Multigrid low-mode averaging

R. Gruber, T. Harris, M. K. Marinkovic

ETH Zürich

December 9, 2024

Table of contents

1. Low-mode averaging (LMA) and its variants

2. Multigrid / Deflation

3. Multigrid low-mode averaging (MG LMA)

4. Where is the variance?

5. Cost

6. Chirality

7. Conclusion

Low-mode averaging (LMA) and its variants

- Idea [Neff et al. hep-lat/0106016, DeGrand and Schaefer hep-lat/0401011, Giusti et al. hep-lat/0402002]: Decompose the quark propagator into two pieces
  - One piece: affordable to do volume averaging
  - Remaining piece: cannot afford volume average exactly
- Determine N_c lowest modes of $D, Q = \gamma^5 D$, eo-preconditioned D, Q
- Write $S = D^{-1} =$ truncated spectral/singular sum + remainder

$$Q^{-1} = \underbrace{\sum_{i=1}^{N_c} \frac{1}{\lambda_i} \phi_i \phi_i^\dagger}_{Q^{-1}_{eigen}} + \underbrace{(1-P)Q^{-1}(1-P)^\dagger}_{Q^{-1}_{rest} = Q^{-1} - Q^{-1}_{eigen}}, \tag{1}$$

with

$$Q\phi_i = \lambda_i \phi_i, \qquad |\lambda_i| = \text{small}, \qquad P = \sum_{i=1}^{N_c} \phi_i \phi_i^\dagger.$$

Motivation - Low-mode averaging (LMA)

- Assume a set of N_c orthonormal low modes $\{\phi_c\}_{c=0}^{N_c-1}$ of $Q = \gamma^5 D$
- Restrictor $R : \mathcal{V}_0 \to \mathcal{V}_1$ as

$$R = \begin{pmatrix} - & \phi_0^\dagger & - \\ & \vdots & \\ - & \phi_{N_c-1}^\dagger & - \end{pmatrix}$$

- Prolongator $T : \mathcal{V}_1 \to \mathcal{V}_0$ as

$$T = R^\dagger = \begin{pmatrix} | & & | \\ \phi_0 & \dots & \phi_{N_c-1} \\ | & & | \end{pmatrix}$$

- Projector $P : \mathcal{V}_0 \to \mathcal{V}_0$ and identity $\hat{\mathbf{1}} : \mathcal{V}_1 \to \mathcal{V}_1$:

$$P = TR, \qquad \hat{\mathbf{1}} = RT$$

- Define the coarse operator $\hat{Q} : \mathcal{V}_1 \to \mathcal{V}_1$ as

$$\hat{Q} = RQT$$

- If ϕ_c are exact modes of $Q \implies$ diagonal $\hat{Q}$ and inverse

$$\hat{Q} = \text{diag}(\lambda_0, \ldots, \lambda_{N_c-1}) \qquad \hat{Q}^{-1} = \text{diag}(\lambda_0^{-1}, \ldots, \lambda_{N_c-1}^{-1})$$

- Decompose the quark propagator

$$Q^{-1} = T\hat{Q}^{-1}R + \left(Q^{-1} - T\hat{Q}^{-1}R\right) \tag{2}$$

$$= \sum_{i=0}^{N_c-1} \frac{1}{\lambda_i} \phi_i \phi_i^\dagger + (1-P)Q^{-1}(1-P^\dagger) \tag{3}$$

Motivation - Low-mode averaging (LMA)

- Define the coarse operator $\hat{Q} : \mathcal{V}_1 \to \mathcal{V}_1$ as

$$\hat{Q} = RQT$$

- If ϕ_c are exact modes of $Q \implies$ diagonal $\hat{Q}$ and inverse

$$\hat{Q} = \text{diag}(\lambda_0, \dots, \lambda_{N_c-1}) \qquad \hat{Q}^{-1} = \text{diag}(\lambda_0^{-1}, \dots, \lambda_{N_c-1}^{-1})$$

- Decompose the quark propagator

$$Q^{-1} = T\hat{Q}^{-1}R + \left(Q^{-1} - T\hat{Q}^{-1}R\right) \tag{2}$$

$$= \sum_{i=0}^{N_c-1} \frac{1}{\lambda_i} \phi_i \phi_i^\dagger + (1-P)Q^{-1}(1-P^\dagger) \tag{3}$$

eigen

- Define the coarse operator $\hat{Q} : \mathcal{V}_1 \to \mathcal{V}_1$ as

$$\hat{Q} = RQT$$

- If ϕ_c are exact modes of $Q \implies$ diagonal $\hat{Q}$ and inverse

$$\hat{Q} = \text{diag}(\lambda_0, \dots, \lambda_{N_c-1}) \qquad \hat{Q}^{-1} = \text{diag}(\lambda_0^{-1}, \dots, \lambda_{N_c-1}^{-1})$$

- Decompose the quark propagator

$$Q^{-1} = T\hat{Q}^{-1}R + \left(Q^{-1} - T\hat{Q}^{-1}R\right) \tag{2}$$

$$= \underbrace{\sum_{i=0}^{N_c-1} \frac{1}{\lambda_i} \phi_i \phi_i^\dagger}_{\text{eigen}} + \underbrace{(1-P)Q^{-1}(1-P^\dagger)}_{\text{rest}} \tag{3}$$

- Two-point connected light-quark vector correlator
- In the time-momentum representation [Bernecker and Meyer 1107.4388] (local-local), $S = D^{-1}$

$$G(t) = \frac{1}{|\Omega_0|} \sum_{y \in \Omega_0} \sum_{\vec{x} \in \Sigma_0} C(y_0 + t, \vec{x}|y), \quad (4)$$

$$C(x|y) = \mathrm{tr}\left[\Gamma_1 S(x|y) \Gamma_2 S(x|y)^\dagger\right], \quad (5)$$

- Stochastic sources: introduce extra noise
- Point sources: cost $= O(V)$
- Ideally, but unrealistic: full lattice volume average

- Plug in decomposition of propagator

$$G(t) = G_{ee}(t) + \underbrace{G_{re}(t) + G_{er}(t)}_{G_\times(t)} + G_{rr}(t) \quad (6)$$

- Get 3-4 terms: eigen-eigen, cross (rest-eigen + eigen-rest), rest-rest

$G_{ee}(t)$: exact, volume-averaged, at its gauge noise
$G_{rr}(t)$: little variance contribution $\rightarrow$ few sources
$G_\times(t)$: typically significant contribution to total noise $\gg$ gauge noise

- Plug in decomposition of propagator

$$G(t) = G_{ee}(t) + \underbrace{G_{re}(t) + G_{er}(t)}_{G_\times(t)} + G_{rr}(t) \tag{6}$$

- Get 3-4 terms: eigen-eigen, cross (rest-eigen + eigen-rest), rest-rest

$G_{ee}(t)$: exact, volume-averaged, at its gauge noise ✔

$G_{rr}(t)$: little variance contribution $\rightarrow$ few sources

$G_\times(t)$: typically significant contribution to total noise $\gg$ gauge noise

- Plug in decomposition of propagator

$$G(t) = G_{ee}(t) + \underbrace{G_{re}(t) + G_{er}(t)}_{G_\times(t)} + G_{rr}(t) \tag{6}$$

- Get 3-4 terms: eigen-eigen, cross (rest-eigen + eigen-rest), rest-rest

$G_{ee}(t)$: exact, volume-averaged, at its gauge noise ✓
$G_{rr}(t)$: little variance contribution $\rightarrow$ few sources ✓
$G_\times(t)$: typically significant contribution to total noise $\gg$ gauge noise

- Plug in decomposition of propagator

$$G(t) = G_{ee}(t) + \underbrace{G_{re}(t) + G_{er}(t)}_{G_{\times}(t)} + G_{rr}(t) \tag{6}$$

- Get 3-4 terms: eigen-eigen, cross (rest-eigen + eigen-rest), rest-rest

$G_{ee}(t)$: exact, volume-averaged, at its gauge noise ✓

$G_{rr}(t)$: little variance contribution $\rightarrow$ few sources ✓

$G_{\times}(t)$: typically significant contribution to total noise $\gg$ gauge noise 💩

The limits of LMA

1. V^2-problem: number of required low modes scales $O(V)$ with the volume, on state-of-the-art lattices at the physical point
   - 1000-6000 eigenmodes [Kuberski 2312.13753, Blum et al. 1801.07224, Borsanyi et al. 1711.04980, Blum et al. 1512.09054]
   - Memory requirements
   - Storage and I/O requirements (people don't store them anymore!)

The limits of LMA

1. V^2-problem: number of required low modes scales $O(V)$ with the volume, on state-of-the-art lattices at the physical point
   - 1000-6000 eigenmodes [Kuberski 2312.13753, Blum et al. 1801.07224, Borsanyi et al. 1711.04980, Blum et al. 1512.09054]
   - Memory requirements
   - Storage and I/O requirements (people don't store them anymore!)

Note

Number of eigenmodes are limited by memory / resources.

1. V^2-problem: number of required low modes scales $O(V)$ with the volume, on state-of-the-art lattices at the physical point
   - 1000-6000 eigenmodes [Kuberski 2312.13753, Blum et al. 1801.07224, Borsanyi et al. 1711.04980, Blum et al. 1512.09054]
   - Memory requirements
   - Storage and I/O requirements (people don't store them anymore!)

Note

Number of eigenmodes are limited by memory / resources.

2. Cross-term-problem: Cross term has lots of noise $\rightarrow$ expensive!
   - Method 1: all-mode averaging, AMA, [Blum et al. 1208.4349, Shintani et al. 1402.0244, Blum et al. 1801.07224, Blum et al. 1512.09054]
   - Method 2: truncated solver method (TSM) + bias correction [Kuberski 2312.13753, Borsanyi et al. 1711.04980]
   - Method 3: stochastically evaluate the rest-rest + rest-eigen piece [Lynch and DeTar (2023), Bazavov et al. 2301.08274]
   - Other methods ...

Multigrid / Deflation

- Low modes of Dirac operator are locally coherent [Lüscher 0706.2298]

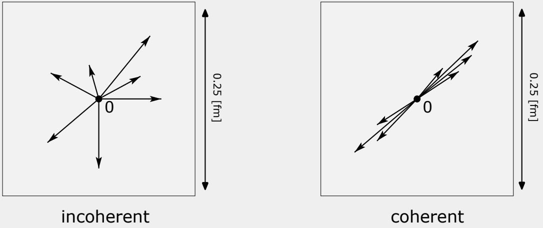


Figure: (Local) coherence of low modes (taken from Ref. [Lüscher 1002.4232]).

- Low modes of Dirac operator are locally coherent [Lüscher 0706.2298]

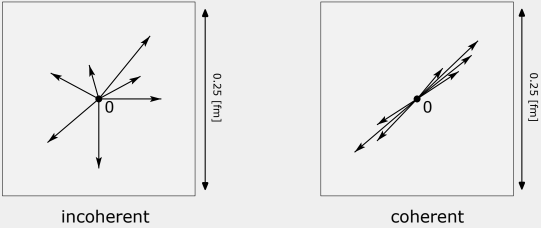


Figure: (Local) coherence of low modes (taken from Ref. [Lüscher 1002.4232]).

Conclusion

Using domain decomposition / coarsening on 10-100 low modes is enough to span the $O(V)$ low-mode space!

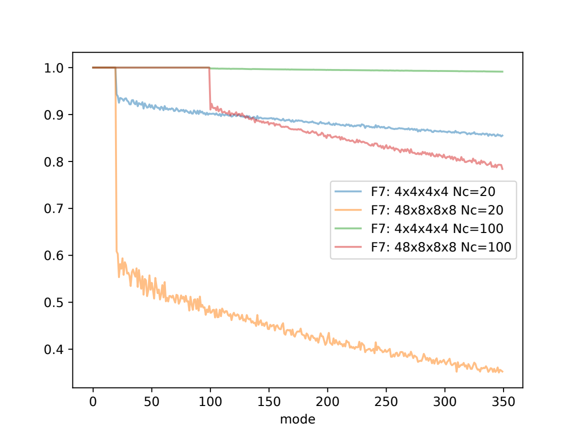
- Assume we have a few lowest modes ϕ_c of D - say $N_c = 20$
- Project to a block decomposition

$$\phi_c^{B_{\hat{y}}}(x) = \begin{cases} \phi_c(x) & \text{if } x \in B_y \\ 0 & \text{else} \end{cases} \quad (7)$$

where the block $B_{\hat{y}}$ is indexed by a new coarse coordinate $\hat{y} \in \hat{\Lambda}$

- Reorthonormalize
- Gives a basis $\mathcal{B}$ of size $N_c|\hat{\Lambda}|$
- Local coherence $\implies$ span of $N_c|\hat{\Lambda}|$ fields has a large overlap with the space of $N \gg N_c$ low modes!
- Lüscher: local coherence [Lüscher 0706.2298], Multigrid: weak approximation property [Brezina et al. (2005), Babich et al. 1005.3043]
- Used in Krylov solvers with deflation/multigrid preconditioning

LOCAL COHERENCE IN MORE DETAIL II



- x-axis: mode number $c = 0, \ldots 349$
- y-axis: $\|P\phi_c\| / \|\phi_c\|$ where $P = \sum_{\phi \in \mathcal{B}} \phi \phi^\dagger$.

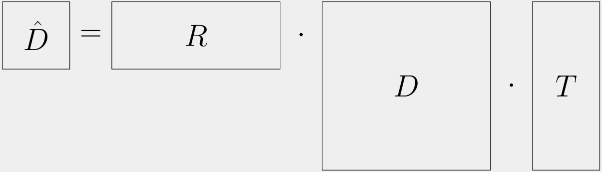
Multigrid / Deflation

- Setup subspace(s) as in the previous slide (domain-decomposed low modes)
- Define restrictors R and prolongators $T = R^\dagger$ from/to these subspaces

$$R\colon\ \psi \mapsto \theta, \quad \theta(i) = \langle \phi_i | \psi \rangle\,, \tag{8}$$

$$T\colon\ \theta \mapsto \psi = \sum_i \theta(i)\phi_i, \tag{9}$$

- Define the coarse-grid Dirac operator(s) as $\hat{D} = RDT$



- Connection to solver: sloppy $\hat{D}^{-1}$ as preconditioner for the Dirac equation

$$LD\psi = L\eta \quad \text{with} \quad L = T\hat{D}^{-1}R \quad \text{(left preconditioning)}$$

- Setup subspace(s) as in the previous slide (domain-decomposed low modes)
- Define restrictors R and prolongators $T = R^\dagger$ from/to these subspaces

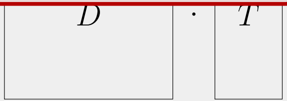
$$R\colon\ \psi \mapsto \theta\,, \quad \theta(i) = \langle \phi_i | \psi \rangle\,, \tag{8}$$

(9)

- De

Main message

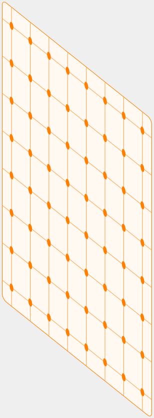
Coarse-grid operator $\hat{D}$ has smaller dimension, smaller condition number and is thus **cheaper to invert**!



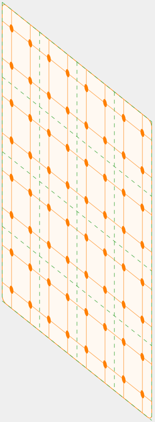
- Connection to solver: sloppy $\hat{D}^{-1}$ as preconditioner for the Dirac equation

$$LD\psi = L\eta \quad \text{with} \quad L = T\hat{D}^{-1}R \quad \text{(left preconditioning)}$$

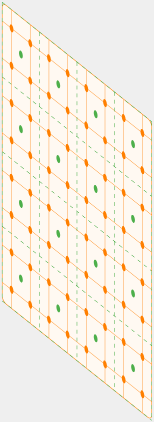
Multigrid / Deflation

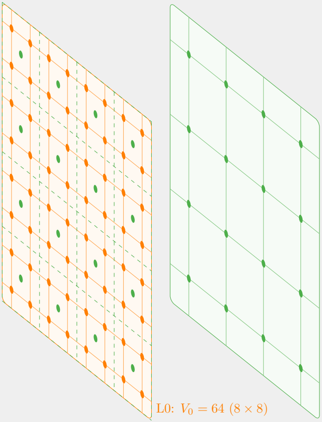


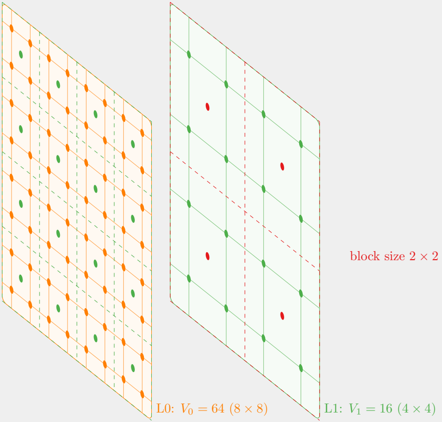
L0: $V_0 = 64\ (8 \times 8)$


block size 2×2
L0: $V_0 = 64\ (8 \times 8)$

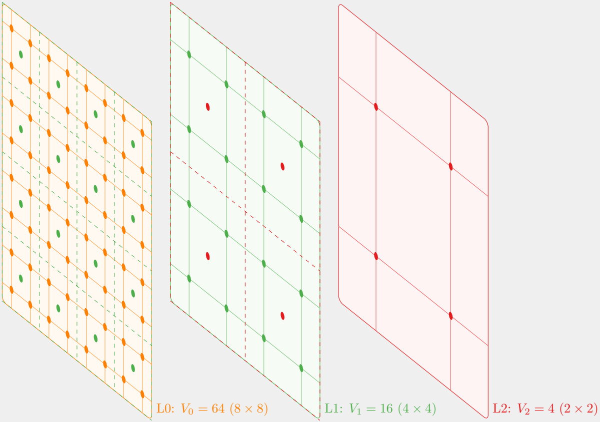
Multigrid / Deflation

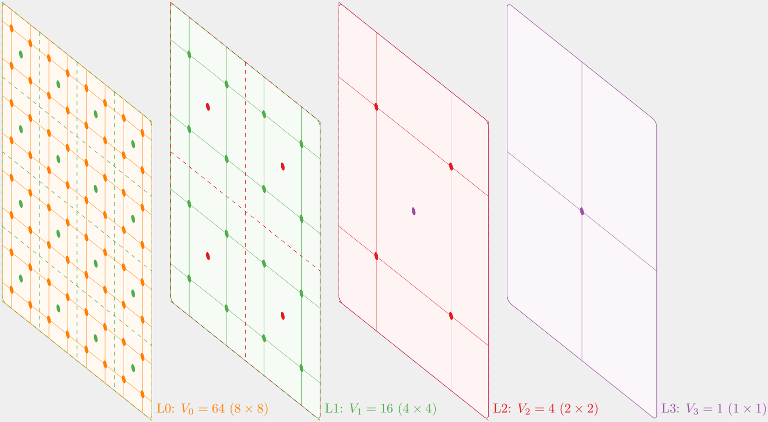


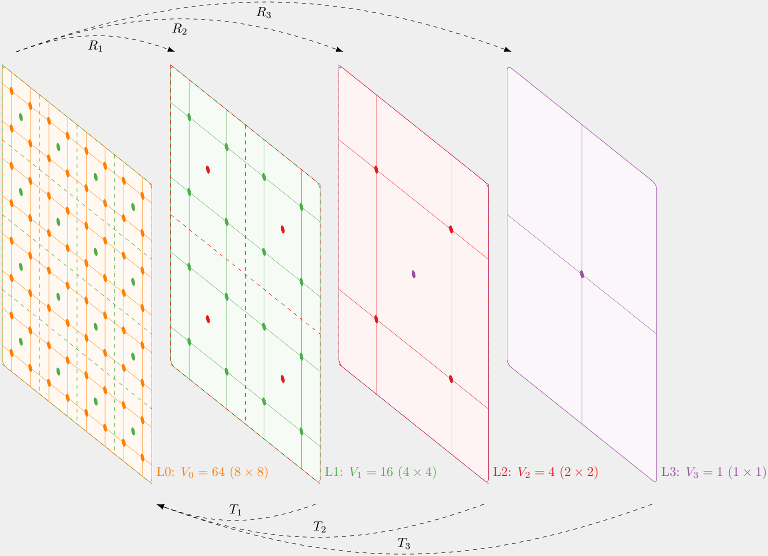

L0: $V_0 = 64\ (8 \times 8)$
L1: $V_1 = 16\ (4 \times 4)$


block size 2×2
L0: $V_0 = 64$ (8×8)
L1: $V_1 = 16$ (4×4)

Multigrid / Deflation




L0: $V_0 = 64$ (8×8)
L1: $V_1 = 16$ (4×4)
L2: $V_2 = 4$ (2×2)
L3: $V_3 = 1$ (1×1)


R_1
R_2
R_3
L0: $V_0 = 64\ (8 \times 8)$
L1: $V_1 = 16\ (4 \times 4)$
L2: $V_2 = 4\ (2 \times 2)$
L3: $V_3 = 1\ (1 \times 1)$
T_1
T_2
T_3

Multigrid low-mode averaging (MG LMA)

- Decompose the quark propagator $S = D^{-1}$ using the coarsenings

$$S = \sum_{i=0}^{N-1} S_i = \underbrace{S - K_1}_{=S_0} + \underbrace{K_1 - K_2}_{=S_1} + \underbrace{K_2 - K_3}_{=S_2} + \cdots + \underbrace{K_{N-1}}_{S_{N-1}}, \qquad (10)$$

$K_i = T_i(\hat{D}_i)^{-1} R_i, \quad S_i =$ deflated propagator on level i.

- Each level is defined by a different domain decomp./coarse grid

- Plug into the correlator
- For the correlator we find a matrix of correlators:

$$C_{ij}(x,y) = \mathrm{tr}\left[\Gamma_1 S_i(x|y)\Gamma_2 S_j(y|x)\right], \quad C = \sum_{i,j} C_{ij}. \tag{11}$$

- $i,j = 0,\ldots,N-1$ correspond to MG-level (with L0 the fine grid)
- Grouping the N^2 correlators into levels (see figure on next slide) gives us

$$G(t) = \sum_{k=0}^{N-1} G_{Lk}(t). \tag{12}$$

Grouping of correlators

C_{00}	C_{10}	C_{20}	C_{30}
C_{01}	C_{11}	C_{21}	C_{31}
C_{02}	C_{12}	C_{22}	C_{32}
C_{03}	C_{13}	C_{23}	C_{33}

$$G = G_{L0} + G_{L1} + G_{L2} + G_{L3}$$

C_{rr}	C_{re}
C_{er}	C_{ee}

$$G = \underbrace{G_{rr} + G_{\times}}_{G_{L0}} + \underbrace{G_{ee}}_{G_{L1}}$$

- Each level-contribution can be evaluated with a different strategy, i.e. number and type of sources!

Main message

Evaluating G_{Lk} requires inversions of the Dirac operator $\hat{D}_k$ on level k and coarser, but not finer levels!

- If dimension small enough $\dim_{\mathbb{C}}(\mathcal{V}_{N-1}) \sim 10^4$

$$G_{\mathrm{L}(N-1)}(t) = \frac{1}{TL^3} \sum_{y_0=0}^{L_0-a} \mathrm{tr}\left\{\widehat{\Gamma_1}(y_0+t)\hat{Q}^{-1}\widehat{\Gamma_2}(y_0)\hat{Q}^{-1}\right\}, \tag{13}$$

with

$$\widehat{\Gamma_{1,2}}(x_0)_{ij} = \sum_{\vec{x}} \phi_i^\dagger(x)\Gamma_{1,2}\phi_j(x) \tag{14}$$

- Full lattice volume averaged correlator:

$$G(t) = \frac{1}{TL^2} \sum_{y\in\Lambda}\sum_{\vec{x}\in\Lambda} \mathrm{tr}\left[\Gamma_1 S(y_0+t,\vec{x}|y)\Gamma_2 S(y_0+t,\vec{x}|y)^\dagger\right] \tag{15}$$

Where is the variance?

Name	Size $[T \times L^3]$	L [fm]	m_π L
E7[1]	64×32^3	2.1 fm	3.2
F7[2]	96×48^3	3.2 fm	4.8
G7[1]	128×64^3	4.2 fm	6.4
H7[1]	192×96^3	6.3 fm	9.6

Table: All ensembles have a pion mass $m_\pi = 270$ MeV and a lattice spacing of $a = 0.0658$ fm with $N_f = 2$ $O(a)$-improved Wilson fermions.

[1]Generated by Tim Harris using openQCD 2.4.2 [Lüscher et al. (2012-2023)]
[2]CLS lattice from Ref. [CLS (2012-2023)]

Relative variances: G7

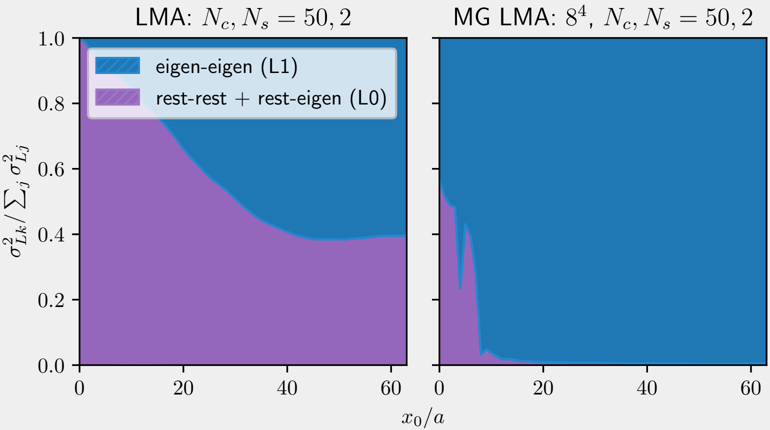


Figure: Relative variance for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term.

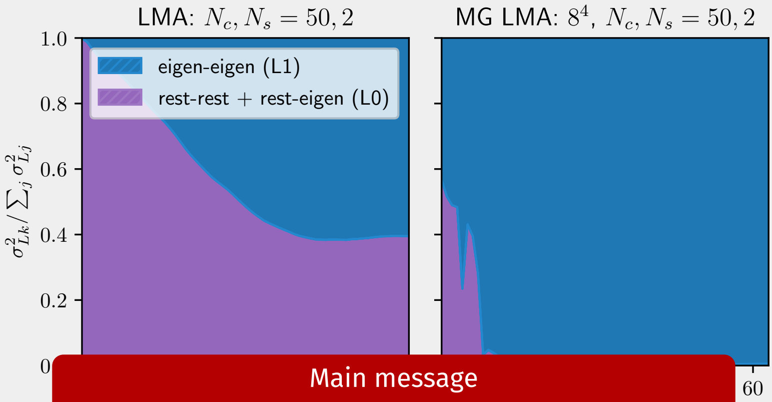


Figure: F
correlat

Main message

We observe a significant variance contribution from the cheap-to-evaluate L1-term w.r.t LMA.

Absolute variances: G7

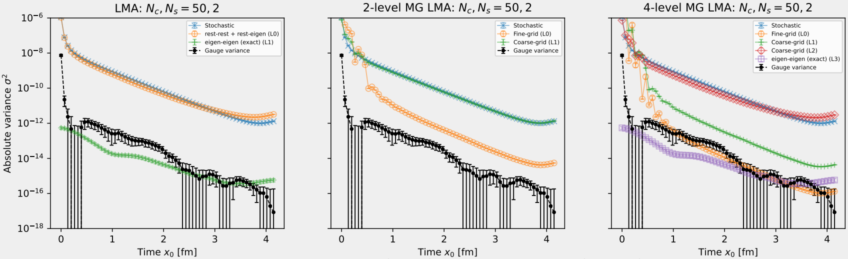


Figure: Absolute variances for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term. The black line is the gauge variance.

ABSOLUTE VARIANCES: G7

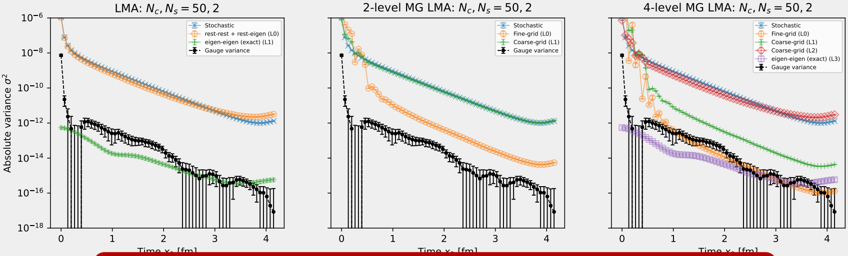


Figure: A… …r correlat… …e gauge variance…

Main message

We are able to push the remaining L0 noise down to the gauge noise using only a few stochastic sources.

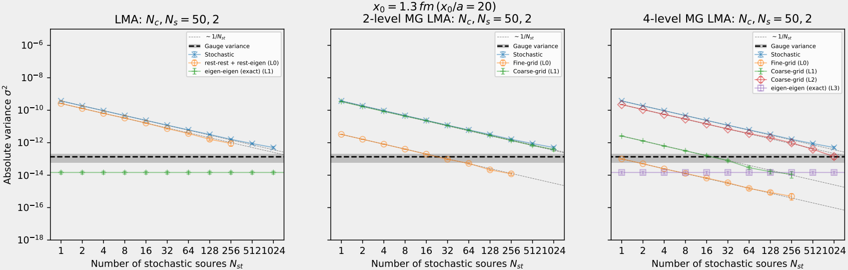


Figure: Absolute variances for LMA (left) and MG LMA (right) against number of stochastic sources N_{st}. The black line is the gauge variance.

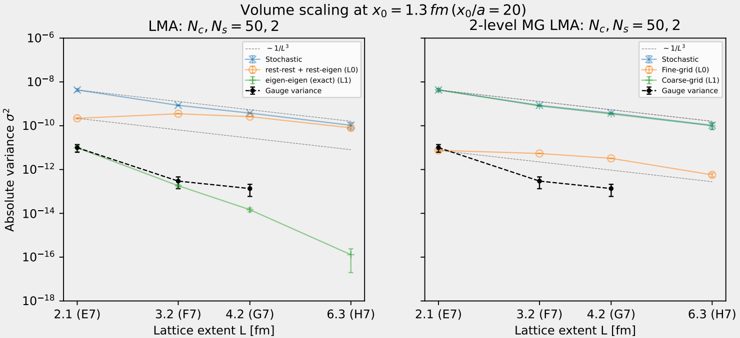


Figure: Absolute variances for LMA (left) and MG LMA (right) against the lattice extent L. The black line is the gauge variance.

Variance vs. volume

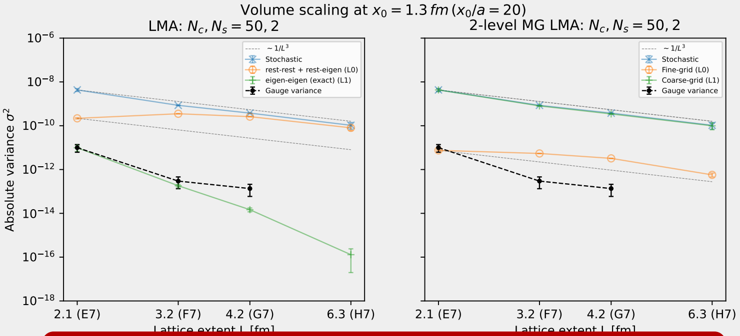


Figure: A … lattice extent L …

Main message

MG LMA with a constant number of low modes scales well with the volume.

Cost

Table: Cost breakdown to reach the gauge variance for G7 (4.2 fm).

Estimator	# modes	# sources	meas. cost[1]	model cost[1]
Stochastic	0	L0: 4096	16384	16384
LMA[2]	50	L0: 2048	8192	8192
2-lvl MG LMA[2]	50	L0: 16* L1: 2048***	557.8	80.7
4-lvl MG LMA[2]	50	L0: 1* L1: 16** L2: 1024***	466.7	14.4

My 💩 implementation:

$\star$	fine-grid	128×64^3	inv: 11.1 ± 0.4 sec	(iter: 46.53 ± 0.23)
$\star\star$	coarse-grid	32×16^3	inv: 37.3 ± 2.4 sec	(iter: 1417 ± 22)
$\star\star\star$	coarse-grid	16×8^3	inv: 0.667 ± 0.041 sec	(iter: 502.1 ± 5.8)

[1]Unit = fine-grid inversions.
[2]Cost of determination of low modes not included (or add 100 - 200 to the cost).

Chirality

Requirements

Requirement 1) Coarse operators should be better conditioned than fine ones:

$$\kappa(\hat{K}) \leq \kappa(K) \tag{16}$$

for $K = D, \gamma^5 D, D^\dagger, D\gamma^5, \ldots$

Requirement 2) Variance contribution of coarse levels should dominate

Fine-grid lattice

- Spacetime points: $x \in \Lambda$
- Colours: $N_c = 3$
- Spins: $N_s = 4$
  - 2 chiral d.o.f.
  - 2 non-chiral d.o.f.

Coarse-grid lattice

- Spacetime points: $\hat{x} \in \hat{\Lambda}$
- Colours: $N_c =$ number of low modes
- Spins: $N_s = 1$

- Spectral decomposition of Q

$$Q = \sum_i \lambda_i \phi_i \phi_i^\dagger \qquad \text{where} \qquad Q\phi_i = \lambda_i \phi_i \tag{17}$$

- Singular value decomposition of D

$$D = \sum_i |\lambda_i| \tilde{\phi}_i \phi_i^\dagger \qquad \text{where} \qquad \tilde{\phi}_i = \mathrm{sign}(\lambda_i)\gamma^5 \phi_i \tag{18}$$

$$(\tilde{Q}\tilde{\phi}_i = \lambda_i \tilde{\phi}_i \qquad \text{where} \qquad \tilde{Q} = D\gamma^5)$$

- Low modes of Q are linear combinations of low modes of D (not the case for low modes of $\tilde{Q}$)

$$\|D\phi_i\| \sim |\lambda_i| \qquad \left\|D\tilde{\phi}_i\right\| \gg |\lambda_i| \tag{19}$$

This suggest to use:

- right singular vectors of D for prolongation: T
- left singular vectors of D for restriction: $R = T^\dagger \gamma^5$

Thus the coarse operator is (Petrov-Galerkin approach; $R \neq T^\dagger$)

$$T^\dagger \gamma^5 D T \tag{20}$$

This suggest to use:

- right singular vectors of D for prolongation: T
- left singular vectors of D for restriction: $R = T^\dagger \gamma^5$

Thus the coarse operator is (Petrov-Galerkin approach; $R \neq T^\dagger$)

$$T^\dagger \gamma^5 DT \tag{20}$$

Notice

This is just $\hat{Q} = RQT$ in the Galerkin-approach ($R = T^\dagger$)

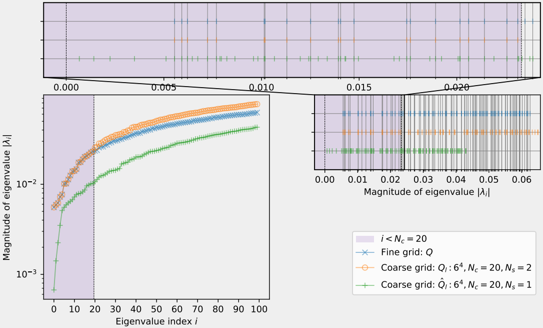


Figure: Lower left: lowest 100 eigenvalues (in magnitude) of the Hermitian Dirac operator $\gamma^5 D$ and $\widehat{\gamma^5 D}$. **Lower right** eigenvalues plotted on the x-axis with a gray vertical line at every fine grid eigenvalue (blue). **Upper panel**: zoom. $N_c = 20$.

Problem of coarsening Q

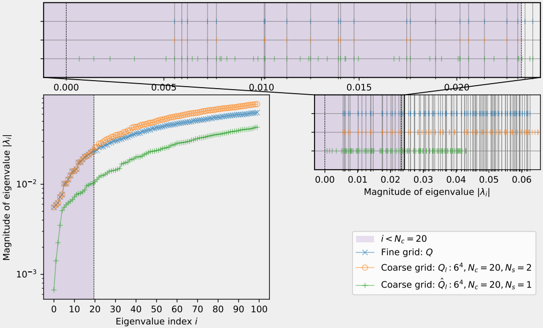


Figure: Lower left: lowest 100 eigenvalues (in magnitude) of the Hermitian Dirac operator $\gamma^5 D$ and $\widehat{\gamma^5 D}$. **Lower right** eigenvalues plotted on the x-axis with a gray vertical line at every fine grid eigenvalue (blue). **Upper panel**: zoom. $N_c = 20$.

- ✗ Requirement 1) $\kappa(\hat{K}) \leq \kappa(K)$, unless $N_s = 2$
- ✓ Requirement 2) Coarse level variance dominates

Problem of coarsening D

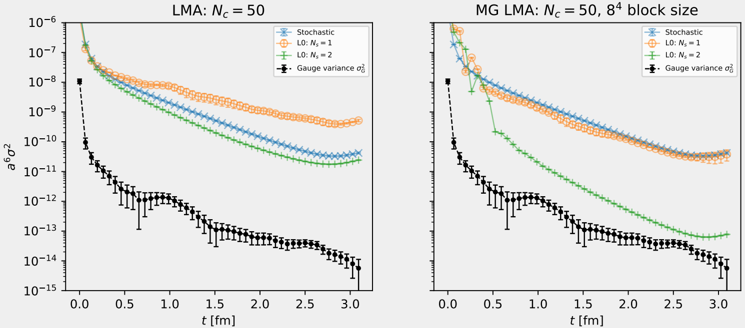


Figure: Variance contribution for LMA (left) and MG LMA (right) without chirality preservation (yellow) and with chirality preservation (green).

Problem of coarsening D

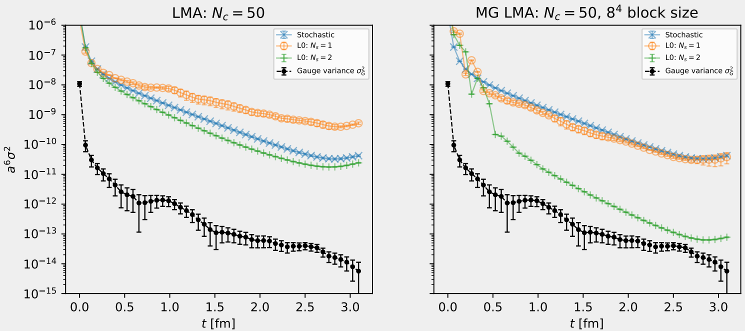


Figure: Variance contribution for LMA (left) and MG LMA (right) without chirality preservation (yellow) and with chirality preservation (green).

- ✓ Requirement 1) $\kappa(\hat{K}) \leq \kappa(K)$
- ✗ Requirement 2) Coarse level variance dominates, unless $N_s = 2$

- The Dirac operator is pseudo-Hermitian w.r.t. γ^5: $(\gamma^5 D)^\dagger = \gamma^5 D$
- We may retain this property on the coarse grid by imposing $[P, \gamma^5] = 0$:

$$(\widehat{\gamma^5}\hat{D})^\dagger = \widehat{\gamma^5}\hat{D} \tag{21}$$

$$\widehat{\gamma^5 D} = \widehat{\gamma^5}\hat{D} \tag{22}$$

where $P = TR$ and $\widehat{\gamma^5} = R\gamma^5 T$

- The Dirac operator is pseudo-Hermitian w.r.t. γ^5: $(\gamma^5 D)^\dagger = \gamma^5 D$
- We may retain this property on the coarse grid by imposing $[P, \gamma^5] = 0$:

$$(\widehat{\gamma^5}\hat{D})^\dagger = \widehat{\gamma^5}\hat{D} \tag{21}$$

$$\widehat{\gamma^5 D} = \widehat{\gamma^5}\hat{D} \tag{22}$$

where $P = TR$ and $\widehat{\gamma^5} = R\gamma^5 T$

Conclusion

How to coarsen? By keeping both fine-grid chirality indices explicit on the coarse subspace.

Explicit chiral d.o.f. can be implemented in different ways:

$$\{\phi_i\}_{i=0}^{N_c-1} \longmapsto \{\phi_i\}_{i=0}^{N_c-1} \cup \{\gamma_5 \phi_i\}_{i=0}^{N_c-1} \tag{23}$$

$$\{\phi_i\}_{i=0}^{N_c-1} \longmapsto \{P_- \phi_i\}_{i=0}^{N_c-1} \cup \{P_+ \phi_i\}_{i=0}^{N_c-1} \tag{24}$$

where $P_\pm = \frac{1}{2}(1 \pm \gamma_5)$ are the chiral projectors.

- Eq. (23) are left and right singular vectors of D
- Eq. (24): coarse $\widehat{\gamma^5}$ and fine γ^5 have the same structure

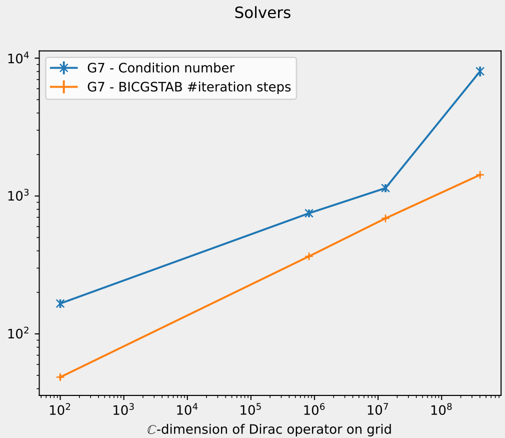


Figure: G7 (4.2 fm): $\mathbb{C}$-dimension of the Dirac operator (x-axis) vs. Number of BiCGSTAB iterations and condition number (y-axis). The rightmost operator is the fine-grid, the others ones are different coarse-grid Dirac operators.

Conclusion

If we impose explicit chiral d.o.f. (i.e. chiral doubling of the modes), requirements are met. No matter which operator we coarsen!

- ✔ Requirement 1) Coarse operators should be better conditioned than fine ones:

$$\kappa(\hat{K}) \leq \kappa(K) \tag{25}$$

  for $K = D, \gamma^5 D, D^\dagger, D\gamma^5, \ldots$

- ✔ Requirement 2) Variance contribution of coarse levels should dominate

Conclusion

Concluding summary

- Subspaces based on domain-decomposed / coarsened low modes
- Propagator decomposition $\rightarrow$ Correlator decomposition into-MG levels
- Method can be defined recursively
- Every level-contribution $\rightarrow$ separate statistics
- 50 low modes capture all the variance (independent of the lattice volume!)
- Fewer low modes & more variance contribution than LMA

- Subspaces based on domain-decomposed / coarsened low modes
- Propagator decomposition → Correlator decomposition into-MG levels
- Method can be defined recursively
- Every level-contribution → separate statistics
- 50 low modes capture all the variance (independent of the lattice volume!)
- Fewer low modes & more variance contribution than LMA

Key idea

Hierarchical evaluation: noisy part is cheaper to evaluate!

References I

[1] L. Giusti et al., **"Frequency-splitting estimators of single-propagator traces"**, Eur. Phys. J. C **79**, 586 (2019), arXiv:1903.10447 [hep-lat].

[2] F. Knechtli et al., **"Optimizing creation operators for charmonium spectroscopy on the lattice"**, Phys. Rev. D **106**, 034501 (2022), arXiv:2205.11564 [hep-lat].

[3] L. N. Bushnaq, **"Exploring efficient methods for precision QCD calculations on the lattice"**, PhD thesis (Trinity Coll., Dublin, 2023).

[4] M. A. Clark et al., **"Multi-Grid Lanczos"**, EPJ Web Conf. **175**, edited by M. Della Morte et al., 14023 (2018), arXiv:1710.06884 [hep-lat].

[5] H. Neff et al., **"On the low fermionic eigenmode dominance in QCD on the lattice"**, Phys. Rev. D **64**, 114509 (2001), arXiv:hep-lat/0106016.

[6] T. A. DeGrand and S. Schaefer, **"Improving meson two point functions in lattice QCD"**, Comput. Phys. Commun. **159**, 185–191 (2004), arXiv:hep-lat/0401011.

[7] L. GIUSTI ET AL., **"LOW-ENERGY COUPLINGS OF QCD FROM CURRENT CORRELATORS NEAR THE CHIRAL LIMIT"**, JHEP **04**, 013 (2004), arXiv:hep-lat/0402002.

[8] T. BLUM ET AL., **"NEW CLASS OF VARIANCE-REDUCTION TECHNIQUES USING LATTICE SYMMETRIES"**, Phys. Rev. D **88**, 094503 (2013), arXiv:1208.4349 [hep-lat].

[9] E. SHINTANI ET AL., **"COVARIANT APPROXIMATION AVERAGING"**, Phys. Rev. D **91**, 114511 (2015), arXiv:1402.0244 [hep-lat].

[10] S. KUBERSKI, **"MUON $g-2$: LATTICE CALCULATIONS OF THE HADRONIC VACUUM POLARIZATION"**, PoS **LATTICE2023**, 125 (2024), arXiv:2312.13753 [hep-lat].

[11] T. BLUM ET AL. (RBC, UKQCD), **"CALCULATION OF THE HADRONIC VACUUM POLARIZATION CONTRIBUTION TO THE MUON ANOMALOUS MAGNETIC MOMENT"**, Phys. Rev. Lett. **121**, 022003 (2018), arXiv:1801.07224 [hep-lat].

[12] S. BORSANYI ET AL. (BUDAPEST-MARSEILLE-WUPPERTAL), **"HADRONIC VACUUM POLARIZATION CONTRIBUTION TO THE ANOMALOUS MAGNETIC MOMENTS OF LEPTONS FROM FIRST PRINCIPLES"**, Phys. Rev. Lett. **121**, 022002 (2018), arXiv:1711.04980 [hep-lat].

[13] T. BLUM ET AL., **"CALCULATION OF THE HADRONIC VACUUM POLARIZATION DISCONNECTED CONTRIBUTION TO THE MUON ANOMALOUS MAGNETIC MOMENT"**, Phys. Rev. Lett. **116**, 232002 (2016), arXiv:1512.09054 [hep-lat].

[14] M. LÜSCHER, **"LOCAL COHERENCE AND DEFLATION OF THE LOW QUARK MODES IN LATTICE QCD"**, JHEP **07**, 081 (2007), arXiv:0706.2298 [hep-lat].

[15] M. BREZINA ET AL., **"ADAPTIVE SMOOTHED AGGREGATION (α SA) MULTIGRID"**, SIAM review **47**, 317–346 (2005).

[16] R. BABICH ET AL., **"ADAPTIVE MULTIGRID ALGORITHM FOR THE LATTICE WILSON-DIRAC OPERATOR"**, Phys. Rev. Lett. **105**, 201602 (2010), arXiv:1005.3043 [hep-lat].

[17] D. BERNECKER AND H. B. MEYER, **"VECTOR CORRELATORS IN LATTICE QCD: METHODS AND APPLICATIONS"**, Eur. Phys. J. A **47**, 148 (2011), arXiv:1107.4388 [hep-lat].

[18] CLS, ***Coordinated lattice simulations***, (2012-2023) https://wiki-zeuthen.desy.de/CLS/.

[19] M. Lüscher, **“Computational Strategies in Lattice QCD”**, in Les Houches Summer School: Session 93: Modern perspectives in lattice QCD: Quantum field theory and high performance computing (Feb. 2010), pp. 331–399, arXiv:1002.4232 [hep-lat].

[20] T. Whyte et al., **“Optimizing shift selection in multilevel Monte Carlo for disconnected diagrams in lattice QCD”**, Comput. Phys. Commun. **294**, 108928 (2024), arXiv:2212.04430 [hep-lat].

[21] M. Lüscher et al., ***OpenQCD, simulation programs for lattice QCD***, (2012-2023) https://luscher.web.cern.ch/luscher/openQCD/.

[22] M. Lynch and C. DeTar, **“Contrasting low-mode noise reduction techniques for light HISQ meson propagators”**, in The 39th international symposium on lattice field theory, (2023), p. 252.

References V

[23] A. Bazavov et al. (Fermilab Lattice, HPQCD,, MILC), **"Light-quark connected intermediate-window contributions to the muon g-2 hadronic vacuum polarization from lattice QCD"**, Phys. Rev. D **107**, 114514 (2023), arXiv:2301.08274 [hep-lat].

Backup slides: Table of contents

8 Gauge variance estimator
9 LMA - V^2-problem
10 LMA - Cross-term-problem
11 MG LMA - Problems solved
12 MG LMA vs. LMA with more low modes
13 LMA as a special case of MG LMA
14 Frequency splitting
15 RBC
16 Distillation
17 Scaling with pion mass
18 Scaling with number of low modes
19 Scaling with block size
20 Performance model with multiple RHS
21 Chirality on the subspace
22 Detailed setups
23 Variance contribution - All ensembles
24 Absolute variance - All ensembles
25 Variance vs. sources - All ensembles
26 Cost - All ensembles
27 Optimizations
28 Comparison - LMA vs. MG LMA
29 Multigrid Multilevel Monte Carlo (MGMLMC)
30 Multilevel Monte Carlo

Gauge variance estimator

Backup slide: Gauge variance estimator I

We define the gauge variance as the minimum variance (N_{st} = number of stochastic sources)

$$\sigma^2_{\text{vol}} = \lim_{N_{st}\to\infty} \sigma^2_{\text{appx}} \tag{26}$$

where

$$\sigma^2_{\text{appx}} = \langle G^2_{\text{appx}} \rangle - \langle G_{\text{appx}} \rangle^2, \tag{27}$$

$$G_{\text{appx}} = \frac{1}{N_{st}} \sum_{i=1}^{N_{st}} G_i, \tag{28}$$

$$G_i = -\frac{a^6}{L^6} \sum_{\vec{x}',\vec{x},\vec{y}} \eta_i^\dagger(x') S(x',x) \Gamma S(y,x) \Gamma \eta_i(x) \tag{29}$$

Backup slide: Gauge variance estimator II

Thus our estimator is

$$\sigma^2_{\text{vol}} \approx \frac{1}{L_0^2}\frac{1}{N_{st}}\sum_{\{x_0\}}\Bigg[\frac{1}{N_{st}-1}\sum_{i\neq j}\Big\{\langle G_i(x_0,y_0)G_j(x_0,y_0)\rangle_U - \langle G_i(x_0,y_0)\rangle_U\langle G_j(x_0,y_0)\rangle_u\Big\} + \sum_{\{x_0'\}}(1-\delta_{x_0,x_0'})\frac{1}{N_{st}}\sum_{i,j}\Big\{\langle G_i(x_0,y_0)G_j(x_0',y_0')\rangle_U - \langle G_i(x_0,y_0)\rangle_U\langle G_j(x_0',y_0')\rangle_U\Big\}\Bigg].$$

Backup slide: Gauge variance estimation

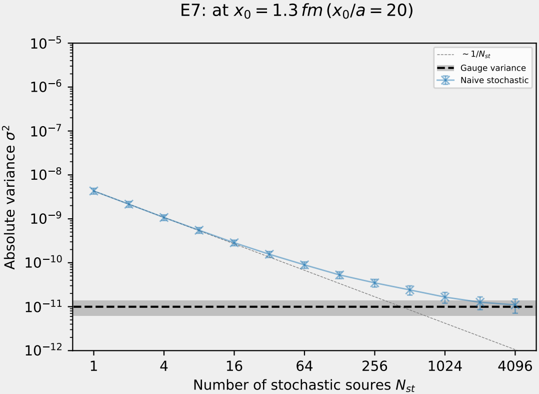
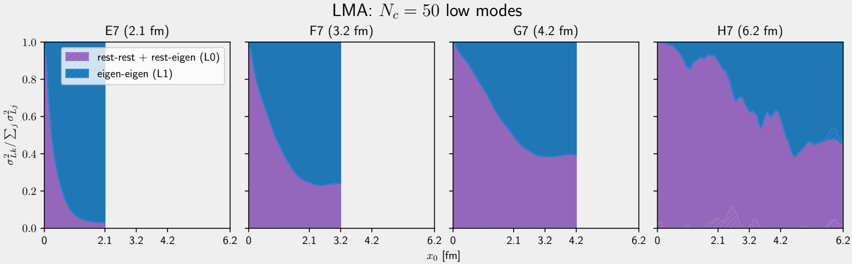


Figure: Absolute variance vs. number of stochastic noise sources. The gauge variance is reached.

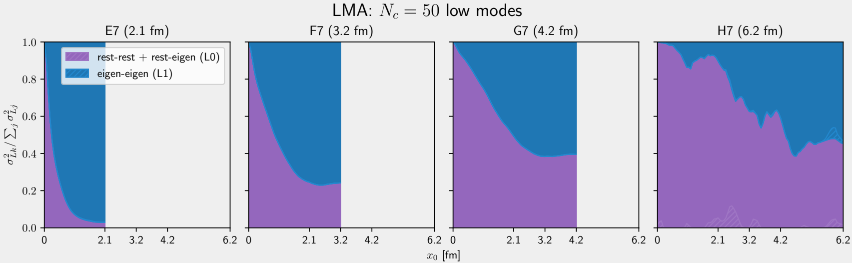
LMA - V^2-PROBLEM

Backup slide: V^2-problem of LMA



- Low mode contribution (L1) diminishes with larger volume

Backup slide: V^2-problem of LMA



- Low mode contribution (L1) diminishes with larger volume

V^2-problem of LMA

Need more low modes with larger volume.

LMA - Cross-term-problem

Backup slide: the cross-term problem in more detail

2. Cross-term-problem: Cross term has lots of noise contribution.
   - Method 1: all-mode averaging, AMA, [Blum et al. 1208.4349, Shintani et al. 1402.0244, Blum et al. 1801.07224, Blum et al. 1512.09054]

$$Q^{-1} = \underbrace{\sum_{i=1}^{N_c} \frac{1}{\lambda_i} \xi_i \xi_i^\dagger + P_n(Q)P}_{S_{AMA}} + \underbrace{Q^{-1} - S_{AMA}}_{S_{rest}}, \quad P_n = \text{TSM-poly. of deg. n.}$$

   - Method 2: truncated solver method (TSM) + bias correction [Kuberski 2312.13753, Borsanyi et al. 1711.04980]
     - Very similar to AMA
     - Needs 1 inversion per mode per gamma-matrix:

$$G_\times(x,y) = \frac{1}{|\Lambda_0|} \sum_{i=1}^{N_c} \frac{1}{\lambda_i} \sum_{\vec{x}} \left\langle \chi_j^{\Gamma_2}(y_0+t,\vec{x}) \middle| \Gamma_1 \gamma^5 \xi_i(y_0+t,\vec{x}) \right\rangle$$

$$\chi_j^\Gamma = D^{-1} \Gamma \xi_j$$

     - Needs 1 inversion for every mode for every gamma-matrix!
   - Method 3: stochastically evaluate the rest-eigen piece
   - ...

Backup slide: the cross-term problem in more detail

2. Cross-term-problem: Cross term has lots of noise contribution.
   - Method 1: all-mode averaging, AMA, [Blum et al. 1208.4349, Shintani et al. 1402.0244, Blum et al. 1801.07224, Blum et al. 1512.09054]

$$Q^{-1} = \sum^{N_c} \frac{1}{\lambda_i} \xi_i \xi_i^\dagger + \boxed{P_n(Q)P} + \underbrace{Q^{-1} - S_{AMA}}, \quad P_n = \text{TSM-poly. of deg. n.}$$

Expensive cross-term treatment is hidden in here.

   - Method 2: truncated solver method (TSM) + bias correction [Kuberski 2312.13753, Borsanyi et al. 1711.04980]
     - Very similar to AMA
     - Needs 1 inversion per mode per gamma-matrix:

$$G_\times(x,y) = \frac{1}{|\Lambda_0|} \sum_{i=1}^{N_c} \frac{1}{\lambda_i} \sum_{\vec{x}} \left\langle \chi_j^{\Gamma_2}(y_0+t,\vec{x}) \middle| \Gamma_1 \gamma^5 \xi_i(y_0+t,\vec{x}) \right\rangle$$

$$\chi_j^\Gamma = D^{-1}\Gamma\xi_j$$

     - Needs 1 inversion for every mode for every gamma-matrix!
   - Method 3: stochastically evaluate the rest-eigen piece
   - ...

Backup slide: the cross-term problem in more detail

2. Cross-term-problem: Cross term has lots of noise contribution.
   - Method 1: all-mode averaging, AMA, [Blum et al. 1208.4349, Shintani et al. 1402.0244, Blum et al. 1801.07224, Blum et al. 1512.09054]

$$Q^{-1} = \sum^{N_c} \frac{1}{\lambda_i} \xi_i \xi_i^\dagger + \boxed{P_n(Q)P} + \underbrace{Q^{-1} - S_{AMA}}, \quad P_n = \text{TSM-poly. of deg. n.}$$

Expensive cross-term treatment is hidden in here.

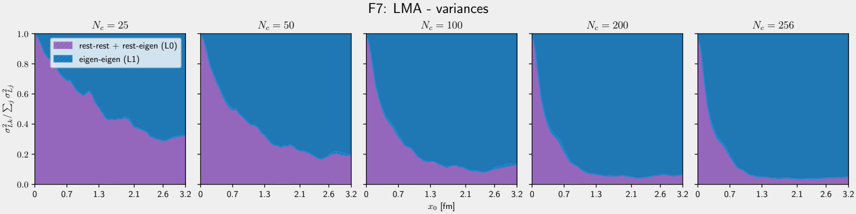
   - Method 2: truncated solver method (TSM) + bias correction [Kuberski 2312.13753, Borsanyi et al. 1711.04980]
     - Very similar to AMA
     - Needs 1 inversion per mode per gamma-matrix:

$$G_\times(x,y) = \frac{1}{|\Lambda_0|} \sum_{i=1}^{N_c} \frac{1}{\lambda_i} \sum_{\vec{x}} \left\langle \chi_j^{\Gamma_2}(y_0+t,\vec{x}) \middle| \Gamma_1 \gamma^5 \xi_i(y_0+t,\vec{x}) \right\rangle$$

$$\boxed{\chi_j^\Gamma = D^{-1}\Gamma\xi_j}$$ many inversions!

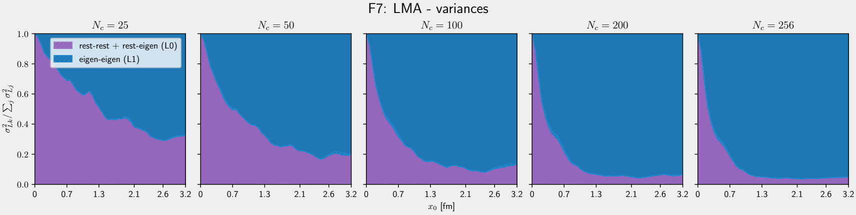
     - Needs 1 inversion for every mode for every gamma-matrix!
   - Method 3: stochastically evaluate the rest-eigen piece
   - ...

Backup slide: Cross-term-problem of LMA



- Increasing number of low modes doesn't push L0-noise contribution to gauge noise (dotted line)

Backup slide: Cross-term-problem of LMA



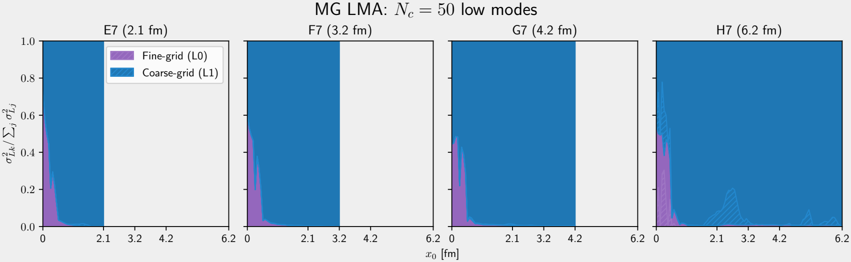
- Increasing number of low modes doesn't push LO-noise contribution to gauge noise (dotted line)

Cross-term-problem of LMA

Reminiscent variance of LO-term (rest-rest + rest-eigen) doesn't vanish.

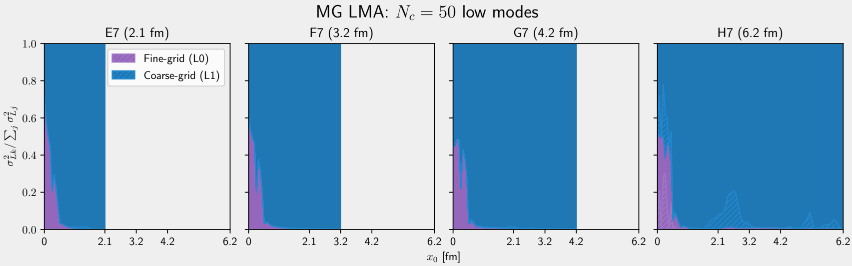
MG LMA - Problems solved

Backup slide: Problems solved using MG LMA



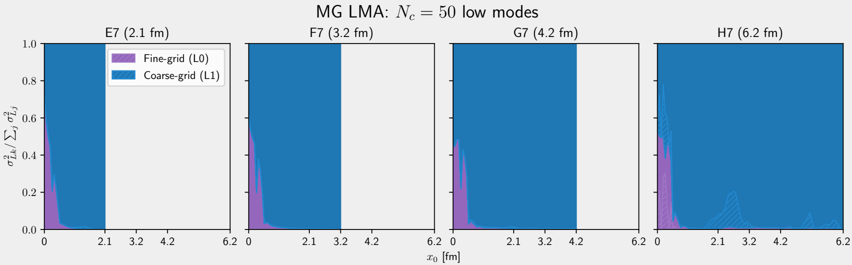
- L1-grid: generated with $B = 8^4$ block size

Backup slide: Problems solved using MG LMA



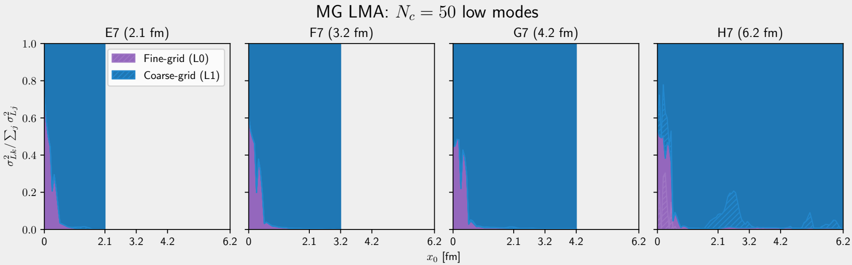
- L1-grid: generated with $B = 8^4$ block size
  $\Longrightarrow$ Coarse operator dimension increases with volume (same ratio)

Backup slide: Problems solved using MG LMA



- L1-grid: generated with $B = 8^4$ block size
  $\implies$ Coarse operator dimension increases with volume (same ratio)
- Constant number of low modes even when increasing the volume

BACKUP SLIDE: PROBLEMS SOLVED USING MG LMA

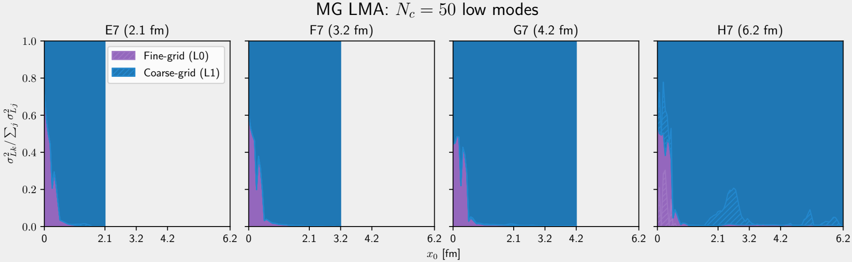


- L1-grid: generated with $B = 8^4$ block size

  $\implies$ Coarse operator dimension increases with volume (same ratio)
- Constant number of low modes even when increasing the volume

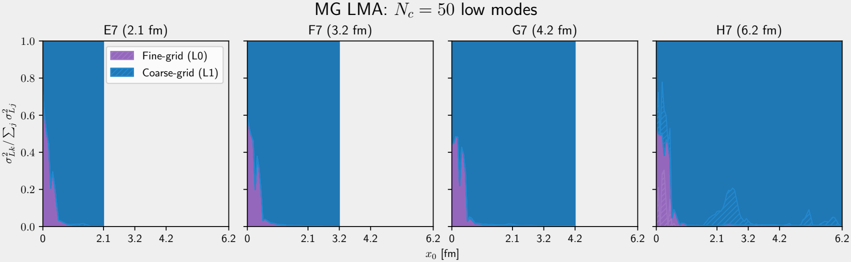
  $\implies$ V^2-problem solved!

Backup slide: Problems solved using MG LMA



- L1-grid: generated with $B = 8^4$ block size
  $\implies$ Coarse operator dimension increases with volume (same ratio)
- Constant number of low modes even when increasing the volume
  $\implies$ V^2-problem solved!
- L0-noise is negligible

Backup slide: Problems solved using MG LMA



- L1-grid: generated with $B = 8^4$ block size

  $\implies$ Coarse operator dimension increases with volume (same ratio)
- Constant number of low modes even when increasing the volume

  $\implies$ V^2-problem solved!
- L0-noise is negligible

  $\implies$ Cross-term-problem solved!

MG LMA vs. LMA with more low modes

BACKUP SLIDE: MG LMA VS. LMA WITH MORE LOW MODES

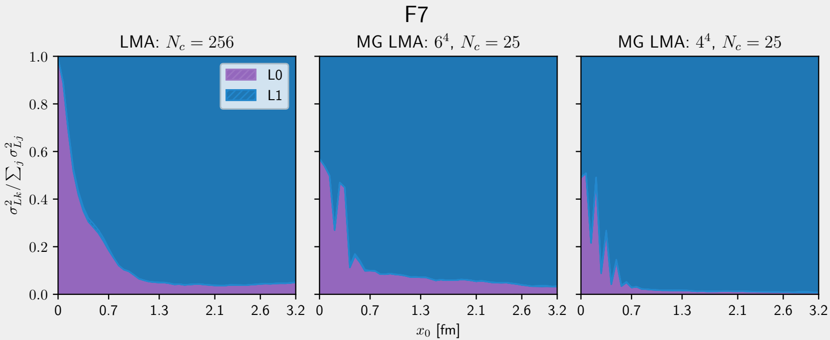


Figure: LMA (left) with $N_c = 256$ low modes vs. MG LMA (center and right) with $N_c = 25$ low modes. The variance contributions are comparable.

LMA as a special case of MG LMA

Backup slide: L1-propagator vs. eigen-eigen piece

- LMA: eigen-eigen propagator

$$S_{LMA} = \sum_{i=0}^{N_c-1} \frac{1}{\lambda_i} \xi_i \xi_i^\dagger \gamma^5 \tag{30}$$

- MG LMA: L1-propagator (propagator restricted to coarse grid)

$$S_1 = T_1 \hat{D}^{-1} R_1 = \sum_{i,j=0}^{N_b N_c N_s - 1} (\hat{D}^{-1})_{ij} \phi_i \phi_j^\dagger \tag{31}$$

$\Longrightarrow$ LMA is a special case of MG LMA

Frequency splitting

Backup slide: Frequency splitting

- Decomposition of propagator [Giusti et al. 1903.10447]

$$S = M_{2n,m} + D_m^{-1} H_m^{2n}, \tag{32}$$

$$M_{2n,m} = (D_{ee} + D_{oo})^{-1} \sum_{k=0}^{2n-1} H_m^k, \tag{33}$$

$$H_m = -(D_{eo} D_{oo}^{-1} + D_{oe} D_{ee}^{-1}). \tag{34}$$

- Frequency splitting
  - Split away the low frequency modes (high eigenmodes of D)
  - When the variance is in the high end of the spectrum (i.e. disconnected diagrams which get most of their contributions from short distances), because the large mass doesn't affect the even larger energy scales close to 1/a.
- Low-mode averaging
  - Split away the high frequency modes (low eigenmodes of D)
  - Beneficial when the variance is in the low end of the spectrum (i.e. long-distance two-point functions)

RBC

Backup slide: RBC storage of low modes

- Storage of low modes [Clark et al. 1710.06884]
- Creation of multigrid subspace using $N \approx 200-400$ exact low modes of Dirac operator
- Determination of further $1k-2k$ low modes in the coarse grid subspace (coarse grid is significantly smaller than fine grid)
- Storage of $1k-2k$ low modes in the coarse grid basis $\implies$ smaller I/O and memory footprint
- Contraction is done in the coarse grid subspace using coarse grid modes
- Like applying LMA to coarse grid operator

Distillation

Backup slide: Distillation

- Distillation [Knechtli et al. 2205.11564] is a Smearing technique (alters absolute value of correlation function)
- Used for spectroscopy
- Determination of low modes of the spatial Laplacian
- Smearing operator as projector to distillation subspace
- Improves the overlap of operators with hadronic states
- Was explored as variance reduction technique à la LMA in Ref. [Bushnaq (2023)]

Scaling with pion mass

Backup slide: Variance vs. pion mass

Name	Size $[T \times L^3]$	Pion mass	a [fm]	L [fm]
G7	128×64^3	270 MeV	0.065 fm	4.2 fm
G8	128×64^3	180 MeV	0.065 fm	4.2 fm

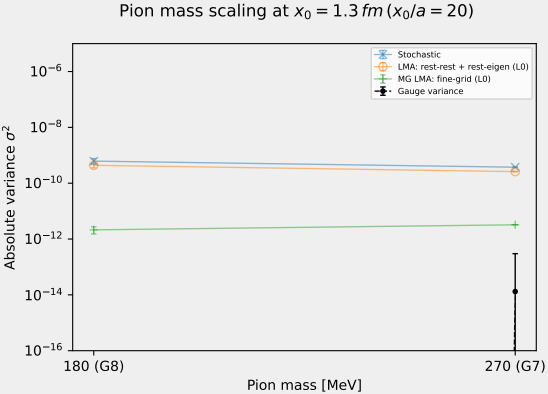


Figure: Absolute variance of L0-terms vs. pion mass. We used $N_c = 50$ low modes on both lattices and the same MG setup.

Scaling with number of low modes

Backup slide: Variance vs. number of low modes

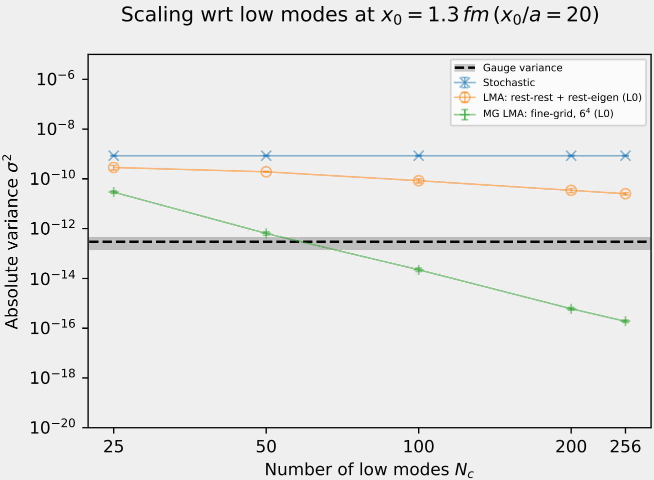


Figure: F7 (3.2 fm): Absolute variance (y-axis) for LMA (yellow) and MG LMA (green) to the vector correlator with varying number of low modes (x-axis) (block size held constant, 6^4). Blue is the stochastic estimator (for reference), the dashed black line is the gauge variance.

Scaling with block size

Backup slide: Variance vs. block size

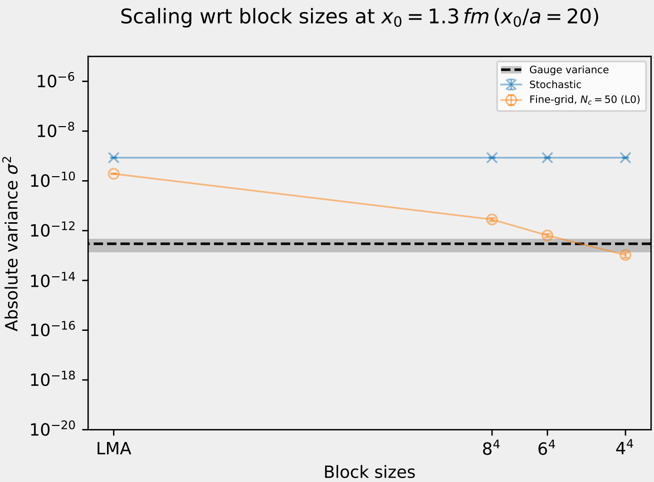


Figure: F7 (3.2 fm): Absolute variance (y-axis) for L0-term (yellow) to the vector correlator with varying block sizes (x-axis) (number of low modes held constant, $N_c = 50$). Blue is the stochastic estimator (for reference), the dashed black line is the gauge variance.

Performance model with multiple RHS

Backup slide: Why multiple RHS? I

Problem

- Time for one application of Dirac operator is linear in its memory footprint
  - Fine-grid Dirac operator D: $(4V \cdot 9 \cdot 2 + 2V \cdot 36)$ floats $\implies 2304V$ bytes.
  - Coarse-grid Dirac operator $\hat{D}$: $9N_c^2 N_s^2 N_b$ complex floats $\implies 144(N_c N_s)^2 N_b$ bytes
- Might happen that mem($\hat{D}$) > mem(D), but still $\dim(D) \gg \dim(\hat{D})$.
  $\implies$ coarse-grid operator more expensive than fine-grid ✗
- Also we have cond$(D) >$ cond$(\hat{D})$
- Krylov solvers: number of iterations $=$ iter$(D) \sim$ cond(D)
  $\implies$ fewer iterations for coarse grid operators ✓

Backup slide: Why multiple RHS? II

Solution

- Memory bandwidth of one spinor field (fine-grid):

$$\text{mem}(\psi) = \dim(D) = 12V \text{ complex floats}$$

- Memory bandwidth of one spinor field (coarse-grid):

$$\text{mem}(\psi) = \dim(\hat{D}) = N_c N_s N_b \text{ complex floats}$$

- Memory bandwidth (Operator with one RHS):

$$\text{mem}(Op) + 2\dim(Op)$$

- Memory bandwidth (Operator with N_{rhs} RHS):

$$\text{mem}(Op) + 2N_{rhs}\dim(Op)$$

Backup slide: Speedup

- We define the speedup as

$$Sp(N_{rhs}) = \frac{\text{iter}(\text{D})}{\text{iter}(\hat{D})} \frac{\text{mem}(D) + 2N_{rhs}\dim(D)}{\text{mem}(\hat{D}) + 2N_{rhs}\dim(\hat{D})}. \tag{35}$$

- With one RHS:

$$Sp(1) \approx \frac{\text{iter}(\text{D})}{\text{iter}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})} \approx \frac{\text{cond}(\text{D})}{\text{cond}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})}. \tag{36}$$

- With many RHS:

$$Sp(\infty) \approx \frac{\text{iter}(\text{D})}{\text{iter}(\hat{D})} \frac{\dim(D)}{\dim(\hat{D})} \approx \underbrace{\frac{\text{cond}(\text{D})}{\text{cond}(\hat{D})}}_{>1} \underbrace{\frac{\dim(D)}{\dim(\hat{D})}}_{30-500} \gg 1. \tag{37}$$

Backup slide: Speedup

- We define the speedup as

$$Sp(N_{rhs}) = \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\text{mem}(D) + 2N_{rhs}\dim(D)}{\text{mem}(\hat{D}) + 2N_{rhs}\dim(\hat{D})}. \tag{35}$$

- With one RHS:

with comparable solvers

$$Sp(1) \approx \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})} \approx \frac{\text{cond(D)}}{\text{cond}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})}. \tag{36}$$

- With many RHS:

$$Sp(\infty) \approx \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\dim(D)}{\dim(\hat{D})} \approx \underbrace{\frac{\text{cond(D)}}{\text{cond}(\hat{D})}}_{>1} \underbrace{\frac{\dim(D)}{\dim(\hat{D})}}_{30-500} \gg 1. \tag{37}$$

Backup slide: Speedup

- We define the speedup as

$$Sp(N_{rhs}) = \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\text{mem}(D) + 2N_{rhs}\dim(D)}{\text{mem}(\hat{D}) + 2N_{rhs}\dim(\hat{D})}. \tag{35}$$

- With one RHS:

with comparable solvers

$$Sp(1) \approx \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})} \approx \frac{\text{cond(D)}}{\text{cond}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})}. \tag{36}$$

- With many RHS:

$$Sp(\infty) \approx \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\dim(D)}{\dim(\hat{D})} \approx \underbrace{\frac{\text{cond(D)}}{\text{cond}(\hat{D})}}_{>1} \underbrace{\frac{\dim(D)}{\dim(\hat{D})}}_{30-500} \gg 1. \tag{37}$$

$N_{rhs} \to \infty$

Backup slide: Speedup

- We define the speedup as

$$Sp(N_{rhs}) = \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\text{mem}(D) + 2N_{rhs}\dim(D)}{\text{mem}(\hat{D}) + 2N_{rhs}\dim(\hat{D})}. \quad (35)$$

- With one RHS:

with comparable solvers

$$Sp(1) \approx \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})} \approx \frac{\text{cond(D)}}{\text{cond}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})}. \quad (36)$$

- With many RHS:

with comparable solvers

$$Sp(\infty) \approx \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\dim(D)}{\dim(\hat{D})} \approx \underbrace{\frac{\text{cond(D)}}{\text{cond}(\hat{D})}}_{>1} \underbrace{\frac{\dim(D)}{\dim(\hat{D})}}_{30-500} \gg 1. \quad (37)$$

$N_{rhs} \to \infty$

Backup slide: Speedup

- We define the speedup as

$$Sp(N_{rhs}) = \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\text{mem}(D) + 2N_{rhs}\dim(D)}{\text{mem}(\hat{D}) + 2N_{rhs}\dim(\hat{D})}. \tag{35}$$

- With one RHS:

with comparable solvers

$$Sp(1) \approx \boxed{\frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})}} \approx \frac{\text{cond(D)}}{\text{cond}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})}. \tag{36}$$

My 💩 implementation

- With many RHS:

with comparable solvers

$$Sp(\infty) \approx \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\dim(D)}{\dim(\hat{D})} \approx \underbrace{\frac{\text{cond(D)}}{\text{cond}(\hat{D})}}_{>1} \underbrace{\frac{\dim(D)}{\dim(\hat{D})}}_{30-500} \gg 1. \tag{37}$$

$N_{rhs} \to \infty$

Backup slide: Speedup

- We define the speedup as

$$Sp(N_{rhs}) = \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\text{mem}(D) + 2N_{rhs}\dim(D)}{\text{mem}(\hat{D}) + 2N_{rhs}\dim(\hat{D})}. \quad (35)$$

- With one RHS:

with comparable solvers

$$Sp(1) \approx \boxed{\frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})}} \approx \frac{\text{cond(D)}}{\text{cond}(\hat{D})} \frac{\text{mem}(D)}{\text{mem}(\hat{D})}. \quad (36)$$

My 💩 implementation

- With many RHS:

with comparable solvers

$$Sp(\infty) \approx \frac{\text{iter(D)}}{\text{iter}(\hat{D})} \frac{\dim(D)}{\dim(\hat{D})} \approx \boxed{\underbrace{\frac{\text{cond(D)}}{\text{cond}(\hat{D})}}_{>1} \underbrace{\frac{\dim(D)}{\dim(\hat{D})}}_{30-500}} \gg 1. \quad (37)$$

$N_{rhs} \to \infty$

Ideal implementation

Backup slide: Scaling w.r.t. dimension

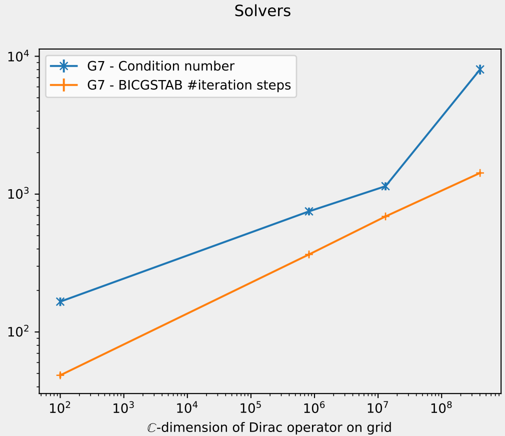


Figure: G7 (4.2 fm): $\mathbb{C}$-dimension of the Dirac operator (x-axis) vs. Number of bicgstab iterations and condition number (y-axis). The rightmost operator is the fine-grid, the others ones are different coarse-grid Dirac operators.

Backup slide: Memory bandwidth model

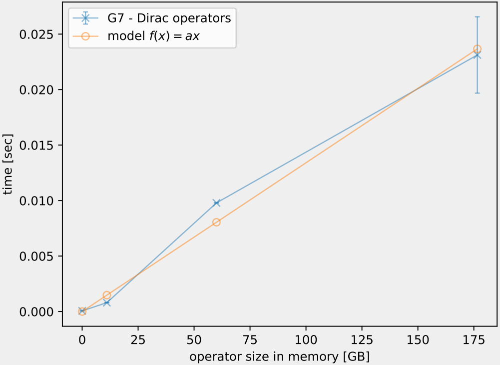


Figure: G7 (4.2 fm): Time for one application of a fine- or coarse-grid Dirac operator (y-axis) vs. its memory footprint.

Chirality on the subspace

Backup slide: Why preserve chirality I

- Chirality preservation on the subspace (γ^5-Hermiticity on the subspace):

$$[P, \gamma^5] = 0 \implies \widehat{\gamma^5}\hat{D} = \left(\widehat{\gamma^5}\hat{D}\right)^\dagger . \quad (38)$$

where $\widehat{\gamma^5} = R\gamma^5 T$ and $P = TR$

- Or number of remaining spin d.o.f., $N_s = 1, 2, 4$ on the coarse subspace
- When generating the subspace basis from eigenmodes ϕ_i

$$\{\phi_i\}_{i=1}^{N_c} \longmapsto \{\phi_i\}_{i=1}^{N_c} \cup \{\gamma^5 \phi_i\}_{i=1}^{N_c} \quad (39)$$

$$\{\phi_i\}_{i=1}^{N_c} \longmapsto \{P_+\phi_i\}_{i=1}^{N_c} \cup \{P_-\phi_i\}_{i=1}^{N_c} \quad (40)$$

where $P_\pm = \frac{1}{2}\left(1 \pm \gamma^5\right)$.

Backup slide: Why preserve chirality II

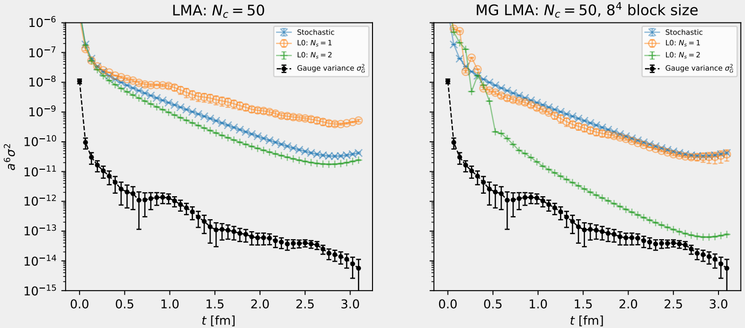


Figure: Variance contribution for LMA (left) and MG LMA (right) without chirality preservation (yellow) and with chirality preservation (green).

Backup slide: Why preserve chirality III

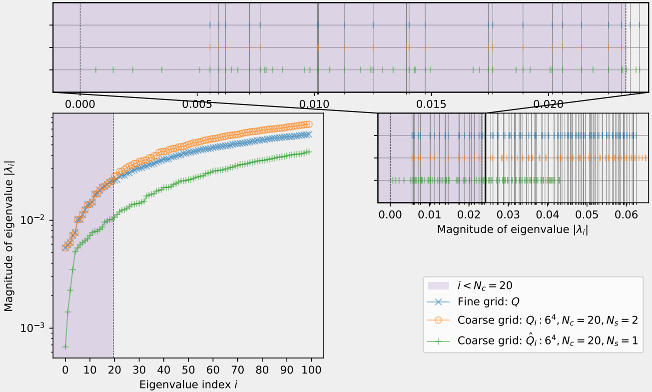


Figure: Lower left: lowest 100 eigenvalues (in magnitude) of the Hermitian Dirac operator $\gamma^5 D$ and $\widehat{\gamma^5 D}$. **Lower right** eigenvalues plotted on the x-axis with a gray vertical line at every fine grid eigenvalue (blue). **Upper panel**: zoom. $N_c = 20$.

Detailed setups

Backup slide: Detailed setups

Estimator	# modes	Sources	Levels
Stochastic	N/A	semwall	L0: only fine-grid
LMA	50	semwall exact	L0: (rest-rest + rest-eigen) L1: (eigen-eigen)
2-level MG LMA	50	semwall	L0: fine-grid L1: block size 8^4
3-level MG LMA	50	semwall exact	L0: fine-grid L1: block size 8^4 L2: (eigen-eigen)
4-level MG LMA	50	semwall exact	L0: fine-grid L1: block size 4^4 L2: block size 8^4 L3: (eigen-eigen)

Variance contribution - All ensembles

Relative variances: E7

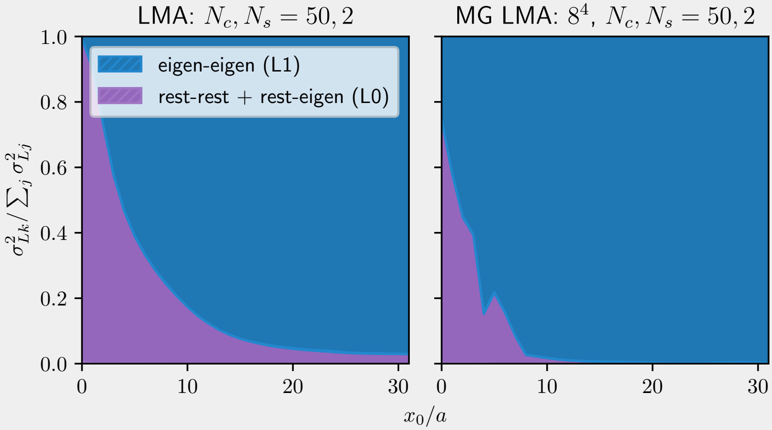
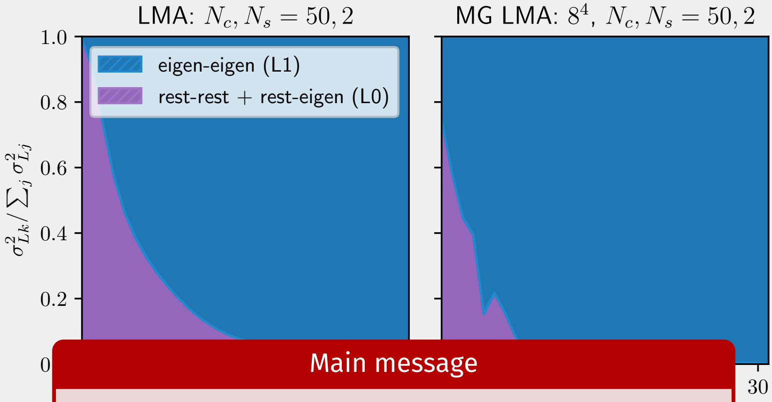


Figure: Relative variance for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term.



Main message

We observe a significant variance contribution from the cheap-to-evaluate L1-term w.r.t LMA.

Figure: F... correlat... with one stochastic source for each term.

Relative variances: F7

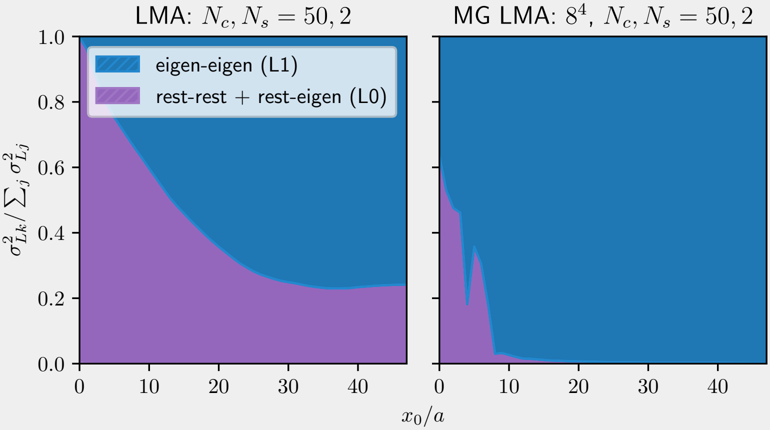
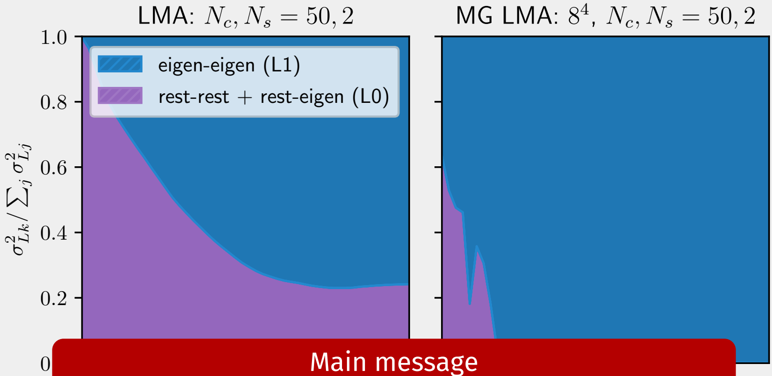


Figure: Relative variance for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term.



Main message

We observe a significant variance contribution from the cheap-to-evaluate L1-term w.r.t LMA.

Figure: F… correlat… one stochastic source for each term.

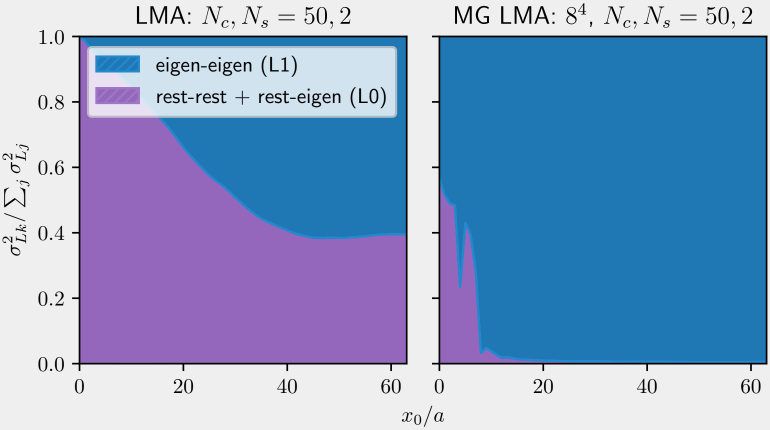


Figure: Relative variance for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term.

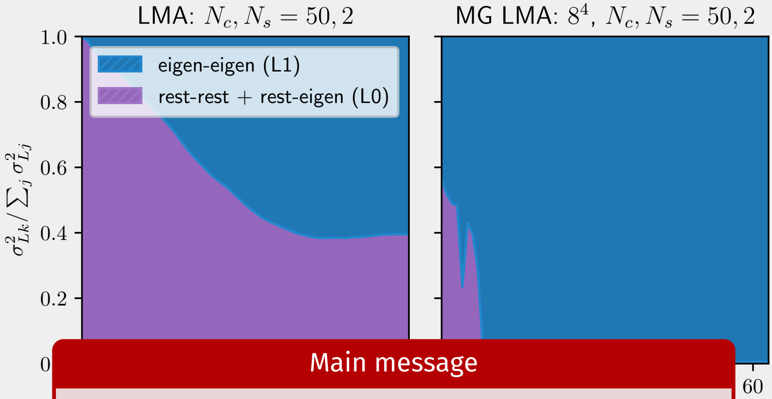


Figure: F… correlato… with one stochastic source for each term.

Main message

We observe a significant variance contribution from the cheap-to-evaluate L1-term w.r.t LMA.

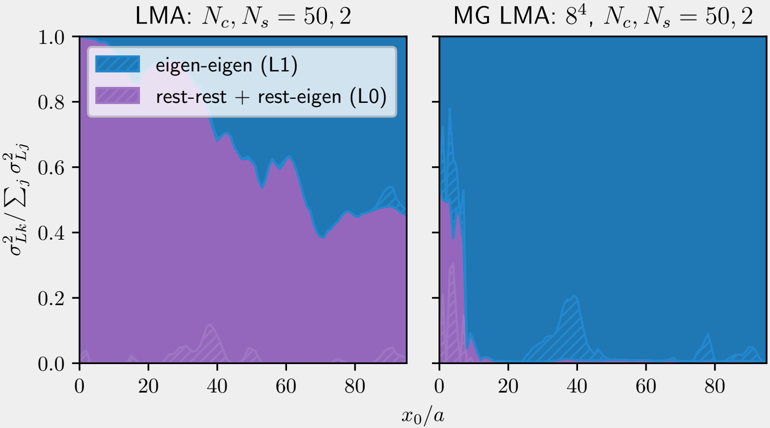
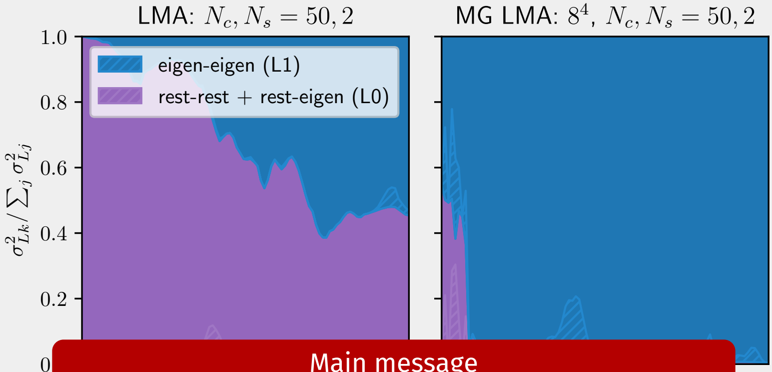


Figure: Relative variance for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term.



Main message

We observe a significant variance contribution from the cheap-to-evaluate L1-term w.r.t LMA.

Figure: F… correlato… with one stochastic source for each term.

Absolute variance - All ensembles

Absolute variances: E7

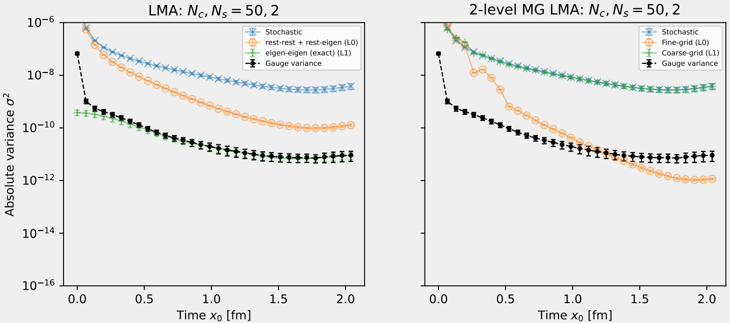


Figure: Absolute variances for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term. The black line is the gauge variance.

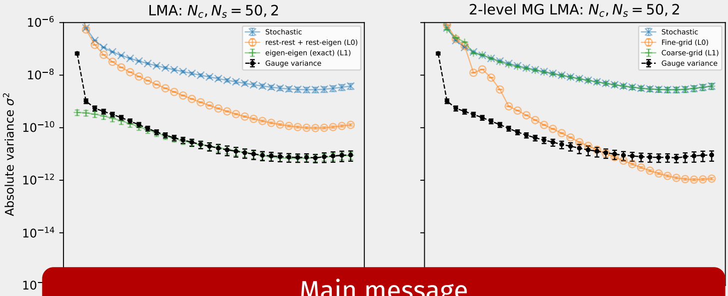


Figure: A ... r
correlat ... e gauge
variance

Main message

We are able to push the remaining L0 noise down to the gauge noise using only a few stochastic sources.

Absolute variances: F7

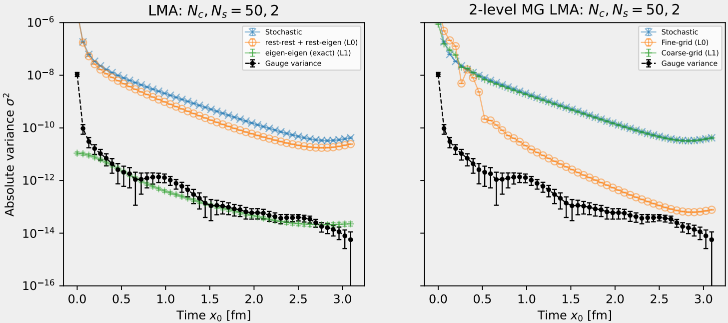


Figure: Absolute variances for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term. The black line is the gauge variance.

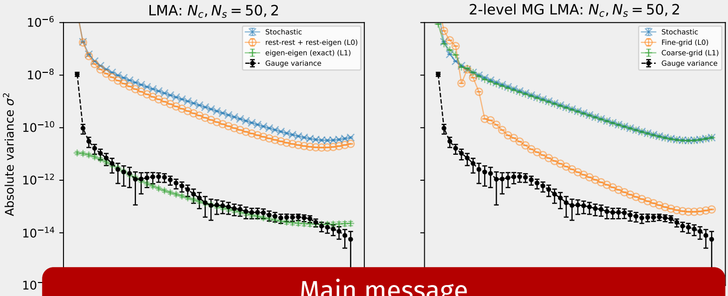


Figure: A … r
correlat … e gauge
variance …

Main message

We are able to push the remaining L0 noise down to the gauge noise using only a few stochastic sources.

Absolute variances: G7

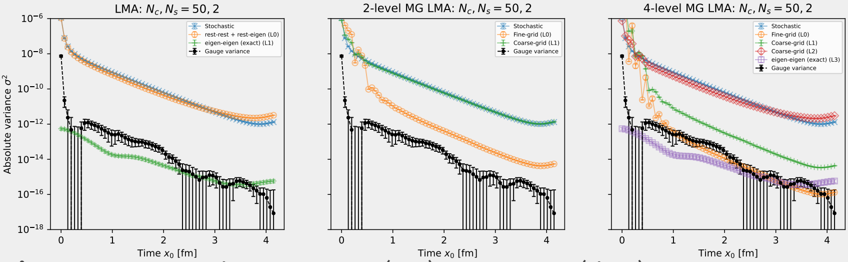


Figure: Absolute variances for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term. The black line is the gauge variance.

Absolute variances: G7

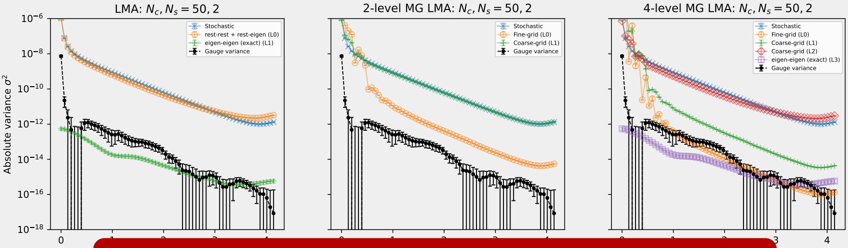


Figure: A … r
correlat… e gauge
variance…

Main message

We are able to push the remaining L0 noise down to the gauge noise using only a few stochastic sources.

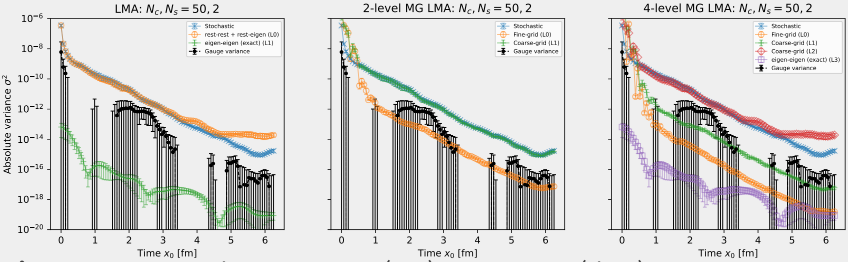


Figure: Absolute variances for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term. The black line is the gauge variance.

Absolute variances: H7

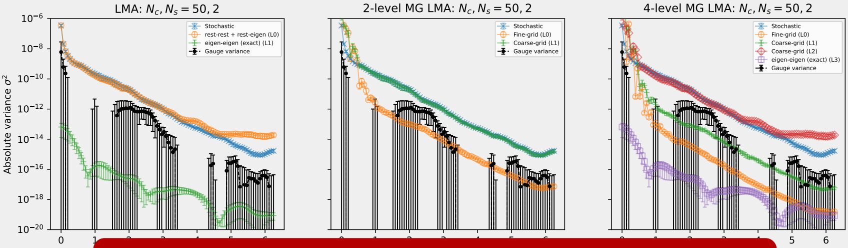


Figure: A ... r
correlat ... e gauge
variance ...

Main message

We are able to push the remaining L0 noise down to the gauge noise using only a few stochastic sources.

Variance vs. sources - All ensembles

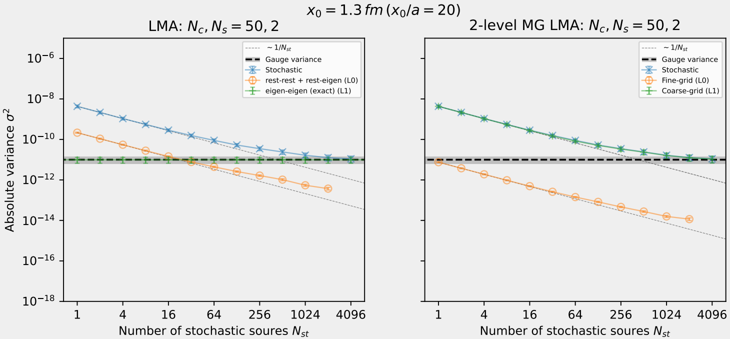


Figure: Absolute variances for LMA (left) and MG LMA (right) against number of stochastic sources N_{st}. The black line is the gauge variance.

Variance vs. sources: F7

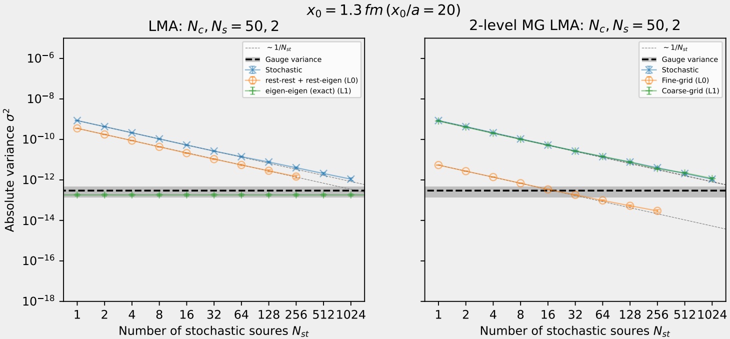


Figure: Absolute variances for LMA (left) and MG LMA (right) against number of stochastic sources N_{st}. The black line is the gauge variance.

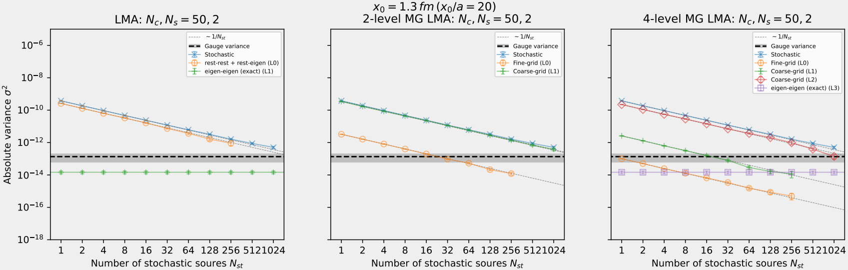


Figure: Absolute variances for LMA (left) and MG LMA (right) against number of stochastic sources N_{st}. The black line is the gauge variance.

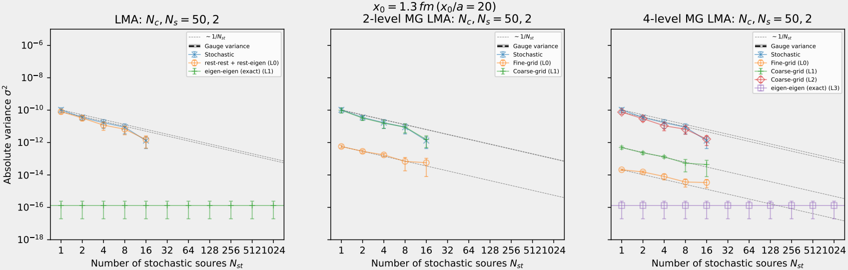


Figure: Absolute variances for LMA (left) and MG LMA (right) against number of stochastic sources N_{st}. The black line is the gauge variance.

Cost - All ensembles

Cost - E7

Table: Cost breakdown to reach the gauge variance for E7 (2.1 fm).

Estimator	# modes	# sources	meas. cost[1]	model cost[1]
Stochastic	0	L0: 1024	4096	4096
LMA[2]	50	L0: 16	64	64
2-lvl MG LMA[2]	50	L0: $1^{\star}$ L1: $1024^{\star\star}$	100.4	12.3
3-lvl MG LMA[2]	50	L0: $1^{\star}$ L1: $16^{\star\star}$	5.5	4.1

My 💩 implementation:

$\star$	fine-grid	64×32^3	inv: 5.32 ± 0.03 sec	(iter: 35.65 ± 0.15)
$\star\star$	coarse-grid	8×4^3	inv: 0.125 ± 0.000 sec	(iter: 140.5 ± 0.3)

[1] Unit = fine-grid inversions.

[2] Cost of determination of low modes not included (or add 100 - 200 to the cost).

Cost - F7

Table: Cost breakdown to reach the gauge variance for F7 (3.2 fm).

Estimator	# modes	# sources	meas. cost[1]	model cost[1]
Stochastic	0	L0: 2048	8192	8192
LMA[2]	50	L0: 1024	4096	4096
2-lvl MG LMA[2]	50	L0: 16* L1: 2048**	462.3	80.7
3-lvl MG LMA[2]	50	L0: 16* L1: 1024**	263.2	72.3

My 💩 implementation:

$\star$	fine-grid	96×48^3	inv: 8.42 ± 0.04 sec	(iter: 43.77 ± 0.15)
$\star\star$	coarse-grid	12×6^3	inv: 0.409 ± 0.002 sec	(iter: 337.6 ± 1.3)

[1] Unit = fine-grid inversions.
[2] Cost of determination of low modes not included (or add 100 - 200 to the cost).

Table: Cost breakdown to reach the gauge variance for G7 (4.2 fm).

Estimator	# modes	# sources	meas. cost[1]	model cost[1]
Stochastic	0	L0: 4096	16384	16384
LMA[2]	50	L0: 2048	8192	8192
2-lvl MG LMA[2]	50	L0: $16^{\star}$ L1: $2048^{\star\star\star}$	557.8	80.7
4-lvl MG LMA[2]	50	L0: $1^{\star}$ L1: $16^{\star\star}$ L2: $1024^{\star\star\star}$	466.7	14.4

My 💩 implementation:

$\star$	fine-grid	128×64^3	inv: 11.1 ± 0.4 sec	(iter: 46.53 ± 0.23)
$\star\star$	coarse-grid	32×16^3	inv: 37.3 ± 2.4 sec	(iter: 1417 ± 22)
$\star\star\star$	coarse-grid	16×8^3	inv: 0.667 ± 0.041 sec	(iter: 502.1 ± 5.8)

[1] Unit = fine-grid inversions.

[2] Cost of determination of low modes not included (or add 100 - 200 to the cost).

Optimizations

Optimisations

1. Relax the precision of the low-modes (we used 10^{-12} precise modes)
2. Investigate different source types (we used time-diluted spin-diagonal random wall-sources)
3. Do contractions on the coarse grid(s) (we prolongate the coarse propagators to fine grid)
4. Spin-sources on fine grid $\implies$ 4 inversions/source (4 Spins) $\rightarrow$ Coarse grid: 1-2 Spin d.o.f. $\implies$ only 1-2 coarse inversions/source? (we do 4 inversions on the coarse grid)
5. Implementation: solid coarse-grid solver

Comparison - LMA vs. MG LMA

Comparison

Comparison: LMA vs. MG LMA		
Metric	LMA	MG LMA
# low modes	$O(1000)$	$O(50)$
Memory footprint	TBytes[1]	GBytes
LO treatment	very complicated	simple
Complexity of contraction code	different contraction for every term	same contraction code for all
LO-Inversions	O(1000)	few (1-10)
L1-term treatment	exact evaluation	usually stochastic with hierarchical evaluation or exact

[1] G7-like lattice (Wilson) with $N_c = 2000$ low modes $\rightarrow$ 24 TB

Multigrid Multilevel Monte Carlo (MGMLMC)

Multigrid Multilevel Monte Carlo (MGMLMC)

- Estimation of traces of matrix inverses [Whyte et al. 2212.04430], $\mathrm{tr}\left(A^{-1}\right)$, with
$$A^{-1} = A^{-1} - PA_c^{-1}P^\dagger + PA_c^{-1}P^\dagger \quad (41)$$
- A_c is a coarse operator from multigrid
- They only apply if to full inverse matrix traces, no real disconnected diagrams $\rightarrow$ final goal
- Frequency splitting $\rightarrow$ noise in disconnected diagrams comes from the high modes $\rightarrow$ MGMLMC is not expected to be very beneficial

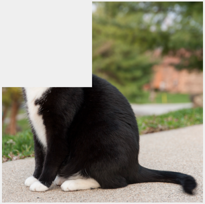
Multilevel Monte Carlo

Multilevel Monte Carlo

- Multigrid LMA propagator decomposition:

 $=$  $+\,$correction. (42)

- Multilevel Monte Carlo propagator decomposition:

 $=$ $\underbrace{\quad}_{local}$  $+$ $\underbrace{\quad}_{non-local}$  (43)