

PIPELINED, FLEXIBLE KRYLOV SUBSPACE METHODS*

P. SANAN[†], S.M. SCHNEPP[‡], AND D.A. MAY[‡]

Abstract. We present variants of the conjugate gradient (CG), conjugate residual (CR), and generalized minimal residual (GMRES) methods which are both *pipelined* and *flexible*. These allow computation of inner products and norms to be overlapped with operator and nonlinear or nondeterministic preconditioner application. The methods are hence aimed at hiding network latencies and synchronizations which can become computational bottlenecks in Krylov methods on extreme-scale systems or in the strong-scaling limit. The new variants are not arithmetically equivalent to their base flexible Krylov methods, but are chosen to be similarly performant in a realistic use case, the application of strong nonlinear preconditioners to large problems which require many Krylov iterations. We provide scalable implementations of our methods as contributions to the PETSc package and demonstrate their effectiveness with practical examples derived from models of mantle convection and lithospheric dynamics with heterogeneous viscosity structure. These represent challenging problems where multiscale nonlinear preconditioners are required for the current state-of-the-art algorithms, and are hence amenable to acceleration with our new techniques. Large-scale tests are performed in the strong-scaling regime on a contemporary leadership supercomputer, where speedups approaching, and even exceeding, $2\times$ can be observed. We conclude by analyzing our new methods with a performance model targeted at future exascale machines.

Key words. Krylov methods, FCG, GCR, FGMRES, parallel computing, latency hiding, global communication

AMS subject classifications. 65F08, 65F10, 65Y05, 68W10

DOI. 10.1137/15M1049130

1. Introduction. The current high performance computing (HPC) paradigm involves computation on larger and larger clusters of individual compute nodes connected through a network. Since single-core performance has plateaued, increased parallelism is used to increase total performance. As systems scale in this way, new performance bottlenecks emerge in algorithms popularized before massive parallelism became relevant. While peak compute power is abundant on modern high performance clusters, utilizing them at full capacity requires modifications to many well-established algorithms. We investigate modifications to flexible Krylov subspace methods to better exploit computational resources available from modern clusters. As parallelism increases, collective operations are more likely to become computational bottlenecks. For instance, distributed dot products can become performance limiting due to network latency exposed by synchronized all-to-all communication. Similar concerns apply to hypothetical future machines and to current machines in the strong-scaling limit when small amounts of local work expose bottlenecks in collective operations.

*Submitted to the journal's Software and High Performance Computing section November 23, 2015; accepted for publication (in revised form) June 3, 2016; published electronically September 1, 2016.

<http://www.siam.org/journals/sisc/38-5/M104913.html>

Funding: This work was supported by the Swiss University Conference and the Swiss Council of Federal Institutes of Technology through the Platform for Advanced Scientific Computing (PASC) program.

[†]Advanced Computing Laboratory, Institute of Computational Science, Università della Svizzera italiana (USI), Via Giuseppe Buffi 13, 6904 Lugano, Switzerland, and Institute of Geophysics, ETH Zürich, Sonneggstrasse 5, 8092 Zürich, Switzerland (patrick.sanan@usi.ch, patrick.sanan@erdw.ethz.ch).

[‡]Institute of Geophysics, ETH Zürich, Sonneggstrasse 5, 8092 Zürich, Switzerland (mail@saschaschnepp.net, dave.may@erdw.ethz.ch).

One approach to mitigating these bottlenecks is to overlap these collective operations with local work, hiding the latency. This is a nontrivial task due to data dependencies which forbid arbitrary rearrangements of operations.

1.1. Pipelining flexible Krylov subspace methods. A classical way to use a system with independent hardware resources at higher efficiency is to *pipeline*. Broadly speaking, this technique can accelerate throughput of a repeated multistage process when different stages require different resources. Concurrent work is performed on multiple iterations of the process, overlapped such that multiple resources are used simultaneously. This induces latency as the pipeline is filled as well as other overheads related to rearranging an algorithm and allowing simultaneous progress on multiple tasks.

We consider pipelining a particular class of algorithms, namely Krylov subspace methods for the solution of large, sparse, linear systems $Ax = b$ [34, 39]. Recent research has provided algorithms which loosen data dependencies in these algorithms, allowing communication- and computation-oriented resources to operate concurrently [12, 13, 23].

Krylov subspace methods for scientific applications are typically preconditioned [40]. This is essential for scalability in many cases. For instance, when the operator A is a discretized elliptic operator, a multilevel preconditioner [38] is required for algorithmic scalability. Preconditioners are often not available as assembled matrices. They commonly involve nested approximate solves, an important case being the application of another Krylov method [28]. The application of a nested Krylov method to a given tolerance and/or iteration count is a nonlinear operation; the optimal polynomial chosen at each iteration depends on A , b , and x_0 . This requires the “outer” Krylov method to be *flexible*, that is, able to operate with a nonlinear preconditioner. Flexible Krylov methods do not involve Krylov subspaces but approximations to them. In comparison with their nonflexible counterparts, flexible Krylov methods typically have higher storage requirements, as they cannot exploit the structure of true Krylov spaces.

In this work, we explore, analyze, and implement Krylov methods which are both pipelined and flexible. As shown in sections 5–7, promising applications of the new methods presented here involve preconditioners with nested inexact solves used over many Krylov iterations. For the overlap of reductions to be useful, systems must involve large numbers of nodes, or one must be in a strong-scaling regime where the time for a global reduction is comparable to the time spent performing local and neighborwise work, or when local processing times are very variable.

Our methods are applicable to the solution of any system of linear equations but are expected to be useful when the number of processing units is large, relative to the local problem size, and when strong, complex, nonlinear (or even nondeterministic) preconditioners are applied to systems difficult enough to require enough Krylov iterations to amortize the overhead of filling a pipeline. Thus, for the examples presented in section 6, we focus on applications in geophysics, in particular mantle convection and lithospheric dynamics with heterogeneous viscosity structure. These represent challenging problems, where strong multiscale nonlinear preconditioners are typically required for acceptable convergence.

Nonlinear preconditioning is still an active area of research, and newly developed nonlinear methods [6] may provide additional use cases for the methods presented here. The advent of nondeterministic and randomized preconditioning techniques, promising for use with hybrid or heterogeneous clusters featuring accelerators, may

provide future use cases for algorithms of the type presented here; these allow overlap of communication and computation and also loosen synchronization requirements, which is crucial when local processing times have heavy-tailed distributions [30].

1.2. Notation and algorithmic presentation. We use similar notation to analyze variants of the conjugate gradient (CG), conjugate residual (CR), and generalized minimal residual (GMRES) methods, as described in Tables 1 and 2 in sections 2 and 4, respectively. This allows similar presentation of the methods, despite the fact that CG and CR methods are written as left preconditioned, while flexible GMRES methods are right preconditioned. Operators are denoted by uppercase letters, vectors by lowercase letters, and scalars by Greek letters. Algorithms are presented using 0-based iteration counts. For readability and analysis, main iteration loops are arranged such that variables indexed by the current index are updated. Convergence tests and optimizations related to initialization, elimination of intermediates, and loop ordering are omitted for readability; the interested reader is referred to section 5 where full open-source implementations are discussed. We adopt the following labeling scheme for operations involving vectors. Scalar or “no operation” operations are not labeled.

(i) A dash (–) denotes an algorithmic step which is “local,” meaning that no internode communication is required in a typical parallel implementation.

(ii) A plus sign (+) denotes an algorithmic step which operates “neighborwise,” that is, only involving communication between $O(1)$ neighboring nodes interspersed with local work. Sparse matrix-vector multiplication typically falls into this category.

(iii) A bullet (•) denotes an algorithmic step which is “global,” that is, involving global communication, exposing the full latency of the network. These operations, which include computing vector dot products and norms, can become the computational bottleneck in massively parallel systems.

(iv) An asterisk (*) is used for preconditioner application, which may be local (e.g., applying a Jacobi preconditioner), neighborwise (e.g., applying a sparse approximate inverse), or global (e.g., applying a nested Krylov method).

1.3. Contributions.

(i) We present pipelined variants of flexible methods, allowing for inexact and variable preconditioning (sections 2–4).

(ii) We provide open-source, scalable implementations of three solvers as contributions to the PETSc package (section 5).

(iii) We provide analysis of a novel modification of a “naive” pipelining which gives unacceptable convergence behavior in typical applications. This leads to useful, pipelined variants of the flexible CG (FCG) and generalized CR (GCR) methods (section 2.2).

(iv) We demonstrate the use of the methods with applications to solving linear systems from challenging problems in lithospheric dynamics (section 6), and with performance models to extrapolate to exascale (section 7).

2. Pipelined flexible conjugate gradient methods.

Notation. We base the notation used in this section, collected in Table 1, on standard notation for CG methods and its pipelined variant [13].

2.1. Review of conjugate gradient methods.

2.1.1. The method of conjugate directions. A general class of methods known as *conjugate direction (CD) methods* [11, 19] has been investigated to compute (approximate) numerical solutions to the linear system $Ax = b$, where $A = A^H$

TABLE 1
Notation for CG and related methods.

A	linear operator
M^{-1}	linear, symmetric positive definite, left preconditioner
B	nonlinear left preconditioner
x	true solution vector $A^{-1}b$
x_i	approximate solution vector at iteration $i = 0, 1, \dots$
e_i	error $x - x_i$ at iteration i
r_i	residual $b - Ax_i = Ae_i$ at iteration i
u_i	preconditioned residual $B(r_i)$ or $M^{-1}r_i$
\tilde{u}_i	approximation to u_i , exact for a linear preconditioner
w_i	pipelining intermediate Au_i or $A\tilde{u}_i$
m_i	pipelining intermediate $B(w_i)$ or $M^{-1}w_i$
n_i	pipelining intermediate Am_i
p_i	search direction or basis vector
s_i	transformed search direction Ap_i
q_i	pipelining intermediate $B(s_i)$ or $M^{-1}s_i$
z_i	pipelining intermediate Aq_i
α_i	scalar weight in solution and residual update
β_i	scalar weight in computation of new search direction or basis vector
η_i	squared norm $\langle Ap_i, p_i \rangle = \ p_i\ _A^2$
δ_i	squared norm $\langle Au_i, u_i \rangle = \ u_i\ _A^2$
γ_i	L_2 -inner product involving preconditioned or unpreconditioned residuals depending on the method under consideration ($\langle r_i, r_i \rangle$, $\langle u_i, u_i \rangle$, or $\langle u_i, r_i \rangle$)

and A is positive definite. Presume one has a set of A -orthogonal (“conjugate”) vectors $\{p_j\}$, $j = 0, \dots, n-1$, available, that is a set of n vectors with the property $\langle p_j, p_k \rangle_A = \langle p_j, Ap_k \rangle = 0$ when $j \neq k$. At each step of the algorithm, the CD method computes an approximate solution x_i with the property that $\|x - x_i\|_A^2$ is minimal for $x_i - x_0 \in \text{span}(p_0, \dots, p_{i-1})$. The resulting process is described in Algorithm 1. Due to the unspecified nature of the A -orthogonal directions $\{p_j\}$, the class of conjugate direction methods includes a wide range of algorithms, including Gaussian elimination and the family of CG methods, which are the main subject of this section.

We note an important property of all CD methods. By the minimization property, we have

$$(1) \quad e_i \perp_A \text{span}(p_0, \dots, p_{i-1}) \iff r_i \perp \text{span}(p_0, \dots, p_{i-1}).$$

That is, the residual is orthogonal to the previously explored space in the standard norm.¹

The directions p_j need not be specified before the algorithm begins; they can be computed as needed, based on the progress of the algorithm.

2.1.2. Flexible conjugate gradient methods. To ensure rapid convergence of the CD algorithm, it is desirable at each iteration for the new search direction p_i to be well aligned with the remaining error $e_i = x - x_i$. As the true error is unknown, one available option is the residual $r_i = Ae_i$, A -orthogonalized against the other search directions.

The residual can be seen as a transformed error or as a gradient descent direction for the function $\frac{1}{2}\|x\|_A^2 - \langle b, x \rangle$, a minimizer of which is a solution of $Ax = b$. Preconditioning can be seen as an attempt to obtain a direction better aligned with e_i by instead A -orthogonalizing $u_i \doteq B(r_i) = B(Ae_i)$ against the previous search directions

¹Other norms may be used, as described by Hestenes [18].

Algorithm 1 Conjugate directions [19].

```

1: function CD( $A, b, x_0, p_0, \dots, p_{n-1}$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
- 3:    $p_0 \leftarrow r_0$ 
+ 4:    $s_0 \leftarrow Ap_0$ 
• 5:    $\gamma_0 \leftarrow \langle p_0, r_0 \rangle$ 
• 6:    $\eta_0 \leftarrow \langle p_0, s_0 \rangle$ 
7:    $\alpha_0 \leftarrow \gamma_0 / \eta_0$ 
8:   for  $i = 1, 2, \dots, n$  do
- 9:      $x_i \leftarrow x_{i-1} + \alpha_{i-1} p_{i-1}$ 
-10:     $r_i \leftarrow r_{i-1} - \alpha_{i-1} s_{i-1}$ 
•11:     $\gamma_i \leftarrow \langle p_i, r_i \rangle$ 
+12:     $s_i \leftarrow Ap_i$ 
•13:     $\eta_i \leftarrow \langle s_i, p_i \rangle$ 
14:     $\alpha_i \leftarrow \gamma_i / \eta_i$ 
    
```

p_0, \dots, p_{i-1} . If B is an approximate inverse of A , it is effective by this criterion, and many preconditioners are motivated as such.

The process just described, found in Algorithm 2, describes the FCG method [31], with complete orthogonalization; i.e., the new search direction is found by A -orthogonalizing the residual against all previous directions in every iteration. Full A -orthogonalization requires potentially excessive memory usage and computation. Hence, it is a common approach to only A -orthogonalize against a number ν_i of previous directions. Note, however, that only for full orthogonalization ($\nu_i = i$) is a true A -orthogonal basis computed by the Gram–Schmidt process in lines 14–16 of Algorithm 2, defining a CD method in the strict sense.

One can observe from Algorithm 2 that no overlapping of reductions and operator or preconditioner application is possible; the dot products depend on the immediately preceding computations.

2.1.3. Preconditioned conjugate gradients. Let $\mathcal{K}^j(A, b)$ represent the j th Krylov subspace $\text{span}(b, Ab, \dots, A^{j-1}b)$. If B is a linear operator $B(v) \equiv M^{-1}v$, then line 15 of Algorithm 2 evaluates to zero for all but the most recent previous direction, and the search directions lie in Krylov subspaces $\mathcal{K}^i(M^{-1}A, M^{-1}b)$. We thus recover the preconditioned conjugate gradient (PCG) method [14, 18, 37] shown in Algorithm 3. Again, communication and computation cannot be overlapped using the algorithm as written, and in a parallel implementation, two reductions must be performed per iteration.

2.1.4. Pipelined preconditioned conjugate gradients. Any rearrangement of the PCG algorithm beyond trivial options must appeal to a notion of algorithmic equivalence more general than producing identical floating-point results. These notions include algebraic rearrangements, which typically do not produce the same iterates in finite precision arithmetic, but which are equivalent in exact arithmetic. One such operation is to rearrange the CG algorithm to involve a single reduction, requiring less communication and synchronization overhead on parallel systems. This was done by Chronopoulos and Gear [8], as described in Algorithm 4. The rearrangement is also particularly important in the context of implementing efficient Krylov methods for current hybrid systems involving high throughput coprocessors for which

Algorithm 2 Flexible conjugate gradients [31].

```

1: function FCG( $A, B, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
* 3:    $u_0 \leftarrow B(r_0)$ 
- 4:    $p_0 \leftarrow u_0$ 
+ 5:    $s_0 \leftarrow Ap_0$ 
• 6:    $\gamma_0 \leftarrow \langle u_0, r_0 \rangle$ 
• 7:    $\eta_0 \leftarrow \langle p_0, s_0 \rangle$ 
8:    $\alpha_0 \leftarrow \gamma_0 / \eta_0$ 
9:   for  $i = 1, 2, \dots$  do
-10:     $x_i \leftarrow x_{i-1} + \alpha_{i-1} p_{i-1}$ 
-11:     $r_i \leftarrow r_{i-1} - \alpha_{i-1} s_{i-1}$ 
*12:     $u_i \leftarrow B(r_i)$ 
•13:     $\gamma_i \leftarrow \langle u_i, r_i \rangle$ 
14:    for  $k = i - \nu_i, \dots, i - 1$  do
•15:      $\beta_{i,k} \leftarrow \frac{-1}{\eta_k} \langle u_i, s_k \rangle$ 
-16:     $p_i \leftarrow u_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k} p_k$ 
+17:     $s_i \leftarrow Ap_i$ 
•18:     $\eta_i \leftarrow \langle p_i, s_i \rangle$ 
19:     $\alpha_i \leftarrow \gamma_i / \eta_i$ 

```

Algorithm 3 Preconditioned conjugate gradients [18].

```

1: function PCG( $A, M^{-1}, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
* 3:    $u_0 \leftarrow M^{-1} r_0$ 
- 4:    $p_0 \leftarrow u_0$ 
+ 5:    $s_0 \leftarrow Ap_0$ 
• 6:    $\gamma_0 \leftarrow \langle u_0, r_0 \rangle$ 
• 7:    $\eta_0 \leftarrow \langle s_0, p_0 \rangle$ 
8:    $\alpha_0 \leftarrow \gamma_0 / \eta_0$ 
9:   for  $i = 1, 2, \dots$  do
-10:     $x_i \leftarrow x_{i-1} + \alpha_{i-1} p_{i-1}$ 
-11:     $r_i \leftarrow r_{i-1} - \alpha_{i-1} s_{i-1}$ 
*12:     $u_i \leftarrow M^{-1} r_i$ 
•13:     $\gamma_i \leftarrow \langle u_i, r_i \rangle$ 
14:     $\beta_i \leftarrow \gamma_i / \gamma_{i-1}$ 
-16:     $p_i \leftarrow u_i + \beta_i p_{i-1}$ 
+17:     $s_i \leftarrow Ap_i$ 
•18:     $\eta_i \leftarrow \langle s_i, p_i \rangle$ 
19:     $\alpha_i \leftarrow \gamma_i / \eta_i$ 

```

kernel fusion is beneficial [32].

This rearrangement is possible because CG relies on the geometric structure induced by interpreting symmetric positive definite operators as inner products. Noting that η_i is defined as the squared A -norm of p_i and that p_i is constructed by A -orthogonalizing u_i with respect to p_{i-1} , one can use the Pythagorean theorem to write

$$\eta_i = \|p_i\|_A^2 = \|u_i + \beta p_{i-1}\|_A^2 = \|u_i\|_A^2 - |\beta|^2 \|p_{i-1}\|_A^2 = \|u_i\|_A^2 - |\beta|^2 \eta_{i-1},$$

where terms involving products of directions at different iteration numbers are zero because of the A -orthogonality of p_i and p_{i-1} . The quantity $\delta_i \doteq \|u_i\|_A^2 = \langle u_i, Au_i \rangle$ does not depend on p_i and can thus be computed at the same time as γ_i . The quantity $w_i \doteq Au_i$ needs to be computed, which apparently adds another matrix multiply to the overall iteration. However, a second type of structure, namely linear structure, can be exploited to reuse this computation and remove the existing matrix multiply with a recurrence relation

$$(2) \quad s_i = Ap_i = Au_i + \beta_i Ap_{i-1} = w_i + \beta s_{i-1}.$$

These rearrangements require additional storage (one extra vector w in this case) and floating-point computations and come with potential numerical penalties as the computed norm could become negative (“norm breakdown”) in finite precision arithmetic. These breakdowns are most easily handled through algorithm restarts. Also, errors can accumulate in recursively computed variables.

As η_i is only used to calculate α_i , one can avoid computing it and update α_i

Algorithm 4 Chronopoulos–Gear conjugate gradients [8].

```

1: function CGCG( $A, M^{-1}, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
* 3:    $u_0 \leftarrow M^{-1}r_0$ 
+ 4:    $w_0 \leftarrow Au_0$ 
5:
6:
• 7:    $\gamma_0 \leftarrow \langle u_0, r_0 \rangle$ 
• 8:    $\delta_0 \leftarrow \langle u_0, w_0 \rangle$ 
9:    $\eta_0 \leftarrow \delta_0$ 
-10:   $p_0 \leftarrow u_0$ 
-11:   $s_0 \leftarrow w_0$ 
12:
13:
14:   $\alpha_0 \leftarrow \gamma_0 / \eta_0$ 
15:  for  $i = 1, 2, \dots$  do
-16:     $x_i \leftarrow x_{i-1} + \alpha_{i-1}p_{i-1}$ 
-17:     $r_i \leftarrow r_{i-1} - \alpha_{i-1}s_{i-1}$ 
18:
19:
*20:     $u_i \leftarrow M^{-1}r_i$ 
+21:     $w_i \leftarrow Au_i$ 
•22:     $\delta_i \leftarrow \langle u_i, w_i \rangle$ 
•23:     $\gamma_i \leftarrow \langle u_i, r_i \rangle$ 
24:     $\beta_i \leftarrow \gamma_i / \gamma_{i-1}$ 
25:     $\eta_i \leftarrow \delta_i - |\beta_i|^2 \eta_{i-1}$ 
26:     $\alpha_i \leftarrow \gamma_i / \eta_i$ 
-27:     $p_i \leftarrow u_i + \beta_i p_{i-1}$ 
-28:     $s_i \leftarrow w_i + \beta_i s_{i-1}$ 
29:
30:
    
```

Algorithm 5 Pipelined conjugate gradients [13].

```

1: function PIPECG( $A, M^{-1}, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax$ 
* 3:    $u_0 \leftarrow M^{-1}r_0$ 
+ 4:    $w_0 \leftarrow Au_0$ 
* 5:    $m_0 \leftarrow M^{-1}w_0$ 
+ 6:    $n_0 \leftarrow Am_0$ 
• 7:    $\gamma_0 \leftarrow \langle u_0, r_0 \rangle$ 
• 8:    $\delta_0 \leftarrow \langle u_0, w_0 \rangle$ 
9:    $\eta_0 \leftarrow \delta_0$ 
-10:   $p_0 \leftarrow u_0$ 
-11:   $s_0 \leftarrow w_0$ 
-12:   $q_0 \leftarrow m_0$ 
-13:   $z_0 \leftarrow n_0$ 
14:   $\alpha_0 \leftarrow \gamma_0 / \eta_0$ 
15:  for  $i = 1, 2, \dots$  do
-16:     $x_i \leftarrow x_{i-1} + \alpha_{i-1}p_{i-1}$ 
-17:     $r_i \leftarrow r_{i-1} - \alpha_{i-1}s_{i-1}$ 
-18:     $u_i \leftarrow u_{i-1} - \alpha_{i-1}q_{i-1}$ 
-19:     $w_i \leftarrow w_{i-1} - \alpha_{i-1}z_{i-1}$ 
*20:     $m_i \leftarrow M^{-1}w_i$ 
+21:     $n_i \leftarrow Am_i$ 
•22:     $\gamma_i \leftarrow \langle u_i, r_i \rangle$ 
•23:     $\delta_i \leftarrow \langle u_i, w_i \rangle$ 
24:     $\beta_i \leftarrow \gamma_i / \gamma_{i-1}$ 
25:     $\eta_i \leftarrow \delta_i - |\beta_i|^2 \eta_{i-1}$ 
26:     $\alpha_i \leftarrow \gamma_i / \eta_i$ 
-27:     $p_i \leftarrow u_i + \beta_i p_{i-1}$ 
-28:     $s_i \leftarrow w_i + \beta_i s_{i-1}$ 
-29:     $q_i \leftarrow m_i + \beta_i q_{i-1}$ 
-30:     $z_i \leftarrow n_i + \beta_i z_{i-1}$ 
    
```

directly as

$$\alpha_i = \frac{\gamma_i}{\eta_i} = \frac{\gamma_i}{(\delta_i - |\beta_i|^2 \eta_{i-1})} = \frac{\gamma_i}{\delta_i - \left(\frac{\gamma_i}{\gamma_{i-1}}\right)^2 \left(\frac{\gamma_{i-1}}{\alpha_{i-1}}\right)} = \frac{\gamma_i}{\delta_i - \bar{\beta}_i \frac{\tilde{\gamma}_i}{\alpha_{i-1}}}.$$

This recurrence is more concise, but we retain explicit computation of η_i in the algorithms as presented here to allow for an easier comparison.

The Chronopoulos–Gear CG method cannot overlap the preconditioner and sparse matrix multiply with the reductions, so some further rearrangement is desirable to allow concurrent use of computational resources. Ghysels and Vanroose [13] developed a further variant of the Chronopoulos–Gear algorithm to accomplish this, as described in Algorithm 5. The algorithm takes advantage of multiple “unrollings” using the linear structure of the variable updates, as in (2). For example, since the computation of $u_i = M^{-1}r_i$ on line 20 of Algorithm 4 blocks the subsequent inner

product involving u_i , one can use the identity

$$(3) \quad u_i = M^{-1}r_i = M^{-1}r_{i-1} - \alpha_{i-1}M^{-1}s_i = u_{i-1} - \alpha_{i-1}q_i,$$

which shifts the application of M^{-1} from r_i to s_i , to compute a new variable $q_i \doteq M^{-1}s_i$. Similarly, q_i is rewritten until computation, for m_i in Algorithm 5, can be overlapped with the dot products in the previous iteration. The identity in (3) relies on the linearity of M^{-1} . This rearrangement allows for more concurrent use of computational resources at the cost of storing more vectors, performing more floating-point operations, and reducing numerical stability [13]. Examples as well as a performance model [12, 13] show performance gains on current and future parallel systems.

The pipelined CG method as presented in Algorithm 5 requires storing 10 vectors compared to 6 for PCG and 7 for Chronopoulos–Gear CG.

2.2. Pipelined flexible conjugate gradient methods. The FCG algorithm can be modified in the same manner used to produce the Chronopoulos–Gear CG algorithm. This does not involve any assumption of linearity of B , so the new single reduction FCG algorithm described in Algorithm 6 is equivalent to FCG in exact arithmetic, for an arbitrary preconditioner.

Algorithm 6 Single reduction flexible conjugate gradients.

```

1: function CGFCG( $A, B, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
* 3:    $u_0 \leftarrow B(r_0)$ 
+ 4:    $w_0 \leftarrow Ap_0$ 
• 5:    $\gamma_0 \leftarrow \langle u_0, r_0 \rangle$ 
• 6:    $\delta_0 \leftarrow \langle u_0, w_0 \rangle$ 
- 7:    $p_0 \leftarrow u_0$ 
- 8:    $s_0 \leftarrow w_0$ 
   9:    $\eta_0 \leftarrow \delta_0$ 
  10:    $\alpha_0 \leftarrow \gamma_0 / \eta_0$ 
  11:   for  $i = 1, 2, \dots$  do
-12:      $x_i \leftarrow x_{i-1} + \alpha_{i-1}p_{i-1}$ 
-13:      $r_i \leftarrow r_{i-1} - \alpha_{i-1}s_{i-1}$ 
*14:      $u_i \leftarrow B(r_i)$ 
+15:      $w_i \leftarrow Au_i$ 
•16:      $\gamma_i \leftarrow \langle u_i, r_i \rangle$ 
  17:     for  $k = i - \nu_i, \dots, i - 1$  do
  18:        $\beta_{i,k} \leftarrow \frac{-1}{\eta_k} \langle u_i, s_k \rangle$ 
•19:      $\delta_i \leftarrow \langle u_i, w_i \rangle$ 
-20:      $p_i \leftarrow u_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}p_k$ 
-21:      $s_i \leftarrow w_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}s_k$ 
  22:      $\eta_i \leftarrow \delta_i - \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}^2 \eta_k$ 
  23:      $\alpha_i \leftarrow \gamma_i / \eta_i$ 

```

One can now attempt to pipeline FCG in the same manner as pipelined variants of CG are obtained—this leads to the variant described in Algorithm 7. The quantity \tilde{u} is used to reinforce the idea that the preconditioned search directions differ in general from those computed by FCG.

Algorithm 7 Pipelined FCG (naive).

```

1: function PIPEFCG?( $A, B, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
* 3:    $\tilde{u}_0 \leftarrow B(r_0)$ 
- 4:    $p_0 \leftarrow \tilde{u}_0$ 
+ 5:    $w_0 \leftarrow Ap_0$ 
* 6:    $m_0 \leftarrow B(w_0)$ 
+ 7:    $n_0 \leftarrow Am_0$ 
• 8:    $\gamma_0 \leftarrow \langle \tilde{u}_0, r_0 \rangle$ 
• 9:    $\delta_0 \leftarrow \langle \tilde{u}_0, w_0 \rangle$ 
-10:   $s_0 \leftarrow w_0$ 
-11:   $q_0 \leftarrow m_0$ 
-12:   $z_0 \leftarrow n_0$ 
13:   $\eta_0 \leftarrow \delta_0$ 
14:   $\alpha_0 \leftarrow \gamma_0 / \eta_0$ 
15:  for  $i = 1, 2, \dots$  do
-16:     $x_i \leftarrow x_{i-1} + \alpha_{i-1} p_{i-1}$ 
-17:     $r_i \leftarrow r_{i-1} - \alpha_{i-1} s_{i-1}$ 
-18:     $\tilde{u}_i \leftarrow \tilde{u}_{i-1} - \alpha_{i-1} q_{i-1}$ 
-19:     $w_i \leftarrow w_{i-1} - \alpha_{i-1} z_{i-1}$ 
•20:     $\gamma_i \leftarrow \langle \tilde{u}_i, r_i \rangle$ 
21:    for  $k = i - \nu_i, \dots, i - 1$  do
•22:       $\beta_{i,k} \leftarrow \frac{-1}{\eta_k} \langle \tilde{u}_i, s_k \rangle$ 
•23:     $\delta_i \leftarrow \langle \tilde{u}_i, w_i \rangle$ 
*24:     $\mathbf{m}_i \leftarrow \mathbf{B}(\mathbf{w}_i)$ 
+25:     $n_i \leftarrow Am_i$ 
-26:     $p_i \leftarrow \tilde{u}_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k} p_k$ 
-27:     $s_i \leftarrow w_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k} s_k$ 
-28:     $q_i \leftarrow m_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k} q_k$ 
-29:     $z_i \leftarrow n_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k} z_k$ 
30:     $\eta_i \leftarrow \delta_i - \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}^2 \eta_k$ 
31:     $\alpha_i \leftarrow \gamma_i / \eta_i$ 

```

Algorithm 8 Pipelined FCG.

```

1: function PIPEFCG( $A, B, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
* 3:    $\tilde{u}_0 \leftarrow B(r_0)$ 
- 4:    $p_0 \leftarrow \tilde{u}_0$ 
+ 5:    $w_0 \leftarrow Ap_0$ 
* 6:    $m_0 \leftarrow B(w_0)$ 
+ 7:    $n_0 \leftarrow Am_0$ 
• 8:    $\gamma_0 \leftarrow \langle u_0, r_0 \rangle$ 
• 9:    $\delta_0 \leftarrow \langle u_0, w_0 \rangle$ 
-10:   $s_0 \leftarrow w_0$ 
-11:   $q_0 \leftarrow m_0$ 
-12:   $z_0 \leftarrow n_0$ 
13:   $\eta_0 \leftarrow \delta_0$ 
14:   $\alpha_0 \leftarrow \gamma_0 / \eta_0$ 
15:  for  $i = 1, 2, \dots$  do
-16:     $x_i \leftarrow x_{i-1} + \alpha_{i-1} p_{i-1}$ 
-17:     $r_i \leftarrow r_{i-1} - \alpha_{i-1} s_{i-1}$ 
-18:     $\tilde{u}_i \leftarrow \tilde{u}_{i-1} - \alpha_{i-1} q_{i-1}$ 
-19:     $w_i \leftarrow w_{i-1} - \alpha_{i-1} z_{i-1}$ 
•20:     $\gamma_i \leftarrow \langle \tilde{u}_i, r_i \rangle$ 
21:    for  $k = i - \nu_i, \dots, i - 1$  do
•22:       $\beta_{i,k} \leftarrow \frac{-1}{\eta_k} \langle \tilde{u}_i, s_k \rangle$ 
•23:     $\delta_i \leftarrow \langle \tilde{u}_i, w_i \rangle$ 
*24:     $\mathbf{m}_i \leftarrow \tilde{\mathbf{u}}_i + \mathbf{B}(\mathbf{w}_i - \mathbf{r}_i)$ 
+25:     $n_i \leftarrow Am_i$ 
-26:     $p_i \leftarrow \tilde{u}_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k} p_k$ 
-27:     $s_i \leftarrow w_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k} s_k$ 
-28:     $q_i \leftarrow m_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k} q_k$ 
-29:     $z_i \leftarrow n_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k} z_k$ 
30:     $\eta_i \leftarrow \delta_i - \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}^2 \eta_k$ 
31:     $\alpha_i \leftarrow \gamma_i / \eta_i$ 

```

A variant analogous to Gropp's asynchronous CG [16] could also be defined. We do this for the corresponding method based on conjugate residuals in section 3.2.

2.2.1. Issues with naive pipelining and variable preconditioning. Convergence of Algorithm 7 typically stagnates for nonlinear preconditioners B . In cases where the preconditioner has an associated “noise level” or tolerance which characterizes its variability, stagnation is typically observed at a relative error comparable to this quantity. This is illustrated in Figure 1, plotting convergence for a diagonal system with a preconditioner which simply adds Gaussian noise to the residual.

Analysis of preconditioned CG methods typically makes use of the preconditioner as an inner product. The CG method can be seen as a Ritz–Galerkin approach to a Krylov subspace method where approximations in successive Krylov subspaces are chosen to satisfy the condition that the resulting residual is orthogonal to the Krylov subspace [39, section 4.1]. Choosing a different definition of orthogonality, based

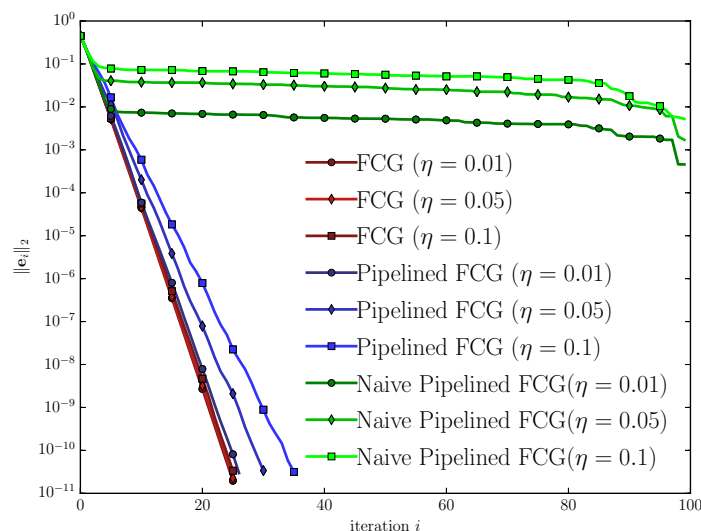


FIG. 1. A toy diagonal system, $n = 100$, with condition number 5, equally spaced eigenvalues, and $b = 1/\sqrt{n}$, solved with FCG, naively pipelined FCG (Algorithm 7), and our PIPEFCG (Algorithm 8) with $\nu_{\max} = n$. The pathological preconditioner B adds Gaussian noise of magnitude $\eta\|r\|$ to the residual r . Note that convergence for the naively pipelined algorithm stagnates at a relative error of approximately η .

on the metric induced by M^{-1} , produces the PCG method. This is equivalent to the notion of simply replacing the L_2 -inner product with a different inner product, wherever it appears in the algorithm. This freedom to choose norms was recognized soon after the development of the CG method [18]. In terms of the PCG, Algorithm 3,

$$(4) \quad \gamma_i = \langle r_i, r_i \rangle_{M^{-1}} = \langle u_i, r_i \rangle.$$

Here, the nonlinearity of the preconditioner precludes direct use of this idea. We opt to consider B in a more general sense of a preconditioner, as an approximate inverse to A . Specifically, noting that $Ae_i = r_i$, we shall later require that $\|B(r_i) - e_i\|/\|e_i\|$ be controlled.

Naively replacing the linear preconditioner M^{-1} in (3) with a nonlinear one, we obtain the unrolling used in Algorithm 7,

$$(5) \quad u_i = B(r_i) \approx B(r_{i-1}) - B(\alpha_{i-1}s_i) \approx u_{i-1} - \alpha_{i-1}q_i,$$

which is inexact even in exact arithmetic. This unrolling is not a priori unacceptable, as FCG is designed to cope with varying and inexact preconditioning, and the rearrangement amounts to a different nonlinear preconditioner, albeit one derived using intuition from the linear case. However, unlike the unrolling (3), the above leads to an accumulation of errors in the preconditioned residual \tilde{u}_i , with respect to $u_i = B(r_i)$. This invalidates the equivalence (4), which PCG structurally relies on for projecting out search directions from the residual.

Note that when introducing a nonlinear preconditioner with FCG² the basis

²We assume $\nu_i = i$, that is, complete orthogonalization, for all of the analysis in this section.

vectors p_j remain A -orthogonal with this rearrangement, yet the residuals r_i , which are M^{-1} -orthogonal when $B(v) \equiv M^{-1}v$, are now orthogonal with respect to an equivalent linear operator \hat{B} , using Notay's concept and notation of an *equivalent linear operator* [31]. This operator is defined by the relations $u_i = B(r_i) = \hat{B}r_i$, depends on b and x_0 , and is unique except in degenerate cases. It seems to be the case that the equivalent linear operator \hat{B} implicitly defined (cf. (3)) by the update on line 18 of Algorithm 7 is unbounded as $i \rightarrow \infty$, mapping residuals of arbitrarily small magnitude to preconditioned residuals of the same order, making it an unacceptable approximation for any given linear operator M^{-1} bounded as $n \rightarrow \infty$.

Thus, it will be useful to distinguish between the equivalent linear operators as induced by the base and pipelined methods, \hat{B} and $\hat{\hat{B}}$, respectively. The operator $\hat{\hat{B}}$ is defined by the relations $\tilde{u}_i = \hat{\hat{B}}r_i$. The quantities α_i , in a CD method like FCG, are chosen to minimize $\|x - (x_{i+1} + \alpha_{i-1}u_i)\|_A$. However, in the pipelined variant, the effective update is $x_i \leftarrow x_{i-1} + \alpha_{i-1}\tilde{u}_i$. Thus, deviation from the base method induced by pipelining can of course be characterized by the difference between u and \tilde{u} or, equivalently, the difference between \hat{B} and $\hat{\hat{B}}$. The advantage of this second view is that it suggests a repair procedure. Namely, one should respect the already-defined part of \hat{B} as much as possible when computing \tilde{u} , hence partially defining $\hat{\hat{B}}$. One approach is to note that the operation of the equivalent linear operator in direction r_i has already been defined, so one can project this direction out of the quantity to which B is applied. Thus, the algorithm can be modified to compute the vector

$$(6) \quad m_i \leftarrow \theta_i \tilde{u}_{i-1} + B(w_{i-1} - \theta_i r_{i-1}), \quad \theta_i \doteq \frac{\langle r, w \rangle}{\|r\|^2}.$$

Indeed, numerical experiments show that this modification can restore convergence in some cases. This modification is, unfortunately, not useful as it stands for the purposes of improving the naive pipelined FCG algorithm, as it requires additional blocking reductions. A natural question, however, is how one might estimate θ_i . One useful property is that as the preconditioner approaches A^{-1} , θ_i approaches 1 for all i . This motivates the choice of $\theta_i \equiv 1$ as at least an “unbiased” approximation.

2.2.2. Faithful preconditioners. To be more precise, we restrict our attention to a subclass of nonlinear preconditioners.

DEFINITION 1. A family of functions $B_n : \mathbb{C}^n \mapsto \mathbb{C}^n$, $n = 1, 2, \dots$, represents a faithful preconditioner with respect to a family of linear operators $A_n \in L(\mathbb{C}^n, \mathbb{C}^n)$ if $B_n(A_n(\cdot)) - I_n(\cdot)$ is a uniformly bounded sequence of operators, with respect to the standard norm on \mathbb{C}^n . That is, there exists $c \in \mathbb{R}$ such that

$$\|B_n(A_n v) - v\| < c\|v\| \quad \forall v \in \mathbb{C}^n \setminus \{0\}, \quad \forall n \in \{1, 2, \dots\}.$$

Remark 2. We will typically abuse notation and omit the dependence on n .

Remark 3. Applied to the special case where $v = e_i \doteq x - x_i$, the error in a particular iteration of a Krylov method, noting that $B(Ae_i) = B(r_i) = u_i$, the condition above implies that

$$\|u_i - e_i\| < c\|e_i\|,$$

which is to say that the preconditioned residual is a *uniformly accurate approximation to the error*.

Remark 4. The class of sparse approximate inverse preconditioners is a special case of faithful preconditioners if the approximation quality is uniform in the problem size.

To compute θ_i in (6) requires additional inner products, so instead we seek to cheaply estimate it. Note that θ_i minimizes the distance $\|w_i - \theta_i r_i\|$. If we assume our preconditioner to be faithful with a small constant, so that the operator $AB(\cdot) - I(\cdot)$ is bounded with a small constant, then $\theta_i \equiv 1$ is a reasonable estimate, as $\|w - r\|/\|r\| = \|AB(r) - r\|/\|r\|$ is then small.

Note that $\theta_i \equiv 0$ corresponds to the naive unrolling as in (5).

2.2.3. Defining an improved pipelined flexible conjugate gradient algorithm. In a modified version of the naive pipelined FCG, Algorithm 7, we take advantage of the estimate $\theta_i \equiv 1$ to mitigate the observed stagnation of convergence.

Motivated by the considerations above, we observe great improvements employing the approximation

$$(7) \quad \tilde{u}_i = B(r_i) = B(r_{i-1} - \alpha_{i-1} A p_{i-1}) \approx (1 - \alpha_{i-1}) B(r_{i-1}) - \alpha_{i-1} B(A p_{i-1} - r_{i-1})$$

instead of the recurrence relation (5). This leads to a replacement of the algorithmic step $m_i \leftarrow B(w_i)$ in line 24 of Algorithm 7 by

$$(8) \quad m_i \leftarrow \tilde{u}_i + B(w_i - r_i).$$

For a linear preconditioner these are equivalent in exact arithmetic.

We proceed to show that this can be interpreted in terms of improved control of the accumulation of perturbations in \tilde{u}_i with respect to $B(r_i)$, assuming the preconditioner is sufficiently faithful.

Note that $w_i = A\tilde{u}_i$ in exact arithmetic and rewrite (8) as

$$m_i \leftarrow \tilde{u}_i + B(A\tilde{u}_i - r_i) = \tilde{u}_i + B(A(\tilde{u}_i - e_i)).$$

With $q_i = m_i + \sum_k \beta_k q_k$ we write out the recurrence relation

$$\begin{aligned} \tilde{u}_i &= \tilde{u}_{i-1} - \alpha_{i-1} \left(\tilde{u}_{i-1} + B(A(\tilde{u}_{i-1} - e_{i-1})) + \sum_k \beta_k q_k \right) \\ &= (1 - \alpha_{i-1}) \tilde{u}_{i-1} - \alpha_{i-1} \left(B(A(\tilde{u}_{i-1} - e_{i-1})) + \sum_k \beta_k q_k \right). \end{aligned}$$

Rearranging terms we obtain

$$(9) \quad \tilde{u}_i = \underbrace{\left[(1 - \alpha_{i-1}) \tilde{u}_{i-1} - \alpha_{i-1} \sum_k \beta_k q_k \right]}_I - \underbrace{\left[\alpha_{i-1} B(A(\tilde{u}_{i-1} - e_{i-1})) \right]}_{II},$$

where term II is affected by the nonlinear preconditioner but term I is not. Using Definition 1 and Remark 3 we rewrite II as

$$\alpha_{i-1} B(A(\tilde{u}_{i-1} - e_{i-1})) = \alpha_{i-1} \tilde{I}(\tilde{u}_{i-1} - e_{i-1}) \approx \alpha_{i-1} (\tilde{u}_{i-1} - e_{i-1}),$$

with the operator \tilde{I} close to the identity operator using the notion of a faithful preconditioner. Comparing magnitudes of I and this approximation of II we expect that

$$(10) \quad \|(1 - \alpha_{i-1}) \tilde{u}_{i-1} - \alpha_{i-1} \sum_k \beta_k q_k\| > \|\alpha_{i-1} (\tilde{u}_{i-1} - e_{i-1})\|$$

as a faithful preconditioner transforms the residual such that it is similar to the true error e . If instead we use $B(w)$ as in the naive Algorithm 7, then

$$\tilde{u}_i = \underbrace{\tilde{u}_{i-1} - \alpha_{i-1} \sum_k \beta_k q_k}_{\text{III}} - \underbrace{B(A\tilde{u}_{i-1})}_{\text{IV}}.$$

Following the same line of arguments we obtain that $B(A\tilde{u}_{i-1}) = \tilde{I}(\tilde{u}_{i-1}) \approx \tilde{u}_{i-1}$. Hence, terms III and IV are of similar magnitude. While for the modified formulation the application of a nonlinear preconditioner leads to small perturbations of the preconditioned residual and, consequently, of the next search directions, this does not hold true for the naive formulation. Through recurrence of the preconditioned residual the accumulation of perturbations eventually leads to effectively random search directions and stagnation of convergence. This is observed in Figure 1. Once the “noise level” of the preconditioner is reached, convergence stagnates.

By employing the modification to the computation of the quantity m , analyzed in the previous section, we arrive at a useful pipelined variant of the FCG method, Algorithm 8.

Remark 5. Algorithm 8 satisfies several desiderata:

- (i) In exact arithmetic, with a linear preconditioner, it is equivalent to CG and FCG.
- (ii) As $B \rightarrow A^{-1}$, the operator becomes increasingly “faithful,” and hence the algorithm approaches the behavior of FCG.
- (iii) No new reductions are introduced.
- (iv) A single reduction phase per iteration is used, which can be overlapped with application of the operator and preconditioner.

The pipelined FCG method requires storing $(4\nu_{\max} + 11)$ vectors, compared to $(\nu_{\max} + 6)$ vectors for FCG (Algorithm 2) and $(2\nu_{\max} + 7)$ for single reduction FCG (Algorithm 6).

3. Pipelined generalized conjugate residuals methods. The CR family of Krylov subspace methods is applicable to systems with a Hermitian but not necessarily positive definite operator. While CG methods enforce A -conjugacy of search directions and minimize $\|e\|_A$, CR methods construct $A^H A$ -orthogonal search directions. In some derivations of the method, such as in Saad’s textbook [34], iterates minimize $\|e_i\|_{A^H A} = \|r_i\|$ over a Krylov subspace, which puts these methods in close relation to MINRES [39, section 6.4] and GMRES [33, 35]. However, in some other derivations [25], these methods are described as minimizing $\|e\|_A$, as does CG. Here, we consider methods of the first kind, that is, minimal residual methods.

3.1. Review of conjugate residuals methods. There is also an ambiguity regarding the terminology within the related flexible Krylov subspace methods. We note that the generalized conjugate residual (GCR) method [10], which this section is concerned with, is closely related to FCG. However, despite its name, the GCR method is less closely related to Axelsson’s generalized conjugate gradient (GCG) method [2, 1]. While FCG and GCR both employ a multiterm Gram–Schmidt orthogonalization of new search directions against previous ones, GCG additionally applies old directions in the update of the solution and residual vectors.

As in the preceding section, we describe the flexible method (GCR) and proceed to show its reduction to preconditioned CR (PCR) for constant preconditioners and extension to pipelined CR [13]. We introduce pipelined GCR in section 3.2.

Algorithm 9 Generalized conjugate residuals.

```

1: function GCR( $A, B, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
* 3:    $u_0 \leftarrow B(r_0)$ 
- 4:    $p_0 \leftarrow u_0$ 
+ 5:    $s_0 \leftarrow Ap_0$ 
• 6:    $\gamma_0 \leftarrow \langle u_0, s_0 \rangle$ 
• 7:    $\eta_0 \leftarrow \langle s_0, s_0 \rangle$ 
8:    $\alpha_0 \leftarrow \gamma_0/\eta_0$ 
9:   for  $i = 1, 2, \dots$  do
-10:     $x_i \leftarrow x_{i-1} + \alpha_{i-1}p_{i-1}$ 
-11:     $r_i \leftarrow r_{i-1} - \alpha_{i-1}s_{i-1}$ 
*12:     $u_i \leftarrow B(r_i)$ 
13:    for  $k = i - \nu_i, \dots, i - 1$  do
•14:      $\beta_{i,k} \leftarrow \frac{-1}{\eta_k} \langle Au_i, s_k \rangle$ 
-15:     $p_i \leftarrow u_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}p_k$ 
+16:     $s_i \leftarrow Ap_i$ 
•17:     $\gamma_i \leftarrow \langle u_i, s_i \rangle$ 
•18:     $\eta_i \leftarrow \langle s_i, s_i \rangle$ 
19:     $\alpha_i \leftarrow \gamma_i/\eta_i$ 

```

Algorithm 10 Preconditioned conjugate residuals.

```

1: function PCR( $A, M^{-1}, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
* 3:    $u_0 \leftarrow M^{-1}r_0$ 
- 4:    $p_0 \leftarrow u_0$ 
+ 5:    $s_0 \leftarrow Ap_0$ 
• 6:    $\gamma_0 \leftarrow \langle u_0, s_0 \rangle$ 
• 7:    $\eta_0 \leftarrow \langle s_0, s_0 \rangle$ 
8:    $\alpha_0 \leftarrow \gamma_0/\eta_0$ 
9:   for  $i = 1, 2, \dots$  do
-10:     $x_i \leftarrow x_{i-1} + \alpha_{i-1}p_{i-1}$ 
-11:     $r_i \leftarrow r_{i-1} - \alpha_{i-1}s_{i-1}$ 
*12:     $u_i \leftarrow M^{-1}r_i$ 
•13:     $\gamma_i \leftarrow \langle u_i, s_i \rangle$ 
14:     $\beta_i \leftarrow \gamma_i/\gamma_{i-1}$ 
-15:     $p_i \leftarrow u_i + \beta_i p_{i-1}$ 
+16:     $s_i \leftarrow Ap_i$ 
17:
•18:     $\eta_i \leftarrow \langle s_i, s_i \rangle$ 
19:     $\alpha_i \leftarrow \gamma_i/\eta_i$ 

```

The GCR method is shown in Algorithm 9. From a comparison of the GCR and FCG methods (Algorithms 9 and 2), it is straightforward to see that they employ $A^H A$ - and A -inner products, respectively. As in section 2.1.3, we recover the nonflexible (or nongeneralized) PCR method by restricting to linear preconditioners. In this case all but one β_k are zero, giving Algorithm 10.

3.2. Pipelined generalized conjugate residual methods. GCR and PCR do not allow any overlapping of global reductions in their standard formulation. Reformulating the algorithms using the Chronopoulos–Gear trick and introducing unrollings and pipelining intermediates leads to two pipelined GCR variants, shown in Algorithms 11 and 12. These employ the same modification to the computation of m_i as discussed in section 2.2.3.

In Algorithm 11 the quantity w_i is computed by performing the sparse matrix-vector product Au_i in line 15, whereas in Algorithm 12 w is computed with a recurrence relation. Note that the overlapping properties of the variants are different, as in Algorithm 11 the application of the preconditioner overlaps the reductions, while in Algorithm 12 a sparse matrix-vector product also contributes to the overlapping. The performance of both variants is modeled in section 7. The suitability of one or the other algorithm depends on the cluster architecture, in particular the speed of the connecting network, and the available memory. Algorithm 12 has a larger memory footprint due to the storage of the extra intermediate variable z and its history of length ν_i .

The original GCR method [10] as presented in Algorithm 9 requires a total of $(2\nu_{\max} + 7)$ vectors to be stored. The pipelined variants of Algorithms 11 and 12 require $(\nu_{\max} + 2)$ and $(2\nu_{\max} + 4)$ additional vectors to be stored.

Algorithm 11 Pipelined generalized conjugate residuals.

```

1: function PIPEGCR( $A, B, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
* 3:    $\tilde{u}_0 \leftarrow B(r_0)$ 
- 4:    $p_0 \leftarrow u_0$ 
+ 5:    $s_0 \leftarrow Ap_0$ 
* 6:    $q_0 \leftarrow B(s_0)$ 
7:
• 8:    $\gamma_0 \leftarrow \langle r_0, Ar_0 \rangle$ 
• 9:    $\eta_0 \leftarrow \|s_0\|$ 
10:   $\alpha_0 \leftarrow \gamma_0/\eta_0$ 
11:  for  $i = 1, 2, \dots$  do
-12:     $x_i \leftarrow x_{i-1} + \alpha_{i-1}p_{i-1}$ 
-13:     $r_i \leftarrow r_{i-1} - \alpha_{i-1}s_{i-1}$ 
-14:     $u_i \leftarrow \tilde{u}_{i-1} - \alpha_{i-1}q_{i-1}$ 
+15:     $w_i \leftarrow Au_i$ 
*16:     $m_i \leftarrow B(w_i - r_i) + \tilde{u}_i$ 
17:
•18:     $\gamma_i \leftarrow \langle r_i, w_i \rangle$ 
•19:     $\delta_i \leftarrow \langle w_i, w_i \rangle$ 
20:    for  $k = i - \nu_i, \dots, i - 1$  do
•21:       $\beta_{i,k} \leftarrow \frac{-1}{\eta_k} \langle w_i, s_k \rangle$ 
-22:       $p_i \leftarrow \tilde{u}_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}p_k$ 
-23:       $s_i \leftarrow w_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}s_k$ 
-24:       $q_i \leftarrow m_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}q_k$ 
25:
26:       $\eta_i \leftarrow \delta_i - \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}^2 \eta_k$ 
27:       $\alpha_i \leftarrow \gamma_i/\eta_i$ 

```

Algorithm 12 Pipelined generalized conjugate residuals (w unrolled).

```

1: function PIPEGCR_w( $A, B, b, x_0$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
* 3:    $\tilde{u}_0 \leftarrow B(r_0)$ 
- 4:    $p_0 \leftarrow \tilde{u}_0$ 
+ 5:    $s_0 \leftarrow Ap_0$ 
* 6:    $q_0 \leftarrow B(s_0)$ 
+ 7:    $z_0 \leftarrow Aq_0$ 
• 8:    $\gamma_0 \leftarrow \langle r_0, Ar_0 \rangle$ 
• 9:    $\eta_0 \leftarrow \|s_0\|$ 
10:   $\alpha_0 \leftarrow \gamma_0/\eta_0$ 
11:  for  $i = 1, 2, \dots$  do
-12:     $x_i \leftarrow x_{i-1} + \alpha_{i-1}p_{i-1}$ 
-13:     $r_i \leftarrow r_{i-1} - \alpha_{i-1}s_{i-1}$ 
-14:     $u_i \leftarrow \tilde{u}_{i-1} - \alpha_{i-1}q_{i-1}$ 
-15:     $w_i \leftarrow w_{i-1} - \alpha_{i-1}z_{i-1}$ 
*16:     $m_i \leftarrow B(w_i - r_i) + \tilde{u}_i$ 
+17:     $n_i \leftarrow Am_i$ 
•18:     $\gamma_i \leftarrow \langle r_i, w_i \rangle$ 
•19:     $\delta_i \leftarrow \langle w_i, w_i \rangle$ 
20:    for  $k = i - \nu_i, \dots, i - 1$  do
•21:       $\beta_{i,k} \leftarrow \frac{-1}{\eta_k} \langle w_i, s_k \rangle$ 
-22:       $p_i \leftarrow \tilde{u}_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}p_k$ 
-23:       $s_i \leftarrow w_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}s_k$ 
-24:       $q_i \leftarrow m_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}q_k$ 
-25:       $z_i \leftarrow n_i + \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}z_k$ 
26:       $\eta_i \leftarrow \delta_i - \sum_{k=i-\nu_i}^{i-1} \beta_{i,k}^2 \eta_k$ 
27:       $\alpha_i \leftarrow \gamma_i/\eta_i$ 

```

4. Pipelined flexible GMRES methods. GMRES methods use explicitly orthonormalized basis vectors and as such do not rely on the same identities which were so easily disturbed by algorithmic rearrangement in the FCG and GCR cases above. Thus, deriving a usable pipelined FGMRES method is a relatively straightforward extension of pipelined GMRES, as predicted with the introduction of that method [12].

4.1. Notation. We base our notation in Table 2 on standard notation for GMRES methods and pipelined GMRES. It is distinct from the notation used for CG and its variants because FGMRES methods naturally use right preconditioning. The meanings of symbols p_i , r_i , e_i , A , B , and M^{-1} are identical to those in Table 1.

4.2. Review of the flexible GMRES method. The FGMRES method [33] modifies the GMRES method [35] to admit variable right preconditioning. It accomplishes this by finding approximate solutions x_i with $x_i - x_0$ having minimal residual 2-norm $\|b - Ax_i\|_2$ in the subspace $\text{span}(b, u_1, \dots, u_{i-1})$, where $u_i \doteq B(p_i)$ and p_i

TABLE 2
Notation for GMRES and related methods.

u_i	preconditioned basis vector $B(p_i)$ or $M^{-1}p_i$
\tilde{u}	approximation to u_i , exact if B is linear
z_i	transformed and preconditioned basis vector Au_i or $A\tilde{u}_i$
\tilde{z}_i	preconditioned, transformed, and shifted basis vector $Au_i - \sigma_i p_i$ or $A\tilde{u}_i - \sigma_i p_i$
\tilde{q}_i	pipelining intermediate $B(\tilde{z}_i)$
\tilde{w}_i	pipelining intermediate $A\tilde{q}_i$

is formed by orthonormalizing $z_{i-1} \doteq Au_{i-1} = AB(p_{i-1})$ with p_0, \dots, p_{i-1} .³ This Arnoldi process can be summarized as $AU = PH$, where P is a matrix with i th column p_i , U is a matrix with i th column u_i , and H is an upper Hessenberg matrix. Let P be a matrix with i th column p_i . If $B = M^{-1}$ for some fixed linear operator, GMRES is recovered: the solution is found in the Krylov space $\mathcal{K}^i(AM^{-1}, b)$ and the relation $AU = AM^{-1}P = PH$ is available to avoid the need to explicitly store the vectors u_i . Algorithm 13 describes the FGMRES method. Replacing line 15 with

$$x \leftarrow x_0 + M^{-1}Py$$

in the case of a linear preconditioner recovers the right preconditioned GMRES algorithm. The FGMRES method may be preferable even for a constant preconditioner if this additional application of the preconditioner (“unwinding”) is expensive and the required additional memory is available. Restarting and convergence checks (including on-the-fly QR decomposition of H and happy breakdown checks) are not included for simplicity. Similarly, checks for the (rare in practice) possibility that the Hessenberg matrix H is singular are omitted.

Algorithm 13 Flexible GMRES [33].

```

1: function FGMRES( $A, B, b, x_0, m$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
• 3:    $p_0 \leftarrow r_0 / \|r_0\|_2$ 
* 4:    $u_0 \leftarrow B(p_0)$ 
+ 5:    $z_0 \leftarrow Au_0$ 
6:   for  $i = 1, \dots, m$  do
7:     for  $j = 0, \dots, i-1$  do
• 8:        $h_{j,i-1} \leftarrow \langle p_j, z_{i-1} \rangle$ 
- 9:      $\bar{p} \leftarrow z_{i-1} - \sum_{k=0}^{i-1} p_k h_{k,i-1}$ 
•10:     $h_{i,i-1} \leftarrow \|\bar{p}\|_2$ 
-11:     $p_i \leftarrow \bar{p} / h_{i,i-1}$ 
*12:     $u_i \leftarrow B(p_i)$ 
+13:     $z_i \leftarrow Au_i$ 
-14:     $y \leftarrow \operatorname{argmin} \|H_{m+1,m}y - \|r_0\|_2 e_0\|_2$ 
-15:     $x \leftarrow x_0 + Uy$ 

```

4.3. Review of the pipelined GMRES method. The GMRES algorithm can be pipelined with the strategy outlined by Ghysels et al. [12]. They observe that at

³For improved numerical stability, the same subspace can be spanned by orthogonalizing $\tilde{z}_{i-1} \doteq Au_{i-1} - \sigma_i p_{i-1}$ [12].

the cost of degraded numerical stability, GMRES can be modified to involve only one global reduction (“ l^1 -GMRES” in their nomenclature). This parallels the modification of PCG (Algorithm 3) to derive Chronopoulos–Gear CG (Algorithm 4). They use the identity $p_i = Ap_{i-1} - \sum_{k=1}^{i-1} \langle Ap_{i-1}, p_k \rangle p_k$ and note that if $z_i \doteq Ap_{i-1}$ is already available, this update can be performed while $Az_i = A^2 p_{i-1}$ is being computed. Using the now available value of p_i , one can (locally) compute $z_{i+1} \doteq Ap_i$.

4.4. Pipelining the flexible GMRES method. The same procedure used to pipeline GMRES can also be used to produce a single reduction variant of FGMRES, as shown in Algorithm 14, and a pipelined version of FGMRES, as shown in Algorithm 15.

This algorithm can suffer “norm breakdown” requiring a restart and refilling of the pipeline. Tuning of the shift parameters can help avoid this. We allow for a distinct shift σ_i to be used at each iteration, but note that if this value is constant, the expression on line 22 of Algorithm 15 simplifies to

$$(11) \quad \bar{z}_i \leftarrow \left(\bar{w}_{i-1} - \sum_{k=0}^{i-1} h_{k,i-1} \bar{z}_k \right) / h_{i,i-1}.$$

A practical choice, and the one we use in our implementations discussed in sections 5 and 6, is to set σ_i to a constant value which approximates the largest eigenvalue of the preconditioned operator.

Algorithm 14 Single reduction FGMRES.

```

1: function CGFGMRES( $A, B, b, x_0, m$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
• 3:    $p_0 \leftarrow r_0 / \|r_0\|_2$ 
* 4:    $u_0 \leftarrow B(p_0)$ 
+ 5:    $\bar{z}_0 \leftarrow Au_0 - \sigma_0 p_0$ 
6:   for  $i = 1, \dots, m$  do
7:     for  $j = 0, \dots, i-2$  do
• 8:        $\bar{h}_{j,i-1} \leftarrow \langle p_j, \bar{z}_{i-1} \rangle$ 
9:        $h_{j,j-1} \leftarrow \bar{h}_{j,i-1}$ 
•10:       $\bar{h}_{i-1,i-1} \leftarrow \langle p_{i-1}, \bar{z}_{i-1} \rangle$ 
11:       $h_{i-1,i-1} \leftarrow \bar{h}_{i-1,i-1} + \sigma_{i-1}$ 
•12:       $t \leftarrow \|\bar{z}_{i-1}\|_2^2 - \sum_{k=0}^{i-1} \bar{h}_{k,i-1}^2$ 
13:      if  $t < 0$  then
14:        BREAKDOWN
15:       $\bar{h}_{i,i-1} \leftarrow \sqrt{t}$ 
16:       $h_{i,i-1} \leftarrow \bar{h}_{i,i-1}$ 
17:      if  $i = m$  then break
-18:       $p_i \leftarrow (\bar{z}_{i-1} - \sum_{k=0}^{i-1} \bar{h}_{k,i-1} p_k) / \bar{h}_{i,i-1}$ 
*19:       $u_i \leftarrow B(p_i)$ 
+20:       $\bar{z}_i \leftarrow Au_i - \sigma_i p_i$ 
-21:       $y \leftarrow \operatorname{argmin} \|H_{m+1,m} y - \|r_0\|_2 e_0\|_2$ 
-22:       $x \leftarrow x_0 + Uy$ 

```

In exact arithmetic, the iterates produced by Algorithm 15 are not equivalent to the FGMRES algorithm because of the approximation in computing \tilde{u}_i on line 21.

Algorithm 15 Single-stage pipelined FGMRES.

```

1: function PIPEFGMRES( $A, B, b, x_0, m$ )
+ 2:    $r_0 \leftarrow b - Ax_0$ 
• 3:    $p_0 \leftarrow r_0 / \|r_0\|_2$ 
* 4:    $u_0 \leftarrow B(p_0)$ 
+ 5:    $\bar{z}_0 \leftarrow Au_0 - \sigma_0 p_0$ 
* 6:    $\bar{q}_0 \leftarrow B(\bar{z}_0)$ 
+ 7:    $\bar{w}_0 \leftarrow A\bar{q}_0$ 
8:   for  $i = 1, \dots, m$  do
9:     for  $j = 0, \dots, i - 2$  do
•10:       $\bar{h}_{j,i-1} \leftarrow \langle p_j, \bar{z}_{i-1} \rangle$ 
11:       $h_{j,i-1} \leftarrow \bar{h}_{j,i-1}$ 
•12:       $\bar{h}_{i-1,i-1} \leftarrow \langle p_{i-1}, \bar{z}_{i-1} \rangle$ 
13:       $h_{i-1,i-1} \leftarrow \bar{h}_{i-1,i-1} + \sigma_{i-1}$ 
•14:       $t \leftarrow \|\bar{z}_{i-1}\|_2^2 - \sum_{k=0}^{i-1} \bar{h}_{k,i-1}^2$ 
15:      if  $t < 0$  then
16:        BREAKDOWN
17:       $\bar{h}_{i,i-1} \leftarrow \sqrt{t}$ 
18:       $h_{i,i-1} \leftarrow \bar{h}_{i,i-1}$ 
19:      if  $i = m$  then break
-20:       $p_i \leftarrow (\bar{z}_{i-1} - \sum_{k=0}^{i-1} \bar{h}_{k,i-1} p_k) / \bar{h}_{i,i-1}$ 
-21:       $\tilde{u}_i \leftarrow (\bar{q}_{i-1} - \sum_{k=0}^{i-1} u_k \bar{h}_{k,i-1}) / \bar{h}_{i,i-1}$ 
-22:       $\bar{z}_i \leftarrow (\bar{w}_{i-1} - \sum_{k=0}^{i-1} (\bar{z}_k + \sigma_k p_k) \bar{h}_{k,i-1}) / \bar{h}_{i,i-1} - \sigma_i p_i$  ▷ see (11)
*23:       $\bar{q}_i \leftarrow B(\bar{z}_i)$ 
+24:       $\bar{w}_i \leftarrow A\bar{q}_i$ 
-25:    $y \leftarrow \operatorname{argmin} \|H_{m+1,m} y - \|r_0\|_2 e_0\|_2$ 
-26:    $x \leftarrow x_0 + \tilde{U}y$ 

```

Recall that $u_i \doteq B(p_i) = B((v_i - \sum_{k=1}^{i-1} h_{k,i-1} p_k) / h_{i,i-1})$. If B were linear, then the approximation \tilde{u}_i on line 21 would be arithmetically equivalent. PIPEFGMRES requires storage of $4m+2$ vectors, as opposed to $2m+2$ for FGMRES and CGFGMRES.

5. Implementation. As a practical consequence of our choice of pipelined algorithms which are not arithmetically equivalent to their base flexible methods, useful preconditioners for the base solver, for a given problem, will not necessarily translate into effective solvers with the pipelined variant. In particular, for the pipelined FCG and GCR methods, the preconditioner must be compatible with the approximation $\theta_i \equiv 1$, as described in section 2.2.1. Thus, experimentation is required, as it is in general when choosing preconditioners for complex systems where canonical solutions or analysis tools are not available.

To allow for this experimentation to be performed easily, and to allow for usage in highly distributed settings by way of MPI, we implement PIPEFCG (Algorithm 8), PIPEGCR (Algorithm 11), PIPEGCR.w (Algorithm 12), and PIPEFGMRES (Algorithm 15) within the KSP class in PETSc [3, 4]. The implementations are identified as KSPPIPEFCG, KSPPIPEGCR, and KSPPIPEFGMRES.⁴

An important practical point is that although many residual norms may be used to monitor convergence, only the natural norm for the CG- and CR-based solvers may be used without introducing extra computations and reductions. Similarly, the right

⁴See bitbucket.org/pascgeopc/petsc for current information.

preconditioned residual norm for PIPEFGMRES should be used when performance is important. Other norms can and should be used to assess convergence behavior.

All of the pipelined algorithms may suffer from norm breakdown, although in practice this is typically observed only near convergence. Our implementations detect norm breakdown and respond by restarting the algorithm, flushing the pipeline in the process.

Remark 6 (truncation strategies for FCG and GCR methods). In [31] Notay proposes a truncation–restart technique for determining the number of old directions ν_i to be used in the current iteration i as $\nu_i = \max(1, \text{mod}(i, \nu_{\max} + 1))$. With this technique, after every restart there are two consecutive iterations using only the respective last search direction. Consequently, the first two steps after a restart are actually standard CG iterations. In our implementations we implement the rule $\nu_i = \text{mod}(i, \nu_{\max}) + 1$. This employs $n + 1$ directions in the n th iteration after a restart. If $\nu_{\max} = 1$, we recover the IPCG algorithm as analyzed by Golub and Ye [15]. We also provide a standard truncation procedure, $\nu_i = \min(i, \nu_{\max})$.

Remark 7 (shift parameters for KSPPIPEFGMRES). Our implementation covers only the constant- σ case, as discussed in section 4.4.

Remark 8 (scaling). We observe that scaling the system matrix A is often essential to allow for a faithful preconditioner (see the survey by Wathen [40] for an exposition on scenarios which may otherwise profit from diagonal scaling). This is due to the fact that an effective preconditioner for A may not be faithful, as we have defined it in Definition 1. Indeed, a Krylov method will typically converge rapidly when the preconditioned operator has a small number of tight clusters of eigenvalues, a property which is not affected by scaling of the preconditioner or operator. Operators close to the identity frequently exhibit this property, but this property is neither necessary nor sufficient.

6. Numerical experiments.

6.1. Blocking and nonblocking reductions. The MPI-3 standard [29] provides an API for asynchronous operations which *may* allow for overlap of communication and computation, as required by the algorithms described in this paper. This is still an active area of software development [5, 21, 41]. The latency involved in a global reduction on a given massively parallel machine is difficult to characterize, particularly if using only a portion of the nodes, and performance models are not always predictive [22]. It should be noted that the only reliably reproducible quantity when testing message passing routines on complex machines is the *minimum* time required for an operation, in the limit of a large number of tests, though of course users are more affected by typical or average times [17].

Here and in the following tests, we report a variety of data and statistics to allow for *interpretability* [20] of our experiments, even in the absence of access to the specific hardware and software which were available to the authors at the time of this writing. Note that the performance model discussed in section 7 can also provide additional insight into the performance characteristics of the new solvers.

In order to assess the potential time savings and performance enhancements by using pipelined vs. nonpipelined preconditioned Krylov solvers, empirical data are presented, comparing the time to execute `MPI_Allreduce` and `MPI_Iallreduce` [29]. The reductions are paired with identical computational work local to each rank, consisting of arbitrary floating-point operations on individual elements of a small local array. In the former case we get the combined time of the sequence of local work and the blocking reduction, while in the latter case we time the sequence of starting the

reduction, doing the local work and finishing with `MPI.Wait`. The reduction size is 32 double values. Table 3 lists the timing results for Piz Daint, a Cray XC30 cluster at the Swiss National Supercomputing Centre. The machine includes 5272 nodes, each with an 8-core Intel Xeon E5-2670 CPU, 32 GB DDR3-1600 RAM, and Aries routing and communications ASIC. The cluster features the Dragonfly network topology [24]. As a reference we also include timings only for performing the reduction and the local work. We used five warmup rounds and 50 trial runs. The timings presented are the minimum and maximum times across ranks averaged over the trial runs.

It can be confirmed that `MPI.Iallreduce` functionality is in place. While the time for performing the local work and a blocking reduction just exceeds the cumulative time for the individual operations, the reduction time is hidden when using nonblocking reductions. From the min/max values we also see that usually timing is consistent across the communicator sizes we considered.

Table 3 is also instructive in understanding when a performance gain can be expected by using a pipelined solver, as these methods require computing and updating of additional intermediate variables.

TABLE 3

Timing for blocking and nonblocking reductions in milliseconds. We report both the minimum (upper rows) and maximum (lower rows) CPU times over all ranks within the MPI communicator. The times for the nonblocking reductions with local work have to be compared to only the maximum times of the local work, as the reduction can terminate only once all ranks finish with the local work.

#Ranks	256	1024	2048	4096	8192	16384	24576	32768
Reduction only	0.188	0.577	0.613	0.670	0.787	0.843	0.918	0.894
	0.198	0.588	0.622	0.689	0.821	0.876	0.956	0.929
Local work only	4.88	4.88	4.88	4.87	4.88	4.88	4.88	4.88
	4.95	4.95	4.97	4.97	4.98	5.68	5.67	5.68
+ blocking reduction	5.56	5.75	5.81	5.89	6.03	6.12	6.13	6.13
	5.57	5.76	5.83	5.90	6.05	6.14	6.15	6.15
+ nonblocking reduction	4.90	4.88	4.90	4.89	4.93	5.65	5.64	5.66
	4.98	4.98	4.99	4.99	4.98	5.71	5.69	5.71

6.2. Three-dimensional strong-scaled coarse grid solver example. A key motivation for the development of pipelined Krylov methods for use on current supercomputers is the scenario in which the amount of local work per iteration in a Krylov method is very low, and hence requiring time comparable to that consumed by reductions. This is the case when attempting to strong-scale past the point of the current reduction bottleneck. The same situation arises when a Krylov method is employed as a coarse grid solver within a multigrid hierarchy.

In the Piz Daint environment described in section 6.1, we use Cray CLE 5.2.40, Cray-MPICH 7.2.2, a PETSc development branch based on PETSc 3.6 `maint`,⁵ and a PTATIN3D [26] development branch. This allows us to evaluate our solvers with real application software to solve the variable viscosity Stokes equations using $Q_2 - P_1^{\text{disc}}$ mixed finite elements.

We examine a three-dimensional “viscous sinker” scenario with a spherical inclusion of higher density ($1.2\times$) and viscosity ($100\times$) in a cubic domain. This is a useful test case for challenging Stokes flow problems, as it presents a tunable, non-grid-aligned viscosity contrast. Tests correspond to a highly distributed coarse grid solve

⁵commit 2c5c660747c89d9a265191735ba48020b7a0dd72 at <https://bitbucket.org/petsc/petsc>.

or an extreme strong-scaling. The former case is a particularly interesting potential application of our methods, as it allows some adjustment of the grid size to balance communication and allows computation to be overlapped. 4096 MPI ranks on 1024 nodes hold a single Q_2 finite element for 16^3 elements and 107,811 degrees of freedom. The resulting linear systems are solved with the new KSPFCG and KSPPIPEFCG solvers to a relative tolerance of 10^{-8} in the natural residual norm. In each case, $\nu_{\max} = 30$ and standard truncation is employed. FCG and PIPEFCG are compared for two nonlinear preconditioners. Sixty tests were run for each of the two solves in each test, with 10 at a time per batch job. The first run is considered warmup and is reported separately. Sets of three batch jobs were submitted at a time.

6.2.1. Block Jacobi preconditioner. A per-rank block Jacobi preconditioner is employed, performing five iterations of Jacobi-preconditioned CG on each block. Figure 2 shows the residual norm convergence behavior as a function of the iteration count (left) and CPU time (right) for a typical experiment. Despite the increase in iterations required for convergence when using PIPEFCG, the time to solution is $\sim 2\times$ faster compared to FCG. Table 4 collects statistical variations obtained from multiple runs (indicated via the “Samples” column). The “First solve” column indicates whether the solve was the first call to the PETSc KSPSolve routine after defining the system; successive solves were run with no intermediate calls apart from those used for timing. Note the relatively large variability in timings. Interestingly, the maximum speedup per iteration can *exceed* $2\times$. While this might initially seem impossible to accomplish with an algorithm which aims to overlap two processes, it should be noted that there is also some benefit to be gained in relaxing the synchronization required, beneficial in the presence of heavy-tailed distributions of local processing times. For more, see the recent work by Morgan et al. [30].

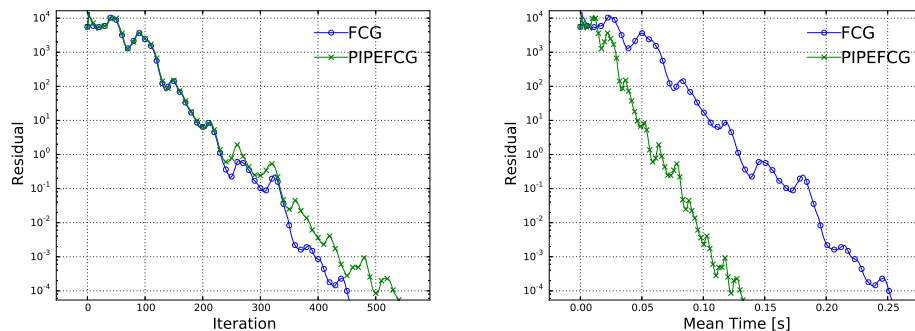


FIG. 2. Convergence behavior for the viscous sinker system using block Jacobi preconditioning.

6.2.2. RASM-1 preconditioner. A rankwise RASM-1 [7] preconditioner is used, with five iterations of Jacobi-preconditioned CG as a subdomain solver. Figure 3 shows the residual norm convergence behavior as a function of the iteration count (left) and mean CPU time (right) for a typical experiment. Note that, as expected, choosing a more faithful preconditioner allows recovery of comparable convergence behavior between the pipelined and base methods. Here, the speedup is not as pronounced as in the previous example, as the problem size was tuned as an example of a coarse grid solve with respect to the block Jacobi preconditioner. This example shows that speedup is practically achievable even given nested MPI calls defining neighbor-

TABLE 4

Statistics for multiple runs of the PTATIN3D viscous sinker system with a block Jacobi preconditioner.

Solver	First solve	Samples	Its.	Time / Krylov iteration [s]		
				Mean (std. dev.)	Min.	Max.
FCG	yes	6	454	6.38E-04 (1.38E-04)	4.48E-04	7.89E-04
PIPEFCG	yes	6	540	4.21E-04 (7.20E-05)	3.31E-04	5.41E-04
FCG	no	54	454	5.58E-04 (1.11E-04)	3.73E-04	9.13E-04
PIPEFCG	no	54	540	2.45E-04 (3.15E-05)	1.94E-04	3.73E-04

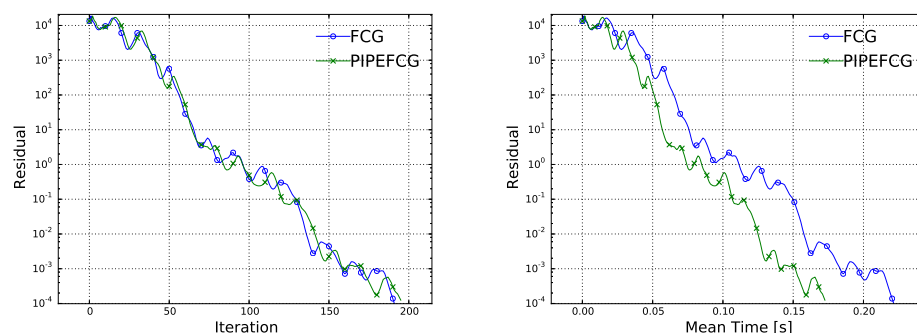


FIG. 3. Convergence behavior for the viscous sinker system using RASM-1 preconditioning.

wise communication patterns in the preconditioner. Table 5 presents statistics of the multiple solves performed.

6.3. Two-dimensional Stokes flow test. As a relevant yet still comparatively simple test case of a full solve, we ran extensive tests using a PETSc tutorial,⁶ which solves the incompressible, variable viscosity Stokes equation in two spatial dimensions. The discretization employs $Q_1 - Q_1$ elements, stabilized with Bochev's polynomial projection method [9] and free slip boundary conditions on all sides. The viscosity structure is a circular viscous sinker, with density contrast giving a buoyant force and a viscosity contrast of $25\times$. All tests were run on 4096 MPI ranks on 1024 nodes of Piz Daint. We experiment with all three new solvers as subsolvers within an upper-triangular block preconditioner. The discrete Stokes problem is of the form

$$\begin{bmatrix} K & G \\ D & C \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix},$$

which is right preconditioned with

$$B = \begin{bmatrix} \hat{K} & G \\ 0 & \hat{S} \end{bmatrix}^{-1},$$

where \hat{S} is a Jacobi preconditioner derived from the pressure mass matrix scaled by the local (elementwise) inverse of the viscosity, and \hat{K} is an inexact flexible pipelined Krylov solve applied to the viscous block of the Stokes system, preconditioned with

⁶www.mcs.anl.gov/petsc/petsc-current/src/ksp/ksp/examples/tutorials/ex43.c.html.

TABLE 5

Statistics for multiple runs of the PTATIN3D viscous sinker system with restricted ASM-1 preconditioner.

Solver	First solve	Samples	Its.	Time / Krylov iteration [s]		
				Mean (std. dev.)	Min.	Max.
FCG	yes	6	191	1.28E-03 (1.28E-04)	1.11E-03	1.41E-03
PIPEFCG	yes	6	195	1.21E-03 (1.18E-04)	1.10E-03	1.36E-03
FCG	no	54	191	1.16E-03 (1.48E-04)	9.08E-04	1.55E-03
PIPEFCG	no	54	195	8.85E-04 (9.89E-05)	6.99E-04	1.10E-03

a rankwise block Jacobi preconditioner with five iterations of Jacobi-preconditioned CG as subsolves. The convergence criterion is that the unpreconditioned norm of an outer FGMRES method is reduced by a factor of 10^6 . Inner solves are limited to 100 iterations. As they best exploit the structure of the suboperator, we would expect FCG and PIPEFCG to be more performant. Table 6 shows the resulting timings. We see significant speedup in all cases as the reduction bottlenecks in the base solvers are overcome. In the case where conjugate residuals are employed as inner solvers, we observe variability in the number of iterations to convergence, due to the degradation in conditioning arising from use of the $A^H A$ -inner product.

TABLE 6

Statistics for nine runs of a two-dimensional Stokes flow problem with an upper block triangular preconditioner employing various inner solvers, as described in section 6.3.

Inner solver	Outer its.	Solve time [s]		
		Mean (std. dev.)	Min.	Max.
FCG	8	3.80 (0.29)	3.47	4.37
PIPEFCG	8	1.80 (0.18)	1.52	2.03
GCR	18–23	10.02 (1.51)	8.26	12.76
PIPEGCR	12–16	3.08 (0.46)	2.59	4.01
FGMRES	19	6.70 (0.53)	6.10	7.71
PIPEFGMRES	19	5.04 (0.27)	4.73	5.65

7. Performance model and extrapolation to exascale. We describe performance models to predict the performance of our methods at exascale. For ease of comparison we consider the same hypothetical exaflop machine used to analyze pipelined GMRES [12], based on a 2010 report [36]. The specifications of the hypothetical machine are listed in Table 7. Rank-local computation times are obtained by counting flops and multiplying by t_c . This makes the significant assumption that any given computation can attain the peak flop rate of the machine, without being memory bandwidth limited.

Global reductions. Dot products are typically computed with a reduction tree of height $\lceil \log_r(P_n) \rceil$, with r the tree radix. Present day networks make use of a hierarchy of interconnects where more closely located nodes, e.g., in a cabinet, are connected through a tree with a high effective radix. A job scheduler typically assigns compute nodes, giving the user limited influence on the specific compute nodes used. Our performance model makes the simplifying assumption of a tree radix $r = 8$ for the entire machine [12, 36].

In order to obtain an expression for the global reduction communication time

TABLE 7
Properties of a hypothesized exascale machine.

Property	Symbol	Value
Nodes	P_n	2^{20}
Cores per node	C_n	2^{10}
Compute cores	$P_c = P_n C_n$	2^{30}
Word size	w	32 B
Machine node interconnect bandwidth	BW	100 GB/s
Tree radix	r	8
Per word transfer time	t_w	w/BW
Latency (startup) time	t_s	$1\ \mu\text{s}$
Time per flop	t_c	$2^{30}/10^{18}$

$T_{\text{red,comm}}$, we assume the model $t_s + mt_w$ consisting of a constant latency or startup time t_s and a message size-dependent part involving the number of words m and the time t_w required to exchange a word. We neglect time for intranode communication leading to the cost of $2\lceil\log_r(P_n)\rceil(t_s + mt_w)$ for a reduction and subsequent broadcast across the entire machine. In the context of pipelined methods, global reduction costs include only the excess time (if any) which is not overlapped by other work.

Global reductions also involve floating-point operations, accounted for in the computation time $T_{\text{red,calc}}$. Adding up the partial reduction results across the reduction tree leads to $\lceil\log_r(P_n)\rceil$ flops along the longest path, giving the cost function shown in Table 8. While this is the maximum number of flops performed by a single process, all other processes wait for this summation to finish. This might appear to be an insignificant contribution to the overall computational cost, but it can become nonnegligible in the strong-scaling limit.

Sparse matrix-vector products (SpMV). The cost of a sparse matrix-vector multiplication also includes both communication and computation. The system matrix is assumed to represent a stencil-based approximation of the underlying partial differential equations, with n_{nz} nonzero entries per row. For a standard second order finite difference discretization in three-dimensional space, for example, $n_{\text{nz}} = 7$. Denoting by $N = N_x N_y N_z$ the total problem size and by $n_{\text{loc}} = N/P_c$ the portion local to a process, we obtain the compute costs $T_{\text{SpMV,calc}} = 2n_{\text{nz}}n_{\text{loc}}t_c$. Neighbor communication costs are $T_{\text{SpMV,comm}} = 6(t_s + (N/P_n)^{2/3}t_w)$ and can mostly be overlapped. Consequently, the SpMV cost function is $T_{\text{SpMV}} = \max(T_{\text{SpMV,calc}}, T_{\text{SpMV,comm}})$.

TABLE 8
Communication time cost functions for fundamental operations with n_{loc} as defined above, n_{nz} the number of nonzero entries per row, and w the number of values (words) to be broadcast.

Method	Operation	Communication time cost function
RED	for $k = 1..w : \beta_k \leftarrow \langle p, u_k \rangle$	$T_{\text{red,comm}}(w) = 2\lceil\log_r(P_n)\rceil(t_s + wt_w)$
SpMV	$s \leftarrow Ap$	$T_{\text{SpMV,comm}} = 6(t_s + (N/P_n)^{2/3}t_w)$

Preconditioner application. We augment the model with the application of a preconditioner. We model the performance of the RASM preconditioner used in section 6, defining T_{PC} to account for five CG-Jacobi iterations on subdomains overlapping by one degree of freedom and for an internode communication step with the same pattern as an SpMV. We note that scalable multigrid preconditioners may also be designed which do not involve any reductions [27] and thus could be attractive for

extreme-scale solves. Nested pipelined Krylov methods could be used as coarse grid solvers within these preconditioners.

Total cost per iteration. The total cost of one iteration comprises the time for performing flops, T_{calc} , involving the cost of sparse matrix-vector products, T_{SpMV} ; the time for applying the preconditioner, T_{PC} , and fundamental operations listed in Table 9; and the (excess of the) reduction communication time, $T_{\text{red,comm}}$. We do not take into account fused operations and vector units to establish an easier comparison with previous work [12].

The cost per iteration for flexible methods depends on the number of previous directions to orthogonalize against and hence on the truncation strategy and the iteration count. In our model we use a factor representing an average number of previous directions being used in the Gram–Schmidt orthogonalization as $\nu_{\text{avg}} = k_{\text{avg}} \cdot \nu_{\text{max}}$. Operations involving scalar values only are not taken into account. This leads to the models for the single iteration cost functions listed in Table 10.

TABLE 9

Compute time cost functions for fundamental operations, with $n_{\text{loc}} = N/P_c$ denoting the vector size local to the process.

Method	Operation	Compute time cost function
AXPY	$x \leftarrow x + \alpha p$	$T_{\text{AXPY}} = t(n_{\text{loc}}(\text{mlt}) + n_{\text{loc}}(\text{add})) = 2n_{\text{loc}}t_c$
MAXPY	$p \leftarrow u + \sum_m \beta_m p_m$	$T_{\text{MAXPY}}(m) = t(mn_{\text{loc}}(\text{mlt}) + (m-1)n_{\text{loc}}(\text{add}) + n_{\text{loc}}(\text{add})) = 2mn_{\text{loc}}t_c$
RED	$\langle r, u \rangle$	$T_{\text{red,calc}} = t(n_{\text{loc}}(\text{mlt}) + (n_{\text{loc}}-1)(\text{add}) + \lceil \log_r(P_n) \rceil (\text{add})) \approx (2n_{\text{loc}} + \lceil \log_r(P_n) \rceil)t_c$

Problem and solver specifications. For evaluating the cost functions, we consider a strong-scaling experiment (cf. [12]), with a total problem size $N = 2000^3$ and $n_{\text{nz}} = 7$ nonzero entries per row, and evaluate for a varying number of node counts. For PIPEFCG and PIPEGCR we assume $\nu_{\text{max}} = 30$ and $k_{\text{avg}} = 0.8$; for PIPEFGMRES we assume a restart parameter of 30.

Evaluation. In Figure 4 we show a comparison of the anticipated performance of all solvers for this problem. Eventually all methods are limited by latency time, but the graphs indicate that the pipelined methods scale up to the 10^6 nodes of the assumed exascale machine, while their standard counterparts level off about one order of magnitude earlier. The performance crossover regarding pipelined vs. non-pipelined methods for this particular example occurs around $8 \cdot 10^4$ nodes. The example of PIPEGCR vs. PIPEGCR_w clearly shows the effect of trading local work load for better overlapping. Note how PIPEGCR (explicit computation of w , overlapping by preconditioner application) performs slightly better at lower node counts, while PIPEGCR_w (recurrence of w , additional AXPY and MAXPY, overlapping by preconditioner and SpMV) shows a more significant gain in efficiency for larger numbers of nodes.

Figure 5 shows a breakdown of the anticipated total iteration time for FCG and PIPEFCG. For the test case considered here the model predicts that global reductions will be entirely overlapped up to approximately $5 \cdot 10^5$ nodes.

Sensitivity to parameters. The performance models are insensitive to the specific value of most of the parameters when chosen within reasonable bounds. Among the insensitive parameters are the interconnect bandwidth, word size, the tree radix, and assumptions such as ignoring fused AXPY operations. It may appear unintuitive that the interconnect bandwidth is not a sensitive parameter, but latency is dominant in the communication time model.

TABLE 10

Performance models for standard and pipelined methods for the truncation strategy. Again, $n = N/P_c$, $\nu_{\text{avg}} = k_{\text{avg}} \cdot \nu_{\text{max}}$, where k_{avg} corresponds to a factor accounting for the average number of previous search directions being used. If the number of iterations required is $\gg \nu_{\text{max}}$, this factor approaches one for a standard truncation strategy. Communication time for multiple reductions is fused whenever possible. GMRES-type models do not include the time for manipulating the Hessenberg matrix.

Krylov method	Total cost
FCG	$T_{\text{calc}} = 2T_{\text{AXPY}} + T_{\text{MAXPY}}(\nu_{\text{avg}}) + (\nu_{\text{avg}} + 2)T_{\text{red,calc}} + T_{\text{SpMV}} + T_{\text{PC}}$ $T_{\text{red}} = T_{\text{red,comm}}(\nu_{\text{avg}} + 1) + T_{\text{red,comm}}(1)$
PIPEFCG	$T_{\text{calc}} = 4T_{\text{AXPY}} + 4T_{\text{MAXPY}}(\nu_{\text{avg}}) + (\nu_{\text{avg}} + 2)T_{\text{red,calc}} + T_{\text{SpMV}}$ $+ T_{\text{PC}} + 2nt_c$ $T_{\text{red}} = \max(0, T_{\text{red,comm}}(\nu_{\text{avg}} + 2) - [T_{\text{SpMV}} + T_{\text{PC}} + 2nt_c])$
GCR	$T_{\text{calc}} = 2T_{\text{AXPY}} + T_{\text{MAXPY}}(\nu_{\text{avg}}) + (\nu_{\text{avg}} + 2)T_{\text{red,calc}} + T_{\text{SpMV}} + T_{\text{PC}}$ $T_{\text{red}} = T_{\text{red,comm}}(\nu_{\text{avg}}) + T_{\text{red,comm}}(2)$
PIPEGCR	$T_{\text{calc}} = 3T_{\text{AXPY}} + 3T_{\text{MAXPY}}(\nu_{\text{avg}}) + (\nu_{\text{avg}} + 2)T_{\text{red,calc}} + T_{\text{SpMV}}$ $+ T_{\text{PC}} + 2nt_c$ $T_{\text{red}} = \max(0, T_{\text{red,comm}}(\nu_{\text{avg}} + 2) - [T_{\text{PC}} + 2nt_c])$
PIPEGCR_w	$T_{\text{calc}} = 4T_{\text{AXPY}} + 4T_{\text{MAXPY}}(\nu_{\text{avg}}) + (\nu_{\text{avg}} + 2)T_{\text{red,calc}} + T_{\text{SpMV}}$ $+ T_{\text{PC}} + 2nt_c$ $T_{\text{red}} = \max(0, T_{\text{red,comm}}(\nu_{\text{avg}} + 2) - [T_{\text{SpMV}} + T_{\text{PC}} + 2nt_c])$
FGMRES	$T_{\text{calc}} = T_{\text{MAXPY}}(\nu_{\text{avg}}) + t_cn + (\nu_{\text{avg}} + 1)T_{\text{red,calc}} + T_{\text{SpMV}} + T_{\text{PC}}$ $T_{\text{red}} = T_{\text{red,comm}}(\nu_{\text{avg}}) + T_{\text{red,comm}}(1)$
PIPEFGMRES	$T_{\text{calc}} = \nu_{\text{avg}}T_{\text{AXPY}} + 3T_{\text{MAXPY}}(\nu_{\text{avg}}) + (\nu_{\text{avg}} + 5)t_cn + (\nu_{\text{avg}} + 2)T_{\text{red,calc}}$ $+ T_{\text{SpMV}} + T_{\text{PC}}$ $T_{\text{red}} = \max(0, T_{\text{red,comm}}(\nu_{\text{avg}} + 2) - [T_{\text{SpMV}} + T_{\text{PC}}])$

The models are most obviously sensitive to the latency time but also to the number of cores per node, as more or fewer cores per node implies a different number of nodes for obtaining exaflop performance. Another sensitive parameter is the total problem size N . A mild dependency of the average fill factor k_{avg} , the maximum number of previous directions ν_{max} , and the number of nonzero entries n_{nz} per row is observed. Varying any of these values does not cause a change of the general shape of the curves but rather induces a shift of the crossover point.

For the hypothesized exascale machine, a latency time of $1\mu s$ is assumed, which is significantly lower than times observed on current systems, including Piz Daint, as in Table 3. Consequently, pipelined Krylov methods can lead to significant performance improvements for much smaller numbers of nodes on current systems, as confirmed by the tests in section 6.

8. Conclusions and outlook. Bottlenecks related to global communication and synchronization are increasingly common in high performance computing, as parallelism and hybridism are used to increase peak performance. Traditional algorithms can often be improved with this consideration in mind. Given the increasing relative impact of latency and synchronization required with global all-to-all communication patterns, it becomes increasingly attractive to tailor algorithms to mitigate these costs. This requires introducing additional costs in terms of memory footprint, local computational work, memory traffic, numerical stability, and allowing asynchronous operations.

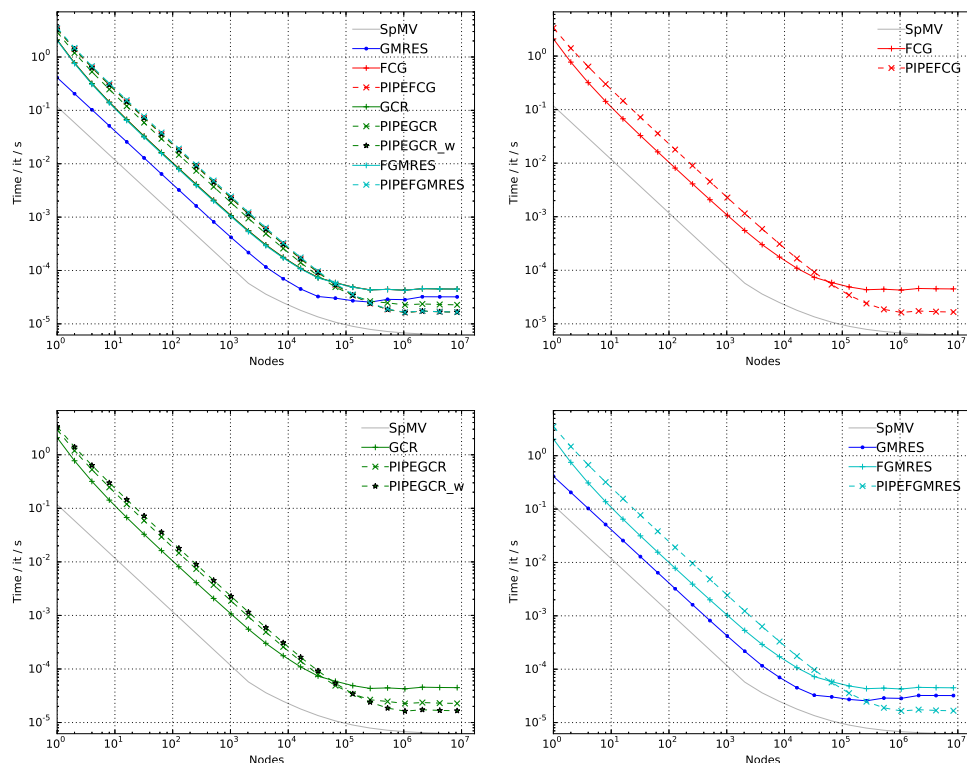


FIG. 4. Evaluation of the performance models of Table 10 for predicting iteration times on the hypothesized exascale machine of Table 7. The top left panel includes all methods, and the others compare FCG-, GCR-, and GMRES-type methods. All graphs include the time for performing an SpMV operation as a reference. The number of nodes scales beyond the 1M nodes of the assumed machine to better depict the leveling at high node counts. While all methods are eventually limited by latency time, the pipelined methods outperform their standard counterparts on high node counts.

We introduce variants of the FCG, GCR, and FGMRES flexible Krylov methods, allowing overlap of global reductions with other work through operation pipelining. The base methods are among the most commonly used flexible Krylov methods for Hermitian positive definite, Hermitian, and general operators, respectively. Future work includes extending and implementing the approaches here for other useful Krylov methods. A crucial consideration in this process, as demonstrated here, is the effect of allowing algorithmic rearrangements which modify the effective preconditioner. Naive rearrangement is not always effective, and effective rearrangements may impose conditions on the class of useful preconditioners beyond those common in previous Krylov methods. We propose and analyze an approach to provide methods which are increasingly effective for stronger preconditioners.

The proposed methods are shown to be effective on a current leadership supercomputing platform with a fast network, where speedups could exceed $2\times$, challenging common assumptions that pipelined Krylov methods are only effective at future machine scales, and that the speedup of single-stage pipelining is limited to $2\times$. Additionally, we develop analytic models for anticipating performance on future exascale systems.

All methods have been implemented in the PETSc package, and open-source software is available for all methods.

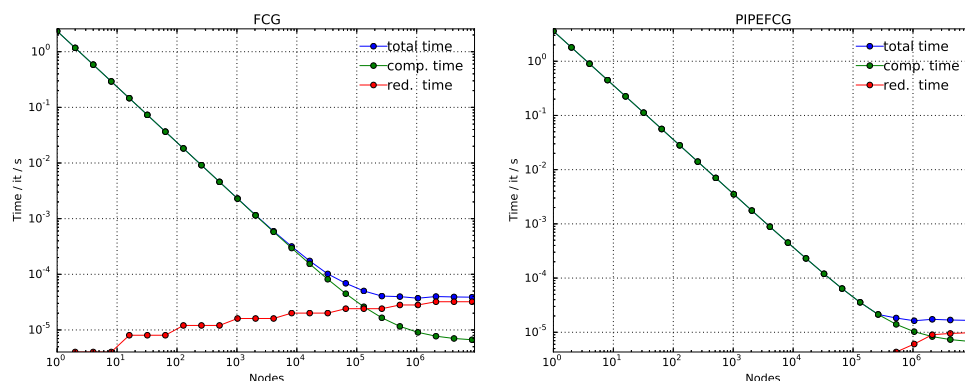


FIG. 5. Breakdown of the total iteration time into computation and reduction times for FCG and PIPEFCG. For the test case considered, it is anticipated that the reductions will be entirely overlapped up to approximately $5 \cdot 10^5$ nodes for the pipelined method.

Deeper pipelining. The same rearrangements used in the single-stage pipelined methods described here could be pursued to allow a reduction to overlap more than one iteration's worth of other work. This has been explored for the GMRES method [12]. Our focus has been on the “faithful” preconditioners discussed above, which require a substantial amount of work to apply and likely would not benefit from further pipelining. Additionally, numerical instability becomes more of a concern with deeper pipelining—in the case of GMRES, norm breakdown becomes more common—and we anticipate that the required stabilization for nonlinear preconditioners would also be nontrivial. Investigation of deeper pipelining is of course of interest from a mathematical point of view, but it is unclear whether exascale systems will be developed exhibiting the extreme reduction latency required to justify overlapping inner products with multiple nonlinear preconditioner applications. From a practical point of view, current MPI implementations do not prioritize operations like multiple overlapping reductions, which makes the testing of these methods problematic.

Outlook. The methods described in this paper are expected to become more and more relevant as nondeterministic, randomized, nested, highly distributed, and nonlinear preconditioning techniques come into greater use. These trends are encouraged by an increasing level of parallelism, hybridism, and hierarchy in computational machinery, as well as intense research into randomized, finely grained parallel, and asynchronous techniques for approximate and exact solvers.

REFERENCES

- [1] O. AXELSSON, *A generalized conjugate gradient, least square method*, Numer. Math., 51 (1987), pp. 209–227, doi:10.1007/BF01396750.
- [2] O. AXELSSON AND P. S. VASSILEVSKI, *A black box generalized conjugate gradient solver with inner iterations and variable-step preconditioning*, SIAM J. Matrix Anal. Appl., 12 (1991), pp. 625–644, doi:10.1137/0612048.
- [3] S. BALAY, S. ABHYANKAR, M. F. ADAMS, J. BROWN, P. BRUNE, K. BUSCHELMAN, L. DALCIN, V. EIJKHOUT, W. D. GROPP, D. KAUSHIK, M. G. KNEPLEY, L. C. MCINNES, K. RUPP, B. F. SMITH, S. ZAMPINI, AND H. ZHANG, *PETSc Users Manual*, Tech. Report ANL-95/11, Revision 3.6, Argonne National Laboratory, 2015, <http://www.mcs.anl.gov/petsc>.
- [4] S. BALAY, S. ABHYANKAR, M. F. ADAMS, J. BROWN, L. DALCIN, T. ISAAC, M. G. KNEPLEY, K. RUPP, J. SARICH, B. F. SMITH, S. ZAMPINI, AND H. ZHANG, *Portable, Extensible Toolkit for Scientific Computation (PETSc), Version 3.7*, 2016, <http://www.mcs.anl.gov/petsc>.

- [5] M. T. BRUGGENCATE AND D. ROWETH, *DMAPP—an API for one-sided program models on Baker systems*, in Proceedings of the Cray User Group Conference, 2010, pp. 1–7, https://cug.org/5-publications/proceedings.attendee.lists/CUG10CD/pages/1-program/final_program/CUG10_Proceedings/pages/authors/01-5Monday/03B-tenBruggencate-Paper-2.pdf.
- [6] P. R. BRUNE, M. G. KNEPLEY, B. F. SMITH, AND X. TU, *Composing scalable nonlinear algebraic solvers*, SIAM Rev., 57 (2015), pp. 535–565, doi:10.1137/130936725.
- [7] X.-C. CAI AND M. SARKIS, *A restricted additive Schwarz preconditioner for general sparse linear systems*, SIAM J. Sci. Comput., 21 (1999), pp. 792–797, doi:10.1137/S106482759732678X.
- [8] A. T. CHRONOPOULOS AND C. W. GEAR, *s-step iterative methods for symmetric linear systems*, J. Comput. Appl. Math., 25 (1989), pp. 153–168, doi:10.1016/0377-0427(89)90045-9.
- [9] C. R. DOHRMANN AND P. B. BOCHEV, *A stabilized finite element method for the Stokes problem based on polynomial pressure projections*, Int. J. Numer. Methods Fluids, 46 (2004), pp. 183–201, doi:10.1002/flid.752.
- [10] S. C. EISENSTAT, H. C. ELMAN, AND M. H. SCHULTZ, *Variational iterative methods for non-symmetric systems of linear equations*, SIAM J. Numer. Anal., 20 (1983), pp. 345–357, doi:10.1137/0720023.
- [11] L. FOX, H. D. HUSKEY, AND J. H. WILKINSON, *Notes on the solution of algebraic linear simultaneous equations*, Quart. J. Mech. Appl. Math., 1 (1948), pp. 149–173, doi:10.1093/qjmam/1.1.149.
- [12] P. GHYSELS, T. J. ASHBY, K. MEERBERGEN, AND W. VANROOSE, *Hiding global communication latency in the GMRES algorithm on massively parallel machines*, SIAM J. Sci. Comput., 35 (2013), pp. C48–C71, doi:10.1137/12086563X.
- [13] P. GHYSELS AND W. VANROOSE, *Hiding global synchronization latency in the preconditioned conjugate gradient algorithm*, Parallel Comput., 40 (2014), pp. 224–238, doi:10.1016/j.parco.2013.06.001.
- [14] G. H. GOLUB AND D. P. O’LEARY, *Some history of the conjugate gradient and Lanczos algorithms: 1948–1976*, SIAM Rev., 31 (1989), pp. 50–102, doi:10.1137/1031003.
- [15] G. H. GOLUB AND Q. YE, *Inexact preconditioned conjugate gradient method with inner-outer iteration*, SIAM J. Sci. Comput., 21 (1999), pp. 1305–1320, doi:10.1137/S1064827597323415.
- [16] W. GROPP, *Update on Libraries for Blue Waters*, Joint Laboratory for PetaScale Computation, Bordeaux, France, 2010, <http://jointlab-pc.ncsa.illinois.edu/events/workshop3/pdf/presentations/Gropp-Update-on-Libraries.pdf>.
- [17] W. GROPP AND E. L. LUSK, *Reproducible measurements of MPI performance characteristics*, in Proceedings of the 6th European PVM/MPI Users’ Group Meeting on Recent Advances in Parallel Virtual Machine and Message Passing Interface, Lecture Notes in Comput. Sci. 1697, Springer, Berlin, 1999, pp. 11–18, doi:10.1007/3-540-48158-3_2.
- [18] M. R. HESTENES, *The conjugate-gradient method for solving linear systems*, in Proceedings of the Sixth Symposium in Applied Mathematics, McGraw-Hill, New York, 1956, pp. 83–102.
- [19] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436.
- [20] T. HOEFLER AND R. BELLI, *Scientific benchmarking of parallel computing systems: Twelve ways to tell the masses when reporting performance results*, in Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC ’15, ACM, New York, 2015, pp. 73:1–73:12, doi:10.1145/2807591.2807644.
- [21] T. HOEFLER AND A. LUMSDAINE, *Message progression in parallel computing—to thread or not to thread?*, in Proceedings of the 2008 IEEE International Conference on Cluster Computing, IEEE Press, Piscataway, NJ, 2008, pp. 213–222, doi:10.1109/CLUSTER.2008.4663774.
- [22] T. HOEFLER, T. SCHNEIDER, AND A. LUMSDAINE, *Accurately measuring collective operations at massive scale*, in IPDPS Miami 2008—Proceedings of the IEEE International Symposium on Parallel and Distributed Processing, IEEE Press, Piscataway, NJ, 2008, pp. 1–8, doi:10.1109/IPDPS.2008.4536494.
- [23] M. HOEMMEN, *Communication-Avoiding Krylov Subspace Methods*, Ph.D. thesis, University of California at Berkeley, 2010.
- [24] J. KIM, W. J. DALLY, S. SCOTT, AND D. ABTS, *Technology-driven, highly-scalable dragonfly topology*, in Proceedings of the 35th International Symposium on Computer Architecture (ISCA), IEEE Press, Piscataway, NJ, 2008, pp. 77–88, doi:10.1109/ISCA.2008.19.
- [25] D. G. LUENBERGER, *The conjugate residual method for constrained minimization problems*, SIAM J. Numer. Anal., 7 (1970), pp. 390–398, doi:10.1137/0707032.
- [26] D. A. MAY, J. BROWN, AND L. LE POURHET, *pTatin3D: High-performance methods for long-term lithospheric dynamics*, in Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC ’14), IEEE Press, Piscataway, NJ, 2014, pp. 274–284, doi:10.1109/SC.2014.28.

- [27] D. A. MAY, J. BROWN, AND L. LE POURHET, *A scalable, matrix-free multigrid preconditioner for finite element discretizations of heterogeneous Stokes flow*, Comput. Methods Appl. Mech. Engrg., 290 (2015), pp. 496–523, doi:10.1016/j.cma.2015.03.014.
- [28] L. C. MCINNES, B. SMITH, H. ZHANG, AND R. T. MILLS, *Hierarchical Krylov and nested Krylov methods for extreme-scale computing*, Parallel Comput., 40 (2014), pp. 17–31, doi:10.1016/j.parco.2013.10.001.
- [29] MESSAGE PASSING INTERFACE FORUM, *MPI: A Message-Passing Interface standard, Version 3.0*, 2012, <http://www.mpi-forum.org/docs/mpi-3.0/mpi30-report.pdf>.
- [30] H. MORGAN, M. G. KNEPLEY, P. SANAN, AND L. R. SCOTT, *A stochastic performance model for pipelined Krylov methods*, Concurrency and Computation: Practice and Experience, published online March 31, 2016, doi:10.1002/cpe.3820.
- [31] Y. NOTAY, *Flexible conjugate gradients*, SIAM J. Sci. Comput., 22 (2000), pp. 1444–1460, doi:10.1137/S1064827599362314.
- [32] K. RUPP, J. WEINBUB, A. JÜNGEL, AND T. GRASSER, *Pipelined Iterative Solvers with Kernel Fusion for Graphics Processing Units*, preprint, arXiv:1410.4054, 2014.
- [33] Y. SAAD, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469, doi:10.1137/0914028.
- [34] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003, doi:10.1137/1.9780898718003.
- [35] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869, doi:10.1137/0907058.
- [36] J. SHALF, S. DOSANJH, AND J. MORRISON, *Exascale computing technology challenges*, in High Performance Computing for Computational Science (VECPAR), Lecture Notes in Comput. Sci. 6449, Springer, Berlin, 2011, pp. 1–25, doi:10.1007/978-3-642-19328-6_1.
- [37] J. R. SHEWCHUK, *An Introduction to the Conjugate Gradient Method without the Agonizing Pain*, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1994, <http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf>.
- [38] B. SMITH, P. BJORSTAD, AND W. GROPP, *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press, Cambridge, UK, 2004.
- [39] H. VAN DER VORST, *Iterative Krylov Methods for Large Linear Systems*, Cambridge University Press, Cambridge, UK, 2003.
- [40] A. J. WATHEN, *Preconditioning*, Acta Numer., 24 (2015), pp. 329–376, doi:10.1017/S0962492915000021.
- [41] M. WITTMANN, G. HAGER, T. ZEISER, AND G. WELLEIN, *Asynchronous MPI for the Masses*, preprint, arXiv:1302.4280, 2013.