Variance reduction for the lattice HVP

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

ETH Zürich

September 11, 2024

Table of contents

Low-mode averaging (LMA) and its variants

Motivation - Low-mode averaging (LMA)

- 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: should contain most of the variance
  - Other piece: negligible variance
- Determine N_c lowest modes of $D, Q = \gamma^5 D$, eo-preconditioned D, Q
- Write $S = D^{-1} =$ truncated spectral/singular sum + remainder

$$Q^{-1} = \underbrace{\sum_{i=1}^{N_c} \frac{1}{\lambda_i} \xi_i \xi_i^\dagger}_{Q^{-1}_{LMA}} + \underbrace{P Q^{-1} P^\dagger}_{Q^{-1}_{rest} = Q^{-1} - Q^{-1}_{LMA}} , \qquad (1)$$

with

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

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

$$G(t) = \frac{1}{|\Omega_0|} \sum_{y \in \Omega_0} \sum_{\vec{x} \in \Sigma_0} C(y_0 + t, \vec{x}|y), \tag{2}$$

$$C(x|y) = \mathrm{tr}\left[\Gamma_1 S(x|y) \Gamma_2 S(x|y)^\dagger\right], \tag{3}$$

- Stochastic sources: introduce extra noise
- Point sources: costs L^3
- Ideally, but unrealistic: full lattice volume average

- Plug in decomposition of propagator

$$G(t) = G_{ee}(t) + \underbrace{G_{re}(t) + G_{er}(t)}_{G_\times(t)} + G_{rr}(t) \qquad (4)$$

- 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
$G_\times(t)$: 10-30% contribution to total noise $\gg$ gauge noise

The two-point correlator with LMA

- Plug in decomposition of propagator

$$G(t) = G_{ee}(t) + \underbrace{G_{re}(t) + G_{er}(t)}_{G_\times(t)} + G_{rr}(t) \tag{4}$$

- 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
$G_\times(t)$: 10-30% contribution to total noise $\gg$ gauge noise

- Plug in decomposition of propagator

$$G(t) = G_{ee}(t) + \underbrace{G_{re}(t) + G_{er}(t)}_{G_{\times}(t)} + G_{rr}(t) \tag{4}$$

- 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 ✔
$G_{\times}(t)$: 10-30% contribution to total noise $\gg$ gauge noise

The two-point correlator with LMA

- Plug in decomposition of propagator

$$G(t) = G_{ee}(t) + \underbrace{G_{re}(t) + G_{er}(t)}_{G_\times(t)} + G_{rr}(t) \tag{4}$$

- 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	✓
$G_\times(t)$:	10-30% contribution to total noise $\gg$ gauge noise	💩

The various problems 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 various problems 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-eigen piece
   - ...

Multigrid / Deflation

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

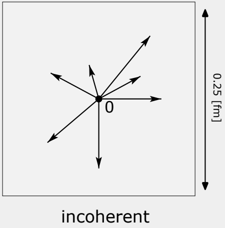


Figure: (Local) coherence of low modes (taken from Ref. [Luscher 1002.4232]).

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

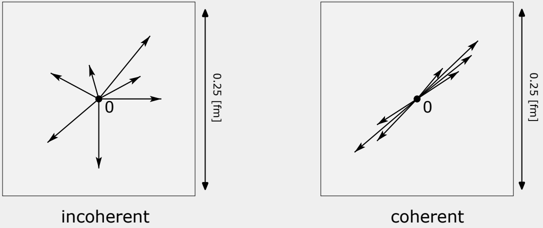


Figure: (Local) coherence of low modes (taken from Ref. [Luscher 1002.4232]).

Conclusion

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

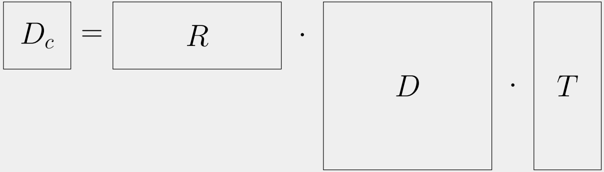
Multigrid / Deflation

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

$$R: \ \psi \mapsto \theta, \quad \theta(i) = \langle \phi_i | \psi \rangle \,, \tag{5}$$

$$T: \ \theta \mapsto \psi = \sum_i \theta(i) \phi_i, \tag{6}$$

- Define the coarse-grid Dirac operator(s) as $D_c = RDT$



- Connetion to solver: sloppy D_c^{-1} as preconditioner for the Dirac equation

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

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

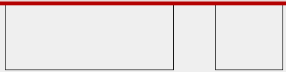
$$R:\ \psi \mapsto \theta, \quad \theta(i) = \langle \phi_i | \psi \rangle\,, \tag{5}$$

$$T:\ \theta \mapsto \psi = \sum \theta(i)\phi_i, \tag{6}$$

- De

Main message

Coarse-grid operator has smaller dimension, smaller condition number and is thus **cheaper to invert**!



- Connetion to solver: sloppy D_c^{-1} as preconditioner for the Dirac equation

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

Multigrid low-mode averaging (MG LMA)

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

$$S = \sum_{i=0}^{N-1} S_i = \underbrace{S - K_1}_{=S_0} + \underbrace{K_1 - K_2}_{=S_1} + \underbrace{K_2 - K_3}_{=S_2} + \cdots + \underbrace{K_{N-1}}_{S_{N-1}}, \qquad (7)$$

$K_i = T_i(D_{c,i})^{-1} R_i, \quad S_i =$ 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:

$$C_{ij}(x,y) = \mathrm{tr}\left[\Gamma_1 S_i(x|y)\Gamma_2 S_j(y|x)\right], \quad C = \sum_{i,j} C_{ij}. \tag{8}$$

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

$$G(t) = \sum_{k=0}^{N-1} G_{Lk}(t). \tag{9}$$

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 $D_{c,k}$ on level k and coarser, but not finer levels!

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

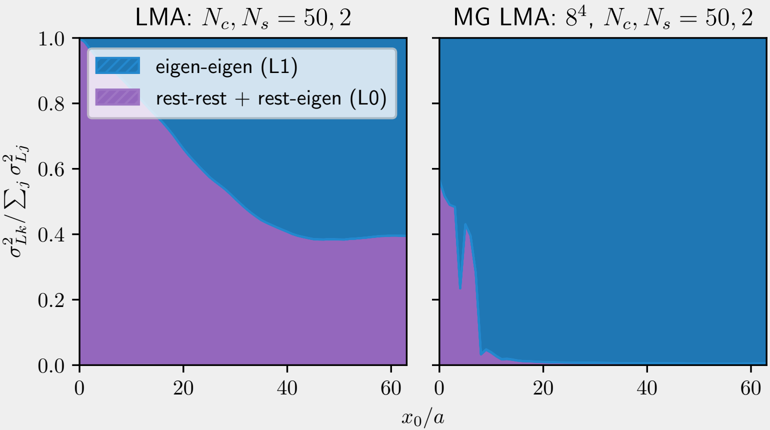


Figure: Relative variance for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term.

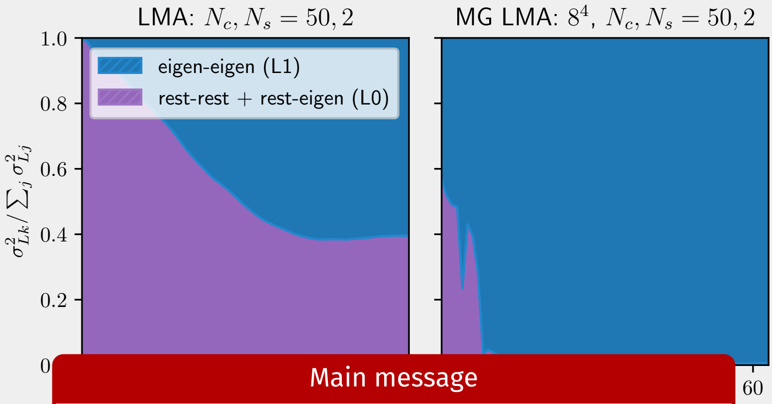


Figure: F
correlat

Main message

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

Absolute variances: G7

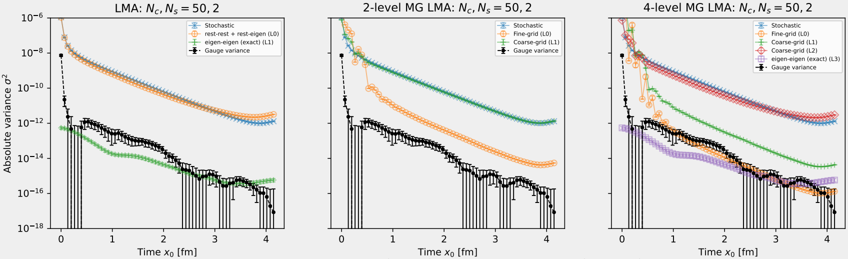


Figure: Absolute variances for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term. The black line is the gauge variance.

Absolute variances: G7

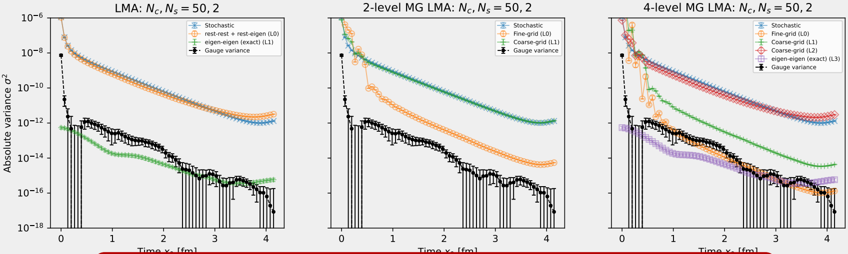


Figure: A … r
correlat … e gauge
variance …

Main message

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

Variance vs. sources: G7

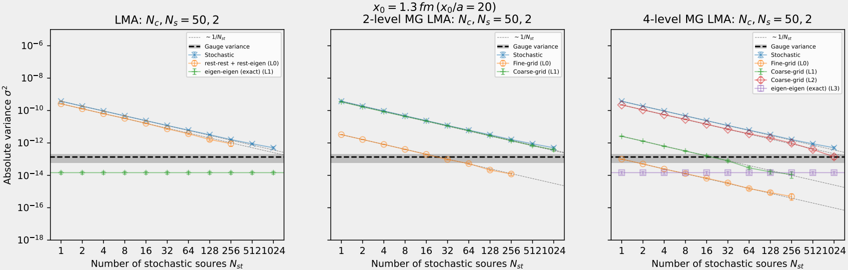


Figure: Absolute variances for LMA (left) and MG LMA (right) against number of stochastic sources N_{st}. The black line is the gauge variance.

Variance vs. volume

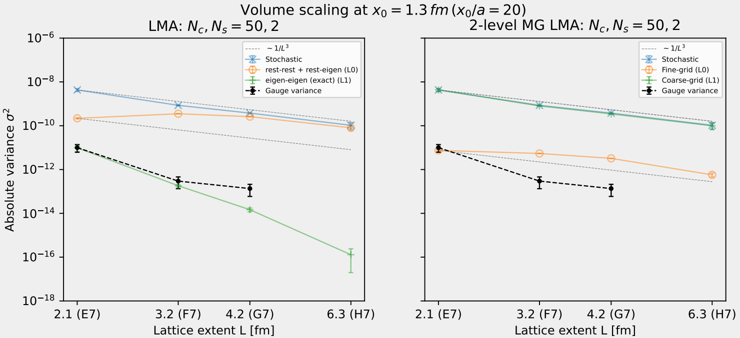


Figure: Absolute variances for LMA (left) and MG LMA (right) against the lattice extent L. The black line is the gauge variance.

Variance vs. volume

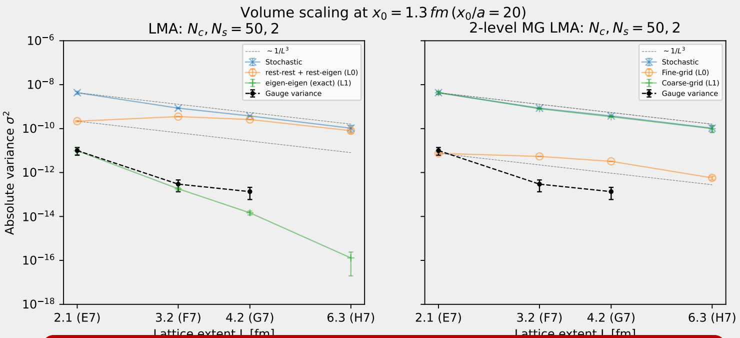


Figure: A ... lattice extent L

Main message

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

Cost

Cost - G7

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:

*	fine-grid	128×64^3	inv: 11.1 ± 0.4 sec	(iter: 46.53 ± 0.23)
**	coarse-grid	32×16^3	inv: 37.3 ± 2.4 sec	(iter: 1417 ± 22)
***	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).

CONCLUSION

Concluding summary

- Subspaces based on domain-decomposed / coarsened low modes
- 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
- 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

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.

References II

[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].

References III

[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. Luscher, **"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].

References IV

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

[19] M. Luscher, **"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/.

Backup slides: Table of contents

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

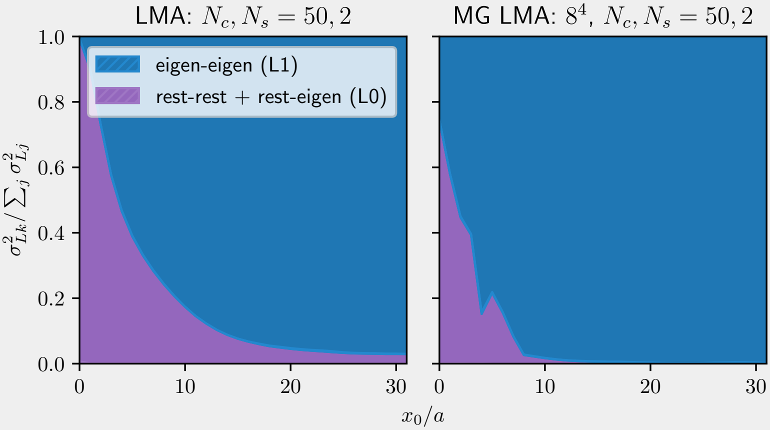


Figure: Relative variance for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term.

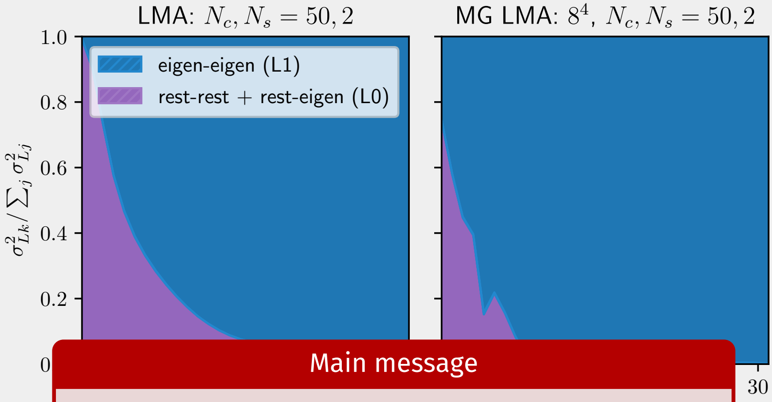


Figure: F... correlat... with one stochastic source for each term.

Main message

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

Relative variances: F7

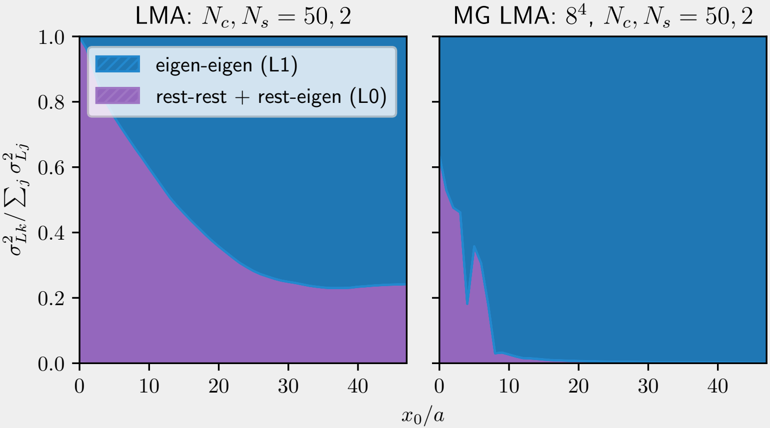


Figure: Relative variance for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term.

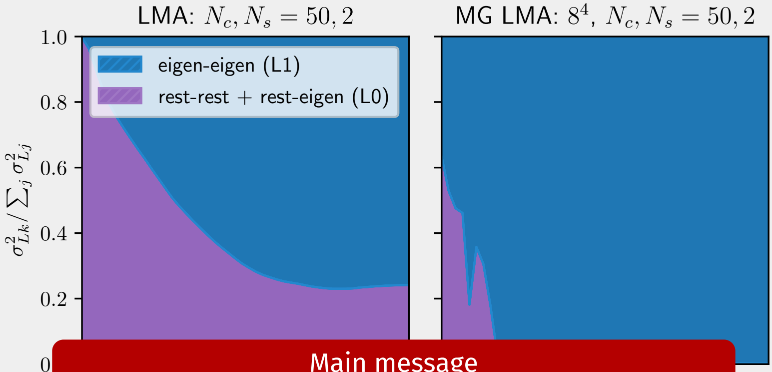


Figure: F

correlato

Main message

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

Relative variances: G7

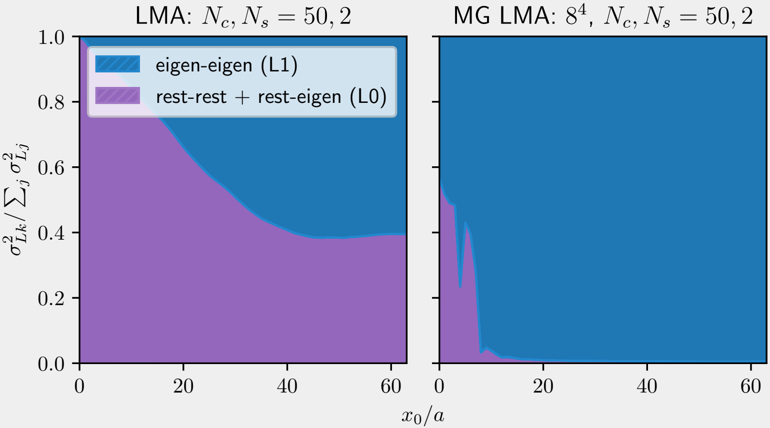


Figure: Relative variance for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term.

Relative variances: G7

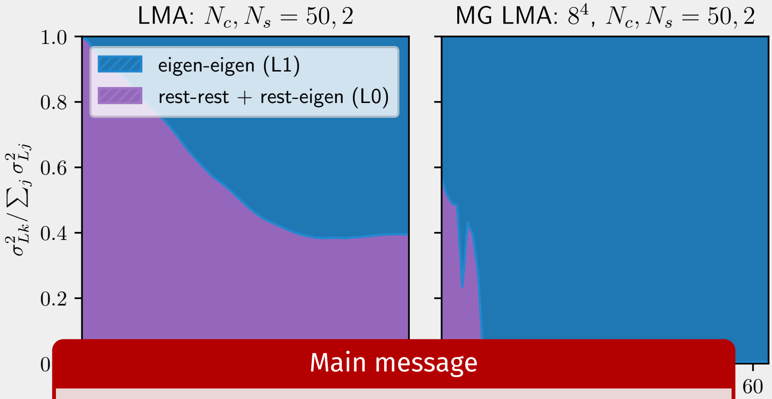


Figure: F… correlat… with one stochastic source for each term.

Main message

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

Relative variances: H7

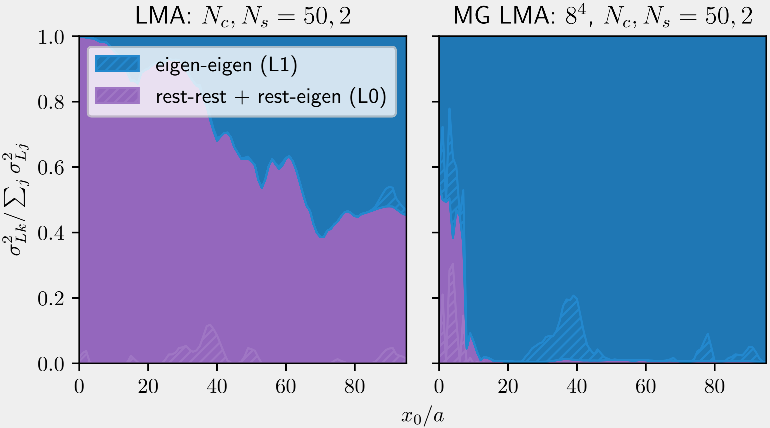


Figure: Relative variance for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term.

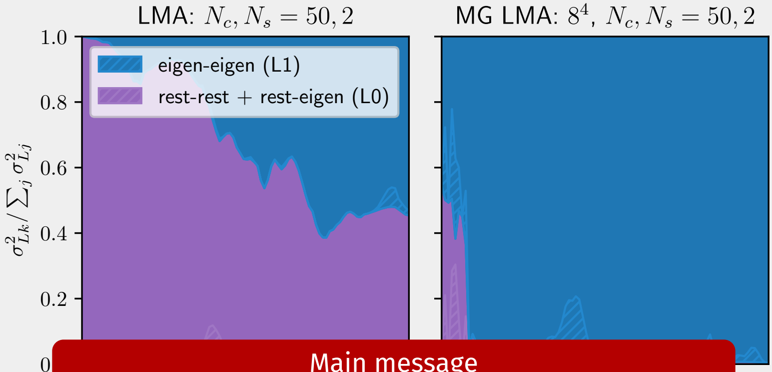


Figure: F… correlator with one stochastic source for each term.

Main message

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

Absolute variance - All ensembles

Absolute variances: E7

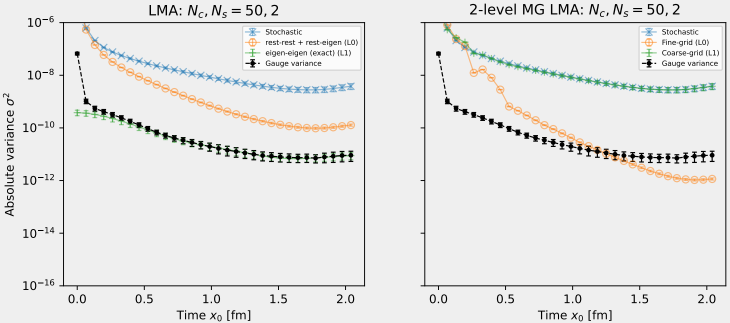


Figure: Absolute variances for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term. The black line is the gauge variance.

ABSOLUTE VARIANCES: E7

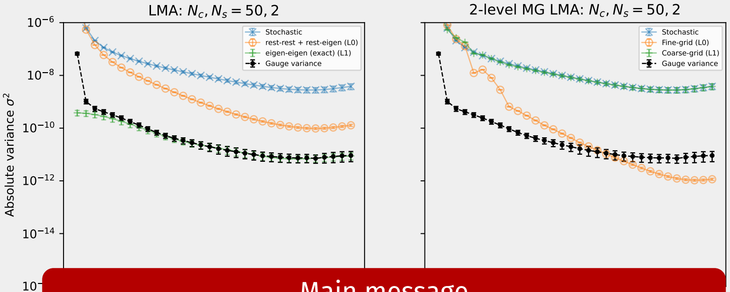


Figure: A ... r correlat ... e gauge variance

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

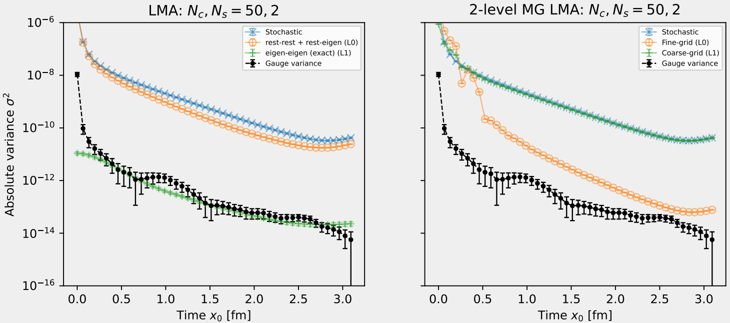


Figure: Absolute variances for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term. The black line is the gauge variance.

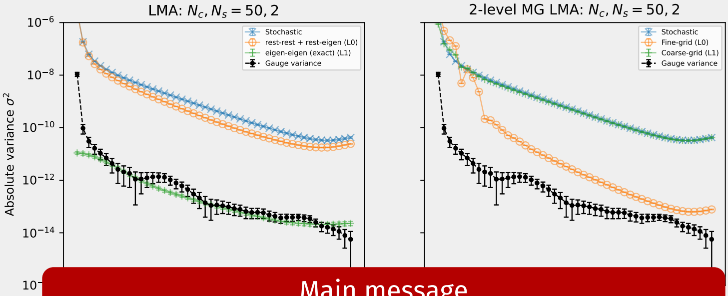


Figure: A ... r correlat... e gauge variance...

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

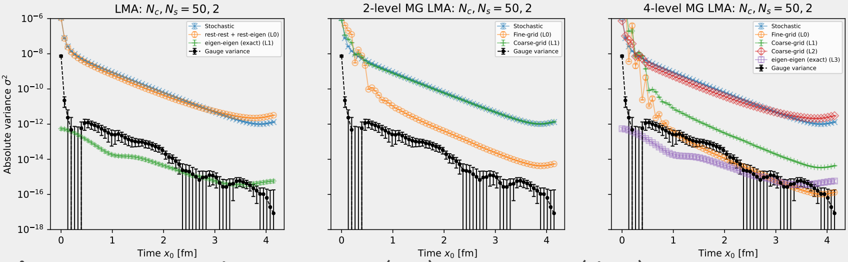


Figure: Absolute variances for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term. The black line is the gauge variance.

ABSOLUTE VARIANCES: G7

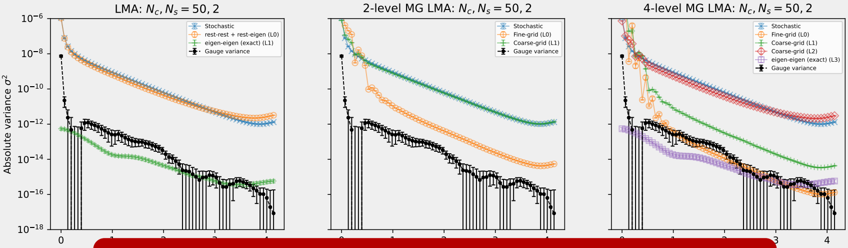


Figure: A ... r
correlat ... e gauge
varianc

Main message

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

Absolute variances: H7

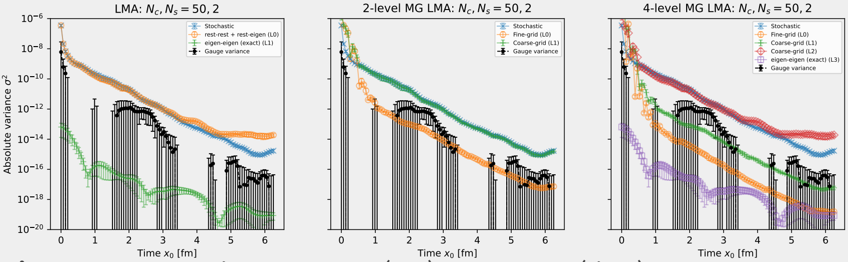


Figure: Absolute variances for LMA (left) and MG LMA (right) to the vector correlator with one stochastic source for each term. The black line is the gauge variance.

Absolute variances: H7

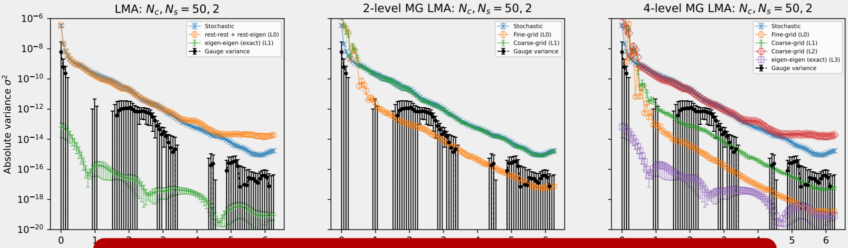


Figure: A … or correlat … e gauge variance …

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

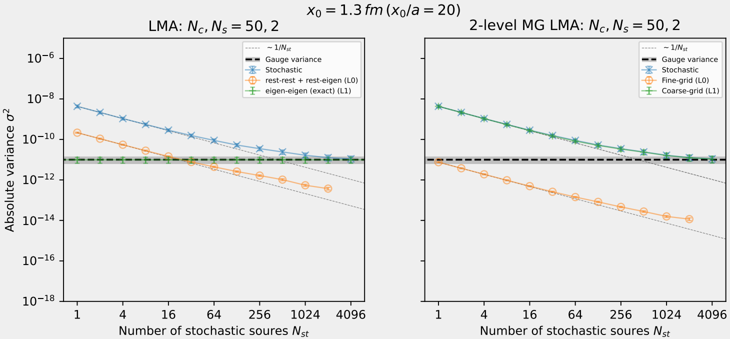


Figure: Absolute variances for LMA (left) and MG LMA (right) against number of stochastic sources N_{st}. The black line is the gauge variance.

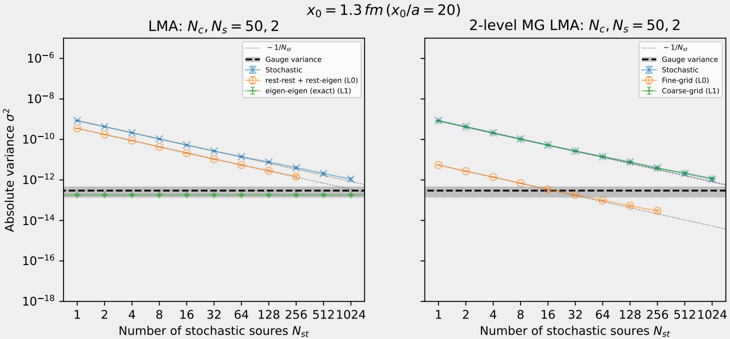


Figure: Absolute variances for LMA (left) and MG LMA (right) against number of stochastic sources N_{st}. The black line is the gauge variance.

Variance vs. sources: G7

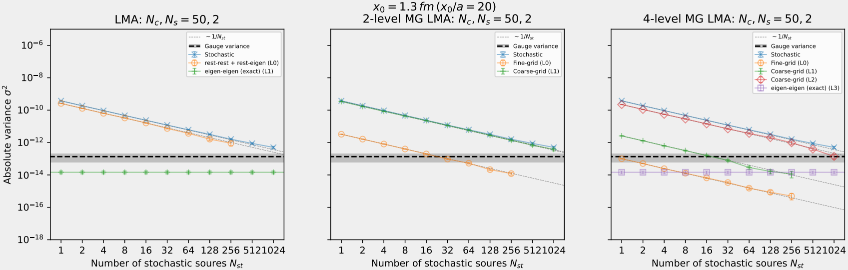


Figure: Absolute variances for LMA (left) and MG LMA (right) against number of stochastic sources N_{st}. The black line is the gauge variance.

Variance vs. sources: H7

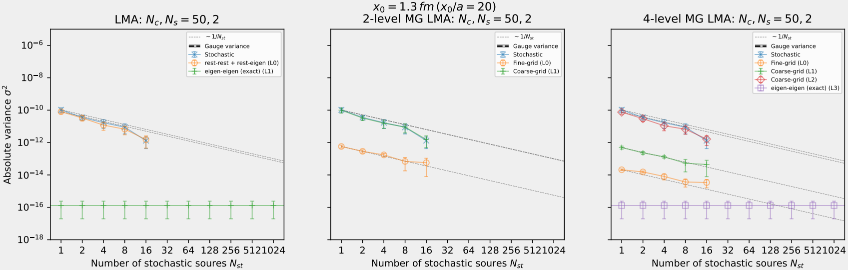


Figure: Absolute variances for LMA (left) and MG LMA (right) against number of stochastic sources N_{st}. The black line is the gauge variance.

Cost - All ensembles

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: $16^{\star}$ L1: $2048^{\star\star}$	462.3	80.7
3-lvl MG LMA[2]	50	L0: $16^{\star}$ L1: $1024^{\star\star}$	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* L1: 2048***	557.8	80.7
4-lvl MG LMA[2]	50	L0: 1* L1: 16** L2: 1024***	466.7	14.4

My 💩 implementation:

*	fine-grid	128×64^3	inv: 11.1 ± 0.4 sec	(iter: 46.53 ± 0.23)
**	coarse-grid	32×16^3	inv: 37.3 ± 2.4 sec	(iter: 1417 ± 22)
***	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).