μ-BFBT PRECONDITIONER FOR STOKES FLOW PROBLEMS WITH HIGHLY HETEROGENEOUS VISCOSITY*

JOHANN RUDI†, GEORG STADLER‡, AND OMAR GHATTAS†§

Abstract. We present μ-BFBT, an approximation for the inverse Schur complement of a Stokes system with highly heterogeneous viscosity. When used as part of a Schur complement-based Stokes system preconditioner, we observe robust convergence rates for Stokes problems with smooth but highly varying (up to 10 orders of magnitude) viscosities, optimal algorithmic scalability with respect to mesh refinement, and merely a mild dependence on the polynomial order of high-order finite element discretizations ($\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\text{disc}}$, order $k \geq 2$). For certain problems, μ-BFBT significantly improves Stokes solver convergence over the widely used inverse viscosity weighted pressure mass matrix approximation of the Schur complement. Using detailed numerical experiments, we discuss modifications to μ-BFBT at Dirichlet boundaries, which decrease the number of iterations. The overall algorithmic performance of the Stokes solver is governed by the efficacy of μ-BFBT as a Schur complement approximation and, in addition, by our parallel hybrid spectral-geometric-algebraic multigrid (HMG) method, used for approximating the inverses of the viscous block and variable-coefficient pressure Poisson operators within μ-BFBT. Building on the scalability of HMG, our Stokes solver achieves parallel weak scalability of 90 % for a more than 600-fold increase from 48 to all 30,000 cores of TACC's Lonestar 5 supercomputer.

1. Introduction and key ideas. Many problems in science and engineering involve creeping non-Newtonian fluids. Important examples can be found in geophysical fluid flows, where the incompressible Stokes equation with power-law rheology has become a prototypical continuum mechanical description for creeping flows occurring in mantle convection, magma dynamics, and ice flow. The linearization of the nonlinear momentum equations, for instance within a Newton method, leads to flow equations (here, incompressible Stokes equations) with spatially-varying and highly heterogeneous viscosity fields.

In particular, simulations of earth's mantle convection at global scale [17] exhibit extreme computational challenges due to a highly heterogeneous and anisotropic viscosity stemming from temperature and strain-rate dependence and sharp gradients in narrow regions modeling tectonic plate boundaries (six orders of magnitude drop in $\sim$5 km) [3, 15]. This leads to a wide range of spatial scales since small localized features at plate boundaries of size $\mathcal{O}(1\,\text{km})$ influence plate motion at continental scales of $\mathcal{O}(1000\,\text{km})$. The complex character of the flow presents severe computational challenges for iterative solvers due to poor conditioning of linear systems that arise.

Since this paper focuses on preconditioning linearized Stokes problems, we can simplify our problem setup by taking the viscosity as independent of the strain rate, but otherwise exhibiting severe spatial heterogeneity. Given a bounded domain $\Omega \subset \mathbb{R}^3$, right-hand side forcing $\boldsymbol{f}$, and viscosity $\mu(\boldsymbol{x}) \geq \mu_{\min} > 0$ for all $\boldsymbol{x} \in \Omega$, consider the incompressible Stokes equations

$$-\nabla \cdot \left[\mu(\boldsymbol{x})\,(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^\top)\right] + \nabla p = \boldsymbol{f} \quad \text{in } \Omega, \tag{1.1a}$$

$$\nabla \cdot \boldsymbol{u} = 0 \quad \text{in } \Omega, \tag{1.1b}$$

$$\boldsymbol{u} = 0 \quad \text{on } \partial\Omega, \tag{1.1c}$$

where $\boldsymbol{u}$ and p are the unknown velocity and pressure fields, respectively. Discretizing (1.1) leads to a sequence of linear systems of equations of the form

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}, \tag{1.2}$$

*Submitted to Student Paper Competition of 14-th Copper Mountain Conference on Iterative Methods.

†Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, USA (johann@ices.utexas.edu, omar@ices.utexas.edu).

‡Courant Institute of Mathematical Sciences, New York University, USA (stadler@cims.nyu.edu).

§Jackson School of Geosciences and Dept. of Mechanical Engineering, The University of Texas at Austin, USA.

where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{B}^\top$ are matrices corresponding to viscous stress, discrete divergence, and gradient operators, respectively. The discretization is carried out by high-order finite elements on (possibly aggressively adaptively refined) hexahedral meshes with velocity-pressure pairings $\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\text{disc}}$ of polynomial order $k \geq 2$ with a continuous, nodal velocity space $\mathbb{Q}_k$ and a discontinuous, modal pressure space $\mathbb{P}_{k-1}^{\text{disc}}$. These pairings yield optimal asymptotic convergence of the finite element approximation to the infinite dimensional solution with decreasing mesh element size, are inf-sup stable, and have the advantage of preserving mass locally at the element level due to the discontinuous pressure space [8]. While these properties are important for geophysics applications, the high-order discretization, adaptivity, and the discontinuous pressure space add to the computational challenges for iterative solvers. The level of difficulty is further raised by demands for large-scale parallelism, i.e., efficient execution of the solvers on millions of threads with billions of degrees of freedom (DOF).

While exhibiting the previously mentioned challenges, the applications we are targeting demand robust linear Stokes preconditioners for (1.2) with nearly optimal algorithmic and parallel scalability. This paper describes the design of such preconditioners that are suitable for highly heterogeneous viscosities μ as occurring, e.g., in mantle convection simulations. The proposed Schur complement approximation, called μ-BFBT, significantly outperforms state of the art methods when both robustness and algorithmic scalability are crucial. Further, the efficacy of the preconditioner is by design independent of the number of parallel threads and parallel efficiency is largely unaffected when scaling out to millions of threads.

It is widely accepted that an effective approximation of the Schur complement $\mathbf{S} \coloneqq \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\top$ is an essential ingredient for attaining fast convergence. More precisely, a sufficiently good approximation of the inverse Schur complement $\tilde{\mathbf{S}}^{-1} \approx \mathbf{S}$ is sought, which, together with an approximation of the inverse viscous block, $\tilde{\mathbf{A}}^{-1} \approx \mathbf{A}$, is used in an iterative scheme with right preconditioning based on an upper triangular block matrix:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{B}^\top \\ \mathbf{0} & \tilde{\mathbf{S}} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}. \tag{1.3}$$

Note that the original solution to (1.2) is recovered by applying the preconditioner once to the solution of (1.3). For the preconditioned Stokes system (1.3), we use GMRES as the Krylov subspace solver. This particular combination of Krylov method and preconditioner is known to converge in just two iterations for optimal choices of $\tilde{\mathbf{A}}^{-1}$ and $\tilde{\mathbf{S}}^{-1}$ [2].

The established state of the art approximation of the Schur complement is the inverse viscosity-weighted pressure mass matrix, $\mathbf{M}_p(1/\mu)$, with entries $[\mathbf{M}_p(1/\mu)]_{i,j} = \int_\Omega q_i(\boldsymbol{x})\, q_j(\boldsymbol{x})/\mu(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$, where $q_i, q_j \in \mathbb{P}_{k-1}^{\text{disc}}$ are basis functions of the finite dimensional space $\mathbb{P}_{k-1}^{\text{disc}}$. In practice $\mathbf{M}_p(1/\mu)$ is diagonalized via lumping to simplify its inversion. Provided that μ is sufficiently smooth, $\mathbf{M}_p(1/\mu)$ can be an effective approximation of $\mathbf{S}$ in numerical experiments [4, 11] and spectral equivalence can be shown [10]. However, it has been observed in applications with highly heterogeneous viscosities μ (e.g., mantle convection [14, 15]) that convergence slows down significantly due to a poor Schur complement approximation by $\mathbf{M}_p(1/\mu)$. Therefore, we propose a new approximation, called μ-BFBT, that remains robust when $\mathbf{M}_p(1/\mu)$ fails. The possibly dramatic improvement in convergence is demonstrated in Figure 1.1.

Preconditioners based on BFBT approximations for the Schur complement were initially proposed in [6] for the Navier-Stokes equations. Over the years, these ideas were refined and extended in [16, 12, 7, 9] to arrive at a class of closely related Schur complement approximations: pressure convection-diffusion (PCD), BFBT, and Least Squares Commutator (LSC). The underlying principle is that one seeks a commutator matrix $\mathbf{X}$ such that the commutator $\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top - \mathbf{B}^\top\mathbf{X} \approx \mathbf{0}$ nearly vanishes (given an appropriate diagonal scaling matrix $\mathbf{D}^{-1}$). This can be converted into the following

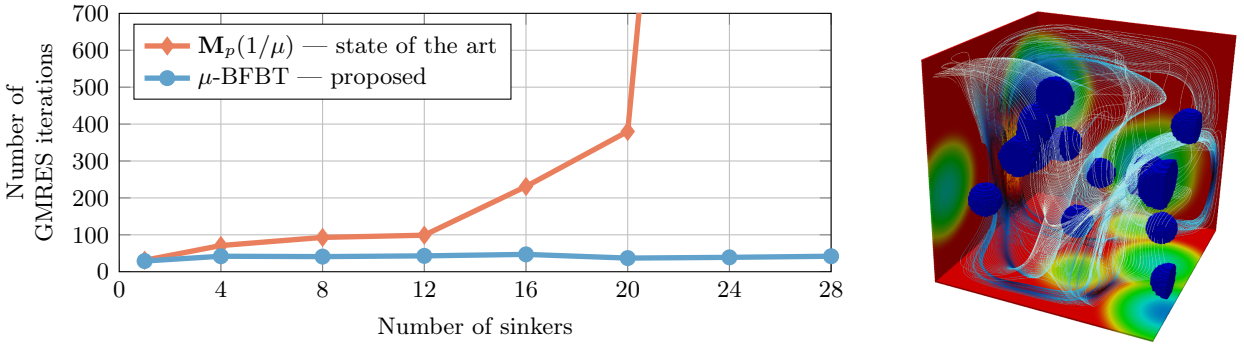


FIGURE 1.1. *Left image shows the improvement in convergence obtained with the proposed μ-BFBT preconditioner over a preconditioner using the inverse viscosity-weighted pressure mass matrix as Schur complement approximation. Number of randomly placed sinkers (i.e., inclusions with high viscosity in low viscosity medium) is increased along x-axis. The number of GMRES iterations required for a residual reduction by 10^6 is reported for the state of the art $\mathbf{M}_p(1/\mu)$ and the proposed μ-BFBT preconditioner. Fixed problem parameters are the dynamic ratio $\mathrm{DR}(\mu) = \max(\mu)/\min(\mu) = 10^8$, discretization order $k = 2$, and the mesh refinement level $\ell = 7$, resulting in 128^3 finite elements. Right image shows an example viscosity with 16 sinkers (blue spheres) and the stream lines of the computed velocity field.*

least-squares minimization problem: $\min_{\mathbf{X}} \left\|\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top\mathbf{e}_j - \mathbf{B}^\top\mathbf{X}\mathbf{e}_j\right\|^2_{\mathbf{C}^{-1}}$ for all j, where $\mathbf{e}_j$ is the j-th Cartesian unit vector and $\mathbf{C}$ is symmetric and positive definite. The solution is given by $\mathbf{X} = (\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top)^{-1}(\mathbf{B}\mathbf{C}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top)$. Then the BFBT approximation of the inverse Schur complement can be derived by algebraic rearrangement: $\tilde{\mathbf{S}}^{-1}_{\mathrm{BFBT}} \coloneqq \left(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top\right)^{-1}\left(\mathbf{B}\mathbf{C}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B}^\top\right)\left(\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^\top\right)^{-1}$. In the literature cited above (which targets preconditioning of the Navier-Stokes equations), the diagonal scaling matrices are chosen as $\mathbf{C} = \mathbf{D} = \tilde{\mathbf{M}}_{\boldsymbol{u}}$, i.e., a lumped version of the velocity mass matrix; hence we call this the $\mathbf{M}_{\boldsymbol{u}}$-BFBT approximation of the Schur complement. In [14] and later in [15], the authors recognize that BFBT can provide an effective Schur complement approximation for Stokes problems with highly varying viscosity. This was realized by different choices for the diagonal scaling matrices based on entries of $\mathbf{A}$, e.g., $\mathbf{C} = \mathbf{D} = \mathrm{diag}(\mathbf{A})$ in [15]; hence we refer to this approach as diag($\mathbf{A}$)-BFBT.

However, even diag($\mathbf{A}$)-BFBT can fail to achieve fast convergence for some problems, as shown below. Moreover, choosing the scaling matrices as diag($\mathbf{A}$) builds on heuristic algebraic arguments and is problematic for high-order discretizations, where diag($\mathbf{A}$) is a very poor approximation of $\mathbf{A}$. These drawbacks led us to develop a new BFBT-type approximation for the Schur complement,

$$\tilde{\mathbf{S}}^{-1}_{\mu-\mathrm{BFBT}} \coloneqq \left(\mathbf{B}\mathbf{C}_\mu^{-1}\mathbf{B}^\top\right)^{-1}\left(\mathbf{B}\mathbf{C}_\mu^{-1}\mathbf{A}\mathbf{D}_\mu^{-1}\mathbf{B}^\top\right)\left(\mathbf{B}\mathbf{D}_\mu^{-1}\mathbf{B}^\top\right)^{-1}, \tag{1.4}$$

where $\mathbf{C}_\mu = \mathbf{D}_\mu = \tilde{\mathbf{M}}_{\boldsymbol{u}}(\sqrt{\mu})$ are lumped velocity mass matrices that are weighted by the square root of the viscosity.

After defining a set of benchmark problems, we compare the convergence obtained with different Schur complement approximations. We study when μ-BFBT is advantageous to $\mathbf{M}_p(1/\mu)$, and discuss boundary modifications for μ-BFBT that accelerate the convergence. Finally, we summarize an algorithm for μ-BFBT-based Stokes preconditioning, which uses hybrid spectral-geometric-algebraic multigrid (HMG), and we show near-optimal algorithmic and parallel scalability.

2. Class of benchmark problems. The design of suitable benchmark problems is critical to conduct studies that can give useful convergence estimates for challenging applications. We seek complex geometrical structures in the viscosity that generate irregular, nonlocal, multi-scale flow fields. Additionally, the viscosity should exhibit sharp gradients and its dynamic ratio $\mathrm{DR}(\mu) \coloneqq \max(\mu)/\min(\mu)$ (also commonly referred to as viscosity contrast) can be six orders of magnitude

or higher in demanding applications and it is important that $\min(\mu) \ll 1$ and $\max(\mu) \gg 1$. It has been observed, by the authors and in [13], that the viscosity arising from a multi-sinker setup with randomly positioned inclusions (e.g., as in Figure 1.1, *right image*) is a possible candidate to model such challenging, highly heterogeneous coefficients.

We define the viscosity coefficient $\mu(\boldsymbol{x}) \in [\mu_{\min}, \mu_{\max}]$, $0 < \mu_{\min} < \mu_{\max} < \infty$, with dynamic ratio $\mathrm{DR}(\mu) = \mu_{\max}/\mu_{\min}$ by means of rescaling a C^∞ indicator function $\chi_n(\boldsymbol{x}) \in [0,1]$ that accumulates n sinkers via a product of modified Gaussian functions:

$$\mu(\boldsymbol{x}) \coloneqq (\mu_{\max} - \mu_{\min})(1 - \chi_n(\boldsymbol{x})) + \mu_{\min} \quad \text{with} \quad \chi_n(\boldsymbol{x}) \coloneqq \prod_{i=1}^{n} 1 - \exp\left(-d \max\left(0, |\boldsymbol{c}_i - \boldsymbol{x}| - \frac{w}{2}\right)^2\right)$$

where $\boldsymbol{c}_i \in \Omega$, $i = 1, \ldots, n$, are the center points of the sinkers, $d > 0$ is exponential decay, and $w \geq 0$ is the width of a sinker where $\mu_{\max}$ is attained. Throughout the paper, we fix $\Omega = [0,1]^3$, $d = 200$, $w = 0.1$, and draw from the same set of precomputed random points $\boldsymbol{c}_i$ in all numerical experiments. Two parameters are varied: (i) the number of sinkers n at random positions (e.g., the label S16-rand indicates $n = 16$) and (ii) the dynamic ratio $\mathrm{DR}(\mu)$ which in turn determines $\mu_{\min} \coloneqq \mathrm{DR}(\mu)^{-1/2}$ and $\mu_{\max} \coloneqq \mathrm{DR}(\mu)^{1/2}$.

3. Comparison of Schur complement approximations. We compare convergence of the Stokes solver using the Schur complement approximation $\mathbf{M}_p(1/\mu)$ with $\mathrm{diag}(\mathbf{A})$-BFBT and with the proposed μ-BFBT. The problem parameters are held fixed to S16-rand and $\mathrm{DR}(\mu) = 10^8$. The numerical experiments were carried out using different levels of mesh refinement $\ell = 5, \ldots, 7$ (fixed order $k = 2$) and different discretization orders $k = 2, \ldots, 5$ (fixed level $\ell = 5$).

The results are presented in Figure 3.1. In the *left* two plots, the poor Schur complement approximation by $\mathbf{M}_p(1/\mu)$ for this problem setup can be observed clearly. Convergence is stagnating severely after about 25 GMRES iterations (similar results are found in [13, 15]). Since refining the mesh or increasing the order does not influence convergence behavior, we assume the features of the viscosity to be sufficiently resolved by the discretization.

The case $\mathrm{diag}(\mathbf{A})$-BFBT (Figure 3.1, *middle*) is able to achieve fast convergence for discretization order $k = 2$. A limitation of $\mathrm{diag}(\mathbf{A})$-BFBT is a strong dependence on the order k. It can be explained as a consequence of the reduction of diagonal dominance in the viscous block $\mathbf{A}$ with k, i.e., for higher k the approximation of $\mathbf{A}$ by $\mathrm{diag}(\mathbf{A})$ deteriorates. Note that numerical experiments with $\mathbf{M}_{\boldsymbol{u}}$-BFBT are not presented, because it performs poorly in the presence of spatially varying viscosities. This leads to the conclusion that the choice of the scaling matrices $\mathbf{C}, \mathbf{D}$ in $\tilde{\mathbf{S}}^{-1}_{\mathrm{BFBT}}$ crucially affects the quality of the Schur complement approximation.

The μ-BFBT approximation delivers convergence that is nearly as fast as in the $\mathrm{diag}(\mathbf{A})$-BFBT, $k = 2$ case, however without deterioration when k is increased (see Figure 3.1, *right*). Thus μ-BFBT exhibits the robustness of $\mathrm{diag}(\mathbf{A})$-BFBT and additionally shows superior algorithmic scalability with respect to k. Having shown the efficacy of μ-BFBT under certain problem parameters, we continue by classifying the influence of crucial parameters on convergence.

4. Robustness classification of μ-BFBT over established state of the art. This section compares the widely used Schur complement approximation $\mathbf{M}_p(1/\mu)$ with μ-BFBT by means of varying the two problem parameters: (i) the number of randomly placed sinkers n and (ii) the dynamic ratio $\mathrm{DR}(\mu)$. The parameter n influences the geometric complexity of the viscosity μ while $\mathrm{DR}(\mu)$ controls the sharpness of gradients.

Tables 4.1a and 4.1b present the number of GMRES iterations for a 10^{-6} residual reduction. We observe that for the S1-rand problem the iteration count is essentially the same for both $\mathbf{M}_p(1/\mu)$ and μ-BFBT, and that it stays stable across all dynamic ratios $\mathrm{DR}(\mu) = 10^4, \ldots, 10^{10}$. Hence for this

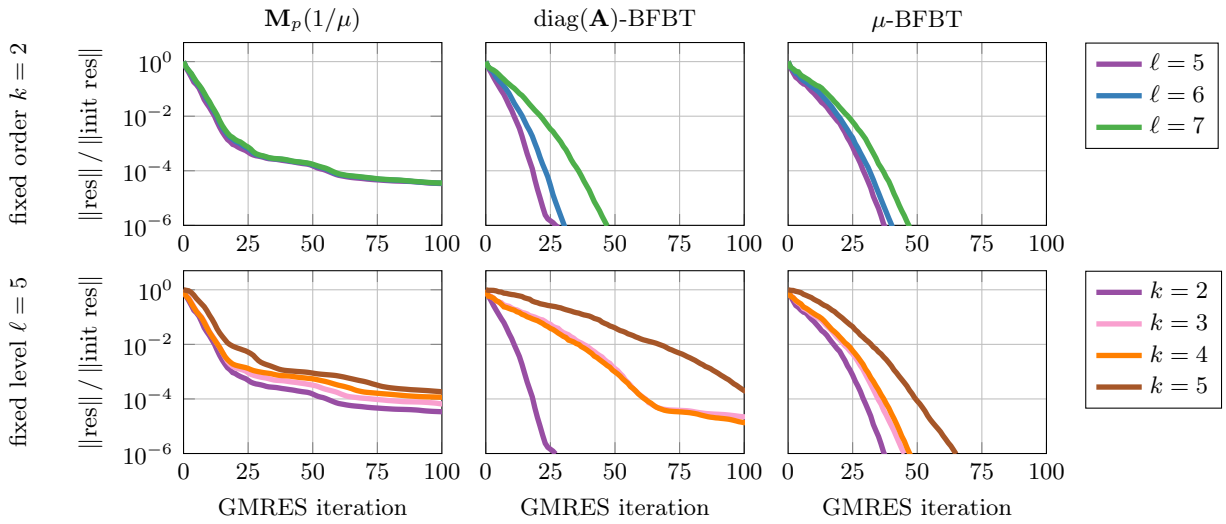


FIGURE 3.1. *Comparison of Stokes solver convergence with* $\mathbf{M}_p(1/\mu)$ *(left column),* diag($\mathbf{A}$)*-BFBT (middle column), and* μ*-BFBT (right column) preconditioning. We fix the problem S16-rand,* $\text{DR}(\mu) = 10^8$ *while varying mesh refinement level* ℓ *(top row) and discretization order* k *(bottom row). This comparison shows that* μ*-BFBT combines robust convergence of* diag($\mathbf{A}$)*-BFBT with improved algorithmic scalability when* k *is increased.*

relatively simple problem, μ-BFBT has no advantages and its additional computational costs make it less efficient than $\mathbf{M}_p(1/\mu)$. This simple setup is often considered in convergence experiments, but it does not reveal the limitations of the $\mathbf{M}_p(1/\mu)$ approach.

The limitations are revealed by increasing the number of randomly positioned sinkers. Two observations for $\mathbf{M}_p(1/\mu)$ can be made from Table 4.1a. First, the number of GMRES iterations rises dramatically with increasing number of sinkers (factor $\sim$80 increase for $n = 1, \ldots, 28$, $\text{DR}(\mu) = 10^8$). Second, in a multi-sinker setup the dependence on $\text{DR}(\mu)$ becomes more severe (factor $\sim$50 increase for $n = 24$, $\text{DR}(\mu) = 10^4, \ldots, 10^{10}$). Over the whole range of parameters n and $\text{DR}(\mu)$, the number of GMRES iterations increases by a factor >300. For the the most extreme setup, S28-rand, $\text{DR}(\mu) = 10^{10}$, the iteration count exceeds 10000. Such convergence behavior shows that $\mathbf{M}_p(1/\mu)$ exhibits poor approximation properties to the Schur complement for certain classes of problems with highly heterogeneous viscosities.

The advantages in robustness of the μ-BFBT preconditioner are demonstrated in Table 4.1b. Compared to $\mathbf{M}_p(1/\mu)$, the number of GMRES iterations is stable and the increase over the whole range of problem parameters is just about a factor of 2. Only 60 iterations were performed for the most extreme parameters, S28-rand, $\text{DR}(\mu) = 10^{10}$, which took >10000 iterations with $\mathbf{M}_p(1/\mu)$.

5. Modifications for Dirichlet boundary conditions. In this final section on Stokes solver convergence with μ-BFBT, we discuss the dependence on mesh refinement level ℓ and discretization order k. The numerical experiments for μ-BFBT in Figure 3.1, *right* did show slightly slower convergence rates when ℓ or k were increased. This behavior is investigated now with the goal to achieve perfect mesh independence and only a mild dependence on the order.

For unbounded domains $\Omega = \mathbb{R}^3$, the commutator relationship that leads to the BFBT formulation is exactly satisfied for the associated differential operators, thus the minimum of the infinite dimensional problem corresponding to the discrete least-squares minimization problem (see Section 1) is zero. In the presence of Dirichlet boundary conditions the commutator, in general, does not vanish at the boundary. Therefore a possible source for deteriorating Schur complement approximation properties of μ-BFBT is a commutator mismatch inside mesh elements that are touching the

TABLE 4.1
Robustness classification for Schur complement approximations (a) $\mathbf{M}_p(1/\mu)$ and (b) μ-BFBT in terms of number of GMRES iterations (10^{-6} residual reduction, GMRES restart at 100). Problem parameters n (number of randomly placed sinkers) and $\mathrm{DR}(\mu)$ (dynamic ratio) are varied. Discretization is fixed to $k = 2$, $\ell = 7$.

(a) $\mathbf{M}_p(1/\mu)$

#sinkers \ DR(μ)	10^4	10^6	10^8	10^{10}
S1-rand	29	31	31	29
S4-rand	53	63	71	80
S8-rand	64	79	93	165
S12-rand	70	86	99	180
S16-rand	85	167	231	891
S20-rand	84	167	380	724
S24-rand	117	286	3279	5983
S28-rand	108	499	2472	>10000

(b) μ-BFBT

#sinkers \ DR(μ)	10^4	10^6	10^8	10^{10}
S1-rand	29	29	29	30
S4-rand	39	41	42	44
S8-rand	38	40	41	44
S12-rand	38	40	43	45
S16-rand	40	45	47	48
S20-rand	34	36	37	38
S24-rand	31	32	39	55
S28-rand	29	31	42	60

boundary $\partial\Omega$. A similar argument is made in [9] where a damping to the scaling matrix $\mathbf{D}^{-1}$ in $\tilde{\mathbf{S}}^{-1}_{\mathrm{BFBT}}$ is introduced to achieve mesh independence ($\mathbf{C}^{-1}$ is not changed). There, damping affects the normal components of the velocity space inside mesh elements touching $\partial\Omega$ and simply a constant damping factor of 1/10 is set regardless of mesh refinement ℓ.

Now, we attempt to enhance our understanding how modifications at the Dirichlet boundary $\partial\Omega$ influence convergence and therefore the efficacy of μ-BFBT as a Schur complement approximation. Let $\Omega_D \coloneqq \bigcup_{e\in D} \Omega_e$, $D \coloneqq \{e \mid \overline{\Omega_e} \cap \partial\Omega \neq \emptyset\}$. Ω_D is the union of all mesh elements Ω_e touching the Dirichlet boundary. Given value $a \geq 1$, extend the definition of $\tilde{\mathbf{M}}_{\boldsymbol{u}}(\sqrt{\mu})$ (the lumped velocity matrix weighted by $\sqrt{\mu}$) to a version with boundary amplification:

$$\tilde{\mathbf{M}}_{\boldsymbol{u}}(w_{\mu,a}) \quad \text{where} \quad w_{\mu,a}(\boldsymbol{x}) \coloneqq \begin{cases} a\sqrt{\mu(\boldsymbol{x})} & \boldsymbol{x} \in \Omega_D \\ \sqrt{\mu(\boldsymbol{x})} & \boldsymbol{x} \notin \Omega_D \end{cases} \tag{5.1}$$

We distinguish between left boundary amplification a_l for $\mathbf{C}_\mu = \tilde{\mathbf{M}}_{\boldsymbol{u}}(w_{\mu,a_l})$ and right boundary amplification a_r for $\mathbf{D}_\mu = \tilde{\mathbf{M}}_{\boldsymbol{u}}(w_{\mu,a_r})$ in (1.4). Note that amplifying of the weighting function at the boundary is similar to damping at the boundary after taking the inverses $\mathbf{C}^{-1}, \mathbf{D}^{-1}$, which was performed in [9].

The Stokes solver convergence under the influence of boundary amplifications a_l, a_r is summarized in Table 5.1 where the number of GMRES iterations for 10^{-6} residual reduction is given. Additionally, the fastest iteration counts +2 are highlighted in color. The highlighting creates a "pattern of fast convergence" showing that the boundary amplification is most effective when performed non-symmetrically, i.e., either $a_l > 1$ or $a_r > 1$ but not both. Further, we learn that with higher mesh refinement level ℓ the boundary amplification should increase roughly proportional to 2^ℓ. Similar observations can be made for the discretization order k. Note that these implications were made after extensive numerical experiments for which Table 5.1 just serves as a brief summary.

6. Parallel hybrid spectral-geometric-algebraic multigrid (HMG) for μ-BFBT. Two aspects of the Stokes preconditioner with μ-BFBT have not been discussed yet. One is the approximation of the inverse viscous block $\tilde{\mathbf{A}}^{-1}$ required in (1.3) and the other is the approximation of inverses $(\mathbf{B}\mathbf{C}_\mu^{-1}\mathbf{B}^\top)^{-1}$ and $(\mathbf{B}\mathbf{D}_\mu^{-1}\mathbf{B}^\top)^{-1}$ in (1.4). For brevity, we introduce the notation $\mathbf{K}_{\mathrm{d}} \coloneqq \mathbf{B}\mathbf{C}_\mu^{-1}\mathbf{B}^\top$. The operator $\mathbf{K}_{\mathrm{d}}$ can be regarded as a discrete variable-coefficient Poisson operator on the discontinuous pressure space $\mathbb{P}^{\mathrm{disc}}_{k-1}$ with Neumann boundary conditions (recall from Section 5 that $\mathbf{C}_\mu = \mathbf{D}_\mu$ when $a_l = a_r$).

TABLE 5.1

Influence of boundary amplifications a_l, a_r on the Stokes solver convergence with μ-BFBT for discretizations: (a) $k = 2, \ell = 5$, (b) $k = 2, \ell = 7$, (c) $k = 5, \ell = 5$. Reported are the number of GMRES iterations for 10^{-6} residual reduction for the problem S16-rand, $\mathrm{DR}(\mu) = 10^6$. Colors highlight lowest iteration count +2. Increase of mesh refinement level ℓ or discretization order k demands larger boundary amplification to maintain fast convergence.

(a) $k = 2, \ell = 5$

$a_l \setminus a_r$	1	2	4	8	16	32
1	33	33	34	34	34	35
2	33	33	34	34	34	34
4	33	34	34	36	38	39
8	34	34	36	39	43	44
16	34	34	38	43	46	49
32	34	34	39	44	49	53

(b) $k = 2, \ell = 7$

$a_l \setminus a_r$	1	2	4	8	16	32
1	45	37	34	34	34	34
2	37	36	35	36	36	36
4	34	36	38	39	40	41
8	34	36	39	42	44	44
16	34	36	40	44	45	46
32	34	36	41	44	46	47

(c) $k = 5, \ell = 5$

$a_l \setminus a_r$	1	2	4	8	16	32
1	63	53	46	43	43	44
2	53	51	51	51	52	53
4	47	51	55	59	62	64
8	44	51	59	65	69	72
16	43	52	62	69	75	78
32	44	53	64	72	78	82

The approximation of the inverse viscous block $\tilde{\mathbf{A}}^{-1}$ is well suited for multigrid V-cycles. In [15], we developed a hybrid spectral-geometric-algebraic multigrid (HMG) for that purpose, which exhibits extreme parallel scalability and retains nearly optimal algorithmic scalability (see also Section 7 for scalability results). While traversing the V-cycle shown in Figure 6.1, HMG initially reduces the discretization order (spectral multigrid); after arriving at order one, it continues by coarsening mesh elements (geometric multigrid); once the degrees of freedom fall below a threshold, algebraic multigrid (AMG) carries out further coarsening until a direct solve can be computed efficiently. Parallel forest-of-octrees algorithms, implemented in the p4est parallel adaptive mesh refinement library, are used for efficient, scalable mesh refinement/coarsening, mesh balancing, and repartitioning [5]. During the geometric coarsening, the number of compute cores and the size of the MPI communicator is reduced successively to minimize communication. Then the transition to AMG is done at a sufficiently small core count. AMG continues to further reduce the number of cores down to a single core for the direct solver.

Additionally, multigrid V-cycles can also also be employed to approximate the pressure Poisson operator $\mathbf{K}_\mathrm{d}$. However, it turned out to be problematic to apply the HMG strategy directly due to the discontinuous, modal discretization of the pressure space. We took a novel approach in [15] by considering the underlying infinite-dimensional variable-coefficient Poisson operator, where the coefficient is derived from the diagonal scaling matrix (here, $\mathbf{C}_\mu^{-1}$ or $\mathbf{D}_\mu^{-1}$). Then we re-discretize with continuous, nodal high-order finite elements in $\mathbb{Q}_k$. This continuous, nodal Poisson operator is then approximately inverted with an HMG V-cycle that is similar to the one described above for the inverse viscous block approximation $\tilde{\mathbf{A}}^{-1}$. Additional smoothing is applied in the pressure space (Figure 6.1, *green level*) to account for modes in a residual that are present in $\mathbb{P}_{k-1}^{\mathrm{disc}}$ but not in $\mathbb{Q}_k$.

In numerical experiments throughout the paper, three smoothing iterations with Chebyshev accelerated point-Jacobi smoother at downward and upward traversal of the V-cycle were performed. PETSc library [1] implementations of Chebyshev acceleration, AMG, direct solver, and GMRES were used.

7. Algorithmic and parallel scalability for HMG + μ-BFBT Stokes preconditioner. After establishing the robustness of the Stokes solver due to μ-BFBT preconditioning in Section 4, we study the scalability of the solver in this section. One component of scalability is algorithmic scalability, which is the dependence of GMRES iterations on the resolution of the mesh and the discretization order. The second component is parallel scalability where runtime is measured on increasing numbers of compute cores. Presenting both components of scalability is required to fully assess the performance of a solver at scale.

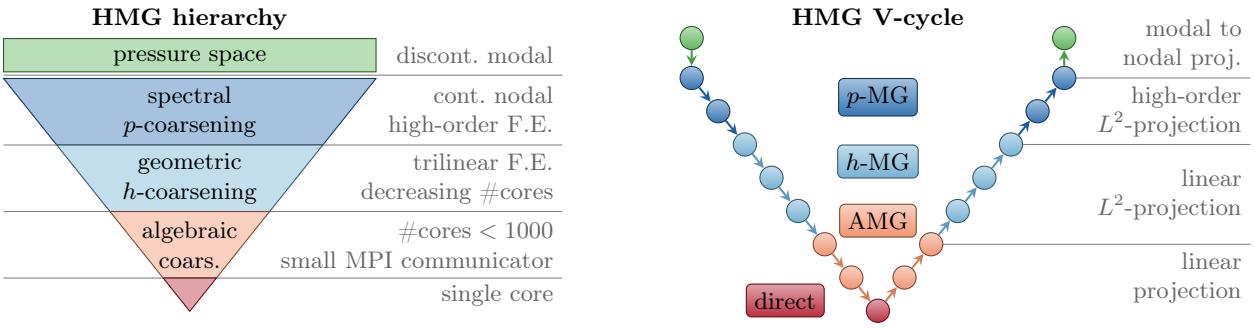


FIGURE 6.1. *Hybrid spectral-geometric-algebraic multigrid (HMG). Left: Illustration of multigrid hierarchy. From top to bottom, first, the multigrid levels are obtained by spectral coarsening (dark blue). Next, the mesh is geometrically coarsened and repartitioned on successively fewer cores to minimize communication (light blue). Finally, AMG further reduces problem size and core count (light red). The multigrid hierarchy for the pressure Poisson operator* $\mathbf{K}_\mathrm{d}$ *additionally involves smoothing in the discontinuous modal pressure space (green). Right: The multigrid V-cycle consists of smoothing at each level of the hierarchy (circles) and intergrid transfer operators (arrows downward for restriction and arrows upward for interpolation). To enhance efficacy of the the V-cycle as a preconditioner, different types of projection operators are employed for these operators depending on the phase within the V-cycle.*

The algorithmic scalability in Table 7.1 shows results for the Stokes solver as well as its individual components by giving iteration numbers for solving the systems $\mathbf{Au} = \mathbf{f}$ and $\mathbf{K}_\mathrm{d}\mathbf{p} = \mathbf{g}$. By studying the individual components we seek to potentially identify poor approximations of the Schur complement. All systems, Stokes, $\mathbf{A}$, and $\mathbf{K}_\mathrm{d}$, are solved with preconditioned GMRES down to a relative tolerance of 10^{-6}. The preconditioners for $\mathbf{A}$ and $\mathbf{K}_\mathrm{d}$ are HMG-V-cycles as described in Section 6. For the μ-BFBT preconditioner, we set a constant left boundary amplification $a_l = 1$ and vary the right boundary amplification a_r according to results from Section 5. The iteration counts show ideal mesh independence when increasing the level of refinement ℓ in Table 7.1a. This holds for both components $\mathbf{A}$ and $\mathbf{K}_\mathrm{d}$ and also the whole Stokes solver, hence the Schur complement approximation by μ-BFBT is also mesh-independent. When the discretization order k grows the iteration counts presented in Table 7.1b increase mildly. The convergence of both components $\mathbf{A}$ and $\mathbf{K}_\mathrm{d}$ exhibits a small dependence on k. Since the increase in number of iterations is sightly larger for the full Stokes solve than for $\mathbf{A}$ and $\mathbf{K}_\mathrm{d}$, we assume a mild deterioration of μ-BFBT as a Schur complement approximation. A reason can be the lumping performed to generate the diagonal scaling matrices $\mathbf{C}_\mu^{-1}$ and $\mathbf{D}_\mu^{-1}$.

Parallel scalability results were obtained on the full Lonestar 5 peta-scale system at the Texas Advanced Computing Center (TACC). Lonestar 5 is a new Cray XC40 system consisting of 1252 compute nodes. Each node is equipped with two Intel Haswell 12-core processors and 64 GBytes of memory. The Lonestar 5 supercomputer entered production in January 2016 and is a new architecture. It is therefore desirable to perform scalability measurements in order to broaden our knowledge of the parallel performance of our Stokes solver (in addition to results in [15]).

The weak scalability (DOF/core fixed to ∼1 million) in Figure 7.1a shows that the Stokes solver with μ-BFBT *(blue curve)* maintains 90 % efficiency over a 618-fold increase in degrees of freedom along with cores. Even for the setup of the Stokes solver *(green curve)*, which mainly involves the HMG hierarchy generation, we observe 71 % efficiency. We consider these to be remarkable results for such a complex implicit multi-level solver with optimal algorithmic performance (when the mesh is refined or nearly algorithmically optimal when the order is increased) and with a convergence rate that is by design independent of the number of cores. Furthermore, extreme scalability results utilizing IBM's BG/Q architecture were demonstrated in [15]. There, we achieved 97 % weak efficiency over a 96-fold core increase up to 1.6 million cores for the solve phase of the

TABLE 7.1

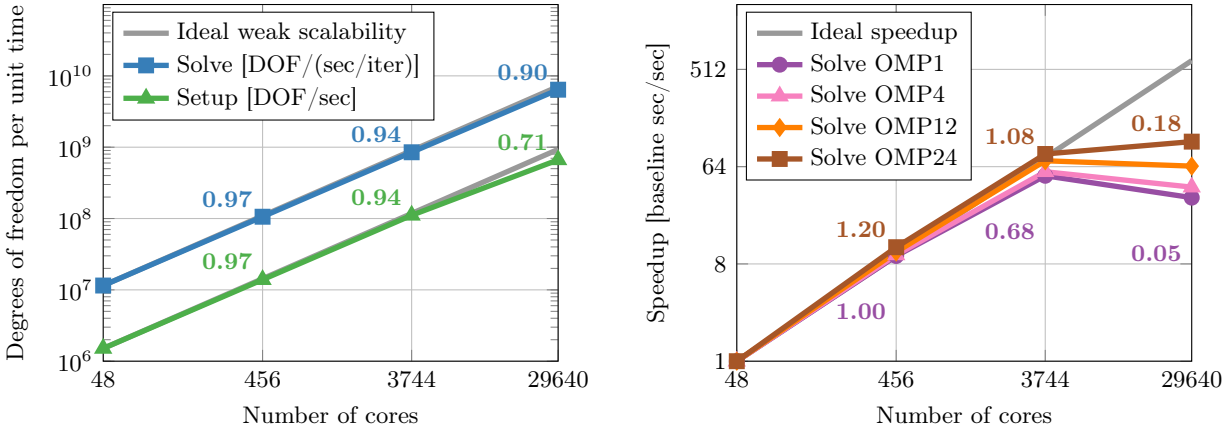
Algorithmic scalability for Stokes solver with HMG+μ-BFBT preconditioning while (a) varying mesh refinement level ℓ and (b) varying discretization order k (problem S16-rand, $\mathrm{DR}(\mu) = 10^6$). Computational cost is expressed in number of GMRES iterations (abbreviated by It.) for full Stokes solve (10^{-6} residual reduction). Left boundary amplification for $\mathbf{C}_\mu$ is fixed to $a_l = 1$; right boundary amplification a_r for $\mathbf{D}_\mu$ varies. Additionally, the number of GMRES iterations for solving only the sub-systems $\mathbf{Au} = \mathbf{f}$ and $\mathbf{K}_\mathrm{d}\mathbf{p} = \mathbf{g}$ are given for demonstration of HMG efficacy.

(a) Algorithmic scalability (fixed order $k = 2$)

ℓ	a_r	$\boldsymbol{u}$-DOF [$\times 10^6$]	It. **A**	p-DOF [$\times 10^6$]	It. $\mathbf{K}_\mathrm{d}$	DOF [$\times 10^6$]	It. **Stokes**
4	1	0.11	18	0.02	8	0.12	40
5	2	0.82	18	0.13	7	0.95	33
6	4	6.44	18	1.05	6	7.49	33
7	8	50.92	18	8.39	6	59.31	34
8	16	405.02	18	67.11	6	472.12	34
9	32	3230.67	18	536.87	6	3767.54	34
10	64	25807.57	18	4294.97	6	30102.53	34

(b) Algorithmic scalability (fixed level $\ell = 5$)

k	a_r	$\boldsymbol{u}$-DOF [$\times 10^6$]	It. **A**	p-DOF [$\times 10^6$]	It. $\mathbf{K}_\mathrm{d}$	DOF [$\times 10^6$]	It. **Stokes**
2	2	0.82	18	0.13	7	0.95	33
3	4	2.74	20	0.32	8	3.07	37
4	8	6.44	20	0.66	7	7.10	36
5	16	12.52	23	1.15	12	13.67	43
6	32	21.56	23	1.84	12	23.40	50
7	64	34.17	22	2.75	10	36.92	54
8	128	50.92	22	3.93	10	54.86	67



(a) Weak scalability ($k = 2$, $\ell = 7, \ldots, 10$)

(b) Strong scalability ($k = 2$, $\ell = 7$)

FIGURE 7.1. *Parallel scalability on Lonestar 5 for Stokes solver with HMG+μ-BFBT preconditioning (problem S16-rand, $\mathrm{DR}(\mu) = 10^6$ as in Table 7.1a). (a) Weak scalability of setup and solve phases (normalized w.r.t. deviations from const. DOF/core). Numbers along the graph lines indicate weak efficiency w.r.t. ideal weak scalability (baseline is 48 cores result). DOF/core is ∼1 million, the largest problem size on 29640 cores has 30 billion DOF. (b) Strong scalability of solve phase for different configurations of OpenMP-threads (OMP) substituting MPI ranks s.t. one thread is assigned to each core. Numbers along the graph lines indicate strong efficiency w.r.t. ideal speedup (baseline is 48 cores result).*

HMG + diag(**A**)-BFBT preconditioned Stokes solver with the additional difficulty of handling highly adaptively refined meshes. Note that μ-BFBT and diag(**A**)-BFBT are similar in terms of parallel scalability, because we measure time per GMRES iteration.

Finally, Figure 7.1b reports strong scalability results (DOF fixed to 59 million) and how the number of OpenMP threads substituting MPI ranks influences speedup. Over the 78-fold increase from 48 to 3744 cores, efficiency reduces moderately, which is 68 % in the worst case OMP1. However, in the largest run with 29,640 cores, the granularity is only ∼2000 DOF/core, which becomes problematic for strong scalability. This behavior is expected for an implicit solver, especially for a multi-level based method that does not sacrifice algorithmic optimality for parallel scalability.

8. Conclusion. The numerical performance of μ-BFBT as an approximation to the inverse Schur complement for problems with variations in the viscosity is very promising. The combination of robustness as well as algorithmic and parallel scalability (when combined with HMG) make it an appealing choice for challenging large-scale Stokes flow applications with highly heterogenous viscosity. Since locally refined meshes are important in many applications, for instance in geophysics, extending the treatment of boundary conditions in μ-BFBT to adaptively refined meshes is highly desirable. This is the subject of our current research. Furthermore, we seek theoretical arguments that can support the efficacy of μ-BFBT as a preconditioner and systematic derivations for modifications at Dirichlet boundaries.

REFERENCES

[1] Satish Balay, Shrirang Abhyankar, Mark F. Adams, Jed Brown, Peter Brune, Kris Buschelman, Lisandro Dalcin, Victor Eijkhout, William D. Gropp, Dinesh Kaushik, Matthew G. Knepley, Lois Curfman McInnes, Karl Rupp, Barry F. Smith, Stefano Zampini, and Hong Zhang, *PETSc users manual*, Tech. Report ANL-95/11 - Revision 3.6, Argonne National Laboratory, 2015.

[2] Michele Benzi, Gene H. Golub, and Jörg Liesen, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.

[3] Carsten Burstedde, Omar Ghattas, Michael Gurnis, Eh Tan, Tiankai Tu, Georg Stadler, Lucas C. Wilcox, and Shijie Zhong, *Scalable adaptive mantle convection simulation on petascale supercomputers*, in Proceedings of SC08, ACM/IEEE, 2008, pp. 1–15. Gordon Bell Prize finalist.

[4] Carsten Burstedde, Omar Ghattas, Georg Stadler, Tiankai Tu, and Lucas C. Wilcox, *Parallel scalable adjoint-based adaptive solution for variable-viscosity Stokes flows*, Comput. Method. Appl. M., 198 (2009), pp. 1691–1700.

[5] Carsten Burstedde, Lucas C. Wilcox, and Omar Ghattas, *p4est: Scalable algorithms for parallel adaptive mesh refinement on forests of octrees*, SIAM J. Sci. Comput., 33 (2011), pp. 1103–1133.

[6] Howard C. Elman, *Preconditioning for the steady-state Navier-Stokes equations with low viscosity*, SIAM J. Sci. Comput., 20 (1999), pp. 1299–1316.

[7] Howard C. Elman, V. E. Howle, John Shadid, Robert Shuttleworth, and Ray Tuminaro, *Block preconditioners based on approximate commutators*, SIAM J. Sci. Comput., 27 (2006), pp. 1651–1668.

[8] Howard C. Elman, David J. Silvester, and Andrew J. Wathen, *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, second ed., 2014.

[9] Howard C. Elman and Ray S. Tuminaro, *Boundary conditions in approximate commutator preconditioners for the Navier-Stokes equations*, Electron. Trans. Numer. Anal., 35 (2009), pp. 257–280.

[10] Piotr P. Grinevich and Maxim A. Olshanskii, *An iterative method for the Stokes-type problem with variable viscosity*, SIAM J. Sci. Comput., 31 (2009), pp. 3959–3978.

[11] Tobin Isaac, Georg Stadler, and Omar Ghattas, *Solution of nonlinear Stokes equations discretized by high-order finite elements on nonconforming and anisotropic meshes, with application to ice sheet dynamics*, SIAM J. Sci. Comput., 37 (2015), pp. B804–B833.

[12] David Kay, Daniel Loghin, and Andrew Wathen, *A preconditioner for the steady-state Navier-Stokes equations*, SIAM J. Sci. Comput., 24 (2002), pp. 237–256.

[13] Dave A. May, Jed Brown, and Laetitia Le Pourhiet, *pTatin3D: High-performance methods for long-term lithospheric dynamics*, in Proceedings of SC14, IEEE Press, 2014, pp. 274–284.

[14] Dave A. May and Louis Moresi, *Preconditioned iterative methods for Stokes flow problems arising in computational geodynamics*, Phys. Earth Planet. In., 171 (2008), pp. 33–47.

[15] Johann Rudi, A. Cristiano I. Malossi, Tobin Isaac, Georg Stadler, Michael Gurnis, Yves Ineichen, Costas Bekas, Alessandro Curioni, and Omar Ghattas, *An extreme-scale implicit solver for complex PDEs: Highly heterogeneous flow in earth's mantle*, in Proceedings of SC15, ACM, 2015, pp. 5:1–5:12. Winner of Gordon Bell Prize.

[16] David J. Silvester, Howard C. Elman, David Kay, and Andrew J. Wathen, *Efficient preconditioning of the linearized Navier-Stokes equations for incompressible flow*, J. Comput. Appl. Math., 128 (2001), pp. 261–279. Numerical analysis 2000, Vol. VII, Partial differential equations.

[17] Georg Stadler, Michael Gurnis, Carsten Burstedde, Lucas C. Wilcox, Laura Alisic, and Omar Ghattas, *The dynamics of plate tectonics and mantle flow: From local to global scales*, Science, 329 (2010), pp. 1033–1038.