EM sea effects in hadron polarizabilities through reweighting
Improved methods and results

Walter Freeman, Andrei Alexandru, Michael Lujan, Frank Lee

The George Washington University

June 26, 2014

Outline

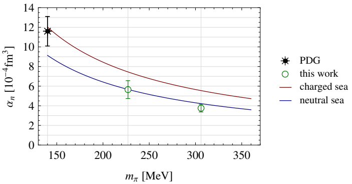
- Review of the background field method (prior talk)
- Charging the sea quarks
- Reweighting, in general
- Perturbative reweighting
- Simulation parameters
- First calculation: $24^3 \times 48$ ensemble (done)
  - Improving the stochastic estimator: offdiagonal element mapping
  - The body-centered hypercubic dilution scheme
  - Results
- Second calculation: $48 \times 24^2 \times 48$ ensemble (in progress)
  - Further improvement: Low-mode subtraction
  - Results so far
  - Future directions and progress

Review: The background field method

- Expect a mass shift equal to $\alpha_E E^2$ in the presence of a uniform background field
- Determine polarizability by measuring neutron correlators with $\vec{E} = 0$ and $\vec{E} = \pm i E_0 \hat{x}$, then fitting them to determine the mass shift
- Use the same value of $\eta_d \equiv a^2 q E_0 = 10^{-4}$ as in the valence-only study
  - Small enough to be well within the perturbative regime
  - Large enough to avoid numerical precision issues and extreme inverter precisions
- When fitting correlators, the zero-field and nonzero-field correlators are correlated
- These correlations result in a much smaller error on ΔM than on the mass measurements themselves

Coupling to the charged sea

- α_E describes the deformation of the hadron in response to an electric field
- As in other cases (spin, σ term), both valence and sea quarks contribute
- A valence-only calculation ignores crucial dynamics
- Potentially large contribution: "stretching of the pion cloud"
- Likely to have large finite volume effects at small volume



- χPT predicts that charged sea quarks account for $\sim 25\%$ of the polarizability for the neutron

Simulation parameters

Same ensembles used for the broader polarizability study (we use the same valence correlators):

- $(24, 48) \times 24^2 \times 48$ lattices, 2 flavors of dynamical nHYP-clover fermions, 300 configs each
- $m_\pi = 306$ MeV
- $a = 0.1245(16)$ fm by Sommer scale r_0
- Periodic boundary conditions used in MD for gauge generation; Dirichlet BC's applied afterwards

Reweighting approach

How do we include the effects of the sea quark charges in the background field approach?

- In principle it's easy: just generate two otherwise identical ensembles, one with a background field and one without
- But this requires unaffordably high statistics, since our two mass measurements now no longer have correlated errors
  - Lose all the information in the "cross-correlation" terms of the covariance matrix
- Reweighting is a technique for extracting physics from a different action than the one used in generation: "retroactively change the ensemble parameters"
- We can use it to generate two correlated ensembles, one with and one without the electric field
- As in the valence-only case, the strength of the correlation between these ensembles drives the overall error

Determining the weight factors

In order to do reweighting, need the weight factors $w_i = e^{-\Delta \mathcal{S}} = \det^{-1} \mathbf{M}_\eta^{-1} \mathbf{M_0}$.

- This weight factor is generally estimated stochastically
- So long as the estimator is unbiased, the result will be too – just with larger error bars
  - For ordinary reweighting calculations, reweighting "succeeds" if the weight factors don't fluctuate much
  - Not true for us: we also require strong correlations between unreweighted and reweighted correlators
  - Possible to have fluctuations large enough to destroy these correlations
- There is a standard stochastic estimator for the inverse determinant
- Far too noisy when reweighting in the background field (standard improvements don't work)
- Try another approach: perturbative reweighting

Perturbative reweighting

- Idea: expand weight factor as a power series in a small parameter, keep only a few orders
- Useful whenever we only want a perturbatively-small shift in the action
  - Shift m_s by perturbatively-small amount $\rightarrow$ compute $\frac{\partial M_N}{\partial m_s}$ for nucleon strangeness
  - Turn on perturbatively-small electric field for the sea quarks
- Easier to estimate $\left.\frac{\partial w_i}{\partial \eta}\right|_{\eta=0}$ and $\left.\frac{\partial^2 w_i}{\partial \eta^2}\right|_{\eta=0}$ than w_i itself?
- Expand w_i in a power series in η up to second order, about $\eta = 0$
  - Linear term in weight factor can combine with linear dependence of $G_N(t)$ on η to give quadratic effect
  - Quadratic term in weight factor by itself can give quadratic effect
- If we can estimate these derivatives instead we can evaluate at any η we choose to get $w_i(\eta)$
- Sea contributions taken into account in a way that is similar in practice to the current-insertion approach of Engelhardt

Derivation of the estimator

For the first derivative, we want $\frac{\partial}{\partial\eta} \frac{\det M_\eta}{\det M_0}\Big|_{\eta=0}$. Rewrite $\det M_\eta$ as a Grassman integral:

$$\begin{aligned}
\frac{\partial}{\partial\eta} \left.\frac{\det M_\eta}{\det M_0}\right|_{\eta=0} &= \frac{1}{\det M_0} \frac{\partial}{\partial\eta} \int d\psi d\bar\psi \, e^{-\bar\psi M \psi} \\
&= \frac{1}{\det M_0} \int d\psi d\bar\psi \; -\bar\psi \frac{\partial M_0}{\partial\eta} e^{-\bar\psi M_0 \psi} \\
&= \mathrm{Tr}\left(\frac{\partial M_0}{\partial\eta} M_0^{-1} \right).
\end{aligned}$$

- Standard stochastic estimator: $\mathrm{Tr}\,\mathcal{O} = \langle \xi | \mathcal{O} | \xi \rangle_\xi$; $\xi \in \mathbb{Z}(4)$

Derivation of the estimator

For the first derivative, we want $\frac{\partial}{\partial\eta}\left.\frac{\det M_\eta}{\det M_0}\right|_{\eta=0}$. Rewrite $\det M_\eta$ as a Grassman integral:

$$\begin{aligned}
\frac{\partial}{\partial\eta}\left.\frac{\det M_\eta}{\det M_0}\right|_{\eta=0} &= \frac{1}{\det M_0}\frac{\partial}{\partial\eta}\int d\psi d\bar{\psi}\, e^{-\bar{\psi}M\psi} \\
&= \frac{1}{\det M_0}\int d\psi d\bar{\psi}\; -\bar{\psi}\frac{\partial M_0}{\partial\eta}e^{-\bar{\psi}M_0\psi} \\
&= \mathrm{Tr}\left(\frac{\partial M_0}{\partial\eta}M_0^{-1}\right).
\end{aligned}$$

- Standard stochastic estimator: $\mathrm{Tr}\,\mathcal{O} = \langle\xi|\mathcal{O}|\xi\rangle_\xi$; $\xi \in \mathbb{Z}(4)$

The second derivative proceeds similarly:

$$\frac{\partial^2}{\partial\eta^2}\left.\frac{\det M_\eta}{\det M_0}\right|_{\eta=0} = \mathrm{Tr}\frac{\partial^2 M}{\partial\eta^2}M_0^{-1} - \mathrm{Tr}\left(\frac{\partial M}{\partial\eta}M_0^{-1}\right)^2 + \left(\mathrm{Tr}\frac{\partial M}{\partial\eta}M_0^{-1}\right)^2$$

Derivation of the estimator

For the first derivative, we want $\frac{\partial}{\partial\eta} \left.\frac{\det M_\eta}{\det M_0}\right|_{\eta=0}$. Rewrite $\det M_\eta$ as a Grassman integral:

$$\begin{aligned}\frac{\partial}{\partial\eta} \left.\frac{\det M_\eta}{\det M_0}\right|_{\eta=0} &= \frac{1}{\det M_0}\frac{\partial}{\partial\eta}\int d\psi d\bar\psi\, e^{-\bar\psi M\psi} \\ &= \frac{1}{\det M_0}\int d\psi d\bar\psi\; -\bar\psi\frac{\partial M_0}{\partial\eta} e^{-\bar\psi M_0\psi} \\ &= \mathrm{Tr}\left(\frac{\partial M_0}{\partial\eta}M_0^{-1}\right).\end{aligned}$$

- Standard stochastic estimator: $\mathrm{Tr}\,\mathcal{O} = \langle\xi|\mathcal{O}|\xi\rangle_\xi$; $\xi\in\mathbb{Z}(4)$

The second derivative proceeds similarly:

$$\frac{\partial^2}{\partial\eta^2} \left.\frac{\det M_\eta}{\det M_0}\right|_{\eta=0} = \mathrm{Tr}\frac{\partial^2 M}{\partial\eta^2}M_0^{-1} - \mathrm{Tr}\left(\frac{\partial M}{\partial\eta}M_0^{-1}\right)^2 + \left(\mathrm{Tr}\frac{\partial M}{\partial\eta}M_0^{-1}\right)^2$$

(Notation: These first two terms nearly cancel; we call them collectively $\tilde w''$. The third is w'^2.)

Derivation of the estimator

For the first derivative, we want $\frac{\partial}{\partial\eta}\,\frac{\det M_\eta}{\det M_0}\Big|_{\eta=0}$. Rewrite $\det M_\eta$ as a Grassman integral:

$$\begin{aligned}\frac{\partial}{\partial\eta}\,\frac{\det M_\eta}{\det M_0}\Big|_{\eta=0} &= \frac{1}{\det M_0}\frac{\partial}{\partial\eta}\int d\psi d\bar\psi\, e^{-\bar\psi M\psi} \\ &= \frac{1}{\det M_0}\int d\psi d\bar\psi\ -\bar\psi\frac{\partial M_0}{\partial\eta}e^{-\bar\psi M_0\psi} \\ &= \mathrm{Tr}\left(\frac{\partial M_0}{\partial\eta}M_0^{-1}\right).\end{aligned}$$

- Standard stochastic estimator: $\mathrm{Tr}\,\mathcal{O} = \langle\xi|\mathcal{O}|\xi\rangle_\xi;\ \xi\in\mathbb{Z}(4)$

The second derivative proceeds similarly:

$$\frac{\partial^2}{\partial\eta^2}\,\frac{\det M_\eta}{\det M_0}\Big|_{\eta=0} = \mathrm{Tr}\frac{\partial^2 M}{\partial\eta^2}M_0^{-1} - \mathrm{Tr}\left(\frac{\partial M}{\partial\eta}M_0^{-1}\right)^2 + \left(\mathrm{Tr}\frac{\partial M}{\partial\eta}M_0^{-1}\right)^2$$

(Notation: These first two terms nearly cancel; we call them collectively $\tilde w''$. The third is w'^2.)

These four terms correspond directly to the four disconnected diagrams that must be evaluated in the fully-perturbative direct evaluation approach.

Unfortunately, stochastic estimators of the traces here are still very noisy, but can be improved in a variety of ways.

Origin of stochastic noise – first order

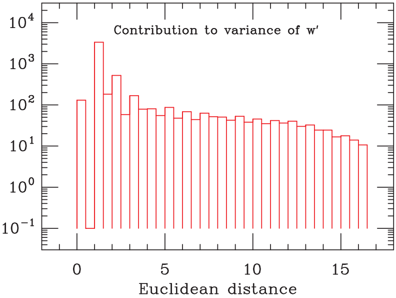
In order to design an improvement technique, we need to understand where the stochastic noise comes from.

Since $\mathrm{var}(\mathrm{Tr}\,\mathcal{O}) = \sum_{i \neq j} |\mathcal{O}_{ij}|^2$; need to understand which offdiagonal elements $\mathcal{O}_{ij}$ contribute most.

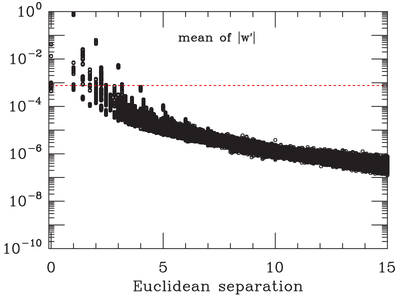
Can't examine them all, but can map a representative sample of them.

Can look at this in two ways:

Contribution by Euclidean distance:



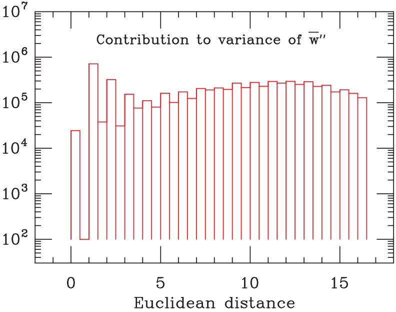
Matrix element size by Euclidean distance:



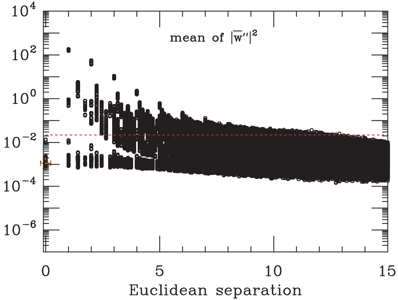
Origin of stochastic noise – second order

A preliminary study indicates that stochastic fluctuations in the second-order term cause a larger overall hit to the error bar, so we should look at it too:

Contribution by Euclidean distance:

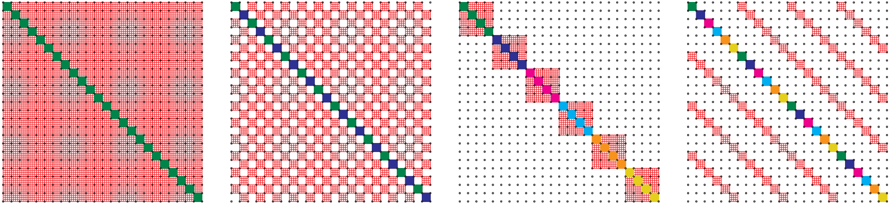


Matrix element size by Euclidean distance:



Designing an improvement technique

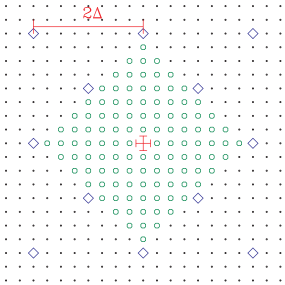
- Our preliminary study used 7^{th}-order hopping parameter expansion
- While this substantially reduces the variance of the near-diagonal elements, it can only take us so far (cost $\sim 14^n$), and we need something different.
- Most of the noise comes from near-diagonal elements; we can eliminate them with dilution.
- Dilution separates the matrix dimension into N subsets and stochastically estimates the trace over each separately
  - Advantage: eliminates noise contributions from offdiagonal terms from different subsets
  - Disadvantage: Requires N operations to cover the lattice; could have reduced noise by factor of $\sqrt{N}$ by simple repetition
  - Only outperforms simple repetition if the offdiagonal matrix elements "kept" are lower than the average



- We should choose a (four-dimensional generalization) of the rightmost scheme to eliminate the large near-diagonal elements

Body-centered hypercubic dilution

- In general, better to use the largest-N dilution scheme on which we can afford a single estimator (dilution is better than repetition)
- In the limit as $N \to 12V$, the estimator variance goes to zero (full dilution)
- Since $\mathcal{O}_{ij}$ decays exponentially away from the diagonal, try to maximize the minimum separation between elements of the same subspace
- In 2D, this is reasonably done with a grid with additional red/black coloring

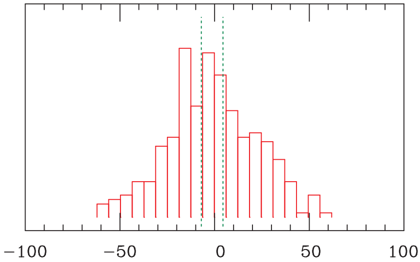


- If the grid spacing here is Δ, the minimum separation is $\sqrt{\Delta}$
- 4D extension: body-centered hypercubic lattice
- If the unit cell is 2Δ, the minimum separation is also 2Δ
- This achieves the same minimum separation at half the cost of a grid
- Must be combined with spin-color dilution

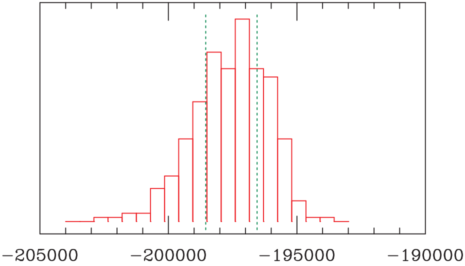
First production run: $24^3 \times 48$ ensemble

- We used a BCHC scheme with $\Delta = 6$, giving $N = 124,416$ per configuration
- We repeated our stochastic estimators 6 times on a single configuration to estimate their noise
- These noise levels correspond well with estimates obtained from doing a “simulated” dilution on the $\mathcal{O}_{ij}$ data
- Benchmark goal for reweighting: $\sigma_{\rm gauge} > \sigma_{\rm stoc}$

Estimates of w'



Estimates of $\tilde{w}''$



Goal handily met for the first-order term; not there yet at second order

Results

Of course, the end goal is to compute the polarizabilities:

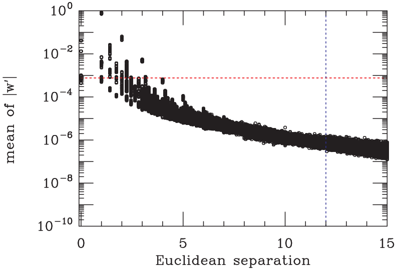
	Valence only	1^{st} order	$\tilde{w}_q''$ only	2^{nd} order
Neutron	2.56(19)	2.60(22)	2.89(55)	2.70(55)
Pion	-0.21(14)	-0.24(14)	0.21(22)	0.22(23)
Kaon	0.14(3)	0.13(3)	0.36(12)	0.38(12)

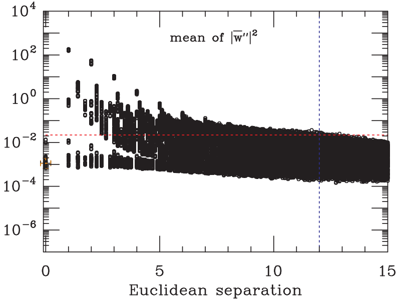
- The first-order reweighting neither affects the errors or central values much
- The second-order reweighting increases the errors significantly; need to do better with estimates of $\tilde{w}''$
- Results still have lower/comparable error bars to other polarizability calculations
- In general the full reweighting doesn't cause any effect
- The exception is the kaon, where the second-order terms cause a sizable shift in the central value
- This is consistent with the sensitivity of the kaon polarizability on the sea behavior seen in the previous talk

Second run: the $48 \times 24^2 \times 48$ ensemble

- Effects of the charged sea expected to be smallest on the $24^3 \times 48$, $m_\pi = 306$ MeV ensemble
  - "Stretching" the pion cloud difficult in such a small volume
- The ultimate goal is to reweight all the ensembles to do the same full volume and chiral study as the valence-only data
- We need a better way to estimate the weight factors, in particular at second order; affording 250k inversions/config on the larger volumes is tough

Look at the off-diagonal element maps again:





Need to do a better job of reducing long-distance contributions!

Strategy: low mode substitution

- We can increase the falloff rate of the off-diagonal elements by removing low modes from M^{-1} and treating them separately: standard LMS idea
  - This is a technique to reduce noise from the long-distance tail
  - Only successful in combination with some other technique (dilution) to deal with the short-distance noise
  - Remove low modes of the Dirac operator, *not* the operator whose trace is being computed
- Subtract the low modes from the inverses in the stochastic estimator, and add their exact traces back in later
- Problem: we're using Wilson quarks, so can't get low modes of M since it's not Hermitian
- ... but the low modes of $\gamma_5 M$ capture the long-distance behavior about as well

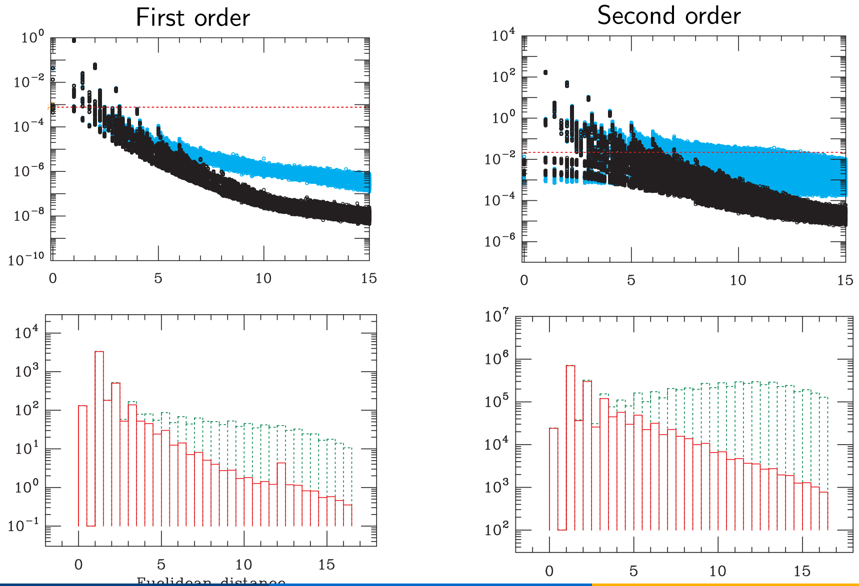
Define $M_l = \sum \lambda_i \gamma_5 |\lambda_i\rangle \langle\lambda_i|$, where $|\lambda_i\rangle$ are the eigenmodes of $\gamma_5 M$, and $M_h = M - M_l$. Then:

$$
\begin{aligned}
\mathrm{Tr}\, M' M^{-1} &= \left\langle \xi | M' M_h^{-1} | \xi \right\rangle + \mathrm{Tr}\, M' M_l^{-1} \\
\mathrm{Tr}\, M'' M^{-1} &= \left\langle \xi | M'' M_h^{-1} | \xi \right\rangle + \mathrm{Tr}\, M'' M_l^{-1} \\
\mathrm{Tr}\, M' M^{-1} M' M^{-1} &= \left\langle \xi | M' M_h^{-1} M' M_h^{-1} | \xi \right\rangle + 2\mathrm{Tr}\, M' M_l^{-1} M' M^{-1} - \mathrm{Tr}\, M' M_l^{-1} M' M_l^{-1}
\end{aligned}
$$

The exact traces can be computed as sums over eigenvectors; cost dominated by N_{ev} inversions (not too bad).

LMS: expected payoff

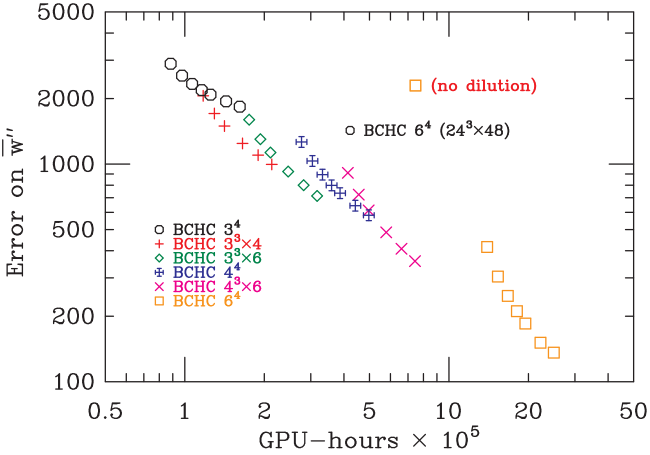
We can use the $\mathcal{O}_{ij}$-mapping technique to examine the benefit of LMS with 2000 eigenvectors:



Performance vs. cost

Some overhead associated with LMS:

- If the linear algebra is done on the CPU, it takes meaningful time
- If the linear algebra is done on the GPU, it hurts scaling from memory requirements



- Based on our resources we chose BCHC $3^3 \times 6$ with 1000 eigenvectors
- This should give us lower errors in less time despite an ensemble with double the volume

Conclusions

- Perturbative reweighting can be used to calculate the sea polarizability (and many other things too)
- Stochastic estimators are very, very hard – consequence of this problem, not perturbative reweighting in general
- Offdiagonal-element mapping can be used to plan and evaluate estimator improvement strategies
- $24^3 \times 48$ ensemble: $N = 124,416$ dilution gives charged-sea polarizabilities with reasonable errors
- Kaon shows significant shift; other particles unaffected
- Sea effects expected to be bigger for other ensembles in our study
- Low-mode substitution along with strong dilution pays off for this problem
- $48 \times 24^2 \times 48$ ensemble: $N = 15,552$ dilution along with 1000-vector LMS should give lower errors in less time
- ... stay tuned for the result