

POSTPRINT

Task-based, GPU-accelerated and Robust Library for Solving Dense Nonsymmetric Eigenvalue Problems

Mirko Myllykoski* | Carl Christian Kjelgaard Mikkelsen

Department of Computing Science and HPC2N, Umeå University, SE-901 87 Umeå, Sweden

Correspondence

*Mirko Myllykoski, Department of Computing Science and HPC2N, Umeå University, SE-901 87 Umeå, Sweden.
Email: mirkom@cs.umu.se

Notice

This is the peer reviewed version of the following article: M. Myllykoski, C.C. Kjelgaard Mikkelsen, *Task-based, GPU-accelerated and Robust Library for Solving Dense Nonsymmetric Eigenvalue Problems*, *Concurrency and Computation: Practice and Experience*, 2020, which has been published in final form at <https://doi.org/10.1002/cpe.5915>. Published under Attribution 4.0 International (CC BY 4.0) license.

Summary

In this paper, we present the StarNEig library for solving dense nonsymmetric standard and generalized eigenvalue problems. The library is built on top of the StarPU runtime system and targets both shared and distributed memory machines. Some components of the library have support for GPU acceleration. The library currently applies to real matrices with real and complex eigenvalues and all calculations are done using real arithmetic. Support for complex matrices is planned for a future release. This paper is aimed at potential users of the library. We describe the design choices and capabilities of the library, and contrast them to existing software such as LAPACK and ScaLAPACK. StarNEig implements a ScaLAPACK compatibility layer which should assist new users in the transition to StarNEig. We demonstrate the performance of the library with a sample of computational experiments.

KEYWORDS:

eigenvalue problem, parallel computing, task-based, numerical library

1 | INTRODUCTION

In this paper, we present the StarNEig library¹ for solving dense nonsymmetric standard and generalized eigenvalue problems. StarNEig differs from existing libraries such as LAPACK² and ScaLAPACK³ by relying on a modern task-based approach (see, for example, Reference⁴) in a manner similar to the already well-established PLASMA library⁵. Specifically, StarNEig is built on top of the StarPU runtime system⁶ developed by the StarPU team at INRIA. This allows StarNEig to target both shared memory and distributed memory machines. Furthermore, some components of StarNEig have support for GPU acceleration. The library is under continuous development. Currently, StarNEig applies to real matrices with real and complex eigenvalues and all calculations are done using real arithmetic.

This paper is addressed to potential users of the StarNEig library. We hope that readers, who are already familiar with ScaLAPACK, will be able to decide if StarNEig is suitable for them. In particular, we wish to communicate the changes needed to integrate existing ScaLAPACK style (or LAPACK style) software with StarNEig. Central to this integration is the ScaLAPACK compatibility layer implemented in StarNEig. This compatibility layer allows users to keep their existing two-dimensional block cyclic distribution of the data and call StarNEig routines directly to perform the computations. The authors hope to start a discussion which will help guide and prioritize the future development of the library.

Dense nonsymmetric eigenvalue problems are usually solved using three phases: reduction to Hessenberg form, reduction to (real) Schur form and computation of eigenvectors. Additionally, a fourth phase, called eigenvalue reordering, can be performed to acquire an invariant subspace that is associated with a given subset of eigenvalues. StarNEig provides shared memory task-based implementations of all four phases for the standard eigenvalue problem while some phases are incomplete for the generalized eigenvalue problem and for distributed memory computations (see Table 2 in Section 5). Performance-wise, the Hessenberg reduction phase in StarNEig is comparable to the LAPACK, ScaLAPACK and MAGMA⁷ libraries, while the Schur reduction and the reordering phases in StarNEig are significantly faster than the ScaLAPACK implementations. Moreover, StarNEig can compute the eigenvectors directly from any real Schur form without suffering from floating-point overflow, i.e., the implementation is robust. The latter functionality simply does not exist in ScaLAPACK and the implementation in StarNEig is significantly faster than the LAPACK implementation in both sequential and parallel settings (parallel BLAS). We refer the reader to the NLA-FET Deliverable Report D2.7⁸ for comprehensive performance and accuracy evaluations.

The rest of this paper is organized as follows: Section 2 discusses existing libraries that provide functionality for solving dense eigenvalue problems. Section 3 provides a brief summary of the solution of dense nonsymmetric eigenvalue problems using the four phases mentioned earlier. Section 4 introduces the task-based approach and explains why the task-based approach can lead to superior performance when compared to older, well-established techniques. Section 5 introduces the reader to some of the inner workings of StarNEig. In particular, the current state of the library and its limitations are explained in this section. Section 6 presents a sample of computational results which demonstrate the expected performance of StarNEig in both shared and distributed memory, with and without GPUs. Finally, Section 7 concludes the paper.

Remark 1. This paper is an extended version of a conference paper by the authors⁹. The original conference paper only discussed extensively two of the four phases (Schur reduction and eigenvalue reordering). This extended paper discusses all four phases. We have also significantly extended the part of the paper that discusses the Schur reduction phase. The added content includes several new illustrations and computational results. The major contribution of this paper is the extended description of the library and its capabilities/limitations.

2 | RELATED WORK

TABLE 1 A comparison between libraries that provide functionality for solving dense eigenvalue problems. From left to right: complex input and output matrices supported, nonsymmetric eigenvalue problems supported, generalized eigenvalue problems supported, multiple CPU cores supported, distributed memory supported, and GPU accelerators supported. IC stands for "Incomplete".

Library	Complex	Nonsymmetric	Generalized	Multicore	Distributed memory	GPUs
LAPACK ²	Yes	Yes	Yes	.	.	.
MAGMA ⁷	Yes	Yes	Yes	IC	.	IC
ScaLAPACK ³	IC	IC	IC	Yes	Yes	.
Elemental ¹⁰	IC	IC	IC	Yes	Yes	.
ELPA ¹¹	Yes	.	Yes	Yes	Yes	.
EigenEXA ¹²	.	.	Yes	Yes	Yes	.
PLASMA ⁵	Yes	.	Yes	Yes	.	.
StarNEig (v0.1.1)	.	Yes	IC	Yes	IC	IC
StarNEig (planned)	Yes	Yes	Yes	Yes	Yes	Yes

Many established and emerging libraries include software for solving dense eigenvalue problems. The most well-known and significant libraries are summarized in Table 1. In particular, the libraries that provide functionality that is comparable to StarNEig are the following:

LAPACK (Linear Algebra PACKage). This is the de facto standard library for dense numerical linear algebra using shared memory machines. LAPACK contains a full software stack for solving dense standard and generalized, symmetric and

nonsymmetric eigenvalue problems. Some LAPACK routines can be parallelized by compiling LAPACK with a parallel BLAS library. This works well for the Hessenberg reduction phase, but poorly for the three remaining phases.

MAGMA (Matrix Algebra on GPU and Multicore Architectures). The MAGMA project seeks to develop a new generation of linear algebra libraries which target heterogeneous computer architectures, starting with current multicore and multi-GPU systems. Currently, MAGMA offers only very limited parallel and GPU-accelerated functionality for dense nonsymmetric eigenvalue problems (Hessenberg reduction) and offloads missing routines to LAPACK.

ScaLAPACK (Scalable LAPACK). This is the de facto standard library for numerical linear algebra using distributed memory machines. ScaLAPACK does not contain a full stack for solving dense eigenvalue problems. In particular, there are no routines for computing the eigenvectors from real Schur forms, and the routines for computing eigenvectors from complex Schur forms are all vulnerable to floating-point overflow.

Elemental. The Elemental project sought to provide a distributed memory and arbitrary precision library for dense and sparse-direct linear algebra, conic optimization, and lattice reduction. The library was never completed and the project was abandoned in 2016. Elemental offloads the Hessenberg and Schur reduction phases to LAPACK and ScaLAPACK. The library contains a multi-shift triangular solver that can be used to compute eigenvectors, but the overflow protection is incomplete.

The SLATE (Software for Linear Algebra Targeting Exascale)¹³ library is being developed to provide software for dense numerical linear algebra on current and future distributed computer systems. Currently, SLATE implements a set of parallel basic linear algebra subroutines (parallel BLAS), as well as high level subroutines for solving linear systems and linear least square problems. The authors state that future work will include nonsymmetric eigenvalue problems¹³.

In summary, LAPACK is the *only* library that can compute eigenvalues and eigenvectors for symmetric and nonsymmetric matrices and matrix pairs. The more recent libraries only support the symmetric eigenvalue problems and the other libraries are incomplete.

3 | SOLUTION OF DENSE NONSYMMETRIC EIGENVALUE PROBLEMS

Given a matrix $A \in \mathbb{R}^{n \times n}$, the standard eigenvalue problem consists of computing eigenvalues $\lambda_i \in \mathbb{C}$ and matching eigenvectors $x_i \in \mathbb{C}^n$, $x_i \neq 0$, such that

$$Ax_i = \lambda_i x_i. \quad (1)$$

Similarly, given matrices $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times n}$ the generalized eigenvalue problem for the matrix pair (A, B) consists of computing generalized eigenvalues $\lambda_i \in \mathbb{C} \cup \{\infty\}$ and matching generalized eigenvectors $x_i \in \mathbb{C}^n$, $x_i \neq 0$, such that

$$Ax_i = \lambda_i Bx_i. \quad (2)$$

If the matrices A and B are sparse, then the well-known SLEPc library¹⁴ is one of the better tools for solving the eigenvalue problems (1) and (2). Similarly, if the matrices A and B are symmetric, then algorithms and software that take advantage of the symmetry are preferred (see the libraries listed in Table 1). Otherwise, if the matrices are both *dense* and *nonsymmetric*, then the route of acquiring the (generalized) eigenvalues and the (generalized) eigenvectors usually includes the following *three phases*:

Hessenberg(-triangular) reduction: The matrix A or the matrix pair (A, B) is reduced to Hessenberg form H or Hessenberg-triangular form (H, R) by a similarity transformation

$$A = Q_1 H Q_1^T \text{ or } (A, B) = Q_1 (H, R) Z_1^T, \quad (3)$$

where H is upper Hessenberg, R is a upper triangular, and $Q_1, Z_1 \in \mathbb{R}^{n \times n}$ are orthogonal matrices[†].

Schur reduction: The Hessenberg matrix H or the Hessenberg-triangular matrix pair (H, R) is reduced to real Schur form S or generalized real Schur form (S, T) by a similarity transformation

$$H = Q_2 S Q_2^T \text{ or } (H, R) = Q_2 (S, T) Z_2^T, \quad (4)$$

[†]For $X, Y, Q, Z \in \mathbb{R}^{n \times n}$, we define $Q(X, Y) Z^T = (QXZ^T, QYZ^T)$.

where S is upper quasi-triangular with 1×1 and 2×2 blocks on the diagonal, T is upper triangular, and $Q_2, Z_2 \in \mathbb{R}^{n \times n}$ are orthogonal matrices. The eigenvalues or the generalized eigenvalues can be determined from the diagonal blocks of S or (S, T) . In particular, the 2×2 blocks on the diagonal of S correspond to the complex conjugate pairs of (generalized) eigenvalues.

Eigenvectors: Finally, we solve for vectors $y_i \in \mathbb{C}^n$ from

$$(S - \lambda_i I)y_i = 0 \quad \text{or} \quad (S - \lambda_i T)y_i = 0 \quad (5)$$

and backtransform to the original basis by

$$x_i = Q_1 Q_2 y_i \quad \text{or} \quad x_i = Z_1 Z_2 y_i. \quad (6)$$

Additionally, a *fourth phase* can be performed to acquire an invariant subspace of A or (A, B) that is associated with a given subset of eigenvalues or a given subset of generalized eigenvalues:

Eigenvalue reordering: The real Schur form S or the generalized real Schur form (S, T) is reordered, such that a selected set of eigenvalues or generalized eigenvalues appears among the leading diagonal blocks of an updated real Schur form \hat{S} or an updated generalized real Schur form (\hat{S}, \hat{T}) , by a similarity transformation

$$S = Q_3 \hat{S} Q_3^T \quad \text{or} \quad (S, T) = Q_3 (\hat{S}, \hat{T}) Z_3^T, \quad (7)$$

where $Q_3, Z_3 \in \mathbb{R}^{n \times n}$ are orthogonal matrices.

A detailed explanation of the underlying mathematical theory can be found in the textbook by Golub and van Loan¹⁵.

4 | A CASE FOR THE TASK-BASED APPROACH

A task-based algorithm functions by cutting the computational work into self-contained tasks that all have a well defined set of inputs and outputs. In particular, StarNEig divides the matrices into disjoint (square) tiles[‡] and each task takes a set of tiles as its input and produces/modifies a set of tiles as its output. The main difference between tasks and regular function/subroutine calls is that a task-based algorithm does not call the associated computational kernels directly. Instead, StarNEig inserts the tasks into a runtime system which then derives the data dependences between the tasks from the supplied input and output information (i.e., the task graph is constructed implicitly). The runtime system then schedules the tasks to computational resources, such as CPUs and GPUs, in a sequentially consistent order as dictated by the data dependences.

The main benefit of the task-based approach is that as long as the cutting is carefully done, the underlying parallelism is exposed automatically as the runtime system gradually traverses the resulting task graph. In particular, the runtime system can detect and tap into previously unexplored avenues of parallelism that are hidden within the task graphs. This leads to significantly more powerful algorithms that are able to adapt to different inputs and changing hardware configurations. Other benefits of the task-based approach include, for example, better load balancing and resource utilization due to dynamic scheduling, task priorities and implicit MPI communications. In particular, since all MPI nodes have a (partial) copy of the task graph and are aware of the data distribution, the runtime system is able to derive the necessary MPI communication pattern from the provided information. Some explicit communication is required in a few special cases, for example, when the main thread accesses the data before a data-dependent branching point. StarNEig also helps the runtime system by setting communication priorities. Otherwise most of the communication is offloaded to the runtime system.

The following subsections briefly discuss the four phases in the algorithm stack from the point of view of task parallelism. We do not attempt to explain the details of each algorithm, rather we focus on the key steps and explain how they benefit from task parallelism. We will use the standard eigenvalue problem as an illustration.

4.1 | GPU-accelerated Hessenberg reduction

We begin by discussing the Hessenberg reduction phase. We emphasize that this phase does not benefit as much from the task-based approach at the other phases. Now, the so-called standard algorithm¹⁶ for reducing a nonsymmetric matrix A to upper

[‡]The StarPU runtime system does not differentiate between tiles that are stored contiguously and tiles that are stored in a strided manner. The latest version of StarNEig (v0.1.1) does not rearrange the data in memory before registering it with StarPU but the feature may be introduced in a future release of the library.

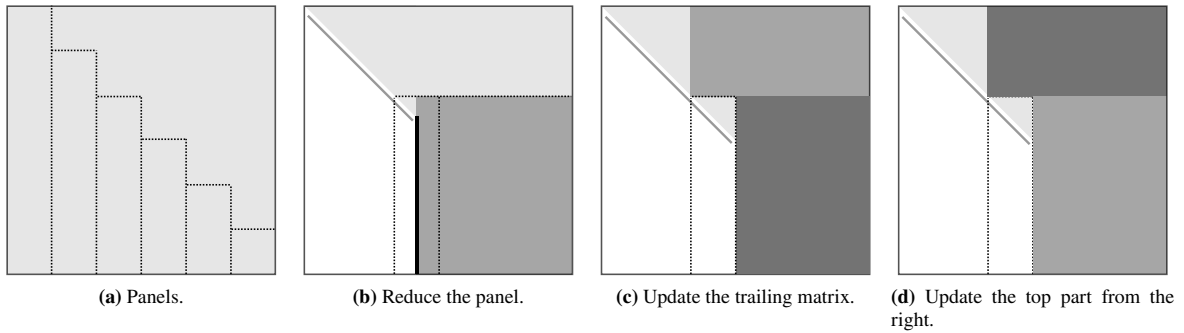


FIGURE 1 An illustration of the standard algorithm for reducing a dense matrix to upper Hessenberg form.

Hessenberg form H first divides the matrix A into disjoint panels as illustrated in Figure 1a. Each panel is then reduced to upper Hessenberg form as summarized below:

1. *Reduce the panel.* The i th column in the panel is reduced by constructing and applying a suitable Householder reflector $I - \tau_i v_i v_i^T$, where $\tau_i \in \mathbb{R}$ and $v_i \in \mathbb{R}^n$. All involved reflectors are initially applied only inside the current panel and accumulated into a compact WY representation $I - V_i T_i V_i^T$, where $T_i \in \mathbb{R}^{l \times l}$ is upper triangular and $V_i = [v_1 v_2 \dots v_i]$. In tandem, one of the necessary intermediate results is also accumulated into a matrix $Y_i = A V_i T_i$. We emphasize that the formula $Y_i \leftarrow A V_i T_i$ is never explicitly evaluated. Instead, the matrix Y_i is constructed column by column using an equivalent update formula

$$Y_i \leftarrow [Y_{i-1} \mid \tau_i (A - Y_{i-1} V_{i-1}^T) v_i]. \quad (8)$$

The construction of the matrix Y_i allows us to reduce the entire panel without updating the other sections of the matrix. However, the update formula (8) includes a large matrix-vector multiplication involving the matrix A and the vector v_i . Fortunately, only lower part of the matrix Y_i is needed and the matrix-vector multiplication therefore only involves the section of the matrix A that trails the current column (see the shaded area in Figure 1b) and the non-zero part of the vector v_i . Although these BLAS-2 matrix-vector multiplications constitute approximately only 20% of the total number of flops of the algorithm, they are significantly more expensive than the remaining 80% due to the fact that matrix-vector multiplication is a memory-bound operation and the remaining 80% are mostly compute-bound BLAS-3 operations. This is an important factor to consider when analysing the performance.

2. *Update the trailing matrix.* Update the section of the matrix that trails the current panel (see the shaded area in Figure 1c) utilizing the update formula

$$A \leftarrow (I - V T V^T)^T (A - Y V^T), \quad (9)$$

where $I - V T V^T$ and Y are the final compact WY representation and the final intermediate result matrix from the panel reduction step, respectively.

3. *Update the top part from the right.* Update the section of the matrix above the current panel (see the shaded area in Figure 1d) utilizing the update formula

$$A \leftarrow A (I - V T V^T). \quad (10)$$

Note that the above description is a simplification that is only meant to highlight the key computational steps, including the memory-bound component. The formulas are very close to the ones used in the actual implementation but adding additional details would make the description unnecessarily long. See the paper by Quintana-Ortí and van de Geijn¹⁶ for a complete description of the algorithm.

In StarNEig (version 0.1-beta.2), each of the aforementioned three steps is formulated as a set of tasks. However, for task scheduling overhead related reasons, the *reduce the panel* and *update the trailing matrix* steps are implemented as monolithic tasks. Implementing each column reduction as a set of separate tasks would lead to an excessive number of lightweight tasks and thus to an unmanageable amount of scheduling overhead. The computational resources are divided into two scheduling contexts:

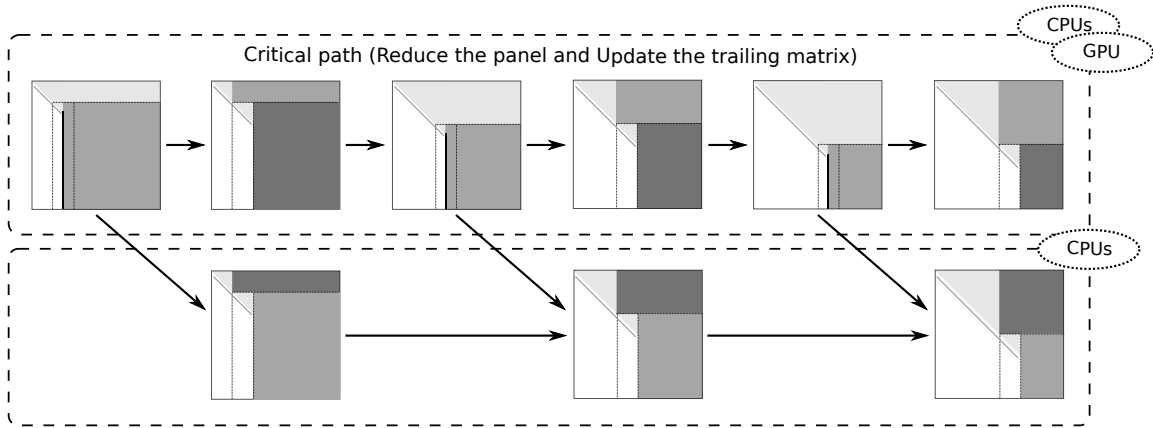


FIGURE 2 An illustration of the two scheduling contexts.

Parallel scheduling context contains a subset of the available CPU cores and a GPU, if one exists in the system. Each inserted task is executed in parallel by all included CPU cores or by the GPU, depending on which resource is predicted to be the best option by the runtime system. In order to accomplish this, the runtime system uses calibrated performance models to predict the execution and data transfer times for each task. The *reduce the panel* and *update the trailing matrix* tasks form the critical path of the algorithm and are scheduled to this scheduling context as illustrated in Figure 2. The CPU implementation of the *reduce the panel* task copies the trailing matrix to memory buffers that are allocated from the local NUMA islands. That is, each involved CPU core has its own local memory buffer and the obtainable memory bandwidth is thus significantly higher compared to a situation where the CPU cores would access the memory across the NUMA islands. Note that the trailing matrix is copied only once at the beginning of each *reduce the panel* task. This is another reason for implementing the *reduce the panel* step as one monolithic task.

Sequential scheduling context contains the remaining CPU cores and will inherit computational resources from the parallel scheduling context once the critical path has been completed. Each inserted task is executed sequentially by one of the available computational resources. The *update the top part from the right* tasks are scheduled to this context as illustrated in Figure 2. Note that the *update the top part from the right* tasks never feed back into the critical path and can therefore be scheduled independently. The tasks that compute the matrix Q_1 in (3) are also scheduled to this context and given lower priority.

The main benefit that comes from the task-based approach here is that the runtime system is allowed to schedule the work to the GPU when it predicts this will improve the performance. In particular, the runtime system usually schedules *reduce the panel* tasks to the GPU because most modern GPUs have a much higher memory bandwidth than CPUs. This means that the time that is spent computing the large matrix-vector multiplications is significantly reduced. The runtime system will handle the necessary data transfer between main memory and GPU memory, including prefetching. The end result is that the task-based implementation will naturally behave very similar to the implementation available in the MAGMA library^{7,17} but provides some additional flexibility as the scheduling decisions are done dynamically. See the NLAfet Deliverable Report D2.6¹⁸ for further information.

4.2 | Schur reduction and eigenvalue reordering

We will now use the Schur reduction and eigenvalue reordering phases to illustrate some of the more notable benefits of the task-based approach. In particular, we demonstrate that StarNEig can successfully merge computational steps, that were previously separated by global synchronization points, and thus improve the concurrency. The modern approach for obtaining a Schur form S of a nonsymmetric matrix A is to apply the multi-shift QR algorithm with Aggressive Early Deflation (AED) to the upper Hessenberg form H (see References^{19,20,21,22} and references therein). The algorithm is a sequence of steps of two types, AED and bulge chasing, as illustrated in Figure 3 and summarized below:

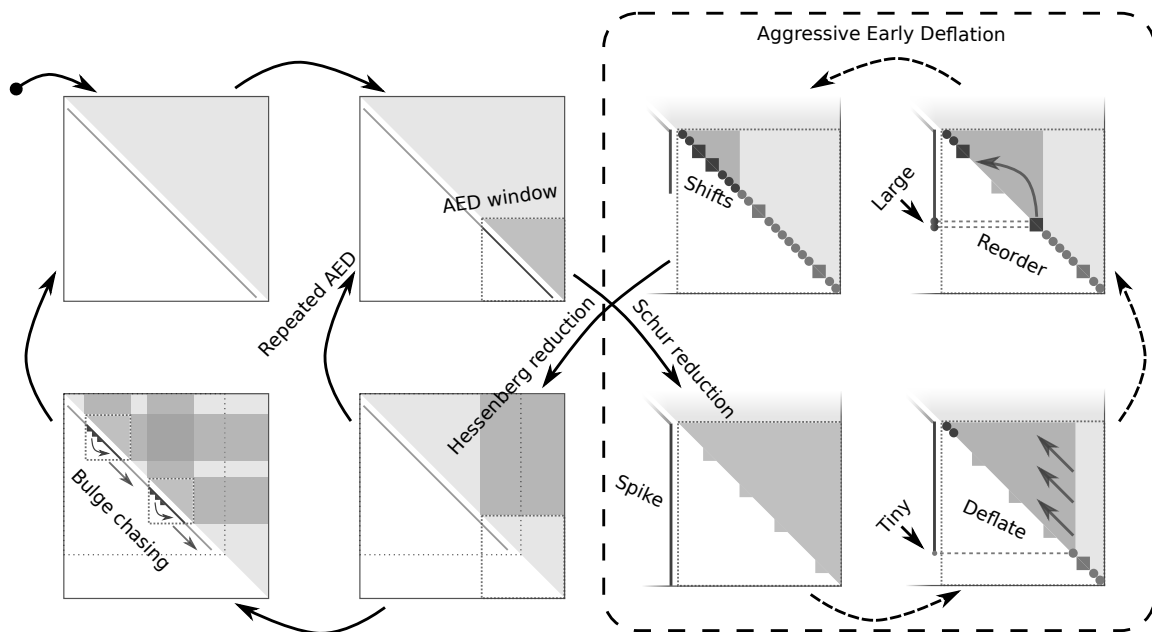


FIGURE 3 An illustration of the multi-shift QR algorithm with aggressive early deflation.

AED step first reduces a small diagonal window (a.k.a AED window) to real Schur form using a recursive application of the QR algorithm. When the sections of the matrix outside the AED window are updated from the left, a spike is induced to the left of the AED window and thus we will temporarily deviate from the upper Hessenberg form. All diagonal blocks in the reduced AED window are then systematically evaluated in order to identify those eigenvalues that can be safely deflated without introducing significant perturbations. If the corresponding element in the spike is found to be small enough (i.e., the so-called deflation condition is satisfied), then the element is set to zero and the eigenvalue that corresponds to the diagonal block is deflated. On the other hand, if the corresponding entry in the spike is found to be too large, then the remaining AED window is reordered such that the diagonal block that failed the deflation check is moved to the upper left corner of the window thus pushing the remaining unevaluated diagonal blocks down along the diagonal. After all diagonal blocks have been evaluated, the remaining spike is eliminated by performing a small-sized Hessenberg reduction. Note that the QR algorithm involves several data-dependent branching points that affect the resulting task graph and thus StarNEig has to wait the completion of certain tasks before continuing the task insertion.

Bulge chasing step chases a set of 3×3 bulges down the diagonal. The eigenvalues of the diagonal blocks that failed the deflation condition, $\lambda_1, \lambda_2, \dots, \lambda_m$, are used as shifts to generate the bulges. That is, the first column of the matrix H is transformed to the first column of the matrix $(H - \lambda_1 I)(H - \lambda_2 I)$ using a small-sized Householder reflector. When applied from the both sides, the reflector creates fill-in in the form of a 3×3 bulge that appears in the upper left corner of the matrix. At this point, the bulge could be eliminated by chasing it down the diagonal with a sequence of overlapping small-sized Householder reflectors and this would complete one implicit QR iteration. However, a multi-shift QR algorithm will instead chase the bulge just enough so that a second bulge can be introduced using the shifts λ_3 and λ_4 . The same procedure is then repeated until a total of $m/2$ bulges have been introduced. The bulges are then chased in groups down the diagonal to complete one pipelined QR iteration. The bulge chasing step is then followed by a second AED step and the same procedure is repeated until all eigenvalues have been deflated or an iteration limit is reached.

Similarly, the eigenvalue reordering phase is based on applying sequences of overlapping Givens rotations and small-sized Householder reflectors to S . The Schur form S is essentially reordered in a bubble sort manner using kernels that swap two adjacent diagonal blocks²³.

If the Givens rotations and small-sized Householder reflectors are applied one by one, then memory is accessed as shown in Figure 4a. This is grossly inefficient for two reasons: i) the transformations are so localized that parallelizing them would not produce any significant speedup and ii) the matrix elements are touched only once leading to very low arithmetic intensity. The modern approach (see References^{19,20,21,22,24,25} and the references therein) groups together a set of local transformations and

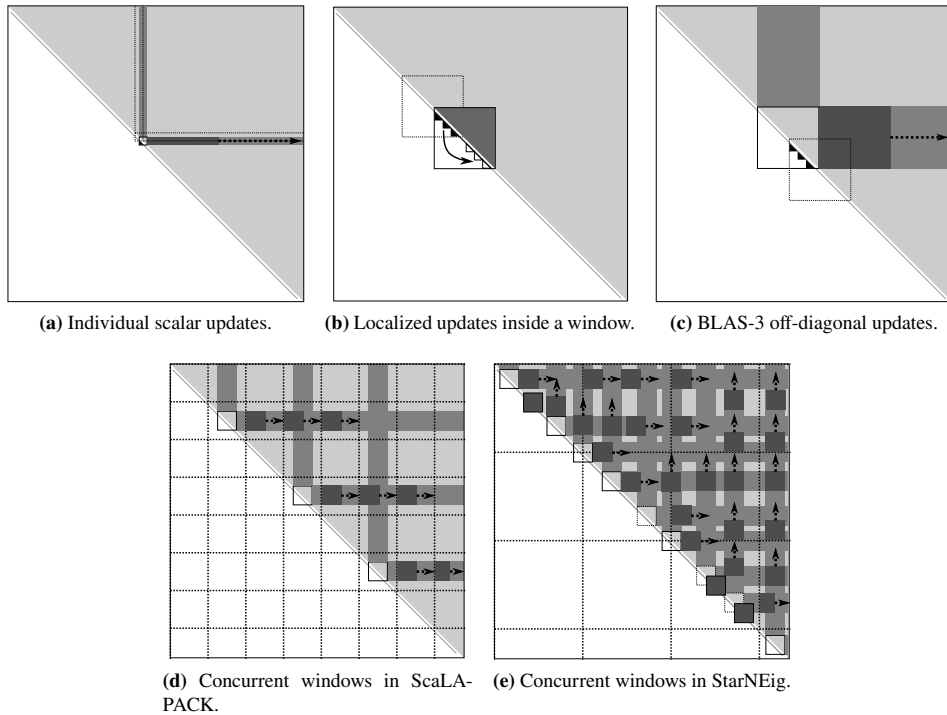


FIGURE 4 Hypothetical snapshots taken during the bulge chasing step of the Schur reduction phase. The currently active regions are highlighted with darker shade and the propagation directions of the off-diagonal updates are marked with arrows. In (a), the overlap between two overlapping transformations is highlighted with dashed lines. In (b) and (c), the overlap between two diagonal windows is highlighted with dashed lines. In (d) and (e), the dashed lines illustrate how the matrix is divided into distributed blocks.

initially applies them to a relatively small diagonal window as shown in Figure 4b. The localized transformations are accumulated into an accumulator matrix and later propagated as BLAS-3 operations acting on the off-diagonal sections of the matrix as shown in Figure 4c. This leads to much higher arithmetic intensity and enables proper parallel implementations as *multiple* diagonal windows can be processed concurrently. These are the main reasons why the multi-shift QR algorithm introduces several 3×3 bulges to the diagonal. The 3×3 bulges are divided into groups and each group is chased separately down the diagonal. A set, or a *chain* (see Reference²⁶), of overlapping diagonal windows is associated with each group. Similarly, several selected diagonal blocks can be grouped and moved together in the eigenvalue reordering phase^{24,25,27}.

The Schur reduction and eigenvalue reordering phases are implemented in ScaLAPACK as the PDHSEQR²¹ and the PDTRSEN²⁵ subroutines, respectively. Following the ScaLAPACK convention, the matrices are distributed in a two-dimensional block cyclic fashion²⁸. The resulting memory access pattern is illustrated in Figure 4d for a 3×3 MPI process mesh. In this example, three diagonal windows can be processed simultaneously. The related BLAS-3 off-diagonal updates require careful coordination since the left and right hand side updates must be performed in a sequentially consistent order. In practice, this means (global or row/column communicator broadcast) synchronization after each set of BLAS-3 off-diagonal updates have been applied. In addition, each AED step introduces a global synchronization point.

In StarNEig, the Schur reduction bulge chasing step and the eigenvalue reordering phase are implemented with the following tasks types:

Window task generates and applies a set of local transformations inside a diagonal window. Takes the intersecting tiles as input, and produces updated tiles and an accumulator matrix as output.

Right update task applies accumulated right-hand side updates using BLAS-3 operations. Takes the intersecting tiles and an accumulator matrix as input, and produces updated tiles as output.

Left update task applies accumulated left-hand side updates using BLAS-3 operations. Takes the intersecting tiles and an accumulator matrix as input, and produces updated tiles as output.

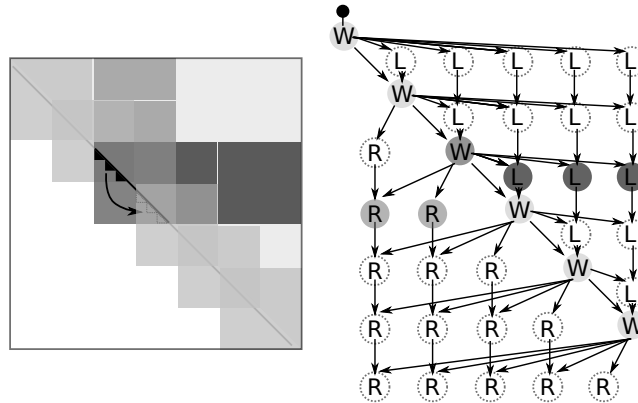


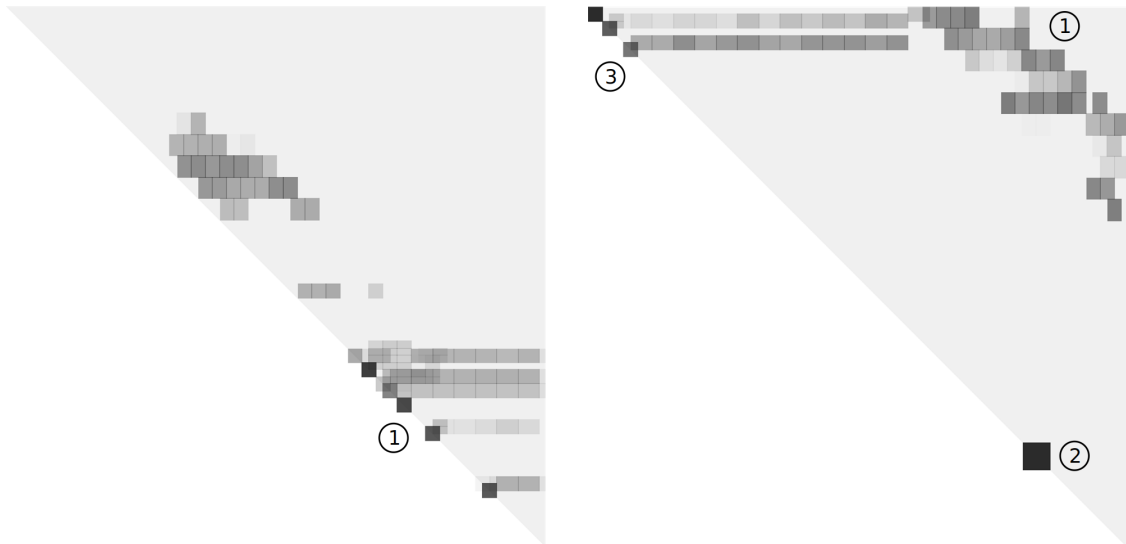
FIGURE 5 A hypothetical task graph arising from a situation where a set of three bulges is chased down the diagonal. We have simplified the graph by omitting dependences between the left (L) and right (R) update tasks as these dependences are enforced through the diagonal bulge chasing tasks (W). Note that the actual bulge chasing step involves *several* sets of bulges and the resulting task graph is therefore significantly more complex than the simplified graph presented here.

The tasks are inserted into the runtime system in a sequentially consistent order and each chain of overlapping diagonal windows leads to a task graph like the one shown in Figure 5. Note that real live task graphs are significantly more complex than shown here, but also enclose more opportunities for parallelism. It is also critical to realize that the runtime system *guarantees* that the tasks are executed in a sequentially consistent order. In particular, there is no need for explicit synchronization and different computational steps are allowed to overlap (see Figure 4e) as the runtime system *merges* the corresponding sub-graphs together. This can lead to a much higher concurrency since idle time can be reduced by delaying low priority tasks until computational resources start becoming idle. This is exactly what can be observed from Figure 6a where a subset of the right-hand side updates are given lower priority and are therefore scheduled only after the diagonal bulge chasing tasks and the left-hand side update tasks cannot saturate all available computational resources. This is possible because, as seen in Figure 5, most right-hand side update tasks do not feed back to the other sections of the task graph and can therefore be scheduled independently.

The AED step can also be overlapped with the bulge chasing steps as shown in Figures 6b and 7. This improves the concurrency significantly compared to ScaLAPACK because the ScaLAPACK implementation will effectively synchronize before and after each AED step. This means that in ScaLAPACK most MPI ranks will become idle while a subset of the MPI ranks are involved in the AED step. Actually, StarNEig can even overlap two bulge chasing steps with each other as seen in Figures 6b. A complete illustration is available on YouTube²⁹. See the NLAfet Deliverable Reports D2.5³⁰, D2.6¹⁸, D2.7⁸ and D6.5³¹ for further information.

4.3 | Robust computation of eigenvectors

Mathematically, the problem of computing a single eigenvector of, say, a quasi-triangular matrix is trivial. The standard algorithm is a variant of substitution. However, substitution is very vulnerable to floating-point overflow. In particular, there exist triangular linear systems which are well-conditioned in the sense of Skeel for which the solution grows exponentially and rapidly exceeds the representational range of any floating-point number system³². We say that an algorithm or a subroutine is robust if all intermediate and final results are in the representational range, i.e., floating point overflow is avoided. In LAPACK there exists a robust family xLATRS of subroutines for solving triangular linear system $Tx = b$ ³³. They dynamically scale the entire right-hand side and return a scaling factor α and a vector x such that $Tx = \alpha b$. The purpose of the scaling factor α is to extend the floating-point representational range. In LAPACK, the solvers for computing eigenvectors from Schur forms are all descended from xLATRS. They are scalar codes which compute the eigenvectors one by one. The back-transformation to the



(a) The first bulge chasing step with delayed right-hand side updates. (b) The second AED step overlapping with two bulge chasing steps.

FIGURE 6 Snapshots taken during the first two iterations of the multi-shift QR algorithm with AED. The numbering matches the numbering in Figure 7. That is, (1) is the bulge chasing step from the first iteration, (2) is the small-sized Hessenberg reduction from the second AED step and (3) is the bulge chasing step from the second iteration. Note that the three steps are merged in (b).

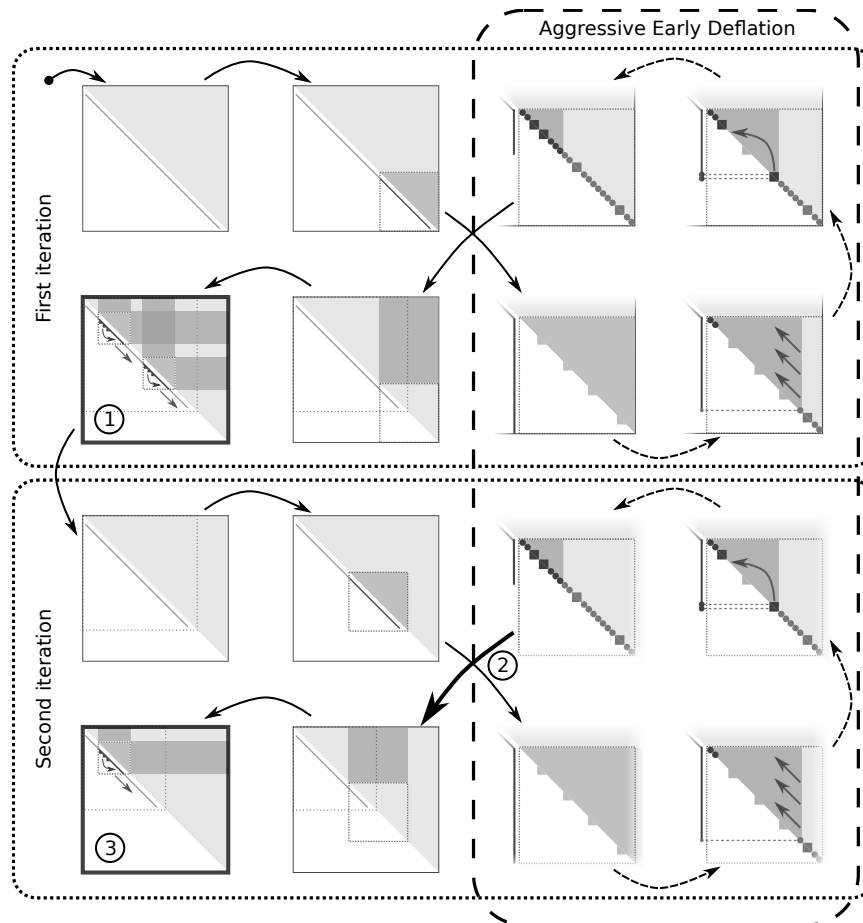


FIGURE 7 An illustration of the first two iterations of the multi-shift QR algorithm with AED. The steps that are concurrently active in Figure 6b are highlighted in bold and numbered correspondingly. Contrast with Figure 3.

original basis can of course be done using BLAS-3 operations. In ScaLAPACK, there is no parallel implementation of xLATRS and the solvers for computing eigenvectors are not robust.

In contrast, StarNEig implements novel algorithms for computing eigenvectors which are tiled, parallel and robust. In StarNEig, each matrix of eigenvectors is partitioned into tiles X_{ij} . Every tile $X = [x_1 \ x_2 \ \dots \ x_k]$ is augmented with a vector of scaling factors $\alpha \in \mathbb{R}^k$ with one scaling factor per column. The augmented tile $\langle \alpha, X \rangle$ represents the matrix $Y = [y_1 \ y_2 \ \dots \ y_k]$ given by $y_j = x_j/\alpha_j$. The StarNEig solvers for computing eigenvectors accept and produce augmented tiles. This allows StarNEig to obtain a representation of the eigenvectors without exceeding the representational range. A final post-processing step ensures that all segments of each eigenvector are consistently scaled. In reality, StarNEig is exploiting a principle which is familiar to every scientist: Two results can be combined if we know how to convert between the different system of measurements, say, the metric system and the imperial system.

The use of augmented tiles is compatible with the linear update $Z \leftarrow Y - TX$ and the vast majority of the arithmetic operations can be completed using BLAS-3 operations. Given a matrix T and two augmented tiles $\langle \alpha, X \rangle$ and $\langle \beta, Y \rangle$, such that $Y - TX$ is defined, StarNEig produces an augmented tile $\langle \gamma, Z \rangle$ which represents the intended result, i.e.,

$$\gamma_j^{-1} z_j = \beta_j^{-1} y_j - \alpha_j^{-1} T x_j. \quad (11)$$

This is achieved as follows. A preprocessing step ensures that each pair (x_j, y_j) of columns of X and Y are not only consistently scaled, but the linear update $z_j = y_j - T x_j$ can be computed without exceeding the overflow threshold. If X, Y and T are $n \times n$ matrices, then this preprocessing step requires $O(n^2)$ operations. Then the linear update $Z \leftarrow Y - TX$ is executed using a single BLAS-3 operation. We emphasize that the cost of the preprocessing step is insignificant compared with the $O(n^3)$ arithmetic operations required for the linear update. Rescaling vectors to obtain a consistent scaling requires that the scaling factors are nonzero. Otherwise, a division-by-zero is attempted. In StarNEig, the possibility of the scaling factors underflowing is significantly reduced by using scaling factors which are integer powers of 2. Currently StarNEig uses at least 32 bit signed integers, for which the smallest scaling factor is $\alpha = 2^{-2^{31}+1} \approx 10^{-6.4646 \cdot 10^8}$, but this can easily be extended to the point where underflow is a practical impossibility using 64 bit unsigned integers.

The main benefits that stem from the task-based approach are the merging of different computational steps and the improved load balancing. Additional information can be found in the existing literature. In particular, the fundamental principles for solving triangular linear systems in parallel without suffering from overflow are discussed in References^{32,34}. The StarNEig solvers for computing standard and generalized eigenvectors are the subjects of separate papers^{35,36}.

5 | STARNEIG LIBRARY

StarNEig is a C-library that runs on top of the StarPU task-based runtime system. StarPU handles low-level operations such as heterogeneous scheduling, data transfers and replication between memory spaces and MPI communication between compute nodes. In particular, StarPU is responsible for managing the various computational resources such as CPU cores and GPUs. The support for GPUs and distributed memory were the main reasons why StarPU was chosen as the runtime system.

StarPU manages a set of worker threads; usually one thread per computational resource. In addition, one thread is responsible for inserting the tasks into StarPU and tracking the state of the machine. If necessary, one additional thread is allocated for MPI communication. For these reasons, StarNEig must be used in a *one process per node* (1ppn) configuration, i.e., several CPU cores should be allocated for each MPI process (a node can be a full physical node, a NUMA island or some other reasonably large collection of CPU cores).

The current status of StarNEig is summarized in Table 2. The *Experimental* status indicates that the software component has not been tested as extensively as those software components that are considered *Complete*. In particular, the GPU functionality requires some additional involvement from the user (performance model calibration). At the time of writing this paper, only real arithmetic is supported and certain interface functions are implemented as LAPACK and ScaLAPACK wrapper functions. However, we emphasize that StarNEig supports real valued matrices that have complex eigenvalues and eigenvectors. Additional distributed memory functionality and support for complex data types are planned for a future release.

Remark 2. StarNEig uses several LAPACK routines as low-level kernels, the most complex one of them being the DHSEQR routine that is used to solve tiny eigenvalue problems sequentially, but StarNEig does not reuse any LAPACK code. StarNEig does not reuse nor call any ScaLAPACK routines, excluding the wrapper functions listed in Table 2 and certain interface functions that are related to the ScaLAPACK compatibility layer (see Subsection 5.2). In particular, StarNEig can be compiled without

TABLE 2 Current status of the StarNEig library (version 0.1.1).

Computational phase	Shared memory	Distributed memory	GPUs (CUDA)
<i>Hessenberg reduction</i>	Complete	ScaLAPACK	Single GPU
<i>Schur reduction</i>	Complete	Complete	Experimental
<i>Eigenvalue reordering</i>	Complete	Complete	Experimental
<i>Eigenvectors</i>	Complete	In progress	.
<i>Hessenberg-triangular reduction</i>	LAPACK	ScaLAPACK	.
<i>Generalized Schur reduction</i>	Complete	Complete	Experimental
<i>Generalized eigenvalue reordering</i>	Complete	Complete	Experimental
<i>Generalized eigenvectors</i>	Complete	In progress	.

ScaLAPACK. The NLAfet Deliverable Report D2.6¹⁸ mentions the reuse of some sequential ScaLAPACK code in early prototype codes, but these components have since been replaced with code that has been written from the scratch in C. We have not used Fortran-to-C compilers.

5.1 | Distributed memory

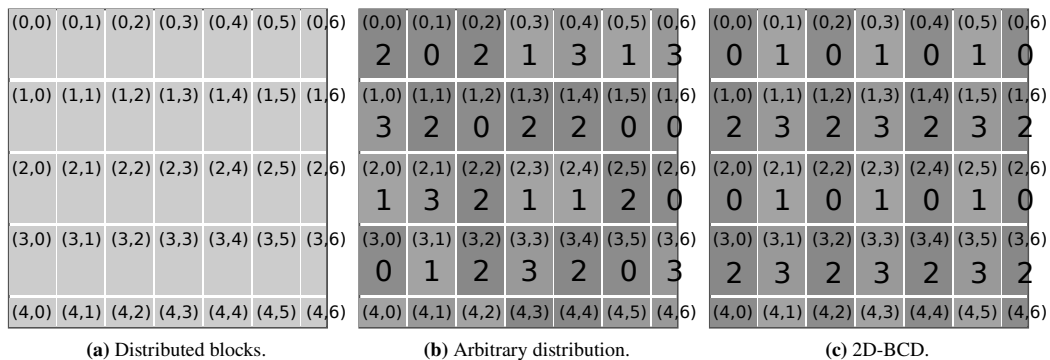


FIGURE 8 Examples of various data distributions supported by StarNEig, including two-dimensional block cyclic distribution (2D-BCD).

StarNEig distributes the matrices in rectangular blocks of uniform size (excluding the last block row and column) as illustrated in Figure 8a. The data distribution, i.e., the mapping from the distributed blocks to the MPI process rank space, can be arbitrary as illustrated in Figure 8b. A user has three options:

1. Use the default data distribution. This is recommended for most users and leads to reasonable performance in most situations.
2. Use a two-dimensional block cyclic distribution (see Figure 8c). In this case, the user may select the MPI process mesh dimensions and the rank ordering.
3. Define a data distribution function $d : \mathbb{Z}^+ \times \mathbb{Z}^+ \rightarrow \mathbb{Z}^+$ that maps the block row and column indices to the MPI rank space. For example, in Figure 8b, rank 0 owns the blocks (0,1), (1,2), (1,5), (1,6), (2,6), (3,0) and (3,5).

The library implements copy, scatter and gather operations that support data redistribution between all supported distributions.

Users who are familiar with ScaLAPACK are likely accustomed to using relatively small distributed block sizes (between 64–256). In contrast, StarNEig functions optimally only if the distributed blocks are relatively large (at least 1000 but preferably much larger). This is due to the fact that StarNEig further divides the distributed blocks into tiles and a tiny tile size leads to

excessive task scheduling overhead because the tile size is closely connected to the task granularity. Furthermore, as mentioned in the preceding section, StarNEig should be used in 1ppn configuration as opposed to a *one process per core* (1ppc) configuration which is common with ScaLAPACK.

5.2 | ScaLAPACK compatibility

StarNEig is fully compatible with ScaLAPACK and provides a ScaLAPACK compatibility layer that encapsulates BLACS contexts and descriptors (see the BLACS User's Guide²⁸) inside transparent objects, and implements a set of bidirectional conversion functions. The conversions are performed in-place and do not modify any of the underlying data structures. Users can mix StarNEig interface functions with ScaLAPACK routines without intermediate conversions. The use of the ScaLAPACK compatibility layer requires the use of either the default data distribution or the two-dimensional block cyclic data distribution. Listing 1 illustrates how a distributed StarNEig matrix is converted to an equivalent BLACS context and a matching local buffer. The matrix is then reduced to upper Hessenberg form by calling the PDGHRD routine from ScaLAPACK. This is actually how the ScaLAPACK wrapper functions are implemented in StarNEig, see Table 2.

Listing 1: An example how a distributed matrix is converted to a BLACS descriptor and a local buffer.

```
// create a 2D block cyclic data distribution (pm x pn process mesh)
starneig_distr_t distr =
    starneig_distr_init_mesh(pm, pn, STARNEIG_ORDER_DEFAULT);

// create an n x n distributed matrix (bn x bn blocks)
starneig_distr_matrix_t dA =
    starneig_distr_matrix_create(n, n, bn, bn, STARNEIG_REAL_DOUBLE, distr);
...

// convert the data distribution to a BLACS context
starneig_blacs_context_t ctx = starneig_distr_to_blacs_context(distr);

// convert the distributed matrix to a BLACS descriptor and a local buffer
starneig_blacs_descr_t descr;
double *ptr;
starneig_distr_matrix_to_blacs_descr(dA, ctx, &descr, (void **)&ptr);

// ScaLAPACK routine for reducing a general distributed matrix to upper
// Hessenberg form
extern void pdghrd_(int const *, int const *, int const *, double *,
    int const *, int const *, starneig_blacs_descr_t const *, double *,
    double *, int const *, int *);

pdghrd_(&n, &ilo, &ihi, ptr, &ia, &ja, &descr, tau, ...);
```

6 | PERFORMANCE EVALUATION

This section demonstrates the performance of the StarNEig library through a set of computational experiments.

Hardware environment

Computational experiments were performed on the Kebnekaise system, located at the High Performance Computing Center North (HPC2N), Umeå University. Kebnekaise is a heterogeneous systems consisting of many different types of compute nodes. The compute node types relevant for this paper are:

Broadwell compute node contains 28 Intel Xeon E5-2690v4 cores organized into 2 NUMA islands with 14 cores each and 128 GB memory. The nodes are connected with FDR Infiniband. All distributed memory experiments were performed on these nodes.

Skylake compute node contains 28 Intel Xeon Gold 6132 cores organized into 2 NUMA islands with 14 cores each and 192 GB memory. All Hessenberg reduction experiments were performed on these nodes.

Large memory node contains 72 Intel Xeon E7-8860v4 cores organized into 4 NUMA islands with 18 cores each and 3072 GB memory. All eigenvector experiments were performed on these nodes.

V100 GPU node contains 28 Intel Xeon Gold 6132 cores organized into 2 NUMA islands with 14 cores each and 192 GB memory. Each NUMA island is connected to a single NVIDIA Tesla V100 GPU. All GPU experiments were performed on these nodes.

Software environment

For the distributed memory and eigenvalue reordering related GPU experiments, the software was compiled with GCC 7.3.0 and linked to OpenMPI 3.1.3, OpenBLAS 0.3.2, ScaLAPACK 2.0.2, CUDA 9.2.88, and StarPU 1.2.8. Since the version of the PDHSEQR routine that exists in ScaLAPACK 2.0.2 is known to be buggy, StarNEig was actually compared against an updated version of PDHSEQR; see Reference²¹. The updated version places far less strict conditions on the distributed block size and thus performs better on modern hardware. For the Hessenberg reduction and eigenvector computation experiments, the software was compiled with ICC 19.0.1.144 and linked to Intel MPI 2018.4.274, Intel MKL 2019.1.144, CUDA 10.1.105 and StarPU 1.2.8. In all experiments, StarPU was compiled with the `--disable-cuda-memcpy-peer` configuration flag enabled as this reduces the CUDA overhead.

StarNEig version

In most experiments, StarNEig version 0.1-beta.2 was used. However, the Schur reduction and eigenvalue reordering experiments in distributed memory were performed using an older and unpublished version of StarNEig. The main difference between this unpublished version and the version 0.1-beta.2 is the deflation condition used in the Schur reduction phase. The deflation condition is used to decide when an eigenvalue can be deflated and thus impacts the convergence rate of the algorithm. The unpublished version uses a deflation condition that is identical to the one used in LAPACK and PDHSEQR whereas the version 0.1-beta.2 uses the so-called norm stable deflation condition^{19,20}. The latter deflation condition is less strict and could thus potentially lead to faster convergence. Newer versions of StarNEig (since version 0.1-beta.4) allow the user to choose between these two deflation conditions.

6.1 | Hessenberg reduction in shared memory with and without a GPU

Table 3 shows how the Hessenberg reduction routine in StarNEig compares against MKL-LAPACK (with parallel BLAS), MKL-ScaLAPACK (multiple MPI ranks on single physical node) and MAGMA^{7,17}. StarNEig does not have multi-GPU support in the Hessenberg reduction phase (as mentioned in Table 2). Therefore, only a single GPU and the associated NUMA island was used in the GPU experiments. The additional cores from the second NUMA island would only have had a minimal effect since the algorithm is memory-bound and the 14 AVX-512 capable CPU cores from a single NUMA island can execute the compute-bound BLAS-3 operations faster than the GPU can execute the memory-bound BLAS-2 operations. Spreading the computations across two NUMA islands might actually have slowed the computations and it is more difficult to configure several NUMA islands together with the GPU. The `STARPU_WORKERS_CPUID` environmental variable was set to `0 14 1 15 2 16 [...]`.

Since the implementations in all four libraries are based on the standard algorithm¹⁶, the performance is limited by the throughput of the matrix-vector multiplication routine. It is therefore important that the memory access pattern is optimized. In particular, a memory access to a NUMA island that is not local to a particular core is very likely to reduce the performance since the obtainable memory bandwidth is significantly lower compared to a local access. A memory access pattern that is

TABLE 3 A run time comparison (in seconds) between parallel MKL-LAPACK, MKL-ScaLAPACK, MAGMA and StarNEig when computing a Hessenberg form in shared memory. Columns 2-4 show the execution times for 28 Intel Xeon Gold 6132 cores and columns 5-6 show the execution times for 14 Intel Xeon Gold 6132 cores paired with a NVIDIA Tesla V100 GPU.

n	28 cores			14 cores + V100 GPU	
	LAPACK	ScaLAPACK	StarNEig	MAGMA	StarNEig
5 000	3.8	4.2	8.4	1.3	2.2
10 000	32	29	40	7.3	8.9
15 000	113	95	117	23	24
20 000	278	223	249	49	53
25 000	561	459	454	90	97
30 000	953	728	755	144	164
40 000	2397	1722	1711	—	—

close to optimal is easy to achieve with ScaLAPACK since each MPI process accesses its own local buffer and this buffer can be allocated from the NUMA island that is closest to the core to which the MPI process is bound. For the other CPU experiments, the memory was allocated in interleaved mode across the two NUMA islands in order to evenly distribute the load. Since the matrix-vector multiplication operation is memory bound, it is not surprising that StarNEig closely matches the performance of MKL-LAPACK, MKL-ScaLAPACK and MAGMA. It is only in the case of the smallest matrix considered where the additional task scheduling related overhead begins to negatively impact StarNEig’s performance. The additional memory locality considerations in StarNEig begin to show their effect when the matrices are reasonably large, leading in some cases to a 40% performance improvement compared to MKL-LAPACK.

6.2 | Schur reduction and eigenvalue reordering in distributed memory

All distributed memory experiments were performed using a square MPI process mesh. This was done because the maximum number of diagonal bulge chasing and reordering windows in the PDHSEQR and PDTRSEN routines is given by $\min(pm, pn)$, where pm and pn are the height and width of the process mesh, respectively. A square process mesh thus maximizes the degree of parallelism in the ScaLAPACK routines. We always map each StarNEig process to a full node (28 cores) and each ScaLAPACK process to a single CPU core, the latter being done for the same reason as the choice to use a square process mesh. Since it is not straightforward to find a CPU core allocation that would lead to a square MPI process mesh in both configurations, the number of CPU cores in each ScaLAPACK experiment is always equal or larger than the number of CPU cores in the corresponding StarNEig experiment. The upper Hessenberg matrices for the Schur reduction experiments were computed from random matrices (entries uniformly distributed over the interval $[-1, 1]$). In the ScaLAPACK experiments, the matrices were distributed in 160×160 blocks, and in the StarNEig experiments, the library default block size was used. In the eigenvalue reordering experiments, 35% of the diagonal blocks were randomly selected. The distributed memory experiments were initially reported in the NLAFFET Deliverable Report D2.7⁸ and the original conference paper⁹.

Comparison against ScaLAPACK

Table 4 shows a comparison between ScaLAPACK and StarNEig. We note that StarNEig is between 1.6 and 2.9 times faster than ScaLAPACK (PDHSEQR) when computing the Schur form and between 2.8 and 5.0 times faster than ScaLAPACK (PDTRSEN) when reordering the Schur form.

Strong scalability

Figures 9a and 9b give some idea of how well the library scales in distributed memory. We note that StarNEig scales reasonably when computing the Schur form and almost linearly when reordering the Schur form. The iterative nature of the QR algorithm makes the Schur reduction results less predictable because different matrices require different numbers of bulge chasing steps. That is, with some matrices, the algorithm ends up performing more consecutive AED steps between the bulge chasing steps, thus leading to a faster convergence rate but weaker scalability as each AED step is performed on a single node.

TABLE 4 A run time comparison (in seconds) between ScaLAPACK and StarNEig in distributed memory. Each node contains 28 Intel Xeon E5-2690v4 cores.

n	CPU cores		Schur reduction		Eigenvalue reordering	
	ScaLAPACK	StarNEig	ScaLAPACK	StarNEig	ScaLAPACK	StarNEig
10 000	36	28	38	18	12	3
20 000	36	28	158	85	72	25
40 000	36	28	708	431	512	180
60 000	121	112	992	563	669	168
80 000	121	112	1667	904	1709	391
100 000	121	112	3319	1168	3285	737
120 000	256	252	3268	1111	2902	581

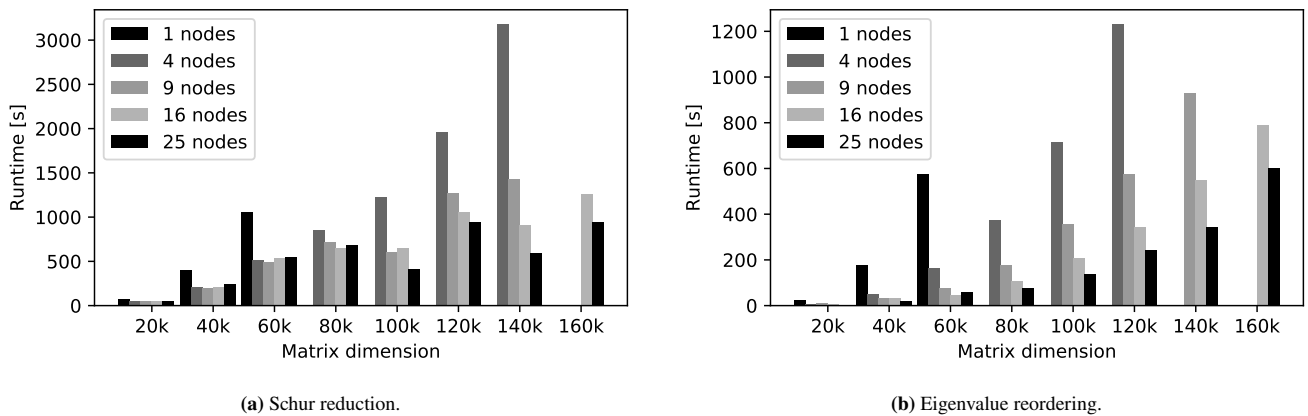


FIGURE 9 Distributed memory scalability of StarNEig when computing and reordering a Schur form. Each node contains 28 Intel Xeon E5-2690v4 cores.

6.3 | Eigenvalue reordering in shared memory with multiple GPUs

Figure 10 demonstrates that StarNEig can indeed take advantage of the available GPUs when reordering a Schur form, as long as the matrices are reasonably large. The introduction of a single V100 GPU gives a speedup of up to 1.65 and the introduction of two V100 GPUs gives a speedup of up to 2.84. In the best case, the speedup when moving from a single V100 GPU to two V100 GPUs is 1.83. This experiment was initially reported in the NLAFFET Deliverable Report D2.7⁸ and the original conference paper⁹.

6.4 | Computation of the eigenvectors in shared memory

Table 5 shows how the eigenvector computation routine in StarNEig compares against MKL-LAPACK (with parallel BLAS). In all experiments, 35% of the eigenvalues were randomly selected. In single core experiments, StarNEig is between 3.4 and 4.9 times faster than MKL-LAPACK. This demonstrates how efficient the tiled approach is compared to the scalar implementation that exists in LAPACK. Furthermore, the multi-core experiments demonstrate the poor scalability of the MKL-LAPACK implementation. In particular, the initial solve phase (5) is executed sequentially and the back transformation phase (6), which is executed in parallel, constitutes only a small fraction of the total execution time. The implementation in StarNEig, on the other hand, demonstrates a very good scalability. In particular, for the larger matrices ($n \geq 40\,000$) the parallel efficiency stays above 70% all the way up to 64 cores. In the most extreme case StarNEig is over 190 times faster than MKL-LAPACK.

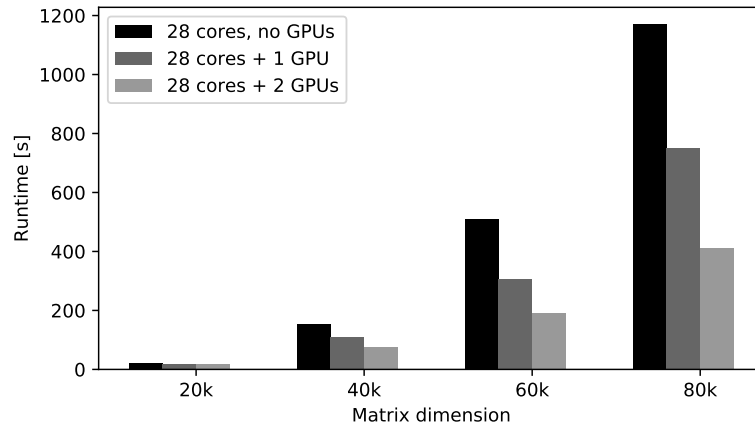


FIGURE 10 GPU performance of StarNEig when reordering a Schur form. Each socket (14 Intel Xeon Gold 6132 cores) is connected to one NVIDIA Tesla V100 GPU.

TABLE 5 A run time comparison (in seconds) between MKL-LAPACK and StarNEig when computing 35% of the eigenvectors. Run times that are longer than four hours are not tabulated.

$n \setminus$ cores	LAPACK					StarNEig				
	1	16	32	48	64	1	16	32	48	64
10 000	95	80	80	80	80	28	2.3	1.4	1.1	1.0
20 000	784	672	694	673	668	194	15	8.1	5.9	5.3
40 000	6731	5826	5681	5608	5743	1364	98	53	38	30
60 000	—	—	—	—	—	4231	315	167	116	90
80 000	—	—	—	—	—	9221	693	359	252	194
100 000	—	—	—	—	—	—	1325	698	483	369
120 000	—	—	—	—	—	—	2226	1194	833	647

7 | SUMMARY

This paper presented a new library called StarNEig. StarNEig aims to provide a complete task-based software stack for solving dense nonsymmetric standard and generalized eigenvalue problems. StarNEig support both shared and distributed memory machines and some routines in the library can take advantage of the available GPUs. The paper is mainly aimed at potential users of the library. Various design choices were explained and contrasted to existing software. In particular, users who are already familiar with ScaLAPACK should know the following:

- StarNEig expect that the matrices are distributed in relatively large blocks compared to ScaLAPACK.
- StarNEig should be used in a *one process per node* (1ppn) configuration as opposed to a *one process per core* (1ppc) configuration which is common with ScaLAPACK.
- StarNEig implements a ScaLAPACK compatibility layer.

The performance of the library was demonstrated with a set of computational experiments. The presented results show the following: In the Hessenberg reduction phase, StarNEig is competitive with LAPACK, ScaLAPACK (multiple MPI ranks on a single physical node) and MAGMA (single GPU). In the Schur reduction phase, StarNEig is between 1.6 and 2.9 times faster than ScaLAPACK. In the eigenvector computation phase, StarNEig’s parallel and robust implementation significantly outperforms LAPACK in both single-core and multi-core settings with recorded speedups as large as 190. In the eigenvalue reordering phase, StarNEig is between 2.8 and 5.0 times faster than ScaLAPACK and scales nearly linearly.

The library is still incomplete. Future work with StarNEig includes the implementation and integration of the missing software components. Support for complex valued matrices is also planned. The GPU support, and the multi-GPU support in particular, are still under active development. The authors hope to start a discussion which would help guide and prioritize the future development of the library.

ACKNOWLEDGMENTS

StarNEig authors

StarNEig has been developed by Mirko Myllykoski (Hessenberg reduction, Schur reduction and eigenvalue reordering; both standard and generalized), Carl Christian Kjelgaard Mikkelsen (generalized eigenvectors and theoretical contributions to the robust computation of eigenvectors), Angelika Schwarz (standard eigenvectors), Lars Karlsson (miscellaneous contributions to the library and theoretical contributions to the robust computation of eigenvectors), and Bo Kågström (coordinator and scientific director of the entire NLAFFET project).

Funding and other contributions

This work is part of a project (NLAFFET) that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 671633. This work was supported by the Swedish strategic research programme eSSSENCE and Swedish Research Council (VR) under Grant E0485301. We thank the High Performance Computing Center North (HPC2N) at Umeå University for providing computational resources and valuable support during test and performance runs. The authors acknowledge the work of Mahmoud Eljammaly and Björn Adlerborn during the NLAFFET project. The authors would also like to thank the StarPU team at INRIA for rapidly providing good answers to several questions related to StarPU. Finally, the authors thank the anonymous reviewers for their valuable feedback.

References

1. StarNEig — A Task-based Library for Solving Nonsymmetric Eigenvalue Problems. <https://nlafet.github.io/StarNEig/>; .
2. Anderson E, Bai Z, Bischof C, et al. *LAPACK Users' Guide*. Philadelphia, PA: Society for Industrial and Applied Mathematics. third ed. 1999.
3. Blackford LS, Choi J, Cleary A, et al. *ScaLAPACK Users' Guide*. Philadelphia, PA: Society for Industrial and Applied Mathematics . 1997.
4. Thibault S. On Runtime Systems for Task-based Programming on Heterogeneous Platforms. Habilitation à diriger des recherches, Université de Bordeaux; 2018.
5. Yarkhan A, Kurzak J, Luszczek P, Dongarra J. Porting the PLASMA Numerical Library to the OpenMP Standard. *Int. J. Parallel Program.* 2017; 45(3): 612–633. doi: 10.1007/s10766-016-0441-6
6. Augonnet C, Thibault S, Namyst R, Wacrenier PA. StarPU: A Unified Platform for Task Scheduling on Heterogeneous Multicore Architectures. *CCPE - Concurrency and Computation: Practice and Experience, Special Issue: Euro-Par 2009* 2011; 23: 187–198. doi: 10.1002/cpe.1631
7. Dongarra J, Gates M, Haidar A, et al. Accelerating Numerical Dense Linear Algebra Calculations with GPUs. *Numerical Computations with GPUs* 2014: 1-26.
8. Myllykoski M, Kjelgaard Mikkelsen CC, Schwarz A, Kågström B. D2.7: Eigenvalue solvers for nonsymmetric problems. tech. rep., Department of Computing Science, Umeå University; SE-901 87 Umeå, Sweden: 2019.
9. Myllykoski M, Kjelgaard Mikkelsen CC. Introduction to StarNEig — A Task-based Library for Solving Nonsymmetric Eigenvalue Problems. In: . 12043 of *Lecture Notes in Computer Science*. Springer Nature Switzerland AG; 2020.
10. Poulson J, Marker B, Geijn v. dRA, Hammond JR, Romero NA. Elemental: A New Framework for Distributed Memory Dense Matrix Computations. *ACM Trans. Math. Softw.* 2013; 39(2). doi: 10.1145/2427023.2427030

11. Marek A, Blum V, Johanni R, et al. The ELPA library: scalable parallel eigenvalue solutions for electronic structure theory and computational science. *Journal of Physics: Condensed Matter* 2014; 26(21): 201–213. doi: 10.1088/0953-8984/26/21/213201
12. EigenExa: High Performance Eigen-Solver. <https://www.r-ccs.riken.jp/labs/lpnctr/en/projects/eigenexa/>; .
13. Gates M, Kurzak J, Charara A, YarKhan A, Dongarra J. SLATE: Design of a Modern Distributed and Accelerated Linear Algebra Library. In: SC'19. Association for Computing Machinery; 2019; New York, NY, USA
14. Hernandez V, Roman JE, Vidal V. SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Trans. Math. Software* 2005; 31(3): 351–362. doi: 10.1145/1089014.1089019
15. Golub GH, Van Loan CF. *Matrix Computations*. Baltimore, MD: Johns Hopkins University Press. 4th ed. 1996.
16. Quintana-Ortí G, Geijn v. dR. Improving the Performance of Reduction to Hessenberg Form. *ACM Trans. Math. Software* 2006; 32(2). doi: 10.1145/1141885.1141887
17. Tomov S, Dongarra J. Accelerating the reduction to upper Hessenberg form through hybrid GPU-based computing. tech. rep., Technical Report 219, LAPACK Working Note; 2009.
18. Myllykoski M, Karlsson L, Kågström B, Eljammaly M, Pranesh S, Zounon M. D2.6: Prototype Software for Eigenvalue Problem Solvers. tech. rep., Umeå University, The University of Manchester; 2018.
19. Braman K, Byers R, Mathias R. The multishift QR algorithm. I. Maintaining well-focused shifts and level 3 performance. *SIAM J. Matrix Anal. Appl.* 2002; 23(4): 929–947. doi: 10.1137/S0895479801384573
20. Braman K, Byers R, Mathias R. The multishift QR algorithm. II. Aggressive early deflation. *SIAM J. Matrix Anal. Appl.* 2002; 23(4): 948–973. doi: 10.1137/S0895479801384585
21. Granat R, Kågström B, Kressner D, Shao M. ALGORITHM 953: Parallel Library Software for the Multishift QR Algorithm with Aggressive Early Deflation. *ACM Trans. Math. Software* 2015; 41(4): 1–23. doi: 10.1145/2699471
22. Granat R, Kågström B, Kressner D, Shao M. ALGORITHM 953: Parallel Library Software for the Multishift QR Algorithm with Aggressive Early Deflation —Electronic Appendix: Derivation of the Performance Model. *ACM Trans. Math. Software* 2015; 41(4). doi: 10.1145/2699471
23. Bai Z, Demmel JW. On swapping diagonal blocks in real Schur form. *Linear Algebra Appl.* 1993; 186: 73–95.
24. Kressner D. Block Algorithms for Reordering Standard and Generalized Schur Forms. *ACM Transactions on Mathematical Software* 2006; 32(4): 521–532. doi: 10.1145/1186785.1186787
25. Granat R, Kågström B, Kressner D. Parallel eigenvalue reordering in real Schur forms. *Concurrency and Computation: Practice and Experience* 2009; 21(9): 1225–1250. doi: 10.1002/cpe.1386
26. Myllykoski M, Kjelgaard Mikkelsen CC, Karlsson L, Kågström B. Task-Based Parallel Algorithms for Reordering of Matrices in Real Schur Form. *NLAFET Working Note WN-11*, 2017. Also as Report UMINF 17.11, Dept. of Computing Science, Umeå University, SE-901 87 Umeå, Sweden.
27. Myllykoski M. A Task-Based Algorithm for Reordering the Eigenvalues of a Matrix in Real Schur Form. In: . 10777 of *Lecture Notes in Computer Science*. Springer International Publishing; 2018: 207–216
28. Dongarra J, Whaley RC. *A User's Guide to the BLACS*. 1997. LAWN 94.
29. Task-based Schur reduction, $N = 20\,000$, shared memory, 28 cores. <https://youtu.be/VPHL0m2tmGI>; .
30. Kjelgaard Mikkelsen CC, Myllykoski M, Adlerborn B, Karlsson L, Kågström . D2.5: Eigenvalue Problem Solvers. tech. rep., Department of Computing Science, Umeå University; SE-901 87 Umeå, Sweden: 2017.
31. Karlsson L, Eljammaly M, Myllykoski M. D6.5 Evaluation of auto-tuning techniques. tech. rep., Department of Computing Science, Umeå University; SE-901 87 Umeå, Sweden: 2019.

32. Kjelgaard Mikkelsen CC, Schwarz AB, Karlsson L. Parallel robust solution of triangular linear systems. *Concurrency and Computation: Practice and Experience* 2018; 0(0): 1–19. doi: 10.1002/cpe.5064
33. Anderson E. LAPACK Working Note No. 36: Robust Triangular Solves for Use in Condition Estimation. Tech. Rep. CS-UT-CS-91-142, University of Tennessee; Knoxville, TN, USA: 1991.
34. Kjelgaard Mikkelsen CC, Karlsson L. Blocked Algorithms for Robust Solution of Triangular Linear Systems. In: Wyrzykowski R, Dongarra J, Deelman E, Karczewski K. , eds. *Parallel Processing and Applied Mathematics* Springer International Publishing; 2018; Cham: 68–78.
35. Schwarz A, Kjelgaard Mikkelsen CC, Karlsson L. Robust Parallel Eigenvector Computation For the Non-Symmetric Eigenvalue Problem. Tech. Rep. UMINF 20.02, Department of Computing Science, Umeå University; SE-901 87 Umeå, Sweden: 2020.
36. Kjelgaard Mikkelsen, C. C. and Myllykoski, M. . Parallel Robust Computation of Generalized Eigenvectors of Matrix Pencils in Real Schur Form. In: . 12043 of *Lecture Notes in Computer Science*. Springer Nature Switzerland AG; 2020.

