Algorithms for Disconnected Diagrams

[ab]Arjun Singh Gambhir, [c]Andreas Stathopoulos, [ab]Kostas Orginos

July 27, 2016, Lattice 2016 Conference, Southampton

[a]Department of Physics, College of William and Mary
[b]Thomas Jefferson National Accelerator Facility
[c]Department of Computer Science, College of William and Mary

Outline

- Introduction
- Deflation Algorithm
- Numerical Experiments
- Application to Nucleon Charges
- Conclusion and Acknowledgements

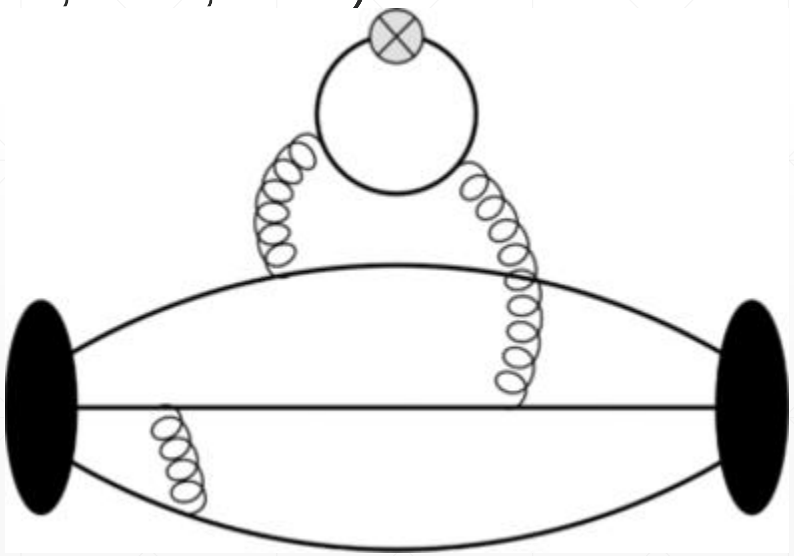
Outline

- Introduction
- Deflation Algorithm
- Numerical Experiments
- Application to Nucleon Charges
- Conclusion and Acknowledgements

Computing Disconnected Contributions

- Need trace of all to all propagator $\mathrm{Tr}(\Gamma A^{-1})$. Stochastic methods need help.
- Variance reduction techniques such as dilution and eigenvalue deflation have been proposed - arXiv:hep-lat/0106016 (H. Neff, et al., 2001), arXiv:hep-lat/0505023 (Foley, et al., 2005), arXiv:0710.5536 (Babich, et al., 2007).
- Hutchinson method, Z_4:

$$
\begin{aligned}
t(A^{-1}) &= \frac{1}{s}\sum_{j=1}^{s} z_j^{\dagger} A^{-1} z_j \\
\mathrm{Var}(t(A^{-1})) &= \|\tilde{A}^{-1}\|_F^2 = \left(\|A^{-1}\|_F^2 - \sum_{i=1}^{N} |A_{i,i}^{-1}|^2\right)
\end{aligned}
$$



Hierarchical Probing

- Classical probing (CP) takes advantage of decay in $A^{-1}_{i,j}$ by discovering structure.
- Dilution: probing based on known structure (red black, spin/color, or timeslice).
- Hierarchical probing (HP) uses nested coloring to approximate CP; quadratures may be re-used.
- HP basis described by reordered Hadamard matrix for lattices of power 2.
- Simple toy model:
- $\begin{matrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{matrix}$ CP: $\begin{matrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{matrix}$ HP: $\begin{matrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{matrix}$; CP: $\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$ HP: $\begin{matrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{matrix}$

HP's successes and limitations

- HP has provided a 10x speedup over pure noise for strange quark contributions - arXiv:1302.4018 (Stathopoulos, Laeuchli, and Orginos, 2013) .
- Estimator made unbiased by element-wise product with random noise.
- At lighter quark masses, A^{-1} dominated by near null space components that don't decay.
- HP is not expected to reduce variance appreciably at these masses.
- Is there a way to induce appropriate decay for HP?

Outline

- Introduction
- Deflation Algorithm
- Numerical Experiments
- Application to Nucleon Charges
- Conclusion and Acknowledgements

Applying SVD Deflation to the Trace Estimator

- Remove low singular value space (SVD) from the matrix.
- Effects of SVD deflation studied in arXiv:1603.05988.
- An improved algorithm may be utilized with two phases.
- Phase 1: obtain the lowest singular triplets.
- Phase 2: employ trace estimator with low modes projected out.

$$\mathrm{Tr}(A^{-1}) \approx \frac{1}{s}\sum_{j=1}^{s}\left(z_j^\dagger A^{-1} z_j - z_j^\dagger V\Sigma^{-1}U^\dagger z_j\right) + \mathrm{Tr}(V\Sigma^{-1}U^\dagger)$$

Generating the SVD Space

- Obtaining lowest singular triplets (phase 1) efficiently is of central importance.
- Eigensolver works on $A^\dagger A$.
- Challenging due to dense low lying spectrum.
- Accelerate eigensolver by a preconditioner with high overlap of low modes.
- Used PReconditioned Iterative MultiMethod Eigensolver (PRIMME) blocked method with relaxed tolerance.
- Algebraic Multigrid (AMG) preconditioner drives PRIMME to low modes.
- Generated 1,000 lower singular triplets.

Deflated Trace Estimation Algorithm

Algorithm 1 $Trace = \text{deflatedHP}(A)$

$[\Lambda, V] = \text{PRIMME}(A^\dagger A)$
$T_D = \text{Tr}(\Lambda^{-1} V^\dagger A^\dagger V)$;
$T_R = 0$;
$z_0 = \text{random}(z_4)$;
for $j = 1 : s$ **do**
  $z_h =$ next vector from Hierarchical Probing or other scheme
  $z = z_0 \odot z_h$
  Solve $Ay = z$
  $T_R = T_R + z^\dagger y - z^\dagger V \Lambda^{-1} V^\dagger (A^\dagger z)$
end for
$Trace = T_R / s + T_D$

Outline

- Introduction
- Deflation Algorithm
- Numerical Experiments
- Application to Nucleon Charges
- Conclusion and Acknowledgements

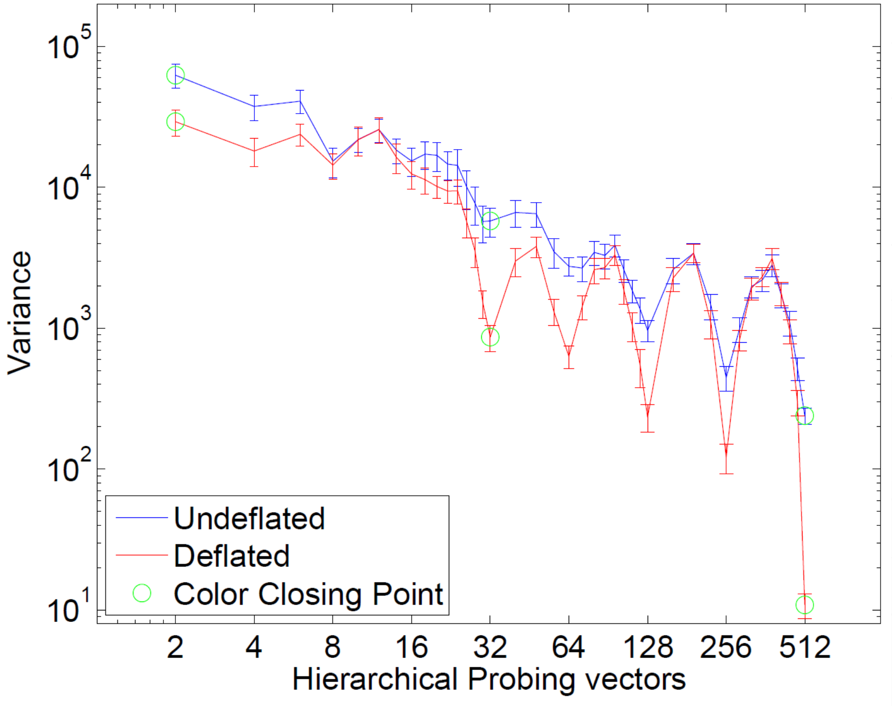
Variance Reduction

- Lattice setup: $32^3 \times 64$, 2+1 flavor Clover-Wilson ensemble $a = 0.081\ fm$ and $m_\pi = 312\ MeV$

- Errors on variance estimated with Jacknife resampling over $n = 40$ samples.

- The deep minima of 2, 32, and 512 (color closing points) give the highest variance reduction.

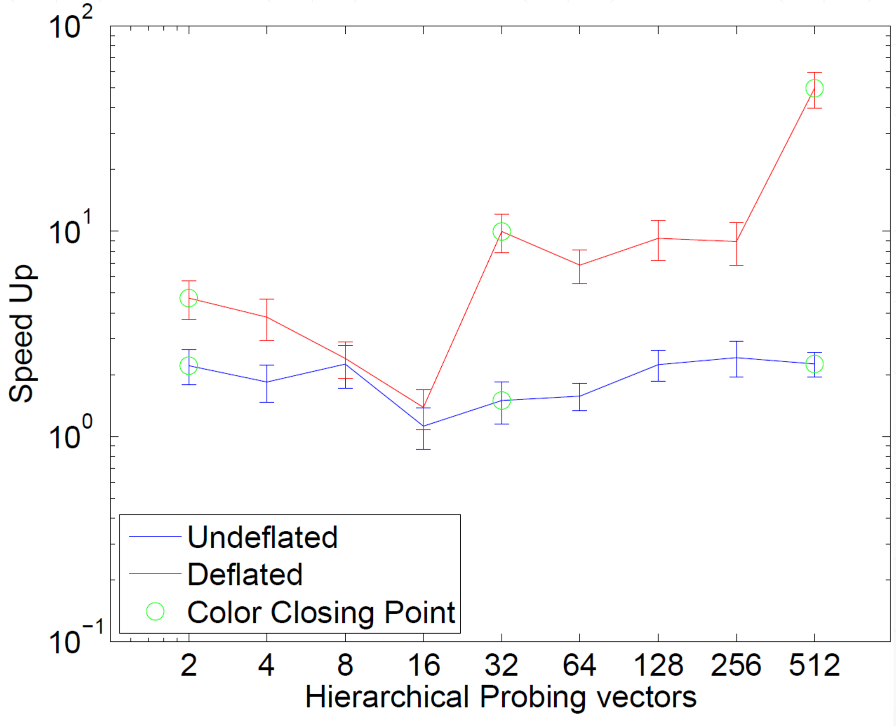
- At 512 probing vectors, HP with SVD Deflation reduces variance by a factor of 20 compared to HP alone.



Speedup

- $\mathrm{SpeedUp} = \frac{V_{stoc}}{V_{hp}(s) \times s}$, where V_{stoc} is variance of a purely stochastic Z_4 estimator, $V_{hp}(s)$ is variance of HP + Deflation for the s^{th} probing vector.

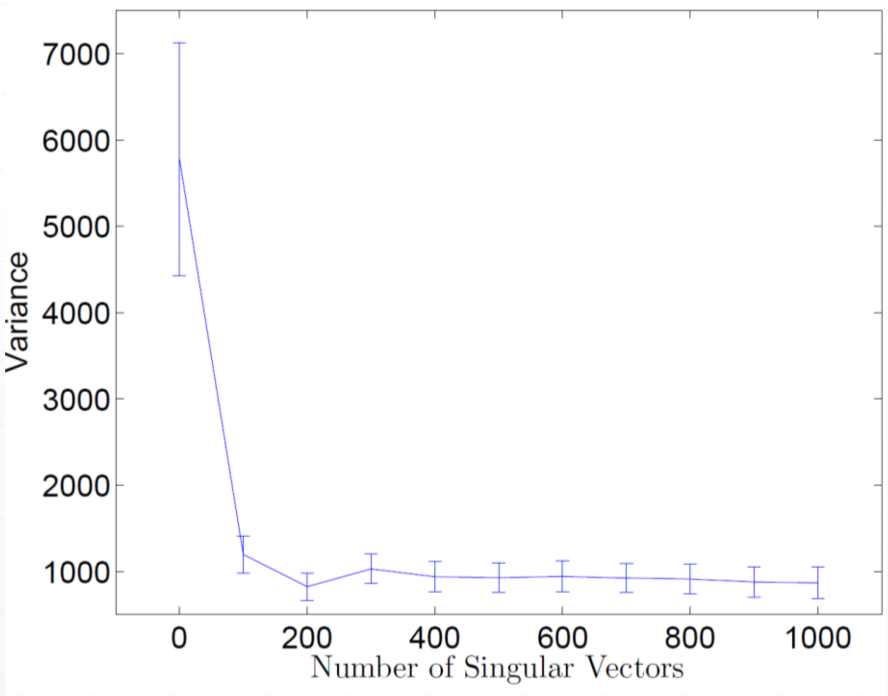
- The factor of s rescales the stochastic variance, since it's error goes down as $\frac{1}{\sqrt{s}}$
- The greatest speedup is observed at the color closing points 2, 32, and 512.
- HP without deflation only gives a 3x speedup.
- At 512 probing vectors, HP with Deflation has a 60x speedup over pure noise.



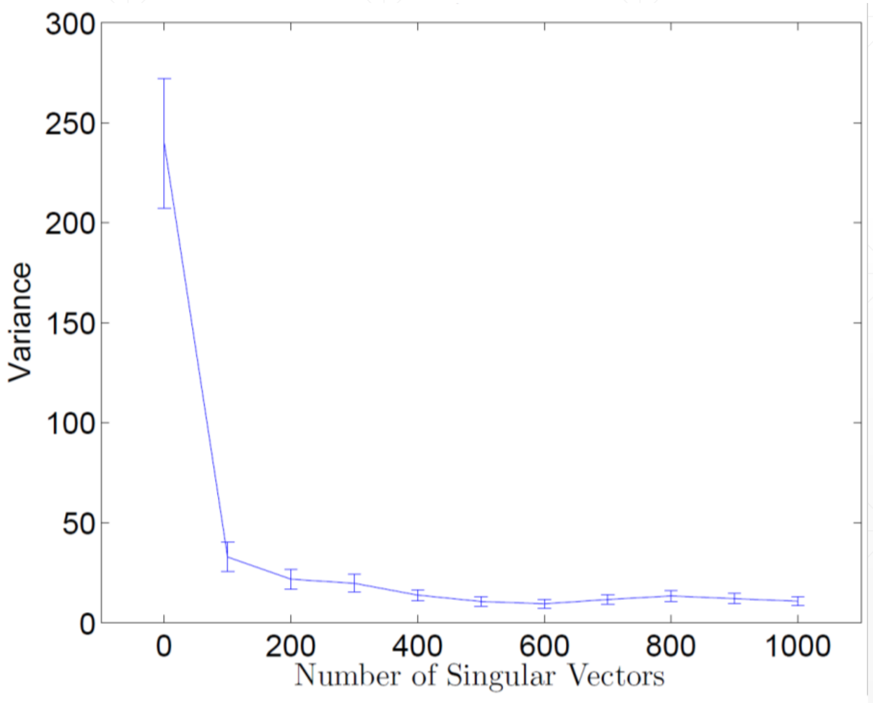
Varying Low Mode Subspace

- The effect of varying number of singular triplets. Not all 1,000 are needed.

Varying SVD Space at 32 HP vectors



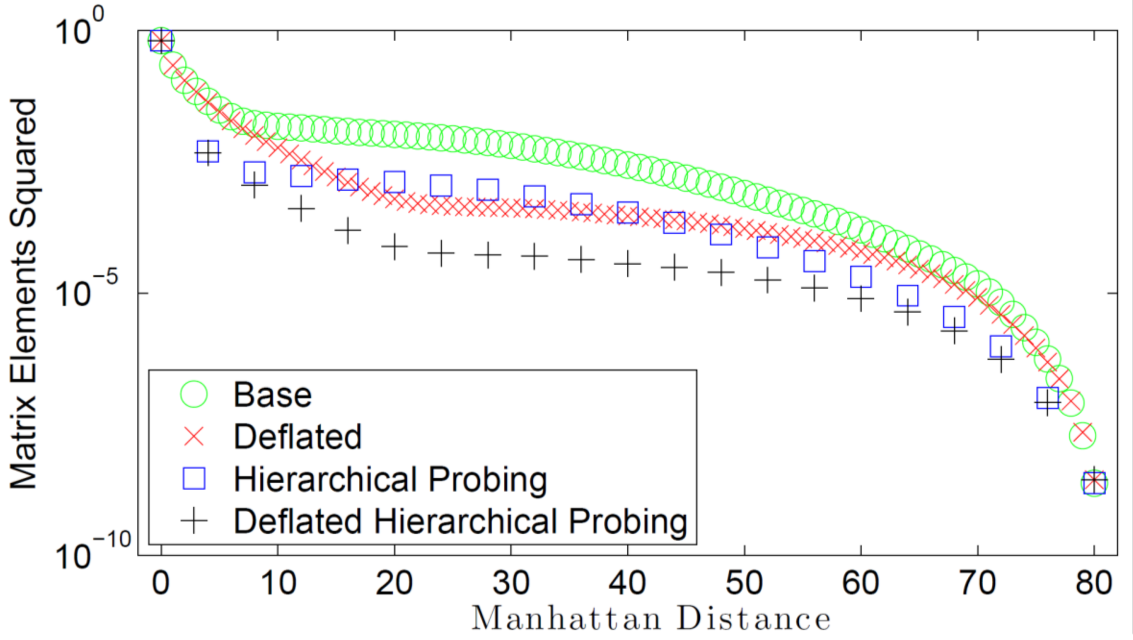
Varying SVD Space at 512 HP vectors



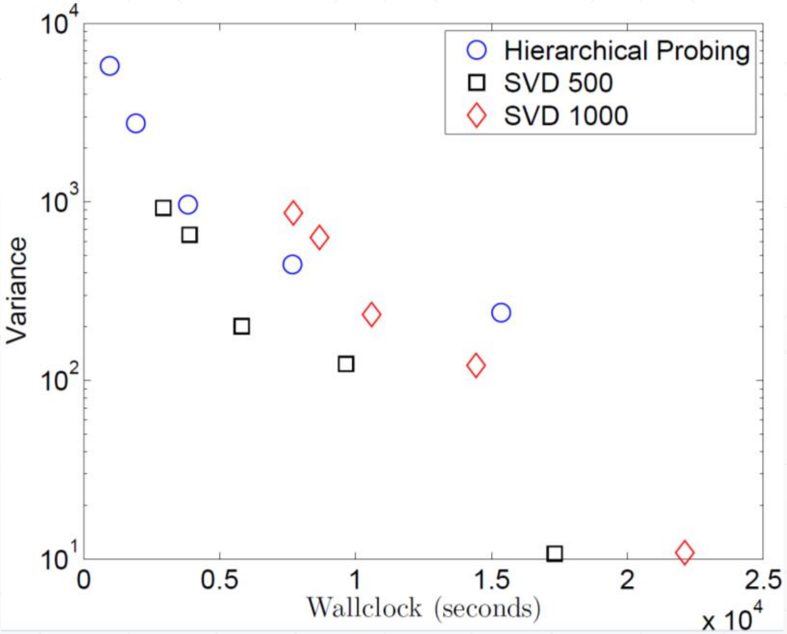
Synergy between Deflation and HP

Elements squared of 10 randomly selected rows of A^{-1}, HP given by:

$$\|(H_{32}H_{32}^{H}) \odot A^{-1}\|_F^2$$



Cost of full calculations for 32, 64, 128, 256, and 512 HP vectors.

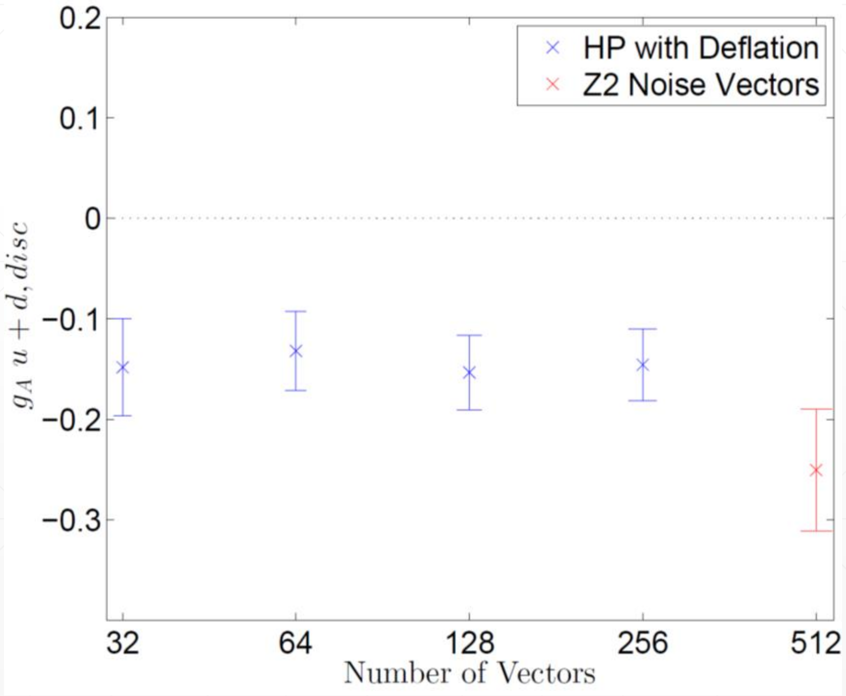


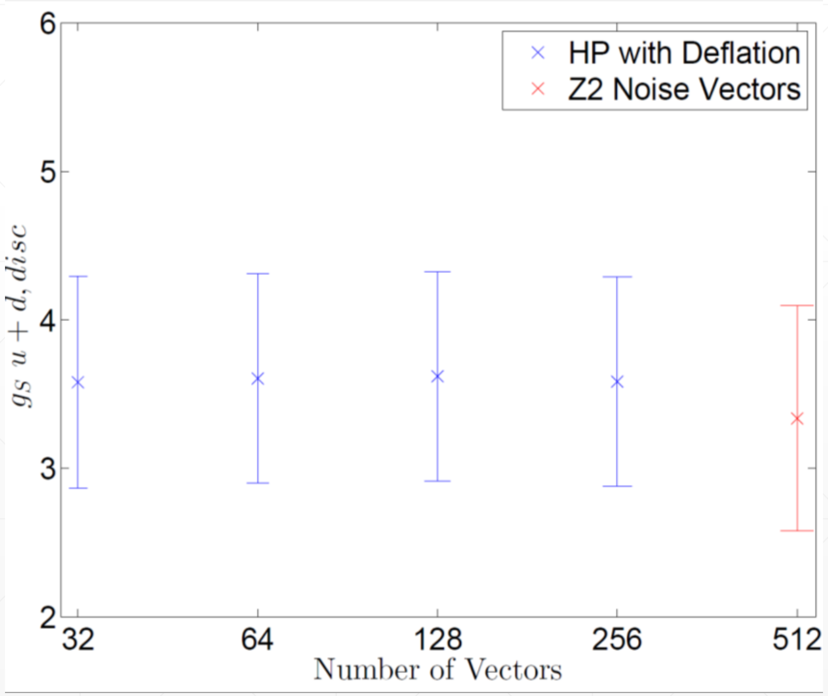
Outline

- Introduction
- Deflation Algorithm
- Numerical Experiments
- Application to Nucleon Charges
- Conclusion and Acknowledgements

Nucleon Charges Comparison

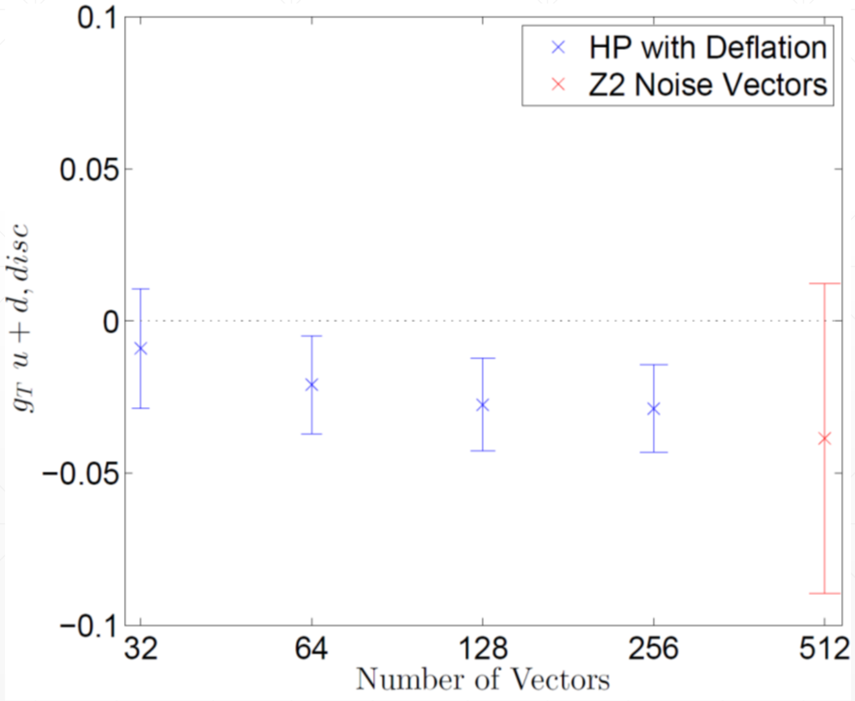
- $32^3 \times 64$, 2+1 flavor Clover-Wilson ensemble $a = 0.081\, fm$ and $m_\pi = 312\, MeV$
- $t_{sep} = 10, \tau = 5$, 100 cfgs
- 500 SVD plus # HP compared to 512 Z_2 noisy sources

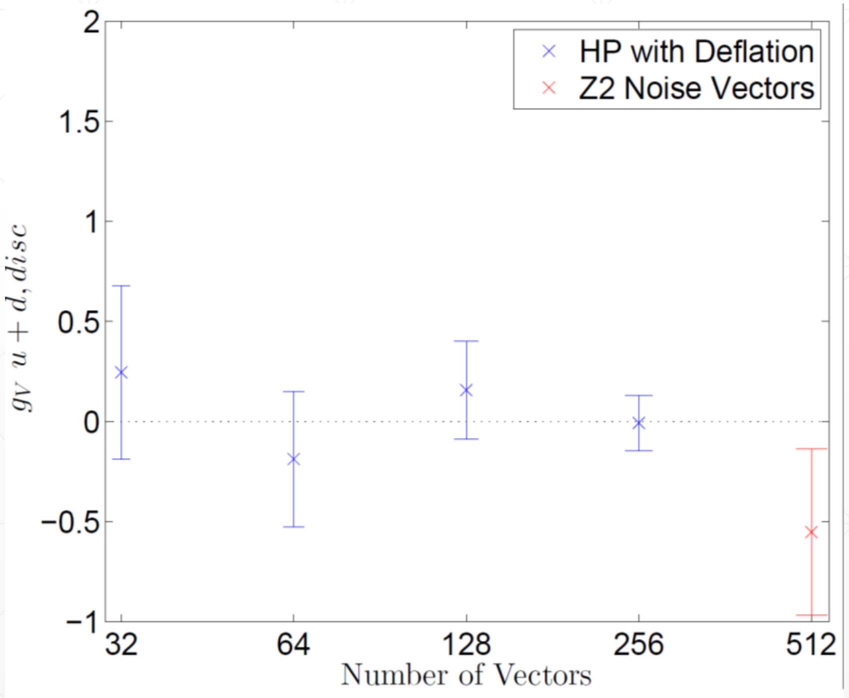




Nucleon Charges Comparison Continued

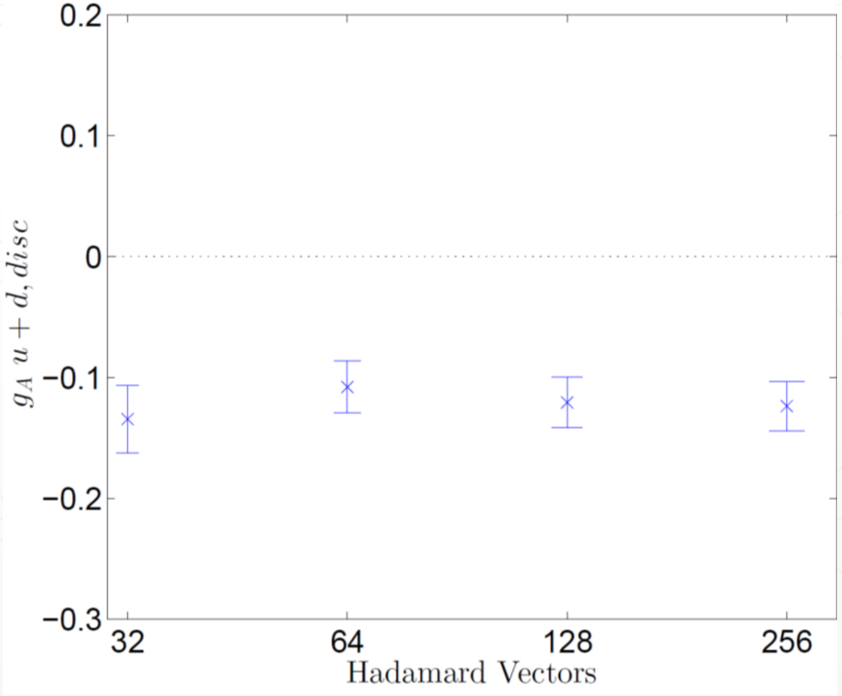
- $32^3 \times 64$, 2+1 flavor Clover-Wilson ensemble $a = 0.081\ fm$ and $m_\pi = 312\ MeV$
- $t_{sep} = 10, \tau = 5$, 100 cfgs
- 500 SVD plus # HP compared to 512 Z_2 noisy sources

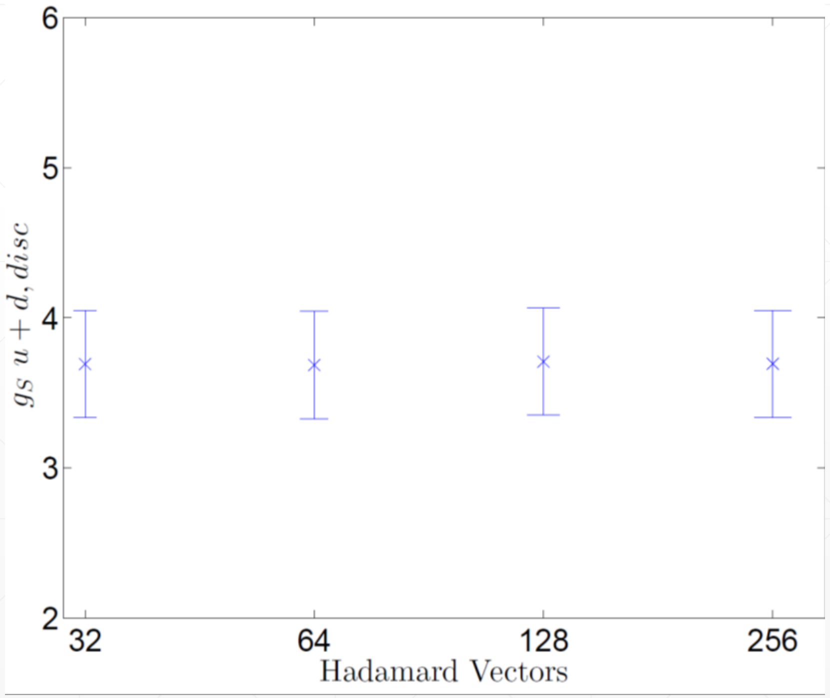




Nucleon Charges (Higher Statistics)

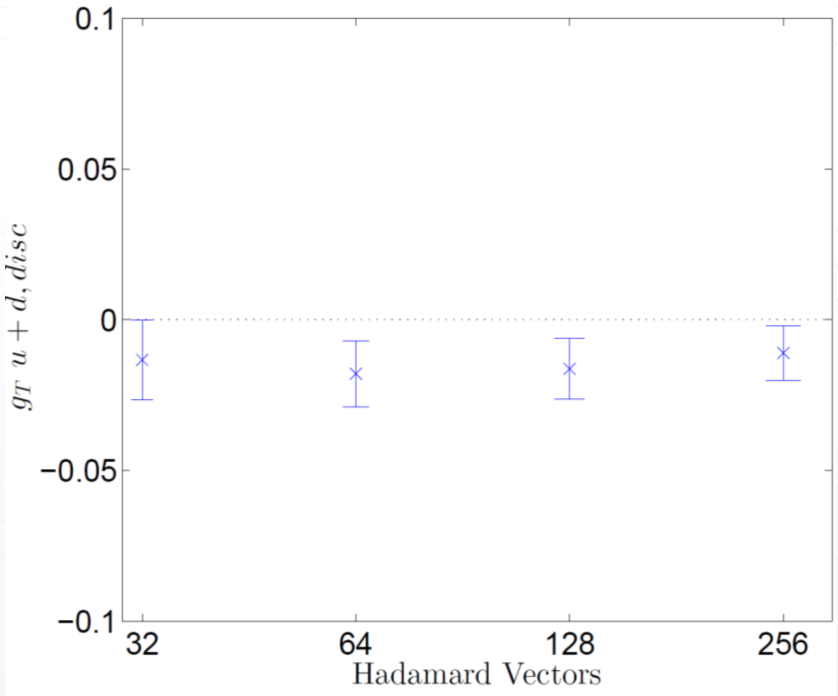
- $32^3 \times 64$, 2+1 flavor Clover-Wilson ensemble $a = 0.081\ fm$ and $m_\pi = 312\ MeV$
- $t_{sep} = 10, \tau = 5$, 400 cfgs
- 500 SVD plus # HP

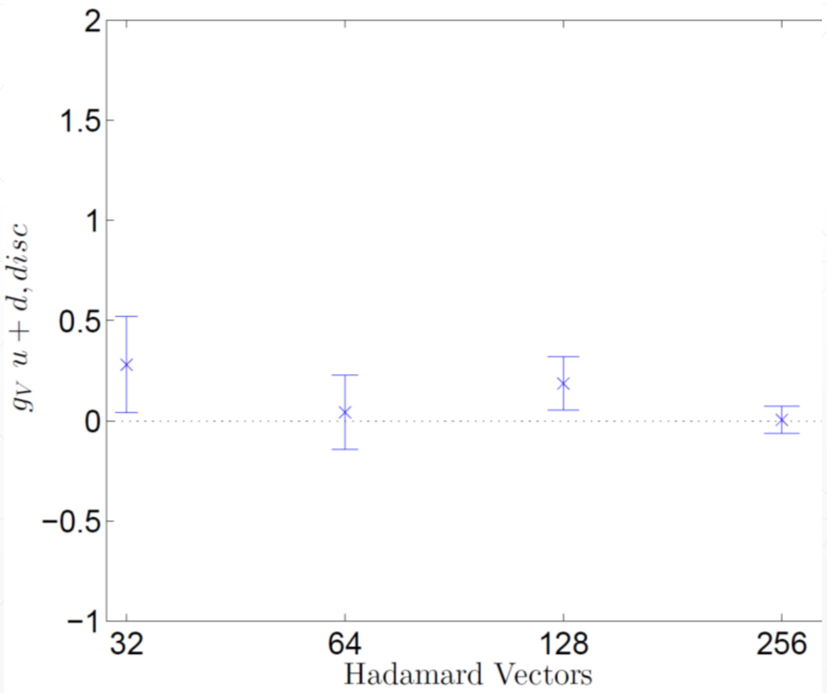




Nucleon Charges (Higher Statistics) Continued

- $32^3 \times 64$, 2+1 flavor Clover-Wilson ensemble $a = 0.081\ fm$ and $m_\pi = 312\ MeV$
- $t_{sep} = 10, \tau = 5$, 400 cfgs
- 500 SVD plus # HP





Outline

- Introduction
- Deflation Algorithm
- Numerical Experiments
- Application to Nucleon Charges
- Conclusion and Acknowledgements

Conclusions and Outlook

- Hierarchical Probing (HP) does not work as well at lighter quark masses.
- HP and SVD deflation provide significant variance reduction over HP alone.
- SVD subspace size must be based on number of HP vectors.
- Between 32 and 64 HP vectors is sufficient for g_A, g_S, and g_T.
- It pays to go to 256 HP vectors (and possibly beyond) for g_V.
- Future: extend the analysis of nucleon charges to form factors.
- This methodology has been used to compute disconnected diagrams of form factors at high momentum (Sergey Syritsyn’s talk).

Acknowledgements

- Sergey Syritsyn for helping optimize our contraction code and helpful discussions.
- Boram Yoon and Rajan Gupta for thoughtful discussions, Z2 noise runs, and 2-pt functions.
- NME collaboration for discussions and providing resources to compute the loops.
- Work supported by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE-AC05-06OR23100.
- This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.
- This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.