V European Conference on Computational Fluid Dynamics
ECCOMAS CFD 2010
J. C. F. Pereira and A. Sequeira (Eds)
Lisbon, Portugal,14-17 June 2010

SCALABLE ROBUST SOLVERS FOR UNSTRUCTURED FE GEODYNAMIC MODELING APPLICATIONS: SOLVING THE STOKES EQUATION FOR MODELS WITH LARGE LOCALIZED VISCOSITY CONTRASTS IN 3D SPHERICAL DOMAINS

T. Geenen [*], M. ur Rehman[†], S. P. MacLachlan[‡], G. Segal[†], C. Vuik[†], A. P. van den Berg[*], W. Spakman[*]

[*] Institute of Earth Science, Utrecht University
Budapestlaan 4, 3584 CD, Utrecht, Netherlands
e-mail: geenen@geo.uu.nl
[†]Faculty EEMCS, Delft Institute of Applied Mathematics, Delft University of Technology
Mekelweg 4, 2628 CD, Delft, Netherlands
[‡]Department of Mathematics, Tufts University
Bromfield-Pearson Building, 503 Boston Avenue, Medford, Massachusetts 02155, USA

Key words: Preconditioners, Incompressible Stokes, Unstructured grid, Algebraic multigrid, Parallel solvers, Geodynamics

Abstract. *The development of scalable robust solvers for unstructured finite element applications related to viscous flow problems in earth sciences is an active research area. Solving high-resolution convection problems with order of magnitude* 10^8 *degrees of freedom requires solvers that scale well, both with respect to the number of degrees of freedom as well as having optimal parallel scaling characteristics on computer clusters. We investigate the use of a smoothed aggregation (SA) algebraic multigrid (AMG) type solution strategy to construct efficient preconditioners for the Stokes equation. We integrate AMG in our solver scheme as a preconditioner to the conjugate gradient method (CG) used during the construction of a block triangular preconditioner (BTR) to the Stokes equation, accelerating the convergence rate of the generalized conjugate residual method (GCR). We abbreviate this procedure as BTA-GCR.*
For our experiments, we use unstructured grids with linear(3D) and quadratic(2D) finite elements, making the model flexible with respect to geometry and topology. We find that AMG type methods scale linearly $(O(n))$*, with respect to the number of degrees of freedom,* n*. Although not all parts of AMG have preferred parallel scaling characteristics, we show that it is possible to tune AMG, resulting in parallel scaling characteristics that we consider optimal, for our experiments with up to 100 million degrees of freedom. Furthermore, AMG-type methods are shown to be robust methods, allowing us to solve very ill-conditioned systems resulting from strongly varying material properties over short distances in the model interior.*

1 INTRODUCTION

Solving the Stokes equation is the main time-consuming computation in mantle convection applications. The transition from 2D to 3D model computations has only worsened this situation, due to the suboptimal scaling of popular solver implementations with the number of degrees of freedom; such as ILU type preconditioned CG or parallel direct solvers[1] (by suboptimal, we mean that the amount of effort spent in solving the system of equations does not scale linearly with the number of degrees of freedom). The use of classical geometric multigrid (GMG) type methods has overcome this issue[2], but suffers from erratic robustness characteristics and constraints with respect to the geometry and topology of the model domain. In its conventional form (e.g. with Gauss-Seidel relaxation and linear interpolation), the performance of the GMG method usually deteriorates drastically when applied to problems more difficult than a constant coefficient Poisson-type equation[3]. AMG-type solution strategies[4] do not suffer from these limitations and allow for arbitrary discretization strategies. This is an advantage when the Stokes equation is solved with the finite element method (FEM), since this method implicitly allows for arbitrary domain geometries.

A key aspect of solving systems of equations, $\mathbf{Ax}=\mathbf{b}$, with an iterative method is the use of a preconditioner, $\mathbf{P}$. Without the use of a suitable preconditioner, the convergence of any solver will be very slow if $\mathbf{A}$ has an unfavorable eigenvalue spectrum. A preconditioner should improve the eigenvalue spectrum properties of $\mathbf{A}$ prior to solving, by applying it as $\mathbf{P^{-1}Ax} = \mathbf{P^{-1}b}$. An additional requirement for a suitable preconditioner is that it must be computationally inexpensive to construct and apply, with respect to the time spent in solving the unpreconditioned system.

To solve problems with a large number of degrees of freedom within a reasonable amount of wall time, the use of parallel computers is necessary. To solve problems on parallel systems, we subdivide the model domain into so-called sub domains, containing mutually disjoint (not overlapping) subsets of finite elements. The parallel efficiency of the method is determined by how well a solution method, executed on a specific hardware configuration, can exploit the parallel layout (domain decomposition) of the problem, with respect to end-to-end runtime reduction.

Several authors have recently presented solution methods to solve the Stokes equation in large-scale mantle-convection applications on parallel computers.[5] use a parallel direct solver,[6,7,8] and[9,10] use geometric multigrid while[11,12] and[13] propose to use a block preconditioned Krylov method.

Here, we present an alternative solution scheme to solve the Stokes equation. We follow a somewhat similar strategy as described in[11,12] and[13], in that we use a block preconditioner, the operation of which is approximated by a limited accuracy solve, to accelerate a Krylov solver for the Stokes equation. Our approach differs from previous work in that each element of the solver scheme has optimal characteristics. The applied Krylov method (GCR) allows for a block triangular preconditioner, making the convergence twice as fast

compared to block diagonal preconditioners[14] , a characteristic of the preconditioner that is independent of the Krylov method, see Table 1. For the preconditioning operator construction, we use AMG as a preconditioner to CG resulting in a robust, scalable subsystem solver. We use spectrally equivalent blocks (with respect to the system to be solved) in the preconditioner, making the convergence independent of the problem size. As a result, our method shows a favorable combination of characteristics, i.e., linear scaling with the number of degrees of freedom and optimal scaling with the number of processing cores as well as being robust for large localized viscosity contrasts.

2 Description of the Solution method

2.1 Mathematical formulation of the problem

For the analysis of the scaling relations of BTA-GCR, we consider thermal convection in both a 2D Cartesian and 3D spherical model setup. We assume the fluid to be incompressible(Boussinesq approximation) and the Prandtl number to be infinite. For the scaling experiments, we focus on solving the non-dimensional Stokes equation and incorporate thermal effects only through a contribution to the right-hand side (rhs), with a given temperature field,

$$\partial_j \eta \left(\partial_j u_i + \partial_i u_j\right) - \partial_i p = RaT\delta_{iz} \tag{1}$$

with the incompressibility constraint

$$\partial_j u_j = 0 \tag{2}$$

Symbols used are defined in Table 2. For a list of abbreviations refer to Table 3.

2.2 Discretization

The solution of the Stokes equation is formulated in a weak form, approximated by the Galerkin formulation, and solved on an unstructured grid consisting of linear isoparametric finite elements.
We use linear stabilized tetrahedral elements (P1-P1), the so called mini element, in 3D and quadratic Taylor-Hood triangular elements (P2-P1) in 2D[15]. The resulting coupled system of equations for the velocity and the pressure with the incompressibility constraint leads to a saddle point system of the form,

$$\mathbf{A}\begin{bmatrix}\mathbf{u}\\ \mathbf{p}\end{bmatrix} = \begin{bmatrix}\mathbf{Q} & \mathbf{G^T}\\ \mathbf{G} & \mathbf{C}\end{bmatrix}\begin{bmatrix}\mathbf{u}\\ \mathbf{p}\end{bmatrix} = \mathbf{b} \tag{3}$$

in which $\mathbf{Q}$, the velocity stiffness matrix, is symmetric positive definite (SPD). $\mathbf{G}$ and $\mathbf{G^T}$ are associated with minus the divergence of the velocity and gradient of the pressure respectively[16]. $\mathbf{C}$ results from the stabilization of the finite elements[17] in 3D. For our 2D experiments with quadratic Taylor-Hood tetrahedral elements $\mathbf{C} = \mathbf{0}$.

	Problem size		
Preconditioner			
BD	16×16	32×32	64×64
Solver			
MINRES	8	8	8
GMRES	9	9	9
Preconditioner			
BTR			
Solver			
GMRES	5	5	5
GCR	5	5	5

Table 1: The number of iterations for the SOLCX experiment of [12]. For a viscosity jump of 10^6. We compare the Block diagonal preconditioner used in [11] with a block triangular preconditioner, our preferred approach. We observe that the number of iterations doubles with BD when compared to BTR.

Symbol	Meaning
Ra	Thermal Rayleigh number $\frac{\rho\alpha\Delta T h^3}{\eta\kappa}$
T	Temperature
p	Pressure
η	viscosity
E	activation energy
R	Gasconstant
u	velocity
$\mathbf{A}$	constrained Stokes system
$\mathbf{Q}$	velocity subsystem
$\mathbf{G}$	gradient operator(FE context)
$\mathbf{G^T}$	divergence operator(FE context)
$\mathbf{M_p}$	pressure mass matrix
ϕ	basis function
Ω	model domain
$\mathbf{P}$	preconditioner
$\mathbf{S}$	Schur complement
$\mathbf{b}$	right hand side
$\mathbf{x}$	solution vector
$\mathbf{r}$	residual
h	element discretization size
n	number of degrees of freedom
E_p^n	Parallel efficiency

Table 2: List of symbols

2.3 BTA-GCR method

We employ a Krylov method combined with a so-called block preconditioner for the saddle point problem (3) [18], based on an incomplete block triangular factorization of the matrix $\mathbf{A}$ of the form,

$$\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{G^T} \\ 0 & -\tilde{\mathbf{S}} \end{bmatrix} \tag{4}$$

With $\tilde{\mathbf{S}}$ an approximation to the Schur complement, $\mathbf{G Q^{-1} G^T} + \mathbf{C}$.
We solve the saddle point problem arising from the constrained Stokes equation (3) with a Krylov method, GCR [19], right preconditioned (postconditioned) with a block triangular preconditioner (BTR) [20].
The GCR algorithm listed in Table 4 contains the block triangular preconditioner $\mathbf{P}$ in the term $\mathbf{P^{-1} r^k}$.
Instead of constructing $\mathbf{P^{-1}}$ explicitly, and applying it to the residual $\mathbf{r} = \mathbf{b} - \mathbf{Ax}$, we solve

$$\begin{aligned} \mathbf{Ps} &= \mathbf{r} \\ \mathbf{s} &= [\mathbf{s_1}; \mathbf{s_2}] \end{aligned} \tag{5}$$

List of abbreviations	
AMG	algebraic multigrid method
BTA-GCR	BTR with CG preconditioned with AMG as inner and GCR as outer solver
BTR	block triangular preconditioner
CG	conjugate gradient, a Krylov solver
FEM	finite element method
FGMRES	flexible generalized minimal residual , a Krylov solver
GCR	generalized conjugate residual, a Krylov solver
GMG	geometrical multigrid method
GMRES	generalized minimal residual , a Krylov solver
LU	lower- and upper-triangular decomposition
ILU	incomplete lower- and upper-triangular decomposition
MINRES	minimal residual , a Krylov solver
MUMPS	Multifrontal Massively Parallel Sparse direct Solver
SPD	symmetric positive definite
SA	smoothed aggregation, an algebraic multigrid method

Table 3: List of abbreviations

$$\mathbf{r} = [\mathbf{r_1}; \mathbf{r_2}]$$

resulting in the distributed solution scheme,

$$\tilde{\mathbf{S}}\mathbf{s_2^{k+1}} = \mathbf{r_2^k} \tag{6}$$

$$\mathbf{Q}\mathbf{s_1^{k+1}} = \mathbf{r_1^k} - \mathbf{G^T}\mathbf{s_2^{k+1}} \tag{7}$$

We take $\mathbf{M_p}$, the pressure mass matrix[16], scaled with the inverse of viscosity as an approximation to the Schur complement $\tilde{\mathbf{S}}$, which is spectrally equivalent.[21] proved this for the isoviscous case. The proof for variable viscosity is provided by[22] . This proof only holds under certain smoothness conditions of the viscosity variations and only for stable elements. To our knowledge, a proof for the spectral equivalence of the Schur complement for stabilized elements with $\mathbf{M_p}$ has not been published. We present numerical support that, for the viscosity contrasts we consider in our experiments, this relation holds for unstable elements. In this case we add $\mathbf{C}$ to $\mathbf{M_p}$. $\mathbf{M_p}$ is scaled with the viscosity during assembly, $M_{pi,j} = \int_\Omega \frac{1}{\eta}\phi_i\phi_j dA$, where ϕ_i are the pressure basis functions. Using the scaled pressure mass matrix, guarantees h (i.e. element size)-independent convergence of the Krylov method for system (3)[23]. The use of the pressure mass matrix is known to be sensitive for elongated computational domains and element shapes, resulting in a larger number of iterations[23]. However, for typical domains and element shapes used in our experiments, this effect is not observed. Experiments by[11] support our observation that this also holds for 3D spherical domains.

$\mathbf{r^0} = \mathbf{b} - \mathbf{A}\mathbf{x^0}$
for $k = 0, 1, ..., ngcr$
  $\mathbf{s^{k+1}} = \mathbf{P^{-1}}\mathbf{r^k}$
  $\mathbf{v^{k+1}} = \mathbf{A}\mathbf{s^{k+1}}$
  for $i = 0, 1, ..., k$
    $\mathbf{v^{k+1}} = \mathbf{v^{k+1}} - (\mathbf{v^i}, \mathbf{v^{k+1}})\mathbf{v^i}$
    $\mathbf{s^{k+1}} = \mathbf{s^{k+1}} - (\mathbf{v^{k+1}}, \mathbf{v^i})\mathbf{s^i}$
  end for
  $\mathbf{v^{k+1}} = \mathbf{v^{k+1}} / \|\mathbf{v^{k+1}}\|_\mathbf{2}$
  $\mathbf{s^{k+1}} = \mathbf{s^{k+1}} / \|\mathbf{v^{k+1}}\|_\mathbf{2}$
  $\mathbf{x^{k+1}} = \mathbf{x^k} + (\mathbf{v^{k+1}}, \mathbf{r^k})\mathbf{s^{k+1}}$
  $\mathbf{r^{k+1}} = \mathbf{r^k} - (\mathbf{v^{k+1}}, \mathbf{r^k})\mathbf{v^{k+1}}$
end for

Table 4: Postconditioned GCR algorithm, taken from [24]

$\mathbf{r^0} = \mathbf{b} - \mathbf{A}\mathbf{x^0}$
$\mathbf{s^0} = \mathbf{P^{-1}}\mathbf{r^0}$
$\mathbf{v^0} = \mathbf{s^0}$
for $k = 0, 1, ..., ncg$
  $\alpha_\mathbf{k} = \frac{\mathbf{r_k^T s_k}}{\mathbf{v_k^T A v_k}}$
  $\mathbf{x_{k+1}} = \mathbf{x_k} + \alpha_\mathbf{k}\mathbf{v_k}$
  $\mathbf{r_{k+1}} = \mathbf{r_k} - \alpha_\mathbf{k}\mathbf{A}\mathbf{v_k}$
  if r_{k+1} **sufficiently small exit**
  $\mathbf{s_{k+1}} = \mathbf{P^{-1}}\mathbf{r_{k+1}}$
  $\beta_\mathbf{k} = \frac{\mathbf{r_{k+1}^T s_{k+1}}}{\mathbf{r_k^T v_k}}$
  $\mathbf{v_{k+1}} = \mathbf{s_{k+1}} - \beta_\mathbf{k}\mathbf{v_k}$
end for

Table 5: Preconditioned CG algorithm the action of $\mathbf{P^{-1}}$ is approximated by an inaccurate solve of $\mathbf{s}$ from $\mathbf{P}\mathbf{s_{k+1}} = \mathbf{r_{k+1}}$

2.3.1 Solving the subsystems for the preconditioner

We employ AMG from the ML library [25] as a preconditioner to CG for the approximate solution of the subsystems during the preconditioner construction phase, Equations (6) and (7). We use AMG as a preconditioner, rather than as a solver, based on the robustness of this approach, resulting in faster convergence [26]. This is especially relevant for localized viscosity anomalies [27]. Using AMG as a preconditioner to CG for the subsystem solution guarantees h-independent convergence of the solver during the preconditioner construction phase. Figure 1 illustrates the fixed number of CG iterations for increasing problem size. The efficiency of AMG as a preconditioner to CG for ill-conditioned symmetric positive definite (SPD) systems, arising in geodynamic applications, on parallel computers was previously shown by [28,29]. One AMG V-cycle is used to precondition $\mathbf{Q}$. Again we do not explicitly construct the preconditioner $\mathbf{P}$, but approximate $\mathbf{P^{-1}}$ by an inaccurate solve of $\mathbf{s}$ from $\mathbf{P}\mathbf{s} = \mathbf{r}$, Table 5. In Figure 3 we present a flow diagram illustrating the steps that are taken to solve the velocity part of system 3, illustrating the relation between the different parts of the solution method.

2.3.2 Solving the BTR preconditioned system

The Krylov method used to solve the preconditioned saddle point problem (3), must be able to handle an asymmetric preconditioner, since the BTR preconditioner is asymmetric. For a thorough discussion on the subject of suitable Krylov methods for BTR preconditioners, we refer to [12] Section 1.1
MINRES, the Krylov solver used by [11] and [13], is not designed to handle asymmetric preconditioners. Their choice to use MINRES with a block diagonal (BD) preconditioner results in approximately twice the number of iterations compared to BTR [14]. Our experiment comparing BD with MINRES and BTR with GCR, Table 1, show that this is a robust feature of the preconditioner

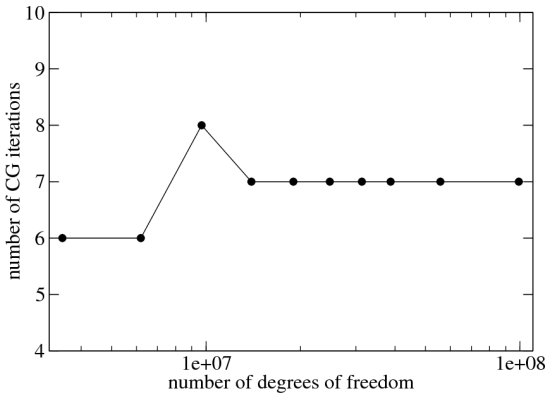

Figure 1: The solution of equation 7 with CG preconditioned with one AMG V-cycle scales optimal with the number of degrees of freedom. A difference of one iteration between experiments with different number of degrees of freedom can be discarded

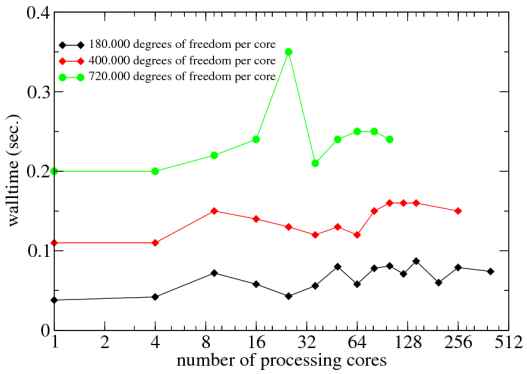

Figure 2: The scaling of CG with increasing number of processing cores. The number of degrees of freedom is kept fixed per processing core. The timing results are for a single CG iteration step. Although we observe a sometimes erratic wall time behavior with respect to the number of processing cores, the general trend is linear up to several hundred processing cores. The maximum numbers of degrees of freedom for these experiments with 180,000, 400,000, and 720,000 degrees of freedom per processing core are 72, 102.4, and 72 million, respectively.

We use GCR, a Krylov method similar to FGMRES[30,24], for the solution of (3) right preconditioned by (4). Since GCR has an increasing storage requirement with increasing number of iterations, unlike CG or MINRES, this method can become impracticable when large number (more than several tens) of iterations are needed to solve the system to an acceptable accuracy. In our proposed solution method, we can keep the number of iterations for GCR low, not more than 30, since we solve the velocity subsystem in the preconditioner phase to a high accuracy, see Table 6 and Table 7. This has the added advantage that we do not have to apply the preconditioner for the velocity subsystem as often as would be the case with a lower accuracy subsystem solve. For a comprehensive analysis of the characteristics of BTR as a preconditioner to GCR for the Stokes equation, we refer to[31,32]. They show that BTR is a robust preconditioner for the SINKER model of[12] and an example from an aluminium extrusion model. They also explain in more detail the way the pressure mass matrix is scaled with viscosity.

3 Numerical experiments and performance tests

We performed experiments to analyze the scaling of BTA-GCR with,

- number of degrees of freedom
- number of parallel processing cores

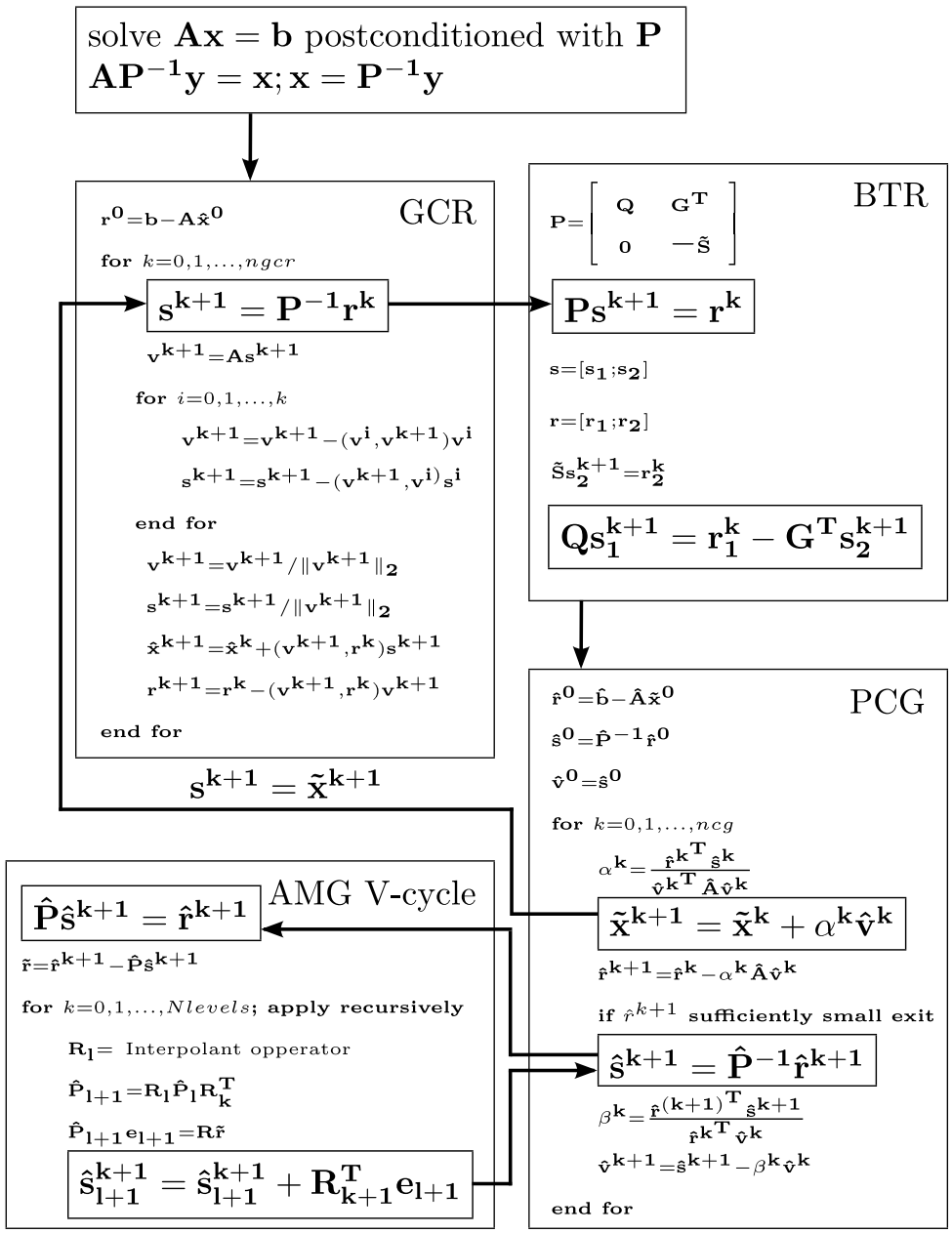

Figure 3: We illustrate the relation between the different solver and preconditioner steps to solve for the velocity part of system 3. We solve for $\mathbf{x}$ by applying a postconditioner $\mathbf{P}$ to system 3 and solving with GCR. the effect of applying the preconditioner $\mathbf{P}$(BTR) during the GCR iterations is approximated by an incomplete solve of $\mathbf{s^{k+1}}$ from $\mathbf{Ps^{k+1}}=\mathbf{r^k}$. This results in the distributed solution scheme 5. We solve the velocity part of this distributed scheme with PCG. during the iteration of PCG the effect of applying the preconditioner $\hat{\mathbf{P}}$ is approximated by applying a single V-cycle of AMG to $\hat{\mathbf{s}}$. The approximate solution for the velocity in PCG is substituted in the part of $\mathbf{s^{k+1}}$ associated with the velocity unknowns in $\mathbf{x}$.

	iterations GCR/CG (walltime in seconds)				
number of elements	$\Delta\eta = 10$	10^3	10^6	10^7	10^8
32×32			5/11		
64×64	18(50)	19(52)	21(56)	28(72)	35(88)
64×64	6/8(31)	4/9(24)	5/11(30)	5/16(36)	5/20(40)
128×128			5/11		
256×256			5/11		

Table 6: We reproduced the SOLCX experiment from [12]. For a viscosity jump of 10^6, the number of iterations of the outer GCR and average number for inner CG solver, preconditioned with AMG, for the solution of the velocity subsystem (7) is h-independent. For increasing viscosity contrast, we observe a mild increase for both inner and outer iterations. For a large viscosity contrast, the amount of inner iterations increases sharply. In brackets we present the total walltime for the GCR iteration, AMG setup, CG iteration and AMG application for a high accuracy inner CG solve for the velocity subsystem (six orders of magnitude relative residual reduction) and for a run where there is no inner CG solve, and we just apply the AMG V-cycle. Solving the velocity subsystem to a high accuracy approximately halves the walltime.

- order of magnitude viscosity contrasts

3.1 Scaling with respect to the number of degrees of freedom

For these scaling experiments, we choose a 2D isoviscous Cartesian setup based on the model by [33], in which a smooth initial temperature depth profile is perturbed by a small amplitude periodic temperature field, driving the thermal convection.
The h-independent scaling characteristics of both CG (preconditioned with one AMG V-cycle for the velocity subsystem of BTR , Equation 7) and GCR preconditioned with BTR, are shown in Figure 1 and Table 6, the number of iterations remain constant for increasing problem size. We observed this scaling relation, both for 2D and 3D model experiments. The 3D model experiments are performed for a similar setup with the exception of the model domain geometry, which consist of a spherical section.

3.2 Scaling with respect to the number of processing cores

For these scaling experiments, we choose the same problem setup as for the previous experiment.
We found that the parallel efficiency $E_p^n = \frac{T_s}{T_p^n}$ (with T_s the walltime for the sequential run and T_p^n walltime for parallel run with n processing cores and n times the number of degrees of freedom) scales almost linearly for both the application of the AMG V-cycle preconditioner as well as CG as shown in Figures 2 and 5. The setup of the AMG V-cycle does not scale linearly due to the non-linear scaling of both the analysis and the factorization part of the parallel direct solver [34] used on the coarsest grid, Figure 4. Fortunately, the AMG setup phase has to be performed only once per time step where

	b=0	b=0.2	b=0.3	b=0.4
accuracy outer/inner	Iterations(inner/outer) (walltime in seconds on 1/100 processing cores)			
10^{-6}/	46(113/264)	59(143/334)	57(138/323)	70(168/394)
$10^{-6}/10^{-2}$	38/6(135/266)	39/7(146/282)	38/8(151/284)	43/9(179/330)
$10^{-6}/10^{-3}$	29/8(117/220)	30/9(127/235)	33/10(146/265)	34/11(157/280)
$10^{-6}/10^{-4}$	27/9(115/213)	29/11(135/241)	30/13(152/263)	31/13(157/272)
$10^{-6}/10^{-5}$	26/11(122/218)	28/13(143/247)	29/15(160/269)	30/16(171/285)
$10^{-6}/10^{-6}$	26/12(128/224)	27/15(149/251)	28/17(166/274)	29/18(178/290)

Table 7: We analyze the robustness of BTA-GCR for sharp, large amplitude viscosity jumps resulting from the relation $T(x,y) = T(y)\left[\left(1-\frac{b}{2}\right)+br\right]$, with r a random distribution between 0 and 1. The walltime is minimal for inner accuracy $10^{-2}-10^{-3}$ for small values of b (in some cases just applying the preconditioner without solving with CG at all results in the lowest walltime). For larger values of b, it pays off to solve the inner problems with an accuracy $10^{-3}-10^{-4}$. In brackets, we present the total walltime for the GCR iteration, AMG setup, CG iteration and AMG application for a run on a single and a run on 100 processing cores with 720.000 degrees of freedom per processing core. This illustrates that it pays of to reduce the number of outer iterations when increasing the number of processing cores due to the (initial) non linear scaling of the preconditioner application phase. This can be achieved by solving the velocity subsystem to a higher accuracy.

the AMG application and CG solve have to be performed for each outer GCR iteration, Figure 3. This weak scaling relation is plotted for a 3D spherical model geometry in Figures 6, 7 and 8

3.2.1 Domain decomposition

For our 3D scaling experiments we look at strong and weak scaling characteristics. For our 3D weak scaling experiments we use parmetis to enhance an initial blockwise domain decomposition. For our strong scaling experiments we look at computational domains with a continuous outer boundary (a complete sphere). In this case we use a domain decomposition method based on the method by[35]. This method finds the minimum energy distribution of charged particles on the outer surface of a sphere. These positions are used to determine the domain boundaries on the outer surface from which volumes are creates by continuation of domain boundaries towards the center of the sphere[36]. We show strong scaling results for this domain decomposition strategies in Figure 9, 10 and 11 and .

3.3 Robustness

To test the robustness of BTA-GCR, we performed a number of experiments with different viscosity contrasts, similar to[33] and[12]. These include the following configurations, (1) a step function for the viscosity in the x-direction across element boundaries, (2) a random viscosity perturbation across element boundaries[33]. We kept the viscosity con-

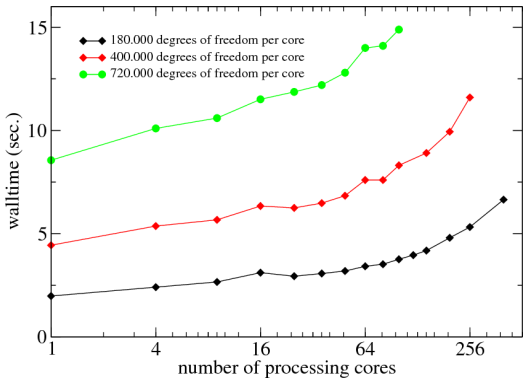

Figure 4: The amount of wall time for the AMG construction phase. With increasing number of processing cores the wall time increases nonlinearly. We can attribute this effect to the analysis phase of the direct solver, used to solve the coarse system, and to a lesser extent to its factorization phase. The other parts of the AMG setup phase have much weaker nonlinear scaling characteristics.

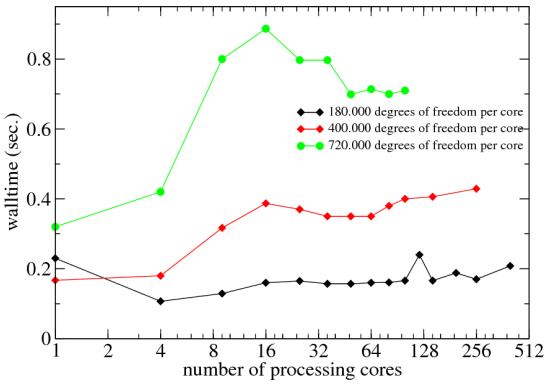

Figure 5: The amount of walltime for a single AMG V-cycle. We observe that after an initial increase in walltime with increasing number of processing cores, this relation flattens out to become roughly linear

stant per element, for all experiments, to prevent steep viscosity gradients in the element interior. Steep gradients in the interior of an element can otherwise be prevented by using an adequate, local, mesh resolution. For the first case, we reproduced the experiment SOLCX[12]. We solve the subsystems for the construction of the BTR with high relative accuracy $\frac{\|r_k\|}{\|r_0\|} \leq 10^{-5}$ and outer GCR solver with 10^{-6}. For increasing viscosity contrast, we observe that the number of outer iterations remains almost constant. The number of inner iterations to solve the BTR subsystems is constant for low viscosity contrasts, increases slightly for a high viscosity contrast, and increases sharply for very high viscosity contrast $\Delta\eta = 10^8$, see Table 6. However, models with extremely high viscosity contrast across a single element will result in inaccurate results for the FE method[37] and should therefore be avoided. The inaccuracy is most notable in high amplitude spurious pressure oscillations in the high viscosity area but is also, though to a much lesser extent, present in the solution for the velocity. Table 6 also shows the constant number of iterations with increasing problem size similar to our experiments with an isoviscous model. This illustrates in a numerical sense the spectral equivalency of the scaled PMM to the Schur complement. This was also shown by[11] and[31] .

The second experiment has the same setup as the scaling experiments in Sections 3.1 and 3.2, but this time the viscosity is temperature dependent through the relation $\eta = \eta_0 e^{\frac{E}{R}\left(\frac{1}{T}-\frac{1}{T_0}\right)}$ with R the gas constant and $E = 101.1 kJmol^{-1}$, the activation energy. A smooth background temperature field is randomly perturbed $T(x,y) = T(y)\left[\left(1-\frac{b}{2}\right)+br\right]$ with b a constant between 0 and 0.4 and r a random distribution between 0 and 1 which gives rise to viscosity jumps of up to 4.4 orders of magnitude across element boundaries.

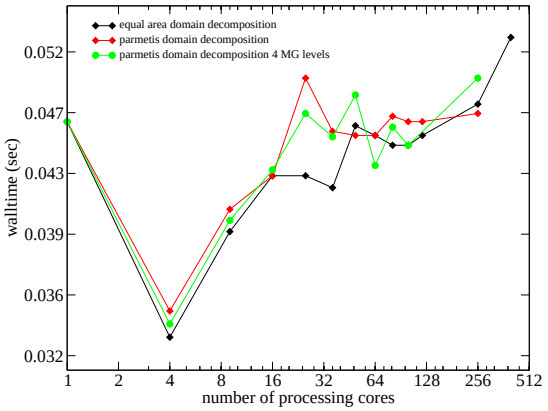

Figure 6: The scaling of CG with increasing number of processing cores in 3D. The number of degrees of freedom is kept fixed per processing core. The timing results are for a single CG iteration step without the preconditioner application step. The maximum numbers of degrees of freedom for these experiments is 55 million.

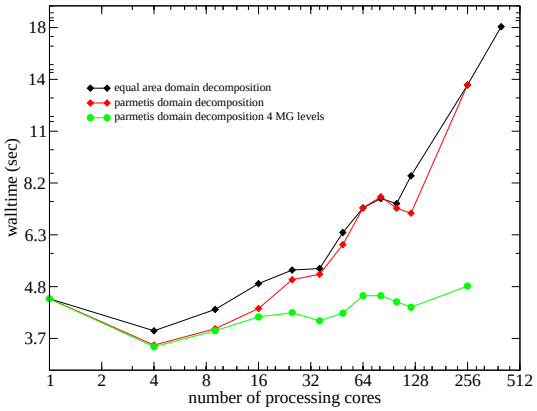

Figure 7: The amount of wall time for the AMG construction phase in 3D. With increasing number of processing cores the wall time increases nonlinearly. This is especially true when the coarse system is large relative to the total problem size. When we increase the number of coarse levels in the AMG V-cycle not only the total walltime decreases put also the parallel scaling characteristics become more favorable.

Table 7 shows that our solution method is only mildly sensitive to increasing viscosity contrasts across element boundaries. This experiment also illustrates the relation between the accuracy of the subsystem solution and the number of outer GCR iterations. For small viscosity jumps, the subsystem can be solved inaccurately (with tolerance $10^{-2} - 10^{-3}$) with a small number of inner iterations, without increasing the number of outer iterations, but for larger jumps the optimum, in terms of walltime, occur with an inner accuracy between $10^{-3} - 10^{-4}$.

4 Discussion and concluding remarks

We showed that BTA-GCR scales linearly with the number of degrees of freedom and has optimal scaling characteristics with increasing number of processing cores. We also showed that our method is robust with respect to large localized viscosity contrasts. In [38] we showed results for 2D model experiments that we extended to 3D in this study.
An essential part of BTA-GCR is the use of AMG as a preconditioner to CG during the preconditioner (BTR) construction phase which is the only scalable method currently known for unstructured grid models.
Recently authors have reported results with block preconditioners for saddle point problems in geodynamical applications, [12] [11] and [13]. The method employed by [12] uses a preconditioner for the pressure part based on the velocity matrix, leading to h-dependent scaling of the number of iterations making the method suboptimal for large scale models. [11] present a method that is the most closely related to our approach; however, their

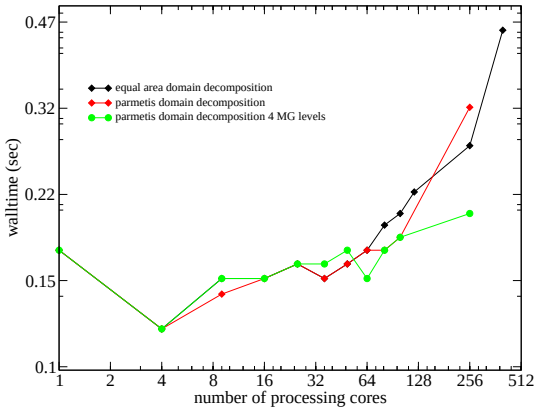

Figure 8: The amount of walltime for a single AMG V-cycle in 3D. We observe that after an initial increase in walltime with increasing number of processing cores, this relation flattens out to become roughly linear for a well balanced domain decomposition.

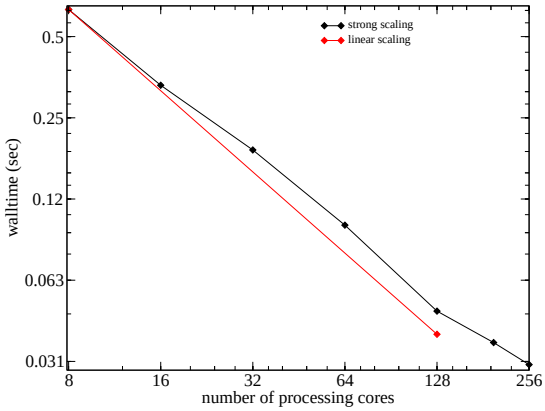

Figure 9: Strong scaling for a single CG iteration in 3D. The model size is 13.5×10^6 degrees of freedom. We observe a linear strong scaling relation up to 128 processing cores. It seems that there is an offset in the scaling between 16 and 32 cores that could be attributed to the number of processes we run per node, 16.

use of a block diagonal rather than a block triangular preconditioner makes the convergence rate on average twice as small for the outer Krylov solver[14], compared to block diagonal preconditioners, Table 1. The only extra operation for BTR preconditioners is one vector update and one matrix vector product, which are negligible compared to the overall solution scheme. Their results are, however, unique in the size of the problem they have solved and the number of processing cores they have employed for their calculation. This illustrates the potential of BTR in combination with AMG (i.e. BTA-GCR) to solve problems with several billion degrees of freedom efficiently on large numbers of processing cores in parallel[39].

[13] present a diagonal preconditioner where the sub-block for the velocity is preconditioned with its diagonal and a lumped pressure mass matrix scaled with the viscosity is used for the pressure block. This class of preconditioners can only be used efficiently for isoviscous models or models with minimal viscosity contrasts in the model interior. We found that for large viscosity contrast convergence of the outer Krylov solver depends critically on the approximation of the BTR preconditioner with respect to the solution of the velocity subsystem, Equation 7.

Our approach does not suffer from any of the limitations of the above mentioned approaches and scales optimally both with the number of degrees of freedom and with the number of processing cores. The method is able to handle large viscosity contrasts, to the extent that the numerical accuracy of the FE method is the limiting factor, not the convergence of the numerical scheme.

Modelling geodynamic processes that incorporate localization phenomena requires a high resolution model with robust scalable solvers. By employing AMG type methods to con-

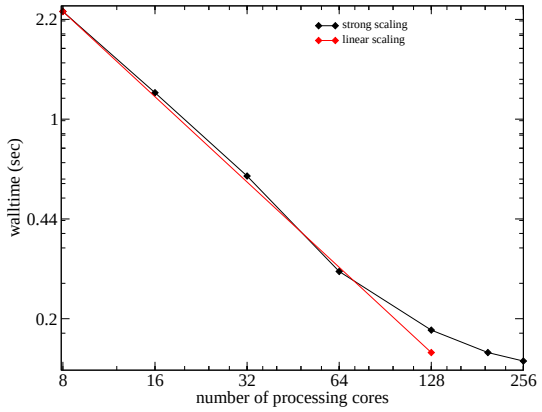

Figure 10: Strong scaling for a single AMG V-cycle in 3D. The model size is 13.5×10^6 degrees of freedom. We observe a linear strong scaling relation up to 64 processing cores. We attribute the drop off to the back substitution phase of the direct solver on the coarsest grid. This phase is performed by a limited number of processes, in our case 16, redundantly. Performing the back substitution on all processes in parallel is not practical considering the size of the coarse system, typically less than 1000 matrix rows.

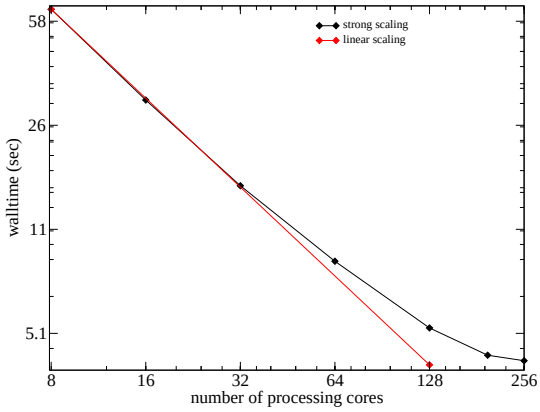

Figure 11: Strong scaling for the AMG construction phase in 3D. The model size is 13.5×10^6 degrees of freedom. We observe a linear strong scaling relation up to 32 processing cores. We attribute the drop off to the analysis and factorization phase of the direct solver on the coarsest grid.

struct a preconditioner to solve the constrained Stokes equation, we are able to solve this kind of problems within a reasonable walltime.

Acknowledgments

Part of this work was funded by NCF grant NRG-2005.05. This work was conducted under the ISES(Integrated Solid Earth Science) program. Computational resources were provided by NCF grant SH-021-07.
The work of Mehfooz ur Rehman was supported by HEC, Pakistan and Nuffic, Netherlands , contract Ref:2-17/PM-OS/Neth/2005.
The work of Scott MacLachlan was supported in part by the European Community's Sixth Framework Programme, through a Marie Curie International Incoming Fellowship, MIF1-CT-2006-021927.

References

[1] Geenen, T. and van den Berg, A. NCF project NRG-2005.05. The implementation of several parallel direct and iterative solvers/preconditioners in TDC, a finite element based thermo chemical convection code. Technical report, University Utrecht, (2006).

[2] Baumgardner, J. A 3-D treatment of convective flow in the earth's mantle. *Journal of Statistical Physics* , 501–511 (1984).

[3] Kettler, R. Analysis and comparison of relaxation schemes in robust multigrid and preconditioned conjugate gradient methods. In *Multigrid Methods*, 502–534. Springer, Heidelberg, Germany (1982).

[4] Vanek, P., Mandel, J., and Brezina, M. Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems. *Computing* **56**(3), 179–196 (1996).

[5] Braun, J., Thieulot, C., Fullsack, P., DeKool, M., Beaumont, C., and Huismans, R. DOUAR: a new three-dimensional creeping flow numerical model for the solution of geological problems. *Physics of the Earth and Planetary Interiors* **171**(1-4), 76–91.

[6] Choblet, G., Cadek, O., Couturier, F., and Dumoulin, C. OE/DIPUS: a new tool to study the dynamics of planetary interiors. *Geophysical Journal International* **170**(1), 9–30 (2007).

[7] Kameyama, M., Kageyama, A., and Sato, T. Multigrid-based simulation code for mantle convection in spherical shell using Yin-Yang grid. *Physics of the Earth and Planetary Interiors* **171**(1-4), 19–32 (2008).

[8] Zhong, S., Yuen, D., Moresi, L., and Schubert, G. Numerical methods for mantle convection. In *Treatise on Geophysics*, 227–252. Elsevier, Amsterdam (2007).

[9] Tackley, P. Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the Yin-Yang grid. *Physics of the Earth and Planetary Interiors* **171**(1-4), 7–18 (2008).

[10] Tackley, P. Effects of strongly temperature-dependent viscosity on time- dependent, three-dimensional models of mantle convection. *Geophysical Research Letters* **20**(20), 2187–2190 (1993).

[11] Burstedde, C., Ghattas, O., Gurnis, M., Stadler, G., Tan, E., Tu, T., Wilcox, L. C., and Zhong, S. Scalable adaptive mantle convection simulation on petascale supercomputers. In *Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, 1–15 (IEEE Press, Austin, Texas, 2008).

[12] May, D. and Moresi, L. Preconditioned iterative methods for Stokes flow problems arising in computational geodynamics. *Physics of the Earth and Planetary Interiors* **171**(1-4), 33–47 (2008).

[13] Schmid, D., Dabrowski, M., and Krotkiewski, M. Evolution of large amplitude 3D fold patterns: A FEM study. *Physics of the Earth and Planetary Interiors* **171**(1-4), 400–408, December (2008).

[14] Elman, H. and Silvester, D. Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations. *SIAM J. Sci. Comput* **17**, 33—46 (1996).

[15] Segal, A. *SEPRAN Manual: Standard Problems.* SEPRA Analysis, Leidschendam, Netherlands, (2005).

[16] Cuvelier, C., Segal, A., and van Steenhoven, A. *Finite element methods and Navier-Stokes equations.* D. Reidel, Dordrecht, Netherlands, (1986).

[17] Pierre, R. Simple C^0 approximations for the computation of incompressible flows. *Comput. Methods Appl. Mech. Eng.* **68**(2), 205–227 (1988).

[18] Benzi, M., Golub, G., and Liesen, J. Numerical solution of saddle point problems. *Acta Numerica* **14**, 1–137 (2005).

[19] van der Vorst, H. and Vuik, C. GMRESR: a family of nested GMRES methods. *Num. Lin. Alg. Appl.* **1**, 369–386 (1994).

[20] Bramble, J. H. and Pasciak, J. E. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Mathematics of Computation* **50**(181), 1–17 (1988).

[21] Verfürth, R. A combined conjugate gradient-multigrid algorithm for the numerical solution of the Stokes problems. *IMA J. NUMER. ANAL.* **4**(4), 441–455 (1984).

[22] Grinevich, P. P. and Olshanskii, M. A. An iterative method for the Stokes-Type problem with variable viscosity. *SIAM J. Sci. Comput.* **31**(5), 3959–3978 (2009).

[23] Janka, A. Smoothed aggregation multigrid for a Stokes problem. *Computing and Visualization in Science* **11**(3), 169–180 (2008).

[24] Vuik, C. New insights in GMRES-like methods with variable preconditioners, (1995).

[25] Gee, M., Siefert, C., Hu, J., Tuminaro, R., and Sala, M. ML 5.0 smoothed aggregation user's guide. Technical Report SAND2006-2649, Sandia National Laboratories, Albuquerque, N. M, (2006).

[26] Oosterlee, C. and Washio, T. An evaluation of parallel multigrid as a solver and a preconditioner for singularly perturbed problems. *SIAM Journal of Scientific Computing* **19**(1), 87–110 (1998).

[27] MacLachlan, S., Tang, J., and Vuik, C. Fast and robust solvers for pressure-correction in bubbly flow problems. *Journal of Computational Physics* **227**(23), 9742–9761 (2008).

[28] Geenen, T., van den Berg, A., and Spakman, W. Large scale, high resolution, mantle dynamics modeling. American Geophysical Union, Fall Meeting 2007, abstract T51B-0546, (2007).

[29] Geenen, T., van den Berg, A., and Spakman, W. Scalable robust solvers for 3d unstructured modeling applications, solving the Stokes equation with large, localized, viscosity contrasts. European Geoscience Union, General Assembly 2008, EGU2008-A-09460; GD1-1MO3O-005, (2008).

[30] Saad, Y. *Iterative methods for sparse linear systems*. SIAM, Philadelphia, Pa., (2003).

[31] ur Rehman, M., Geenen, T., Vuik, C., Segal, G., and MacLachlan, S. P. On iterative methods for the incompressible Stokes problem. *International Journal for Numerical Methods in Fluids* (2010).

[32] ur Rehman, M. *Fast Iterative Methods for The Incompressible Navier-Stokes Equations*. PHD Thesis, Delft University of Technology, The Netherlands, (2010).

[33] Yang, W. and Baumgardner, J. A matrix-dependent transfer multigrid method for strongly variable viscosity infinite Prandtl number thermal convection. *Geophysical and Astrophysical Fluid Dynamics* **92**(3-4), 151–195 (2000).

[34] P.R. Amestoy, I. S. Duff, J. K. and L'Excellent, J. *Multifrontal Massively Parallel Solver (MUMPS Version 4.6.3) Users guide*, (2006). `http://graal.ens-lyon.fr/MUMPS/index.html`.

[35] Hüttig, C. and Stemmer, K. The spiral grid: A new approach to discretize the sphere and its application to mantle convection. *Geochem. Geophys. Geosyst.* **9** (2008).

[36] van Thienen, P. Some tools for the generation of finite element meshes in faulted 3-D domains for geodynamical applications. Technical report, University Utrecht, (2008).

[37] L. Moresi, S. Z. and Gurnis, M. The accuracy of finite element solutions of Stokes flow with strongly varying viscosity. *Physics of the Earth and Planetary Interiors* **97**, 83–94 (1996).

[38] Geenen, T., ur Rehman, M., MacLachlan, S., Segal, G., Vuik, C., van den Berg, A., and Spakman, W. Scalable robust solvers for unstructured FE geodynamic modeling applications: Solving the stokes equation for models with large localized viscosity contrasts. *Geochemistry, Geophysics, Geosystems* **10**(9) (2009).

[39] Burstedde, C., Ghattas, O., Stadler, G., Tu, T., and Wilcox, L. C. Parallel scalable adjoint-based adaptive solution of variable-viscosity Stokes flow problems. *Computer Methods in Applied Mechanics and Engineering* **198**(21-26), 1691–1700, May (2009).