

A staggered-grid multilevel incomplete LU for steady incompressible flows

Sven Baars¹  | Mark van der Kloek² | Jonas Thies³ | Fred W. Wubs¹

¹Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, The Netherlands

²Zernike Institute for Advanced Materials, University of Groningen, Groningen, The Netherlands

³Simulation and Software Technology, German Aerospace Center (DLR), Cologne, Germany

Correspondence

Sven Baars, Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Nijenborgh 9, 9747 AG Groningen, The Netherlands
Email: s.baars@rug.nl

Funding information

Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 657.014.007; Netherlands eScience Center, Grant/Award Number: 027.017.G02

Abstract

Algorithms for studying transitions and instabilities in incompressible flows typically require the solution of linear systems with the full Jacobian matrix. Other popular approaches, like gradient-based design optimization and fully implicit time integration, also require very robust solvers for this type of linear system. We present a parallel fully coupled multilevel incomplete factorization preconditioner for the 3D stationary incompressible Navier-Stokes equations on a structured grid. The algorithm and software are based on the robust two-level method developed by Wubs and Thies. In this article, we identify some of the weak spots of the two-level scheme and propose remedies such as a different domain partitioning and recursive application of the method. We apply the method to the well-known 3D lid-driven cavity benchmark problem, and demonstrate its superior robustness by comparing with a segregated SIMPLE-type preconditioner.

KEYWORDS

F-matrix, incomplete factorization, incompressible Navier-Stokes, lid-driven cavity, multilevel, parallel

1 | INTRODUCTION

The incompressible Navier-Stokes equations accurately describe flow of Newtonian fluids like water and air at low Mach numbers. In this article, we consider the numerical solution of these equations after spatial discretization on a structured grid. The discretization we use does not introduce artificial diffusion and is therefore popular for direct numerical simulations of turbulent flow. In this flow regime, one is typically interested in accurately resolving the temporal evolution of the flow. The relatively small time steps in such a simulation allow simplifications in the solution process, in particular, Picard linearization (instead of Newton's), and a segregated solution scheme that treats the velocities and pressure separately.¹

On the other end of the spectrum are Stokes flows and flows at very low Reynolds numbers. Here it is common to solve the coupled linear systems with a segregated preconditioner; see, for example, Elman et al.² This class of linear solvers tries to reduce the problem to scalar linear systems that can be solved by multigrid methods.

In this article, we are interested in the intermediate flow regimes, where the focus lies on accurately describing flow transitions occurring when model parameters are varied (or uncertain). In such situations, it may be beneficial to directly

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *International Journal for Numerical Methods in Fluids* published by John Wiley & Sons, Ltd.

compute steady state solutions, or at least take very large implicit time steps in order to quickly reach a statistical equilibrium solution. Another application where the solution of linear systems with the full Jacobian (and its adjoint) play a key role is design optimization with gradient-based optimization methods.³ In that application, however, unstructured grids may be preferable, which we do not address here.

To study the stability of steady states in fluid flow problems as a function of a parameter, for example, the Reynolds number (see Charru and Forcrand-Millard⁴ for an introduction to such problems), a continuation method⁵ can be used. To apply such a method to high-dimensional systems, we require an efficient and robust method for solving linear systems associated with the discretized incompressible Navier-Stokes equations. An elegant way of solving these systems is by replacing the complete LU factorization by an accurate incomplete one, which can be used as a preconditioner in an iterative procedure. By an appropriate ordering technique and dropping procedure, one can construct an incomplete LU (ILU) factorization that yields grid independent convergence behavior and approaches an exact factorization as the amount of allowed fill is increased. During the continuation process, this preconditioner can also be used to find approximate smallest eigenvalues and eigenvectors of the Jacobian matrix of the incompressible Navier-Stokes equations Sleijpen and Wubs.⁶ These eigenvalues are of interest since a switch in sign may indicate a bifurcation.

The incompressible Navier-Stokes equations can be written as

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} &= -\nabla p + \frac{1}{\text{Re}} \Delta \mathbf{u}, \\ \nabla \cdot \mathbf{u} &= 0, \end{aligned} \quad (1)$$

where $\text{Re} = \frac{\rho U D}{\mu}$ is the Reynolds number, ρ is the density and μ is the dynamic viscosity, and D and U are characteristic length and velocity scales of the flow. These equations are discretized using a second-order symmetry-preserving finite volume method on an Arakawa C-grid;⁷ see Figure 1. The discretization leads to a system of ordinary differential equations (ODEs)

$$\begin{aligned} M \frac{d\mathbf{u}}{dt} + N(\mathbf{u}, \mathbf{u}) &= -G\mathbf{p} + \frac{1}{\text{Re}} L\mathbf{u} + \mathbf{f}_u, \\ -G^T \mathbf{u} &= \mathbf{f}_p, \end{aligned}$$

where \mathbf{u} and \mathbf{p} now represent the velocity and pressure in each grid point, $N(\cdot, \cdot)$ is the bilinear form arising from the convective terms, L is the discretized Laplace operator, G is the discretized gradient operator, M is the mass matrix, which contains the volumes of the grid cells on its diagonal, and \mathbf{f} contains the known part of the boundary conditions. Our method will exploit the property that the divergence operator is given by $-G^T$.

The term $N(\mathbf{u}, \mathbf{v})$ is the discretized form of $\mathbf{u} \cdot \nabla \mathbf{v}$, and is bilinear. Hence for given \mathbf{u} the expression is linear in \mathbf{v} : there exists a matrix $N_1(\mathbf{u})$ such that $N(\mathbf{u}, \mathbf{v}) = N_1(\mathbf{u})\mathbf{v}$. Similarly, for given \mathbf{v} , there exists a matrix $N_2(\mathbf{v})$ such that $N(\mathbf{u}, \mathbf{v}) = N_2(\mathbf{v})\mathbf{u}$, which is the discretized form of $\mathbf{v}(\nabla \cdot \mathbf{u})$. For the contribution of $N(\mathbf{u}, \mathbf{v})$ to the Jacobian, we consider the derivative of $N(\mathbf{u}, \mathbf{u})$ in the direction $\Delta \mathbf{u}$, which is found from the following expression by taking the limit $\epsilon \rightarrow 0$:

$$[N(\mathbf{u} + \epsilon \Delta \mathbf{u}, \mathbf{u} + \epsilon \Delta \mathbf{u}) - N(\mathbf{u}, \mathbf{u})]/\epsilon = N(\mathbf{u}, \Delta \mathbf{u}) + N(\Delta \mathbf{u}, \mathbf{u}) + \epsilon N(\Delta \mathbf{u}, \Delta \mathbf{u}) = N_1(\mathbf{u})\Delta \mathbf{u} + N_2(\mathbf{u})\Delta \mathbf{u} + \epsilon N(\Delta \mathbf{u}, \Delta \mathbf{u}),$$

where we used only the bilinearity of the expression $N(\cdot, \cdot)$. The last term becomes zero when taking the limit.

Using this notation, the linear system of saddle point type⁸ that has to be solved within each Newton step is given by

$$\begin{pmatrix} N_1(\mathbf{u}) + N_2(\mathbf{u}) - \frac{1}{\text{Re}} L & G \\ G^T & 0 \end{pmatrix} \begin{pmatrix} \Delta \mathbf{u} \\ \Delta \mathbf{p} \end{pmatrix} = - \begin{pmatrix} \mathbf{f}_u \\ \mathbf{f}_p \end{pmatrix}. \quad (2)$$

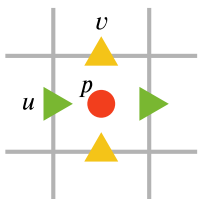


FIGURE 1 Positioning of unknowns in an Arakawa C-grid [Colour figure can be viewed at wileyonlinelibrary.com]

It is known¹ that, on closed domains, dissipation of kinetic energy only occurs by diffusion. Our discretization preserves this property, which has the consequence that for divergence-free \mathbf{u} , the operator $N_1(\mathbf{u})$ is skew-symmetric. A popular simplification is to omit $N_2(\mathbf{u})$, which replaces the Newton process by a Picard iteration.² The approximate Jacobian then becomes negative definite (on the space of discrete divergence-free velocities), which greatly simplifies the solution process since definiteness is a condition under which many standard iterative methods converge. We remark, however, that Picard iteration works well only for fairly low Reynolds numbers and far away from bifurcation points where steady solutions become unstable, and seriously impairs the convergence rate of the nonlinear iteration.^{9,10} Furthermore, some authors use time dependent approaches to study the stability of steady states.¹¹ This approach, however, also requires some tricks to obtain the desired information. Since we want to study precisely the phenomena where the above methods experience difficulties, we would rather use the full Jacobian matrix of the nonlinear equations, applying Newton's method.

There are many methods, based on segregation of physical variables, that can solve the linear equations that arise in every Newton iteration. In this approach the system is split into subproblems of an easier form for which standard methods exist. The segregation can already be done at the discretization level, for example, by doing a time integration and solving a pressure-correction equation independently of the momentum equations.^{1,12} Another class of methods performs the segregation during the linear system solve, often in a preconditioning step. Physics based preconditioners¹³⁻¹⁶ try to split the problem into subsystems which capture the bulk of the physics. The subsystems are again solved by iterative procedures, for example, algebraic multigrid (AMG) for Poisson-like equations. These methods consist of nested loops for: the nonlinear iteration, iterations over the coupled system, and iterations over the subsystems. The stopping criteria of all these different iterations have to be tuned to make the solver efficient. Furthermore, the total number of iterations in the innermost loop is given by the product of the number of iterations performed on all three levels of iteration and thus easily becomes excessive. This is a major problem when trying to achieve extreme parallelism, because each innermost iteration typically requires global communication in the inner products. The number of levels of nested iteration may increase even more if additional physical phenomena are added.^{13,17} Our ultimate aim is to remove the inner iterations altogether and to solve the nonlinear equations using a subspace accelerated inexact Newton method. In Sleijpen and Wubs,⁶ we did this for simple eigenvalue problems using the Jacobi-Davidson method, which is in fact an accelerated Newton method. The method we present here will make this feasible for the 3D Navier-Stokes equations, even at high Reynolds numbers and close to bifurcation points, for which we have so far failed to find other scalable parallel solution methods.

To achieve this, fully aggregated approaches are important. In this category, multigrid and multilevel ILU methods for systems of PDEs exist (see Trottenberg et al.¹⁸ and Wathen¹⁹ and references therein). The former is attractive, but for those methods smoothers may fail due to a loss of diagonal dominance for higher Reynolds numbers, for which a common solution is to use time integration.²⁰ The latter comprise AMG algorithms and multilevel methods like MRILU²¹ and the methods available in ILUPACK.²² ILUPACK has been successful in many fields since it uses a bound to preclude very unstable factorizations. However, this method does not show grid independence for Navier-Stokes problems and is difficult to parallelize.²³ It works well for large problems, but not yet beyond a single shared memory system.

In this article, we present a novel multilevel preconditioning method which is specially designed for the 3D Navier-Stokes equations. In Section 2, we first describe the two-level ILU preconditioner as introduced in Wubs and Thies²⁴ and Thies and Wubs.²⁵ After this, we generalize the two-level method to a multilevel method in Section 3. To make this method work for the 3D Navier-Stokes equations, we introduce a skew partitioning method in Section 4. In Section 5, we discuss the scalability and general performance of the method, and compare it to a popular physics based method, after which we summarize the paper in Section 6.

2 | THE TWO-LEVEL ILU PRECONDITIONER

In Wubs and Thies,²⁴ a robust parallel two-level method was developed for solving

$$A\mathbf{x} = \mathbf{b},$$

with $A \in \mathbb{R}^{n \times n}$ for a class of matrices known as \mathcal{F} -matrices. An \mathcal{F} -matrix is a matrix of the form

$$A = \begin{pmatrix} K & B \\ B^T & 0 \end{pmatrix},$$

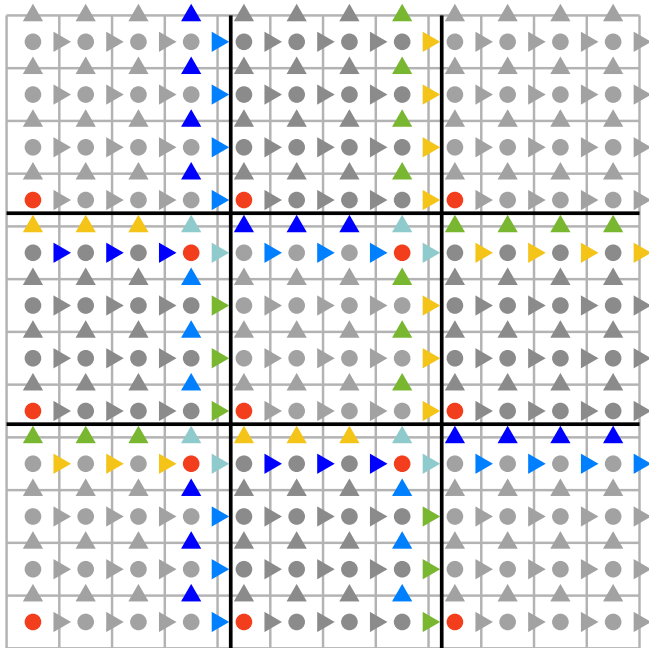


FIGURE 2 Cartesian partitioning of a 12×12 C-grid discretization of the Navier-Stokes equations into 9 subdomains of size $s_x \times s_y$ with $s_x = s_y = 4$. The interiors are shown in gray. Velocities of the same (nongray) color together form a separator group (but only if they point in the same direction and are in neighboring grid cells). The red circles are pressure nodes that are retained [Colour figure can be viewed at wileyonlinelibrary.com]

with K symmetric positive definite and B such that it has at most two entries per row and the entries in each row sum up to 0, as is the case for our gradient matrix G .^{26,27} It has been shown that the two-level preconditioner leads to grid-independent convergence for the Stokes equations on an Arakawa C-grid, and that the method is robust even for the Navier-Stokes equations, which strictly speaking do not yield \mathcal{P} -matrices because K becomes nonsymmetric and possibly indefinite.

Rather than reviewing the method and theory in detail, we only briefly present it here. For details, see Wubs and Thies²⁴ and Thies and Wubs.²⁵

To simplify the discussion, we focus on the special case of the 3D incompressible Navier-Stokes equations in a cube-shaped domain, discretized on an Arakawa C-grid (see Figure 1). We refer to the velocity variables, which are located on the cell faces as V -nodes, and to the pressure, which is located in the cell center, as P -node. The variables are numbered cell-by-cell, that is, the first three variables are the $u/v/w$ -velocity at the north/east/top face of the cell in the bottom south west corner of the domain, and variable four is the P -node in its center. Appropriate boundary conditions (e.g., Dirichlet conditions) are easily implemented in this situation. We remark that the algorithm (and software) can handle more general situations like rectangular domains, interior boundary cells, and so on, and could be generalized to arbitrary domain shapes and partitionings.

First we describe the initialization phase of the preconditioner, which is the necessary preprocessing that has to be done only once for a series of linear systems with matrices with the same sparsity pattern. Then we describe the factorization phase, which has to be done separately for every matrix. Finally, we describe the solution phase, which has to be performed for every application of the preconditioner.

2.1 | Initialization phase

First the system is partitioned into N nonoverlapping subdomains Ω_α , with $\alpha = 1, \dots, N$. These subdomains may be distributed over different processes, which allows for parallelization of the computation. The partitioning may be done based on the graph of the matrix, as implemented for instance in METIS,²⁸ or by a geometric approach, for example, by dividing the domain into rectangular subdomains. An example of a Cartesian partitioning of a square domain is shown in Figure 2. The nonoverlapping subdomains are denoted by the black lines.

Based on the nonoverlapping decomposition, we can define a global minimal set of separator nodes Γ such that it decouples the linear systems associated with the remaining variables in any two subdomains $\Omega_{\alpha,\beta}$. This Γ can straightforwardly be defined based on either geometric information or the graph of the matrix. Let $I_\alpha = \Omega_\alpha \setminus \Gamma$ denote the set of

interior variables of subdomain Ω_α , then we have by construction

$$\forall \alpha \neq \beta, \quad i \in I_\alpha, \quad j \in I_\beta : A_{ij} = 0. \tag{3}$$

Furthermore, we define the set of separator variables associated with subdomain Ω_α as $S_\alpha = \{j \in \Gamma : \exists i \in I_\alpha : A_{ij} \neq 0\}$. Note that the interior variables of two subdomains can be eliminated independently on a parallel computer, and that the union of the sets I_α and S_α defines an overlapping partitioning. Although we use the nonoverlapping domain decomposition to define our incomplete factorization, we formally introduce the overlapping subdomains $\bar{\Omega}_\alpha = I_\alpha \cup S_\alpha$. In the remainder of this article, we will refer to the variables in the sets I_α and $S_\alpha, \alpha = 1, \dots, N$ as *interior nodes*, and *separator nodes*, respectively. One *separator* is defined as a set of separator nodes that are coupled with the same set of subdomains (geometrically, separators comprise faces, edges, and corners). One *separator group* is defined as the variables on the same separator that have the same variable type (u, v, w , or p). Note that because separators are defined by the set of subdomains they are coupled to, separator nodes can never be in multiple separator groups at the same time. In Figure 2, the interior nodes are shown in gray and the different separator groups are denoted by different colors.

We can now reorder the matrix A such that the interiors (I) and separators (S) are grouped. This gives us the system

$$\begin{pmatrix} A_{II} & A_{IS} \\ A_{SI} & A_{SS} \end{pmatrix} \begin{pmatrix} \mathbf{x}_I \\ \mathbf{x}_S \end{pmatrix} = \begin{pmatrix} \mathbf{b}_I \\ \mathbf{b}_S \end{pmatrix},$$

where A_{II} consists of independent diagonal blocks. This submatrix is invertible because on each subdomain we deal with a discretized Navier-Stokes problem on an Arakawa C-grid which is known to be well-posed if the normal and tangential velocities are specified on the boundaries and the level of the pressure is fixed by setting it in one grid point inside the domain. The velocities are indeed specified on the surrounding boundaries and separators of each subdomain and one pressure is fixed in each subdomain. Since A_{II} consists of independent diagonal blocks, applying A_{II}^{-1} is easy and trivial to parallelize. For this reason, we aim to solve

$$\begin{aligned} S\mathbf{x}_S &= \mathbf{b}_S - A_{SI}A_{II}^{-1}\mathbf{b}_I, \\ \mathbf{x}_I &= A_{II}^{-1}\mathbf{b}_I - A_{II}^{-1}A_{IS}\mathbf{x}_S, \end{aligned}$$

where S is the Schur complement given by $S = A_{SS} - A_{SI}A_{II}^{-1}A_{IS}$.

Whether a variable is coupled to a different subdomain could be determined directly from the graph of the matrix, as discussed earlier. In our implementation, however, we determine this geometrically by defining the overlapping subdomains during the partitioning step, and checking what overlapping subdomains a node of the nonoverlapping subdomain belongs to.

There are three types of separators: faces (in 3D), edges and corners. For the Navier-Stokes problem on a C-grid, these separators only consist of V -nodes. The P -nodes are only connected to V -nodes that belong to the same overlapping subdomain, so these should never lie on a separator. We arbitrarily choose one P -node in every interior which we also define to be its own separator group to make sure the Schur complement stays nonsingular.

We want to eliminate velocities on a separator in such a way that the velocities that remain on a separator face provide an approximation of the collective flux through that face. It is therefore important that the variables are correctly scaled before the factorization in a way that they represent physical fluxes. In the matrix in Equation (2) this gives a G -part with entries of constant magnitude. In this case, we can define a Householder transformation which exactly decouples all but one V -node from the P -nodes.²⁴ This transformation is of the form

$$H_j = I - 2 \frac{\mathbf{v}_j \mathbf{v}_j^T}{\mathbf{v}_j^T \mathbf{v}_j}, \tag{4}$$

for some separator group g_j with

$$\mathbf{v}_j^{(k)} = \begin{cases} \mathbf{e}_j^{(k)} + \|\mathbf{e}_j\|_2 & \text{if node } k \text{ is the first node of separator group } g_j \\ \mathbf{e}_j^{(k)} & \text{otherwise} \end{cases} \tag{5}$$

and

$$\mathbf{e}_j^{(k)} = \begin{cases} 1 & \text{if node } k \text{ is in separator group } g_j \\ 0 & \text{if node } k \text{ is not in separator group } g_j, \end{cases}$$

for all $k = 1, \dots, n$. We call the one V -node that is still coupled to the P -nodes a V_Σ node. Since the Householder transformation can be applied independently for every separator group, we may collect all these transformations in one single transformation matrix H .

2.2 | Factorization phase

For every overlapping subdomain $\bar{\Omega}_i, i = 1, \dots, N$, where N is the total number of overlapping subdomains, we can define a local matrix

$$A^{(i)} = \begin{pmatrix} A_{II}^{(i)} & A_{IS}^{(i)} \\ A_{SI}^{(i)} & A_{SS}^{(i)} \end{pmatrix}.$$

After computing the factorization $A_{II}^{(i)} = L_{II}^{(i)} U_{II}^{(i)}$, the local contribution to the Schur complement is given by

$$S_i = A_{SS}^{(i)} - A_{SI}^{(i)} (L_{II}^{(i)} U_{II}^{(i)})^{-1} A_{IS}^{(i)},$$

and the global Schur complement is given by

$$S = \sum_{\Omega_i} A_{SS}^{(i)} - \sum_{\Omega_i} A_{SI}^{(i)} (L_{II}^{(i)} U_{II}^{(i)})^{-1} A_{IS}^{(i)}.$$

Here we take the sum of the $A_{SS}^{(i)}$ contributions over the nonoverlapping subdomains to make sure that contributions from separators are not counted multiple times.

We now apply the Householder transformation

$$\begin{aligned} S_T = HSH^T &= H \left(\sum_{\Omega_i} A_{SS}^{(i)} - \sum_{\Omega_i} A_{SI}^{(i)} (L_{II}^{(i)} U_{II}^{(i)})^{-1} A_{IS}^{(i)} \right) H^T \\ &= \sum_{\Omega_i} H_i A_{SS}^{(i)} H_i^T - \sum_{\Omega_i} H_i A_{SI}^{(i)} (L_{II}^{(i)} U_{II}^{(i)})^{-1} A_{IS}^{(i)} H_i^T, \end{aligned} \quad (6)$$

which can be done separately for every separator group or subdomain, or on the entire Schur complement. We choose to apply the transformation separately for every subdomain, since H is very sparse, and sparse matrix-matrix products are very expensive and memory consuming.

We now drop all connections between V_Σ and non- V_Σ nodes, and between non- V_Σ nodes and non- V_Σ nodes in different separator groups. The dropping that is applied here is what makes our factorization inexact. Not applying the dropping gives us a factorization that can still be partially parallelized, but is also exact.

Our transformed Schur complement is now reduced to a block-diagonal matrix with blocks of non- V_Σ nodes for every separator, and one block for all V_Σ nodes, which we call $S_{\Sigma\Sigma}$. Separate factorizations can again be made for all these diagonal blocks, which can again be done in parallel. For the non- V_Σ blocks, a sequential LAPACK solver can be used, and for $S_{\Sigma\Sigma}$ we can employ a (distributed) sparse direct solver. We denote the factorization that is computed by these solvers as $L_S U_S$.

The full factorization obtained by the method is given by

$$A \approx \begin{pmatrix} L_{II} & 0 \\ A_{SI} U_{II}^{-1} & H^T L_S \end{pmatrix} \begin{pmatrix} U_{II} & L_{II}^{-1} A_{IS} \\ 0 & U_S H \end{pmatrix}.$$

2.3 | Solution phase

After the preconditioner has been computed, it can be applied in each step of a preconditioned Krylov subspace method, for which we use the well-known GMRES method.²⁹ Communication has to occur in the application of A_{IS} and A_{SI} , and when solving the distributed V_{Σ} system. The latter was addressed by using a parallel sparse direct solver in Thies and Wubs,²⁵ but in the next section we propose a different road that does not rely on the availability of such a solver.

3 | THE MULTILEVEL ILU PRECONDITIONER

The main issue with the above two-level ILU factorization that prevents us from scaling up to very large problem sizes in three space dimensions is that, as the problem size increases and the subdomain size remains constant, the size of $S_{\Sigma\Sigma}$ increases steadily, and any (serial or parallel) sparse direct solver eventually limits the feasible problem sizes. Increasing the subdomain size, on the other hand, leads to more iterations and therefore more synchronization points.

One of the strong points, on the other hand, is the fact that it preserves the structure of the original problem in the sense that, when applied to an \mathcal{F} -matrix, it produces a strongly reduced matrix $S_{\Sigma\Sigma}$ which is again an \mathcal{F} -matrix. It is therefore intuitive to try to apply the scheme recursively and avoid the problem of the growing sparse matrix that has to be factorized. From the structure preserving properties of the algorithm, it is expected that such a recursive scheme again leads to grid-independent convergence if the number of recursions is kept constant as the grid size is increased.

From now on we refer to the number of recursions, or the number of times a reduced matrix $S_{\Sigma\Sigma}$ is computed, as the number of levels. Note that this means that the method, which was previously referred to the two-level method is in fact the first level of the multilevel method. Applying a direct solver to S_T from (6) is level zero. In this case the preconditioner is a direct solver and GMRES will converge in 1 iteration.

In order to apply the method to the reduced matrix $S_{\Sigma\Sigma}$, we require a coarser partitioning, in which a subdomain consists of multiple subdomains from the original partitioning. In case we have a regular partitioning like a rectangular partitioning, this may be done by multiplying the *separator length* by a certain *coarsening factor*. Having a coarsening factor of 2, for instance, means that in 3D the separator length is increased by a factor 2, and the number of subdomains is reduced by a factor 8.

As stated in the previous section, we require the velocity variables to be correctly scaled to be able to apply the Householder transformation. However, the V_{Σ} -variables from the previous level that lie on one separator have a different number of variables in their separator groups. In case of a regular partitioning, an edge separator, for instance, consists of V_{Σ} -nodes from two edges and one corner from the previous level. This leads to a different scaling of the V_{Σ} -nodes and thus to nonconstant entries in the G -part of $S_{\Sigma\Sigma}$. Instead of re-scaling the $S_{\Sigma\Sigma}$ matrix on every level, we instead use a test vector \mathbf{t} . The test vector is defined initially as a constant vector of ones, and is multiplied by the Householder transformation H at each consecutive level. The Householder transformation is as defined in Equations (4) and (5), but now with

$$\mathbf{e}_j^{(k)} = \begin{cases} \mathbf{t}^{(k)} & \text{if node } k \text{ is in separator group } g_j \\ 0 & \text{if node } k \text{ is not in separator group } g_j \end{cases},$$

for all $k = 1, \dots, n$. After applying the Householder transformation, we can again apply dropping to remove connections between V_{Σ} and non- V_{Σ} nodes and between non- V_{Σ} nodes and non- V_{Σ} nodes in different separator groups. When the matrix $S_{\Sigma\Sigma}$ is sufficiently small, a direct solver is applied to factorize it.

Instead of just having one separator group per variable per separator, we may also choose to have multiple separator groups, meaning that instead of retaining only one V_{Σ} node per variable per separator we retain multiple V_{Σ} nodes. This in turn means that less dropping is applied, and therefore the factorization is more accurate. Retaining all nodes in this way, possibly only after reaching a certain level, gives us an exact factorization, which, in terms of iterations for the outer iterative solver, should give the same results as using any other direct solver at that level. A visual representation of this process is given in Figure 3.

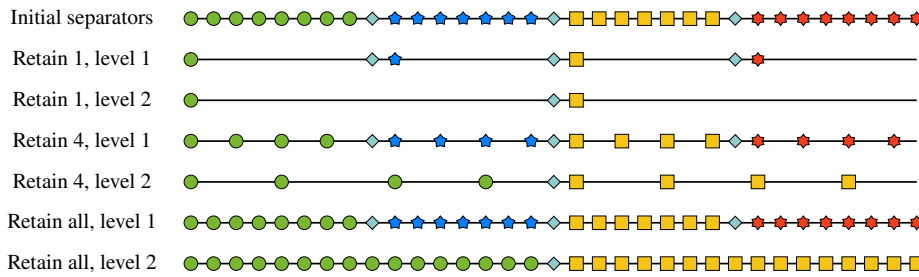


FIGURE 3 One-dimensional representation of the process of retaining multiple nodes per separator. Each separator group has its own color and shape [Colour figure can be viewed at wileyonlinelibrary.com]

4 | SKEW PARTITIONING IN 2D AND 3D

Looking at Figure 2, we see that there are pressures that are located at the corners of the subdomains that are surrounded by velocity separators. This means that if we add these pressures to the interior, as was suggested above, we get a singular interior block. We call these pressure nodes that are surrounded by velocity separators *isolated pressure nodes*. For the two-level preconditioner in 2D, it was possible to solve this problem by adding these pressures to the Schur complement as single-element separator groups. This unfortunately does not work for the multilevel method, since in this case velocity nodes around the isolated pressure nodes get eliminated. It also does not work in 3D because there the isolated pressure nodes also exist inside of the edge separators, where they form tubes of isolated nodes.

A possible way to solve this for the multilevel case and in 3D is to also turn all velocity nodes around the isolated pressure nodes into separate separator groups. This means that they are all treated as V_{Σ} nodes and are never eliminated until they are in the interior of the domain at a later level. This, however, has the downside that a lot more nodes have to be retained at every level, resulting in much larger $S_{\Sigma\Sigma}$ matrices at every level. Furthermore, a lot of bookkeeping is required to keep track of all the extra separator groups, and their elimination leads to long-range connections and thus a more irregular matrix structure.

In 2D, we can resolve these problems by rotating the Cartesian partitioning by 45 degrees. The result is shown in Figure 4A. It is easy to see that in this case, no pressure nodes are surrounded by only velocity separators. We call this partitioning method *skew partitioning*. In Figure 4, we also show the workings of the multilevel method, with all the steps being displayed in the different subfigures.

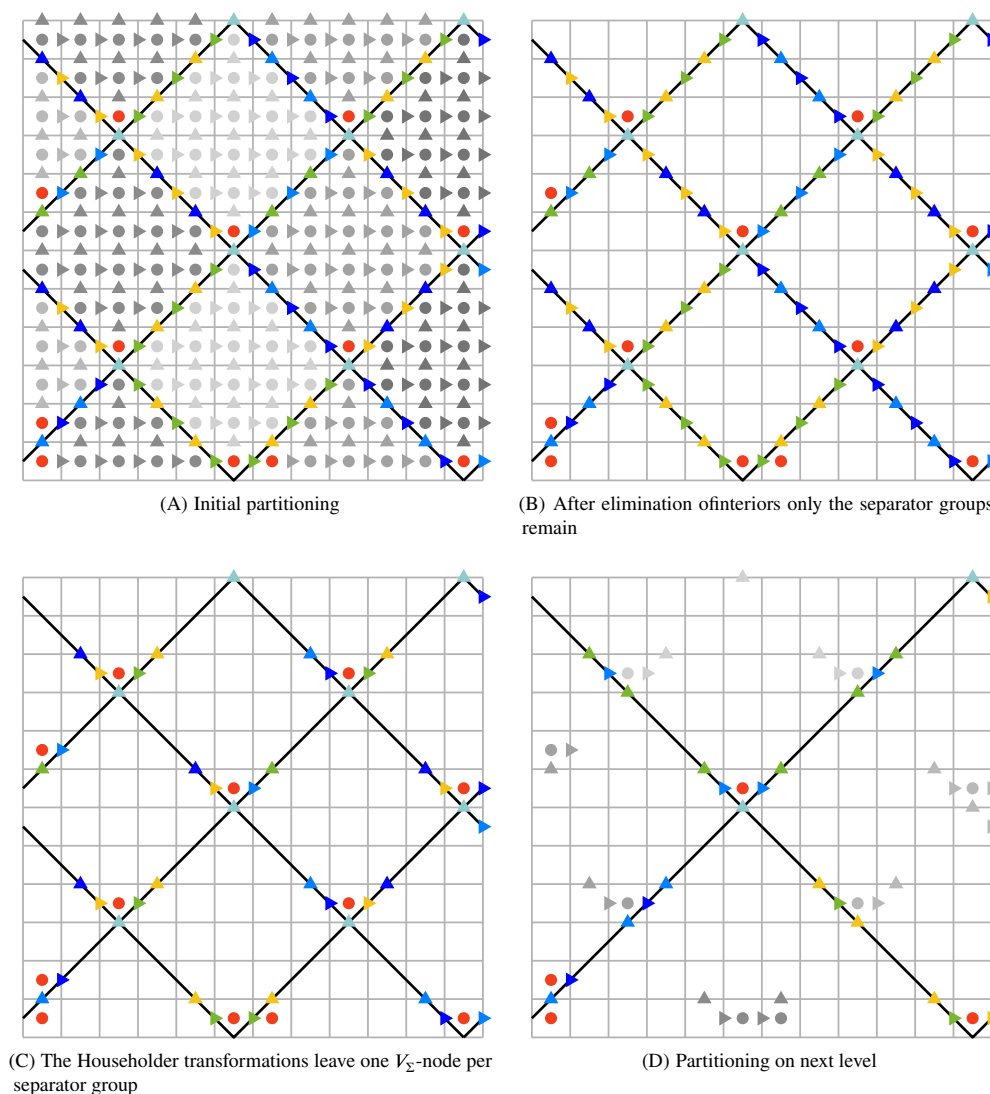
For the skew partitioning to work with our multilevel method, we have two requirements on the shape of the subdomains: (1) it should be *space-filling*, meaning that it can be used to fill the entire domain and (2) it should be *self-similar*, meaning that we can construct a larger subdomain of the same shape from multiple smaller subdomains. It is easy to see that these two properties hold for the 2D skew partitioning.

The most basic idea for generalizing the rotated square shape to a 3D setting is to use octahedral subdomains. Partitioning with this design turned out to be unsuccessful, but it is still briefly discussed here since it led to some new insights. Since regular octahedra (the dual to cubes, having their vertices at the centers of the cube faces) by themselves are not space-filling, the octahedra can be slightly distorted to make them fit within a single cubic repeat unit. The resulting subdomains are space-filling, but only by using three mutually orthogonal subdomain types. The fact that it requires the use of three types of subdomains increases the programming efforts significantly since it introduces a lot of edge cases that should be considered, for example, for subdomains located at the boundary of the domain.

The major problem with the octahedral subdomains, however, is that they are not self-similar, meaning that we cannot construct a larger octahedral subdomain from multiple smaller octahedral subdomains. However, self-similarity can be achieved by splitting the octahedra into four smaller tetrahedra, of which six different types are required to fill 3D space. This introduces additional separator planes that are similar to the 2D skew case and hence it increases the risk of isolating a pressure node when such planes intersect. Especially planar intersections which are parallel to any of the Cartesian axes have a high risk of producing isolated pressure nodes.

We did indeed not manage to find any self-similar tiling using tetrahedra that would not give rise to any isolated pressure nodes. Moreover, we would like to have a single subdomain shape that we can use instead of six, since this would greatly simplify the implementation. A lesson we learned is that isolated pressure nodes always seem to arise when having faces that are aligned with the grid. Therefore, we looked into a rotated parallelepiped shape that does not have

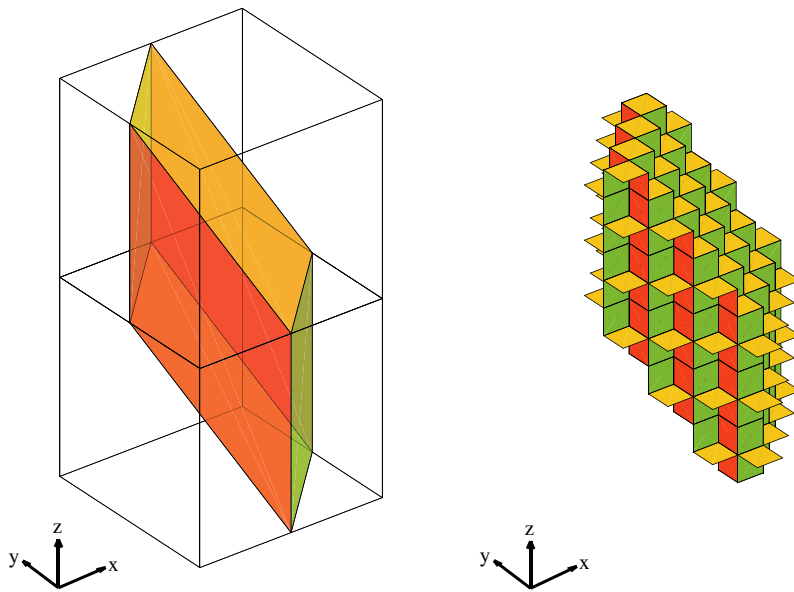
FIGURE 4 Skew partitioning of a 12×12 C-grid discretization of the Navier-Stokes equations into 12 subdomains. The interiors are shown in gray. Velocities of the same (nongray) color together form a separator group (but only if they point in the same direction and are in neighboring grid cells). The red circles are pressure nodes that are retained. This example uses a coarsening factor of 2, that is, the separators on the next level have twice the length of those on the previous level. (A) Initial partitioning. (B) After elimination of interiors only the separator groups remain. (C) The householder transformations leave one V_{Σ} -node per separator group. (D) Partitioning on next level [Colour figure can be viewed at wileyonlinelibrary.com]



any faces that are aligned with the grid.³⁰ This shape is shown in Figure 5A, where the cubes represent a set of $s_x \times s_x \times s_x$ grid cells. A welcome property of this domain is that its central cross section resembles the proposed 2D skew partitioning method.

A schematic view of the position of the separator nodes is shown in Figure 5B. One important detail to note is that on the side that is facing toward us, only half of the w -nodes are displayed. This is because the w -nodes have to be positioned in an alternating fashion on the inside and outside of the nonoverlapping subdomain to prevent isolated pressure nodes from appearing. A consequence of this alternating property is that the w -planes have to be divided into two separate separator groups; one for the w -nodes that are located inside the nonoverlapping subdomain, and one for w -nodes that are only present in the overlapping subdomain.

Another advantage of using a skew domain partitioning is that the amount of communication that is required is reduced when compared to a square partitioning. In Bisseling,³¹ it is estimated that for the Laplace problem, the communication is asymptotically reduced by a factor of $\sqrt{2}$ for the 2D diamond shape. If we instead compare the diamond shape to a rectangular domain with the same number of nodes (having the same number of nodes with a square domain is impossible), we find that communication is reduced by a factor of $\frac{3}{2}$. In the same manner, we can compare a 3D domain of size $s_x \times s_x \times s_x/2$ to the rotated parallelepiped, and find a factor of $\frac{4}{3}$. We remark that the truncated octahedron that is used in Bisseling³¹ for the 3D domain and has a factor of 1.68 cannot be used for our multilevel method, since truncated octahedra are not self-similar.



(A) Shape of the parallelepiped inside of two cubes that consist of $m \times m \times m$ grid cells

(B) Schematic view of the position of the separator nodes from the same point of view as the figure on the left. The u -nodes are depicted as red faces, the v -nodes are depicted in green and the w -nodes in yellow.

FIGURE 5 Parallelepiped shape for partitioning the domain. (A) Shape of the parallelepiped inside of two cubes that consist of $m \times m \times m$ grid cells. (B) Schematic view of the position of the separator nodes from the same point of view as the figure on the left. The u -nodes are depicted as red faces, the v -nodes are depicted in green, and the w -nodes in yellow [Colour figure can be viewed at wileyonlinelibrary.com]

5 | NUMERICAL RESULTS

A common benchmark in fluid dynamics is the lid-driven cavity problem. We consider an incompressible Newtonian fluid in a square three-dimensional domain of unit length, with a lid at the top which moves at a constant speed U . The equations are given by Equation (1). No-slip boundary conditions are applied at the walls, meaning that they are Dirichlet boundary conditions, and the equations are discretized on an Arakawa C-grid as described before. We take $n_x = n_y = n_z$ grid cells in every direction.

At first, however, we will only look at the Stokes equations of the form

$$\begin{aligned}\Delta \mathbf{u} - \nabla p &= \mathbf{f}, \\ \nabla \cdot \mathbf{u} &= 0,\end{aligned}$$

where we take \mathbf{f} to be random. This is because our preconditioner is constructed in such a way that memory usage and time cost for both computation and application of the preconditioner should not be influenced by inclusion of the convective term. After this, we further investigate the behavior of the method on the lid-driven cavity problem for increasing Reynolds numbers, which constitute harder problems. Therefore we expect an increase in iterations of the iterative solver, but otherwise the same behavior.

For obtaining the exact memory usage, we developed a custom library which overrides all memory allocation routines when linking against it. The library contains a hash table in which the amount of memory that is allocated is stored by its memory address. We keep track of the total amount of memory that is allocated, which is increased on memory allocation, and reduced by the amount that is stored in the hash table when memory is freed. The reason we developed this library is that existing methods rely on rough estimates of the memory usage of the used data structures, use the data that is available from `/proc/meminfo`, which is inaccurate, or actually count memory usage in a similar way as we do (e.g., `valgrind`), but have a large performance impact.

We perform every experiment twice: once to determine the memory usage, and once to determine the run time, without linking to the memory usage library. This means that when reporting timing results, we are not affected by the performance impact of tools to determine memory usage. The reason that we are still concerned about their performance impact, even though we developed our own library, is that it adds roughly a constant amount of time per process, which

impacts scalability results. The memory usage that we report is the exact difference in memory usage before and after a certain action is performed, for example, before and after the construction of the preconditioner.

For the timing results, we do not include the time it takes to compute the partitioning, because we did not optimize this step. The partitioning is computed by first constructing one full overlapping subdomain of the prescribed subdomain size sequentially, which is then mapped to the correct position of every overlapping subdomain to determine the interiors and separator groups. However, since the initial full overlapping subdomain contains all (possibly already eliminated) nodes, every time we increase the number of levels by 1, the amount of time required to compute the partitioning increases by a factor 8, which is the worst possible scenario. The reason we do not include this in the timing results is that this may be resolved by only computing the partitioning for nodes that are still present in the Schur complement at that level. This requires a full rewrite of the existing partitioning code, and will be addressed in a future version of the software. It does, however, not have any impact on the timing results of the remainder of the implementation, since this is completely decoupled. This means that even though the partitioning method does not scale at all, the preconditioning method itself can be studied reliably without including the timing of the partitioning.

For the implementation of the preconditioner and solver, we use libraries from the Trilinos project.³² The libraries we use are Epetra (vector and matrix data structures), IFPACK (direct solver and preconditioning interfaces),³³ Amesos (direct solvers)³⁴ and Belos (iterative solvers).³⁵ As iterative solver we use GMRES(250),²⁹ as parallel sparse direct solver on the coarsest level we use SuperLU_DIST 6.1,³⁶ and as direct solver for the interior blocks we use KLU³⁷ with the fill-reducing ordering from Niet and Wubs.²⁷ The implementation of our preconditioner can be found on GitHub.*

The benchmark is performed on the SuperMUC-NG cluster at LRZ, Munich.[†] The nodes contain two Intel Xeon “SkyLake” Platinum 8174 processors with 24 cores and have 96 GB of memory. For all experiments, we disable multithreading inside the MPI processes, since the use of OpenMP in Epetra is not compatible with our need for very small objects (e.g., sparse matrices living on a very small subdomain).

5.1 | Weak scalability

First, we look at results obtained when increasing the grid size n_x at the same rate as the number of used cores n_p , that is, the problem size on each core is kept constant. The exact configurations that we use are 1 core for $n_x = 16$, 1 core for $n_x = 32$, 8 cores for $n_x = 64$, 64 cores for $n_x = 128$, 512 cores for $n_x = 256$, and 4096 cores for $n_x = 512$. The size of the subdomains (the size of the cubes in Figure 5A) at the first level is $s_x = 8$, while we choose the coarsening factor to be 2, meaning that we increase s_x by a factor of 2 at each level. We perform experiments where we keep the number of levels constant at $L = 2$, and experiments where we increase the number of levels by 1 when doubling the grid size. For the latter, we look at three cases where we retain a different number of separator nodes starting at level 2: 1, 4, and all. Since we start retaining more nodes after two levels, results with only two levels ($n_x = 16$) will be the same for all configurations. When no result is shown for $n_x = 512$, this means the method ran out of memory, unless otherwise stated.

For the fixed number of levels, we expect the number of iterations of the iterative solver to converge to a constant number as the grid is refined. For the case where we increase the number of levels as the domain size increases, we expect the number of iterations to only increase mildly, and we expect that retaining more separator nodes starting at level 2 decreases the number of iterations until it again becomes constant as we retain more and more nodes per separator.

The results are shown in Figure 6. We see that indeed the number of iterations becomes constant when fixing the number of levels or when retaining all separator nodes. When increasing the number of levels with the grid size, we see that the number of iterations keeps increasing gradually. What is interesting is that when we retain 4 instead of 1 separator node per separator group after level 2, the number of iterations that is required decreases drastically, even though this does not have a significant impact on the memory usage as we will see later.

The computational time of both computing the preconditioner, as well as the application of the preconditioner would ideally become constant when keeping the problem size at each core constant while increasing the problem size. However,

*<https://github.com/nlesc-smcm/hymls>

†<https://www.top500.org/system/179566/>

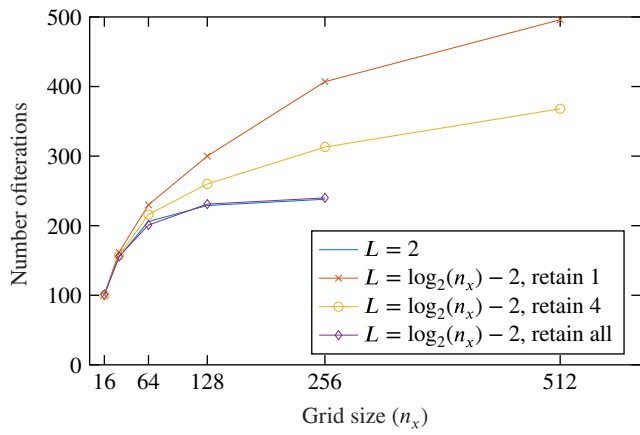


FIGURE 6 Number of iterations of GMRES for the Stokes problem on a grid of size $n_x \times n_x \times n_x$, while keeping the number of levels at $L = 2$, and when increasing the number of levels by 1 when n_x is doubled. A relative residual of 10^{-8} was used as convergence tolerance [Colour figure can be viewed at wileyonlinelibrary.com]

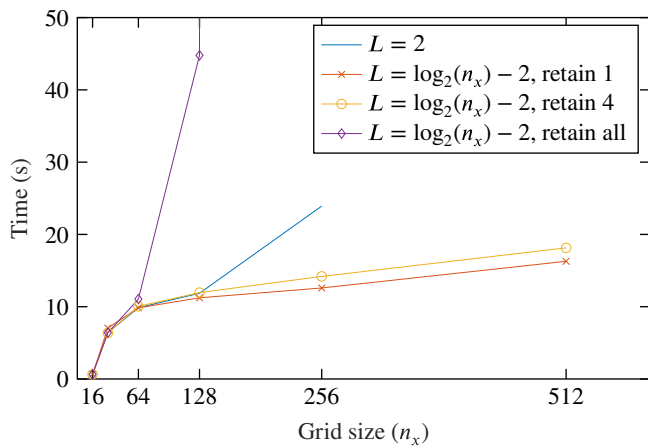


FIGURE 7 Time to compute the preconditioner for the Stokes problem on a grid of size $n_x \times n_x \times n_x$, while keeping the number of levels at $L = 2$, and when increasing the number of levels by 1 when n_x is doubled [Colour figure can be viewed at wileyonlinelibrary.com]

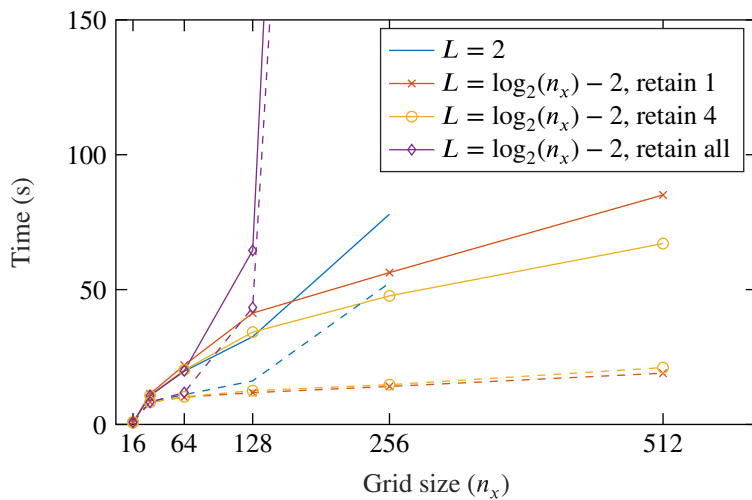
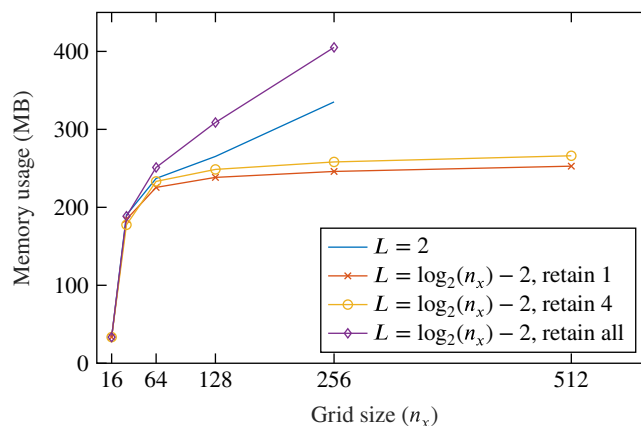


FIGURE 8 Time to solve the linear system with GMRES (lines), and time of 200 applications of the preconditioner (dashed lines). This is for the Stokes problem on a grid of size $n_x \times n_x \times n_x$, while keeping the number of levels at $L = 2$, and when increasing the number of levels by 1 when n_x is doubled. A relative residual of 10^{-8} was used as convergence tolerance [Colour figure can be viewed at wileyonlinelibrary.com]

in practice this is not possible since increasing the number of cores also increases the required amount of communication. The results are shown in Figures 7 and 8.

When computing the preconditioner, we see that if we increase the number of levels with the grid size, the computational time rises only slightly. Ideally, there would be no rise at all, but since an increased the number of computational nodes also means more communication, there will always be an increase in practice. In our case, the communication mainly happens at the point where contributions of neighboring subdomains have to be added up in the Schur complement. Since retaining more nodes per level means an increase in the amount of communication, we also see that retaining

FIGURE 9 Memory usage of the preconditioner per core for the Stokes problem on a grid of size $n_x \times n_x \times n_x$, while keeping the number of levels at $L=2$, and when increasing the number of levels by 1 when n_x is doubled [Colour figure can be viewed at wileyonlinelibrary.com]



4 or all nodes takes more time than retaining only 1 node per separator at every level. It must be noted that due to system variability, we also found results that were about 10% better or worse than the results displayed here. This could be improved by disabling frequency adaptation in the CPUs, but we are aware that at very large core counts the synchronous nature of our algorithm may become a scalability problem.

If we keep the number of levels constant, the computational time required to compute the factorization at the last level increases rapidly, and for larger grid sizes, the amount of memory that is required for the factorization is too large for the computation to complete. We also notice that retaining all nodes after level 2 is much less efficient than just using SuperLU_DIST at level 2, meaning that our preconditioner performs very poorly as a direct solver. This is mainly due to the fact that the Schur complement at the last level is quite large and full. The last level Schur complement for a grid of size 256^3 has size $20,961 \times 20,961$, and its factorization is filled with 72% nonzeros. Computing the factorization of this matrix and using it in a forward/backward substitution is therefore very expensive. Using a parallel dense solver instead of SuperLU_DIST might help to make it more efficient. Moreover, since we choose $n \sim n_p$, the cost of computing the factorization by a sparse direct solver grows at best with $\mathcal{O}(n^2/n_p) = \mathcal{O}(n) = \mathcal{O}(n_x^3)$,³⁶ and we expect that the cost of computing the preconditioner when keeping the number of levels constant, or when retaining all separator nodes, increases with n_x^3 .

In Figure 8, we show both the time required to solve the linear system after computation of the preconditioner, and the time of 200 applications of the preconditioner, which is not influenced by the number of iterations of the iterative solver and does not include the time spent on, for instance, matrix-vector products and orthogonalization. First of all, we again observe that retaining all separator nodes is a bad idea, since the computational time goes off the chart. For the case where we only use 2 levels, we also see the unwanted behavior of a time that keeps increasing linearly for the total solution time, and superlinearly for the application of the preconditioner, where we would actually expect $\mathcal{O}(n^{4/3}/n_p) = \mathcal{O}(n_x)$ behavior from the triangular solve.³⁶ This may be caused by disabling multithreading support, which results in an increased amount of communication inside of SuperLU_DIST.

For both cases where we retain 1 and 4 separator nodes after 2 levels, we see that the computational time only slightly increases for larger grid sizes. We also note that for the case where we retain 4 nodes, the application of the preconditioner is slightly slower, but the total solution time is much smaller due to the lower number of iterations that is required, as can also be seen in Figure 6, and the fact that applying the preconditioner is relatively cheap.

In Figure 9, we see the average memory usage of the preconditioner per core. Here we again observe that the 2 level case, and the case where we retain all separator nodes after level 3, perform poorly. The cases where we retain 1 or 4 separator nodes per separator group after level 2 show similar behavior in terms of memory usage, and the memory usage becomes constant as expected.

5.2 | Strong scalability

In this section, we look at a problem of size $n_x = 128$, with 6 levels and retaining only one node per separator group. We use 1 to 128 cores with steps of a factor of 2 for all cases except the application time, where we use 2 to 128 cores. The

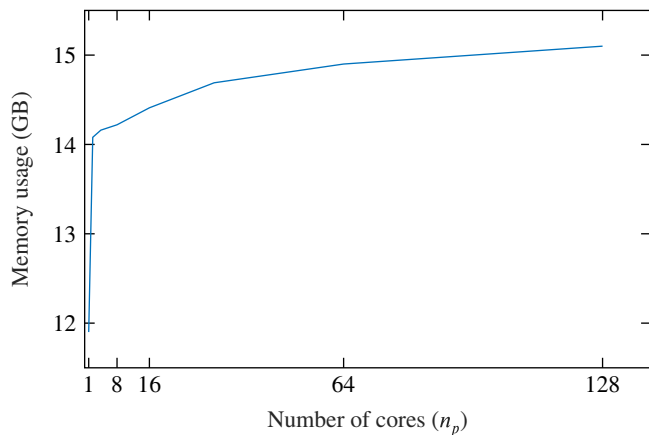


FIGURE 10 Total memory usage of the preconditioner for the Stokes problem on a grid of size 128^3 , with 1 to 128 cores [Colour figure can be viewed at wileyonlinelibrary.com]

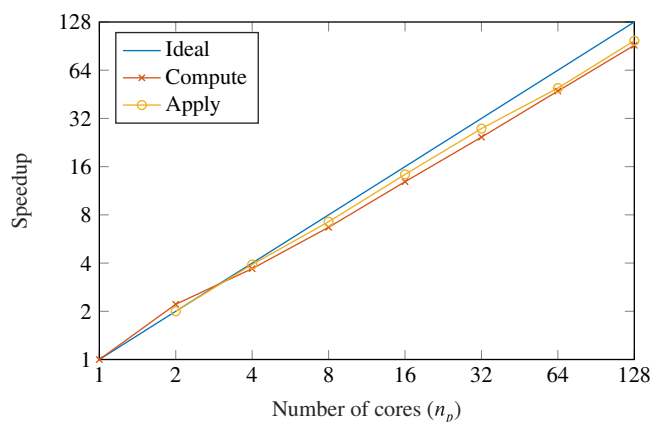


FIGURE 11 Speedup of computation and application of the preconditioner for the Stokes problem on a grid of size 128^3 , with 2 to 128 cores [Colour figure can be viewed at wileyonlinelibrary.com]

reason we do not look at 1 core for this timing is that this configuration caused a memory allocation error in the iterative solver (Belos).

We first look at the total memory usage in Figure 10. We observe that there is a large difference between the memory usage on one core, and the memory usage on two cores. This is because especially Epetra uses different implementations of its data structures for serial and parallel computations.

We also observe that the total memory usage increases slightly when using more cores. This can be explained by the overlap that exists between different processes. Since the difference in communication between a cubical domain and a parallelepipedal domain is a constant factor of $\frac{3}{4}$, we may instead look at a cubical domain to explain this behavior. If we have a subdomain of size $s_x \times s_y \times s_z$, then communication is required for $2s_x s_y + 2s_x s_z + 2s_y s_z$ grid cells. If the subdomain is split in the x -direction, communication for $2s_x s_y + 2s_x s_z + 4s_y s_z$ grid cells is required. For the case $s_x = s_y = s_z$, communication increases with a factor $\frac{4}{3}$ when doubling the number of cores. This explains the slight increase in memory usage we measure.

In Figure 11, we look at the speedup when using more cores. Ideally, we would see that using twice the number of cores would mean half of the computational time. We plotted this ideal line for reference. We see that the speedup of both the computation and application of the preconditioner is very close to this ideal line. There is one outlier at 2 cores which is above the ideal line, which may be explained by system variability. Other than that the behavior is as expected.

5.3 | Lid-driven cavity

In the previous sections, we determined the weak and strong scalability properties of the preconditioner, which means that we can now continue with the robustness of the solver on the lid-driven cavity problem with increasing Reynolds numbers. We perform a continuation with steps of $Re = 100$ starting from the solution of the Stokes problem. We show the results of the first iteration of Newton at $Re = 500$ and $Re = 2000$. The reason we go up to $Re = 2000$ is that a Hopf

FIGURE 12 Number of iterations of GMRES for the lid-driven cavity problem at $Re = 500$ on a grid of size $n_x \times n_x \times n_x$, while keeping the number of levels at $L = 2$, and when increasing the number of levels by 1 when n_x is doubled. A relative residual of 10^{-8} was used as convergence tolerance. As initial guess we used the solution at $Re = 400$ [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

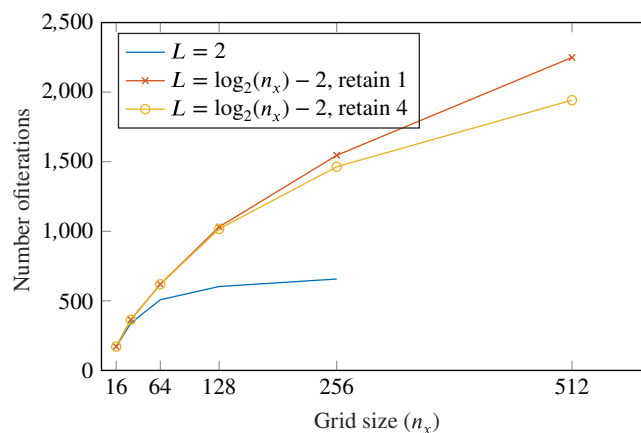
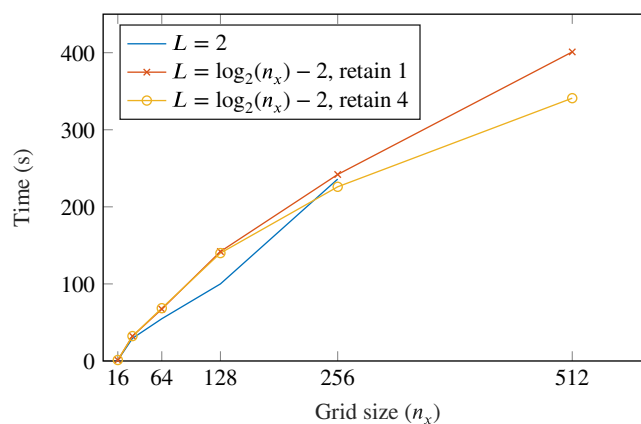


FIGURE 13 Time for GMRES to converge for the lid-driven cavity problem at $Re = 500$ on a grid of size $n_x \times n_x \times n_x$, while keeping the number of levels at $L = 2$, and when increasing the number of levels by 1 when n_x is doubled. A relative residual of 10^{-8} was used as convergence tolerance. As initial guess we used the solution at $Re = 400$ [Colour figure can be viewed at [wileyonlinelibrary.com](#)]



bifurcation is located between $Re = 1900$ and $Re = 2000$,^{20,38} which is of interest because it changes the qualitative behavior of the solution and makes the Jacobian indefinite.

In Figures 12 and 13, we show the results at Reynolds number 500, which show good correspondence with the results on the Stokes problem. The main difference is that more iterations are needed, since higher Reynolds numbers constitute harder problems. We see that the number of iterations that is required when keeping the number of levels constant actually decreases, which is due to the grid refinement having a positive effect on the mesh Péclet number, that is, the coefficients of the diffusion increase with respect to those of the convection.

The results at Reynolds number 2000 are shown in Figures 14 and 15. We again observe that the number of iterations is much larger. What is odd, however, is that retaining more nodes now actually gives worse convergence. This may be because we use GMRES(250) instead of GMRES due to memory limitations, and therefore do not preserve the convergence properties of GMRES. This effect is more prevalent for this problem because of the large number of iterations that is required. We did confirm that for the first 250 iterations, retaining 4 nodes after level 2 gives rise to better convergence, as we expected. For Reynolds number 2000, we do not show results with a grid of size 512^3 , because the method does not converge within 10,000 GMRES iterations, which we set as the maximum. It does converge, however, and extrapolating the results, we expect that around 15,000 iterations are required to meet the tolerance. Alternatively, instead of increasing the number of levels with the grid size, we could also have used 6 levels as we did for a grid of size 256^3 , in which case we would have expected it to actually use fewer iterations due to the effect of grid refinement on the mesh Péclet number as explained earlier.

In Table 1, we show results for Reynolds number 500 using the LSC-SIMPLEC block preconditioner implemented in Teko.¹⁵ We also tried preconditioners from other packages, as well as other preconditioning strategies implemented in the Teko package, but those unfortunately converged even slower or did not yield convergence at all. The stopping criterion of the linear solver is 10^{-8} , as before.

Compared to our method, we see that Teko has much more difficulty with the grid refinement, leading to a huge computational cost at a grid size of only 64^3 . A crude computation shows that per grid point the method becomes about 65

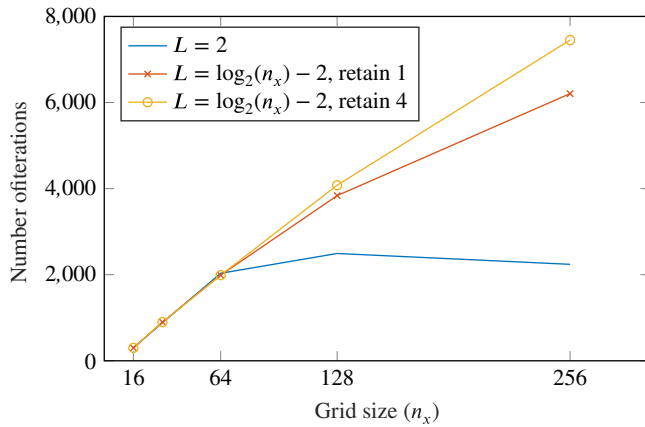


FIGURE 14 Number of iterations of GMRES for the lid-driven cavity problem at $Re = 2000$ on a grid of size $n_x \times n_x \times n_x$, while keeping the number of levels at $L = 2$, and when increasing the number of levels by 1 when n_x is doubled. A relative residual of 10^{-8} was used as convergence tolerance. As initial guess we used the solution at $Re = 1900$ [Colour figure can be viewed at wileyonlinelibrary.com]

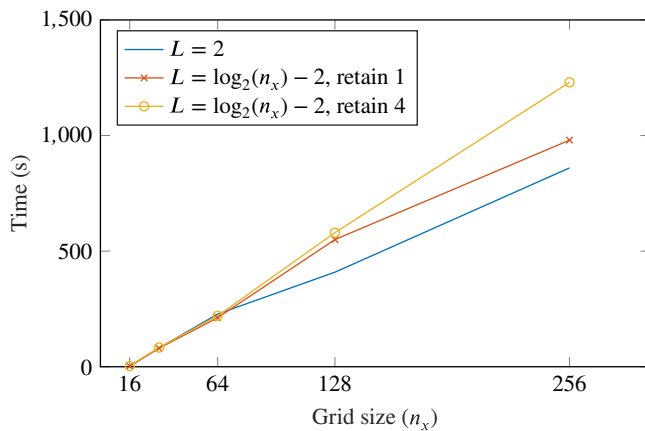


FIGURE 15 Time for GMRES to converge for the lid-driven cavity problem at $Re = 2000$ on a grid of size $n_x \times n_x \times n_x$, while keeping the number of levels at $L = 2$, and when increasing the number of levels by 1 when n_x is doubled. A relative residual of 10^{-8} was used as convergence tolerance. As initial guess we used the solution at $Re = 1900$ [Colour figure can be viewed at wileyonlinelibrary.com]

n_p	n_x	L	its	t_c (s)	t_s (s)
1	16	2	142	0.12	0.77
1	32	2	187	9.88	18.80
8	64	3	245	1511.12	313.00

Note: Here n_p is the number of cores, L is the number of levels, its is the number of iterations, t_c is the time to compute the preconditioner and t_s is the time for solving the linear system.

TABLE 1 Performance of Teko with LSC-SIMPLEC preconditioner for the lid-driven cavity problem at $Re = 500$ on a grid of size $n_x \times n_x \times n_x$

times more expensive per iteration. This must be attributed to slow convergence of algebraic multigrid on the subblocks. One of these blocks is the $L + N$ block from Equation (2), with N indefinite, and this is something that is difficult for a standard AMG method. We choose the number of levels to be 2 for grid sizes 16^3 and 32^3 , and 3 levels for 64^3 since these seem to give the optimal results. For Reynolds number 2000, we do not observe convergence past the 16^3 grid.

6 | SUMMARY AND DISCUSSION

We presented a robust method for solving the steady, incompressible Navier-Stokes equations, which makes use of parallel piped shaped overlapping subdomains. The interiors of these overlapping subdomains can be eliminated in parallel. On the interfaces of the subdomains, Householder transformations are applied to decouple all but one velocity node from the pressure nodes, after which all decoupled nodes can also be eliminated in parallel. The key to the multilevel approach is the resulting reduced Schur complement matrix, which has the same structure as the original matrix. We can therefore recursively apply our method to this matrix. The shape of the subdomains makes sure that pressure nodes are not isolated in the factorization process, which would result in a singular Schur complement matrix.

Our weak scalability experiments show that if the number of levels of the multilevel approach is kept constant while increasing the problem size, grid independent convergence is obtained. We also show that by increasing the number of levels and processors as the problem size increases, we only require a small amount of additional time and memory for both the factorization and solution process. Moreover, by increasing the number of nodes that is retained per separator after level 2, we can further decrease the time that is required to solve the system.

Our strong scalability experiments show that the time that is required to both compute and apply the preconditioner scales very well. The same holds for the memory usage, which behaves as expected.

We also showed that the preconditioner leads to convergence of GMRES for the lid-driven cavity problem at high Reynolds numbers, where other methods, such as the LSC-SIMPLEC preconditioner that is implemented in Teko, fail to do so.

This leads us to conclude that we implemented a robust solution method for the Navier-Stokes equations on staggered grids which shows good parallel scalability. In this article, we only showed results for Arakawa C-grids. We have already succeeded to apply the idea to discretization on other staggered grids, for example, Arakawa B-grids and mixtures of B- and C-grids as used in oceanography. This will allow us to use it as preconditioner in ocean-climate models, which is our application goal. The approach is, however, not limited to structured grids. The essence of the method is that we can iterate in the divergence-free space in a cheap way. As far as we can see this is the case when we have discrete conservation of mass. The C-grid discretization leads to one mass conservation law per subdomain (the B-grid has two). By uniting two neighbouring subdomains, we can find, by a simple combination of the two laws, a single new one for the united subdomains. This is also possible for finite volume discretizations on unstructured grids. For this reason, we think that the method can be applied to such discretizations in general and to some (discontinuous) Galerkin methods that also have the discrete conservation property. In the future, we may generalize the approach to unstructured grid discretizations for which we will need graph based partitioning methods.

ACKNOWLEDGMENTS

This work is part of the Mathematics of Planet Earth research program with project number 657.014.007, which is financed by the Netherlands Organization for Scientific Research (NWO) and the SMCM project of the Netherlands eScience Center (NLeSC) with project number 027.017.G02 (SB and FW). Access to the SuperMUC-NG system during the “friendly user phase” was kindly granted by LRZ Munich via a test account. We would like to thank Eric Cyr for his help during our attempts of getting the solvers in Teko to work as expected.

ORCID

Sven Baars  <https://orcid.org/0000-0002-4228-6479>

REFERENCES

- Verstappen R, Veldman A. Symmetry-preserving discretization of turbulent flow. *J Comput Phys*. 2003;187(1):343-368. [https://doi.org/10.1016/s0021-9991\(03\)00126-8](https://doi.org/10.1016/s0021-9991(03)00126-8).
- Elman H, Silvester D, Wathen A. *Finite Elements and Fast Iterative Solvers*. Oxford, UK: Oxford University Press; 2014.
- Brandenburg C, Lindemann F, Ulbrich M, Ulbrich SA. Continuous adjoint approach to shape optimization for Navier Stokes flow. *Optimal Control of Coupled Systems of Partial Differential Equations 158 of International Series of Numerical Mathematics*. Basel: Birkhäuser; 2009:35-56.
- Charru F, Forcrand-Millard DP. *Hydrodynamic Instabilities*. Cambridge, MA: Cambridge University Press; 2011.
- Keller HB. Numerical solution of bifurcation and nonlinear eigenvalue problems. In: Rabinowitz PH, ed. *Applications of Bifurcation Theory*. New York, NY: Academic Press; 1977:359-384.
- Sleijpen GLG, Wubs FW. Exploiting multilevel preconditioning techniques in eigenvalue computations. *SIAM J Sci Comput*. 2004;25(4):1249-1272. <https://doi.org/10.1137/s1064827599361059>.
- Arakawa A, Lamb VR. Computational design of the basic dynamical processes of the UCLA general circulation model. Chang J, *Methods in Computational Physics: Advances in Research and Applications*. Vol 17. New York, NY: Academic Press; 1977:173-265.
- Benzi M, Golub GH, Liesen J. Numerical solution of saddle point problems. *Acta Numerica*. 2005;14:1-137. <https://doi.org/10.1017/s0962492904000212>.
- Carey G, Krishnan R. Convergence of iterative methods in penalty finite element approximation of the Navier-Stokes equations. *Comput Methods Appl Mech Eng*. 1987;60(1):1-29. [https://doi.org/10.1016/0045-7825\(87\)90127-7](https://doi.org/10.1016/0045-7825(87)90127-7).
- Elman HC, Howle VE, Shadid JN, Tuminaro RSA. Parallel block multi-level preconditioner for the 3D incompressible Navier-Stokes equations. *J Comput Phys*. 2003;187(2):504-523. [https://doi.org/10.1016/s0021-9991\(03\)00121-9](https://doi.org/10.1016/s0021-9991(03)00121-9).
- Dijkstra HA, Wubs FW, Cliffe AK, et al. Numerical bifurcation methods and their application to fluid dynamics: analysis beyond simulation. *Commun Comput Phys*. 2014;15(01):1-45. <https://doi.org/10.4208/cicp.240912.180613a>.

12. Verstappen R, Dröge MA. Symmetry-preserving Cartesian grid method for computing a viscous flow past a circular cylinder. *Comptes Rendus Mécanique*. 2005;333(1):51-57. <https://doi.org/10.1016/j.crme.2004.09.021>.
13. De Niet A, Wubs F, Terwisscha van Scheltinga A, Dijkstra HA. A tailored solver for bifurcation analysis of ocean-climate models. *J Comput Phys*. 2007;227(1):654-679. <https://doi.org/10.1016/j.jcp.2007.08.006>.
14. Klaij CM, Vuik C. SIMPLE-type preconditioners for cell-centered, colocated finite volume discretization of incompressible reynolds-averaged Navier–Stokes equations. *Int J Numer Methods Fluids*. 2012;71(7):830-849. <https://doi.org/10.1002/fld.3686>.
15. Cyr EC, Shadid JN, Stabilization TRS. Scalable block preconditioning for the Navier–Stokes equations. *J Comput Phys*. 2012;231(2):345-363. <https://doi.org/10.1016/j.jcp.2011.09.001>.
16. He X, Vuik C, Klaij CM. Combining the augmented Lagrangian preconditioner with the simple schur complement approximation. *SIAM J Sci Comput*. 2018;40(3):A1362-A1385. <https://doi.org/10.1137/17m1144775>.
17. Thies J, Wubs F, Dijkstra HA. Bifurcation analysis of 3D ocean flows using a parallel fully-implicit ocean model. *Ocean Modell*. 2009;30(4):287-297. <https://doi.org/10.1016/j.ocemod.2009.07.005>.
18. Trottenberg U, Oosterlee CW, Schüller A. *Multigrid*. London, UK: Academic Press; 2000.
19. Wathen AJ. Preconditioning. *Acta Numerica*. 2015;24:329-376. <https://doi.org/10.1017/s0962492915000021>.
20. Feldman Y, Gelfgat AY. Oscillatory instability of a three-dimensional lid-driven flow in a cube. *Phys Fluids*. 2010;22(9):093602. <https://doi.org/10.1063/1.3487476>.
21. Botta EFF, Wubs FW. Matrix renumbering ILU: an effective algebraic multilevel ILU preconditioner for sparse matrices. *SIAM J Matrix Anal Appl*. 1999;20(4):1007-1026. <https://doi.org/10.1137/s0895479897319301>.
22. Bollhöfer M, Saad Y. Multilevel pre-conditioners constructed from inverse-based ILUs. *SIAM J Sci Comput*. 2006;27(5):1627-1650. <https://doi.org/10.1137/040608374>.
23. Aliaga JI, Bollhöfer M, Martín AF, Quintana-Ortí ES. *Design, Tuning and Evaluation of Parallel Multilevel ILU Preconditioners*. Berlin Heidelberg/Germany: Springer; 2008:314-327.
24. Wubs FW, Thies J. A robust two-level incomplete factorization for Navier–Stokes saddle point matrices. *SIAM J Matrix Anal Appl*. 2011;32(4):1475-1499. <https://doi.org/10.1137/100789439>.
25. Thies J, Wubs F. Design of a parallel hybrid direct/iterative solver for CFD problems. Paper presented at: Proceedings of the 2011 IEEE 7th International Conference on eScience; Stockholm, Sweden: 2011. <https://doi.org/10.1109/escience.2011.60>
26. Tuma M. A note on the LDLT decomposition of matrices from saddle-point problems. *SIAM J Matrix Anal Appl*. 2002;23(4):903-915. <https://doi.org/10.1137/s0895479897321088>.
27. De Niet AC, Wubs FW. Numerically stable LDLT-factorization of F-type saddle point matrices. *IMA J Numer Anal*. 2008;29(1):208-234. <https://doi.org/10.1093/imanum/drn005>.
28. Karypis G, Kumar V. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM J Sci Comput*. 1998;20(1):359-392. <https://doi.org/10.1137/s1064827595287997>.
29. Saad Y, Schultz MH. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J Sci Stat Comput*. 1986;7(3):856-869. <https://doi.org/10.1137/0907058>.
30. Van der Kloek, M. *Skew Partitioning for the Hybrid Multilevel Solver* [Bachelor's thesis], University of Groningen; 2017.
31. Bisseling RH. *Parallel Scientific Computation*. Oxford, UK: Oxford University Press; 2004.
32. Heroux MA, Phipps ET, Salinger AG, et al. An overview of the Trilinos project. *ACM Trans Math Softw*. 2005;31(3):397-423. <https://doi.org/10.1145/1089014.1089021>.
33. Sala M, Heroux MA. *Robust algebraic preconditioners using IFPACK 3.0. Technical Report*. Albuquerque, NM: Sandia National Laboratories; 2005.
34. Sala M, Stanley KS, Heroux MA. On the design of interfaces to sparse direct solvers. *ACM Trans Math Softw*. 2008;34(2):1-22. <https://doi.org/10.1145/1326548.1326551>.
35. Bavier E, Hoemmen M, Rajamanickam S, Thornquist H. Amesos2 and Belos: direct and iterative solvers for large sparse linear systems. *Sci Program*. 2012;20(3):241-255. <https://doi.org/10.1155/2012/243875>.
36. Liu Y, Jacquelin M, Ghysels P, Li XS. Highly scalable distributed-memory sparse triangular solution algorithms. Paper presented at: Proceedings of the 2018 7th SIAM Workshop on Combinatorial Scientific Computing; Bergen, Norway: 2018:87-96. <https://doi.org/10.1137/1.9781611975215.9>
37. Davis TA, Natarajan EP. Algorithm 907. *ACM Trans Math Softw*. 2010;37(3):1-17. <https://doi.org/10.1145/1824801.1824814>.
38. Liberzon A, Feldman Y, Gelfgat AY. Experimental observation of the steady-oscillatory transition in a cubic lid-driven cavity. *Phys Fluids*. 2011;23(8):084106. <https://doi.org/10.1063/1.3625412>.

How to cite this article: Baars S, van der Kloek M, Thies J, Wubs FW. A staggered-grid multilevel incomplete LU for steady incompressible flows. *Int J Numer Meth Fluids*. 2021;93:909–926. <https://doi.org/10.1002/fld.4913>