The University of Manchester

Towards Parallel 2D R-Matrix Propagation Codes for Multicore Architectures

Maksims Abaļenkovs*   David Stevens†   Andrew Sunderland‡   Stan Scott§

*Alan Turing Building 2.233, School of Mathematics, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

†Department of Geography, The University of Sheffield, Sheffield S10 2TN, UK

‡The Hartree Centre, STFC Daresbury Laboratory, Sci-Tech Daresbury, Keckwick, Daresbury, Warrington WA4 4AD, UK

§School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, University Road, Belfast BT7 1NN, UK

`m.abalenkovs@manchester.ac.uk`   `d.r.stevens@sheffield.ac.uk`   `andrew.sunderland@stfc.ac.uk`   `ns.scott@qub.ac.uk`

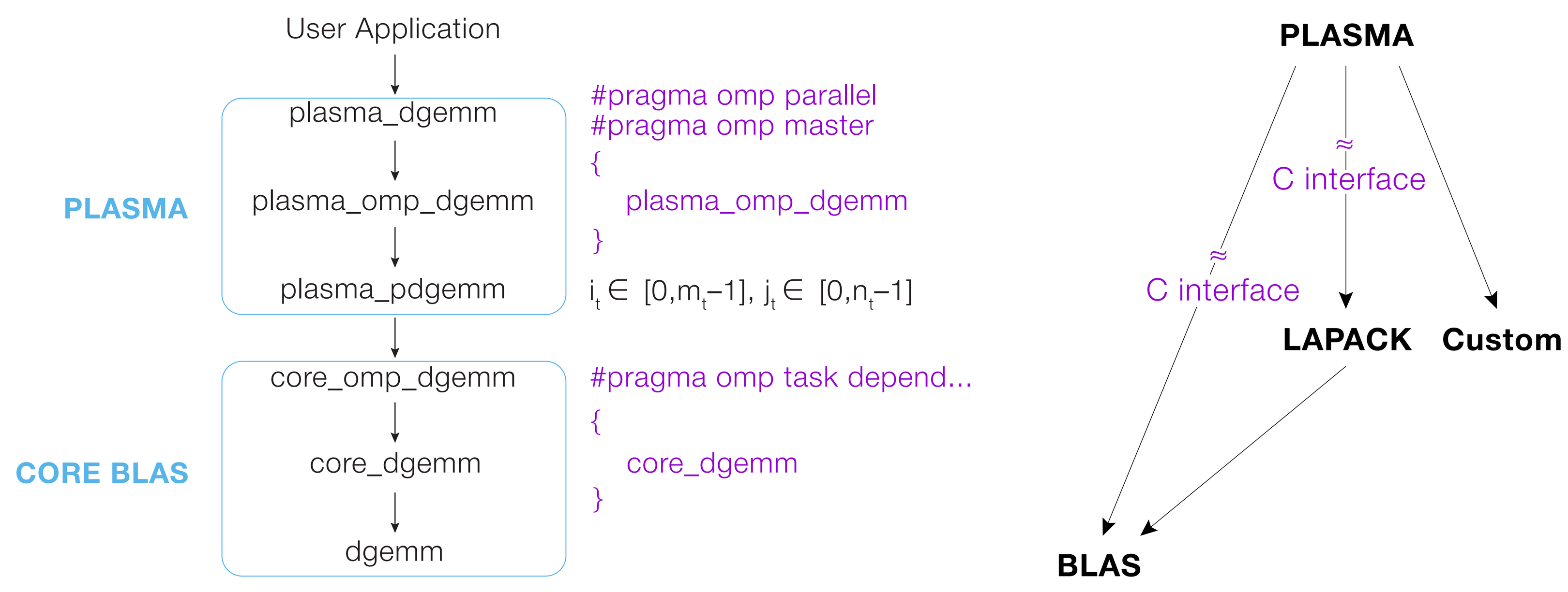
Abstract

The work presented in this publication proposes a major update to the 2DRMP software. The code has been instrumented with the ability to apply novel PLASMA routines alternatively to the standard LAPACK package. Performance of PLASMA was compared against industrial libraries—Intel MKL and ARMPL. Two types of experiments (i) artificial linear algebra kernel runs and (ii) realistic 2DRMP scenario runs were conducted on two different supercomputers—Intel Skylake-based MareNostrum4 and ARM Cavium-based Wombat.

2D R-Matrix Propagation Codes (2DRMP)

- Aimed at study of electron scattering from H-like atoms and ions at intermediate energies [1].
- Based on two-dimensional R-matrix propagation theory.
- Constructs local R-matrices within each subregion to propagate a global R-matrix.
- Relies on three linear algebra *kernels:* eigenvalue decomposition `DSYEVD`, linear system solution `DGESV` and matrix multiplication `DGEMM`.

Parallel Linear Algebra for Scalable Multi-core Architectures (PLASMA)



(a) Routine stack   (b) Structure

Figure 1: PLASMA routine stack and structure

- Open-source dense linear algebra package for shared-memory architectures [2].
- Research project aimed at developing new algorithms and schedulling techniques.
- Operates on matrix *tiles*—rectangular blocks of original matrices.
- Implements tile-based algorithms via OpenMP `task depend` construct.
- Relies on OpenMP runtime for dynamic, task-based schedulling.

Numerical Results

Simulation Settings

- Two sets of experiments were run: one on the linear algebra kernels and another on the 2DRMP method.
- Three linear algebra kernels were tested: eigenvalue decomposition `DSYEVD`, linear system solution `DGESV` and matrix multiplication `DGEMM`.
- Performance of PLASMA v18.11.1 was compared with vendor optimised mathematical libraries: Intel Math Kernel Library (MKL) v2019.3 and ARM Performance Library (ARMPL) v19.1.
- PLASMA was tuned for optimal tile size `nb`, inner block size `ib` and maximum number of threads per panel `mtpf` for every matrix order `n`.
- Realistic 2DRMP testing scenarios were obtained by increasing the maximum principal quantum number of the basis orbitals $n_{max} \in \{25, 30, 35\}$.

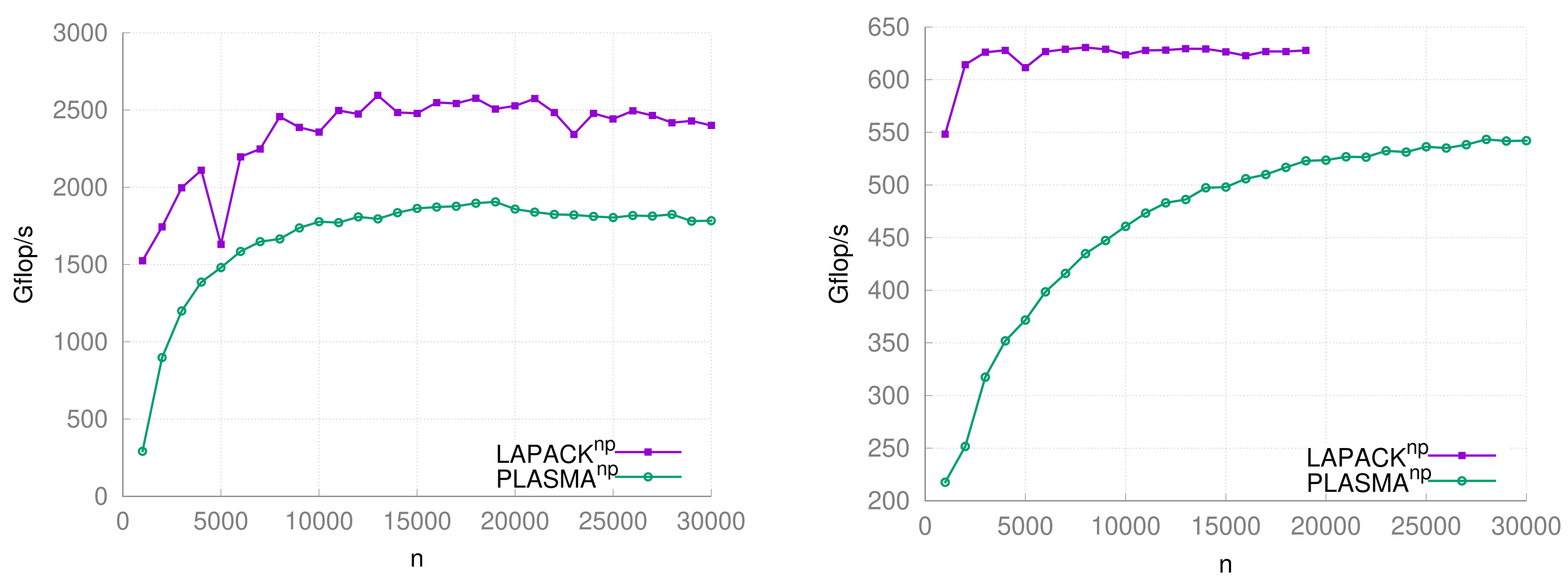
Hardware Platforms

- Numerical experiments were run on a single node of two different HPC systems:
  - MareNostrum4 node: 2 × 24 core Intel Xeon Platinum CPU 8160 (Skylake) @ 2.1 GHz, $n_{thr} = 48$.
  - Wombat node: 2 × 28 core ARM Cavium ThunderX2 CPU @ 2.0 GHz, $n_{thr} = 56$.
- The kernel and 2DRMP source codes were compiled with:
  - GNU `gfortran` v8.1 using `-O3 -fopenmp -march=skylake-avx512` flags.
  - ARM `armflang` v19.1 using `-O3 -fopenmp -armpl=parallel -mcpu=thunderx2t99` flags.
- Execution time was measured with the `omp_get_wtime()` routine.
- Performance results denoted in the plots by superscript letters "np" were launched via `numactl --interleave=all` with OpenMP processor binding (`OMP_PROC_BIND=true`).

Acknowledgements

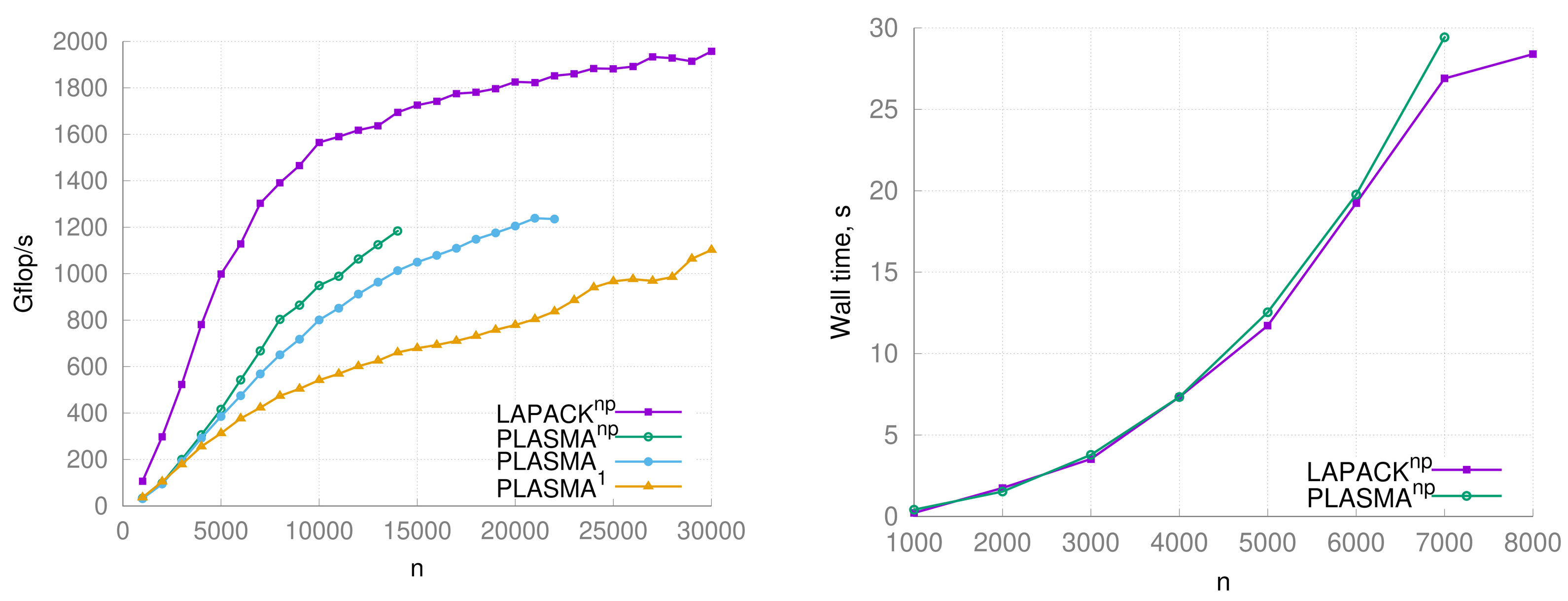
- The EPSRC grant (EP/M01147X/1) for the Scale-free, Energy-aware, Resilient and Transparent Adaptation (SERT) project.
- The Horizon 2020 FET–HPC grant (671633) for the Parallel Numerical Linear Algebra for Extreme Scale Systems (NLAFET) project.
- The Horizon 2020 HPC–Europa3 programme (INFRAIA-2016-1-730897) and Barcelona Supercomputing Center (BSC) for access to the MareNostrum4 supercomputer.

Kernel Tests



(a) MareNostrum   (b) Wombat

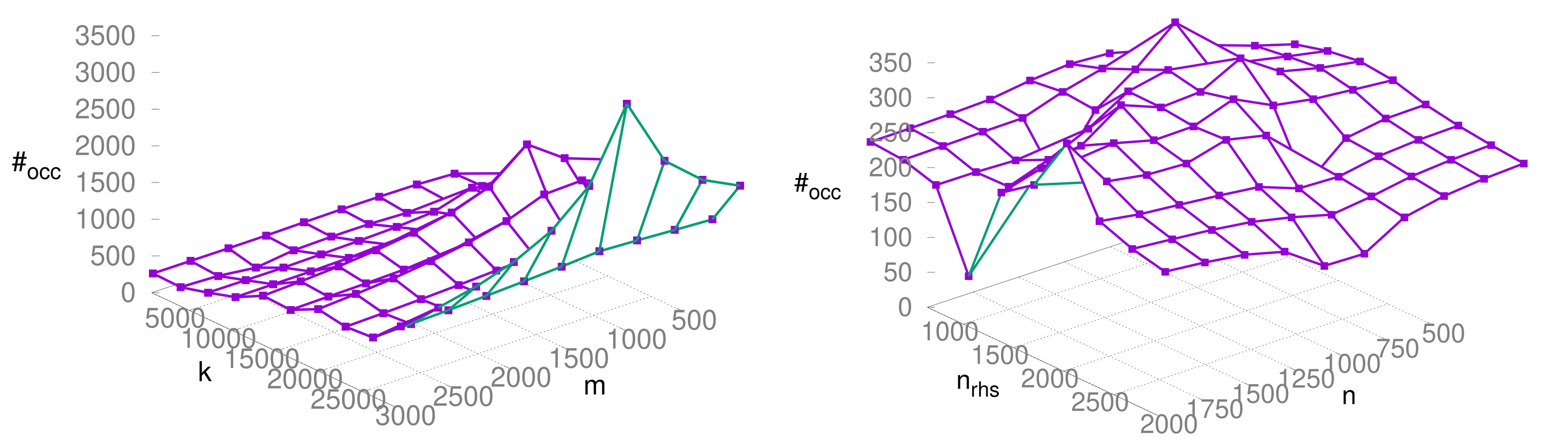
Figure 2: Matrix multiplication kernel `DGEMM`



(a) MareNostrum   (b) Wombat

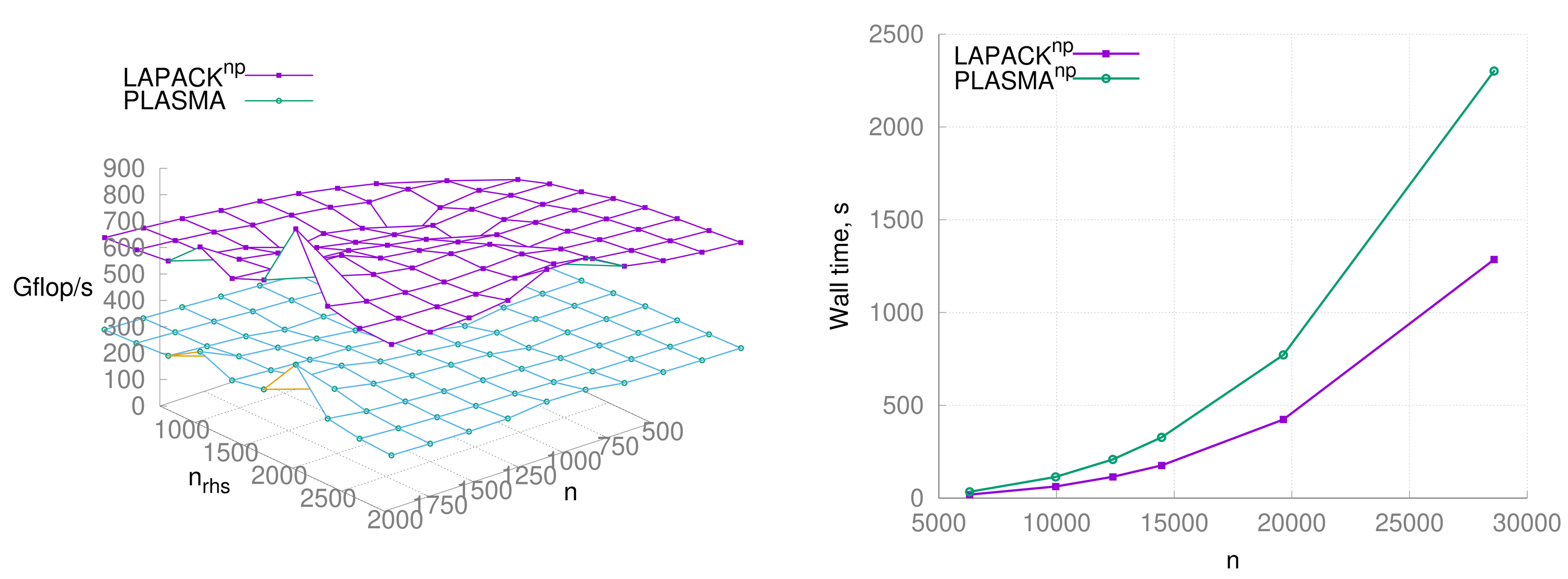
Figure 3: Linear system solution and eigenvalue decomposition kernels `DGESV` and `DSYEVD`

2DRMP Tests



(a) Matrix multiplication `DGEMM`   (b) Linear system solution `DGESV`

Figure 4: Profile of matrix sizes in 2DRMP



(a) MareNostrum   (b) Wombat

Figure 5: Linear system solution and eigenvalue decomposition kernels `DGESV` and `DSYEVD`

Conclusion and Future Work

- Industrial libraries with low-level Assembly optimisations outperform tile-based PLASMA in common linear algebra operations—matrix multiplication `DGEMM` and linear system solution `DGESV`.
- Expand 2DRMP parallelisation to distributed memory systems via the promising Software for Linear Algebra Targeting Exascale (SLATE) package.
- Implement, apply and test *non-tiled* version of PLASMA algorithms.

References

[1] N. S. Scott, M. P. Scott, P. G. Burke, T. Stitt, V. Faro-Maza, C. Denis, and A. Maniopoulou, "2DRMP: A suite of two-dimensional R-matrix propagation codes," *Computer Physics Communications*, vol. 180, no. 12, pp. 2424–2449, 2009.

[2] J. Dongarra, M. Gates, A. Haidar, J. Kurzak, P. Luszczek, P. Wu, I. Yamazaki, A. Yarkhan, M. Abalenkovs, N. Bagherpour, J. Šístek, D. Stevens, M. Zounon, and S. D. Relton, "PLASMA: Parallel linear algebra software for multicore using OpenMP," *ACM Trans. Math. Softw.*, vol. 45, no. 2, pp. 16–35, 2019.