THE BROWN MUCK OF B^0 AND B_S^0 MIXING: BEYOND THE STANDARD MODEL

BY

CHRISTOPHER MICHAEL BOUCHARD

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Kevin Pitts, Chair
Professor Aida El-Khadra, Director of Research
Professor John Stack
Professor Michael Stone

Abstract

Standard Model contributions to neutral B meson mixing begin at the one loop level where they are further suppressed by a combination of the GIM mechanism and Cabibbo suppression. This combination makes B meson mixing a promising probe of new physics, where as yet undiscovered particles and/or interactions can participate in the virtual loops.

Relating underlying interactions of the mixing process to experimental observation requires a precise calculation of the non-perturbative process of hadronization, characterized by hadronic mixing matrix elements. This thesis describes a calculation of the hadronic mixing matrix elements relevant to a large class of new physics models. The calculation is performed via lattice QCD using the MILC collaboration's gauge configurations with $2 + 1$ dynamical sea quarks.

To Amy, Max, Claire, Mom, and Dad. Thanks for everything.

Acknowledgments

Thanks to Aida and Andreas for advising and mentoring me. Some students are lucky enough to have a good advisor. I've had two.

To Chris, Vishnu, Kevin, Todd, Paco, Elvira, Elizabeth, Gordan, Alejandro, and Yuhsin – a large chunk of the physics I know I've learned over beer and coffee with you.

Table of Contents

List of Tables

List of Figures

Chapter 1

The Standard Model

1.1 Particles and gauge symmetries

The Standard Model (SM) is a $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge theory that describes, with great success, observed electroweak and strong interactions. The fermions of the SM come in three generations, or flavors, labeled here by index $i = 1, 2, 3$

$$q^I_{Li}(3,2)_{\frac{1}{6}}, \quad u^I_{Ri}(3,1)_{\frac{2}{3}}, \quad d^I_{Ri}(3,1)_{-\frac{1}{3}}, \quad l^I_{Li}(1,2)_{-\frac{1}{2}}, \quad e^I_{Ri}(1,1)_{-1} \tag{1.1}$$

where, for example, the left handed quark, $q^I_{Li}(3,2)_{\frac{1}{6}}$, is a triplet under $SU(3)_C$, a doublet under $SU(2)_L$, and has hypercharge $1/6$. The superscript I indicates an eigenstate of the interaction basis.

To preserve invariance under the SM gauge groups, the kinetic terms in the SM Lagrangian density are built with the gauge covariant derivative

$$D^\mu = \partial^\mu + ig_s G^\mu_a L_a + igW^\mu_b T_b + ig' B^\mu Y, \tag{1.2}$$

where g_s, g, and g' are the strong, weak, and hypercharge gauge couplings, respectively. The eight G^μ_a are the $SU(3)_C$ octet of gluons responsible for mediating the strong interaction and L_a are generators of $SU(3)_C$ (3×3 Gell-Mann matrices for the $SU(3)_C$ triplet quark fields, and 0 for $SU(3)_C$ singlets). The three W^μ_b are the triplet of weak gauge bosons responsible for mediating the unbroken $SU(2)_L$ gauge interaction and T_b are generators of $SU(2)_L$ (2×2 Pauli matrices for the $SU(2)_L$ doublet fields, 0 for the right-handed $SU(2)_L$ singlets). B^μ is the hypercharge gauge boson responsible for mediating the unbroken $U(1)_Y$ gauge interaction and Y is the hypercharge of Eq. (1.1). The resulting kinetic terms in the Lagrangian density are

given by

$$\begin{aligned}\mathcal{L}_{\rm SM}^{\rm kinetic} &= i\bar{q}^I_{Li}\gamma_\mu\left(\partial^\mu + ig_s G^\mu_a L_a + igW^\mu_b T_b + \frac{i}{6}g'B^\mu\right)q^I_{Li} \\ &+ i\bar{l}^I_{Li}\gamma_\mu\left(\partial^\mu + igW^\mu_b T_b - \frac{i}{2}g'B^\mu\right)l^I_{Li} + \cdots \end{aligned} \tag{1.3}$$

where the ellipses contain kinetic terms for bosonic SM particles.

In addition to the fermions and gauge bosons, we consider here the standard, minimal extension to the SM to generate spontaneous electroweak symmetry breaking (EWSB) by introducing a Higgs $SU(2)_L$ doublet, $\phi(1,2)_{\frac{1}{2}}$. The presence of a fundamental scalar to achieve EWSB has not been confirmed and alternatives (*eg.* technicolor theories [1, 2] in which a dynamically broken, strongly coupled gauge theory generates a pseudo Nambu Goldstone boson that plays the role of the Higgs) seem more likely with each update to the Higgs exclusion bounds [3, 4]. Independent of the mechanism, EWSB due to the presence of a *Higgs-like* particle or condensate is generally accepted. In the framework of the minimally extended SM considered here, all particles are massless prior to EWSB.

Higgs and Yukawa interactions in the SM (those interactions that will, upon spontaneous symmetry breaking, be responsible for generating SM masses) are encoded in the following terms of the SM Lagrangian density

$$\begin{aligned}\mathcal{L}_{\rm SM}^{\rm Higgs,Yukawa} &= \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \\ &- Y^u_{ij}\bar{q}^I_{Li}\epsilon\phi^* u^I_{Rj} - Y^d_{ij}\bar{q}^I_{Li}\phi d^I_{Rj} + \text{h.c.} \\ &- Y^l_{ij}\bar{l}^I_{Li}\phi e^I_{Rj} + \text{h.c.}\end{aligned} \tag{1.4}$$

$Y^{u,d}$ are 3×3 complex matrices with 10 physical parameters corresponding to six quark masses and four parameters that will form the CKM matrix, and ϵ is the antisymmetric 2×2 matrix. Y^l is a 3×3 matrix with three physical parameters corresponding to the masses of the three charged leptons.

1.2 Electroweak symmetry breaking

The Higgs gets a vacuum expectation value, $\langle\phi\rangle = (0,\ v/\sqrt{2})$ where $v \approx 246$ GeV, spontaneously breaking the SM gauge group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \longrightarrow SU(3)_C \otimes U(1)_{EM}. \tag{1.5}$$

The gluons of the unbroken $SU(3)_C$ remain massless. Of the four gauge bosons, W_b^μ and B^μ, representing $SU(2)_L \otimes U(1)_Y$, one admixture (the photon) corresponding to the unbroken $U(1)_{EM}$ remains massless. Admixtures of gauge bosons corresponding to the three broken gauge degrees of freedom form the massive $W_\mu^\pm$ and Z_μ^0, mediators of the weak interaction.

In addition to the three gauge bosons, the quarks and charged leptons also acquire masses from EWSB. Focusing on the quarks, the mass matrices that results from Eq. (1.4) are in the basis of weak eigenstates and are not diagonal. Diagonalizing the $Y^{u,d}$ with the unitary matrices, $U_{L,R}^{u,d}$, gives the diagonal mass matrices

$$M^{u,d} = U_L^{u,d}\ Y^{u,d}\ (U_R^{u,d})^\dagger\ v/\sqrt{2} \tag{1.6}$$

and quark mass eigenstates

$$q_{Li} = (U_L^q)_{ij} q_{Lj}^I, \qquad \text{and} \qquad q_{Ri} = (U_R^q)_{ij} q_{Rj}^I, \qquad \text{where} \qquad q = u, d. \tag{1.7}$$

Interactions between quarks and $SU(2)_L \otimes U(1)_Y$ gauge bosons of Eq. (1.3) are governed by

$$\mathcal{L}_{\rm SM}^{q^I - W_\mu, B_\mu} = -\bar{q}_{Li}^I \left(g\gamma_\mu W_b^\mu T_b + \frac{g'}{6} B^\mu \right) q_{Li}^I. \tag{1.8}$$

After EWSB, we rewrite the $SU(2)_L \otimes U(1)_Y$ gauge bosons, W_b^μ and B^μ, in linear combinations that correspond to the massless photon and the three massive weak gauge bosons, $W_\mu^\pm$ and Z_μ^0. In the mass basis, the charged current interactions of the quarks and the $W_\mu^\pm$ are

$$\mathcal{L}_{\rm SM}^{q - W_\mu^\pm} = -\frac{g}{\sqrt{2}} \bar{u}_{Li} \gamma^\mu (U_L^u U_L^{d\dagger})_{ij} d_{Lj} W_\mu^+ + h.c. \tag{1.9}$$

1.3 The CKM matrix

The unitary 3×3 matrix $U_{uL} U_{dL}^\dagger$ is defined as the Cabibbo-Kobayashi-Maskawa (CKM) matrix,

$$V_{\rm CKM} = U_L^u U_L^{d\dagger} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{1.10}$$

An alternate parameterization, called the Wolfenstein parameterization, makes explicit the fact that there are only four physical parameters in V_{CKM}, illustrates its hierarchical structure, and enforces the SM assumption

of unitarity. Following PDG [5] notation for the Wolfenstein parameterization, we define V_{CKM} in terms of the four parameters λ, A, ρ, and η,

$$V_{\rm CKM} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{1.11}$$

where $\lambda = |V_{us}|/\sqrt{|V_{ud}|^2 + |V_{us}|^2} \approx 0.23$, $\bar{\rho} + i\bar{\eta} = -(V_{ud}V_{ub}^*)/(V_{cd}V_{cb}^*)$, and $A = |V_{cb}|/(\lambda|V_{us}|)$.

V_{CKM} permits charged-current, flavor-changing interactions mediated by the $W_\mu^\pm$. Interactions between the quarks and neutral Z_μ^0, for left-handed up type quarks in the mass basis, are governed by

$$\begin{aligned} \mathcal{L}_{\rm SM}^{q-Z_\mu^0} &= -\frac{g}{\cos\theta_W}\left(\frac{1}{2} - \frac{2}{3}\sin^2\theta_W\right)\bar{u}_{Li}\gamma^\mu u_{Lj}(U_L^u)_{ik}(U_L^{u\dagger})_{kj}\ Z_\mu^0 \\ &= -\frac{g}{\cos\theta_W}\left(\frac{1}{2} - \frac{2}{3}\sin^2\theta_W\right)\bar{u}_{Li}\gamma^\mu u_{Li}\ Z_\mu^0 \end{aligned} \tag{1.12}$$

where unitarity of the $U_{R,L}^{u,d}$ gives $(U_L^u)_{ik}(U_L^{u\dagger})_{kj} = \delta_{ij}$, preventing flavor-changing neutral current (FCNC) interactions. This result holds for the down type quarks as well.

Unitarity of V_{CKM} ($V_{CKM}V_{CKM}^\dagger = 1$) implies that $\text{row}_i \cdot \text{row}_j^* = \text{col}_i \cdot \text{col}_j^* = \delta_{ij}$. Of particular significance to B mixing are the relations

$$V_{uq}V_{ub}^* + V_{cq}V_{cb}^* + V_{tq}V_{tb}^* = 0, \qquad \text{where } q = d, s. \tag{1.13}$$

These expressions, three complex numbers adding to zero, can be expressed graphically as triangles in the complex plane. The resultant triangles are the so-called *Unitarity Triangles*. One triangle, for $q = d$, has sides of roughly the same length and the greatest number of relevant experimental measurements. It is typically referred to as "the Unitary Triangle" and is shown, in the complex $\bar{\rho} - \bar{\eta}$ plane, in Fig. 1.1 with angles given by

$$\alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \text{and} \quad \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{1.14}$$

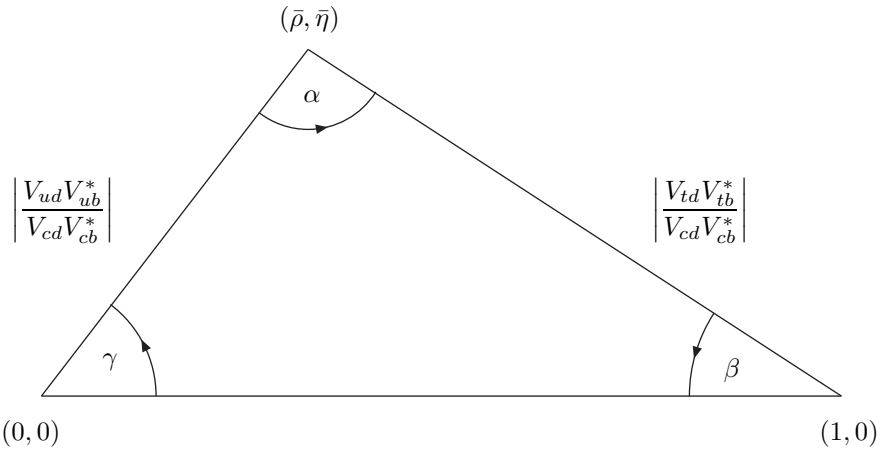


Figure 1.1: The Unitarity Triangle (*ie.* $V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0$) in the complex $\bar{\rho}$ – $\bar{\eta}$ plane. Figure adapted from Ref. [6].

The magnitudes of the elements of V_{CKM} have been determined to be [5]

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97425 \pm 0.00022 & 0.2252 \pm 0.0009 & (3.89 \pm 0.44) \times 10^{-3} \\ 0.230 \pm 0.011 & 1.023 \pm 0.036 & (40.6 \pm 1.3) \times 10^{-3} \\ (8.4 \pm 0.6) \times 10^{-3} & (38.7 \pm 2.1) \times 10^{-3} & 0.88 \pm 0.07 \end{pmatrix}. \quad (1.15)$$

Entries for $|V_{td}|$ and $|V_{ts}|$ come from B mixing measurements (*ie.* Eq. (2.4)) combined with lattice QCD SM input via Eq. (3.15), and the assumption that $|V_{tb}| = 1$ (roughly equivalent to assuming third row unitarity). All other entries are free of unitarity constraints. In particular, the reported value for $|V_{tb}|$ has been determined directly from top quark decays and is $\sim 1\sigma$ less than the value implied by $|V_{td}|$, $|V_{ts}|$, and unitarity.

The general success of the CKM description of SM flavor physics means new physics scenarios (*eg.* EWSB via other than a fundamental scalar) are tightly constrained. They must reproduce observed levels of unitarity in flavor-changing charged- and suppression in flavor-changing neutral-current interactions. The dilema this poses for generic new physics scenarios is referred to as the new physics flavor problem [6]. Better precision for nonperturbative input to new physics theory predictions means more leverage for flavor physics to sort out what lies beyond the SM.

Chapter 2

Motivation for improved precision

2.1 Theory

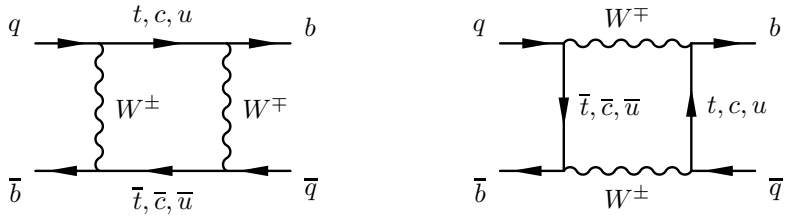


Figure 2.1: Leading order Feynman diagrams for SM B_q^0 mixing, where $q \in \{d, s\}$.

B mixing calculations and experimental measurements are motivated by the simple fact that the SM contribution to B mixing is small. The SM contribution is small because it begins at the loop level where its further suppressed by a combination of the GIM mechanism and Cabibbo suppression. This opens the door for the possibility of relatively sizeable new physics contributions.

Fig. 2.1 shows the leading order, one loop, Feynman diagrams for SM B mixing. Loop suppression is realized by the fact that the amplitude for this diagram is proportional to $\alpha_W^2 \sim 10^{-3}$.

The amplitude for each diagram in Fig. 2.1 can generically be written

$$\mathcal{A} \propto \left[\sum_{i=u,c,t} f(m_i) V_{ib} V_{iq}^* \right]^2 \tag{2.1}$$

where $f(m) \propto m^2/M_W^2$ [6]. Expanding the sum and using V_{CKM} unitarity to make the replacement $V_{ub}V_{uq}^* = -V_{cb}V_{cq}^* - V_{tb}V_{tq}^*$, we have

$$\mathcal{A} \propto \frac{1}{M_W^4} \left[(m_c^2 - m_u^2) V_{cb} V_{cq}^* + (m_t^2 - m_u^2) V_{tb} V_{tq}^* \right]^2. \tag{2.2}$$

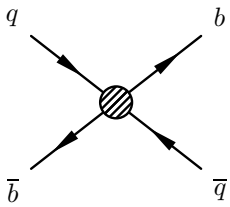


Figure 2.2: The effective four-quark interaction for B_q^0 mixing.

Given the relative size of the quark masses, $m_u \ll m_c \ll m_t$,

$$\mathcal{A} \propto \frac{m_t^4}{M_W^4} \left| V_{tb} V_{tq}^* \right|^2 . \tag{2.3}$$

Together with suppression by factors of m_u^4/M_W^4 and m_c^4/M_W^4, this partial cancellation of the amplitude due to the unitarity of the CKM matrix, known as the GIM mechanism[1], effectively kills all but the top quark contribution to B mixing and, naively, attributes a suppression factor of 1/3. The role of the GIM mechanism in B mixing goes far beyond its meager contribution to suppressing the SM amplitude. Our ability to describe B mixing via an effective, local, four quark mixing operator, as in Fig. 2.2, arises from the fact that only the top quark and $W_\mu^\pm$ contribute significantly.

The last source of SM B mixing suppression comes from the presence of $|V_{tb}V_{tq}^*|^2$ in Eq. (2.3). Using the values from Eq. (1.15), $|V_{tb}V_{td}^*|^2 \approx 5.5 \times 10^{-3}$ and $|V_{tb}V_{ts}^*|^2 \approx 1.2 \times 10^{-3}$. This is referred to as Cabibbo suppression. Combining the suppression factors, and $(m_t/M_W)^4$ enhancement, the amplitude for SM B mixing is suppressed by an overall factor of $\sim 10^{-6}$.

Examples of new physics scenarios whose contributions to B mixing have been studied in detail (*eg.* Buras *et al.* [10] and Lenz *et al.* [11]) are supersymmetric flavor models, little Higgs models with extended weak gauge groups, and Randall-Sundrum warped extra dimension models. The beauty of the effective theory description of B mixing is that it provides a model-independent way to characterize the hadronic contribution to BSM B mixing. Our analysis is therefore free of the details related to these, or any other, specific BSM scenario. Provided new physics flavor changing interactions are mediated by a particle at least as massive as the $W^\pm$, the effective theory picture of B mixing via a local, four quark interaction is valid.

[1]Note that if the quark masses are degenerate, the amplitude vanishes. The fact that experimental limits on $\Gamma(K^0 \to \mu^+\mu^-)$ were much less than calculated decay rates prompted Glashow, Iliopolous, and Maiani [7] to postulate the existence of the charm quark to facilitate this cancellation. They proposed this in 1970 – four years before the discovery of the charm quark [8, 9].

2.2 Experiment

More important than the mere *idea* that small SM contributions may make possible significant BSM contributions, there are *actual* experimental hints of new physics related to B mixing.

The past four years have seen a persistent, growing, tension in the unitarity triangle [12, 13, 14, 15, 16]. These analyses perform global fits to multiple experimental and lattice inputs to extract best fit values of the sides and angles of the unitarity triangle. Tension in the fits indicates best fit values fail to form a closed triangle. By removing a set of inputs related to a specific physical process (*eg.* B mixing) and redoing the fit, the sensitivity of the tension to the process can be assessed. The most recent analysis [17] reports a 3σ tension and attributes it to B mixing.

The UT*fit* collaboration performed a model independent analysis of B_s^0 mixing–related measurements and found evidence for new physics at the level of 3σ [18].

CP even initial state $p\bar{p}$ collisions at the Tevatron occasionally produce B_q^0, $\bar{B}_q^0$ pairs which, with a roughly 10% probability, undergo semi-leptonic decay to a final state μ^+ or μ^-, respectively. If one of the B_q^0 mesons oscillates prior to decay, a same sign dimuon final state will result. SM CP violation results in a non-zero, though small, same sign dimuon charge asymmetry corresponding to about 0.03% more $\mu^-\mu^-$ than $\mu^+\mu^+$ events [19]. D0 has measured the same sign dimuon charge asymmetry in B_q^0 mixing and found deviation with the SM at the level of 3.9σ [19].

The SM branching fraction $\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{\rm SM} = (3.2 \pm 0.2) \times 10^{-9}$ [20]. CDF has recently measured $\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{\rm expt.} = (1.8^{+1.1}_{-0.9}) \times 10^{-8}$ [21], nearly 6 times the SM expectation – tantalizing evidence of new physics in the B_s^0 system. Recent work [22] has correlated new physics in B mixing and $B_s \to \mu^+\mu^-$ in the framework of several new physics scenarios. Additionally, if combined with a corresponding measurement of $\mathcal{B}(B_d^0 \to \mu^+\mu^-)_{\rm expt.}$, this would allow an optimum way to determine the amount of BSM contributions to $\Delta M_s/\Delta M_d$ [20].

Lastly, though no less important, is the precision of the experimental measurements of the B mixing oscillation frequencies,

$$\begin{aligned} \Delta M_d &= 0.507 \pm 0.003(\text{stat}) \pm 0.003(\text{syst})\ \text{ps}^{-1} \quad [5], \text{ and} \\ \Delta M_s &= 17.77 \pm 0.10(\text{stat}) \pm 0.07(\text{syst})\ \text{ps}^{-1} \quad [23]. \end{aligned} \tag{2.4}$$

These measurements are extremely precise, $< 1\%$. With theory errors at least an order of magnitude larger, there is plenty of room for improvement.

Chapter 3

Theoretical description of B mixing

We start with a quantum mechanical description of B mixing at the meson level, then transition to the quark level by describing the mesons via composite operators and utilizing the operator product expansion (OPE). The end result is an expression for the B mixing oscillation frequency to be evaluated by a combination of lattice QCD and perturbation theory.

3.1 Meson level

Imagine observing, in its rest frame, the life of a neutral B-meson. As we watch, it oscillates between the particle and anti-particle version of itself (*ie.* between its weak eigenstates). During the 10^{-12} s of its life[1] the B_d^0 (B_s^0) oscillates, on average, once (27 times) before decaying. We now develop an expression for the frequency of oscillation, ΔM.

We begin with a quantum mechanical, meson–level description, where the effects of decay are accounted for by the Wigner–Weisskopf approximation [24]. To span the Hilbert space we'd have to include $|B^0\rangle$, $|\bar{B}^0\rangle$, and all states accessible via decay. The list of possible decay products is long. In lieu of specifying them all we limit our description to states of interest, $|B^0\rangle$ and $|\bar{B}^0\rangle$, and account for the non–unitarity of our theory by adding an anti–Hermitian term to the Hamiltonian as a catch–all for the cumulative effect of decay to unspecified states,

$$i\frac{d}{dt}\begin{pmatrix} |B^0(t)\rangle \\ |\bar{B}^0(t)\rangle \end{pmatrix} = \left(M - \frac{i}{2}\Gamma\right)\begin{pmatrix} |B^0(t)\rangle \\ |\bar{B}^0(t)\rangle \end{pmatrix}. \tag{3.1}$$

The 2×2 Hermitian matrices M and Γ account for mixing and decay, respectively. Hermiticity implies real-valued diagonal elements, $M_{21} = M_{12}^*$, and $\Gamma_{21} = \Gamma_{12}^*$, while CPT invariance requires $M_{22} = M_{11}$ and

[1] The B_d^0 has a mean life of $(1.525 \pm 0.009) \times 10^{-12}$ s and the B_s^0 a mean life of $(1.472^{+0.024}_{-0.026}) \times 10^{-12}$ s.

$\Gamma_{22} = \Gamma_{11}$. The Hamiltonian is therefore

$$M - \frac{i}{2}\Gamma = \begin{pmatrix} M_{11} - \frac{i}{2}\Gamma_{11} & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{12}^* - \frac{i}{2}\Gamma_{12}^* & M_{11} - \frac{i}{2}\Gamma_{11} \end{pmatrix}, \tag{3.2}$$

with eigenvalues

$$\begin{aligned} \omega_H &= (M_{11} - \tfrac{i}{2}\Gamma_{11}) + \sqrt{|M_{12}|^2 - \tfrac{1}{4}|\Gamma_{12}|^2 - iRe(M_{12}\Gamma_{12}^*)} \text{ and} \\ \omega_L &= (M_{11} - \tfrac{i}{2}\Gamma_{11}) - \sqrt{|M_{12}|^2 - \tfrac{1}{4}|\Gamma_{12}|^2 - iRe(M_{12}\Gamma_{12}^*)} \end{aligned} \tag{3.3}$$

and mass eigenstates, labeled as heavy and light, $|B_H(t)\rangle$ and $|B_L(t)\rangle$. The weak eignenstates, $|B^0\rangle$ and $|\bar{B}^0\rangle$, are not eigenstates of the Hamiltonian. They mix with frequency ΔM and have a decay width difference $\Delta\Gamma$[2],

$$\Delta M = 2Re\Big(\sqrt{|M_{12}|^2 - \tfrac{1}{4}|\Gamma_{12}|^2 - iRe(M_{12}\Gamma_{12}^*)}\ \Big) \text{ and} \tag{3.4}$$

$$\Delta\Gamma = 2Im\Big(\sqrt{|M_{12}|^2 - \tfrac{1}{4}|\Gamma_{12}|^2 - iRe(M_{12}\Gamma_{12}^*)}\ \Big). \tag{3.5}$$

The weak eigenstates are related to the mass eigenstates by

$$|B_{H/L}\rangle = p|B^0\rangle \pm q|\bar{B}^0\rangle, \tag{3.6}$$

where $|p|^2 + |q|^2 = 1$ and

$$(\Delta M)^2 - \tfrac{(\Delta\Gamma)^2}{4} = 4|M_{12}|^2 - |\Gamma_{12}|^2, \quad \Delta M\Delta\Gamma = -4\ Re\ (M_{12}\Gamma_{12}^*), \text{ and } \tfrac{q}{p} = \tfrac{\Delta M + i\Delta\Gamma}{2M_{12} - i\ \Gamma_{12}}. \tag{3.7}$$

ΔM is determined by the off diagonal element of the Hamiltonian, $M_{12} - \frac{i}{2}\Gamma_{12}$, which represents the time-independent probability amplitude for a $B^0 \leftrightarrow \bar{B}^0$ transition.

Introducing a mixing Hamiltonian density[3], $\mathcal{H}_{\rm mix}$, we have

$$M_{12} - \frac{i}{2}\Gamma_{12} = \langle B^0|\mathcal{H}_{\rm mix}|\bar{B}^0\rangle. \tag{3.8}$$

[2] An ambiguity exists in the definition of the sign of $\Delta\Gamma$, which has yet to be experimentally determined.

[3] Dimensional arguments, beginning with Eq. (3.1) and including a requirement that meson states be consistently normalized before and after the OPE, require this to be a Hamiltonian density.

Figure 3.1: Diagrams representing the effective theory of mixing at the meson level via (*left*) a $\Delta b = 2$ interaction and (*right*) two $\Delta b = 1$ interactions with intermediate state $|n\rangle$.

Next, we perturbatively expand $\mathcal{H}_{\rm mix}$, in powers of the Fermi coupling G_F, giving

$$\mathcal{H}_{\rm mix} = \mathcal{H}^{\Delta b=0} + \mathcal{H}^{\Delta b=1} + \cdots \tag{3.9}$$

where $\mathcal{H}^{\Delta b=n} \propto G_F^n$ induces a change in heavy quark flavor, $\Delta b = n$. Stationary state perturbation theory and a strict flavor changing selection–rule (the total change of flavor in B mixing is $\Delta b = 2$) give

$$M_{12} - \frac{i}{2}\Gamma_{12} = \langle B^0|\mathcal{H}^{\Delta b=2}|\bar{B}^0\rangle + \sum_n \frac{\langle B^0|\mathcal{H}^{\Delta b=1}|n\rangle\langle n|\mathcal{H}^{\Delta b=1}|\bar{B}^0\rangle}{E_n - M_B}, \tag{3.10}$$

depicted graphically in Fig. 3.1. The $\Delta b = 2$ interaction is referred to as *absorptive* and the pair of $\Delta b = 1$ interactions *dispersive*. The use of flavor to specify $\mathcal{H}_{\rm mix}$ allows us to take advantage of the strict flavor changing selection rule and facilitates a transition to a quark level description.

3.2 Quark level

To transition to a description of meson mixing in which quarks, not mesons, are the fundamental degrees of freedom, we express $|B^0\rangle$, $|\bar{B}^0\rangle$, $\mathcal{H}^{\Delta b=2}$, and $\mathcal{H}^{\Delta b=1}$ as composite operators. We take $|B_q^0\rangle = \bar{b}\gamma_5 q|\Omega\rangle$, and similarly for $|\bar{B}^0\rangle$, with $|\Omega\rangle$ the vacuum. Note the composite operator $\bar{b}\gamma_5 q$ has the quantum numbers of the B_q^0 meson. The $\Delta b = 2$ term in the expansion of $\mathcal{H}_{\rm mix}$ is the four quark operator $(\bar{b}\Gamma_2 q)_y\ (\bar{b}\Gamma_1 q)_x$ that annihilates a $\bar{B}_q^0$ at spacetime point x then creates a B_q^0 meson at y. The $\Delta b = 1$ term involves four quark operators like $(\bar{q}''\Gamma_2 q')_y\ (\bar{b}\Gamma_1 q)_x$ that annihilate a $\bar{B}_q^0$ meson at x then create an intermediate state $q'\bar{q}''$ meson at y. The SM spin structures Γ_i are fixed by the $V - A$ structure of the weak interaction. The local, four quark, BSM mixing operators will be discussed in detail in Ch. 4.

For a $\Delta b = n$ interaction, the OPE yields

$$\langle B^0|\mathcal{H}^{\Delta b=n}|\bar{B}^0\rangle = \sum_i C_i^{(n)}(\mu)\langle B^0|\mathcal{O}_i^{\Delta b=n}(\mu)|\bar{B}^0\rangle + \cdots, \tag{3.11}$$

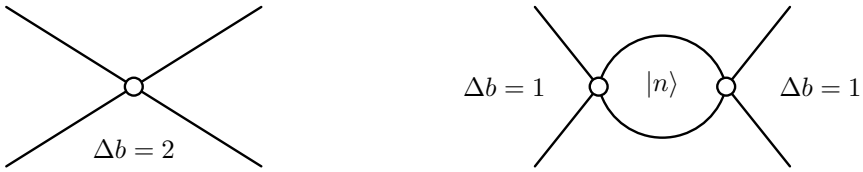


Figure 3.2: Diagrams representing the effective theory of mixing at the quark level via (*left*) a $\Delta b = 2$ interaction and (*right*) two $\Delta b = 1$ interactions with intermediate state $|n\rangle$.

to leading order, *ie.* a mixing operator, $\mathcal{O}_i$, of mass dimension six. The energy scale at which the quantities are evaluated is μ and the ellipses contain matrix elements of mixing operators of dimension greater than six. We will neglect their effects.

Two disparate scales of physics in B mixing factorize under the OPE [25]. The high energy, model-dependent physics of flavor changing interactions occurs at a scale set by the mass of the particle responsible for mediating the interactions, at least as large as the scale of EWSB, ~ 100 GeV. This physics is characterized by, generally, perturbatively calculable Wilson coefficients $C_i^{(n)}(\mu)$. The low energy physics of hadronization occurs at the scale $\Lambda_{QCD} \sim 500$ MeV, and is characterized by the hadronic mixing matrix elements. Under the OPE Eq. (3.10) becomes

$$M_{12} - \frac{i}{2}\Gamma_{12} = \sum_i C_i^{(2)}(\mu)\langle B^0|\mathcal{O}_i^{\Delta b=2}(\mu)|\bar{B}^0\rangle + \sum_n \sum_{j,k} \frac{C_j^{(1)}(\mu)C_k^{(1)}(\mu)\langle B^0|\mathcal{O}_j^{\Delta b=1}(\mu)|n\rangle\langle n|\mathcal{O}_k^{\Delta b=1}(\mu)|\bar{B}^0\rangle}{E_n - M_B}, \tag{3.12}$$

depicted graphically in Fig. 3.2.

In SM B mixing, the GIM mechanism results in intermediate states composed of top quarks or $W^{\pm}$'s and the denominator of the $\Delta b = 1$ term in Eq.'s (3.10) and (3.12) is sufficiently large to allow the long distance contribution from intermediate states to be neglected. This long distance piece corresponds to mixing via intermediate states whose presence in the analysis is due to of our inclusion of $i\Gamma/2$ in the Hamiltonian, Eq. (3.1). Generically for B mixing, $|M_{12}| \sim 10^3\ |\Gamma_{12}|$ [5]. Therefore, we can safely neglect the contributions to B mixing from $\Delta b = 1$ interactions and write

$$\Delta M = 2|M_{12}|\left[1 + \mathcal{O}\left(\left|\frac{\Gamma_{12}}{M_{12}}\right|^2\right)\right], \tag{3.13}$$

where, to a very good approximation, M_{12} is given by

$$M_{12} = \sum_i C_i^{(2)}(\mu) \langle B^0 | \mathcal{O}_i^{\Delta b=2}(\mu) | \bar{B}^0 \rangle. \tag{3.14}$$

3.3 Role of lattice QCD

For the SM we have [26]

$$\Delta M_q = \left(\frac{G_F^2 M_W^2 S_0}{4\pi^2} \right) \eta_B(\mu) \left| V_{tb} V_{tq}^* \right| \langle B_q^0 | \mathcal{O}^{\Delta b=2}(\mu) | \bar{B}_q^0 \rangle \tag{3.15}$$

where S_0 is the Inami–Lim function [27] and η_B encodes perturbative QCD corrections. This expression highlights the role played by lattice QCD in B mixing calculations. The experimental determination of ΔM_q is combined with known, or perturbatively calculable, quantities and a lattice calculation of the hadronic mixing matrix element to allow extraction of the CKM matrix elements. For generic extensions beyond the SM we have

$$\Delta M = 2 \left| \sum_i C_i^{(2)}(\mu) \langle B^0 | \mathcal{O}_i^{\Delta b=2}(\mu) | \bar{B}^0 \rangle \right|, \tag{3.16}$$

and a similar game is played. A new physics model is placed on the chopping block of comparison with experiment. The Wilson coefficients are calculated from the model and combined with a lattice determination of the hadronic mixing matrix elements to allow the comparison. The better the precision of the hadronic mixing matrix elements, the more of the model's parameter space can be chopped off.

This work calculates all possible BSM hadronic mixing matrix elements, $\langle B^0 | \mathcal{O}_i^{\Delta b=2}(\mu) | \bar{B}^0 \rangle$.

Chapter 4

The mixing operators

In the discussion that follows, we generalize to heavy quark, Q, and light quark, q, as the description is relevant to the mixing of B_q^0, K^0, and D^0. The generic heavy-light meson is $|X^0\rangle$. The effective four-quark, $\Delta Q = 2$ interactions are current-current interactions of two heavy-light quark bilinears. We consider the set of all such interactions that are Lorentz invariant color singlets. Examples are the composite operators $\overline{Q}_i^\alpha \gamma_{\mu,ij} L_{jk} q_k^\alpha \; \overline{Q}_r^\beta \gamma_{\mu,rs} L_{st} q_t^\beta$, referred to as $\mathcal{O}_1$, and $\overline{Q}_i^\alpha L_{ij} q_j^\beta \; \overline{Q}_r^\beta L_{rs} q_s^\alpha$, or $\mathcal{O}_3$. Arabic indices refer to spin and Greek to color. When performing Wick contractions on the quark fields to express the matrix elements of these operators in terms of heavy and light quark propagators it will be necessary to keep track of which quark fields are heavy, which are light, and where spin indices go. Until then things can be simplified a bit by using Takahashi notation [28] where quark fields and spin indices are dropped, color indices are traded for square or curved brackets, and bracket pairing represents pairing of color indices. In this notation $\mathcal{O}_1$ and $\mathcal{O}_3$ become

$$\begin{aligned} \mathcal{O}_1 &= \overline{Q}_i^\alpha \gamma_{\mu,ij} L_{jk} q_k^\alpha \; \overline{Q}_r^\beta \gamma_{\mu,rs} L_{st} q_t^\beta \longrightarrow (\gamma_\mu L)[\gamma_\mu L], \text{ and} \\ \mathcal{O}_3 &= \overline{Q}_i^\alpha L_{ij} q_j^\beta \; \overline{Q}_r^\beta L_{rs} q_s^\alpha \longrightarrow (L)[L). \end{aligned} \tag{4.1}$$

Besides saving room, this notation has the advantage that it emphasizes the two features that distinguish the operators – the spin structure of each bilinear and how color singlets are formed.

To generate all possible Lorentz invariant, color singlet, four-quark operators, we first build all possible Lorentz invariant spin structure pairs (a space that, in four space-time dimensions, is spanned by ten elements [29]) then consider color. The space of all possible single spin structures (*ie.* the space of 4×4 complex matrices) is spanned by the chiral basis[1] $\{R, L, \gamma_\mu L, \gamma_\mu R, \sigma_{\mu\nu}\}$. Forming all possible Lorentz invariant tensor products of chiral basis elements yields

$$\{L\otimes L,\; L\otimes R,\; R\otimes L,\; R\otimes R,\; \gamma_\mu L\otimes\gamma_\mu L,\; \gamma_\mu L\otimes\gamma_\mu R,\; \gamma_\mu R\otimes\gamma_\mu L,\; \gamma_\mu R\otimes\gamma_\mu R,\; \epsilon_{\mu\nu\rho\tau}\,\sigma_{\mu\nu}\otimes\sigma_{\rho\tau},\; \sigma_{\mu\nu}\otimes\sigma_{\mu\nu}\}. \tag{4.2}$$

[1] The Dirac basis, $\{1, \gamma_5, \gamma_\mu, \gamma_\mu\gamma_5, \sigma_{\mu\nu}\}$ (*cf.* Section 3.4 of [30]) is perhaps more common, but the chiral basis is a more convenient choice given the chiral nature of the current-current interactions of the bilinears.

Next, we consider color. For each of these pairs there are two ways to form color singlets. Either each bilinear is a color singlet as in $\mathcal{O}_1$ or color indices from one bilinear are tied to the other as in $\mathcal{O}_3$ (*eg.* $L \otimes L \longrightarrow (L)[L]$, $(L][L)$). This gives 20 composite operators:

$$\begin{aligned}
\mathcal{O}_1 &= (\gamma_\mu L)[\gamma_\mu L] & \mathcal{O}_2 &= (L)[L] \\
\mathcal{O}_3 &= (L][L) & \mathcal{O}_4 &= (L)[R] \\
\mathcal{O}_5 &= (L][R) & \mathcal{O}_6 &= (\gamma_\mu L)[\gamma_\mu R] \\
\mathcal{O}_7 &= (\gamma_\mu L][\gamma_\mu R) & \mathcal{O}_8 &= (R][R] \\
\mathcal{O}_9 &= (R][R) & \mathcal{O}_{10} &= (\gamma_\mu R)[\gamma_\mu R] \\
\mathcal{O}_{11} &= (\gamma_\mu R][\gamma_\mu R) & \mathcal{O}_{12} &= (\gamma_\mu L][\gamma_\mu L) \\
\mathcal{O}_{13} &= (\sigma_{\mu\nu})[\sigma_{\mu\nu}] & \mathcal{O}_{14} &= (\sigma_{\mu\nu}][\sigma_{\mu\nu}) \\
\mathcal{O}_{15} &= (\gamma_\mu R)[\gamma_\mu L] & \mathcal{O}_{16} &= (\gamma_\mu R][\gamma_\mu L) \\
\mathcal{O}_{17} &= (R)[L] & \mathcal{O}_{18} &= (R][L) \\
\mathcal{O}_{19} &= \epsilon_{\mu\nu\rho\tau}(\sigma_{\mu\nu})[\sigma_{\rho\tau}] & \mathcal{O}_{20} &= \epsilon_{\mu\nu\rho\tau}(\sigma_{\mu\nu}][\sigma_{\rho\tau})
\end{aligned} \tag{4.3}$$

The mixing matrix elements generated by these 20 operators are not, however, linearly independent. There are relationships between the operators themselves (Fierz transformations and re-arrangement by commuting quark fields) and relationships between the matrix elements of the operators (parity and time-reversal).

After studying the symmetries we'll find there are only five independent hadronic mixing matrix elements, the so-called SUSY basis. The remaining 15 matrix elements can be simply expressed in terms of these five. The derived relationships give us multiple ways to generate physically equivalent data – data that may be statistically distinguishable. In these cases, we can combine multiple sets of data, effectively increasing the statistics associated with the Monte Carlo evaluation of the path integral used to evaluate the mixing matrix elements.

In the remainder of this chapter we'll derive relationships between the mixing matrix elements, and evaluate the level of agreement between physically equivalent data. We'll find data sets that are physically equivalent due to symmetry, yet are distinguishable at a level that makes it advantageous to average them.

4.1 Symmetries of operators

4.1.1 Re-arrangement by commuting quark fields

For a generic mixing operator, the quark fields can be commuted to write $\overline{Q}_i^\alpha(\Gamma_1)_{ij}q_j^\beta\ \overline{Q}_r^\beta(\Gamma_2)_{rs}q_s^\alpha = \overline{Q}_r^\beta(\Gamma_2)_{rs}q_s^\alpha\ \overline{Q}_i^\alpha(\Gamma_1)_{ij}q_j^\beta$. This re-arrangement, trivially satisfied when $\Gamma_1 = \Gamma_2$, allows us to equate the following operators:

$$\begin{aligned} \mathcal{O}_4 &\overset{\mathcal{R}}{=} \mathcal{O}_{17} & \mathcal{O}_5 &\overset{\mathcal{R}}{=} \mathcal{O}_{18} \\ \mathcal{O}_6 &\overset{\mathcal{R}}{=} \mathcal{O}_{15} & \mathcal{O}_7 &\overset{\mathcal{R}}{=} \mathcal{O}_{16} \end{aligned} \tag{4.4}$$

where $\overset{\mathcal{R}}{=}$ indicates equality by re-arrangement. Operators equal by re-arrangement generate identical[2] data.

4.1.2 Chiral Fierz identities

Fierz symmetries provide additional relationships among the operators. For Γ in the chiral basis and $\tilde{\Gamma}$ in its dual basis, $\{R, L, \gamma_\mu R, \gamma_\mu L, \frac{1}{2}\sigma_{\mu\nu}\}$, the chiral Fierz identities are [31]

$$\begin{aligned} (\Gamma_A)[\Gamma_B] &= -\sum_{C,D}\frac{1}{4}Tr[\Gamma_A\tilde{\Gamma}_C\Gamma_B\tilde{\Gamma}_D](\Gamma_D)[\Gamma_C), \text{ and} \\ (\Gamma_A][\Gamma_B) &= -\sum_{C,D}\frac{1}{4}Tr[\Gamma_A\tilde{\Gamma}_C\Gamma_B\tilde{\Gamma}_D](\Gamma_D)[\Gamma_C]. \end{aligned} \tag{4.5}$$

Using the Mathematica program FeynCalc [32] to aid in evaluating traces of gamma matrices, we performed Fierz transformations on these 20 operators, generating the following relationships:

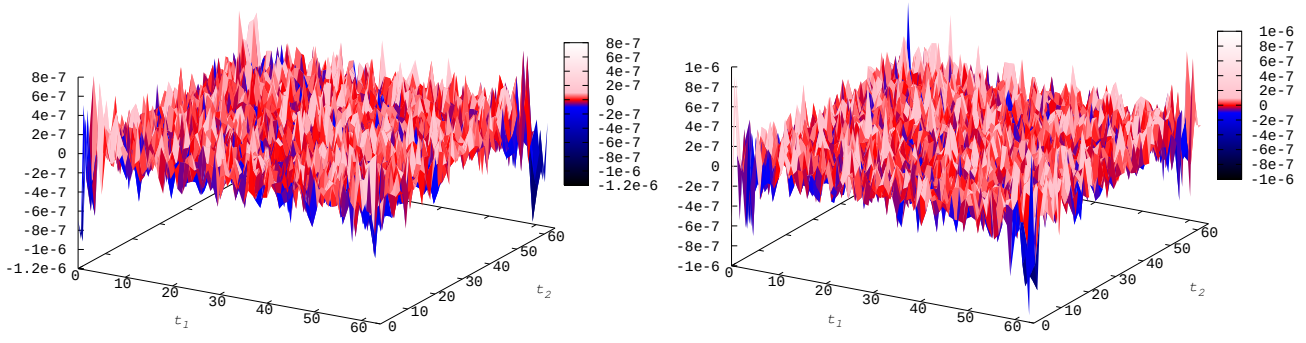
$$\begin{aligned} \mathcal{O}_1 &\overset{\mathcal{F}}{=} \mathcal{O}_{12} & 8\mathcal{O}_2 &\overset{\mathcal{F}}{=} -4\mathcal{O}_3 - \mathcal{O}_{14} + \mathcal{O}_{20} \\ 8\mathcal{O}_3 &\overset{\mathcal{F}}{=} -4\mathcal{O}_2 - \mathcal{O}_{13} + \mathcal{O}_{19} & 2\mathcal{O}_4 &\overset{\mathcal{F}}{=} -\mathcal{O}_{16} \\ 2\mathcal{O}_5 &\overset{\mathcal{F}}{=} -\mathcal{O}_{15} & \mathcal{O}_6 &\overset{\mathcal{F}}{=} -2\mathcal{O}_{18} \\ \mathcal{O}_7 &\overset{\mathcal{F}}{=} -2\mathcal{O}_{17} & 8\mathcal{O}_8 &\overset{\mathcal{F}}{=} -4\mathcal{O}_9 - \mathcal{O}_{14} - \mathcal{O}_{20} \\ 8\mathcal{O}_9 &\overset{\mathcal{F}}{=} -4\mathcal{O}_8 - \mathcal{O}_{13} - \mathcal{O}_{19} & \mathcal{O}_{10} &\overset{\mathcal{F}}{=} \mathcal{O}_{11} \\ 2\mathcal{O}_{13} &\overset{\mathcal{F}}{=} -6\left(\mathcal{O}_3 + \mathcal{O}_9\right) + \mathcal{O}_{14} & 2\mathcal{O}_{14} &\overset{\mathcal{F}}{=} -6\left(\mathcal{O}_2 + \mathcal{O}_8\right) + \mathcal{O}_{13} \\ 2\mathcal{O}_{19} &\overset{\mathcal{F}}{=} 6\left(\mathcal{O}_3 - \mathcal{O}_9\right) + \mathcal{O}_{20} & 2\mathcal{O}_{20} &\overset{\mathcal{F}}{=} 6\left(\mathcal{O}_2 - \mathcal{O}_8\right) + \mathcal{O}_{19} \end{aligned} \tag{4.6}$$

where $\overset{\mathcal{F}}{=}$ indicates equality by Fierz identity.

To determine the utility of the Fierz transformations in generating additional data, the data for the matrix element of each operator and the matrix element of its Fierz transformation are compared. The

[2]Comparison of data for $\mathcal{O}_5$ and $\mathcal{O}_{18}$ on the coarse 0.1ms ensemble revealed a single point in the t_1, t_2 plane where the difference is non-zero. However, the disagreement at this point is at the level of $10^{-15}\sigma$.

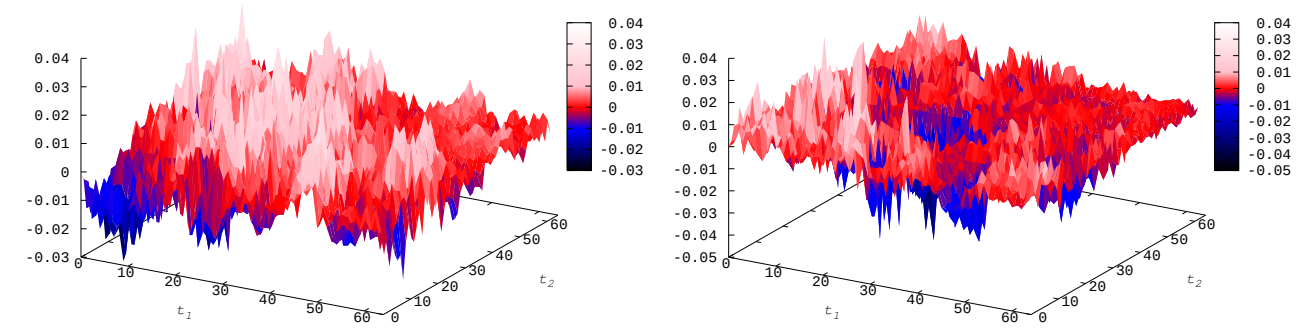
comparison is carried out by calculating the Z-statistic, described in the Appendix, Section E.4. Data for $\mathcal{O}_1$ is identical to data for $\mathcal{F}(\mathcal{O}_1)$ and data for $\mathcal{O}_{10}$ is identical to that for $\mathcal{F}(\mathcal{O}_{10})$. Data generated by operators $\mathcal{O}_{4-7,13,14}$ differs at the $10^{-6}\sigma$ level, as shown in Fig. 4.1, from the corresponding data generated by their Fierz transformations. Data generated by operators $\mathcal{O}_{2,3,8,9,19,20}$ differ at the level of $10^{-2}\sigma$, as shown in Fig. 4.2, from the Fierz transformed data. Though not identically satisfied, as in the case of re-arrangement symmetries, Fierz symmetries are satisfied to such a high degree that related data are essentially equivalent and the achievable statistical gain is insignificant.



(a) Difference in σ between $\mathcal{O}_4$ and $\mathcal{F}(\mathcal{O}_4)$.

(b) Difference in σ between $\mathcal{O}_{13}$ and $\mathcal{F}(\mathcal{O}_{13})$.

Figure 4.1: The difference between data generated for operators $\mathcal{O}_{4,13}$ and data generated for their Fierz transformations is nearly exact, with differences around one-millionth of a σ. The same is true for $\mathcal{O}_{5-7,14}$.



(a) Difference in σ between $\mathcal{O}_2$ and $\mathcal{F}(\mathcal{O}_2)$.

(b) Difference in σ between $\mathcal{O}_9$ and $\mathcal{F}(\mathcal{O}_9)$.

Figure 4.2: The difference between data generated for operators $\mathcal{O}_{2,9}$ and data generated for their Fierz transformations differs by $\sim 10^{-2}\sigma$. The same is true for $\mathcal{O}_{3,8,19,20}$.

4.2 Symmetries of mixing matrix elements

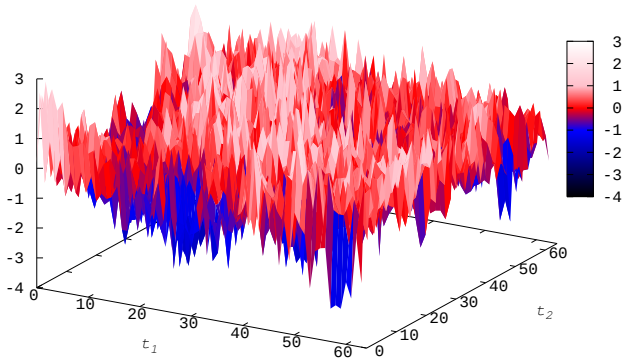
The hadronic mixing matrix elements include only the effects of QCD. Therefore, the hadronic mixing matrix elements separately conserve parity and time-reversal. To determine the effects of parity, P, and time-reversal, T, on the mixing matrix elements we must start with the transformations of quark and gauge fields, build up to transformations of quark propagators, and finally develop the transformation laws for the mixing matrix elements. This work is presented in App. B, with the transformations of the mixing matrix elements under P and T given in Eq's. (B.7) and (B.8), respectively.

4.2.1 Parity

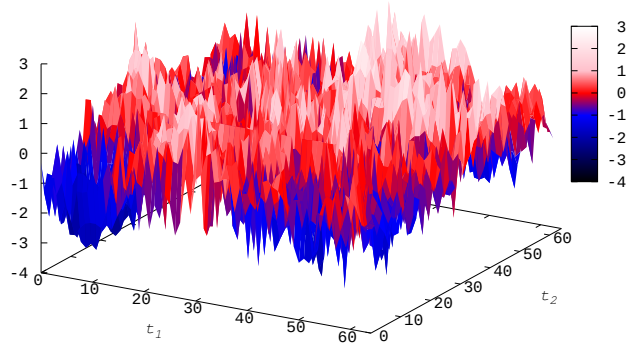
With the exception of those operators containing a $\sigma_{\mu\nu}$, the effect of a parity transformation is effectively the interchange $L \leftrightarrow R$. Using Eq. (B.7), the parity transformations for each of the $\mathcal{O}_{1-20}$ are:

$$
\begin{aligned}
P: \qquad \langle X^0|\mathcal{O}_1([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} \langle X^0|\mathcal{O}_{10}([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_2([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} \langle X^0|\mathcal{O}_8([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_3([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} \langle X^0|\mathcal{O}_9([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_4([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} \langle X^0|\mathcal{O}_{17}([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_5([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} \langle X^0|\mathcal{O}_{18}([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_6([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} \langle X^0|\mathcal{O}_{15}([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_7([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} \langle X^0|\mathcal{O}_{16}([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_{11}([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} \langle X^0|\mathcal{O}_{12}([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_{13}([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} \langle X^0|\mathcal{O}_{13}([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_{14}([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} \langle X^0|\mathcal{O}_{14}([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_{19}([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} -\langle X^0|\mathcal{O}_{19}([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p} \\
\langle X^0|\mathcal{O}_{20}([U])|\bar{X}^0\rangle_{x_1,x_2} &\stackrel{P}{=} -\langle X^0|\mathcal{O}_{20}([U]^p)|\bar{X}^0\rangle_{x_1^p,x_2^p}
\end{aligned}
\tag{4.7}
$$

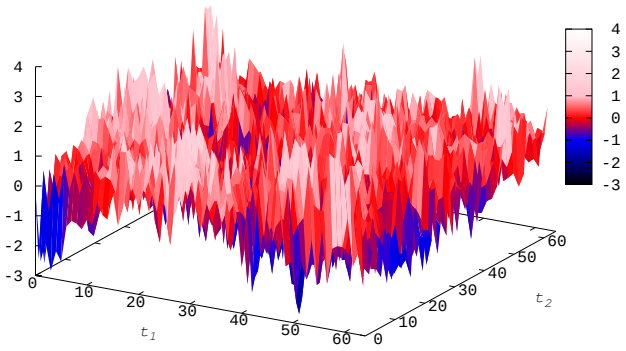
where $\stackrel{P}{=}$ indicates equality by parity symmetry. The parity transformations for $\mathcal{O}_{4-7}$ are equivalent to the re-arrangement symmetries of Eq. (4.4), while matrix elements of operators $\mathcal{O}_{13}$ and $\mathcal{O}_{14}$ are symmetric under parity. Therefore, the parity transformations of $\mathcal{O}_{4-7,13,14}$ generate physically equivalent but statistically indistinguishable data. To assess the utility of parity transformations of the remaining operators to generate useful additional data, the data for matrix elements of the operators and their parity transformations are



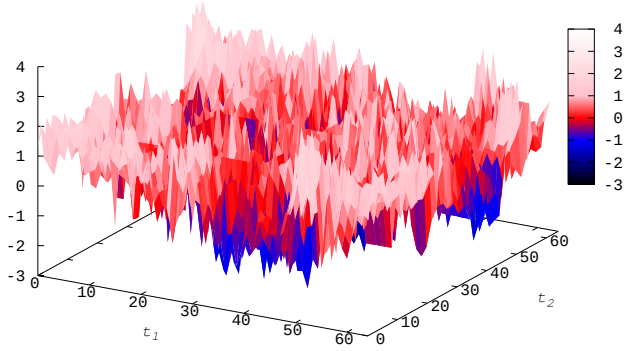
(a) Difference in σ between $\mathcal{O}_1$and $P(\mathcal{O}_1)$.



(b) Difference in σ between $\mathcal{O}_3$ and $P(\mathcal{O}_3)$.



(c) Difference in σ between $\mathcal{O}_{11}$ and $P(\mathcal{O}_{11})$.



(d) Difference in σ between $\mathcal{O}_{20}$ and $P(\mathcal{O}_{20})$. By parity, $\mathcal{O}_{20} = 0$.

Figure 4.3: The difference between data generated for operators $\mathcal{O}_{1,3,11,20}$ and data generated for their parity transformations differs by $\lesssim 4\sigma$. The same is true for $\mathcal{O}_{2,19}$.

compared via the Z-statistic, as described in the App. E.4. Under parity $\mathcal{O}_{1-3,11,19,20}$ generate data that are physically equivalent, but differ by $\lesssim 4\sigma$. This indicates that, for these operators, we can effectively increase our statistics by averaging $\mathcal{O}$ and $P(\mathcal{O})$.

Role of parity equivalent data

Here we discuss the manner in which parity transformations provide additional data. Let $[U]$ be a gauge configuration, $[U]^p$ its parity transformation, and $\mathcal{U}$ the ensemble of gauge configurations used to generate correlation functions. If $\mathcal{U}$ is not closed under parity (*ie.* $\exists$ a $[U] \in \mathcal{U}$ such that $[U]^p \notin \mathcal{U}$) then parity is not a symmetry of $\mathcal{U}$. Inclusion of parity-equivalent data, however, closes $\mathcal{U}$ under parity. To see this, consider

the parity transformation of $\mathcal{O}_3$ from Eq. (4.7),

$$\begin{aligned} \langle X^0|\mathcal{O}_3([U])|\bar{X}^0\rangle &\overset{P}{=} \langle X^0|\mathcal{O}_9([U]^p)|\bar{X}^0\rangle \\ \Rightarrow \langle X^0|\mathcal{O}_9([U])|\bar{X}^0\rangle &\overset{P}{=} \langle X^0|\mathcal{O}_3([U]^p)|\bar{X}^0\rangle. \end{aligned} \tag{4.8}$$

By using the parity averaged matrix element

$$\begin{aligned} \langle \mathcal{O}_3\rangle &= \langle X^0|\ \mathcal{O}_3([U]) + \mathcal{O}_9([U])\ |\bar{X}^0\rangle\ /\ 2 \\ &= \langle X^0|\ \mathcal{O}_3([U]) + \mathcal{O}_3([U]^p)\ |\bar{X}^0\rangle\ /\ 2 \end{aligned} \tag{4.9}$$

we include $[U]^p$ for each $[U] \in \mathcal{U}$, effectively expanding our ensemble from $\mathcal{U} = \{[U]\}$ to $\mathcal{U}^p = \{[U], [U]^p\}$. $\mathcal{U}^p$ is closed under parity, enforcing parity symmetry of the expanded ensemble of gauge configurations, $\mathcal{U}^p$.

We have seen evidence of a systematic, $< 1\sigma$, shift in data generated with $\mathcal{U}$ as compared to data generated with $\mathcal{U}^p$. The possibility of such a bias, induced by an ensemble of parity violating gauge configurations, provides additional motivation to utilize parity equivalent data sets.

4.2.2 Time-reversal

Time-reversal transformations for each of the $\mathcal{O}_{1-20}$ are calculated using Eq. (B.8):

$$\begin{aligned} T: \quad \langle X^0|\mathcal{O}_1([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} \langle X^0|\mathcal{O}_{10}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_2([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} \langle X^0|\mathcal{O}_8([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_3([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} \langle X^0|\mathcal{O}_9([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_4([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} \langle X^0|\mathcal{O}_{17}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_5([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} \langle X^0|\mathcal{O}_{18}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_6([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} \langle X^0|\mathcal{O}_{15}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_7([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} \langle X^0|\mathcal{O}_{16}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_{11}([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} \langle X^0|\mathcal{O}_{12}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_{13}([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} \langle X^0|\mathcal{O}_{13}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_{14}([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} \langle X^0|\mathcal{O}_{14}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_{19}([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} -\langle X^0|\mathcal{O}_{19}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \\ \langle X^0|\mathcal{O}_{20}([U])|\bar{X}^0\rangle_{x_1,x_2} &\overset{T}{=} -\langle X^0|\mathcal{O}_{20}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \end{aligned} \tag{4.10}$$

Note that time-reversal couples the same pairs of operators as parity and there are, therefore, no new relationships to exploit. We do, however, utilize a combined parity and time-reversal transformation in a process called folding and discussed in Sec. 6.5.

4.3 Exploiting parity and Fierz symmetries: *increasing statistics*

Imposing parity (we could alternatively use time-reversal), Fierz, and re-arrangement symmetries – and using the shorthand of simply referring to the operator instead of the matrix element to which parity symmetry actually applies, $\mathcal{O}_{6-20}$ can be eliminated in favor of $\mathcal{O}_{1-5}$:

$$\begin{array}{lll}
\mathcal{O}_6 \overset{\mathcal{RF}}{=} -2\mathcal{O}_5 & \mathcal{O}_{11} \overset{\mathcal{FP}}{=} \mathcal{O}_1 & \mathcal{O}_{16} \overset{\mathcal{F}}{=} -2\mathcal{O}_4 \\
\mathcal{O}_7 \overset{\mathcal{RF}}{=} -2\mathcal{O}_4 & \mathcal{O}_{12} \overset{\mathcal{F}}{=} \mathcal{O}_1 & \mathcal{O}_{17} \overset{\mathcal{R}}{=} \mathcal{O}_4 \\
\mathcal{O}_8 \overset{P}{=} \mathcal{O}_2 & \mathcal{O}_{13} \overset{\mathcal{FP}}{=} -8\mathcal{O}_3 - 4\mathcal{O}_2 & \mathcal{O}_{18} \overset{\mathcal{R}}{=} \mathcal{O}_5 \\
\mathcal{O}_9 \overset{P}{=} \mathcal{O}_3 & \mathcal{O}_{14} \overset{\mathcal{FP}}{=} -8\mathcal{O}_2 - 4\mathcal{O}_3 & \mathcal{O}_{19} \overset{P}{=} 0 \\
\mathcal{O}_{10} \overset{P}{=} \mathcal{O}_1 & \mathcal{O}_{15} \overset{\mathcal{F}}{=} -2\mathcal{O}_5 & \mathcal{O}_{20} \overset{P}{=} 0
\end{array} \tag{4.11}$$

Relationships that don't rely on re-arrangement or Fierz symmetries allow us to effectively generate additional statistics for $\mathcal{O}_{1-5}$ by averaging the equivalent operators. Denoting by $\langle\mathcal{O}\rangle$ the statistically improved mixing matrix element after averaging equivalent, yet statistically distinguishable, operators:

$$\begin{aligned}
\langle\mathcal{O}_1\rangle &= (\mathcal{O}_1 + \mathcal{O}_{10})/2 \\
\langle\mathcal{O}_2\rangle &= (\mathcal{O}_2 + \mathcal{O}_8)/2 \\
\langle\mathcal{O}_3\rangle &= (\mathcal{O}_3 + \mathcal{O}_9)/2
\end{aligned} \tag{4.12}$$

4.4 The SUSY basis of operators

The space of effective four-quark, mass dimension six, $\Delta Q = 2$, hadronic mixing matrix elements in and beyond the SM is spanned by the matrix elements of the SUSY basis of operators [33]. The name, SUSY basis, is purely historical.

The SUSY basis of operators, with quark fields and indices explicit, are the $\mathcal{O}_{1-5}$ found above:

$$\begin{aligned}
\mathcal{O}_1 &= \overline{Q}_i^\alpha \gamma_{\mu,ij} L_{jk} q_k^\alpha \; \overline{Q}_r^\beta \gamma_{\mu,rs} L_{st} q_t^\beta \\
\mathcal{O}_2 &= \overline{Q}_i^\alpha L_{ij} q_j^\alpha \; \overline{Q}_r^\beta L_{rs} q_s^\beta \\
\mathcal{O}_3 &= \overline{Q}_i^\alpha L_{ij} q_j^\beta \; \overline{Q}_r^\beta L_{rs} q_s^\alpha \\
\mathcal{O}_4 &= \overline{Q}_i^\alpha L_{ij} q_j^\alpha \; \overline{Q}_r^\beta R_{rs} q_s^\beta \\
\mathcal{O}_5 &= \overline{Q}_i^\alpha L_{ij} q_j^\beta \; \overline{Q}_r^\beta R_{rs} q_s^\alpha
\end{aligned} \tag{4.13}$$

The hadronic mixing matrix elements of operators $\mathcal{O}_{1,2,3}$ are needed for SM calculations. $\mathcal{O}_1$ has the SM's required $V-A \times V-A$ structure and operators $\mathcal{O}_{2,3}$ enter via operator mixing under renormalization. Therefore, generic extensions beyond the SM require only the hadronic matrix elements of two operators, $\mathcal{O}_{4,5}$, and the seemingly sweeping statement ending Sec. 3.3 can be focused considerably. This work calculates the BSM hadronic mixing matrix elements, $\langle B_q^0 | \mathcal{O}_4 | \bar{B}_q^0 \rangle$ and $\langle B_q^0 | \mathcal{O}_5 | \bar{B}_q^0 \rangle$, for $q = d, s$.

Chapter 5

The lattice calculation

5.1 Lattice basics

We begin in the continuum with the Euclidean, path integral formulation of the QCD vacuum expectation value (VEV) of, a typically composite, operator A,

$$\langle\Omega|A|\Omega\rangle = \frac{\int[d\psi][d\bar{\psi}][dG_\mu]\ A[\psi,\bar{\psi},G_\mu]\ e^{-S_{QCD}[\psi,\bar{\psi},G_\mu]}}{\int[d\psi][d\bar{\psi}][dG_\mu]\ e^{-S_{QCD}[\psi,\bar{\psi},G_\mu]}}, \tag{5.1}$$

where

$$S_{QCD}[\psi,\bar{\psi},G_\mu] = \int d^4x\ \bar{\psi}(\not{D}+m)\psi + S[G_\mu], \tag{5.2}$$

and G_μ are the gluons of the QCD gauge group, $SU(3)_C$. The perturbative approach expands the integrand in powers of the strong coupling to yield analytically calculable integrals, a valid approach provided the coupling remains small enough to justify the expansion – a condition not met at energies below or near Λ_{QCD}. In the lattice approach we first perform the Berezin integration over quark fields,

$$\langle\Omega|A|\Omega\rangle = \frac{\int[dG_\mu]\ A[(\not{D}+m)^{-1},G_\mu]\ e^{-S[G_\mu]+\ln\left[det(\not{D}+m)\right]}}{\int[dG_\mu]\ e^{-S[G_\mu]+\ln\left[det(\not{D}+m)\right]}}, \tag{5.3}$$

where $(\not{D}+m)^{-1}$ is the quark propagator[1]. The gauge field integration that remains is not (yet) solvable analytically and, as we've discussed, can only be tackled with perturbation theory at energy scales for which the strong coupling is small. This is where lattice QCD enters. We discretize space-time, providing explicit UV and IR regulation, and approximate $SU(3)_C$ gauge group integration by a Monte Carlo sum over possible gauge field configurations, allowing numerical evaluation of the VEV of A.

Discretization requires transcription of fields and operators to the lattice. We place quark fields on lattice sites and represent gauge fields by link variables, $U_{x,\mu} \in SU(3)_C$, connecting those sites. Our gauge action is a functional of the gauge links, $S[U_\mu]$. Gauge covariant derivatives are built from finite differences and

[1] Quark fields appearing in A are combined, via Wick contraction, to produce propagators.

the parallel transport action of $U_{\mu,x}$,

$$\begin{aligned} \psi(x) &\rightarrow \psi_x \\ G_\mu(x) &\rightarrow U_{x,\mu} = \mathcal{P}\exp\left\{ig\int_x^{x+a\hat{\mu}} G_\nu(y)dy_\nu\right\} \\ D_\mu\psi(x) &\rightarrow \Delta_\mu\psi_x = \frac{1}{2a}\left(U_{\mu,x}\psi_{x+a\hat{\mu}} - U^\dagger_{\mu,x-a\hat{\mu}}\psi_{x-a\hat{\mu}}\right) \end{aligned} \tag{5.4}$$

where a is the lattice spacing and g the bare lattice coupling.

The collection of gauge field configurations used to evaluate the Monte Carlo sums are generated with a Boltzmann-like probability distribution $\exp(-S[U_i] + \ln\left[det(\not{\Delta} + m)\right])$ from Eq. (5.3), a process referred to as importance sampling. On the lattice, Eq. (5.3) is then

$$\langle\Omega|A|\Omega\rangle \approx \frac{1}{N}\sum_{i=1}^{N} A\Big(\{U_i\}, (\not{\Delta} + m)^{-1}\Big), \tag{5.5}$$

allowing the calculation of non-perturbative processes associated with arbitrarily strong coupling.

5.2 Extrapolating to reality

The value for $\langle\Omega|A|\Omega\rangle$ obtained from Eq. 5.5 is specific to a discrete universe with lattice spacing a of Eq. 5.4. To relate this result to reality we must extrapolate from the lattice spacings used in our simulations (approximately 0.09 - 0.12 fm) to zero. In addition, the quark masses used in the simulation are typically too heavy (in the case of an up or down quark) or only approximately equal to (in the case of the strange quark) their physical counterparts. As a result, we simulate over a range of light quark masses to permit extrapolation to the physical light quark mass and interpolation to the physical strange quark mass. The simultaneous extrapoltion to physical light quark mass and the continuum is outlined in more detail in Ch. 8. Table 5.1 summarizes the ensembles used in this analysis.

5.3 Simulation details

The MILC gauge configurations [34] used in this simulation have Symanzik [35, 36, 37] tadpole improved [38] gluons, with lattice artifacts beginning at $\mathcal{O}(\alpha_s a^2)$ and $\mathcal{O}(a^4)$. App. C outlines this action and its improvements.

Sea quarks are incorporated in the gauge configurations through the determinant in Eq. (5.3), where we use the fourth-root trick [39] to reduce the effect of doublers in the staggered, $\mathcal{O}(a^2)$ and tadpole improved

$\approx a$ [fm]	$\left(\frac{L}{a}\right)^3 \times \frac{T}{a}$	$\beta = \frac{10}{g^2}$	κ	am_l	am_s	am_x	$N_{\rm cfg}$
0.12	$24^3\times64$	6.760	0.0901	0.005	0.050	(0.0050, 0.0070, 0.0100, 0.0200, 0.0300, 0.0349, 0.0415, 0.0500)	2099
0.12	$20^3\times64$	6.760	0.0901	0.007	0.050	(0.0050, 0.0070, 0.0100, 0.0200, 0.0300, 0.0349, 0.0415, 0.0500)	2110
0.12	$20^3\times64$	6.760	0.0901	0.010	0.050	(0.0050, 0.0070, 0.0100, 0.0200, 0.0300, 0.0349, 0.0415, 0.0500)	2259
0.12	$20^3\times64$	6.790	0.0918	0.020	0.050	(0.0050, 0.0070, 0.0100, 0.0200, 0.0300, 0.0349, 0.0415, 0.0500)	2052
0.09	$40^3\times96$	7.045	0.0976	0.0031	0.0310	(0.0031, 0.0047, 0.0062, 0.0093, 0.0124, 0.0261, 0.0310)	1015
0.09	$32^3\times96$	7.085	0.0977	0.00465	0.0310	(0.0031, 0.0047, 0.0062, 0.0093, 0.0124, 0.0261, 0.0310)	984
0.09	$28^3\times96$	7.090	0.0979	0.0062	0.0310	(0.0031, 0.0047, 0.0062, 0.0093, 0.0124, 0.0261, 0.0310)	1931
0.09	$28^3\times96$	7.110	0.0982	0.0124	0.0310	(0.0031, 0.0047, 0.0062, 0.0093, 0.0124, 0.0261, 0.0310)	1996

Table 5.1: Summary of MILC ensembles and quark masses used in this analysis. The bare lattice coupling of the gauge action is set by β, *cf.* Eq. (C.3). The heavy valence quark is tuned to the physical bottom quark mass and is represented by κ, Eq. (D.8). The light and strange sea quark masses, in lattice units, are am_l and am_s, respectively. The light valence quark mass, in lattice units, is am_x. At $a \approx 0.12$ fm, the physical strange quark mass, in lattice units, is $am_s^{\rm phy} \approx 0.0349$ and at $a \approx 0.09$ fm its $am_s^{\rm phy} \approx 0.0261$. $N_{\rm cfg}$ is the number of gauge configurations used to evaluate the sum in Eq. (5.5).

asqtad action [40, 41, 42, 43]. A discussion of the difficulties associated with discretizing fermions is given in App. D, followed by a description of the asqtad action and the fourth-root trick in App. D.3.

The asqtad action is also used to generate quark propagators used in the simulation of light valence quarks. Heavy valence quarks are simulated with quark propagators generated by the Fermilab action [44], an $\mathcal{O}(a)$ improved Wilson action [45] with a heavy quark effective theory inspired interpretation of improvement coefficients. Lattice artifacts enter via the Fermilab action at $\mathcal{O}(\alpha_s a\Lambda_{QCD})$ and $\mathcal{O}(a^2\Lambda_{QCD}^2)$ while heavy quark effective theory power counting reveals heavy quark errors that enter at $\mathcal{O}(\alpha_s\Lambda_{QCD}/m_Q)$ and $\mathcal{O}(\Lambda_{QCD}^2/m_Q^2)$, where m_Q is the heavy quark mass. The Fermilab action is discussed in App. D.2.

Chapter 6

Generating Data

To generate lattice data for the hadronic mixing matrix elements of Eq. (3.16), we begin with their description in terms of composite operators with quark degrees of freedom, as discussed in Sec. 3.2, *ie.* the combination of composite operators $\langle B^\dagger \mathcal{O} \bar{B} \rangle$[1]. On the discrete spacetime of the lattice we generate data for this three point correlation function and extract the hadronic mixing matrix element from its amplitude. In the remainder of this chapter we again generalize the discussion to a generic heavy light neutral meson, X^0, made of a light and heavy quark, q and Q, respectively. The relevant combination of composite operators, for mixing operator $\mathcal{O}_N$, is then $\langle X^\dagger \mathcal{O}_N \bar{X} \rangle$, where $N \in \{1..5\}$ specifies which operator of Sec. 4.4 is responsible for the mixing.

6.1 Quark propagators

Wick contraction of quark fields in $\langle X^\dagger \mathcal{O}_N \bar{X} \rangle$ generates quark propagators, quantities that can be calculated on the lattice and used as building blocks to make correlation function data. In terms of Wick contracted fields, the light (L) and heavy (H) quark propagators are

$$L_{ij}^{\alpha\beta}(x,y) \equiv \langle q_j^\alpha(x) \overline{q}_k^\beta(y) \rangle \qquad \text{and} \qquad H_{ij}^{\alpha\beta}(x,y) \equiv \langle Q_j^\alpha(x) \overline{Q}_k^\beta(y) \rangle \tag{6.1}$$

with indices as specified in the beginning of Ch. 4. A quark propagator is built by inverting the quadratic operator in the action (*eg.* the Dirac operator, $\not{\Delta} + m$), a multi-dimensional matrix with spacetime, spin, and color indices. This matrix is numerically inverted, a costly step in data production. In practice, the generation of quark propagators is part of a large script developed by the FNAL-MILC collaboration and referred to as the "superscript".

[1] Adopting the shorthand $\langle \bullet \rangle = \langle \Omega | \bullet | \Omega \rangle$, where $|\Omega\rangle$ is the vacuum.

6.2 Building three point correlation functions

Hadronic matrix elements are extracted from the three point functions. In this section we develop the relationship that governs how quark propagators computed on the lattice are combined to generate data for the three point correlation functions. We work in the meson rest frame, create the anti-meson at x_1, place the mixing operator at the spacetime origin, and annihilate the meson at x_2. We build neutral meson states from interpolating operators $|\bar{X}^0\rangle = \overline{q}\gamma_5 Q|\Omega\rangle$ and $\langle X^0| = \langle\Omega|\overline{q}\gamma_5 Q$. The correlation function is then the time-ordered VEV

$$\langle (X^0)_{\vec{x}_2,t_2} | (\mathcal{O}_N)_{\vec{0},0} | (\bar{X}^0)_{\vec{x}_1,t_1}\rangle = \langle T\{(\overline{q}\gamma_5 Q)_{\vec{x}_2,t_2} (\mathcal{O}_N)_{\vec{0},0} (\overline{q}\gamma_5 Q)_{\vec{x}_1,t_1}\}\rangle, \tag{6.2}$$

where time ordering is ensured by taking $t_2 > 0 > t_1$. Writing $\mathcal{O}_1$, $\mathcal{O}_2$ and $\mathcal{O}_4$ (operators with color indices not mixed between bilinears) generically as $\mathcal{O}_{\text{nomix}}$ and $\mathcal{O}_3$ and $\mathcal{O}_5$ (operators with color indices mixed between bilinears) as $\mathcal{O}_{\text{mix}}$, we have

$$\mathcal{O}_{\text{nomix}} = \overline{Q}_i^\alpha \Gamma_{1,ij} q_j^\alpha \; \overline{Q}_r^\beta \Gamma_{2,rs} q_s^\beta \qquad \text{and} \qquad \mathcal{O}_{\text{mix}} = \overline{Q}_i^\alpha \Gamma_{1,ij} q_j^\beta \; \overline{Q}_r^\beta \Gamma_{2,rs} q_s^\alpha, \tag{6.3}$$

where the form of Γ_1 and Γ_2 identifies the mixing operator.

Inserting the field content of the mixing operators and mesons gives

$$\langle X^0 | \mathcal{O}_{\text{nomix}} | \bar{X}^0\rangle_{x_1,x_2} = \langle (\overline{q}_i^\alpha \gamma_{5,ij} Q_j^\alpha)_{x_2} \; (\overline{Q}_k^\beta \Gamma_{1,kl} q_l^\beta \; \overline{Q}_m^\gamma \Gamma_{2,mn} q_n^\gamma)_0 \; (\overline{q}_o^\delta \gamma_{5,op} Q_p^\delta)_{x_1}\rangle \tag{6.4}$$

and

$$\langle X^0 | \mathcal{O}_{\text{mix}} | \bar{X}^0\rangle_{x_1,x_2} = \langle (\overline{q}_i^\alpha \gamma_{5,ij} Q_j^\alpha)_{x_2} \; (\overline{Q}_k^\beta \Gamma_{1,kl} q_l^\gamma \; \overline{Q}_m^\gamma \Gamma_{2,mn} q_n^\beta)_0 \; (\overline{q}_o^\delta \gamma_{5,op} Q_p^\delta)_{x_1}\rangle. \tag{6.5}$$

There are four possible Wick contractions

$$\begin{aligned}
\langle X^0 | \mathcal{O}_{\text{nomix}} | \bar{X}^0\rangle_{x_1,x_2} \;=\; & \langle (\overline{q}_i^\alpha \gamma_{5,ij} Q_j^\alpha)_{x_2} \; (\overline{Q}_k^\beta \Gamma_{1,kl} q_l^\beta \; \overline{Q}_m^\gamma \Gamma_{2,mn} q_n^\gamma)_0 \; (\overline{q}_o^\delta \gamma_{5,op} Q_p^\delta)_{x_1}\rangle \\
& + \langle (\overline{q}_i^\alpha \gamma_{5,ij} Q_j^\alpha)_{x_2} \; (\overline{Q}_k^\beta \Gamma_{1,kl} q_l^\beta \; \overline{Q}_m^\gamma \Gamma_{2,mn} q_n^\gamma)_0 \; (\overline{q}_o^\delta \gamma_{5,op} Q_p^\delta)_{x_1}\rangle \\
& + \langle (\overline{q}_i^\alpha \gamma_{5,ij} Q_j^\alpha)_{x_2} \; (\overline{Q}_k^\beta \Gamma_{1,kl} q_l^\beta \; \overline{Q}_m^\gamma \Gamma_{2,mn} q_n^\gamma)_0 \; (\overline{q}_o^\delta \gamma_{5,op} Q_p^\delta)_{x_1}\rangle \\
& + \langle (\overline{q}_i^\alpha \gamma_{5,ij} Q_j^\alpha)_{x_2} \; (\overline{Q}_k^\beta \Gamma_{1,kl} q_l^\beta \; \overline{Q}_m^\gamma \Gamma_{2,mn} q_n^\gamma)_0 \; (\overline{q}_o^\delta \gamma_{5,op} Q_p^\delta)_{x_1}\rangle
\end{aligned} \tag{6.6}$$

and

$$\begin{aligned}\langle X^0|\mathcal{O}_{\rm mix}|\overline{X}^0\rangle_{x_1,x_2} &= \langle(\overline{q}_i^\alpha\gamma_{5,ij}Q_j^\alpha)_{x_2}\ (\overline{Q}_k^\beta\Gamma_{1,kl}q_l^\gamma\ \overline{Q}_m^\gamma\Gamma_{2,mn}q_n^\beta)_0\ (\overline{q}_o^\delta\gamma_{5,op}Q_p^\delta)_{x_1}\rangle \\ &+\langle(\overline{q}_i^\alpha\gamma_{5,ij}Q_j^\alpha)_{x_2}\ (\overline{Q}_k^\beta\Gamma_{1,kl}q_l^\gamma\ \overline{Q}_m^\gamma\Gamma_{2,mn}q_n^\beta)_0\ (\overline{q}_o^\delta\gamma_{5,op}Q_p^\delta)_{x_1}\rangle \\ &+\langle(\overline{q}_i^\alpha\gamma_{5,ij}Q_j^\alpha)_{x_2}\ (\overline{Q}_k^\beta\Gamma_{1,kl}q_l^\gamma\ \overline{Q}_m^\gamma\Gamma_{2,mn}q_n^\beta)_0\ (\overline{q}_o^\delta\gamma_{5,op}Q_p^\delta)_{x_1}\rangle \\ &+\langle(\overline{q}_i^\alpha\gamma_{5,ij}Q_j^\alpha)_{x_2}\ (\overline{Q}_k^\beta\Gamma_{1,kl}q_l^\gamma\ \overline{Q}_m^\gamma\Gamma_{2,mn}q_n^\beta)_0\ (\overline{q}_o^\delta\gamma_{5,op}Q_p^\delta)_{x_1}\rangle. \end{aligned} \tag{6.7}$$

Anticommuting the quark fields to the left and identifying the quark propagators defined in Eq. (6.1), we have[2]

$$\begin{aligned}\langle X^0\,|\mathcal{O}_{\rm nomix}|\,\bar{X}^0\rangle_{x_1,x_2} &= H^{\beta\alpha}_{ri}(0,x_2)^*\gamma_{5,rk}\Gamma_{1,kl}L^{\beta\alpha}_{li}(0,x_2)\Gamma_{2,mn}L^{\gamma\delta}_{no}(0,x_1)H^{\gamma\delta}_{to}(0,x_1)^*\gamma_{5,tm} \\ &-H^{\beta\alpha}_{ri}(0,x_2)^*\gamma_{5,rk}\Gamma_{1,kl}L^{\beta\delta}_{lo}(0,x_1)\Gamma_{2,mn}L^{\gamma\alpha}_{ni}(0,x_2)H^{\gamma\delta}_{to}(0,x_1)^*\gamma_{5,tm} \\ &-H^{\gamma\alpha}_{ri}(0,x_2)^*\gamma_{5,rm}\Gamma_{1,kl}L^{\beta\alpha}_{li}(0,x_2)\Gamma_{2,mn}L^{\gamma\delta}_{no}(0,x_1)H^{\beta\delta}_{to}(0,x_1)^*\gamma_{5,tk} \\ &+H^{\gamma\alpha}_{ri}(0,x_2)^*\gamma_{5,rm}\Gamma_{1,kl}L^{\beta\delta}_{lo}(0,x_1)\Gamma_{2,mn}L^{\gamma\alpha}_{ni}(0,x_2)H^{\beta\delta}_{to}(0,x_1)^*\gamma_{5,tk} \end{aligned} \tag{6.8}$$

and

$$\begin{aligned}\langle X^0\,|\mathcal{O}_{\rm mix}|\,\bar{X}^0\rangle_{x_1,x_2} &= H^{\beta\alpha}_{ri}(0,x_2)^*\gamma_{5,rk}\Gamma_{1,kl}L^{\gamma\alpha}_{li}(0,x_2)\Gamma_{2,mn}L^{\beta\delta}_{no}(0,x_1)H^{\gamma\delta}_{to}(0,x_1)^*\gamma_{5,tm} \\ &-H^{\beta\alpha}_{ri}(0,x_2)^*\gamma_{5,rk}\Gamma_{1,kl}L^{\gamma\delta}_{lo}(0,x_1)\Gamma_{2,mn}L^{\beta\alpha}_{ni}(0,x_2)H^{\gamma\delta}_{to}(0,x_1)^*\gamma_{5,tm} \\ &-H^{\gamma\alpha}_{ri}(0,x_2)^*\gamma_{5,rm}\Gamma_{1,kl}L^{\gamma\alpha}_{li}(0,x_2)\Gamma_{2,mn}L^{\beta\delta}_{no}(0,x_1)H^{\beta\delta}_{to}(0,x_1)^*\gamma_{5,tk} \\ &+H^{\gamma\alpha}_{ri}(0,x_2)^*\gamma_{5,rm}\Gamma_{1,kl}L^{\gamma\delta}_{lo}(0,x_1)\Gamma_{2,mn}L^{\beta\alpha}_{ni}(0,x_2)H^{\beta\delta}_{to}(0,x_1)^*\gamma_{5,tk}. \end{aligned} \tag{6.9}$$

We isolate a recurring component in each term – called the *open meson propagator* – that's independent of $\Gamma_{1,2}$. We generate it once and store it for later use. This is cost effective, as the open meson propagators require less space than the quark propagators[3]. Identifying the open meson propagators

$$E^{\alpha\beta}_{ij}(x,y) \equiv L^{\alpha\gamma}_{ik}(x,y)H^{\beta\gamma}_{lk}(x,y)^*\gamma_{5,lj}, \tag{6.10}$$

[2]We utilize here γ_5-Hermiticity, Eq. (B.5), a symmetry of Euclidean heavy quark propagators: $H(y,x) = \gamma_5 H^\dagger(x,y)\gamma_5$.

[3]For a $20^3\times 64$, $a = 0.12$fm lattice an open meson propagator is 150kb while a light quark propagator is 37Mb and a heavy quark propagator is 590Mb! The open meson propagator requires only 0.02% of the disk space required to store quark propagators.

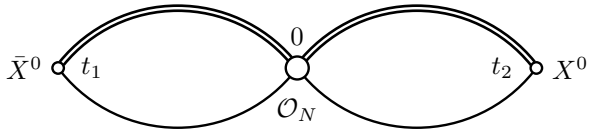


Figure 6.1: The anti-meson is created at time t_1, mixing occurs (via mixing operator $\mathcal{O}_N$) at time 0 converting the anti-meson into a meson, and the meson is annihilated at time t_2, where $t_1 < 0 < t_2$. The heavy (light) quark propagator is indicated by a double (single) line. To extract the hadronic mixing matrix element we need the three point correlation function depicted here and the two point correlation functions corresponding to the propagation of the $\bar{X}^0$ from t_1 to 0 and of the X^0 from 0 to t_2.

summing over space to set $\vec{p} = 0$, and specifying spin and color traces, tr_s and tr_c, respectively, we have

$$\begin{aligned}\langle X^0|\mathcal{O}_{\mathrm{nomix}}|\bar{X}^0\rangle_{t_1,t_2} &= \sum_{\vec{x}_1,\vec{x}_2} \{\mathrm{tr}_s\left[\Gamma_1 \mathrm{tr}_c E(0,x_2)\right]\mathrm{tr}_s\left[\Gamma_2 \mathrm{tr}_c E(0,x_1)\right] - \mathrm{tr}_s\mathrm{tr}_c\left[\Gamma_1 E(0,x_1)\ \Gamma_2 E(0,x_2)\right] \\ &\quad -\mathrm{tr}_s\mathrm{tr}_c\left[\Gamma_1 E(0,x_2)\ \Gamma_2 E(0,x_1)\right] + \mathrm{tr}_s\left[\Gamma_1 \mathrm{tr}_c E(0,x_1)\right]\mathrm{tr}_s\left[\Gamma_2 \mathrm{tr}_c E(0,x_2)\right]\} \quad (6.11)\end{aligned}$$

and

$$\begin{aligned}\langle X^0|\mathcal{O}_{\mathrm{mix}}|\bar{X}^0\rangle_{t_1,t_2} &= \sum_{\vec{x}_1,\vec{x}_2} \{\mathrm{tr}_c\left[\mathrm{tr}_s\left[\Gamma_1 E(0,x_2)\right]\mathrm{tr}_s\left[\Gamma_2 E(0,x_1)\right]\right] - \mathrm{tr}_s\left[\Gamma_1 \mathrm{tr}_c E(0,x_1)\ \Gamma_2 \mathrm{tr}_c E(0,x_2)\right] \\ &\quad -\mathrm{tr}_s\left[\Gamma_1 \mathrm{tr}_c E(0,x_2)\ \Gamma_2 \mathrm{tr}_c E(0,x_1)\right] + \mathrm{tr}_c\left[\mathrm{tr}_s\left[\Gamma_1 E(0,x_1)\right]\mathrm{tr}_s\left[\Gamma_2 E(0,x_2)\right]\right]\}. \quad (6.12)\end{aligned}$$

The open meson propagators are generated by the superscript. A separate code then ties them together with the appropriate Γ's to generate the time-dependent three point correlation function data, depicted in Fig. 6.1. The three point correlation function data for mixing operator $\mathcal{O}_N$ will be denoted by $C_N^{3pt}(t_1,t_2)$,

$$C_N^{3pt}(t_1,t_2) = \begin{cases} \langle X^0|\mathcal{O}_{\mathrm{nomix}}|\bar{X}^0\rangle_{t_1,t_2}, & \text{if } N = 1,2,4 \\ \langle X^0|\mathcal{O}_{\mathrm{mix}}|\bar{X}^0\rangle_{t_1,t_2}, & \text{if } N = 3,5. \end{cases} \quad (6.13)$$

6.3 Building two point correlation functions

In order to extract the hadronic mixing matrix elements from the three point correlation function data, we need the meson two point correlation functions as well. We start with the spacetime dependent correlation

functions, written as time-ordered VEV's complete with spin and color indices

$$\begin{aligned}\langle\left(X^0\right)_{\vec{x},t}|(X^0)_{\vec{0},0}\rangle &= \langle T\{\overline{q}_i^\alpha(x)\gamma_{5,ij}Q_j^\alpha(x)\overline{Q}_k^\beta(0)\gamma_{5,kl}q_l^\beta(0)\}\rangle \text{ and} \\ \langle(\bar{X}^0)_{\vec{0},0}|\left(\bar{X}^0\right)_{\vec{x},t}\rangle &= \langle T\{\overline{Q}_i^\alpha(0)\gamma_{5,ij}q_j^\alpha(0)\overline{q}_k^\beta(x)\gamma_{5,kl}Q_l^\beta(x)\}\rangle, \end{aligned} \tag{6.14}$$

then perform the Wick contractions to get

$$\begin{aligned}\langle\left(X^0\right)_{\vec{x},t}|(X^0)_{\vec{0},0}\rangle &= \overbrace{\langle\overline{q}_i^\alpha(x)\gamma_{5,ij}\overbrace{Q_j^\alpha(x)\overline{Q}_k^\beta(0)}\gamma_{5,kl}q_l^\beta(0)\rangle} \\ &= -\langle q_l^\beta(0)\overline{q}_i^\alpha(x)\rangle\gamma_{5,ij}\langle Q_j^\alpha(x)\overline{Q}_k^\beta(0)\rangle\gamma_{5,kl} \\ &= -L_{li}^{\beta\alpha}(0,x)\gamma_{5,ij}H_{jk}^{\alpha\beta}(x,0)\gamma_{5,kl} \\ &= -E_{lk}^{\beta\beta}(0,x)\gamma_{5,kl} \\ &= -\mathrm{tr}_s\left[\mathrm{tr}_c E(0,x)\gamma_5\right] \end{aligned} \tag{6.15}$$

and for the anti-meson

$$\begin{aligned}\langle(\bar{X}^0)_{\vec{0},0}|\left(\bar{X}^0\right)_{\vec{x},t}\rangle &= \overbrace{\langle\overline{Q}_i^\alpha(0)\gamma_{5,ij}\overbrace{q_j^\alpha(0)\overline{q}_k^\beta(x)}\gamma_{5,kl}Q_l^\beta(x)\rangle} \\ &= -\langle Q_l^\beta(x)\overline{Q}_i^\alpha(0)\rangle\gamma_{5,ij}\langle q_j^\alpha(0)\overline{q}_k^\beta(x)\rangle\gamma_{5,kl} \\ &= -H_{li}^{\beta\alpha}(x,0)\gamma_{5,ij}L_{jk}^{\alpha\beta}(0,x)\gamma_{5,kl} \\ &= -E_{ji}^{\alpha\alpha}(0,x)\gamma_{5,ij} \\ &= -\mathrm{tr}_s\left[\mathrm{tr}_c E(0,x)\gamma_5\right]. \end{aligned} \tag{6.16}$$

We find the meson two point correlation function from 0 to t is equal to the anti-meson two point correlation function from t to 0, as it must by CPT invariance. Finally, we sum over all space to set $\vec{p}=0$, giving the expression used to generate time-dependent two point correlation function data,

$$\langle\left(X^0\right)_t|(X^0)_0\rangle = \langle(\bar{X}^0)_0|\left(\bar{X}^0\right)_t\rangle = -\sum_{\vec{x}}\mathrm{tr}_s\left[\mathrm{tr}_c E(0,x)\gamma_5\right], \tag{6.17}$$

referred to as $C^{2pt}(t)$.

6.4 Operator improvement

6.4.1 Heavy quark smearing

The composite meson interpolating operator acts on the vacuum to produce a linear combination of all states, subject to quantum number selection rules. To increase the overlap of the generated state with the ground state we convolve the heavy quark interpolating field with a 1S wavefunction based on work [46] in the charmonium system. This is referred to as *1S-smearing* and can, in principle, be turned on or off independently at the creation or annihilation of each heavy quark field. If 1S-smearing is not used, the operator is referred to as local[4]. In this work, we use 1S-smearing at the creation and annihilation of heavy quark fields in the two point correlation function data (*ie.* at times 0 and t) and at times t_1 and t_2 in the generation of three point correlation function data. We do not 1S-smear the heavy quark fields in the mixing operator (*ie.* at $t = 0$ in the three point correlation function data) as this is the spacetime point at which we extract the hadronic mixing matrix element. Any smearing would have to be deconvolved to resurrect the correctly normalized value.

6.4.2 Heavy quark rotation

We do, however, improve the heavy quark fields of the mixing operator to improve the matching of lattice heavy quark spinors and continuum relativistic heavy quark spinors [44]. The normalization of the lattice spinor is corrected to remove an $\mathcal{O}(a\vec{p})$ discretization error by a heavy quark spatial "rotation"

$$Q^{\text{rot}} = \left(1 + ad_1\vec{\gamma}\cdot\vec{D}\right)Q, \tag{6.18}$$

where d_1 is tuned to

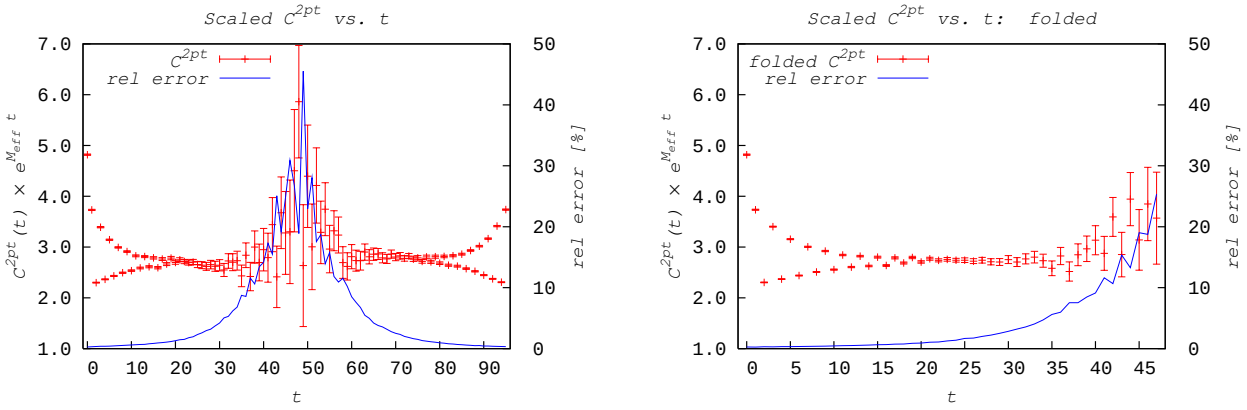
$$d_1 = \frac{1}{2+am_0} - \frac{1}{2(1+am_0)} \tag{6.19}$$

to remove the $\mathcal{O}(a\vec{p})$ discretization error. Here am_0 is the tadpole improved bare heavy quark mass,

$$am_0 = \frac{1}{2u_0}\left(\frac{1}{\kappa} - \frac{1}{\kappa_{\text{crit}}}\right) \tag{6.20}$$

expressed in terms of the hopping parameter, κ, and the critical hopping parameter, κ_{crit}, at which a heavy-heavy "pion" is massless.

[4] *Local* as in a spatial point source, equivalent to maximal spread of momenta internal to the meson and therefore maximal overlap with all states subject to quantum number selection rules.



(a) Unfolded data.

(b) Data folded to enforce a combined PT symmetry.

Figure 6.2: Two point correlation function data for the $a \approx 0.09$ fm, $40^3 \times 96$ ensemble with valence heavy quark $\kappa = 0.0976$ and light quark $am = 0.0261$. This valence mass combination corresponds to the B_s^0 meson. The scaling factor M_{eff} removes leading exponential decay. The relative error of the folded data is a factor of $\sqrt{2}$ smaller than the relative error of the unfolded data. Temporal oscillations in the data are discussed in detail in App. D.4.

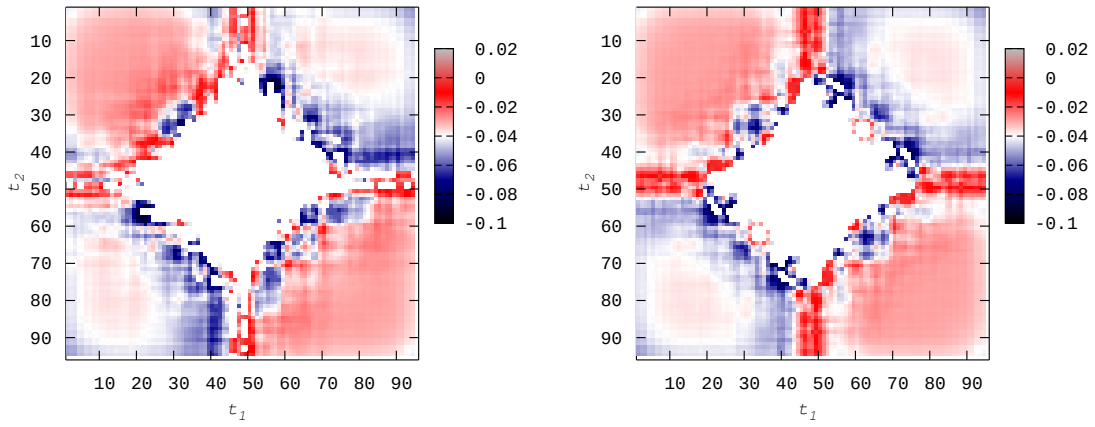
6.5 Exploiting time-reversal and parity: *folding the data*

We take advantage of the derived parity and time-reversal transformations to combine the data in a way that differs from the data combinations performed in Eq. (4.12) based solely on parity. When performing simulations, we use periodic boundary conditions on the lattice. This means the creation of a meson at $t = 0$ also creates a meson at $t = T - 1$. We can identify the signal for a meson traveling backward in time from $t = T - 1$ with the signal for our meson traveling forward in time from $t = 0$, allowing us to *fold* the data about the midpoint $t = T/2$, effectively doubling the data. Using the P and T transformations of the open meson propagators, Eq. (B.6), we can transform the two point correlation function data of Eq. (6.17) to find

$$C^{2pt}(t) = PT(C^{2pt}(t)). \tag{6.21}$$

The backwards propagating meson is physically equivalent to the forward propagating meson, evaluated on parity transformed gauge links. This equivalence is evidenced in the data for a representative two point correlation function, scaled to remove leading exponential decay, shown in Fig. 6.2a. By folding the data we effectively increase statistics associated with the forward propagating two point correlation function. The decrease in relative error, shown in Fig. 6.2b, corresponds to a factor of $1/\sqrt{2}$. App. E summarizes terms used throughout this work to discuss statistical errors and analyses.

We similarly fold the three point correlation function data. For a concrete example, consider the unfolded three point correlation function data for mixing operator $\mathcal{O}_4$ on the $a \approx 0.09$ fm, $32^3 \times 96$ ensemble: $\langle X^0|\mathcal{O}_4([U])|\bar{X}^0\rangle_{t_1,t_2} = \sum_{\vec{x}_1,\vec{x}_2}\langle X^0|\mathcal{O}_4([U])|\bar{X}^0\rangle_{x_1,x_2}$, shown in Fig. 6.3a . Note the explicit $t_1 \leftrightarrow t_2$ sym-



(a) Unfolded data has an explicit $t_1 \leftrightarrow t_2$ symmetry. (b) Data folded to enforce a combined PT symmetry.

Figure 6.3: Data for the scaled three point correlation function $\langle B_s^0|\mathcal{O}_4|\bar{B}_s^0\rangle_{t_1,t_2}$ on the $a \approx 0.09$ fm, $32^3 \times 96$ ensemble with valence heavy quark $\kappa = 0.0977$ and light quark $am_x = 0.0261$. This valence mass combination corresponds to the B_s^0 meson.

metry that follows from the form of Eq.'s (6.11) and (6.12) – the data is a mirror image of itself across the diagonal running from the upper-left to the lower-right. Under time-reversal,

$$\langle X^0|\mathcal{O}_4([U])|\bar{X}^0\rangle_{x_1,x_2} \overset{T}{=} \langle X^0|\mathcal{O}_{17}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau}. \tag{6.22}$$

Enforcing parity symmetry of the matrix element we have

$$\langle X^0|\mathcal{O}_{17}([U]^\tau)|\bar{X}^0\rangle_{x_1^\tau,x_2^\tau} \overset{P}{=} \langle X^0|\mathcal{O}_4([U]^{\tau,p})|\bar{X}^0\rangle_{x_1^{p,\tau},x_2^{p,\tau}} \ , \tag{6.23}$$

giving the combined PT transformation

$$\langle X^0|\mathcal{O}_4([U])|\bar{X}^0\rangle_{x_1,x_2} \overset{PT}{=} \langle X^0|\mathcal{O}_4([U]^{\tau,p})|\bar{X}^0\rangle_{x_1^{p,\tau},x_2^{p,\tau}}, \tag{6.24}$$

analogous to what we found for the two point correlation function data in Eq. (6.21). Data for $\mathcal{O}_4$ is physically equivalent to backward propagating data for the same operator evaluated on parity transformed gauge configurations. This result holds for each of the $\mathcal{O}_{1-5}$. By folding the time-reversed data (folding across the diagonal running from the lower-left to the upper-right) we effectively double our statistics. The folded data are shown in Fig. 6.3b.

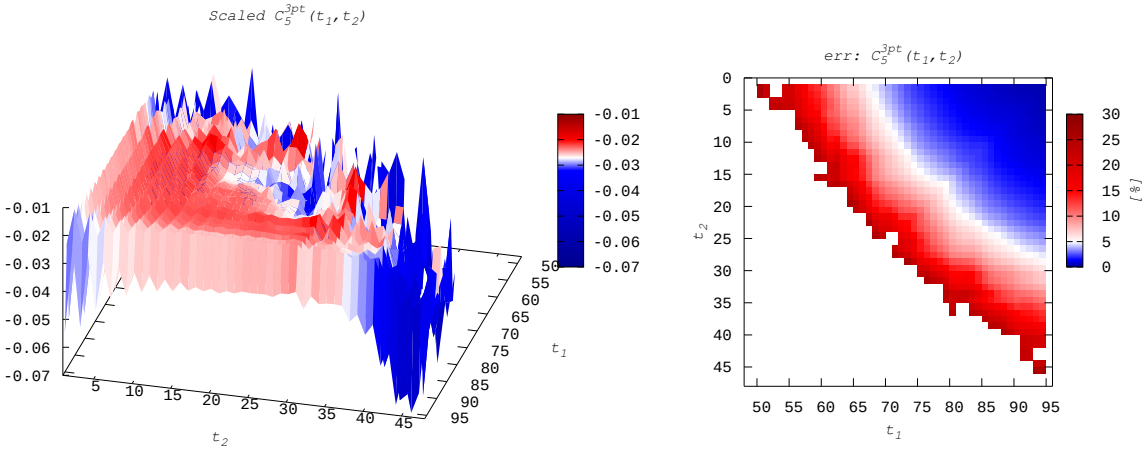
The requirement to time order the VEV of the three point correlation function, Eq. (6.2), is $t_1 < 0 < t_2$.

The progression of times that satisfy this requirement is: $\ldots, t_1 = -2, t_1 = -1, 0, t_2 = 1, t_2 = 2, \ldots$ Given the periodic boundary conditions, this corresponds to

$$\ldots, t_1 = T - 2, t_1 = T - 1, 0, t_2 = 1, t_2 = 2, \ldots \tag{6.25}$$

Therefore, we analyze the upper right hand quadrant of Fig. 6.3b.

An example of three point correlation function data, scaled to remove the leading exponential decay, is shown in Fig. 6.4a and its relative error is plotted in Fig. 6.4b. Aside from providing a visual display of the data, the plots in Fig.'s 6.2b, 6.4a, and 6.4b aide with the fitting procedure outlined in Sec. 7.2.



(a) Scaled data.

(b) Relative error of data.

Figure 6.4: Three point correlation function data for $\mathcal{O}_5$ on the $a \approx 0.09$ fm, $40^3 \times 96$ ensemble with valence heavy quark $\kappa = 0.0976$, and light quark $am_x = 0.0261$. This valence mass combination corresponds to the B_s^0 meson.

Chapter 7

Analyzing correlation functions

7.1 Preliminary analyses

7.1.1 Autocorrelations

Gauge configurations used to evaluate the Monte Carlo sum in Eq. (5.5) are generated by the MILC collaboration via a Molecular Dynamics based algorithm [34]. In this process each gauge configuration is generated from the previous configuration. In such processes, referred to as Markov processes, configurations are necessarily related to each other. The resulting correlations between gauge configurations are referred to as *autocorrelations* and their presence results in an underestimate of the statistical error associated with the Monte Carlo sum. In an attempt to decrease autocorrelations, only every n^{th} configuration is saved (with a typical value of $n = 6$). Despite this effort, autocorrelations may still remain in data built from the gauge configurations. A first step in the lattice calculation of a particular observable[1] is an assessment of autocorrelations.

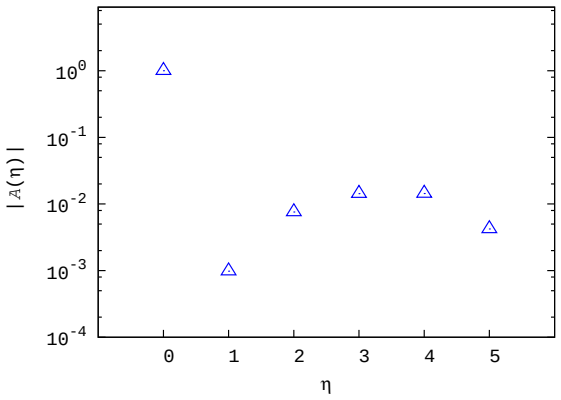
Using methods outlined in App. E.3.2, autocorrelations in B meson two point correlation function data are evaluated by calculating the *autocorrelation function*, $\mathcal{A}(\eta)$,

$$\mathcal{A}(\eta) = \frac{1}{N}\sum_i \frac{(C_{i+\eta}(t) - \overline{C(t)})(C_i(t) - \overline{C(t)})}{\sigma_t^2} \tag{7.1}$$

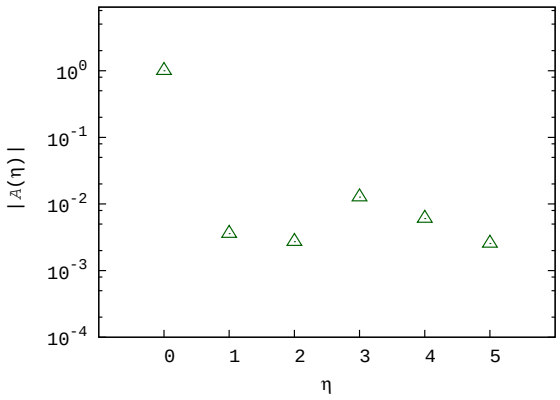
which, by construction, satisfies $\mathcal{A}(0) = 1$ and $-1 < \mathcal{A}(\eta) < 1$.

Fig. 7.1a shows $\mathcal{A}(\eta)$ calculated directly from the two point correlation function data while Fig. 7.1b shows $\mathcal{A}(\eta)$ calculated using the principal component analysis. In each case an *autocorrelation time*, τ, can be extracted from the plots by assuming $|\mathcal{A}(\eta)| = e^{-\eta/\tau}$ [47]. Both plots (using points at $\eta = 0$ and $\eta = 1$) give $\tau \approx 0.15$. Following [47], we skip 2τ configurations to ensure our configurations are independent. For $\tau \approx 0.15$, the configurations are, essentially, independent. That is, the level of autocorrelation present in the data is negligible.

[1] The degree of autocorrelation depends, to some extent, on the observable being calculated.



(a) Two point data correlation function data analysis.



(b) Principal component analysis.

Figure 7.1: Time slice averaged $\mathcal{A}(\eta)$ for the $a \approx 0.12$ fm, $am_l/am_s = 0.007/0.050$ ensemble with $am_x = 0.0415$. Each plot indicates an autocorrelation time of roughly $\tau \approx 0.15$.

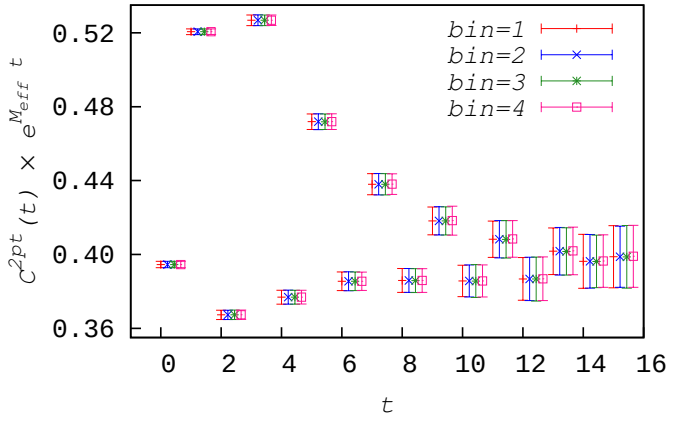


Figure 7.2: The scaled two point correlation function, and its standard error, are plotted for bin sizes ranging from one to four. The data are from the $a \approx 0.12$ fm, $am_l/am_s = 0.007/0.050$ ensemble with $am_x = 0.0415$ and $\kappa = 0.0901$. The standard error is largely unchanged by bin size, indicating negligible autocorrelations.

Again using two point correlation function data, we test for autocorrelations via the third method of App. E.3.2, *binning*. Multiple adjacent configurations are averaged into *bins*, effectively reducing the number of configurations in favor of greater separation in configuration time. Then, by observing the change in relative error with respect to bin size, we can discern the degree of autocorrelation. No change implies no autocorrelation. Fig. 7.2 shows the data for a two point correlation function, scaled to remove leading exponential decay, for bin sizes ranging from one (no binning) to four. The relative error is observed to be constant with respect to bin size, indicating no autocorrelation.

Based on the autocorrelation function, principal component, and binning studies presented here we have determined the B meson data contain negligible autocorrelations. Therefore, we do not bin the data.

7.1.2 Effective mass and amplitude

Several plots of correlation function data in this work are scaled to remove the leading exponential decay, simply for the purpose of making the plots easier to look at. The leading exponential decay comes from a quantity referred to as the *effective mass*, M_{eff}. In addition to scaling data for plots, the effective mass plays an important role in the fitting procedure outlined in Sec. 7.2.3.

In a Euclidean metric, the eigenstate for a free particle with four momentum p has the characteristic behavior, $|p\rangle \sim e^{-px}$. In the particle's rest frame this simplifies to $|M\rangle \sim e^{-Mt}$. The local[2] interpolating operators introduced in Sec. 6.2, however, have overlap with an infinite tower of excited states and the resultant meson correlation function is described by an infinite sum of decaying exponentials $C(t) \sim Z_0 e^{-M_0 t} + Z_1 e^{-M_1 t} + \ldots$, where $M_0 < M_1 < \ldots$ We define $M_{\text{eff}}(t)$ such that $C(t) = Z_{\text{eff}} e^{-M_{\text{eff}} t}$ and extract its value from the correlation function data,

$$M_{\text{eff}}(t) = \frac{1}{n} \log \left[\frac{C(t)}{C(t+n)} \right]. \tag{7.2}$$

This gives us a sort of average over the masses of all states contributing at time t. The expression for the effective mass can be tweaked slightly to account for periodic boundary conditions used in our simulations. Defining M_{eff} such that $C(t) = Z_{\text{eff}} \left(e^{-M_{\text{eff}} t} + e^{-M_{\text{eff}}(T-t)} \right)$ gives

$$M_{\text{eff}}(t) = \frac{1}{n} \cosh^{-1} \left[\frac{C(t+n) + C(t-n)}{2C(t)} \right]. \tag{7.3}$$

Due the exponential decay that results from a Euclidean metric, $M_{\text{eff}}(t) \xrightarrow{t \to \infty} M_0$. Fig. 7.3 shows a representative plot of $M_{\text{eff}}(t)$ obtained using Eq. (7.3) with $n = 2$. The asymptotic M_{eff} value used to remove leading exponential decay in correlation function data comes from a representative point on the plateau beginning near $t = 6$.

With a value for M_{eff} in hand we write the effective amplitude as $Z_{\text{eff}} = C(t) / \left(e^{-M_{\text{eff}} t} + e^{-M_{\text{eff}}(T-t)} \right)$, precisely the quantity plotted in Fig. 6.2 and referred to there as the scaled two point correlation function. We use M_{eff} and Z_{eff} to guide our choice of priors in the constrained fits described in Sec. 7.2.2.

[2] In the next chapter we'll discuss a process called smearing to improve overlap with the ground state. We'll also put the correlation function time dependence on a firmer footing by including effects of staggered light quarks – a complication that does not affect our definitions of M_{eff} or Z_{eff}.

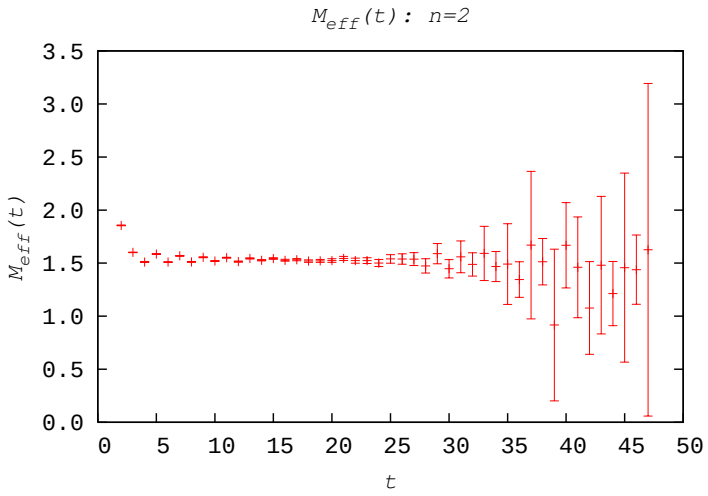


Figure 7.3: The effective mass plot from the $a \approx 0.09$ fm, $am_l/am_s = 0.0031/0.031$ ensemble with $am_x = 0.0031$ and $\kappa = 0.0976$. This lattice has temporal extent $T = 96$ and the data have been folded about the midpoint. The effective mass is calculated using Eq. (7.3) with $n = 2$.

7.2 Fitting correlation functions

7.2.1 Fit functions

The correlation function data generated per Ch. 6, $C^{2pt}(t)$ and $C_N^{3pt}(t_1, t_2)$, are modeled by the fit functions $\mathbb{C}^{2pt}(t)$ and $\mathbb{C}_N^{3pt}(t_1, t_2)$. We build the fit functions starting with Eq.'s (6.14) and (6.2). Working in the B meson rest frame where $E_n = M_n$, we shift quark fields to the Heisenberg picture, *eg.* $q(t) = e^{Ht} q e^{-Ht}$, and insert the complete sets of energy eigenstates,

$$1 = \sum_{n=}^{\infty} \frac{|B_n\rangle\langle B_n|}{2M_n} = \sum_{n=0}^{\infty} \frac{|\bar{B}_n\rangle\langle \bar{B}_n|}{2M_n}, \tag{7.4}$$

where $|B_n\rangle$ is the n^{th} excited state of the B^0 meson. This gives the infinite sums of exponentials,

$$\mathbb{C}^{2pt}(t) \;=\; \sum_{n=0}^{\infty} \frac{Z_n^2}{2M_n} (-1)^{n(t+1)} \left(e^{-M_n t} + e^{-M_n (T-t)} \right) \tag{7.5}$$

$$\mathbb{C}_N^{3pt}(t_1, t_2) \;= \sum_{n,m=0}^{\infty} \frac{\langle B_n | \mathcal{O}_N | \bar{B}_m \rangle Z_n Z_m}{4 M_n M_m} (-1)^{n(t_1+1)+m(t_2+1)} \left(e^{-M_n |t_1|} + e^{-M_n (T-|t_1|)} \right) \left(e^{-M_m t_2} + e^{-M_m (T-t_2)} \right) \tag{7.6}$$

where the real-valued amplitude $Z_n = \langle \overline{q} \gamma_5 Q | B_n \rangle$, $H|\Omega\rangle = 0$, and $t_1 < 0 < t_2$. Time oscillating opposite parity states from staggered light valence quarks (detailed in App. D.4) and the effect of periodic boundary conditions are explicitly accounted for in Eq.'s (7.5) and (7.6). Note the sums separate the parity odd

(expected pseudoscalar) and even (opposite parity oscillating) states according to whether or not the index being summed over is even or odd, *ie.* M_1 is the lowest lying state for the oscillating opposite parity state.

In practice, we limit the time range of data included in the fit ($t_{\text{min}} \leq t \leq t_{\text{max}}$) and truncate the sums ($\sum_{n=0}^{N^{2pt}-1}$ and $\sum_{n,m=0}^{N^{3pt}-1}$), where $N^{2,3pt}$ is the number of states used in the fit. The selection of time ranges and numbers of states is discussed in Sec. 7.2.3.

A comparison of correlation function data to the fit functions allows us to extract the hadronic mixing matrix element (as well as the masses and amplitudes). The best fit values of these parameters are obtained by minimizing the goodness-of-fit parameter, χ^2,

$$\chi^2 = \sum_{t_1,t_2} [f_{t_1}(\{p\}) - \bar{d}_{t_1}] \; (\sigma^2_{t_1 t_2})^{-1} \; [f_{t_2}(\{p\}) - \bar{d}_{t_2}] \tag{7.7}$$

where f is the fit function, $\{p\}$ are parameters to be extracted from the fit, $\bar{d}_t$ are correlation function data at time t averaged over configurations, and $\sigma^2_{t_1 t_2}$ is an element of the covariance matrix, Eq. (E.25).

7.2.2 Constrained fitting

A characteristic feature of meson two point (Fig. 6.2b) and three point (Fig. 6.4b) correlation function data is the exponential growth of relative error with time. This comes from the fact that a large component of the noise in our correlation function data is due to intermediate states composed of the lightest quarks in the simulations, that is, the noise comes from pions [48]. Because the pions are lighter than our heavy-light mesons they decay slower and the signal to noise ratio decays exponentially.

The cleanest, most desirable, data are therefore those at smallest times. However, those data also have the greatest relative contribution from excited states. Including cleaner data comes at the expense of having to include more terms in the sums of decaying exponentials in our fit functions, a notoriously hard problem [49].

A solution to this problem is the inclusion of prior knowledge to prevent the fit from searching unphysical regions of parameter space. For example, the masses of the odd and even parity excited states are ordered, $M_0 < M_2 < M_4 < \ldots$ and $M_1 < M_3 < M_5 < \ldots$, so if we parameterize them by mass differences, *eg.*

$$M_2 = M_0 + e^{\Delta_2}, \tag{7.8}$$

and fit Δ_2, we constrain the fit by preventing it from considering masses such that $M_2 \leq M_0$ in its search of parameter space.

Another application of this type of constraint limits the search of parameter space for the amplitudes. From Eq.'s (7.5) and (7.6) we can assume, without loss of generality, that $Z_n > 0$. The sign associated with

any negative amplitudes can be absorbed in the three point matrix element, a process that doesn't affect the groundstate matrix element we're after. We therefore write

$$\mathcal{Z}_n^2 = \frac{Z_n}{\sqrt{2M_n}} \tag{7.9}$$

and determine the best fit value for $\mathcal{Z}_n$.

The final instance of parametric constraint utilized in our fits is associated with the hadronic mixing matrix elements. CP symmetry of hadronic matrix elements, combined with the fact that our matrix elements are real-valued, gives $\langle B_n|\mathcal{O}_N|\bar{B}_m\rangle = \langle B_m|\mathcal{O}_N|\bar{B}_n\rangle$. This reduces the number of unique entries in the $N^{3pt} \times N^{3pt}$ matrix representing the elements $\langle B_n|\mathcal{O}_N|\bar{B}_m\rangle$ from $(N^{3pt})^2$ to $N^{3pt}(N^{3pt}+1)/2$. We write the groundstate matrix element as[3]

$$O^{(N)} = \frac{\langle B_0|\mathcal{O}_N|\bar{B}_0\rangle}{M_0} \tag{7.10}$$

and define the remaining unique elements, $\{O_i^{(N)} \mid i = 1, \ldots, N^{3pt}(N^{3pt}+1)/2 - 1\}$, to include the matrix element, masses, and amplitudes. In terms of these reparameterizations, and defining

$$\begin{aligned} g_n(t) &= e^{-M_n t} + e^{-M_n(T-t)} \\ &= 2e^{-M_n T/2}\cosh(M_n(T/2-t)), \end{aligned} \tag{7.11}$$

the fit functions are given by

$$\begin{aligned} \mathbb{C}^{2pt}(t) &= \mathcal{Z}_0^2 g_0(t) + \mathcal{Z}_1^2(-1)^{t+1} g_1(t) + \ldots \\ \mathbb{C}_N^{3pt}(t_1,t_2) &= \tfrac{1}{2}O^{(N)}\mathcal{Z}_0^4 g_0(|t_1|)g_0(t_2) + \tfrac{1}{2}O_1^{(N)}\Big[(-1)^{t_2+1}g_0(|t_1|)g_1(t_2) + (-1)^{t_1+1}g_1(|t_1|)g_0(t_2)\Big] \\ &\quad + \tfrac{1}{2}O_2^{(N)}(-1)^{t_1+t_2}g_1(|t_1|)g_1(t_2) + \ldots \end{aligned} \tag{7.12}$$

These simple applications of prior knowledge can be extended in a less trivial way by giving a *preference* for certain values of the parameters to be fit. For example, the plot of M_{eff} in Fig. 7.3 suggests the ground state mass M_0 is probably between 1.0 and 2.0, in fact its probably pretty close to 1.5. The important part of our prior knowledge isn't $M_0 \approx 1.5$, but rather that its *not* Avogadro's number or -1000. Unless we somehow give the fit the information we have from the effective mass plot, it will consider these possibilities with the same probability as values around 1.5. We inform the fit by adding a term to χ^2 that grows as the fit values for M_0 stray too far from 1.5. This is referred to as Bayesian, or constrained, curve fitting [50].

[3] We occasionally denote the light quark content of the B meson in the ground state mixing matrix element by subscript x.

	$\hat{p} \pm \hat{\sigma}$	
fit parameter	$a \approx 0.12$ fm	$a \approx 0.09$ fm
M_0	1.9 ± 0.15	1.55 ± 0.15
M_1	2.15 ± 0.15	1.75 ± 0.15
Δ_2	-0.91 ± 0.35	-1.16 ± 0.4
$\Delta_{i>2}$	-1.23 ± 1.0	-1.48 ± 1.0
$\mathcal{Z}_0$	1.30 ± 0.1	1.25 ± 0.15
$\mathcal{Z}_1$	0.85 ± 0.4	0.85 ± 0.4
$\mathcal{Z}_2$	0.65 ± 0.65	0.62 ± 0.62
$\mathcal{Z}_{i>2}$	0.45 ± 0.45	0.45 ± 0.45
$O^{(N)}$	0 ± 1.0	0 ± 1.0
$O^{(N)}_{i>0}$	0 ± 0.5	0 ± 0.5

Table 7.1: Priors for correlation function fits.

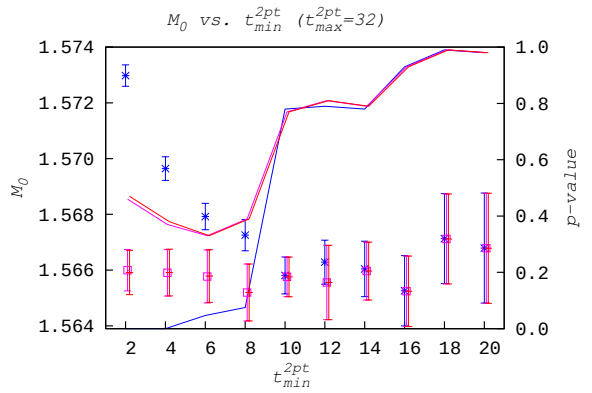
We use a software package developed by G. P. Lepage that augments χ^2 with prior knowledge of the fit parameters $\{p_i\}$, encoded in *priors* $\{\hat{p}_i\}$ and *prior widths* $\{\hat{\sigma}_i\}$. The augmented χ^2_{aug} is written

$$\chi^2_{aug} = \chi^2 + \sum_i \frac{(p_i - \hat{p}_i)^2}{\hat{\sigma}_i^2}. \tag{7.13}$$

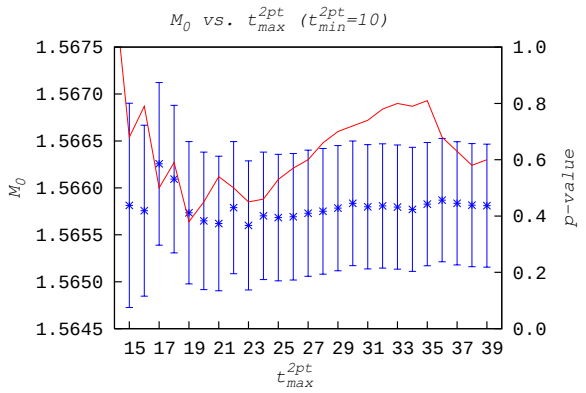
Our choice of priors and prior widths must be given careful consideration to avoid unduly influencing fit results. We routinely check that fit results are contained within the range of values defined by the prior and prior width, $\hat{p} \pm \hat{\sigma}$. Priors for mass splittings are derived from the Particle Data Group [5] or other works [51, 52]. The 1S-smearing of Sec. 6.4.1 results in an overlap with the tower of possible excited states that, in general, decreases with increasing excitation energy, *ie.* $\mathcal{Z}_{n+1} \leq \mathcal{Z}_n$, a fact reflected in our choice of priors for $\mathcal{Z}_n$. Priors for other fit parameters are obtained from subsets of the data or preliminary analyses like the effective mass and amplitude plots. Table 7.1 gives the priors and prior widths used in the constrained fits to the correlation function data.

7.2.3 Fit procedure

Fits using data at small t must account for increased excited state contributions by including an adequate number of states, made possible by constrained curve fitting. To select optimum values of N, t_{min} and t_{max} for the two and three point correlation function fits, we follow a systematic procedure that allows us to achieve consistent and stable fits, with a suitable choice of time range, for $N^{2,3pt} = 2, 4$, and 6. Plots of scaled correlation function data guide our selection of time ranges for study. For t_{min} we generally consider from $t_{\text{min}} = 2$ until excited state contributions have significantly decreased. Though increasing t_{max} utilizes more



(a) Two point correlation function fits versus t^{2pt}_{min}.



(b) Two point correlation function fits versus t^{2pt}_{max}.

Figure 7.4: Two point correlation function fits for the $a \approx 0.09$ fm, $40^3 \times 96$, $am_l/am_s = 0.0031/0.031$ ensemble with $am_x = 0.0261$ and $\kappa = 0.0976$. Fit results reveal a common plateau for $N^{2pt} = 2$ (blue burst), 4 (pink square) and 6 (red dash) fits. Solid lines indicate p-values for the fits.

data, it also introduces an increasing level of noise and can reduce our ability to determine the covariance matrix.

Two point correlation function fits

For the two point correlation function of Fig. 6.2 we consider the ranges $t^{2pt}_{\rm min} = 2, \ldots, 20$, $t^{2pt}_{\rm max} = 15, \ldots, 39$, and $N^{2pt} = 2, 4, 6$. We fit for each combination of N^{2pt}, $t^{2pt}_{\rm min}$ and $t^{2pt}_{\rm max}$ and select a representative fit from the common plateau, ensuring stability with respect to our choice of N^{2pt}, $t^{2pt}_{\rm min}$ and $t^{2pt}_{\rm max}$. Fig. 7.4a (7.4b) shows fit results for the the ground state mass as a function of $t^{2pt}_{\rm min}$ ($t^{2pt}_{\rm max}$). We generate similar plots for the ground state amplitude. Though we've performed cursory excited state studies our primary concern is the ground state mixing matrix element. Our focus is on ground state quantities that directly affect its determination.

From Fig. 7.4a we select the following based on a combination of fit error, central value relative to the common plateau, and p-value:

$$\begin{aligned} N^{2pt} &= 2; \quad & t^{2pt}_{\rm min} &= 10 \\ N^{2pt} &= 4; \quad & t^{2pt}_{\rm min} &= 2 \\ N^{2pt} &= 6; \quad & t^{2pt}_{\rm min} &= 2 \end{aligned} \tag{7.14}$$

For these choices of $t^{2pt}_{\rm min}$, we generate plots of fit results versus $t^{2pt}_{\rm max}$. The $N^{2pt} = 2$ plot is shown in Fig. 7.4b, from which we select a value of $t^{2pt}_{\rm max}$ based on the same criteria used in the selection of $t^{2pt}_{\rm min}$. Our choices for

this two point correlation function are:

$$\begin{aligned} N^{2pt} &= 2; \quad t^{2pt}_{\text{min}} = 10, \quad t^{2pt}_{\text{min}} = 33 \\ N^{2pt} &= 4; \quad t^{2pt}_{\text{min}} = 2, \quad t^{2pt}_{\text{max}} = 33 \\ N^{2pt} &= 6; \quad t^{2pt}_{\text{min}} = 2, \quad t^{2pt}_{\text{max}} = 33 \end{aligned} \tag{7.15}$$

Three point correlation function fits

In the context of testing new physics scenarios via Eq. (3.16), the hadronic mixing matrix element is a natural choice for parameterizing the low energy QCD contribution to B mixing. In the SM, and historically for BSM work, other parameterizations are often used. The most prevalent alternative BSM parameterization uses the *bag parameter*, B_N, defined by

$$O^{(N)} = \mathfrak{c}_N f^2 B_N \tag{7.16}$$

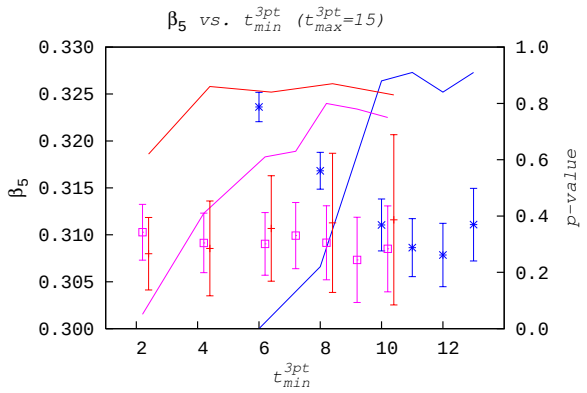
where f is the B meson decay constant and $\mathfrak{c}_N$ are given by: $\mathfrak{c}_1 = 2/3$, $\mathfrak{c}_2 = -5/12$, $\mathfrak{c}_3 = 1/12$, $\mathfrak{c}_4 = 1/2$, and $\mathfrak{c}_5 = 1/6$. In SM B mixing analyses its common to speak of the *mixing parameter*, β_N,

$$\beta_N = f\sqrt{B_N}. \tag{7.17}$$

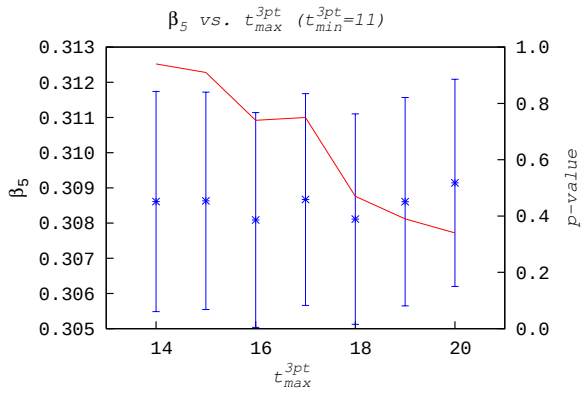
The three point correlation function fit results in this work are reported in terms of either $O_0^{(N)}$ or β_N.

To extract $O^{(N)}$ from Eq. (7.12) we perform a simultaneous fit to the two and three point correlation function data. This is done in an analogous way to the fit procedure outlined above for the two point correlation function. Scaled correlation function data and relative error plots guide selection of time ranges to consider. For the data in Fig.'s 6.4a and 6.4b we choose to study $t^{3pt}_{\text{min}} = 2, \ldots, 13$, $t^{3pt}_{\text{max}} = 14, \ldots, 20$, and $N^{3pt} = 2, 4, 6$. Experience shows that for $a \approx 0.12$ fm data a relative error of $\sim 5\%$ is indicative of the optimum t^{3pt}_{max} while for $a \approx 0.09$ fm data, $\sim 3\%$ is typical. When performing simultaneous fits we match the numbers of states used in the two and three point fit functions ($N^{2pt} = N^{3pt}$) and fix the two point correlation function time ranges per Eq. (7.15).

From resultant simultaneous fits depicted in Fig. 7.5a, we select $N^{3pt} = 2$ and $t^{3pt}_{\text{min}} = 11$ based on the combination of fit error, central value consistency, and p-value. Then, fixing $t^{3pt}_{\text{min}} = 11$, we plot the $N^{3pt} = 2$ fit results versus t^{3pt}_{max} as shown in Fig. 7.5b. From this plot we select $t^{3pt}_{\text{max}} = 15$. The value of $O_0^{(5)}$ at this combination of $N^{2,3pt}$ and $t^{2,3pt}_{\text{min,max}}$ represents the fit result for mixing operator $\mathcal{O}_5$ on the $a \approx 0.09$ fm, $40^3 \times 96$, $am_l/am_s = 0.0031/0.0310$ ensemble with valence quarks $am_x = 0.0261$ and $\kappa = 0.0976$. This analysis is repeated, for $\mathcal{O}_4$ and $\mathcal{O}_5$, on the 60 light valence quark mass and MILC ensemble combinations



(a) Three point correlation function fits versus t^{3pt}_{min} reveal a common plateau for $N^{3pt} = 2$ (blue burst), 4 (pink square) and 6 (red dash) fits.



(b) Three point correlation function N^{3pt} fit results (blue bursts) versus t^{3pt}_{max}.

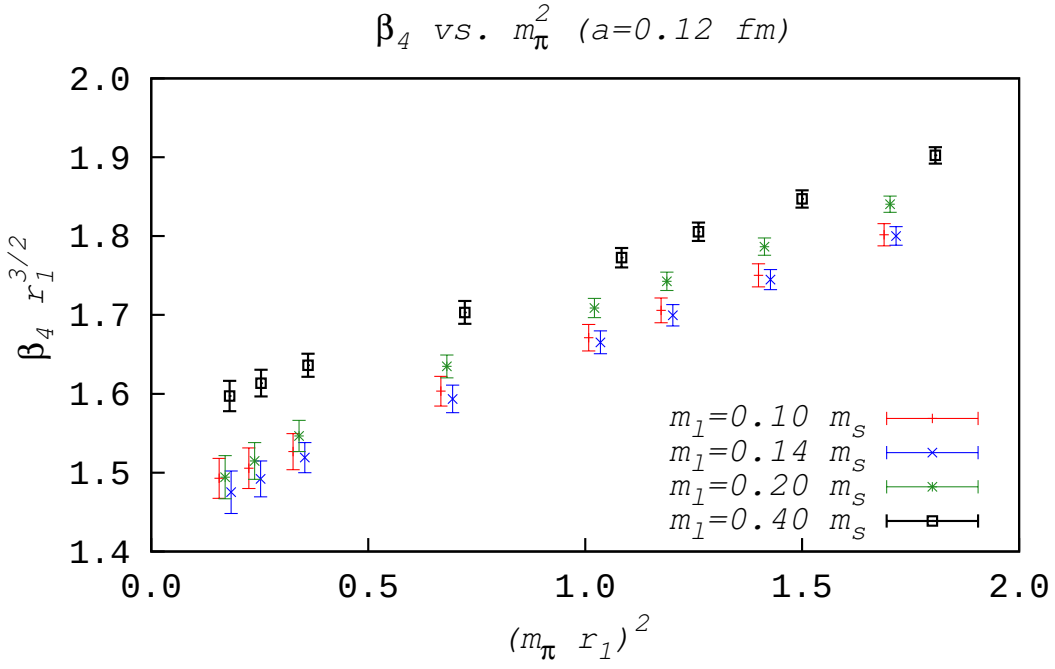
Figure 7.5: Three point correlation function fits for the $a \approx 0.09$ fm, $40^3 \times 96$, $am_l/am_s = 0.0031/0.031$ ensemble with $am_x = 0.0261$ and $\kappa = 0.0976$. Solid lines indicate p-values for the fits.

of Table 5.1. The results of these analyses are shown in Fig. 7.6 for mixing operator $\mathcal{O}_4$ and Fig. 7.7 for mixing operator $\mathcal{O}_5$.

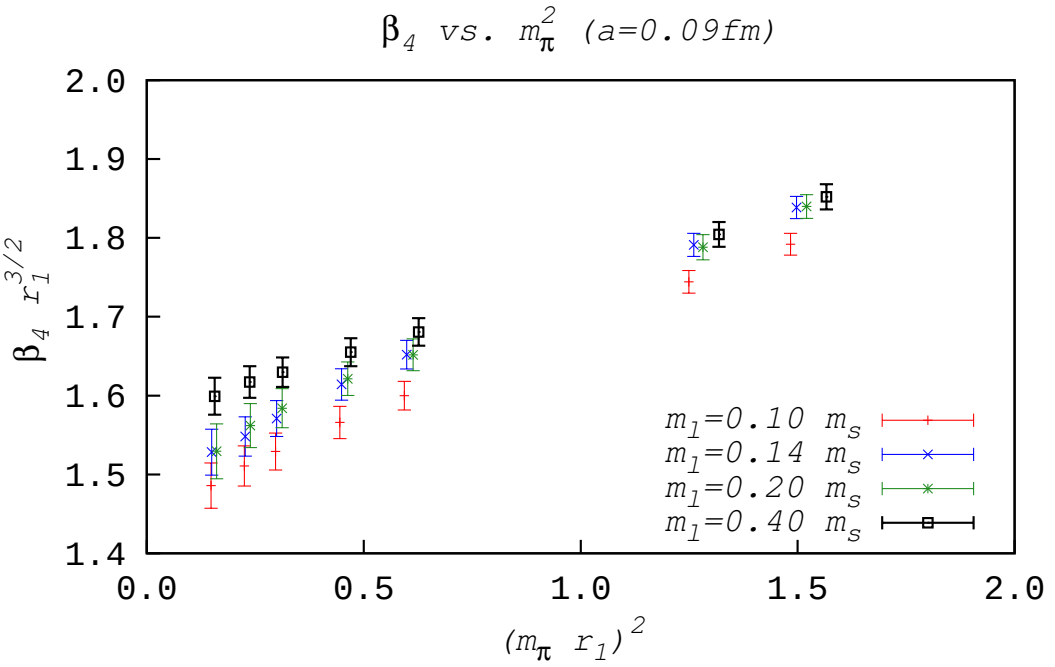
Results are plotted as a function of squared pion mass, with a pion built from degenerate light valence quarks (with mass am_x from Table 5.1). This is in preparation for a subsequent stage of analysis where chiral effective theory, with pions as fundamental degrees of freedom, guides extrapolation to physical pion (and by proxy, quark) masses. Values of $\beta_{4,5}$ extracted from fits, as well as the pion masses, are in lattice units. We use the scale conversion factor r_1/a to convert results to so called "r_1 units", free of lattice spacing dependence. These points are discussed in the context of the chiral extrapolations in Sec.'s 8.2 and 8.3.3, respectively.

Fit errors

We perform the least squares fit of Eq. (7.7) using a covariance matrix calculated from the data. Errors returned by the fit should therefore account for correlations. If the covariance matrix is poorly determined (see App. E.3.3), errors from the fit may not adequately account for correlations. Additionally, fit errors are parametrically dependent on the form of the fit function in Eq.'s (7.5) and (7.6). By using the bootstrap method, outlined in App. E.2.1, we generate a non-parametric estimate of the error by studying the distribution of fit results to 600 bootstrap resamples. The error bars in Fig.'s 7.6 and 7.7 correspond to width of the middle 68% of the bootstrap distribution. We observe excellent agreement between naive fit and bootstrap errors.

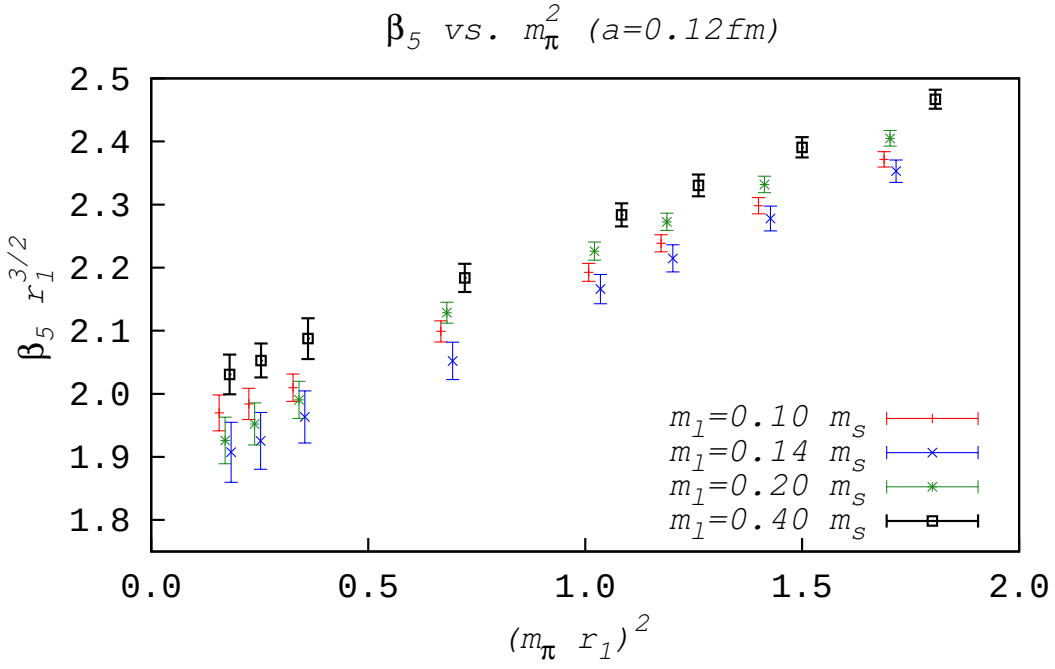


(a) Fit results for $a \approx 0.12$ fm.

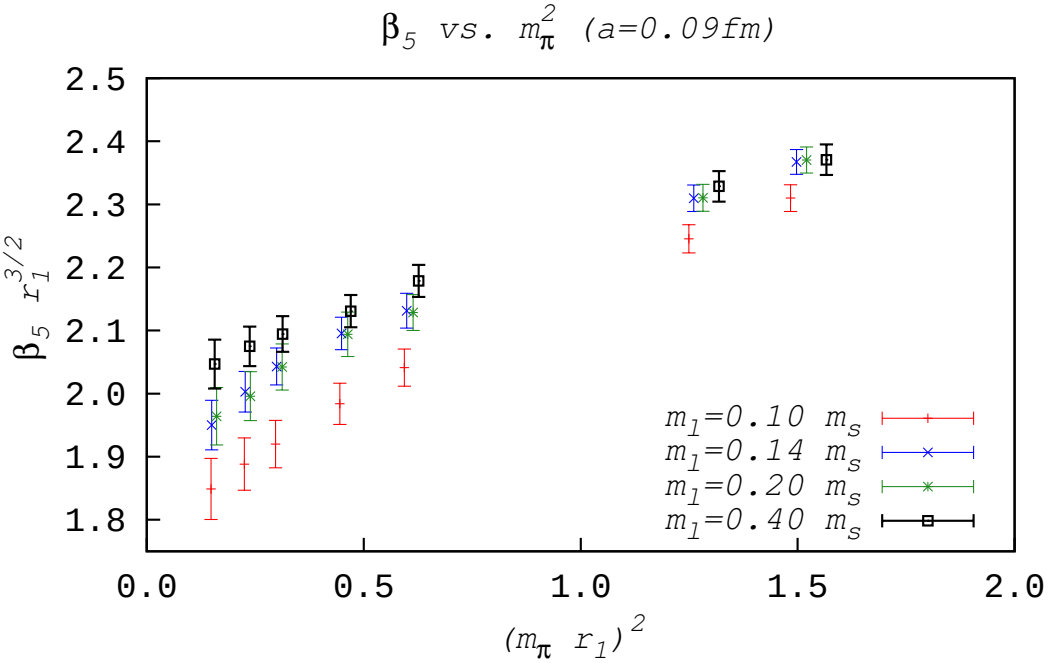


(b) Fit results for $a \approx 0.09$ fm.

Figure 7.6: Simultaneous two and three point correlation function fit results for mixing operator $\mathcal{O}_4$.



(a) Fit results for $a \approx 0.12$ fm.



(b) Fit results for $a \approx 0.09$ fm.

Figure 7.7: Simultaneous two and three point correlation function fit results for mixing operator $\mathcal{O}_5$.

Chapter 8

Relating lattice results to experiment

Correlation function data are built at finite lattice spacing with unphysical quark masses. To compare with experiment, we must first correct for this.

8.1 Renormalization and matching

Lattice calculations are explicitly UV regulated by the lattice spacing, $p_{\max} = \pi/a$. Results obtained at different lattice spacings in Fig.'s 7.6 and 7.7 are effectively evaluated at different energies. In order to perform the necessary extrapolations we need all our data to be evaluated at the same scale. Also, our results for the hadronic mixing matrix elements will be combined with perturbatively calculated quantities per Eq. (3.16). This combination requires consistency with the continuum scheme and scale.

At one loop the BSM operators mix only with each other under renormalization[1]

$$\langle B|\mathcal{O}_4(m_b)|\bar{B}\rangle^R = [1+\alpha_s\zeta_{44}(m_b)]\,\langle B|\mathcal{O}_4|\bar{B}\rangle^{\rm lat} + \alpha_s\zeta_{45}(m_b)\langle B|\mathcal{O}_5|\bar{B}\rangle^{\rm lat} \tag{8.1}$$

$$\langle B|\mathcal{O}_5(m_b)|\bar{B}\rangle^R = [1+\alpha_s\zeta_{55}(m_b)]\,\langle B|\mathcal{O}_5|\bar{B}\rangle^{\rm lat} + \alpha_s\zeta_{54}(m_b)\langle B|\mathcal{O}_4|\bar{B}\rangle^{\rm lat} \tag{8.2}$$

where the ζ_{ij} run lattice results from their value at the energy scale set by the lattice spacing, $\langle\bullet\rangle^{\rm lat}$, to their value at the energy scale used to evaluate perturbatively calculated Wilson coefficients, $C_i^{(2)}(\mu)$, of Eq. (3.16), $\langle\bullet(\mu)\rangle$. For B mixing the scale used is the mass of the bottom quark, $\mu = m_b$. The superscript "R" indicates renormalized matrix elements while "lat" indicates matrix elements obtained at finite lattice spacing. The coefficients ζ_{ij} are calculated to one loop in lattice perturbation theory by collaboration members Elvira Gámiz and Andreas Kronfeld and are listed in Table 8.1.

The QCD coupling constant, α_s, should be evaluated at a scale indicative of the typical gluon loop momentum. Following [38] and [55] we use the "V scheme" and define α_s in terms of the static quark

[1] From [53] and the Fierz transformations of Sec. 4.1.2.

ζ_{ij}	$a \approx 0.12$ fm	$a \approx 0.09$ fm
ζ_{44}	1.7352	1.4329
ζ_{45}	-0.4886	-0.4388
ζ_{54}	0.2532	0.0660
ζ_{55}	0.1958	0.1809

Table 8.1: BSM mixing operator matching coefficients, calculated to one loop [54].

potential $V(q)$. Schematically, we take $\alpha_s = \alpha_V$, where α_V is defined by the static quark potential,

$$V(q) = \frac{4}{3}\frac{4\pi\alpha_V(q)}{q^2}. \tag{8.3}$$

Combined with a lattice measurement of the static quark potential, this allows us to extract a value for $\alpha_V(q)$. This value is then run to the desired scale, $\alpha_V(\mu)$, using the renormalization group. In practice, we take the third-order expansion for $\alpha_{\overline{MS}}(q)$ in terms of $\alpha_V(qe^{5/6})$ from [56] as an all orders definition of $\alpha_V(q)$. A lattice measurement of the static quark potential [55] yields a value for $\alpha_V(q)$ which is run to the desired scale $\alpha_V(\mu)$ by

$$q^2\frac{d\alpha_V(q)}{dq^2} = -\beta_0\alpha_V(\mu)^2 - \beta_1\alpha_V(\mu)^3 - \beta_2\alpha_V(\mu)^4 - \beta_3\alpha_V(\mu)^5 - \cdots \tag{8.4}$$

where the perturbative coefficients, β_i, are determined from the results of [57]. We choose $\mu = 2/a$ as the typical scale of gluon loop momentum, a choice consistent with the optimal choices of [38] and [58].

8.2 Chiral perturbation theory

Extrapolation to physical light quark mass is guided by chiral perturbation theory χPT – the low energy, low mass effective theory of QCD [59, 60, 61, 62]. In the limit of zero light quark (u, d, s) mass, QCD gains a global $SU(3)_L \otimes SU(3)_R$ chiral symmetry, believed to be broken spontaneously to $SU(3)_V$ with eight conserved Noether currents, $j^a_\mu = \overline{q}\gamma_\mu\lambda^a q$ (λ^a are the 3×3 Gell-Mann matrices). The particles associated with $SU(3)_V$ are, after spontaneous symmetry breaking, the massive vector mesons (ρ, K^*, ω, ϕ). The broken $SU(3)_A$ has eight massless, pseudoscalar, Nambu Goldstone bosons, referred to generically as pions, π^a. The pions of $SU(3)_A$ are associated with the observed pseudoscalar octet (π, K, η). In the massless limit, the axial vector currents, $j^a_{\mu 5} = \overline{q}\gamma_\mu\gamma_5\lambda^a q$, are also conserved, $p_\mu\langle\Omega|j^a_{\mu 5}|\pi^b\rangle = \delta^{ab}M_\pi^2 f_\pi = 0$ (massless pions). For light, not massless, quarks chiral symmetry is softly broken. The pions are no longer massless and are referred to as *pseudo* Nambu Goldstone bosons and the axial vector current is only partially conserved.

Pions dominate low energy dynamics and are the fundamental degrees of freedom in χPT. They're packaged to build the field,

$$\Sigma = \exp\left(\frac{i\pi^a\lambda^a}{f_\pi}\right) \tag{8.5}$$

where

$$\pi^a\lambda^a = \begin{pmatrix} \pi^0 + \frac{1}{\sqrt{3}}\eta & \sqrt{2}\pi^+ & \sqrt{2}K^+ \\ \sqrt{2}\pi^- & -\pi^0 + \frac{1}{\sqrt{3}}\eta & \sqrt{2}K^0 \\ \sqrt{2}K^- & \sqrt{2}\bar{K}^0 & -\frac{2}{\sqrt{3}}\eta \end{pmatrix}. \tag{8.6}$$

Under $SU(3)_L \otimes SU(3)_R$, Σ transforms as $\Sigma \to U_L \Sigma U_R$, with $U_{L/R} \in SU(3)_{L/R}$.

The effective lagrangian is built from chiral invariant interactions, making it a low mass effective theory. Its also a low energy effective theory in that its interactions are built from an expansion in powers of momentum. Lowest order interactions have the fewest powers of momentum. The leading order effective lagrangian encodes this by starting with the most general, Lorentz and chiral invariant term with the fewest number of derivatives,

$$\begin{aligned} \mathcal{L}_{\text{eff}}^{\text{LO}} &= \frac{f_\pi^2}{4} Tr\left(\partial_\mu \Sigma \partial_\mu \Sigma^\dagger\right) \\ &= \tfrac{1}{2}\partial_\mu \pi^a \partial_\mu \pi^a + \mathcal{L}_{\text{int}}, \end{aligned} \tag{8.7}$$

where the second line comes from expanding Σ and grouping all higher order terms in $\mathcal{L}_{\text{int}}$. Higher order terms correspond to higher powers of momenta and are suppressed by powers of p/Λ_χ, where $\Lambda_\chi \sim 4\pi f_\pi \sim 1$ GeV.

The lagrangian of Eq. (8.7) contains no quark (and therefore pion) masses. We incorporate non zero quark mass by adding the explicit chiral symmetry breaking term

$$\mathcal{L}_{\text{sb}} = \frac{f_\pi^2 B_0}{2} Tr(M\Sigma^\dagger + \Sigma M^\dagger) \tag{8.8}$$

where $M = \text{diag}(m_u, m_d, m_s)$ and B_0 is a second (f_π was the first) free parameter added to the theory. The validity of χPT extends to small but nonzero $m_{u,d,s}$ giving small but nonzero pion masses. The pion mass dependence guides the extrapolation of data in Fig.'s 7.6 and 7.7 to physical masses.

It is computationally cheaper to take an existing ensemble and generate data for different valence quark masses than to generate a new ensemble[2]. This results in non-degenerate sea and valence quarks of the same flavor. Chiral effective theory, modified [63] to allow this generalization, is referred to as *partially quenched* χPT. Notationally, we specify valence quarks by m_x and sea quarks by $m_{u,d,s}$. In practice, we use degenerate

[2]This is the reason for the proliferation of valence quark masses for each ensemble in Table 5.1.

light sea quarks, $m_u = m_d$ (and occasionally refer to the light sea quark mass as m_l). These mass differences are accounted for by the flavor splitting, $\delta_{xf} \sim (m_f - m_x)$, for sea quark flavor $f \in \{l, s\}$.

The presence of a heavy b quark in the B meson allows additional constraints, from heavy quark spin and flavor symmetries, to be incorporated in χPT via heavy quark effective theory [64]. The continuum, partially quenched, heavy meson χPT for the $\Delta b = 2$ flavor changing interactions relevant to B mixing was developed by Detmold and Lin [65]. Heavy-light pseudoscalar mesons, via pion exchange, can generate virtual heavy-light vector mesons, an interaction characterized by the coupling constant g. The mass difference between the vector and pseudoscalar B meson is the hyperfine splitting, Δ^*.

The incorporation of rooted, staggered light quarks to partially quenched, heavy meson χPT was worked out by Aubin and Bernard [66, 67]. Within this framework, Laiho and Van de Water [68] extended the work of Detmold and Lin to include rooted, staggered quarks. The resulting rooted, staggered, heavy meson, χPT for $\Delta b = 2$ flavor changing interactions in B mixing is the starting point for the extrapolation performed in this work. An important component of staggered χPT is the incorporation of lattice spacing dependence, *eg.*

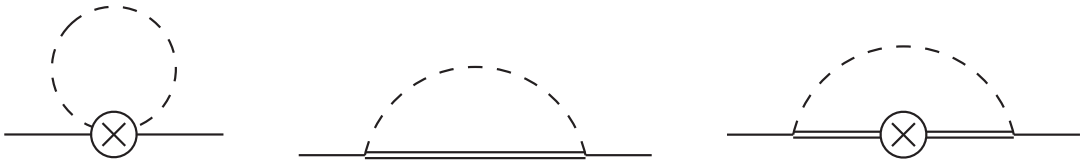
$$m^2_{ij,\Xi} = \mu(m_i + m_j) + a^2\Delta_\Xi, \tag{8.9}$$

where $m_{ij,\Xi}$ is a taste Ξ pion composed of flavor i and j quarks. By explicitly incorporating leading order discretization (taste breaking) effects in the fit function the effects are systematically removed from the fit results. The low energy constant μ and taste splittings Δ_Ξ are fixed [69]. In addition to removing leading order discretization (taste breaking) effects, the incorporation of lattice spacing dependence allows us to perform a simultaneous continuum and chiral extrapolation.

8.3 Chiral Extrapolation

We treat the fit results for the matrix elements (at different sea and valence masses and lattice spacings) as data points to which the chiral expression for the matrix element is fit, allowing extrapolation to physical quark masses and the continuum. We outline the elements of the chiral expression in the context of performing a chiral fit.

The next to leading order (NLO) expressions for the rooted, staggered, heavy meson, chiral expansions



(a) Tadpole diagram, $\mathcal{T}$. (b) Wavefunction diagram, $\mathcal{W}$. (c) Sunset diagram, $\mathcal{Q}$.

Figure 8.1: Next to leading order χPT diagrams. Dashed lines are pions, solid lines are B meson, double lines are B^* mesons, and $\otimes$ is the mixing operator.

of the mixing matrix elements[3] are [68]

$$O_x^{(1)} = \beta_1 \left(1 + \mathcal{T}_x^1 + \mathcal{W}_x + \mathcal{Q}_x\right) + \text{analytic terms} \tag{8.10}$$

$$O_x^{(N)} = \beta_N \left(1 + \mathcal{T}_x^i + \mathcal{W}_x\right) + \beta'_N \mathcal{Q}_x + \text{analytic terms}, \ N \in \{2..5\} \tag{8.11}$$

where β_N and β'_N are constants to be fit, representing the leading order value of the matrix elements.

The "analytic terms" in Eq.'s (8.11) and (8.11) encompass NLO contributions from $\mathcal{L}_{\text{sb}}$ and $\mathcal{O}(a^2)$ taste violations. These contributions are parameterized by [68]

$$\text{analytic terms} = c_0 m_x + c_1(2m_u + m_s) + c_2 a^2 \tag{8.12}$$

where c_0, c_1, and c_2 are low energy constants to be determined from the fit. Prior selection for NLO analytic terms is guided by [70].

The tadpole diagram term, $\mathcal{T}_x^i$, of Fig. 8.1a includes a sum over the three flavors of sea quarks with the degeneracy $m_u = m_d$. The factors of four and six associated with the taste multiplets indicate the sum runs over each element of the multiplet. At order a^2, taste splittings are degenerate within a multiplet, and sum results in four or six identical copies. The sum over all 16 tastes and factor of $\frac{1}{16}$ gives the average taste contribution. The tadpole diagram term for the SM mixing operators is [68]

$$\begin{aligned}
\mathcal{T}_x^{1,2,3} = \frac{-i}{f^2}\Bigg\{ & \frac{1}{16} \sum_{\substack{j=xu,xd,xs \\ \Xi=I,P,4V,4A,6T}} \mathcal{I}_{j\Xi} + \frac{1}{16} \sum_{\Xi=I,P,4V,4A,6T} \mathcal{I}_{xx\Xi} + \frac{2}{3}\Bigg[R_{X_I}(\{M_{X_I}\};\{\mu_I\}) \left(\frac{\partial \mathcal{I}_{X_I}}{\partial m_{X_I}^2}\right) \\
& - \sum_{j\in\{M_{X_I}\}} D_{j,X_I}(\{M_{X_I}\};\{\mu_I\})\mathcal{I}_j \Bigg] + a^2\delta'_V \Bigg[R_{X_V}(\{M_{X_V}\};\{\mu_V\}) \left(\frac{\partial \mathcal{I}_{X_V}}{\partial m_{X_V}^2}\right) \\
& - \sum_{j\in\{M_{X_V}\}} D_{j,X_V}(\{M_{X_V}\};\{\mu_V\})\mathcal{I}_j \Bigg] + (V \to A) \Bigg\}
\end{aligned} \tag{8.13}$$

[3]Field normalization from heavy meson χPT results in matrix elements that are physically equivalent to the quantity $O^{(N)}$, referred to as the ground state matrix element and defined in Eq. (7.10).

and the expression for the BSM mixing operators is [68]

$$\mathcal{T}_x^{4,5} = \frac{-i}{f^2}\left[\frac{1}{16}\sum_{\substack{j=xu,xd,xs\\ \Xi=I,P,4V,4A,6T}} \mathcal{I}_{j\Xi} - \frac{1}{16}\sum_{\Xi=I,P,4V,4A,6T} \mathcal{I}_{xx\Xi}\right] \tag{8.14}$$

where $\delta'_{V(A)}$ is a *hairpin vertex factor* arising from flavor singlet pions [67] that we treat as a fit parameter. The MILC collaboration determination of $\delta'_{V(A)}$ guides prior selection. The pseudoscalar decay constant, f, is fixed to $f = 132$ MeV as in [65].

The wave function diagram of Fig. 8.1b is given by [68]

$$\begin{aligned}\mathcal{W}_x = \frac{ig^2}{f^2}\Bigg\{&\frac{1}{16}\sum_{\substack{j=xu,xd,xs\\ \Xi=I,P,4V,4A,6T}} \mathcal{H}_{j\Xi}^{\Delta^*+\delta_j} + \frac{1}{3}\Bigg[R_{X_I}(\{M_{X_I}\};\{\mu_I\})\left(\frac{\partial \mathcal{H}_{X_I}^{\Delta^*}}{\partial m_{X_I}^2}\right)\\ &- \sum_{j\in\{M_{X_I}\}} D_{j,X_I}(\{M_{X_I}\};\{\mu_I\})\mathcal{H}_j^{\Delta^*}\Bigg] + a^2\delta'_V\Bigg[R_{X_V}(\{M_{X_V}\};\{\mu_V\})\left(\frac{\partial \mathcal{H}_{X_V}^{\Delta^*}}{\partial m_{X_V}^2}\right)\\ &- \sum_{j\in\{M_{X_V}\}} D_{j,X_V}(\{M_{X_V}\};\{\mu_V\})\mathcal{H}_j^{\Delta^*}\Bigg] + (V\to A)\Bigg\}\end{aligned} \tag{8.15}$$

where g is the πBB^* coupling. We use several works [65, 71, 72, 73] to guide prior selection and treat g as a fit parameter. The hyperfine, Δ^*, and flavor, δ_{xf}, splittings are fixed per [72].

The sunset diagram in Fig. 8.1c is evaluated by [68]

$$\begin{aligned}\mathcal{Q}_x = \frac{-ig^2}{f^2}\Bigg\{&\frac{1}{16}\sum_{\Xi=I,P,4V,4A,6T} \mathcal{H}_{xx\Xi}^{\Delta^*} + \frac{1}{3}\Bigg[R_{X_I}(\{M_{X_I}\};\{\mu_I\})\left(\frac{\partial \mathcal{H}_{X_I}^{\Delta^*}}{\partial m_{X_I}^2}\right)\\ &- \sum_{j\in\{M_{X_I}\}} D_{j,X_I}(\{M_{X_I}\};\{\mu_I\})\mathcal{H}_j^{\Delta^*}\Bigg]\Bigg\}.\end{aligned} \tag{8.16}$$

Per Ref. [74] splittings here should be Δ^* (the heavy-light meson in the loop is a vector meson so there should be a hyperfine splitting to adjust its mass, and there are no sea quark loops to give flavor splitting) in contrast to [65].

8.3.1 Summary of chiral functions

In the evaluation of these diagrams a common set of functions is encountered. The residue functions R and D, the integrals $\mathcal{H}$ and $\mathcal{I}$, and their derivatives $\frac{\partial \mathcal{H}}{\partial m^2}$ and $\frac{\partial \mathcal{I}}{\partial m^2}$ are discussed below. In this work we neglect finite volume effects, expected to be $< 0.1\%$ [71], but give the necessary expressions. Hyperfine and flavor splittings, motivated [72] by an expected improvement to the determination of finite volume effects, are also

neglected.

The residue function R appears from single poles in the flavor singlet hairpin propagator and is defined in the appendix of [75]

$$R_j(\{m\};\{\mu\}) = \frac{\prod_{a=1}^{k}(\mu_a^2 - m_j^2)}{\prod_{\substack{i=1 \\ i\neq j}}^{n}(m_i^2 - m_j^2)} \tag{8.17}$$

where m_j is the pole mass in the propagator. The lists $\{m\}$ and $\{\mu\}$ have n and k elements, respectively.

The *double pole* residue function D arises from double poles in the flavor singlet hairpin propagator. Also defined in the appendix of [75], D is

$$D_{j,l}(\{m\};\{\mu\}) = -\frac{d}{dm_l^2} R_j(\{m\};\{\mu\}) \tag{8.18}$$

where $m_l \in \{m\}$. Note that for $l \neq j$,

$$D_{j,l}(\{m\};\{\mu\}) = \frac{R_j(\{m\};\{\mu\})}{(m_l^2 - m_j^2)}, \tag{8.19}$$

while for $l = j$,

$$D_{j,j}(\{m\};\{\mu\}) = R_j(\{m\};\{\mu\}) \left(\sum_{a=1}^{k} \frac{1}{\mu_a^2 - m_j^2} - \sum_{i=1, i\neq j}^{n} \frac{1}{m_i^2 - m_j^2} \right). \tag{8.20}$$

The *light meson integral* $\mathcal{I}_m$ is defined in two forms in the appendix of [65], depending on whether finite volume effects are included (FV) or the infinite volume limit is taken (IV). I'll specify the case by indexing $\mathcal{I}_m$, *eg.* the finite volume version of $\mathcal{I}_m$ will be denoted by $\mathcal{I}_m^{\mathrm{FV}}$. UV divergences from loop integration using dimensional regularization have been removed by subtracting

$$\overline{\lambda} = \frac{2}{4-d} - \gamma_E + \log(4\pi) + 1. \tag{8.21}$$

The simplest form of $\mathcal{I}_m$ neglects finite volume effects

$$\mathcal{I}_m^{\mathrm{IV}} = \frac{-im^2}{16\pi^2} \log\left(\frac{m^2}{\Lambda_\chi^2}\right) \tag{8.22}$$

where Λ_χ is the chiral symmetry breaking scale and m is the relevant meson mass. Finite volume effects would be incorporated by adding a correction term. That is,

$$\mathcal{I}_m^{\mathrm{FV}}(L) = \mathcal{I}_m^{\mathrm{IV}} + \delta\mathcal{I}_m^{\mathrm{FV}}(L) \tag{8.23}$$

where the finite volume correction term is given by

$$\begin{aligned}\delta\mathcal{I}_m^{\rm FV}(L) &= \frac{-im}{4\pi^2}\sum_{\vec{n}\neq\vec{0}}\frac{1}{nL}K_1(nmL)\\ &\xrightarrow{mL\gg 1} \frac{-i}{4\pi^2}\sum_{\vec{n}\neq\vec{0}}\sqrt{\frac{m\pi}{2nL}}\left(\frac{1}{nL}\right)e^{-nmL}\left\{1+\frac{3}{8nmL}-\frac{15}{128(nmL)^2}+\mathcal{O}\left(\left[\frac{1}{nmL}\right]^3\right)\right\}\end{aligned} \tag{8.24}$$

with L the spatial extent of the lattice, K_1 the modified Bessel function of the second kind of order one, and $n=|\vec{n}|$. The derivative of Eq. (8.24) is

$$\begin{aligned}\frac{\partial}{\partial m^2}\,\delta\mathcal{I}_m^{\rm FV}(L) &= \frac{\partial}{\partial m^2}\left[\frac{-i}{4\pi^2}\sum_{\vec{n}\neq\vec{0}}\sqrt{\frac{m\pi}{2nL}}\left(\frac{1}{nL}\right)e^{-nmL}\left\{1+\frac{3}{8nmL}-\frac{15}{128(nmL)^2}+\mathcal{O}\left(\left[\frac{1}{nmL}\right]^3\right)\right\}\right]\\ &= \frac{i}{8\pi^2}\sum_{\vec{n}\neq\vec{0}}\sqrt{\frac{\pi}{2mnL}}\,e^{-nmL}\left\{1-\frac{1}{8nmL}+\frac{9}{128(nmL)^2}+\mathcal{O}\left(\left[\frac{1}{nmL}\right]^3\right)\right\}.\end{aligned} \tag{8.25}$$

The *heavy meson integral*, $\mathcal{H}_m^{\Delta}$, is defined in two forms in the appendix of [65], based on whether or not finite volume effects are included. We use the same notation used for the light meson integrals to specify the case (*ie.* a superscript IV or FV.). Neglecting finite volume effects,

$$\mathcal{H}_m^{\Delta;\,\rm IV}=\frac{3i}{16\pi^2}\frac{\partial}{\partial\Delta}\Big\{\log\Big(\frac{m^2}{\Lambda_\chi^2}\Big)\Big(m^2-\frac{2\Delta^2}{3}\Big)\Delta+\Big(\frac{10\Delta^2}{9}-\frac{4m^2}{3}\Big)\Delta+\frac{2m(\Delta^2-m^2)}{3}R\Big(\frac{\Delta}{m}\Big)\Big\} \tag{8.26}$$

where

$$R(x)=\sqrt{x^2-1}\log\left(\frac{x-\sqrt{x^2-1+i\epsilon}}{x+\sqrt{x^2-1+i\epsilon}}\right). \tag{8.27}$$

Evaluating the derivative gives,

$$\begin{aligned}\mathcal{H}_m^{\Delta;\,\rm IV} &= \frac{i}{8\pi^2}\left\{\frac{3}{2}\log\left(\frac{m^2}{\Lambda_\chi^2}\right)(m^2-2\Delta^2)+5\Delta^2-2m^2+2m\Delta R\Big(\frac{\Delta}{m}\Big)\right.\\ &\left.+(\Delta^2-m^2)\left[\frac{\Delta/m}{\sqrt{\Delta^2/m^2-1}}\log\left(\frac{\Delta/m-\sqrt{\Delta^2/m^2-1+i\epsilon}}{\Delta/m+\sqrt{\Delta^2/m^2-1+i\epsilon}}\right)-2\right]\right\}\end{aligned} \tag{8.28}$$

In the limit of zero splitting this reduces to

$$\lim_{\Delta\to 0}\mathcal{H}_m^{\Delta;\,\rm IV}=-3\,\mathcal{I}_m^{\rm IV}. \tag{8.29}$$

Because we neglect hyperfine and flavor splittings, we make the replacement $\mathcal{H}_m^{\Delta;\,\rm IV}\to -3\,\mathcal{I}_m^{\rm IV}$.

Finite volume effects are incorporated by adding a correction term,

$$\mathcal{H}_m^{\Delta;\ \mathrm{FV}}(L) = \mathcal{H}_m^{\Delta;\ \mathrm{IV}} + \delta\mathcal{H}_m^{\Delta;\ \mathrm{FV}}(L) \tag{8.30}$$

where the finite volume correction term is given by

$$\delta\mathcal{H}_m^{\Delta;\ \mathrm{FV}}(L) = i(m^2-\Delta^2)\frac{\partial J_m^\Delta(L)}{\partial\Delta} - 2\Delta i J_m^\Delta(L) - \delta\mathcal{I}_m^{\mathrm{FV}}(L) \tag{8.31}$$

and the function $J_m^\Delta(L)$ is, for $mL \gg 1$,

$$J_m^\Delta(L) = \sum_{\vec{n}\neq\vec{0}} \left(\frac{1}{8\pi nL}\right) e^{-nmL}\mathcal{A}_{m,n}^\Delta(L) \tag{8.32}$$

where

$$\begin{aligned}\mathcal{A}_{m,n}^\Delta(L) &= e^{z^2}\big[1-\mathrm{Erf}(z)\big] + \left(\frac{1}{nmL}\right)\left[\frac{1}{\sqrt{\pi}}\left(\frac{z}{4}-\frac{z^3}{2}\right) + \frac{z^4}{2}e^{z^2}\big[1-\mathrm{Erf(z)}\big]\right] \\ &\quad -\left(\frac{1}{nmL}\right)^2\left[\frac{1}{\sqrt{\pi}}\left(\frac{9z}{64}-\frac{5z^3}{32}+\frac{7z^5}{16}+\frac{z^7}{8}\right) - \left(\frac{z^6}{2}+\frac{z^8}{8}\right)e^{z^2}\big[1-\mathrm{Erf(z)}\big]\right] + \mathcal{O}\left(\left[\frac{1}{nmL}\right]^3\right),\end{aligned} \tag{8.33}$$

and

$$z = \left(\frac{\Delta}{m}\right)\sqrt{\frac{nmL}{2}}. \tag{8.34}$$

8.3.2 Masses and splittings

Quark masses used in the lattice simulation are combined with LEC's and taste splittings to build the needed meson masses. In the code, the resultant meson masses are packaged in lists that coincide with the sums in $\mathcal{T}$, $\mathcal{W}$, and $\mathcal{Q}$, and the arguments of R and D. The meson masses that appear in $\{M_{X_I}\}$, $\{M_{X_{V,A}}\}$, and $\{\mu_{I,V,A}\}$ are

$$\begin{aligned} m_{\eta_I}^2 &= \frac{1}{3}\big(m_{U_I}^2 + 2m_{S_I}^2\big) \\ m_{X_\Xi}^2 &= 2\mu m_x + a^2\Delta_\Xi \\ m_{\eta_{V,A}}^2 &= \frac{1}{2}\Big(m_{U_{V,A}}^2 + m_{S_{V,A}}^2 + \frac{3}{4}a^2\delta'_{V,A} - Z_{V,A}\Big) \\ m_{\eta'_{V,A}}^2 &= \frac{1}{2}\Big(m_{U_{V,A}}^2 + m_{S_{V,A}}^2 + \frac{3}{4}a^2\delta'_{V,A} + Z_{V,A}\Big) \end{aligned} \tag{8.35}$$

with the hairpin splittings

$$Z_{V,A} = \sqrt{(m^2_{S_{V,A}} - m^2_{U_{V,A}})^2 - \frac{a^2\delta'_{V,A}}{2}(m^2_{S_{V,A}} - m^2_{U_{V,A}}) + \frac{9(a^2\delta'_{V,A})^2}{16}}. \tag{8.36}$$

The index $\Xi \in \{I, P, V, A, T\}$ labels the meson taste. Additional masses enter as subscripts of $\mathcal{I}$ and $\mathcal{H}$ (*eg.* $\mathcal{I}_{xu,\Xi}$ with valence x and sea u) and are calculated using Eq. (8.9).

The lists of meson masses used as arguments for the residue functions are $\{M_{X_I}\}$, $\{M_{X_{V,A}}\}$, and $\{\mu_{I,V,A}\}$, given by

$$\begin{aligned} \{M_{X_I}\} &= \{m_{\eta_I}, m_{X_I}\}, \\ \{M_{X_{V,A}}\} &= \{m_{\eta_{V,A}}, m_{\eta'_{V,A}}, m_{X_{V,A}}\}, \text{ and} \\ \{\mu_{I,V,A}\} &= \{m_{U_{I,V,A}}, m_{S_{I,V,A}}\}. \end{aligned} \tag{8.37}$$

The other meson mass lists needed to evaluate the sums $\mathcal{T}$, $\mathcal{W}$, and $\mathcal{Q}$ are

$$\begin{aligned} \{M_{xu,\Xi}\} &= \{m_{xu,I}, m_{xu,P}, m_{xu,V}, m_{xu,A}, m_{xu,T}\}, \\ \{M_{xs,\Xi}\} &= \{m_{xs,I}, m_{xs,P}, m_{xs,V}, m_{xs,A}, m_{xs,T}\}, \text{ and} \\ \{M_{xx,\Xi}\} &= \{m_{xx,I}, m_{xx,P}, m_{xx,V}, m_{xx,A}, m_{xx,T}\}. \end{aligned} \tag{8.38}$$

There are three types of mass splittings. Based on power counting arguments of Ref. [72], the first affects only the pion masses and not the heavy-light meson pole (*ie.* the Δ that enters $\mathcal{H}$):

- Mesons with taste $\Xi \neq P$ have a taste-dependent mass splitting, Δ_Ξ, as in Eq. (8.9). Values for $\Delta_{P,V,T,A,I}$ are obtained from [69].

The remaining two splittings affect heavy-light meson masses and are related to the fact that heavy meson χPT field redefinitions result in meson propagators with masses defined relative to a heavy-light pseudoscalar composed of valence quarks. These two splittings enter $\mathcal{H}$ and its derivatives and are neglected in this work:

- The hyperfine splitting (difference between the vector and pseudoscalar), $m_{B^*_x} - m_{B_x}$ is Δ^*. This splitting is present in continuum χPT and is due to the presence of a virtual vector meson in the sunset and wave function diagrams. Per Ref. [72], $\Delta^*(B) = 45.8$ MeV.

- In a partially quenched calculation, a splitting is possible due a difference in sea, $f \in \{l, s\}$, and valence, x, quarks of the same flavor, δ_{xf}. This splitting only appears in the connected wave function

renormalization diagram. The flavor splitting is calculated by

$$\delta_{xf} = 2\mu\lambda_1(m_f - m_x) \tag{8.39}$$

where $\lambda_1(B) = 0.192\text{GeV}^{-1}$ [69].

8.3.3 Units

Dimensionful quantities, like masses and matrix elements extracted from correlation function fits, appear in the simulation in dimensionless, lattice spacing dependent combinations (*eg.* $am_x = 0.007$). These numbers are said to be in "lattice units". Prior to performing the chiral fits, we convert all dimensionful quantities from lattice units to "r_1 units", where r_1 is defined by the force between static quarks, $r_1^2 F(r_1) = 1.0$ [76, 77]. The ratio r_1/a is measured for each ensemble by the MILC collaboration and the values tabulated in [69]. The latest MILC collaboration determination of r_1 [78] and a useful ratio relating fm to GeV are

$$r_1 = 0.3117(6)(^{+12}_{-31})\text{ fm} \tag{8.40}$$

$$1\text{ GeV}\cdot\text{fm} = 5.06773(12) \tag{8.41}$$

All parameters and data are scaled by appropriate powers of r_1/a and/or r_1 and converted to r_1 units. For example, the decay constant $f = 0.132$ GeV, becomes

$$\begin{aligned} r_1 f &= 0.132\text{ GeV} \times \frac{5.06773}{\text{GeV}\cdot\text{fm}} \times 0.3117\text{ fm} \\ &= 0.2085 \end{aligned} \tag{8.42}$$

With all data and parameters in r_1 units, the chiral fits are performed simultaneously for data at differing lattice spacing. Results of the extrapolation are converted to physical units (*eg.* GeV). Error associated with the measurement of r_1 (setting the scale) comprises part of the error budget for any lattice calculation.

8.3.4 Summary of parameters

Table 8.2 lists parameters determined by the fit and Table 8.3 lists parameters fixed prior to the fit. Fixed parameters, data used in the fit, and priors and widths are converted to r_1 units per the discussion of Sec. 8.3.3.

parameter	$\hat{p}$	$\hat{\sigma}$
g	0.47	0.2
β_4	1	1
β_5	1	1
β'_4	1	1
β'_5	1	1
c_0	0	s
c_1	0	s
c_2	0	s
$a^2 r_1^2 \delta'_V$	c:0, f:0	c:0.07, f:0.02
$a^2 r_1^2 \delta'_A$	c:-0.28, f:-0.098	c:0.06, f:0.021
d_0	0	s^2
d_2	0	s^2

Table 8.2: Parameters to be fit, priors $\hat{p}$, and widths $\hat{\sigma}$. Priors for $\delta'_{V/A}$ depend on lattice spacing and values are given for " c "oarse $a \approx 0.12$ fm and " f "ine $a \approx 0.09$ fm ensembles. Widths for NLO and NNLO LEC's are chosen such that the associated term is limited to a range defined by the characteristic size of NLO chiral logarithms [71, 70], $s = (\frac{m_\pi}{4\pi f})^2 \sim 0.3$.

parameter	value
f	0.132 GeV
Δ_I	c:0.0712222, f:0.0145911
Δ_P	c: 0, f: 0
Δ_V	c:0.0582393, f:0.0107817
Δ_A	c:0.0272273, f:0.00499694
Δ_T	c:0.0433517, f:0.00816435
μ	c:2.270327, f:1.697543
Λ_χ	1 GeV
r_1/a	†
r_1	$0.3117(6)(^{+12}_{-31})$ fm

Table 8.3: Parameters fixed prior to the fit. Where necessary, values are given for " c "oarse $a \approx 0.12$ fm and " f "ine $a \approx 0.09$ fm ensembles. †Values for r_1/a vary with ensemble and lattice spacing, see [69] for values used in this analysis.

8.3.5 Higher orders

At NLO we've found reliable fit results are only possible for relatively light valence quark masses, $(m_\pi r_1)^2 \lesssim 1$. As a result, we've studied fits using combinations of the NNLO analytic terms,

$$d_0 m_x a^2, \quad d_1(2m_u + m_s)a^2, \quad d_2 m_x(2m_u + m_s), \quad d_3 m_x^2, \quad d_4(2m_u + m_s)^2, \quad \text{and} \quad d_5 a^4. \tag{8.43}$$

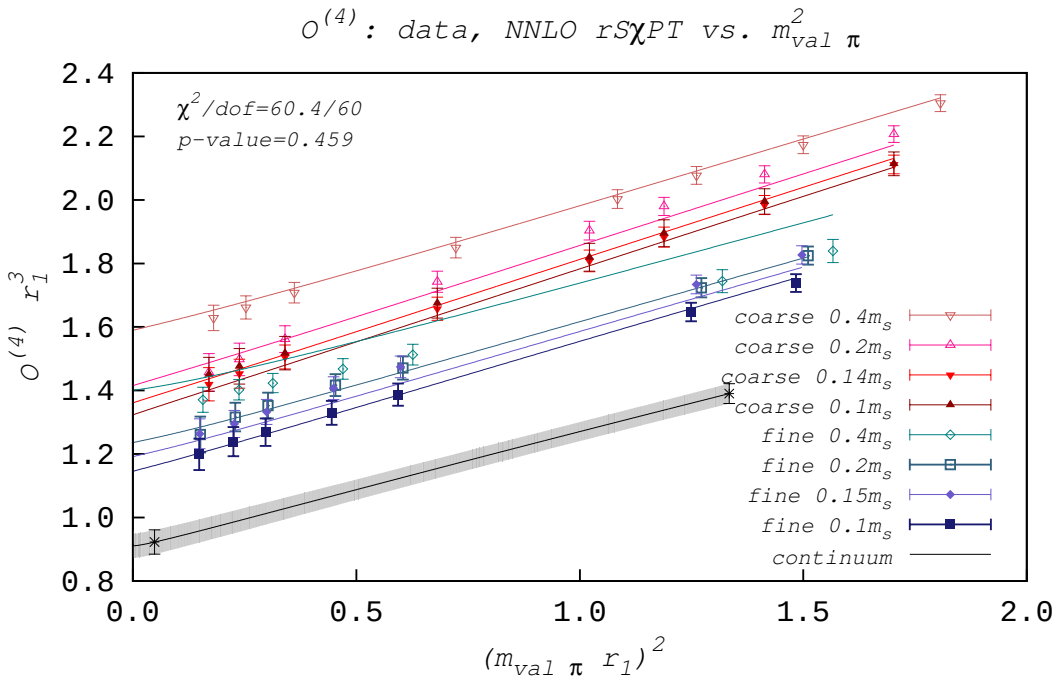
That NNLO terms are required to fit the full range of quark masses has been noted in previous analyses [71, 78, 79]. Selection of priors for NNLO LEC's is guided by [70]. Chiral fits displayed in Fig.'s 8.2 and 8.3 use the NLO expression plus NNLO terms needed to fit the full range of quark masses, $d_0 m_x a^2$ and $d_2 m_x(2m_u + m_s)$.

8.3.6 Chiral fit results

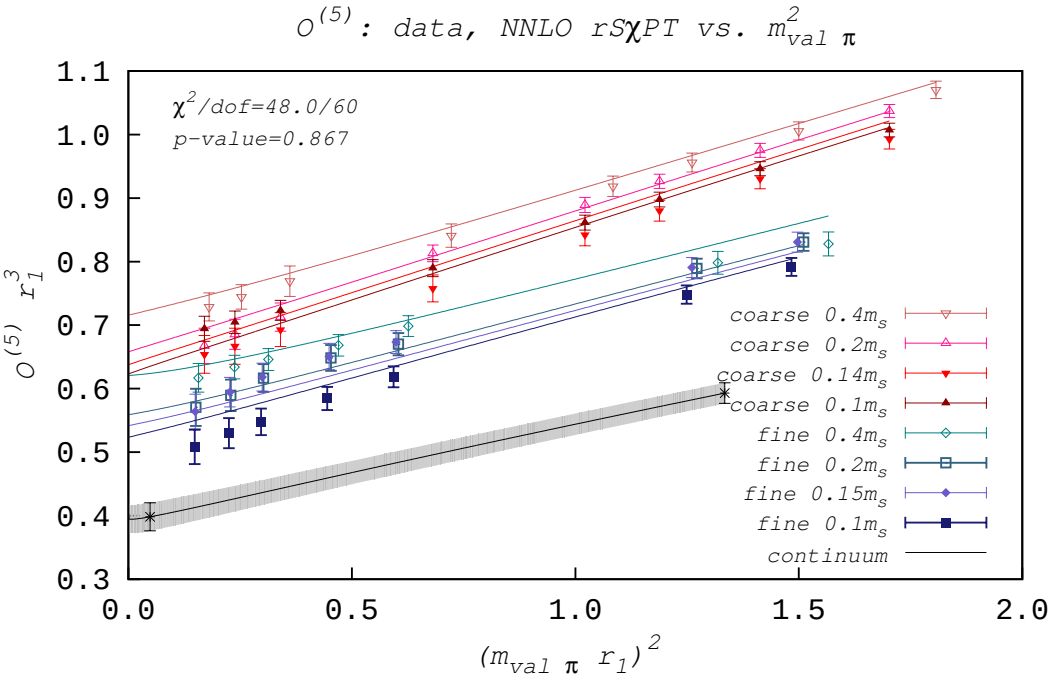
The parameters of Table 8.2 are determined by performing the chiral fits. With these values, the NNLO chiral expression is used to extrapolate (interpolate) to the physical d (s) quark mass, and extrapolate to the continuum.

Fig. 8.2 shows the the data and chiral fit versus valence pion mass. Data points in shades of red represent the $a \approx 0.12$ fm coarse data and shades of blue the $a \approx 0.09$ fm fine data, with darker shades corresponding to more chiral ensembles. The chiral fits are shown fore each ensemble to allow a comparison of the fit and data. The continuum and sea quark extrapolation is given by the black curve, labeled continuum. By performing the chiral fit on each of the 600 bootstrap fit results for the matrix elements, we generate a bootstrap distribution of chiral fit results. The middle 68% of this distribution gives the one sigma error on the chiral fit and generates the gray band. The black bursts to the left (right) in the gray band represent the physical value of the matrix element for the B_d (B_s) meson.

Fig. 8.3 shows the chiral fit and a subset of data versus sea pion mass. The data points in the upper (lower) half of the plot have simulated light valence quark mass closest to the physical s (d) quark mass. Red squares correspond to the $a \approx 0.12$ fm coarse data and blue circles to $a \approx 0.09$ fm fine data. The curves in the upper (lower) half of the plot are evaluated at the physical light valence s (d) quark mass. The black curves represent the light valence quark inter(extra)polation and continuum extrapolation. The gray band represents the one sigma error on the chiral fit. The black bursts to the left in the gray band represent chiral extrapolation to physical light sea quark mass.

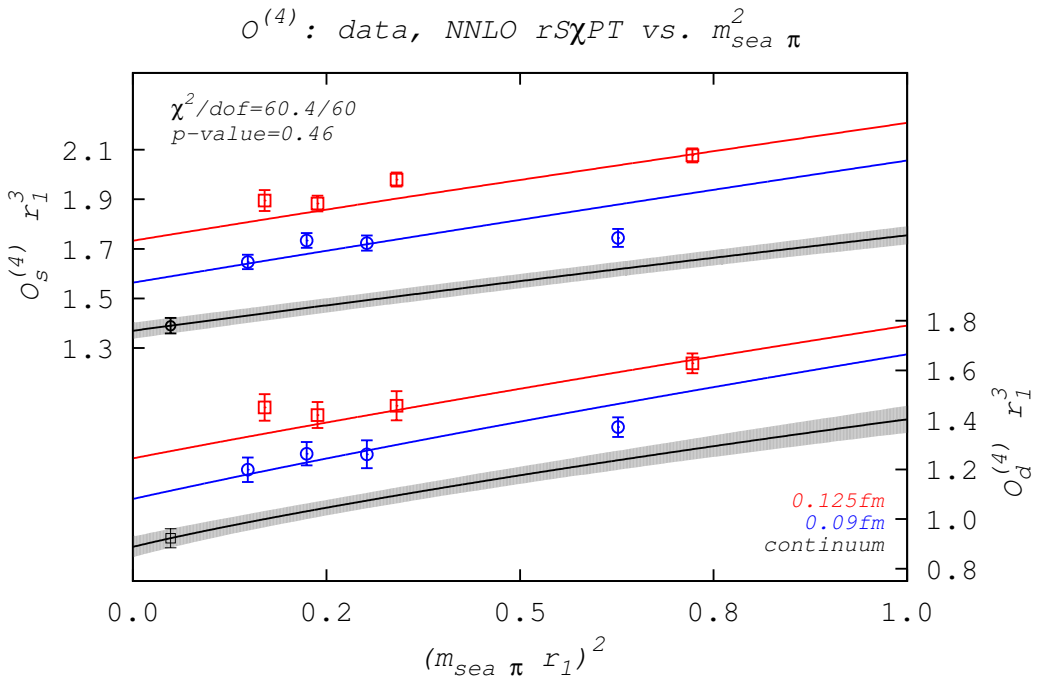


(a) Chiral fit results for mixing operator $\mathcal{O}_4$.

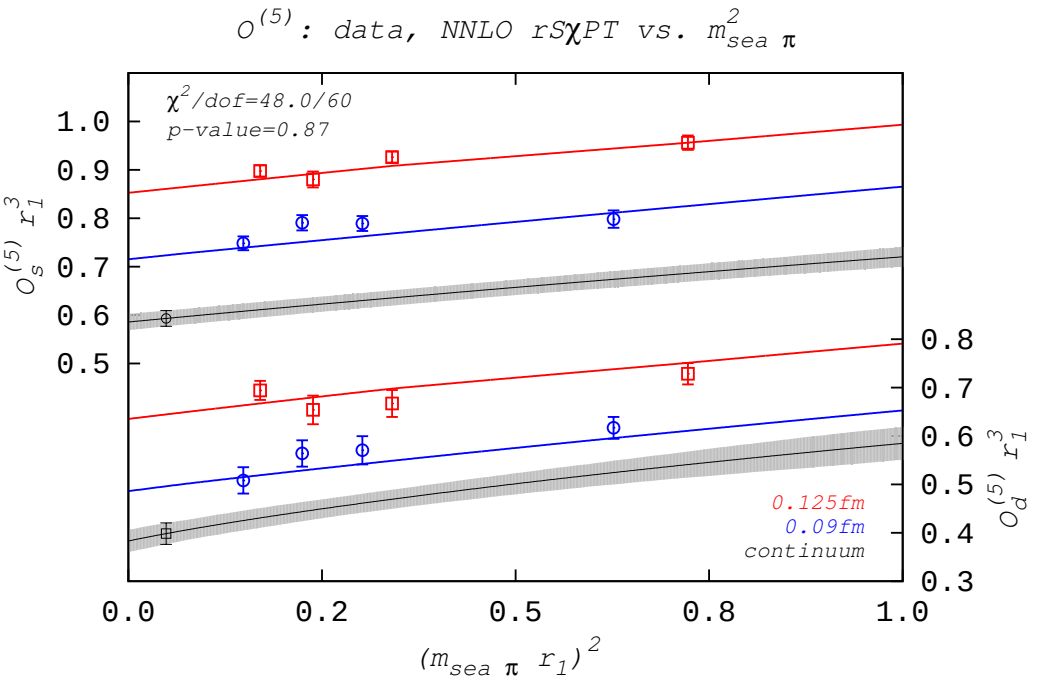


(b) Chiral fit results for mixing operator $\mathcal{O}_5$.

Figure 8.2: All data, chiral fits, and the chiral and continuum extrapolations versus valence pion mass. Data and fit results are shown for the ground state matrix element as defined in Eq. (7.10). Additional details related to the plot can be found in the text.



(a) Chiral fit results for mixing operator $\mathcal{O}_4$.



(b) Chiral fit results for mixing operator $\mathcal{O}_5$.

Figure 8.3: Subset of data and chiral extrapolations versus sea pion mass. Data and fit results are shown for the ground state matrix element as defined in Eq. (7.10). Additional details of the plot can be found in the text.

Chapter 9

Results and outlook

Preliminary chiral and continuum extrapolated values for the hadronic mixing matrix elements are summarized in Table 9.1. The results are preliminary as we intend to add additional data at lattice spacings of $a \approx 0.06$ fm and $a \approx 0.045$ fm. The analysis outlined in Ch.'s 5–8 is expected to remain, largely, unchanged. The addition of data at finer lattice spacings is expected to improve most errors, notably, those associated with perturbative matching and heavy quark discretization errors.

9.1 Error budget

In Table 9.2 we provide an estimated budget of systematic errors to augment the bootstrap errors associated with the results of Table 9.1, with conservative estimates based on the addition of data at lattice spacing of $a \approx 0.06$ fm. These error estimates are made by way of comparison with our collaboration's previous SM B mixing studies [80, 81], where results for errors on the mixing parameter β_1 have been converted (multiplied by two) to give corresponding errors on $O^{(N)} = \mathfrak{c}_N \beta_N^2$.

9.1.1 Sources of error

Statistics

The most straightforward, and computationally expensive, way to reduce statistical errors is to increase the number of gauge field configurations used in the Monte Carlo evaluation of the VEV in Eq. (5.5). We utilize two methods to effectively increase the number of correlation functions available for measurement at minimal computational expense. The first involves packing multiple correlation functions in a single configuration.

	$O^{(4)}\ r_1^3$	$O^{(5)}\ r_1^3$
B_d	0.923(38)	0.398(22)
B_s	1.390(31)	0.593(16)

Table 9.1: Preliminary BSM mixing operator matrix element results, in r_1 units, with bootstrap errors.

Source of Error [%]	Previous	Expected
scale (r_1)	6.0	2.1
light quark masses	0.6	0.6
$g_{B^*B\pi}$	0.6	< 0.6
κ_b	2.2	~ 1.7
χPT $\oplus$ light quark discretization	0.8	< 0.8
heavy quark discretization	4.0	~ 2.7
matching (1-loop perturbation theory)	~ 8.0	~ 5.0
finite volume effects	≤ 1.0	< 1.0
Total	11.1	~ 6.5

Table 9.2: Systematic error budget for the hadronic mixing matrix element, $O^{(N)}$. Previous column entries are derived from previous FNAL–MILC calculations for SM B mixing [80, 81].

The temporal extent of the lattices are large enough that multiple correlation functions "fit" on the lattice associated with each configuration. For each of the ensembles of Table 5.1 four correlation functions are used per configuration. The four correlation functions on a given configuration are averaged together to generate the $N_{\rm cfg}$ configurations used in the analysis. This effectively increases statistics by a factor of two, neglecting autocorrelations. A newly implemented procedure, in place for the production runs of Table 5.1, decreases autocorrelations by randomizing the spatial location of the sources in each configuration. The lack of observed autocorrelations in Sec. 7.1.1 suggests this effort is successful.

Systematics

The analysis of systematic errors is not as straightforward as with statistical errors. Sources of systematic error include errors associated with input quantities, light and heavy quark discretization errors, one loop perturbative matching, and finite volume effects. Several of these are evaluated by varying the chiral fit function.

A lattice QCD calculation, just like a perturbative QCD calculation, requires a host of input values, each of which comes with it's own uncertainty. These errors are typically propagated via the chiral and continuum extrapolation. We vary the inputs according to their error and treat the range of extrapolated results as the resultant error. For example, the error associated with the scale r_1 is estimated by varying r_1 over its error [78] in the scaling of each dimensionful parameter and observing the range of final results from the chiral and continuum extrapolation. A similar procedure is carried out for the light quark masses and the heavy meson χPT coupling, $g_{B^*B\pi}$.

The bottom quark mass is represented by the hopping parameter κ_b (see Eq. (D.8)), tuned to reproduce

the B_s meson mass [82]. Errors associated with this tuning are then translated to errors in the fit parameters by studying the dependence of the fit parameter over a range of κ's [71, 70], *ie.* $\sigma_O^2 = \left(\frac{dO}{d\kappa}\right)^2 \ \sigma_\kappa^2$.

Light quark discretization effects are incorporated in rooted staggered χPT discussed in Sec. 8.2 and the addition of NNLO analytic terms of Sec. 8.3.5. Light quark discretization errors are therefore contained in a systematic error associated with varying the form of the chiral fit function (*ie.* variation of extrapolated results based on the combination of NNLO terms in the fit). The role of NNLO terms in the chiral fits is dependent on the data used in the fit. For example, with data at two lattice spacings the inclusion of the NNLO term, $d_5 a^4$, is unjustified as it would be equivalent to introducing a third unknown to describe two data points. We delay this analysis until after the $a \approx 0.06$ fm data has been analyzed and incorporated in the chiral extrapolation.

Heavy quark discretization effects are separately estimated by power counting arguments. Following the results of the discussion in App. D.2, heavy quark lattice cutoff effects begin at $\mathcal{O}(\alpha_s a\Lambda_{\rm QCD}, (a\Lambda_{\rm QCD})^2)$. The relative size of these effects can be estimated by considering the ratio of lattice spacings, $a \approx 0.09$ fm (the finest lattice used in [80, 81]) and $a \approx 0.06$ fm, giving an improvement factor of $\sim \frac{2}{3}$.

Errors associated with one loop perturbative matching are estimated via conservative estimate of the two loop term in the perturbative expansion of the matrix element, generically written

$$O^{(N)} = O_{LO}^{(N)}(1 + C_{NLO}\ \alpha_s + C_{NNLO}\ \alpha_s^2 + \dots). \tag{9.1}$$

For example, if $C_{NNLO}\ \alpha_s^2 = 0.1$, we'd estimate the one loop truncation error at 10%. In our estimate of the two loop term, we assume $C_{NNLO} \sim 1$ and use the value $\alpha_s = 0.225$ on the $a \approx 0.06$ fm ensembles [54], where α_s is defined and evaluated per the discussion of Sec. 8.1. This estimate yields an error of $\sim 5\%$.

Finite volume effects have typically been negligible in B physics simulations performed on the MILC ensembles [71, 73, 82], *ie.* $< 1.0\%$. By incorporating the finite volume (with hyperfine and flavor splittings) terms in the chiral fit functions, as described in Sec. 8.3, finite volume errors will be incorporated in the bootstrap errors obtained from the chiral and continuum extrapolation.

9.2 Outlook

In the near future we intend to include MILC ensembles with lattice spacings of $a \approx 0.06$ fm and $a \approx 0.045$ fm. The $a \approx 0.06$ fm data should allow us to realize the improvements in the errors discussed above and the inclusion of $a \approx 0.045$ fm data promises to further improve upon those errors. We've coordinated with collaboration members to facilitate a simultaneous analysis of B meson decay constants and the hadronic

mixing matrix elements to allow extraction of the bag parameter (see Eq. (7.16)) in a manner that accurately accounts for correlations in the two analyses. The framework of this analysis is generally applicable to calculating the short distance contributions to heavy-light neutral meson mixing. A planned study of D meson mixing should benefit from this work.

This represents the first calculation of the hadronic contribution to BSM B mixing in nearly a decade [33] and the first ever calculation including the effects of dynamical sea quarks.

Appendix A

Conventions

A.1 Gamma matrices

The representation of euclidean gamma matrices used in the generation of two- and three-point correlation function data is referred to as the MILC, or Degrand-Rossi, basis. The MILC basis gamma matrices and some of their relevant properties are listed here:

$$\gamma_0 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \ \gamma_1 = \begin{bmatrix} 0 & 0 & 0 & i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ -i & 0 & 0 & 0 \end{bmatrix}, \ \gamma_2 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix},$$

$$\gamma_3 = \begin{bmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \\ -i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{bmatrix}, \ \gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{A.1}$$

where $\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}$, $\{\gamma_\mu, \gamma_5\} = 0$, and $\gamma_\mu^\dagger = \gamma_\mu$.

The right/left projection operators are defined by $R/L = \frac{1}{2}(1 \pm \gamma_5)$, and we define $\sigma_{\mu\nu} = \frac{i}{2}[\gamma_\mu, \gamma_\nu]$ so that $\sigma_{\mu\nu}^\dagger = \sigma_{\mu\nu}$. Some useful relations involving the right/left projection operators, $\sigma_{\mu\nu}$, and the gamma

matrices are:

$$\begin{aligned}
\gamma_5 L = L\gamma_5 &= -L \\
\gamma_5 R = R\gamma_5 &= R \\
\gamma_\mu R/L &= L/R\gamma_\mu \\
\gamma_5 \sigma_{\mu\nu} &= \sigma_{\mu\nu}\gamma_5 \\
\gamma_\rho \sigma_{\mu\nu} &= \begin{cases} \sigma_{\mu\nu}\gamma_\rho, & \text{if } \rho \neq \mu, \nu; \\ -\sigma_{\mu\nu}\gamma_\rho, & \text{else} \end{cases} \\
Tr(\gamma_\mu\gamma_\nu\gamma_\rho\gamma_\sigma\gamma_5) &= -4\epsilon_{\mu\nu\rho\sigma}.
\end{aligned} \tag{A.2}$$

A.2 Fourier transformations

In the continuum, with four spacetime dimensions, the Fourier transformations of the fermion field $\psi(x)$ are given by

$$\psi(x) = \frac{1}{(2\pi)^4}\int_{-\infty}^{\infty} d^4k\ \psi(k)e^{ikx} \quad \text{and} \quad \bar{\psi}(x) = \frac{1}{(2\pi)^4}\int_{-\infty}^{\infty} d^4k\ \bar{\psi}(k)e^{-ikx} \tag{A.3}$$

and the inverse transformations of $\psi(k)$ by

$$\psi(k) = \int_{-\infty}^{\infty} d^4x\ \psi(x)e^{-ikx} \quad \text{and} \quad \bar{\psi}(k) = \int_{-\infty}^{\infty} d^4x\ \bar{\psi}(x)e^{ikx}. \tag{A.4}$$

The normalization is consistent with the definition of the Dirac delta function,

$$\int_{-\infty}^{\infty} d^4k\ e^{ik(x-x')} = (2\pi)^4\delta^{(4)}(x-x'). \tag{A.5}$$

The discrete spacetime Fourier transformations replace the integrals by sums over all the spacetime points or accessible wavenumbers,

$$\frac{1}{(2\pi)^4}\int_{-\infty}^{\infty} d^4k \to \frac{1}{V}\sum_{k\in B} \quad \text{and} \quad \int_{-\infty}^{\infty} d^4x \to \sum_{x\in V} \tag{A.6}$$

where V is the spacetime volume. In a simple discrete theory[1], there's a minimum wavelength that can fit on the lattice, $\lambda_{min} = 2a$, and an associated $|k|_{max} = \pi/a$. There's also a maximum wavelength that can fit in the lattice. Ignoring the trivial state with infinite wavelength, the maximum wavelength is set by the

[1] With twisted boundary conditions [83], wavenumbers between zero and π/L are possible.

extent L of the lattice, $\lambda_{max} = 2L$. The associated $|k|_{min} = \pi/L$. The corresponding space of all possible wavenumbers, called the Brillouin zone and denoted B, is given by

$$B = \{k; -\pi \leq ak_\mu \leq \pi, \forall \mu\}. \tag{A.7}$$

Fourier transformations of the fermion field ψ_x, on a lattice are given by

$$\psi_x = \frac{1}{V} \sum_{k \in B} \psi_k \; e^{ikx} \quad \text{and} \quad \bar{\psi}_x = \frac{1}{V} \sum_{k \in B} \bar{\psi}_k \; e^{-ikx} \tag{A.8}$$

and the inverse transformations of ψ_k by

$$\psi_k = \sum_{x \in V} \psi_x \; e^{-ikx} \quad \text{and} \quad \bar{\psi}_k = \sum_{x \in V} \bar{\psi}_x \; e^{ikx}. \tag{A.9}$$

The normalization is consistent with the definition of the Kronecker delta function,

$$\sum_k e^{ik(x-x')} = V \delta_{xx'}. \tag{A.10}$$

Appendix B

Discrete symmetries on the lattice

B.1 Symmetries of Wilson quarks, gauge fields, and propagators

Our mixing operators act on valence quarks. The valence quarks are either naive (in the case of light quarks) or Fermilab heavy quarks (Wilson action with the mass-independent Fermilab interpretation of the Symanzik improvement program). Naive valence quarks are a special case of Wilson quarks. We can, therefore, generically consider the discrete symmetries of the Wilson quarks. The action of the euclidean, discrete lattice symmetries P, T, and C on Wilson quark fields, $\psi(x)$, and gauge links, $U_\mu(x)$, is [84]

$$
\begin{aligned}
P: \quad \psi(x) &\rightarrow \gamma_0 \psi(x^p) \\
\overline{\psi}(x) &\rightarrow \overline{\psi}(x^p)\gamma_0 \\
U_0(x) &\rightarrow U_0(x^p) \\
U_j(x) &\rightarrow U_{-j}(x^p) \qquad \text{(B.1)} \\
T: \quad \psi(x) &\rightarrow \gamma_0\gamma_5 \psi(x^\tau) \\
\overline{\psi}(x) &\rightarrow \overline{\psi}(x^\tau)\gamma_0\gamma_5 \\
U_0(x) &\rightarrow U_{-0}(x^\tau) \\
U_j(x) &\rightarrow U_j(x^\tau) \qquad \text{(B.2)} \\
C: \quad \psi(x) &\rightarrow \mathcal{C}\overline{\psi}^T(x) \\
\overline{\psi}(x) &\rightarrow -\psi^T(x)\mathcal{C}^{-1} \\
U_\mu(x) &\rightarrow U_\mu^*(x) \qquad \text{(B.3)}
\end{aligned}
$$

where x^p is the parity inverted space-time point $x^p = (x_0, -\vec{x})$, x^τ is the time reversed space-time point $x^\tau = (-x_0, \vec{x})$, and $\mathcal{C}\gamma_\mu\mathcal{C} = -\gamma_\mu^* = -\gamma_\mu^T$ where $\mathcal{C} = \gamma_5\gamma_2$ (note the difference here from [84] due to a different choice of basis for the gamma matrices).

When applied to a quark propagator, G, between space-time points x and y, and evaluated on gauge

configuration $[U]$, $G(x,y,[U])$,

$$\begin{aligned} P: \qquad & G(x,y,[U]) \;\rightarrow\; \gamma_0 G(x^p,y^p,[U]^p)\gamma_0 \\ T: \qquad & G(x,y,[U]) \;\rightarrow\; \gamma_0\gamma_5 G(x^\tau,y^\tau,[U]^\tau)\gamma_5\gamma_0 \\ C: \qquad & G(x,y,[U]) \;\rightarrow\; \mathcal{C}G^T(x,y,[U]^c)\mathcal{C}^{-1} \end{aligned} \tag{B.4}$$

where $[U]^p$, $[U]^\tau$, and $[U]^c$ are the P, T, and C transformed gauge configurations, respectively.

There are additional discrete symmetries of Wilson quark propagators related to "γ_5-*hermiticity*", H,

$$\begin{aligned} H: \qquad & G(x,y,[U]) \;\rightarrow\; \gamma_5 G^\dagger(y,x,[U])\gamma_5 \\ CH: \qquad & G(x,y,[U]) \;\rightarrow\; \mathcal{C}\gamma_5 G^*(x,y,[U]^c)\gamma_5 \end{aligned} \tag{B.5}$$

B.2 Symmetries of three point correlation function data

Applying the P and T transformations[1] of Eq. (B.4) to the definition of the open meson propagator in Eq. (6.10) yields

$$\begin{aligned} P: \qquad & E_{ij}^{\alpha\beta}(x,y,[U]) \;\rightarrow\; -\gamma_{0,ik}\; E_{kl}^{\alpha\beta}(x^p,y^p,[U]^p)\;\gamma_{0,lj} \\ T: \qquad & E_{ij}^{\alpha\beta}(x,y,[U]) \;\rightarrow\; -(\gamma_0\gamma_5)_{ik}\; E_{kl}^{\alpha\beta}(x^\tau,y^\tau,[U]^\tau)\;(\gamma_5\gamma_0)_{lj} \end{aligned} \tag{B.6}$$

and finally, applying these transformations to the expressions for $\langle B_q^0|\mathcal{O}_{(\mathrm{no})\mathrm{mix}}|\bar{B}_q^0\rangle_{t_1,t_2}$, we arrive at the P and T transformed expressions for the mixing correlation functions:

[1] Charge conjugation transformations do not produce relationships useful for generating additional data.

$$
\begin{aligned}
P: \quad \langle B_q^0|\mathcal{O}_{\rm nomix}|\bar{B}_q^0\rangle_{t_1,t_2} \;\rightarrow\; & \sum_{\vec{x}_1^p,\vec{x}_2^p} \{{\rm tr}_s\left[\gamma_0\Gamma_1\gamma_0\ {\rm tr}_c E(0,x_2^p,[U]^p)\right]{\rm tr}_s\left[\gamma_0\Gamma_2\gamma_0\ {\rm tr}_c E(0,x_1^p,[U]^p)\right] \\
& -{\rm tr}_s{\rm tr}_c\left[\gamma_0\Gamma_1\gamma_0\ E(0,x_1^p,[U]^p)\ \gamma_0\Gamma_2\gamma_0\ E(0,x_2^p,[U]^p)\right] \\
& -{\rm tr}_s{\rm tr}_c\left[\gamma_0\Gamma_1\gamma_0\ E(0,x_2^p,[U]^p)\ \gamma_0\Gamma_2\gamma_0\ E(0,x_1^p,[U]^p)\right] \\
& +{\rm tr}_s\left[\gamma_0\Gamma_1\gamma_0\ {\rm tr}_c E(0,x_1^p,[U]^p)\right]{\rm tr}_s\left[\gamma_0\Gamma_2\gamma_0\ {\rm tr}_c E(0,x_2^p,[U]^p)\right]\} \\
\\
\langle B_q^0|\mathcal{O}_{\rm mix}|\bar{B}_q^0\rangle_{t_1,t_2} \;\rightarrow\; & \sum_{\vec{x}_1^p,\vec{x}_2^p} \{{\rm tr}_c\left[{\rm tr}_s\left[\gamma_0\Gamma_1\gamma_0\ E(0,x_2^p,[U]^p)\right]{\rm tr}_s\left[\gamma_0\Gamma_2\gamma_0\ E(0,x_1^p,[U]^p)\right]\right] \\
& -{\rm tr}_s\left[\gamma_0\Gamma_2\gamma_0\ {\rm tr}_c E(0,x_2^p,[U]^p)\ \gamma_0\Gamma_1\gamma_0\ {\rm tr}_c E(0,x_1^p,[U]^p)\right] \\
& -{\rm tr}_s\left[\gamma_0\Gamma_1\gamma_0\ {\rm tr}_c E(0,x_2^p,[U]^p)\ \gamma_0\Gamma_2\gamma_0\ {\rm tr}_c E(0,x_1^p,[U]^p)\right] \\
& +{\rm tr}_c\left[{\rm tr}_s\left[\gamma_0\Gamma_2\gamma_0\ E(0,x_2^p,[U]^p)\right]{\rm tr}_s\left[\gamma_0\Gamma_1\gamma_0\ E(0,x_1^p,[U]^p)\right]\right]\}
\end{aligned}
\tag{B.7}
$$

$$
\begin{aligned}
T: \quad \langle B_q^0|\mathcal{O}_{\rm nomix}|\bar{B}_q^0\rangle_{t_1,t_2} \;\rightarrow\; & \sum_{\vec{x}_1,\vec{x}_2} \{{\rm tr}_s\left[\gamma_0\gamma_5\Gamma_1\gamma_5\gamma_0\ {\rm tr}_c E(0,x_2^\tau,[U]^\tau)\right]{\rm tr}_s\left[\gamma_0\gamma_5\Gamma_2\gamma_5\gamma_0\ {\rm tr}_c E(0,x_1^\tau,[U]^\tau)\right] \\
& -{\rm tr}_s{\rm tr}_c\left[\gamma_0\gamma_5\Gamma_1\gamma_5\gamma_0\ E(0,x_1^\tau,[U]^\tau)\ \gamma_0\gamma_5\Gamma_2\gamma_5\gamma_0\ E(0,x_2^\tau,[U]^\tau)\right] \\
& -{\rm tr}_s{\rm tr}_c\left[\gamma_0\gamma_5\Gamma_1\gamma_5\gamma_0\ E(0,x_2^\tau,[U]^\tau)\ \gamma_0\gamma_5\Gamma_2\gamma_5\gamma_0\ E(0,x_1^\tau,[U]^\tau)\right] \\
& +{\rm tr}_s\left[\gamma_0\gamma_5\Gamma_1\gamma_5\gamma_0\ {\rm tr}_c E(0,x_1^\tau,[U]^\tau)\right]{\rm tr}_s\left[\gamma_0\gamma_5\Gamma_2\gamma_5\gamma_0\ {\rm tr}_c E(0,x_2^\tau,[U]^\tau)\right]\} \\
\\
\langle B_q^0|\mathcal{O}_{\rm mix}|\bar{B}_q^0\rangle_{t_1,t_2} \;\rightarrow\; & \sum_{\vec{x}_1,\vec{x}_2} \{{\rm tr}_c\left[{\rm tr}_s\left[\gamma_0\gamma_5\Gamma_1\gamma_5\gamma_0\ E(0,x_2^\tau,[U]^\tau)\right]{\rm tr}_s\left[\gamma_0\gamma_5\Gamma_2\gamma_5\gamma_0\ E(0,x_1^\tau,[U]^\tau)\right]\right] \\
& -{\rm tr}_s\left[\gamma_0\gamma_5\Gamma_2\gamma_5\gamma_0\ {\rm tr}_c E(0,x_2^\tau,[U]^\tau)\ \gamma_0\gamma_5\Gamma_1\gamma_5\gamma_0\ {\rm tr}_c E(0,x_1^\tau,[U]^\tau)\right] \\
& -{\rm tr}_s\left[\gamma_0\gamma_5\Gamma_1\gamma_5\gamma_0\ {\rm tr}_c E(0,x_2^\tau,[U]^\tau)\ \gamma_0\gamma_5\Gamma_2\gamma_5\gamma_0\ {\rm tr}_c E(0,x_1^\tau,[U]^\tau)\right] \\
& +{\rm tr}_c\left[{\rm tr}_s\left[\gamma_0\gamma_5\Gamma_2\gamma_5\gamma_0\ E(0,x_2^\tau,[U]^\tau)\right]{\rm tr}_s\left[\gamma_0\gamma_5\Gamma_1\gamma_5\gamma_0\ E(0,x_1^\tau,[U]^\tau)\right]\right]\}
\end{aligned}
\tag{B.8}
$$

These expressions allow calculation of the P and T transformations of the three point correlation function data. As can be seen from Eq's. (B.7) and (B.8) the calculation that must be done is a manipulation of the gamma matrices $\gamma_0\Gamma_{1,2}\gamma_0$ for P and $\gamma_0\gamma_5\Gamma_{1,2}\gamma_5\gamma_0$ for T. The relations of Eq. (A.2) are useful here.

Appendix C

Gluons on the lattice

Gluons are represented on the lattice by gauge links, $U_{x,\mu}$. Gauge links transform under gauge transformation, V_x, according to

$$U_{x,\mu} \to V_x U_{x,\mu} V^\dagger_{x+a\hat{\mu}}. \tag{C.1}$$

An important consequence of this is the fact that the trace of a product of gauge links around any closed path, a *Wilson loop*, is gauge invariant. Wilson's original [85] lattice gauge action, $S_W[U_\mu]$, is built of Wilson loops on the simplest closed paths on the lattice, squares of side a. The oriented product of link variables needed to construct this path (a 1×1 Wilson loop) is called a plaquette, $P_{x;\mu\nu}$,

$$P_{x;\mu\nu} = U_{x,\mu} U_{x+a\hat{\mu},\nu} U^\dagger_{x+a\hat{\nu},\mu} U^\dagger_{x,\nu} \tag{C.2}$$

where $U^\dagger_{x,\mu} = U_{x+a\hat{\mu},-\mu}$. Built from these plaquettes the Wilson action is

$$S_W[U_\mu] = \frac{2}{g^2} \sum_x \sum_{\mu<\nu} Re\ \mathrm{tr}(1 - P_{x;\mu\nu}) \tag{C.3}$$

where expansion in powers of a reproduces the Yang-Mills action plus $\mathcal{O}(a^2)$ terms.

In the 37 years since Wilson introduced the lattice gauge action, much effort has gone into improving it. A systematic improvement program, developed by Symanzik [35, 36, 37], reduces $\mathcal{O}(a^2)$ effects by adding extended Wilson loops to the gauge action with coefficients computed to one loop in lattice perturbation theory [86, 87]. This allows effects from physics at distances shorter than the lattice spacing to be incorporated.

Perturbation theory plays an important role in lattice calculations. In addition to its role in the Symanzik improvement program, quantities calculated on the lattice are often perturbatively run to an energy scale appropriate for matching with continuum calculations (*cf.* Eq. 3.14). The convergence of lattice perturbation theory is greatly improved by a method referred to as tadpole improvement [38]. Perturbation theory performed as an expansion about the bare lattice coupling converges slowly because of the renormalization

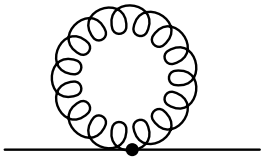


Figure C.1: Gluon tadpole diagram.

of the gauge link, $U_{x,\mu}$, defined in Eq. 5.4, and given approximately by

$$\begin{aligned} U_\mu &\approx e^{iagG_\mu} \\ &\approx 1 + iagG_\mu - a^2 g^2 G^2 + \cdots \end{aligned} \tag{C.4}$$

The G^2 term in the small a expansion induces a $\psi G^2 \bar{\psi}$ vertex that, at the loop level, generates the gluon tadpole diagram of Fig. C.1. This diagram is quadratically divergent but regulated by the lattice cutoff and is therefore proportional to a^{-2}. This cancels the a^2 suppression in Eq. C.4, results in large tadpole contributions to renormalization, and slows convergence. Perturbation theory formulated this way is formally correct but poorly executed. The solution of [38] factors out the UV contribution to the gauge link and replaces it by the mean field value, effectively integrating out the UV modes. This is equivalent to evaluating the coupling constant at a momentum scale at which the tadpole diagrams are suppressed and the theory behaves more like its continuum counterpart, *ie.* its equivalent to not expanding about the bare coupling. The mean field value of the gauge link is defined by the measured value of the plaquette

$$u_0 \equiv \langle \Omega | \tfrac{1}{3} tr P | \Omega \rangle^{1/4} \tag{C.5}$$

and used to rescale all gauge links in our lattice simulation, $U_\mu \to U_\mu / u_0$.

The MILC gauge configurations used in this simulation have Symanzik, tadpole improved glue with lattice artifacts beginning at $\mathcal{O}(\alpha_s a^2)$ and $\mathcal{O}(a^4)$.

Appendix D

Fermions on the lattice

There's a famous (infamous, really) no-go theorem by Nielsen and Ninomiya [88, 89] that plays a significant role in the way quarks are dealt with on the lattice. The cocktail party version of the theorem states that you can't have fermions on the lattice that are simultaneously local, chiral, and free of what's referred to as the *doubling problem*, unwanted fermionic degrees of freedom. The degree to which the theory remains local or chiral can be blurred, giving different approaches to tackling the problem. In the end, either you have doublers (naive and staggered fermions), lose (Wilson fermions) or alter what is meant by (overlap fermions) chirality, or alter what is meant by locality (domain wall fermions).

Poles in the fermion propagator occur when the fermion field satisfies the Dirac equation. Consider a continuum, non-interacting theory in Euclidean metric. We examine the momentum dependence of the propagator's pole structure by Fourier transforming (App. A.2) the field

$$\begin{aligned} 0 &= (\gamma_\mu \partial_\mu + m)\ \psi(x) \\ &= (\gamma_\mu \partial_\mu + m) \int_{-\infty}^{\infty} \frac{d^4k}{(2\pi)^4}\ \psi(k)\ e^{ikx} \\ &= \int_{-\infty}^{\infty} \frac{d^4k}{(2\pi)^4}\ \psi(k)(i\gamma_\mu k_\mu + m)\ e^{ikx} \end{aligned} \tag{D.1}$$

yielding the Euclidean on-shell condition, $-i\not{k} = m$. We discretize the theory by introducing lattice spacing a and replacing the derivative with a finite difference, $\partial_\mu \psi(x) \to \Delta_\mu \psi_x = (\psi_{x+a\hat{\mu}} - \psi_{x-a\hat{\mu}})/2a$, and do the analogous calculation

$$\begin{aligned} 0 &= (\gamma_\mu \Delta_\mu + m)\ \psi_x \\ &= (\gamma_\mu \Delta_\mu + m) \frac{1}{V} \sum_{k \in B} \psi_k\ e^{ikx} \\ &= \frac{1}{V} \sum_{k \in B} \psi_k \left(\frac{\gamma_\mu}{2a} (e^{ik_\mu a} - e^{-ik_\mu a}) + m \right)\ e^{ikx} \\ &= \frac{1}{V} \sum_{k \in B} \psi_k \left(i\gamma_\mu \frac{\sin(k_\mu a)}{a} + m \right)\ e^{ikx} \end{aligned} \tag{D.2}$$

we find the discrete version of the on-shell condition, $-i\gamma_\mu \sin(k_\mu a) = ma$. Taking the magnitude squared of both sides gives

$$\sin^2(k_\mu a) = (ma)^2. \tag{D.3}$$

Within the range of accessible wavenumbers (see App. A.2) there are multiple potential solutions, provided $\sin(\pi a/L) \leq am \leq 1$. Simplifying to one dimension, we have $\sin(ka) = \pm ma$. Fig. D.1 shows the location

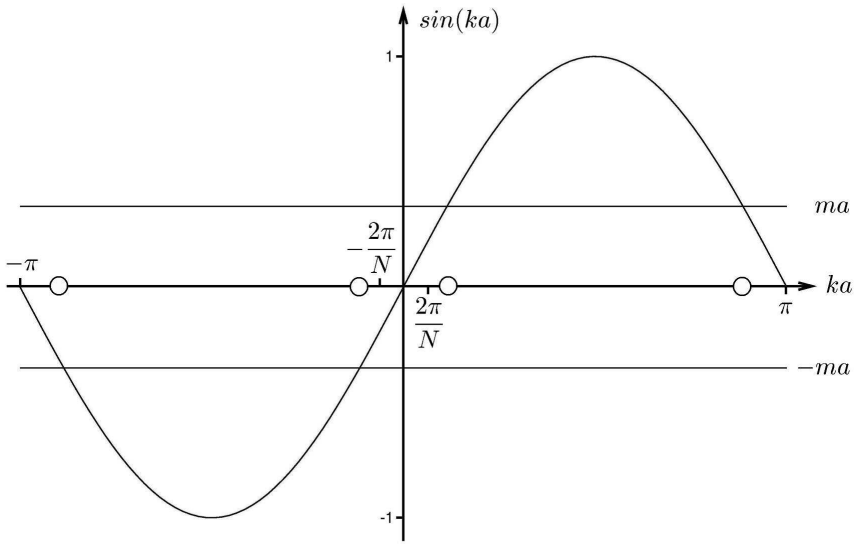


Figure D.1: Poles, circles on the ka-axis, in the one dimensional, discrete fermion propagator reveal the appearance of extra fermions, *doublers*, at large values momenta near the lattice cutoff, $a|k| \sim \pi$.

of continuum-like poles at small $|k|$ and two discrete spacetime specific poles at large $|k|$. In d dimensions, there are a total of 2^d *tastes* of fermions that live at the corners of the Bruillon zone.

D.1 Naive action

The naive discretized fermion action simply replaces the covariant derivative in the Dirac operator with a covariant finite difference operator,

$$\begin{aligned} S_{\text{niave}} &= \sum_x \bar{\psi}_x(\gamma_\mu \Delta_\mu + m)\psi_x \\ &= \sum_x \bar{\psi}_x \left[\sum_\mu \frac{\gamma_\mu}{2a} \left(U_{\mu,x}\psi_{x+a\hat{\mu}} - U^\dagger_{\mu,x-a\hat{\mu}}\psi_{x-a\hat{\mu}} \right) + m\psi_x \right] \end{aligned} \tag{D.4}$$

which suffers from the doubling problem.

D.2 Fermilab action

Wilson recognized the appearance of doublers and proposed a solution [45]. So called Wilson fermions are obtained by adding a second derivative, dimension five term to the naive action,

$$\begin{aligned} S_{\text{Wilson}} &= S_{\text{niave}} - \frac{ar}{2}\sum_x \bar{\psi}_x \sum_\mu \nabla_\mu \psi_x \\ &= S_{\text{niave}} - \frac{r}{2a}\sum_x \bar{\psi}_x \sum_\mu \Big[U_{\mu,x}\psi_{x+a\hat{\mu}} + U^\dagger_{\mu,x-a\hat{\mu}}\psi_{x-a\hat{\mu}} - 2\psi_x\Big], \end{aligned} \tag{D.5}$$

where r is a free parameter typically set to one. This operator, which goes to zero in the $a \to 0$ limit, adds to the doublers' mass a term proportional to a^{-1}. In the continuum limit, the doublers become infinitely massive and decouple from the theory. An additive mass term, however, explicitly breaks chiral symmetry. The chiral dynamics of light quarks, namely the protection of quark masses from additive renormalization, is lost when using the Wilson fermion action. Wilson fermions, and their variants discussed below, are therefore typically used to simulate heavy quarks.

The additional term, along with other dimension five operators generated via operator mixing under renormalization, generates lattice artifacts at $\mathcal{O}(a)$. To remove these artifacts and improve the Wilson fermion action to $\mathcal{O}(a^2)$, Sheikholeslami and Wohlert [90] added a second dimension five operator with coefficient, c_{SW}, calculated [91, 92] to eliminate the $\mathcal{O}(a)$ effects. The Wilson fermion action plus this second dimension five operator, is typically referred to as the *clover action* because of the clover leaf pattern of plaquettes used to build the lattice field strength tensor, $\mathcal{F}_{\mu\nu,x}$,

$$S_{\text{clover}} = S_{\text{Wilson}} + \frac{iag}{4} c_{SW} \sum_x \bar{\psi}_x \sigma_{\mu\nu} \mathcal{F}_{\mu\nu,x} \psi_x. \tag{D.6}$$

With c_{SW} calculated to one-loop in lattice perturbation theory, discretization effects enter at $\mathcal{O}(a^2, a\alpha_s)$. The clover action is typically simulated in the *hopping parameter* form

$$S_{\text{clover}} = \sum_n \bar{\psi}_n \Bigg\{ \psi_n - \kappa \sum_\mu \Big[(r-\gamma_\mu)U_{\mu,n}\psi_{n+\hat{\mu}} + (r+\gamma_\mu)U^\dagger_{\mu,n-\hat{\mu}}\psi_{n-\hat{\mu}}\Big] + i\kappa\, c_{SW} \sum_{\mu\nu} \sigma_{\mu\nu}\mathcal{F}_{\mu\nu,n}\psi_n \Bigg\} \tag{D.7}$$

where

$$\psi_n = \frac{a^{3/2}}{\sqrt{2\kappa}}\psi_x \qquad \text{and} \qquad \kappa = \frac{1}{2(am_0 + 1 + 3r)}. \tag{D.8}$$

The clover action is tadpole improved, as outlined in App. C. In addition to scaling the gauge links, the clover action terms are rescaled by $\kappa \to u_0\kappa$ and $c_{SW} \to c_{SW}/u_0^3$. In our simulations, the hopping parameter, κ, is a dimensionless number representing the bare heavy quark mass, am_0.

The final modification to our treatment of heavy quarks is referred to as the Fermilab interpretation [44]. It uses heavy quark effective theory arguments to incorporate short distance, heavy quark effects in improvement coefficients via a systematic expansion in powers of $\Lambda_{\rm QCD}/m_Q$, where m_Q is the heavy quark mass. In general, spatial and temporal directions as well as chromomagnetic and chromoelectric components are treated separately, resulting in improvement coefficients κ_t, κ_s, r_s, c_B, and c_E,

$$\begin{aligned} S_{\rm FNAL} \;=\; & \sum_n \Big\{ \bar\psi_n\psi_n - \kappa_t\Big[\bar\psi_n(1-\gamma_0)U_{0,n}\psi_{n+\hat 0} + \bar\psi_n(1+\gamma_0)U^\dagger_{0,n-\hat 0}\psi_{n-\hat 0}\Big] \\ & -\kappa_s\sum_i\Big[\bar\psi_n(r_s-\gamma_i)U_{i,n}\psi_{n+\hat i} + \bar\psi_n(r_s+\gamma_i)U^\dagger_{i,n-\hat i}\psi_{n-\hat i}\Big] \\ & +\frac{i}{2}\kappa_s c_B\sum_{ijk}\epsilon_{ijk}\bar\psi_n\sigma_{ij}B_{k,n}\psi_n + i\kappa_s c_E\sum_i\bar\psi_n\sigma_{0i}E_{i,n}\psi_n\Big\}. \end{aligned} \tag{D.9}$$

The coefficients are bounded functions of am_0 and undergo a smooth transition between light and heavy quark limits.

We must tune the simulation parameters so our simulated b quark corresponds to the physical bottom quark. Details of the the tuning procedure are given in [82] and summarized here. The dispersion relation for a heavy meson on the lattice [44]

$$E^2(\vec p) = M_1^2 + \frac{M_1}{M_2}\vec p^{\,2} + \cdots \tag{D.10}$$

gives rise to lattice *rest* and *kinetic* meson masses, given by

$$M_1 = E(0) \qquad \text{and} \qquad M_2^{-1} = \left.\frac{\partial E(\vec p)}{\partial \vec p^{\,2}}\right|_{\vec p=0}, \tag{D.11}$$

respectively. In the limit $am_0 \to 0$, $M_1 = M_2$. For the b quark, however, $am_0 \sim 1$ and $M_1 \neq M_2$. While it is possible to tune the improvement coefficients to achieve $M_1 = M_2$, observables of interest are independent of M_1 [93]. We therefore take $\kappa = \kappa_t = \kappa_s$ and $c_{SW} = c_E = c_B$ and tune κ to achieve a value for M_2 consistent with the experimentally observed B_s meson mass. M_1 governs the exponential decay of correlation function

data and the quantity determined by the effective mass analysis of Sec. 7.1.2.

In the Fermilab action, cutoff effects begin at $\mathcal{O}(\alpha_s a\Lambda_{\rm QCD}, (a\Lambda_{\rm QCD})^2)$. For heavy quarks such that $am_Q \sim 1$, heavy quark discretization errors begin at $\mathcal{O}(\alpha_s \Lambda_{\rm QCD}/m_Q, (\Lambda_{\rm QCD}/m_Q)^2)$.

D.3 Asqtad staggered action

To simulate light quarks in a manner that preserves chiral dynamics we use staggered fermions [40, 41, 42, 43]. The staggered action is obtained by first spin-diagonalizing naive fermions, $\psi \to \psi'$, via the Kawamoto-Smit [94] transformation

$$\psi_x = \Gamma_x \psi'_x, \qquad \bar{\psi}_x = \bar{\psi}'_x \Gamma^\dagger_x \tag{D.12}$$

where the transformation matrix

$$\Gamma_x = \gamma_0^{(x_0/a)} \gamma_1^{(x_1/a)} \gamma_2^{(x_2/a)} \gamma_3^{(x_3/a)} \tag{D.13}$$

is hermitian, $\Gamma_x \Gamma^\dagger_x = 1$, and reduces gamma matrices to a phase

$$\eta_{\mu,x} \equiv \Gamma^\dagger_x \gamma_\mu \Gamma_{x+a\hat{\mu}} = (-1)^{(x_0+\cdots+x_{\mu-1})/a}. \tag{D.14}$$

Under this transformation, the naive fermion action becomes the Kogut-Susskind action

$$S_{\rm KS} = \sum_x \bar{\psi}'_x (\eta_{\mu,x} \Delta_\mu + m) \psi'_x \tag{D.15}$$

with Δ_μ defined as in Eq. (D.4). Note that $\eta_{\mu,x}$, and therefore $S_{\rm KS}$, is spin-diagonal. Like ψ, the fermion field ψ' is a four component Dirac spinor. However, the spinor components of ψ' don't interact with one another. We reduce the number of doublers by a factor of four by considering only one component. The one component field is the staggered fermion, χ. The staggered action

$$S_{\rm staggered} = \sum_x \bar{\chi}_x (\eta_{\mu,x} \Delta_\mu + m) \chi_x \tag{D.16}$$

then has four doubler degrees of freedom called *tastes*.

The chiral dynamics of staggered quarks protects them from additive mass renormalization at the expense of these extra degrees of freedom. This chiral behavior is guaranteed by a combination of an even-odd

$U(1)_e \times U(1)_o$ remnant of the continuum chiral symmetry [94, 95, 96],

$$\begin{aligned} \chi_x &\rightarrow e^{i\alpha_e}\chi_x, \quad \bar{\chi}_x \rightarrow \bar{\chi}_x e^{-i\alpha_o} \quad \text{for } \sum_\mu (x_\mu/a) \text{ even, and} \\ \chi_x &\rightarrow e^{i\alpha_o}\chi_x, \quad \bar{\chi}_x \rightarrow \bar{\chi}_x e^{-i\alpha_e} \quad \text{for } \sum_\mu (x_\mu/a) \text{ odd} \end{aligned} \tag{D.17}$$

and shift symmetry [97, 98]

$$\chi_x \rightarrow \rho_{\mu,x}\chi_{x+a\hat{\mu}}, \quad \bar{\chi}_x \rightarrow \rho_{\mu,x}\bar{\chi}_{x+a\hat{\mu}}, \quad U_{\nu,x} \rightarrow U_{\nu,x+a\hat{\mu}} \tag{D.18}$$

where $\rho_{\mu,x} = (-1)^{x_{\mu+a}+\cdots+x_3}$.

If taste degrees of freedom didn't interact with one another, we could isolate a single taste and effectively remove all doubler degrees of freedom. However, there are $\mathcal{O}(a^2)$ taste-changing interactions from large momentum gluons. These taste-changing interactions are suppressed by smearing the gauge links in the covariant finite difference operator Δ_μ in Eq. (D.16), effectively suppressing the coupling of staggered fermions to large momentum gluons. Remaining $\mathcal{O}(a^2)$ taste violations are removed by the inclusion of a Naik term [99]. The resulting action is improved to $\mathcal{O}(\alpha_s a^2)$ and $\mathcal{O}(a^4)$ and is referred to as the asq staggered action. Finally, tadpole improvement, outlined in App. C, produces the asqtad staggered action.

In Eq. (5.3) we got a factor of $det(\not{D} + m)$ from the Berezin integration of the single species of fermion field considered. Integration over multiple fermion fields, *ie.* $\int \cdots \int \Pi_{i=1}^{N} [d\psi_i][d\bar{\psi}_i]$, each with Dirac operator $\not{D} + m_i$, would generate a factor of $\Pi_{i=1}^{N} det(\not{D} + m_i)$. To remove the effects of the the four tastes of fermions in the asqtad action, we take the fourth-root of this factor - effectively reducing the number of sea quarks in our QCD vacuum from four per flavor (the four tastes) to one per flavor. Fermion tastes are not degenerate in mass due to taste violating effects. Use of the asqtad action suppresses these differences to $\mathcal{O}(a^2)$. The fourth-root trick is controversial (*cf.* [100], [101], and references therein).

D.4 Heavy-light mesons

We simulate heavy-light mesons using the Fermilab action for the heavy bottom quark and the asqtad action for the light (down or strange) quark. The effect of staggered light quarks on the heavy-light operators has been studied in detail [102]. Here we summarize the salient points of that analysis, focusing on the time-dependence. Minor notational differences exist between the discussion here and that in [102]. These differences are largely due to an effort to maintain consistency between this discussion and that in App. D.3 and choice of Fourier transformation conventions.

First, we focus on some relevant properties of naive fermions. From the transformations of Eq. (D.12), the propagators for the four component naive spinor (ψ) and the single component staggered spinor (χ) can be related,

$$G_\psi(x,y) = \Gamma_x \Gamma_y^\dagger G_\chi(x,y), \tag{D.19}$$

allowing us to easily switch between (theoretically simple and easy to interpret) naive fermions when constructing heavy-light bilinears and (cheaper to simulate due to fewer degrees of freedom) staggered fermions in the simulation.

To develop the time-dependence we work with free, naive fermion action

$$S_0 = \sum_{k\in B} \bar{\psi}_k \left(\frac{i\gamma_\mu}{a} \sin(k_\mu a) + m \right) \psi_k \tag{D.20}$$

and begin with a symmetry of naive fermions called *doubling* symmetry,

$$\psi_x \to e^{ix\pi_g} M_g \psi_x \quad \text{and} \quad \bar{\psi}_x \to e^{ix\pi_g} \bar{\psi}_x M_g^\dagger, \tag{D.21}$$

where the four vector

$$a\pi_g \in \{(0,0,0,0), (\pi,0,0,0), \ldots, (\pi,\pi,\pi,\pi)\} \tag{D.22}$$

represents one of the 16 corners of the four dimensional Brillouin zone. The transformation matrix is defined by the product of gamma matrices

$$M_g = \prod_{\mu\in g} i\gamma_5\gamma_\mu \tag{D.23}$$

where g is an element of the set of ordered lists of indices used to build the collection of $a\pi_g$. To be specific, the values of g corresponding to Eq. (D.22) are

$$g \in \{\emptyset, (0), \ldots, (0,1,2,3)\}, \tag{D.24}$$

ie. where a π goes. In what follows we specify the "low momentum" central region of the Brillouin zone, $B_\emptyset = \{k; -\pi \le 2ak_\mu \le \pi, \forall\mu\}$. Note that of all the π_g, only $\pi_\emptyset \in B_\emptyset$. Using the π_g, Eq. (D.20) is re-written as a sum over the momenta in $B_\emptyset$, with remaining momenta in B included by summing over momentum shifts via π_g,

$$S_0 = \sum_g \sum_{k\in B_\emptyset} \bar{\psi}_{k+\pi_g} \left(\frac{i\gamma_\mu}{a} \sin([k+\pi_g]_\mu a) + m \right) \psi_{k+\pi_g}. \tag{D.25}$$

By introducing new momentum space spinors q^g with taste g,

$$q_k^g = M_g \psi_{k+\pi_g} \quad \text{and} \quad \bar{q}_k^g = \bar{\psi}_{k+\pi_g} M_g^\dagger \tag{D.26}$$

and using

$$M_g \gamma_\mu M_g^\dagger \sin([k+\pi_g]_\mu a) = \gamma_\mu \sin(k_\mu a) \tag{D.27}$$

the naive fermion action can be re-written

$$S_0 = \sum_g \sum_{k \in B_\emptyset} \bar{q}_k^g \left(\frac{i\gamma_\mu}{a} \sin(k_\mu a) + m \right) q_k^g. \tag{D.28}$$

making apparent the contributions from different tastes g.

Now we introduce a heavy quark field Ψ from which we build the heavy-light meson, $\mathcal{W} = \bar{\Psi}\gamma_5\psi$,

$$\begin{aligned} \mathcal{W}_{\vec{x},t} &= \frac{1}{L^3} \sum_{\vec{k} \in B_s} \mathcal{W}_{\vec{k},t}\, e^{i\vec{k}\cdot\vec{x}} \\ &= \frac{1}{L^3} \sum_{\vec{k} \in B_s} \bar{\Psi}_{\vec{k},t} \gamma_5 \psi_{\vec{k},t}\, e^{i\vec{k}\cdot\vec{x}} \end{aligned} \tag{D.29}$$

where $B_s = \{\vec{k}; -\pi \leq a k_i \leq \pi, \forall i\}$ is the three momentum subspace of B. Trading the sum over all momenta for a sum over the central region with an associated sum over momentum shifts, we get

$$\mathcal{W}_{\vec{x},t} = \frac{1}{L^3} \sum_{g_s} \sum_{\vec{k} \in B_{s,\emptyset}} \bar{\Psi}_{\vec{k}+\vec{\pi}_{g_s},t} \gamma_5 \psi_{\vec{k}+\vec{\pi}_{g_s},t}\, e^{i(\vec{k}+\vec{\pi}_{g_s})\cdot\vec{x}}. \tag{D.30}$$

Focusing on the time-dependence of the light quark in this expression, we write

$$\begin{aligned} \psi_{\vec{k}+\vec{\pi}_{g_s},t} &= \frac{1}{T} \sum_{k_0 \in B_t} e^{ik_0 t}\, \psi_{\vec{k}+\vec{\pi}_{g_s},k_0} \\ &= \frac{1}{T} \sum_{k_0 \in B_{t,\emptyset}} [e^{ik_0 t}\, \psi_{\vec{k}+\vec{\pi}_{g_s},k_0} + e^{i(k_0+\pi/a)t}\, \psi_{\vec{k}+\vec{\pi}_{g_s},k_0+\pi/a}] \end{aligned} \tag{D.31}$$

where the sum over the two possible π_{g_t} shifts, 0 and π/a, is written explicitly. Using the transformations of Eq. (D.26), we write

$$\psi_{\vec{k}+\vec{\pi}_{g_s},k_0} = M_{g_s}^\dagger q_{\vec{k},k_0}^{g_s} \quad \text{and} \quad \psi_{\vec{k}+\vec{\pi}_{g_s},k_0+\pi/a} = M_{g_t}^\dagger M_{g_s}^\dagger q_{\vec{k},k_0}^{g_t,g_s}. \tag{D.32}$$

Plugging these light quark expressions into Eq. (D.31) and then plugging the result into the expression for

the heavy-light meson in Eq. (D.30) gives

$$\mathcal{W}_{\vec{x},t} = \frac{1}{V}\sum_{g_s}\sum_{k\in B_\emptyset} \bar{\Psi}_{\vec{k}+\vec{\pi}_{g_s},t}\gamma_5\Big[e^{i(\vec{k}+\vec{\pi}_{g_s})\cdot\vec{x}}\ \Big(e^{ik_0 t}\ M^\dagger_{g_s} q^{g_s}_{\vec{k},k_0} + e^{i(k_0+\pi/a)t}\ M^\dagger_{g_t} M^\dagger_{g_s} q^{g_t,g_s}_{\vec{k},k_0}\Big)\Big]. \tag{D.33}$$

Doubling symmetry allows us to recast a light quark with shifted momentum as a light quark of different taste and unshifted momentum. This is not the case for our non-staggered heavy quark, where the momentum shift pushes the heavy quark far off shell. Therefore, up to lattice artifacts of $\mathcal{O}(\alpha_s a^2)$, we can neglect the $g_s \neq \emptyset$ contributions and write,

$$\mathcal{W}_{\vec{x},t} = \frac{1}{V}\sum_{k\in B_\emptyset} \bar{\Psi}_{\vec{k},t}\gamma_5\Big(q_{\vec{k},k_0} + (-1)^t M^\dagger_{g_t} q^{g_t}_{\vec{k},k_0}\Big)e^{ikx}, \tag{D.34}$$

with temporal oscillations characteristic of staggered fermions.

Appendix E

Statistics

E.1 Basic error estimators

Starting with a collection of N measurements $\{x_i \mid i = 1, \ldots, N\}$ the basic statistical quantities of *sample mean* and *sample variance* are:

$$\begin{aligned} \bar{x} &= \frac{1}{N}\sum_{i=1}^{N} x_i \\ \sigma_x^2 &= \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \end{aligned} \tag{E.1}$$

and the *standard deviation* of the $\{x_i\}$, σ_x, is given by the square root of the sample variance. Note that σ_x has the same units as x and provides an absolute measure of the spread of the x's about $\bar{x}$. This must be kept in mind when comparing standard deviations (or errors) of different measurements. To allow an adequate comparison of standard deviations its useful to talk of the relative standard deviation, ϵ. For our x's, the relative standard deviation is

$$\epsilon_x = \frac{\sigma_x}{\bar{x}} \tag{E.2}$$

and can be reported as a percentage since relative standard deviation has no units. The standard deviation is useful if what you're interested in is the spread of the x's about the sample mean. However, if you're interested in how close the mean value of your measurements is to "truth", then the standard deviation is an over-estimate of this. We'll discuss how to approximate the error of our estimate of the mean below.

Next we generalize and say we have a collection of N measurements of each of a bunch of quantities, say $\{x_i, y_i, \ldots\}$. We could then speak separately of the sample means, sample variances and standard deviations of the x's, the y's and so on. If, however, the y's were nothing more than another set of measurements of the same thing the x's measured, and the same thing for the z's, and on and on, then we could either choose to just treat all the x's, y's, *etc.*, together or, and this will be the subject of Sec.'s E.3 and E.3.2, we could first average the x's, then average the y's, *etc.*, then treat $\{\bar{x}, \bar{y}, \ldots\}$ as a collection of measurements and do

our statistical analysis on them.

Suppose we want to take our collection of measurements and turn them into the quantity we're really after, some function of what we measured, $f_i = f(x_i, y_i, \dots)$. We need to understand how uncertainties in our measurements translate into an uncertainty in f. We start by considering the mean[1] and variance of f.

$$\begin{aligned} \bar{f} &= \frac{1}{N}\sum_{i=1}^{N} f_i \\ \sigma_f^2 &= \frac{1}{N-1}\sum_{i=1}^{N}(f_i - \bar{f})^2 \end{aligned} \tag{E.3}$$

The variance of f is what we're after and to get explicit dependence on what we actually measure, we Taylor expand f_i about its mean

$$f_i - \bar{f} = (x_i - \bar{x})\frac{\partial \bar{f}}{\partial x} + (y_i - \bar{y})\frac{\partial \bar{f}}{\partial y} + \cdots \tag{E.4}$$

where $\bar{f}$ and lack of indices in the partial derivatives mean we differentiate and then evaluate the result at $\{x_i, y_i, \dots\} = \{\bar{x}, \bar{y}, \dots\}$. Now we take this Taylor expansion and plug it into our expression for σ_f^2. After some algebra we get

$$\sigma_f^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[(x_i - \bar{x})^2\left(\frac{\partial \bar{f}}{\partial x}\right)^2 + (y_i - \bar{y})^2\left(\frac{\partial \bar{f}}{\partial y}\right)^2 + \cdots + 2(x_i - \bar{x})(y_i - \bar{y})\frac{\partial \bar{f}}{\partial x}\frac{\partial \bar{f}}{\partial y} + \cdots\right]. \tag{E.5}$$

Using the expression for the variance in Eq. (E.1) we re-write this as

$$\sigma_f^2 = \sigma_x^2\left(\frac{\partial \bar{f}}{\partial x}\right)^2 + \sigma_y^2\left(\frac{\partial \bar{f}}{\partial y}\right)^2 + \cdots + \sum_{i=1}^{N}\left[2\frac{(x_i - \bar{x})(y_i - \bar{y})}{N-1}\frac{\partial \bar{f}}{\partial x}\frac{\partial \bar{f}}{\partial y}\right] + \cdots \tag{E.6}$$

We call (part of) the cross-term the *covariance* of x and y,

$$\sigma_{xy}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}), \tag{E.7}$$

it measures for us the degree to which the x's and y's are interrelated. Now we have our answer about the error in f due to errors in our measurements of $\{x_i, y_i, \dots\}$, its called the *error propagation equation*

$$\sigma_f^2 = \sigma_x^2\left(\frac{\partial \bar{f}}{\partial x}\right)^2 + \sigma_y^2\left(\frac{\partial \bar{f}}{\partial y}\right)^2 + \cdots + 2\sigma_{xy}^2\frac{\partial \bar{f}}{\partial x}\frac{\partial \bar{f}}{\partial y} + \cdots \tag{E.8}$$

We now add a third item, in addition to the standard deviation and relative error, to the list of quantities

[1] We could alternatively have defined $\bar{f} = f(\bar{x}, \bar{y}, ...)$.

we can use when reporting estimates of the errors in our measurements. The *standard error*, $\sigma_{\bar{x}}$, is an estimate of the error of the mean in which we use the error propagation formula with $f = \bar{x}$ and make a few assumptions. First, we drop all the higher order terms represented by the ellipses. Second, we assume the covariance terms are all zero. This gives

$$\sigma_{\bar{x}}^2 \approx \sum_{i=1}^{N} \sigma_{i,x}^2 \left(\frac{\partial \bar{x}}{\partial x_i} \right)^2 \tag{E.9}$$

where $\sigma_{i,x}^2$ is the variance of x for the i^{th} configuration. Now if we assume all the $\{x_i\}$ have the same variance, σ_x^2 and use the expression for the mean to evaluate the derivative we get

$$\sigma_{\bar{x}} \approx \frac{\sigma_x}{\sqrt{N}}. \tag{E.10}$$

Plugging in $\bar{x}$ for f might seem like nonsense if you go back and look at Eq. (E.3), since there we take the difference $f_i - \bar{f}$ and that seems like it must be zero if f is an average value. What this means is there's some *actual* mean value, known only to nature (sometimes referred to as the population mean and written μ_x). The population mean is $\bar{f}$ in Eq. (E.3). Then we come along and do our experiment to try and figure out what μ_x is, but the best we can do is $\bar{x}$, often referred to in this context as the sample mean. The standard error gives us an idea of how close $\bar{x}$ might be to μ_x. An example to illustrate the difference between population and sample mean is the average height of humans on Earth. We could never get around to making all those measurements and really finding out the population mean. Its more feasible to imagine we measure several hundred people from each country and get an average for each country. Then we'd treat the country averages as our $\{x_i\}$ and, assuming the $\{\sigma_i\}$ for the countries were all the same, estimate the error on our sample mean using the standard error, $\sigma_{\bar{x}}$.

Note that standard error, like standard deviation, is an absolute error. That is, it has the units of x. We could also talk about the *relative standard error* and label it $\epsilon_{\bar{x}}$,

$$\epsilon_{\bar{x}} = \frac{\sigma_{\bar{x}}}{\bar{x}\sqrt{N}}. \tag{E.11}$$

We could follow this same approach and treat the standard deviation as the random variable (*ie.* set

error estimator	symbol	formula
standard deviation	σ_x	$\sqrt{\frac{\sum(x_i-\bar{x})^2}{N-1}}$
relative standard deviation	ϵ_x	$\frac{\sigma_x}{\bar{x}}$
standard error	$\sigma_{\bar{x}}$	$\frac{\sigma_x}{\sqrt{N}}$
relative standard error	$\epsilon_{\bar{x}}$	$\frac{\sigma_x}{\bar{x}\sqrt{N}}$
error on the error	σ_{σ_x}	$\frac{\sigma_x}{\sqrt{N(N-1)}}$

Table E.1: Different types of errors and how they're related.

$f = \sigma_{\bar{x}}$) to estimate the *error on the error*. We begin by finding the derivative of $\sigma_{\bar{x}}$ with respect to x_j.

$$\begin{aligned}
\frac{\partial \sigma_{\bar{x}}}{\partial x_j} &= \frac{1}{\sqrt{N}} \frac{\partial}{\partial x_j} \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N-1}} \\
&= \frac{1}{\sqrt{N}} \frac{1}{2} \left[\frac{\sum_i (x_i - \bar{x})^2}{N-1} \right]^{-1/2} \frac{2}{N-1} \sum_i (x_i - \bar{x})(\delta_{ij} - \frac{1}{N}) \\
&= \frac{1}{\sqrt{N}} \frac{1}{(N-1)\sigma_x} \sum_i \left(x_i \delta_{ij} - \bar{x}\delta_{ij} - \frac{x_i}{N} + \frac{\bar{x}}{N} \right) \\
&= \frac{1}{\sqrt{N}} \frac{x_j - \bar{x}}{(N-1)\sigma_x}
\end{aligned} \tag{E.12}$$

Then by assuming the x's all have the same standard deviation, $\sigma_{x_i} = \sigma_x$, we arrive at the estimate of the error on the error

$$\begin{aligned}
\sigma^2_{\sigma_x} &\approx \frac{1}{N} \sum_j \sigma_x^2 \frac{(x_j - \bar{x})^2}{(N-1)^2 \sigma_x^2} \\
&\approx \frac{\sigma_x^2}{N(N-1)}, \\
\Rightarrow \sigma_{\sigma_x} &\approx \frac{\sigma_x}{\sqrt{N(N-1)}}.
\end{aligned} \tag{E.13}$$

Table E.1 summarizes the different types of basic errors and how they're related.

E.2 Non-parametric error estimators

Error estimates made via the error propagation formula, Eq. (E.8), are inherently dependent on our knowledge of the functional form of the data, f. If we know what f is, this isn't a problem. Though there are still issues with correctly accounting for covariance, dropping higher order terms, *etc.* If we're trying to assess the error associated with fit results, then f is our best guess at a fit function for the data, meaning we don't

really know what f is. Plus, there are still the issues of correctly accounting for covariance, *etc.* To deal with these issues we can use so-called *non-parametric* error estimators [103]. These methods (bootstrap and jackknife) don't depend on any knowledge of f and they account for covariance. They have their own limitations, but in cases where we don't know f (like fitting) the benefits outweigh the limitations. As an aside, when we do the fit to extract values for our fit parameters, the fitting program typically provides estimates for the errors associated with the fit – though we don't trust the errors returned because of correlations in the data and/or the fact that we had to choose an f, so we continue on to bootstrap.

E.2.1 Bootstrap

Here we start with N measurements of something (the data in our fit). Our original collection of measurements, called "boot-zero", will be denoted by

$$d^0 = \{d_i^0 \mid i = 1, \ldots, N\}. \tag{E.14}$$

We then randomly sample, with replacement, from these N measurements building a new collection of N measurements, called d^1 (bootstrap resample number one),

$$d^1 = \{d_i^1 \mid i = 1, \ldots, N\}. \tag{E.15}$$

The difference between this new collection of measurements and our original collection is that the new collection has some repeats and is missing some of the originals (because we sampled with replacement). We do this over and over, creating a large number, N_B, of d's. This collection of d's is called our ensemble of bootstrap resamples, $\{d^1, d^2, \ldots, d^{N_B}\}$.

Now we do our fit on each of the d's and collect the fit results. Let's say a fit returns a collection of fit parameters, $\vec{p}$, where I've called it a vector simply because we might be fitting more than one thing. If we collect all our fit results together $\{\vec{p}^{\,0};\ \vec{p}^{\,1}, \vec{p}^{\,2}, ..., \vec{p}^{\,N_B}\}$, where I've used a semi-colon to separate boot-zero from the ensemble of boostrap fit results, we can study the distribution of fit results. If we plot a histogram of a particular fit parameter over the N_B results, say $\{p_i^0;\ p_i^1, p_i^2, ..., p_i^{N_B}\}$, we'll get a spread of measurements that, hopefully, is centered about our boot-zero result, p_i^0. By picking the values of p_i above and below which 16% of the results lie, we have a reliable estimate of the $\pm 1\sigma$ errors. These errors could, in principle, be asymmetric, though if this were the case we'd have to think about why that happened and what it might mean before reporting an asymmetric error.

This is the strength of the bootstrap method. It gives a reliable estimate of the distribution of fit results

about the mean in a way that accounts for correlations and is independent of our specific choice of fit function.

Note the value of N_B needs to be large enough that the error reaches a steady-state value. Common choices are typically of order 1000.

In a multi-step analysis (*eg.* feeding our fit results into a subsequent χPT fit) we use $\{\vec{p}^{\,0};\ \vec{p}^{\,1}, \vec{p}^{\,2}, ..., \vec{p}^{\,N_B}\}$ to generate a parametric independent estimate of the error in the χPT fit. The $\{\vec{p}^{\,0};\ \vec{p}^{\,1}, \vec{p}^{\,2}, ..., \vec{p}^{\,N_B}\}$, or perhaps a subset of the fit parameters, are treated as data in the χPT fit. Then we do the χPT fit for each of data sets and generate a collection of χPT fit results, $\{\vec{\pi}^{\,0};\ \vec{\pi}^{\,1}, \vec{\pi}^{\,2}, ..., \vec{\pi}^{\,N_B}\}$. We can study the distribution of these fit results to get a robust error estimate.

Note that the distribution of χ^2 for the bootstrap ensemble of fit results is generally centered about a mean that is roughly twice $\chi^2_{\text{boot-zero}}$, [104].

E.2.2 Jackknife

Whereas bootstrap gives a reliable estimate of the distribution, and from that we extract the error, jackknife does not. It gives a reliable estimate of the mean and allows a reliable estimate of the error by scaling the variance of the distribution.

We once again start with N measurements of something (the data in our fit). Here I'll call this collection of measurements "jack-zero", though it is identical to what I called boot-zero in the last section,

$$d^0 = \{d_i \mid i = 1, \ldots, N\}. \tag{E.16}$$

We then generate a new collection of $N - m$ measurements by omitting the first m elements of d^0,

$$d^1 = \{d_i \mid i = m + 1, \ldots, N\}, \tag{E.17}$$

and a second collection of $N - m$ measurements by omitting the second m elements,

$$d^2 = \{d_i \mid i = 1, m + 2, \ldots, N\}, \tag{E.18}$$

We repeat this a total of N/m times (assuming this divides without remainder) and arrive at a collection of N/m jackknife resamples, $\{d^1, d^2, \ldots, d^{N/m}\}$.

Note there is no randomness in the generation of the jackknife resample. If done correctly, it generates exactly the same answer each time. That its exactly reproducible is is a strength of the jackknife method,

especially if the analysis needs to be verified independently by multiple people.

Again, we do our fit on each of the d's and collect the fit results, $\{\vec{p}^{\,0};\ \vec{p}^{\,1}, \vec{p}^{\,2}, \ldots, \vec{p}^{\,N/m}\}$. We could histogram these fit results, but the distribution is too narrow, by roughly a factor of $\sqrt{N/m}$ (this is too rough an estimate to use as our final result, just a ballpark estimate of how much too narrow the jackknife distribution is). The jackknife method is unable to reliably determine the distribution of fit results about the mean, a weakness of the method. The mean of the distribution is, however, reliable.

Despite not being able to reproduce the distribution, we can get a reliable estimate of the error on one of the fit parameters by doing a quick calculation on the collection of fit results, $\{p_i^0;\ p_i^1, p_i^2, \ldots, p_i^{N/m}\}$

$$\sigma_{p_i} = \sqrt{\frac{N/m-1}{N/m}\sum_{n=1}^{N/m}\left(\langle p_i^n\rangle - \langle p_i^0\rangle\right)^2}. \tag{E.19}$$

This is where a factor of $N/m-1$ is inserted relative to a naive calculation of the standard error. The covariance matrix passed to the fitter for a fit of a single jackknife sample, d^n, is calculated in the "standard" way, Eq. (E.7).

Note the value of m is chosen based, largely, on autocorrelations in the data. The more autocorrelated the data, the larger m should be. If autocorrelations are negligible $m = 1$ is adequate.

For the subsequent χPT fit, we again use our collection of fit results, $\{\vec{p}^{\,0};\ \vec{p}^{\,1}, \vec{p}^{\,2}, \ldots, \vec{p}^{\,N/m}\}$, to generate a parametric independent estimate of the error in the χPT fit. As with the bootstrap we use, perhaps a subset of, the fit resutls $\{\vec{p}^{\,0};\ \vec{p}^{\,1}, \vec{p}^{\,2}, \ldots, \vec{p}^{\,N/m}\}$ as data for the χPT fit. Then we collect our jackknife ensemble of χPT fit results, $\{\vec{\pi}^{\,0};\ \vec{\pi}^{\,1}, \vec{\pi}^{\,2}, \ldots, \vec{\pi}^{\,N/n}\}$. From this collection we can get a reliable estimate of the mean, but not the distribution. The robust error estimate for χPT fit parameter π_i is given by

$$\sigma_{\pi_i} = \sqrt{\frac{N/m-1}{N/m}\sum_{n=1}^{N/m}\left(\langle \pi_i^n\rangle - \langle \pi_i^0\rangle\right)^2}. \tag{E.20}$$

E.3 Multivariate Correlations

We now consider a two-point correlation function on an ensemble of gauge configurations. The correlation function is, time slice by time slice, the average value of N configurations of individual correlation functions. So we have T different values of the correlation function and each of the T values is itself an average of N things. This is the type of generalization we made near the beginning of Sec. E.1. We have $\{x_i, y_i, ...\}$ where the $\{x_i\}$ are N values of the correlation function at time slice t_1, the $\{y_i\}$ are N values of the correlation function at time slice t_2, and so on. Now we want to talk about the average values $\{\bar{x}, \bar{y}, ...\}$, the variances

$\{\sigma_x^2, \sigma_y^2, ...\}$, the correlations between the x's and the y's and the correlations between the different $\{x_i\}$.

E.3.1 Correlations Among Time Slices

The covariance we found in Eq. (E.7) measured for us the correlation between the $\{x_i\}$ and the $\{y_i\}$, each of which was a collection of different variables. This is exactly what we need in order to talk about the covariance between our correlation function at different time slices. To be concrete, in lieu of $\{x_i\}$ and $\{y_i\}$, let's talk about a two point correlation function, $C_i(t)$. The index i runs over all the configurations and t runs over each time slice. The collection of all the configuration's two-point correlation functions could then be represented by the collection of N T-dimensional row vectors

$$\begin{aligned} C_1(t) &= (C_1(t_1), C_1(t_2), ..., C_1(t_T)) \\ C_2(t) &= (C_2(t_1), C_2(t_2), ..., C_2(t_T)) \\ &\vdots \\ C_N(t) &= (C_N(t_1), C_N(t_2), ..., C_N(t_T))\,. \end{aligned} \tag{E.21}$$

To get to the covariance matrix, we start by subtracting from each $C_i(t_j)$ the mean value of the correlation function at that time slice, $\overline{C(t_j)} = \sum_{i=1}^{N} C_i(t_j)/N$, giving the following vectors for each configuration

$$\begin{aligned} C_1(t) - \overline{C(t)} &= \left(C_1(t_1) - \overline{C(t_1)}, C_1(t_2) - \overline{C(t_2)}, ..., C_1(t_T) - \overline{C(t_T)}\right) \\ C_2(t) - \overline{C(t)} &= \left(C_2(t_1) - \overline{C(t_1)}, C_2(t_2) - \overline{C(t_2)}, ..., C_2(t_T) - \overline{C(t_T)}\right) \\ &\vdots \\ C_N(t) - \overline{C(t)} &= \left(C_N(t_1) - \overline{C(t_1)}, C_N(t_2) - \overline{C(t_2)}, ..., C_N(t_T) - \overline{C(t_T)}\right), \end{aligned} \tag{E.22}$$

then we take the outer product of each configuration's vector with itself. For the i^{th} configuration this looks like

$$(C_i(t) - \overline{C(t)})^T (C_i(t) - \overline{C(t)}) =$$

$$\begin{pmatrix} \left(C_i(t_1) - \overline{C(t_1)}\right)^2 & \cdots & \left(C_i(t_1) - \overline{C(t_1)}\right)\left(C_i(t_T) - \overline{C(t_T)}\right) \\ \vdots & \ddots & \vdots \\ \left(C_i(t_T) - \overline{C(t_T)}\right)\left(C_i(t_1) - \overline{C(t_1)}\right) & \cdots & \left(C_i(t_T) - \overline{C(t_T)}\right)^2 \end{pmatrix} \tag{E.23}$$

Finally, we sum over all the configurations and normalize by $N-1$, giving us the covariance matrix, σ^2

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(C_i(t)-\overline{C(t)})^T(C_i(t)-\overline{C(t)}) = \tag{E.24}$$

$$\frac{1}{N-1}\sum_{i=1}^{N}\begin{pmatrix} \left(C_i(t_1)-\overline{C(t_1)}\right)^2 & \cdots & \left(C_i(t_1)-\overline{C(t_1)}\right)\left(C_i(t_T)-\overline{C(t_T)}\right) \\ \vdots & \ddots & \vdots \\ \left(C_i(t_T)-\overline{C(t_T)}\right)\left(C_i(t_1)-\overline{C(t_1)}\right) & \cdots & \left(C_i(t_T)-\overline{C(t_T)}\right)^2 \end{pmatrix}$$

Using Eq.'s (E.1) and (E.7) we identify the elements of this matrix as either variances or covariances. The diagonals are variances, $\sigma^2_{tt} = (N-1)^{-1}\sum_i(C_i(t)-\overline{C(t)})^2$, and the off-diagonals are covariances, $\sigma^2_{tt'} = (N-1)^{-1}\sum_i(C_i(t)-\overline{C(t)})(C_i(t')-\overline{C(t')})$. The larger the correlation between the data at times t and t' the larger the value of $\sigma^2_{tt'}$. In terms of the variances and covariances the covariance matrix is then

$$\sigma^2 = \begin{pmatrix} \sigma^2_{t_1t_1} & \sigma^2_{t_1t_2} & \cdots & \sigma^2_{t_1t_T} \\ \sigma^2_{t_2t_1} & \sigma^2_{t_2t_2} & \cdots & \sigma^2_{t_2t_T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_{t_Tt_1} & \sigma^2_{t_Tt_2} & \cdots & \sigma^2_{t_Tt_T} \end{pmatrix}. \tag{E.25}$$

A few observations about the covariance matrix are in order. First, if our correlation functions are real so is the covariance matrix. Actually, for a given configuration the correlation functions have a non-zero imaginary component but when averaged over all configurations the imaginary parts average to zero. The same thing happens for the elements of the covariance matrix. Second, its symmetric so there are only $T(T+1)/2$ unique elements and not T^2. Lastly, its elements have units that are the units of $C(t)$ squared. For this last reason, its often more convenient to work with the *correlation matrix*, ρ. It's obtained from the covariance matrix by rescaling each entry by the appropriate standard deviation(s),

$$\rho = \begin{pmatrix} 1 & \frac{\sigma^2_{t_1t_2}}{\sigma_{t_1}\sigma_{t_2}} & \cdots & \frac{\sigma^2_{t_1t_T}}{\sigma_{t_1}\sigma_{t_T}} \\ \frac{\sigma^2_{t_2t_1}}{\sigma_{t_2}\sigma_{t_1}} & 1 & \cdots & \frac{\sigma^2_{t_2t_T}}{\sigma_{t_2}\sigma_{t_T}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma^2_{t_Tt_1}}{\sigma_{t_T}\sigma_{t_1}} & \frac{\sigma^2_{t_Tt_2}}{\sigma_{t_T}\sigma_{t_2}} & \cdots & 1 \end{pmatrix}. \tag{E.26}$$

E.3.2 Correlations Among Configurations

Gauge configurations are generated via a Markov-chain of potential configurations by using one configuration to generate the next, generating a set of inherently correlated gauge configurations. To try to reduce these correlations we keep, say, every sixth one. To analyze residual correlations among configurations we might follow the procedure of the previous section and try to build a $N \times N$ correlation matrix from the T, N-dimensional data vectors

$$\begin{aligned} C_i(t_1) &= (C_1(t_1), C_2(t_1), ..., C_N(t_1)) \\ C_i(t_2) &= (C_1(t_2), C_2(t_2), ..., C_N(t_2)) \\ &\vdots \\ C_i(t_T) &= (C_1(t_T), C_2(t_T), ..., C_N(t_T)) \,. \end{aligned} \tag{E.27}$$

The problem with this approach is that our correlation matrix would have $N(N+1)/2$ unique elements built from T underlying time slices of data. With $\mathcal{O}(1000)$ configurations and $\mathcal{O}(100)$ time slices, we don't have enough time slices to build a correlation matrix to study correlations among the configurations. There are, however, other ways we can study correlations among configurations.

Autocorrelation function

The first method calculates the autocorrelation function, $\mathcal{A}(\eta)$, a measure of the degree to which correlations exist as a function of separation in *configuration time*. The configurations are naturally ordered according to their generation. By preserving that ordering we can speak meaningfully of a configuration time, *eg.* an index specifying the order in which they were created. Let η be the number of configurations that separate the two configurations whose correlations we wish to analyze. The autocorrelation function is [47]

$$\mathcal{A}(\eta) = \frac{1}{N} \sum_i \frac{(C_{i+\eta}(t) - \overline{C(t)})(C_i(t) - \overline{C(t)})}{\sigma_t^2}, \tag{E.28}$$

where a value of zero indicates no correlation, a value of (minus) one indicates the data are 100% (anti)correlated. By evaluating $\mathcal{A}(\eta)$ at different values of η we can estimate the degree to which adjacent configurations are correlated and estimate the necessary number to skip in order to generate nearly independent configurations.

Principle component analysis

A potentially more sensitive measure of autocorrelations is obtained from the principal component of the covariance matrix, Eq. (E.25), built from the data for a two point correlation function. The eigenvector corresponding to the largest eigenvalue of the covariance matrix is called the first principal component, $\vec{pc}_1$. We project the scaled vector of correlation function data for configuration i,

$$\vec{z}_i = \left(\frac{C_i(t_1) - \overline{C(t_1)}}{\sigma_{t_1 t_1}}, \ldots, \frac{C_i(t_T) - \overline{C(t_T)}}{\sigma_{t_T t_T}} \right) \tag{E.29}$$

onto $\vec{pc}_1(t)$ to generate the projection,

$$proj_i = \vec{z}_i \cdot \vec{pc}_1. \tag{E.30}$$

Then, we study the autocorrelation function of this projection using $\mathcal{A}(\eta)$.

The output of this analysis, whether based directly on correlation function data or the principal component of the covariance matrix, is a measure of autocorrelation as a function of separation in configuration time. From this we determine the minimum number of configurations that need to be skipped[2] to reduce autocorrelations to an acceptable level. One way of doing this is by assuming an exponential decay [47],

$$|\mathcal{A}(\eta)| = e^{-\eta/\tau}, \tag{E.31}$$

and determining the autocorrelation time, τ by comparing Eq. (E.31) to calculated values of $\mathcal{A}(\eta)$.

Binning

The autocorrelation coefficient and principal component analysis provide ways of studying autocorrelations and determining the needed separation in configuration time in order to reduce the effects of autocorrelations. *Binning*, on the other hand, is the method by which one actually removes the the autocorrelations.

Let $\{x_i\}$ be a collection of data measured on configuration i. Correlations among the $\{x_i\}$ are reduced by forming a new, smaller, collection of data $\{y_j \mid j = 1, \ldots, N/b\}$ where b is the size of the bins and the $\{y_j\}$ are the averages

$$y_j = \frac{x_{b(j-1)+1} + \cdots + x_{jb}}{b}. \tag{E.32}$$

We can then relate the mean, variance, standard deviation and standard error of the $\{y_j\}$ to those of the $\{x_i\}$ by plugging Eq. (E.32) into the appropriate expressions from App. E.1. For instance, the mean of the

[2] In practice we don't throw out configurations, but average multiple adjacent configurations together in a procedure called binning and described in App. E.3.2

collection of $\{y_j\}$ is

$$\begin{aligned}
\bar{y} &= \frac{1}{N/b}\sum_{j=1}^{N/b} y_j \\
&= \frac{1}{N/b}\sum_{j=1}^{N/b} \frac{x_{b(j-1)+1}+\cdots+x_{jb}}{b} \\
&= \frac{(x_1+\cdots+x_b)+(x_{b+1}+\cdots+x_{2b})+\cdots+(x_{N-b+1}+\cdots+x_N)}{N} \\
&= \bar{x}.
\end{aligned} \tag{E.33}$$

Binning doesn't change the answer we get. It should, however, change the error we report if it does its job of removing correlations among the data. Let's calculate the variance of the binned sample σ_y^2 and see how it differs from the unbinned case σ_x^2

$$\begin{aligned}
\sigma_y^2 &= \frac{1}{N/b-1}\sum_{j=1}^{N/b}(y_j-\bar{y})^2 \\
&= \frac{1}{N/b-1}\sum_{j=1}^{N/b}\left(\frac{x_{b(j-1)+1}+\cdots+x_{jb}}{b}-\bar{x}\right)^2 \\
&= \frac{1}{N/b-1}\left[\left(\frac{x_1+\cdots+x_b}{b}-\bar{x}\right)^2+\left(\frac{x_{b+1}+\cdots+x_{2b}}{b}-\bar{x}\right)^2+\cdots+\left(\frac{x_{N-b+1}+\cdots+x_N}{b}-\bar{x}\right)^2\right] \\
&= \frac{1}{b(N-b)}\left[(x_1+\cdots+x_b-b\bar{x})^2+(x_{b+1}+\cdots+x_{2b}-b\bar{x})^2+\cdots+(x_{N-b+1}+\cdots+x_N-b\bar{x})^2\right] \\
&= \frac{1}{b(N-b)}\left[[(x_1-\bar{x})+\cdots+(x_b-\bar{x})]^2+\cdots+[(x_{N-b+1}-\bar{x})+\cdots+(x_N-\bar{x})]^2\right] \\
&= \frac{1}{b(N-b)}\left[(x_1-\bar{x})^2+\cdots+(x_N-\bar{x})^2+2\sum_{bins}(x_i-\bar{x})(x_j-\bar{x})\Big|_{i,j\in bin}^{i\neq j}\right] \\
&= \sigma_x^2\frac{N-1}{b(N-b)}\left[1+\frac{2}{N-1}\sum_{bins}\frac{(x_i-\bar{x})(x_j-\bar{x})}{\sigma_x^2}\Big|_{i,j\in bin}^{i\neq j}\right].
\end{aligned} \tag{E.34}$$

The overall factor out front accounts for the change in number of samples used to determine the variance. It's there whether there are correlations or not and is an artifact of the change in the sample size due to binning. The term not equal to 1 in the square brackets measures the autocorrelation of the data within each bin. By using the definition of the autocorrelation $\mathcal{A}(\eta)$ of Eq. (E.28) with η and the sum restricted so as to keep the data all in the same bin,

$$\mathcal{A}_b(\eta) = \frac{1}{b-\eta}\sum_{i=1}^{b-\eta}\frac{(x_{i+\eta}-\bar{x})(x_i-\bar{x})}{\sigma_x^2} \tag{E.35}$$

we can express this term in the square brackets in terms of the $\mathcal{A}_b(\eta)$ for each bin

$$\sigma_{\bar{y}}^2 = \sigma_x^2 \frac{N-1}{b(N-b)} \left[1 + \frac{2}{N-1} \sum_{bins} \sum_{\eta=1}^{b-\eta} (b-\eta) \mathcal{A}_b(\eta) \right], \tag{E.36}$$

allowing us to relate the unbinned sample variance to a binned sample variance.

If the data are completely uncorrelated the $\mathcal{A}_b(\eta)$ are all zero and we have

$$\sigma_y = \sigma_x \sqrt{\frac{N-1}{b(N-b)}} \tag{E.37}$$

which goes as $1/\sqrt{b}$ for large N.

If the data are correlated, we make the approximation that the correlations within each bin are the same and the sum over bins gives a factor of N/b

$$\sigma_{\bar{y}}^2 \approx \sigma_x^2 \left(\frac{N-1}{b(N-b)} \right) \left[1 + \frac{2N}{b(N-1)} \sum_{\eta=1}^{b-\eta} (b-\eta) \mathcal{A}_b(\eta) \right]. \tag{E.38}$$

If the data are strongly correlated then each of the $\mathcal{A}_b(\eta)$ are almost one and we have

$$\begin{aligned} \sigma_y &\approx \sigma_x \sqrt{\frac{N-1}{b(N-b)}} \left[1 + \frac{2N}{b(N-1)} \sum_{\eta=1}^{b-\eta} (b-\eta) \right]^{\frac{1}{2}} \\ &\approx \sigma_x \sqrt{\frac{N-1}{b(N-b)}} \left[1 + \frac{2N}{b(N-1)} \frac{b(b-1)}{2} \right]^{\frac{1}{2}} \\ &\approx \sigma_x \sqrt{\frac{Nb-1}{b(N-b)}} \end{aligned} \tag{E.39}$$

which approaches one for large N.

When considering the standard error, Eq. (E.10), the leading effects of bin size are accounted for (for the binned variance you divide by N/b while for the unbinned variance you divide by N). Any remaining difference between $\sigma_{\bar{y}}$ and $\sigma_{\bar{x}}$ is due almost entirely to correlations,

$$\sigma_{\bar{y}} = \sigma_{\bar{x}} \times \begin{cases} \sqrt{\frac{N-1}{N-b}} \approx 1, & 0\ \% \text{ correlation;} \\ \sqrt{\frac{Nb-1}{N-b}} \approx \sqrt{b}, & 100\% \text{ correlation,} \end{cases} \tag{E.40}$$

providing a third way of looking for autocorrelations. If the data are correlated and you bin, you get a larger relative error. If the data are uncorrelated and you bin, you get no change in relative error. The standard

procedure is to bin the data by increasing bin size until the relative stops increasing. At that point the bin size is large enough that the reduced number, N/b, of configurations, $\{y_i\}$, are uncorrelated.

E.3.3 Ability to determine the correlation matrix

Calculating the correlation matrix is an integral part of the fitting procedure of Sec. 7.2 if one wishes to account for the fact that data at different time slices are, in fact, correlated. We calculate the correlation matrix directly from the data, so our ability to adequately determine its elements is based on the amount of data available. Ignoring autocorrelations (*ie.* assuming the N configurations are independent) one would naively expect the NT data elements provided by N configurations, each with T timeslices, must exceed the $T(T-1)/2$ unique entries in the correlation matrix. This gives the constraint

$$N > \frac{T-1}{2} \tag{E.41}$$

However, this not only ignores the fact there aren't NT independent data elements, due to autocorrelation, but also the error on the calculated entries of the correlation matrix due to finite N.

A determination of the autocorrelation time, τ, could be used to estimate the effective number of independent configurations by binning. Following Montvay and Munster's advice [47] we use binsize $2\tau + 1$

$$N_{\text{eff}} = \frac{N}{2\tau + 1} \tag{E.42}$$

By replacing NT data elements with $N_{\text{eff}}T$ we have

$$N > (2\tau + 1)\frac{T-1}{2}, \tag{E.43}$$

a fairly weak constraint on N. Given the autocorrelation time $\tau = 0.15$, and a reasonably large number of time slices, say $T = 35$, we have $N > 22$. A simultaneous fit of four two point correlation functions, each with $T = 35$, requires only $N > 90$. In our analysis we perform a simultaneous fit to a two and three point correlation function. By symmetry, the three point correlation function provides $T_3(T_3+1)/2$ independent data points when fitting T_3 time slices. When combined with T_2 two point correlation function timeslices this gives $T = T_3(T_3+1)/2+T_2$ total data elements. Using $T_2 = 35$ and $T_3 = 15$, typical values, this requires $N > 142$.

To consider the error associated with finite N, we demand the relative error on an arbitrary element,

$c_{tt'}$, of the correlation matrix be small,

$$\frac{\sigma_{c_{tt'}}}{c_{tt'}} < \delta \tag{E.44}$$

where δ defines what small means. To determine what this says about N we start with $c_{tt'}$

$$c_{tt'} = \frac{\sigma^2_{tt'}}{\sigma_{tt}\sigma_{t't'}} \tag{E.45}$$

where $\sigma_{tt'}$ is given by

$$\sigma^2_{tt'} = \frac{1}{N-1}\sum_{i=1}^{N}(x_{it} - \bar{x}_t)(x_{it'} - \bar{x}_{t'}). \tag{E.46}$$

We estimate $\sigma_{c_{tt'}}$ via the error propagation formula[3]

$$\sigma^2_{c_{tt'}} \approx \sum_{i=1}^{N}\sigma^2_{i,tt}\left(\frac{\partial c_{tt'}}{\partial x_{it}}\right)^2 + \sum_{i=1}^{N}\sigma^2_{i,t't'}\left(\frac{\partial c_{tt'}}{\partial x_{it'}}\right)^2 + 2\sum_{i=1}^{N}\sigma^2_{i,tt'}\left(\frac{\partial c_{tt'}}{\partial x_{it}}\right)\left(\frac{\partial c_{tt'}}{\partial x_{it'}}\right) \tag{E.47}$$

where $\sigma^2_{i,tt'}$ is the covariance $\sigma^2_{tt'}$ for the i^{th} configuration. First we calculate $\partial c_{tt'}/\partial x_{jt}$, then, noting that $c_{tt'}$ is symmetric under $t \leftrightarrow t'$, obtain $\partial c_{tt'}/\partial x_{jt'}$ by interchanging $t \leftrightarrow t'$,

$$\begin{aligned}\frac{\partial c_{tt'}}{\partial x_{jt}} &= \frac{\partial}{\partial x_{jt}}\left[\frac{\sum_i (x_{it} - \bar{x}_t)(x_{it'} - \bar{x}_{t'})}{\sqrt{\sum_i (x_{it} - \bar{x}_t)^2}\sqrt{\sum_i (x_{it'} - \bar{x}_{t'})^2}}\right] \\ &= \frac{1}{(N-1)\sigma_{tt}\sigma_{t't'}}\left[(x_{jt'} - \bar{x}_{t'}) - \frac{\sigma^2_{tt'}}{\sigma^2_{tt}}(x_{jt} - \bar{x}_t)\right].\end{aligned} \tag{E.48}$$

By assuming the covariance $\sigma^2_{i,tt'}$ is the same for all configurations, $\sigma^2_{i,tt'} = \sigma^2_{tt'}$, this gives

$$\begin{aligned}\sigma^2_{c_{tt'}} \approx\ & \frac{1}{(N-1)^2\sigma^2_{tt}\sigma^2_{t't'}}\sum_{j=1}^{N}\left\{\sigma^2_{tt}\left[(x_{jt'} - \bar{x}_{t'})^2 - 2\frac{\sigma^2_{tt'}}{\sigma^2_{tt}}(x_{jt'} - \bar{x}_{t'})(x_{jt} - \bar{x}_t) + \frac{\sigma^4_{tt'}}{\sigma^4_{tt}}(x_{jt} - \bar{x}_t)^2\right]\right. \\ &+ \sigma^2_{t't'}\left[(x_{jt} - \bar{x}_t)^2 - 2\frac{\sigma^2_{tt'}}{\sigma^2_{t't'}}(x_{jt} - \bar{x}_t)(x_{jt'} - \bar{x}'_t) + \frac{\sigma^4_{tt'}}{\sigma^4_{t't}}(x_{jt'} - \bar{x}'_t)^2\right] \\ &\left. + 2\sigma^2_{tt'}\left[(x_{jt'} - \bar{x}_{t'})(x_{jt} - \bar{x}_t) - \frac{\sigma^2_{tt'}}{\sigma^2_{t't'}}(x_{jt'} - \bar{x}_{t'})^2 - \frac{\sigma^2_{tt'}}{\sigma^2_{tt}}(x_{jt} - \bar{x}_t)^2 + \frac{\sigma^4_{tt'}}{\sigma^2_{tt}\sigma^2_{t't'}}(x_{jt} - \bar{x}_t)(x_{jt'} - \bar{x}_{t'})\right]\right\}.\end{aligned} \tag{E.49}$$

[3] This calculation follows Bevington and Robinson's calculation of the error on the mean, $\sigma_\mu \approx \sigma/\sqrt{N}$, pp. 53-54 in [105].

Using $\sum(x_{it} - \bar{x}_t)(x_{it'} - \bar{x}_{t'}) = (N-1)\sigma^2_{tt'}$ this becomes

$$\begin{aligned}
\sigma^2_{c_{tt'}} &\approx \frac{1}{(N-1)\sigma^2_{tt}\sigma^2_{t't'}}\left\{\sigma^2_{tt}\left[\sigma^2_{t't'} - 2\frac{\sigma^2_{tt'}}{\sigma^2_{tt}}\sigma^2_{tt'} + \frac{\sigma^4_{tt'}}{\sigma^4_{tt}}\sigma^2_{tt}\right] + \sigma^2_{t't'}\left[\sigma^2_{tt} - 2\frac{\sigma^2_{tt'}}{\sigma^2_{t't'}}\sigma^2_{tt'} + \frac{\sigma^4_{tt'}}{\sigma^4_{t't'}}\sigma^2_{t't'}\right]\right. \\
&\quad \left. + 2\sigma^2_{tt'}\left[\sigma^2_{tt'} - \frac{\sigma^2_{tt'}}{\sigma^2_{t't'}}\sigma^2_{t't'} - \frac{\sigma^2_{tt'}}{\sigma^2_{tt}}\sigma^2_{tt} + \frac{\sigma^4_{tt'}}{\sigma^2_{tt}\sigma^2_{t't'}}\sigma^2_{tt'}\right]\right\} \\
&\approx 2\left(1 - 2\frac{\sigma^4_{tt'}}{\sigma^2_{tt}\sigma^2_{t't'}} + \frac{\sigma^8_{tt'}}{\sigma^4_{tt}\sigma^4_{t't'}}\right) \\
&\approx \frac{2}{N-1}\left(1 - c^2_{tt'}\right)^2
\end{aligned} \tag{E.50}$$

where Eq. (E.45) was used in the last line. From this we arrive at a constraint on the required number of configurations in terms of an arbitrary element of the correlation matrix, $c_{tt'}$, and the precision with which we want it determined, δ,

$$N > 1 + \frac{2}{\delta^2}\left(\frac{1 - c^2_{tt'}}{c_{tt'}}\right)^2. \tag{E.51}$$

This constraint is more severe than Eq. (E.43). If we want to be able to resolve a minimum correlation of 10% with 30% precision (*ie.* $c_{tt'} = 0.1$ and $\delta = 0.3$) we need $N > 2179$. Given the number of configurations used in our analysis (see Tab. 5.1) and the fact that we use four meson creation operators per configuration (*ie.* we effectively have $\sim 4N$ configurations) we satisfy this criteria for reasonable choices of δ and $c_{tt'}$.

E.4 Comparing two measurements: the Z-statistic

Given two measurements, x_1 and x_2, with errors σ_1 and σ_2, respectively, and covariance σ^2_{12}, the Z-statistic allows a quantitative evaluation of the level of disagreement between x_1 and x_2, incorporating the effects of σ_1, σ_2, and σ_{12}. The Z-statistic is given by

$$Z = \frac{x_1 - x_2}{\sqrt{\sigma^2_1 + \sigma^2_2 - 2\sigma^2_{12}}}. \tag{E.52}$$

Provided x_1 and x_2 differ only because of statistical fluctuations, Z is normally distributed with $\sigma_Z = 1$. Because Z is normally distributed it is easy to interpret the level of disagreement between x_1 and x_2. A value of $Z = 1$ means x_1 is one sigma larger than x_2.

References

[1] S. Weinberg. Implications of dynamical symmetry breaking: An addendum. *Phys. Rev.*, **D19**(4):1277–1280, 1979. [doi:10.1103/PhysRevD.19.1277].

[2] Leonard Susskind. Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory. *Phys. Rev.*, **D20**(10):2619–2625, 1979. [doi:10.1103/PhysRevD.20.2619].

[3] The CDF Collaboration, The D0 Collaboration, and The Tevatron New Physics Higgs Working Group. Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 6.7 fb-1 of Data. *arXiv:1007.4587v1 [hep-ex]*, Jul. 2010.

[4] The CDF Collaboration, The D0 Collaboration, and The Tevatron New Physics Higgs Working Group. Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 8.2 fb-1 of Data. *arXiv:1103.3233v2[hep-ex]*, Mar. 2011.

[5] K. Nakamura *et al.* (Particle Data Group). *J. Phys.*, **G37**:075021, 2010. [http://pdg.lbl.gov/].

[6] Y. Grossman. Introduction to flavor physics. *arXiv:1006.3534*, Jun. 2010.

[7] S.L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev.*, **D2**:1285, Oct. 1970. [doi:10.1103/PhysRevD.2.1285].

[8] J. J. Aubert *et al.* Experimental Observation of a Heavy Particle J. *Phys. Rev. Lett.*, **33**:1404–1406, Nov. 1974. [doi:10.1103/PhysRevLett.33.1404].

[9] J.-E. Augustin *et al.* Discovery of a Narrow Resonance in e^+e^- Annihilation. *Phys. Rev. Lett.*, **33**:1406–1408, Nov. 1974. [doi:10.1103/PhysRevLett.33.1406].

[10] A. Buras and D. Guadagnoli. On the consistency between the observed amount of CP violation in the K- and B_d-systems within minimal flavor violation. *Phys. Rev.*, **D79**:053010, 2009. [arXiv:0901.2056[hep-ph]].

[11] A. Lenz, U. Nierste, J. Charles, S. Descotes-Genon, A. Jantsch, C. Kaufhold, H. Lacker, S. Monteil, V. Niess, and S. T'Jampens. Anatomy of New Physics in B-Bbar mixing. *Phys. Rev.*, **D83**:036004, Feb. 2011. [arXiv:1008.1593[hep-ph]].

[12] E. Lunghi and A. Soni. Footprints of the Beyond in flavor physics: Possible role of the Top Two Higgs Doublet Model. *JHEP*, **0709**:053, 2007. [arXiv:0707.0212[hep-ph]].

[13] E. Lunghi and A. Soni. Possible Indications of New Physics in B_d-mixing and in sin(2 beta) Determinations. *Phys. Lett.*, **B666**:162–165, 2008. [arXiv:0803.4340[hep-ph]].

[14] A. Buras and D. Guadagnoli. Correlations among new CP violating effects in Delta F = 2 observables. *Phys. Rev.*, **D78**:033005, 2008. [arXiv:0805.3887[hep-ph]].

[15] M. Bona, M. Ciuchini, E. Franco, V. Lubicz, G. Martinelli, F. Parodi, M. Pierini, C. Schiavi, L. Silvestrini, V. Sordini, A. Stocchi, C. Tarantino, and V. Vagnoni. An Improved Standard Model Prediction Of BR(B $\to$ tau nu) And Its Implications For New Physics. 2009. [arXiv:hep-lat/0403012].

[16] E. Lunghi and A. Soni. Hints for the scale of new CP-violating physics from B-CP anomalies. *JHEP*, **08**:051, 2009. [arXiv:0903.5059[hep-ph]].

[17] J. Laiho, E. Lunghi, and R. S. Van de Water. Lessons for new physics from CKM studies. 2011. [arXiv:1102.3917[hep-ph]].

[18] UT*fit* Collaboration: M. Bona, M. Ciuchini, E. Franco, V. Lubicz, G. Martinelli, F. Parodi, M. Pierini, C. Schiavi, L. Silvestrini, V. Sordini, A. Stocchi, and V. Vagnoni. First Evidence of New Physics in $b \leftrightarrow s$ Transitions. *PMC Phys.*, **A3**:6, 2009. [arXiv:0803.0659[hep-ph]].

[19] D0 Collaboration: V. Abazov *et al.* Measurement of the anomalous like-sign dimuon charge asymmetry with 9 fb^{-1} of p pbar collisions. 2011. [arXiv:1106.6308[hep-ex]].

[20] A. Buras, M.V. Carlucci, S. Gori, and G. Isidori. Higgs–mediated FCNCs: Natural Flavour Conservation vs. Minimal Flavour Violation. *JHEP*, **10**:009, 2010. [arXiv:1005.5310 [hep-ph]].

[21] CDF Collaboration: T. Aaltonen *et al.* Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ Decays with CDF II. 2011. [arXiv:1107.2304 [hep-ex]].

[22] Eugene Golowich, JoAnne Hewett, and Sandip Pakvasa. Relating B_s MIxing and $B_s \to \mu^+\mu^-$ with New Physics. *Phys. Rev.*, **D83**:114017, Jul. 2011. [doi:10.1103/PhysRevD.83.114017].

[23] CDF Collaboration. Observation of Bs - Bsbar Oscillations. *Phys. Rev. Lett.*, **97**:242003, 2006. [arXiv:hep-ex/0609040].

[24] V.V. Weisskopf and E.P. Wigner. Calculation of the natural line width due to the Dirac theory of light. *Z. Phys.*, **63**:54, 1930. [doi:10.1007/BF01336768].

[25] H. Galić. Factorization and its Applicability in Weak Nonleptonic Processes. *Phys. Rev.*, **D24**:2441–2456, 1981. [doi:10.1103/PhysRevD.24.2441].

[26] G. Buchalla, A. Buras, and M. Lautenbacher. Weak Decays Beyond Leading Logarithms. *Rev. Mod. Phys.*, **68**:1125, 1996. [arXiv:hep-ph/9512380].

[27] T. Inami and C.S. Lim. Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes $K_L \to \mu\bar{\mu}$, $K^+ \to \pi^+\nu\bar{\nu}$ and $K^0 \leftrightarrow \bar{K}^0$. *Prog. Theor. Phys.*, **65**(1):297–314, 1981. [doi:10.1143/PTP.65.297].

[28] Y. Takahashi. The Fierz Identities. In H. Ezawa and K. Kamefuchi, editors, *Progress in Quantum Field Theory*, page 121. North-Holland, Amsterdam, 1986.

[29] Stefan Herrlich and Ulrich Nierste. Evanescent Operators, Scheme Dependences and Double Insertions. *Nucl. Phys.*, **B445**:39–58, 1995. [arXiv:hep-ph/9412375].

[30] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Addison Wesley, 1995.

[31] C.C. Nishi. Simple derivation of general Fierz-type identities. *Am. J. Phys.*, **73**(12):1160–1163, Dec. 2005. [doi:10.1119/1.2074087].

[32] Rolf Mertig. Mathematica 3.0 in High Energy Physics: FeynCalc 3.0. *Mathematica in Education and Research*, **5**:56–58, 1996.

[33] D. Becirevic, V. Gimenez, G. Martinelli, M. Papinutto, and J. Reyes. B-parameters of the complete set of matrix elements of $\Delta B = 2$ operators from the lattice. *JHEP*, **0204**:025, 2002. [arXiv:hep-lat/0110091].

[34] A. Bazavov, D. Toussaint, C. Bernard, J. Laiho, C. DeTar, L. Levkova, M.B. Oktay, Steven Gottlieb, U.M. Heller, J.E. Hetrick, P.B. Mackenzie, R. Sugar, and R.S. Van de Water. Nonperturbative QCD simulations with 2+1 flavors of improved staggered quarks. *Rev. Mod. Phys.*, **82**:1349–1417, 2010. [doi:10.1103/RevModPhys.82.1349].

[35] K. Symanzik. Cutoff Dependence in Lattice ϕ_4^4 Theory. In G. 't Hooft *et al.*, editor, *Recent Developments in Gauge Thoeries*, pages 313–330. Plenum, New York, 1980.

[36] K. Symanzik. Continuum Limit and Improved Action in Lattice Theories (I). Principles and ϕ_4^4 Theory. *Nucl. Phys.*, **B226**:187–204, Sep. 1983. [doi:10.1016/0550-3213(83)90468-6].

[37] K. Symanzik. Continuum Limit and Improved Action in Lattice Theories (II). O(n) Non-Linear Sigma Model in Perturbation Theory. *Nucl. Phys.*, **B226**:205–227, 1983. [doi:10.1016/0550-3213(83)90469-8].

[38] G. Peter Lepage and Paul B. Mackenzie. Viability of lattice perturbation theory. *Phys. Rev.*, **D48**:2250–2264, Sep. 1993. [doi:10.1103/PhysRevD.48.2250].

[39] E. Marinari, G. Parisi, and C. Rebbi. Monte Carlo Simulation of the Massive Schwinger Model. *Nucl. Phys.*, **B190**:734–750, Jun. 1981. [doi:10.1016/0550-3213(81)90048-1].

[40] John Kogut and Leonard Susskind. Hamiltonian formulation of Wilson's lattice gauge theories. *Phys. Rev.*, **D11**:395–408, 1975. [doi:10.1103/PhysRevD.11.395].

[41] T. Banks, Leonard Susskind, and John Kogut. Strong-coupling calculations of lattice gauge theories: (1+1)-dimensional exercises. *Phys. Rev.*, **D13**:1043–1053, 1976. [doi:10.1103/PhysRevD.13.1043].

[42] T. Banks, S. Raby, L. Susskind, J. Kogut, D.R.T. Jones, P.N. Scharbach, and D.K. Sinclair. Strong-coupling calculations of the hadron spectrum of quantum chromodynamics. *Phys. Rev.*, **D15**:1111–1127, 1977. [doi:10.1103/PhysRevD.15.1111].

[43] Leonard Susskind. Lattice fermions. *Phys. Rev.*, **D16**:3031–3039, 1977. [doi:10.1103/PhysRevD.16.3031].

[44] Aida X. El-Khadra, Andreas S. Kronfeld, and Paul B. Mackenzie. Massive Fermions in Lattice Gauge Theory. *Phys. Rev.*, **D55**:3933–3957, 1997. [doi:10.1103/PhysRevD.55.3933].

[45] K.G. Wilson. In A. Zichichi, editor, *New Phenomena In Subnuclear Physics. Part A. Proceedings of the First Half of the 1975 International School of Subnuclear Physics, Erice, Sicily, July 11 - August 1, 1975*, pages 13–32. Plenum Press, New York, 1975.

[46] M. Di Pierro, A.X. El-Khadra, Steven Gottlieb, A.S. Kronfeld, P.B. Mackenzie, D.P. Menscher, M.B. Oktay, and J.N. Simone. Charmonium with three flavors of dynamical quarks. *POS (Lattice 2002)*, **119**, Oct. 2002. [arXiv:hep-lat/0210051].

[47] István Montvay and Gernot Münster. *Cambridge Monographs on Mathematical Physics: Quantum Fields on a Lattice*. Cambridge University Press, 1994.

[48] Thomas DeGrand and Carlton DeTar. *Lattice Methods For Quantum Chromodynamics*. World Scientific, 2006.

[49] Forman S. Acton. *Numerical Methods that (Usually) Work*. The Mathematical Association of America, 4th edition, 1990.

[50] G. Peter Lepage, B. Clark, Christine T.H. Davies, K. Hornbostel, Paul B. Mackenzie, and C. Morningstar. Constrained Curve Fitting. *arXiv:hep-lat/0110175*, Oct. 2001.

[51] M. Di Pierro and E. Eichten. Excited Heavy-Light Systems and Hadronic Transitions. *Phy. Rev.*, **D64**:114004, Oct. 2001. [arXiv:hep-ph/0104208].

[52] D. Ebert, R.N. Faustov, and V.O. Galkin. Heavy-llight meson spectroscopy and Regge trajectories in the relativstic quark model. *Eur. Phys. J.*, **C66**:197–206, 2009. [arXiv:0910.5612 [hep-ph]].

[53] Elvira Gámiz, Junko Shigemitsu, and Howard Trottier. Four Fermion Operator Matching with NRQCD Heavy and AsqTad Light Quarks. *Phys. Rev.*, **D77**:114505, 2008. [doi:10.1103/PhysRevD.77.114505].

[54] E. Gámiz. Private communication.

[55] Q. Mason, H.D. Trottier, C.T.H. Davies, K. Foley, A. Gray, G.P. Lepage, M. Nobes, and J. Shigemitsu. Accurate Determinations of α_s from Realistic Lattice QCD. *Phys. Rev. Lett.*, **95**:052002, 2005. [doi:10.1103/PhysRevLett.95.052002].

[56] Y. Schröder. The static potential in QCD to two loops. *Phys. Lett.*, **B447**:321–326, 1999. [arXiv:hep-ph/9812205].

[57] T. van Ritbergen, J.A.M. Vermaseren, and S.A. Larin. The four loop β function in quantum chromodynamics. *Phys. Lett.*, **B400**:379–384, May 1997.T. van Ritbergen, 1997. [doi:10.1016/S0370-2693(97)00370-5].

[58] K. Hornbostel, G.P. Lepage, and C. Morningstar. Scale setting for alpha_s beyond leading order. *Phys. Rev.*, **D67**, 2003. [arXiv:hep-ph/0208224].

[59] S. Weinberg. Phenomenological lagrangians. *Physica*, **96A**:327–340, 1979. [doi:10.1016/0378-4371(79)90223-1].

[60] J. Gasser and H. Leutwyler. Chiral perturbation theory to one looop. *Ann. Phys.*, **158**:142–210, 1984. [doi:10.1016/0003-4916(84)90242-2].

[61] J. Gasser and H. Leutwyler. Chiral perturbation theory: Expansions in the mass of the strange quark. *Nucl. Phys.*, **B250**:465–516, 1985. [doi:10.1016/0550-3213(85)90492-4].

[62] H. Leutwyler. On the Foundations of Chiral Perturbation Theory. *Ann. Phys.*, **235**:165–203, 1994. [doi:10.1006/aphy.1994.1094].

[63] A. Morel. Chiral logarithms in quenched QCD. *J. Physique (Paris)*, **48**:1111, 1987.

[64] Aneesh V. Manohar and Mark B. Wise. *Heavy Quark Physics*. Cambridge University Press, 2000.

[65] William Detmold and C.-J. David Lin. Matrix elements of the complete set of $\Delta B = 2$ and $\Delta C = 2$ operators in heavy meson chiral perturbation theory. *Phys. Rev.*, **D76**:014501, 2007. [doi:10.1103/PhysRevD.76.01450].

[66] C. Aubin and C. Bernard. Staggered Chiral Perturbation Theory with Heavy-Light Mesons. *Nucl. Phys.*, **B**:491–493, Mar. 2005. [doi:10.1016/j.nuclphysbps.2004.11.167].

[67] C. Aubin and C. Bernard. Staggered Chiral Perturbation Theory for Heavy-Light Mesons. *Phys. Rev.*, **D73**:014515, 2006. [arXiv:hep-lat/0510088].

[68] Jack Laiho and R.S. Van de Water. Staggered chiral perturbation theory for B–$\overline{B}$ mixing. Technical report, FNAL Lattice Collaboration, May 2011.

[69] http://physics.wustl.edu/cb/Fermilab-MILC.

[70] J.A. Bailey, A. Bazavov, C. Bernard, C.M. Bouchard, C. DeTar, M. Di Pierro, R.T. Evans, E. Freeland, E. Gámiz, Steven Gottlieb, U.M. Heller, J.E. Hetrick, R. Jain, A.X. El-Khadra, A.S. Kronfeld, J. Laiho, L. Levkova, P.B. Mackenzie, E. Neil, M.B. Oktay, J.N. Simone, R. Sugar, D. Toussaint, and R.S. Van de Water. B- and D-meson decay constants from three-flavor lattice QCD. *In preparation.*

[71] Richard Todd Evans. *Determination of the B^0 and B_s^0 Mixing Matrix Elements in 2+1 Lattice QCD*. PhD thesis, University of Illinois at Urbana-Champaign, 2008.

[72] C. Bernard. Notes on including B-B^* and flavor splittings for heavy-light decay constants in SχPT. Technical report, FNAL Lattice and MILC Collaboration, Feb. 2010.

[73] Jon A. Bailey, C. Bernard, C. DeTar, M. Di Pierro, A.X. El-Khadra, R.T. Evans, E.D. Freeland, E. Gamiz, Steven Gottlieb, U.M. Heller, J.E. Hetrick, A.S. Kronfeld, J. Laiho, L. Levkova, P.B. Mackenzie, M. Okamoto, J.N. Simone, R. Sugar, D. Toussaint, and R.S. Van de Water. The $B \to \pi l \nu$ semileptonic form factor from three-flavor lattice QCD: A model-independent determination of $|V(ub)|$. *Phys. Rev.*, **D79**:054507, Nov. 2008. [arXiv:hep-lat/0811.3640].

[74] Daniel Arndt and C.-J. David Lin. Heavy meson chiral perturbation theory in finite volume. *Phys. Rev.*, **D70**:014503, 2004. [arXiv:hep-lat/0403012].

[75] Jack Laiho and Ruth S. Van de Water. $B \to D^* l \nu$ and $B \to D l \nu$ form factors in staggered chiral perturbation theory. *Phys. Rev.*, **D73**:054501, 2005. [arXiv:hep-lat/0512007].

[76] R. Sommer. A New Way to Set the Energy Scale in Lattice Gauge Theories and its Application to the Static Force and α_s in SU(2) Yang–Mills Theory. *Nucl. Phys.*, **B411**:839–854, 1994. [arXiv:hep-lat/931002].

[77] C. Bernard, T. Burch, K. Orginos, D. Toussaint, T.A. DeGrand, C.E. DeTar, S. Gottlieb, U.M. Heller, J.E. Hetrick, and R.L. Sugar. The static quark potential in three flavor QCD. *Phy. Rev.*, **D62**:034503, 2000. [arXiv:hep-lat/0002028].

[78] A. Bazavov, C. Bernard, C. DeTar, X. Du, W. Freeman, Steven Gottlieb, Urs M. Heller, J.E. Hetrick, J. Laiho, L. Levkova, M.B. Oktay, J. Osborn, R. Sugar, D. Toussaint, and R.S. Van de Water. Results from the MILC collaboration's SU(3) chiral perturbation theory analysis. *POS (Lattice 2009)*, **079**, Oct. 2009. [arXiv:0910.3618[hep-lat]].

[79] A. Bazavov, C. Bernard, C. DeTar, X. Du, W. Freeman, Steven Gottlieb, U.M. Heller, J.E. Hetrick, J. Laiho, L. Levkova, M.B. Oktay, R. Sugar, D. Toussaint, and R.S. Van de Water. Staggered chiral perturbation theory in the two-flavor case and SU(2) analysis of the MILC data. Nov. 2010. [arXiv:1011.1792[hep-lat]].

[80] R. Todd Evans, Aida X. El-Khadra, and Elvira Gámiz. A determination of the B_s^0 and B_d^0 mixing matrix elements using 2+1 lattice QCD. *POS (Lattice 2008)*, **052**, 2008.

[81] R. Todd Evans, Elvira Gámiz, Aida X. El-Khadra, and Andreas S. Kronfeld. $B_q^0 - \bar{B}_q^0$ Mixing and Matching with Fermilab Heavy Quarks. *POS (Lattice 2009)*, **245**, 2009.

[82] C. Bernard, C. DeTar, M. Di Pierro, A.X. El-Khadra, R.T. Evans, E.D. Freeland, E. Gámiz, Steven Gottlieb, U.M. Heller, J.E. Hetrick, A.S. Kronfeld, J. Laiho, L. Levkova, P.B. Mackenzie, J.N. Simone, R. Sugar, D. Toussaint, and R.S. Van de Water. Tuning Fermilab Heavy Quarks in $2+1$ Flavor Lattice QCD with Application to Hyperfine Splittings. *Phys. Rev.*, **D83**:034503, 2011. [doi:10.1103/PhysRevD.83.034503].

[83] Paulo F. Bedaque. Aharonov-Bohm effect and nucleon-nucleon phase shifts on the lattice. *Phys. Lett.*, **B593**:82–88, Jul. 2004. [doi:10.1016/j.physletb.2004.04.045].

[84] C. Bernard. Weak Matrix Elements on and off the Lattice. In Thomas DeGrand and D. Toussaint, editors, *From Actions to Answers: Procedings of the 1989 TASI in Elementary Particle Physics*, pages 233–292. World Scientific, June 1989.

[85] K.G. Wilson. Confinement of Quarks. *Phys. Rev.*, **D10**:2445–2459, 1974. [doi:10.1103/PhysRevD.10.2445].

[86] M. Lüscher and P. Weisz. Computation of the action for on-shell improved lattice gauge theories at weak coupling. *Phys. Lett.*, **B158**:250–254, Aug. 1985. [doi:10.1016/0370-2693(85)90966-9].

[87] M. Lüscher and P. Weisz. On-shell improved lattice gauge theories. *Commun. Math. Phys.*, **97**:59–77, 1985. [doi:10.1007/BF01206178].

[88] N.B. Nielsen and M. Ninomiya. A no-go theorem for regularizing chiral fermions. *Phys. Lett.*, **B105**:219–223, Oct. 1981. [doi:10.1016/0370-2693(81)91026-1].

[89] N.B. Nielsen and M. Ninomiya. Absence of neutrinos on a lattice : (I). Proof by homotopy theory. *Nucl. Phys.*, **B185**:20–40, 1981. [doi:10.1016/0550-3213(81)90361-8].

[90] B. Sheikholeslami and R. Wohlert. Improved Continuum Limit Lattice Action for QCD with Wilson Fermions. *Nucl. Phys.*, **B259**:572–596, Sep. 1985. [doi:10.1016/0550-3213(85)90002-1].

[91] M. Lüscher and P. Weisz. O(a) improvement of the axial current in lattice QCD to one-loop order of perturbation theory. *Nucl. Phys.*, **B479**:429–458, Nov. 1996. [doi:10.1016/0550-3213(96)00448-8].

[92] M. Lüscher, S. Sint, R. Sommer, and P. Weisz. Chiral symmetry and O(a) improvement in lattice QCD. *Nucl. Phys.*, **B478**:365–397, Oct. 1996. [doi:10.1016/0550-3213(96)00378-1].

[93] A.S. Kronfeld. Application of heavy-quark effective theory to lattice QCD: I. Power Corrections. *Phys. Rev.*, **D62**:014505, Jun. 2000. [arXiv:hep-lat/0002008].

[94] N. Kawamoto and J. Smit. Effective lagrangian and dynamical symmetry breaking in strongly coupled lattice QCD. *Nucl. Phys.*, **B192**:100–124, 1981. [doi:10.1016/0550-3213(81)90196-6].

[95] H. Kluberg-Stern, A. Morel, O. Napoly, and B. Petersson. Spontaneous chiral symmetry breaking for a $U(N)$ gauge theory on a lattice. *Nucl. Phys.*, **B190**:504–518, Oct. 1981. [doi:10.1016/0550-3213(81)90445-4].

[96] H. Kluberg-Stern, A. Morel, and B. Petersson. Spectrum of lattice gauge theories with fermions from a 1/d expansion at strong coupling. *Nucl. Phys.*, **B215**:527–554, Apr. 1983. [doi:10.1016/0550-3213(83)90259-6].

[97] C. van den Doel and J. Smit. Dynamical symmetry breaking in two flavor $SU(N)$ and $SO(N)$ lattice gauge theories. *Nucl. Phys.*, **B228**:122–144, Nov. 1983. [doi:10.1016/0550-3213(83)90401-7].

[98] M.F.L. Golterman and J. Smit. Self-energy and flavor interpretation of staggered fermions. *Nucl. Phys.*, **B245**:61–88, 1984. [doi:10.1016/0550-3213(84)90424-3].

[99] S. Naik. On-shell improved action for QCD with Susskind fermions and the asymptotic freedom scale. *Nucl. Phys.*, **B316**:238–268, Apr. 1989. [doi:10.1016/0550-3213(89)90394-5].

[100] M. Creutz. Comments on staggered fermions / Panel discussion. *POS (Confinement8)*, **16**, Oct. 2008. [arXiv:0810.4526[hep-lat]].

[101] Gordon. C. Donald, Christine T.H. Davies, Eduardo Follana, and Andreas S. Kronfeld. Staggered fermions, zero modes, and flavor-singlet mesons. *arXiv:1106.2412v1 [hep-lat]*, Jun. 2011.

[102] Matthew Wingate, Junko Shigemitsu, Christine T.H. Davies, G. Peter Lepage, and Howard D. Trottier. Heavy-Light Mesons with Staggered Light Quarks. *Phys. Rev.*, **D67**:054505, 2003. [doi:10.1103/PhysRevD.67.054505].

[103] Bradley Efron. *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics, 6th edition, 1994.

[104] D. Toussaint. Private communication.

[105] Philip R. Bevington and D. Keith Robinson. *Data Reduction and Error Analysis for the Physical Sciences.* McGraw-Hill, 3rd edition, 2003.