Multigrid Preconditioning for the Overlap Operator

James Brannick[†], Andreas Frommer[⋆], Karsten Kahl[⋆], Björn Leder[⋆], **Matthias Rottmann**[⋆], Marcel Schweitzer[⋆], and Artur Strebel[⋆]

[⋆] Bergische Universität Wuppertal
[†] Pennsylvania State University

June 23, 2014

Motivation

Task: Find solution of $D_N \varphi = \eta$ where

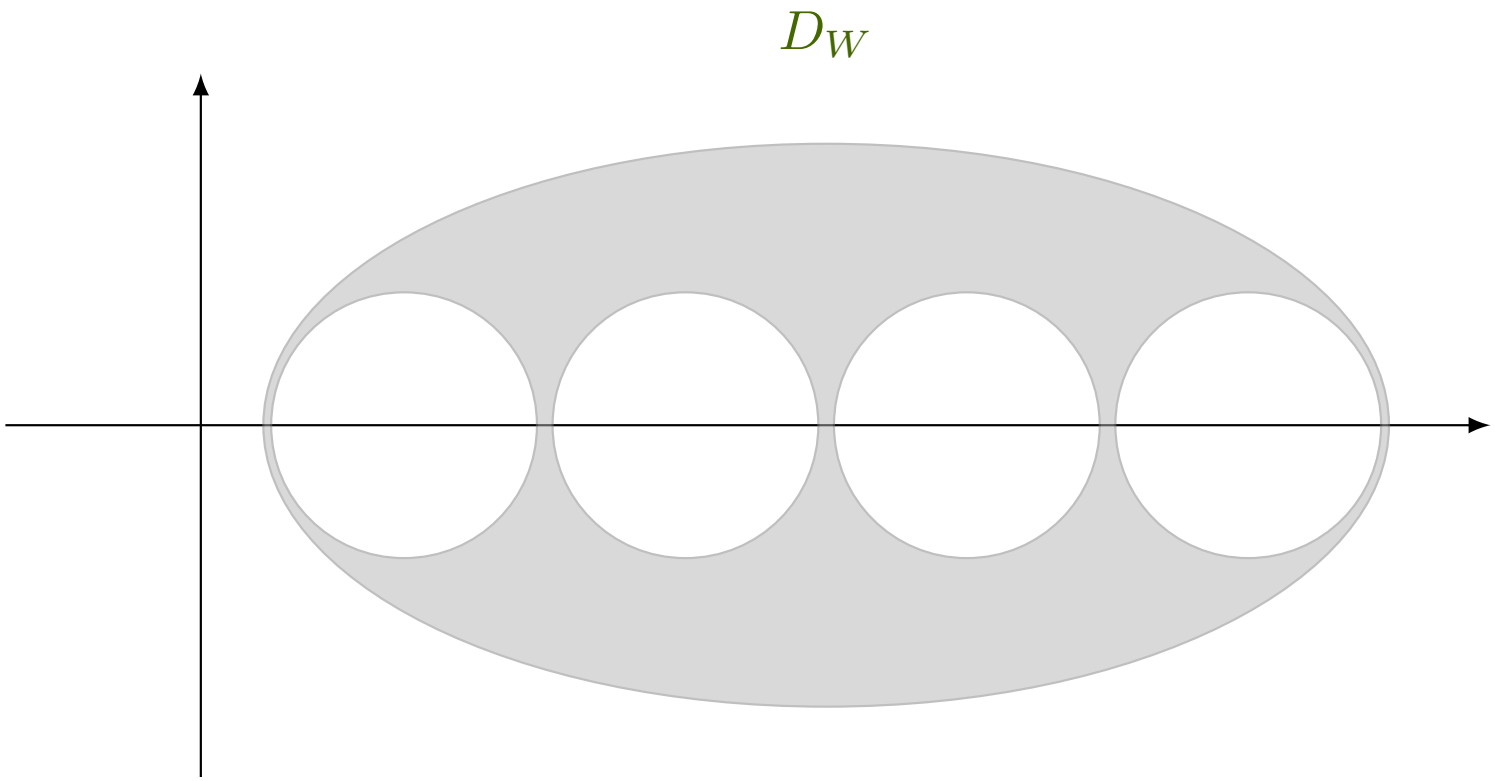
$$\begin{aligned} D_N =& (m_0^N - \frac{m}{2})(1 + \gamma_5 \operatorname{sign}(\overbrace{\gamma_5 (D_W - m_0^N)}^{H_W})) + m \\ =& (m_0^N - \frac{m}{2})(1 + \gamma_5 (H_W^\dagger H_W)^{-\frac{1}{2}} H_W) + m \end{aligned}$$

Challenges:

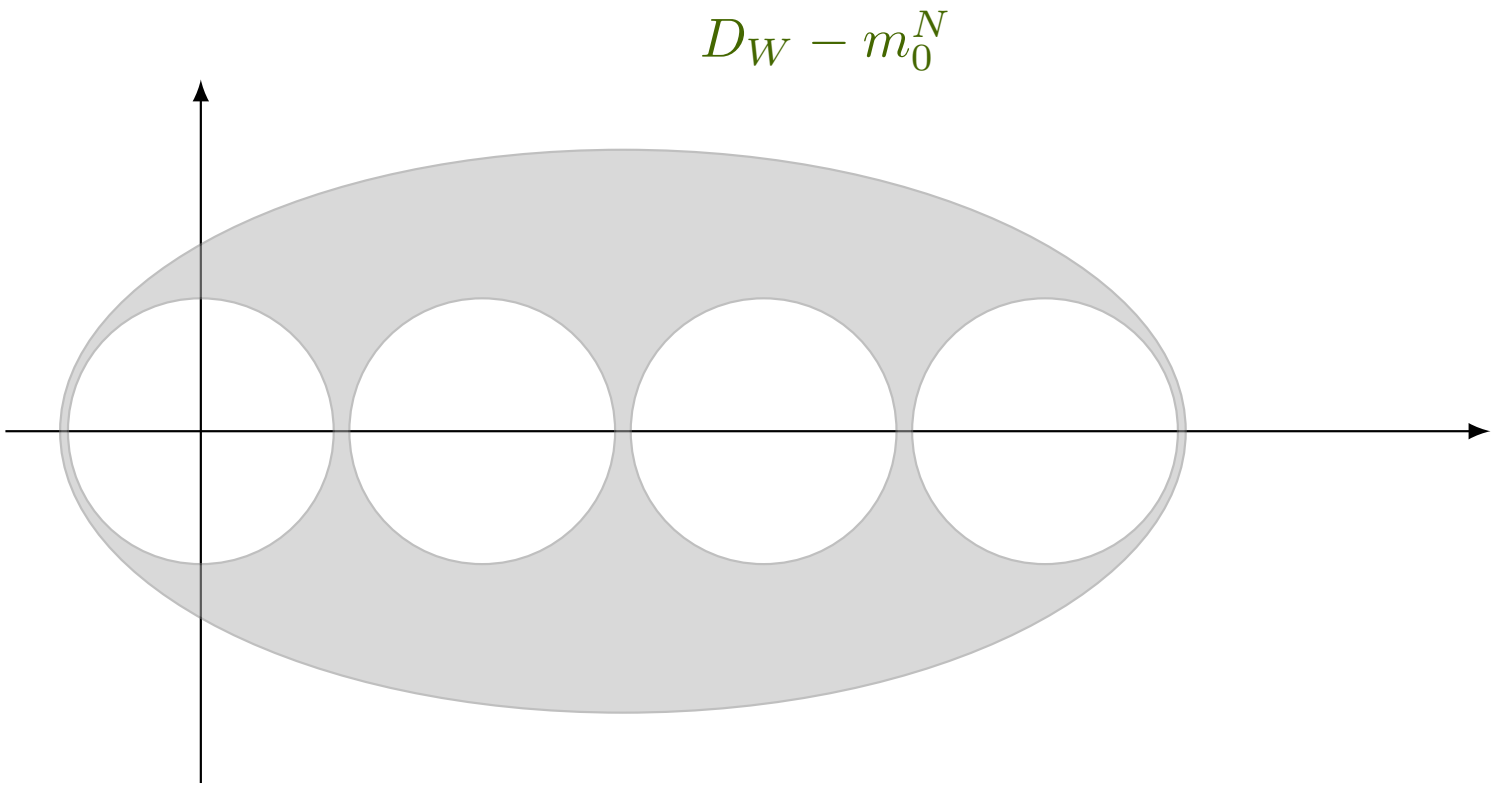
i) Evaluating $(H_W^\dagger H_W)^{-\frac{1}{2}} x$ is quite costly

ii) Iteration counts of $\mathcal{O}(1{,}000)$ for $D_N \varphi = \eta$

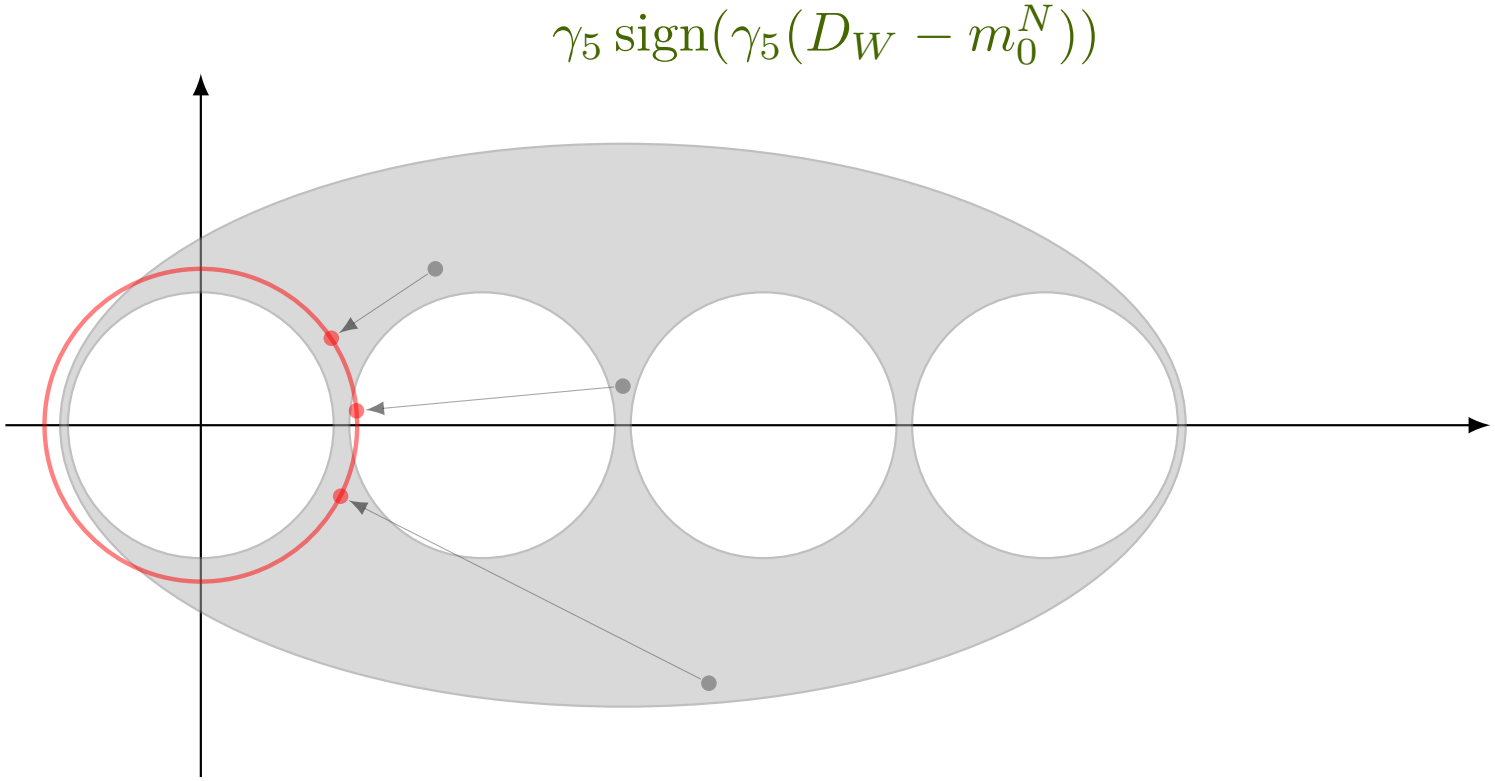
Overlap construction – assuming normality of D_W



Overlap construction – assuming normality of D_W

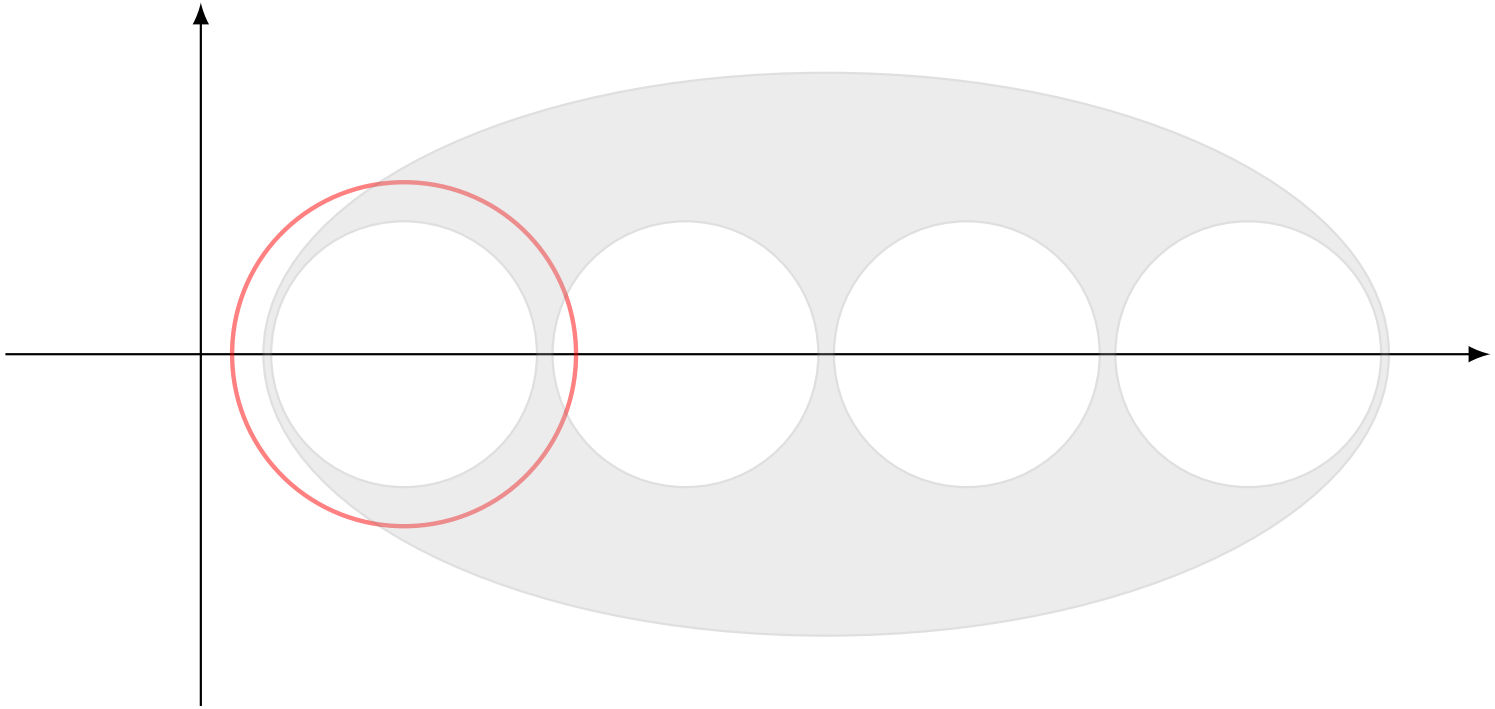


Overlap construction – assuming normality of D_W



Overlap construction – assuming normality of D_W

$$D_N = (m_0^N - \tfrac{m}{2})(1 + \gamma_5 \operatorname{sign}(\gamma_5(D_W - m_0^N)) + m$$



Preconditioning with the Wilson-Dirac operator

Challenge ii): Reduce number of iterations for $D_N \varphi = \eta$

Idea: Preconditioning (what else?)

Preconditioning with the Wilson-Dirac operator

Challenge ii): Reduce number of iterations for $D_N \varphi = \eta$

Idea: Preconditioning (what else?)

What is a suitable preconditioner for D_N?

Preconditioning with the Wilson-Dirac operator

Challenge ii): Reduce number of iterations for $D_N \varphi = \eta$

Idea: Preconditioning (what else?)

What is a suitable preconditioner for D_N?

The kernel operator of D_N, i.e., the Wilson-Dirac operator D_W

$$D_N D_W^{-1} \psi = \eta \text{ with } \varphi = D_W^{-1} \psi$$

- Computing D_W^{-1} is done by DD-αAMG [arXiv:1303.1377]
- D_W^{-1} is cheap

Why is D_W a good preconditioner for D_N?

Assuming normality of D_W (i.e., $D_W^\dagger D_W = D_W D_W^\dagger$) we find

Relation between low modes of D_W and D_N

Let λ be a small eigenvalue of D_W, i.e., $D_W x = \lambda x$ with $|\lambda|$ small. W.l.o.g. assume $m = 0$. Then

$$\begin{aligned} D_N x &= m_0^N \big(1 + \gamma_5 \operatorname{sign}(\gamma_5 (D_W - m_0^N))\big) x \\ &= m_0^N \big(1 + \gamma_5 ((D_W - m_0^N)^\dagger (D_W - m_0^N))^{-\frac{1}{2}} \\ &\quad \cdot (\gamma_5 (D_W - m_0^N))\big) x \\ &= m_0^N x + m_0^N (\lambda - m_0^N) \big((\lambda - m_0^N)(\overline{\lambda - m_0^N})\big)^{-\frac{1}{2}} x \\ &= m_0^N \big(1 + \operatorname{sign}(\lambda - m_0^N)\big) x \end{aligned}$$

Deviation from normality of non-chiral discretizations

Quality of preconditioner depends on normality?!

Measure for the deviation of normality

$$\delta_N := ||D_W^\dagger D_W - D_W D_W^\dagger||_F,$$

where $||X||_F^2 = \sum_{i,j=1}^n x_{ij}^2,\ X \in \mathbb{C}^{n\times n}$.

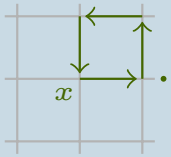
Theorem

The deviation of normality of D_W is given by

$$\delta_N = 16 \sum_x \sum_{\mu>\nu} \mathrm{Re}(\mathrm{tr}(I - Q_x^{\mu,\nu})),$$

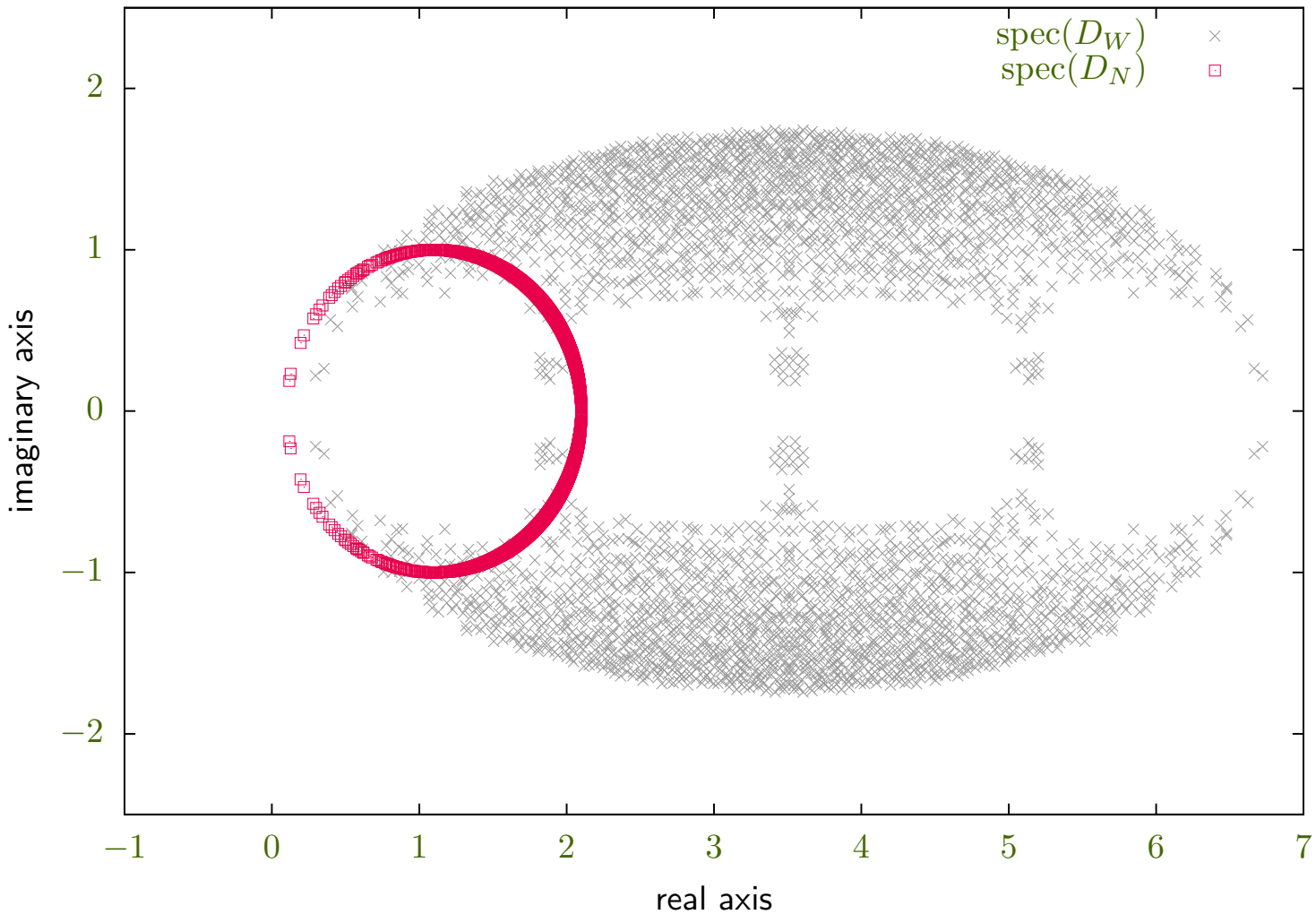
where $Q_x^{\mu,\nu}$ is the plaquette defined by

$$Q_x^{\mu,\nu} = U_\nu(x) U_\mu(x+\hat{\nu}) U_\nu^H(x+\hat{\mu}) U_\mu^H(x) =$$

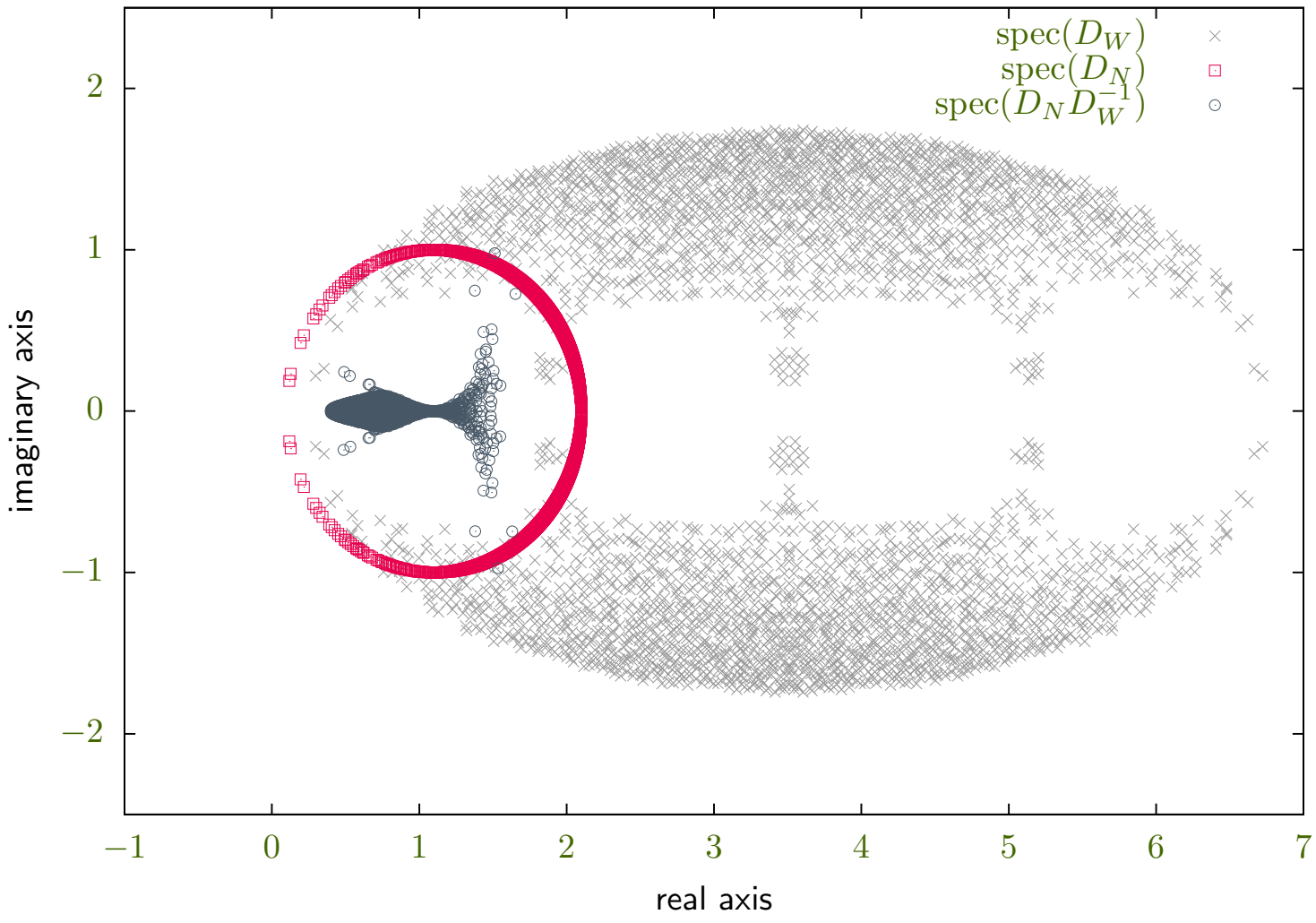


.

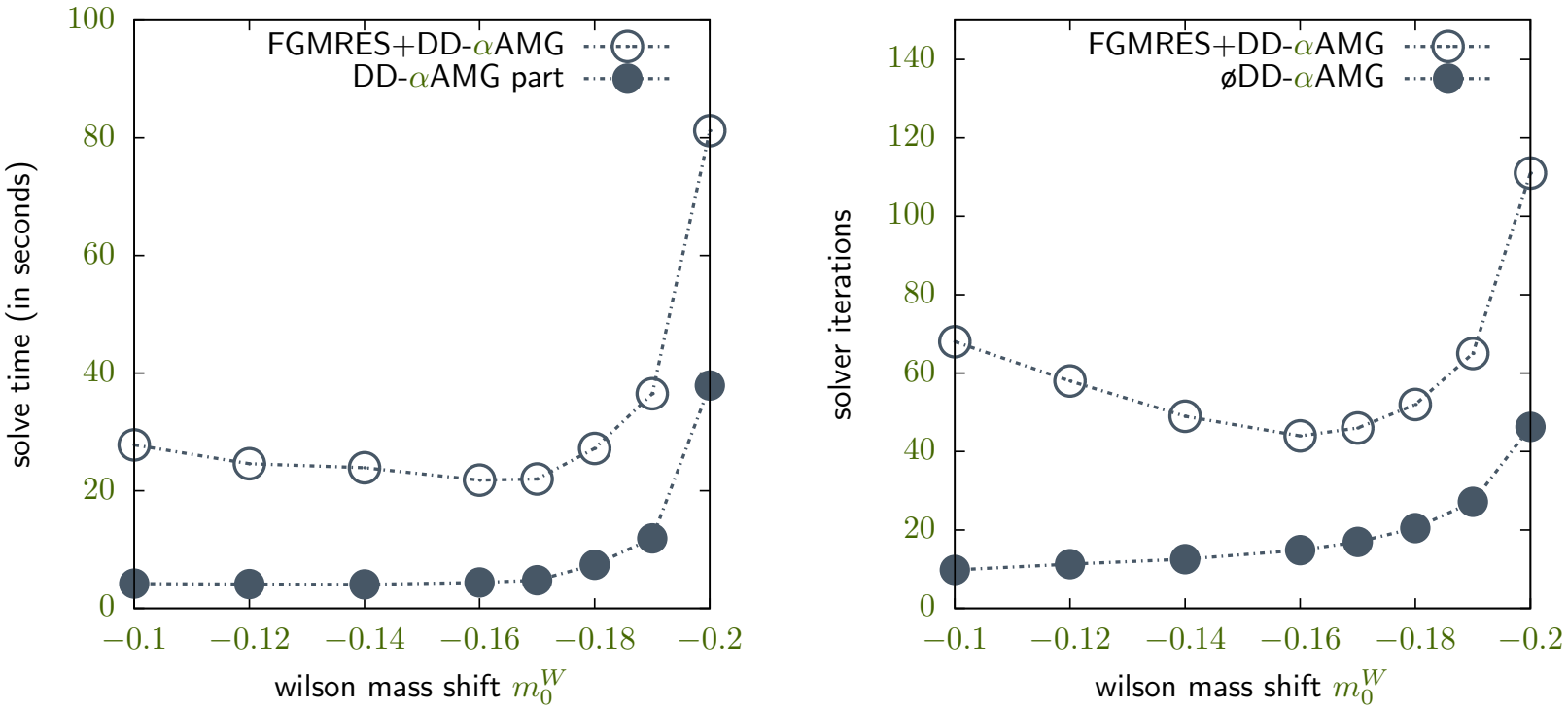
Why is D_W a good preconditioner for D_N?



Why is D_W a good preconditioner for D_N?

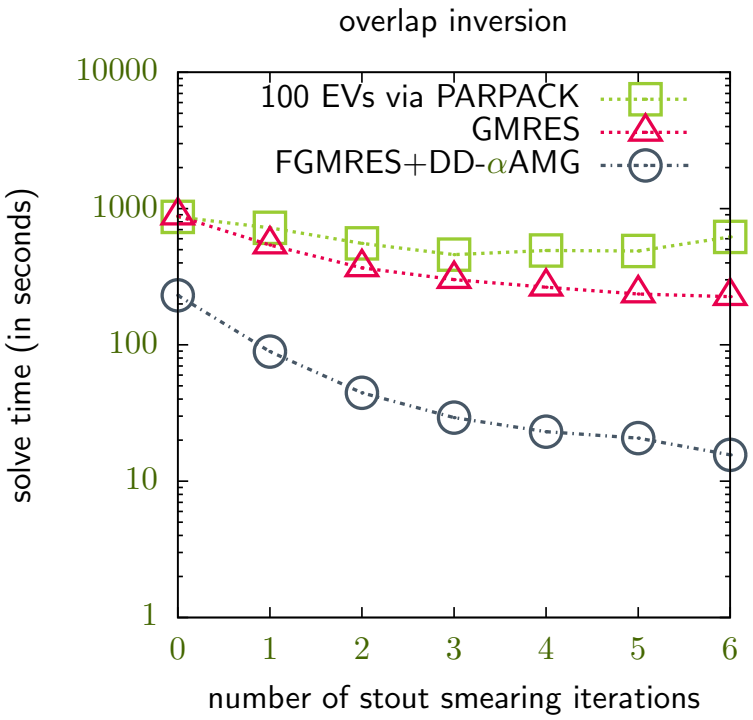


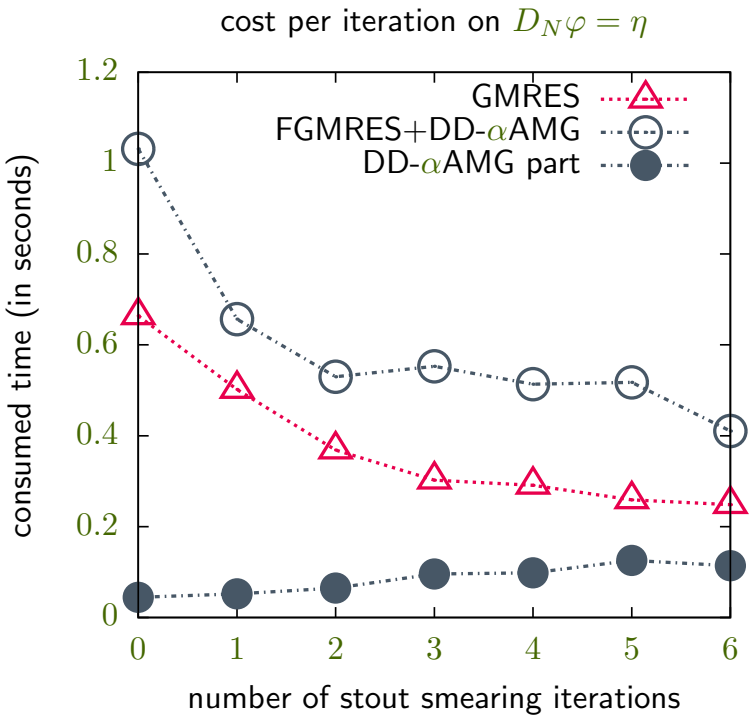
Scanning the Optimal Wilson Preconditioner Mass Shift m_0^W



- 32^4 lat, 3HEX smeared BMW-c cnfg (unpublished), $m_\pi \approx 350$ MeV, $1{,}024$ processes
- overlap tol 10^{-8}, Wilson tol 10^{-2}, sign fct with explicit deflation and relaxed tol
- optimal $m_0^W \approx 0.16$

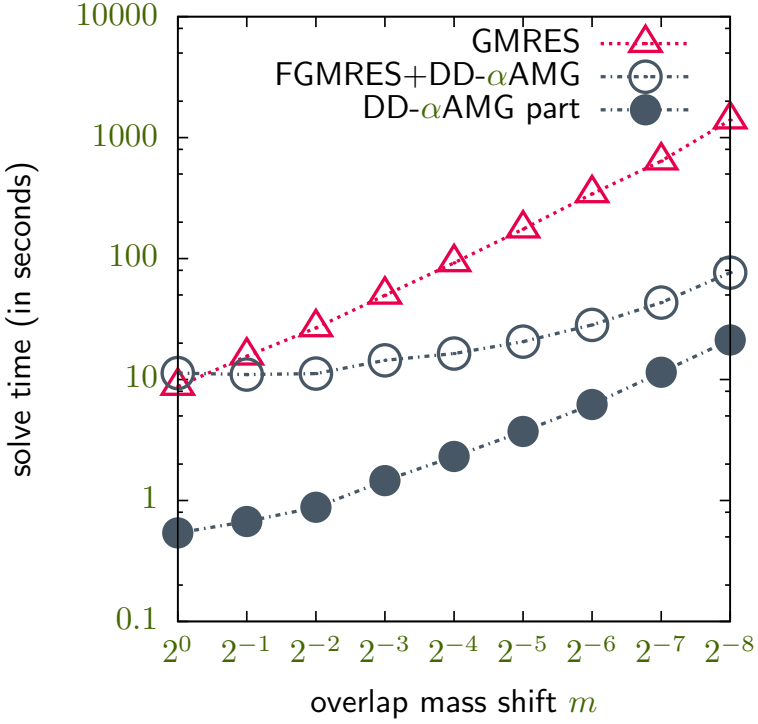
Smearing Study

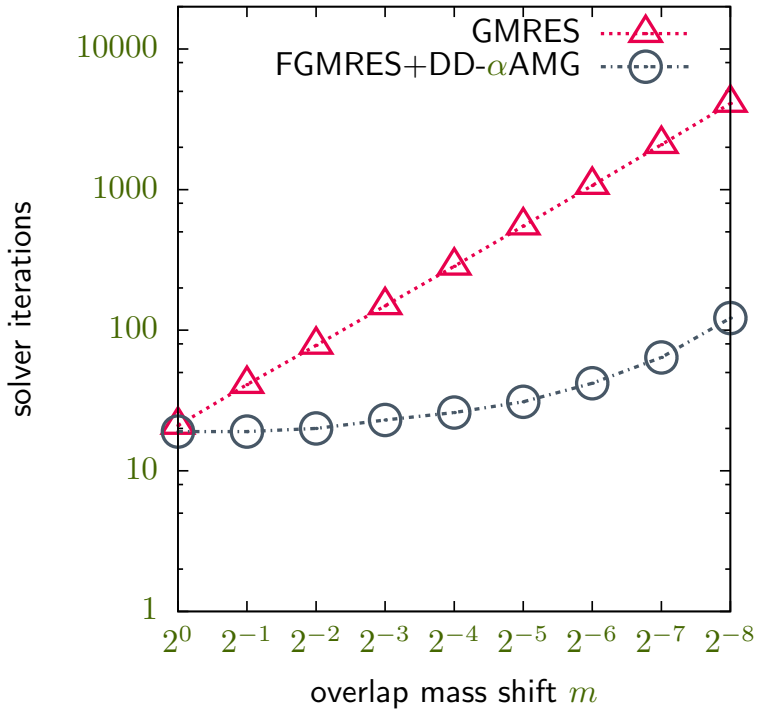




- 32^4 lattice, $1{,}024$ processes
- no smearing $\rightarrow$ $\times 5$ speedup
- 3–6 steps of stout smearing $\rightarrow$ $\times 20$–$\times 30$ speedup
- cost per iteration for preconditioned method only slightly higher
- preconditioner cost almost negligible

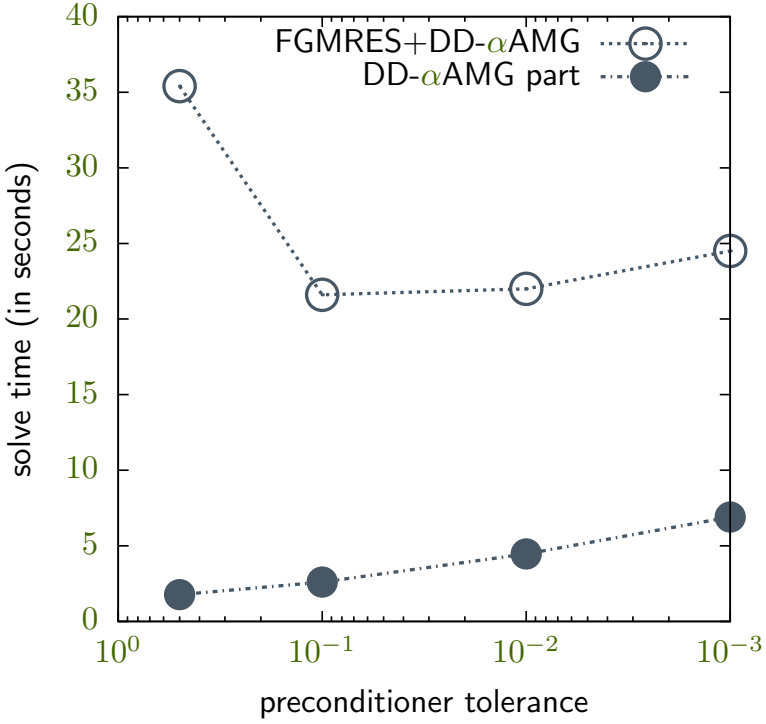
Scaling with the Overlap Mass Shift

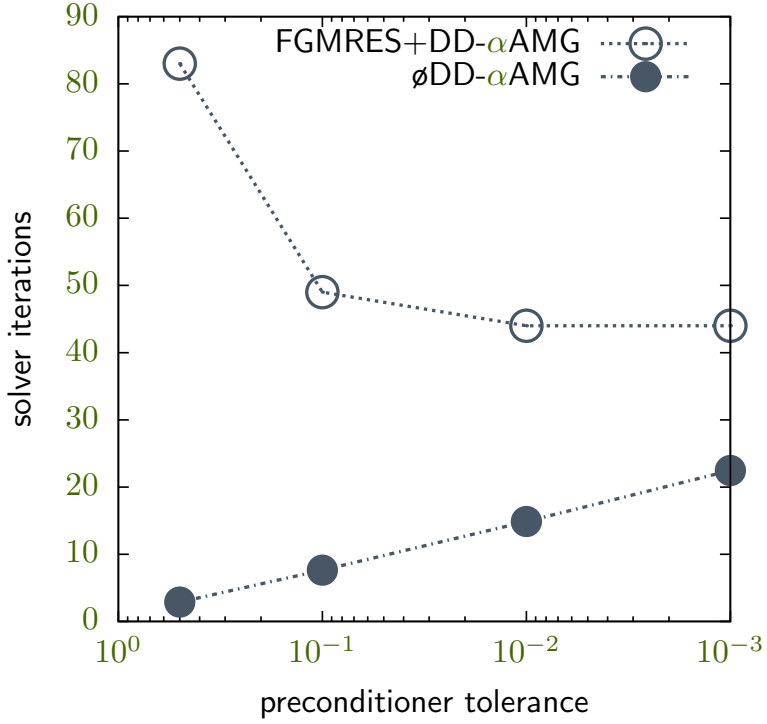




- 32^4 lat, 3HEX smeared BMW-c cnfg
- cnfg generated at approx. $m = 2^{-6}$ ($m_\pi \approx 350$ MeV)
- smaller masses $\longrightarrow$ bigger gain

Influence of the Preconditioner Accuracy





- 32^4 lat, 3HEX smeared BMW-c cnfg
- $tol = 10^{-2}$ optimal in terms of iteration count
- $tol = 10^{-1}$ optimal in terms of solve time

Solving the inverse square root

Challenge i): Evaluating $(H_W^\dagger H_W)^{-\frac{1}{2}} x$

Good convergence without explicit calculation of low modes of H_W?

Idea: Use implicit low mode information via *thick restarts* (cf. [Eiermann, Ernst, Güttel 2011]).

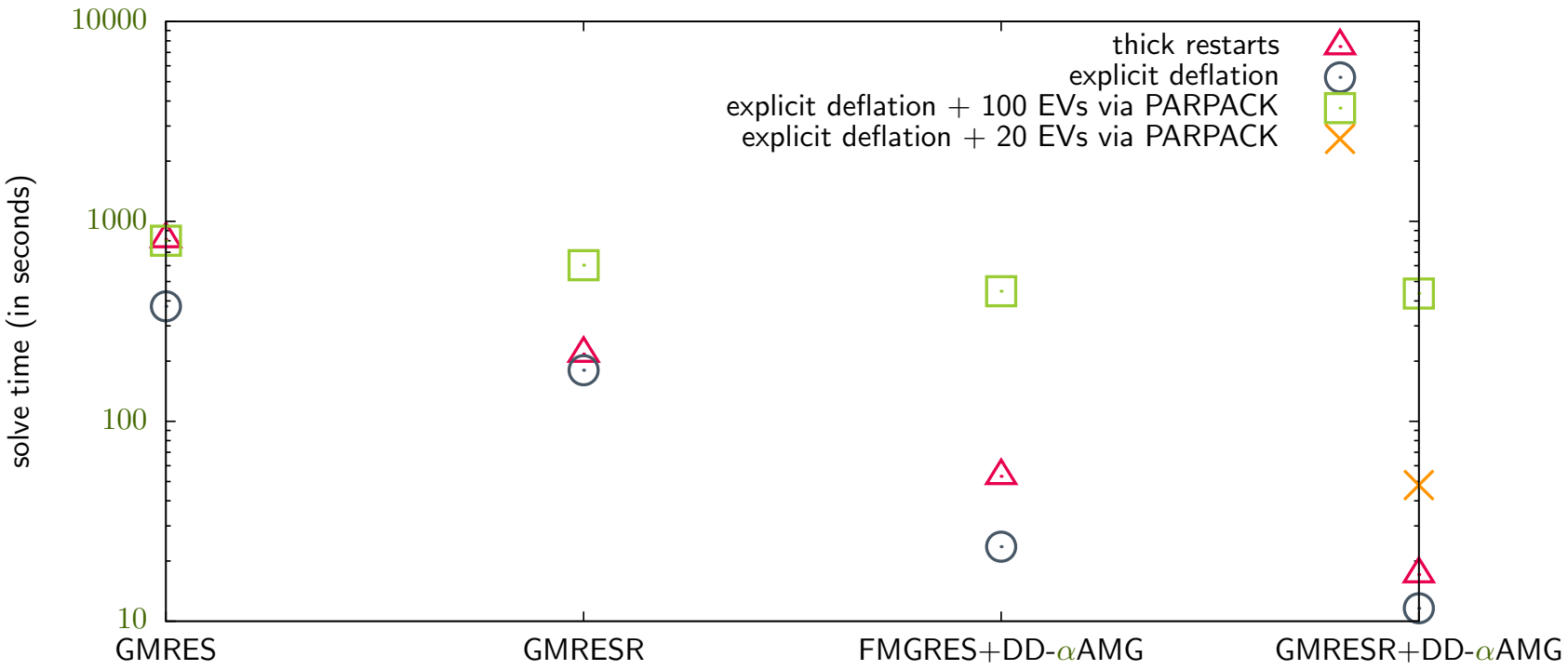
With the Cauchy integral representation

$$f(A) = A^{-\frac{1}{2}} = \int_\Gamma g(t)(tI - A)^{-1}\,\mathrm{d}t$$

and a Lanczos decomposition of $H_W^\dagger H_W$ we can compute the k-th error propagator by numerical quadrature (cf. [Frommer, Güttel, Schweitzer 2014]):

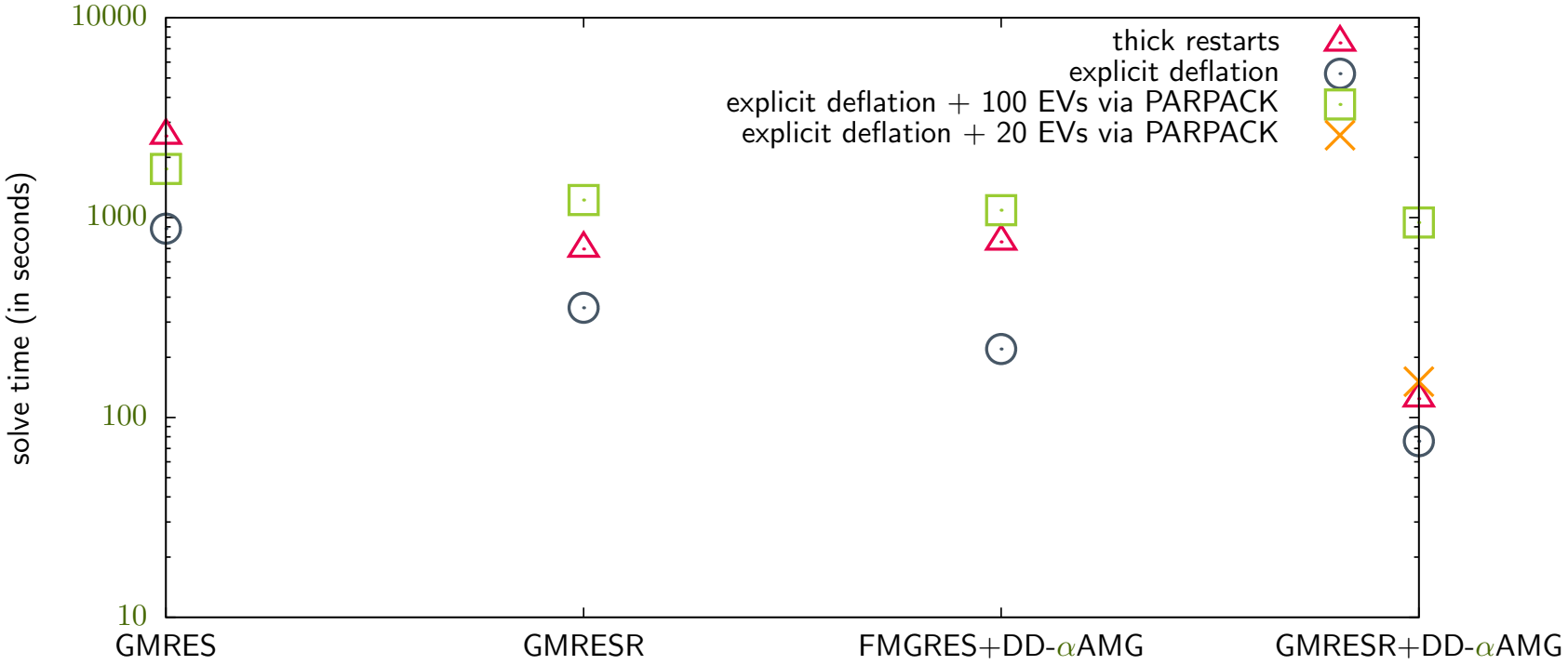
$$e^{(k)}(T) = c \sum_{i=1}^{l} \rho(T, x_i) \frac{\omega_i}{-\beta(1 - x_i) - T(1 + x_i)}$$

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

Thick Restarts and Explicit Deflation



- 32^4 lat, no smearing, $1{,}024$ cores
- one RHS: preconditioning + thick restarts
- many RHS: preconditioning + explicit deflation

Summary & Outlook

Summary:

- Preconditioning overlap equation leads to fewer iterations for the solution of $D_N \varphi = \eta$
- Preconditioner is cheap
- Efficiency of preconditioner improves
  - when approaching normality
  - for smaller masses
- For few RHS: thick restarts instead of EV computation

Outlook:

- Incoorporate solver into production codes of collaborators
- Further optimization of preconditioner
- Overall performance improvement of the method

Acknowledgments

All results computed on JUROPA at
Jülich Supercomputing Centre (JSC)

Partially funded by Deutsche Forschungsgemeinschaft (DFG),
Transregional Collaborative Research Centre 55 (SFB TR 55)

All configurations provided by BMW-c