Variance Reduction with Probing and Multilevel Monte Carlo in Lattice QCD

Jose Jimenez-Merchan,

Andreas Frommer, Bruno Lang, Mario Papace, Gustavo Ramirez-Hidalgo*

Bergische Universität Wuppertal

Forschung Zentrum Jülich*

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

FOR 5269

BERGISCHE UNIVERSITÄT WUPPERTAL

Trace Estimation in LQCD

The Hutchinson's method (Stochastic sources)

Trace Estimation in LQCD

The Hutchinson's method (Stochastic sources)

$$\mathrm{tr}(D^{-1}) \approx \hat{\mathrm{tr}}(D^{-1}) = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)})^{\dagger} D^{-1} x^{(i)}$$

Entries of $x^{(i)}$ are e.g. $\mathbb{Z}_2$, $\mathbb{Z}_4$, or $\mathcal{N}(0,1)$

Trace Estimation in LQCD

The Hutchinson's method (Stochastic sources)

$$\mathrm{tr}(D^{-1}) \approx \hat{\mathrm{tr}}(D^{-1}) = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)})^{\dagger} D^{-1} x^{(i)}$$

Entries of $x^{(i)}$ are e.g. $\mathbb{Z}_2$, $\mathbb{Z}_4$, or $\mathcal{N}(0,1)$

Problem: For target acc. ϵ

$$N \propto \frac{\|\mathrm{offdiag}(D^{-1})\|_F^2}{\epsilon^2}$$

Trace Estimation in LQCD

The Hutchinson's method (Stochastic sources)

$$\mathrm{tr}(D^{-1}) \approx \hat{\mathrm{tr}}(D^{-1}) = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)})^{\dagger} D^{-1} x^{(i)}$$

Entries of $x^{(i)}$ are e.g. $\mathbb{Z}_2$, $\mathbb{Z}_4$, or $\mathcal{N}(0,1)$

Problem: For target acc. ϵ

$$N \propto \frac{\|\mathrm{offdiag}(D^{-1})\|_F^2}{\epsilon^2}$$

A solution: Deflation!*

$$\mathbb{V}[(x^{(i)})^{\dagger} D^{-1} x^{(i)}] = \|\mathrm{offdiag}(D^{-1})\|_F^2 = \left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2} - \sum_{i=1}^{n}|D_{ii}^{-1}|^2\right)$$

Trace Estimation in LQCD

The Hutchinson's method (Stochastic sources)

$$\mathrm{tr}(D^{-1}) \approx \hat{\mathrm{tr}}(D^{-1}) = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)})^\dagger D^{-1}x^{(i)}$$

Entries of $x^{(i)}$ are e.g. $\mathbb{Z}_2$, $\mathbb{Z}_4$, or $\mathcal{N}(0,1)$

Problem: For target acc. ϵ

$$N \propto \frac{\|\mathrm{offdiag}(D^{-1})\|_F^2}{\epsilon^2}$$

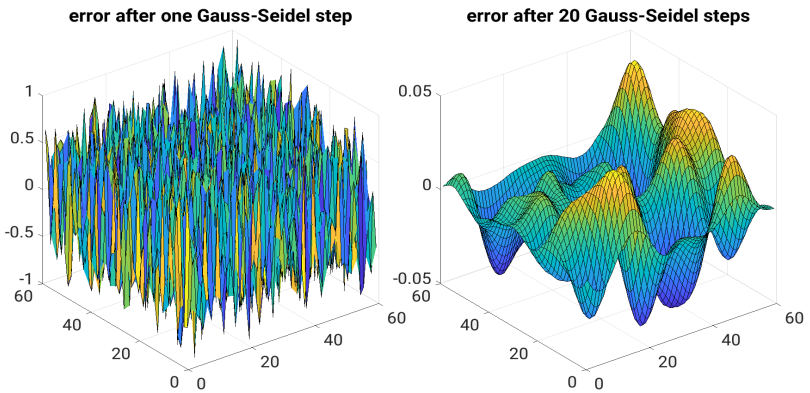
A solution: Deflation!*

- * *Gambhir et al. (2017)*
- Frequency splitting *Giusti et al. (2019)*
- Multigrid MLMC *Frommer et al. (2022)*

$$\mathbb{V}[(x^{(i)})^\dagger D^{-1}x^{(i)}] = \|\mathrm{offdiag}(D^{-1})\|_F^2 = \left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2} - \sum_{i=1}^{n}|D_{ii}^{-1}|^2\right)$$

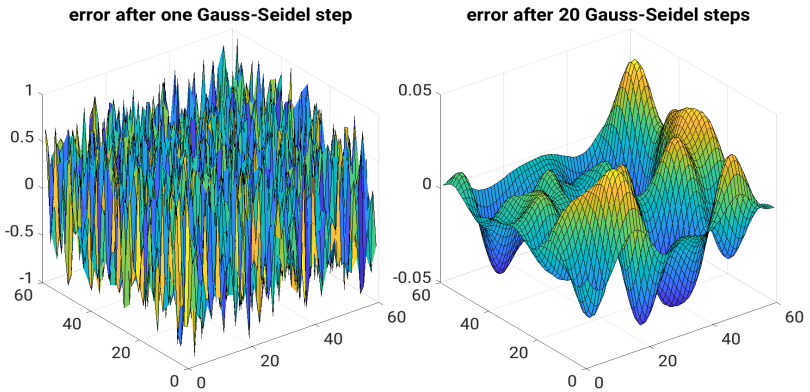
Algebraic Multigrid (AMG)

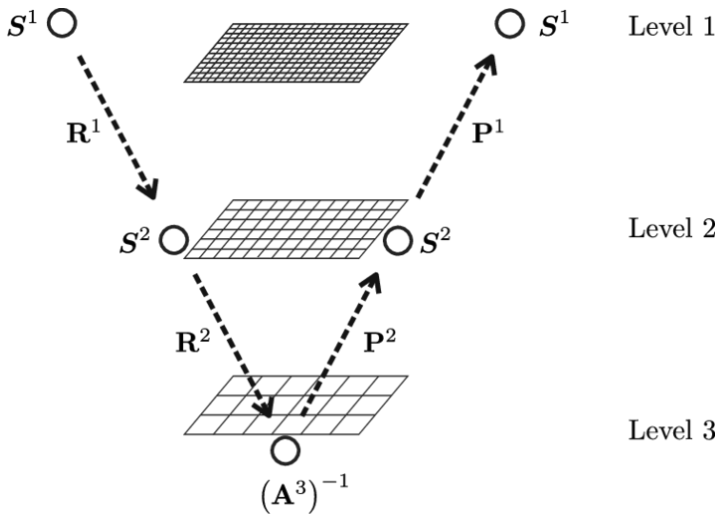
Domain Decomposition AMG (DDαAMG)



Algebraic Multigrid (AMG)

Domain Decomposition AMG (DDαAMG)



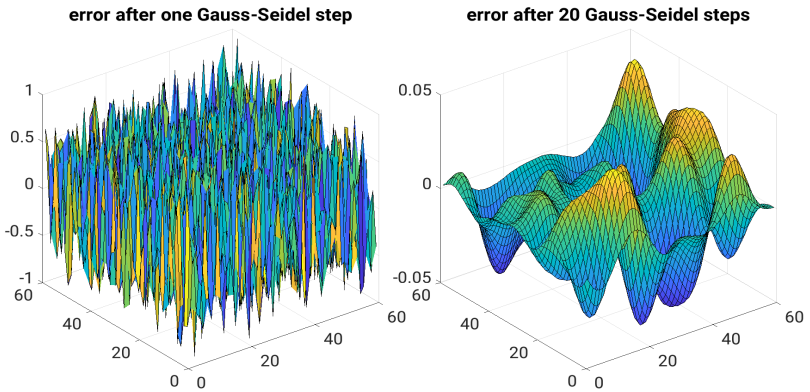


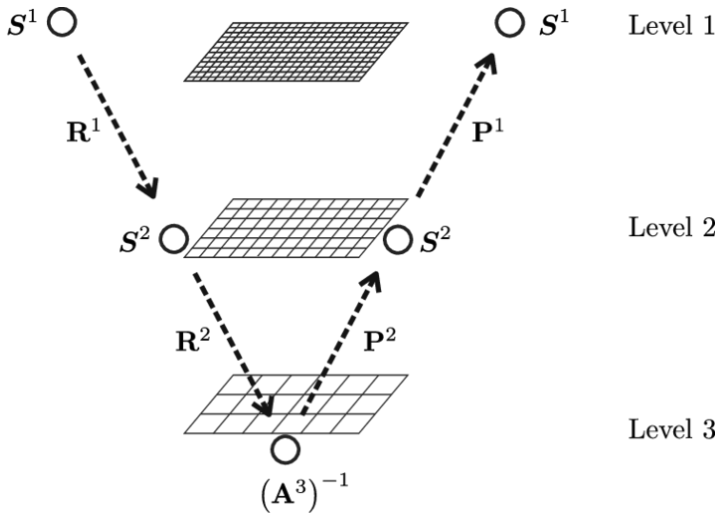
- **Coarse-grid operator:**

$$D_{\ell+1} = R_\ell D_\ell P_\ell, \quad R = P^H$$

Algebraic Multigrid (AMG)

Domain Decomposition AMG (DDαAMG)





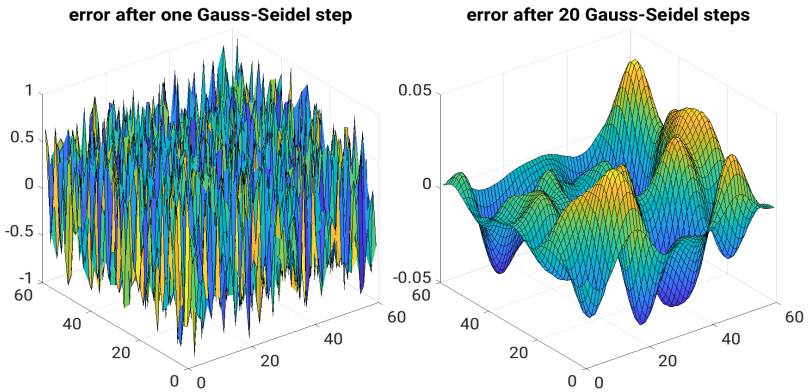
- **Coarse-grid operator:**

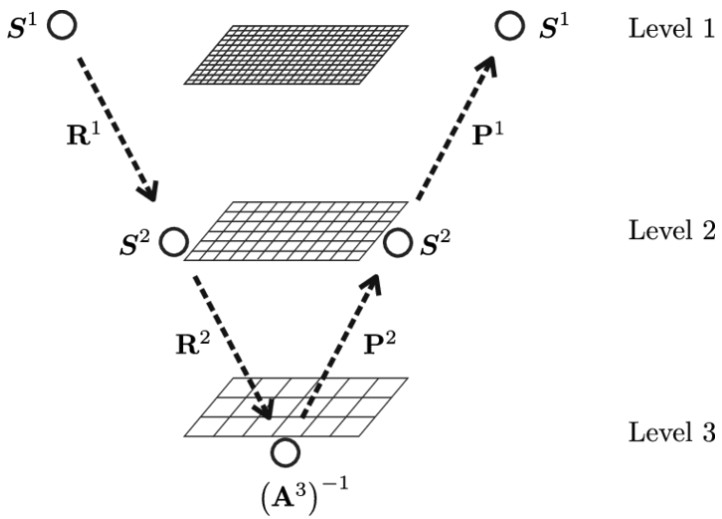
$$D_{\ell+1} = R_\ell D_\ell P_\ell, \quad R = P^H$$

- $\mathrm{range}(P)$ **contains the near kernel**

Algebraic Multigrid (AMG)

Domain Decomposition AMG (DDαAMG)





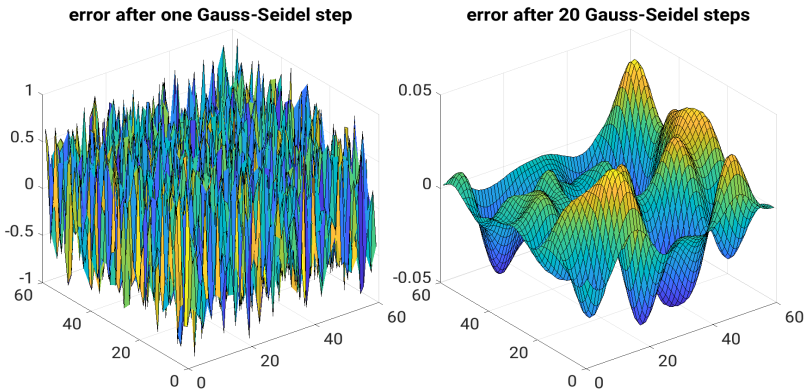
- **Coarse-grid operator:**

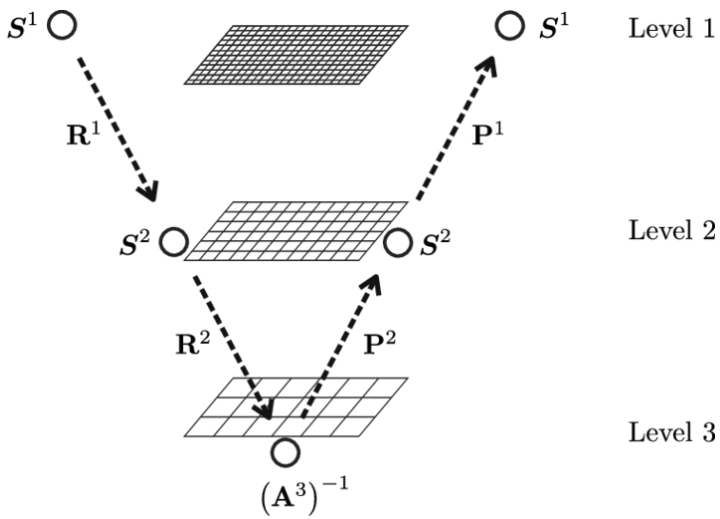
$$D_{\ell+1} = R_\ell D_\ell P_\ell, \quad R = P^H$$

- $\mathrm{range}(P)$ **contains the near kernel**
- P_ℓ from **few approx** eigenmodes

Algebraic Multigrid (AMG)

Domain Decomposition AMG (DDαAMG)





- **Coarse-grid operator:**

$$D_{\ell+1} = R_\ell D_\ell P_\ell, \quad R = P^H$$

- $\mathrm{range}(P)$ **contains the near kernel**
- P_ℓ from **few approx** eigenmodes
- $\widehat{P}_\ell = P_1 \cdots P_{\ell-1} \in \mathbb{C}^{n \times n_\ell}$
  $P_1 = I$

Multigrid Multilevel Monte Carlo

Deflation space for "free"

- **Idea:** since $P_l D_{l+1}^{-1} P_l^H \approx D_l^{-1}$

Multigrid Multilevel Monte Carlo

Deflation space for "free"

- **Idea:** since $P_l D_{l+1}^{-1} P_l^H \approx D_l^{-1}$

  $\rightarrow$ *use deflation* $M_l := D_l^{-1} - P_l D_{l+1}^{-1} P_l^H$

Multigrid Multilevel Monte Carlo

Deflation space for "free"

- **Idea:** since $P_l D_{l+1}^{-1} P_l^H \approx D_l^{-1}$

  $\rightarrow$ *use deflation* $M_l := D_l^{-1} - P_l D_{l+1}^{-1} P_l^H$

- Multigrid MLMC construction *Frommer et al. (2022)*

$$D^{-1} = \sum_{\ell=1}^{L-1} \left(\widehat{P}_\ell D_\ell^{-1} \widehat{R}_\ell - \widehat{P}_{\ell+1} D_{\ell+1}^{-1} \widehat{R}_{\ell+1} \right) + \widehat{P}_L D_L^{-1} \widehat{R}_L$$

Multigrid Multilevel Monte Carlo

Deflation space for "free"

- **Idea:** since $P_l D_{l+1}^{-1} P_l^H \approx D_l^{-1}$

  $\rightarrow$ *use deflation* $M_l := D_l^{-1} - P_l D_{l+1}^{-1} P_l^H$

- Multigrid MLMC construction *Frommer et al. (2022)*

$$D^{-1} = \sum_{\ell=1}^{L-1} \left(\widehat{P}_\ell D_\ell^{-1} \widehat{R}_\ell - \widehat{P}_{\ell+1} D_{\ell+1}^{-1} \widehat{R}_{\ell+1} \right) + \widehat{P}_L D_L^{-1} \widehat{R}_L$$

- $P_l^H P_l = I$ $\rightarrow$ series of smaller operators

$$\widehat{\mathrm{tr}}(D^{-1}) = \sum_{l=1}^{L-1} \widehat{\mathrm{tr}}(M_l) + \widehat{\mathrm{tr}}(D_L^{-1})$$

Applications of Multigrid MLMC

- Variant of this applied to two-point connected light-quark vector correlator *Gruber et al. (2025)*

Applications of Multigrid MLMC

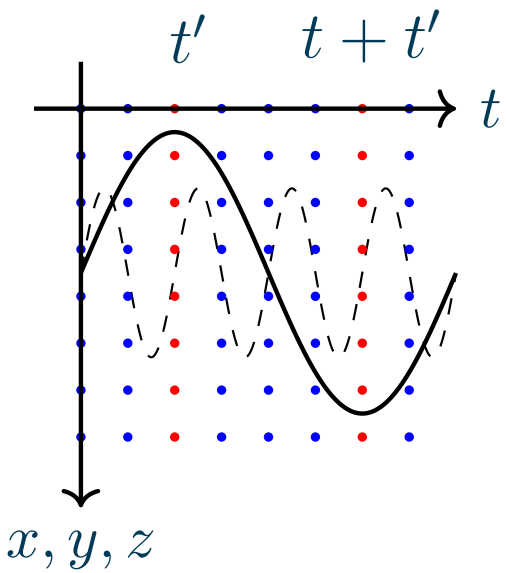
- Variant of this applied to two-point connected light-quark vector correlator *Gruber et al. (2025)*
- Test on **pseudoscalar connected two-point function**

$$G(t) = \frac{1}{T} \sum_{t'} \mathrm{tr}[D^{-1}(t+t', t') \, \Gamma_5 \, D^{-1}(t', t+t') \, \Gamma_5]$$

Applications of Multigrid MLMC

- Variant of this applied to two-point connected light-quark vector correlator *Gruber et al. (2025)*
- Test on **pseudoscalar connected two-point function**

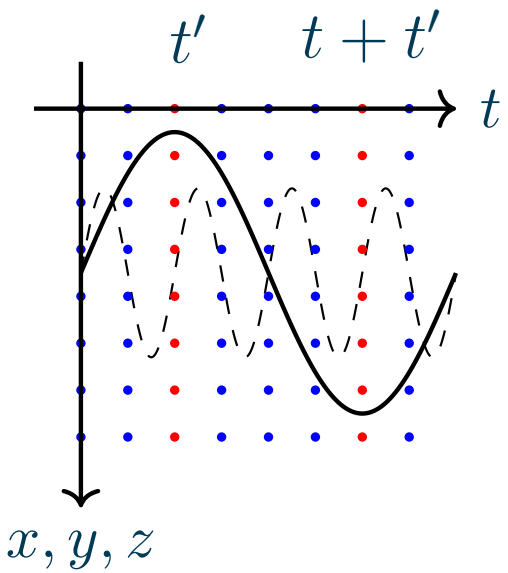
$$G(t) = \frac{1}{T} \sum_{t'} \mathrm{tr}[D^{-1}(t+t', t')\, \Gamma_5\, D^{-1}(t', t+t')\, \Gamma_5]$$



Applications of Multigrid MLMC

- Variant of this applied to two-point connected light-quark vector correlator *Gruber et al. (2025)*
- Test on **pseudoscalar connected two-point function**

$$G(t) = \frac{1}{T}\sum_{t'} \mathrm{tr}[D^{-1}(t+t',t')\,\Gamma_5\,D^{-1}(t',t+t')\,\Gamma_5]$$



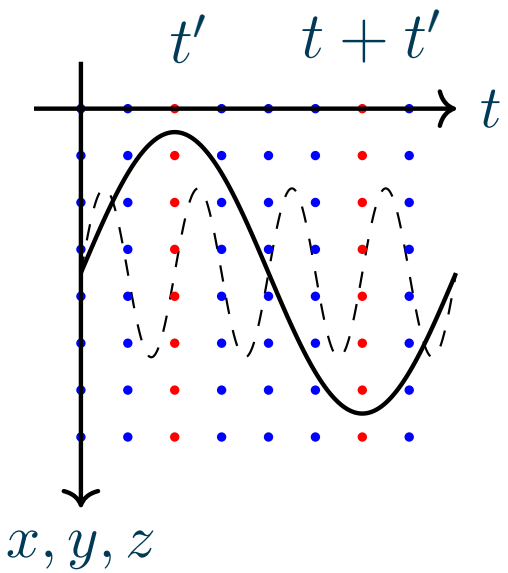
Insert:

$$D^{-1} = \sum_{\ell=1}^{L-1}\left(\widehat{P}_\ell D_\ell^{-1}\widehat{R}_\ell - \widehat{P}_{\ell+1}D_{\ell+1}^{-1}\widehat{R}_{\ell+1}\right) + \widehat{P}_L D_L^{-1}\widehat{R}_L$$

Applications of Multigrid MLMC

- Variant of this applied to two-point connected light-quark vector correlator *Gruber et al. (2025)*
- Test on **pseudoscalar connected two-point function**

$$G(t) = \frac{1}{T} \sum_{t'} \mathrm{tr}[D^{-1}(t+t', t')\, \Gamma_5\, D^{-1}(t', t+t')\, \Gamma_5]$$



Insert:

$$D^{-1} = \sum_{\ell=1}^{L-1} \left(\widehat{P}_\ell D_\ell^{-1} \widehat{R}_\ell - \widehat{P}_{\ell+1} D_{\ell+1}^{-1} \widehat{R}_{\ell+1} \right) + \widehat{P}_L D_L^{-1} \widehat{R}_L$$

$$\rightarrow G(t) = \sum_{i,j=1}^{L} G_{i,j}(t)$$

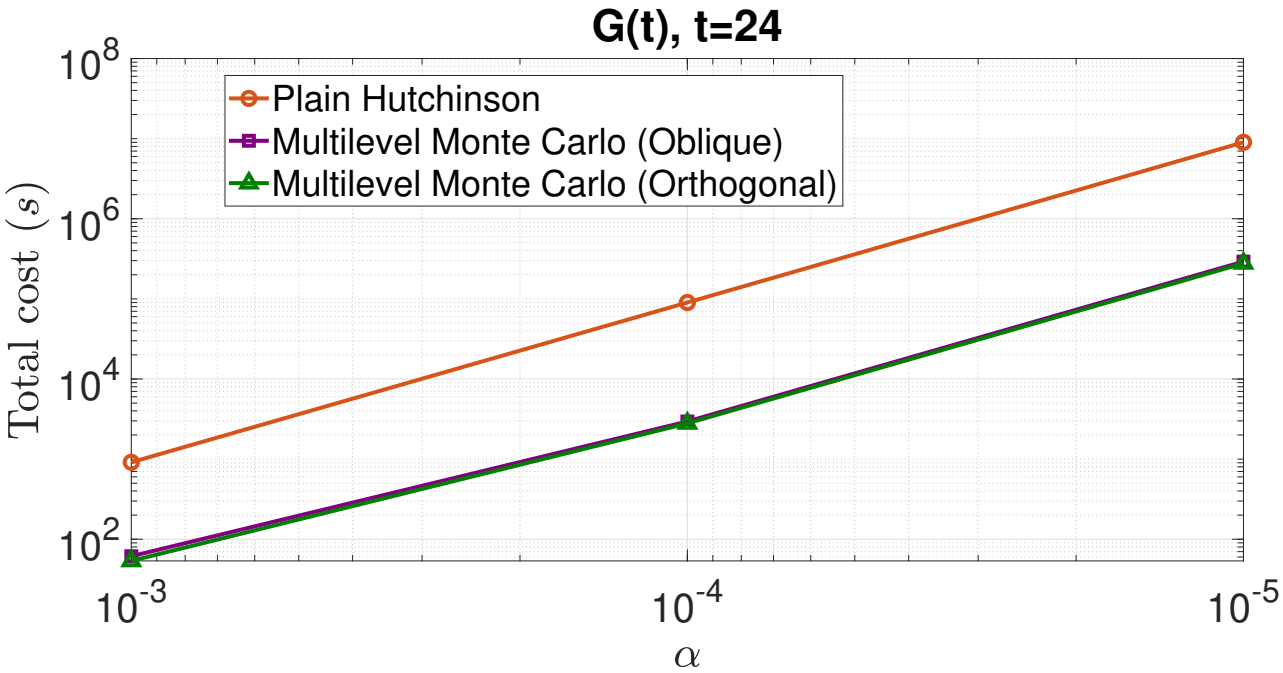
Applications of Multigrid MLMC

Variance and cost reduction — connected two-point function

$$G(t) = \sum_{i,j=1}^{L} G_{i,j}(t)$$

Frommer et al. (2025)

`arxiv: 2509.11424`



$64 \times 32^3, \quad m_\pi = 295\,\text{MeV}, \quad a = 0.071\,\text{fm}$

Applications of Multigrid MLMC

Variance and cost reduction — connected two-point function

$$G(t) = \sum_{i,j=1}^{L} G_{i,j}(t)$$

Frommer et al. (2025)

`arxiv: 2509.11424`

Question:

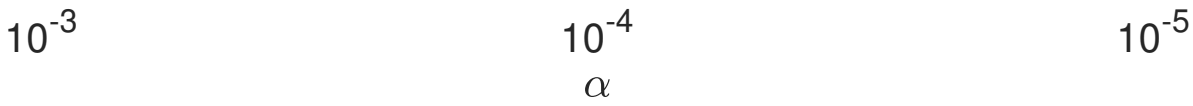
What about local-in-time operators?

Answer:

This does not work

Alternative:

Probing



$64 \times 32^3, \quad m_\pi = 295\,\text{MeV}, \quad a = 0.071\,\text{fm}$

Probing (for a time-slice)

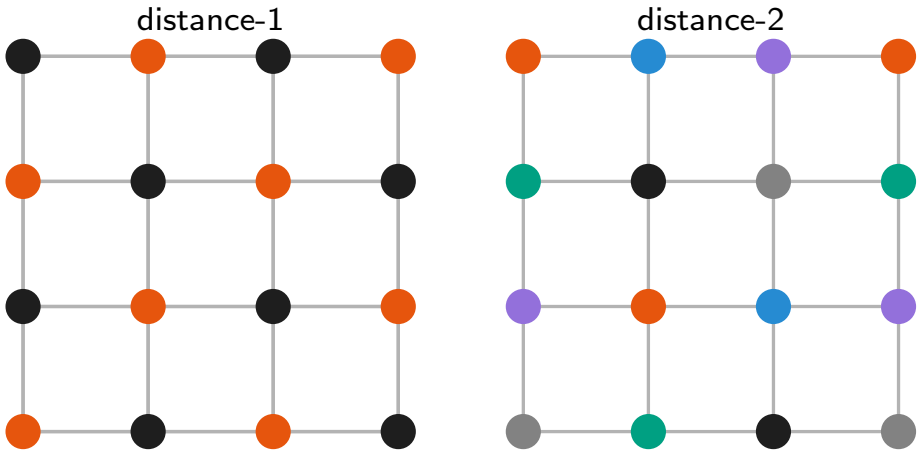
with distance-d coloring of the graph associated to D

samples $N \propto \dfrac{\|\mathrm{offdiag}(A^{-1})\|_F^2}{\epsilon^2}$

Probing (for a time-slice)

with distance-d coloring of the graph associated to D

samples $N \propto \dfrac{\|\text{offdiag}(A^{-1})\|_F^2}{\epsilon^2}$

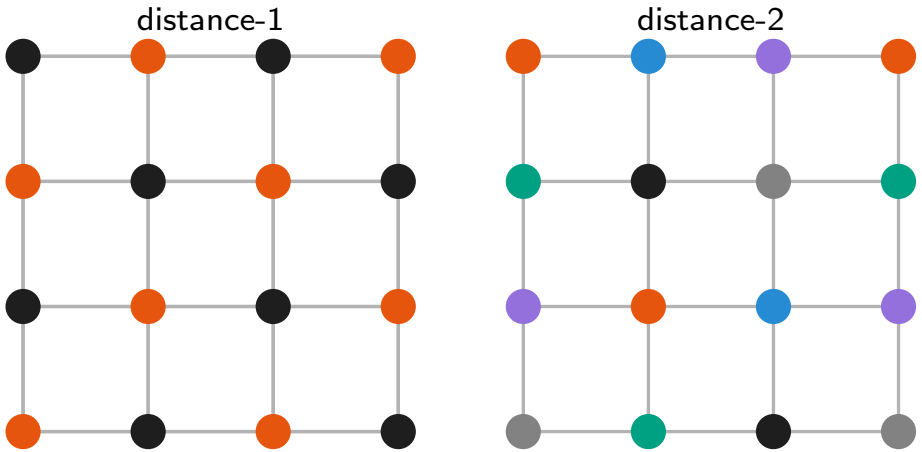


$v = [\pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0]$

Probing (for a time-slice)

with distance-d coloring of the graph associated to D

samples $N \propto \dfrac{\|\mathrm{offdiag}(A^{-1})\|_F^2}{\epsilon^2}$



$v = [\pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0]$

Hierarchical probing

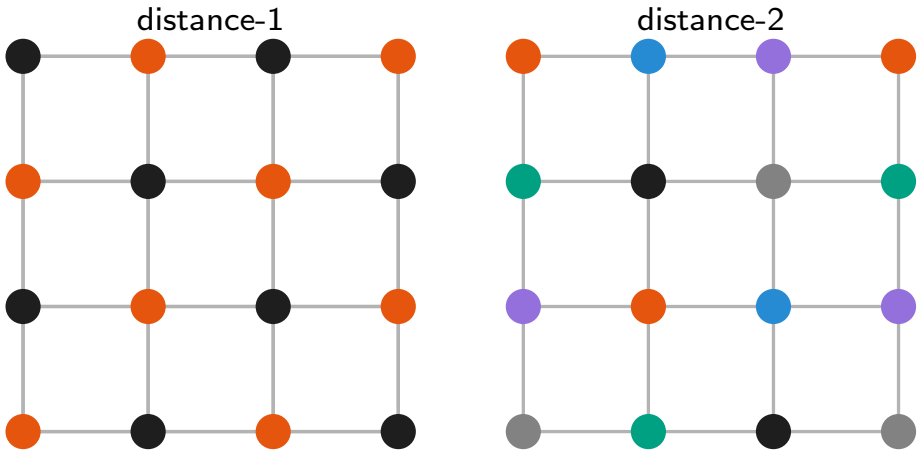
Stathopoulos et al. (2013)

$$n_{\mathrm{HP}}(d) = 2^{\mathrm{Dim}\left\lceil \log_2\left(2d+1\right)\right\rceil}$$

Probing (for a time-slice)

with distance-d coloring of the graph associated to D

samples $N \propto \dfrac{\|\mathrm{offdiag}(A^{-1})\|_F^2}{\epsilon^2}$



$v = [\pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0]$

Hierarchical probing

Stathopoulos et al. (2013)

$$n_{\mathrm{HP}}(d) = 2^{\mathrm{Dim}\left\lceil \log_2\left(2d+1\right)\right\rceil}$$

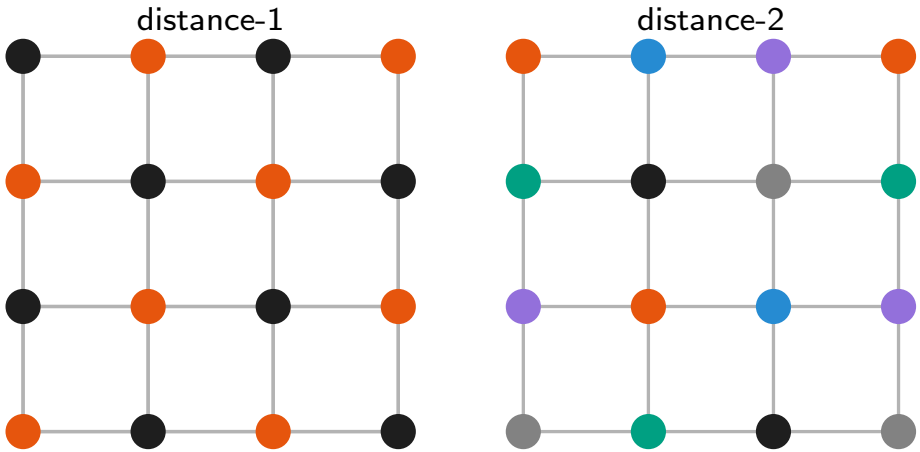
Alternative: Target coloring to 3D torus

$$\mathrm{col}(x) = \left(\sum_{i=1}^{D} \sigma_i x_i\right) \bmod n_c$$

Probing (for a time-slice)

with distance-d coloring of the graph associated to D

samples $N \propto \dfrac{\|\text{offdiag}(A^{-1})\|_F^2}{\epsilon^2}$



$v = [\pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0]$

Hierarchical probing

Stathopoulos et al. (2013)

$$n_{\text{HP}}(d) = 2^{\text{Dim}\left\lceil \log_2\left(2d+1\right)\right\rceil}$$

Alternative: Target coloring to 3D torus

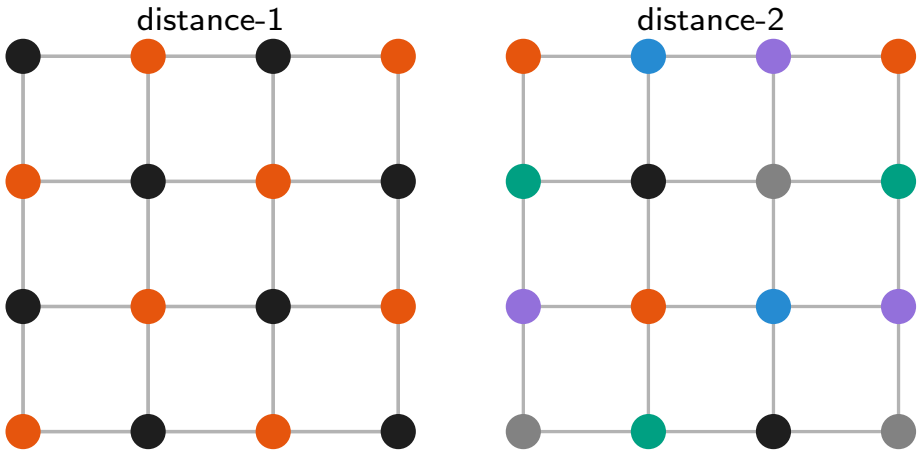
$$\text{col}(x) = \left(\sum_{i=1}^{D} \sigma_i x_i\right) \bmod n_c$$

d	$n_c^{(\text{HP})}$ $(Dim = 3)$	n_c
1	2	2
2	16	8
3	–	16
4	128	32
5	–	88
6	–	128
7	–	176
8	1024	297

Probing (for a time-slice)

with distance-d coloring of the graph associated to D

samples $N \propto \dfrac{\|\mathrm{offdiag}(A^{-1})\|_F^2}{\epsilon^2}$



$v = [\pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0\ \pm 1\ 0]$

Hierarchical probing

Stathopoulos et al. (2013)

$$n_{\mathrm{HP}}(d) = 2^{\mathrm{Dim}\left\lceil \log_2\left(2d+1\right)\right\rceil}$$

Alternative: Target coloring to 3D torus

$$\mathrm{col}(x) = \left(\sum_{i=1}^{D} \sigma_i x_i\right) \bmod n_c$$

d	$n_c^{(\mathrm{HP})}$ $(Dim = 3)$	n_c
1	2	2
2	16	8
3	–	16
4	128	32
5	–	88
6	–	128
7	–	176
8	1024	297

Local: Dilute DoF!

Applications of probing

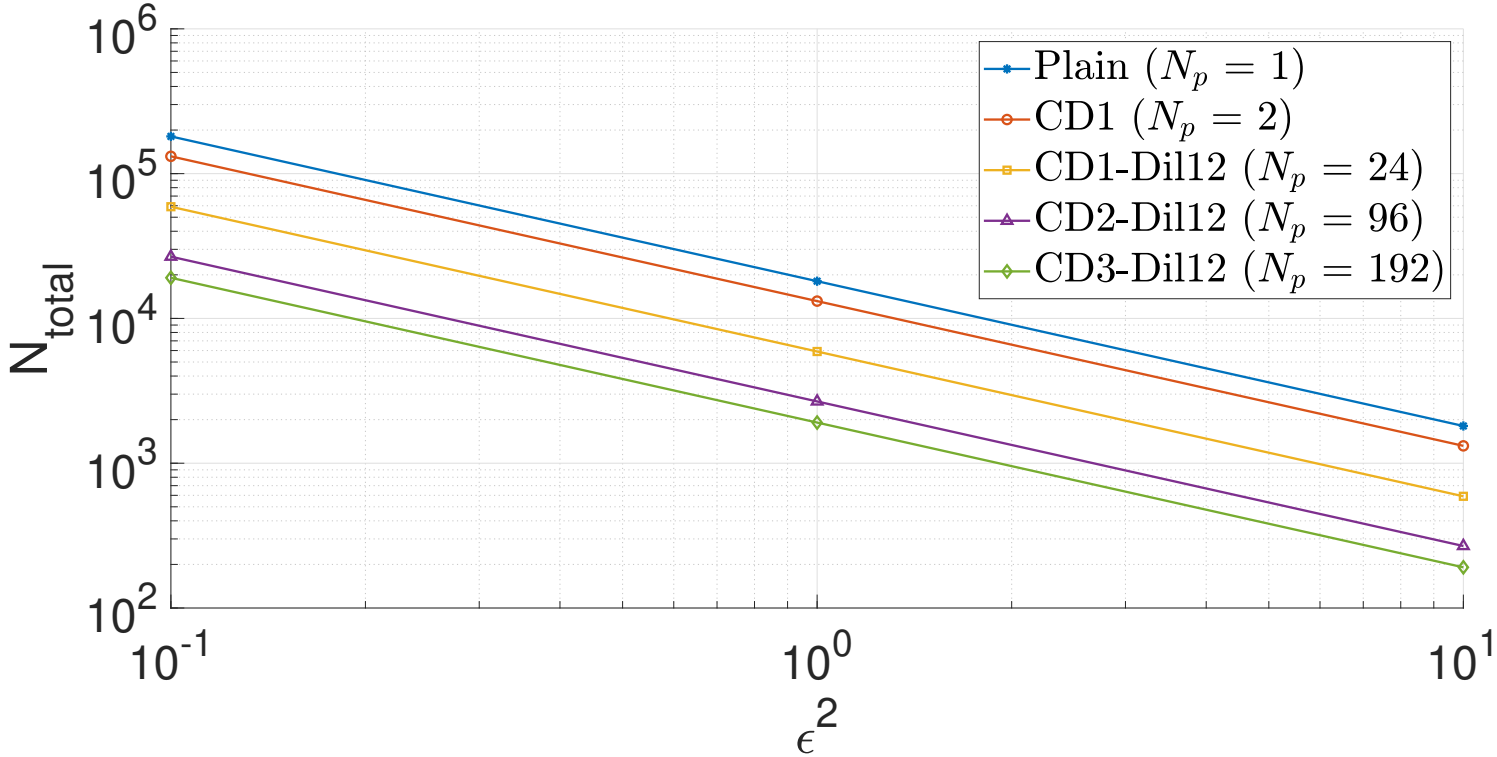
Disconnected 2-point correlation function

$$C(\tau) = \left\langle \frac{1}{T} \sum_{t=0}^{T-1} O(t) O(t+\tau) \right\rangle_g \qquad O(t) = \mathrm{tr}[\Gamma_5(t,t) D^{-1}(t,t)]$$

Applications of probing

Disconnected 2-point correlation function

$$C(\tau) = \left\langle \frac{1}{T} \sum_{t=0}^{T-1} O(t) O(t+\tau) \right\rangle_g \qquad O(t) = \mathrm{tr}[\Gamma_5(t,t) D^{-1}(t,t)]$$



Conclusions and Outlook

Cost reduction for both methods:
factor of 10 (variance reduction: $\sim 10^4$)

Conclusions and Outlook

Cost reduction for both methods:
factor of 10 (variance reduction: $\sim 10^4$)

- **Large distances:**
  Use **Multigrid MLMC**

Conclusions and Outlook

Cost reduction for both methods:
factor of 10 (variance reduction: $\sim 10^4$)

- **Large distances:**
  Use **Multigrid MLMC**
- **On a time slice:**
  Use **probing** with coloring targeted to 3D torus

Conclusions and Outlook

Cost reduction for both methods:
factor of 10 (variance reduction: $\sim 10^4$)

- **Large distances:**
  Use **Multigrid MLMC**
- **On a time slice:**
  Use **probing** with coloring targeted to 3D torus
- **Still to explore:**
  - **Probing:** Compare 3D torus coloring with Hierarchical probing

Conclusions and Outlook

Cost reduction for both methods:
factor of 10 (variance reduction: $\sim 10^4$)

- **Large distances:**
  Use **Multigrid MLMC**
- **On a time slice:**
  Use **probing** with coloring targeted to 3D torus
- **Still to explore:**
  - **Probing:** Compare 3D torus coloring with Hierarchical probing
  - **Multigrid MLMC:** Use exact/more low eigenmodes?

Conclusions and Outlook

Cost reduction for both methods:
factor of 10 (variance reduction: $\sim 10^4$)

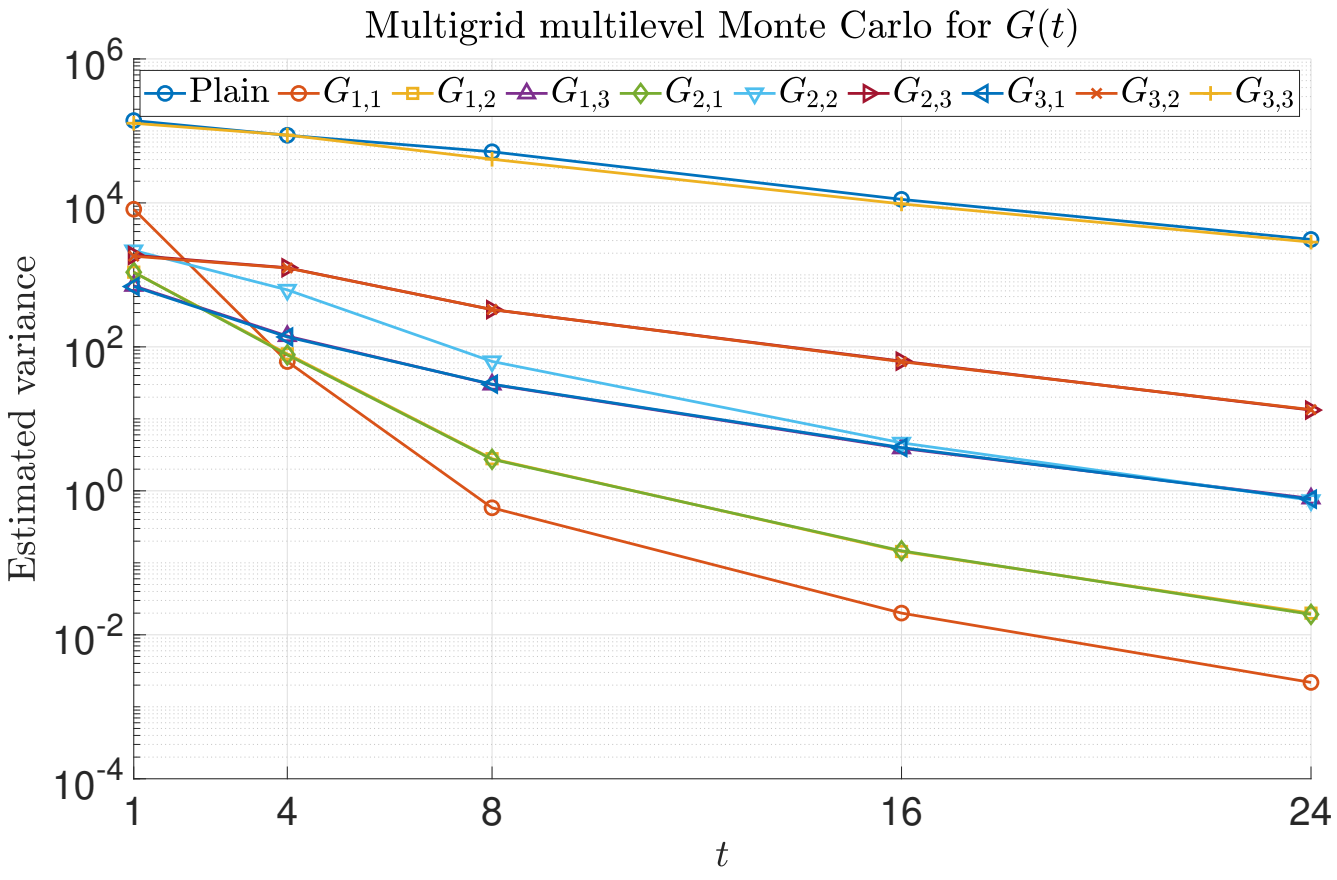
- **Large distances:**
  Use **Multigrid MLMC**
- **On a time slice:**
  Use **probing** with coloring targeted to 3D torus
- **Still to explore:**
  - **Probing:** Compare 3D torus coloring with Hierarchical probing
  - **Multigrid MLMC:** Use exact/more low eigenmodes?
  - **Combine?** multigrid MLMC with new probing approach

Gracias!

- **A.S. Gambhir, A. Stathopoulos, K. Orginos** *Deflation as a method of variance reduction for estimating the trace of a matrix inverse*, *SIAM J. Sci. Comput.* (2017)
- **A. Stathopoulos, J. Laeuchli, K. Orginos** *Hierarchical Probing for Estimating the Trace of the Matrix Inverse on Toroidal Lattices*, *SIAM J. Sci. Comput.* (2013)
- **A. Frommer, M.N. Khalil, G. Ramírez-Hidalgo** *A Multilevel Approach to Variance Reduction in the Stochastic Estimation of the Trace of a Matrix*, *SIAM J. Sci. Comput.* (2022)
- **R. Gruber, T. Harris, M.K. Marinković** *Multigrid Low-Mode Averaging*, *Phys. Rev. D* (2025)
- **A. Frommer, J.M. Jimenez-Merchan, F. Knechtli, T. Korzec, G. Ramírez-Hidalgo** *Using orthogonal projectors in multigrid multilevel Monte Carlo for trace estimation in lattice QCD*, *arXiv:2509.11424* (2025)

Applications of Multigrid MLMC

Variance point of view



Remember... cost reduction: **10x**