Multigrid for Lattice QCD

– Solvers –

Andreas Frommer

Bergische Universität Wuppertal

June 24, 2014

Outline

Geometric and Algebraic MG

Algebraic Multigrid for Lattice QCD

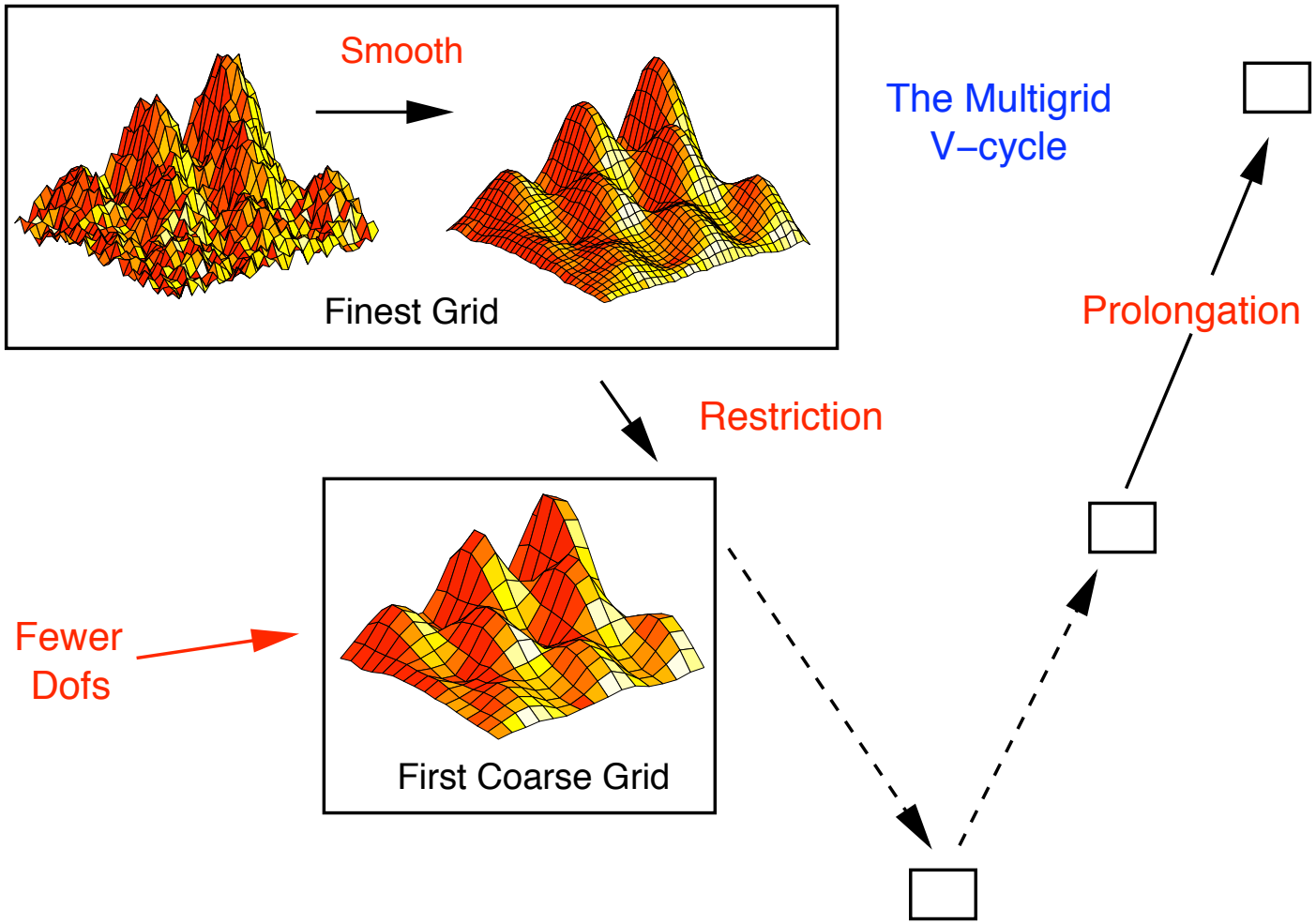
Challenges and opportunities

Geometric and Algebraic MG

Algebraic Multigrid for Lattice QCD

Challenges and opportunities

Multigrid in a Nutshell

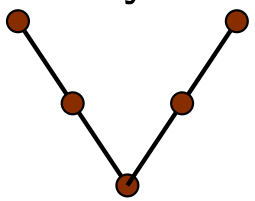


Geometric Multigrid: Fedorenko 1961, Brandt, Hackbusch 1970s, ...

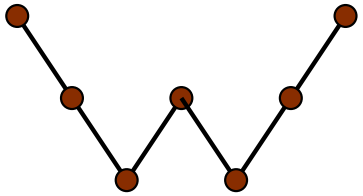
Ingredients

- elliptic PDE $\mathcal{L}(u) = b$
- discretization scheme (finite elements)
  $\longrightarrow$ hierarchy of systems $D_\ell x_\ell = b_\ell$
  $\longrightarrow$ intergrid operators $P_\ell^{\ell+1}, R_{\ell+1}^\ell$
- iterative methods S_ℓ a.k.a. smoothers

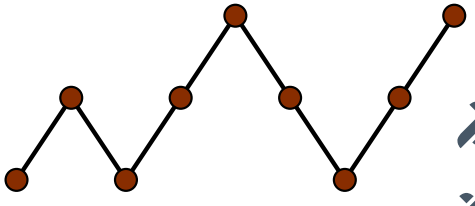
V-cycle



W-cycle



F-cycle



Operations

- smoothing:
  $x_\ell \leftarrow x_\ell - M_\ell^{-1}(D_\ell x_\ell - b_\ell)$
- coarse grid correction:
  $x_\ell \leftarrow x_\ell - P_\ell^{\ell+1} D_{\ell+1}^{-1} R_{\ell+1}^\ell (D_\ell x_\ell - b_\ell)$

The multigrid promise

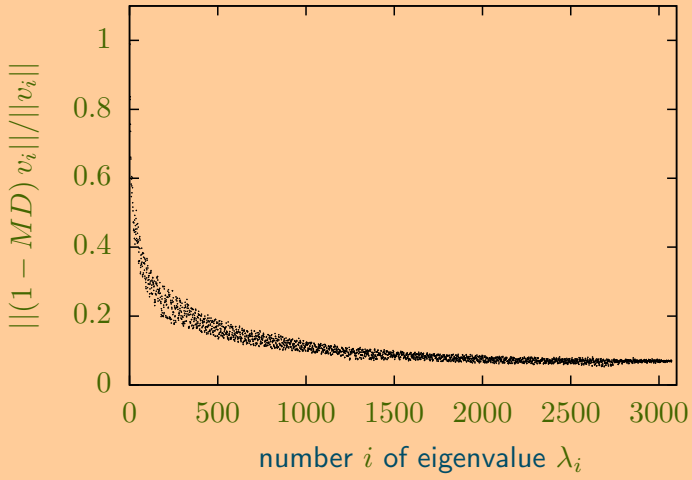
- optimal complexity with F-cycles:
  $\mathcal{O}(n)$ operations for solution accuracy $\sim$ discretization error

Multigrid: the better way to deflate

Smoother: $I - MD$

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

Smoother



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

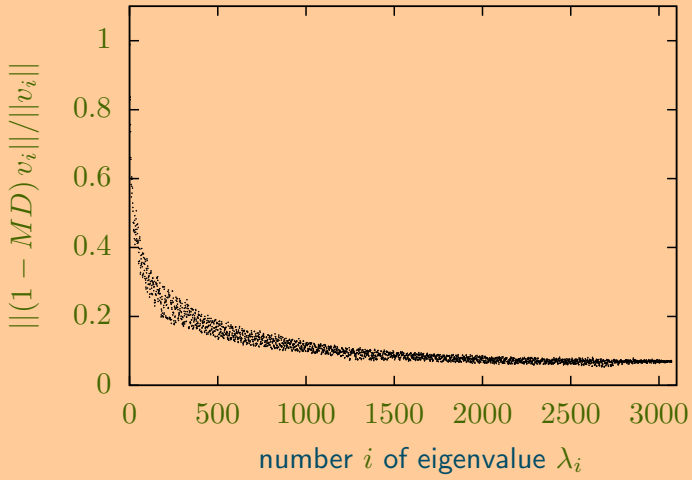
Multigrid: the better way to deflate

Coarse-grid correction: $I - PD_c^{-1}RD$

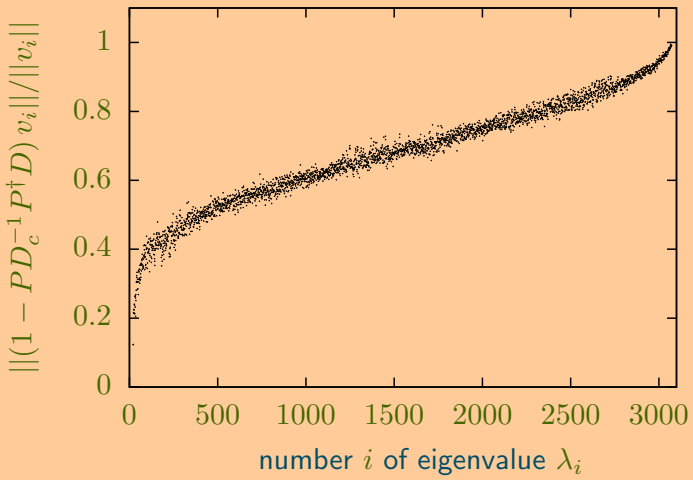
- small eigenvectors built into interpolation P

  $\Rightarrow$ Effective on small eigenvectors

Smoother



Coarse grid correction



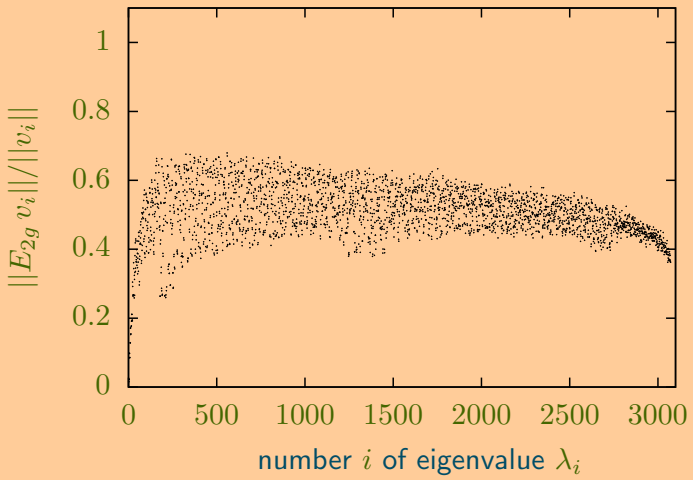
$Dv_i = \lambda_i v_i$ with $|\lambda_1| \leq \ldots \leq |\lambda_{3072}|$

Multigrid: the better way to deflate

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

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

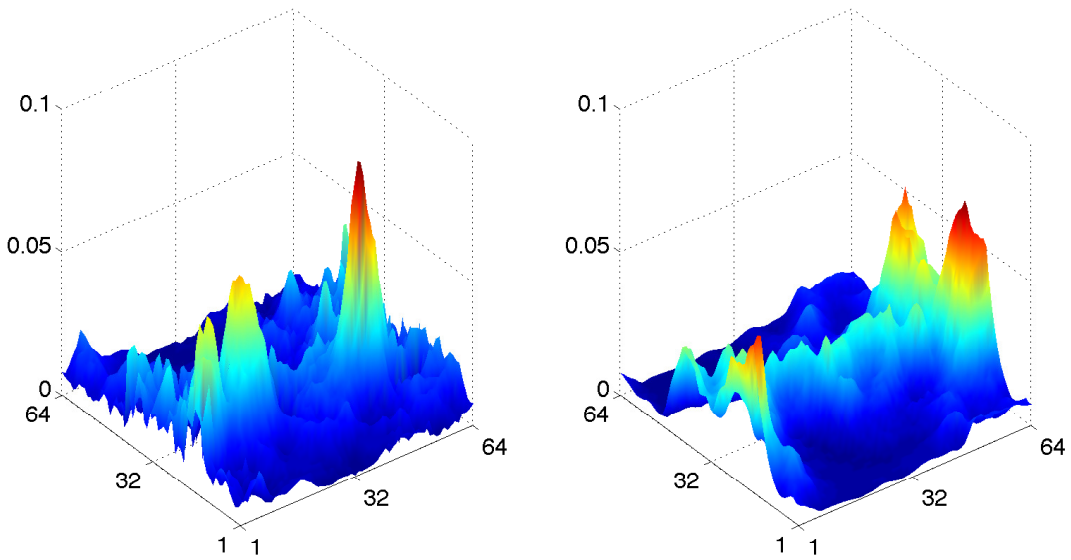
Multigrid



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

A paradoxon

In lattice QCD: **smoothed vectors are not smooth**



Algebraic Multigrid (AMG): Brandt, McCormick, Ruge 1982

Given:
- $Dx = b$
- Iterative method S a.k.a. smoother

Wanted:
- Hierarchy of spaces (grids) $\mathcal{V}_\ell, \quad \ell = 0, \ldots, L$
- Intergrid transfer operators
  $P_{\ell+1}^{\ell} : \mathcal{V}_{\ell+1} \longrightarrow \mathcal{V}_\ell, \;\; R_\ell^{\ell+1} : \mathcal{V}_\ell \longrightarrow \mathcal{V}_{\ell+1}$

Result:
- Hierarchy of systems
  $D_\ell x_\ell = b_\ell$ with $D_{\ell+1} = R_{\ell+1}^{\ell} D_\ell P_\ell^{\ell+1}$ (Petrov-Galerkin)
- smoothers S_ℓ

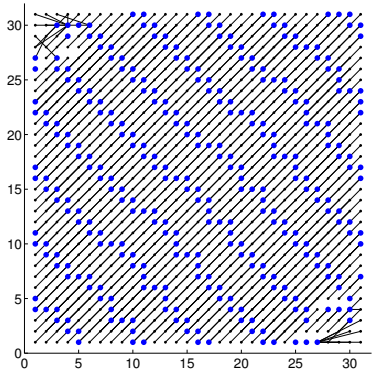
Guidelines:
- smooth vectors: $\|Dv\| \ll \|v\|$
- complementarity of smoother and coarse grid correction:
  v smooth $\Rightarrow$ v well approximated in $\mathrm{range}(P)$

AMG: Hierarchy of spaces and intergrid operators I

Hermitian case: $D = D^\dagger$. Take $R = P^\dagger$

C-F-splitting: Identify coarse variables as a subset $\mathcal{C}$ of all variables $\mathcal{C} \cup \mathcal{F}$

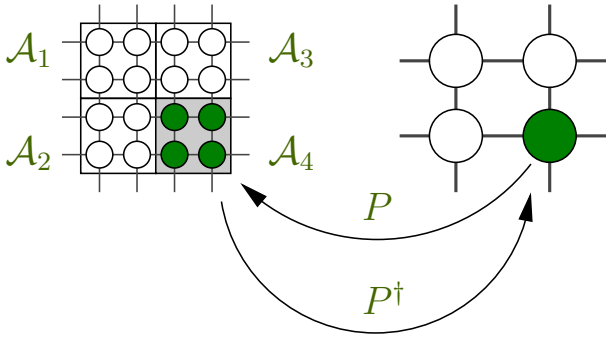
- Geometric coarsening
- Strength of connection (Ruge-Stüben '85, Chow '03, Brannick et al. '06, . . .)
- Compatible relaxation (Brandt '00, Brannick-Falgout '10, . . .)



Interpolation for C-F-splitting:

- For each $i \in \mathcal{F}$ determine set $\mathcal{C}_i$ from which i interpolates.
- Preserve smooth vectors: $Dv \approx 0 \Leftrightarrow P(v_f) \approx v$.

AMG: Hierarchy of spaces and intergrid operators II



Aggregation:

- Group several variables into one coarse aggregate $\mathcal{A}$ (Braess '95, Vanek, Mandel, Brezina, '94, '96, ...)

Interpolation for aggregation:

- piecewise constant, $P = \sum_{\mathcal{A}} \mathbf{1}_{\mathcal{A}}$ (!)
- smoothed aggregation, $P = \sum_{\mathcal{A}} D \mathbf{1}_{\mathcal{A}}$

AMG: Building interpolation using test vectors

Recall: smooth vectors are to be well approximated in $\mathrm{range}(P)$.

Given: test vectors $v^{(1)}, \ldots, v^{(k)} \in \mathbb{C}^n$ representing low modes

Wanted: interpolation P accurate for test vectors $v^{(s)}$

C-F-splittings: Least Squares Interpolation (Kahl '09)

$$\mathcal{L}_{\mathcal{C}_i}(p_i) = \sum_{s=1}^{k} \omega_s \left(v_i^{(s)} - \sum_{j \in \mathcal{C}_i} (p_i)_j v_j^{(s)} \right)^2 \to \min_{p_i}$$

Aggregates: Distribute test vecs over aggregates (Brezina et al '04)

$$(v^{(1)}, \ldots, v^{(k)}) = \begin{Vmatrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{Vmatrix} = \begin{pmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \\ \vdots \\ \mathcal{A}_s \end{pmatrix} \to P = \begin{pmatrix} \mathcal{A}_1 & & & \\ & \mathcal{A}_2 & & \\ & & \ddots & \\ & & & \mathcal{A}_s \end{pmatrix}$$

AMG: Adaptive setups I

How to get test vectors?

- Known from the problem: rigid body modes in mechanics, e.g.
- Adaptively:
  αSA (Brezina, Falgout, MacLachlan, Manteuffel, McCormick, Ruge '04)
  Bootstrap AMG (Brandt, Brannick, Kahl, Livshitz '10)

Adaptivity in αSA

Adaptive Algebraic Multigrid (αSA)

"Iteratively test and improve the current method until good enough"

Initialize $\mathcal{M}$ to be the smoothing iteration
Initialize random test vector x
Apply $\mathcal{M}$ to $Dx = 0$
$\rightarrow$ smoothed vector $\widetilde{x}$, convergence speed θ
while $\theta > tol$ **do**
  Update set of test vectors $\mathcal{U} = \mathcal{U} \cup \widetilde{x}$
  Construct multigrid method M based on $\mathcal{U}$
  $\mathcal{M} = M$
  Choose new random x
  Apply $\mathcal{M}$ to $Dx = 0$
  $\rightarrow$ smoothed vector $\widetilde{x}$, convergence speed θ
end while

[Brezina et. al. 04]

AMG: Adaptive setups III

Bootstrap Algebraic Multigrid

"Continuous updating components of the MG hierarchy using practical tools and measures built from the evolving MG solver"

- smoother action known, initial test vectors

$$u^{(s)} = S^{\eta}\widetilde{u}^{(s)}, \quad \widetilde{u}^{(s)} \text{ random}$$

- observation $(P_\ell = P_1^0 \cdots P_\ell^{\ell-1},\ D_\ell = P_\ell^\dagger D_0 P_\ell,\ T_\ell = P_\ell^\dagger P_\ell)$

$$\frac{\langle v_\ell, v_\ell\rangle_{D_\ell}}{\langle v_\ell, v_\ell\rangle_{T_\ell}} = \frac{\langle P_\ell v_\ell, P_\ell v_\ell\rangle_D}{\langle P_\ell v_\ell, P_\ell v_\ell\rangle_2}$$

Bootstrap Idea

Eigenpairs (v_ℓ, λ_ℓ) of (D_ℓ, T_ℓ) $\longrightarrow$ Eigenpairs $(P_\ell v_\ell, \lambda_\ell)$ of D + interpolation error

[Brandt, Brannick, Kahl, Livshits '10, Manteuffel, McCormick, Park, Ruge '10]

Geometric and Algebraic MG

Algebraic Multigrid for Lattice QCD

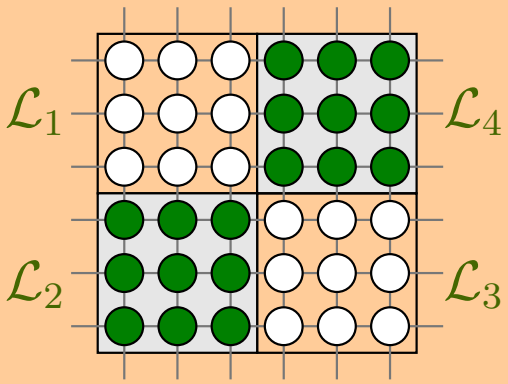
Challenges and opportunities

Multigrid for lattice QCD

From now on: $D\psi = \eta$, D (clover improved) Wilson-Dirac operator, periodic, anti-periodic or open bc

SAP: Schwarz Alternating Procedure, aka Multiplicative Schwarz

Two color decomposition of $\mathcal{L}$



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

in: ψ, η, ν – out: ψ
for $k = 1$ to ν **do**
  $r \leftarrow \eta - D\psi$
  for all *green* $\mathcal{L}_i$ **do**
    $\psi \leftarrow \psi + B_{\mathcal{L}_i} r$
  end for
  $r \leftarrow \eta - D\psi$
  for all *white* $\mathcal{L}_i$ **do**
    $\psi \leftarrow \psi + B_{\mathcal{L}_i} r$
  end for
end for

- $B_{\mathcal{L}_i}$ inverted approximately
- Preconditioner to GCR or FGMRES

(Schwarz 1870, Lüscher '03)

Local coherence and the inexact deflation method

Local coherence of low quark modes (Lüscher '07):

- Locally, all low quark modes are well approximated by just a few (experimental result).
- Approximations to low quark modes can be obtained via **inverse iteration** for D
- $\longrightarrow$ "Inexact deflation" method (Lüscher '07)

Inexact deflation method

- subdivide lattice into "subdomains" ($=$ aggregates)
- setup: compute test vectors via bootstrap approach
- interpolation P : defined as in aggregation based AMG
- coarse system $D_c = P^\dagger D P$ ("little Dirac"): solved with SAP $+$ standard deflation
- solve $D\pi_R\psi = \eta$ with $\pi_R = I - PD_c^{-1}P^\dagger D$

Transfer of αSA to Lattice QCD

Babich, Brannick, Brower, Clark, Manteuffel, McCormick, Osborn, Rebbi '10, . . . :

- 4d Wilson-Dirac system $D\psi = \eta$
- 2 aggregates per 4^4-lattice $\times$ colors $\times$ spins to preserve γ_5-symmetry[1)]
- GMRES as smoother
- 3 levels, W-cycles
- αSA setup

  or

  modified αSA setup: Works with several vectors a time
- used as preconditioner for GCR

[1)] see appendix

Current AMG solvers for D_W

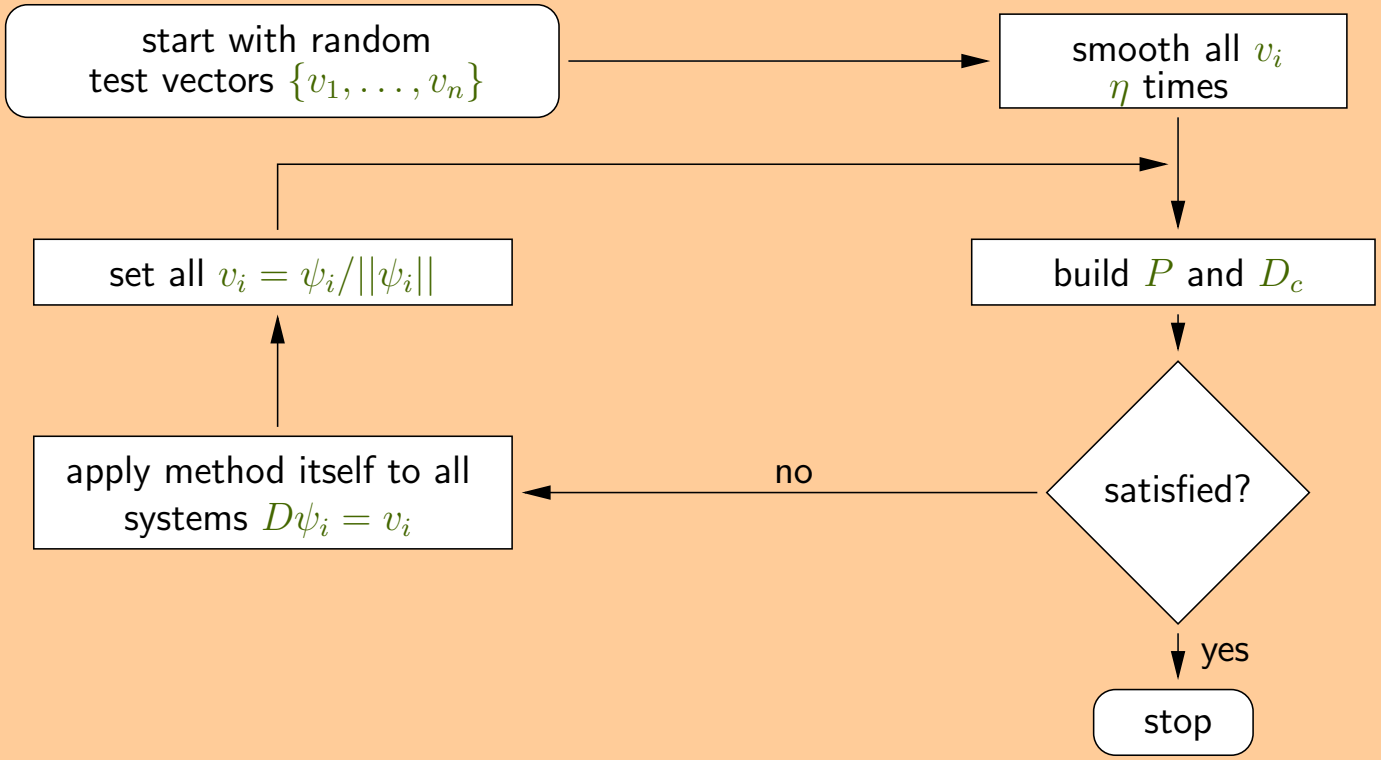
	QOPQDP	OpenQCD	DD-αAMG
clover term	included	included	included
mixed precision	yes	yes	yes
smoother	GMRES	SAP	SAP
aggregation	γ_5-comp.	arbitrary	γ_5-comp.
setup	1)	2)	3)
typ. # test vecs (N)	20	30	20
# vars / coarse site	$2N$	N	$2N$
cycling	K-cycle	n.a.	K-cycle

1) inverse iterations with GMRES on sequence of test vecs
2) repeated inverse iteration with emerging solver on all test vecs at once
3) modification of 2)

DD-αAMG: Frommer, Kahl, Krieg, Leder, Rottmann '13

Setup in DD-αAMG

Bootstrapping process



Snapshots on performance: 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

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.

Snapshots on performance: oe-BiCGStab vs DD-αAMG

	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.

Snapshots on performance: mass scaling and levels

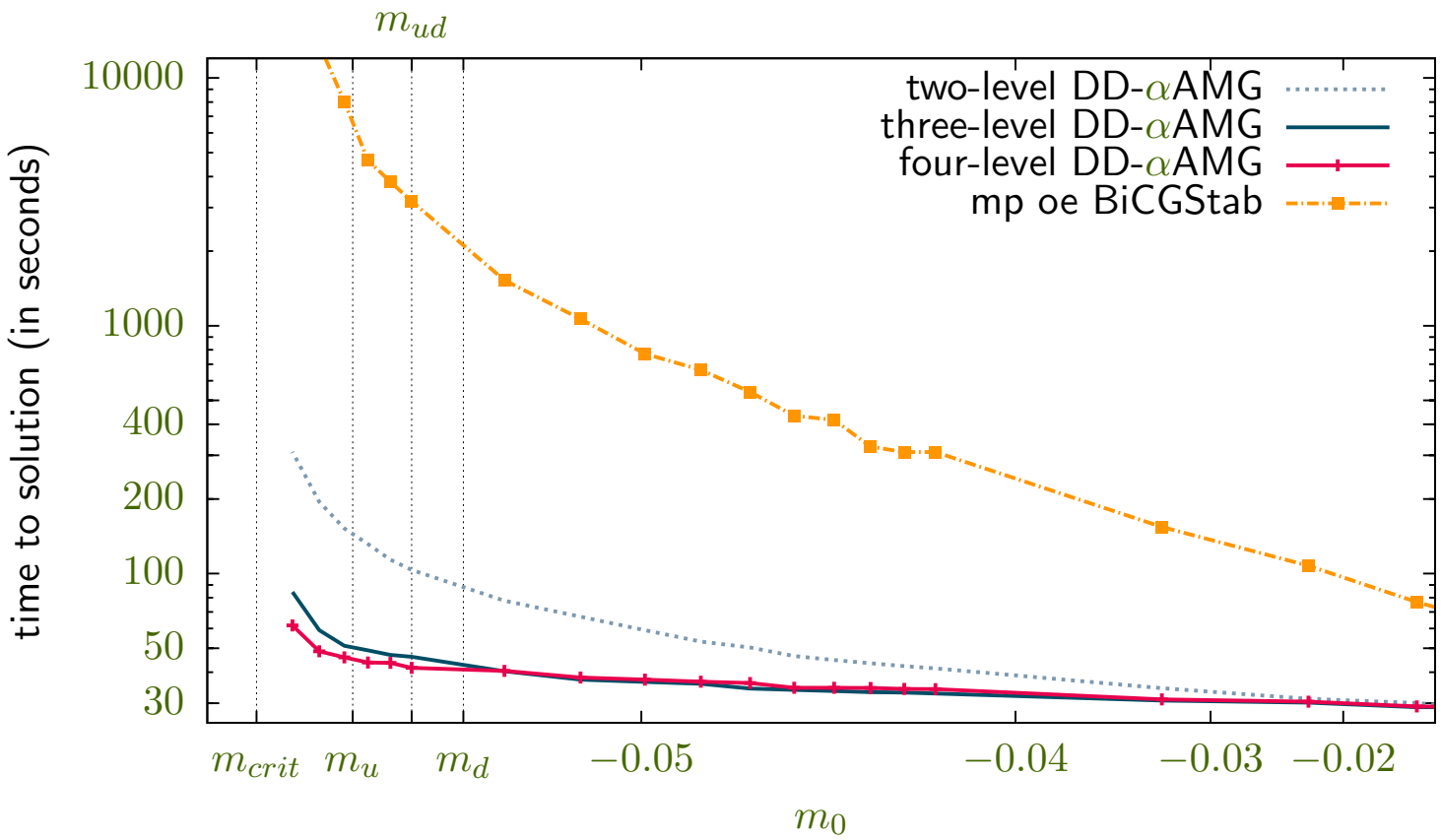


Figure : Mass scaling of 2, 3 and 4 level DD-αAMG, 64^4 lattice, configuration id 5, restart length $n_{kv} = 10$, 128 cores

Geometric and Algebraic MG

Algebraic Multigrid for Lattice QCD

Challenges and opportunities

Challenge I: Improving the setup

Robustness: A "default" setup should work well on all configurations

Role of heuristics: Setup should be more supported by mathematical theory $\longrightarrow$ AMG for non-symmetric systems (Brezina, Manteuffel, McCormick, Ruge, Sanders '10)

Investigate:

- R and P from singular vecs rather than eigenvecs
  $Dv_i = \sigma_i u_i, v_i$ orthogonal, u_i orthogonal, $\sigma_i > 0$
- deviation from normality:
  - continuum operator is normal, $\mathcal{D}\mathcal{D}^\dagger = \mathcal{D}^\dagger\mathcal{D}$
  - smearing makes D more normal,
  - D more normal towards the continuum limit or with other discretization

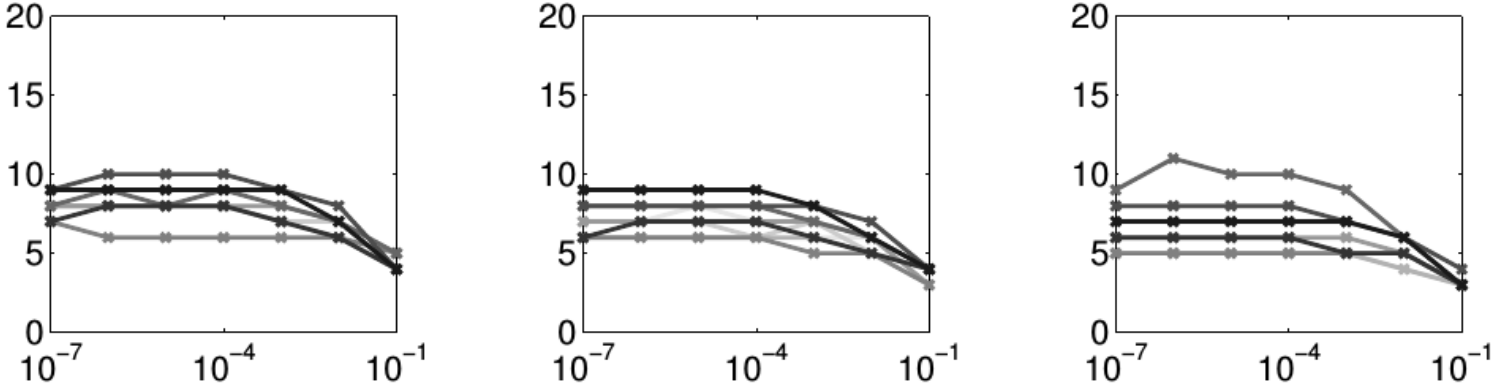
Challenge I: Improving the setup

Recent work by Brannick and Kahl for Schwinger model ('13):

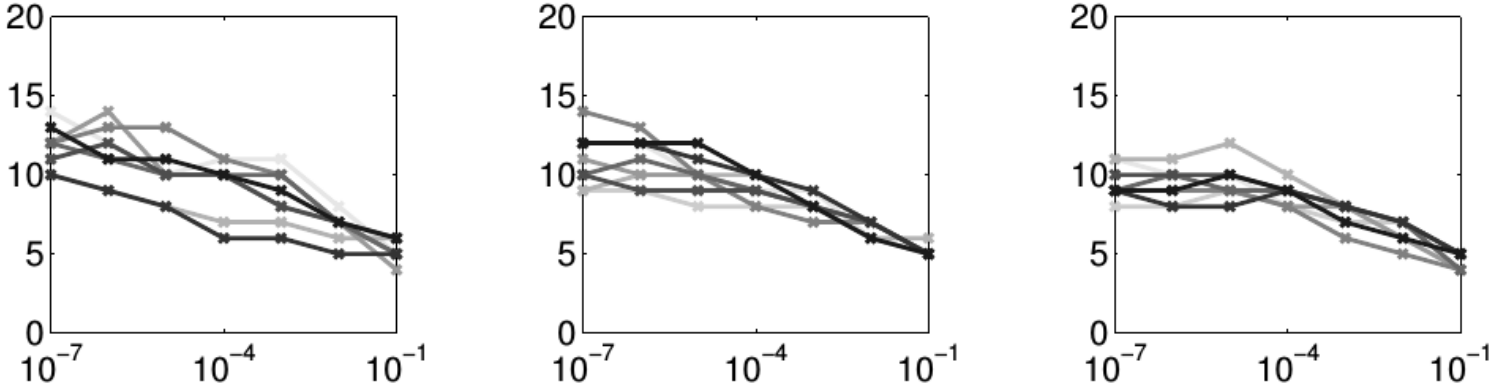
- Singular vecs for D are related to eigenvecs of $\gamma_5 D$
- Smoothing with Kaczmarz for D
  $\longrightarrow$ only right sing. vecs of D matter
- justification for $R = P^\dagger$ in terms of sing. vecs approximation
- Bootstrap setup for $\gamma_5 D$ gives approx. left and right singular vecs for D
- geometric C-F splittings, least squares interpolation
- W-cycle
- all for oe-reduced system

Results for Schwinger model

$N = 128$



$N = 256$



$\beta = 3$ $\beta = 6$ $\beta = 10$

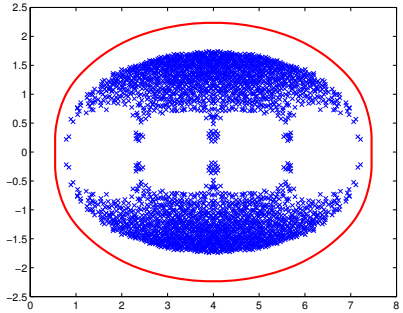
Challenge II: Deviation from normality

Field of values:
$\mathcal{F}(D) = \{\psi^\dagger D \psi : \psi^\dagger \psi = 1\}$

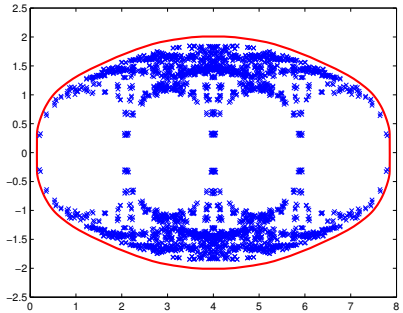
Property: $\mathcal{F}(P^\dagger D P) \subset \mathcal{F}(D)$

If D were normal

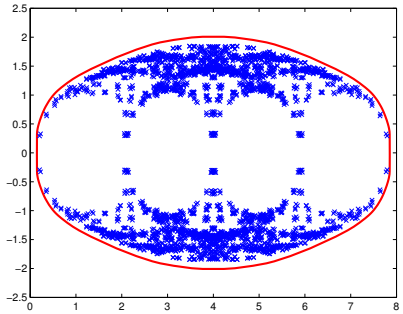
- Eigenvecs and singular vecs coincide
- $\mathcal{F}(D) =$ convex hull of spectrum
- Spectrum of coarse grid operator falls 'inside" spectrum of fine grid operator



no smoothing



1 HYP



2 HYP

Challenge III: System hierarchy from modelling

Idea: move towards "geometric" multigrid

- find coupled hierarchy of discretized Dirac equations
  $\longrightarrow$ finite elements?
- This fixes the coarse grid system and the prolongations
- obtain smoother geometrically rather than algebraically
- Example: MG for Maxwell's equations

Opportunities

"One setup, many solves"
strategies:

- Wilson-Dirac preconditioner for the overlap operator ($\rightarrow$ M. Rottmann, Mo 15:15)
- Updating of P, R in HMC (OpenQCD, $\rightarrow$ M. Lin, Mo 14:35)

Similar idea, other operators:

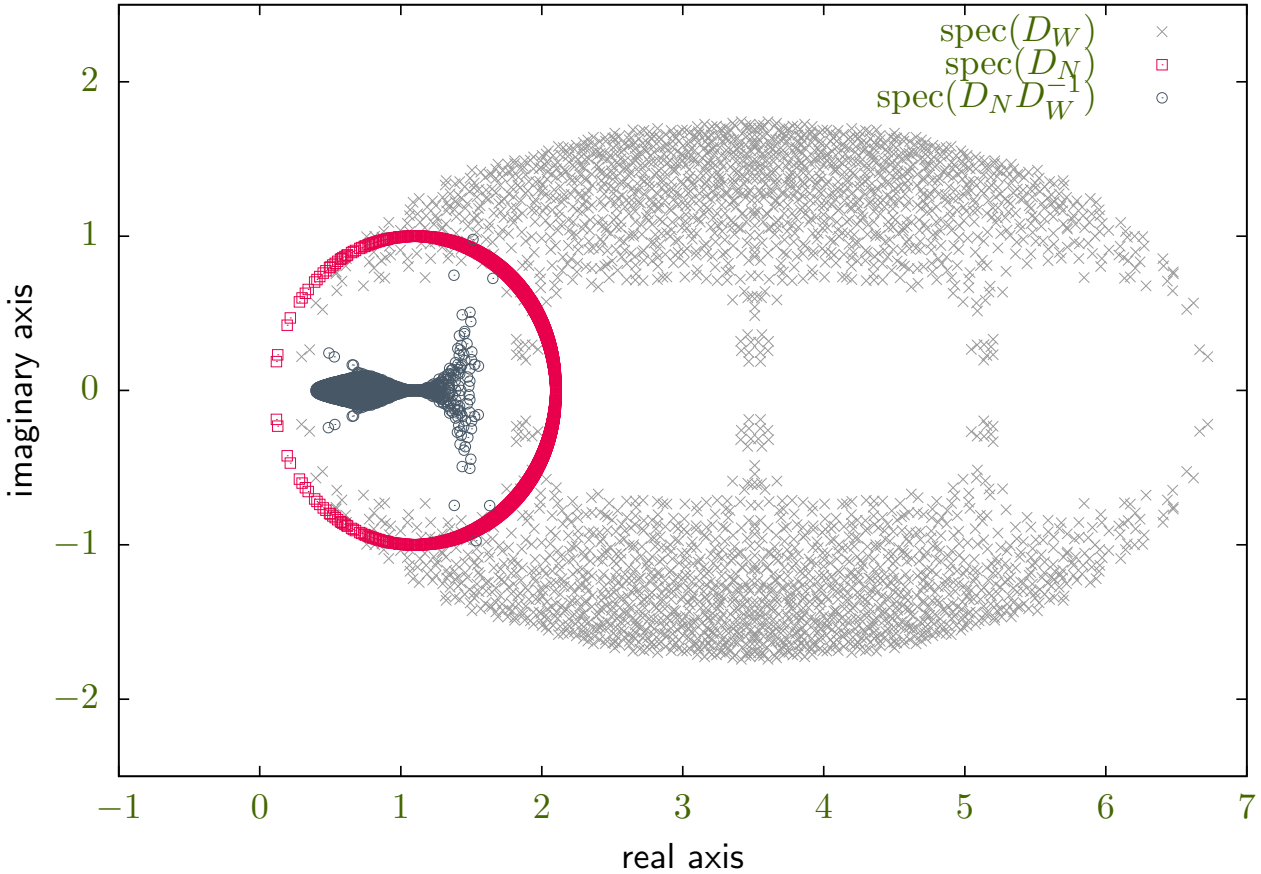
- Domain wall: aggregate 5th dimension (Cohen et al '10)
- Domain wall: recursive "in-exact deflation" (Boyle '14)

Implementations:

- in QUDA for GPUs ($\rightarrow$ M. Clark, Mo 16:50)[1)]
- QPACE 2 ($\rightarrow$ T. Wettig, Sa 09:00)

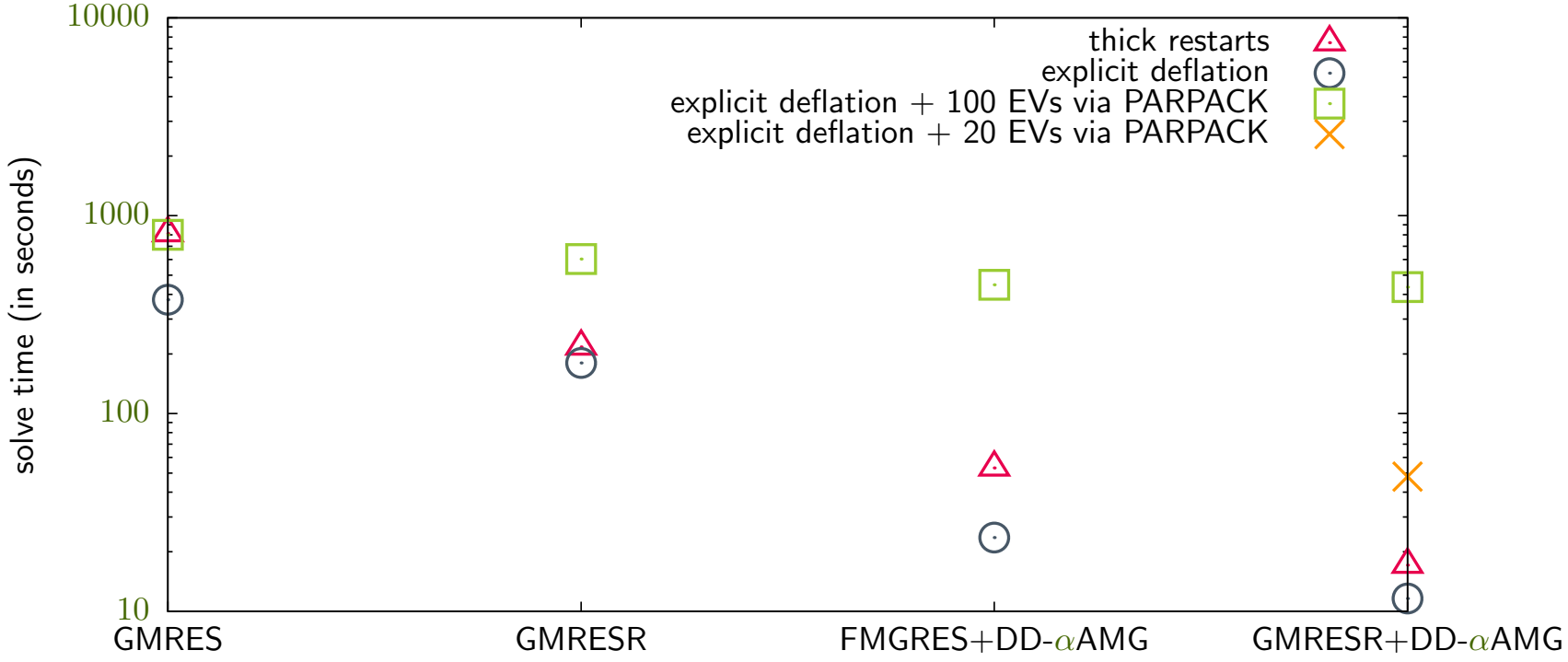
[1)] heterogeneous / additive AMG

Preconditioning the overlap operator



Brannick, Frommer, Kahl, Rottmann, Strebel: work in progress

Thick Restarts and Explicit Deflation



- 32^4 lat, 3HEX smeared BMW-c cnfg, $1{,}024$ cores
- GMRESR := FGMRES-64bit + GMRES-32bit
- GMRESR+DD-αAMG := FGMRES-64bit + FGMRES-32bit + DD-αAMG

Conclusions

state-of-the-art

- adaptive AMG works for lattice QCD
- fairly robust
- best for multiple r.h.s.
- parallel efficiency depends on no. of levels
- software is available

To do

- further improve setup
- singular vecs instead of eigenvecs, normality
- HMC
- ...

Thanks to: James Brannick, Karsten Kahl, Stefan Krieg, Björn Leder, Matthias Rottmann, Marcel Schweitzer, Artur Strebel, Regensburg and Wuppertal

Complementarity of smoother and coarse grid correction **revisited**

$$D_W \neq D_W^\dagger, \text{ but } \gamma_5 D_W = (\gamma_5 D_W)^\dagger$$

- coarse grid correction $I - P(RDP)^{-1}RD$ projects onto $\mathrm{range}(RD)^\perp$ along $\mathrm{range}(P)$
- $\mathrm{range}(P)$ should well approximate smooth vectors
- $\mathrm{range}(RD)^\perp$ should well approximate non-smooth-vectors

Consequence:

- P built from approximate right evs
- R built from approximate left evs
- Suggestion: take $R = (\gamma_5 P)^\dagger$

Additional feature

- aggregate positive and negative spin components separately
  $\longrightarrow \mathrm{range}(P) = \mathrm{range}(\gamma_5 P)$
  $\longrightarrow$ take $R = P^\dagger$

(Babich et al. '10)