]), one might attempt to determine the value of Q from the masses of the charged and neutral kaon and pion.

Since we intend to use (82) to predict the isospin splittings in QCD (i.e. without electromagnetism), it seems more advisable to build on the long tradition in the phenomenological literature to determine Q from the rate for $\eta \to 3\pi$ decays or from the branching ratio of $\psi' \to \psi\pi^0$ versus $\psi' \to \psi\eta$ decays. The former amplitude seems particularly interesting, as it violates isospin, while being barely affected by electromagnetic corrections [60]. Evidently, this renders it sensitive to the effect of $m_d - m_u \neq 0$, which is exactly what we are interested in. In the following, we restrict ourselves to the dispersive treatment of the $\eta \to 3\pi$ amplitude, as given by Kambor-Wiesendanger-Wyler [61], Anisovich-Leutwyler [62], and Colangelo-Lanz-Passemar [63]. In the first place we note that the central value found in these works has been remarkably consistent over one and a half decades. Let us also emphasize that a dispersive treatment is, conceptually, as much from first principles as a lattice computation – dispersion theory rests exclusively on the axioms of quantum field theory. In a world with perfect experimental data, this would be the complete story. However, with presently available data, additional input is required (see e.g. [63]). To account for such provisional effects, Leutwyler has assigned his estimate $Q = 22.3(8)$ [64] a much larger error bar than claimed in some of the publications it is based on. In our view this is the most accurate value available, if one is not willing to resort to model calculations, and we shall thus stay content with its rather conservative error bar.

We now extend our lattice determinations of m_{ud} and m_s/m_{ud} to all three quark masses, using this dispersive information. Upon rewriting (82, 83) in the form

$$\frac{1}{Q^2} = 4\,\Big(\frac{m_{ud}}{m_s}\Big)^2\,\frac{m_d - m_u}{m_d + m_u} \tag{84}$$

it follows that the above-mentioned value of Q and our lattice result

$$\frac{m_s}{m_{ud}} = 27.53(20)(08) \tag{85}$$

yield the light quark mass asymmetry parameter

$$\frac{m_d - m_u}{m_d + m_u} = 0.381(05)(27) \tag{86}$$

where the error on Q is considered a systematic error. As an aside we mention that this asymmetry parameter is equivalent to $m_u/m_d = 0.448(06)(29)$. Combining (86) with our result $m_{ud} = 3.503(48)(49)\,\mathrm{MeV}$, we obtain

$$m_u = 2.17(04)(10)\,\mathrm{MeV}, \quad m_d = 4.84(07)(12)\,\mathrm{MeV} \tag{87}$$

with all masses given in the RI/MOM scheme at the scale $\mu = 4\,\mathrm{GeV}$. These values and our original results for m_s and m_{ud} (along with their counterparts in the RGI and $\overline{\mathrm{MS}}$ schemes) are summarized in Tab. 5 (see Sec. 15) and quoted in [1].

To summarize the technical part, let us say that we have determined m_u and m_d, based on our lattice value of m_{ud}, our lattice value of the ratio m_s/m_{ud} and the dispersive treatment of Q. Given that our simulation points bridge the physical values of m_{ud} and m_s (cf. Sec. 5), the chiral framework is no longer needed in the first two quantities, and the use of ChPT is thus limited to a subdominant contribution in a mostly dispersive framework to determine Q.

13.3 Physics implication, robustness issues and precision outlook

Physicswise, an important conclusion is that our result (86) for the light quark mass asymmetry parameter excludes a vanishing up-quark mass by 22.1 standard deviations. This is a consequence of the dispersive determination of Q being entirely inconsistent with 13.8, the value of Q which relation (84) and our result for m_s/m_{ud} would enforce if $m_u = 0$. As can be seen from (84), the asymmetry parameter depends strongly on the ratio m_s/m_{ud}, which is the quantity that we have determined to sub-percent precision. The bottom line is that our precise lattice results and the dispersive processing of phenomenological information which excludes very large corrections to Dashen's theorem, when combined, rule out the simplest proposed solution to the so-called "strong CP problem". This corroborates previous findings [59].

Note that the way in which we have used phenomenological information is designed to make sure that the so-called "Kaplan-Manohar ambiguity" is circumvented in our derivation of m_u and m_d. This ambiguity expresses the fact that a redefinition of the quark condensate and of certain low-energy constants allows one to move on Leutwyler's ellipse [65]. It represents an accidental symmetry of those Green's functions in the effective theory which determine pseudoscalar masses, scattering amplitudes and matrix elements of the vector and axial-vector currents [59]. However, the *aspect ratio* of Leutwyler's ellipse is not affected by this ambiguity, and it is this shape information[5] which is encoded in Q. In consequence, relation (84) ensures

[5] We remark that Q as defined in (82) picks up, under a Kaplan-Manohar transformation, terms of order NNLO and a change proportional to $m_d - m_u$. The latter "deficiency" could be cured by defining $Q_1^2 = (m_s^2 - m_d^2)/(m_d^2 - m_u^2)$ [66]. Note, however, that the numerical difference between Q and Q_1 [or Q_2, the quantity that shows up in (84)] is about one permil, i.e. more than an order of magnitude smaller than the uncertainty that we have assigned to Q.

that the high precision that we have reached in m_s/m_{ud}, together with the robust value of Q that we use, leads to a determination of the asymmetry parameter (86) and thus of the individual u and d quark masses which is unaffected by the Kaplan-Manohar ambiguity.

We stress that, in our view, there is not much conceptual difference between using only M_π, M_K as input quantities versus including Q, too. To compute m_{ud}, m_s, we needed two (polished) experimental input numbers to adjust the average light and the strange quark masses, apart from M_Ω to set the overall scale (cf. Sec. 4). To compute m_u, m_d, m_s, evidently, we need a third one, and we are well advised to choose one which is sensitive to the effect we want to quantify. We select Q for its large sensitivity to QCD-induced isospin breaking, thus requiring very little theoretical polishing, and for this little bit resting on dispersion theory which is well founded. Still, there is room for improvement, as can be seen from the fact that our value of m_{ud} had 2% precision, while m_u and m_d have only 5% and 3% accuracy, respectively. The problem is that the current value of Q determines the asymmetry parameter (86) to only about 7% precision. While improvements on the value of Q obtained in this way may be possible [63], reaching accuracies of m_u, m_d below the few percent level will most probably require a different approach, even more heavily based on lattice field theory. Indeed, once simulations become available with $N_f = 1+1+1$ physical quark flavors (i.e. with non-degenerate up, down, and strange quark masses, each of which is taken at its physical value) and with an additional abelian gauge field[6] to account for electromagnetic interactions, it will become possible to take full advantage of the very accurately known K^+ and K^0 masses to determine m_u and m_d with even higher precision.

14 Assessment of systematic errors

Our approach is to establish one global fit to interpolate our $11+12+9+9+6 = 47$ simulations at 5 different lattice spacings (cf. Tab. 1) to the physical mass point (i.e. physical M_π and M_K) and to extrapolate to zero lattice spacing (i.e. $a \to 0$). In order to obtain a reliable estimate of the systematic error involved, we repeat the entire analysis with a large selection of interpolation formulae, mass cuts, discretization terms, fit ranges, and renormalization procedures.

In order to extrapolate or interpolate a given quantity to the physical quark mass point, one needs to expand it around some pion and kaon mass point. Often the $N_f = 2$ or $N_f = 3$ chiral limit is chosen as an expansion point and hence SU(2) or SU(3) ChPT [37, 38] as the theoretical framework. Expressing the dependence on the light quark mass as a dependence on M_π, this kind of expansion leads, for a quantity which vanishes in the chiral limit, to a quadratic term $\propto M_\pi^2$ and higher order chiral logs, e.g. $\propto M_\pi^4 \log(M_\pi^2/\Lambda^2)$, with known prefactors but unknown scale Λ. In many cases the practical usefulness of knowing the prefactors is limited, since they contain other quantities (e.g. the axial coupling g_A for octet baryons) which may not be available from the same simulation and which one may not want to borrow from phenomenology. Furthermore, it is rather difficult for a fit to tell, e.g., a pure M_π^4 contribution from

[6] For recent progress in this field see e.g. [67].

an $M_\pi^4 \log(M_\pi^2/\Lambda^2)$ chiral log. Accordingly, choosing an expansion point for an interpolation somewhere in the middle of the region where one has data (or in the middle of the region defined by the data points and the target point in case of an extrapolation) and using a simple Taylor expansion in M_π^2 leads to rather similar results [2].

To flesh out the meaning of these statements, let us consider the quantities of interest, m_{ud} and m_s. In our analysis we use the NLO mass formulae (45) from SU(2) ChPT [37], albeit in reversed form, so that it expresses m_{ud} as a function of M_π. To the order we are working at, this can be done in several ways [the difference is an NNLO effect]; we use the relations

$$m_{ud} = \frac{M_\pi^2}{2B} \cdot \Big\{1 - \frac{1}{2}\frac{M_\pi^2}{(4\pi F_\pi)^2} \log(\frac{M_\pi^2}{\Lambda_3^2})\Big\} \cdot \Big(1 + c_s \Delta\Big) \tag{88}$$

$$m_{ud} = \frac{M_\pi^2}{2B} \Big/ \Big\{1 + \frac{1}{2}\frac{M_\pi^2}{(4\pi F_\pi)^2} \log(\frac{M_\pi^2}{\Lambda_3^2})\Big\} \cdot \Big(1 + c_s \Delta\Big) \tag{89}$$

where we have introduced a hadronic quantity

$$\Delta = 2M_K^2 - M_\pi^2 - [2M_K^2 - M_\pi^2]^{\rm phys} \tag{90}$$

to parametrize the small deviation of our strange quark mass from its physical value [2]. Alternatively, for the light quark mass we use a Taylor expansion of the form

$$m_{ud} = c_1 + c_2 M_\pi^2 + c_3 M_\pi^4 + c_4 \Delta \tag{91}$$

while the strange quark mass is always parametrized as

$$m_s = c_5 + c_6 M_\pi^2 + c_7 \Delta + c_8 \Delta^2 \ . \tag{92}$$

We have tried to augment these formulas by higher order terms, both in M_π^2 and Δ, but we found those coefficients to be consistent with zero, with the given precision of our data. This yields 3 options for the mass interpolation or extrapolation of the pseudoscalars. Similarly, for the Ω baryon that serves to set the scale, a Taylor ansatz in M_π^2 and $2M_K^2 - M_\pi^2$ is used (cf. Sec. 4). In total we have 3 functional ansaetze to interpolate our data.

A standard way to test the functional ansatz is to prune the data with mass cuts. We use $M_\pi < \{380, 480\}\,\mathrm{MeV}$ for the scale setting and $M_\pi < \{340, 380\}\,\mathrm{MeV}$ for the quark mass determination, thus a total of 4 mass cuts.

A source of error which, in practice, often proves dominant is the contamination of the ground state in the two-point correlator by excited states. To reduce this contamination we use a Gaussian source and sink with a fixed width of about $0.32\,\mathrm{fm}$. We tested 1-state and 2-state fits, and found complete agreement if the 1-state fits start at $t_{\rm min} \simeq 0.7\,\mathrm{fm}$ for the PP, PA_4, A_4P, A_4A_4 meson channels and from $t_{\rm min} \sim 0.8\,\mathrm{fm}$ for the Ω. In lattice units this amounts to $at_{\rm min} = \{6, 8, 9, 11, 13\}$ for $\beta = \{3.31, 3.5, 3.61, 3.7, 3.8\}$ (and $\sim 20\%$ later for baryons). In order to estimate any remaining excited state effects, we repeated our analysis with an even more conservative meson fit range (starting at $at_{\rm min} = \{7, 9, 11, 13, 15\}$ and again

cen. val.	$\sigma_{\rm stat}$	$\sigma_{\rm syst}$	plateau	scale set	fit form	mass cut	renorm.	cont.
3.503	0.048	0.049	0.330	0.034	0.030	0.157	0.080	0.926
96.43	1.13	1.47	0.207	0.005	0.031	0.085	0.085	0.970
27.531	0.196	0.083	0.513	0.200	0.023	0.320	—	0.771

Table 4: *Split-up of the total systematic uncertainty of $m_{ud}^{\rm phys}$, $m_s^{\rm phys}$ and $m_s^{\rm phys}/m_{ud}^{\rm phys}$ (from top to bottom) into the various contributions. Entries in columns 1-3 are in MeV and refer to the RI/MOM scheme at $\mu = 4\,\mathrm{GeV}$. Columns 4-10 indicate the relative share of the systematic error given in column 3 (the squares of these numbers add up to 1). The headers of these columns refer to the plateau range in the primary observables, the overall scale setting, the interpolation ansatz to tune to the physical mass point, the cut in the pion mass, the details of the renormalization procedure (read-off scale, chiral extrapolation), and the continuum extrapolation.*

$\sim 20\%$ later for baryons). The end of the fit interval was always chosen to be $at_{\rm max} = 2.7 \times at_{\rm min}$ or $T/2 - 1$ for lattices with a time extent shorter than $5.4 \times at_{\rm min}$. In all cases, the fits were performed in a correlated way. In total this gives 2 different fit ranges to make sure that contamination by excited states is under control.

As a result of the tree-level value $c_{\rm SW} = 1$ our action has formally $O(\alpha a)$ cut-off effects. However, due to the smearing the coefficient in front of this term is small, and the formally subleading $O(a^2)$ contributions might numerically dominate over the $O(\alpha a)$ part. To account for this we augment our global fit by cutoff terms which stipulate either $O(\alpha a)$ or $O(a^2)$ deviation from the continuum. This ambiguity comes into play in the evolution function (69) and in the continuum extrapolation of the quark masses in the RI/MOM scheme, which yields 4 options.

Besides the variations described above, we consider 3 options in the nonperturbative renormalization procedure (scale μ, massless versus massive intermediate scheme), see Sec. 11.

All of this serves the goal of quantifying potential systematic effects on our final results. In addition, there are standard methods to assess the size of the statistical error. Apart from the autocorrelation analysis detailed in Sec. 7, we used different blocking sizes on our ensembles, ranging from 1 to 10 configurations, where two adjacent configurations are separated by ten $\tau = 1$ MD updates (cf. Sec. 5). Last but not least, we found that artificial thermalization cuts (where we ignore the first 20-100 configurations of the thermalized ensembles) induce no noticeable change in our results, and therefore we conclude that possible residual thermalization effects are irrelevant for the error analysis.

Putting everything together we have 3 ansaetze for the interpolation of the quark masses to the physical point, 4 mass cuts in the scale setting and the quark mass determination, 2 different fit intervals for the primary observables, 4 ansaetze for the continuum extrapolation, and 3 ways of doing the RI/MOM renormalization. This gives a total of $3 \cdot 4 \cdot 2 \cdot 4 \cdot 3 = 288$ analyses.

In order to quote a final result, we follow the procedure used in [2]. It is essential to note that we have no *a priori* reason to favor one of these fits over another. Therefore the analysis method should represent the full spread of all reasonable and theoretically justified treatments of our data. In other words, we use all 288 procedures, and weigh the results by the quality of

	m_s	m_{ud}	m_u	m_d
RI/MOM$(4\,\mathrm{GeV})$	96.4(1.1)(1.5)	3.503(48)(49)	2.17(04)(10)	4.84(07)(12)
RGI	127.3(1.5)(1.9)	4.624(63)(64)	2.86(05)(13)	6.39(09)(15)
$\overline{\mathrm{MS}}(2\,\mathrm{GeV})$	95.5(1.1)(1.5)	3.469(47)(48)	2.15(03)(10)	4.79(07)(12)

Table 5: *Renormalized quark masses in the RI/MOM scheme at $\mu = 4\,\mathrm{GeV}$, and after conversion to RGI and the $\overline{\mathrm{MS}}$ scheme at $\mu = 2\,\mathrm{GeV}$. The RI/MOM values are fully nonperturbative, so the first line is our main result. The first two columns emerge directly from our lattice calculation, the last two build, in addition, on dispersive information on Q. The precision of m_s and m_{ud} is somewhat below the 2% level, for m_u and m_d it is about 5% and 3%, respectively. The ratio $m_s/m_{ud} = 27.53(20)(08)$ is independent of the scheme and accurate to better than 1%.*

fit $Q = \Gamma(n/2, \chi^2/2)$ to form a histogram. Next, we compute the mean and standard deviation of the distribution, and this yields the central value and the systematic error which we quote. Finally, we repeat this extensive procedure on 2000 bootstrap samples. The standard bootstrap error of the mean gives the statistical error.

An additional benefit of our method to treat systematic effects is that we can temporarily suppress one of the variations considered (i.e. abandon one of the factors who's product leads to the 288 procedures) to learn about the contribution of this individual factor to the total error. The total "error budget" compiled in this way is shown in Tab. 4. Evidently, it exhibits the cut-off effects as the dominant source of systematic uncertainty in our results.

All together, our procedure to assess both statistical and systematic errors is an extended frequentist method [20] which was already used in [2].

15 Summary

We have carried out a precise determination of the average light quark mass $m_{ud} = (m_u + m_d)/2$ and of the strange quark mass m_s, using nonperturbative $N_f = 2+1$ lattice QCD and nonperturbative renormalization throughout. Our data cover 5 lattice spacings in the range $0.054 - 0.116\,\mathrm{fm}$, with pion masses down to $\sim 120\,\mathrm{fm}$ and box sizes up to $6\,\mathrm{fm}$. This allows for a safe extrapolation to the continuum ($a \to 0$) and to infinite volume ($L \to \infty$).

We have devised a number of innovative methods, most notably a scheme to exploit the different renormalization pattern of Wilson and PCAC quark masses with tree-level $O(a)$-improved clover quarks and a procedure to overcome the RI/MOM window problem by taking a separate continuum limit of the running of the scalar density $R_S(\mu, \mu')$.

Our main result, m_s and m_{ud} in the RI/MOM scheme at renormalization scale $\mu = 4\,\mathrm{GeV}$ (cf. Tab. 5), is from first principles and fully nonperturbative. To ease comparison with the literature, these values are converted to RGI conventions and, subsequently, to the $\overline{\mathrm{MS}}$ scheme. In this step reference to perturbation theory is unavoidable, but we do this in a controlled way, since we show that the 4-loop anomalous dimension of the scalar density is consistent with our

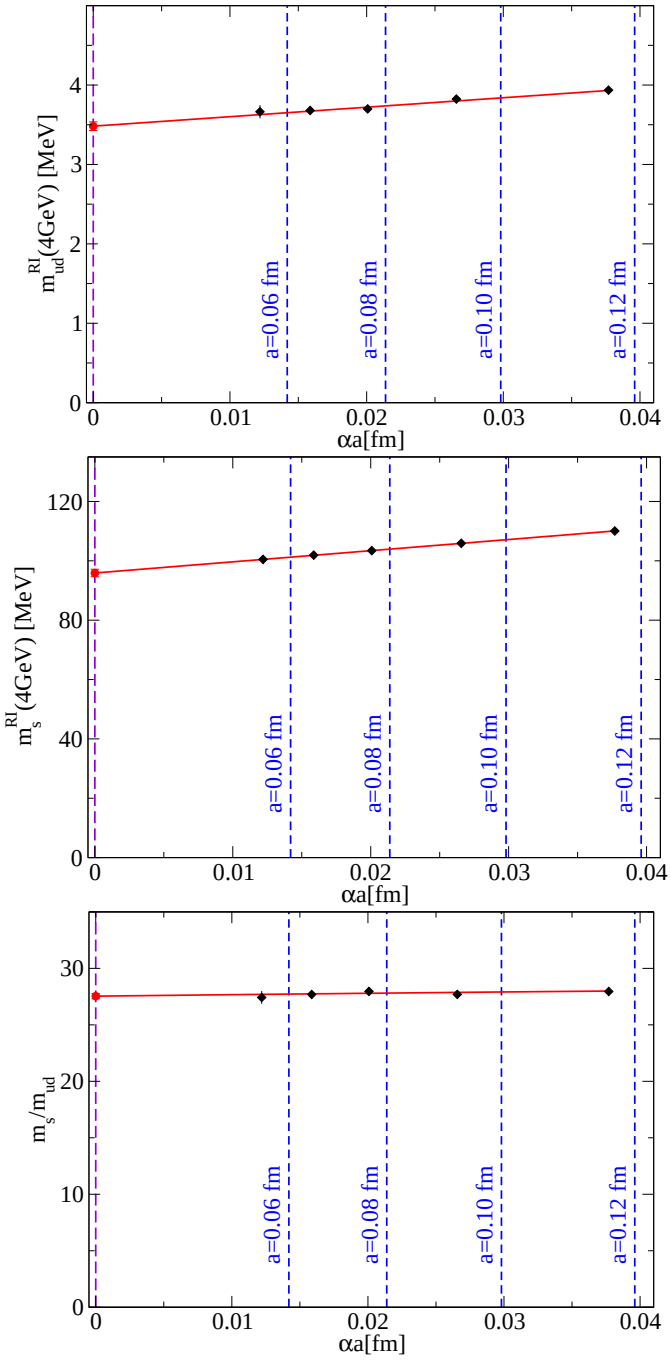

Figure 14: *Continuum extrapolation of m_{ud} (top), m_s (middle), m_s/m_{ud} (bottom) versus αa, for one of our 288 analyses with a good fit quality (cf. discussion in Sec. 14).*

nonperturbative running for $\mu \gtrsim 4\,\mathrm{GeV}$. The ratio m_s/m_{ud} is scheme and scale invariant. It turns out that our action entails favorable scaling properties not just for hadron masses, but also for

renormalized quark masses, as the plot of a representative continuum extrapolation in Fig. 14 shows. The combination of using pion masses down to (and even below) the physical value and having precise and fully nonperturbative renormalization factors allows us to determine m_s and m_{ud} with a precision of better than 2%, and their ratio to better than 1%.

A determination of the individual light quark masses m_u and m_d by lattice methods alone is beyond the scope of this paper. Nevertheless, the precision of our values of m_{ud} and m_s/m_{ud} allows for a fruitful use of the result of the dispersive analysis of the double ratio Q (cf. discussion in Sec. 13). By combining these pieces of information, we obtain values of the individual quark masses m_u and m_d with a precision of 5% and 3%, respectively (cf. Tab. 5).

Acknowledgments

We used HPC resources from FZ Jülich and from GENCI-[IDRIS/CCRT] grant 52275, as well as clusters at Wuppertal and CPT. This work is supported in part by EU grants I3HP, FP7/2007-2013/ERC n^o 208740, MRTN-CT-2006-035482 (FLAVIAnet), DFG grant FO 502/2, SFB-TR 55, CNRS GDR 2921 and PICS 4707.

References

[1] S. Durr *et al.* [BMW Collaboration], arXiv:1011.2403.

[2] S. Durr *et al.* [BMW Collaboration], Science **322**, 1224 (2008).

[3] G. Colangelo, S. Durr and C. Haefeli, Nucl. Phys. B **721**, 136 (2005) [hep-lat/0503014].

[4] M. Luscher and P. Weisz, Phys. Lett. B **158**, 250 (1985).

[5] B. Sheikholeslami and R. Wohlert, Nucl. Phys. B **259**, 572 (1985).

[6] A. Hasenfratz and F. Knechtli, Phys. Rev. D **64**, 034504 (2001) [hep-lat/0103029].

[7] C. Morningstar and M. J. Peardon, Phys. Rev. D **69**, 054501 (2004) [hep-lat/0311018].

[8] S. Capitani, S. Durr and C. Hoelbling, JHEP **0611**, 028 (2006) [hep-lat/0607006].

[9] S. Durr *et al.* [BMW Collaboration], Phys. Rev. D **79**, 014501 (2009) [0802.2706, hep-lat].

[10] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, Phys. Lett. B **195**, 216 (1987).

[11] T. A. DeGrand and P. Rossi, Comput. Phys. Commun. **60**, 211 (1990).

[12] J. C. Sexton and D. H. Weingarten, Nucl. Phys. B **380**, 665 (1992).

[13] M. Hasenbusch, Phys. Lett. B **519**, 177 (2001) [hep-lat/0107019].

[14] T. Takaishi and P. de Forcrand, Phys. Rev. E **73**, 036706 (2006) [hep-lat/0505020].

[15] M. A. Clark and A. D. Kennedy, Phys. Rev. Lett. **98**, 051601 (2007) [hep-lat/0608015].

[16] C. Urbach, K. Jansen, A. Shindler and U. Wenger, Comput. Phys. Commun. **174**, 87 (2006) [hep-lat/0506011].

[17] S. Aoki *et al.* [PACS-CS Collaboration], Phys. Rev. D **79**, 034503 (2009) [0807.1661, hep-lat].

[18] A. Hasenfratz, R. Hoffmann and S. Schaefer, JHEP **0705**, 029 (2007) [hep-lat/0702028].

[19] W. Kamleh, D. B. Leinweber and A. G. Williams, Phys. Rev. D **70**, 014502 (2004) [hep-lat/0403019].

[20] K. Nakamura *et al.* [Particle Data Group], J. Phys. G **37**, 075021 (2010).

[21] G. Colangelo *et al.* [FLAG working group], "Review of lattice results concerning low-energy particle physics", forthcoming.

[22] C. Aubin *et al.* [MILC Collaboration], Phys. Rev. D **70**, 114501 (2004) [hep-lat/0407028].

[23] I. Montvay, G. Munster, *Quantum fields on a lattice* (Cambridge Univ. Pr., Cambridge, 1994).

[24] L. Del Debbio, L. Giusti, M. Luscher, R. Petronzio and N. Tantalo, JHEP **0602**, 011 (2006) [hep-lat/0512021].

[25] B. Alles, G. Boyd, M. D'Elia, A. Di Giacomo and E. Vicari, Phys. Lett. B **389**, 107 (1996) [hep-lat/9607049].

[26] B. Alles *et al.*, Phys. Rev. D **58**, 071503 (1998) [hep-lat/9803008].

[27] S. Schaefer, R. Sommer and F. Virotta, PoS **LAT2009**, 032 (2009) [0910.1465, hep-lat].

[28] M. Luscher, Commun. Math. Phys. **104**, 177 (1986).

[29] M. Luscher, Commun. Math. Phys. **105**, 153 (1986).

[30] M. Luscher, Nucl. Phys. B **354**, 531 (1991).

[31] M. Luscher, Nucl. Phys. B **364**, 237 (1991).

[32] J. Gasser and H. Leutwyler, Phys. Lett. B **184**, 83 (1987).

[33] J. Gasser and H. Leutwyler, Phys. Lett. B **188**, 477 (1987).

[34] J. Gasser and H. Leutwyler, Nucl. Phys. B **307**, 763 (1988).

[35] A. Ali Khan *et al.* [QCDSF-UKQCD Collaboration], Nucl. Phys. B **689**, 175 (2004) [hep-lat/0312030].

[36] G. Colangelo, A. Fuhrer and S. Lanz, Phys. Rev. D **82**, 034506 (2010) [1005.1485, hep-lat].

[37] J. Gasser and H. Leutwyler, Annals Phys. **158**, 142 (1984).

[38] J. Gasser and H. Leutwyler, Nucl. Phys. B **250**, 465 (1985).

[39] G. Colangelo, J. Gasser and H. Leutwyler, Nucl. Phys. B **603**, 125 (2001) [hep-ph/0103088].

[40] M. Gockeler, R. Horsley, A. C. Irving, D. Pleiter, P. E. L. Rakow, G. Schierholz and H. Stuben [QCDSF Collaboration and UKQCD Collaboration], Phys. Lett. B **639**, 307 (2006) [hep-ph/0409312].

[41] P. E. L. Rakow, Nucl. Phys. Proc. Suppl. **140**, 34 (2005) [hep-lat/0411036].

[42] T. Bhattacharya, R. Gupta, W. Lee, S. R. Sharpe and J. M. S. Wu, Phys. Rev. D **73**, 034504 (2006) [hep-lat/0511014].

[43] G. Martinelli, C. Pittori, C. T. Sachrajda, M. Testa and A. Vladikas, Nucl. Phys. B **445**, 81 (1995) [hep-lat/9411010].

[44] K. G. Chetyrkin and A. Retey, Nucl. Phys. B **583**, 3 (2000) [hep-ph/9910332].

[45] D. B. Leinweber, J. I. Skullerud, A. G. Williams and C. Parrinello [UKQCD collaboration], Phys. Rev. D **58**, 031501 (1998) [hep-lat/9803015].

[46] T. Bakeyev *et al.* [QCDSF-UKQCD Collaboration], Phys. Lett. B **580**, 197 (2004) [hep-lat/0305014].

[47] G. Martinelli, C. T. Sachrajda and A. Vladikas, Nucl. Phys. B **358**, 212 (1991).

[48] G. Martinelli, S. Petrarca, C. T. Sachrajda and A. Vladikas, Phys. Lett. B **311**, 241 (1993) [Erratum-ibid. B **317**, 660 (1993)].

[49] D. Becirevic, V. Gimenez, V. Lubicz and G. Martinelli, Phys. Rev. D **61**, 114507 (2000) [hep-lat/9909082].

[50] S. Capitani, M. Gockeler, R. Horsley, H. Perlt, P. E. L. Rakow, G. Schierholz and A. Schiller, Nucl. Phys. B **593**, 183 (2001) [hep-lat/0007004].

[51] G. Martinelli, G. C. Rossi, C. T. Sachrajda, S. R. Sharpe, M. Talevi and M. Testa, Nucl. Phys. B **611**, 311 (2001) [hep-lat/0106003].

[52] V. Maillart and F. Niedermayer, arXiv:0807.0030 [hep-lat].

[53] J. Garden, J. Heitger, R. Sommer and H. Wittig [ALPHA and UKQCD Collaboration], Nucl. Phys. B **571**, 237 (2000) [hep-lat/9906013].

[54] R. Sommer, Nucl. Phys. B **411**, 839 (1994) [hep-lat/9310022].

[55] S. Aoki *et al.* [JLQCD Collaboration], Phys. Rev. Lett. **82**, 4392 (1999) [hep-lat/9901019].

[56] S. Durr and C. Hoelbling, Phys. Rev. D **72**, 071501 (2005) [hep-ph/0508085].

[57] S. Aoki *et al.* [CP-PACS Collaboration], Phys. Rev. D **67**, 034503 (2003) [hep-lat/ 0206009].

[58] R. F. Dashen, Phys. Rev. **183**, 1245 (1969).

[59] H. Leutwyler, Phys. Lett. B **378**, 313 (1996) [hep-ph/9602366].

[60] C. Ditsche, B. Kubis and U. G. Meissner, Eur. Phys. J. C **60**, 83 (2009) [0812.0344, hep-ph].

[61] J. Kambor, C. Wiesendanger and D. Wyler, Nucl. Phys. B **465**, 215 (1996) [hep-ph/9509374].

[62] A. V. Anisovich and H. Leutwyler, Phys. Lett. B **375**, 335 (1996) [hep-ph/9601237].

[63] G. Colangelo, S. Lanz and E. Passemar, PoS **CD09**, 047 (2009), [0910.0765, hep-ph].

[64] H. Leutwyler, PoS **CD09**, 005 (2009) [0911.1416, hep-ph].

[65] D. B. Kaplan and A. V. Manohar, Phys. Rev. Lett. **56**, 2004 (1986).

[66] H. Leutwyler, private communication.

[67] T. Blum, T. Doi, M. Hayakawa, T. Izubuchi, S. Uno, N. Yamada and R. Zhou, arXiv:1006.1311 [hep-lat].