Internship offer:

Adaptive compression methods for the solution of sparse linear systems, with applications to EDF industrial simulations

Contact:

Olivier Boiteau (olivier.boiteau@edf.fr), Théo Mary (theo.mary@lip6.fr)

Context of the internship:

EDF guarantees the technical and economic control of its electricity production, transport and distribution resources, from design to end of life. Safety and availability requirements require supporting its operating decisions with digital simulation. Predicting and analyzing the behavior of infrastructures and sites involves, in particular, mastering a whole range of field physics models. To do this, EDF develops several simulation codes and platforms: code_aster [2a], code_carmel [2b], TELEMAC [2c], Salomé [2d]…

However, these tools, often based on the finite element method, remain complex and computationally demanding, in particular due to the time and memory consumption of a recurring algorithmic step: that of solving large systems of equations. Indeed, it is often this linear algebra step that determines the machine requirements and the computation times of simulations, or even determines their feasibility. To solve it efficiently, EDF codes are often based on dedicated and optimized external packages, such as the MUMPS product [1]. The latter is used daily by hundreds of R&D and Engineering users at EDF via simulations of its codes.

Main objectives:

The main objective of the internship is to improve the performance of MUMPS to enable much more efficient system solutions and to address much larger models, without sacrificing the requirements of robustness, precision, scope of use and ergonomics.

To do this, the starting point will be the so-called “Block Low-Rank” (BLR) compression method [3] currently used by MUMPS to reduce its computing time and memory footprint. The objective will be to develop new compression methods to further improve performance on medium-sized problems and to scale up to very large problems currently beyond the reach of the BLR method.

Several avenues can be explored during this internship:

- Algorithmic compression: several compression methods theoretically more compact than BLR have been proposed in the literature, for example multilevel methods [4], shared bases [5], or butterfly matrices [6]. Their real potential as alternatives to BLR in the context of MUMPS is to be studied but seems promising for large problems.
- Floating-point compression: low-precision arithmetic formats (e.g., 32 or 16 bits) can significantly reduce the size of the data and speed up calculations. We have previously [7] proposed a mixed-precision BLR compression strategy that mixes several formats to maintain a guaranteed level of precision. To benefit from both this computational compression and the use of other algorithmic compression methods more compact than BLR mentioned above, it will therefore be necessary to extend this mixed-precision strategy to these new methods.

- Adaptive strategies: there is no compression method that is systematically better than the others. We will therefore seek to optimize compression by combining different methods at different locations in the code. In particular, MUMPS is based on the multifrontal method, which processes a tree of increasingly larger matrices. It seems natural to use more compact formats (better asymptotic complexity) for the levels at the top of the tree where the matrices are the largest. In addition, within the same level, we will also study the relevance of using different formats for different sub-blocks of the matrices depending on their distance from the diagonal.

Work environment:
The internship will take place on the EDF Lab Paris-Saclay site in close collaboration with the MUMPS product development teams: Mumps Technologies (ENS Lyon) and the LIP6 (Paris) and IRIT (Toulouse) laboratories.

- Internship location: EDF Lab Paris-Saclay in Palaiseau (91) with weekly meetings at LIP6, and potentially a stay or two in Lyon and/or Toulouse.
- Duration: 5/6 months.
- Remuneration: to be specified according to the training.
- Required knowledge:
  - Master's level Research or final year of engineering school.
  - FORTRAN/C and MPI/OpenMP programming, UNIX environment.
  - HPC, linear algebra.

Continuation with a PhD thesis:
Depending on the results of the avenues raised by this internship, a CIFRE thesis on the same subject and with the same team could be proposed in continuity.

References:

[1] http://www.mumps-solver.org
[2a] http://www.code-aster.org
[2b] http://code-carmel.univ-lille1.fr
[2c] http://opentelemac.org
[2d] http://salome-platform.org
[3] Amestoy et al., Performance and Scalability of the Block Low-Rank Multifrontal Factorization on Multicore Architectures, *ACM Trans. Math. Softw.* (2019). https://hal.science/hal-01955766
[4] Amestoy et al., Bridging the Gap between Flat and Hierarchical Low-rank Matrix Formats: the Multilevel Block Low-Rank Format, *SIAM J. Sci. Comput.* (2019). https://hal.science/hal-01774642
[5] Ashcraft et al., Block Low-Rank Matrices with Shared Bases: Potential and Limitations of the BLR^2 Format, *SIAM J. Matrix Anal. Appl.* (2021). https://hal.science/hal-03070416
[6] Liu et al., Sparse approximate multifrontal factorization with butterfly compression for high-frequency wave equations, *SIAM J. Sci. Comput.* (2021). https://arxiv.org/abs/2007.00202
[7] Amestoy et al., Mixed Precision Low Rank Approximations and their Application to Block Low Rank LU Factorization, *IMA J. Numer. Anal.* (2022). https://hal.science/hal-03251738