

Using PLASMA in the NAG library

Joseph Dobson

Monday 20th May 2013

Introduction

Who am I?

- ▶ I am a computer science Graduate (not Maths)
- ▶ Currently pursuing a KTP between NAG and the University of Manchester
 - ▶ Knowledge Transfer Partnership

Introduction

Who am I?

- ▶ I am a computer science Graduate (not Maths)
- ▶ Currently pursuing a KTP between NAG and the University of Manchester
 - ▶ Knowledge Transfer Partnership
 - ▶ Academic supervisor: David Silvester
 - ▶ Support academic: Nick Higham
 - ▶ Lead academic: Jack Dongarra
 - ▶ Company supervisor: Craig Lucas

Numerical Algorithms Group

Numerical Algorithms Group (*NAG*):

- ▶ Not for profit software company
- ▶ NAG's major products are:
 - ▶ HPC services
 - ▶ NAG library.
- ▶ Based in Oxford with offices in Manchester and around the world.

Numerical Algorithms Group

The NAG library is:

- ▶ 1700 numerical routines
- ▶ Available for many different programming languages including Fortran, C, Matlab & Java
- ▶ Engine code is written in Fortran
- ▶ OpenMP is used for parallelism both internally and via the vendor libraries
- ▶ Relies heavily upon LAPACK for dense linear algebra.

PLASMA

Who are behind PLASMA:

- ▶ Jack Dongara is the academic lead of PLASMA at ICL and has a position here at UoM.
- ▶ ICL in Knoxville Tennessee; also involved with:
 - ▶ LAPACK
 - ▶ MAGMA
 - ▶ ScaLAPACK...

PLASMA

What is PLASMA:

- ▶ LAPACK functionality, dense linear algebra.

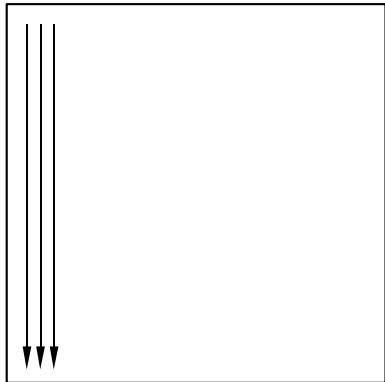
PLASMA

What is PLASMA:

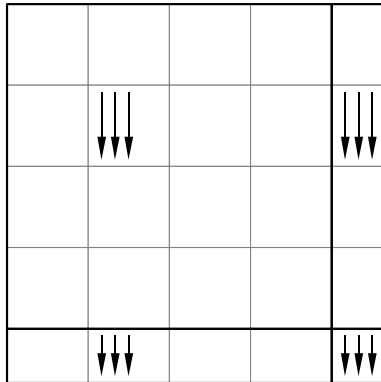
- ▶ LAPACK functionality, dense linear algebra.
- ▶ Tile storage format

PLASMA - Tile Major

LAPACK (CM)



PLASMA (CCRB)



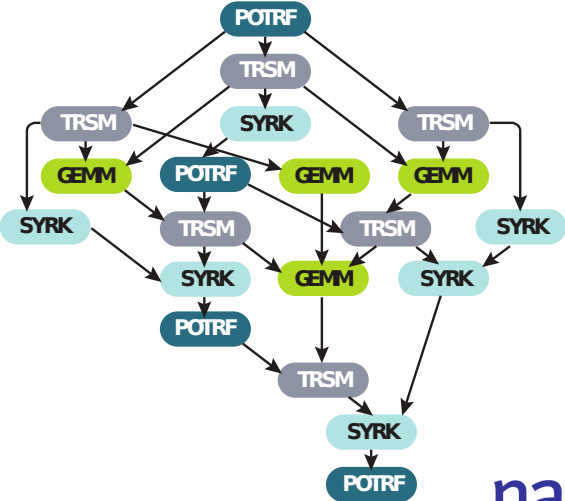
PLASMA

What is PLASMA:

- ▶ LAPACK functionality, dense linear algebra.
- ▶ Tile Major Storage format
- ▶ Directed Acyclic Graph (*DAG*)
 - ▶ Edges are dependencies
 - ▶ Nodes are sequentially executed tasks operating tiles

PLASMA - DAG

The DAG for a Cholesky factorisation.

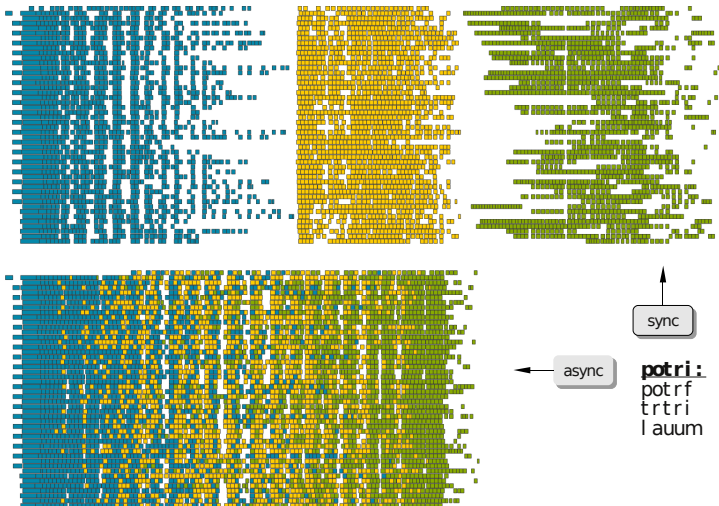


PLASMA

What is PLASMA:

- ▶ LAPACK functionality, dense linear algebra.
- ▶ Tile Major Storage format
- ▶ Directed Acyclic Graph (*DAG*)
 - ▶ Edges are dependencies
 - ▶ Nodes are sequentially executed tasks operating tiles
- ▶ Asynchronous interface

PLASMA - Asynchronous Interface



Interoperability:

- ▶ The NAG library already uses OpenMP to achieve parallelism, it is critical that PLASMA can interoperate with it.
- ▶ We are currently working with ICL to iron out issues we are having with OpenMP runtimes.

Issues

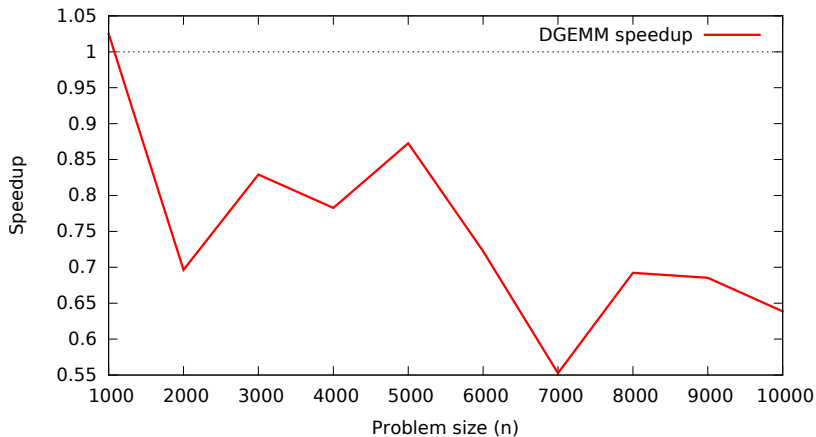
PLASMA has an overhead, caused by:

- ▶ Matrix format conversion
- ▶ Thread spawning and scheduling overhead
- ▶ Allocation

We will use crossover points to switch between the vendor library and PLASMA depending on problem size.

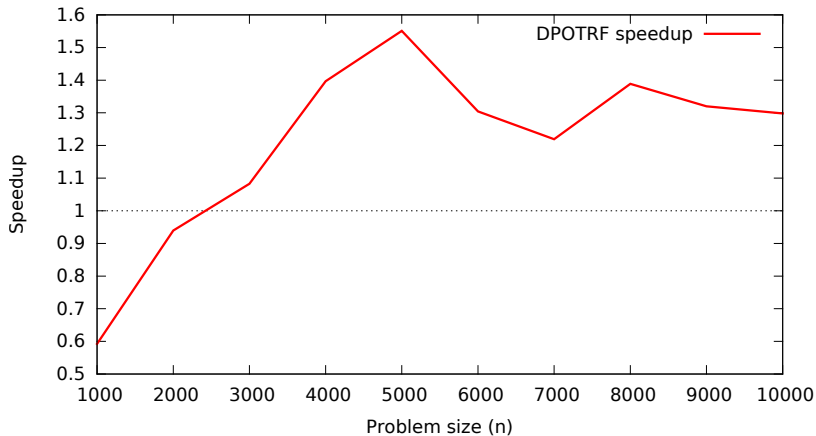
Performance

DGEMM Speedup, PLASMA against MKL with 48 cores

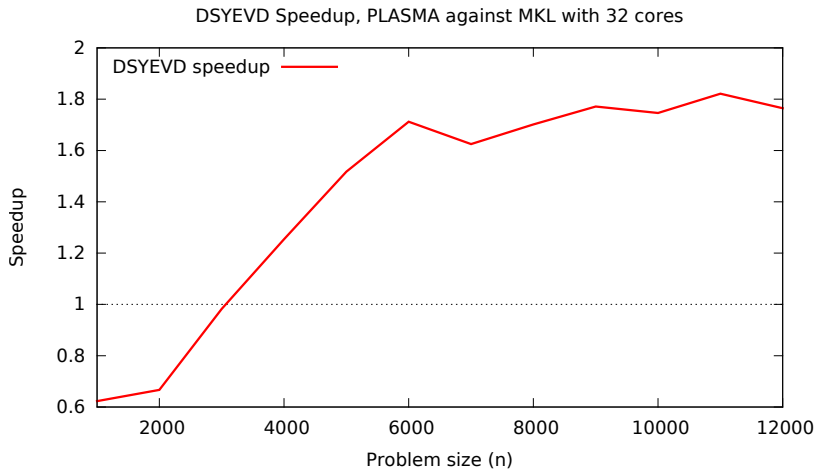


Performance

DPOTRF Speedup, PLASMA against MKL with 32 cores



Performance



Autotuning

The addition of PLASMA to the NAG library adds new tunable performance parameters:

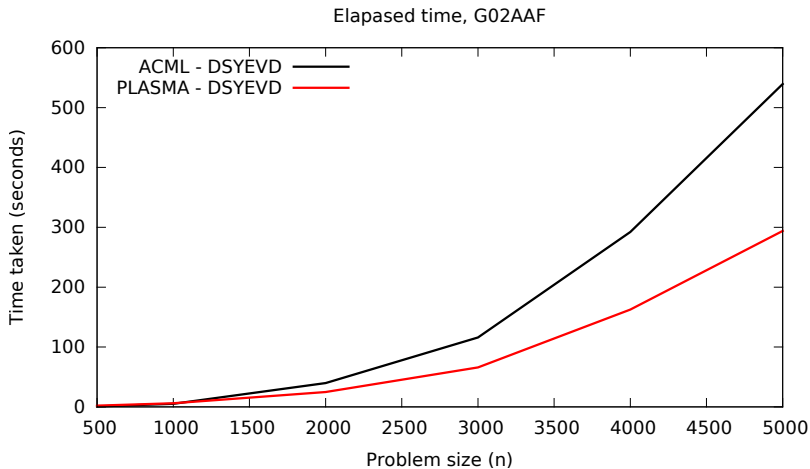
- ▶ The specific crossover points will vary between machines
- ▶ The internal PLASMA parameters such as tile and inner block sizes will also vary between different architectures.

Autotuning

Optimal parameters vary depending on the machine it is running, therefore tuning is desirable.

- ▶ Autotuning will require a program to be run on the target system
- ▶ A new element to the NAG library
- ▶ This is an extension to the original project
- ▶ An active research area for the PLASMA team

Prospective performance



References

NAG <http://www.nag.co.uk/>

UoM <http://www.maths.manchester.ac.uk/>

ICL <http://icl.cs.utk.edu/>

PLASMA <http://icl.cs.utk.edu/plasma/>