Geochemistry Geophysics Geosystems G3
AN ELECTRONIC JOURNAL OF THE EARTH SCIENCES
Published by AGU and the Geochemical Society

Technical Brief
Volume 10, Number 9
3 September 2009
Q09002, doi:10.1029/2009GC002526
ISSN: 1525-2027

Scalable robust solvers for unstructured FE geodynamic modeling applications: Solving the Stokes equation for models with large localized viscosity contrasts

T. Geenen
Institute of Earth Science, Utrecht University, Budapestlaan 4, NL-3584 CD, Utrecht, Netherlands (geenen@geo.uu.nl)

M. ur Rehman
Faculty EEMCS, Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, NL-2628 CD, Delft, Netherlands

S. P. MacLachlan
Department of Mathematics, Tufts University, Bromfield-Pearson Building, 503 Boston Avenue, Medford, Massachusetts 02155, USA

G. Segal and C. Vuik
Faculty EEMCS, Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, NL-2628 CD, Delft, Netherlands

A. P. van den Berg and W. Spakman
Institute of Earth Science, Utrecht University, Budapestlaan 4, NL-3584 CD, Utrecht, Netherlands

[1] 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, with respect to both 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 quadratic finite elements, making the model flexible with respect to geometry and topology and $O(h^3)$ accurate. 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.

Components: 6738 words, 4 figures, 6 tables.

Keywords: preconditioners; incompressible Stokes; unstructured grid; algebraic multigrid; parallel solvers; viscosity contrast.

Index Terms: 0535 Computational Geophysics: Hardware solutions; 0545 Computational Geophysics: Modeling (4255); 0560 Computational Geophysics: Numerical solutions (4255); 3225 Mathematical Geophysics: Numerical approximations and analysis (4260); 8121 Tectonophysics: Dynamics: convection currents, and mantle plumes.

Received 27 March 2009; **Revised** 7 July 2009; **Accepted** 27 July 2009; **Published** 3 September 2009.

Geenen, T., M. ur Rehman, S. P. MacLachlan, G. Segal, C. Vuik, A. P. van den Berg, and W. Spakman (2009), Scalable robust solvers for unstructured FE geodynamic modeling applications: Solving the Stokes equation for models with large localized viscosity contrasts, *Geochem. Geophys. Geosyst.*, *10*, Q09002, doi:10.1029/2009GC002526.

1. Introduction

[2] Solving the Stokes equation is the main time-consuming computation in mantle convection applications. The transition from 2-D to 3-D model computations has only worsened this situation, owing 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 [*Geenen and van den Berg*, 2006] (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 [*Baumgardner*, 1985], 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 [*Kettler*, 1982]. AMG-type solution strategies [*Vaněk et al.*, 1996] 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.

[3] 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 for matrices that are ill-conditioned or have an unfavorable eigenvalue spectrum. By ill-conditioned, we refer to the presence of both large and small eigenvalues in $\mathbf{A}$. The ratio of the smallest to the largest eigenvalue as well as the eigenvalue spectrum determines to a large extent how well an iterative solver converges. Even a single small eigenvalue can stall convergence for many iterations. A preconditioner is an operator that improves the condition and eigenvalue spectrum properties of the matrix by reducing the extreme eigenvalues prior to solving, by applying it as $\mathbf{P}^{-1}\mathbf{Ax} = \mathbf{P}^{-1}\mathbf{b}$. A requirement of 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.

[4] To solve problems with a large number of degrees of freedom within a reasonable amount of time, the use of parallel computers is necessary. To solve problems on parallel systems, we subdivide the model domain into so-called subdomains, 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.

[5] Several authors have recently presented solution methods to solve the Stokes equation in large-scale mantle convection applications on parallel computers. *Braun et al.* [2008] use a parallel direct solver, *Choblet et al.* [2007], *Kameyama et al.* [2008], *Zhong et al.* [2007] and *Tackley* [1993, 2008] use geometric multigrid while *Burstedde et al.* [2008], *May and Moresi* [2008] and *Schmid et al.* [2008] propose to use a block preconditioned Krylov method.

[6] Here, we present an alternative solution scheme to solve the Stokes equation. We follow a somewhat similar strategy as described by *Burstedde et al.* [2008], *May and Moresi* [2008] and *Schmid et al.* [2008], 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 con-

Table 1. Number of Iterations for the SOLCX Experiment of *May and Moresi* [2008][a]

	Problem Size		
	6 × 16	32 × 32	64 × 64
	Preconditioner: BD		
Solver			
MINRES	8	8	8
GMRES	9	9	9
	Preconditioner: BTR		
Solver			
GMRES	5	5	5
GCR	5	5	5

[a] For a viscosity jump of 10^6. We compare the block diagonal preconditioner used by *Burstedde et al.* [2008] with a block triangular preconditioner, our preferred approach. We observe that the number of iterations doubles with BD when compared to BTR.

vergence twice as fast compared to block diagonal preconditioners [*Elman and Silvester*, 1996], 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.

[7] In this paper, we first give a description of the model problem used for our experiments; next, we give a brief overview of common solver strategies that can handle unstructured grids, followed by a detailed description of our solver implementation (BTA-GCR), in section 3.3. Finally, we present scaling results, illustrating linear scaling characteristics of the BTA-GCR scheme with the number of degrees of freedom, as expected from a (algebraic) multigrid method, as well as optimal scaling with the number of processing cores, and show its robustness for sharp and localized viscosity contrasts of up to seven orders of magnitude in two dimensions.

2. Description of the Solution Method

2.1. Mathematical Formulation of the Problem

[8] For the analysis of the scaling relations of BTA-GCR, we consider thermal convection in a 2-D Cartesian box. 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 nondimensional Stokes equation and incorporate thermal effects only through a contribution to the right-hand side (rhs), with a given temperature field,

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

and 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

[9] 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 quadratic isoparametric finite elements.

[10] We use quadratic third-order accurate Taylor-Hood tetrahedral elements (P2-P1) [*Segal*, 2005]. 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,

$$\mathbf{A}\begin{bmatrix}\mathbf{u}\\ \mathbf{p}\end{bmatrix} = \begin{bmatrix}\mathbf{Q} & \mathbf{G}^{\mathbf{T}}\\ \mathbf{G} & 0\end{bmatrix}\begin{bmatrix}\mathbf{u}\\ \mathbf{p}\end{bmatrix} = \mathbf{b} \tag{3}$$

Table 2. Symbols

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	gas constant
u	velocity
$\mathbf{A}$	constrained Stokes system
$\mathbf{Q}$	velocity subsystem
$\mathbf{G}$	gradient operator (within a FE context)
$\mathbf{G}^{\mathbf{T}}$	divergence operator (within a 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 3. Abbreviations

Abbreviation	Meaning
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

in which $\mathbf{Q}$, the velocity stiffness matrix, is symmetric positive definite. $\mathbf{G}$ and $\mathbf{G}^T$ are associated with minus the divergence of the velocity and gradient of the pressure, respectively [*Cuvelier et al.*, 1986].

3. Solution Methods for the Stokes Equation

[11] In order to put the BTA-GCR solution scheme in perspective with solution techniques that have been successfully applied to the Stokes equation with up to 10^6 degrees of freedom on parallel computers, we review several popular solution methods and explain why they become less favorable for practical use when the number of degrees of freedom increases.

[12] We will consider three classes of solvers for the saddle point problem (3): (1) integrated, or holistic methods either using an incomplete lower and upper triangular decomposition (ILU)-type preconditioned Krylov methods or a penalty approach with a direct solver, (2) multigrid-type methods and (3) block triangular preconditioners, using AMG preconditioned Krylov methods during the preconditioner construction phase, our preferred solution method.

3.1. Direct Solvers, Krylov Methods, and Penalty Approach

[13] For problems with $O(10^5)$ degrees of freedom, both a parallel direct solver such as MUMPS [*Amestoy et al.*, 2000], and ILU preconditioned Krylov-type methods [*Saad*, 2003] are efficient solvers. In combination with direct solvers, the zero block resulting from the incompressibility constraint, equation (3), can be effectively eliminated with the penalty method [*Cuvelier et al.*, 1986], while at the same time reducing the problem size. However the increase in the condition number of the matrix, resulting from the penalty method, renders iterative solution strategies impossible, restricting this approach to direct solution methods. A straightforward and effective method to eliminate the effect of the zero block of equation (3) for iterative methods is to renumber the degrees of freedom in such a way that pressure unknowns are put at the end. This gives the ILU preconditioner the opportunity to fill the diagonal entries of the zero block during the factorization phase [*Dahl and Wille*, 1992; *ur Rehman et al.*, 2008].

[14] For classical block Jacobi domain decomposition [*Saad*, 2003], ILU-type preconditioners deteriorate significantly with increasing number of subdomains. To overcome this, Additive Schwartz-type domain decomposition [*Saad*, 2003] can be used. This class of preconditioners implements a predetermined amount of overlap between subdomains to limit the deterioration of the preconditioner when a significant amount of interconnectivity between subdomains is present.

[15] We found this strategy to be effective in preserving the quality of the ILU decomposition with increasing number of subdomains [*Geenen and van den Berg*, 2006]. The quality of the preconditioner is determined, in general, by the ratio of iterations of the iterative solver for the preconditioned and unpreconditioned system.

[16] However, another property of ILU-type preconditioners is the deterioration of the quality with increasing number of degrees of freedom. The degree of deterioration is determined by the capability of the ILU preconditioner to reduce the error associated with the longest wavelengths in the solution. How well this can be achieved is highly problem dependent [*Saad*, 2003]. We found that, for our system of equations using quadratic finite elements in two dimensions, ILU preconditioned methods break down for over 5×10^5 degrees of freedom. With large viscosity contrasts, the amount of ILU fill-in needed to guarantee reasonable convergence of the Krylov solver increases, which reduces the efficiency of the method considerably.

[17] The limitation in the number of degrees of freedom, for parallel direct solvers, results from the

suboptimal scaling with increasing number of subdomains, and the substantial amounts of memory needed to do a complete LU decomposition. For problems with a large number of degrees of freedom, order 10^6 and larger, Multigrid-type methods become favorable.

3.2. Multigrid

[18] Multigrid-type methods ideally scale like $O(n)$ with the number of degrees of freedom, n [*Wesseling*, 1992], making them the preferred solver method for many problems with over roughly 1 million degrees of freedom. Multigrid methods are efficient solution schemes because they exploit the fact that (cheap) iterative solvers can reduce the error in the short wavelength part of the solution very efficiently. By mapping the error to subsequent coarser meshes, the long wavelength part of the fine-grid error can be represented as short wavelengths on subsequent coarser grids. This guarantees fast convergence of the iterative solver on all resolution levels and therefore for all wavelengths in the solution.

[19] Multigrid solvers come in different types which can be divided into two main groups, geometric multigrid(GMG) and algebraic multigrid (AMG). Geometric and algebraic methods differ in the way they construct coarse grids and interpolation operators between grids. Geometric methods use geometric information to construct interpolation operators and coarse grids, while AMG methods only use information contained in the fine grid operator, **A**, without explicitly using any geometric information. For isoviscous models, GMG is arguably the fastest solution method for problems with over 10 million degrees of freedom [*MacLachlan et al.*, 2008]. However for problems with several orders of magnitude viscosity contrasts, classical GMG has the tendency to break down [*Kettler*, 1982]. The reasons why GMG might break down are related to the interpolation between coarse grids and the construction of these grids. The reason why interpolation between grids can break down for sharp coefficient variations is illustrated by *Falgout* [2006]. This issue was first addressed by *Alcouffe et al.* [1981].

[20] *Alcouffe et al.* [1981] proposed to use the continuity of $\mathbf{n} \cdot \eta\nabla\mathbf{u}$ rather than that of $\mathbf{n} \cdot \nabla\mathbf{u}$ for the interpolation between grids, which is similar to applying the methodology of AMG for the interpolation between grids and GMG for the coarse-grid construction [*Tackley*, 2008].

[21] This is an effective way of dealing with strong coefficient jumps in the model interior. However, this method can still break down when contrast in viscosity are captured at the finest resolution but not well represented at the coarser levels. Sharp viscosity contrasts are a marked example. In such cases, the construction of the coarse grids should also reflect the distribution of viscosity contrasts and hence should be constructed on the basis of the fine-grid operator.

[22] AMG constructs coarse grids on the basis of so-called strong connections [*Vaněk et al.*, 1996], where a connection is called strong if a_{ij} is significant relative to $\sqrt{a_{ii}a_{jj}}$. The coarse grid is formed by an aggregation strategy based on these strong connections [*Gee et al.*, 2006]. This ensures that parts of the solution domain that exhibit strong interdependence are represented on the coarse grids, providing a proper representation of the problem on subsequent coarse grids.

[23] The use of AMG provides us with a solution strategy that is robust for fine-scale large viscosity contrasts, associated with localization phenomena, when the use of parallel direct solvers become prohibitive.

[24] A further advantage of AMG-type methods is their ability to handle unstructured grids with arbitrary geometry, which is especially relevant for unstructured finite elements applications.

[25] A disadvantage of using AMG, thus refraining from using geometric information altogether, is the limitation to solve only problems with a single physical unknown (in our case, velocity or pressure). Therefore, any coupled system of equations such as the constrained Stokes equation must be decoupled and solved for the velocity and pressure unknowns separately. Recent attempts to use AMG methods for the integrated solution of the saddle point problem arising from the Stokes equation [*Janka*, 2008; *Webster*, 2007] are promising, but have so far only been applied successfully to well-conditioned isoviscous test problems.

3.3. BTA-GCR Method

[26] An efficient approach to decouple velocity and pressure unknowns is by applying a Krylov method combined with a so-called block preconditioner for the saddle point problem (3) [*Benzi et al.*, 2005], based on an incomplete block triangular factorization of the matrix **A** of the form,

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

With $\tilde{\mathbf{S}}$ an approximation to the Schur complement, $\mathbf{GQ}^{-1}\mathbf{G}^{\mathbf{T}}$.

[27] We solve the saddle point problem arising from the constrained Stokes equation (3) with a Krylov method, GCR [*van der Vorst and Vuik*, 1994], right preconditioned (postconditioned) with a block triangular preconditioner (BTR) [*Bramble and Pasciak*, 1988].

[28] The GCR algorithm (taken from *Vuik et al.* [2000]) contains the block triangular preconditioner $\mathbf{P}$ in the term $\mathbf{P}^{-1}\mathbf{r}^{\mathbf{k}}$:

$\mathbf{r}^{\mathbf{0}} = \mathbf{b} - \mathbf{A}\mathbf{x}^{\mathbf{0}}$
for $k = 0, 1, \ldots, ngcr$
  $\mathbf{s}^{\mathbf{k+1}} = \mathbf{P}^{-1}\mathbf{r}^{\mathbf{k}}$
  $\mathbf{v}^{\mathbf{k+1}} = \mathbf{A}\mathbf{s}^{\mathbf{k+1}}$
  for $i = 0, 1, \ldots, k$
    $\mathbf{v}^{\mathbf{k+1}} = \mathbf{v}^{\mathbf{k+1}} - (\mathbf{v}^{\mathbf{i}}, \mathbf{v}^{\mathbf{k+1}})\mathbf{v}^{\mathbf{i}}$
    $\mathbf{s}^{\mathbf{k+1}} = \mathbf{s}^{\mathbf{k+1}} - (\mathbf{v}^{\mathbf{k+1}}, \mathbf{v}^{\mathbf{i}})\mathbf{s}^{\mathbf{i}}$
  end for
  $\mathbf{v}^{\mathbf{k+1}} = \mathbf{v}^{\mathbf{k+1}}/\|\mathbf{v}^{\mathbf{k+1}}\|_2$
  $\mathbf{s}^{\mathbf{k+1}} = \mathbf{s}^{\mathbf{k+1}}\|\mathbf{v}^{\mathbf{k+1}}\|_2$
  $\mathbf{x}^{\mathbf{k+1}} = \mathbf{x}^{\mathbf{k}} + (\mathbf{v}^{\mathbf{k+1}}, \mathbf{r}^{\mathbf{k}})\mathbf{s}^{\mathbf{k+1}}$
  $\mathbf{r}^{\mathbf{k+1}} = \mathbf{r}^{\mathbf{k}} - (\mathbf{v}^{\mathbf{k+1}}, \mathbf{r}^{\mathbf{k}})\mathbf{v}^{\mathbf{k+1}}$
end for

[29] Instead of constructing $\mathbf{P}^{-1}$ explicitly, and applying it to the residual $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$, we solve

$$\begin{aligned} \mathbf{P}\mathbf{s} &= \mathbf{r} \\ \mathbf{s} &= [\mathbf{s}_1; \mathbf{s}_2] \\ \mathbf{r} &= [\mathbf{r}_1; \mathbf{r}_2] \end{aligned} \tag{5}$$

resulting in the distributed solution scheme,

$$\tilde{\mathbf{S}}\mathbf{s}_2^{\mathbf{k+1}} = \mathbf{r}_2^{\mathbf{k}} \tag{6}$$

$$\mathbf{Q}\mathbf{s}_1^{\mathbf{k+1}} = \mathbf{r}_1^{\mathbf{k}} - \mathbf{G}^{\mathbf{T}}\mathbf{s}_2^{\mathbf{k+1}} \tag{7}$$

We take $\mathbf{M_p}$, the pressure mass matrix [*Cuvelier et al.*, 1986], scaled with the inverse of viscosity as an approximation to the Schur complement $\tilde{\mathbf{S}}$, which is spectrally equivalent. *Verfurth* [1984] proved this for the isoviscous case. The proof for variable viscosity is provided by M. Olshanskii and P. Grinevich (An iterative method for the Stokes type problem with variable viscosity, submitted to *SIAM Journal on Scientific Computing*, 2008). This proof only holds under certain smoothness conditions of the viscosity variations. We present numerical support that for the viscosity contrasts we consider in our experiments this relation holds. The pressure mass matrix 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) [*Janka*, 2008]. 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 [*Janka*, 2008]. However, for typical domains and element shapes used in our experiments, this effect is not observed. Experiments by *Burstedde et al.* [2008] show that this effect is also not observed in 3-D spherical domains.

Table 4. Settings Used for the AMG V Cycle Construction and Application Phase[a]

Setting	Value/Description
Coarsening	uncoupled
Strong threshold	0.25
P damping factor	1.33
Smoother drop tolerance	0.25
Number of PDEs	2
Smoother	SOR
Smoother relaxation	1.01
Coarse grid solver	redundant
Redundant subsolver	MUMPS
Size of redundant blocks	16

[a] We use the ML library through the PETSc interface (S. Balay et al., http://www.mcs.anl.gov/petsc).

3.3.1. Solving the Subsystems for the Preconditioner

[30] We employ AMG from the ML library [*Gee et al.*, 2006] as a preconditioner to CG for the approximate solution of the subsystems during the preconditioner construction phase, equations (6) and (7); refer to Table 4 for ML settings. We use AMG as a preconditioner, rather than as a solver, based on the robustness of this approach, resulting in faster convergence [*Oosterlee and Washio*, 1998]. This is especially relevant for localized viscosity anomalies [*MacLachlan et al.*, 2008]. 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.

[31] 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 [*Geenen et al.*, 2007, 2008].

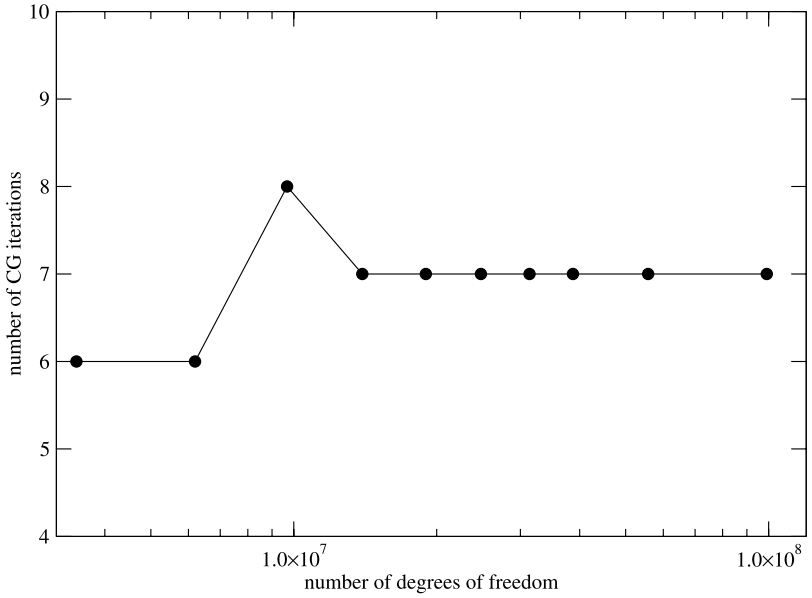

Figure 1. The solution of equation (7) with CG preconditioned with one AMG V cycle scales optimally with the number of degrees of freedom. A difference of one iteration between experiments with different number of degrees of freedom can be discarded.

3.3.2. Solving the BTR Preconditioned System

[32] 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 *May and Moresi* [2008, section 1.1].

[33] MINRES, the Krylov solver used by *Burstedde et al.* [2008] and *Schmid et al.* [2008], 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 [*Elman and Silvester*, 1996]. Our experiment comparing BD with MINRES and BTR with GCR (Table 1) shows that this is a robust feature of the preconditioner.

[34] We use GCR, a Krylov method similar to FGMRES [*Saad*, 2003; *Vuik*, 1995], 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 Tables 5 and 6. 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. In our example, we can perform on average 10 CG iterations within the same amount of wall time as

Table 5. Scaling of Outer, GCR, and Inner, CG, Iterations for Increasing Number of Degrees of Freedom and Viscosity Contrast[a]

Number of Elements	Iterations GCR/CG(Wall Time in Seconds)				
	$\Delta\eta$ =10	$\Delta\eta$ =10^3	$\Delta\eta$ =10^6	$\Delta\eta$ =10^7	$\Delta\eta$ = 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		

[a] We reproduced the SOLCX experiment from *May and Moresi* [2008]. 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 parentheses we present the total wall time 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 wall time.

Table 6. Effect of Increased Accuracy for the Inner Solver With Increasing Viscosity Contrast on the Wall Time to Solve the Velocity Subsystem, Equation (7)[a]

	Iterations (Inner/Outer) (Wall Time in Seconds on 1/100 Processing Cores)			
Accuracy Outer/Inner	$b = 0$	$b = 0.2$	$b = 0.3$	$b = 0.4$
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)

[a] We analyze the robustness of BTA-GCR for sharp, large-amplitude viscosity jumps resulting from the relation $T(x, y) = T(y)[(1 - \frac{b}{2}) + br]$, with r a random distribution between 0 and 1. The wall time 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 wall time). For larger values of b, it pays off to solve the inner problems with an accuracy of 10^{-3}–10^{-4}. In parentheses, we present the total wall time 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 off to reduce the number of outer iterations when increasing the number of processing cores due to the (initial) nonlinear scaling of the preconditioner application phase. This can be achieved by solving the velocity subsystem to a higher accuracy.

a single AMG V cycle. This means that if we can reduce the number of GCR iterations by one by doing less than 10 CG iterations for the velocity subsystem, we will gain in terms off wall time. For a comprehensive analysis of the characteristics of BTR as a preconditioner to GCR for the Stokes equation, we refer to M. ur Rehman et al. (On iterative methods for the incompressible Stokes problem, submitted to *International Journal for Numerical Methods in Fluids*, 2009). They show that BTR is a robust preconditioner for the SINKER model of *May and Moresi* [2008] and an example from an aluminium extrusion model. They also explain in more detail the way the pressure mass matrix is scaled with viscosity.

4. Numerical Experiments and Performance Tests

[35] We performed experiments to analyze the scaling of BTA-GCR with (1) number of degrees of freedom, (2) number of parallel processing cores, and (3) order of magnitude viscosity contrasts.

4.1. Scaling With Respect to the Number of Degrees of Freedom

[36] For these scaling experiments, we choose a 2-D isoviscous Cartesian setup based on the model by *Yang and Baumgardner* [2000], in which a smooth initial temperature depth profile is perturbed by a small-amplitude periodic temperature field, driving the thermal convection.

[37] 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 5, the number of iterations remain constant for increasing problem size.

4.2. Scaling With Respect to the Number of Processing Cores

[38] For these scaling experiments, we choose the same problem setup as for the previous experiment.

[39] We found that the parallel efficiency $E_p^n = \frac{T_s}{T_p^n}$ (with T_s the wall time for the sequential run and T_p^n wall time 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 3. The setup of the AMG V cycle does not scale linearly partly because of the sequential nature of the analysis part of the parallel direct solver we use on the coarsest grid [*Amestoy et al.*, 2006] as well as the nonlinear scaling of the factorization for the parallel direct solver (Figure 4). With the next release of MUMPS, the analysis phase will be made parallel, possibly making this part of the computation more efficient. Fortunately, the AMG setup phase has to be performed only once per time step where the AMG application and CG solve have to be performed for each outer GCR iteration.

4.3. Robustness

[40] To test the robustness of BTA-GCR, we performed a number of experiments with different viscosity contrasts, similar to *Yang and Baumgardner* [2000] and *May and Moresi*

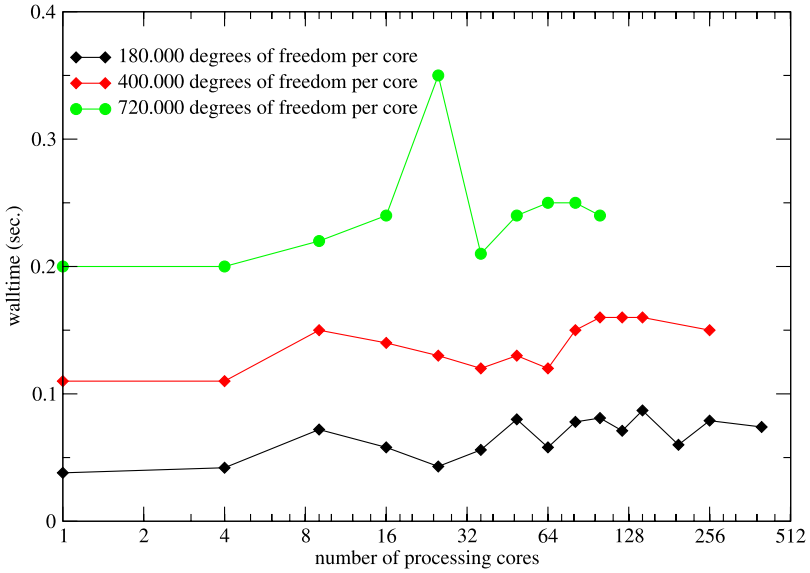

Figure 2. The scaling of CG with increasing number of processing cores. The number of degrees of freedom is kept fixed per processing core. The timing results are for a single CG iteration step. Although we observe a sometimes erratic wall time behavior with respect to the number of processing cores, 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.

[2008]. These include the following configurations: (1) a step function for the viscosity in the x direction across element boundaries and (2) a random viscosity perturbation across element boundaries [*Yang and Baumgardner*, 2000]. We kept the viscosity constant 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 [*May and Moresi*, 2008]. 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

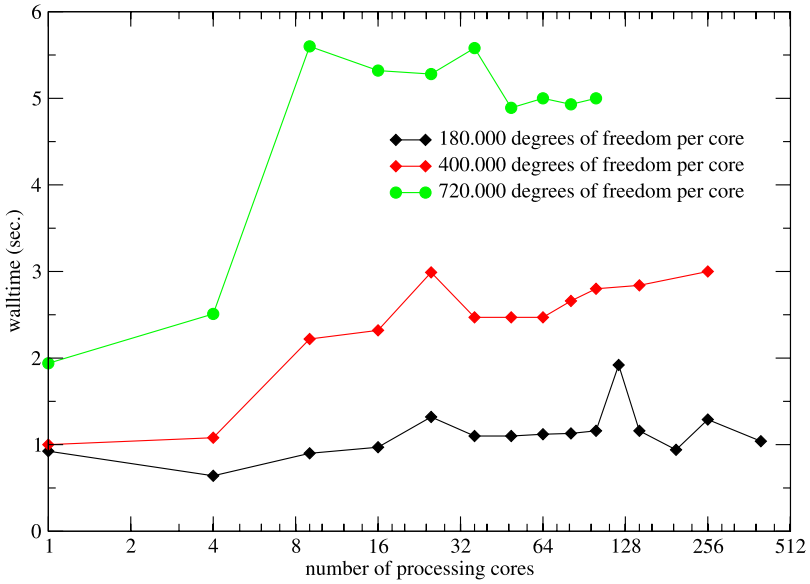

Figure 3. The amount of wall time for one AMG V cycle. We observe that after an initial increase in wall time with increasing number of processing cores, this relation flattens out to become roughly linear.

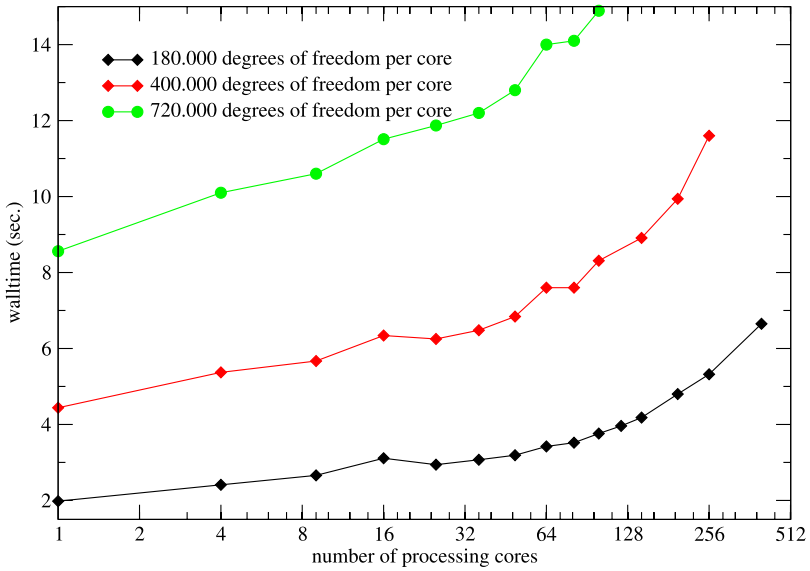

Figure 4. The amount of wall time for the AMG construction phase. With increasing number of processing cores the wall time increases nonlinearly. We can attribute this effect to the analysis phase of the direct solver, used to solve the coarse system, and to a lesser extent to its factorization phase.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

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 5. However, models with extremely high viscosity contrast across a single element will result in inaccurate results for the FE method [*Moresi et al.*, 1996] 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 5 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 *Burstedde et al.* [2008] and ur Rehman et al. (submitted manuscript, 2009).

[41] The second experiment has the same setup as the scaling experiments in sections 4.1 and 4.2, 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 $T(x, y) = T(y)[(1 - \frac{b}{2}) + br]$ 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. Table 6 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 wall time, occur with an inner accuracy between 10^{-3}–10^{-4}. This results from the higher computational costs of doing an outer GCR iterations (which involves applying an AMG V cycle) compared to a CG iterations which is relatively computationally inexpensive.

4.4. Domain Decomposition

[42] For the aggregation phase of the AMG preconditioner, we apply an uncoupled strategy. This reduces the parallel overhead of the AMG setup phase but could lead to so called bad aggregates. We performed a number of tests to assess the effect of the domain decomposition with respect to the viscosity configuration on the quality of the AMG preconditioner.

[43] To this end, we put a narrow high-viscosity band (six orders of magnitude viscosity contrast) in

a unit square domain, that is subdivided in four equal area subdomains. We investigated the effect of the position of the high-viscosity band, relative to the subdomain boundary, on the convergence of the solver, compared to the single domain, sequential, solution. We found that the number of iterations is constant for all configurations, illustrating the robustness of our implementation with respect to domain decomposition strategies.

5. Discussion and Concluding Remarks

[44] 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.

[45] 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.

[46] Recently authors have reported results with block preconditioners for saddle point problems in geodynamical applications [*May and Moresi*, 2008; *Burstedde et al.*, 2008; *Schmid et al.*, 2008]. The method employed by *May and Moresi* [2008] 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. *Burstedde et al.* [2008] present a method that is the most closely related to our approach; however, their 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 [*Elman and Silvester*, 1996], 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. They apply AMG directly to the velocity subsystem rather than using it as preconditioner to CG, which clearly scales suboptimally for increasing number of processing cores [*Burstedde et al.*, 2008, Figure 8] (Tables 5 and 6). 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 [*Burstedde et al.*, 2009, Table 3].

[47] *Schmid et al.* [2008] present a diagonal preconditioner where the subblock 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).

[48] 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.

[49] Modeling geodynamic processes that incorporate localization phenomena requires a high-resolution model with robust scalable solvers. By employing AMG-type methods to construct a preconditioner to solve the constrained Stokes equation, we are able to solve this kind of problems within a reasonable wall time.

Acknowledgments

[50] 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

Alcouffe, R., A. Brandt, J. E. Dendy, and J. Painter (1981), The multigrid method for the diffusion equation with strongly discontinuous coefficients, *SIAM J. Sci. Stat. Comput.*, *2*, 430–454.

Amestoy, P., I. Duff, and J. L'Excellent (2000), Multifrontal parallel distributed symmetric and unsymmetric solvers, *Comput. Meth. Appl. Mech. Eng.*, *184*, 501–520.

Amestoy, P. R., J. Koster, I. S. Duff, and J. L'Excellent (2006), *MUltifrontal Massively Parallel Solver (MUMPS Version 4.6.3) Users' Guide*, CERFACS, Toulouse, France. (Available at http://graal.ens-lyon.fr/MUMPS/index.html)

Baumgardner, J. (1985), Three-dimensional treatment of convective flow in the Earth's mantle, *J. Stat. Phys.*, *39*, 501–511.

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

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

Braun, J., C. Thieulot, P. Fullsack, M. DeKool, C. Beaumont, and R. Huismans (2008), DOUAR: A new three-dimensional creeping flow numerical model for the solution of geological problems, *Phys. Earth Planet. Inter.*, *171*(1–4), 76–91, doi:10.1016/j.pepi.2008.05.003.

Burstedde, C., O. Ghattas, M. Gurnis, G. Stadler, E. Tan, T. Tu, L. C. Wilcox, and S. Zhong (2008), Scalable adaptive mantle convection simulation on petascale supercomputers, in *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*, pp. 1–15, IEEE Press, Austin, Tex., doi:10.1145/1413370.1413434.

Burstedde, C., O. Ghattas, G. Stadler, T. Tu, and L. Wilcox (2009), Parallel scalable adjoint-based adaptive solution for variable-viscosity Stokes flows, *Comput. Meth. Appl. Mech. Eng.*, *198*, 1691–1700.

Choblet, G., O. Čadek, F. Couturier, and C. Dumoulin (2007), dipus: A new tool to study the dynamics of planetary interiors, *Geophys. J. Int.*, *170*(1), 9–30.

Cuvelier, C., A. Segal, and A. van Steenhoven (1986), *Finite Element Methods and Navier-Stokes Equations*, D. Reidel, Dordrecht, Netherlands.

Dahl, O., and S. Ø. Wille (1992), An ILU preconditioner with coupled node fill-in for iterative solution of the mixed finite element formulation of the 2D and 3D Navier-Stokes equations, Int. J, *Numer. Meth. Fluids*, *15*(5), 525–544, doi:10.1002/fld.1650150503.

Elman, H., and D. Silvester (1996), Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations, *SIAM J. Sci. Comput.*, *17*, 33–46, doi:10.1.1.25.265.

Falgout, R. (2006), An introduction to algebraic multigrid computing, *Comput. Sci. Eng.*, *8*(6), 24–33.

Gee, M., C. Siefert, J. Hu, R. Tuminaro, and M. Sala (2006), Ml 5.0 smoothed aggregation user's guide, technical report, Sandia Natl. Lab., Albuquerque, N. M.

Geenen, T., and A. P. van den Berg (2006), 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, Univ. Utrecht, Utrecht, Netherlands.

Geenen, T., A. P. van den Berg, and W. Spakman (2007), Large scale, high resolution, mantle dynamics modeling, *Eos Trans. AGU*, *88*(52), Fall Meet. Suppl., Abstract T51B-0546.

Geenen, T., A. P. van den Berg, and W. Spakman (2008), Scalable robust solvers for 3D unstructured modeling applications, solving the Stokes equation with large, localized, viscosity contrasts, abstract EGU2008-A-09460 presented at European Geosciences Union General Assembly 2008, Vienna.

Janka, A. (2008), Smoothed aggregation multigrid for a Stokes problem, *Comput. Visual. Sci.*, *11*(3), 169–180.

Kameyama, M., A. Kageyama, and T. Sato (2008), Multigrid-based simulation code for mantle convection in spherical shell using Yin-Yang grid, *Phys. Earth Planet. Inter.*, *171*(1–4), 19–32.

Kettler, R. (1982), *Analysis and Comparison of Relaxation Schemes in Robust Multigrid and Preconditioned Conjugate Gradient Methods*, pp. 502–534, Springer, Heidelberg, Germany.

MacLachlan, S., J. Tang, and C. Vuik (2008), Fast and robust solvers for pressure correction in bubbly flow problems, *J. Comput. Phys.*, *227*, 9742–9761.

May, D., and L. Moresi (2008), Preconditioned iterative methods for Stokes flow problems arising in computational geodynamics, *Phys. Earth Planet. Inter.*, *171*(1–4), 33–47.

Moresi, L., S. Zhong, and M. Gurnis (1996), The accuracy of finite element solutions of Stokes' flow with strongly varying viscosity, *Phys. Earth Planet. Inter.*, *97*, 83–94.

Oosterlee, C., and T. Washio (1998), An evaluation of parallel multigrid as a solver and a preconditioner for singularly perturbed problems, *SIAM J. Sci. Comput.*, *19*(1), 87–110.

Saad, Y. (2003), *Iterative Methods for Sparse Linear Systems*, Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Schmid, D., M. Dabrowski, and M. Krotkiewski (2008), Evolution of large amplitude 3D fold patterns: A FEM study, *Phys. Earth Planet. Inter.*, *171*(1–4), 400–408, doi:10.1016/j.pepi.2008.08.007.

Segal, A. (2005), *SEPRAN Manual: User Manual*, *SEPRA Analysis*, Ingenieursbur. Sepra, Leidschendam, Netherlands.

Tackley, P. (1993), Effects of strongly temperature-dependent viscosity on time-dependent, three-dimensional models of mantle convection, *Geophys. Res. Lett.*, *20*(20), 2187–2190.

Tackley, P. (2008), Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid, *Phys. Earth Planet. Inter.*, *171*(1–4), 7–18.

ur Rehman, M., C. Vuik, and G. Segal (2008), A comparison of preconditioners for incompressible Navier-Stokes solvers, *Int. J. Numer. Meth. Fluids*, *57*(12), 1731–1751.

van der Vorst, H., and C. Vuik (1994), GMRESR: A family of nested GMRES methods, *Numer. Linear Algebra Appl.*, *1*, 369–386.

Vaněk, P., J. Mandel, and M. Brezina (1996), Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems, *Computing*, *56*(3), 179–196, doi:10.1007/BF02238511.

Verfurth, R. (1984), A combined conjugate gradient-multigrid algorithm for the numerical solution of the Stokes problems, *IMA J. Numer. Anal.*, *4*(4), 441–455.

Vuik, C. (1995), New insights in GMRES-like methods with variable preconditioners, *J. Comput. Appl. Math.*, *61*, 189–204.

Vuik, C., A. Saghir, and G. Boerstoel (2000), The Krylov accelerated SIMPLE(R) method for flow problems in industrial furnaces, *Int. J. Numer. Meth. Fluids*, *33*, 1027–1040.

Webster, R. (2007), Algebraic multigrid and incompressible fluid flow, *Int. J. Numer. Meth. Fluids*, *53*(4), 669–690.

Wesseling, P. (1992), *An Introduction to Multigrid Methods*, John Wiley, Chichester, N. Y.

Yang, W., and J. Baumgardner (2000), A matrix-dependent transfer multigrid method for strongly variable viscosity infinite Prandtl number thermal convection, *Geophys. Astrophys. Fluid Dyn.*, *92*(3–4), 151–195.

Zhong, S., D. Yuen, L. Moresi, and G. Schubert (2007), *Numerical Methods for Mantle Convection*, pp. 227–252, Elsevier, Amsterdam.