Recursive Pizza

Tiled Pizza

Inner-blocking Pizza

Azzam Haidar ICL Friday talk, August 24, 2012

Click to add title

Azzam Haidar

ICL Friday talk, August 24, 2012

General Overview: the linear algebra algorithms

Two categories:

1. One sided algorithms
   - Cholesky, QR decomposition, LU factorisation.

2. Two sided algorithms
   - Eigenvalue and Singular value problems.

     Eigenvalues, eigenvectors and eigenspaces are the properties of a matrix.
     - *Eigendecomposition have their origin in physics*
     - Stress and strain problems
     - Differential equations and quantum mechanics
     - *Weather forecast*
     - *Electronics simulation*
     - *Image processing*
     - *Material chemistry*
     - *Data storage*
     - *Web analysis*
     - *etc…*

General Overview: the Eigenproblem algorithms

- Symmetric EVP $Ax = \lambda x$
  - Tri-Diagonalization Reduction + solve + back transformation.

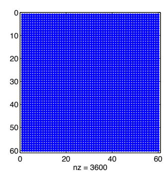
- Generalized EVP $Ax = \lambda Bx$ or $ABx = \lambda x$
  - Cholesky + Tri-Diagonalization Reduction + solve + back transformation.

- Singular Value Decomposition $A = U\Sigma V^T$
  - Bi-Diagonalization Reduction + solve + back transformation.

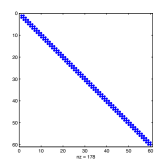
General Overview: the Eigenproblem algorithms

- **Symmetric EVP** $Ax = \lambda x$ **meaning compute** $A = Z \lambda Z^{*}$
  where λ **are the Eigenvalues and** Z **are the eigenvectors.**

1. **Tri-Diagonalization Reduction:** transform A to nice form ☺



==>



$$A = Q\,T\,Q^{*}$$

2. **Solve:** compute the Eigenvalue and Eigenvectors of the tridiagonal

$$T = E\,\lambda\,E^{*}$$

3. **Back transformation:** update the computed Eigenvectors.

$$Z = Q * E$$

General Overview: the Eigenproblem algorithms

- 90% if only eigenvalues
- 50% if eigenvalues and eigenvectors

- Symmetric EVP $Ax = \lambda x$
  - Tri-Diagonalization Reduction + solve + back transformation.
- Generalized EVP $Ax = \lambda Bx$ or $ABx = \lambda x$
  - Cholesky + Tri-Diagonalization Reduction + solve + back transformation.
- Singular Value Decomposition $A = U\Sigma V^T$
  - Bi-Diagonalization Reduction + solve + back transformation.

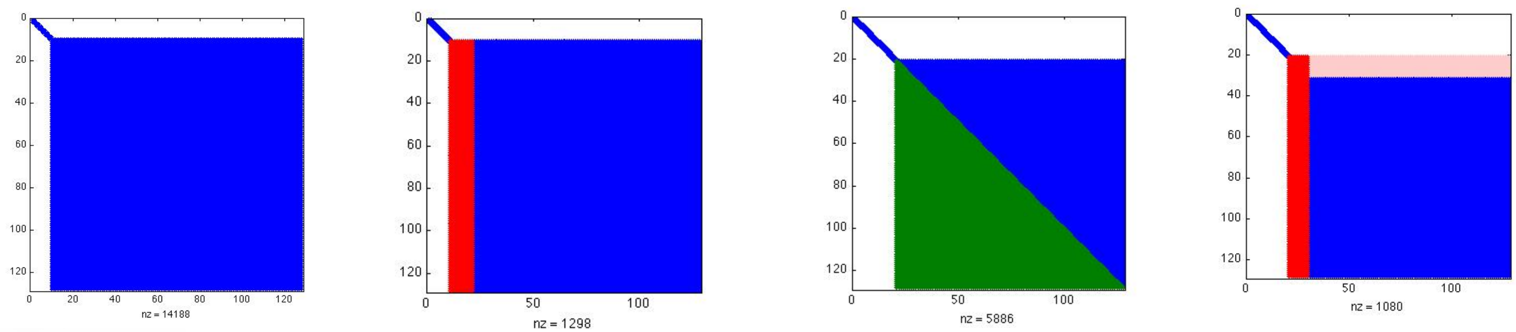
General Overview: the Eigenproblem algorithms

There are **two paths** to tridiagonal form

1. The standard LAPACK algorithm.

2. A new technique based on multi-stage algorithm.
   Christian Bischof , Bruno Lang , Xiaobai Sun (94) proposed multiple-stage implementation called Successive Band Reductions to reduce a matrix to tridiagonal.

The standard Tridiagonal reduction xSYTRD



step *k*: $Q\ A\ Q^H$ then update → step *k+1*

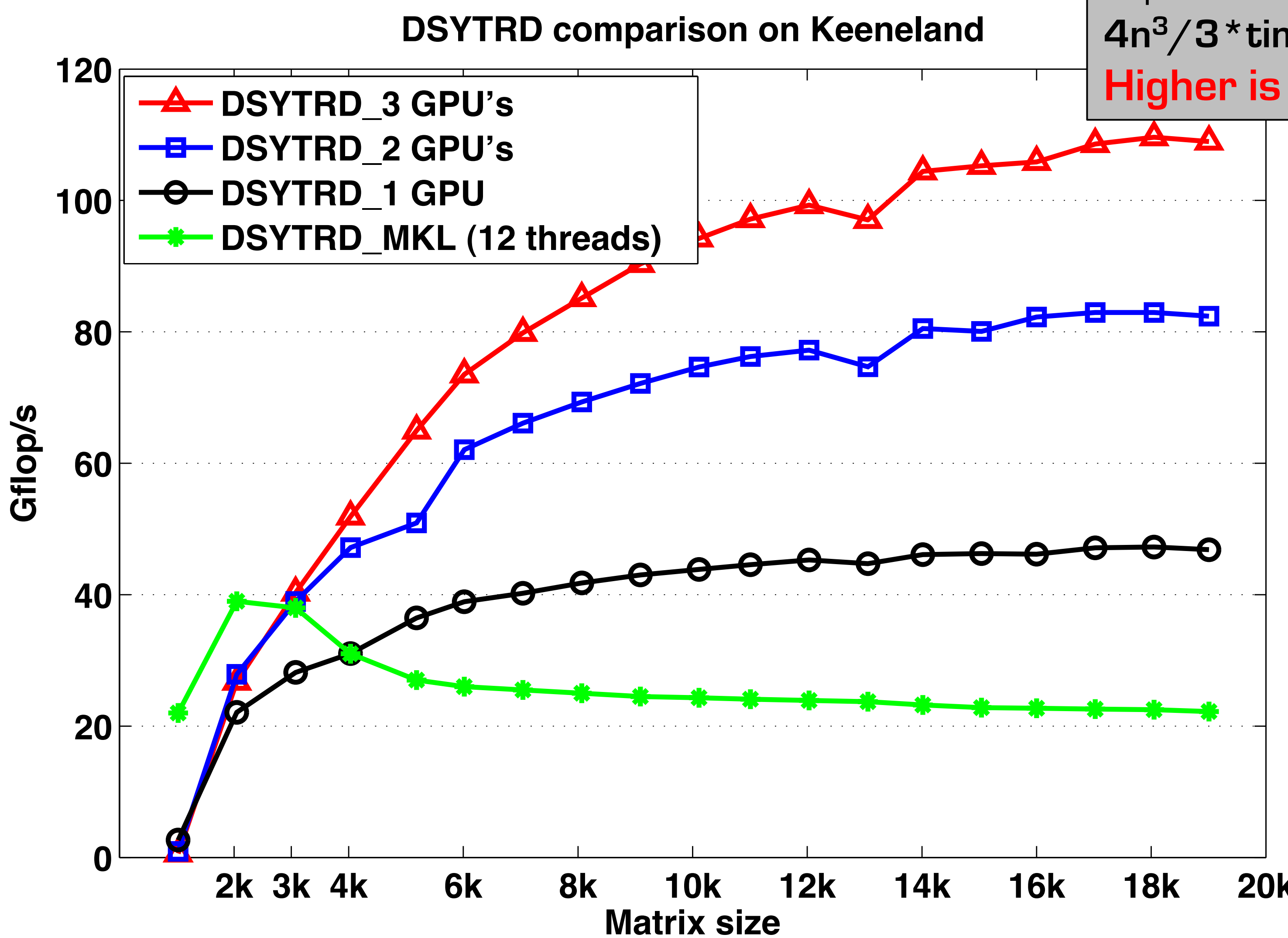
✷ Characteristics

- Too many Blas-2 op,
- Relies on panel factorization,
- Total cost $4n^3/3$,
- →Bulk sync phases,
- →Memory bound algorithm.

The MAGMA full reduction to tridiagonal

Performance comparison

flops formula:
$4n^3/3 * time$
Higher is faster



Results from Ichi

Experiments on socket of six-Intel Xeon X5650 + Fermi M2090.

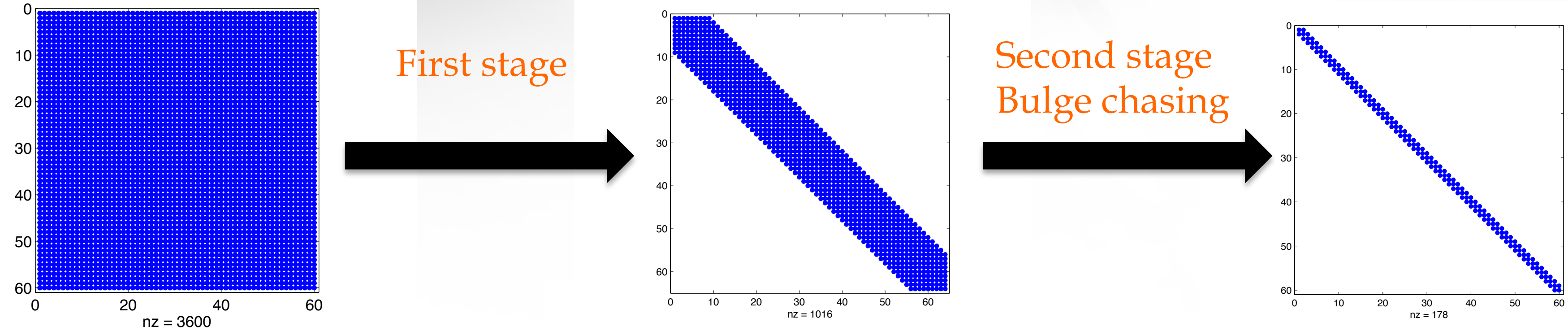
ICL UT

The PLASMA reduction: 2 stage algorithm

Idea:

- The idea is to cast expensive memory operations, occurring during the panel factorization into fast compute intensive ones.
- Redesign the algorithm in a new fashion which increase the cache reuse.
- Design new cache friendly kernels to overcomes the memory bound limitation.
- Extract parallelism and schedule task in an asynchronous order.

The PLASMA reduction: 2 stage algorithm



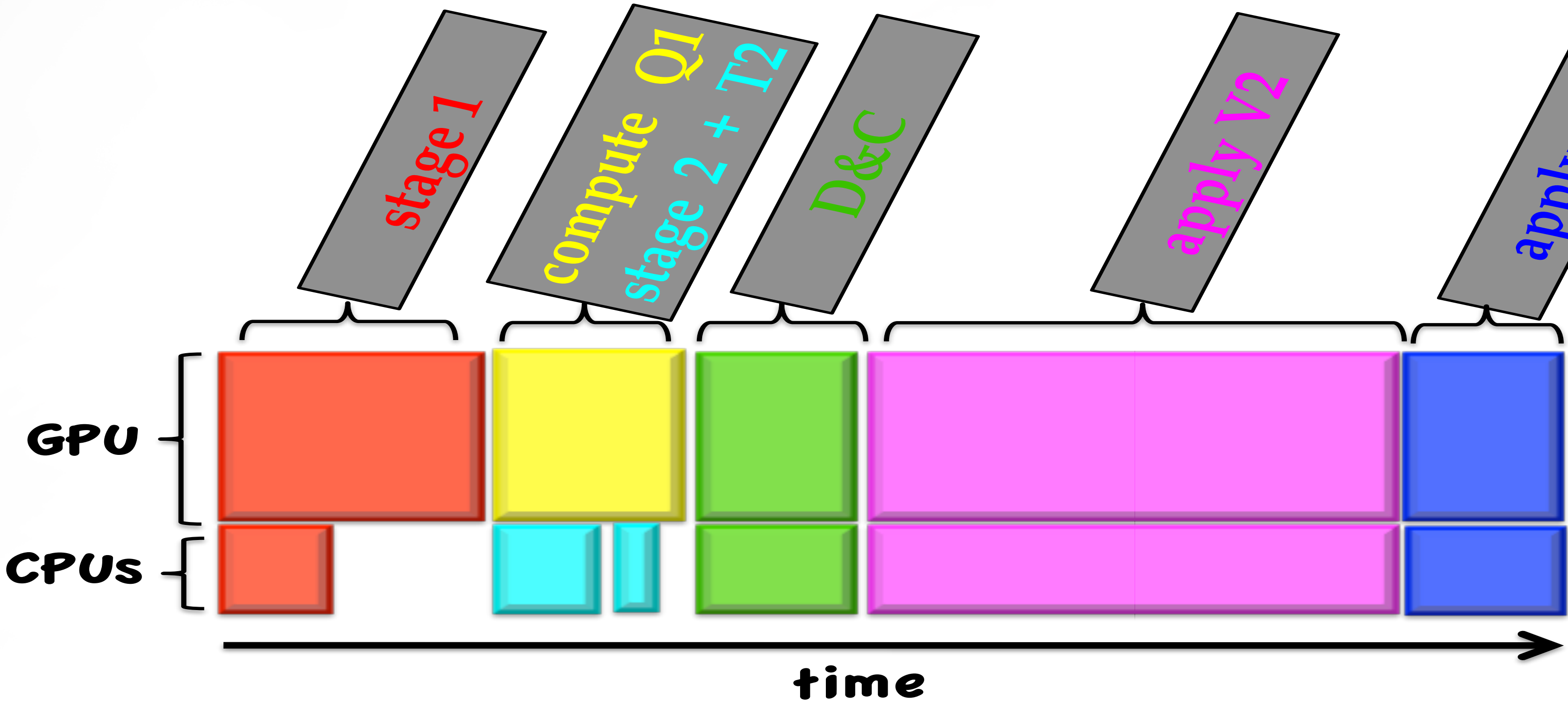
✸ Characteristics

- **Stage 1:**
  - BLAS-3,
  - one shot reduction,
  - asynchronous execution,
- **Stage2:**
  - BLAS-1.5,
  - element-wise/column-wise,
  - asynchronous execution,
  - new cache friendly kernel.

MAGMA: the Eigenproblem algorithms

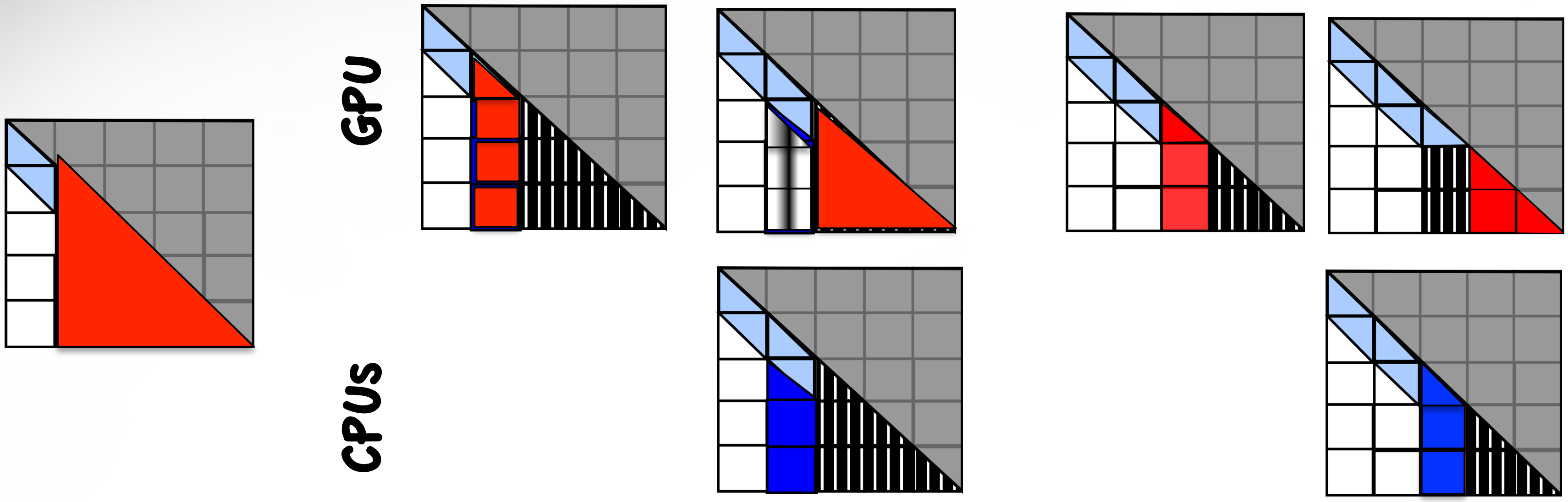
Idea:

- Develop similar approach for hybrid architectures (CPU+GPU)
- The idea is to dump expensive operations into GPU and try to overlap with the CPU.



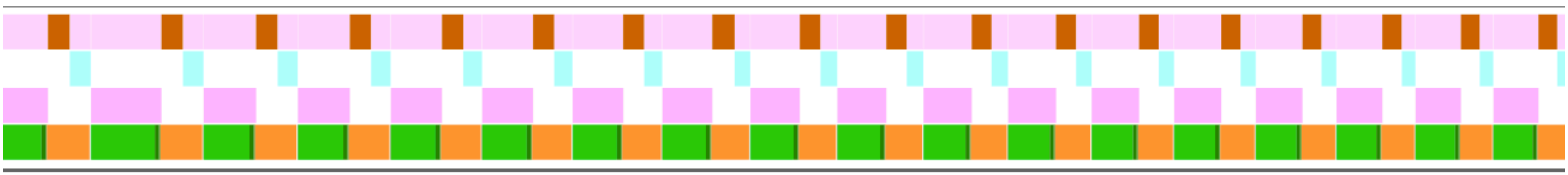
MAGMA: the Eigenproblem algorithms

The reduction:



MAGMA: the Eigenproblem algorithms

The reduction:

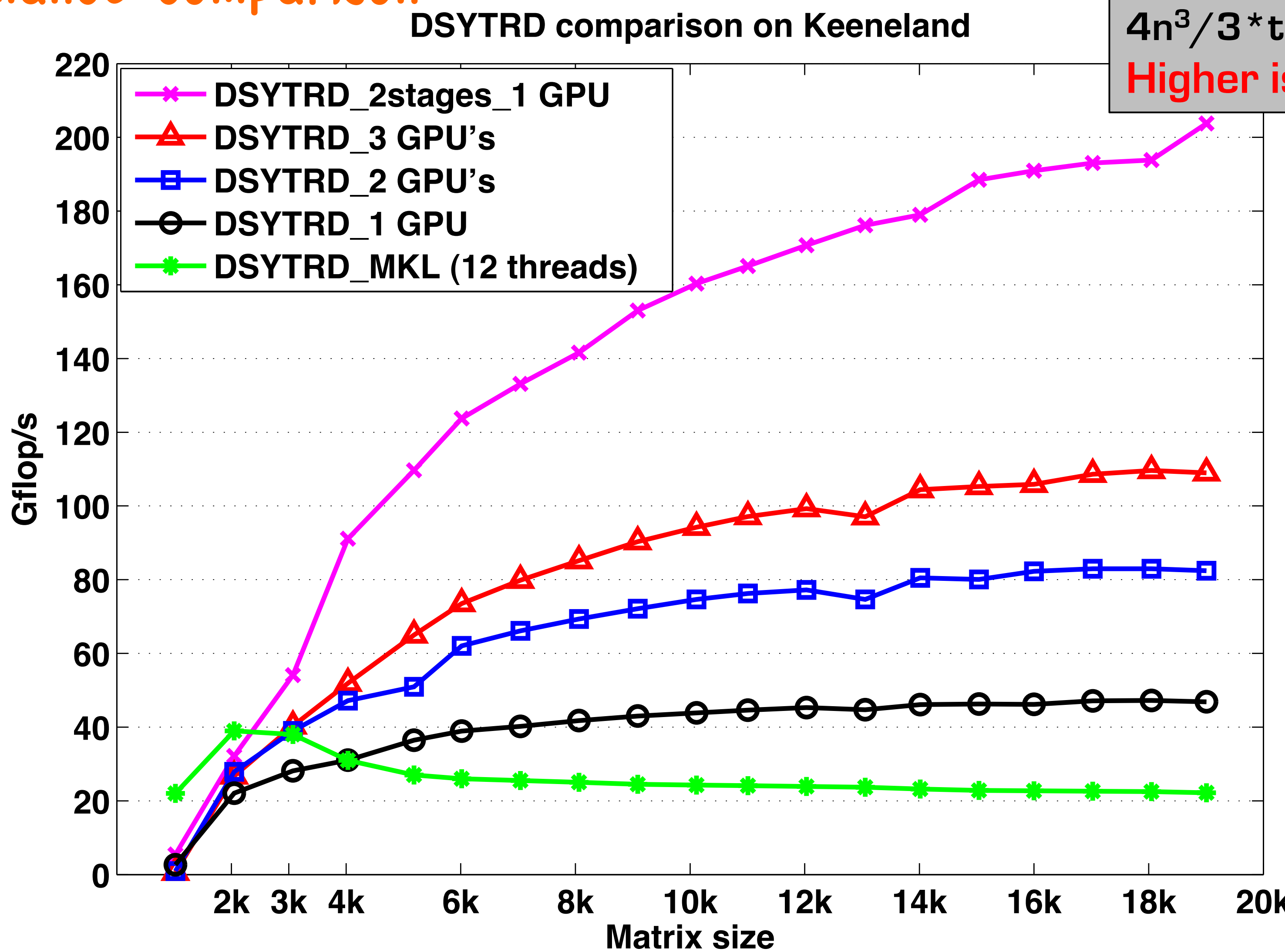


courtesy from Mark Gates

The MAGMA full reduction to tridiagonal

Performance comparison

flops formula:
$4n^3/3 * time$
Higher is faster



Experiments on socket of six-Intel Xeon X5650 + Fermi M2090.

PLASMA: the Eigenproblem algorithms

1. Symmetric EVP $Ax = \lambda x$
   - Tri-Diagonalization Reduction + solve + back transformation.
2. Generalized EVP $Ax = \lambda Bx$ or $ABx = \lambda x$
   - Cholesky + Tri-Diagonalization Reduction + solve + back transformation.
3. Singular Value Decomposition $A = U\Sigma V^T$
   - Bi-Diagonalization Reduction + solve + back transformation.

PLASMA: the Eigenproblem algorithms

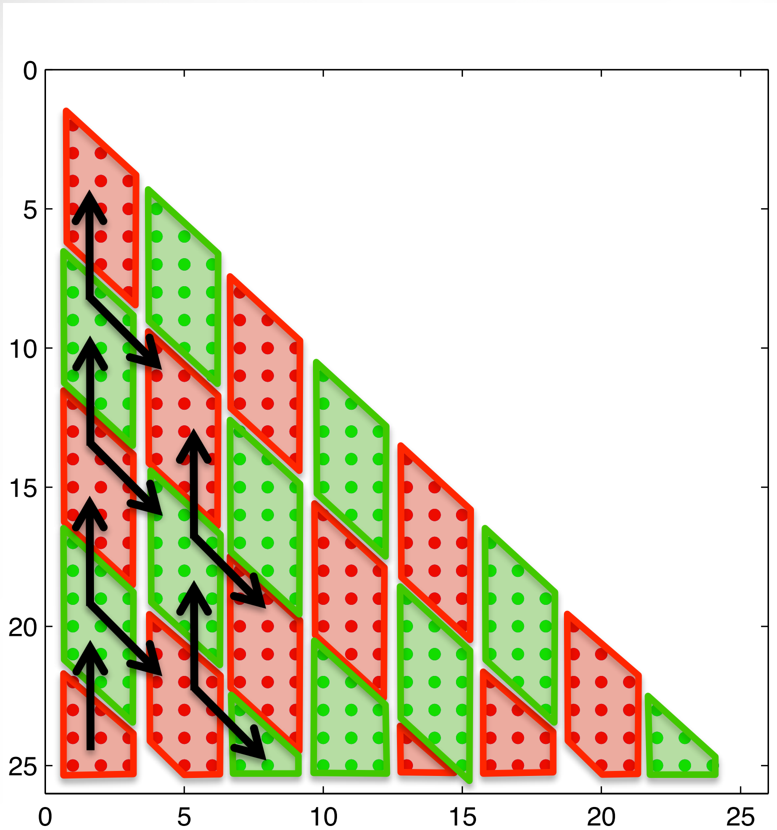
Characteristics

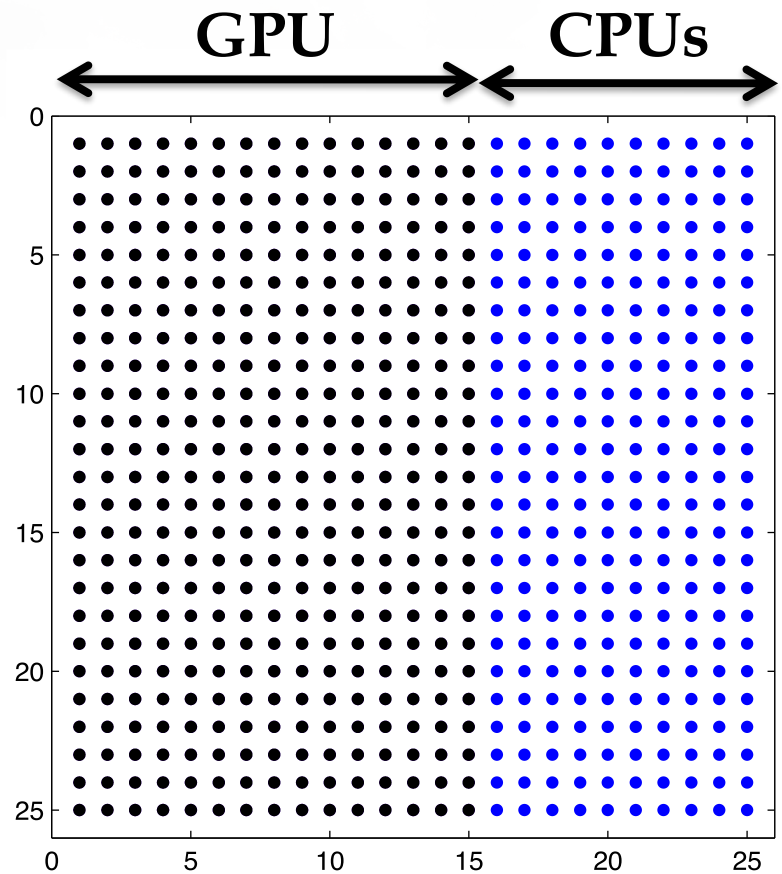
- Algorithm is more challenging, we will have 2 back transformations to apply,
- Allow independent parallelism,
- Deal with different layout of storage,
- Requires new kernels to improve cache reuse,
- is $2n^3$ more expensive.

MAGMA: the Eigenproblem algorithms

Back transformation:

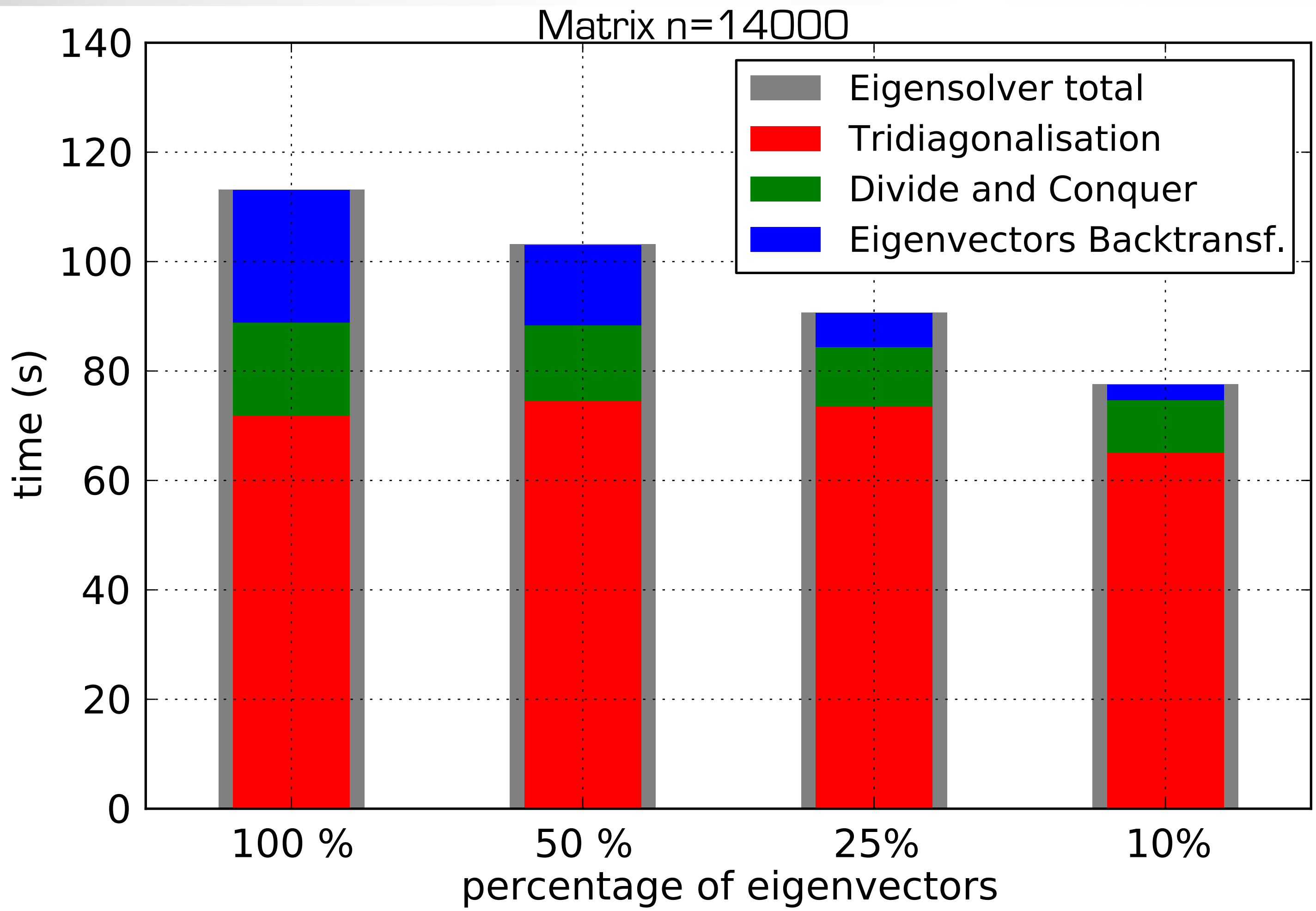
- Allow CPUs to contribute





MAGMA: the Eigenproblem algorithms

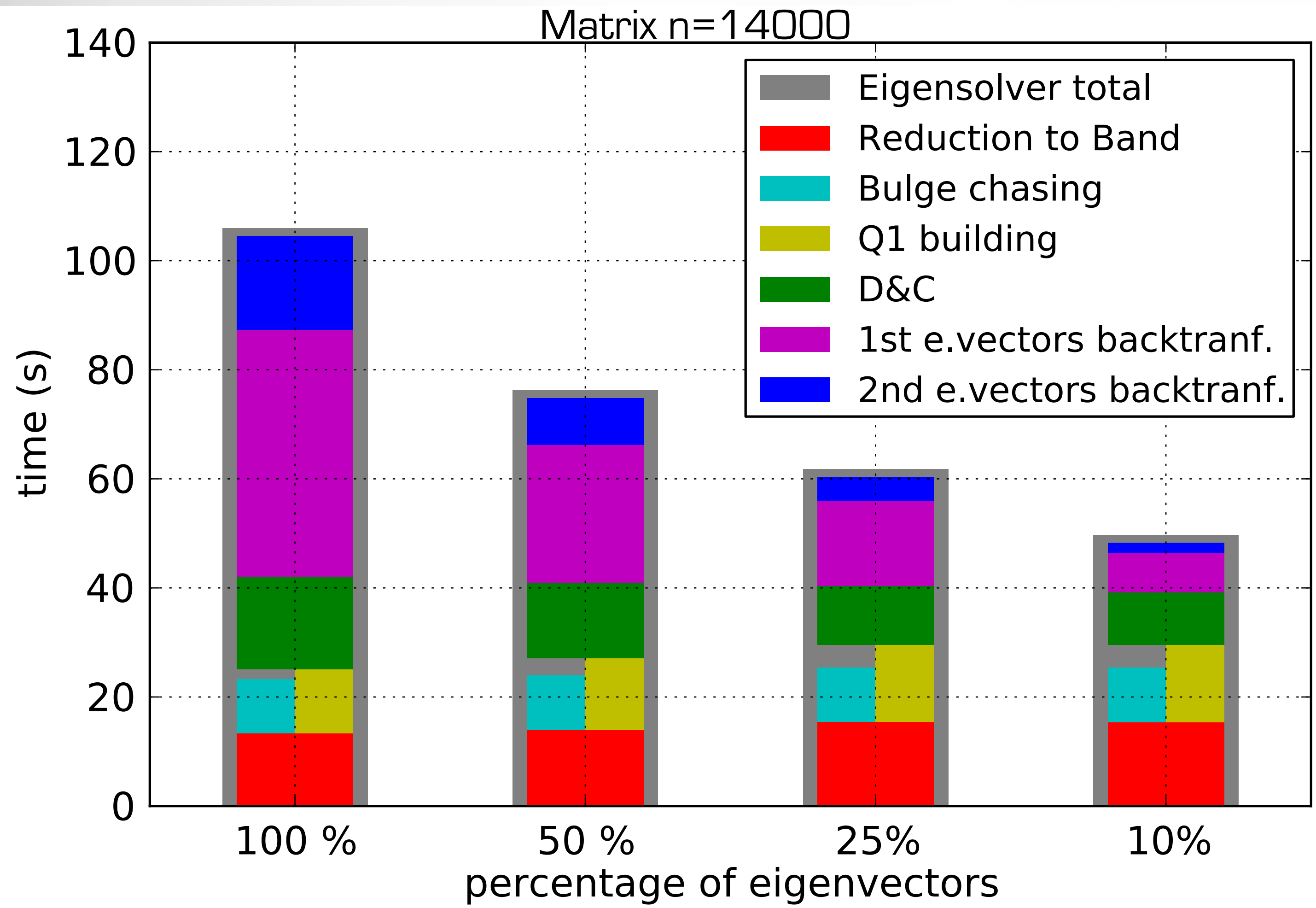
Performance comparison



Experiments on socket of six-Intel Xeon X5650 + Fermi M2090.

MAGMA: the Eigenproblem algorithms

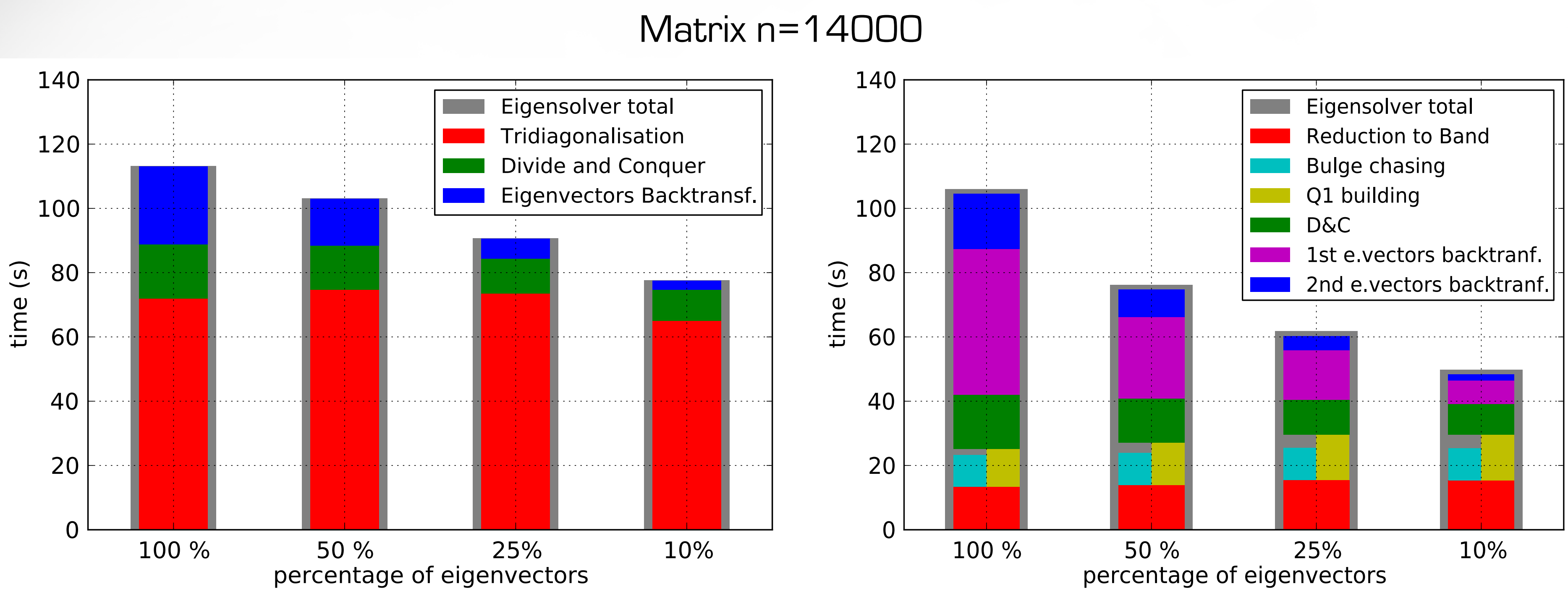
Performance comparison



Experiments on socket of six-Intel Xeon X5650 + Fermi M2090.

MAGMA: the Eigenproblem algorithms

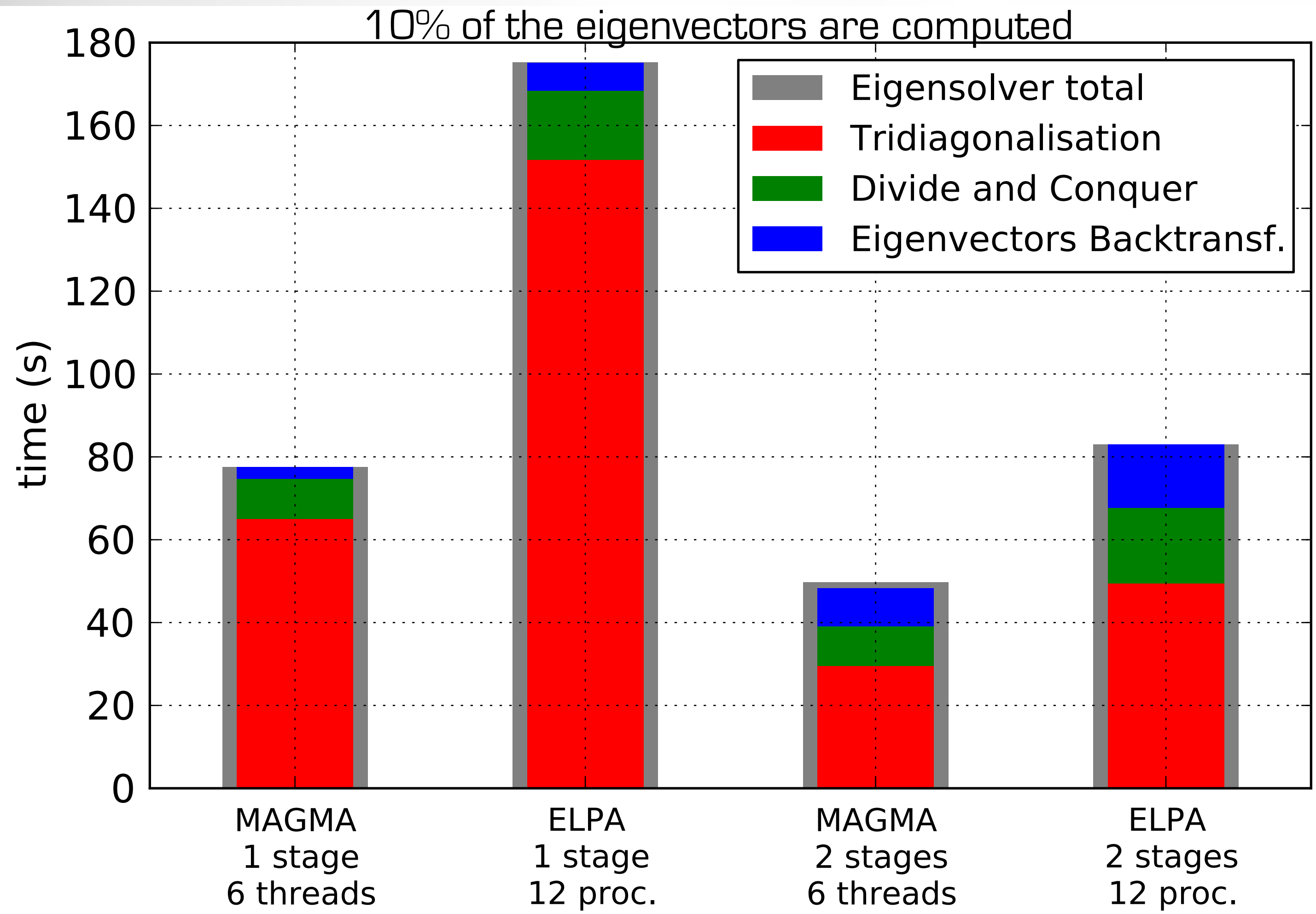
Performance comparison



Experiments on socket of six-Intel Xeon X5650 + Fermi M2090.

MAGMA: the Eigenproblem algorithms

Performance comparison



Experiments on socket of six-Intel Xeon X5650 + Fermi M2090.

Future work

- **Road map and open questions:**
  - Develop similar approach for SVD (ongoing integration).
  - Effort might be made on the eigensolvers.
  - Hessenberg, (Piotr, Hatem)
    - Bulge chasing
    - Gaussian reduction
    - Sign functions

Future work

- Road map and open questions:

  - Developing a multi-GPU/multicore/ditributed version of the algorithm (ongoing).
  - Develop the tridiagonal reduction in the context of distributed memory architecture.
  - Evaluate a band divide and conquer for the band matrices.

Thank you for your attention

ICL UT