Multigrid Methods for Linear Systems with Stochastic Entries Arising in Lattice QCD

Andreas Frommer

SFB TRR55
SONDERFORSCHUNGSBEREICH
HADRON PHYSICS
FROM LATTICE QCD

MATHEMATICAL MODELLING,
ANALYSIS AND
COMPUTATIONAL MATHEMATICS

BERGISCHE
UNIVERSITÄT
WUPPERTAL

Collaborators

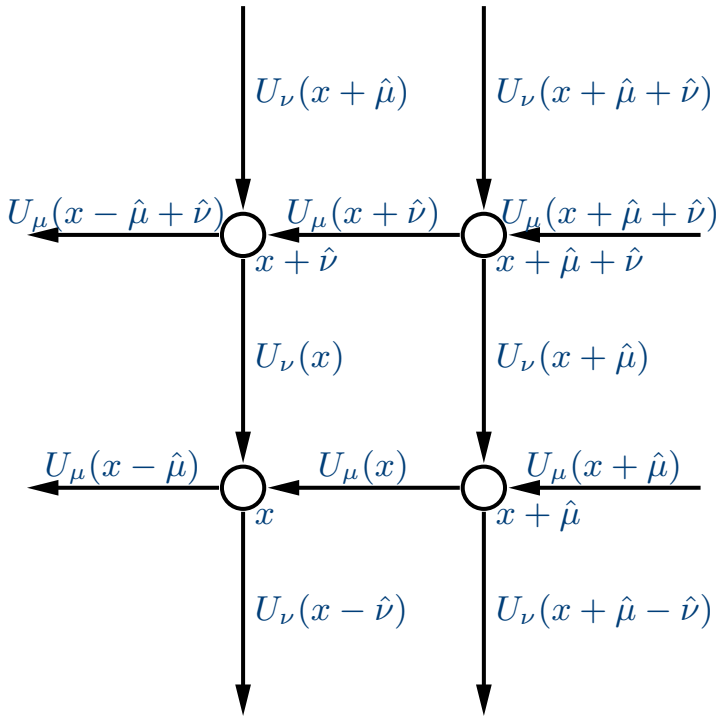
- James Brannick, *Penn State University*
- Björn Leder, *Humboldt Universität Berlin*
- Stephan Krieg, *Jülich Research Centre*
- Karsten Kahl, Matthias Rottmann, Artur Strebel, *Universität Wuppertal*

BERGISCHE UNIVERSITÄT WUPPERTAL

What we learned so far

Setting:

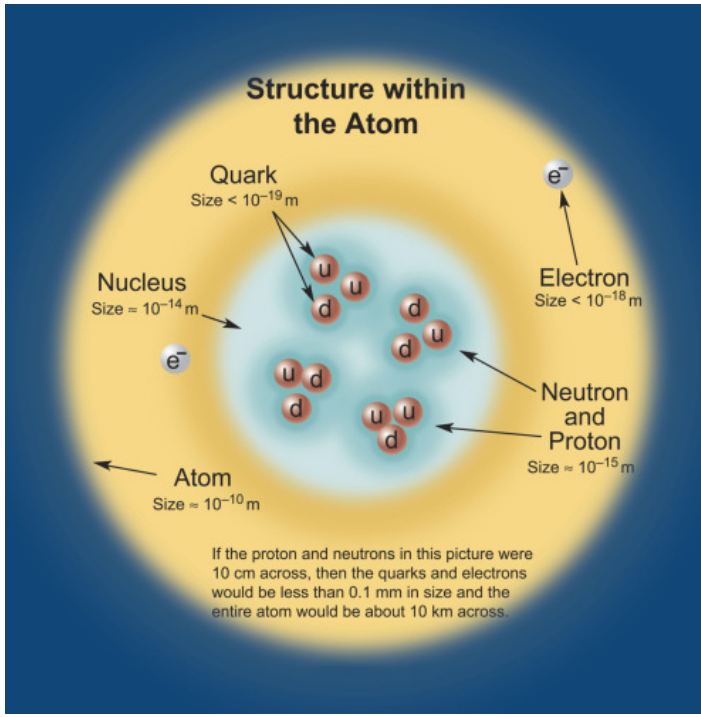
- Statistical mechanics: Observables are expected values of operators depending on random variables
- Lattice QCD: 4d lattice $\mathcal{L}$ in space-time
- We need configurations $\mathcal{U} = \{U_\mu(x) : x \in \mathcal{L}, \mu = 1, \ldots, 4\}$ of gauge links
- These are obtained from fixed point distribution via Markov chain MC
- Computationally: importance sampling via HMC



BERGISCHE UNIVERSITÄT WUPPERTAL

QCD is

- the standard model of the strong interaction between quarks as the constituents of matter
- a quantum field theory
- described by the Dirac operator depending on a background field



Lattice QCD is

- a discretization of QCD
- requiring a discretized Dirac operator and ...
- ... a discrete version of the background field: a configuration of gauge links

BERGISCHE UNIVERSITÄT WUPPERTAL

The Dirac operator

$$(\mathcal{D} + mI)\psi = \eta,$$

where

- $\psi = \psi(x)$ and $\eta = \eta(x)$ represent quark fields
- $x = (x_0, x_1, x_2, x_3)$ point in space-time
- m (implicitly) sets the quark mass
- gluons are represented in the Dirac operator $\mathcal{D}$

$$\mathcal{D} = \sum_{\mu=0}^{3} \gamma_\mu \otimes (\partial_\mu + A_\mu)$$

BERGISCHE UNIVERSITÄT WUPPERTAL

$$\mathcal{D} = \sum_{\mu=0}^{3} \gamma_\mu \otimes (\partial_\mu + A_\mu)$$

- $\partial_\mu = \partial/\partial x_\mu$
- A is the gluon (background) gauge field, $A_\mu(x) \in \mathfrak{su}(3)$ (3×3, skew-Hermitian and traceless)
- γ_μ generators of Clifford algebra

$$\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2 \cdot \delta_{\mu\nu} \quad \text{for } \mu, \nu = 0, 1, 2, 3.$$

Consequences:

- $\psi(x) \in \mathbb{C}^{12} \equiv \mathbb{C}^{3\times 4}$, 3 colors, 4 spins
- $\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3$ satisfies

$$\gamma_5\gamma_\mu = -\gamma_\mu\gamma_5, \quad \mu = 0, 1, 2, 3.$$

BERGISCHE UNIVERSITÄT WUPPERTAL

The Dirac-Wilson matrix D_W

Idea: Approximate via covariate centralized finite differences and regularize with 2nd order term.

- $\psi = \psi(x)$ with $x \in \mathcal{L}$ discrete quark field
- $U_\mu(x) \in SU(3)$ (unitary, $\det(U_\mu(x)) = 1$)

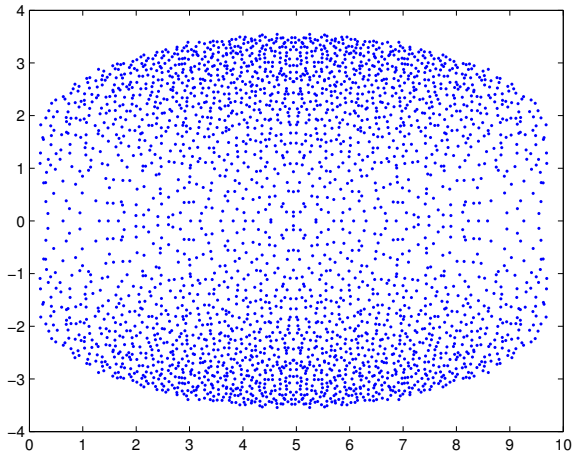
$$
\begin{aligned}
(D_W\psi)(x) \;=\; & \frac{m_0+4}{a}\psi(x) - \frac{1}{2a}\sum_{\mu=0}^{3}\left((I_4-\gamma_\mu)\otimes U_\mu(x)\right)\psi(x+\hat{\mu}) \\
& - \frac{1}{2a}\sum_{\mu=0}^{3}\left((I_4+\gamma_\mu)\otimes U_\mu^H(x-\hat{\mu})\right)\psi(x-\hat{\mu}),
\end{aligned}
$$

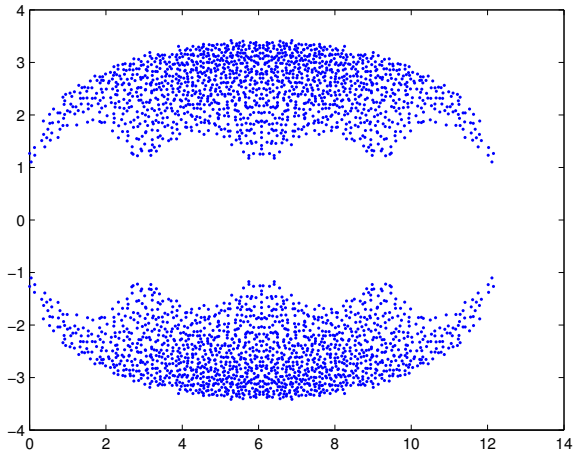
m_0 sets the quark mass

BERGISCHE UNIVERSITÄT WUPPERTAL

Properties of the Dirac-Wilson matrix D_W

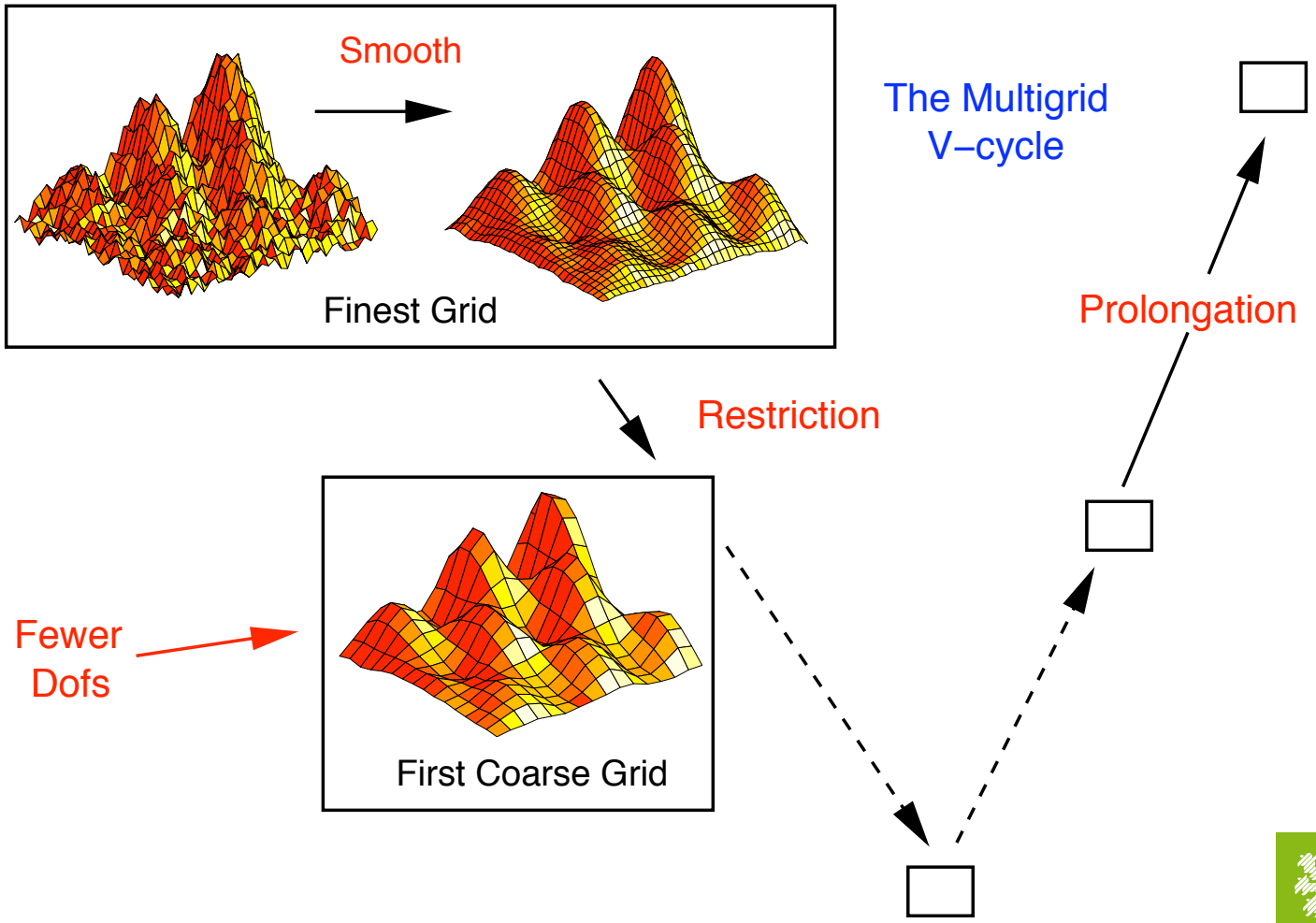
- D_W is non-normal
- spectral gaps
- m_0 is set s.t. spectrum approaches the imaginary axis from the right
- $\gamma_5 D_W$ is Hermitian
- (λ, x) right eigenpair $\Leftrightarrow$ $(\overline{\lambda}, \gamma_5 x)$ left eigenpair
- $\mathcal{L}$ of size 64^4: $D_W \in \mathbb{C}^{192M \times 192M}$





GISCHE
VERSITÄT
PPERTAL

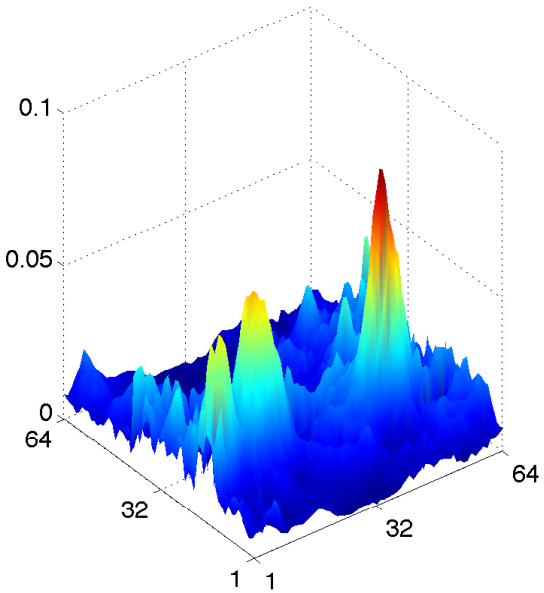
Multigrid in a Nutshell

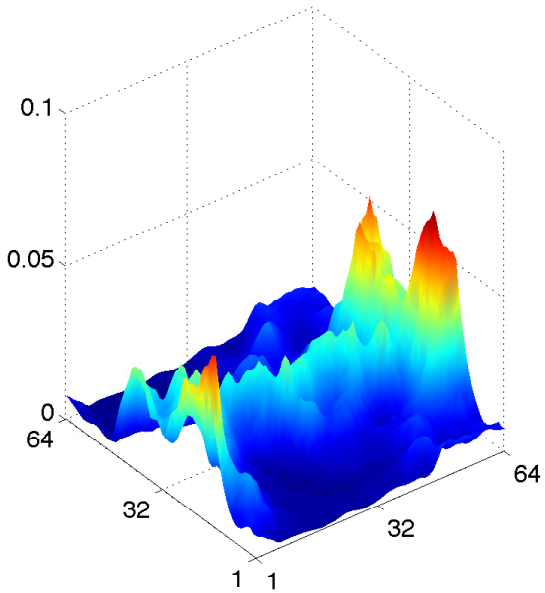


Challenges for Wilson-Dirac I

The Wilson-Dirac operator is special:

- We only know the gauge links $U_\mu(x)$ for a given lattice
- These are random variables, not just perturbed entries
- D_W has a non-trivial symmetry: $\gamma_5 D_W = D_W^\dagger \gamma_5$
- Eigenvectors to small eigenvalues are not geometrically smooth





BERGISCHE UNIVERSITÄT WUPPERTAL

Challenges for Wilson-Dirac II

Consequences:

- Smoothing ($=$ reducing large eigenmodes) is easy (Jacobi, GMRES, ...)
- Coarse grid system D_W^c must be constructed algebraically
- Using a Galerkin ansatz: $D_W^c = P^\dagger D_W P$ with a prolongation operator P, we aim at
  - preserving γ_5 symmetry
  - preserving locality
  - range(P) should approximate eigenvectors to small eigenvalues

There is hope: Local coherence of small eigenmodes [Lüscher 2008]

BERGISCHE UNIVERSITÄT WUPPERTAL

Two-grid error propagator for ν steps of pre-smoothing

$$E_{2g}^{(\nu)} = \underbrace{(I - PD_c^{-1}P^\dagger D)}_{\text{coarse grid correction}} \underbrace{(I - MD)^\nu}_{\text{smoother}}, \quad \underbrace{D_c := P^\dagger DP}_{\text{coarse operator}}$$

- low accuracy for D_c^{-1} and M is sufficient
- introduce recursive construction for $D_c \rightarrow$ multigrid

To Do: Define interpolation P and smoother M

DD-αAMG

M: Schwarz Alternating Procedure (SAP)

P: Aggregation Based Interpolation

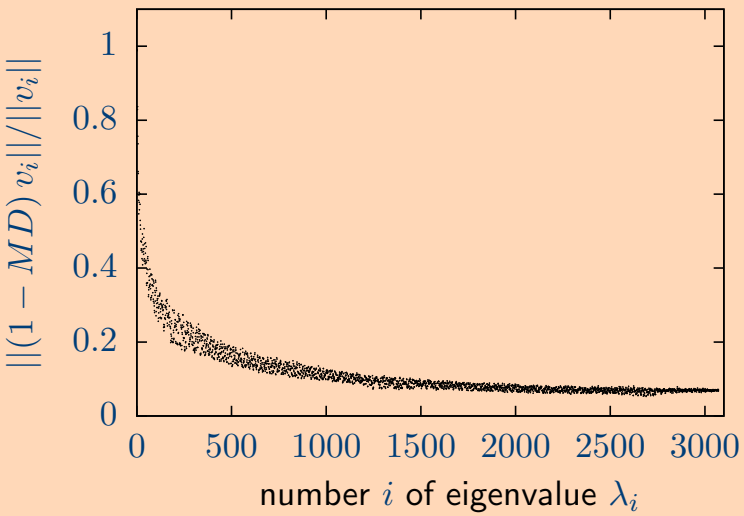
BERGISCHE UNIVERSITÄT WUPPERTAL

Preview: The multigrid principle for Dirac-Wilson

Smoother: $I - MD$

- Effective on "large" eigenvectors
- "small" eigenvectors remain

SAP



$Dv_i = \lambda_i v_i$ with $|\lambda_1| \leq \ldots \leq |\lambda_{3072}|$

BERGISCHE UNIVERSITÄT WUPPERTAL

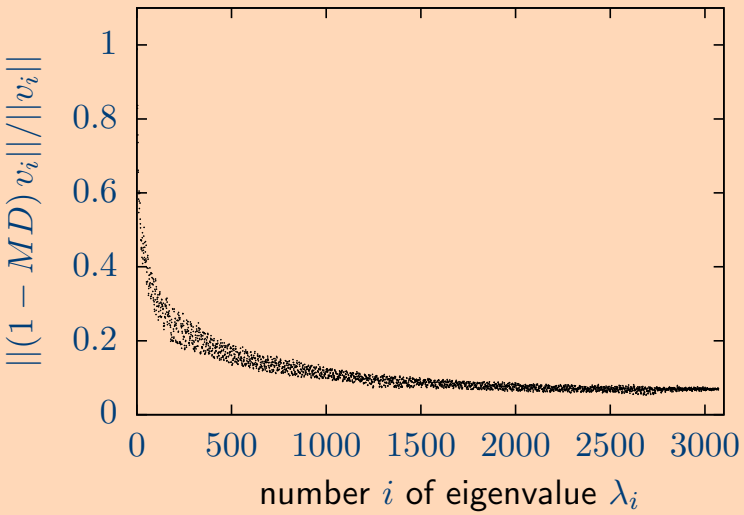
Preview: The multigrid principle for Dirac-Wilson

Coarse-grid correction: $I - P D_c^{-1} P^\dagger D$

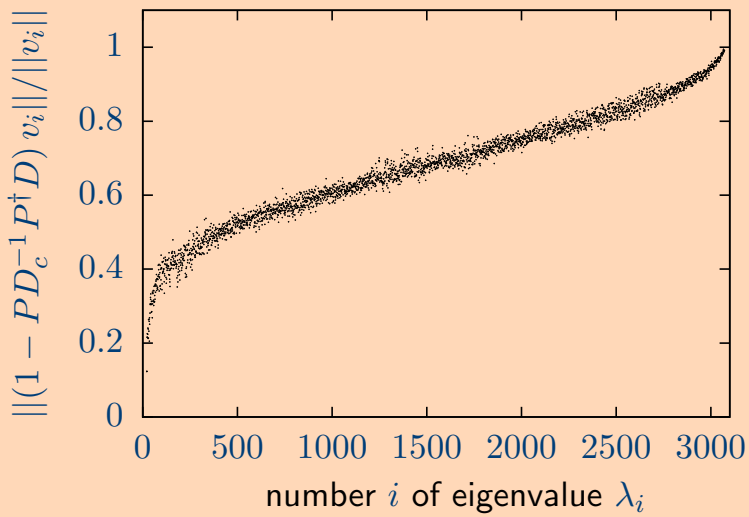
- small eigenvectors built into interpolation P

  $\Rightarrow$ Effective on small eigenvectors

SAP



Aggregation + Low Modes



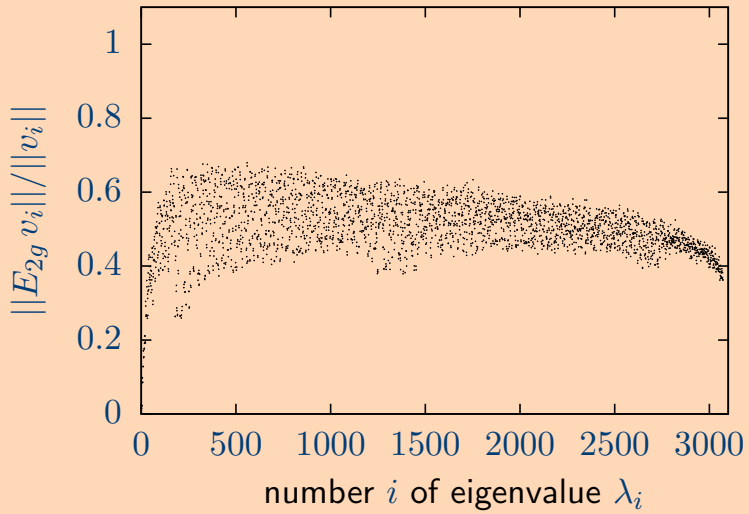
$Dv_i = \lambda_i v_i$ with $|\lambda_1| \leq \ldots \leq |\lambda_{3072}|$

Preview: The multigrid principle for Dirac-Wilson

Two-grid method: $E_{2g} = (I - PD_c^{-1}P^\dagger(I - MD)^\nu D)$

- Complementarity of smoother and coarse-grid correction
- Effective on all eigenvectors!

DD-αAMG

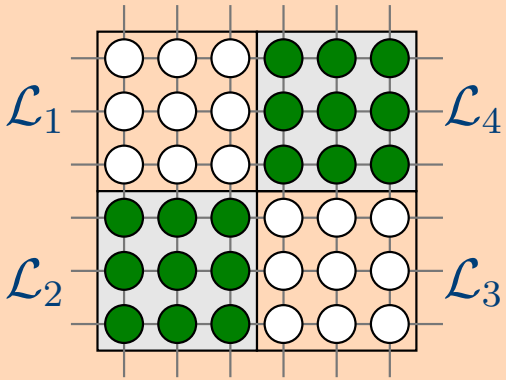


$Dv_i = \lambda_i v_i$ with $|\lambda_1| \leq \ldots \leq |\lambda_{3072}|$

BERGISCHE UNIVERSITÄT WUPPERTAL

SAP: Schwarz alternating procedure

Two color decomposition of $\mathcal{L}$



- canonical injections
  $$\mathcal{I}_{\mathcal{L}_i} : \mathcal{L}_i \to \mathcal{L}$$
- block restrictions
  $$D_{\mathcal{L}_i} = \mathcal{I}^\dagger_{\mathcal{L}_i} D \mathcal{I}_{\mathcal{L}_i}$$
- block inverses
  $$B_{\mathcal{L}_i} = \mathcal{I}_{\mathcal{L}_i} D^{-1}_{\mathcal{L}_i} \mathcal{I}^\dagger_{\mathcal{L}_i}$$

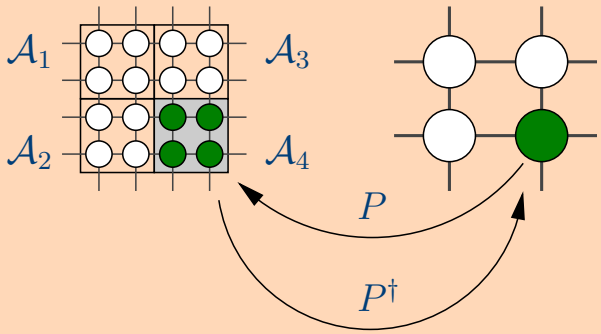
1: in: ψ, η, ν – out: ψ
2: **for** $k = 1$ to ν **do**
3: $\quad r \leftarrow \eta - D\psi$
4: $\quad$ **for all** *green* $\mathcal{L}_i$ **do**
5: $\quad\quad \psi \leftarrow \psi + B_{\mathcal{L}_i} r$
6: $\quad$ **end for**
7: $\quad r \leftarrow \eta - D\psi$
8: $\quad$ **for all** *white* $\mathcal{L}_i$ **do**
9: $\quad\quad \psi \leftarrow \psi + B_{\mathcal{L}_i} r$
10: $\quad$ **end for**
11: **end for**

BERGISCHE UNIVERSITÄT WUPPERTAL

Aggregation based interpolation

Construction:

- Define aggregates: domain decomposition $\mathcal{A}_1, ..., \mathcal{A}_s$



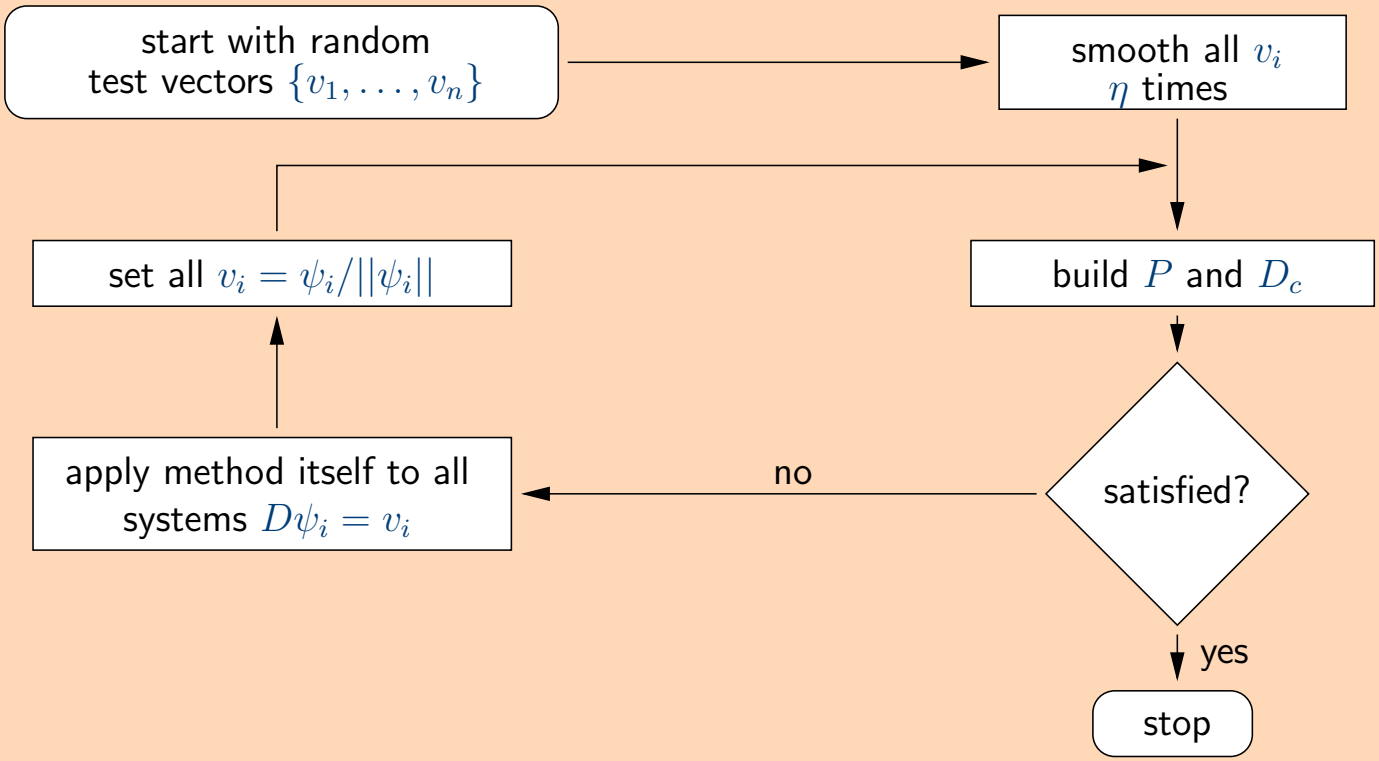
- Calculate test vectors $w_1, ..., w_N$
- Decompose test vectors over aggregates $\mathcal{A}_1, ..., \mathcal{A}_s$

$$(v^{(1)}, \ldots, v^{(k)}) = \left|\begin{matrix}\\ \\ \\ \\ \\ \end{matrix}\right| = \begin{matrix}\mathcal{A}_1\\ \mathcal{A}_2\\ \vdots\\ \mathcal{A}_s\end{matrix} \rightarrow P = \begin{pmatrix}\mathcal{A}_1 & & & \\ & \mathcal{A}_2 & & \\ & & \ddots & \\ & & & \mathcal{A}_s\end{pmatrix}$$

BERGISCHE UNIVERSITÄT WUPPERTAL

Setup: how to obtain test vectors

Bootstrapping process



BERGISCHE UNIVERSITÄT WUPPERTAL

Some history

- **Adaptive algebraic multigrid αAMG:** Brezina, Falgout, Manteuffel, MacLachlan, McCormick, Ruge 2004
- **Inexact deflation method:** Lüscher 2007.
  Solves
  $$D(I - PD_c^{-1}P^\dagger D)\psi = \eta$$
  using SAP as a preconditioner.
- **αAMG for lattice QCD:** Babich, Brannick, Brower, Clark, Manteuffel, McCormick, Osborn, Rebbi 2010.
- **DD-αAMG:** F., Kahl, Leder, Krieg, Rottmann 2013
- 2016: Alexandrou, Bacchio, Finkenrath, F., Kahl, Rottmann: extension to "twisted mass" Dirac operator
- currently: further improvemet for setup

Parameters

	parameter		default
setup	number of iterations	n_{inv}	6
	number of test vectors	N	20
	size of lattice-blocks for aggregates on level 1		4^4
	size of lattice-blocks for aggregates on level $\ell > 1$		2^4
	coarse system relative residual tolerance (stopping criterion for the coarse system)$^{(*)}$	ϵ	$5 \cdot 10^{-2}$
solver	restart length of FGMRES	n_{kv}	10
	relative residual tolerance (stopping criterion)	tol	10^{-10}
	number of post-smoothing steps$^{(*)}$	ν	5
	size of lattice-blocks in SAP$^{(*)}$		2^4
	number of Minimal Residual (MR) iterations to solve the local systems in SAP$^{(*)}$		3
K-cycle	maximal length$^{(*)}$		5
	maximal restarts$^{(*)}$		2
	relative residual tolerance (stopping criterion)$^{(*)}$		10^{-1}

$(*)$: same in solver and setup

Configurations

id	lattice size $N_t \times N_s^3$	pion mass m_π [MeV]	CGNR iterations	shift m_0	clover term c_{sw}	provided by
1	48×16^3	250	7,055	-0.095300	1.00000	BMW-c
2	48×24^3	250	11,664	-0.095300	1.00000	BMW-c
3	48×32^3	250	15,872	-0.095300	1.00000	BMW-c
4	48×48^3	135	53,932	-0.099330	1.00000	BMW-c
5	64×64^3	135	84,207	-0.052940	1.00000	BMW-c
6	128×64^3	270	45,804	-0.342623	1.75150	CLS
7	128×64^3	190	88,479	-0.33485	1.90952	CLS

Table: Ensembles used

Snapshots on performance: setup time vs solve time

number of setup steps n_{inv}	average setup timing	average iteration count	lowest iteration count	highest iteration count	average solver timing	average total timing
1	2.08	149	144	154	6.42	8.50
2	3.06	59.5	58	61	3.42	6.48
3	4.69	34.5	33	36	2.37	7.06
4	7.39	27.2	27	28	1.95	9.34
5	10.8	24.1	24	25	1.82	12.6
6	14.1	23.0	23	23	1.89	16.0
8	19.5	22.0	22	22	2.02	21.5
10	24.3	22.5	22	23	2.31	26.6

Table: Evaluation of DD-αAMG-setup($n_{inv}, 2$), 48^4 lattice, configuration id 4), $2{,}592$ cores, averaged over 20 runs.

BERGISCHE UNIVERSITÄT WUPPERTAL

Comparison of 2, 3 & 4 Level DD-αAMG

	configuration	4	5	6	7
	lattice size	48×48^3	64×64^3	128×64^3	128×64^3
	pion mass m_π	135 MeV	135 MeV	270 MeV	190 MeV
two levels	setup time	316s	736s	630s	701s
	solve time	48.6s	130s	113s	141s
three levels	setup time	374s	744s	719s	948s
	solve time	42.6s	75.2s	74.0s	79.0s
four levels	setup time	–	806s	755s	1,004s
	solve time	–	79.8s	75.7s	79.1s
	processes	81	128	256	256
local lattice	level 1	16×16^3	32×16^3	32×16^3	32×16^3
	level 2	4×4^3	8×4^3	8×4^3	8×4^3
	level 3	2×2^3	4×2^3	4×2^3	4×2^3
	level 4	–	2×1^3	2×1^3	2×1^3

- Conf. 1 slight speed up by using 3rd level

Snapshots on performance: oe-BiCGStab vs DD-αAM

	BiCGStab	DD-αAMG	speed-up factor	coarse grid
setup time		22.9s		
solve iter	13,450	21		$3{,}716^{(*)}$
solve time	91.2s	3.15s	29.0	2.43s
total time	91.2s	26.1s	3.50	

Table: BiCGStab vs. DD-αAMG with default parameters, configuration id 5, 8,192 cores, $(*)$: coarse grid iterations summed up over all iterations on the fine grid.

BERGISCHE UNIVERSITÄT WUPPERTAL

Snapshots on performance: mass scaling and levels

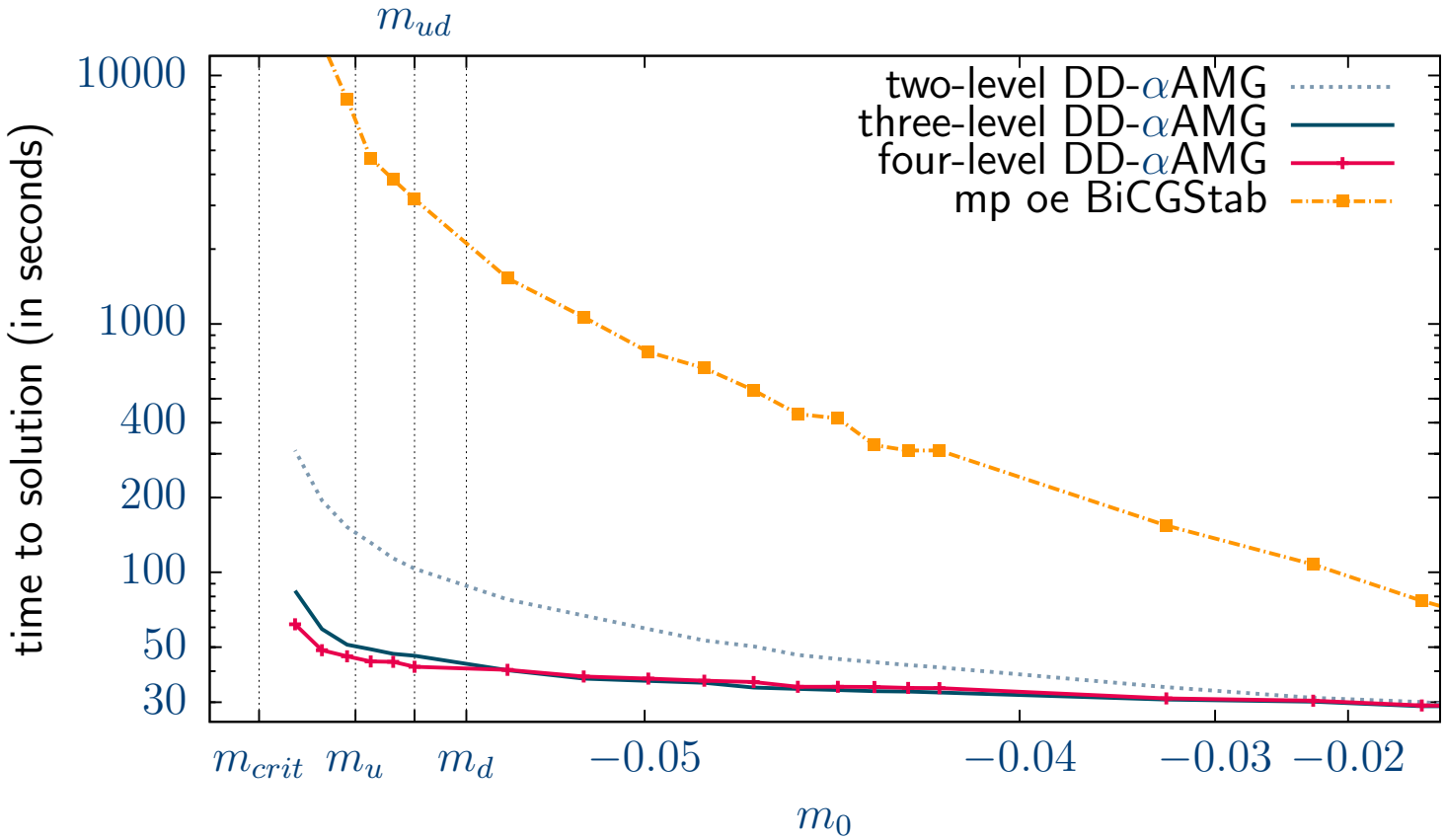


Figure: Mass scaling of 2, 3 and 4 level DD-αAMG, 64^4 lattice, configuration id 5, restart length $n_{kv} = 10$, 128 cores

BERGISCHE UNIVERSITÄT WUPPERTAL

Snapshots on performance: mass scaling and levels

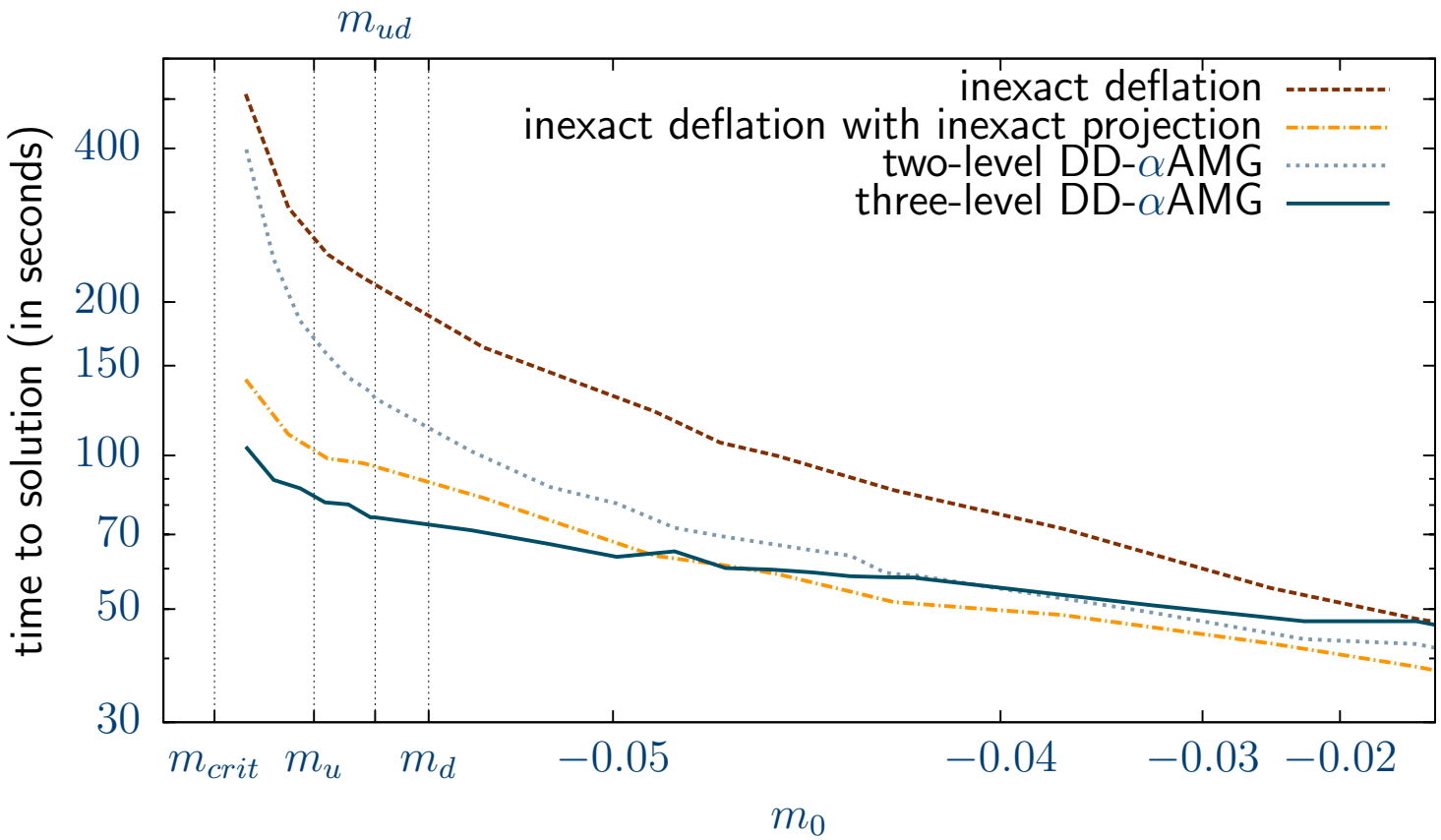


Figure: Mass scaling of 2, 3 and 4 level DD-αAMG, 64^4 lattice, configuration id 5, restart length $n_{kv} = 10$, 128 cores

BERGISCHE UNIVERSITÄT WUPPERTAL

Conclusions

- Stochastic entries require algebraic multigrid
- Adaptive setup is mandatory
- Setup is relatively costly
- Aggregation based coarsening justified because of local coherence
- Other adaptive setups are possible: bootstrap AMG [Brandt, Brannick, Kahl, Livshits 2011]
- Underlying assumption: a small set of small eigenmodes locally represents all small eigenmodes, at least approximately